



UNIVERSITÄT KARLSRUHE
INSTITUT FÜR NACHRICHTENTECHNIK
Prof. Dr.rer.nat. Friedrich Jondral



**Zur Eignung invarianter Momente für die
Klassifikation von Objekten in Bildern**

Diplomarbeit von

Michael Grimm

Hauptreferent : Prof. Dr.-Ing. Kristian Kroschel
Betreuer : Dr.-Ing. Gunther Grasmann

Beginn : 01.10.02
Abgabe : 01.04.03

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Diplomarbeit selbständig und ohne unzulässige fremde Hilfe angefertigt habe. Die verwendeten Literaturquellen sind im Literaturverzeichnis vollständig zitiert.

Karlsruhe, den 01.04.2003

Michael Grimm
Danngrabenweg 7
77815 Bühl

Danksagung

Hiermit möchte ich mich bei allen Personen bedanken, die mir bei der Entstehung dieser Arbeit geholfen haben.

Mein Dank gilt insbesondere Herrn Prof. Dr.-Ing. Kristian Kroschel, der mir diese Diplomarbeit ermöglicht hat, und Herrn Dr.-Ing. Gunther Grasemann für die sehr gute Betreuung. Weiterhin danke ich neben allen Mitarbeitern des Instituts auch meiner Familie und meiner Freundin für die Unterstützung.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Aufgabenstellung und Zielsetzung	1
1.2	Kontext der Arbeit	1
1.3	Stand der Technik	2
1.4	Überblick	4
2	Grundlagen der Momententheorie	5
2.1	Formbeschreibende Merkmale	5
2.1.1	Umrissorientierte Merkmale	5
2.1.2	Flächenorientierte Merkmale	6
2.2	Momente	7
2.2.1	Geometrische Momente	8
2.2.2	Zernike-Momente	9
2.2.3	Weitere Momente	10
2.3	Invariante Momente	12
2.3.1	Anforderungen an invariante Merkmale	12
2.3.2	Invariante Momente nach HU	13
2.3.3	Invariante Zernike-Momente	14
3	Bildaufbereitung	17
3.1	Das PM-Dateiformat	17
3.2	NAG-Bilder	18
3.3	Konturpunktlisten	19
3.4	Segmentierte Konturpunktlisten	20
4	Ergebnisse	21
4.1	Bildmaterial	21
4.2	Vergleich der invarianten Momente	22
4.2.1	Invariante Momente bei der Transformation eines Bildes	23
4.2.2	Invariante Momente bei verschiedenen Realisierungen eines Objektes	25
4.2.3	Invariante Momente bei unterschiedlichen Objekten	26
4.3	Klassifikation	27
4.3.1	Binäre Detektion	28
4.3.2	Multiple Detektion	34

Inhaltsverzeichnis

4.3.3	Mehrrerferenzen-Klassifikation	35
4.3.4	Dynamikbetrachtungen	36
4.4	Erweiterungen des Verfahrens	39
4.4.1	Segmentierung	39
4.4.2	Randunterdrückung	42
4.4.3	Kachelung	42
5	Zusammenfassung und Ausblick	47
5.1	Zusammenfassung und Beurteilung des Verfahrens	47
5.2	Ausblick	48
A	Ergänzungen zu der Momententheorie	51
A.1	Berechnung von Zernike-Momenten aus geometrischen Momenten	51
A.2	Implementierung der geometrischen Momente für eine diskrete Bildmatix	53
B	Ergänzungen zu den invarianten Zernike-Momenten	57
B.1	Berechnung der invarianten Zernike-Momente aus den Zernike-Momenten	57
B.2	Beispiele für invariante Zernike-Momente	58
B.2.1	Invariante Zernike-Momente bei transformierten Bildern	59
B.2.2	Invariante Zernike-Momente bei Bildern mit gleichem Objekt	60
B.2.3	Invariante Zernike-Momente bei Bildern mit unterschiedlichen Objekten	64
C	Verwendetes Bildmaterial	65
C.1	Bundesadler	65
C.2	Coca-Cola-Logos	68
C.3	China-Flaggen	71
C.4	Mercedessterne	73
C.5	UNO-Flaggen	75

Abbildungsverzeichnis

1.1	Prinzip des <i>content-based image retrieval</i>	2
2.1	Affine geometrische Abbildungen	13
2.2	Original und transformierte Bilder mit dem Baden-Württemberg-Wappen	16
3.1	Ebenen eines NAG-Bildes	19
3.2	Darstellung einer Konturpunktliste	20
3.3	Darstellung der einzelnen Segmentbilder	20
4.1	Objekte in den Bildern der Datenbank	22
4.2	Original und transformierte Bilder mit dem Bundesadler	24
4.3	Invariante Zernike-Momente des Bundesadlers und seiner transformierten Versionen	24
4.4	Verschiedene Bilder mit dem Objekt Bundesadler	25
4.5	Invariante Zernike-Momente für die verschiedenen Bilder mit dem Bundesadler in Abbildung 4.4	26
4.6	Invariante Zernike-Momente bei unterschiedlichen Objekten	27
4.7	Binäre Detektion	28
4.8	Bundesadler-Klassifikation bei $\gamma = 0,7$	29
4.9	Akzeptanzrate bei der Bundesadler-Klassifikation	30
4.10	Rückweisungsrate bei der Bundesadler-Klassifikation	30
4.11	ROC des Bundesadler-Klassifikators	31
4.12	Konturbild der China-Flagge	32
4.13	ROCs aller betrachteter Klassifikatoren	33
4.14	Multiple Detektion	34
4.15	NN-Klassifikation aller Bilder	35
4.16	ROC des Mehrreferenzen-Klassifikators	36
4.17	Gewichtungsfaktoren für ein gewichtetes Abstandsmaß	38
4.18	ROC bei verschiedenen Dynamikberücksichtigungen	38
4.19	Verteilung der Segmentlängen	40
4.20	KPL-Bilder zu Abbildung 4.19	40
4.21	Akkumulierte Segmentbilder für unterschiedliche minimale Segmentlängen	41
4.22	Invariante Zernike-Momente bei unterschiedlichen minimalen Segmentlängen	41
4.23	Randunterdrückung bei einem Testbild mit der China-Flagge	42

Abbildungsverzeichnis

4.24	Invariante Zernike-Momente für Bilder mit Randunterdrückung	43
4.25	Prinzip der überlappenden Kachelung	44
4.26	Bundesadler mit weiterem Objekt (gesamtes Bild)	44
4.27	Bundesadler mit weiterem Objekt (Kacheln)	45
4.28	Invariante Zernike-Momente bei der Kachelung eines Bildes (Bundesadler)	45
B.1	Verschiedene Bilder mit dem Objekt UNO-Emblem	61
B.2	Invariante Zernike-Momente der Bilder mit der UNO-Flagge in Abbildung B.1 . .	62
C.1	Bilder der Bundesadler-Datenbank	65
C.2	Bilder der Bundesadler-Datenbank (Fortsetzung)	66
C.3	Bilder der Bundesadler-Datenbank (Fortsetzung)	67
C.4	Bilder der Coca-Cola-Datenbank	68
C.5	Bilder der Coca-Cola-Datenbank (Fortsetzung)	69
C.6	Bilder der Coca-Cola-Datenbank (Fortsetzung)	70
C.7	Bilder der China-Flaggen-Datenbank	71
C.8	Bilder der China-Flaggen-Datenbank (Fortsetzung)	72
C.9	Bilder der Mercedesstern-Datenbank	73
C.10	Bilder der Mercedesstern-Datenbank (Fortsetzung)	74
C.11	Bilder der UNO-Flaggen-Datenbank	75

Tabellenverzeichnis

2.1	Beispiel für die invarianten Momente nach Hu	14
2.2	Anzahl der invarianten Zernike-Momente in Abhängigkeit von der Ordnung	15
2.3	Invariante Zernike-Momente der transformierten Bilder mit dem Wappen des Landes Baden-Württemberg in Abbildung 2.2	16
2.4	Normierter Euklidischer Abstand der verschiedenen invarianten Momente eines transformierten Bildes zum Original (Baden-Württemberg-Wappen)	16
4.1	Euklidischer Abstand der invarianten Zernike-Momente eines transformierten Bildes zum Original (Bundesadler)	23
4.2	Euklidischer Abstand zwischen invarianten Zernike-Momenten bei unterschiedlichen Bildern mit dem Bundesadler	25
4.3	Euklidischer Abstand zwischen den invarianten Zernike-Momenten der Referenzbilder	27
4.4	Euklidischer Abstand der invarianten Zernike-Momente eines gekachelten Bildes zum Original (Bundesadler)	45
A.1	Koeffizienten zur Berechnung der Zernike-Momente aus geometrischen Momenten	54
B.1	Invariante Zernike-Momente des Bundesadler-Referenzbildes und dessen transformierte Versionen in Abbildung 4.2	59
B.2	Invariante Zernike-Momente für die verschiedenen Bilder mit dem Bundesadler in Abbildung 4.4	60
B.3	Invariante Zernike-Momente der Bilder mit der UNO-Flagge in Abbildung B.1	63
B.4	Invariante Zernike-Momente der Bilder mit unterschiedlichen Objekten in Abbildung 4.1	64

Abkürzungsverzeichnis

Akronym	Beschreibung
BMP	Bitmap
CBIR	Content-Based Image Retrieval
DFT	Diskrete Fourier-Transformation
ERS	Abteilung Erkennungssysteme
FFT	Fast Fourier Transform
FhG	Fraunhofer-Gesellschaft
FT	Fourier-Transformation
GIF	Graphics Interchange Format
IITB	Fraunhofer-Institut für Informations- und Datenverarbeitung
JISC	Joint Information Systems Committee
JPG	Joint Photographic Experts Group
JTAP	JISC Technology Applications Program
KPL	Konturpunktliste
LPT	Log-Polar-Transformation
NAG	Normierte Ableitung der Gaußfunktion
NaN	Not a Number
NN	Nearest Neighbour
RGB	Rot Grün Blau
ROC	Receiver Operating Characteristic
ROI	Region-of-Interest
TBIR	Text-Based Image Retrieval
TIFF	Tagged Image File Format
WT	Wavelet-Transformation
ZMI	Zernike Moment Invariants

Verwendete Formelzeichen

\mathbb{N}	Menge der natürlichen Zahlen
\mathbb{R}	Menge der reellen Zahlen
\mathbb{R}^+	Menge der positiven reellen Zahlen
\mathbb{Z}	Menge der ganzen Zahlen
\mathbb{C}	Menge der komplexen Zahlen
$d(\mathbf{x}, \mathbf{y})$	Euklidischer Abstand der Vektoren \mathbf{x} und \mathbf{y}
i	komplexe Einheit
l	Wiederholungsindex von Zernike-Momenten
m_{pq}	Geometrisches Moment der Ordnung $p + q$
n	Ordnungsindex von Zernike-Momenten
p	Ordnungsindex von Geometrischen Momenten (in x-Richtung)
q	Ordnungsindex von Geometrischen Momenten (in y-Richtung)
$s(x, y)$	Grauwert an der diskreten Stelle (x, y)
$\tilde{s}(x, y)$	Zweidimensionale, kontinuierliche Bildfunktion
\bar{x}	x-Koordinate des Bildschwerpunktes
\bar{y}	y-Koordinate des Bildschwerpunktes
A_{nl}	Zernike-Moment der Ordnung n und Wiederholung l
\hat{A}_{nl}	Normiertes Zernike-Moment der Ordnung n und Wiederholung l
B_{nlk}	Koeffizient der radialen Zernike-Polynome
C_{pq}	Komplexes Moment der Ordnung $p + q$
G_{pq}	Geometrisches Moment der Ordnung $p + q$ über dem Einheitskreis
\hat{G}_{pq}	Normiertes geometrisches Moment der Ordnung $p + q$ über dem Einheitskreis
\tilde{G}_{pq}^{Tr}	Normiertes, zentriertes geometrisches Moment der Ordnung $p + q$ über dem Einheitskreis
H_i	Hypothese für Ereignis M_i
K	Anzahl der Objektklassen
K	Freier Parameter eines K NN-Klassifikators
M	Anzahl der Bildzeilen
M_{pq}	Verallgemeinertes Moment der Ordnung $p + q$
M_i	i . Ereignis
M	Anzahl der Referenzbilder
N	Anzahl der Bildspalten

Verwendete Formelzeichen

N	Anzahl der Bilder in der Datenbank
$P_n(x)$	Legendre-Polynom der Ordnung n
$P(H_i M_j)$	Wahrscheinlichkeit für H_i unter der Bedingung M_j
P_E	Entdeckungswahrscheinlichkeit
P_F	Falschalarmwahrscheinlichkeit
$R_{nl}(r)$	Radiales Zernike-Polynom der Ordnung n und Wiederholung l
$\tilde{R}_{nl}(r)$	Radiales Pseudo-Zernike-Polynom der Ordnung n und Wiederholung l
R	Anzahl Bildzeilen und -spalten der erweiterten Bildmatrix
S_{nlk}	Koeffizient der radialen Pseudo-Zernike-Polynome
$V_{nl}(r, \theta)$	Zernike-Polynom der Ordnung n und Wiederholung l
Z_i	i . invariantes Zernike-Moment
H	Filterkern zur Gradientenbildung
R_i	Entscheidungsraum für Ereignis M_i
S	Bildmatrix
Z	Vektor der invarianten Zernike-Momente
Z₀	Vektor der invarianten Zernike-Momente des Referenzbildes
γ	Exponent zur Normierung zentrierter Momente
γ	Schwelle bei binärer Detektion
δ_{ij}	Kronecker-Symbol
ϵ	Minimale Segmentlänge
ϕ_i	i . invariantes Moment nach Hu
ϕ	Vektor der invarianten Momente nach Hu
η_{pq}	Normiertes, zentriertes Moment der Ordnung $p + q$
η	Rückweisungsschranke bei multipler Detektion
κ	Kachelungsindex
λ_{pq}	Legendre-Moment der Ordnung $p + q$
μ_{pq}	Zentriertes Moment der Ordnung $p + q$
ν	Anzahl der invarianten Zernike-Momente pro Merkmalvektor
ρ_i	Gewichtungsfaktor für das i . invariante Moment
Δ	Offset-Vektor bei Randunterdrückung
$\Psi(x, y)$	Kern des verallgemeinerten Momentenbegriffs

1 Einleitung

1.1 Aufgabenstellung und Zielsetzung

Im Rahmen dieser Arbeit sollte ein Klassifikator entworfen werden, mit dem eine Aussage über die Existenz eines Objektes in einem Bild getroffen werden kann. Dafür sollten in diesem Ansatz Momente als charakteristische Merkmale für Objekte herangezogen werden. Momente sind Funktionen, die nur die Farb- bzw. Grauwerte der einzelnen Bildpunkte benötigen. Man spricht deshalb auch von *content-based image retrieval* (CBIR). Außerdem sollen die Ergebnisse der Klassifikation unabhängig von der Position oder Größe des Objektes im Bild sein. Aus den berechneten Momenten lassen sich solche invarianten Merkmale ableiten, die deshalb als *invariante Momente* bezeichnet werden. Diese sollen schließlich als Merkmalvektor für ein Bild bei der Klassifikation Verwendung finden. In dieser Arbeit wird die Eignung der invarianten Momente für die Klassifikationsaufgabe untersucht.

Das Ziel des Verfahrens besteht darin, unabhängig von der Position und Größe eines Objektes im Bild eine verlässliche Aussage über die Existenz des Objektes zu treffen. Damit soll es möglich sein, auch große Datenbanken mit Bildern nach einzelnen Objekten zu durchsuchen. Bei der Suche kann dann die Menge der möglicherweise relevanten Bilder wesentlich verkleinert werden.

Ein typisches Beispielszenario ist die musterbasierte Suche eines Bildes in einer Clipart-Sammlung. Oft hat der Anwender bereits ein Bild mit dem gewünschten Objekt vorliegen, er sucht aber weitere Bilder mit dem selben Inhalt und einer anderen Auflösung oder einer leicht unterschiedlichen Realisierung.

1.2 Kontext der Arbeit

Die vorliegende Arbeit wurde am Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB) durchgeführt. Es liegt dabei kein konkreter Projektbezug vor, die Thematik gliedert sich aber gut in die Forschungsgebiete der Abteilung Erkennungssysteme (ERS) ein. Dies sind vorwiegend die Bild- und Signalauswertung sowie die Objekt- und Situationserkennung.

Die gezielte Suche nach Bildern in großen Sammlungen nach vorgegebenen Kriterien (Bildwiedergewinnung, *Image Retrieval*) ist seit den 70er Jahren ein intensiv behandeltes Forschungsthema. Jeden Tag wachsen die Datenbanken um eine Vielzahl von Bildern an. Diese Datenmengen sind jedoch nur nutzbar, wenn die Bilder strukturiert abgelegt werden oder so indiziert werden,

1 Einleitung

dass eine Suche mit geeigneten Strategien schnell zu der gewünschten Auswahl führt.

Die ersten Strategien verwendeten textbasierte Datenbank-Management-Systeme (*text-based image retrieval*, TBIR), um Bilder über angehängte Schlüsselwörter (*keywords*) zu finden. Diese Technik trägt jedoch trotz vieler Erweiterungen, z.B. der multi-dimensionalen Indizierung, zwei große Nachteile: Zum einen bedeutet es bei großen Sammlungen sehr viel Arbeit, jedem Bild geeignete Schlüsselwörter beizufügen. Zum anderen ist die Wahl der Schlüsselwörter von der subjektiven Betrachtung der Person abhängig, die die Schlüsselwörter vergibt. Denn bei gleichem Bildinhalt können unterschiedliche Betrachter durchaus unterschiedliche Schlüsselwörter als geeignet betrachten.

In den 90er Jahren wurden als Abhilfe Suchstrategien vorgeschlagen, die an Stelle von textbasierten Beschreibungsformen den Bildinhalt selbst verwenden (*content-based image retrieval*). Diese Art des Image Retrieval wird seither durch viele Forschungsarbeiten beständig vorangetrieben. Sowohl aus dem Anfragebild (*query*) als auch aus den Bildern der Datenbank (Bildbasis) werden die relevanten Merkmalvektoren bestimmt und schließlich verglichen (siehe Abbildung 1.1), was zu einer Aussage führt, welche Bilder der Datenbank das Objekt des Referenzbildes enthalten.

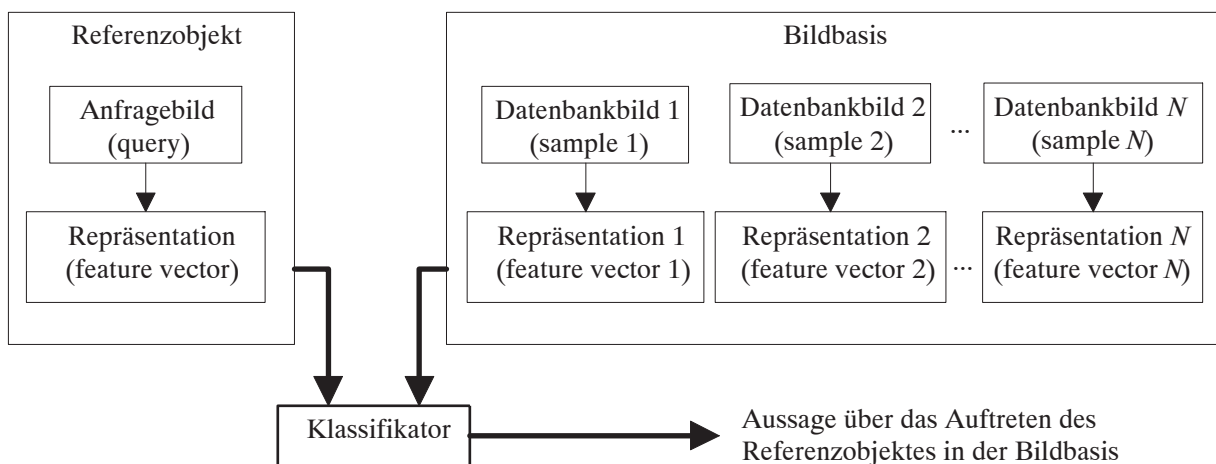


Abbildung 1.1: Prinzip des *content-based image retrieval*

1.3 Stand der Technik

Bisherige Verfahren zur automatischen Bildwiedergewinnung verwenden farb-/grauwert-, textur- oder formbasierte Merkmale, die für einen Ähnlichkeitsvergleich im Klassifikator herangezogen werden. Die Ansätze operieren teilweise auf Bildpyramiden, um durch Mehrfachauflösung der Skalierungsproblematik zu begegnen.

Farb-/grauwertbasierte Merkmale: Die Farbinformation ist eines der am häufigsten verwendeten Kriterien beim Image Retrieval. Das Histogramm der Farbwerte ist unabhängig

von der Bildgröße und -orientierung. Bei den farb- bzw. grauwertbasierten Merkmalen werden deshalb für das ganze Bild oder lokal begrenzt mehrere statistische Kennzahlen generiert. Diese bestehen z.B. aus Mittelwert, Standardabweichung und Momenten unterschiedlicher Ordnung des Histogramms. Zwei Bilder gelten als ähnlich, wenn sie in gewissen Toleranzen ähnliche Kennzahlen aufweisen. Die Kennzahlen lassen sich zwar leicht berechnen, besitzen jedoch in vielen Fällen wenig Aussagekraft.

Texturbasiert Merkmale: Unter einer Textur versteht man ein Pixelmuster, das ein gewisses Maß an Homogenität erfüllt, obwohl es nicht monochrom ist [RHC99]. Texturen charakterisieren damit vor allem Oberflächenstrukturen. Auch bei texturbasierten Merkmalen werden statistische Kennzahlen berechnet, z.B. Kontrast und Entropie, häufig mit Hilfe der Co-Occurance-Matrix (Grauwertübergangsmatrix). In neueren Arbeiten werden die statistischen Kennzahlen aus den Wavelet-Subbändern extrahiert und die biorthogonale Wavelet-Transformation oder die Gabor-Transformation verwendet.

Formbasierte Merkmale: Formmerkmale können sich auf den Umriss und auf die Fläche beziehen. In der Regel ist eine Segmentierung, oder zumindest eine Kantendetektion Voraussetzung für die Bestimmung von Formmerkmalen. Auf diese Art von Repräsentation eines Bildes wird in Abschnitt 2.1 genauer eingegangen, da die invarianten Momente, die in dieser Arbeit untersucht werden, Formmerkmale sind.

Kombination der Merkmale: Durch geeignete Kombination von Farb-, Textur- und Formmerkmalen werden bei aktuellen Image Retrieval-Systemen die unterschiedlichen Vorteile verknüpft. Die Schwierigkeit liegt dabei bei der sinnvollen Gewichtung der einzelnen Parameter.

Eine Übersicht über gegenwärtige Methoden und Systeme des Image Retrieval bieten die Quellen [EG99, VC01, RHC99].

In der Abteilung ERS des Fraunhofer-Instituts für Informations- und Datenverarbeitung wurde bereits eine Patentrecherche auf dem Gebiet des Image Retrieval durchgeführt. Die Patente lassen sich grob in zwei Gruppen unterteilen. Zum einen handelt es sich um Patente, die auf das Datenbanksystem, die Schlüsselwortverarbeitung und die Bedienoberfläche ausgerichtet sind:

- EP 1100024 A2, JP 2001142890 AA: Method and apparatus for image retrieval
- EP 1244025 A1, US 2001004739 A1, WO 2001040995 A1: Image retrieval system and image retrieval method
- EP 1130541 A2/A3: Data storage and retrieval playback apparatus for a still image receiver
- EP 1172741 A2, JP 2002091990 AA, US 2002048054 A1: On-demand image delivery server, image resource database, client terminal, and method of displaying retrieval results
- EP 1131954 A1, GB 9824783 A0, WO 2000028741 A1: Method and device for retrieving customized moving image data from a remote source and a moving image data source for remote retrieval

1 Einleitung

- EP 1217823 A2/A3, JP 2002245161 AA, US 2002105658 A1: Methods for image retrieval
- EP 1221659 A1, JP 2002202979 AA, US 2002092027 A1: Automatic image retrieval system
- JP 2001167095 AA: System and method for retrieving image

Zum anderen existieren Patente für die unterschiedlichen Methoden, um die Merkmale wie Textur-, Farb- oder Formparameter aus einem Bild zu extrahieren, insbesondere auch ein Patent, das die Berechnung von Zernike-Momenten beinhaltet:

- EP 1145148 A2/A3, WO 2001013263 A2/A3: System and method for performing region-based image retrieval using color-based segmentation
- EP 1221662 A2, JP 2002319024 AA, US 2002122587 A1: Image retrieval method
- EP 1202187 A2, JP 2002197117 AA: Image retrieval system and methods with semantic and feature based relevance feedback
- EP 1107136 A1, JP 2001209803 AA, US 2002010704 A1: Content-based image retrieval system and method for retrieving images using the same
- EP 1159709 A1, US 2001031103 A1, WO 2001041071 A1: Texture description method and texture-based image retrieval method using gabor filter in frequency domain
- EP 1208515 A1, WO 2001033480 A1: Method for extracting Zernike/Pseudo-Zernike moment

1.4 Überblick

Die Arbeit ist folgendermaßen gegliedert: Zunächst werden in Kapitel 2 die theoretischen Grundlagen von Momenten dargestellt. Darin wird auch auf die Ableitung invarianter Momente detailliert eingegangen. Im Anschluss wird in Kapitel 3 die Bildaufbereitung besprochen. Nach den dort dargestellten Methoden werden die Bilder für die weiteren Untersuchungen vorverarbeitet. Kapitel 4 enthält schließlich die Ergebnisse der betrachteten Klassifikationen mit invarianter Momenten. Nach einer Präsentation der charakteristischen Ergebniskurven folgt deren Diskussion und einige Ansätze zur Verbesserung und Erweiterung des Verfahrens. Die Arbeit schließt in Kapitel 5 mit einer Zusammenfassung und einem Ausblick.

2 Grundlagen der Momententheorie

In diesem Kapitel wird der mathematische Hintergrund der Momententheorie in den für diese Arbeit wichtigen Elementen beschrieben. Nach einem Abschnitt über die Beschreibungsmöglichkeiten geometrischer Formen (Abschnitt 2.1) werden zunächst verschiedene Momente betrachtet (Abschnitt 2.2), gefolgt von den daraus abgeleiteten invarianten Merkmalen in Abschnitt 2.3.

2.1 Formbeschreibende Merkmale

Bevor auf die Verwendung von invarianten Momenten eingegangen wird, soll an dieser Stelle eine Einführung in formbeschreibende Merkmale (*shape descriptors*) erfolgen. Sie zeigt die möglichen Alternativen für die im Folgenden eingesetzten Momente, wenn man Beschreibungsformen verwenden will, die insbesondere die geometrische Form von Objekten berücksichtigen. Eine kompakte Übersicht bietet [Rau94]. Die folgende Auflistung berücksichtigt einige der wichtigsten Verfahren zur Beschreibung von Objekten an Hand ihrer Form. In vielen Fällen wird auch eine Kombination dieser Merkmale verwendet, z.B. in [MPA98].

2.1.1 Umrissorientierte Merkmale

Alle umrissorientierte Merkmale setzen eine Kantendetektion voraus. Darauf aufbauend gibt es verschiedene Beschreibungen für das Bild.

Chain-Codierung: Das Grundprinzip der Chain-Codierung besteht darin, die Kontur eines Segmentes durch Richtungen zu beschreiben, denen man beim Wandern entlang der Kontur folgt [Hab95]. Nimmt man eine 8-Nachbarschaft an, d.h. genau acht Pixel in der Umgebung eines Pixels werden als dessen Nachbarn betrachtet, so sind die Richtungen Vielfache von 45° und mit 3 Bit codierbar. Für nicht zusammenhängende Konturen werden virtuelle Ketten verwendet, die genau wie Start- und Endpunkt durch spezielle Codefolgen eingeleitet werden.

Signatur: Mit der Signatur wird die zweidimensionale Beschreibung eines Segmentes auf eine eindimensionale Funktion abgebildet. Eine einfache Umsetzung wird erreicht, wenn jeweils der Abstand zum Schwerpunkt von jedem Pixel des Umrisses aus berechnet wird.

2 Grundlagen der Momententheorie

Fourier-Deskriptoren: Fasst man die (x, y) -Koordinaten eines Pixels in der Bildebene als Real- und Imaginärteil einer komplexen Zahl $z = x + iy \in \mathbb{C}$ auf, so werden die Pixelkoordinaten einer geschlossenen Umrisslinie zu einer periodischen, komplexen Zahlenfolge. Als Fourier-Deskriptoren werden dann die einzelnen Werte der Fourier-Transformierten dieser Folge bezeichnet, die man im Allgemeinen mit der FFT berechnet [GW93]. Eine einfache Implementierung der Fourier-Deskriptoren basierend auf einer 8-Nachbarschaft enthält [BKB99].

2.1.2 Flächenorientierte Merkmale

Die folgenden Merkmale beschreiben eine Form nicht über den Umriss, sondern basierend auf der gesamten Bildfläche.

Momente: Momente sind globale charakteristische Größen eines Bildes. Sie werden mit Summationen über die Grauwerte berechnet, bei logischen Bildern (Binärbilder) erstrecken sich die Summationen über die Fläche der gesetzten Pixel. Die verschiedenen Momente werden im folgenden Abschnitt detailliert beschrieben.

