

UNIVERSITÄT LEIPZIG
FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
Institut für Informatik

**Momentbasierte Methoden
zur Schriftzeichenerkennung**

Diplomarbeit

Frank-Michael Schleif

Leipzig, den 26. Juli 2002

Kurzzusammenfassung

Die Ermittlung von diskriminanten Merkmalen aus Schriftzeichen zum Zweck der anschließenden Klassifikation stellt ein schwieriges Gebiet der optischen Schriftzeichenerkennung (OCR) dar.

In dieser Diplomarbeit werden verschiedene merkmalsgenerierende, statistische Momente analysiert und deren Eignung für diesen Anwendungsfall vergleichend untersucht. Die effiziente Berechnung günstiger Merkmale und eine geeignete Parametrisierung der verschiedenen Zwischenschritte während ihrer Erfassung und Weiterverarbeitung stehen ebenso im Mittelpunkt. Es werden aber auch Aspekte der Merkmalsreduktion berücksichtigt.

Es zeigte sich, daß alle betrachteten Verfahren eine prinzipielle Eignung aufweisen, im Detail aber teilweise deutliche Unterschiede, vor allem in der Diskriminanzfähigkeit und damit späteren Anwendbarkeit für den Klassifikationsprozeß, auftraten.

Am Beispiel der Klassifikation von handschriftlichen, numerischen Symbolen werden die verschiedenen Verfahren getestet und Empfehlungen für eine geeignete Merkmalsextraktion gegeben.

Inhaltsverzeichnis

1	Einführung	1
2	Vorverarbeitung	4
2.1	Rauschreduktion	4
2.1.1	Morphologische Operationen	5
2.1.2	Schwellwertverfahren	5
2.2	Normalisierung und Normierung	5
3	Verfahren der Merkmalsgewinnung	7
3.1	Allgemeine Prinzipien der Merkmalsgewinnung	7
3.2	Invariante Merkmale	8
3.3	Strukturelle Verfahren der Merkmalsgewinnung	10
3.3.1	Template-Matching	10
3.3.2	Unitäre Transformationen	11
3.4	Histogrammprojektionen	11
3.5	Statistische Verfahren der Merkmalsgewinnung	12
3.5.1	Statistische Momente	12
3.5.2	Zeitreihenexpansionen	12
3.5.3	Neural Network Classifiers	13
3.6	Verwendete Verfahren	13
3.7	Trainingsdatengewinnung	15
4	Momente - eine Einführung	16
4.1	Momente und algebraische Invarianten	16
4.1.1	Algebraische Formen und Invarianten	17
4.1.2	Erweiterung des Fundamentaltheorems auf n-Dimensionen	18
4.2	Momente einer Fläche	20
4.3	Invarianz von Momenten	21
4.4	Momentinvarianten von Hu und Reiss	26
4.4.1	Momentinvarianten von Hu	26
4.4.2	Affine Momentinvarianten von Reiss	27
4.5	Gewichtete zentrale Momente	29
4.6	Tsirikolias-Mertzios Momente	29

5	Orthogonale Momente	31
5.1	Legendre-Momente	31
5.2	Zernike-Momente	33
5.2.1	Skalierungsinvarianz	36
5.2.2	Rotationsinvarianz	37
5.2.3	Konstruktion von Invarianten mit Zernike-Momenten	38
5.3	Pseudo-Zernike-Momente	38
5.4	Orthogonale Fourier-Mellin-Momente	39
5.5	Wavelet Momente	39
6	Schnelle und effiziente Berechnung von Momenten	42
6.1	Greens-Theorem nach Philips	43
6.2	Momentberechnung nach Flusser	44
6.3	Momentberechnung nach Mertzios	45
6.4	Einschränkungen	48
6.4.1	Numerische- und Berechnungsprobleme	48
6.4.2	Informationstheoretische Einschränkungen	48
7	Klassifikation	50
7.1	Überwachte Klassifikation	51
7.2	Unüberwachte Klassifikation	52
7.3	Merkmalsselektion	53
7.4	Verwendete Klassifikatoren	54
7.4.1	Linearer Klassifikator mit euklidischer Distanz	54
7.4.2	Kreuzkorrelation	57
7.4.3	Diskriminanzkostenfunktion (DKF)	57
7.4.4	Neuronale Netze als Klassifikator	59
8	Experimente	62
8.1	Untersuchungen der Momente von Hu	63
8.1.1	Allgemeine statistische Daten	63
8.1.2	Diskriminanzanalyse	63
8.1.3	Anwendung verschiedener Normierungen und Klassifikatoren	64
8.2	Untersuchungen zu geometrischen Momenten	69
8.2.1	Untersuchungen zu Tsirikolias-Mertzios-Momenten	70
8.2.2	Untersuchungen zu normalisierten Momenten	74
8.3	Untersuchungen der orthogonalen Momente	79
8.3.1	Untersuchungen zu Legendre-Momenten	79
8.3.2	Untersuchungen zu Orthogonal Fourier-Mellin-Momenten	82
8.3.3	Untersuchungen zu Pseudo-Zernike-Momenten	85
8.3.4	Untersuchungen zu Zernike-Momenten	88
8.3.5	Untersuchungen zu Wavelet-Momenten	92
8.4	Klassifikation mit kombinierten Momenten	96

8.5	Zusammenfassung	99
8.5.1	Normalisierte vs. Tsirikolias-Mertzios-Momente	99
8.5.2	Orthogonale Momente	99
8.5.3	Wie viele Merkmale sind nötig ?	100
8.5.4	Vergleichbarkeit mit anderen Arbeiten	101
8.5.5	Aufwand allgemein - Empfehlungen	103
9	Zusammenfassung	105
A	Deskriptive Statistik zu verschiedenen Momenttypen	107
	Literatur	121

Abbildungsverzeichnis

1.1	Vereinfachtes Schema der OCR [63]	2
3.1	Histogrammprojektionen zum Symbol 3.2(f)	11
3.2	Beispielsymbole aus der verwendeten NIST Datenbank [45]	15
4.1	Darstellung der Auswirkung einer fehlerhaften Boundingbox	24
5.1	Bilder abgeleitet von Zernike Momenten. <i>Zeilen 1-2:</i> Eingabebild der Ziffer '4', und Beträge der Zernike-Momente $I(x, y)$ von Ordnung 1-13. Die Bilder wurden bzgl. Histogramm korrigiert. <i>Zeilen 3-4:</i> Eingabebild der Ziffer '4' und Bildrekonstruktion durch die Zernike-Momente bis zur 13. Ordnung	35
5.2	Bilder abgeleitet von Zernike Momenten. <i>Zeilen 1-2:</i> Eingabebild der Ziffer '5', und Beträge der Zernike-Momente $I(x, y)$ von Ordnung 1-13. Die Bilder wurden bzgl. Histogramm korrigiert. <i>Zeilen 3-4:</i> Eingabebild der Ziffer '5' und Bildrekonstruktion durch die Zernike-Momente bis zur 13. Ordnung	35
6.1	Symbol - links original und rechts mit Ω_- (rot) und Ω_+ (blau)	45
7.1	Schematische Darstellung eines MLP	59
7.2	Schema eines Neurons	60
7.3	Logistische Sigmoiden mit $a = 1$	60
8.1	Scatterplot der Merkmale Hu3 und Hu5	64
8.2	Kanonische Analyse der Merkmale von Hu	64
8.3	Analyse des Einflusses von ϵ bei k -NN und Momenten von Hu	67
8.4	Analyse des Einflusses von k bei k NN und Momenten von Hu mit Normierung 2 und $\epsilon = 2.0$	68
8.5	konkrete Analyse des Einflusses von k nächsten Nachbarn (über $k = 15, \epsilon = 2.0$)	69
8.6	Kanonische Analyse der 40 diskriminantesten TM-Momente für DS1 und DS2	73
8.7	Klassifikationsrate mit TM-Momenten für DS1 und DS2	73

8.8	Kanonische Analyse der 40 diskriminantesten NM-Momente für DS1 und DS2	77
8.9	Klassifikationsrate mit NM-Momenten für DS1 und DS2	77
8.10	Kanonische Analyse der 40 diskriminantesten Legendre-Momente für DS1 und DS2	81
8.11	Klassifikationsrate mit Legendre-Momenten für DS1 und DS2	81
8.12	Kanonische Analyse der 40 diskriminantesten OFM-Momente für DS1 und DS2	84
8.13	Klassifikationsrate mit Orthogonal-Fourier-Mellin-Momenten für DS1 und DS2	84
8.14	Kanonische Analyse der 40 diskriminantesten Pseudo-Zernike-Momente für DS1 und DS2	87
8.15	Klassifikationsrate mit Pseudo-Zernike-Momenten für DS1 und DS2	87
8.16	Kanonische Analyse der 40 diskriminantesten Zernike-Momente für DS1 und DS2	90
8.17	Klassifikationsrate mit Zernike-Momenten für DS1 und DS2	90
8.18	Kanonische Analyse der 40 diskriminantesten Wavelet-Momente für DS1 und DS2	95
8.19	Klassifikationsrate mit Wavelet-Momenten für DS1 und DS2	95
8.20	Kanonische Analyse der 40 diskriminantesten kombinierten moment-basierten Merkmale für DS1 und DS2	97
8.21	Klassifikationsrate bei Verwendung von kombinierten Merkmalen für DS1 und DS2	97

Tabellenverzeichnis

3.1	Merkmalsextraktions- und Projektionsmethoden	9
7.1	Merkmalsselektionsmethoden	55
8.1	Diskriminanzwerte zu Momenten von Hu	65
8.2	Erkennungsraten mit Hu-Momenten bei unterschiedlichen Klassifikatoren - unnormiert	65
8.3	Erkennungsraten mit Hu-Momenten bei unterschiedlichen Klassifikatoren und Normierungen	66
8.4	Einfluß der Translation, Skalierung oder Rotation auf Tsirikolias-Mertzios-Momente	71
8.5	Diskriminanzwerte zu Tsirikolias-Mertzios-Momenten	71
8.6	Klassifikationsergebnisse für DS1 mit Tsirikolias-Mertzios-Momenten	72
8.7	Klassifikationsergebnisse für DS2 mit Tsirikolias-Mertzios-Momenten	72
8.8	Validität der k NN Klassifikation bei Nutzung von Tsirikolias-Mertzios-Momenten	74
8.9	Maximale Ordnung der Parameter p und q für Tsirikolias-Mertzios-Momente	74
8.10	Einfluß der Translation, Skalierung oder Rotation auf normalisierte Momente	75
8.11	Diskriminanzwerte zu normalisierten Momenten	75
8.12	Klassifikationsergebnisse für DS1 mit normalisierten Momenten . . .	75
8.13	Klassifikationsergebnisse für DS2 mit normalisierten Momenten . . .	76
8.14	Maximale Ordnung der Parameter p und q für normalisierte Momente	76
8.15	Validität der k NN-Klassifikation bei Nutzung von normalisierten Momenten	76
8.16	Einfluß der Translation, Skalierung oder Rotation auf Legendre-Momente	79
8.17	Diskriminanzwerte zu Legendre-Momenten	80
8.18	Klassifikationsergebnisse für DS1 mit Legendre-Momenten	80
8.19	Klassifikationsergebnisse für DS2 mit Legendre-Momenten	80
8.20	Maximale Ordnung der Parameter p und q für Legendre-Momente . .	82
8.21	Validität der k NN Klassifikation bei Nutzung von Legendre-Momenten	82
8.22	Einfluß der Translation, Skalierung oder Rotation auf OFM-Momente	83
8.23	Diskriminanzwerte zu OFM-Momenten	83

8.24	Klassifikationsergebnisse für DS1 mit OFM-Momenten	83
8.25	Klassifikationsergebnisse für DS2 mit OFM-Momenten	85
8.26	Maximale Ordnung der Parameter p und q für OFM-Momente	85
8.27	Validität der k NN-Klassifikation bei Nutzung von OFM-Momente	85
8.28	Einfluß der Translation, Skalierung oder Rotation auf Pseudo-Zernike-Momente	86
8.29	Diskriminanzwerte zu Pseudo-Zernike-Momenten	86
8.30	Klassifikationsergebnisse für DS1 mit Pseudo-Zernike-Momenten	86
8.31	Klassifikationsergebnisse für DS2 mit Pseudo-Zernike-Momenten	88
8.32	Validität der k NN-Klassifikation bei Nutzung von Pseudo-Zernike-Momenten	88
8.33	Maximale Ordnung der Parameter p und q für Pseudo-Zernike-Momente	88
8.34	Einfluß der Translation, Skalierung oder Rotation auf Zernike-Momente	89
8.35	Diskriminanzwerte zu Zernike-Momenten	89
8.36	Klassifikationsergebnisse für DS1 mit Zernike-Momenten	91
8.37	Klassifikationsergebnisse für DS2 mit Zernike-Momenten	91
8.38	Validität der k NN-Klassifikation bei Nutzung von Zernike-Momenten	91
8.39	Maximale Ordnung der Parameter p und q für Zernike-Momente	92
8.40	Diskriminanzwerte zu Wavelet-Momenten	92
8.41	Einfluß der Translation, Skalierung oder Rotation auf Wavelet-Momente	93
8.42	Klassifikationsergebnisse für DS1 mit Wavelet-Momenten	93
8.43	Klassifikationsergebnisse für DS2 mit Wavelet-Momenten	93
8.44	Validität der k NN-Klassifikation bei Nutzung von Wavelet-Momenten	94
8.45	Diskriminanzwerte zu kombinierten Momenten	96
8.46	Klassifikationsergebnisse für DS1 bei Verwendung kombinierter Momente	96
8.47	Klassifikationsergebnisse für DS2 bei Verwendung kombinierter Momente	98
8.48	Validität der k NN-Klassifikation bei kombinierten Momenten	98
8.49	Auflistung der Ergebnisse anderer Arbeiten	102
A.1	Deskriptive Statistik zu Momenten von Hu	107
A.1	Deskriptive Statistik zu Momenten von Hu	108
A.2	Deskriptive Statistik zu den Zernike - Momenten	109
A.3	Deskriptive Statistik zu Pseudo-Zernike Momenten	110
A.4	Deskriptive Statistik zu normalisierten Momenten	111
A.5	Deskriptive Statistik zu Legendre Momenten	112
A.6	Deskriptive Statistik zu Wavelet Momenten	113
A.7	Deskriptive Statistik zu Tsirikolias - Momenten	114
A.8	Deskriptive Statistik zu Orthogonal Fourier-Mellin Momenten	115

Kapitel 1

Einführung

Optische Schriftzeichenerkennung oder verallgemeinert auch die Erkennung von Symbolen und Objekten ist einer der wesentlichsten Anwendungsbereiche der Mustererkennung. Die OCR hat in den letzten Jahren erhebliche Fortschritte gemacht, ist aber noch immer Gegenstand intensiver Forschung.

Die heutigen, für normale Nutzer verfügbaren Programme sind in der Lage, hochqualitativen oder akkurat geschriebenen handschriftlichen Text weitestgehend korrekt zu erkennen. Allerdings gibt es noch erhebliche Defizite, wenn man normal geschriebenen handschriftlichen Text erkennen möchte oder versucht, komplexe Dokumente mit aufwendigen Strukturierungen und Formatierungen, wie z.B. mit Tabellen oder eingebetteten Bilddaten, korrekt zu verarbeiten.

Von einem papierlosen Büro ist man somit noch weit entfernt und noch immer werden Myriaden von Papier mit Schriftzeichen bedruckt. Auch wenn nur ein Bruchteil davon wieder in den PC zurückholt werden soll, bleibt immer noch ein sehr großes Anwendungsfeld für Zeichenerkennungsprogramme [15].

Bei der OCR liegt der kritischste Punkt bei der Erkennung von Symbolen im Bereich der Segmentierung und der Merkmalsextraktion. Jeder Bereich der OCR (Datenerfassung, Vorverarbeitung, Segmentierung, Merkmalsgewinnung, Klassifikation, Nachverarbeitung) beeinflusst den jeweils folgenden Verarbeitungsschritt.

Nach Trier et al.. [67] können in einem OCR-System die folgenden Komponenten identifiziert werden:

1. Einscannen der Daten mit Graustufen bei einer Auflösung von 300-1000 dpi;
2. Vorverarbeitung;
 - a Umwandlung in Binärbilder (Schwellwertverfahren) mit globalen oder lokal adaptiven Verfahren;
 - b Segmentierung zu isolierten, individuellen Zeichen;
 - c optionale Anpassung der Zeichendarstellung (Skelettisierung oder Konturkurven);

3. Merkmalsgewinnung;
4. Erkennung unter Verwendung eines oder mehrerer Klassifikatoren;
5. Kontextbezogene Prüfung der Ergebnisse (Wörterbuchmethoden) oder Nachverarbeitung;

Dabei wird in dieser Arbeit allerdings kein vollständiges OCR-System wie im obigen Sinne angestrebt, so daß vor allem die Segmentierung und kontextbezogene Nachverarbeitung nur am Rand Berücksichtigung findet.

Zum besseren Verständnis jedoch an dieser Stelle einige wenige Worte dazu. Nach Erfassung der Symbole durch einen Scanner liegt die Folge von Symbolen in digitaler Form vor. Nach eventuellen Vorverarbeitungsschritten, wie z.B. Rauschreduktion werden die einzelnen Zeichen segmentiert, um dann später für die Merkmalsgewinnung als Eingabedaten nutzbar zu sein.

Zur Segmentierung von Texten und Schriftzeichen aus Bilddaten gibt es sehr viele unterschiedliche Ansätze. Beginnend mit den einfachen projektiven Verfahren, die z.B. auf der Basis von Histogrammauswertungen arbeiten, bis hin zu komplexeren Verfahren wie z.B. mit Fuzzy Entscheidungsregeln und Entscheidungsbäumen [17]. Ein aktuelles Verfahren ist z.B. in der Arbeit von [46] beschrieben und zeigt, daß die exakte Segmentierung von Schriftzeichen noch immer ein nicht vollständig gelöstes Problem ist. Vergleichende Studien zu den Verfahrensschritten, die vor bzw. nach der Merkmalsgewinnung ablaufen, können in [4] nachgelesen werden.

In dieser Arbeit wird allerdings von bereits vorsegmentierten Daten ausgegangen und die Analysen sind auf die Merkmalsgewinnung und anschließende Klassifikation ausgerichtet. Kurze Anmerkungen zur Vorverarbeitung und Segmentierung sind im 2 Kapitel angegeben.

Eine Übersicht zur automatischen Dokumentverarbeitung, die sich obiger Probleme annimmt, kann in [64] nachgelesen werden. Für die Analysen in dieser Arbeit soll jedoch das Schema nach Trier in etwas vereinfachter Form, wie in Abbildung 1 dargestellt, Anwendung finden.



Abbildung 1.1: Vereinfachtes Schema der OCR [63]

Bei der OCR muß man, von den obigen Verarbeitungsschritten einmal abgesehen, generell auch noch zwischen zwei Arten unterscheiden.

1. Online Recognition
2. Offline Recognition

Mit Online-Erkennung ist dabei die Erfassung des geschriebenen Wortes z.B. über ein Touchscreen gemeint, wie es heutzutage in vielen PDAs ¹ Anwendung findet. Die dabei erfaßten Daten sind Zeitreihen, also Meßpunkte der Zeit $f(t, x, y)$ die von t und den Koordinaten x und y abhängig sind. Für derartige Verfahren sind grundsätzlich andere Erkennungsstrategien möglich (siehe auch [50] bis hin zur Rekonstruktion des Schreibflusses). In der vorliegenden Arbeit werden derartige Ansätze nicht betrachtet, sondern die Merkmalsextraktion im Fall der Offline-Erkennung.

Die Offline OCR beschreibt die Symbolerkennung im klassischen Stil, wobei aus einem gegebenen Eingabebild B mit den obig beschriebenen Schritten der im Bild vorhandene Text extrahiert und erkannt wird. In [50] wird dieser Schritt als Transformation der Sprache, repräsentiert in Form von grafischen Darstellungen, in eine symbolische Darstellung bezeichnet. Typischer Weise erfolgt diese Transformation dabei in den 8-bit ASCII Zeichensatz oder auch den 16-bit Unicode Zeichensatz.

Somit ergeben sich diverse interessante Probleme und Fragestellungen bei der Arbeit an Schriftzeichenerkennungen, wobei in dieser Arbeit allerdings primär die Merkmalsextraktion, spezialisiert auf statistische Momente und deren Nutzen für die OCR, untersucht wird.

Im Kapitel 3 werden verschiedene Methoden zur Erfassung von Merkmalen über Bilddaten angeführt und die Relevanz der statistischen Momente aufgezeigt. Die theoretischen Grundlagen und einfache geometrische Momente werden in Kapitel 4 vorgestellt. Dabei wird zum einen die Theorie für Momentinvarianten angegeben, wie sie z.B. von Reiss und Hu [53] genutzt wurden, als auch am Beispiel der geometrischen Momente die benötigten Invarianzeigenschaften eingeführt. Im Kapitel 5 werden darauf aufbauend verschiedene orthogonale Momente angegeben, die im Experimenteteil sehr gute Ergebnisse zeigten. Nach diesen einleitenden Abschnitten folgen in Kapitel 6 Ausführungen zu effizienten Algorithmen zur Berechnung von statistischen Momenten. Es werden dabei verschiedene Ansätze skizziert und die nach neueren Arbeiten z.B. von Flusser [18] als vorteilhaft angesehenen Algorithmen beschrieben und deren Nutzen verdeutlicht. Die eigentliche Merkmalsgewinnung und Klassifikation wird im Kapitel 7 betrachtet, dabei werden verschiedene Klassifikationsansätze beschrieben, die in den späteren Analysen Anwendung finden.

Im Anschluß erfolgen im Kapitel 8 weitgehende Analysen der verschiedenen Momenttypen und Aussagen über zu favorisierende Verfahren und deren Parametrisierung. Die Arbeit schließt mit den Schlußbemerkungen zu den endgültigen Ergebnissen und offenen Problemen.

¹Personal Digital Assistant - handheld computer

Kapitel 2

Vorverarbeitung

Als einer der ersten Schritte nach der Datenerhebung müssen die Daten geeignet vorverarbeitet werden. Dieser Schritt beginnt genaugenommen schon mit der Erfassung der Daten, da bereits an dieser Stelle diverse Probleme auftreten können, die die spätere Verarbeitung erschweren oder gar unmöglich machen können. So wurden die Daten bereits mit grösster Sorgfalt erfasst, um Probleme wie Rauschen, Rotationen und andere Störgrößen zu minimieren.

Trotz der Sorgfalt bei der Datenerfassung ergeben sich stets diverse notwendige Vorverarbeitungsschritte, die nicht vermieden werden können. Im wesentlichen sind dies die folgenden drei Schritte [4]

1. Rauschreduktion
2. Normalisierung der Daten
3. Kompression des Informationsgehaltes

2.1 Rauschreduktion

Wie angedeutet, ist stets ein gewisser Anteil von Rauschen in den Bilddaten nach dem Einscannen festzustellen. Diese Störungen sind oft durch Verschmutzungen der einzuscannenden Materialien verursacht und müssen z.B. durch einen Lowpassfilter reduziert werden [23], um die Merkmalsextraktion nicht stärker zu behindern. Da bei der nachfolgenden Merkmalsextraktion lediglich binäre Bilder betrachtet wurden, wurde durch das verwendete Schwellwertverfahren bereits eine gute Rauschreduktion erreicht. Für die Bildverbesserung sind in der Literatur eine Vielzahl von Verfahren vorgeschlagen worden, es werden in dieser Arbeit allerdings keine dahingehenden Analysen vorgenommen, sondern von bereits weitestgehend korrigierten Daten ausgegangen.

2.1.1 Morphologische Operationen

In einigen Artikeln zur Schriftzeichenerkennung (z.B. [32]) wird eine, über die Rauschreduktion hinausgehende Vorverarbeitung der Daten mittels morphologischer Operatoren vorgeschlagen. Die Idee hinter der Verwendung von morphologischen Operatoren ist die Filterung des Dokuments (Musters), wobei die Faltungsoperation durch logische Operationen ersetzt wird. Es wurden z.B. verschiedene morphologische Operatoren zur Verbindung von unterbrochenen Linien, zur Verdünnung von Zeichen oder zum Rändersuchen entwickelt. Dies kann zur Normierung der Strichdicke, der Symbole und zur Reduktion von Störungen, wie unterbrochenen Linien eingesetzt werden. So wurden morphologische Operatoren zum Thinning, also Ausdünnen von Linien verwendet, es zeigte sich allerdings, daß eine generelle Anwendung morphologischer Operatoren auf jedes Muster nicht empfehlenswert sind, so daß letztlich darauf verzichtet wurde. Insbesondere bei Berechnung der momentbasierten Merkmale, mittels Greens-Theorem vgl. Seite 43, ist dieser Verfahrensschritt auch unnötig.

2.1.2 Schwellwertverfahren

Um die Verarbeitungsgeschwindigkeit zu erhöhen, ist es häufig akzeptabel, die Grauwert- oder Farbbilder als Binärbilder darzustellen. Dazu wird ein Schwellwertverfahren verwendet, um die Anzahl der Farben zu reduzieren. In dieser Arbeit wurden alle Daten als Binärbilder verarbeitet und mit einem globalen Schwellwertverfahren von Graustufen in Binärbilder umgewandelt¹, so daß dann nur noch die Vordergrundpixel mit der Intensität 1 für Berechnungen verwendet werden mußten. Dies ist unter anderem auch deshalb von Vorteil, da für Binärdaten auch sehr effiziente Berechnungsalgorithmen entwickelt wurden.

Letztlich wurde auf den Rohdaten ein Schwellwertverfahren zur Binarisierung, eine Größennormierung und eine Invertierung angewandt, auf eine Skalierung der Bilddaten bezüglich *center of gravity* wurde verzichtet, da ein ebensolcher Effekt durch Verwendung zentraler Momente realisiert werden kann (siehe Abschnitt 4). Diese entsprechend vorverarbeiteten Daten wurden dann im Lern- bzw. Klassifikationsschritt verwendet.

2.2 Normalisierung und Normierung

Wenn die, aus den Mustern generierten Merkmale den gewünschten Anforderungen entsprechen, wird meist dennoch ein Normierungsschritt vor der eigentlichen Klassifikation eingefügt, um Dominanzeffekte einzelner Merkmale zu reduzieren oder allgemein die Daten in einen bestimmten Wertebereich zu transformieren. Im folgenden werden einige gängige Normierungsschritte aufgeführt und näher erläutert.

¹Diese Umwandlung wurde bei der Datenerhebung durch NIST [22] durchgeführt

Die einfachste Normalisierung über den Merkmalsvektoren kann wie folgt erzielt werden, man ermittelt den Mittelwert und die Varianz über den Merkmalsvektoren einer Klasse und berechnet den normierten Merkmalsvektor: [11]:

$$\hat{f}(x) = \frac{f(x) - \mu_j}{\sigma_j} \quad j = j^{\text{tes}} \text{ Merkmal}$$

Leider haben die verschiedenen Normierungsverfahren nicht nur Vorteile. So führt obiges Verfahren zwar dazu, daß die Dominanz einzelner Merkmale oder Merkmalsblöcke minimiert wird, leider wird allerdings auch die Zwischenklassenvarianz reduziert. Dies zeigte sich auch in einigen Tests mit dieser Normierung, wo zwar für einige Symbole deutlich verbesserte Erkennungsraten auftraten, im Mittel jedoch eine Verschlechterung der Erkennungsraten festzustellen war. Dies deckt sich auch mit Aussagen wie sie von de Sa in [12] gemacht werden, der obiges Normierungsverfahren nur dann empfiehlt, wenn die Merkmalsvarianzen lediglich auf Rauschen zurückzuführen sind.

De Sa schlägt stattdessen die Verwendung von Skalierungsfaktoren vor, wie sie auch im Artikel von Chim [71] genannt werden. Allerdings gibt es auch hierfür keine festen Regeln, wie diese Faktoren zu wählen sind, so daß diese meist empirisch ermittelt werden müssen.

Eine weitere Möglichkeit der Normierung ist, die Daten in den Funktionsbereich $[0, 1]$ zu transformieren. Dies kann mittels folgender Gleichung einfach erfolgen und ergab sehr gute Ergebnisse:

$$\hat{f}(x) = \frac{f(x) - \min_j}{\max_j - \min_j} \quad j = j^{\text{tes}} \text{ Merkmal}$$

Natürlich sind auch nicht lineare Normierungen wie logarithmische Normierungen denkbar. Wenn die Merkmalsklassen im Merkmalsraum allerdings bereits klar separiert sind, kann die Anwendung nicht linearer Verfahren auch zu einer Verschlechterung in der Separierung der Klassen führen. Aufgrund dieser Probleme wurden die Analysen auch jeweils mit und ohne Normalisierungen durchgeführt und die Ergebnisse verglichen.

Im Abschnitt Analysen wurde die Klassifikation der Merkmalsvektoren mit verschiedenen Normierungen untersucht.

Kapitel 3

Verfahren der Merkmalsgewinnung

Wenn die Vorverarbeitung und Segmentierung der Daten erfolgt ist, schließt sich die Merkmalsgewinnung an. Auch in diesem, stark beforschten Bereich gibt es verschiedene Ansätze. Prinzipiell können dabei zwei wesentliche Ansätze unterschieden werden:

1. statistische Verfahren der Merkmalsgewinnung
2. strukturelle Verfahren der Merkmalsgewinnung

Beide Ansätze besitzen verschiedene Ausprägungen und werden im folgenden im Kontrast zu den statistischen Momenten, die in dieser Arbeit primär analysiert werden, betrachtet. Die, im folgenden vorgestellten Methoden sind in ihrer Anwendung nicht ausschließlich auf die OCR beschränkt, sondern können meist allgemein zur Merkmalsgewinnung aus Bild oder gar Volumenbilddaten eingesetzt werden.

3.1 Allgemeine Prinzipien der Merkmalsgewinnung

Devijer und Kittler definieren die Merkmalsextraktion wie folgt [13]:

Definition 1 (Merkmalsextraktion)

Die Merkmalsextraktion ist das Problem der Ermittlung von für die Klassifikation diskriminierenden Merkmalen aus den Rohdaten. Im Sinne einer Minimierung der Innerklassenvarianz und Vergrößerung der Zwischenklassenvarianz der zu klassifizierenden Muster.

Dabei kann die Merkmalsextraktion auch als eine spezielle Form der Datenreduktion verstanden werden, um eine „günstige“ Menge von Variablen zu ermitteln. Dabei sollte beachtet werden, daß verschiedene Verfahren der Merkmalsgewinnung dieses Prinzip unterschiedlich gut erfüllen und ein Verfahren durchaus domainspezifisch geeignet oder ungeeignet sein kann. Die Wahl der zu extrahierenden Merkmale ist dabei ebenfalls domainbezogen. So ist z.B. die Invarianz der Merkmale bezüglich bestimmter Transformationen nur in manchen Fällen erwünscht. Ein Beispiel ist die Erkennung

von Flugzeugformen, bei denen die Rotationsinvarianz nötig ist, wohingegen im Bereich der Texterkennung diese Invarianz nur zum Teil Anwendung findet.

Ein weiteres, von Trier angeführtes Problem ist die *curse of dimensionality*, welche das Problem eines zu geringen Trainingsset bezeichnet, bei dem dann die Anzahl der Merkmale nicht zu hoch sein darf, um eine statistische Klassifikation nicht zu erschweren [67]. Eine diesbezüglich allgemeine Regel lautet: 5 – bis 10mal so viele Trainingsdaten für jede Klasse im Vergleich zur Anzahl der Merkmale zu verwenden. In der Praxis ist somit meist eine problemspezifische Auswahl der Merkmale bzw. der Merkmalsgenerierer sinnvoll.

Ein weiteres Problem ist die Frage der Klassenkonstruktion, die eng mit den gewählten Merkmalen in Bezug steht. So kann die unterschiedliche Ausprägung von bestimmten Symbolen, die eigentlich zur selben Klasse gehören, die Bildung von zwei oder mehr Klassen rechtfertigen (vgl. dazu auch [67], Seite 2).

In der kurzen Übersichtstabelle auf der folgenden Seite aus [30] bzw. [67] sind verschiedene Merkmalsextraktionsmethoden dargestellt.

Die Wahl der Merkmale bestimmt in einem hohen Maße die Effektivität der nachfolgenden Verfahren, auch muß berücksichtigt werden ob die gewählten Merkmale mit den nachfolgenden Klassifikatoren überhaupt verarbeitet werden können, da es bei den Klassifikatoren verschiedene Anforderungen an die Eingabedaten geben kann. So unterscheidet man unter anderem zwischen statistischen, grammatikbasierten oder strukturell syntaktischen Klassifikatoren.

In neueren Ansätzen zur Optimierung von Klassifikationsverfahren wird auch die Verwendung mehrerer Klassifikatoren und ein sich anschließendes Ranking vorgeschlagen [1]. Dies bietet im Regelfall den Vorteil einer höheren Erkennungsrate bei allerdings höherem Berechnungs- und Zeitaufwand. Insbesondere bei den statistischen Klassifikatoren, die reellwertige Eingangsdaten erwarten, ist ein solches mehrschichtiges Klassifikationsverfahren einfach realisierbar. Allerdings können bei entsprechendem Ranking und zusätzlichem Zeitaufwand auch die Ergebnisse nicht statistischer Klassifikatoren mit denen statistischer Klassifikatoren gemischt werden.

Die, in dieser Arbeit verwendeten Klassifikatoren sind ausschließlich statistischer Natur, da auch die Eingangsdaten, die durch die Momentgeneratoren erzeugt werden, stets reelle Werte sind bzw. als solche nach Norm-Bildung weiterverarbeitet werden.

3.2 Invariante Merkmale

Bei der Schriftzeichenerkennung ist das primäre Ziel, die Klassifikation ähnlicher Schriftzeichen zur gleichen Klasse, d.h. trotz teilweise erheblicher Varianzen in der Ausprägung der einzelnen Symbole eine korrekte Klassifikation vorzunehmen. Um dieses Ziel zu erreichen, müssen die gewählten Merkmale invariant gegenüber bestimmter affiner Transformationen, wie z.B. Translation/Verschiebung, Skalierung, Ro-

Methode	Eigenschaften	Kommentar
PCA	Lineare Abb.; schnell; Eigenvektorbasiert	Traditionelle Methode, auch als KLE bekannt (bei Gaussverteilung)
SOM	nichtlinear; iterativ	Neuronengitter; Extraktion niedrigdimensionaler Räume
Grauwert Teilbild	einfach; pixelbasiert	Gewinnung der Merkmale z.B. durch Skelettisierung
Template-Matching	rechenaufwendig	Verwendung von Muster-Templates in verschiedenen Darstellungen
deformierbare Templates	rechenaufwendig	deformierbare Templates; Anwendung z.B. in der Medizin
Histogrammprojek- tionen	diskrete Merkmale; erfassen Konturprofile	einfach
statistische Momente	einfach; standardisiert; effizient	Erfassen sowohl globaler Merkmale als auch von Feinstrukturen

Tabelle 3.1: Merkmalsextraktions- und Projektionsmethoden

tation, Rauschen sein. Mit Invarianz ist dabei gemeint, daß annähernd dieselben Merkmalswerte für Symbole aus ein- und derselben Klasse ermittelt werden, selbst wenn diese Symbole durch affine Transformationen von den Klassenrepräsentanten abweichen. Dabei kann für bestimmte Symbole die Invarianz bzgl. z.B. Rotation oder Skalierung auch nachteilig in der Klassifikation sein, da die Unterscheidung zwischen einem 'm' oder einem 'w' bzw. einem 'o' oder einem 'O' damit verhindert würde. Deshalb sollte man für die Klassifikation solcher Symbole andere Merkmale verwenden bzw. müssen die Merkmalsvektoren geeignet gewählt werden, um dennoch auch eine Klassifikation dieser spezifischen Symbole zu ermöglichen. Dies kann z.B. dadurch geschehen, daß auch nichtrotations- und skalierungsinvariante Merkmale zu dem Merkmalsvektor hinzugefügt werden.

In einigen Fällen ist es auch nützlich, die Rohdaten geeignet vorzuverarbeiten, um z.B. eine Normierung bezüglich Größe und Position zu erhalten. Dies wurde z.B. auch in dieser Arbeit durchgeführt, indem alle Bilddaten auf eine einheitliche Größe skaliert wurden (siehe auch Kapitel 4) ¹.

3.3 Strukturelle Verfahren der Merkmalsgewinnung

Zunächst werden im folgenden Abschnitt strukturelle Verfahren der Merkmalsgewinnung kurz dargestellt, um deren Problematik im Bereich der OCR deutlich zu machen und die Verwendung statistischer Merkmale zu motivieren.

3.3.1 Template-Matching

Eines der ersten Verfahren zur Merkmalsgewinnung ist das Template-Matching. Dabei wird das, zu klassifizierende Objekt als ein Merkmalsvektor verstanden (man könnte auch sagen, daß eigentlich gar keine Merkmalsextraktion stattfindet) und eine Kostenfunktion zwischen dem zu klassifizierenden Bild und dem Template (Prototypen - ebenfalls ein Bild, aufgefaßt als Merkmalsvektor) berechnet. Das Bild wird dabei zu der Klasse des Templates klassifiziert, welches zur niedrigsten Kostenfunktion führt bzw. die größte Ähnlichkeit mit dem zu klassifizierenden Bild besitzt. Dieses Verfahren ist zwar sehr einfach, aber nicht sehr effektiv, da hierbei kaum affine Transformationen berücksichtigt werden können. So ist z.B. bei der Rotation keine adäquate Klassifikation zu erwarten. Eine häufig angewandte Lösung ist dann, das Template in verschiedenen Rotationen über das zu klassifizierende Bild zu matchen, allerdings ist dies nur sinnvoll, wenn lediglich eine geringe Menge an Rotationen zu erwarten ist (also z.B. 45, 90, 135, 180, ... Grad). Aus diesem Grund wird dieser Ansatz im Bereich der OCR nur selten verwendet.

Weitere Probleme bei dieser Form der Erkennung treten durch Rauschen oder unterschiedliche Ausleuchtung der Templates auf. Eine genauere Beschreibung von Tem-

¹Diese Normierung war in den NIST-Daten bereits gegeben



Abbildung 3.1: Histogrammprojektionen zum Symbol 3.2(f)

plates kann in [67] und [4] nachgelesen werden und wird hier nicht weiter behandelt, da Template-Matching für die OCR nur von sehr geringer Relevanz ist.

3.3.2 Unitäre Transformationen

Bei diesem Ansatz wird das zu klassifizierende Bildmaterial in einen anderen Raum unitär transformiert, um dann dort eine Selektion der relevanten Merkmale durchzuführen. Ein Ansatz dazu wurde z.B. von Andrews [3] vorgeschlagen, bei dem das Bild unitär transformiert und dann in diesem Raum eine Sortierung der Pixel nach Varianz durchgeführt wird. Die Pixel mit der höchsten Varianz werden dann als Merkmale genutzt. Die dabei verwendeten Transformationen sind oft die Karhunen-Loeve Transformation [37], die Fouriertransformation [37] und die Walshtransformation [47].

3.4 Histogrammprojektionen

Bei der Histogrammprojektion [67] werden die Merkmale aus dem Histogramm des Bildes gewonnen, in welchem die Grauwertverteilungen abgelesen werden können. Für binäre Bilder gestaltet sich dies besonders einfach, da in diesem Fall lediglich pro Position auf der Abszisse bzw. Ordinate die Anzahl der schwarzen Bildpunkte aufgetragen ist. Wird dieses Histogramm dann auf eine Koordinatenachse projiziert, spricht man von einer Histogrammprojektion. Dieses Verfahren wird heutzutage meist als eine einfache Form der Segmentierung eingesetzt, kann aber auch zur Merkmalsgewinnung verwendet werden. Die erhaltenen Merkmale sind allerdings sehr sensibel bzgl. Rauschen und affiner Transformationen, so daß sie nur von geringem Nutzen sind. In Abbildung 3.1 ist eine entsprechenden Projektion gegeben.

Zoning Beim Zoning wird der Bereich, der das Zeichen enthält, in verschiedene, sich überlappende oder nichtüberlappende Regionen unterteilt. Die Dichten, Punkte oder einige Merkmale in den verschiedenen Regionen werden analysiert und zu einem Merkmalsvektor zusammengefaßt. So kann z.B. die Konturrichtung des Zeichens als Merkmal erfaßt werden, indem das Bild in rechteckige und diagonale Bereiche

unterteilt wird und dann die Histogramme der Chaincodes dieser Bereiche berechnet werden (vgl. Trier et al. [67]).

3.5 Statistische Verfahren der Merkmalsgewinnung

3.5.1 Statistische Momente

Ein wesentliches statistisches Verfahren zur Merkmalsgewinnung ist die Merkmalsextraktion mit statistischen Momenten. Auch hier gibt es sehr unterschiedliche Ansätze bzw. Ausprägungen mit unterschiedlichen Eigenschaften bzgl. informationstheoretischer Überlegungen, Invarianz und Berechnungskosten. Da die statistischen Momente den Hauptteil dieser Arbeit ausmachen, sind ihnen mehrere Kapitel gewidmet, so daß ich an dieser Stelle nur auf das Kapitel 4 ff. für weitere Informationen zu statistischen Momenten verweisen möchte. Im folgenden werden zu Vergleichszwecken weitere Verfahren kurz vorgestellt, die zur statistischen Merkmalsgewinnung verwendet werden können.

3.5.2 Zeitreihenexpansionen

Ein kontinuierliches Signal enthält im allgemeinen mehr Information als benötigt wird, um es z.B. für eine Klassifikation darzustellen [4]. Dies gilt sowohl für kontinuierliche wie auch diskrete Signale. Eine Möglichkeit ein Signal darzustellen besteht in der Zerlegung des Signals in eine Folge von Linearkombinationen einfacherer Funktionen. Die Koeffizienten dieser Linearkombination sind eine kompakte Kodierung, die auch als Zeitreihenexpansion bezeichnet wird und auf Fourier zurückgeht. Deformationen wie Translation und Rotation sind invariant unter globalen Transformationen und Reihenexpansion. Dabei häufig auftretende Expansionsmethoden sind die folgenden:

- **Fouriertransformation (Fourierdeskriptoren)**
Das allgemeine Verfahren ist gegeben durch die Wahl des Betragsspektrums (Magnitudenspektrum) des gemessenen Vektors als Merkmal in einem n -dimensionalen euklidischen Raum. Dabei ist eine der angenehmsten Eigenschaften der Fouriertransformation, die Erkennung, verschobener Zeichen, wenn man nur den Betrag im Spektrum betrachtet und die Phase ignoriert. Die Fouriertransformation findet deshalb sehr oft ihre Anwendung in der OCR.
- **Gabortransformation**
Dies ist eine Abwandlung der (windowed) Fouriertransformation. In diesem Fall ist das verwendete Fenster nicht diskret in der Größe, sondern durch eine Gaussfunktion definiert.
- **Wavelets**
Die Wavelettransformation ist eine Reihenexpansionstechnik, welche die Darstellung des Signals in verschiedenen Auflösungen gestattet. Die Segmente des

Dokumentbildes, die zu Buchstaben in einem Wort korrespondieren, sind dabei durch Waveletkoeffizienten bezüglich verschiedener Auflösungsebenen repräsentiert. Diese Koeffizienten werden dann vom Klassifikator verwendet. Die Darstellung des Signals in Multiresolutionanalysis mit niedriger Auflösung kann dabei genutzt werden, um lokale Variationen in der Handschrift, wie man sie z.B. bei Hochauflösung feststellt, zu eliminieren. Jedoch kann die Repräsentation mit niedriger Auflösung in manchen Fällen auch zum Verlust der relevanten Information führen.

3.5.3 Neural Network Classifiers

Mehrschichtige Feedforward Netzwerke (MLP - multilayer perceptron, im folgenden kurz NN) sind ein sehr approbates Mittel sowohl zur Merkmalsgewinnung, als auch zur Klassifikation und finden eine breite Anwendung in der OCR. Eine gute Einführung in das Thema der Neuronalen Netze ist in [12] angegeben. Neuronale Netze erzeugen dabei Entscheidungsschranken innerhalb des Merkmalsraums, so daß eine Klassifikation entsprechend der gewählten Grenzen erfolgen kann. Die Verwendung von NN als Klassifikator ist in Kapitel 7 näher beschrieben. NN können allerdings auch in einem gewissen Sinne als Merkmalsextraktoren angesehen werden. Dabei wird das NN als hierarchischer Merkmalsextraktor verstanden (vgl. Le Chun [36]). Jeder Knoten „sieht“ ein Fenster in der vorigen Schicht und kombiniert die Merkmale einer niedrigeren Stufe in diesem Fenster mit Merkmalen einer höheren Stufe. Somit ergibt sich die Situation, daß je höher die Netzwerkschicht ist, um so globaler ist das ermittelte Merkmal. Der letzte finale Abstraktionsschritt ist dann z.B. das Merkmal '0', ..., '9', also die zu erkennenden Symbole. Von besonderer Bedeutung ist auch hier ein ausreichendes Training des Klassifikators bzw. Merkmalsextraktors mit einer großen Zahl von Trainingsdaten, um adäquate Resultate zu erzielen. Leider sind NN häufig sehr schwierig zu trainieren und es treten Probleme der Überanpassung und Rauschempfindlichkeit auf. Allerdings zeigen Arbeiten von Petersen et al. [16], daß das NN mit zu den am besten geeigneten Klassifikatoren in der OCR gehört.

3.6 Verwendete Verfahren

Wie bereits in den Vorbemerkungen zu ersehen, gibt es eine Vielzahl von Ansätzen zur Merkmalsgewinnung, jeder mit eigenen Vor- und Nachteilen. Die Auswahl, welches Verfahren verwendet werden soll, richtet sich dabei vor allem nach dem Anwendungszweck und den Rohdaten. In dieser Arbeit werden segmentierte handschriftliche Symbole berücksichtigt und somit sind Verfahren, die eine zu starke Standardisierung, wie z.B. bzgl. Größe oder Illumination erwarten, eher ungeeignet, da diese sicher keine geeigneten Merkmale extrahieren würden, um die Varianzen in den Rohdaten adäquat zu erfassen. Ein weiterer Aspekt ist die Berechnungsdauer. So ist eine Zielsetzung der Arbeit, die Erstellung einer möglichst schnellen Implementierung, die auch eine

online-Erkennung (im Sinne von nicht Batchbetrieb vgl. Abschnitt 1 on/offline Erkennung) ermöglicht. Aus diesem Grund scheiden unter anderem z.B. Verfahren wie das Template-Matching generell aus.

Die Auswahl eines einzelnen Verfahrens wird in der Literatur jedoch kritisiert (z.B. bei Trier [67]), so daß es sich empfiehlt, nicht nur einen einzigen Merkmalsextraktor zu favorisieren. Dies ist nicht zuletzt deshalb nötig, da einige Merkmalsextraktoren nur in einem bestimmten Sinn invariante Merkmale erzeugen und wie auf Seite 21 beschrieben, sind in der OCR sowohl invariante als auch nicht invariante Merkmale von Nutzen. Die Fourierdeskriptoren können in unserem Fall auch nicht sinnvoll verwendet werden, da diese nur sehr schlecht Symbole mit unterbrochenen Konturen beschreiben. Somit sind von den hier betrachteten für den Bereich der OCR am ehesten geeigneten Merkmalsgeneratoren diejenigen, welche auf statistischen Momenten basieren, da diese sowohl verschiedenste Invarianzeigenschaften realisieren, als auch effizient berechnet werden können.

Wenn die reine Verarbeitungsgeschwindigkeit nicht primär im Vordergrund steht, weisen neuere Arbeiten (z.B. [25]) darauf hin, daß eventuell bessere Resultate durch Verwendung bzw. Kombination mit strukturellen Merkmalen zu erwarten sind. Dabei stellt sich allerdings das Problem der Kombination der statistischen und strukturellen Merkmale sowie der nachfolgenden Klassifikation. In einer Arbeit von Heute et al. [25] wird dieses Problem für Handschriftenerkennung analysiert und ein entsprechendes Verfahren zur Erzeugung von Merkmalsvektoren vorgestellt.