Log-Polar-Transformation: Bei der Log-Polar-Transformation (LPT) wird ein Bildpunkt (x, y) des kartesischen Koordinatensystems in Polarkoordinaten (r, ϕ) transformiert. Der komplexe Logarithmus der komplexen Zahl $z = x + iy = |z| \exp(i \arg z)$ ist definiert als

$$\ln z = \ln |z| + i \arg z + i \cdot 2k\pi, \quad k \in \mathbb{Z}. \quad (2.1)$$

Fasst man nun jeden Bildpunkt (x, y) als komplexe Zahl $z = x + iy \in \mathbb{C}$ auf, so lautet die Transformationsvorschrift

$$LPT : (x, y) \mapsto (\ln |z|, \arg z). \quad (2.2)$$

Diese Transformation ist eine Anlehnung an die Biologie des menschlichen Auges, wo eine analoge Abbildung von der Netzhaut (Retina) in die Hirnrinde (Kortex) stattfindet [Sch77].

Die Konsequenz aus dieser Abbildung ist, dass eine Rotation oder Skalierung im kartesischen Koordinatensystem auf eine einfache Translation im Log-Polar-Raum abgebildet wird. Durch die nicht-uniforme Abtastung entsteht eine vom Bildmittelpunkt aus abnehmende Auflösung.

Wavelet-Parameter: Die Bilder in Datenbanken werden oft in einer komprimierten Form abgespeichert, die Wavelet-Koeffizienten verwendet. Deshalb bieten Wavelet-Koeffizienten eine gute Basis, um formbeschreibende, invariante Merkmale abzuleiten. Die Wavelet-Transformation (WT) der kontinuierlichen Funktion $f(t)$ ist definiert als [Kro96]

$$F(a, b) = \int_{-\infty}^{\infty} f(t) \Psi_{a,b}^*(t) dt \quad (2.3)$$

mit dem Wavelet

$$\Psi_{a,b}(t) = a^{-1/2} \cdot \Psi\left(\frac{t-b}{a}\right), \quad a \in \mathbb{R}^+, b \in \mathbb{R}. \quad (2.4)$$

Die Wavelet-Transformation zerlegt ein Bild in eine Pyramide von Teilbildern mit unterschiedlicher Auflösung entsprechend dem Skalierungsfaktor a . In [MPA98] wurde herausgestellt, dass sich dann zwei bestimmte Merkmale, die aus der Statistik der Hochpass-Teilbilder berechnet werden, gut als formbeschreibende Merkmale eignen. Diese Merkmale basieren auf den Parametern α, β, γ einer verallgemeinerten Gaußschen Dichtefunktion

$$f_X(x) = \alpha \exp(-|\beta x|^\gamma), \quad (2.5)$$

mit der das Histogramm der Hochpass-Teilbilder näherungsweise dargestellt (*matched*) wird. In [DAV98] werden Momente von Maxima der Wavelet-Transformierten als formbeschreibende Merkmalvektoren vorgeschlagen.

2.2 Momente

Momente sind vor allem aus der Wahrscheinlichkeitstheorie und der Mechanik bekannt. So ist beispielsweise das erste Moment einer Zufallsgröße der Erwartungswert und das zweite zentrierte Moment die Varianz. Die meisten der höheren Momente entziehen sich der Anschauung. In der Bildverarbeitung kann eine ähnliche Interpretation verwendet werden, wenn die kontinuierliche oder diskrete Bildfunktion in Analogie zur Wahrscheinlichkeitsdichtefunktion der Zufallsgröße betrachtet wird.

Im Bereich der Bildanalyse werden Momente in vielen Bereichen verwendet, um Objekte oder Bildsegmente zu charakterisieren. Der Vorteil von Momenten ist, dass einfach invariante Merkmale daraus extrahiert werden können, wie in Abschnitt 2.3 beschrieben wird. Es soll an dieser Stelle betont werden, dass die Momente selbst im Allgemeinen nicht invariant sind.

Es existieren in der Literatur verschiedene Arten von Momenten, die entsprechend ihren Vorteilen in der Praxis eingesetzt werden. Eine Übersicht der grundlegenden Eigenschaften bietet [TC88]. Darin wird außerdem untersucht, inwiefern ein Bild aus den jeweiligen Momenten rekonstruiert werden kann und wie sich die Momente bezüglich Rauschempfindlichkeit und Redundanz aus informationstheoretischer Sicht verhalten. Eine ausführliche Diskussion der Momententheorie ist in [MR98] zu finden.

Grundsätzlich ist mit allen beschriebenen Momenten eine vollständige, fehlerfreie Rekonstruktion des digitalen Bildes möglich, falls die Ordnung der Momente groß genug ist. Dann entspricht die Beschreibung des Bildes mit Momenten lediglich einer Transformation der Basisvektoren.

2 Grundlagen der Momententheorie

2.2.1 Geometrische Momente

Das *Moment der Ordnung* $p + q$ einer zweidimensionalen, reellwertigen, kontinuierlichen Funktion $\tilde{s}(x, y)$ ist für alle $p, q = 0, 1, 2, \dots$ nach [Hab95, GW93] definiert als

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q \tilde{s}(x, y) dx dy. \quad (2.6)$$

Die Funktion $\tilde{s}(x, y)$ enthält die Farb- bzw. Grauwertinformation für jeden Punkt (x, y) der Bildebene. Momente können damit als Projektion dieser Bildfunktion auf die Monome $\{x^p y^q\}$ interpretiert werden. Zur Unterscheidung von anderen Momenten spricht man auch von *geometrischen Momenten*. In der diskreten Form für ein digitalisiertes Bild $\mathbf{S} = ((s(x, y)))$ der Dimension $M \times N$ berechnet sich das geometrische Moment der Ordnung $p + q$ zu

$$m_{pq} = \sum_{x=1}^M \sum_{y=1}^N x^p y^q s(x, y). \quad (2.7)$$

Die Integration bzw. Summation erstreckt sich dabei über das gesamte Bild. Daraus lassen sich mit den Centroiden

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad (2.8)$$

und

$$\bar{y} = \frac{m_{01}}{m_{00}} \quad (2.9)$$

die *zentrierten Momente der Ordnung* $p+q$ berechnen:

$$\mu_{pq} = \sum_{x=1}^M \sum_{y=1}^N (x - \bar{x})^p (y - \bar{y})^q s(x, y). \quad (2.10)$$

Die *normierten, zentrierten Momente der Ordnung* $p+q$ berechnen sich schließlich zu

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (2.11)$$

mit

$$\gamma = \frac{p+q}{2} + 1, \quad p+q \geq 2. \quad (2.12)$$

Durch die Verschiebung des Koordinatenursprungs in den Schwerpunkt des Bildes und die Normierung erhält man mit (2.11) formbeschreibende Merkmale, die sowohl translations- als auch skalierungsinvariant sind.

Für die Berechnung der geometrischen Momente $\{m_{pq}\}$ existieren schnelle Algorithmen basierend auf den Green'schen Formeln [YA94]. Dabei wird die zweifache Integration über die Fläche eines Bildsegments mit einer einfachen Integration entlang der Kontur ermittelt, was eine Einschränkung an die Bildkomposition bedingt. Ein anderer, häufig verfolgter Ansatz interpretiert die Momente als Filterergebnis einer Hintereinanderausführung mehrerer Single-Pole-Filter nach Hatamian [Hat86]. Grundlegende Überlegungen zur Berechnung der Momente über den Rand geschlossener Flächen findet man in [Leu91, Flu98].

Geometrische Momente sind häufig eine Vorstufe in der Berechnung invarianter Merkmale (siehe Abschnitt 2.3) oder sind Grundlage für die im Folgenden beschriebenen anderen Momente.

2.2.2 Zernike-Momente

Zernike-Momente werden häufig zur Mustererkennung und Bildanalyse eingesetzt. An dieser Stelle und für die weitere Darstellung wird die Notation nach [TC88] verwendet. Das komplexe Zernike-Moment der Ordnung n mit Wiederholung l ist definiert als

$$A_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{nl}(r, \phi)]^* \tilde{s}(r \cos \phi, r \sin \phi) r dr d\phi = A_{n,-l}^*, \quad (2.13)$$

wobei $n = 0, 1, 2, \dots$ und l ganzzahlig mit den Bedingungen $(n - |l|)$ gerade und $|l| \leq n$ ist. Die Bildung des konjugiert Komplexen ist durch den Stern * gekennzeichnet. Die Indices n und l sind dabei nur eine Umdefinition der bisher verwendeten Indices p und q . Die komplexwertigen Zernike-Polynome

$$V_{nl}(r, \phi) = R_{nl}(r) \exp(il\phi) \quad (2.14)$$

bilden eine vollständige, orthogonale Basis innerhalb des Einheitskreises:

$$\int_0^{2\pi} \int_0^1 [V_{nl}(r, \phi)]^* V_{mk}(r, \phi) r dr d\phi = \frac{\pi}{n+1} \delta_{mn} \delta_{kl} \quad (2.15)$$

(δ_{ij} ist das Kronecker-Symbol). Die reelwertigen, radialen Polynome $R_{nl}(r)$ sind schließlich definiert als

$$R_{nl}(r) = \sum_{s=0}^{(n-|l|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|l|}{2} - s\right)! \left(\frac{n-|l|}{2} - s\right)!} r^{n-2s} = \sum_{\substack{k=|l| \\ (n-k) \text{ gerade}}}^n B_{n|l|k} r^k \quad (2.16)$$

mit den Koeffizienten [MR95]

$$B_{nlk} = (-1)^{\frac{n-k}{2}} \frac{\left(\frac{n+k}{2}\right)!}{\left(\frac{n-k}{2}\right)! \left(\frac{k+l}{2}\right)! \left(\frac{k-l}{2}\right)!} \quad (2.17)$$

und genügen der Orthogonalitätsbedingung

$$\int_0^1 R_{nl}(r) R_{ml}(r) r dr = \frac{1}{2(n+1)} \delta_{mn}. \quad (2.18)$$

In dieser Arbeit werden Zernike-Momente als Grundlage für die Ableitung invarianter Merkmale gewählt, da sie die besten Eigenschaften bezüglich der Bildrepräsentation mit Momenten besitzen [MR98].

Zernike-Momente können einfach durch die geometrischen Momente dargestellt werden. Für $l \geq 0$ gilt [TC88]:

$$A_{nl} = \frac{n+1}{\pi} \sum_{k=l}^n \sum_{j=0}^q \sum_{m=0}^l (-i)^m \binom{q}{j} \binom{l}{m} B_{nlk} \cdot \mu_{k-2j-m, 2j+m} \quad (2.19)$$

$(n-k)$ gerade

2 Grundlagen der Momententheorie

mit $q = \frac{1}{2}(k - l)$. Da diese Äquivalenz nicht unmittelbar einsichtig, aber für die praktische Implementierung von großer Bedeutung ist, wird die Herleitung zu (2.19) in Anhang A.1 durchgeführt. Damit lassen sich auch für die Berechnung von Zernike-Momenten die schnellen Algorithmen zur Berechnung der geometrischen Momente verwenden. In [ARY02] sind detaillierte Geschwindigkeitsangaben zu finden für einen Rechner mit einem CPU-Takt von 350MHz. Für ein Bild der Dimension 256×256 Pixel werden die Zernike-Momente bis zur Ordnung 12 beispielsweise in 0.085 Sekunden berechnet.

Die ersten Zernike-Momente lauten in Abhängigkeit von den zentrierten geometrischen Momenten $\{\mu_{pq}\}$ nach (2.10)

$$\begin{aligned}
 A_{00} &= \frac{\mu_{00}}{\pi} = \frac{1}{\pi} \\
 A_{11} &= 0 \\
 A_{1,-1} &= 0 \\
 A_{20} &= \frac{3}{\pi}(\mu_{20} + 2\mu_{02} - 1) \\
 A_{22} &= \frac{3}{\pi}(\mu_{02} - \mu_{20} - 2i\mu_{11}) \\
 A_{2,-2} &= \frac{3}{\pi}(\mu_{02} - \mu_{20} + 2i\mu_{11}) \\
 A_{31} &= \frac{12}{\pi}[\mu_{03} + \mu_{21} - i(\mu_{30} + \mu_{12})] \\
 A_{3,-1} &= \frac{12}{\pi}[\mu_{03} + \mu_{21} + i(\mu_{30} + \mu_{12})] \\
 A_{33} &= \frac{4}{\pi}[\mu_{03} - 3\mu_{21} + i(\mu_{30} - 3\mu_{12})] \\
 A_{3,-3} &= \frac{4}{\pi}[\mu_{03} - 3\mu_{21} - i(\mu_{30} - 3\mu_{12})],
 \end{aligned} \tag{2.20}$$

wobei dies lediglich eine Interpretation von (2.19) für $n \leq 3$ darstellt.

Auf Grund der Orthogonalitätseigenschaft sind die Zernike-Momente ab $n = 2$ linear unabhängig und aus informationstheoretischer Sicht redundanzfrei.

Für die praktische Implementierung ist bei Zernike-Momenten das Integrationsgebiet besonders zu beachten. Im Gegensatz zu den geometrischen Momenten wird über den Einheitskreis integriert, d.h. das Bild muss vorher auf diese Größe skaliert werden. Eine detaillierte Betrachtung dieses Problems enthält Anhang A.2.

2.2.3 Weitere Momente

In der Literatur wird häufig auch ein verallgemeinerter Momentenbegriff mit dem Kern $\Psi_{pq}(x, y)$ und unterschiedlichen Integrationsgebieten verwendet:

$$M_{pq} = \int \int \Psi_{pq}(x, y) \tilde{s}(x, y) dx dy. \tag{2.21}$$

Die wichtigsten Realisierungen dieser Gleichung neben den bereits Besprochenen sollen an dieser Stelle der Vollständigkeit wegen aufgeführt werden. Eine detaillierte Diskussion bietet [MR98].

Legendre-Momente: Nach [Tea80] ist das orthogonale Legendre-Moment der Ordnung $(p + q)$ von $\tilde{s}(x, y)$ definiert als

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^1 \int_{-1}^1 P_p(x)P_q(y)\tilde{s}(x, y)dxdy \quad (2.22)$$

mit $p, q = 0, 1, 2, \dots$. Das Integrationsgebiet ist dabei $-1 \leq x, y \leq 1$ und die Bildfunktion muss in diesen Bereich skaliert werden. Die Legendre-Polynome $\{P_n(x)\}$ mit

$$P_n(x) = \sum_{j=0}^n a_{nj}x^j = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n \quad (2.23)$$

bilden eine vollständige, orthogonale Basis im Intervall $[-1, 1]$:

$$\int_{-1}^1 P_p(x)P_q(x)dx = \frac{2}{2p+1} \delta_{pq}. \quad (2.24)$$

Legendre-Momente können einfach durch geometrische Momente dargestellt werden und umgekehrt. Der Vorteil gegenüber den geometrischen Momenten liegt in der Orthogonalitätseigenschaft. Allerdings zeigen Legendre-Momente eine höhere Rauschempfindlichkeit als andere Momente.

Pseudo-Zernike-Momente: Pseudo-Zernike-Momente sind aus den Zernike-Momenten abgeleitet und enthalten ebenfalls Polynome in r . Die radialen Polynome $\tilde{R}_{nl}(r)$ erlauben auch $(n - |l|)$ *ungerade* und sind folgendermaßen gegenüber (2.16) modifiziert:

$$\tilde{R}_{nl}(r) = \sum_{s=0}^{n-|l|} (-1)^s \frac{(2n+1-s)!}{s!(n-|l|-s)!(n+|l|+1-s)!} r^{n-s} = \sum_{k=|l|}^n S_{n|l|k} r^k. \quad (2.25)$$

Damit besteht die Menge der Pseudo-Zernike-Momente der maximalen Ordnung n aus $(n+1)^2$ linear unabhängigen Polynomen im Gegensatz zu den $\frac{1}{2}(n+1)(n+2)$ Zernike-Momenten.

Pseudo-Zernike-Momente werden eingesetzt, wenn eine geringe Rauschempfindlichkeit gefordert ist [TC88].

Komplexe Momente: Insbesondere für verallgemeinerte Betrachtungen der Momententheorie werden die komplexen Momente verwendet. Mit $p, q = 0, 1, 2, \dots$ ist das komplexe Moment der Ordnung $(p + q)$ definiert als

$$C_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+iy)^p (x-iy)^q \tilde{s}(x, y) dxdy \quad (2.26)$$

2 Grundlagen der Momententheorie

Es kann gezeigt werden, dass die komplexen Momente Linearkombinationen der geometrischen Momente $\{m_{pq}\}$ sind, jedoch mit komplexen Koeffizienten. Auch zu den Zernike- und Pseudo-Zernike-Momenten besteht eine einfache Abhängigkeit.

Komplexe Momente zeigen die größte Resistenz gegen Rauschen.

2.3 Invariante Momente

In diesem Abschnitt sollen nun invariante Merkmale definiert und auf der Basis der beschriebenen Momente genauer betrachtet werden. Die Wichtigsten sind aus den geometrischen Momenten oder aus den Zernike-Momenten abgeleitet. Für spezielle Anwendungsfälle gibt es weitere invariante Merkmale, die z.B. gegen Beleuchtungsschwankungen invariant sind [MPA98].

Jedes invariante Merkmal ist durch eine funktionale Abhängigkeit von den Momenten für jedes Bild berechenbar. Alle berechneten invarianten Merkmale eines Bildes werden dann zu einem Merkmalvektor (*feature vector*) zusammengefasst und als Punkt im Merkmalraum (*feature space*) interpretiert.

2.3.1 Anforderungen an invariante Merkmale

Ein für eine Bildmatrix S berechnetes Merkmal heißt invariant, wenn es den selben Wert annimmt, auch wenn das Bild eine affine geometrische Transformation erfährt. Eine allgemeine affine Transformation wird durch die Variablentransformation

$$S'(x, y) = S(a_0 + a_1x + a_2y, b_0 + b_1x + b_2y) \quad (2.27)$$

beschrieben. Dazu zählen insbesondere (siehe Abbildung 2.1)

- Translation mit dem Verschiebungsvektor $(\Delta x, \Delta y)^T$: $S'(x, y) = S(x + \Delta x, y + \Delta y)$,
- Rotation mit dem Drehwinkel θ : $S'(x, y) = S(x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$ und
- Skalierung um den Faktor λ : $S'(x, y) = S(\lambda x, \lambda y)$.

Diese Anforderungen werden von den Momenten selbst bereits teilweise erfüllt. So sind die normierten, zentrierten geometrischen Momente nach (2.11) invariant bei einer Translation oder Skalierung des Bildes, nicht aber bei einer Rotation. Die Beträge der Zernike-Momente nach (2.13) sind auf Grund der Struktur der Zernike-Momente bereits rotationsinvariant, aber nicht invariant bei einer Translation oder Skalierung des Bildes.

Mit Hilfe der allgemeinen algebraischen Invariantentheorie können nun aus den Momenten invariante Merkmale abgeleitet werden. Da diese Merkmale auf den Momenten des Bildes basieren, werden sie synonym als *invariante Momente* bezeichnet.

Für die meisten Anwendungen genügen Merkmale, die invariant sind bezüglich Translation, Rotation und Skalierung. Viele der Merkmale sind zusätzlich invariant bei einer Spiegelung des Bildes. Der Fall der allgemeinen affinen Transformation nach (2.27) wird in [FS93] diskutiert.

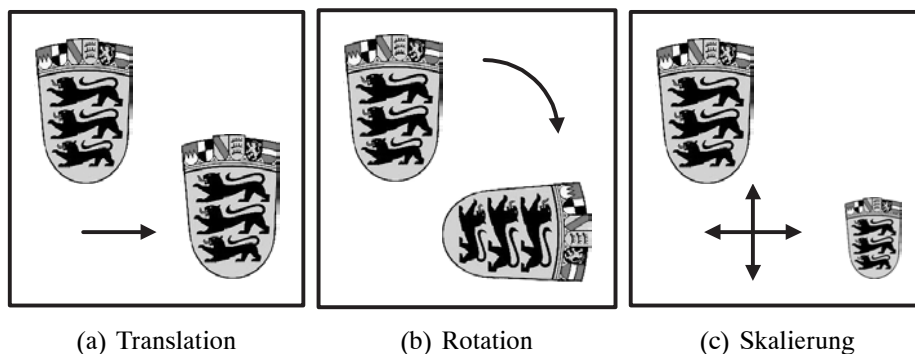


Abbildung 2.1: Affine geometrische Abbildungen

2.3.2 Invariante Momente nach HU

Hu leitete bereits 1962 aus den normierten, zentrierten Momenten $\{\eta_{pq}\}$ sieben Merkmale (*Hu moment invariants*) ab, die gegenüber den wichtigsten Transformationen invariant sind [Hu62]:

$$\begin{aligned}
 \phi_1 &= \eta_{20} + \eta_{02} \\
 \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
 \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
 \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
 \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
 &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
 \phi_6 &= (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
 \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
 &\quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2].
 \end{aligned} \tag{2.28}$$

Diese invarianten Momente nach Hu sind jedoch linear abhängig [Flu00] und nicht invariant im Falle der allgemeinen affinen Transformation (2.27) [Rei91]. Trotzdem werden sie in vielen Anwendungen eingesetzt. Die wichtigsten Gebiete sind die Erkennung von Flugzeugen [DBM77] und die Schriftzeichenerkennung [CH87, TMRSSG00]. Die einzelnen invarianten Momente werden als Vektor $\phi = (\phi_1, \phi_2, \dots, \phi_7)^T$ zusammengefasst.

Als Beispiel werde ein Bild mit dem Wappen des Landes Baden-Württemberg betrachtet. Abbildung 2.2 zeigt neben dem Original auch die transformierten Versionen des Bildes.

Tabelle 2.1 zeigt die Ergebnisse der Gleichung (2.28) für die einzelnen Bilder in Abbildung 2.2. Zum direkten Vergleich der invarianten Momente ist in Tabelle 2.1 zusätzlich der Euklidische Abstand zwischen den Merkmalvektoren des transformierten Bildes und dem Original angegeben:

$$d(\phi, \phi_{Original}) = \|\phi - \phi_{Original}\| = \sqrt{\sum_{i=1}^7 (\phi_i - \phi_{i,Original})^2}. \tag{2.29}$$

2 Grundlagen der Momententheorie

	Original	Skaliert 125%	Skaliert 75%	Verschieben 20 Pixel	Gedreht 5° im UZS	Gedreht 90° im UZS	Gespiegelt horizontal
ϕ_1	-0.36186	-0.12235	-0.52711	-0.36186	-0.40351	-0.36416	-0.35883
ϕ_2	1.4563	1.9403	1.2479	1.4563	1.3952	1.4665	1.4651
ϕ_3	1.6566	2.313	1.1337	1.6566	1.4772	1.6604	1.6554
ϕ_4	2.146	2.8533	1.6159	2.146	1.9571	2.1425	2.1556
ϕ_5	5.0238	6.5468	3.9404	5.0238	4.6631	4.9881	5.0575
ϕ_6	3.5231	4.5533	2.8733	3.5231	3.3138	3.5057	3.5513
ϕ_7	6.2572	7.6223	5.5347	6.2572	5.7628	6.2944	6.3619
$\ \phi - \phi_{Original}\ $		2.5431	1.6562	0	0.70118	0.055677	0.11433

Tabelle 2.1: Beispiel für die invarianten Momente nach Hu

Die Werte der einzelnen invarianten Momente liegen jeweils im gleichen Bereich und unterscheiden sich nicht für jede Transformation genauso stark. Eine Vergrößerung des Bildes zieht den größten Fehler bei der Momentenberechnung mit sich, während der Fehler bei der Translation in diesem Beispiel sogar den Wert Null annimmt.

Invariante Momente nach Hu bilden damit die historische Grundlage für die Arbeit mit invarianten Momenten in der Bildverarbeitung, werden aber bei dem hier entworfenen Klassifikator nicht als Unterscheidungskriterium für Bilder betrachtet.

2.3.3 Invariante Zernike-Momente

Die Basis für die Berechnung der invarianten Zernike-Momente bildet Gleichung (2.13) bzw. in der Realisierung mit geometrischen Momenten Gleichung (2.20). Das Bild muss bereits vor der Berechnung der Zernike-Momente $\{A_{nl}\}$ skaliert und so transformiert werden, dass der Bildschwerpunkt in der Bildmitte liegt [KH99].

Die Beträge $|A_{nl}|$ der Zernike-Momente sind schon invariant gegenüber einer Rotation des Bildes, da der Drehwinkel nur in einer veränderten Phase der radialen Polynome (2.14) resultiert. Merkmale, die zusätzlich gegenüber einer Skalierung oder Translation des Bildes invariant sind, können für jede Ordnung konstruiert werden. Die $\nu = 6$ invarianten Zernike-Momente $\mathbf{Z}^T = (Z_1, Z_2, \dots, Z_\nu)$ bis zur Ordnung 3 lauten [Tea80]

$$\begin{aligned}
 Z_1 &= A_{20} \\
 Z_2 &= |A_{22}|^2 \\
 Z_3 &= |A_{33}|^2 \\
 Z_4 &= |A_{31}|^2 \\
 Z_5 &= A_{33}(A_{31}^*)^3 + [A_{33}(A_{31}^*)^3]^* \\
 Z_6 &= (A_{31})^2 A_{22}^* + [(A_{31})^2 A_{22}^*]^*
 \end{aligned} \tag{2.30}$$

Weitere invariante Momente sind dem Anhang B.1 zu entnehmen. In [MR98] werden die invarianten Momente zusätzlich unterteilt in *primäre* und *sekundäre invariante Momente*. Für jede

Ordnung setzen sich die invarianten Merkmale damit zusammen aus allen primären invarianten Momenten, die den Gleichungen

$$Z^{(1)} = A_{n0} \quad \text{oder} \quad Z^{(2)} = |A_{nl}|^2 \quad (2.31)$$

genügen und allen sekundären invarianten Momenten nach

$$Z^{(3)} = (A_{nl})^* A_{rs}^m + [(A_{nl})^* A_{rs}^m]^*, \quad (2.32)$$

wobei l durch s teilbar sein muss, $m \geq 1$ und $m = l/s$ gilt. Der Term $(A_{nl})^* A_{rs}^m$ in (2.32) ist selbst bereits invariant, durch die Addition des konjugiert Komplexen wird aber ein reellwertiges Ergebnis erreicht.

Die Wahl der Ordnung bestimmt damit die Anzahl der invarianten Merkmale, die verwendet werden können (siehe Tabelle 2.2). Es ist dabei allerdings insofern ein Kompromiss einzugehen, als mit größerer Ordnung zwar die Bildrepräsentation durch die Erfassung von mehr Bilddetails genauer wird, gleichzeitig jedoch die Rauschempfindlichkeit zunimmt.

In dieser Arbeit wird als maximale Ordnung $n_{max} = 7$ gewählt, sodass für jedes Bild ein Merkmalvektor $\mathbf{Z}^T = (Z_1, Z_2, \dots, Z_{32}) \in \mathbb{R}^{32}$ bestehend aus 32 Elementen berechnet wird. In der Literatur wird beispielsweise für detailreiche Landschaftsaufnahmen eine maximale Ordnung $n_{max} = 12$ gewählt [ARY02], zur Erkennung von Flugzeugen $n_{max} = 6$ [BSA91] und bei der Schriftzeichenerkennung $n_{max} = 8$ [Tea80].

Maximale Ordnung	n_{max}	2	3	4	5	6	7	8	9	10
Anzahl der Merkmale	ν	2	6	11	17	24	32	41	51	62

Tabelle 2.2: Anzahl der invarianten Zernike-Momente in Abhängigkeit von der Ordnung

Für das Bild 2.2(a) und dessen transformierte Versionen lassen sich nun diese invarianten Merkmale beispielhaft berechnen. Um den direkten Vergleich mit den invarianten Momenten nach Hu zu ermöglichen, werden an dieser Stelle ebenfalls nur Momente bis zur maximalen Ordnung $n_{max} = 3$ betrachtet, für die Klassifikation (siehe Kapitel 4) werden alle 32 invarianten Zernike-Momente herangezogen. Tabelle 2.3 enthält die einzelnen Werte $\mathbf{Z}^T = (Z_1, Z_2, \dots, Z_6)$. Am Euklidischen Abstand der Vektoren der transformierten Bilder zu dem Vektor des Originalbildes, $d(\mathbf{Z}, \mathbf{Z}_{Original}) = \|\mathbf{Z} - \mathbf{Z}_{Original}\|$, ist ersichtlich, dass der Fehler sehr kleine Werte annimmt. Der Unterschied zum Vektor des Originalbildes fällt bei einer Drehung um 5° am stärksten, bei einer Verschiebung des Bildes am geringsten aus.

Ein Vergleich des Fehlers mit dem der invarianten Momente nach Hu zeigt, dass bei äquivalentem $n_{max} = 3$ die Verwendung von Zernike-Momenten wesentlich besser ist. Selbst wenn die Größenordnung der Vektoren berücksichtigt wird, gilt entsprechend Tabelle 2.4 für alle Transformationen

$$\frac{d(\mathbf{Z}, \mathbf{Z}_{Original})}{\|\mathbf{Z}_{Original}\|} \leq \frac{d(\phi, \phi_{Original})}{\|\phi_{Original}\|}. \quad (2.33)$$

2 Grundlagen der Momententheorie

	Original	75% skaliert	125% skaliert	Vershoben	5° gedreht	90° gedreht	Ggespiegelt
Z_1	-0.55527	-0.56188	-0.55452	-0.55527	-0.59327	-0.55487	-0.55507
Z_2	0.067186	0.065911	0.063325	0.067186	0.060125	0.066755	0.067127
Z_3	0.030937	0.031125	0.031449	0.030937	0.027367	0.030818	0.031121
Z_4	0.053195	0.050134	0.054382	0.053195	0.047271	0.053664	0.052778
Z_5	8.4216e-006	6.7637e-006	9.5167e-006	8.4216e-006	5.7809e-006	-8.5325e-006	8.2526e-006
Z_6	0.00037078	0.0003227	0.00036944	0.00037078	0.00025949	0.00037534	0.0003656
$\ \mathbf{Z} - \mathbf{Z}_{Original}\ $		0.007401	0.0041404	0	0.039264	0.0007605	0.00050007

Tabelle 2.3: Invariante Zernike-Momente der transformierten Bilder mit dem Wappen des Landes Baden-Württemberg in Abbildung 2.2

	Verkleinert	Vergrößert	Vershoben	Gedreht	Gekippt	Gespiegelt
$d(\phi, \phi_{Original}) / \ \phi\ $	0.2736	0.1782	0	0.0754	0.0060	0.0123
$d(\mathbf{Z}, \mathbf{Z}_{Original}) / \ \mathbf{Z}\ $	0.0132	0.0074	0	0.0698	0.0014	0.0009

Tabelle 2.4: Normierter Euklidischer Abstand der verschiedenen invarianten Momente eines transformierten Bildes zum Original (Baden-Württemberg-Wappen)

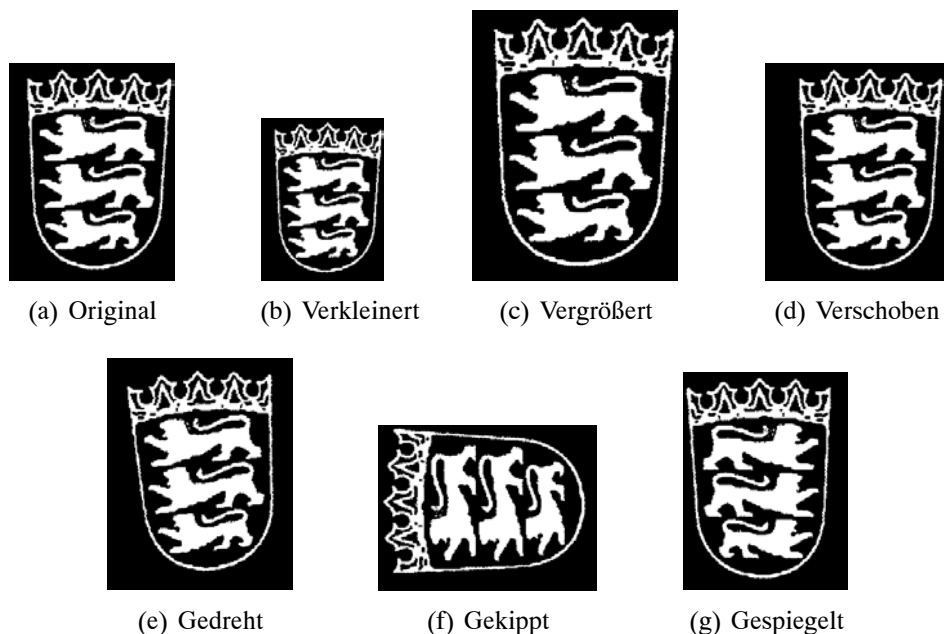


Abbildung 2.2: Original und transformierte Bilder mit dem Wappen des Landes Baden-Württemberg

3 Bildaufbereitung

Die Bilder für die vorliegende Klassifikationsaufgabe von Objekten in Bildern können in verschiedener Größe und Farbtiefe, sowie in unterschiedlichen Dateiformaten vorhanden sein. Um die invarianten Momente als Elemente eines vergleichbaren Merkmalvektors berechnen zu können, müssen daher die Daten in eine fest definierte, gleiche Ausgangsform gebracht werden. Die Bildgröße braucht nicht vereinheitlicht zu werden, da mit invarianten Momenten die Objekte unabhängig von ihrer Größe erkannt werden. Als Dateiformat wird das sehr einfach aufgebaute *PM-Format* gewählt (Abschnitt 3.1). Oft kommen dieselben Objekte in unterschiedlicher Farbdarstellung in Bildern vor. Deshalb wird mit Hilfe der am IITB entwickelten Software-Tools zur Bildverarbeitung eine Bildaufbereitung in zwei Schritten durchgeführt: Zunächst wird ein Gradientenbildes berechnet (Abschnitt 3.2), und danach werden daraus die Konturen im Bild extrahiert (Abschnitt 3.3).

3.1 Das PM-Dateiformat

Das PM-Format ist ein einfaches Dateiformat für Bilddaten, das an der University of Pennsylvania, USA, entwickelt wurde. Es besteht aus einem 28 Byte langen Header, gefolgt von den Grauwerten jedes Bildpunktes. Im Anschluss daran folgt optional ein Kommentar beliebiger Länge. Das Format unterstützt mehrere Bänder (*bands*) und mehrere Ebenen (*plains*), wie sie bei dem Gradientenbild (siehe Abschnitt 3.2), bei Bildfolgen oder bei den drei Farbkanälen von RGB-Bildern verwendet werden.

Der Header setzt sich aus folgenden jeweils 4 Byte großen Bestandteilen zusammen:

- Eindeutige Codefolge als Kennung für eine Datei im PM-Format (*WEIV* oder *VIEW*)
- Anzahl der Ebenen (*plains*)
- Anzahl der Bildspalten (*columns*)
- Anzahl der Bildzeilen (*rows*)
- Anzahl der Bänder (*bands*)
- Datentyp der Grauwerte (*unsigned char*, *float* und andere)
- Länge des Kommentars

Jedes gängige Bildformat (tif, gif, bmp, jpg, ...) kann in das PM-Format konvertiert werden.

3 Bildaufbereitung

3.2 NAG-Bilder

Ein NAG-Bild ist eine Darstellung, bei der durch eine Normierte Ableitung der Gaußfunktion ein Gradientenbild berechnet wird. Die Gradientenbildung wird durch eine Filterung im Ortsbereich durchgeführt. Die Filterkerne zur einfachen partiellen Differentiation in Zeilen- und Spaltenrichtung

$$\mathbf{H}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \text{ und } \mathbf{H}_y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.1)$$

sind zur Kantenextraktion zu störungsanfällig [Hab95]. Stattdessen wird häufig der *Sobel-Operator* verwendet:

$$\mathbf{H}_x = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \text{ und } \mathbf{H}_y = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix}. \quad (3.2)$$

Bei dem NAG-Operator können mit dem Parameter σ der Gaußfunktion verschiedene Filterkerne generiert werden. Für $\sigma = 1$ gilt

$$\mathbf{H}_x = \begin{pmatrix} 0,004609 & 0,015923 & 0,024070 & 0,015923 & 0,004609 \\ 0,024258 & 0,083798 & 0,126676 & 0,083798 & 0,024258 \\ 0,041899 & 0,144739 & 0,218800 & 0,144739 & 0,041899 \\ 0,000000 & 0,000000 & 0,000000 & 0,000000 & 0,000000 \\ 0,041899 & 0,144739 & 0,218800 & 0,144739 & 0,041899 \\ 0,024258 & 0,083798 & 0,126676 & 0,083798 & 0,024258 \\ 0,004609 & 0,015923 & 0,024070 & 0,015923 & 0,004609 \end{pmatrix} \text{ und } \mathbf{H}_y = \mathbf{H}_x^T. \quad (3.3)$$

Die Verwendung des NAG-Operators hat gegenüber gewöhnlichen Filterkernen zur Gradientenbildung den Vorteil, dass auch mit geringen Kontrastwerten gearbeitet werden kann. Als IITB-internes Software-Tool liegt ein Programm vor, das diese Berechnung für ein Grauwerte-Bild im PM-Format durchführt, und das Ergebnis als Multi-Plane-Bild im PM-Format bereitstellt. Es besteht aus fünf Ebenen (umgekehrte Aufzählungsreihenfolge):

- Originalbild (*ImagePlane*)
- Betrag des Gradienten (*MagGradPlane*)
- Richtungen des Gradienten (*DirGradPlane*)
- Maxima des Gradientenbetrags (*MagMaxPlane*)
- Richtungen des Gradienten in den Maxima (*DirMaxPlane*)

Abbildung 3.1 zeigt beispielhaft die fünf Ebenen eines NAG-Bildes vom Wappen des Landes Baden-Württemberg.

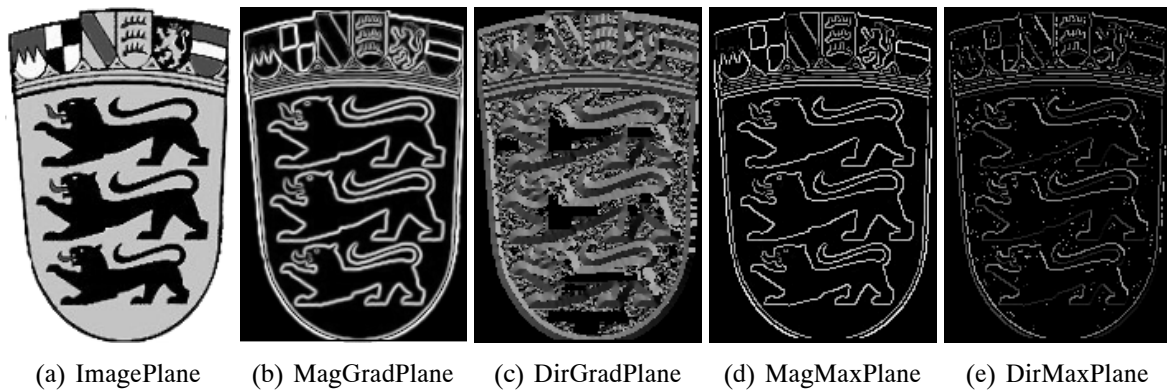


Abbildung 3.1: Ebenen eines NAG-Bildes

3.3 Konturpunktlisten

Um die Kanten aus einem Bild zu extrahieren, wird ein weiteres Software-Tool verwendet, das die Konturen aus einem NAG-Bild berechnet. Das Ergebnis liegt als sortierte Konturpunktliste (*KPL*) vor. Als Konturpunkte in diesem Sinne sind hier Bildpunkte des Betragsmaximabandes des NAG-Bildes gemeint, deren Gradientenbetrag eine feste Schwelle überschreitet. Untereinander benachbarte Konturpunkte werden in einer Konturpunktkette zusammengefasst und es werden nur Konturen in der Ergebnisdatei abgelegt, die eine vorgegebene Mindestlänge überschreiten.

Die Ergebnisdatei enthält als Kopfinformation die ID der Konturpunktliste, die Anzahl der abgelegten Konturpunkte, sowie die Kommandozeile, mit der das Programm gestartet wurde. Es folgt pro Zeile ein Konturpunkt mit folgender Beschreibung:

- Kennzeichnung der Zeile als Konturpunkt durch das Codewort `siob_kpl`¹
- x-Koordinate, y-Koordinate
- Typ (Kettenanfang, Kette, Kettenende, Knickpunkt)
- Winkel zum Nachfolger
- Grauwertbetrag

Die Konturpunktlinie ist direkt in ein logisches Binärbild umsetzbar (*KPL-Bild*). Abbildung 3.2 zeigt beispielhaft ein Originalbild und dessen logisches Binärbild der Konturpunktliste (Wappen des Landes Baden-Württemberg mit Wappentieren Hirsch und Greif).

¹Integration modularer neuronaler und wissensbasierter Komponenten zur *sichtgestützten Inspektion von Objekten in einer industriellen Umgebung* ist ein abgelaufenes Projekt am IITB, in dessen Rahmen das Tool zur Generierung von Konturpunktlisten erstellt wurde.

3 Bildaufbereitung

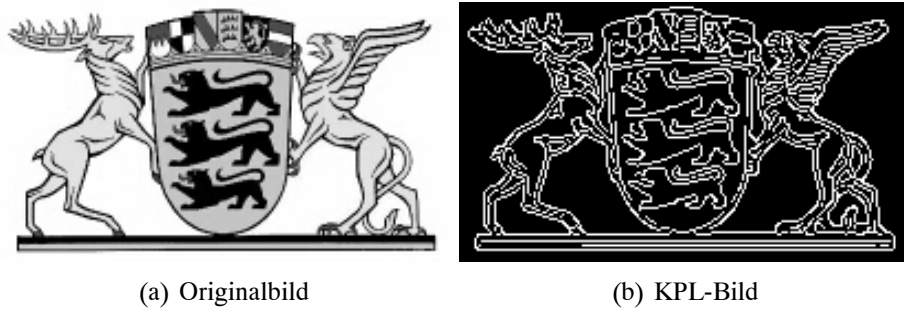


Abbildung 3.2: Darstellung einer Konturpunktliste

3.4 Segmentierte Konturpunktlisten

Am IITB existiert ein Software-Tool, das eine Konturpunktliste nach Abschnitt 3.3 in einzelne zusammenhängende Teilkonturen logisch zerlegt. Es werden damit jeweils benachbarte Bildpunkte, die als Konturpunkt erfasst sind, gruppiert und als Segment gekennzeichnet. Die Ergebnisdatei liegt im gleichen Format wie die Konturpunktliste vor. Sie enthält alle Konturpunkte in sortierter Reihenfolge mit einleitender Kommentarzeile zu Beginn eines neuen Segments. Segmentierte Konturpunktlisten werden in Unterabschnitt 4.4.1 zur Verbesserung der Klassifikationsergebnisse herangezogen.

Abbildung 3.3 zeigt alle Segmente, die in der Kontur von Abbildung 3.3(u) enthalten sind.

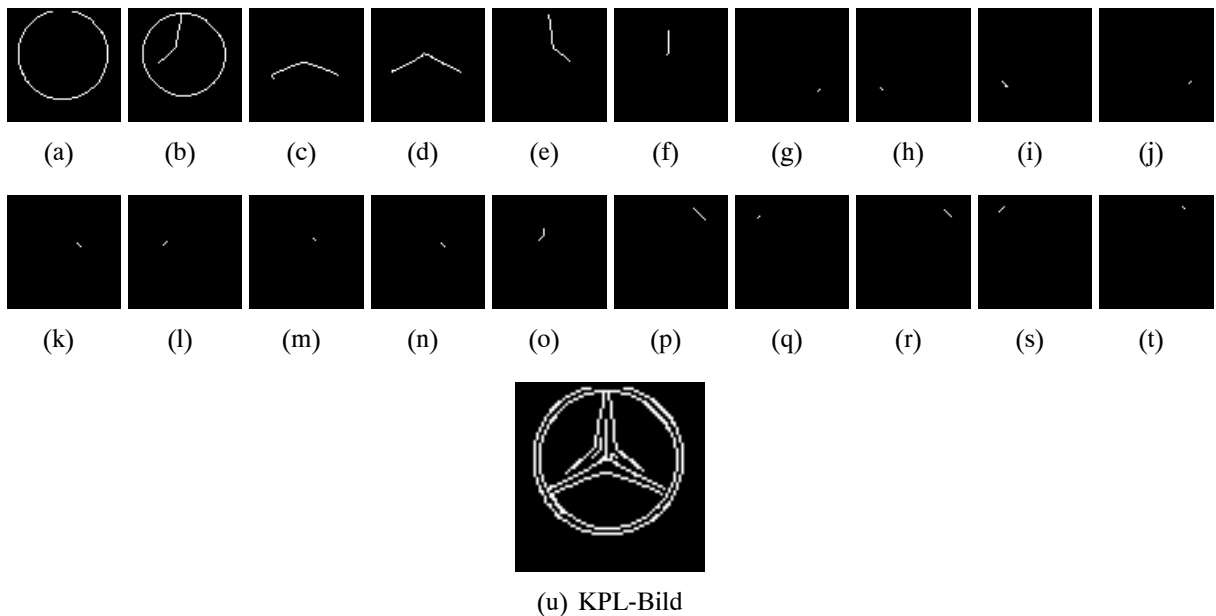


Abbildung 3.3: Darstellung der einzelnen Segmentbilder

4 Ergebnisse

4.1 Bildmaterial

Für die Untersuchung von invarianten Momenten für die Klassifikation von Objekten in Bildern ist umfangreiches Bildmaterial nötig. Nur wenn eine große Anzahl von Bildern vorhanden ist, kann eine Aussage über die Eignung der Momente mit statistischem Gewicht getroffen werden. Da für die vorliegende Arbeit keine firmeninterne oder kommerzielle Datenbanken zur Verfügung standen, besteht das verwendete Bildmaterial durchgängig aus frei zugänglichen Bildern im Internet.

Als zu klassifizierende Objekte stehen verschiedene Gruppen zunächst im Vordergrund:

- Firmen-Logos: eine Vielzahl von Aufgaben benötigt die Information über das Auftreten von Firmen-Logos, z.B. im Marketing
- Warenzeichen: vor allem geschützte Warenzeichen stellen eine relevante Zielgruppe dar; einer illegalen Verwendung in Bildern könnte entgegen gewirkt werden
- Flaggen und Wappen: aus verschiedenen Gründen spielen Bilder mit repräsentativem Inhalt eines Staates oder Landes eine wichtige Rolle
- Symbole, Embleme: viele Organisationen und Vereinigungen (insbesondere verbotene oder kriminelle) lassen sich an Hand ihres Symbols oder Emblems identifizieren.

Nicht untersucht werden sollen dagegen Bilder von Personen, Tieren und Landschaften, Gesichtern oder Fingerabdrücken, da es dafür andere, angepasste Verfahren zur Klassifikation gibt.

Aus dieser Zusammenstellung relevanter Gruppen werden zunächst fünf Objekte exemplarisch ausgewählt und im folgenden Abschnitt für die Klassifikation mit invarianten Momenten verwendet: der Mercedesstern als Firmen-Logo der DaimlerChrysler AG, das Warenzeichen der Coca-Cola Company, der Bundesadler der Bundesrepublik Deutschland, die Flagge der Volksrepublik China und das Emblem der Vereinten Nationen (UNO). Abbildung 4.1 zeigt jeweils eine Realisierung dieser Klassen. Diese Bilder werden im Folgenden als Referenzbilder bezeichnet.

Verteilt auf die einzelnen Objekte wurde insgesamt eine Datenbank erstellt, die 154 Bilder umfasst. Eine Schwierigkeit bei der Zusammenstellung des Bildmaterials aus dem Internet besteht allerdings darin, dass häufig dasselbe Bild in identischer Form an vielen Stellen zu finden ist. In

4 Ergebnisse

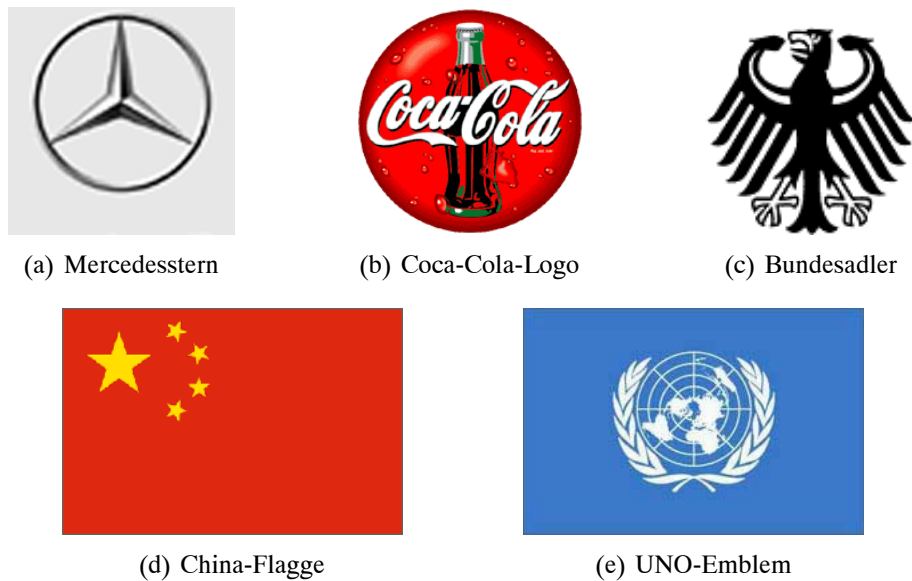


Abbildung 4.1: Objekte in den Bildern der Datenbank

der vorliegenden Datenbank wurde darauf geachtet, dass kein Bild doppelt oder mehrfach auftritt. Die zusammengestellten Bilder enthalten zum größten Teil neben dem Objekt der Klasse keine weitere relevante Bildinformation. Auf diese Beschränkung sowie Ansätze zur Aufhebung wird im Abschnitt 4.4 mit mehreren Verfahren näher eingegangen.

4.2 Vergleich der invarianten Momente

In einem ersten Schritt werden nun die Bilder der Datenbank als Konturbilder nach Kapitel 3 aufbereitet. Für diese Konturbilder werden dann die invarianten Zernike-Momente berechnet. Für eine erfolgreiche Klassifikation müssen die invarianten Momente folgende Eigenschaften erfüllen:

- Sie müssen auch in der praktischen Implementierung bei einer Transformation des Bildes nach Gl. (2.27) gleich bleiben.
- Sie sollten bei unterschiedlichen Realisierungen gleicher Objekte möglichst ähnliche Werte besitzen.
- Sie sollten bei unterschiedlichen Objekten deutlich unterscheidbare Werte besitzen.

Alle invarianten Zernike-Momente haben gemeinsam, dass die Dynamik mit der Ordnung stark zunimmt. Die Beträge bleiben bis zu einer Ordnung 7 bei allen betrachteten Bildern immer kleiner als 10 und gehen für manche Vektorelemente sehr nahe an den Nullpunkt (Grenze der Rechengenauigkeit). Das ist durch die Potenzierung der Zernike-Momente mit Beträgen kleiner als

eins in (2.32) zu erklären. Große Beträge (Peaks) sind vor allem bei den Elementen des Merkmalvektors zu erwarten, die die primären invarianten Zernike-Momente enthalten. Das Vorzeichen ist bei den Vektorelementen mit zu beachten.

4.2.1 Invariante Momente bei der Transformation eines Bildes

Die invarianten Zernike-Momente erfüllen näherungsweise die Bedingungen der Invarianz bei den affinen Transformationen Skalierung, Translation, Rotation und Spiegelung. In Unterabschnitt 2.3.3 wurde dies bereits für ein Beispiel im Vergleich zu den invarianten Momenten nach Hu gezeigt. Auch bei den für den Klassifikator verwendeten Bildern nehmen die invarianten Zernike-Momente die gleichen Werte an, selbst wenn die Bilder transformiert werden.

Für das Bundesadler-Referenzbild und dessen transformierte Versionen in Abbildung 4.2 enthält Tabelle B.1 die invarianten Zernike-Momente, Abbildung 4.3 zeigt sie auch graphisch als Trajektorie der sequenziellen Werte $\{Z_i\}_{i=1..32}$. Der Abstand der entsprechenden Vektoren zum Referenzvektor des nicht transformierten Originalbildes fällt sehr gering aus (siehe Tabelle 4.1), sodass die invarianten Zernike-Momente tatsächlich auch in der praktischen Implementierung nahezu unveränderlich sind, wenn das Bild vergrößert, verkleinert, verschoben, gedreht, gekippt oder gespiegelt wird. Die Unterschiede in den einzelnen Vektorelementen sind vor allem durch Rechenungenauigkeiten zu erklären, wenn man bedenkt, dass für die Berechnung der invarianten Zernike-Momente die teilweise sehr kleinen Zahlenwerte der Zernike-Momente noch potenziert werden. Außerdem stehen bei dem verwendeten Bildmaterial keine kontinuierlichen Bildfunktionen zu Verfügung, sondern jeweils digitale Realisierungen, die gerade bei einer Skalierung nicht verlustfrei transformiert werden können.

Wenn nun Bilder an Hand der invarianten Momente klassifiziert werden sollen, so wird die Klassifikation bei den Bildern sehr einfach gelingen, die das Referenzbild unmittelbar oder in einer der angesprochenen transformierten Versionen enthalten. Die Voraussetzungen für die Verwendung von invarianten Momenten sind damit erfüllt.

Als nächster Schritt werden nun Bilder mit dem selben Inhalt betrachtet, die aber *nicht* durch affine Transformation aus einander hervorgehen.

	Verkleinert auf 75%	Vergrößert auf 120%	$(10, 10)^T$ Pixel verschoben	Gedreht um 5°	Gedreht um 90°	Horizontal gespiegelt
$d(\mathbf{Z}, \mathbf{Z}_{Original})$	0.0853	0.0433	0.0248	0.0200	0.0248	0.0143

Tabelle 4.1: Euklidischer Abstand der invarianten Zernike-Momente eines transformierten Bildes zum Original (Bundesadler)

4 Ergebnisse

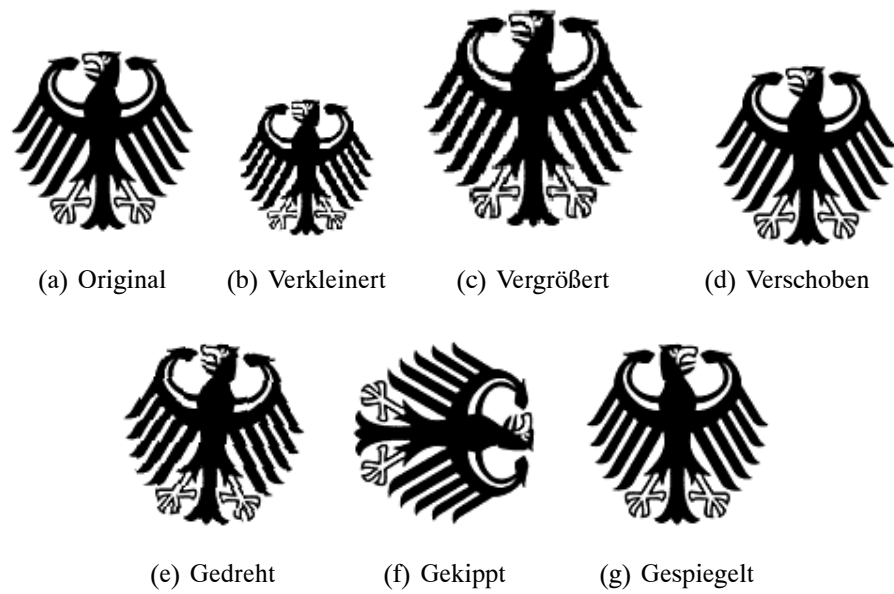


Abbildung 4.2: Original und transformierte Bilder mit dem Bundesadler

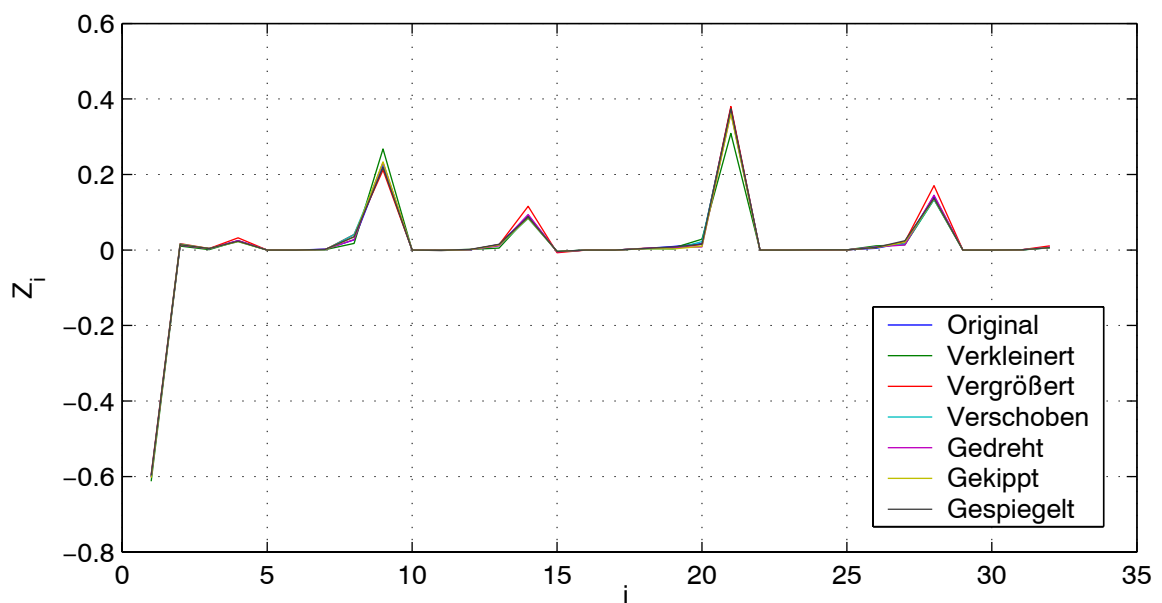


Abbildung 4.3: Invariante Zernike-Momente des Bundesadlers und seiner transformierten Versionen

4.2.2 Invariante Momente bei verschiedenen Realisierungen eines Objektes

Bei der vorliegenden Klassifikationsaufgabe ist es wichtig, dass die Merkmalvektoren von Objekten innerhalb einer Klasse große Ähnlichkeit aufweisen. Die Distanz dieser Vektoren hat einen entscheidenden Einfluss auf das Klassifikationsergebnis, da daraus direkt ein Entscheidungskriterium gebildet wird. Diese Anforderung an die Merkmalvektoren wird von den invarianten Zernike-Momenten erfüllt, solange das Objekt in den Bildern sich nicht zu sehr von dem Referenzobjekt unterscheidet.

Als Beispiel wird erneut der Bundesadler der Bundesrepublik Deutschland betrachtet. Abbildung 4.4 zeigt fünf verschiedene Realisierungen dieses Objektes. Tabelle B.2 enthält die zugehörigen invarianten Zernike-Momente $\mathbf{Z} = (Z_1, \dots, Z_{32})^T$, die in der grafischen Auswertung in Abbildung 4.5 als sequenzielle Werte aufgeführt sind. Die Merkmalvektoren zeigen eine große Ähnlichkeit zu einander, obwohl es sich um unterschiedliche Bilder handelt, und sind offensichtlich als Klassifikationsparameter zur Erkennung des Objekts gut geeignet. Das zeigt auch der Euklidische Abstand $d(\mathbf{Z}_i, \mathbf{Z}_j)$ in Tabelle 4.2.