3.7 Trainingsdatengewinnung

Bei der Erkennung von Schriftzeichen tritt die Frage der Rohdaten in zwei Kontexten auf. Zum einen benötigt man Prototypen (Trainingsdaten) von allen zu erkennenden Schriftzeichen, die im späteren Klassifikationsprozeß zur Zuordnung der Symbole zu den jeweiligen Klassen Anwendung finden. Dabei sollte jede Symbolklasse durch eine Vielzahl von Repräsentanten definiert werden, um eine möglichst große Varianz für jedes Symbol zu erfassen. Zum anderen werden die Rohdaten im eigentlichen OCR Prozess in Form von Testdaten genutzt.

Die numerischen Rohdaten entstammen dabei der frei verfügbaren Onlinedatenbank des NIST [45]. Die NIST - Daten waren bereits vorsegmentiert, vorverarbeitet (Rauschreduktion, Rotationskorrektur usw.) und sind in einer standardisierten Form (NIST - Format) im Netz verfügbar.

Von der Menge der Daten wurden $\frac{2}{3}$ als Prototypen im Lernschritt eingesetzt und das verbleibende $\frac{1}{3}$ für die spätere Testphase genutzt.

Insgesamt standen 3471 numerische Symbole zur Verfügung. In der folgenden Abbildung 3.2 ist exemplarisch eines der verwendeten Sets (Symbole 0 – 9) der NIST-Datenbank [45] dargestellt.

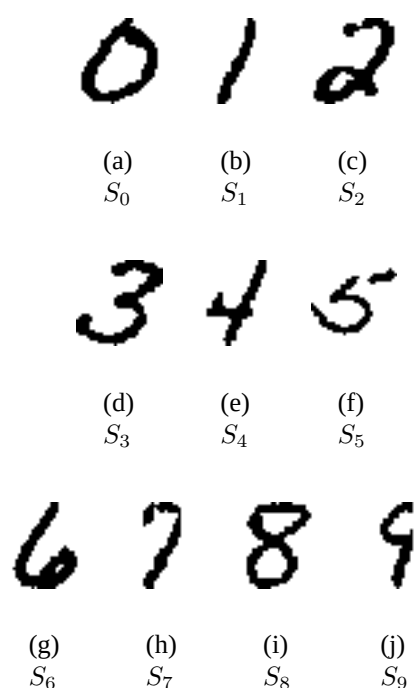


Abbildung 3.2: Beispielsymbole aus der verwendeten NIST Datenbank [45]

Kapitel 4

Momente - eine Einführung

Statistische Momente sind in den letzten 30 Jahren zu einer Standardmethode der Objekterkennung geworden, sind jedoch noch immer Gegenstand aktueller Forschung. Sie wurden erstmalig von Hu [26] für die Erkennung von Mustern, basierend auf dem von Hu eingeführten „Fundamentaltheorem für Momentinvarianten“ eingesetzt. Im folgenden nun eine Einführung in die dazu gehörige Theorie, wie sie im Artikel von Hu [26] und in einem Artikel von Mamistvalov [42] gegeben ist.

4.1 Momente und algebraische Invarianten

Die zweidimensionalen Momente der $(p+q)$ ten Ordnung einer Dichtefunktion $f(x, y)$ sind in Form von Riemannintegralen definiert als:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots$$

Wenn man annimmt, daß die Dichtefunktion $f(x, y)$ eine stückweise kontinuierlich beschränkte Funktion ist und nur in einem endlichen Abschnitt der xy -Ebene Werte auftreten, die von Null verschieden sind, dann existieren die Momente einer beliebigen Ordnung und das folgende Eindeutigkeitstheorem gilt [26]:

Theorem 1 (Uniqueness theorem)

Die Momentfolge m_{pq} ist eindeutig durch $f(x, y)$ bestimmt. Umgekehrt gilt auch, daß die Dichtefunktion $f(x, y)$ eindeutig durch eine Momentfolge m_{pq} bestimmt ist.

Es ist laut des Eindeigkeitstheorems also möglich jedes beliebige Bild (mit den obigen Bedingungen), welches einer Dichtefunktion entspricht, aus einer unendlichen Momentfolge zu rekonstruieren. Ebenso ist dadurch gegeben, daß die Momente die Information über das Bild beinhalten, was im folgenden Mustererkennungsansatz genutzt wird. Das Integral in Gleichung 1 kann als eine 1:1 Abbildung der kontinuierlich, beschränkten Bildfläche auf eine unendliche diskrete Matrix von Momenten $\mathbf{M}$ mit den Einträgen m_{pq} verstanden werden.

Wie im folgenden noch genauer beschrieben wird, sollten die berechneten Momente gegenüber bestimmter Transformationen über der Dichtefunktion (in unserem Fall dem Muster) invariant sein. So kann z.B. die Translationsinvarianz durch Einführung von zentralen Momenten erreicht werden.

4.1.1 Algebraische Formen und Invarianten

Invariante Momente basieren auf dem theoretischen Ansatz der algebraischen Invariantentheorie. Im folgenden sind die, für die Herleitung des Fundamentaltheorems der Momenteninvarianten notwendigen Begriffe angegeben.

Das folgende homogene Polynom mit 2 Variablen u und v und den Koeffizienten $(a_{p0}, \dots, a_{0p})$

$$f = a_{p0}u^p + \binom{p}{1}a_{p-1,1}u^{p-1}v + \binom{p}{2}a_{p-2,2}u^{p-2}v^2 + \dots + \binom{p}{p-1}a_{1,p-1}uv^{p-1} + a_{0,p}v^p$$

wird als binäre algebraische Form oder vereinfacht als Binärform der Ordnung p bezeichnet. Dies kann unter Verwendung der Notation von Cayley wie folgt geschrieben werden [42]:

$$f \equiv (a_{p0}; a_{p-1,1}; \dots; a_{1,p-1}; a_{0,p})(u, v)^p \quad (4.1)$$

Ein homogenes Polynom $I(a)$ mit den Koeffizienten $a = \{a_{p0}, \dots, a_{0,p}\}$ ist eine algebraische Invariante mit Gewichtung ω , wenn folgende Bedingung gilt:

$$I(a'_{p,0}, \dots, a'_{0,p}) = \Delta^\omega I(a_{p,0}, \dots, a_{0,p})$$

wobei $a'_{p,0}, \dots, a'_{0,p}$ die neuen Koeffizienten des Polynoms sind, die sich ergeben, wenn man die folgende Transformation in die Gleichung 4.1 einsetzt.

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix}, \Delta = \begin{vmatrix} \alpha & \gamma \\ \beta & \delta \end{vmatrix} \neq 0$$

Im folgenden ein kurzes Beispiel aus [52]. Man betrachte die algebraische Invariante 2. Ordnung mit $\omega = 2$ $I(A, B, C) = AC - B^2$. Bei Skalierung der Koeffizienten $(a_{p0}, \dots, a_{0p})$ um einen konstanten Faktor β wird die Invariante der 2. Ordnung wie folgt beeinflusst:

$$I(\beta A, \beta B, \beta C) = \beta^2 AC - (\beta B)^2 = \beta^2(AC - B^2) = \beta^2 I(A, B, C)$$

Wenn $\omega = 0$, dann wird die Invariante als absolute, andernfalls als relative Invariante bezeichnet. Wie bei Hu weiter ausgeführt, kann unter Verwendung der obig eingeführten Feststellungen das Fundamentaltheorem der Momentinvarianten abgeleitet werden. Im folgenden wird die revidierte Fassung des Fundamentaltheorems nach Reiss [52] angegeben:

Theorem 2 (Fundamentaltheorem der Momentinvarianten)

Wenn die Binärform der p^{ten} Ordnung eine algebraische Invariante mit Gewicht ω und Ordnung k besitzt:

$$I(a'_{p0}, \dots, a'_{0p}) = \Delta^\omega I(a_{p0}, \dots, a_{0p})$$

dann haben die Momente der p^{ten} Ordnung die gleichen Invarianten, aber mit der Hinzunahme eines Faktors $|J|^k$

$$\begin{aligned} I(\mu'_{p0}, \dots, \mu'_{0p}) &= |J| \Delta^\omega I(|J| \mu_{p0}, \dots, |J| \mu_{0p}) \\ &= |J|^k \Delta^\omega I(\mu_{p0}, \dots, \mu_{0p}) \end{aligned}$$

Dabei sei $|J|$ der Absolutbetrag der Determinante Δ der Bildtransformation.

Basierend auf diesen Theoremen können nun unter Zuhilfenahme der algebraischen Invariantentheorie verschiedene momentbasierte Invarianten abgeleitet werden (vgl. 7 Invarianten von Hu), die dann für die Musterkennung bzw. Erfassung invarianter Merkmale nutzbar sind.

4.1.2 Erweiterung des Fundamentaltheorems auf n-Dimensionen

Das von Hu eingeführte Fundamentaltheorem der Momentinvarianten ist für den zweidimensionalen Fall definiert. Von Mamistvalov [42] wurde eine Erweiterung der Definitionen für den n -dimensionalen Fall vorgestellt, wie sie z.B. auch bei der Berechnung von Momenten über dreidimensionalen Daten nützlich ist. Die n -dimensionalen Momente der p^{ten} Ordnung einer Intensitätsfunktion $f(x_1, \dots, x_n) = f(x)$ können in Form eines Riemannintegrals wie folgt definiert werden:

$$m_{p_1, \dots, p_n} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{p_1} \dots x_n^{p_n} f(x) dx_1 \dots dx_n \quad 0 \leq p_i < \infty$$

Wenn $f(x)$ als stückweise kontinuierliche beschränkte Funktion angenommen wird und es nur in einem endlichen Bereich von R^n von Null verschiedene Werte gibt, dann existieren die Momente jeder beliebigen Ordnung. Die n -dimensionalen zentralen Momente können dann als

$$\mu_{p_1, \dots, p_n} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_1 - \bar{x}_1)^{p_1} \dots (x_n - \bar{x}_n)^{p_n} f(x) dx_1 \dots dx_n$$

definiert werden mit $\bar{x}_1 = \frac{m_{1\dots 0}}{m_{0\dots 0}}, \dots, \bar{x}_n = \frac{m_{0\dots 1}}{m_{0\dots 0}}$ und $p_1 + \dots + p_n = p, 0 \leq p \leq \infty$

Die zentralen Momente sind dabei invariant gegenüber Translation. Erweitert man nun noch die Definitionen für algebraische Invarianten für n Dimensionen entsprechend der Definitionen von Hu (vgl. [42], S.820/821), kann man das verallgemeinerte Fundamentaltheorem der Momentinvarianten wie folgt angeben:

Theorem 3 (Verallgemeinertes Fundamentaltheorem der Momenteninvarianten)

Wenn das n -äre Quantik der Ordnung p eine algebraische Invariante mit dem Gewicht ω und der Ordnung k besitzt:

$$I(\acute{a}) = \Delta^\omega I(a)$$

dann haben die n -dimensionalen Momente der Ordnung p die selben Invarianten, allerdings mit einem zusätzlichen Faktor $|J|^k$, wobei $|J|$ die Determinante der Jakobimatrix der auf der Dichtefunktion angewandten Transformation ist.

$$I(\mu'_{p,0\dots 0}, \mu'_{p-1,1\dots 0}, \dots, \mu'_{0,0\dots p}) = \Delta^\omega |J|^k I(\mu_{p,0\dots 0}, \mu_{p-1,1\dots 0}, \dots, \mu_{0,0\dots p})$$

Eine angenehme Eigenart von Momenten ist die Fähigkeit, die Form eines Objektes zu beschreiben - das ist es, was Momente für die Mustererkennung interessant werden lässt. Die einfachsten Momente sind dabei Momente von gewöhnlichen Funktionen wie $f(x)$. Damit kann man dann auch bereits die einfachsten Momente ermittelt, da z.B. der Erwartungswert oder die Varianz geeignet sind, die Form eines Objektes (hier einer einfachen Funktion) zu beschreiben.

$$\begin{aligned} E[x] &= \int x g(x) dx && \text{Erwartungswert} \\ E[x^2] &= \int x^2 g(x) dx && \text{Varianz} \end{aligned}$$

Diese beiden Werte werden als *first-moment* (Erwartungswert) bzw. *second-moment* (Varianz) bezeichnet. Verallgemeinert lässt sich aus der obigen Darstellung die allgemeine Form für Momente ableiten

$$m_p = \int f(x) x^p dx \quad (1 - D)$$

Aus den obigen Gleichungen lassen sich nun unendlich viele Werte ableiten, die die Form der betrachteten Funktion beschreiben. Natürlich kann man nicht unendlich viele Momentenwerte verwenden, das wäre für eine effiziente Verarbeitung bzw. online OCR inakzeptabel. Aus diesem Grund werden nur einige wenige, diskriminante Momente verwendet, um die Funktion geeignet zu beschreiben. Mittels dieser wenigen Werte lässt sich die Funktion dann relativ genau rekonstruieren bzw. kompakt kodieren. Die Forschung an Momenten konzentriert sich somit unter anderem auf die Ermittlung einer möglichst geringen Menge an Momenten, die eine gegebene Funktion $f(x)$ möglichst mit geringem Fehler beschreiben, oder invariante Eigenschaften des Musters erfassen. Die verschiedenen Ansätze dazu werden in den folgenden Abschnitten näher beschrieben.

Im obigen Fall wurden nur eindimensionale Funktionen betrachtet, allerdings lassen sich Momente, wie in der Einleitung bereits gezeigt wurde, auch auf mehrdimensionale Funktionen anwenden.

Momente sind also zur Formbeschreibung geeignet - das ist zwar eine sehr schöne Eigenschaft, nun tritt aber bei der Mustererkennung stets ein Problem in den Vordergrund. Objekte ähnlicher Gestalt, z.B. unterschiedlicher Größe sollten stets korrekt als ein und dasselbe Objekt erkannt werden. Es wäre also wünschenswert, wenn die Merkmale gegenüber Transformationen, wie z.B. Skalierung oder Translation invari-

ant wären. Auch sollten die Merkmale gegenüber Rotation meist invariant sein.¹ Ebenso wäre eine Invarianz der extrahierten Merkmale gegenüber gewissen Störungen, wie Rauschen wünschenswert. Invariant bedeutet also, daß eine bestimmte Größe sich bei Durchführung einer bestimmten Transformation nicht verändert. Derartige Probleme werden in den folgenden Abschnitten diskutiert und mit Bezug auf bereits abgeschlossene wissenschaftliche Analysen für jede Form der betrachteten Momente untersucht.

4.2 Momente einer Fläche

Wir betrachten im folgenden eine zweidimensionale Lichtintensitätsfunktion bzw. Bildfunktion (Grauwertbilder), die als visuelle Eingabe vorliegt. Diese Intensitätsfunktion kann man als eine Dichtefunktion auffassen und wir können die Momente für den zweidimensionalen Fall definieren:

$$m_{pq} = \int \int x^p y^q f(x, y) dx dy$$

Man kann nun wieder die Momente beliebiger Ordnung für diese Eingaben berechnen - allerdings genügt auch hier eine, eher geringe Anzahl von Momenten, um das Bild adäquat zu beschreiben.

Nun sind die obigen Funktionen stets für den kontinuierlichen Fall beschrieben. Dies ist für den Fall von Bilddaten als Eingabe natürlich nicht der Fall. Somit betrachten wir obige Funktionen nur für den diskreten Fall und wählen unsere Eingabedaten stets quadratisch als $n \times n$ Muster. Für die Breite und Höhe der Pixel nehmen wir eine Einheit an, die Intensität ergibt sich aus dem Wert von $f(x, y)$ an den entsprechenden Koordinaten des Bildes. Für das pq te Moment erhält man somit:

$$m_{pq} = \sum_{k=1}^n \sum_{l=1}^n x_k^p y_l^q f(x_k, y_l)$$

Für den Fall von Binärbildern genügt es lediglich, die Bildwerte mit Intensität $\equiv 1$ zu betrachten. Die Formel reduziert sich zu:

$$m_{pq} = \sum_A x^p y^q \quad A = \text{Fläche des Bildes}$$

Nun kann man von gegebenen Bilddaten die Merkmalsvektoren auf Basis von Momenten erstellen und über diesen Vektoren die Symbolerkennung durchführen. Problematisch ist allerdings, daß die meisten Momente sehr sensibel gegenüber Rauschen (z.B. Salt-Pepper-Noise) sind.

¹Warum nur meist ? - Insbesondere im Bereich der Texterkennung sollten die extrahierten Merkmale eines Zeichens nicht notwendiger Weise stets rotationsinvariant sein, da z.B. der Buchstabe *m* und *w* bezogen auf Rotation nahezu identische Objekte sind und damit in eine Klasse fallen.

4.3 Invarianz von Momenten

Aus dem vorigen Abschnitt kann man entnehmen, daß das Moment M_{00} die Fläche des Bildes beschreibt, genaugenommen die Anzahl der schwarzen Bildpunkte. Die einfachsten Momente sind die first-order und second-order Momente mit $(p + q = 1)$ bzw $(p + q = 2)$:

$$\begin{aligned} m_{10} &= \sum_{k=1}^n \sum_{l=1}^n x f(x_k, y_l) \\ m_{01} &= \sum_{k=1}^n \sum_{l=1}^n y f(x_k, y_l) \\ m_{20} &= \sum_{k=1}^n \sum_{l=1}^n x^2 f(x_k, y_l) \\ m_{02} &= \sum_{k=1}^n \sum_{l=1}^n y^2 f(x_k, y_l) \\ m_{11} &= \sum_{k=1}^n \sum_{l=1}^n xy f(x_k, y_l) \end{aligned}$$

Diese definieren die Größe und Orientierung des Bildes. Wenn nur diese Momente (bis Ordnung 2) verwendet werden, ist das Originalbild vollständig äquivalent zu einer konstanten Beleuchtungsdichtenellipse mit definierter Größe, Orientierung und Ausdehnung mit Zentrierung zum Bildzentroid [65].

In der Physik gibt es das Konzept des Massezentrums, gegeben als ein Punkt im Objekt, zu dem die gesamte Masse konzentriert werden kann, so daß das Objekt als ein einzelner Punkt betrachtet werden kann. Dieses Konzept kann auf Bilddaten erweitert werden. Der arithmetische Mittelwert des Bildes, bzw. auch *center of gravity* genannt, ergibt sich damit zu

$$\bar{x} = \frac{m_{01}}{m_{00}} \quad \bar{y} = \frac{m_{10}}{m_{00}}$$

Damit ist es nun möglich, das Moment gegenüber Verschiebung (Translation) invariant zu gestalten

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) d(x - \bar{x}) d(y - \bar{y})$$

Beweis 1

Seien x und y die kartesischen Koordinaten des untransformierten Bildes $f(x, y)$, $\acute{x}$, $\acute{y}$ die kartesischen Koordinaten des transformierten Bildes. Mit einer Verschiebung

$$\begin{aligned} \acute{x} &= x + \alpha & \alpha, \beta - \text{konstant} \\ \acute{y} &= y + \beta \end{aligned}$$

Seien die Momente des transformierte Bildes gegeben durch:

$$\begin{aligned} g'_{pq}(\acute{x}, \acute{y}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \alpha)^p (y - \beta)^q f(x, y) dx dy \\ \acute{x} &= \bar{x} - \alpha \\ \acute{y} &= \bar{y} - \beta \end{aligned}$$

Die zentralen Momente unter Verschiebung sind dann gegeben durch:

$$\begin{aligned} \acute{g}_{pq}(\acute{x}, \acute{y}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\acute{x} - \acute{x})^p (\acute{y} - \acute{y})^q \acute{f}(\acute{x}, \acute{y}) d\acute{x} d\acute{y} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \alpha - (\bar{x} - \alpha))^p (y - \beta - (\bar{y} - \beta))^q \acute{f}(\acute{x}, \acute{y}) d\acute{x} d\acute{y} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \cdot 1 dx dy \\ &= \bar{g}_{pq}(x, y) \quad \text{mit Jakobian} = 1 \end{aligned}$$

□

Die zentralen Momente sind dabei äquivalent zu den regulären Momenten eines Bildes, welches so verschoben wurde, daß das Zentrum des Bildes mit dem Ursprung übereinstimmt. Daraus folgt, daß die zentralen Momente sich unter Verschiebung der Koordinaten nicht verändern, also invariant unter Translation sind. Außerdem gilt $\mu_{01} \equiv \mu_{10} \equiv 0$. Die zentralen Momente stehen mit den regulären geometrischen Momenten wie folgt in Beziehung

$$\mu_{pq} = \sum_{k=0}^p \sum_{l=0}^q \binom{p}{k} \binom{q}{l} (-1)^{k+l} m_{p-k, q-l} m_{10}^k m_{01}^l m_{00}^{-(k+l)}$$

Betrachtet man nun die Standardabweichung in x - bzw. y -Richtung

$$\begin{aligned} \sigma_x &= \sqrt{\frac{1}{M} \sum_{k=1}^M (x_k - \bar{x})^2} \quad M = m_{00} \\ \sigma_x &= \sqrt{\frac{\bar{m}_{20}}{m_{00}}} \end{aligned}$$

erhält man die Invarianz bzgl. Verzerrungen (Stauen, Stretchen, Auslöschung) wie folgt ²:

$$\eta_{pq} = \sum_A \left(\frac{x - \hat{x}}{\sigma_x} \right)^p \left(\frac{y - \hat{y}}{\sigma_y} \right)^q \quad (4.2)$$

²Der Beweis zu obiger Aussage ist äquivalent zum Beweis auf Seite 30

Diese Momente werden als normalisierte Momente bezeichnet. Die Normalisierung bezüglich Skalierung kann allerdings auch wie folgt erzielt werden [53]. Sei $\acute{f}(\acute{x}, \acute{y})$ das um α skalierte Bild mit folgender Transformation

$$\begin{vmatrix} \acute{x} \\ \acute{y} \end{vmatrix} = \begin{vmatrix} \alpha & 0 \\ 0 & \alpha \end{vmatrix} \begin{vmatrix} x \\ y \end{vmatrix}$$

dann ist $\acute{f}(\acute{x}, \acute{y}) = f(\alpha x, \alpha y) = f(x, y)$ und $\acute{x} = \alpha x, \acute{y} = \alpha y$. Man erhält

$$\begin{aligned} m'_{pq} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \acute{x}^p \acute{y}^q \acute{f}(\acute{x}, \acute{y}) d\acute{x} d\acute{y} \\ &= \alpha^{p+q+2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \end{aligned}$$

Dann kann die Skalierungsinvarianz wie folgt erreicht werden. Man erhält für die skalierten Momente:

$$\begin{aligned} m'_{pq} &= \alpha^{p+q+2} m_{pq} \\ \mu'_{pq} &= \alpha^{p+q+2} \mu_{pq} \end{aligned}$$

es folgt $\mu'_{00} = \alpha^{p+q+2} \mu_{00}$ und somit

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \gamma = \frac{p+q}{2} + 1 \quad p+q = 2, 3, \dots$$

Damit ist die Skalierungsinvarianz erreicht, da:

Beweis 2

$$\acute{\eta}_{pq} = \frac{\acute{\mu}_{pq}}{\acute{\mu}_{00}^\gamma} = \frac{\alpha^{p+q+2} \mu_{pq}}{\alpha^{2\gamma} \mu_{00}^\gamma} = \frac{\mu_{pq}}{\mu_{00}^\gamma} = \eta_{pq}$$

□

Die Skalierungsinvarianz kann also bei Momenten durch Division der Momente durch eine normalisierte Funktion erzielt werden, die den Skalierungseffekt neutralisiert. Ein weiteres Problem ist die Position des interessierenden Objektes innerhalb der Bildmatrix. Wird z.B. die Fläche der Bildmatrix bei gleichbleibender Größe des Musters verändert, so ändern sich meist auch die Koordinaten dieses Objektes (insbesondere, wenn die Bildmatrix z.B. in $[-1, 1]$ skaliert ist). Dies verändert die berechneten Momente und kann durch eine Boundingbox-Korrektur behoben werden (vgl. 5.2.1). Im folgenden nur ein kurzes Beispiel, welches das Problem verdeutlichen soll. Die Abbildung auf der nächsten Seite zeigt in den Teilbildern 4.1(a) und 4.1(c) die Zahl 5 mit korrekter Boundingbox (32×32 Pixel) bzw. mit inkorrektter Boundingbox (256×256 Pixel) [Die Zahl 5 nutzt dabei in beiden Fällen etwa 32×32 Pixel und ist jeweils

zum Bildmittelpunkt zentriert]. In den Bildern 4.1(b) und 4.1(c) folgt eine jeweils mittels Zernike-Momenten aus den Momenten rekonstruierte Darstellung des Symbols. Exemplarisch sind hier noch die Werte für ZM_{20} angegeben:

$$ZM_{20 \text{ korrekt}} = -0.841947 \quad ZM_{20 \text{ inkorrekt}} = -2.96035$$

Wie zu sehen ist, kann eine sehr fehlerhafte Boudingbox zu völlig anderen Moment-

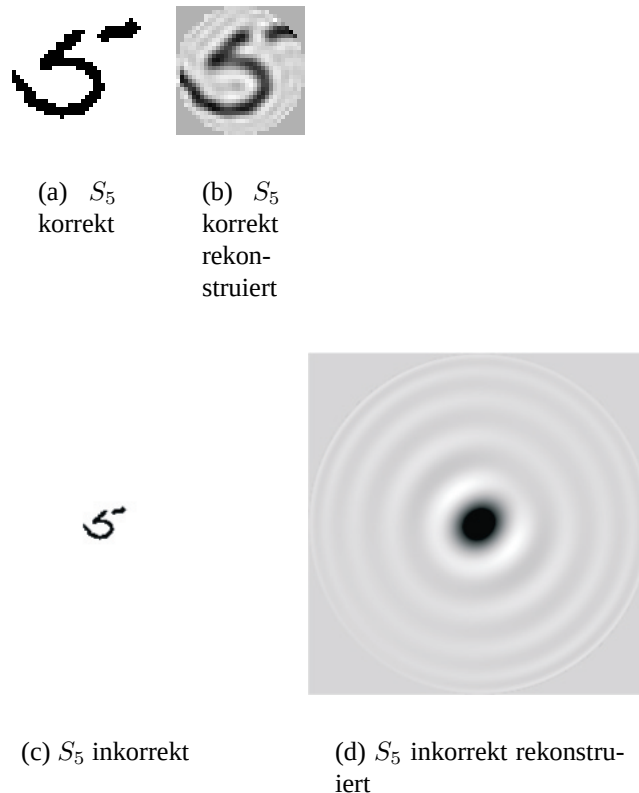


Abbildung 4.1: Darstellung der Auswirkung einer fehlerhaften Boundingbox

werten führen und damit auch eine, später auf diesen Momenten basierende Klassifikation massiv stören oder sogar unmöglich machen.