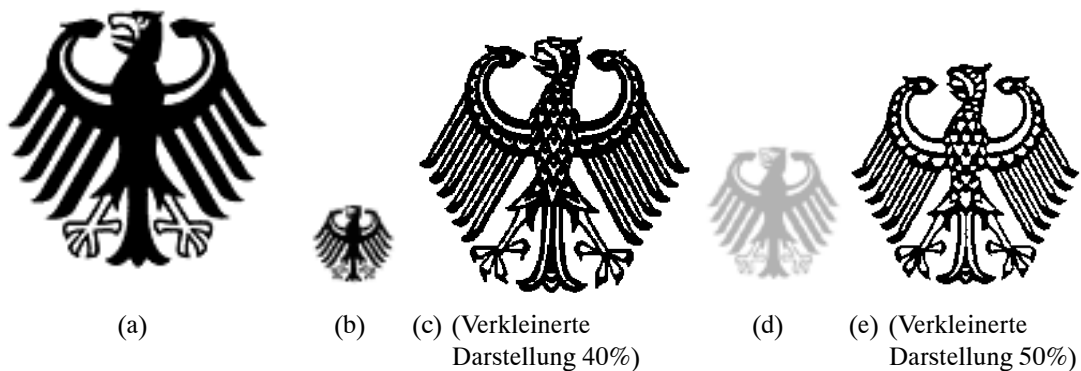


Abbildung 4.4: Verschiedene Bilder mit dem Objekt Bundesadler

	Bild 4.4(a)	Bild 4.4(b)	Bild 4.4(c)	Bild 4.4(d)	Bild 4.4(e)
Bild 4.4(a)		0.1395	0.1417	0.1502	0.1753
Bild 4.4(b)	0.1395		0.1654	0.2317	0.1814
Bild 4.4(c)	0.1417	0.1654		0.1298	0.0728
Bild 4.4(d)	0.1502	0.2317	0.1298		0.1441
Bild 4.4(e)	0.1753	0.1814	0.0728	0.1441	

Tabelle 4.2: Euklidischer Abstand zwischen invarianten Zernike-Momenten bei unterschiedlichen Bildern mit dem Bundesadler

4 Ergebnisse

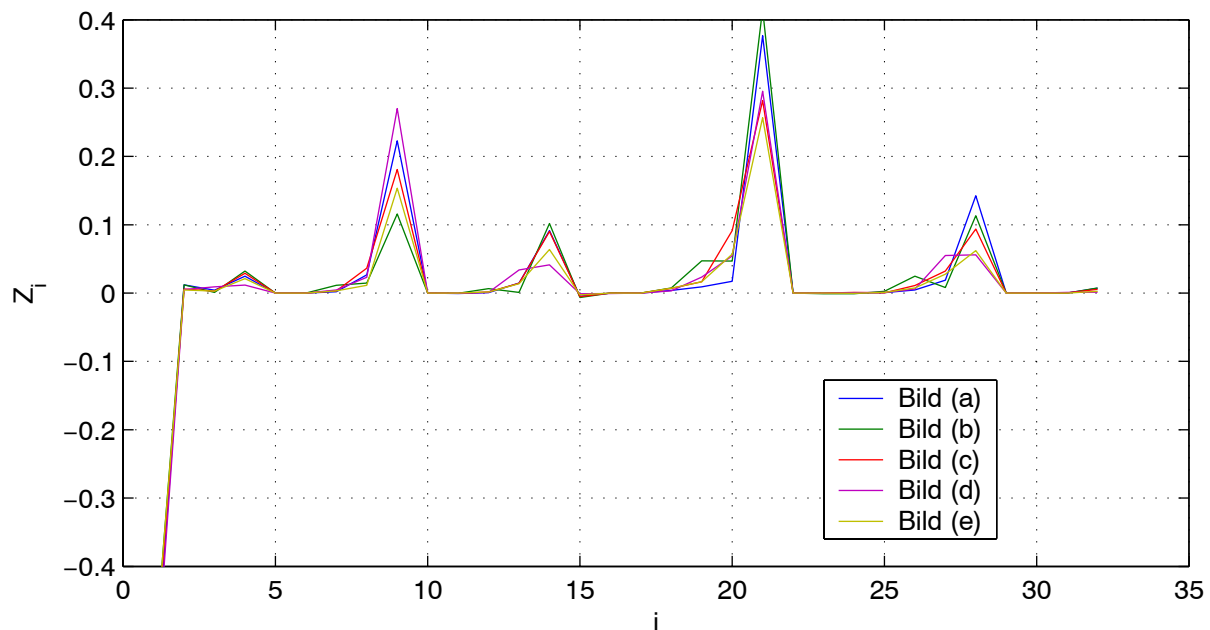


Abbildung 4.5: Invariante Zernike-Momente für die verschiedenen Bilder mit dem Bundesadler in Abbildung 4.4

4.2.3 Invariante Momente bei unterschiedlichen Objekten

Im Gegensatz zum vorhergehenden Abschnitt ist es bei Bildern mit verschiedenen Objekten wünschenswert, einen deutlichen Unterschied bei den Werten der Zernike-Momente zu erhalten. Diese Unterscheidbarkeit wird bei den betrachteten Objektklassen erreicht, da gerade die Elemente der invarianten Zernike-Momente, die große Beträgen aufweisen, verschiedene Werte annehmen oder ein anderes Vorzeichen aufweisen.

Im Folgenden werden die invarianten Zernike-Momente der fünf Referenzbilder in Abbildung 4.1 betrachtet (Tabelle B.4). Abbildung 4.6 zeigt deren sequenzielle Darstellung. Die fünf Kurven sind deutlich von einander zu unterscheiden.

Trotzdem ist offensichtlich, dass es an Hand der Merkmale vergleichsweise einfacher sein wird, ein Objekt zu erkennen, als es von anderen zu unterscheiden. Diese Erkenntnis wird sich auch in Abschnitt 4.3 als Klassifikationsergebnis bestätigen.

Tabelle 4.3 zeigt den Euklidischen Abstand zwischen den invarianten Zernike-Momenten der Referenzbilder. Mit Ausnahme des Abstandes zwischen den Objekten Bundesadler und Coca-Cola-Logo sind alle Abstände größer als 1, und 50% der Abstände nehmen Werte größer als 3 an. Im Vergleich zu den kleinen Abständen bei transformierten Bildern oder unterschiedlichen Bildern mit gleichen Objekten sind diese Abstände damit ausreichend groß für eine Unterscheidung. Vergleicht man diese Objekte mit Bildern, die eine völlig andere Struktur besitzen, z.B. Landschaftsaufnahmen oder Portraits, so erhält man noch größere Abstände zwischen den Merkmalvektoren.

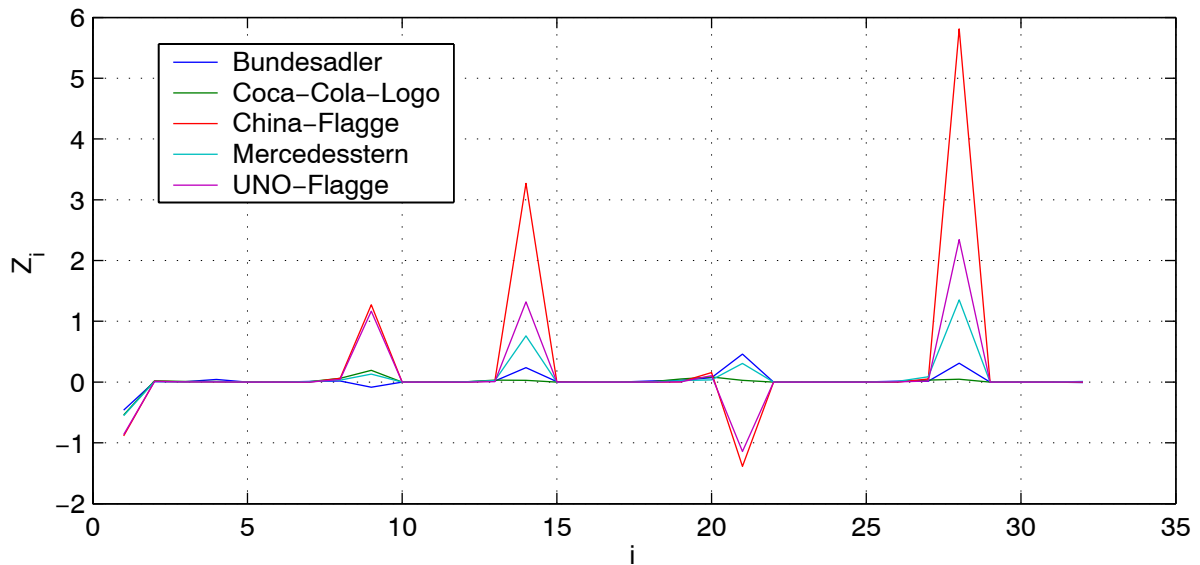


Abbildung 4.6: Invariante Zernike-Momente bei unterschiedlichen Objekten

	Bundesadler	Coca-Cola-Logo	China-Flagge	Mercedesstern	UNO-Flagge
Bundesadler		0.6239	6.7016	1.203	3.0984
Coca-Cola-Logo	0.6239		6.8594	1.5269	3.0607
China-Flagge	6.7016	6.8594		5.5213	3.9891
Mercedesstern	1.203	1.5269	5.5213		2.1376
UNO-Flagge	3.0984	3.0607	3.9891	2.1376	

Tabelle 4.3: Euklidischer Abstand zwischen den invarianten Zernike-Momenten der Referenzbilder

4.3 Klassifikation

Bei der Klassifikation von Bildern sind grundsätzlich zwei Denkansätze möglich, basierend auf unterschiedlichen Aufgabenstellungen:

- Eine Bilderdatenbank soll dahingehend untersucht werden, ob ein vorgegebenes Objekt darin enthalten ist. Für jedes Bild in der Datenbank muss die Entscheidung gefällt werden, ob es das Objekt beinhaltet oder nicht. Es handelt sich um eine binäre Detektionsaufgabe. Diese Aufgabenstellung ist für die Verwendung von invarianten Momenten die wichtigere und wird deshalb ausführlicher betrachtet.
- Eine Bilderdatenbank soll in mehrere Kategorien unterteilt werden, wobei jede Kategorie der Klasse eines Objektes entspricht. Es ist für jedes Bild die Entscheidung zu treffen,

4 Ergebnisse

welchem Objekt der Bildinhalt am ähnlichsten ist. In diesem Fall liegt eine multiple Detektionsaufgabe vor.

In beiden Fällen ist für die Klassifikation eines Testbildes eine Lernstichprobe als Referenzparameter und darüber hinaus ein Klassifikationskriterium nötig. Die Klassifikationsaufgabe kann natürlich insbesondere für den Sonderfall gelöst werden, dass nur ein einziges Bild untersucht werden soll. Für eine Aussage über die Güte der Verfahren ist es jedoch sinnvoller, eine Datenbank mit statistisch relevanter Mächtigkeit zu betrachten.

4.3.1 Binäre Detektion

Für die binäre Detektion gibt es bei unserer Anwendung die zwei Ereignisse

M_1 : Das gesuchte Objekt ist im Bild enthalten und
 M_2 : Das gesuchte Objekt ist im Bild nicht enthalten.

Der Klassifikator muss sich in Form einer Hypothese für eines dieser beiden Ereignisse entscheiden (siehe Abbildung 4.7):

H_1 : Akzeptanz, d.h. Hypothese dafür, dass M_1 eingetreten ist, oder
 H_2 : Rückweisung, d.h. Hypothese dafür, dass M_2 eingetreten ist.

Eine richtige Entscheidung wird getroffen, wenn bei Ereignis M_1 für H_1 und bei Ereignis M_2 für

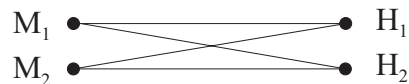


Abbildung 4.7: Binäre Detektion

H_2 entschieden wird. Falsche Entscheidungen treten ein, wenn bei Ereignis M_1 die Entscheidung für H_2 oder bei Ereignis M_2 für H_1 fällt. Für jede Hypothese existiert ein Entscheidungsraum \mathbf{R}_i , $i \in \{1, 2\}$. Dieser wird bei der binären Detektion durch eine Schwelle γ definiert. Ist der Abstand $d(\mathbf{Z}_0, \mathbf{Z}_i)$ zwischen dem Referenzvektor \mathbf{Z}_0 und dem Testvektor \mathbf{Z}_i eines Datenbank-Bildes \mathbf{S}_i , $1 \leq i \leq N$, kleiner als diese Schwelle, so wird für H_1 entschieden, andernfalls für H_2 :

$$d(\mathbf{Z}_0, \mathbf{Z}_i) \begin{cases} \leq \gamma \\ > \gamma \end{cases} \Rightarrow \begin{cases} H_1 \\ H_2 \end{cases} \quad (4.1)$$

In unserem Fall sind die Parameter die 32 Elemente der invarianten Zernike-Momente bis zur Ordnung 7 und \mathbf{R}_1 wird damit zu einer Hyperkugel in \mathbb{R}^{32} mit Radius γ .

Es wird nun zunächst die binäre Klassifikation für das Objekt Bundesadler mit $\gamma = 0,7$ betrachtet. Der Klassifikator wird mit 40 Bildern, die das Objekt enthalten (M_1), und 114 Bildern, die das Objekt nicht enthalten (M_2), getestet. Abbildung 4.8(a) zeigt nun das Ergebnis für die

Bilder, die das Objekt enthalten. 30 Bilder (75%) werden richtig klassifiziert, 10 Bilder (25%) werden zurückgewiesen, da für H_2 trotz M_1 entschieden wird. Dazu ist auch Abbildung 4.8(b) zu betrachten, die die Verhältnisse bei der Klassifikation der Bilder zeigt, die das Objekt nicht enthalten. 72 Bilder (63%) werden richtig zurückgewiesen, da der Abstand die Schranke überschreitet; bei 42 Bildern (37%) wurde falsch für H_1 entschieden.

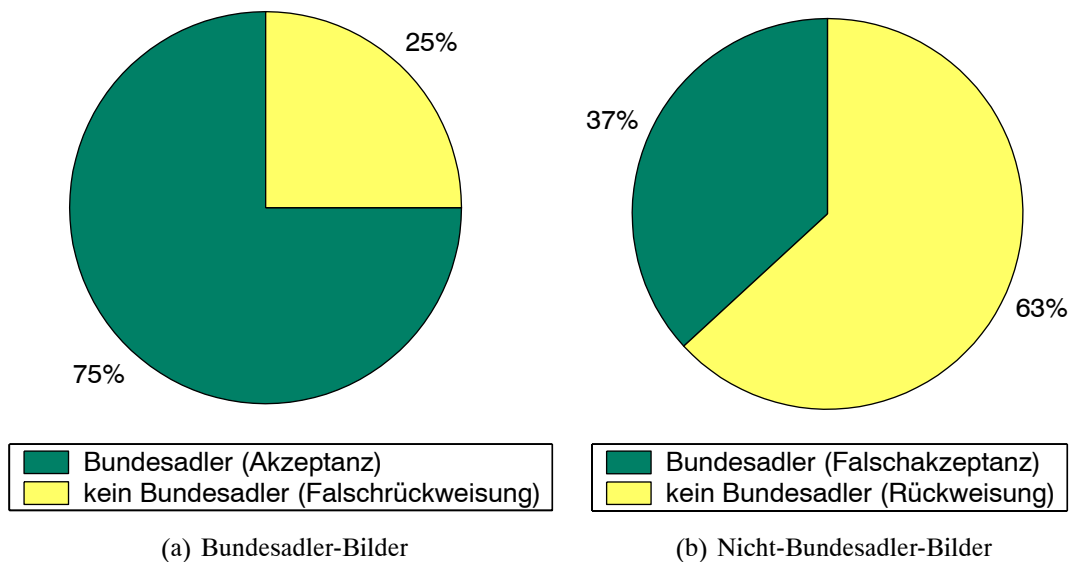


Abbildung 4.8: Bundesadler-Klassifikation bei $\gamma = 0,7$

In keinem Fall ist es möglich, gleichzeitig die Anzahl richtig klassifizierter Bilder, die das Objekt beinhalten, und die Anzahl richtig klassifizierter Bilder, die das Objekt nicht beinhalten, zu maximieren. Insofern ist bei dieser Klassifikation immer ein Kompromiss einzugehen. Der freie Parameter dieser Entscheidung ist die Schwelle γ .

Unter Verwendung der Notation für bedingte Wahrscheinlichkeiten lässt sich für die Akzeptanzanteile sagen, dass zwar bei kleiner Schwelle die Wahrscheinlichkeit $P(H_1|M_2)$ gering ist, aber damit verbunden auch die Wahrscheinlichkeit $P(H_1|M_1)$. Genauso wird man bei einer großen Schwelle eine große Wahrscheinlichkeit $P(H_1|M_1)$ erreichen, muss diese aber mit einer ebenfalls großen Wahrscheinlichkeit $P(H_1|M_2)$ erkaufen. Abbildung 4.9 zeigt diese Verhältnisse bei dem Bundesadler-Klassifikator für $0,1 < \gamma \leq 8,1$.

Dasselbe gilt natürlich in umgekehrter Weise für die richtige und falsche Rückweisung, denn es gilt $P(H_2|M_2) = 1 - P(H_1|M_2)$ und $P(H_2|M_1) = 1 - P(H_1|M_1)$. Eine hohe Rate an richtigen Rückweisungen von Bildern, die das Objekt nicht enthalten, muss mit einer hohen Rate an falschen Rückweisungen von Bildern, die das Objekt enthalten, erkaufte werden (siehe Abbildung 4.10). Für steigendes γ wird man weniger falsche Rückweisungen $P(H_2|M_1)$ erhalten, aber auch eine kleinere Rate an richtigen Rückweisungen $P(H_2|M_2)$, da der Entscheidungsraum R_1 vergrößert wird.

Entscheidend für einen guten Klassifikator ist nun, dass der Anteil der richtigen Akzeptanz für

4 Ergebnisse

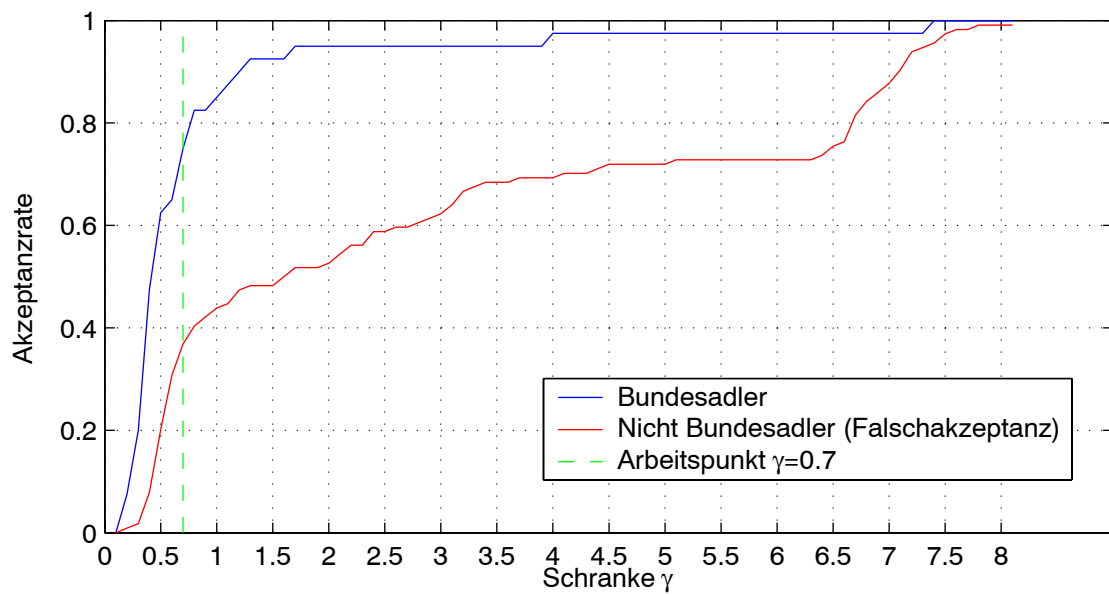


Abbildung 4.9: Akzeptanzrate bei der Bundesadler-Klassifikation

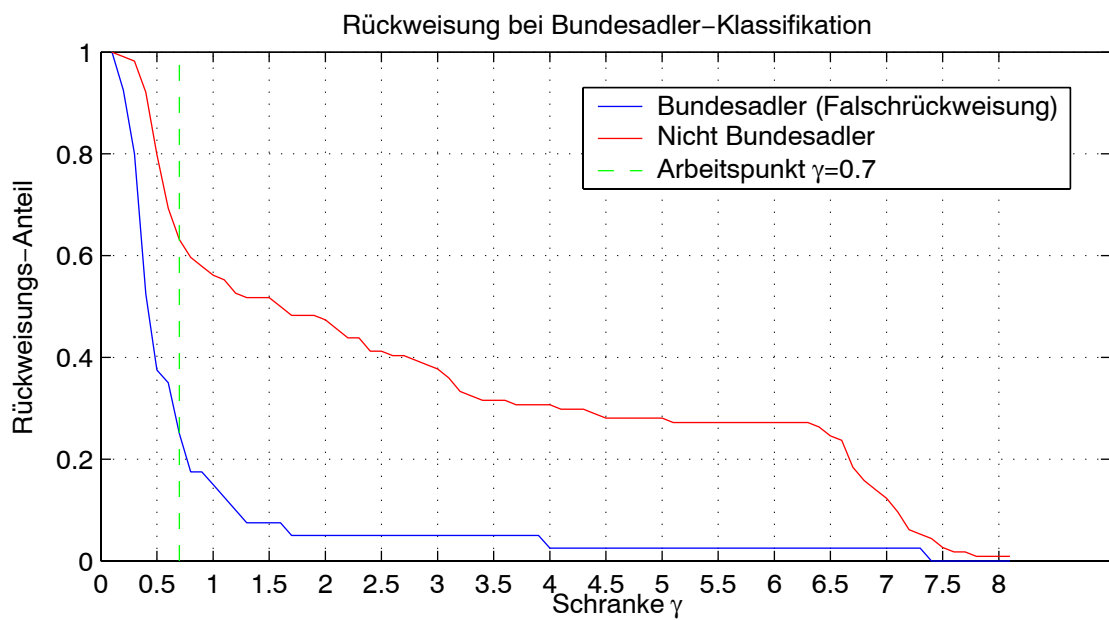


Abbildung 4.10: Rückweisungsrate bei der Bundesadler-Klassifikation

alle Schwellen γ über dem der Falschakzeptanz liegt und dass auch für alle γ der Anteil der richtigen Rückweisung größer ist als der der falschen Rückweisung. Dieses Kriterium wird für den Bundesadler-Klassifikator erfüllt.

Verwendet man nun die Begriffe *Entdeckungswahrscheinlichkeit* $P_E = P(H_1|M_1)$ für eine richtige Akzeptanz und *Falschalarmwahrscheinlichkeit* $P_F = P(H_1|M_2)$ für eine falsche Akzeptanz nach [Kro96], so lässt sich die Charakteristik des Klassifikators als Funktion $P_E(P_F)$ in einer Kurve darstellen. Diese Kurve wird *Empfängerarbeitscharakteristik* genannt (ROC: receiver operating characteristic) und zeigt das Detektionsverhalten des Klassifikators für alle γ .

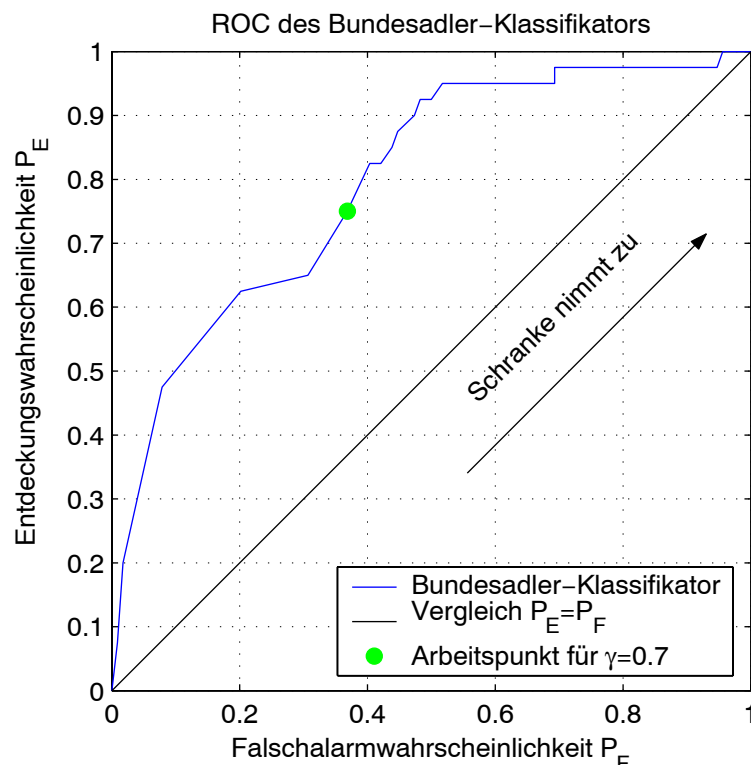


Abbildung 4.11: ROC des Bundesadler-Klassifikators

Abbildung 4.11 enthält diese Kurve für den Bundesadler-Klassifikator. Für steigende Werte der Schwelle γ wird die Kurve vom Punkt $(P_F = 0, P_E = 0)$ in Richtung $(P_F = 1, P_E = 1)$ durchlaufen. Die Funktionswerte sind monoton steigend und liegen oberhalb der Geraden $P_E = P_F$. Diese Empfängerarbeitscharakteristik ist somit deutlich besser als eine „blinde“, gleichverteilte Entscheidung für H_1 und H_2 unter der Annahme gleicher Auftrittswahrscheinlichkeiten von M_1 und M_2 .

Je nach Anwendungsfall kann nun auf dieser Kurve ein Arbeitspunkt gewählt werden. Dieser sollte möglichst nahe an dem Punkt $(P_F = 0, P_E = 1)$ liegen. In diesem Punkt wäre eine völlig fehlerfreie Detektion möglich. Der eingetragene Arbeitspunkt für $\gamma = 0,7$ enthält gerade die

4 Ergebnisse

Verhältnisse aus Abbildung 4.8 mit einer Entdeckungsrate von 75% und einer Falschalarmrate von 37%.

Als Gütekriterium kann auch der Abstand des Arbeitspunktes zu der Geraden $P_E = P_F$ gewählt werden. In einem optimalen Arbeitspunkt nimmt der Abstand dann gerade seinen maximalen Wert an. In manchen Anwendungen, z.B. der Bildsuche in Datenbanken, wird auch eine feste Entdeckungsrate P_E gefordert sein und die Falschalarmrate P_F spielt eine untergeordnete Rolle. In jedem Fall ist es möglich, nach dem Trainieren des Klassifikators die entsprechende Empfängerarbeitscharakteristik zu bestimmen und daraus einen sinnvollen Wert für γ zu entnehmen.

Eine Auswertung in gleicher Art und Weise wurde auch für die Klassifikatoren der Objekte Coca-Cola-Logo, China-Flagge und Mercedesstern erstellt. Zu dem Objekt UNO-Emblem konnten nicht genügend Bilder für eine ROC mit statistischem Gewicht zusammen getragen werden. Die Empfängercharakteristika zeigt Abbildung 4.13. Alle Kurven liegen oberhalb der Geraden $P_E = P_F$ und stellen damit erfolgreiche Klassifikatoren dar im Sinne der Verbesserung gegenüber „blinder“ Entscheidung. Ein optimaler Arbeitspunkt befindet sich auch hier jeweils an der Stelle des größten Abstandes zu dieser Geraden.

Der beste der betrachteten Klassifikatoren ist der China-Flaggen-Klassifikator. Er hebt sich deutlich von den anderen ab und verzeichnet bei Falschalarmwahrscheinlichkeiten von unter 10% schon Entdeckungswahrscheinlichkeiten oberhalb der 90%-Marke. Dieses sehr gute Ergebnis ist mit dem einfachen geometrischen Aufbau des Objektes zu erklären. Das Konturbild des Referenzobjektes für diese Klasse zeigt Abbildung 4.12. Seine Bildkomposition besteht aus wenigen, aber relativ langen Konturen und keine pixeligen Punktstörungen.

Als schlechtester Klassifikator ist in diesem Vergleich derjenige für Objekte der Klasse Mercedesstern zu sehen. Das liegt vor allem daran, dass es wenige Bilder in der Datenbank gibt, die eine sehr große Ähnlichkeit mit dem Referenzbild besitzen, dafür aber viele merklich unterschiedliche Realisierungen dieses Objektes.



Abbildung 4.12: Konturbild der China-Flagge

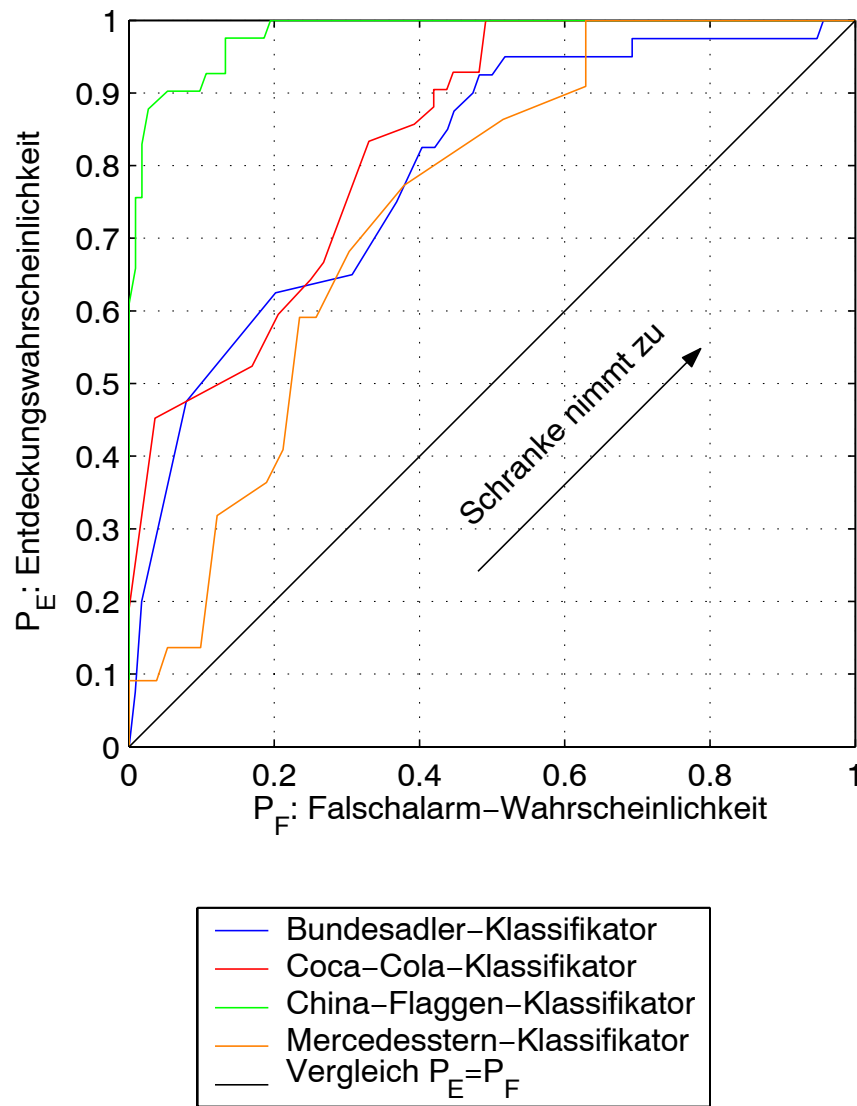


Abbildung 4.13: ROCs aller betrachteter Klassifikatoren

4 Ergebnisse

4.3.2 Multiple Detektion

Die zweite der angesprochenen Aufgabenstellungen betrifft die multiple Detektion. Es stehen dem Klassifikator damit mehr als zwei mögliche Ausgänge zur Verfügung. Will man K Objekt-Klassen unterscheiden, so bezeichnet man das Ereignis, dass in dem Bild ein Objekt der Klasse k , $1 \leq k \leq K$, enthalten ist, als M_k . Die zugehörige Hypothese des Entscheiders wird H_k genannt (siehe Abbildung 4.14(a)). Für diese Klassifikationsaufgabe ist es auch denkbar, eine Rückweisungsklasse mit der Hypothese

H_0 : Es ist keines der Objekte im Bild enthalten

zu definieren (Abbildung 4.14(b)).

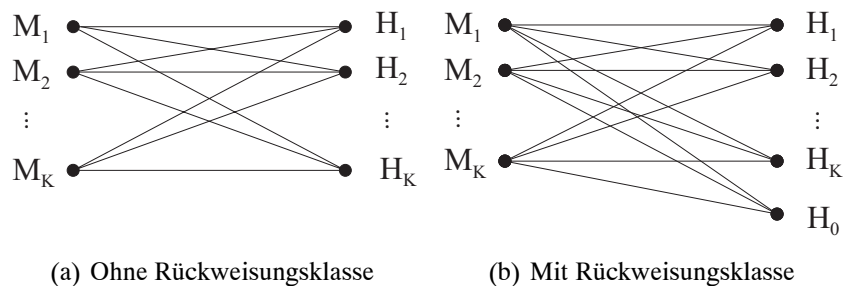


Abbildung 4.14: Multiple Detektion

Für die multiple Detektion wird ein *NN-Klassifikator* (NN: nearest neighbour) verwendet. Damit werden zunächst für alle Referenzbilder die invarianten Zernike-Momente \mathbf{Z}_{0k} , $1 \leq k \leq K$, berechnet und als Merkmalvektoren abgelegt, zusammen mit der Klassenzugehörigkeit. Dieser Vorgang entspricht dem *Trainieren* des Klassifikators an Hand einer Lernstichprobe. Dann kann für jedes Testbild \mathbf{S}_i , $1 \leq i \leq N$, ebenfalls der Merkmalvektor \mathbf{Z}_i berechnet und mit Hilfe eines Abstandsmaßes der nächste Nachbar bestimmt werden. Als nächster Nachbar des Merkmalvektors \mathbf{Z}_i wird der Merkmalvektor \mathbf{Z}_{0k} bezeichnet, der den geringsten Abstand $d(\mathbf{Z}_{0k}, \mathbf{Z}_i)$ aufweist. Die Klassenzugehörigkeit dieses Nachbarn entscheidet dann über die Klassenzuordnung des Testbildes:

$$k = \arg \min_{1 \leq k \leq K} d(\mathbf{Z}_{0k}, \mathbf{Z}_i) \Rightarrow H_k. \quad (4.2)$$

Falls für eine vorgegebene Rückweisungsschranke η gilt, dass $d(\mathbf{Z}_{0k}, \mathbf{Z}_i) > \eta \forall k$, so kann das Testbild der Rückweisungsklasse zugewiesen werden.

Abbildung 4.15 zeigt das Klassifikationsergebnis ohne Rückweisungsklasse für die gesamte Datenbank mit einem Umfang von 154 Testbildern. Bei dieser multiplen Detektion wurden 90 Bilder (58%) den richtigen Klassen zugeordnet, 64 Bilder (42%) wurden falsch klassifiziert. Bei den falsch klassifizierten Bildern zeigen sich ähnliche Verhältnisse für alle Objekte. Die China-Flaggen erreichen mit 6% den kleinsten Anteil und bringen den besten Beitrag zur Gesamtklassifikation. Vergleicht man diese Ergebnisse wieder mit der „blinden“ Entscheidung basierend

auf der Annahme gleicher Auftrittswahrscheinlichkeiten der einzelnen Klassen, so ist ein deutlicher Gewinn zu erkennen. In diesem Fall wäre bei den vier verwendeten Klassen mit einem Mittelwert von 25% bei den richtigen Entscheidungen zu rechnen bei der Annahme gleicher *A-priori*-Wahrscheinlichkeiten. Damit ist mit dem vorgestellten Verfahren eine Verbesserung um 132% erreicht worden.

Trotzdem besteht mit einer richtigen Entscheidungsrate von 60% ein großer Unterschied zu den Ergebnissen, die bei einer binären Detektion erreicht werden. Das ist vor allem damit zu erklären, dass die invarianten Zernike-Momente verschiedener Objekte trotz deutlicher Unterscheidbarkeit eine gewisse ähnliche Struktur besitzen, zumindest was die Größenordnung der einzelnen Elemente betrifft (siehe Unterabschnitt 4.2.3). Deshalb sind invariante Momente für die multiple Detektionsaufgabe nicht so gut geeignet wie für die binäre Detektion.

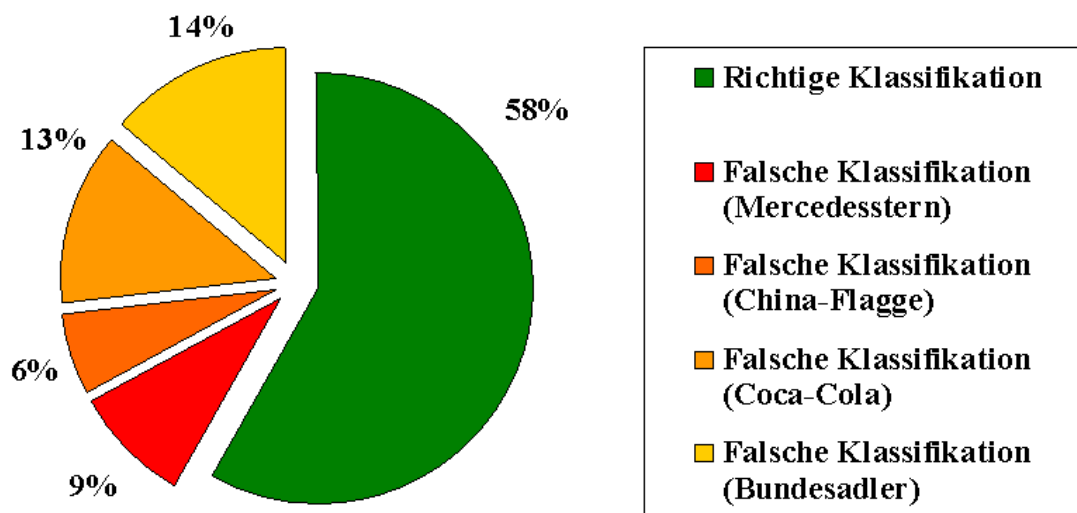


Abbildung 4.15: NN-Klassifikation aller Bilder

4.3.3 Mehrreferenzen-Klassifikation

Für die gleiche Datenbank wurde auch ein K NN-Klassifikator mit $K = 5$ verwendet und die Anzahl der Referenzbilder für jede Klasse auf 5 erhöht. Dabei werden statt einem nächsten Nachbarn die K nächsten Nachbarn berücksichtigt. Das Testbild wird dann derjenigen Klasse zugeordnet, der die Mehrzahl der K nächsten Nachbarn angehören. Im Idealfall sind alle K nächsten Nachbarn des Testbildes Elemente der selben Klasse.

4 Ergebnisse

In Abbildung 4.16 wird das Ergebnis einer Mehrreferenzen-Klassifikation für den Bundesadler-Klassifikator mit der Empfängerarbeitscharakteristik dargestellt. Die Kurve für einen Mehrreferenzen-Klassifikator liegt zum größten Teil oberhalb der Vergleichskurve für den Einreferenzen-Klassifikator und stellt damit eine weitere Verbesserung des Verfahrens dar. Das Ausmaß der Verbesserung ist stark vom verwendeten Bildmaterial abhängig. So ist es auch möglich, dass sich das Ergebnis leicht verschlechtert, wenn die verwendeten Referenzbilder einer Klasse zu große Unterschiede aufweisen und deshalb im Merkmalraum weit auseinander liegen. Dieses Verhalten kann dadurch weitgehend kompensiert werden, dass man die Zahl der Referenzbilder wesentlich erhöht und für jede Klasse damit mehr Vektoren im Merkmalraum hat, die als nächste Nachbarn des Testbildes in Betracht kommen. Dann wird der Einsatz eines Mehrreferenzen-Klassifikators in den meisten Fällen eine Verbesserung des Verfahrens bringen.

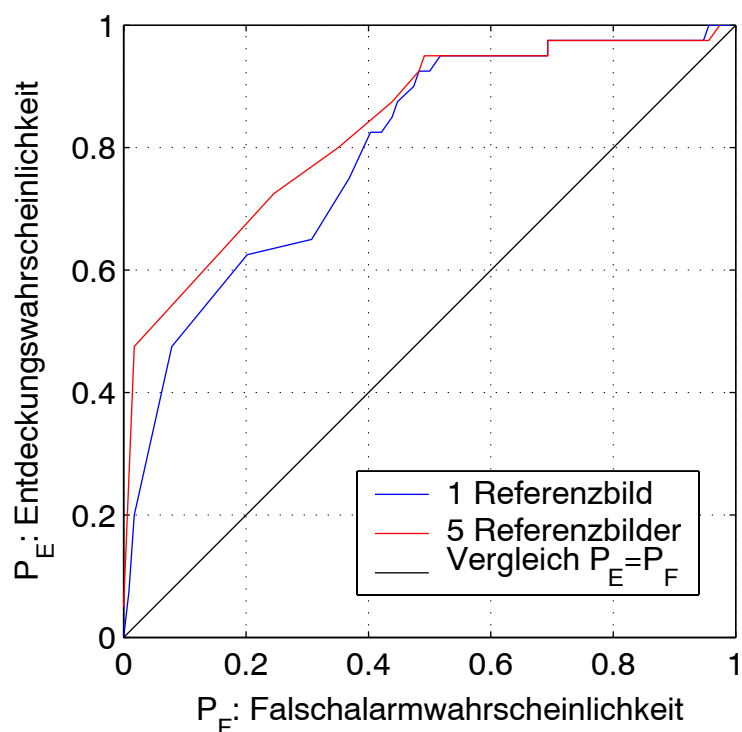


Abbildung 4.16: ROC des Mehrreferenzen-Klassifikators

4.3.4 Dynamikbetrachtungen

Der Wertebereich in den einzelnen Elementen der Merkmalvektoren ist recht groß mit Werten im Bereich um 10 bis zu sehr kleinen Werten in der Nähe der Rechengenauigkeit. Aus diesem Grund werden an dieser Stelle zwei mögliche Verfahren betrachtet, die diese Dynamik bei der Klassifikation berücksichtigt.

Zunächst ist es möglich, ein gewichtetes Abstandsmaß einzuführen. Verwendet man M Referenzvektoren $\{\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(M)}\}$, die jeweils ν invariante Zernike-Momente als Vektorelemente beinhalten, $\mathbf{Z}^{(m)} = (Z_1^{(m)}, \dots, Z_\nu^{(m)})^T$, $1 \leq m \leq M$, so lässt sich für jeden Testvektor $\mathbf{Z} = (Z_1, \dots, Z_\nu)^T$ der gewichtete Euklidische Abstand nach der Gleichung

$$d(\mathbf{Z}, \mathbf{Z}^{(m)}) = \sqrt{\sum_{i=1}^{\nu} \rho_i (Z_i - Z_i^{(m)})^2} \quad (4.3)$$

angeben [MR98]. Darin wird für die Gewichtungsfaktoren ρ_i ein Varianzmaß verwendet:

$$\rho_i = \frac{M}{\sum_{m=1}^M (Z_i^{(m)} - \bar{Z}_i)^2} \text{ mit } \bar{Z}_i = \frac{\sum_{m=1}^M Z_i^{(m)}}{M}. \quad (4.4)$$

Die Gewichtungsfaktoren basierend auf 5 Referenzvektoren der Bundesadler-Klasse zeigt Abbildung 4.17. Bei der Klassifikation mit diesem modifizierten Abstandsmaß stellt sich jedoch heraus, dass nun zwar die Dynamik der einzelnen Summanden reduziert wurde, dafür aber der Wertebereich des Abstandes d wesentlich größer ist. Damit variiert also auch der Wertebereich der Schranke γ wesentlich stärker als bei der Verwendung des Euklidischen Abstandes. Die Empfängercharakteristik ist in Abbildung 4.18 enthalten und zeigt, dass die gewichtete Abstandsbildung nicht in jedem Bereich einen Vorteil bringt. Das liegt vor allem an den teilweise sehr großen Gewichtungsfaktoren ρ_i , die durch die Multiplikation mit sehr kleinen Werten der invarianten Zernike-Momente zu wesentlich ungenaueren Ergebnissen führen können.

Ein weiterer Ansatz reduziert den Dynamikbereich, indem die invarianten Zernike-Momente aus normierten Zernike-Momenten berechnet werden [BSA91]. Darin werden die Momente auf Momente niederer Ordnung bezogen:

$$\hat{A}_{nl} = \begin{cases} A_{nl}/A_{n-2,l}, & \text{falls } A_{n-2,l} \neq 0 \text{ und } l < p \\ A_{nl}, & \text{falls } A_{n-2,l} = 0 \text{ oder } l = p \end{cases} \quad (4.5)$$

Damit kann allerdings nur bei einzelnen Elementen die Größenordnung angepasst werden, nicht bei allen. Abbildung 4.18 zeigt schließlich den Vergleich.

4 Ergebnisse

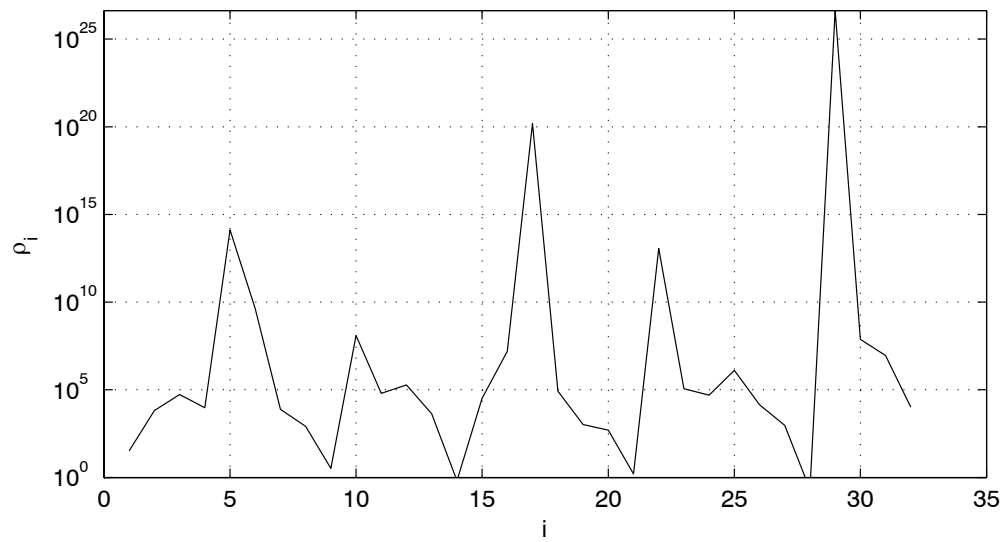


Abbildung 4.17: Gewichtungsfaktoren für ein gewichtetes Abstandsmaß

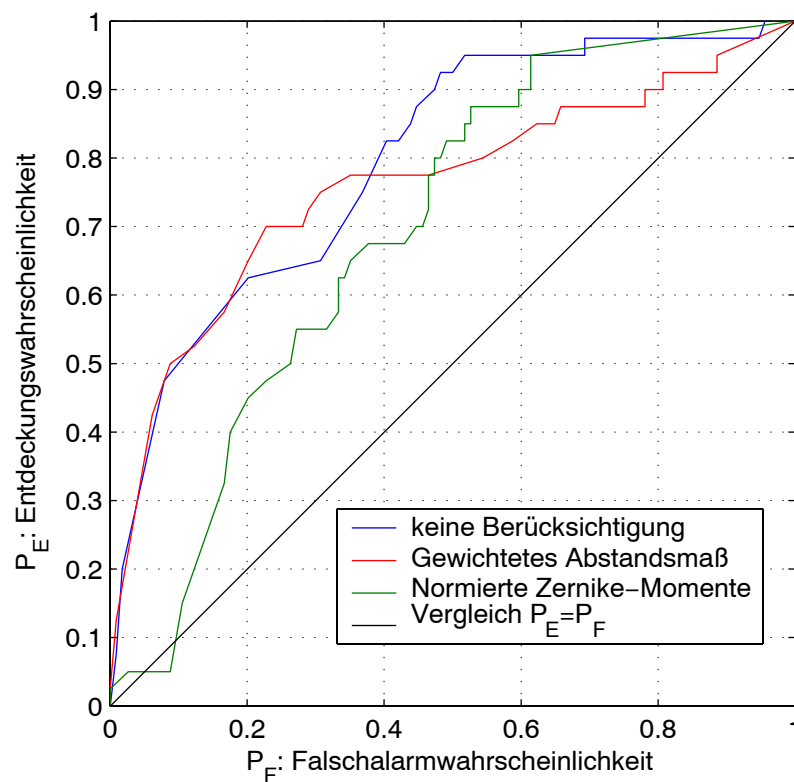


Abbildung 4.18: ROC des Bundesadler-Klassifikators bei verschiedenen Dynamikberücksichtigungen

4.4 Erweiterungen des Verfahrens

4.4.1 Segmentierung

In Abschnitt 3.4 wurde bereits auf die Möglichkeit eingegangen, Konturpunktlisten in Segmente zu zerlegen. Dies bietet nun den Vorteil, einzelne Segmente aus dem Bild zu entfernen, um eine bessere Klassifikation zu erreichen. Es ist auffällig, dass jede Konturpunktliste eine große Anzahl an Segmenten beinhaltet, die nur wenige Pixel groß sind. Diese kleinen Segmente können als Punktstörungen betrachtet werden. Entfernt man alle kleinen Segmente aus dem Konturbild, so zeigt sich ein besseres Klassifikationsergebnis.

Den Hintergrund zu dieser Überlegung bietet die Betrachtung zur Rauschempfindlichkeit der verschiedenen Momente in [TC88]. Es wird darin gezeigt, dass die Rauschempfindlichkeit von Zernike-Momenten mit der Ordnung steigt und über derjenigen von geometrischen oder komplexen Momenten liegt. In [Hew96] wird insbesondere auf die Verschiebung des Bildschwerpunktes eingegangen, die durch einzelne Punktstörungen im Randbereich entstehen.

Abbildung 4.19 zeigt absteigend sortiert die Segmentlängen aller Segmente in den sechs verschiedenen segmentierten KPL-Bildern von Abbildung 4.20. Die Darstellung der Segmentlängen ist nach oben auf 100 Pixel begrenzt, um Bilder unterschiedlicher Größe vergleichbar zu machen. Unabhängig vom Objekt im Bild ergibt sich eine charakteristische Verteilung der Segmentlängen in der Kontur: mehr als die Hälfte aller Segmente besitzen eine Segmentlänge von 15 Pixel oder weniger, bei Abbildung 4.20(e) sind sogar 84% aller Segmente kleiner als 10 Pixel. Außerdem gibt es in den betrachteten Segmentlisten höchstens einen Anteil von 31%, für den die Segmentlänge 50 Pixel überschreitet (8% bei Abbildung 4.20(e)). Deshalb kann durch die Wahl einer sinnvollen Untergrenze ϵ erreicht werden, dass irrelevante Information in der Erscheinung von Punktstörungen aus dem Bild entfernt werden. Bei den betrachteten Bildern liegt diese Schranke in dem Intervall [10 Pixel, 50 Pixel], im allgemeinen Fall kann sie auch über die Stelle der globalen maximalen Krümmung der Segmentlängenverteilung bestimmt werden.

Alle Segmente, deren Länge diese Schranke ϵ unterschreitet, werden aus der Segmentliste gelöscht und damit nicht in das Segmentbild übernommen, aus dem die invarianten Zernike-Momente als Elemente des Merkmalvektors berechnet werden.

Damit wird in die Bildaufbereitung ein weiterer Schritt eingefügt: Aus der Segmentliste muss ein Segmentbild erstellt werden, das die nicht entfernten Segmente akkumuliert. Dies kann bei Binärbildern einfach durch bitweise Oder-Verknüpfung erreicht werden.

Um beispielhaft zu zeigen, dass sich durch die Elimination von Punktstörungen tatsächlich eine Verbesserung der Klassifikation ergibt, werden die invarianten Zernike-Momente eines sehr stark gestörten Bildes mit einem Mercedesstern betrachtet. Abbildung 4.21 zeigt die akkumulierten Segmentbilder für unterschiedliche minimale Segmentlängen ϵ , und Abbildung 4.22 enthält sequenziell dargestellt die Elemente des jeweils zugehörigen Merkmalvektors $\mathbf{Z}(\epsilon)$. Für $\epsilon = 1$ sind alle Konturen im Bild enthalten und die invarianten Zernike-Momente zeigen nicht sehr deutlich die charakteristischen Peaks für den Mercedesstern. Für $\epsilon = 20$ wird das beste Ergebnis erreicht,

4 Ergebnisse

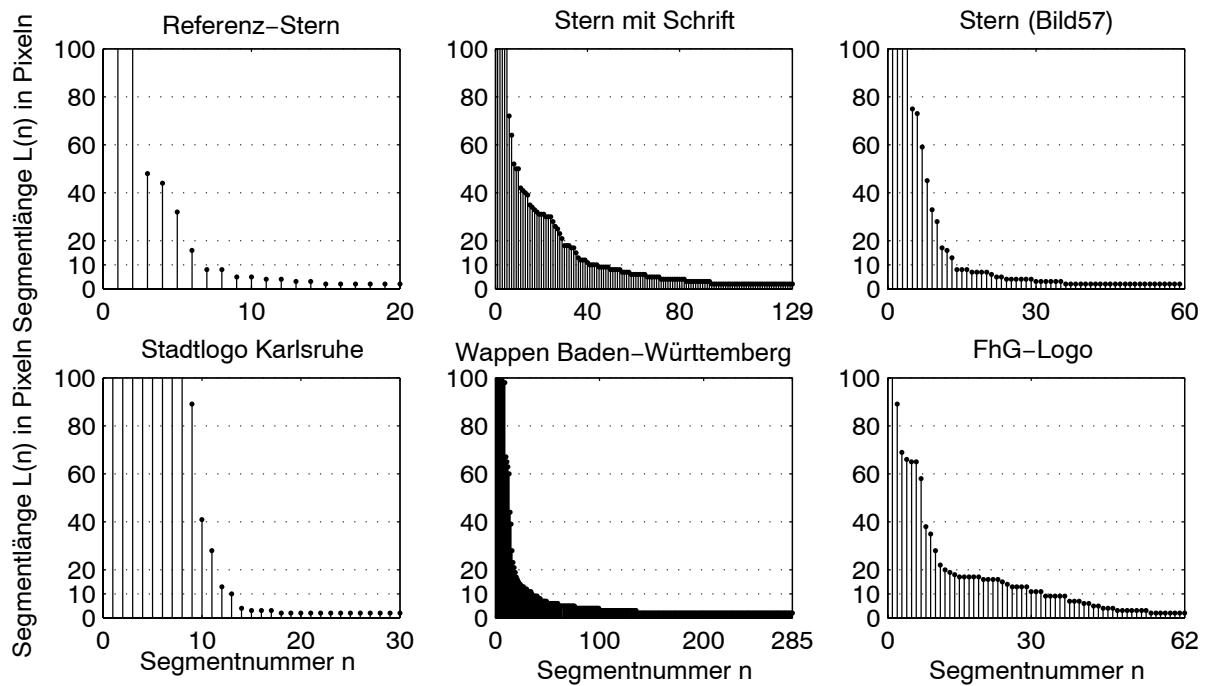


Abbildung 4.19: Verteilung der Segmentlängen

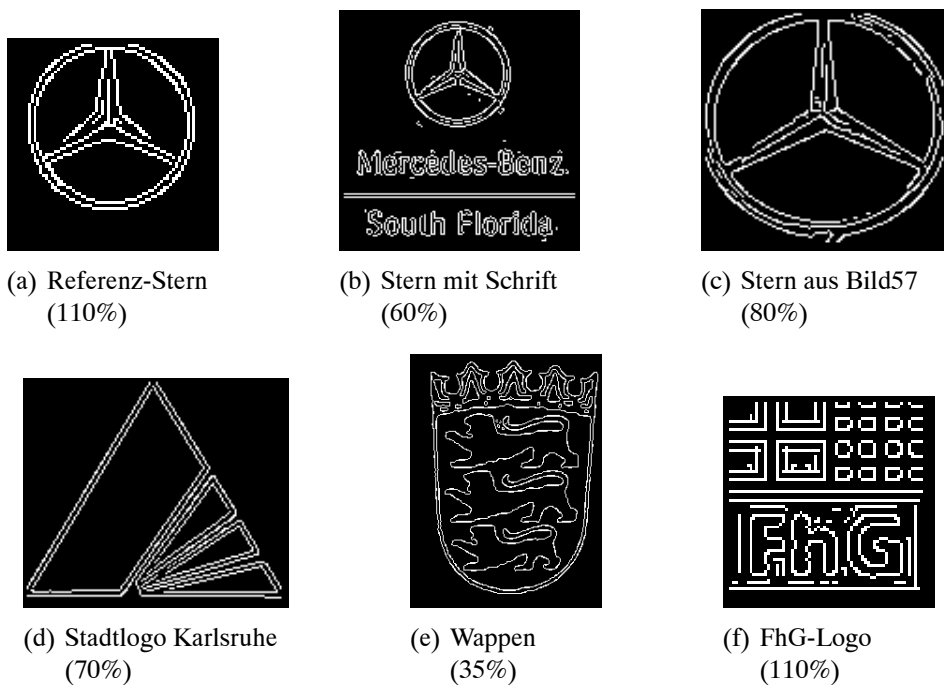


Abbildung 4.20: KPL-Bilder zu Abbildung 4.19

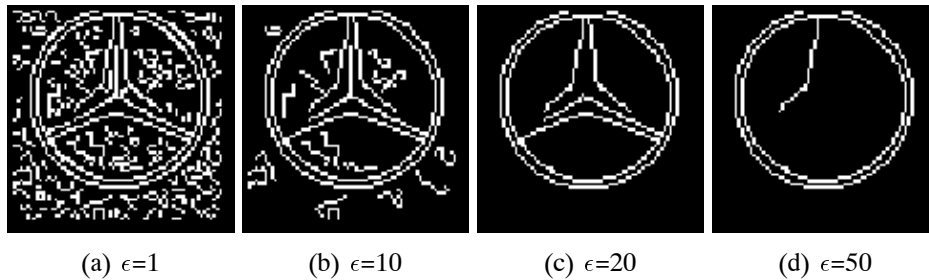


Abbildung 4.21: Akkumulierte Segmentbilder für unterschiedliche minimale Segmentlängen

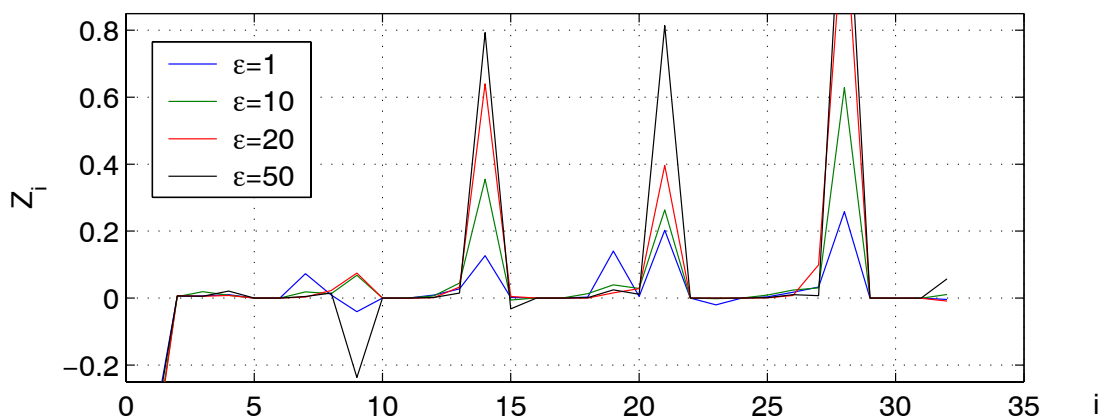


Abbildung 4.22: Invariante Zernike-Momente bei unterschiedlichen minimalen Segmentlängen

während für $\epsilon = 50$ bereits relevante Elemente der Kontur verloren gehen. Dies bestätigt auch die Berechnung des Abstandes zum Referenzvektor der Bundesadler-Klasse $d(\mathbf{Z}_0, \mathbf{Z}(\epsilon))$: Bei der Verwendung der gesamten Kontur ergibt sich $d(\mathbf{Z}_0, \mathbf{Z}(\epsilon = 1)) = 1,2958$, bei den in der Praxis sinnvollen Werten ergibt sich $d(\mathbf{Z}_0, \mathbf{Z}(\epsilon = 10)) = 0,8367$ bzw. $d(\mathbf{Z}_0, \mathbf{Z}(\epsilon = 20)) = 0,2545$, während für ein zu großes ϵ der Abstand wieder ansteigt zu $d(\mathbf{Z}_0, \mathbf{Z}(\epsilon = 50)) = 0,6494$.