Damit sind bereits zwei wesentliche Invarianzen erreicht:

1. Invarianz bezüglich Größenänderung.
2. Invarianz gegenüber Verschiebung.

Eine weitere meist wünschenswerte Invarianz ist die Rotation. Die Rotation eines Bildes entspricht der folgenden Koordinatentransformation:

$$\begin{vmatrix} \dot{x} \\ \dot{y} \end{vmatrix} = \begin{vmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{vmatrix} \begin{vmatrix} x \\ y \end{vmatrix}$$

Von der Gruppentheorie ergibt sich, daß die einzigen eindimensionalen Invarianten für 2D Rotationen durch den Betrag von $h_k(r)$ gegeben sind, wobei die zirkulären Fouriertransformation eines Bildes $f(r, \theta)$ in Polarkoordinaten durch die folgende Gleichung gegeben ist:

$$h_k(r) = \int_0^{2\pi} f(r, \theta) e^{jk\theta} d\theta, \quad k \in \mathbb{N}$$

Diese Invarianz kann auf sehr verschiedene Arten erzielt werden, so sind z.B. die später vorgestellten Zernike-Momente betragsmäßig, wie in Gleichung 4.3 angedeutet, rotationsinvariant. Es ist allerdings auch möglich, die Hauptachsen eines Musters auf Basis von niedrigen Momenten zu bestimmen, und damit den Rotationswinkel θ des Musters zu berechnen. Ist dieses θ bekannt, kann man in einem Vorverarbeitungsschritt das Muster entsprechend rotieren und die Muster bezüglich Rotation normieren. Ein solches Verfahren wird bei Hu in Abschnitt IV/B vorgestellt ([26], Seite 184).

Es sind noch weitere Normalisierungen der Merkmale möglich, um z. B. Invarianz bzgl. Verzerrung oder Kontrast und Helligkeitsunterschieden zu erreichen. Auf diese Ansätze wird hier jedoch nicht weiter eingegangen. Es sei an dieser Stelle auf Arbeiten von Reiss [53], Rothe et al. [55] und Flusser et al. [21] verwiesen.

4.4 Momentinvarianten von Hu und Reiss

4.4.1 Momentinvarianten von Hu

Die Momente von Hu nehmen eine Sonderstellung unter den Momenten ein, daß sie die ersten Momente waren, die in der Schrifterkennung Anwendung fanden und alle anderen Arbeiten zu Momenten auf ihnen und dem von Hu erarbeiteten Fundamentaltheorem der Momenteninvarianten (auf Seite 18) aufbauen. Da das Fundamentaltheorem von Hu einen kleinen Fehler enthielt, wurde 1992 von Reiss eine revidierte Fassung angegeben, die in dieser Arbeit anstelle der Originaldefinition angegeben wurde.

Die bereits 1962 von Hu vorgeschlagenen Momente zeichnen sich durch Invarianz bezüglich Position, Größe und Orientierung (Rotation) aus. Diese Momente sind einfache statistische Maße für die Pixelverteilung um das Massezentrum des Bildes bzw. Schriftzeichens $f(x, y)$ und gestatten die Erfassung der globalen Form des Zeichens.

Die 7 Momente von Hu basieren auf geometrischen Momenten und sind gegeben durch die folgenden Gleichungen.

$$\begin{aligned}
 \phi(1) &= \mu_{20} + \mu_{02} \\
 \phi(2) &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\
 \phi(3) &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\
 \phi(4) &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\
 \phi(5) &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12}) \\
 &\quad [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - 3\mu_{03})(\mu_{21} + \mu_{03}) \\
 &\quad [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\
 \phi(6) &= (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\
 \phi(7) &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\
 &\quad - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]
 \end{aligned}$$

Dabei sei μ_{pq} gegeben durch:

$$\begin{aligned}
 \mu_{pq} &= \frac{1}{\left(\sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{i=1}^N (y_i - \bar{y})^2\right)^{\frac{p+q}{2}+1}} \cdot \sum_{i=1}^N (x_i - \bar{x})^p (y_i - \bar{y})^q f(x, y) \\
 \bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \\
 N &= \text{Anzahl der schwarzen Bildpunkte}
 \end{aligned}$$

Die Momente von Hu lassen sich am einfachsten durch Verwendung komplexer Momente c_{pq} herleiten, die erstmalig von Davis [27] gegeben wurden. Komplexe Momen-

te sind dabei definiert als:

$$c_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + jy)^p (x - jy)^q dx dy \quad \text{kartesisch} \quad (4.3)$$

$$c_{pq} = \int_0^{2\pi} \int_0^{\infty} r^{p+q+1} e^{j(p-q)\theta} f(r, \theta) dr d\theta \quad \text{polar} \quad (4.4)$$

$$= \int_0^{\infty} r^{p+q+1} \left[\int_0^{2\pi} f(r, \theta) e^{j(p-q)\theta} d\theta \right] dr \quad k = p - q \quad (4.5)$$

$$= \int_0^{\infty} r^{p+q+1} h_k(r) dr \quad (4.6)$$

Daran erkennt man, daß $|c_{pq}|$ rotationsinvariant ist und daß c_{pq} die Projektion der zirkulären Fouriertransformation des Bildes $h_{p-q}(r)$ auf die Funktion r^{p+q+1} ist. Anders gesagt, wird die kontinuierliche eindimensionale Funktion h_{p-q} auf einen endlichen Vektor mit den Elementen c_{pq} abgebildet. Besitzt ein Bild n -fache Rotationssymmetrie sind alle c_{pq} s, für die $p - q$ nicht durch n teilbar, ist identisch 0 [53]. Zum Beispiel „I“ ist zweifach rotationssymmetrisch. Von Gleichung 4.3 und der Definition der Momente ergibt sich für ein zentriertes Bild:

$$\begin{aligned} c_{11} &= \mu_{20} + j\mu_{02} \\ c_{21} &= (\mu_{30} + \mu_{12})^2 + j(\mu_{21} + \mu_{03})^2 \end{aligned}$$

Ebenso ergibt sich aus der Gleichung 4.3, daß die Verwendung regulärer oder zentraler Momente zu einer Einschränkung auf radiale Momente der $p + q + 1^{\text{ten}}$ Ordnung führt. Wenn Merkmale gewünscht sind, die rotationsinvariant sind, aber nicht invariant bezüglich Verschiebung, dann kann die Einschränkung auf Potenzen von r entfallen und man kann jede beliebige radiale Gewichtsfunktion verwenden, da der Betrag von $h_k(r)$ rotationsinvariant ist.

Neuere Untersuchungen der Hu Momente von Flusser [19] zeigen allerdings, daß nur ein Teil der Hu Momente orthogonal zueinander ist und die restlichen sich als Linearkombination aus den anderen ergeben. So gilt z.B. folgende Beziehung

$$\phi_3 = \frac{\phi_5^2 + \phi_7^2}{\phi_4^3}$$

Dies bedeutet, daß einige der Hu-Momente keinen echten Zugewinn an Informationen ergeben und informationstheoretisch redundant sind (siehe auch [69]).

4.4.2 Affine Momentinvarianten von Reiss

Auch von Reiss wurde eine Menge von vier Momentinvarianten in [52] angegeben. Diese wurden in Abschnitt 8 in Anlehnung an die Arbeit von Chim et al. [71] mit den

Momenten von Hu in einem Merkmalsvektor kombiniert. Die affinen Invarianten von Reiss sind invariant unter affinen Transformationen und gegeben durch:

$$\begin{aligned}
 I_1 &= \frac{1}{\mu_{00}^4} (\mu_{20}\mu_{02} - \mu_{11}^2) \\
 I_2 &= \frac{1}{\mu_{00}^{10}} (\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 + 4\mu_{03}\mu_{21}^3 - 3\mu_{21}^2\mu_{12}^2) \\
 I_3 &= \frac{1}{\mu_{00}^7} (\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) + \mu_{02}(\mu_{30}\mu_{21} - \mu_{21}^2)) \\
 I_4 &= \frac{1}{\mu_{00}^{11}} (\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} \\
 &\quad + 9\mu_{20}^2\mu_{02}\mu_{12}^2 + 12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} + 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} \\
 &\quad - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} - 8\mu_{11}^3\mu_{30}\mu_{03} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} \\
 &\quad + 9\mu_{20}\mu_{20}^2\mu_{21}^2 + 12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} \\
 &\quad + \mu_{02}^3\mu_{30}^2)
 \end{aligned}$$

In den letzten Jahren sind noch weitere Momentdefinitionen entstanden, die auf geometrischen Momenten aufbauen. Im folgenden sind die Definitionen zu *gewichteten zentralen Momenten* und zu *Tsirikolias-Mertzios Momenten* angegeben, die ebenfalls in dieser Arbeit berücksichtigt wurden.

4.5 Gewichtete zentrale Momente

Diese Momente gehen auf Arbeiten von Balslev, Doring und Eriksen [5] zurück und sind durch folgendes Problem motiviert. In vielen Mustererkennungsprozessen werden zentrale Momente verwendet, um Translationsinvarianz zu erreichen. Es zeigt sich jedoch, daß die Klassifikation von Mustern, deren Formvariation sich vor allem im *center of gravity* zeigt, für zentrale Momente höherer Ordnung schnell ungenau wird. Auch treten Probleme bezüglich Rauschen auf, die durch die Einführung von Gewichtungen reduziert werden können. Die bei den gewichteten zentralen Momenten (GZM) eingeführte Gewichtung hat deshalb zum Ziel, die Bildinformation nahe dem *center of gravity* höher zu gewichten und Rauschen im Randbereich entsprechend niedriger. Die GZMs sind dabei wie folgt definiert:

$$\mu_{pq}^* = \sum_{\mathbb{R}^2} F(x, y) \left[x - \frac{m_{10}}{m_{00}} \right]^p \left[y - \frac{m_{01}}{m_{00}} \right]^q$$

Mit dem Ziel, die Gewichte der Regionen nahe zum center of gravity zu erhöhen, wird eine Funktion $F(x, y)$ mit der Form einer 2D Lorentzian verwendet:

$$F(x, y) = \frac{1}{1 + \alpha^2((x - \bar{x})^2 + (y - \bar{y})^2)}$$

α ist dabei ein noch zu bestimmender Parameter, der typischer Weise im Bereich $0 < \alpha < \frac{10}{R_G}$ gewählt wird, mit R_G als Radius der Drehung $R_G = \sqrt{\frac{\mu_{20} + \mu_{02}}{\mu_{00}}}$.

4.6 Tsirikolias-Mertzios Momente

Diese Momente (TMM) sind erstmalig in einem Artikel von Tsirikolias und Mertzios [68] angegeben und wurden in dem OCR System von Chim et al. [71] verwendet. Die TMMs wurden im Gegensatz zu den normalisierten Momenten im Abschnitt auf Seite 22 bezüglich des Mittelwertes und der Standardabweichung normalisiert. Sie sind laut Tsirikolias et al. [68] im Vergleich mit den normalisierten Momenten (NM) weniger sensibel bezüglich Rauschen und zeigen bessere Klassifikationseigenschaften. Die Definition ist wie folgt gegeben:

$$tm_{pq} = \frac{1}{LM} \sum_{i=1}^L \sum_{j=1}^M \left[\frac{x_i - \bar{x}}{\sigma_x} \right]^p \left[\frac{y_j - \bar{y}}{\sigma_y} \right]^q f(x_i, y_j)$$

Dabei ist die Standardabweichung definiert als:

$$\sigma_x = \sqrt{\left(\frac{1}{LM} \left(\sum_{i=1}^L \sum_{j=1}^M (x_i - \bar{x})^2 f(x_i, y_j) \right) \right)}$$

$$\sigma_y = \sqrt{\left(\frac{1}{LM} \left(\sum_{i=1}^L \sum_{j=1}^M (y_j - \bar{y})^2 f(x_i, y_j) \right) \right)}$$

und der Mittelwert wie üblich als:

$$\bar{x} = \frac{1}{L} \sum_{i=1}^L x_i \quad \bar{y} = \frac{1}{M} \sum_{j=1}^M y_j$$

Die TMMs sind sowohl translations- als auch skalierungsinvariant

Beweis 3 (Translationsinvarianz der TMMs)

Sei tm'_{pq} das Moment des um α bzw. β verschobenen Bildobjektes, dann gilt:

$$\begin{aligned} tm'_{pq} &= \frac{1}{LM} \sum_{i=1}^L \sum_{j=1}^M \left[\frac{x'_i - \bar{x}'_i}{\sigma'_x} \right]^p \left[\frac{y'_j - \bar{y}'_j}{\sigma'_y} \right]^q f(x', y') \\ &= \frac{1}{LM} \sum_{i=1}^L \sum_{j=1}^M \left[\frac{x_i + \alpha - (\bar{x}_i + \alpha)}{\sigma_x} \right]^p \left[\frac{y_j + \beta - (\bar{y}_j + \beta)}{\sigma_y} \right]^q f(x, y) \cdot 1 \\ &= tm_{pq} \end{aligned}$$

mit Jacobian der Transformation = 1 und

$$\begin{aligned} \bar{x}' &= \bar{x} + \alpha && \text{Eigenschaft des Mittelwerts} \\ \text{Var}(x') &= \text{Var}(x) && \text{Eigenschaft der Varianz} \end{aligned}$$

□

Beweis 4 (Skalierungsinvarianz der TMMs)

Sei tm'_{pq} das Moment des um α bzw. β skalierten Bildobjektes dann gilt:

$$\begin{aligned} tm'_{pq} &= \frac{1}{L'M'} \sum_{i=1}^{L'} \sum_{j=1}^{M'} \left[\frac{x'_i - \bar{x}'_i}{\sigma'_x} \right]^p \left[\frac{y'_j - \bar{y}'_j}{\sigma'_y} \right]^q f(x', y') \\ &= \frac{1}{LM\alpha\beta} \sum_{i=1}^{L\alpha} \sum_{j=1}^{M\beta} \left[\frac{\alpha(x_i - \bar{x}_i)}{\alpha\sigma_x} \right]^p \left[\frac{\beta(y_j - \bar{y}_j)}{\beta\sigma_y} \right]^q f(x, y) \cdot \beta \cdot \alpha \\ &= tm_{pq} \end{aligned}$$

mit Jacobian der Transformation = $\alpha\beta$ und

$$\text{Var}(\alpha \cdot x) = \alpha^2 \text{Var}(x) = \sigma'^2 \quad \text{E}(\alpha x) = \alpha \text{E}(x)$$

□

Kapitel 5

Orthogonale Momente

Teague [65] führte die Idee der orthogonalen Momente ein, um Rotationsinvarianz zu erhalten [52]. Reguläre Momente und somit zentrale Momente können als eine Projektion des Bildes auf monomiale Polynome $x^p y^q$ angesehen werden. Die, von diesen Funktionen gebildete Basis ist jedoch nicht orthogonal (Weierstrass Approximationstheorem), was im Sinne einer Datenreduktion (Minimalität der Basis) wünschenswert wäre. Alternativ können deshalb auch andere Funktionen zur Projektion verwendet werden.

5.1 Legendre-Momente

Die Legendre-Momente gehen z.B. auf Arbeiten von Teague zurück [65] und werden durch eine Projektion auf eine orthogonale Menge von Legendre-Polynomen berechnet. Dabei sind die p^{ten} Legendre-Polynome definiert als:

$$L_p(x) = \frac{1}{2^p p!} \frac{d^p}{dx^p} (x^2 - 1)^p, x \in [-1, 1]$$

Da die Legendre-Momente über dem Inneren des Einheitskreises definiert sind, muß die Bildfunktion über der die Momente berechnet werden sollen, entsprechend in $-1 < x, y < 1$ skaliert werden. Für die Legendre-Polynome läßt sich für die Orthogonalität das folgende Skalarprodukt angeben:

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

Wird nun die Bildfunktion $f(x, y)$ als stückweise, kontinuierlich beschränkte Funktion definiert, ergeben sich die Legendre-Momente $(p+q)^{\text{ter}}$ Ordnung zu:

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L_p(x) L_q(y) f(x, y) dx dy$$

Mit dieser orthogonalen Basis läßt sich die Anzahl zu speichernder Koeffizienten reduzieren, indem die meisten Terme höherer Ordnung entfallen können (die Koeffizienten höherer Ordnung bieten nur einen geringen Informationsgehalt und sind für die Form des Objektes unerheblich) Außerdem beeinflußt die Eliminierung höherer Koeffizienten die anderen Koeffizienten nicht.

Die Legendre-Momente können wie folgt zu den normalen geometrischen Momenten (m_{jk}) in Bezug gebracht werden. Zunächst lassen sich die Legendre-Polynome in folgender Form darstellen:

$$P_p(x) = \sum_{j=0}^p C_{pj} x^j$$

Dabei sind die Koeffizienten C_{pj} in Referenz [10] gegeben. Man erhält dann für die diskrete Form der Legendre-Polynome die folgende Darstellung:

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{4} \sum_{j=0}^p \sum_{k=0}^q C_{pj} C_{qk} m_{jk}$$

Die Funktionen $P_p(x)$ bilden eine vollständig orthogonale Basis über dem Einheitskreis und die entsprechende Bildfunktion $f(x, y)$ kann auch hier wieder durch eine abgeschnittene Reihe von ausreichend vielen Legendre-Momenten approximiert werden.

$$f(x, y) = \sum_{p=0}^{N_{\max}} \sum_{q=0}^p \lambda_{p-q,q} P_{p-q}(x) P_q(y)$$

Für die Berechnung der Legendre-Momente wurde in dieser Arbeit eine schnelle, rekursive Berechnungsmethode nach Mukundan et al. [44] verwendet. Dabei ist λ_{pq} durch die folgende Gleichung definiert:

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{(N-1)^2} \sum_{i=1}^N \sum_{j=1}^N P_p(x_i) P_q(y_j) f(x_i, y_j) \quad p, q = 0, 1, 2, \dots, \infty$$

mit $P_p(x)$ gegeben durch:

$$\begin{aligned} P_p(x) &= [(2p-1)xP_{p-1}(x) - (p-1)P_{p-2}(x)]/p \quad p > 1 \\ P_0(x) &= 1 \\ P_1(x) &= x \end{aligned}$$

Translations und Skalierungsinvarianz Legendre-Momente sind in ihrer ursprünglichen Definition lediglich orthogonal und besitzen keine Invarianzeigenschaften. Dies läßt sich jedoch relativ leicht durch Verwendung zentraler Legendre-Momente und

durch eine Skalierung entsprechend dem Verfahren auf Seite 36 beheben. Im folgenden die zentralen Legendre-Momente:

$$\lambda_{pq} = \frac{(2p+1)(2q+1)}{(N-1)^2} \sum_{i=1}^N \sum_{j=1}^N P_p(x_i - \bar{x}) P_q(y_j - \bar{y}) f(x_i, y_j) \quad p, q = 0, 1, 2, \dots, \infty$$

5.2 Zernike-Momente

Zernike-Momente wurden von Teague [65] bereits 1980 als eine Form von orthogonalen Momenten vorgeschlagen, um sowohl das Problem der Redundanz (vgl. Geometrische Momente), als auch das Problem der Rotationsinvarianz zu lösen. Die Zernike-Momente sind komplexwertige Momente und verwenden als radiale Polynome die Zernike-Polynome (V_{pq}):

$$Z_{pq} = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{pq}(r, \theta)]^* f(r \cos \theta, r \sin \theta) r \, dr \, d\theta$$

mit $p = 0, \dots, \infty$, $|q| \leq p$ und $p - |q|$ gerade. Die Zernike-Polynome sind dabei die einzigen Polynome in x und y , die eine orthogonale Basis für die Menge der komplexwertigen Funktionen über dem Einheitskreis $x^2 + y^2 \leq 1$ definieren, z.B.

$$\langle V_{nm}, V_{pq} \rangle = \int \int_{x^2+y^2 \leq 1} [V_{nm}(x, y)]^* V_{pq}(x, y) \, dx \, dy = \delta_{np} \delta_{mq}$$

Die Funktion muß dabei ein Polynom in x und y sein, um Rotations- und Skalierungs-invarianz zu erreichen. (Translation wird dabei erreicht, indem die regulären Momente durch zentrale Momente ersetzt werden) Im allgemeinen werden die Zernike-Momente allerdings über Polarkoordinaten definiert. Dabei sind die Basisfunktionen der Ordnung p und Repetition q gegeben durch:

$$\begin{aligned} V_{pq}(r, \theta) &= V_{pq}(r \cos \theta, r \sin \theta) = R_{pq}(r) e^{j r \theta} \\ \theta &= \operatorname{atan} \left(\frac{y}{x} \right) \quad r = \sqrt{x^2 + y^2} \end{aligned}$$

und die radialen Polynome $R_{pq}(r)$ sind realwertig wie folgt definiert:

$$R_{pq} = \sum_{s=0}^{(p-|q|)/2} (-1)^s \frac{(p-s)!}{s! \left(\frac{p+|q|}{2} - s\right)! \left(\frac{p-|q|}{2} - s\right)!} r^{p-2s}$$

Es ergeben sich insgesamt $\frac{1}{2}(n+1)(n+2)$ linear unabhängige Polynome vom Grad $\leq n$ auf Grund der Einschränkung $p - |q|$ gerade. [66]

Das Zernike-Moment einer Bildfunktion f mit Ordnung p und Repetition q ist dabei die Länge der orthogonalen Projektion von f auf die Zernike-Basisfunktion V_{pq}

$$Z_{pq}^f = \frac{\langle f, V_{pq} \rangle}{\langle V_{pq}, V_{pq} \rangle}$$

mit

$$\langle f, g \rangle = \int \int_{r \leq 1} f(x, y) g^*(x, y) dx dy$$

Zernike-Momente wurden bereits sehr vielfältig zur Schriftzeichenerkennung von soliden, binären Symbolen eingesetzt. Sie sind allerdings auch bei Grauwertbildern einsetzbar. Es ist damit möglich sowohl rotationsinvariante, als auch rotationsvariante Merkmale zu extrahieren, im letzteren Fall wird dabei der imaginäre Anteil ignoriert. Zernike-Momente sind Projektionen des Eingabebildes auf den Raum, der durch die orthogonalen Vektoren V (wie oben definiert) aufgespannt wird. Für ein digital, diskret vorliegendes Bild können die Zernike-Momente der Ordnung p und Repetition q wie folgt berechnet werden:

$$Z_{pq} = \frac{p+1}{\pi} \sum_x \sum_y f(x, y) [V_{pq}(x, y)]^*$$

Die Zernike-Momente können zu den normalen geometrischen Momenten in Bezug gebracht werden. Zunächst lassen sich die radialen Polynome wie folgt darstellen:

$$R_{pq}(r) = \sum_{k=1}^p B_{pqk} r^k \quad p - q \text{ gerade}$$

Die Koeffizienten B_{pqk} sind in Referenz [10] gegeben. Man erhält dann für die diskrete Form der Zernike-Momente die folgende Darstellung:

$$Z_{pq} = \frac{(p+1)}{\pi} \sum_{k=q}^p \sum_{j=0}^l \sum_{m=0}^q (-i)^m \binom{l}{j} \binom{q}{m} B_{pqk} m_{k-2j-q-m, 2j+q-m} \quad \text{mit } l = \frac{k-q}{2}$$

Dabei sind $m_{k-2j-q-m, 2j+q-m}$ durch die normalisierten Momente gegeben.