In welchem Maße sich durch den weiteren Vorverarbeitungsschritt der Segmentlängenbegrenzung nun das Ergebnis der Klassifikation verbessert, ist stark von dem verwendeten Bildmaterial abhängig. Aber gerade bei Bildern mit vielen kleinen Störungen kann so die Erkennungsrate erhöht werden. Eine Verbesserung ohne großen Mehraufwand kann auch erreicht werden, wenn zumindest für die Referenzbilder beim Trainieren des Klassifikators vor der Berechnung der invarianten Zernike-Momente diese Segmentlängenbegrenzung durchgeführt wird.

4 Ergebnisse

4.4.2 Randunterdrückung

Einige Bilder werden falsch klassifiziert, weil das gesamte Bild von einem Rand umgeben ist. Auch Rahmen, Schatten, abgesetzte Hintergrundfarbe oder dreidimensionale Buttonoptik führen zu Konturen, durch die sich die Zahlenwerte der Momente wesentlich verändern. Aus diesem Grund wird an dieser Stelle ein einfaches Verfahren betrachtet, um auch bei diesen Bildern eine richtige Aussage über den Inhalt treffen zu können.

Die Momente werden richtig berechnet, wenn die Bereiche störender Konturen aus dem Integrationsgebiet der Momentenberechnung herausgenommen werden. Das wird dadurch erreicht, dass ein Teilbild (*frame*) im Bild definiert wird, auf den sich die Momentenberechnung beschränkt. Dazu wird ein *Offset-Vektor* $\Delta = (\Delta_x, \Delta_y)^T$ definiert, der den Abstand des Teilbildes zu den Bildrändern in x- und y-Richtung enthält. Bei Bildern mit ähnlichen Dimensionen ist es sinnvoll, den Offset-Vektor konstant zu wählen, bei sehr unterschiedlichen Bilddimensionen kann auch ein fester Prozentsatz der Bildgröße als maßgebende Größe für den Offset-Vektor verwendet werden.

Abbildung 4.23(a) zeigt ein typisches Beispiel für ein Bild mit Rand und Schattenwurf. Mit dem Offset-Vektor $\Delta = (15 \text{ Pixel}, 15 \text{ Pixel})^T$ wurde daraus das Teilbild 4.23(b) extrahiert. Betrachtet man dazu die invarianten Zernike-Momente, wie sie in Abbildung 4.24 als sequenzielle Werte dargestellt sind, so ist der Unterschied deutlich ersichtlich: Die invarianten Zernike-Momente des gesamten Bildes weisen große Unterschiede zum Referenzvektor¹ auf, während der Merkmalvektor des Teilbildes so nahe an den Referenzvektor herankommt, dass in der Darstellung die Vektorelemente kaum zu unterscheiden sind.

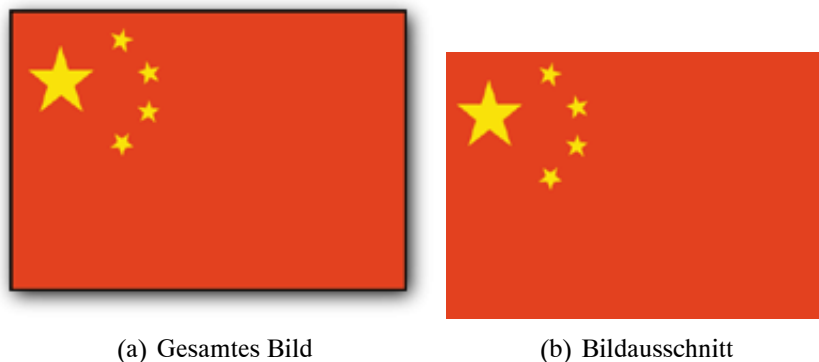


Abbildung 4.23: Randunterdrückung bei einem Testbild mit der China-Flagge

4.4.3 Kachelung

Auch wenn bei den zu untersuchenden Bildern die Segmentierung und die Randunterdrückung berücksichtigt wird, gibt es viele Bilder, bei denen die Klassifikation nicht zum richtigen Ergeb-

¹Vektor der invarianten Zernike-Momente, die aus Bild 4.1(d) berechnet werden

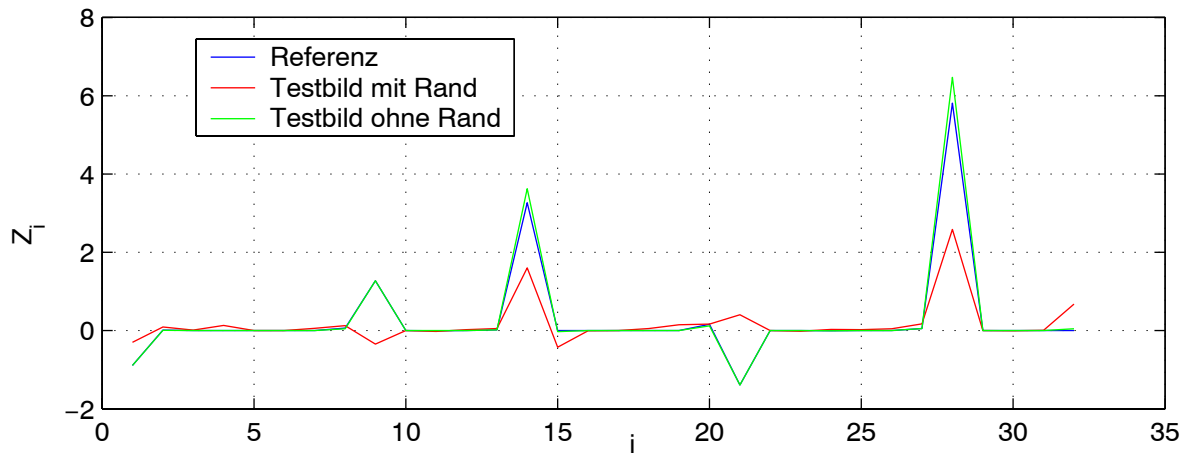


Abbildung 4.24: Invariante Zernike-Momente für die Bilder mit Randunterdrückung in Abbildung 4.23

nis führt. Bei diesen Bildern wird die in Abschnitt 4.1 festgelegte Voraussetzung nicht erfüllt und es existieren neben dem gesuchten Objekt noch weitere relevante Objekte im Bild. Eine richtige Klassifikation mit invarianten Momenten ist dann auch möglich, wenn nur der betreffende Bildausschnitt, in dem sich das relevante Objekt befindet, für die Berechnung der Momente verwendet wird. Über die Position und die Größe des Objektes im Bild wird man im Allgemeinen kein Vorwissen besitzen, deshalb soll mit einer Kachelung (*tiling techniques*) ein möglichst einfaches und robustes Verfahren verwendet werden, das das Gesamtbild in mehrere Teilbilder unterteilt, in denen dann jeweils die invarianten Momente berechnet werden.

Ein solches Kachelungsprinzip sollte zwei Anforderungen erfüllen:

1. Das Objekt sollte an jeder Position im Bild gefunden werden. Dazu wird eine Kachelung mit 50% Überlappung verwendet, sodass der Definitionsbereich des Teilbildes durch das Gesamtbild von links oben bis rechts unten „durchgeschoben“ wird.
2. Das Objekt sollte unabhängig von der Größe im Bild gefunden werden. Deshalb wird die Größe der Kacheln stufenweise reduziert. Nach der Suche im Gesamtbild wird das Bild zunächst in Teilbilder zerlegt, die jeweils ein Viertel der Bildfläche umfassen, danach kann die Suche mit einer noch feineren Unterteilung fortgesetzt werden.

Bei 50% Überlappung in x- und y-Richtung und der vorgeschlagenen Teilbildgröße von 25% ergibt die erste Kachelungsstufe die in Abbildung 4.25 dargestellten neun Teilbilder.

Eine optimale Lösung wäre eine quasi-kontinuierliche Verschiebung des Definitionsbereiches und eine quasi-kontinuierliche Verfeinerung der Teilbildaufösung in Pixelschritten. Das ist in der Praxis wegen zu großer Rechenzeit nicht sinnvoll, aber auch durch das einfache vorgeschlagene Verfahren lassen sich in einigen Fällen bereits die relevanten Objekte finden.

Für die Aufteilung der Kacheln ist es hilfreich, das Wissen über das Referenzobjekt mit einzubringen. So könnte man bei der verwendeten Datenbank bei der Suche nach einem Objekt

4 Ergebnisse

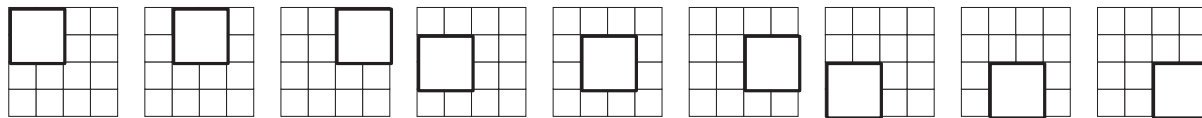


Abbildung 4.25: Prinzip der überlappenden Kachelung

der Klassen Bundesadler, Coca-Cola-Logo oder Mercedesstern möglichst quadratische Kacheln wählen, bei der Suche nach Flaggen eher rechteckige Kacheln mit unterschiedlichen Kantenlängen. Dabei ist allerdings zu bedenken, dass durch die Rotation eines Objektes im Teilbild die optimale Kachelform ebenfalls zu rotieren wäre. Deshalb sind für den allgemeinen Fall ohne Vorwissen eher Kacheln, deren Größe einen fest vorgegebenden Anteil am Gesamtbild ausmachen, eine einfache und sinnvolle Wahl.

Als Beispiel wird ein Bild untersucht, das den Bundesadler und ein zusätzliches Objekt enthält (Abbildung 4.26). Der Abstand des daraus berechneten Merkmalvektors \mathbf{Z}_{gesamt} zum Referenzvektor² \mathbf{Z}_0 beträgt $d(\mathbf{Z}_{gesamt}, \mathbf{Z}_0) = 0.8674$. Wenn nun die vorgeschlagene Kachelung durchgeführt wird (siehe Abbildung 4.27), so gibt es eine Kachel, in der der Bundesadler nahezu vollständig enthalten ist. Der Vergleich der invarianten Momente der einzelnen Kacheln $\mathbf{Z}_{Kachel \kappa}$, $1 \leq \kappa \leq 9$, mit dem Referenzvektor zeigt, dass bei der Kachel, die den Bundesadler enthält, gerade der kleinste Wert $d(\mathbf{Z}_{Kachel 2}, \mathbf{Z}_0) = 0.3002$ angenommen wird. Dieser Wert liegt wohl etwas oberhalb der Werte, die in Unterabschnitt 4.2.2 für Bilder mit gleichen Objekten ermittelt wurden, das ist jedoch durch die Unschärfe des Bundesadlers im Testbild 4.26 zu erklären und nicht auf die Kachelung zurückzuführen. Tabelle 4.4 enthält die Euklidischen Abstände für alle Kacheln $1 \leq \kappa \leq 9$. Mit *NaN* (Not a Number) sind die Werte markiert, bei denen die Kachelung ein leeres Teilbild ergab. In Abbildung 4.28 sind der Übersichtlichkeit wegen nur die invarianten Zernike-Momente für das Referenzbild, das gesamtem Testbild und für die beste Kachel dargestellt.



Abbildung 4.26: Bundesadler mit weiterem Objekt (gesamtes Bild)

²Vektor der invarianten Zernike-Momente, die aus Bild 4.1(c) berechnet werden

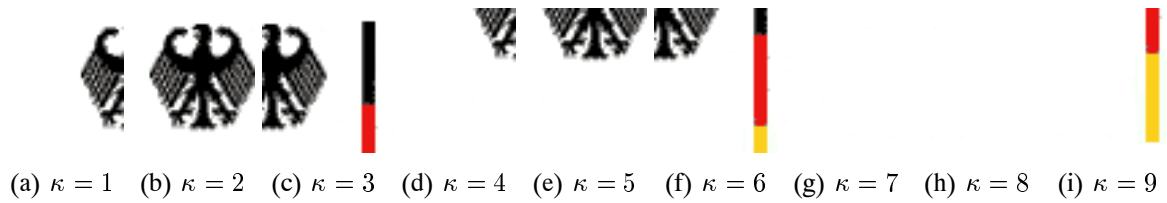


Abbildung 4.27: Bundesadler mit weiterem Objekt (Kacheln)

Kachel	Oben links	Oben mitte	Oben rechts	Mitte links	Mitte mitte	Mitte rechts	Unten links	Unten mitte	Unten rechts
κ	1	2	3	4	5	6	7	8	9
$d(\mathbf{Z}_\kappa, \mathbf{Z}_0)$	1.2778	0.3002	0.5551	2.2660	1.4141	0.8378	<i>NaN</i>	<i>NaN</i>	1.1728

Tabelle 4.4: Euklidischer Abstand der invarianten Zernike-Momente eines gekachelten Bildes zum Original (Bundesadler)

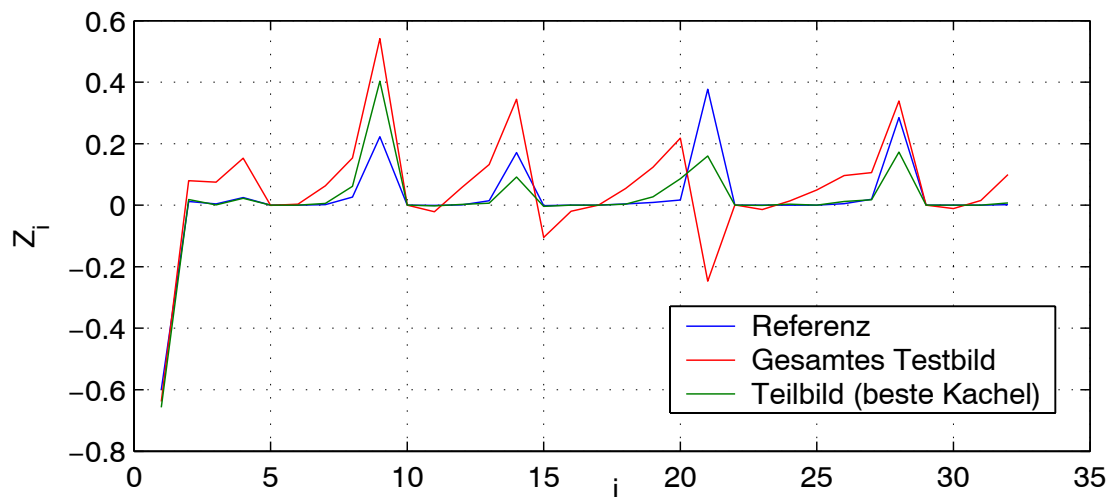


Abbildung 4.28: Invariante Zernike-Momente bei der Kachelung eines Bildes (Bundesadler)

4 Ergebnisse

5 Zusammenfassung und Ausblick

5.1 Zusammenfassung und Beurteilung des Verfahrens

Im Rahmen dieser Diplomarbeit wurde die binäre und multiple Klassifikation von Bildern betrachtet. Als Grundlage für das Klassifikationskriterium wurden die aus dem Bildinhalt bestimmten Momente verwendet. Daraus wurden Merkmale abgeleitet, die unter den affinen Transformationen Skalierung, Rotation, Translation und Spiegelung invariant sind und deshalb als invariante Momente bezeichnet werden. Es wurden dabei zunächst die aus den geometrischen Momenten abgeleiteten invarianten Momente nach Hu betrachtet, bei der Klassifikation wurden dann die invarianten Zernike-Momente verwendet, basierend auf den komplexen Zernike-Momenten.

Die Klassifikation wurde an einer Datenbank mit 154 Bildern aus dem Internet getestet, die die vier Objektklassen Bundesadler, Coca-Cola-Logo, China-Flagge und Mercedesstern enthalten. Einzelne Bilder mit UNO-Flaggen wurden ebenfalls untersucht. In der Vorverarbeitung wurden die Bilder in das PM-Format gewandelt, sowie ein Gradientenbild und daraus ein Konturbild erzeugt. Für jede Objektklasse wurde die Charakteristik des Klassifikators ermittelt und die funktionale Abhängigkeit der Entdeckungswahrscheinlichkeit von der Falschalarmwahrscheinlichkeit dargestellt.

Die Ergebnisse haben gezeigt, dass sich invariante Momente für die binäre Detektion eines Objektes im Bild gut eignen. Insbesondere bei dem geometrisch sehr einfach aufgebauten Objekt der China-Flagge ist eine sehr gute Klassifikation möglich, denn es werden bei einer Entdeckungsrate von 90% nur weniger als 10% der Bilder, die andere Objekte enthalten, fälschlicherweise akzeptiert. Aber auch bei den anderen betrachteten Klassifikatoren konnte eine deutliche Verbesserung im Vergleich zur „blinden“ Entscheidung erreicht werden, was sich in der Empfängerarbeitscharakteristik als großer Abstand zu der Vergleichskurve *Entdeckungswahrscheinlichkeit = Falschalarmwahrscheinlichkeit* gezeigt hat. Bei gleichen Entdeckungsraten von 90% liegen die Falschalarmraten dieser Klassifikatoren im Bereich von 40-45%, nur der Mercedesstern-Klassifikator zeigt ein etwas schlechteres Ergebnis.

Bei der multiplen Detektion wurde getestet, wie gut die einzelnen Objekte mit den invarianten Momenten von einander unterschieden werden können. Dazu wurde ein *K*NN-Klassifikator mit $K = 5$ verwendet. Auch in diesem Fall hat sich ein gutes Klassifikationsergebnis gezeigt. Es wurden ohne jegliches *A-priori*-Wissen 58% der Bilder der richtigen Klasse zugeordnet.

Das Verfahren wurde außerdem mit mehreren Referenzbildern getestet und mit zwei verschiedenen Ansätzen, die die Dynamik in den Elementen der invarianten Momente berücksichtigen. Es konnte so eine leichte Verbesserung des Klassifikationsergebnisses in einzelnen Fällen erreicht

5 Zusammenfassung und Ausblick

werden. Die Verhältnisse von Entdeckungsrate und Falschalarmrate blieben dabei jedoch in der selben Größenordnung.

Damit ist es möglich, allein aus der Bildinformation mit Hilfe der invarianten Momente eine Aussage über die Existenz eines Objektes im Bild zu treffen. Es kann eine Datenbank gezielt nach einzelnen Objekten durchsucht werden, ohne jedes einzelne Bild betrachten zu müssen. Eine große Ähnlichkeit von Datenbankbild und Referenzbild mit dem selben Objekt zeigt sich in richtigen Klassifikationsentscheidungen, Bilder mit sehr ähnlichen Bildkompositionen, aber anderem Inhalt, führen zu Falschentscheidungen des Klassifikators.

Die Klassifikation mit invarianten Momenten zeigte sich problematisch, wenn das Bild von Rauschen in Form von Pixelstörungen überlagert oder das Objekt von einem Rand umgeben ist. Für diese Fälle wurden zwei Erweiterungen des Verfahrens vorgestellt, die die störenden Bildinhalte aus dem Bild entfernen.

Durch einen weiteren Vorverarbeitungsschritt wurde die Kontur des gestörten Bildes in Segmente zerlegt und alle Segmente vernachlässigt, die eine Minimallänge unterschreiten. Eine optimale Minimallänge wurde an dem Punkt bestimmt, bei dem die Verteilung der Segmentlängen die größte Krümmung aufweist. Damit konnte erreicht werden, dass auch stark gestörte Bilder richtig erkannt werden.

Bildränder konnten durch die Definition eines Teilbildes berücksichtigt werden. Die invarianten Momente werden dann nur für dieses Teilbild berechnet. Es konnte gezeigt werden, dass durch die Verwendung eines sinnvollen Randabstandes auch bei diesen Bildern eine gute Erkennung grundsätzlich möglich ist.

Das Verfahren versagt, wenn in einem Bild neben dem Objekt noch weitere Objekte enthalten sind, das Bild einen strukturierten Hintergrund besitzt oder zusätzlich Schrift auftaucht. Dafür wurde abschließend die Kachelung des Bildes mit abnehmender Kachelgröße und 50% Überlappung betrachtet. Es konnte gezeigt werden, dass das relevante Objekt im Bild dann erkannt wird, wenn es sich vollständig in einer der Kacheln befindet.

5.2 Ausblick

Die invarianten Momente wurde bisher an einem für die Aufgabe repräsentativen, aber dennoch eingeschränkten Bildmaterial getestet. Deshalb wäre es wünschenswert, umfangreichere Datenbanken mit Bildern zu untersuchen und die Eignung der invarianten Momente auch für mehr unterschiedliche Objektklassen zu prüfen. Auch eine Implementierung als Java-Applet im Internet wäre denkbar.

Außerdem wäre eine weitere Arbeit hilfreich, die die Objektlokalisierung im Bild untersucht und weitere Verfahren zur Definition der relevanten *Region-of-Interest* (ROI) berücksichtigt. Für die Berechnung der invarianten Momente sollte nur der Bildbereich herangezogen werden, der das Objekt im Bild auch enthält. Deshalb stellt diese Teilbilddefinition eine wichtige Aufgabe dar,

die sich unmittelbar auf das Klassifikationsergebnis auswirken wird.

Möchte man mit invarianten Momenten auch Bilder erkennen, die das zu suchende Objekt in unterschiedlichen Realisierungen enthalten, so kann dafür auf die *Fuzzy Logic* zurückgegriffen werden. Damit können auch Bilder erkannt werden, die bei gleichem Inhalt unterschiedliche geometrische Kompositionen und deshalb große Abstände der Vektoren im Merkmalraum besitzen.

Eine entscheidende Verbesserung in der Zuverlässigkeit der Klassifikationsaussage ließe sich durch die Kombination der invarianten Momente mit anderen Beschreibungsformen für das Bild erreichen. Bei den Momenten handelt es sich um formbeschreibende Merkmale, die bewusst weder Farbe, noch Textur oder Struktur berücksichtigen. In manchen Fällen kann es aber sinnvoll sein, diese anderen Merkmale, die ebenfalls aus der Darstellung der Bildmatrix im Orts- oder Frequenzbereich gewonnen werden können, mit den Momenten zu kombinieren. In [CW02] wird ein Ansatz zur Kombination von verschiedenen Merkmalen mit Hilfe von *Fuzzy Logic* besprochen.

Denkt man noch einmal an die oben erwähnte musterbasierte Suche eines Bildes in einer Clipart-Sammlung, so ist in vielen Fällen neben der eigentlichen Bildinformation jedoch noch eine weitere Beschreibung vorhanden: eine semantische Beschreibung in Form von Dateinamen und Schlüsselwörtern (*keywords*). In [ZH02] wird dieses Wissen mit einfachen Merkmalen aus der Bildinformation gewichtet verknüpft, wobei der Anwender durch die Bewertung der Ergebnisse die Gewichtungsfaktoren beeinflussen kann. In [WSZ02] wird andererseits versucht, aus der Bildinformation auf eine allgemeine semantische Beschreibung des Bildinhalt zu schließen.

So wäre es denkbar, eine Klassifikation in zwei Schritten durchzuführen: Zunächst wird mit Hilfe der invarianten Momente eine Vorauswahl getroffen und die Anzahl der möglicherweise relevanten Bilder wesentlich reduziert. In einem zweiten Schritt werden schließlich die Schlüsselwörter verglichen und damit Bilder unterschieden, die bei ähnlichem geometrischem Aufbau unterschiedliche Inhalte besitzen.

Auf diese Weise ist es möglich, mit Hilfe der invarianten Momente einen wichtigen Beitrag zur Thematik des *Image Retrieval* zu leisten und eine deutliche Verbesserung bei der automatisierten Suche nach Objekten in Datenbanken zu erreichen.

A Ergänzungen zu der Momententheorie

A.1 Berechnung von Zernike-Momenten aus geometrischen Momenten

Da in keiner der angeführten Literaturquellen die Herleitung zu Gleichung (2.19) durchgeführt wird, soll sie an dieser Stelle ausformuliert werden. Dies dient vor allem auch dem besseren Verständnis der Zernike-Momente. Wir betrachten dabei nur die Zernike-Momente A_{nl} für $l \geq 0$ und berechnen die restlichen Momente über die Beziehung $A_{n,-l} = A_{nl}^*$. Jeder Punkt (x, y) der Bildebene wird als komplexe Zahl $z = x + iy = r \cdot e^{i\phi} \in \mathbb{C}$ mit dem Betrag r und dem Argument ϕ betrachtet. Es gilt nach (2.13) und (2.14)

$$A_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{nl}(r, \phi)]^* \tilde{s}(r \cos \phi, r \sin \phi) r dr d\phi \quad (\text{A.1})$$

$$= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 R_{nl}(r) \exp(-il\phi) \cdot \tilde{s}(r \cos \phi, r \sin \phi) r dr d\phi. \quad (\text{A.2})$$

Setzt man darin (2.16) ein, so erhält man

$$A_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 \sum_{\substack{k=l \\ (n-k) \text{ gerade}}}^n B_{nlk} r^k \exp(-il\phi) \cdot \tilde{s}(r \cos \phi, r \sin \phi) r dr d\phi \quad (\text{A.3})$$

$$= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 \sum_{\substack{k=l \\ (n-k) \text{ gerade}}}^n B_{nlk} [r^l e^{-il\phi}] r^{k-l} \tilde{s}(r \cos \phi, r \sin \phi) r dr d\phi. \quad (\text{A.4})$$

A Ergänzungen zu der Momententheorie

Unter der Berücksichtigung der Beziehungen $r \exp(-i\phi) = x - iy$ und $r = \sqrt{x^2 + y^2}$ erhalten wir

$$A_{nl} = \frac{n+1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sum_{k=l}^n B_{nlk} (x-iy)^l (x^2+y^2)^{(k-l)/2} \tilde{s}(x,y) dx dy \quad (\text{A.5})$$

(n-k) gerade

$$= \frac{n+1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sum_{k=l}^n B_{nlk} (x-iy)^l (x^2+y^2)^q \cdot \tilde{s}(x,y) dx dy \quad (\text{A.6})$$

(n-k) gerade

mit $q = (k-l)/2$. An dieser Stelle findet die allgemeine binomische Formel $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$ Verwendung:

$$A_{nl} = \frac{n+1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sum_{k=l}^n B_{nlk} \sum_{m=0}^l \binom{l}{m} (-i)^m x^{l-m} y^m (x^2+y^2)^q \cdot \tilde{s}(x,y) dx dy \quad (\text{A.7})$$

(n-k) gerade

$$= \frac{n+1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sum_{k=l}^n B_{nlk} \sum_{m=0}^l \binom{l}{m} (-i)^m x^{l-m} y^m \cdot$$

(n-k) gerade

$$\sum_{j=0}^q \binom{q}{j} x^{2j} y^{2(q-j)} \cdot \tilde{s}(x,y) dx dy \quad (\text{A.8})$$

$$= \frac{n+1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sum_{k=l}^n B_{nlk} \sum_{m=0}^l \binom{l}{m} (-i)^m x^{l-m} y^m \cdot$$

(n-k) gerade

$$\sum_{j=0}^q \binom{q}{j} x^{2j} y^{2(q-j)} \cdot \tilde{s}(x,y) dx dy \quad (\text{A.9})$$

$$= \frac{n+1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sum_{k=l}^n B_{nlk} \sum_{m=0}^l \sum_{j=0}^q \binom{l}{m} \binom{q}{j} (-i)^m \cdot$$

(n-k) gerade

$$\cdot x^{l-m+2j} y^{m+2q-2j} \cdot \tilde{s}(x,y) dx dy. \quad (\text{A.10})$$

Vertauscht man die Reihenfolge von Summation und Integration, so erhalten wir folgenden Aus-

A.2 Implementierung der geometrischen Momente für eine diskrete Bildmatrix

druck:

$$A_{nl} = \frac{n+1}{\pi} \sum_{k=l}^n \underset{(n-k) \text{ gerade}}{B_{nlk}} \sum_{m=0}^l \sum_{j=0}^q \binom{l}{m} \binom{q}{j} (-i)^m \cdot \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x^{l-m+2j} y^{m+2q-2j} \cdot \tilde{s}(x, y) dx dy \quad (\text{A.11})$$

Darin tauchen die geometrischen Momente $\{m_{pq}\}$ mit dem Integrationsgebiet des Einheitskreises auf:

$$A_{nl} = \frac{n+1}{\pi} \sum_{k=l}^n \underset{(n-k) \text{ gerade}}{B_{nlk}} \sum_{m=0}^l \sum_{j=0}^q \binom{l}{m} \binom{q}{j} (-i)^m \cdot m_{l-m+2j, m+2q-2j} \quad (\text{A.12})$$

$$= \frac{n+1}{\pi} \sum_{k=l}^n \underset{(n-k) \text{ gerade}}{B_{nlk}} \sum_{m=0}^l \sum_{j=0}^q \binom{l}{m} \binom{q}{j} (-i)^m \cdot m_{l-m+2j, m+k-l-2j} \quad (\text{A.13})$$

Diese Gleichung entspricht gerade (2.19). Die praktische Implementierung dieser Formel unter Verwendung der diskreten Bildmatrix \mathbf{S} wird in Anhang A.2 diskutiert.

Die Koeffizienten

$$B_{nlk} = (-1)^{\frac{n-k}{2}} \frac{\binom{n+k}{2}!}{\binom{n-k}{2}! \binom{k+l}{2}! \binom{k-l}{2}!} \quad (\text{A.14})$$

nach Gleichung (2.17) werden dabei rekursiv berechnet [MR98]:

$$B_{nnn} = 1 \quad (\text{A.15})$$

$$B_{nl n} = B_{n(l+2)n} \frac{n+l+2}{n-l} \quad (\text{A.16})$$

$$B_{nlk} = -B_{nl(k+2)} \frac{(k+2)^2 - l^2}{(n+k+2)(n-k)}. \quad (\text{A.17})$$

Wegen der Beziehung $A_{n,-l} = A_{nl}^*$ brauchen die Koeffizienten B_{nlk} nur für $l \geq 0$ berechnet werden. Für die ersten Zernike Momente ergeben sich mit den Restriktionen $k = l, \dots, n$ und $(n-k)$ gerade damit die Werte aus Tabelle A.1.

A.2 Implementierung der geometrischen Momente für eine diskrete Bildmatrix

Für die Berechnung der Zernike-Momente werden nach Gleichung (2.19) die geometrischen Momente verwendet. In keiner der Literaturangaben wurde die Herleitung zu den Formeln mit

A Ergänzungen zu der Momententheorie

n	l	k	Zernike-Moment	Koeffizient
0	0	0	A_{00}	$B_{000} = 1$
1	1	1	A_{11}	$B_{111} = 1$
2	2	2	A_{22}	$B_{222} = 1$
		0	A_{20}	$B_{202} = 2B_{222} = 2$ $B_{200} = -1/2B_{202} = -1$
3	3	3	A_{33}	$B_{333} = 1$
		1	A_{31}	$B_{313} = 3B_{333} = 3$ $B_{311} = -2/3B_{313} = -2$
4	4	4	A_{44}	$B_{444} = 1$
		2	A_{42}	$B_{424} = 4B_{444} = 4$ $B_{422} = -3/4B_{424} = -3$
		0	A_{40}	$B_{404} = 3/2B_{424} = 6$ $B_{402} = -B_{404} = -6$ $B_{400} = -1/6B_{402} = 1$

Tabelle A.1: Koeffizienten zur Berechnung der Zernike-Momente aus geometrischen Momenten

einer diskreten Bildmatrix, wie wir sie verwenden, angegeben. Deshalb soll sie zum besseren Verständnis der Matlab-Implementierung an dieser Stelle angegeben werden. Der wesentliche Unterschied zu der einfachen Formel (2.7) besteht darin, dass das der Summation zugrundeliegende Integrationsgebiet anders definiert ist. Die Verwendung von geometrischen Momenten in (2.19) erfordert damit eine Skalierung des Bildes $\mathbf{S} = ((s_{ij}))$, $1 \leq i \leq M$, $1 \leq j \leq N$, innerhalb des Einheitskreises. Die Matrixindices werden zur Unterscheidung von den Koordinaten x und y im Folgenden mit i und j bezeichnet. Als weiterer Vorbereitungsschritt davor ist das Bild zu normieren und auf quadratische Größe zu ergänzen.

O.b.d.A. sei $N \geq M$. Dann wird die Bildmatrix ergänzt zu

$$s'_{ij} = \begin{cases} s_{ij} & \text{für } 1 \leq i \leq M \quad \text{und } 1 \leq j \leq N \\ 0 & \text{für } M+1 \leq i \leq N \quad \text{und } 1 \leq j \leq N \end{cases} . \quad (\text{A.18})$$

Die Normierung erfolgt mit Hilfe der Bildmasse $m_{00} = \sum_{i=1}^N \sum_{j=1}^N s'_{ij}$ zu

$$s''_{ij} = \frac{1}{m_{00}} s'_{ij} . \quad (\text{A.19})$$

Die Integration über den Einheitskreis $x^2 + y^2 \leq 1$ kann einfach in eine Integration über das Einheitsquadrat $[-1, 1] \times [-1, 1]$ überführt werden, wenn die Bildfunktion außerhalb des Einheitskreises den Wert Null annimmt. Bei der Implementierung muss dabei die Bildmatrix auf die Dimension $R \times R$ mit $R = \lceil \sqrt{2}N \rceil$ erweitert werden mit

$$s'''_{ij} = \begin{cases} s''_{ij} & \text{für } 1 \leq i \leq N \quad \text{und } 1 \leq j \leq N \\ 0 & \text{für } N+1 \leq i \leq R \quad \text{und } N+1 \leq j \leq R \end{cases} . \quad (\text{A.20})$$

A.2 Implementierung der geometrischen Momente für eine diskrete Bildmatrix

Dann können die Matrixelemente s_{ij}''' als Abtastwerte der innerhalb des Einheitskreises definierten kontinuierlichen Bildfunktion $\tilde{s}(x, y)$ interpretiert werden:

$$s_{ij}''' = \tilde{s} \left(-1 + \frac{2}{R-1} (i-1), -1 + \frac{2}{R-1} (j-1) \right) \text{ mit } i, j = 1, \dots, R. \quad (\text{A.21})$$

Es gilt für die geometrischen Momente, die zur besseren Unterscheidung zu der Definition (2.6) mit G_{pq} bezeichnet werden,

$$G_{pq} = \int_{-1}^1 \int_{-1}^1 x^p y^q \tilde{s}(x, y) dx dy. \quad (\text{A.22})$$

Mit der formalen Variablensubstitution

$$\begin{pmatrix} i \\ j \end{pmatrix} = \frac{R-1}{2} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{R+1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} di \\ dj \end{pmatrix} = \frac{R-1}{2} \begin{pmatrix} dx \\ dy \end{pmatrix} \quad (\text{A.23})$$

bzw.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{2}{R-1} \left[\begin{pmatrix} i \\ j \end{pmatrix} - \frac{R+1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right], \quad \begin{pmatrix} dx \\ dy \end{pmatrix} = \frac{2}{R-1} \begin{pmatrix} di \\ dj \end{pmatrix} \quad (\text{A.24})$$

und den Definitionsgebieten

$$\begin{aligned} -1 &\leq x \leq 1 && \Leftrightarrow && 1 &\leq i &\leq R \\ -1 &\leq y \leq 1 && && 1 &\leq j &\leq R \end{aligned} \quad (\text{A.25})$$

ergibt sich

$$\begin{aligned} G_{pq} &= \int_1^R \int_1^R \left(\frac{2}{R-1} \right)^p \left[i - \frac{R+1}{2} \right]^p \left(\frac{2}{R-1} \right)^q \left[j - \frac{R+1}{2} \right]^q \\ &\quad \cdot \tilde{s}(x(i, j), y(i, j)) \left(\frac{2}{R-1} \right)^2 di dj. \end{aligned} \quad (\text{A.26})$$

Der für alle Momente eines Bildes konstante Faktor $\left(\frac{2}{R-1}\right)^2$ wird oft in den Endergebnissen der Literatur weggelassen (z.B. [ARY02]). Das ist nur möglich, wenn eine einheitliche Bildgröße bei allen Bildern verwendet wird. Dann ergibt sich mit der Abkürzung $D = \frac{R+1}{2}$ und dem Übergang auf eine Summation

$$\tilde{G}_{pq} = \sum_{i=1}^R \sum_{j=1}^R \left(\frac{i-D}{D-1} \right)^p \left(\frac{j-D}{D-1} \right)^q s_{ij}'''. \quad (\text{A.27})$$

Auch bei der praktischen Implementierung werden nun zentrierte geometrische Momente als Grundlage für die Berechnung der Zernike-Momente verwendet:

$$\tilde{G}_{pq}^{Tr} = \sum_{i=1}^R \sum_{j=1}^R \left(\frac{i-D}{D-1} - \tilde{G}_{10} \right)^p \left(\frac{j-D}{D-1} - \tilde{G}_{01} \right)^q s_{ij}'''. \quad (\text{A.28})$$

A Ergänzungen zu der Momententheorie

Listing A.1 enthält einen Ausschnitt aus dem Matlab-Programm, das mit einfachen Matrizen-Multiplikationen die ersten zwei geometrischen Momente berechnet. Natürlich können an dieser Stelle auch die in Abschnitt 2.2.1 angesprochenen schnelleren Verfahren, z.B. nach Hatamian [Hat86] eingesetzt werden.

Listing A.1: Matlab-Code zur Berechnung der geometrischen Momente

```
D=(R+1)/2;

i = ones(R,1)*[1:R];
j = [1:R]'*ones(1,R);

G_00 = 1;          % durch Normierung

G_x = sum(sum((i-D).*image/(D-1)));
G_y = sum(sum((j-D).*image/(D-1)));

idash = ((i-D)/(D-1)) - G_x;
jdash = ((j-D)/(D-1)) - G_y;

%*****
%*** Zentrierte geometrische Momente berechnen ***
%*****

% Ordnung 1
G_10 = 0;          % durch Zentrieren
G_01 = 0;

% Ordnung 2
G_20 = sum(sum(image.*idash.*idash));
G_02 = sum(sum(image.*jdash.*jdash));
G_11 = sum(sum(image.*idash.*jdash));

% Ordnung 3
...
```


B Ergänzungen zu den invarianten Zernike-Momenten

B.1 Berechnung der invarianten Zernike-Momente aus den Zernike-Momenten

In Abschnitt 2.3.3 wurden bereits die ersten invarianten Zernike-Momente angegeben. An dieser Stelle soll nun eine Auflistung aller verwendeter Zernike-Momente bis zu der Ordnung $n = 7$ nach [Tea80] folgen.

Ordnung $n = 2$:

$$Z_1 = A_{20} \quad (\text{B.1})$$

$$Z_2 = |A_{22}|^2 \quad (\text{B.2})$$

Ordnung $n = 3$:

$$Z_3 = |A_{33}|^2 \quad (\text{B.3})$$

$$Z_4 = |A_{31}|^2 \quad (\text{B.4})$$

$$Z_5 = A_{33}(A_{31}^*)^3 + [A_{33}(A_{31}^*)^3]^* \quad (\text{B.5})$$

$$Z_6 = A_{31}^2 A_{22}^* + [A_{31}^2 A_{22}^*]^* \quad (\text{B.6})$$

Ordnung $n = 4$:

$$Z_7 = |A_{44}|^2 \quad (\text{B.7})$$

$$Z_8 = |A_{42}|^2 \quad (\text{B.8})$$

$$Z_9 = A_{40} \quad (\text{B.9})$$

$$Z_{10} = A_{44}^* A_{42}^2 + [A_{44}^* A_{42}^2]^* \quad (\text{B.10})$$

$$Z_{11} = A_{42} A_{22}^* + [A_{42} A_{22}^*]^* \quad (\text{B.11})$$

$$(\text{B.12})$$

B Ergänzungen zu den invarianten Zernike-Momenten

Ordnung $n = 5$:

$$Z_{12} = |A_{55}|^2 \quad (\text{B.13})$$

$$Z_{13} = |A_{53}|^2 \quad (\text{B.14})$$

$$Z_{14} = |A_{51}|^2 \quad (\text{B.15})$$

$$Z_{15} = A_{51}^* A_{31} + [A_{51}^* A_{31}]^* \quad (\text{B.16})$$

$$Z_{16} = A_{53}^* A_{33} + [A_{53}^* A_{33}]^* \quad (\text{B.17})$$

$$Z_{17} = A_{55}^* A_{31}^5 + [A_{55}^* A_{31}^5]^* \quad (\text{B.18})$$

Ordnung $n = 6$:

$$Z_{18} = |A_{66}|^2 \quad (\text{B.19})$$

$$Z_{19} = |A_{64}|^2 \quad (\text{B.20})$$

$$Z_{20} = |A_{62}|^2 \quad (\text{B.21})$$

$$Z_{21} = A_{60} \quad (\text{B.22})$$

$$Z_{22} = A_{66}^* A_{33}^2 + [A_{66}^* A_{33}^2]^* \quad (\text{B.23})$$

$$Z_{23} = A_{64}^* A_{44} + [A_{64}^* A_{44}]^* \quad (\text{B.24})$$

$$Z_{24} = A_{62}^* A_{22} + [A_{62}^* A_{22}]^* \quad (\text{B.25})$$

Ordnung $n = 7$:

$$Z_{25} = |A_{77}|^2 \quad (\text{B.26})$$

$$Z_{26} = |A_{75}|^2 \quad (\text{B.27})$$

$$Z_{27} = |A_{73}|^2 \quad (\text{B.28})$$

$$Z_{28} = |A_{71}|^2 \quad (\text{B.29})$$

$$Z_{29} = A_{77}^* A_{31}^7 + [A_{77}^* A_{31}^7]^* \quad (\text{B.30})$$

$$Z_{30} = A_{75}^* A_{55} + [A_{75}^* A_{55}]^* \quad (\text{B.31})$$

$$Z_{31} = A_{73}^* A_{33} + [A_{73}^* A_{33}]^* \quad (\text{B.32})$$

$$Z_{32} = A_{71}^* A_{31} + [A_{71}^* A_{31}]^* \quad (\text{B.33})$$

B.2 Beispiele für invariante Zernike-Momente

Die folgenden Tabellen enthalten die Zahlenwerte der invarianten Zernike-Momente, die für einzelnen Bilder berechnet wurden. Eine Darstellung der zugehörigen Bilder und eine Diskussion dieser Werte ist in dem Abschnitt 4.2 zu finden.

B.2 Beispiele für invariante Zernike-Momente

B.2.1 Invariante Zernike-Momente bei transformierten Bildern

	Original	75% skaliert	120% skaliert	20 Pixel verschoben	5° gedreht	90° gedreht	Horizontal gespiegelt
Z_1	-0.59952	-0.61249	-0.59472	-0.60125	-0.60214	-0.60278	-0.59726
Z_2	0.011966	0.010132	0.016796	0.015919	0.012538	0.014328	0.014161
Z_3	0.0041852	0.0012515	0.0040296	0.0033913	0.0031952	0.0036742	0.0039464
Z_4	0.024576	0.024168	0.031983	0.021983	0.025323	0.022441	0.023232
Z_5	1.0025e-7	1.7099e-8	2.4074e-7	6.6889e-8	-2.7576e-8	-6.4392e-8	7.0704e-8
Z_6	1.4304e-5	1.1358e-5	3.3039e-5	1.4891e-5	1.5124e-5	1.4294e-5	1.5125e-5
Z_7	0.0019549	0.00094575	0.0011651	0.0004694	0.0013672	0.00073776	0.0017312
Z_8	0.026586	0.017387	0.040904	0.04089	0.027587	0.034331	0.034939
Z_9	0.22299	0.26782	0.21182	0.2287	0.23184	0.23353	0.21821
Z_{10}	2.7491e-6	4.4257e-7	1.9666e-6	-1.6649e-7	-2.0788e-6	-4.6424e-8	2.824e-6
Z_{11}	-0.00062987	-0.00034751	-0.0013731	-0.0013005	-0.00069088	-0.00098289	-0.00098786
Z_{12}	0.00065771	0.0017608	0.00097135	0.00113	0.0011302	0.00084847	0.0014652
Z_{13}	0.014588	0.0052244	0.01416	0.011951	0.010875	0.01356	0.015101
Z_{14}	0.091471	0.088683	0.11579	0.083107	0.093901	0.084695	0.086808
Z_{15}	-0.004496	-0.0042866	-0.0074068	-0.0036539	-0.0047557	-0.0038014	-0.0040334
Z_{16}	-0.00012185	-1.1416e-5	-0.00011268	-8.0714e-5	-6.9486e-5	-9.9446e-5	-0.00011913
Z_{17}	7.8002e-12	2.8159e-11	-1.4618e-12	5.8234e-12	1.1167e-11	9.6219e-12	-1.3263e-11
Z_{18}	0.0040284	0.0030652	0.0045095	0.0037138	0.003814	0.0035218	0.0040848
Z_{19}	0.0090708	0.004839	0.0052211	0.0021203	0.0067189	0.0029721	0.0082512
Z_{20}	0.017091	0.028846	0.015118	0.022443	0.0080575	0.0097748	0.015255
Z_{21}	0.37667	0.3087	0.38002	0.36694	0.36422	0.36142	0.37443
Z_{22}	-7.6631e-8	7.6245e-9	-3.7504e-8	-8.2043e-8	-3.2178e-8	-4.6782e-8	-2.1835e-9
Z_{23}	-3.5464e-5	-9.0175e-6	-1.2165e-5	-1.9793e-6	-1.8316e-5	-4.3796e-6	-2.8561e-5
Z_{24}	-9.193e-5	-0.00053104	0.00047256	0.00065696	-9.2846e-5	0.00021161	0.00035343
Z_{25}	0.00048874	0.00014606	0.00050321	0.00046577	0.0004762	0.00015225	0.00024847
Z_{26}	0.0046268	0.010917	0.0070527	0.0078412	0.0081751	0.0063207	0.0065615
Z_{27}	0.018929	0.014081	0.021339	0.015912	0.013458	0.020615	0.024452
Z_{28}	0.14256	0.13414	0.17054	0.13369	0.1451	0.13582	0.13674
Z_{29}	5.19e-15	1.236e-15	-2.5694e-14	-7.1773e-16	-6.258e-15	8.6027e-16	3.6062e-17
Z_{30}	-5.6247e-6	-3.8444e-5	-1.2411e-5	-1.7708e-5	-1.8388e-5	-1.0711e-5	-1.8812e-5
Z_{31}	0.00015241	1.6015e-5	0.00014559	0.00010137	8.6002e-5	0.00014715	0.00019175
Z_{32}	0.0070063	0.0064836	0.010905	0.0058778	0.0073474	0.0060958	0.0063536

Tabelle B.1: Invariante Zernike-Momente des Bundesadler-Referenzbildes und dessen transformierte Versionen in Abbildung 4.2

B.2.2 Invariante Zernike-Momente bei Bildern mit gleichem Objekt

	Bild 4.4(a)	Bild 4.4(b)	Bild 4.4(c)	Bild 4.4(d)	Bild 4.4(e)
Z_1	-0.59952	-0.55353	-0.56883	-0.613	-0.55248
Z_2	0.011966	0.011975	0.0059083	0.0051717	0.0046129
Z_3	0.0041852	0.00099895	0.0028995	0.0090034	0.0030704
Z_4	0.024576	0.032341	0.028974	0.011754	0.021066
Z_5	1.0025e-007	-5.5973e-008	-1.3993e-007	1.6655e-008	-2.5938e-008
Z_6	1.4304e-005	2.4729e-005	-9.8803e-006	-6.2154e-007	2.451e-006
Z_7	0.0019549	0.011517	0.0033327	0.0045918	0.0032153
Z_8	0.026586	0.014692	0.036476	0.022963	0.01127
Z_9	0.22299	0.11592	0.18102	0.27008	0.15366
Z_{10}	2.7491e-006	-4.9663e-006	8.5849e-006	1.6714e-006	-3.7043e-007
Z_{11}	-0.00062987	-0.00034548	-0.00042699	-0.00023309	2.1288e-005
Z_{12}	0.00065771	0.006469	0.0015894	0.0010406	0.0014656
Z_{13}	0.014588	0.00085235	0.014354	0.033934	0.013602
Z_{14}	0.091471	0.10189	0.090239	0.041176	0.063781
Z_{15}	-0.004496	-0.0065901	-0.005229	-0.00096595	-0.0026872
Z_{16}	-0.00012185	-1.4681e-006	-8.3165e-005	-0.00061027	-8.239e-005
Z_{17}	7.8002e-012	-4.043e-010	6.1754e-011	-4.4237e-013	-6.3763e-012
Z_{18}	0.0040284	0.0074506	0.0067362	0.0031035	0.0072686
Z_{19}	0.0090708	0.047324	0.016557	0.023383	0.016499
Z_{20}	0.017091	0.04679	0.091436	0.054241	0.056783
Z_{21}	0.37667	0.4153	0.28168	0.29554	0.25663
Z_{22}	-7.6631e-008	3.4519e-009	-1.0574e-007	-4.3994e-007	-4.7751e-008
Z_{23}	-3.5464e-005	-0.0010901	-0.00011025	-0.00021473	-0.00010415
Z_{24}	-9.193e-005	-0.0010901	0.001057	0.00051943	-0.00038172
Z_{25}	0.00048874	0.0024943	0.00031227	0.0010395	0.00097551
Z_{26}	0.0046268	0.024829	0.0109	0.0057178	0.007324
Z_{27}	0.018929	0.0080614	0.032086	0.055036	0.02776
Z_{28}	0.14256	0.11328	0.093735	0.056092	0.062042
Z_{29}	5.19e-015	-1.607e-013	-9.4796e-015	6.3471e-017	-7.3319e-016
Z_{30}	-5.6247e-006	-0.00032114	-3.4626e-005	-1.1896e-005	-2.0591e-005
Z_{31}	0.00015241	-1.6054e-005	0.00018515	0.00097802	0.00015921
Z_{32}	0.0070063	0.0073251	0.0054294	0.001269	0.0026139

Tabelle B.2: Invariante Zernike-Momente für die verschiedenen Bilder mit dem Bundesadler in Abbildung 4.4

B.2 Beispiele für invariante Zernike-Momente

Der Klassifikator für die Objekte UNO-Flagge wurde von der detaillierten Betrachtung ausgenommen, da nicht genügend Testbilder zusammen getragen werden konnten. An dieser Stelle soll der Vollständigkeit halber die invarianten Zernike-Momente auch für einige Bilder mit der UNO-Flagge aufgeführt werden (Tabelle B.3, Abbildung B.1). Es zeigt sich in Abbildung B.2 die selbe Charakteristik wie bei den anderen Objekten: Die invarianten Zernike-Momente weisen eine große Ähnlichkeit auf.

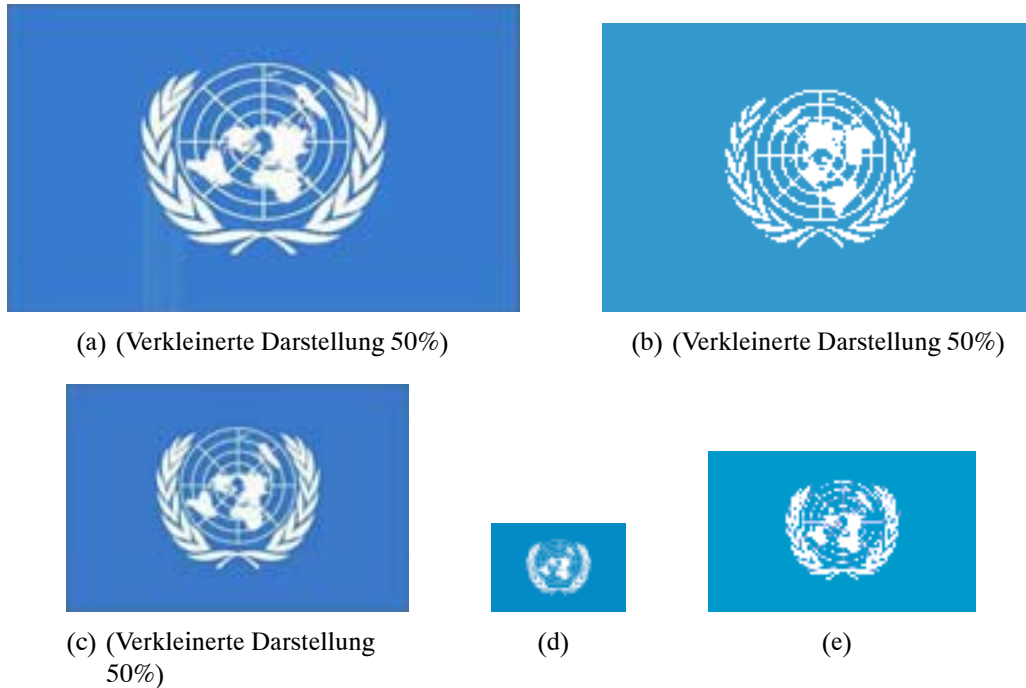


Abbildung B.1: Verschiedene Bilder mit dem Objekt UNO-Emblem

Die Unterschiede der betrachteten fünf Bilder liegen vor allem bei beträchtlichen Größenunterschieden (zwischen 384x230 und 50x33 Bildpunkte). Trotzdem sind die Abstände zwischen den Merkmalvektoren sehr klein.

B Ergänzungen zu den invarianten Zernike-Momenten

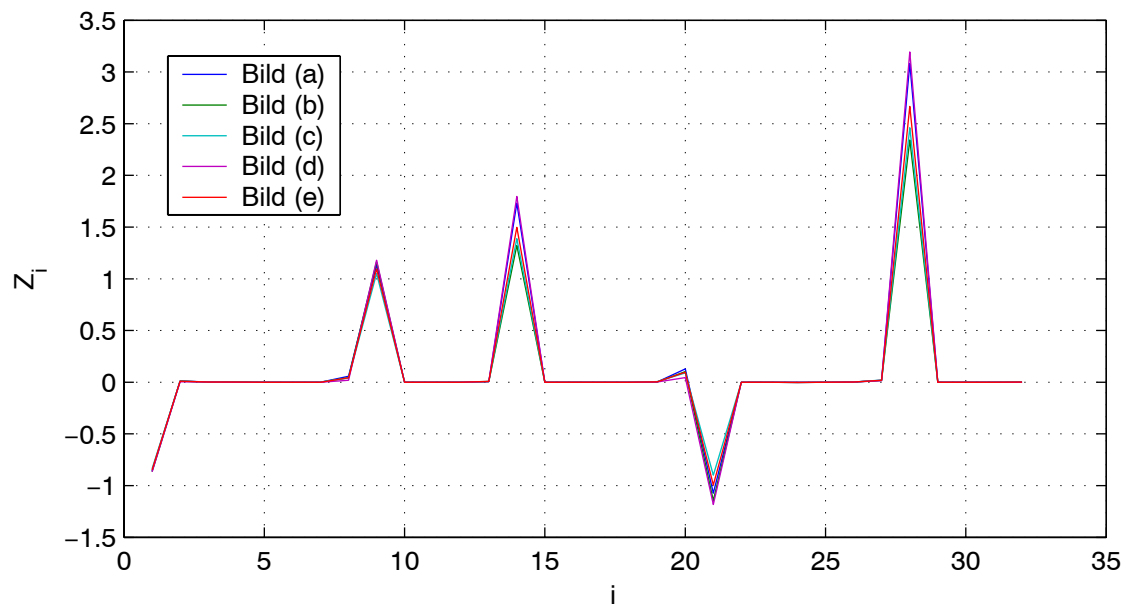


Abbildung B.2: Invariante Zernike-Momente der Bilder mit der UNO-Flagge in Abbildung B.1

B.2 Beispiele für invariante Zernike-Momente

	Bild B.1(a)	Bild B.1(b)	Bild B.1(c)	Bild B.1(d)	Bild B.1(e)
Z_1	-0.85581	-0.86398	-0.83707	-0.86704	-0.84779
Z_2	0.012926	0.01005	0.0096466	0.0044911	0.0099697
Z_3	0.0012861	0.0011632	0.0016354	0.00093188	0.001413
Z_4	0.00047508	0.0011225	0.0016834	5.467e-005	0.00099468
Z_5	2.4677e-013	-9.0332e-013	-7.8686e-012	-2.1279e-016	1.3062e-012
Z_6	-5.4943e-009	-1.7268e-008	-5.4619e-008	-3.8551e-012	-1.5338e-008
Z_7	0.00021567	0.00023992	0.00089421	0.00034789	0.00042791
Z_8	0.056374	0.044588	0.04099	0.01955	0.043292
Z_9	1.1311	1.1651	1.051	1.1798	1.0959
Z_{10}	-1.3559e-006	-9.5255e-007	-2.3596e-006	-2.1638e-007	-1.6034e-006
Z_{11}	-0.0014573	-0.00089624	-0.00079079	-0.00017559	-0.00086321
Z_{12}	0.0001758	0.00018174	0.00030174	0.00011148	0.00020694
Z_{13}	0.0067322	0.0060509	0.008237	0.004884	0.00728
Z_{14}	1.7324	1.3208	1.3876	1.7955	1.4975
Z_{15}	0.00023288	-0.0012665	-0.0008476	0.00013066	-0.0011028
Z_{16}	3.346e-007	-1.0807e-006	-2.2513e-005	-9.6018e-007	-3.2325e-008
Z_{17}	6.1622e-021	-5.5512e-019	-4.6212e-018	-4.5938e-026	-3.9315e-019
Z_{18}	4.9207e-005	3.1992e-005	0.00011992	4.9837e-005	6.1567e-005
Z_{19}	0.0013623	0.001545	0.0054574	0.0021451	0.0025909
Z_{20}	0.12741	0.1036	0.088726	0.044021	0.097157
Z_{21}	-1.0732	-1.1435	-0.89821	-1.1822	-0.99296
Z_{22}	-1.6169e-010	-8.5794e-011	-4.4739e-010	-8.5539e-011	-2.4448e-010
Z_{23}	-5.7784e-007	-7.407e-007	-9.7592e-006	-1.0331e-006	-2.1999e-006
Z_{24}	-0.0032876	-0.0020809	-0.0013409	-0.0003842	-0.00193
Z_{25}	1.2979e-005	5.5094e-006	4.1085e-005	1.2219e-005	1.2849e-005
Z_{26}	0.0011987	0.0013089	0.0021168	0.00079938	0.0014375
Z_{27}	0.018422	0.016359	0.021248	0.013398	0.01944
Z_{28}	3.0829	2.3425	2.4615	3.1929	2.668
Z_{29}	-1.3405e-028	4.0875e-027	3.1235e-024	2.3254e-035	1.8102e-026
Z_{30}	-4.2147e-007	-4.7577e-007	-1.2774e-006	-1.7822e-007	-5.9494e-007
Z_{31}	-1.6425e-006	2.9293e-006	5.8511e-005	2.8271e-006	-1.3045e-007
Z_{32}	-0.00041425	0.0022406	0.0014951	-0.0002324	0.0019685