Der Bereich des Bildes, der sich innerhalb des Einheitskreises (der Projektion) befindet, kann wie folgt approximiert werden:

$$f(x, y) = \lim_{N \rightarrow \infty} \sum_{p=0}^N \sum_q Z_{pq} V_{pq}(x, y)$$

wobei die zweite Summe über all denen $|p| \leq q$ berechnet wird, für die gilt: $p - |q|$ ist gerade. Der Betrag $|Z_{pq}|$ ist rotationsinvariant. Translationsinvarianz kann dabei durch *center of gravity* Skalierung und Skalierungsinvarianz durch die, auf Seite 36 dargestellten beiden Verfahren erzielt werden. Ein Beispiel aus [67], welches die Anwendung der Zernike-Moments verdeutlicht, ist auf der folgenden Seite dargestellt.

Die Bilder $|I_n(x, y)|$ ermittelt als $|I_n(x, y)| = \left| \sum_q Z_{pq} V_{pq}(x, y) \right|$ für die Zeichen '4' und '5' zeigen, daß die extrahierten Merkmale sehr unterschiedlich für die Momente 3. und 4. Ordnung sind. Die Ordnungen 1 und 2 scheinen dabei die Orientierung, Breite und Höhe festzulegen. Dabei zeigt die Rekonstruktion des gleichen Symbols mit verschiedenen Ordnungen, daß erst ab der 6. Ordnung eine ausreichende Differenzierung zu anderen Symbolen erzielt wird und erst ab der 8. - 11. Ordnung die Feinstruktur des Symbol ausreichend erfaßt wird.

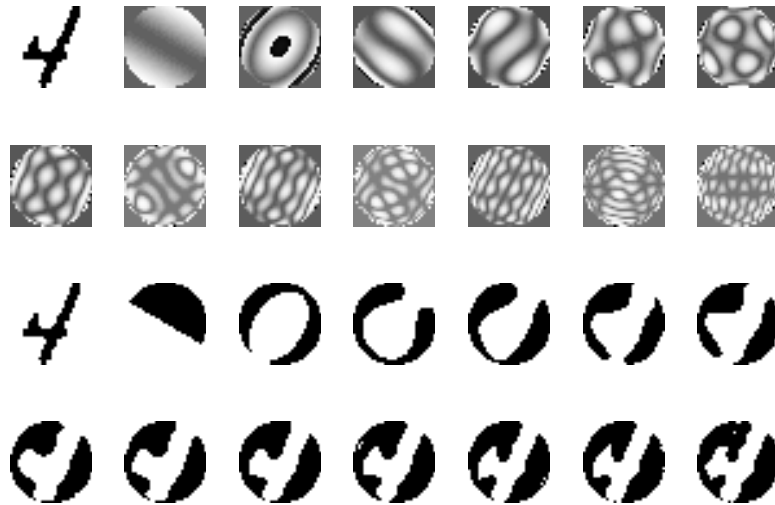


Abbildung 5.1: Bilder abgeleitet von Zernike Momenten. Zeilen 1-2: Eingabebild der Ziffer '4', und Beträge der Zernike-Momente $I(x, y)$ von Ordnung 1-13. Die Bilder wurden bzgl. Histogramm korrigiert. Zeilen 3-4: Eingabebild der Ziffer '4' und Bildrekonstruktion durch die Zernike-Momente bis zur 13. Ordnung

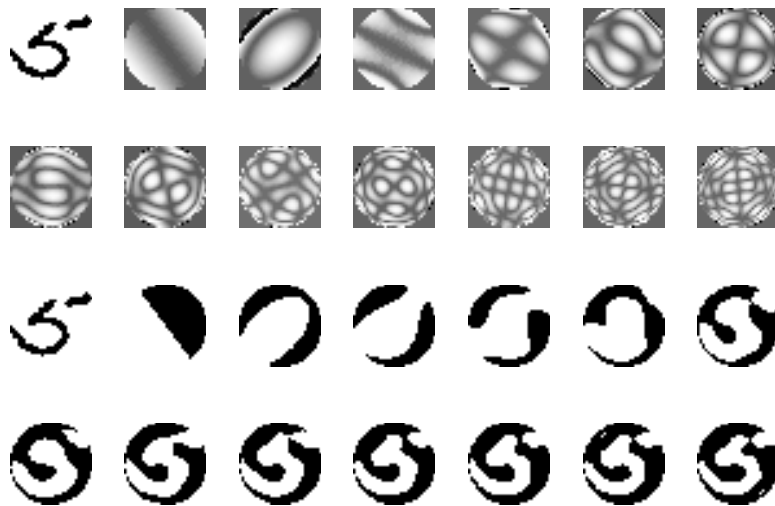


Abbildung 5.2: Bilder abgeleitet von Zernike Momenten. Zeilen 1-2: Eingabebild der Ziffer '5', und Beträge der Zernike-Momente $I(x, y)$ von Ordnung 1-13. Die Bilder wurden bzgl. Histogramm korrigiert. Zeilen 3-4: Eingabebild der Ziffer '5' und Bildrekonstruktion durch die Zernike-Momente bis zur 13. Ordnung

5.2.1 Skalierungsinvarianz

Die erste Methode, um Skalierungsinvarianz zu erreichen, ist die Ausdehnung oder Stauchung jedes Objektes, so daß die Pixelkoordinaten des Objektes in den Bereich des Einheitskreises transformiert werden. Dabei wird die Zentroidbegrenzung (*centroid bounding circle*) des Objektes als Einheitskreis gewählt. Für die Einpassung des Objektes in diesen Kreis wird damit die Berechnung des passenden Radius nötig. Das dafür nötige Verfahren ist im folgenden angegeben [32]

- a Ermittlung des Begrenzungsrahmens (BR) und des Zentroiden (x_c, y_c) des entsprechenden Musters
- b Setze $r_{\max}^2 = 0$
- c Innerhalb des BR wird nun für jede Zeile die folgende Operation durchgeführt:
 - Für jede Zeile y sucht man den am weitesten links bzw. rechts stehenden Objektpixel x_1, x_2
 - Dann berechnet man $x'_i = |x_i - x_c|$, $i = 1, 2$ $x_0 = \max[x'_1, x'_2] + 0.5$
 $y_0 = |y - y_c| + 0.5$ $r^2 = x_0^2 + y_0^2$
 - falls $r^2 > r_{\max}^2$ ist $r_{\max}^2 = r^2$
- d Der maximale Radius ist dann $r_{\max}$, dieser wird als Skalierungsfaktor verwendet

Der zweite Ansatz ist die Fourier-Mellin (FM) Skalierung. Dabei wird ausgenutzt, daß sich Zernike-Momente durch eine Linearkombination von FM-Momenten (FMM) ausdrücken lassen. Wenn ein Muster $f(r, \theta)$ durch einen Faktor k skaliert wird, ändert sich sein FMM wie folgt:

$$F'_{pq} = \int_0^{2\pi} \int_0^\infty r^p f(r/k, \theta) e^{-iq\theta} r dr d\theta = k^{p+2} F_{pq}$$

Die Normalisierung geschieht nun durch Verwendung der dominanten Merkmalskomponente $F_{p,0}$, man erhält:

$$F_{pq}^{\text{normalisiert}} = \frac{F'_{pq}}{F'_{p,0}} = \frac{F_{pq}}{F_{p,0}}$$

Die beiden Methoden sind nicht in jedem Fall gut zur Normierung geeignet. Die erste Methode hat Defizite bei sehr kleinen Objekten, da die Ermittlung des BR in diesem Fall erhebliche Störungen in das Bild und die Merkmale einführt. Das zweite Verfahren ist insbesondere bei starker Variabilität in den Mustern (wie z. B. bei handschriftlichen Symbolen) kritisch, da die Skalierung dann sogar zu einer noch größeren Störung führen kann [32]. In dieser Arbeit wurde die erste Methode zur Skalierung verwendet.

5.2.2 Rotationsinvarianz

Wird ein Muster f um einen Wert α rotiert, entsteht das transformierte Bild $\acute{f}$. In Polarkoordinaten gilt dann die folgende Beziehung:

$$\acute{f}(r, \theta) = f(r, \theta - \alpha)$$

Die Formel zur Berechnung der Zernike-Momente kann in Polarkoordinaten wie folgt beschrieben werden

$$\int_A \int \phi(x, y) dx dy = \int_G \int \phi[p(r, \theta), q(r, \theta)] \frac{\partial(x, y)}{\partial(r, \theta)} dr d\theta$$

dabei definiert $\frac{\partial(x, y)}{\partial(r, \theta)}$ die Jakobideterminante der Transformation mit der Transformationsmatrix:

$$\frac{\partial(x, y)}{\partial(r, \theta)} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix}$$

Für den Fall, das $x = r \cos \theta$ und $y = r \sin \theta$ ergibt sich die Jakobiandeterminante zu r und es gilt:

$$\begin{aligned} Z_{pq} &= \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{pq}(r, \theta)]^* f(r \cos \theta, r \sin \theta) r dr d\theta \\ &= \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 [R_{pq}(r) \exp(-jq\theta)] f(r, \theta) r dr d\theta \end{aligned}$$

Die Zernike-Momente des rotierten Bildes in Polarkoordinaten ergeben sich dann zu ($\theta_1 = \theta - \alpha$):

$$\begin{aligned} Z_{pq} &= \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 [R_{pq}(r) \exp(-jq\theta)] f(r, \theta - \alpha) r dr d\theta \\ &= \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 [R_{pq}(r) \exp(-jq(\theta_1 + \alpha))] f(r, \theta_1) r dr d\theta_1 \\ &= \left[\frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 [R_{pq}(r) \exp(-jq\theta_1)] f(r, \theta_1) r dr d\theta_1 \right] \exp(-jq\alpha) \\ &= Z_{pq} \exp(-jq\alpha) \end{aligned}$$

Somit bedeutet eine Rotation stets nur eine Verschiebung im Phasenraum und man sieht, daß der Betrag des Zernike-Moments des rotierten Bildes identisch mit dem Betrag des Zernike-Moments vor der Rotation ist. Somit kann $|Z_{pq}|$ als rotationsinvariantes Merkmal des betrachteten Bildes verwendet werden. Des weiteren gilt, wenn $Z_{p,-q} = Z_{pq}^*$, dann $|Z_{pq}| = |Z_{p,-q}|$. Damit kann man sich bei den Berechnungen der Merkmale auf positive q ($q \geq 0$) beschränken.

5.2.3 Konstruktion von Invarianten mit Zernike-Momenten

Die genaue Konstruktionsanweisung kann bei Teague [65] auf den Seiten 926-928 nachgelesen werden und wird hier nicht noch einmal aufgeführt. Im folgenden wird lediglich verkürzt angegeben, welche Invarianten verwendet wurden. Die komplexen Zernike-Momente sind als Norm stets rotationsinvariante reelle Merkmale. Man kann also alle $|Z_{pq}|$ als invariante Merkmale (bzgl. der Rotation) verwenden. Allerdings ist das Merkmal 0. Ordnung A_{00} stets konstant 1 (aufgrund der Normierung) und sollte somit nicht verwendet werden. Es lassen sich allerdings auch weitere, invariante Merkmale, als Kombination von anderen, sowie pseudoinvariante Merkmale erzeugen. Dies wird bei Teague entsprechend erklärt. Im folgenden sind exemplarisch die verwendeten Merkmale bis zur 3. Ordnung nach Teague angegeben:

$$\begin{aligned} S_1 &= A_{20} & S_2 &= |A_{22}|^2 & (2. \text{ Ordnung}) \\ S_3 &= |A_{33}|^2 & S_4 &= |A_{31}|^2 & (3. \text{ Ordnung}) \end{aligned}$$

5.3 Pseudo-Zernike-Momente

Im Gegensatz zu den Zernike-Momenten basieren die Pseudo-Zernike-Momente auf einer Abwandlung der ursprünglichen Zernike-Polynome, die in [7] vorgeschlagen wurden und ermöglichen die Berechnung von $(n+1)^2$ linear unabhängigen Polynomen vom Grad $\leq n$. [66]. Die Pseudo-Zernike-Polynome bilden eine vollständige Menge von orthogonalen Funktionen über dem Einheitskreis. Pseudo-Zernike-Momente sind ähnlich wie die Zernike-Momente definiert. Die Pseudo-Zernike-Basisfunktion der p ten Ordnung und Repetition q ist gegeben durch:

$$W_{pq}(r \cos \theta, r \sin \theta) = R_{pq} e^{iq\theta}$$

mit $p \geq 0, q \leq p, p, q \in \mathbb{N}$. Die radialen Polynome können dabei wie folgt berechnet werden:

$$R_{pq} = \sum_{s=0}^{(p-|q|)} (-1)^s \frac{(2 * p + 1 - s)!}{s!(p - |q| - s)!(p + |q| + 1 - s)!} r^{p-s}$$

Des weiteren gilt:

$$\langle W_{nm}, W_{pq} \rangle = \frac{\pi}{n+1} \delta_{np} \delta_{mq}$$

Damit sind die Pseudo-Zernike-Momente definiert als:

$$P_{pq} = \frac{\pi}{n+1} \langle f, W_{pq} \rangle = \frac{\pi}{n+1} \int \int_{\mathbb{R}^2} f(x, y) W_{pq}^*(x, y) dx dy$$

Für ein digital, diskret vorliegendes Bild können die Pseudo-Zernike-Momente der

Ordnung p und Repetition q wie folgt berechnet werden:

$$P_{pq} = \frac{\pi}{n+1} \sum_x \sum_y f(x, y) W_{pq}^*(x, y)$$

Die Formel zum Bezug der Pseudo-Zernike-Momente zu den geometrischen Momenten ist bei Teh et al. [66] gegeben.

5.4 Orthogonale Fourier-Mellin-Momente

Orthogonale Fourier-Mellin-Momente (OFM) wurden ursprünglich von Sheng und Shen eingeführt [58] und im Artikel von Kan [32] im Vergleich mit den Zernike-Momenten untersucht. Die Basisfunktionen von OFMs sind wie folgt definiert:

$$U_{pq}(r, \theta) = Q_p(r) e^{iq\theta}$$

wobei $p \geq 0$, $q \in \mathbb{Z}$, $p \in \mathbb{N}$ und Q_p als radiales Polynom wie folgt definiert wird:

$$Q_p(r) = \sum_{s=0}^p \frac{(-1)^s (2p+1-s)!}{s!(p-s)!(p+1-s)!} r^{p-s}$$

Des weiteren gilt:

$$\langle U_{nm}, W_{pq} \rangle = \frac{\pi}{n+1} \delta_{np} \delta_{mq}$$

Die OFM Momente ergeben sich damit zu:

$$OFM_{pq} = \frac{p+1}{\pi} \langle f, U_{pq} \rangle = \frac{p+1}{\pi} \int \int_{\mathbb{R}^2} f(x, y) U_{pq}^*(x, y) dx dy$$

bzw. für den diskreten Fall vereinfacht zu:

$$OFM_{pq} = \frac{p+1}{\pi} \langle f, U_{pq} \rangle = \frac{p+1}{\pi} \sum_x \sum_y f(x, y) U_{pq}^*(x, y)$$

Die OFM-Momente sind im Vergleich mit den Zernike-Momenten, laut Shen besser geeignet, um sehr kleine Symbole adäquat zu beschreiben. Außerdem benötigen sie eine geringere Ordnung von Momenten zur Erfassung des Symbols. Im Artikel von Shen ist über die komplexen Momente auch ein Bezug zu den geometrischen Momenten gegeben.

5.5 Wavelet Momente

Wavelet Momente wurden 1999 durch Shen et al. [57] eingeführt. Sie bieten gegenüber den bisherigen Ansätzen, die meist nur globale Formeigenschaften des zu analysierenden Objektes berücksichtigen, auch eine gute örtliche Auflösung, welche durch die

Verwendung von Wavelets realisiert wird. Es ist somit möglich sowohl zeitliche, als auch frequenzbezogene Analysen über den Eingabedaten durchzuführen. Die Wavelet-Momente sind als komplexwertige Momente über dem Einheitskreis definiert und sind invariant gegenüber Rotation (Beweis äquivalent zu dem der Zernike-Momente). Rotationsinvariante Momente sind üblicherweise wie folgt definiert [57]:

$$F_{pq} = \int \int f(r, \theta) g_p(r) e^{jq\theta} r \, dr \, d\theta \quad p, q \in \mathbb{N}$$

Dabei ist $g_p(r)$ eine Funktion mit der radialen Variable r . Wird $g_p(r)$ über dem gesamten Bildbereich definiert, können globale andernfalls lokale Merkmale erfaßt werden. Dementsprechend werden mit Zernike-, Li- [39] oder Hu-Momenten eher globale Merkmale erfaßt. Diese sind damit auch anfälliger gegenüber Rauschen und können ähnliche Objekte mit nur minimalen Unterschieden nur schlecht diskriminieren, wie von Shen et al. in [57] gezeigt werden konnte.

Bei den Wavelet-Momenten wurde $g_p(r)$ als Wavelet-Basisfunktion betrachtet:

$$\psi^{a,b}(r) = \frac{1}{\sqrt{a}} \psi\left(\frac{r-b}{a}\right)$$

mit $a \in \mathbb{R}_+$ als Dehnungsparameter und $b \in \mathbb{R}$ als Verschiebungsparameter. Konkret wurden kubische B-spline Wavelets verwendet, da diese optimal im Frequenzraum lokalisiert sind und sich eng an die Form von Li- oder Zernikepolynomen anlehnen. Das Mutterwavelet $\psi(r)$ des kubischen B-Spline ist in Gausscher Annäherung gegeben durch:

$$\psi(r) = \frac{4a^{n+1}}{\sqrt{2\pi(n+1)}} \sigma_w \cos(2\pi f_0(2r-1)) \cdot e^{-\frac{(2r-1)^2}{2\sigma_w^2(n+1)}}$$

mit $n = 3$ $a = 0.697066$ $f_0 = 0.409177$ $\sigma_w^2 = 0.561145$ und für a, b gelte:

$$\begin{cases} a = 0.9^m, & m = 0, 1, 2, 3 \\ b = 0.4 \cdot n \cdot a & n = 0, 1, \dots, 2^{m+1} \end{cases}$$

Somit ergeben sich die Wavelets wie folgt:

$$\psi_{m,n}(r) = 2^{m/2} \psi\left(\frac{1}{0.9^m} r - 0.4n\right)$$

Damit ist es möglich sowohl globale, als auch lokale Informationen in Abhängigkeit von m und n zu ermitteln und die Wavelet-Momente wie folgt zu definieren:

$$||F_{m,n,q}|| = || \int S_q(r) \cdot \psi_{m,n}(r) r \, dr ||$$

mit $m = 0, 1, 2, 3$ $n = 0, 1, \dots, 2^{m+1}$ $q = 0, 1, 2, 3$ $0 \leq r \leq 1$ dabei ist $S_q(r)$ und $\psi_{m,n}(r)$ definiert durch:

$$S_q(r) = \int f(r, \theta) e^{jq\theta} d\theta \quad 0 \leq \theta \leq 2\pi$$

Für den diskreten Fall kann man die Wavelet-Momente wie folgt berechnen:

$$\begin{aligned} \|F_{m,n,q}\| &= \left\| \sum S_q(r) \cdot \psi_{m,n}(r)r \right\| \\ S_q(r) &= \sum f(r, \theta) e^{jq\theta} \quad 0 \leq \theta \leq 2\pi \end{aligned}$$

Wobei die Summe über dem relevanten Bildbereich ermittelt wird. Des weiteren gilt:

$$r = \sqrt{x^2 + y^2} \quad \text{und} \quad \theta = \text{atan2} \left(\frac{y}{x} \right)$$

$F_{m,n,q}$ kann dabei als das erste Moment von $S_q(r)$ in der m^{ten} Skalierungsebene mit Shiftindex n betrachtet werden. Für ein festes r stellt $S_q(r)$ das q^{te} Frequenzmerkmal des Bildobjektes $f(r, \theta)$ im Phasenraum $0 \leq \theta \leq 2\pi$ dar. Damit beschreibt $S_q(r)$ die Merkmalsverteilung des Bildobjektes im Radialraum $0 \leq r \leq 1$.

Für alle in diesem Abschnitt vorgestellten Momente kann die Translationsinvarianz durch Verwendung der *center of gravity* Normierung, wie bei den Legendre-Momenten, angegeben erreicht werden.

Kapitel 6

Schnelle und effiziente Berechnung von Momenten

Die Berechnung von geometrischen Momenten ist in der ursprünglichen Definition sehr aufwendig und hat die Komplexität von $O(pqN^2)$ [18]. In der Literatur wurden verschiedenste Methoden vorgeschlagen, die Berechnung der Momente zu beschleunigen. Es gibt dabei im wesentlichen die folgenden drei Ansätze:

1. schnelle Berechnung durch Approximation des Objektes mit Polygonzügen
2. Berechnung der Momente auf Basis von lookup-Tabellen, bzw. expliziter Formelauflösung
3. schnelle Momentberechnung für Binärbilder unter Verwendung des Theorems von Green

Der erste schnelle Algorithmus zur Berechnung geht auf Zakaria (Delta-Methode) [72] zurück. Die grundlegende Idee seiner „nabla“ Methode ist die Dekomposition des Objektes in individuelle Zeilen von Pixeln. Das Objektmoment ist dann durch Summation aller Zeilenmomente des Objektes gegeben, welches sehr einfach von den Koordinaten der ersten und letzten Pixel berechnet werden kann. Diese Methode funktioniert allerdings nur bei konvexen Formen und ist approximativ.

Bei Ansatz 1 werden die Momente über einem durch Polygonzüge approximierten Bild berechnet, vgl. z.B. [38, 61]. Diese Berechnung geschieht unter Verwendung der Eckpunkte der Polygone. Laut Flusser ist diese Methodik aber lediglich bei sehr einfachen Formen mit wenigen Knoten sinnvoll einsetzbar, da es andernfalls zu erheblichen Ungenauigkeiten in den Ergebnissen kommt [18].

Die Verfahren bei Punkt 3 basieren auf dem Theorem von Green, welches die Berechnung des Doppelintegrals über einem Objekt durch Einzelintegration entlang der Objektgrenzen realisiert [48, 70]. Ein entsprechender Ansatz mit Verwendung des Greenstheorems für den diskreten Fall wurde von Philips [48] angegeben.