Tabelle B.3: Invariante Zernike-Momente der Bilder mit der UNO-Flagge in Abbildung B.1

B.2.3 Invariante Zernike-Momente bei Bildern mit unterschiedlichen Objekten

	Bundesadler	Coca-Cola-Logo	China-Flagge	Mercedesstern	UNO-Flagge
Z_1	-0.46037	-0.54179	-0.88797	-0.55419	-0.86398
Z_2	0.0053985	0.021379	0.013477	0.010542	0.01005
Z_3	0.0065135	0.010769	0.0030344	0.0035689	0.0011632
Z_4	0.043438	0.0088738	0.0010807	0.0057819	0.0011225
Z_5	1.041e-006	-1.3177e-008	7.257e-012	3.5719e-010	-9.0332e-013
Z_6	1.2724e-005	1.5639e-006	-1.8287e-008	-5.6866e-007	-1.7268e-008
Z_7	0.0088296	0.0080343	0.00028854	0.0056742	0.00023992
Z_8	0.018716	0.062687	0.062826	0.032499	0.044588
Z_9	-0.083787	0.19427	1.2711	0.13072	1.1651
Z_{10}	-2.008e-006	-5.1066e-005	2.2124e-006	1.6205e-006	-9.5255e-007
Z_{11}	2.1812e-005	-0.0026286	-0.0016934	-0.00068493	-0.00089624
Z_{12}	0.0022249	0.0010688	7.1729e-005	0.0021021	0.00018174
Z_{13}	0.01893	0.03338	0.017004	0.025538	0.0060509
Z_{14}	0.23766	0.026739	3.2676	0.75879	1.3208
Z_{15}	0.0061509	0.00045837	-8.1172e-005	-0.0062032	-0.0012665
Z_{16}	0.0002108	-0.00012782	2.0831e-005	-4.7667e-005	-1.0807e-006
Z_{17}	6.6163e-010	9.6638e-014	1.2991e-019	1.3677e-015	-5.5512e-019
Z_{18}	0.016857	0.0018932	2.6529e-005	0.00091376	3.1992e-005
Z_{19}	0.028394	0.049854	0.0018137	0.024784	0.001545
Z_{20}	0.073982	0.084093	0.15809	0.036961	0.1036
Z_{21}	0.45881	0.02708	-1.3878	0.30477	-1.1435
Z_{22}	9.4559e-007	3.0022e-007	-2.3369e-010	-2.1291e-008	-8.5794e-011
Z_{23}	-0.00044124	-0.00077856	-9.2867e-007	-0.00023698	-7.407e-007
Z_{24}	0.00068481	0.0014907	-0.0037767	-0.00010095	-0.0020809
Z_{25}	0.0022417	0.0018632	1.0099e-005	0.00041029	5.5094e-006
Z_{26}	0.013485	0.0049691	0.00051229	0.01091	0.0013089
Z_{27}	0.01397	0.03247	0.051454	0.086947	0.016359
Z_{28}	0.30984	0.046315	5.8134	1.3528	2.3425
Z_{29}	-9.812e-013	-5.3823e-018	-1.3081e-026	1.6519e-019	4.0875e-027
Z_{30}	-4.0092e-005	-8.6786e-006	-7.3474e-008	-4.5867e-005	-4.7577e-007
Z_{31}	-0.00018122	0.00019079	-6.2307e-005	0.00014435	2.9293e-006
Z_{32}	-0.0094934	-0.00076045	0.00013317	0.011037	0.0022406

Tabelle B.4: Invariante Zernike-Momente der Bilder mit unterschiedlichen Objekten in Abbildung 4.1

C Verwendetes Bildmaterial

Im Folgenden sind die aus dem Internet heruntergeladenen Bilder aufgelistet, die als Testdatenbank der verschiedenen Klassifikatoren verwendet wurden. Die Auflistung enthält alle Test- und Referenzbilder, sowie vereinzelt weitere Bilder der jeweiligen Objektklasse.

C.1 Bundesadler

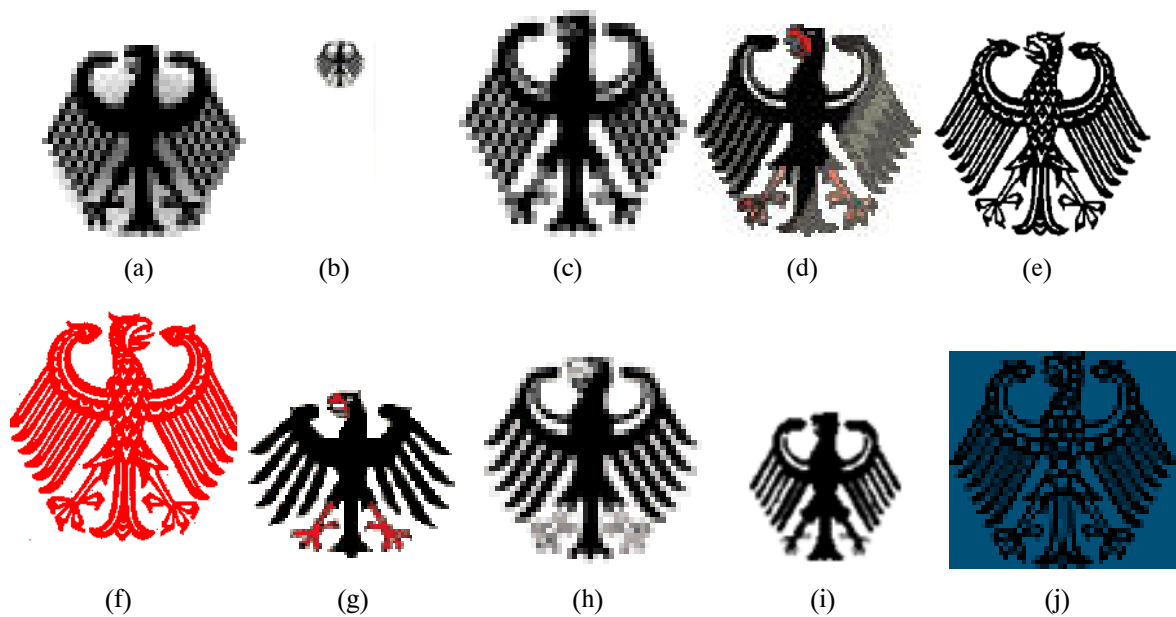


Abbildung C.1: Bilder der Bundesadler-Datenbank

C Verwendetes Bildmaterial



Abbildung C.2: Bilder der Bundesadler-Datenbank (Fortsetzung)

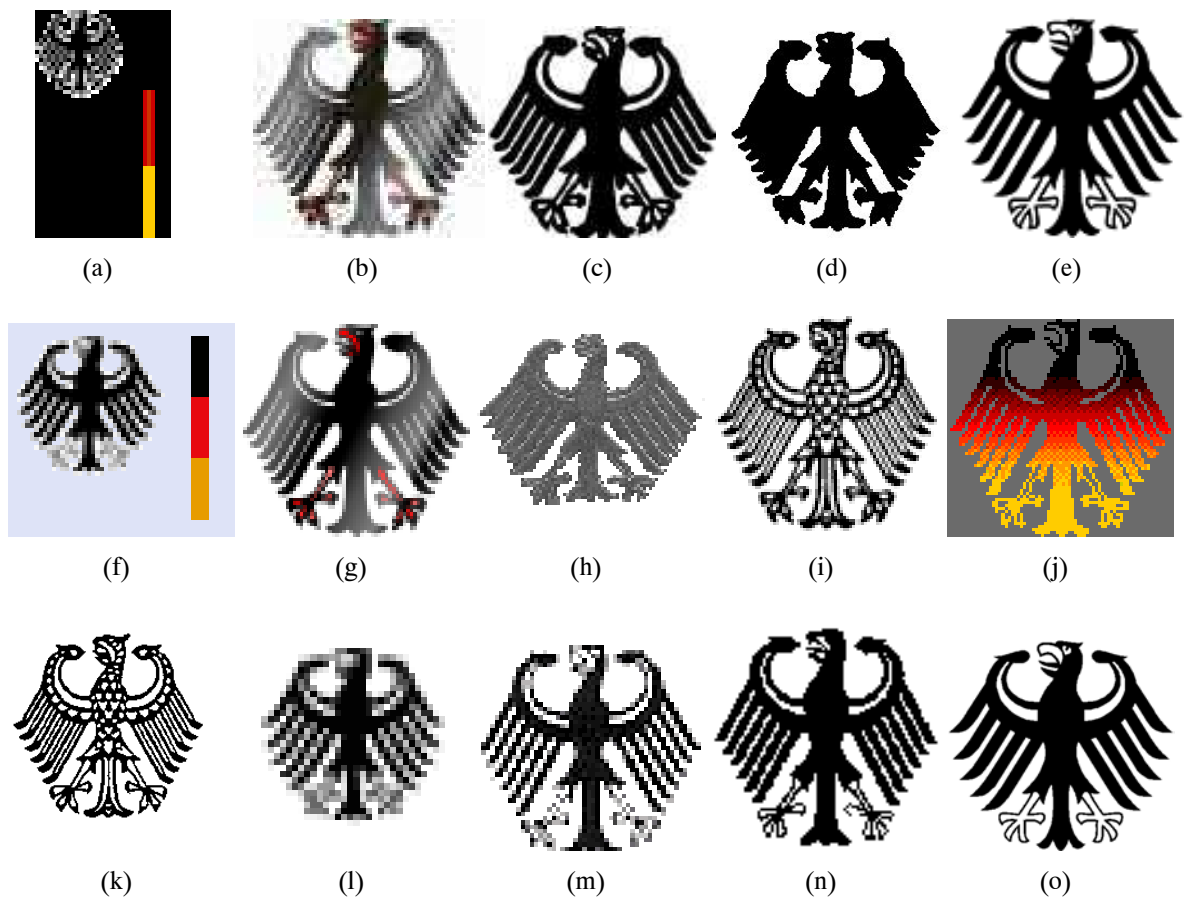


Abbildung C.3: Bilder der Bundesadler-Datenbank (Fortsetzung)

C.2 Coca-Cola-Logos



Abbildung C.4: Bilder der Coca-Cola-Datenbank



Abbildung C.5: Bilder der Coca-Cola-Datenbank (Fortsetzung)

C Verwendetes Bildmaterial



Abbildung C.6: Bilder der Coca-Cola-Datenbank (Fortsetzung)

C.3 China-Flaggen

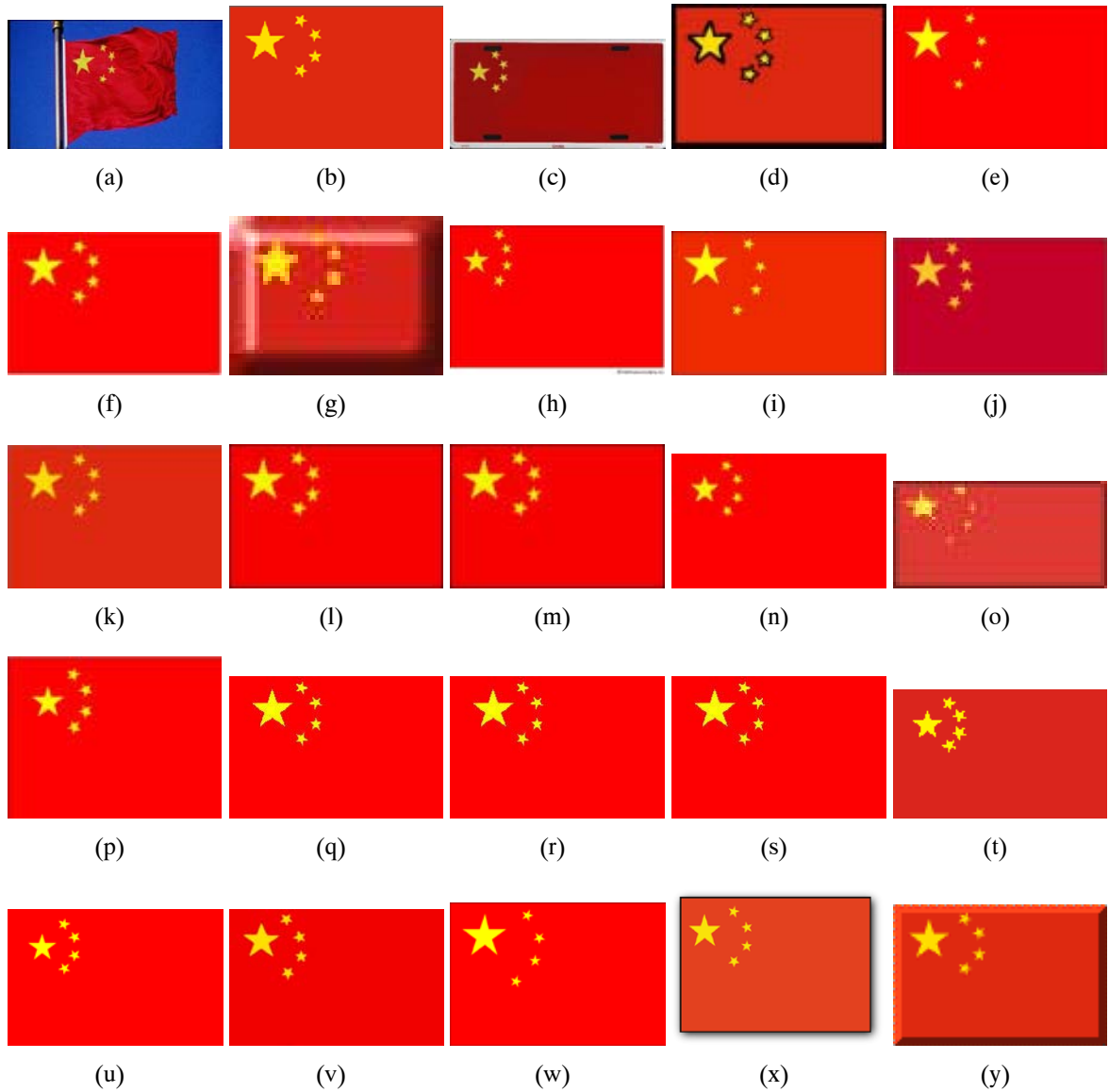


Abbildung C.7: Bilder der China-Flaggen-Datenbank

C Verwendetes Bildmaterial

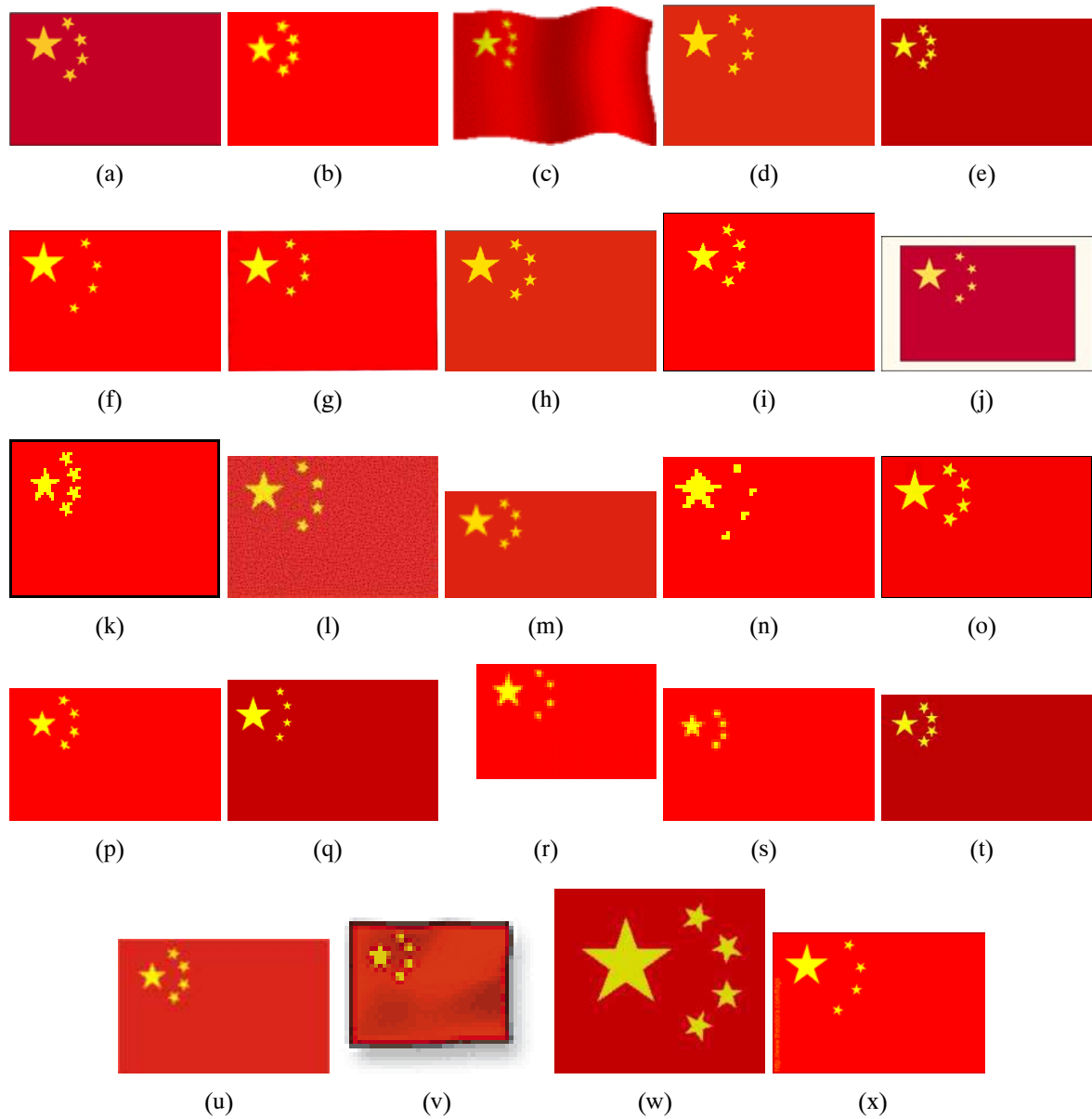


Abbildung C.8: Bilder der China-Flaggen-Datenbank (Fortsetzung)

C.4 Mercedessterne



Abbildung C.9: Bilder der Mercedesstern-Datenbank

C Verwendetes Bildmaterial

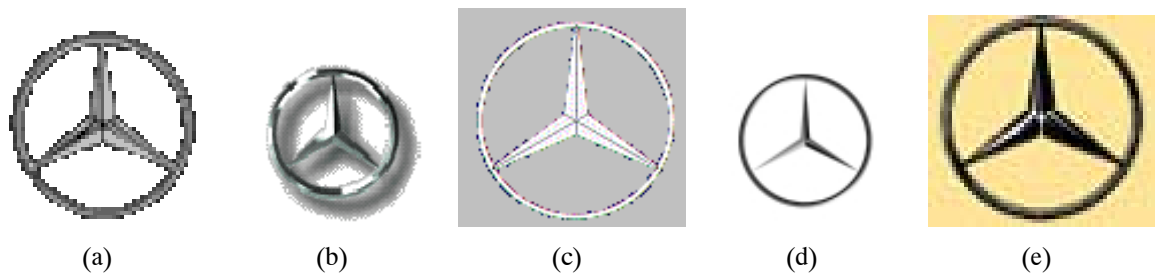


Abbildung C.10: Bilder der Mercedesstern-Datenbank (Fortsetzung)

C.5 UNO-Flaggen

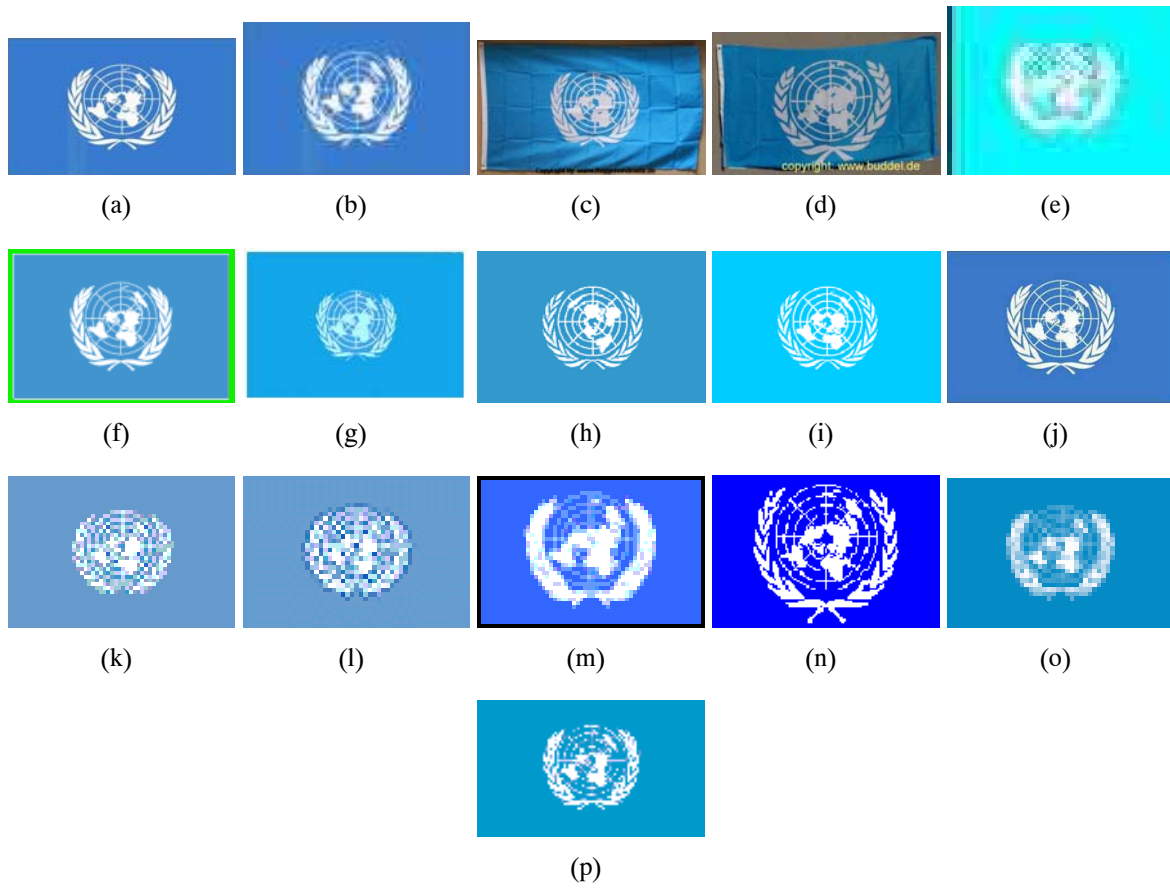


Abbildung C.11: Bilder der UNO-Flaggen-Datenbank

C Verwendetes Bildmaterial

Literaturverzeichnis

- [ARY02] AL-RAWI, M. und J. YANG: *Practical Fast Computation of Zernike Moments*. Journal of Computer Science and Technology, 17(2), März 2002.
- [BKB99] BOUET, M., A. KHENCHAF und H. BRIAND: *Shape Representation for Image Retrieval*. In: *Association for Computing Machinery (ACM) Electronic Proceedings*, 1999.
- [BSA91] BELKASIM, S.O., M. SHRIDHAR und M. AHMADI: *Pattern Recognition with Moment Invariants: A Comparative Study and New Results*. Pattern Recognition, 24(12):1117–1138, 1991.
- [CH87] CASH, G.L. und M. HATAMIAN: *Optical Character Recognition by the Method of Moments*. Computer Vision, Graphics, and Image Processing, 39(3):291–310, 1987.
- [CW02] CHEN, Y. und J.Z. WANG: *A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(9), September 2002.
- [DAV98] DO, M.N., S. AYER und M. VETTERLI: *Invariant Image Retrieval Using Wavelet Maxima Moment*. SSC Doctoral School Project Report, Ecole Polytechnique Fédérale de Lausanne, Schweiz, 1998.
- [DBM77] DUDANI, S., K. BREEDING und R. MCGHEE: *Aircraft Identification by Moment Invariants*. IEEE Transactions on Computers, 26:39–45, 1977.
- [EG99] EAKINS, J.P. und M.E. GRAHAM: *Content-Based Image Retrieval*. Report to the JISC Technology Applications Programme, Institute for Image Data Research, University of Northumbria, Newcastle, Großbritannien, 1999.
- [Flu98] FLUSSER, J.: *Fast Calculation of Geometric Moments of Binary Images*. In: *Pattern Recognition and Medical Computer Vision*, Seiten 265–274. ÖCG, Illmitz, Österreich, 1998.
- [Flu00] FLUSSER, JAN: *On the Independence of Rotation Moment Invariants*. Pattern Recognition, 33(9):1405–1410, 2000.

Literaturverzeichnis

- [FS93] FLUSSER, J. und T. SUK: *Pattern Recognition by Affine Moment Invariants*. Pattern Recognition ICPR94, 26(1):167–174, 1993.
- [GW93] GONZALEZ, R.C. und R.E. WOODS: *Digital Image Processing*. Addison-Wesley Publishing Company, 1993.
- [Hab95] HABERÄCKER, P.: *Praxis der Digitalen Bildverarbeitung und Mustererkennung*. Carl Hanser Verlag München, 1995.
- [Hat86] HATAMIAN, M.: *A Real-Time Two-Dimensional Moment Generating Algorithm and Its Single Chip Implementation*. IEEE Transactions on Acoustics, Speech and Signal Processing, 34(3):546–553, 1986.
- [Hew96] HEW, P.C.: *Zernike Moments – Problems*. Technischer Bericht, Department of Mathematics, The University of Western Australia, 1996.
- [Hu62] HU, M.-K.: *Visual Pattern Recognition by Moment Invariants*. IRE Transactions on Information Theory, 8(2):179–187, 1962.
- [KH99] KHOTANZAD, A. und Y.H. HONG: *Invariant Image Recognition by Zernike Moments*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(5), 1999.
- [Kro96] KROSCHER, K.: *Statistische Nachrichtentheorie: Signal- und Mustererkennung, Parameter- und Signalschätzung*. Springer-Verlag Heidelberg, 1996.
- [Leu91] LEU, J.G.: *Computing A Shape's Moments From Its Boundary*. Pattern Recognition, 24(10):949–957, 1991.
- [MPA98] MANDAL, M.K., S. PANCHANATHAN und T. ABOULNASR: *Illumination Invariant Image Indexing Using Moments and Wavelets*. Journal of Electronic Imaging, April 1998.
- [MR95] MUKUNDAN, R. und K.R. RAMAKRISHNAN: *Fast Computation of Legendre and Zernike Moments*. Pattern Recognition, 28(9):1433–1442, 1995.
- [MR98] MUKUNDAN, R. und K.R. RAMAKRISHNAN: *Moment Functions in Image Analysis: Theory and Application*. World Scientific Publishing Co. Pte. Ltd. Singapore, 1998.
- [Rau94] RAUBER, T.W.: *Inductive Pattern Classification: Methods - Features - Sensors*. Dissertation, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal, 1994.
- [Rei91] REISS, T.H.: *The Revised Fundamental Theorem of Moment Invariants*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(8):830–834, 1991.

- [RHC99] RUI, Y., T. HUANG und S. CHANG: *Image Retrieval: Current Techniques, Promising Directions and Open Issues*. Journal of Visual Communication and Image Representation, 10(4):39–62, 1999.
- [Sch77] SCHWARTZ, E.L.: *Spatial Mapping in Primate Sensory Projection: Analytic Structure and Relevance to Perception*. Biological Cybernetics, 25:181–194, 1977.
- [TC88] TEH, C.-H. und R.T. CHIN: *On Image Analysis by the Methods of Moments*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(4):496 – 513, 1988.
- [Tea80] TEAGUE, M.R.: *Image Analysis via the General Theory of Moments*. Journal of the Optical Society of America, 70(8):920–930, 1980.
- [TMRSSG00] TORRES-MÉNDEZ, L.A., J.C. RUIZ-SUÁREZ, L.E. SUCAR und G. GÓMEZ: *Translation, Rotation, and Scale-Invariant Object Recognition*. IEEE Transactions on System, Man, and Cybernetics, Part C: Applications and Reviews, 30(1):125–130, 2000.
- [VC01] VENTERS, C.C. und M. COOPER: *A Review of Content-Based Image Retrieval Systems*. Report to the JISC Technology Applications Programme, Manchester Visualization Centre, University of Manchester, Großbritannien, 2001.
- [WSZ02] WANG, W., Y. SONG und A. ZHANG: *Semantics Retrieval by Content and Context of Image Regions*. International Conference on Vision Interface, Calgary, Canada, Mai 2002.
- [YA94] YANG, L. und F. ALBREGTSEN: *Fast Computation of Invariant Geometric Moments: a New Method Giving Correct Results*. In: *International Conference on Pattern Recognition*, Band 1, Seiten 201–204, 1994.
- [ZH02] ZHOU, X.S. und T.S. HUANG: *Unifying Keywords and Visual Contents in Image Retrieval*. IEEE MultiMedia, 9(2):23–33, 2002.