In einem Artikel von Bunke und Jiang [31] wird das Objekt zunächst durch Polygonzüge approximiert und dann Greens-Theorem verwendet. Diese Methode ist zwar sehr effizient, jedoch durch die doppelte Approximation nicht sehr genau.

6.1 Greens-Theorem nach Philips

Greens-Theorem besagt, daß wenn eine Funktion $g(x, y)$ über einer einfach verbundenen Menge Ω integrierbar ist und sich als Summe der Ableitungen zweier Funktionen N und M notieren läßt, dann kann das Doppelintegral über Ω als Einfachintegration entlang der Grenze $C = \partial\Omega$ von Ω ermittelt werden. Dabei ist C eine geschlossene Kurve, bestehend aus unendlich vielen glatten Kurven, die die einfach verbundene Menge Ω begrenzt. Die Menge Ω ist die durch das Polygon C eingeschlossene Fläche.

Dabei ist Greens-Theorem für die Momentberechnung nützlich, da die Form eines binären Objektes vollständig durch seine Objektgrenze C bestimmt ist. Es gibt verschiedene Versionen des Theorems von Green. Die von Philips [48] verwendete ist im folgenden angegeben:

Theorem 4 (Greens-Theorem)

$$\int \int_{\Omega} \frac{\partial}{\partial x} g(x, y) dx dy = \oint_{\partial\Omega} g(x, y) dy \quad (6.1)$$

Dabei ist Ω ein zweidimensionaler Bereich und $\partial\Omega$ beschreibt die Kanten (traversiert im Uhrzeigersinn). Für den diskreten Fall wurde von Philips die folgende Variante des Theorems von Green verwendet:

$$\sum_{(x,y) \in \Omega} \nabla_x g(x, y) = \sum_{(x,y) \in \partial\Omega^+} g(x, y) - \sum_{(x,y) \in \partial\Omega^-} g(x, y) \quad (6.2)$$

wobei $\nabla_x g(x, y) = g(x, y) - g(x-1, y)$. In dieser Gleichung ist die Grenze Ω eindeutig bestimmt als $\partial\Omega = \partial\Omega^+ \cup \partial\Omega^-$, mit

$$\begin{aligned} \partial\Omega^+ &= \{(x, y) : (x, y) \in \Omega, (x+1, y) \notin \Omega\} \\ \partial\Omega^- &= \{(x, y) : (x, y) \notin \Omega, (x+1, y) \in \Omega\} \end{aligned}$$

Der Beweis für die Gültigkeit der Gleichung 6.2 kann wie folgt gegeben werden [48]:

Beweis 5

Sei $\Omega_i = (X_i + k, Y_i) : 0 \leq k \leq \hat{N}_i$ und $0 \leq i \leq \hat{M}$ seien die Zeilen des Bildes. Mit dieser Notation und der Definition von $\nabla_x g(x, y)$ ergibt sich für die linke Seite der Gleichung 6.2:

$$\sum_{i=0}^{\hat{M}} \sum_{k=0}^{\hat{N}} (g(X_i + k, Y_i) - g(X_i + k - 1, Y_i)) = \sum_{i=0}^{\hat{M}} g(X_i + \hat{N}_i, Y_i) - \sum_{i=0}^{\hat{M}} g(X_i - 1, Y_i)$$

Mit $\partial\Omega^+ = (X_i + N_i, Y_i) : 0 \leq i \leq \acute{M}$ $\partial\Omega^- = (X_i - 1, Y_i) : 0 \leq i \leq \acute{M}$ ist diese Gleichung äquivalent zur rechten Seite der Gleichung 6.2.

6.2 Momentberechnung nach Flusser

Die Arbeit von Philips wurde nun durch Flusser [18] wie folgt erweitert und verbessert. Die geometrischen Momente sind wie bereits angegeben, in ihrer diskreten Form gegeben durch:

$$m_{pq} = \sum_{i=1}^N \sum_{j=1}^N i^p j^q f(i, j) \quad (6.3)$$

Dies ist allerdings nur eine Approximation der ursprünglichen Berechnung mittels Doppelintegral. Durch Lin [41] wurde eine genauere Approximation durch exakte Integration der Monome $x^p y^q$ vorgeschlagen.

$$\hat{m}_{pq} = \sum_{i=1}^N \sum_{j=1}^N f(i, j) \int \int_{A_{ij}} x^p y^q dx dy \quad (6.4)$$

$$= \frac{1}{(p+1)(q+1)} \sum_{i=1}^N \sum_{j=1}^N f(i, j) \left(\left(i + \frac{1}{2} \right)^{p+1} \right. \quad (6.5)$$

$$\left. - \left(i - \frac{1}{2} \right)^{p+1} \right) \left(\left(j + \frac{1}{2} \right)^{q+1} - \left(j - \frac{1}{2} \right)^{q+1} \right) \quad (6.6)$$

Mit A_{ij} als Bereich der Pixel (i, j) . Ein weiterer, zu berücksichtigender Aspekt ist die Abfolge von Objekten. In der Mustererkennung wird meist nicht nur ein einzelnes Muster betrachtet, sondern eine Folge von Mustern. Diesem Punkt trägt der Ansatz von Flusser Rechnung, in dem bestimmte immer wiederkehrende Berechnungen, die nicht direkt vom Muster abhängig sind, im Vorfeld - einmalig - ausgeführt werden. Aufbauend auf dem Ansatz von Philips wird der Algorithmus zur Berechnung in zwei Teile aufgespalten. In einen gemeinsamen, welcher nur einmal zu Beginn ausgeführt wird und einen zweiten, der für jedes Muster individuell durchgeführt wird. Unter Verwendung der Ω -Definition von Philips können die Gleichungen 6.3 und 6.6 mit $\partial\Omega_-$ and $\partial\Omega_+$ ausdrücken:

$$m_{pq} = \sum_{\partial\Omega_+} y^q \sum_{i=1}^x i^p - \sum_{\partial\Omega_-} y^q \sum_{i=1}^x i^p \quad (6.7)$$

$$\hat{m}_{pq} = \frac{1}{(p+1)(q+1)} \left[\sum_{\partial\Omega_+} \left(x + \frac{1}{2} \right)^{p+1} \left(\left(y + \frac{1}{2} \right)^{q+1} - \left(y - \frac{1}{2} \right)^{q+1} \right) \right. \quad (6.8)$$

$$\left. - \sum_{\partial\Omega_-} \left(x + \frac{1}{2} \right)^{p+1} \left(\left(y + \frac{1}{2} \right)^{q+1} - \left(y - \frac{1}{2} \right)^{q+1} \right) \right] \quad (6.9)$$

Obige Formel lässt sich sehr einfach in Matrixschreibweise notieren. Seien R und S $p_m \times N$ Matrizen mit $p_m \geq p + 1, p_m \geq q + 1$ und wie folgt definiert:

$$R_{ij} = j^{i-1} \quad S_{ij} = \sum_{n=1}^j n^{i-1}$$

Dann ergibt sich 6.7 zu folgender Form:

$$m_{pq} = \sum_{\partial\Omega_+} S_{p+1,x} \cdot R_{q+1,y} - \sum_{\partial\Omega_-} S_{p+1,x} \cdot R_{q+1,y}$$

Sei desweiteren eine P als $p_m \times (N + 1)$ Matrix in der folgenden Form gegeben:

$$P_{ij} = \frac{1}{i} \left(j - \frac{1}{2} \right)^i$$

Dann erhält man für Gleichung 6.8:

$$\hat{m}_{pq} = \sum_{\partial\Omega_+} P_{p+1,x+1} \cdot (P_{q+1,y+1} - P_{q+1,y}) - \sum_{\partial\Omega_-} P_{p+1,x+1} \cdot (P_{q+1,y+1} - P_{q+1,y})$$

Die Matrizen R , S und P sind dabei unabhängig vom Objekt Ω und können einmalig im Vorfeld für eine spezifische Bilddimension N und eine spezifische maximale Ordnung p berechnet werden. Die einzigen, für jedes Bild zu berechnenden Werte sind die Grenzen Ω_- und Ω_+ (vgl. auch Abb. 6.1). Das Verfahren ist sehr schnell, da die

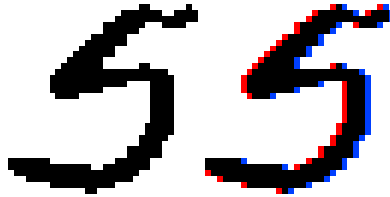


Abbildung 6.1: Symbol - links original und rechts mit Ω_- (rot) und Ω_+ (blau)

Matrizen R , P , S lediglich $O(N \cdot p_m)$ Operationen benötigen und die eigentliche Momentberechnung lediglich K Multiplikationen und $(K - 1)$ Additionen benötigt. Dabei ist K die Anzahl der Pixel in $\partial\Omega_{\pm}$. Dabei wird jedes Moment (mit Ordnung kleiner p_m) gleich schnell berechnet.

6.3 Momentberechnung nach Mertzios

Die von Mertzios in [43] vorgeschlagene Methode zur Beschleunigung der Momentberechnung basiert auf der Repräsentation des Bildes in Form von Bildblöcken (IBR)¹. Damit kann die Berechnungskomplexität von etwa $O(N^2)$ auf $O(L)$ reduziert werden, wobei $(L - 1, L - 1)$ die Ordnung der 2D Momente ist. Die IBR kann dabei wie folgt beschrieben werden:

¹Ein weitere Verbesserung dieses Ansatzes wurde von Flusser in [20] angegeben

- *Definition 1:* Ein Block ist ein rechteckiger Bereich aus gleichfarbigen Pixeln des Bildes, mit Eckpunkten, die parallel zur Bildachse verlaufen.
- *Definition 2:* Ein Binärbild wird als blockrepräsentiert bezeichnet, wenn es als eine Menge von Blöcken auf Objektebene gegeben ist und wenn jeder Pixel des Bildes mit einem Objektwert nur zu einem einzigen Block gehört.

Die IBR ist dabei eine verlustfreie Darstellung. Im folgenden nun ein kurzer Algorithmus zur Erzeugung einer IBR:

1. Man betrachtet jede Zeile y des Bildes f und sucht die zusammenhängenden Objektbereiche innerhalb dieser Zeile y .
2. Es erfolgt dann ein Vergleich von Blöcken und Intervallen, die Pixel in der darüberliegenden Zeile besitzen.
3. Wenn kein Intervall zu einem Block paßt, dann beginnt hier ein neuer Block.
4. Wenn ein Block mit einem Intervall zusammenpasst, ist das Ende dieses Blockes in der Zeile y .

Man erhält damit eine Menge von rechteckigen Bereichen im Bild, die ein Objekt repräsentieren. Das blockrepräsentierte Bild ist dann gegeben durch:

$$f(x, y) = (b_i : i = 0, 1, \dots, k - 1) \quad k - \text{Anzahl der Blöcke}$$

Jeder Block wird durch 4 Koordinaten beschrieben (oben-links (x_1, y_1) , rechts-unten (x_2, y_2))

Berechnung von Momenten basierend auf IBR (binär) Betrachtet man die allgemeine Definition zur Berechnung von geometrischen Momenten (siehe Abschnitt 4) über Binärbildern, dann läßt sich die Berechnung mit IBRs wie folgt notieren:

$$m_{pq} = \sum_{i=0}^{k-1} m_{pq}^{b_i} = \sum_{i=0}^{k-1} \sum_{x=x_{1,b_i}}^{x_{2,b_i}} \sum_{y=y_{1,b_i}}^{y_{2,b_i}} x^p y^q$$

Wobei $x_{1,b_i}, x_{2,b_i}, y_{1,b_i}, y_{2,b_i}$ die Blockkoordinaten des i^{ten} Blocks sind. Unter Berücksichtigung der Rechteckform kann man weiter folgern:

$$\begin{aligned}
 m_{pq}^b &= \sum_{x=x_{1,b_i}}^{x_{2,b_i}} \sum_{y=y_{1,b_i}}^{y_{2,b_i}} x^p y^q = x_{1,b}^p \\
 &\quad \sum_{y=y_{1,b_i}}^{y_{2,b_i}} y^q + (x_{1,b} + 1)^p \sum_{y=y_{1,b_i}}^{y_{2,b_i}} y^q + \dots + x_{2,b}^p \\
 &\quad \sum_{y=y_{1,b_i}}^{y_{2,b_i}} y^q \\
 &= \left[\sum_{x=x_{1,b_i}}^{x_{2,b_i}} x^p \right] \left[\sum_{y=y_{1,b_i}}^{y_{2,b_i}} y^q \right]
 \end{aligned}$$

Damit kann die Komplexität zu $O(N)$ reduziert werden. Eine weitere Verbesserung ergibt sich durch Berechnung der folgenden Summen für Potenzen von x und y .

$$S_{x_{1,b}, x_{2,b}}^p = \sum_{x=x_{1,b}}^{x_{2,b}} x^p, S_{y_{1,b}, y_{2,b}}^q = \sum_{y=y_{1,b}}^{y_{2,b}} y^q \quad x, y, p, q \in \mathbb{Z}$$

Diese Summen lassen sich dann durch explizite Auflösung nach folgender Formel leicht berechnen:

$$\begin{aligned}
 S_{x_{1,b}, x_{2,b}}^p &= \frac{(x_{2,b} + 1)^{p+1} - x_{1,b}^{p+1} - (x_{2,b} - x_{1,b} + 1) - \binom{p+1}{1} S_{x_{1,b}, x_{2,b}}^1}{p+1} \\
 &\quad - \frac{\binom{p+1}{2} S_{x_{1,b}, x_{2,b}}^2 - \dots - \binom{p+1}{p-1} S_{x_{1,b}, x_{2,b}}^{p-1}}{p+1}
 \end{aligned}$$

Eine entsprechende Erweiterung kann für zentrale Momente und dann damit auch für normalisierte Momente erfolgen. Die entsprechenden Formeln sind in [43] zu finden. Entsprechend der Hinweise in [43] wurde die Summationsberechnung wie oben nur für $p, q \leq 4$ durchgeführt.

6.4 Einschränkungen

Bei der Merkmalsgenerierung gibt es eine Menge von Einschränkungen, sowohl mit Blick auf die Genauigkeit der Ergebnisse, als auch bezüglich des informationstheoretischen Nutzens.

6.4.1 Numerische- und Berechnungsprobleme

Bei der Berechnung der Momente weisen die einzelnen Verfahren unterschiedliche Güten bezüglich der numerischen Stabilität des Verfahrens auf. Einige der Probleme sind implementierungsspezifisch und einige sind verfahrensbedingt.

Zunächst ergeben sich bereits Einschränkungen infolge der Diskretisierung der Daten. Da die Momente ursprünglich über kontinuierlichen Daten definiert sind, ergeben sich bei Verwendung diskreter Bilddaten vor allem Limitierungen bezüglich der Genauigkeit (vgl. dazu [66, 40]). So können die Daten bei der Berechnung stets nur in diskreten Schritten durchlaufen werden und auch die Skalierungen bezüglich des Einheitskreises führen zu Problemen. Neben den, nicht vermeidbaren Problemen der Diskretisierungen gibt es auch Probleme in den Berechnungsroutinen selbst. So werden z.B. die Polynome aus Abschnitt 5 mit Fakultäten berechnet, die ab $12!$ nur noch durch Verwendung größerer Datentypen (double) ermittelt werden können und danach relativ schnell zu Ungenauigkeiten führen.

Ein weiteres Problem ist die Einschränkung auf den Einheitskreis, wie sie bei den Fourier-Mellin basierten Momenten auftritt. Dabei zeigt sich, daß z.B. Zernike-Momente bei sehr kleinen Bildgrößen nur noch eine geringe Auflösung über dem Datensatz realisieren können und damit nahezu nutzlos werden. Eine Lösung für dieses Problem wird mit den Wavelet-Momenten von Shen und Horace [57] gegeben, die durch die Definition über Wavelets eine höhere Ortsauflösung selbst bei stark verkleinerten Bilddaten realisieren. Ein weiteres erhebliches Problem liegt in der Anfälligkeit der Verfahren gegenüber Rauschen, so sind die Merkmale von stark verrauschten Bildern meist sehr ungenau, wie es z.B. bei Teh [66] beschrieben wird und können nur eingeschränkt genutzt werden. Dies wird besonders bei der Verwendung von Momenten höherer Ordnung deutlich, die nahezu nur noch das Rauschen erfassen.

6.4.2 Informationstheoretische Einschränkungen

Vom praktischen Standpunkt aus betrachtet, sollten die berechneten Merkmale soweit wie möglich minimal und von einander unabhängig sein, so daß weder unnötige Informationen (z.B. Rauschen) noch redundante Informationen erfaßt werden.

Dieses Ziel kann leider nicht immer erreicht werden. So weisen die Momente von Hu, wie im Artikel von Flusser [19] gezeigt, ein hohes Maß an Redundanz auf. Auch andere Momentdefinitionen sind nicht über orthogonalen Polynomen definiert und führen damit zu Redundanzen. Lösungen für diese Probleme wurden z.B. in den Artikeln [65] und [21] angegeben.

Die ungewünschte Erfassung von Rauschen konnte bis jetzt von keinem hier betrachteten Verfahren vollständig gelöst werden, allerdings weisen die neueren Momentdefinitionen [32] generell bessere Ergebnisse bezüglich ihrer Rauschanfälligkeit auf. Generell bleibt festzuhalten, daß die Generierung von Momenten höherer Ordnung stets auch einen höheren Anteil an Rauschen mit erfaßt, da Momente höherer Ordnung in einem stärkeren Maße die Feinstruktur des Bildes erfassen und somit, wenn möglich, nicht verwendet werden sollten.

Kapitel 7

Klassifikation

Im folgenden Kapitel wird ein weiterer wichtiger Teil eines OCR-Systems behandelt - die Klassifikation. Wie die Wahl der Merkmale, so entscheidet auch die Klassifikation in einem hohen Maße über die Erkennungsrate. Im folgenden werden nicht überwachte und überwachte Klassifikationsmethoden vorgestellt und auf ihre Verwendbarkeit in dieser Arbeit geprüft. Es muß, wie in [11] bereits angegeben wurde, stets berücksichtigt werden, daß die Klassifikation von Daten im allgemeinen ein sehr breiter und noch nicht vollständig erforschter Bereich ist, in dem die Klassifikation von Formen/Symbolen nur ein spezieller Teil ist.

Im folgenden zunächst eine kurze Definition zu Klassifikation, wie sie von Costa gegeben wurde [11]:

Definition 2

Klassifikation ist die Zuweisung von Klassenbezeichnern oder Kategorien zu Datensätzen unter Berücksichtigung der Eigenschaften der Datensätze.

Klassifikation ist ein sehr problematischer Arbeitsschritt, da es im Regelfall keine eindeutige Klassifikation über einer Datenmenge gibt. So kann man z.B. „Gurken und Birnen“ bezüglich Ihrer Farbe oder Größe oder diverser anderer Merkmale kategorisieren. Die „korrekte“ Klassifikation hängt dabei vom spezifischen Kontext ab. Im Bereich der OCR gibt es ebenfalls diverse Merkmale, die über einem Datensatz, hier einem Symbol, erfaßt werden können. Dennoch lassen sich einige allgemein gültige Prinzipien für die Klassifikation formulieren (vgl. dazu auch [11]).

- Klassifikation weist Objekten Klassenbezeichner zu.
- Objekte ein und derselben Klasse besitzen ähnliche Merkmale (Intraklassenähnlichkeit).
- Objekte verschiedener Klassen besitzen unähnliche Merkmale (Interklassenähnlichkeit).
- Klassifikation reduziert Redundanz.

- Klassifikation verlangt die Ordnung des Merkmalsraums in Regionen, die zu den Klassen korrespondieren.
- Klassifikation zwingt zur Auswahl bestimmter Merkmale, Distanzmessungen, Klassifikationskriterien und Parametern - all dies kann zu unterschiedlichen Resultaten führen.
- Es gibt keine exakten Regeln, wie man am Besten klassifiziert.
- Klassifikation ist nicht einfach.

Wie im Abschnitt Merkmalsgewinnung bereits besprochen, sind die, für die hier durchgeführte Symbolerkennung verwendeten Merkmale ausschließlich statistischer Natur, also Meß- bzw. Schätzgrößen, die über den betrachteten Objekten ermittelt wurden. Somit werden im folgenden auch lediglich statistische Klassifikatoren berücksichtigt. Dabei wird das Muster durch d Merkmale in Form eines d -dimensionalen Merkmalsvektors verstanden. Das Ziel bei der Klassifikation ist damit, in dem durch die d -dimensionalen Merkmalsvektoren aufgespannten Merkmalsraum, die entsprechenden Klassengrenzen zu extrahieren. Es existieren dabei - grob - zwei Möglichkeiten der Klassifikation.

7.1 Überwachte Klassifikation

Bei der überwachten Klassifikation existieren ein oder mehrere Trainingsdatensets. Die Prototypen sind in diesem Fall bestimmten bekannten Klassen zugewiesen, um neue Objekte zu klassifizieren. Bei dieser Art der Klassifikation werden normalerweise zwei Schritte durchgeführt. Zum einen die *Lernphase*, wobei die Klassifikationsmethode auf die Prototypen angewandt wird (deren Klassenzugehörigkeit ja bereits bekannt ist) und zum anderen der *Erkennung*, bei der das nun trainierte System zur Klassifikation neuer Objekte eingesetzt wird. Das klassische Beispiel einer solchen Klassifikation sind z.B. neuronale Netze, die ebenso diese zwei Phasen durchlaufen. Der Entscheidungsprozeß in der statistischen Mustererkennung kann dabei wie folgt zusammengefaßt werden [30]: Ein gegebenes Muster, welches durch einen d -dimensionalen Merkmalsvektor repräsentiert ist, soll zu einer der Klassen $\omega_1, \omega_2, \dots, \omega_c$ zugewiesen werden. Dabei wird angenommen, daß die betrachteten Merkmale eine Wahrscheinlichkeitsdichtefunktion (PDF) bzgl. der Musterklassen besitzen.

Ein Vektor x , der zur Klasse ω_i gehört, wird als eine zufällige Stichprobe der entsprechenden PDF $p(x|\omega_i)$ aufgefaßt. Um die dafür nötige Entscheidungsfläche bzw. Grenze zu ermitteln, gibt es eine Vielzahl von Entscheidungsregeln, unter anderem die *Bayessche-Regel*, die *Maximum-Likelihood Regel* und die *Neyman-Pearson Regel*. Die optimale Bayessche Entscheidungsregel zur Minimierung des Risikos (erwarteter Wert der Fehlerfunktion) kann dabei wie folgt definiert werden (vgl. auch Duda & Hart [14]):

Die Zuweisung eines Musters x zur Klasse ω_i mit dem bedingten Risiko

$$R(\omega_i|x) = \sum_{j=1}^c L(\omega_i, \omega_j) \cdot P(\omega_j|x)$$

ist minimal, wenn $L(\omega_i, \omega_j)$ die Kostenfunktion für eine Fehlentscheidung zur Klasse ω_i ist, wobei die korrekte Klasse ω_j war (mit $P(\omega_j|x)$ als posteriori Wahrscheinlichkeit). Im Falle einer 0/1 Kostenfunktion wie in Gleichung 7.1 definiert, wird das bedingte Risiko die bedingte Wahrscheinlichkeit einer Fehlklassifikation:

$$L(\omega_i, \omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \quad (7.1)$$

Für diese Wahl der Kostenfunktion kann die Bayessche Entscheidungsregel wie folgt vereinfacht werden (auch als maximale A posteriori Regel bekannt): Das Eingabemuster x wird der Klasse ω_i zugewiesen, wenn

$$P(\omega_i|x) > P(\omega_j|x) \forall j \neq i$$

Verschiedene Strategien können zur Konstruktion eines Klassifikators verwendet werden, abhängig von der Art der verfügbaren Informationen bzgl. der Klassen-PDFs.

Ideal wäre somit der Bayessche Klassifikator, dieser kann allerdings nur genutzt werden, wenn genügend Informationen über die PDFs der Klassen verfügbar sind. Dies ist in unserem Fall allerdings nicht der Fall, so daß die entsprechenden Parameter entweder aus einer ausreichend großen Stichprobe geschätzt werden (*lernen*) oder parameterfreie Verfahren verwendet werden müssen. Tatsächlich kann das, hier ebenfalls verwendete MLP (siehe Abschnitt 7.4.4) als eine solche, nichtparametrische Methode verstanden werden, welche die Entscheidungsfläche ermittelt.

7.2 Unüberwachte Klassifikation

Bei der offenen oder nicht überwachten Klassifikation übergibt man dem Klassifikator eine Menge von Daten, ohne weitere Vorinformation und läßt die Klassifikation ablaufen, um eine adäquate Klassenzuordnung zu treffen. Die dabei verwendeten Verfahren sind Clusteringverfahren wie z.B. k-means. Die unüberwachte Klassifikation ist meist schwieriger und berechnungsaufwendiger (da es keine nutzbaren Vorinformationen gibt). Sie wird meist explorativ eingesetzt.

Da in dieser Arbeit die Klassifikation von Objekten mit bekannten Klassen angestrebt wird, ist die Anwendung von unüberwachten Verfahren nicht nötig, so daß im folgenden auch nicht weiter darauf eingegangen wird. Eine sehr ausführliche Analyse zu unüberwachten Klassifikationsverfahren (Clustering - Techniken) kann in [29] nachgelesen werden.

Unerheblich, welche überwachte Klassifikationsmethode oder Entscheidungsregel verwendet wird, muß eine Trainingsphase über den vorhandenen Trainingsdaten durchgeführt werden¹. Ein Klassifikator sollte eine gewisse Fähigkeit zur Generalisierung besitzen, so daß auch neu auftretende Muster adäquat erkannt werden können. Somit ist eine Optimierung des Klassifikators zur Maximierung der Leistung über der Trainingsmenge eher ungünstig. Für eine schlechte Generalisierungsfähigkeit des Klassifikators lassen sich im Allgemeinen drei Faktoren angeben [30] :

1. Die Anzahl der Merkmale ist relativ hoch im Vergleich zur Anzahl an Trainingsdaten (*curse of dimensionality (cod)*)
2. Die Anzahl der unbekannten Parameter des Klassifikators ist zu hoch (polynomiale Klassifikatoren, große NN)
3. Ein Klassifikator ist zu stark auf eine bestimmte Trainingsmenge optimiert (*overfitting*)

Die Lösungen für diese Probleme sind recht vielfältig und unter anderem in [30] zitiert, dort ist auch ein sehr schönes Beispiel zu *cod* angegeben. Allgemein empfehlen sich zum Test sehr große Datenmengen, mit Merkmalen hoher Zwischenklassenvarianz. Desweiteren sollte, um dem Problem des *curse of dimensionality* zu begegnen, eine Selektion von k diskriminanten Merkmalen über dem betrachteten Merkmalsset erfolgen (näheres dazu unter 7.3) und soweit wie möglich das Prinzip der Minimalität angestrebt werden. So sollte z.B. die Komplexität (Anzahl von Parametern) des Klassifikators möglichst gering gehalten werden.

7.3 Merkmalsselektion

Wie bereits mehrfach erwähnt, ist die Berechnung von Merkmalen bis zu einer hohen Ordnung zwar prinzipiell leicht möglich, aber für die Erkennung des Objektes meist nicht nötig. Zum einen wird durch die Verwendung einer zu hohen Anzahl von Merkmalen die Zeit für eine Klassifikation unnötig erhöht, zum anderen kann es bei zu vielen Merkmalen auch zu verstärkten Fehlklassifikationen kommen, da die Merkmalsvektoren die Klassen dann nicht mehr ausreichend diskriminieren, sondern generalisierende Effekte zeigen. Aus diesen Gründen wird üblicherweise vor der Klassifikation zunächst eine Merkmalsselektion durchgeführt.

In der Literatur wurden verschiedene Verfahren vorgeschlagen, um eine, für eine optimale Klassifikation geeignete Menge von Merkmalen zu extrahieren. Komplexere Verfahren wie z.B. die Hauptkomponentenanalyse (PCA) können über den ermittelten Vektoren angewandt werden, um die Vektoren entsprechend ihrer Hauptachsen zu transformieren, was zu einer Datenreduktion führt.

¹Um zumindest die Parameter zu schätzen.

Andere Ansätze sind etwas einfacher Natur. Sie versuchen aus dem vorhandenen Merkmalsatz eine, in einem gewissen Maße optimale Menge von Merkmalen zu extrahieren, ohne dabei die Merkmale zu transformieren. Diese Verfahren haben zudem den Vorteil, daß sie einmalig angewandt werden können, so daß dann stets eine fest selektierte Merkmalsmenge Verwendung findet. In Tabelle 7.1 zunächst eine Kurzübersicht zu den Merkmalsselektionsverfahren [30].

In dieser Arbeit wurden verschiedene Merkmalsselektionsmethoden berücksichtigt, darunter SBFS und SFS. Dabei wurde vor allem SFS genutzt, da dieser Algorithmus relativ schnell arbeitet und geeignete Merkmalssets selektiert wurden. Der Algorithmus zu SBFS von Pudil [51] (mit euklidischer Distanz) wurde nur in einigen Vorversuchen genutzt, da die Berechnungen sehr aufwendig sind, obwohl damit Merkmalssets ermittelt werden können, die näher am optimalen Merkmalsset (über einer gegebenen Menge) liegen. Für eine konkrete Beschreibung der Algorithmen sei auf den entsprechenden Artikel [51] verwiesen². Für eine Einführung zum Thema Merkmalsselektion sei auf [28, 35] verwiesen. Es wurden aus den erzeugten Merkmalen pro Merkmalsgenerator jeweils die 30 diskriminantesten Merkmale (Wilks- λ Kriterium) ermittelt und weiter untersucht.

7.4 Verwendete Klassifikatoren

Alle, in dieser Arbeit verwendeten Klassifikatoren sind statistische überwachte Klassifikatoren. Im folgenden werden zunächst die Klassifikatoren beschrieben, die bei Chim et al. [71] verwendet werden. Diese sind zunächst noch relativ schlicht, zeigen aber bei geeignet vorverarbeiteten Daten bereits akzeptable Ergebnisse.

7.4.1 Linearer Klassifikator mit euklidischer Distanz

Die ist der, wohl einfachste mögliche Klassifikator. Es wird mittels euklidischer Distanz der Abstand des Merkmalsvektors (Input) zu den Prototypvektoren bestimmt und der Merkmalsvektor zu der Prototypklasse zugewiesen, die die kleinste Distanz zum Merkmalsvektor aufweist.

$$D_E = \sqrt{\sum_{k=1}^N (F_{P_k} - F_{L_k})^2}$$

Hier ist F_{P_k} das k^{te} Merkmal des Prototypen und F_{L_k} ist das k^{te} Merkmal des Eingabesymbols, welches gerade betrachtet wird. Das Abstandsmaß kann dabei über allen Prototypvektoren oder auch über z.B. Mittelwertvektoren berechnet werden. Eine

²Eine verbesserte Erweiterung zu SBFS wurde in [62] vorgeschlagen, allerdings hier wegen der dabei stark erhöhten Laufzeit nicht weiter berücksichtigt

Methode	Eigenschaft	Kommentar
Vollständige Suche	Auswertung von $\binom{d}{m}$ möglichen Teilmengen	Garantierte Findung des Optimums, zu rechenaufwendig
Branch & Bound Suche	Nur eine Teilmenge aller Möglichkeiten muß durchsucht werden, um das Optimum zu finden	Garantierte Optimumfindung; monoton; worst case - exponentielle Laufzeit
Beste individuelle Merkmale	Einzelauswertung aller m Merkmale; Selektion der n besten Merkmale	Einfach, führt meist nicht zum Optimum
Seq. forward selection (SFS)	selektiere das beste einzelne Merkmal und füge pro Schritt ein weiteres Merkmal hinzu, so daß das Kriterium maximiert wird	Wenn ein Merkmal einmal gelöscht wurde, kann es nicht erneut in der optimalen Teilmenge betrachtet werden; relativ schnell
Seq. backward selection (SBS)	Beginne mit allen d Merkmalen und entferne sukzessive ein Merkmal pro Zeitschritt	wie bei SFS, etwas aufwendiger
„Plus l -take away r “ (PLTR) Auswahl	Zunächst erweitere das Merkmalsset um l Merkmale mit SFS und dann entferne r Merkmale mit SBS	Vermeidet das Problem des „Einbrennens“ in SFS bzw. SBS
Seq. forward/backward floating selection (SF(B)FS)	Verallgemeinerung der PLTR-Methode; l und r werden automatisch bestimmt und angepaßt	Nahe Optimum

Tabelle 7.1: Merkmalsselektionsmethoden

Verbesserung des Verfahrens kann z.B. durch Einführung einer Gewichtung erreicht werden, wie sie in [71] vorgeschlagen wird:

$$D_E = \sqrt{\sum_{k=1}^N \omega_k (F_{P_k} - F_{L_k})^2}$$

An dieser Stelle kann, wie auch in den folgenden aus [71] entnommenen Klassifikatoren für die Gewichtung die Innerklassenvarianz verwendet werden.

$$\omega_k = \frac{1}{\bar{\sigma}_k}$$

$\bar{\sigma}_k$ ermittelt man, indem man den Mittelwert der Standardabweichungen des k^{ten} Merkmals über allen Prototypen jeder Klasse berechnet. Dies geschieht nach folgender Formel:

$$\bar{\sigma} = \frac{1}{n} \sum_{j=1}^n \left(\frac{m \sum_{i=1}^m F_{ji}^2 - (\sum_{i=1}^m F_{ji})^2}{m(m-1)} \right)^{\frac{1}{2}}$$

Die Anzahl der Klassen ist dabei durch n und die Anzahl der Symbole/Klassen durch m gegeben. F_{ji} bezeichnet das Merkmal des i^{ten} Musters aus der j^{ten} Klasse. Es ergibt sich damit, daß Merkmale mit hoher mittlerer Varianz ein entsprechend kleines Klassifikationsgewicht besitzen.

Approximierte nächste Nachbarsuche

Als Erweiterung zur von Chim et al. [71] vorgeschlagenen euklidischen Klassifikation mit Innerklassenvarianz, wurde die Verwendung eines approximierten nächsten Nachbarschemas mit kd-trees untersucht.

Beim nächsten Nachbarproblem betrachtet man eine Menge P von Datenpunkten im d -dimensionalen Raum. Die Daten wurden so vorverarbeitet, daß zu einem Anfragepunkt q der nächste, oder verallgemeinert die k -nächsten Punkte aus P zu q effizient ermittelt werden können. Durch die Verwendung einer Approximation für die ermittelten k nächsten Nachbarn, kann die Suche sehr effizient realisiert werden. Zusätzlich kann diese Klassifikation konkreter analysiert werden, indem die k -nächsten Nachbarn z.B. bezüglich ihrer Beteiligung an der Klassifikation untersucht werden. In Anlehnung an das Prinzip des „relevance feedbacks“ im Bereich des information retrievals wurde mit den k nächsten Nachbarn die Validität des Klassifikationsergebnisses untersucht. Dabei wurde festgehalten, wie viele nächste Nachbarn bei einer Klassifikation die gleiche Klassenzuweisung gewählt haben. Damit kann dann ermittelt werden, ob eine Entscheidung (getroffen durch Mehrheitsentscheid der k nächsten Nachbarn) z.B. sehr sicher erfolgte. Konkret wurden dabei folgende Prüfgrößen berechnet:

- Minimale Validität - der minimale Wert k von gleichentscheidenden, nächsten Nachbarn bei der Entscheidung

- Maximale Validität - der maximale Wert k von gleichentscheidenden, nächsten Nachbarn bei der Entscheidung
- Mittlere Validität - der mittlere Wert k von gleichentscheidenden, nächsten Nachbarn bei der Entscheidung

7.4.2 Kreuzkorrelation

Die normalisierte Kreuzkorrelation R zwischen zwei Merkmalsvektoren ist definiert als [71]:

$$R = \frac{\sum_{i=1}^N F_{P_i} F_{L_i}}{\sqrt{\sum_{j=1}^N F_{P_j}^2 \cdot \sum_{k=1}^N F_{L_k}^2}}$$

F_{P_i} und F_{L_i} sind definiert wie bei 7.4.1. Dem Eingabevektor wird die Klasse zugewiesen, zu der er am höchsten korreliert. Die entsprechend gewichtete normalisierte Kreuzkorrelation kann wie folgt definiert werden³:

$$R = \frac{\sum_{i=1}^N \omega_i F_{P_i} F_{L_i}}{\sqrt{\sum_{j=1}^N \omega_j F_{P_j}^2 \cdot \sum_{k=1}^N \omega_k F_{L_k}^2}}$$

Auch hier führt die Gewichtung zu einer Performanzverbesserung des Klassifikators, in welchem relevantere Merkmale einen höheren Beitrag zur Klassifikation leisten.

7.4.3 Diskriminanzkostenfunktion (DKF)

Der Artikel von Chim et al. [71] baut auf einer Arbeit von Tsirikolias und Mertzios auf, worin der folgende *discrimination cost* Klassifikator vorgeschlagen wird.

$$D(i, v) = \sum_{j=1}^N \left[\frac{\max(p_{vj}, x_{ij})}{\min(p_{vj}, x_{ij}) - 1} \right]^2$$

Das j^{te} Merkmal des v^{ten} Prototypen ist dabei durch p_{vj} gegeben und x_{ij} bezeichnet entsprechend das j^{te} Merkmal des i^{ten} Eingabevektors mit einer Vektorgröße von N Merkmalen. Bei dieser Technik wird die Klasse des Prototypvektors dem Eingabemuster zugewiesen, welcher mit dem Eingabevektor die geringsten Kosten verursacht.

Allgemein empfiehlt es sich, bereits bei den Eingabedaten für den Klassifikator vorsichtig zu sein. So sollten die Daten eine möglichst hohe Innerklassenvarianz und eine geringe Zwischenklassenvarianz aufweisen, um die Überlappung der Merkmalsräume zu vermeiden (vgl. [11], Seite 543). Außerdem müssen die *diskriminierenden* Merkmale so gewählt werden, daß eine adäquate Differenzierung der einzelnen Klassen möglich wird. Um die Berechnungskomplexität akzeptabel zu gestalten, empfiehlt es

³der Gewichtungsfaktor ist in 7.4.1 gegeben

sich die Merkmale so zu wählen, daß sie nicht redundant, nicht überlappend und minimal gewählt sind. Im wesentlichen erfüllen die, durch Momente ermittelbaren Merkmale diese Erwartungen, auch wenn z.B. die Momente von Hu nicht vollständig minimal sind (vgl. Abschnitt 4.4.1)

7.4.4 Neuronale Netze als Klassifikator

Neuronale Netze (NN) können als universelle Funktionsapproximatoren in sehr vielen Bereichen eingesetzt werden. Wie bereits im Abschnitt 3 erwähnt wurde, sind sie auch zur Gewinnung von Merkmalen zur Texterkennung einsetzbar. In dieser Arbeit wurden sie allerdings zur Klassifikation der Merkmalsvektoren eingesetzt.

Es gibt mittlerweile sehr viele unterschiedliche Arten von neuronalen Netzen, so daß auch in anderen Arbeiten zur Mustererkennung verschiedene Konfigurationen auftreten. Im allgemeinen werden aber feedforward Netze - bzw. multilayer perceptrons (MLPs) eingesetzt. Sehr gute Einführungen zu neuronalen Netzen können in [8, 54] nachgelesen werden. Ein sehr guter Übersichtsartikel zur Bildverarbeitung mit NNs wurde von Petersen et al. [16] erstellt. Die Parameterwahl für das MLP erfolgte in Anlehnung an Hinweise aus dem Buch von Haykin [24].

In den meisten neueren Arbeiten zur Mustererkennung werden fast ausschließlich MLPs eingesetzt, die mit dem Backpropagation-Algorithmus nach Rumelhart [56] angelernet wurden. Die Vorteile dieser Netze liegen zum einen in der sehr hohen Erkennungsrate⁴ und zum anderen in der relativ einfachen Parametrisierbarkeit (vgl. z.B. [32]). Die schematische Darstellung eines MLPs ist in Abbildung 7.1 zu sehen. In der

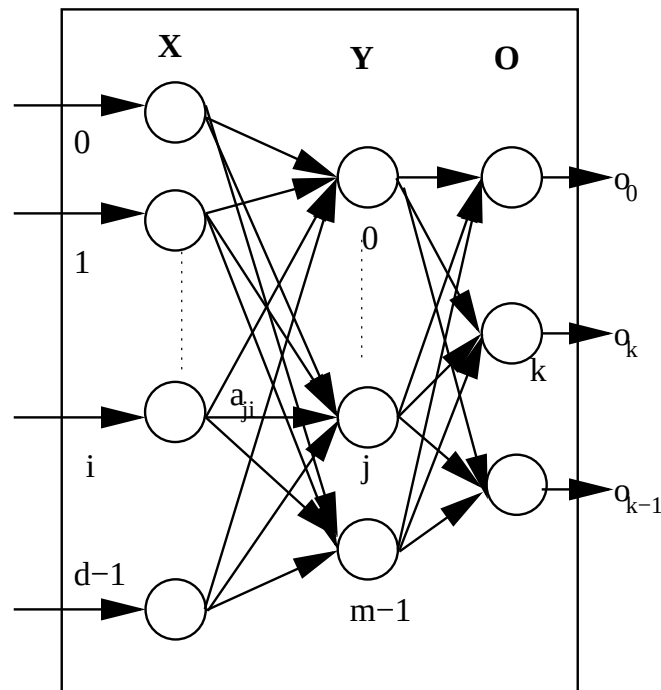


Abbildung 7.1: Schematische Darstellung eines MLP

Grafik 7.1 ist ein vollständig verdrahtetes MLP mit einer Eingangsschicht X , einer

⁴bei geeigneten Daten, insbesondere wurden die Werte der Vektoren auf $[-.025, .025]$ normiert

versteckten Schicht Y und der Ausgabeschicht O dargestellt. Die Anzahl der Neuronen der inneren Schicht ist beliebig, die der Schichten X und O durch die Dimension des Eingangsvektors x bzw. o vorgegeben⁵.

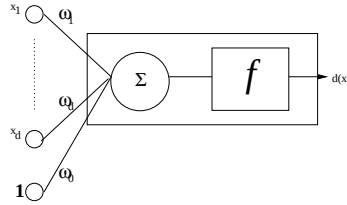


Abbildung 7.2: Schema eines Neurons

Die Aktivierungsfunktion f (siehe Abb. 7.2) der Neuronen wurde dabei mittels der logistischen sigmoiden Funktion definiert⁶.

$$f = \text{sig}(x) = \frac{1}{1 + e^{-ax}} \quad a = 1$$

Bei der logistischen sigmoiden Funktion können zudem die Ausgaben der Neurone als posteriori Wahrscheinlichkeiten interpretiert werden [12]. Im folgenden ein Plot der log. sigmoiden Funktion mit $a = 1$:

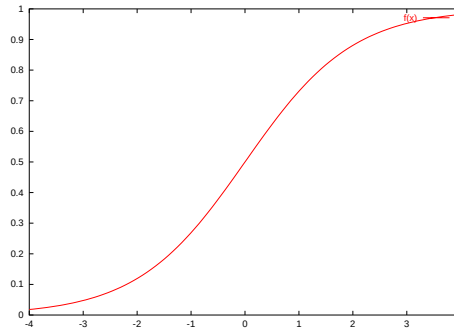


Abbildung 7.3: Logistische Sigmoide mit $a = 1$

In dieser Arbeit wurde ein MLP mit zumindest einer, versteckten vollständig vernetzten Schicht betrachtet. Die Eingangsschicht besitzt dabei so viele Neuronen, wie Merkmale verwendet wurden und die Ausgabeschicht modelliert einen k -dimensionalen Ausgabevektor mit $k = \text{Anzahl der Klassen}$. Wie bei NNs üblich, wird in einer Lernphase zunächst die Gewichtsmatrix des Netzes modelliert. Die Initialisierung des Netzes erfolgt dabei durch Wahl zufälliger Gewichte im Bereich von $[-0.25, 0.25]$. Bei der Lernphase werden dann die Trainingsvektoren durch das Netz propagiert und mit

⁵Davon kann abgewichen werden, wenn die Eingangs- bzw. Ausgangsdaten geeignet kodiert werden.

⁶Andere übliche Funktionen sind z.B. der hyperbolische Tangens.

dem Backpropagationalgorithmus die Gewichtsmatrix des Netzes angepaßt. Da die Trainingsvektoren bezüglich ihrer Klassenzugehörigkeit gekennzeichnet sind, wird im Ausgabevektor an der entsprechenden Stelle i die Klassenposition mit 1 kodiert und die verbleibenden Positionen im Vektor auf 0 gesetzt. Im Verlauf des Backprop-Trainings wird die Gewichtsmatrix unter Propagierung der zufällig gewählten Trainingsvektoren modelliert und nach Abschluß des Trainings gespeichert. Die hier verwendete Netztopologie bestand aus 3 Schichten (einer Eingangsschicht, einer versteckten Schicht und einer Ausgabeschicht). Das Netz wurde dabei mit 200000 Zyklen trainiert. Der Momentumwert wurde auf 0 festgelegt und die Lernrate ϵ im Bereich $[1.0, 0.01]$ variiert.

In der zweiten Phase, der Test bzw. Produktionsphase, bleiben die Gewichte konstant und die Testvektoren werden durch das Netz propagiert. Nach der Propagierung wird das Siegerneuron in der Ausgabeschicht ermittelt, welches den größten Wert zu 1 besitzt, wobei die Werte der Ausgabeneuronen als Wahrscheinlichkeit der Klassifikation aufgefaßt werden.

Das Training kann dabei mit einer festen maximalen Anzahl von Zyklen erfolgen oder entsprechend eines Abbruchkriteriums (wenn die Erkennungsrate hinreichend gut ist z.B. mittlerer quadratischer Fehler) vorzeitig beendet werden.

Ein wesentlicher Vorteil der neuronalen Netze ist dabei, daß in der Produktionsphase die Merkmalsvektoren lediglich nur noch durch das Netz traversiert werden müssen und eine weitere Verwendung der Trainingsdaten wie z.B. beim NN-Klassifikator entfallen kann. Damit sind neuronale Netze, wenn sie einmal trainiert wurden, sehr effizient und schnell in der Klassifikation, da die Klassengrenzen des Merkmalsraums durch die Gewichtsmatrix des Netzes fest modelliert sind. Des weiteren zeigen aktuelle Arbeiten zur Mustererkennung mit neuronalen Netzen sehr gute Klassifikationsergebnisse [63].

Kapitel 8

Experimente

In den folgenden Abschnitten werden zu den, in Kapitel 4 angegebenen Momenttypen Einzel- und Vergleichsanalysen durchgeführt. Dabei wurde der NIST-Datensatz mit 3471 numerischen Symbolen verwendet. Davon wurden, dem Prinzip der Kreuzvalidierung folgend, 2280 als Lerndaten und 1191 als Testdaten genutzt. Aus diesem Datensatz wurden zwei Analysesets abgeleitet, die im folgenden als Datensatz 1 (DS1) bzw. Datensatz 2 (DS2) bezeichnet werden. Diese Datensätze wurden aus den NIST-Daten gewonnen und sind wie folgt beschrieben:

DS1 entsprechend Kapitel 2 vorverarbeitete NIST-Daten mit einer Größe von 32×32 Pixeln.

DS2 entsprechend Kapitel 2 vorverarbeitete NIST-Daten mit einer Größe von 64×64 Pixeln, die bezüglich Rotation und Translation gestört wurden.

Für den Datensatz 2 wurde das Symbol (object of interest) zentriert, in ein Bild der Größe 64×64 Pixel eingebettet und darin zufällig um $\pm\{0, 1, \dots, 16\}$ Pixel verschoben und ebenfalls randomisiert um $\pm\{0, \dots, 45\}$ Grad rotiert.

Je nach Art der Merkmale wurden verschiedene verfahrensbedingte Normierungen, wie z.B. die *center of gravity* Skalierung angewandt, wie sie in den vorigen Abschnitten beschrieben wurden. Zusätzliche Normierungen wie zum Beispiel bzgl. des Erwartungswertes und der Varianz sind bei den jeweiligen Experimenten angegeben.

Die Analysen sind durch eine Vielzahl von Parametern beeinflusst. Dies sind konkret:

- Anzahl und Auswahl der Merkmale, die zur Klassifikation genutzt werden
- Wahl der Parameter p, q bzw. m, n, p für die Momentgenerierung
- Normierung der Merkmalsvektoren
- Art und Parametrisierung des Klassifikators

Zunächst wurden die Momente von Hu zu Voruntersuchungen genutzt, um die Anzahl der zu wählenden Parameter zu minimieren. Darauf aufbauend wurden für jeden Merkmalstyp annähernd äquivalente Experimente durchgeführt und diese Ergebnisse zueinander in Vergleich gesetzt. Die, im Anhang enthaltenen deskriptiven Statistiken zu den verschiedenen Verfahren wurden jeweils für 5 Merkmale notiert und über dem ersten Datensatz DS1 ermittelt.

8.1 Untersuchungen der Momente von Hu

Die Momente von Hu werden auf Basis der normalisierten Momente berechnet und haben damit Translations- und Skalierungsinvarianz zudem sind die, von Hu definierten Merkmale auch rotationsinvariant. Obwohl die 7 Momente von Hu kein vollständig unabhängiges System bilden, werden im folgenden alle 7 Momente betrachtet, um die Vergleichbarkeit z.B. zum Artikel von Chim et al. [71] zu erhalten.

Im folgenden wurden die Momente von Hu bezüglich folgender Aspekte untersucht:

- allgemeine statistische Daten
- Diskriminanzfähigkeit
- Erkennungsrate unter Nutzung verschiedener Klassifikatoren mit verschiedenen Normierungen

Auf Basis der erhaltenen Ergebnisse wurden dann die, später zu verwendenden Klassifikatoren und eine weitestgehend feste Parametrisierung für die nachfolgenden Merkmalsanalysen gewählt.

8.1.1 Allgemeine statistische Daten

Im Anhang auf Seite 108 ist tabellarisch die deskriptive Statistik zu den Momenten von Hu angegeben. Diese wurden über dem Trainingsdatensatz DS1 erstellt. Anhand dieser Daten ist ersichtlich, daß es zum Teil erhebliche Schwankungen der Merkmale für bestimmte Klassen gibt (hohe Innerklassenvarianz), dies wird besonders deutlich, wenn man die Merkmale 3 und 5 betrachtet. In Abbildung 8.1 ist ein Scatterplot zu diesen beiden Merkmalen dargestellt, der sehr gut zeigt, wie ungenügend stark eine allein damit durchgeführte Klassifikation ist, da die Merkmalsräume deutlich überlappen.

8.1.2 Diskriminanzanalyse

Wie im Abschnitt Merkmalsselektion auf Seite 53 bereits angegeben, sind nicht alle Merkmale für die korrekte Klassifikation gleich relevant. Für einen gegebenen Datensatz kann ein optimales Merkmalsset ermittelt werden. Nachfolgend wurde in Abbildung 8.2 mittels einer kanonischen Analyse die Verteilung der Trainingsdaten im

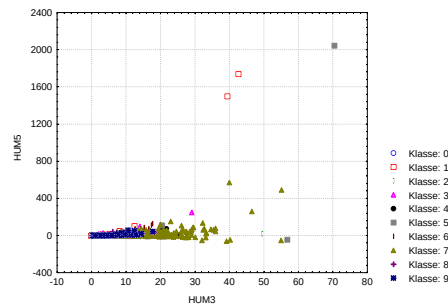


Abbildung 8.1: Scatterplot der Merkmale Hu3 und Hu5

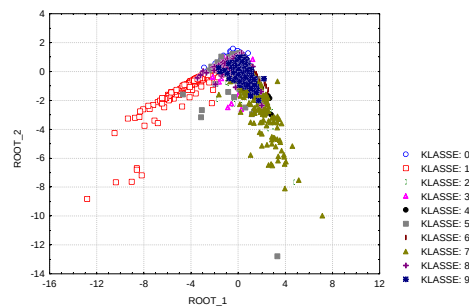


Abbildung 8.2: Kanonische Analyse der Merkmale von Hu

Merkmalsraum dargestellt. Dabei wurden die zwei, am höchsten diskriminierenden kanonischen Funktionen ($root_1$ und $root_2$) zur 2-dimensionalen Darstellung der Verteilung genutzt.

In Abbildung 8.2 wird bereits deutlich, daß die Merkmalsräume teilweise überlappen und die Klassengrenzen nur sehr unscharf existieren.

Obwohl diese Darstellung nicht die tatsächliche Verteilung der Daten im Merkmalsraum widerspiegelt¹, kann man bereits hier erkennen, daß durch die verwendeten Merkmale lediglich die Klasse 1 relativ scharf von den anderen Klassen abgegrenzt werden kann.

Der Wilks/Lambda-Wert des Merkmalssets von Hu ist mit 0.152 noch relativ hoch (0 – perfekte Trennung, 1 – keine Trennung) und nicht geeignet, um die 10 Klassen adäquat voneinander zu trennen.

8.1.3 Anwendung verschiedener Normierungen und Klassifikatoren

In dieser Arbeit wurden vier verschiedenartige Klassifikatoren und drei verschiedene Normierungen berücksichtigt (vgl. Aufzählung auf Seite 66), allerdings sind die Er-

¹Da wir ja eigentlich eine 7dimensionale Verteilung haben

Merkmal	Wilks λ
HUM1	0.169978
HUM2	0.253915
HUM3	0.224212
HUM4	0.184918
HUM5	0.170215
HUM6	0.169235
HUM7	0.164894

Tabelle 8.1: Diskriminanzwerte zu Momenten von Hu

gebnisse dazu nur für die Momente von Hu vollständig angegeben und für die anderen Momentverfahren wurden lediglich die Ergebnisse für die 3 besten Klassifikatoren und die „beste“ Normierung angegeben.

In Anlehnung an die Vorgehensweise von Chim et al.[71] wurden zu Beginn die verschiedenen Klassifikatoren ohne Normierung verglichen. Hier wurde deutlich, daß sich die besten Ergebnisse zunächst für das MLP ergeben. Und auf Platz 2 folgt die Diskriminanzkostenfunktion (DKF). Dies deckt sich nicht mit den Ergebnissen von Chim et al.[71], wo die DKF als schlechtester Klassifikator auftrat.

Klassifikator	Erkennungsrate ungewichtet	Erkennungsrate gewichtet
Kreuzkorrelation	42.68	45.13
Diskriminanzkosten- funktion	46.95	-
Nächster Nachbar mit euklidische Distanz	42.49	46.59
MLP	52.43	-

Tabelle 8.2: Erkennungsraten mit Hu-Momenten bei unterschiedlichen Klassifikatoren - unnormiert

Dieses Ergebnis war zunächst überraschend, läßt sich aber wahrscheinlich dadurch erklären, daß in dieser Arbeit im Gegensatz zu Chim et al.[71] handschriftliche Daten betrachtet wurden und sich damit andere Verteilungen für den Merkmalsraum ergeben.

Bei Chim et al.[71] wird als zusätzliche Variable die Größe der Trainingsmenge variiert, darauf wurde hier verzichtet, da allgemein anerkannt ist, daß sich die Erkennungsrate mit steigender Menge von Trainingsdaten einem Optimum annähert (vgl. [71]).

Entsprechend der Hinweise aus [32] und [12] wurden allerdings auch verschiedene Normierungen geprüft und im Ergebnis die Kreuzkorrelation, die k NN-Klassifikation und das MLP für spätere Analysen verwandt, da damit die besten Klassifikationsergebnisse erzielt wurden. Im folgenden wurden die verwendeten Normierungen wie folgt nummeriert:

0. keine Normierung

$$1. f'(x) = \frac{f(x) - \min}{\max - \min}$$

$$2. f'(x) = \frac{f(x) - \bar{x}}{\sigma_x}$$

$$3. f'(x) = \frac{f(x)}{\max - \min}$$

Klassifikator	Normierung	Erkennungsrate ohne Gewichtung	Erkennungsrate mit Gewichtung
Kreuzkorrelation	0	42.68	45.13
	1	45.90	44.61
	2	51.79	48.43
	3	43.79	40.00
Diskriminanz- kostenfunktion	0	46.95	-
	1	41.24	-
	2	48.32	-
	3	46.95	-
k NN ($k=5, \epsilon = 0$)	0	46.00	-
	1	49.08	-
	2	54.02	-
	3	49.08	-
MLP	0	52.43	-

Tabelle 8.3: Erkennungsraten mit Hu-Momenten bei unterschiedlichen Klassifikatoren und Normierungen

Die Kreuzkorrelation bietet bei Normierung 2 zwar offensichtlich eine gute Erkennungsrate, hat in der verwendeten Form aber den Nachteil, daß der zu klassifizierende Vektor mit allen Trainingsvektoren verglichen werden muß, was somit recht aufwendig ist. Das neuronale Netz weist diesen Nachteil nicht auf, da die nach dem Training erzeugte Gewichtsmatrix sehr schnell und effizient traversiert werden kann. Es wird

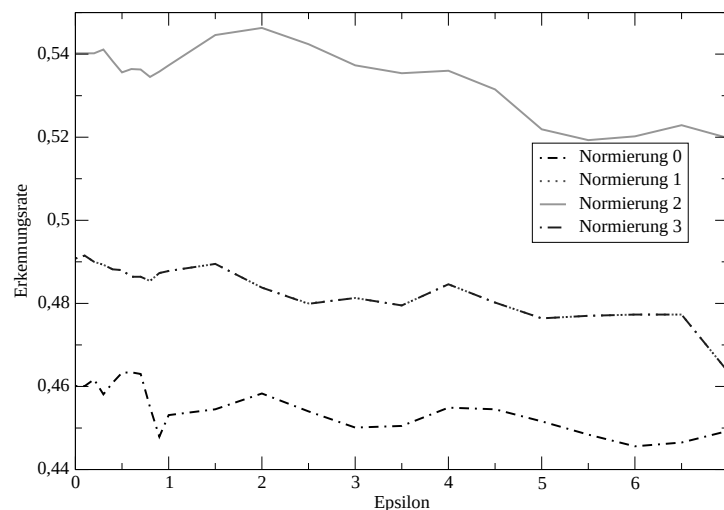


Abbildung 8.3: Analyse des Einflusses von ϵ bei k -NN und Momenten von Hu

ebenso deutlich, daß die Verwendung der gewichteten einfachen euklidischen Klassifikation bzw. der gewichteten Kreuzkorrelation nach Normierung nicht mehr ins Gewicht fällt, so daß im Folgenden auf deren Verwendung zugunsten der Normierung verzichtet wird.

Bei der euklidischen Distanz wurde auch das ANN-Verfahren genutzt, damit ist durch die einmalig erstellte Baumstruktur eine ebenfalls einfache Traversierung der Trainingsdaten möglich, die zudem durch Berücksichtigung der k nächsten Nachbarn eine zusätzliche Möglichkeit der Validierung der Klassifikationsergebnisse bietet. Daher wurden für dieses Verfahren zusätzlich noch weitere Analysen zur Ermittlung des Einflusses der Parameter (k, ϵ) durchgeführt.

Einfluß von ϵ bei der k NN-Klassifikation

Der Parameter ϵ bei der k ANN Suche dient zur approximativen Auffindung der k -nächsten Nachbarn. Dabei werden k unterschiedliche Punkte ermittelt, so daß das Verhältnis des i^{ten} approximativen Punktes dieser k Punkte zum tatsächlichen i^{ten} Nachbar höchstens $1 + \epsilon$ ist. Wird $\epsilon = 0.0$ gewählt, so wird eine exakte k -nächste Nachbarsuche durchgeführt. Im Folgenden wurde der Einfluß von ϵ bei konstant gewähltem $k = 5$ für verschiedene Normierungen untersucht.

Wie in der Abbildung 8.3 zu sehen ist, nimmt die Erkennungsrate bei steigendem ϵ ab, wobei sich eine maximale Erkennungsrate von 54.63% bei Verwendung von Normierung 2 mit $\epsilon = 2.0$ ergibt. Damit ist diese Erkennungsrate noch einmal etwas besser als bei Verwendung von Normierung 2 ohne Nutzung von ϵ ($\epsilon = 0.0$). Allgemein läßt

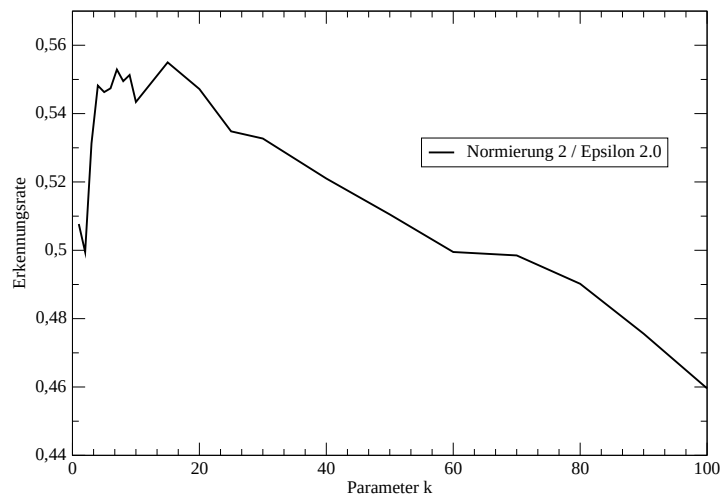


Abbildung 8.4: Analyse des Einflusses von k bei k NN und Momenten von Hu mit Normierung 2 und $\epsilon = 2.0$

sich jedoch sagen, daß der Einfluß des Parameters ϵ auf die Erkennungsrate relativ gering ausfällt.

Einfluß von k bei der k NN-Klassifikation

Die Berücksichtigung von mehr als einem nächsten Nachbarn, stellt eine interessante Möglichkeit dar, die Erkennungsrate zu verbessern. In der Abbildung 8.4 wurde der Einfluß dieses Parameters für die Normierung 2 und $\epsilon = 2.0$ grafisch dargestellt.

Wie zu sehen ist, kann unter Verwendung von 15 nächsten Nachbarn eine weitere Verbesserung der Klassifikationsrate auf 55.5% erreicht werden. Auch läßt sich aus den Daten ableiten, daß die Hinzunahme von weiteren nächsten Nachbarn ab einer bestimmten Schwelle (hier etwa $k = 20$) keine weitere Verbesserung, sondern so gar zu einer Verschlechterung der Klassifikationsrate führt. Dabei wurde bei dieser Form der Klassifikation das Mehrheitsprinzip zur Entscheidung verwendet.

Diesbezüglich wurde die Art der verwendeten nächsten Nachbarn verstärkt untersucht. Es wurde ermittelt, wie oft ein bestimmter nächster Nachbar an einer Entscheidung beteiligt ist und wie oft diese Entscheidung korrekt oder inkorrekt war. In Abbildung 8.5 sind die entsprechenden Kurven dargestellt, diese wurden über der nach den bisherigen Analysen optimalen Parametermenge ($k = 15, \epsilon = 2.0, \text{norm} = 2$) bestimmt.

Man erkennt, daß die ersten nächsten Nachbarn zu etwa 70% korrekt entscheiden (wenn sie am Entscheidungsprozess beteiligt sind), also relativ bedeutend für den Entscheidungsprozess sind, und danach fällt die Rate für eine korrekte Entscheidung allmählich, allerdings nur schwach, ab. Somit wird deutlich, daß im Bereich des „op-

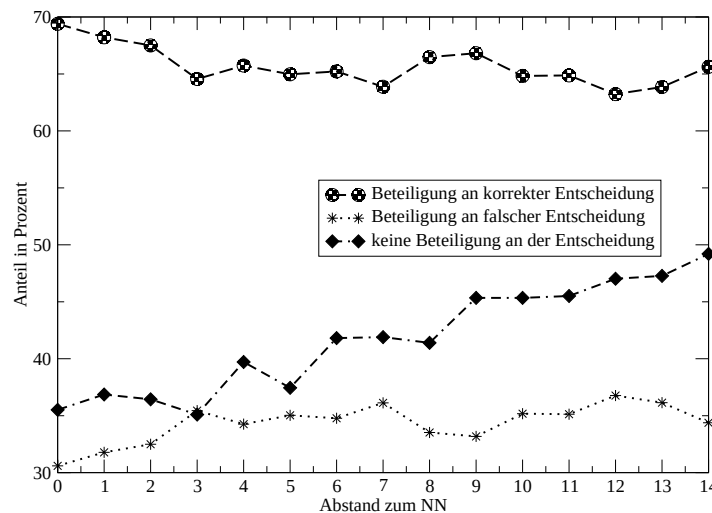


Abbildung 8.5: konkrete Analyse des Einflusses von k nächsten Nachbarn (über $k = 15, \epsilon = 2.0$)

timalen“ k auch die Hinzunahme weiterer Nachbarn das Ergebnis weder wesentlich verbessert noch verschlechtert. Allerdings nimmt die Bedeutung entfernter nächster Nachbarn stark ab, so daß man eventuell auf deren Berechnung zugunsten der Geschwindigkeit verzichten kann.

Insgesamt läßt sich feststellen, daß im Vergleich zur normalen, unnormierten nächsten Nachbar Klassifikation (44.69%), bei Verwendung der Normierung 2 mit $\epsilon = 2.0$ und $k = 15$ nächsten Nachbarn die Klassifikationsrate um 11% gesteigert werden konnte. Dabei haben die Parameter ϵ und k allerdings einen eher geringen Einfluß (ca. 1.5% Verbesserung), so daß sich vor allem die Normierung positiv auswirkt.

Für die 7 Momente von Hu zeigen die Ergebnisse, daß keine ausreichende Trennungsfähigkeit erreicht wurde und die Klassifikation zwar sehr schnell, mit einer maximalen Erkennungsrate von 55.5% jedoch relativ schlecht ist. Dies war allerdings bei dieser geringen Menge von Merkmalen auch nicht unbedingt anders zu erwarten.

In Anlehnung an die Arbeit von Chim et al. [71] wurde auch die Kombination der Merkmale von Hu und von Reiss betrachtet. Dabei ergab sich für die Klassifikation mittels k NN ($k = 15, \epsilon = 2.0, \text{norm} = 2$) eine Klassifikationsrate von 56.94% und bei Nutzung des MLP eine Klassifikationsrate von 57.20%. Diese nur geringe Verbesserung deckt sich mit den Ergebnissen von Chim et al. [71]

8.2 Untersuchungen zu geometrischen Momenten

Im folgenden werden die Momente von Tsirikolias und die normalisierten geometrischen Momente untersucht. Dabei werden zur Vereinfachung die aus dem vorigen Ka-

pitel ermittelten „optimalen“ Parameter genutzt und die Auswahl der Klassifikatoren damit auf k NN ($k = 15$, $\epsilon = 2.0$), die Kreuzkorrelation und das MLP beschränkt. Als Normierung wurde nach einigen Voruntersuchungen die Normierung 1 verwendet, da sich damit im allgemeinen bessere Ergebnisse ergaben. Auf die Nutzung der gewichteten Momente wurde nach einigen Vorversuchen verzichtet, da damit nur ungenügende Erkennungsraten erzielt wurden.

Beide Momenttypen sind entsprechend ihrer Definition invariant bezüglich Translation und Skalierung, jedoch nicht bezüglich Rotation, aus diesem Grund werden sie auch im späteren Vergleich gesondert behandelt.

8.2.1 Untersuchungen zu Tsirikolias-Mertzios-Momenten

Allgemeine statistische Daten

Im Anhang ist auf Seite 114 die deskriptive Statistik tabellarisch dargestellt. Für die verschiedenen Momenttypen wurden auch exemplarische Werte ermittelt, um die Invarianzeigenschaften zu überprüfen. Dabei fand das Symbol '5' aus Darstellung 3.2(f) Anwendung und wurde zentriert mit 32×25 (unkaliert) in einer 64×64 Pixel großen Matrix untersucht. Die Werte dazu sind in den jeweiligen Tabellen gegeben, wobei die Spalte „Grad der Transformation“ die Anzahl der durchgeführten Transformation wie folgt beschreibt²:

Grad 0 - keine Transformation, also der Momentwert ohne eine zusätzliche Bildtransformation

Grad 1 - einfache Translation (maximal um 16 Pixel) oder Skalierung auf $\frac{1}{3}$ der Ursprungsgröße oder Rotierung um 90°

Grad 2 - weitere Translation oder Skalierung auf $\frac{2}{3}$ der Ursprungsgröße oder Rotierung um 45°

Die entsprechenden Werte für die Tsirikolias-Mertzios-Momente sind in Tabelle 8.4 für einige Parametrisierungen³ dargestellt. Wie an den in Tabelle 8.4 gegebenen Werten ersichtlich, ist die Translationsinvarianz vollständig gegeben und auch bezüglich der Skalierungsinvarianz können die Werte als weitestgehend identisch angesehen werden. Für die über rotierten Symbolen ermittelten Momentwerte ergeben sich deutliche Unterschiede, so daß das Fehlen der Rotationsinvarianz offenbar wird. Die Abweichungen bei der Skalierungsinvarianz sind dabei auf den diskreten Zustand der Daten zurückzuführen, da bei zu geringer Größe des Symbols dieses diskretisierungsbedingt verfälscht wird. Alle weiteren Untersuchungen wurden für beide Datensätze DS1 und DS2 durchgeführt.

²Diese Transformationsabstufungen gelten auch für selbige Tabellen der anderen Momenttypen

³Wahl von p und q

Grad der Transformation	Parameter	Translation	Skaliert	Rotiert
0	(1, 1)	-0.5178	-0.5178	-0.5178
1	(1, 1)	-0.5178	-0.5710	0.5178
2	(1, 1)	-0.5178	-0.5014	0.1014
0	(3, 0)	-0.0662	-0.0662	-0.0662
1	(3, 0)	-0.0662	0.0327	0.1002
2	(3, 0)	-0.0662	-0.0554	0.0188
0	(0, 3)	0.10024	0.10024	0.10024
1	(0, 3)	0.10024	0.1438	0.0662
2	(0, 3)	0.10024	0.0230	0.4138

Tabelle 8.4: Einfluß der Translation, Skalierung oder Rotation auf Tsirikolias-Mertzios-Momente

Merkmalsselektion und Diskriminanzanalyse

Für die nachfolgenden Analysen wurden Tsirikolias-Mertzios-Momente bis ($p \leq 20$) berechnet und mittels SFS die 40 diskriminantesten Merkmale extrahiert⁴. Die in Abbildung 8.6(a) durchgeführte kanonische Analyse zeigt bereits eine gute Trennung der 10 Klassen und auch die Wilks-Lambda-Werte, die in Tabelle 8.5 dargestellt wurden, sind relativ niedrig. Hierbei wurden diese Wilks-Lambda-Werte für die mit SFS ermittelten besten 5 Merkmale für die Datensätze DS1 und DS2 dargestellt. Somit weisen die TM-Momente gute Trennfähigkeiten auf und erzeugen diskriminante Merkmale, die zur Nutzung im Klassifikationsprozeß geeignet sind. Für das ebenfalls untersuch-

Merkmal	Wilks λ DS1	Merkmal	Wilks λ DS2
TM1	0.010	TM1	0.195
TM2	0.011	TM1	0.177
TM3	0.007	TM1	0.190
TM4	0.008	TM1	0.207
TM5	0.007	TM1	0.186

Tabelle 8.5: Diskriminanzwerte zu Tsirikolias-Mertzios-Momenten

te zweite Datenset DS2 verschlechtern sich die Wilks-Lambda-Werte erheblich und auch die Trennungsfähigkeit nimmt sichtbar ab. Dies ist vor allem auf die fehlende Rotationsinvarianz zurückzuführen, die im DS2 gefordert ist.

Klassifikationsergebnisse

Die Klassifikation der Testdaten fand mit 40 diskriminanten Sets der Größe $\{1, \dots, 40\}$ für beide Datensätze statt. Dabei erfolgte die Klassifikation mittels k NN-Analyse,

⁴SFS bricht ab einer bestimmten Anzahl von Merkmalen ab, da dann ein Ausschlußkriterium gilt - deshalb sind die Grafiken gelegentlich abgeschnitten

Kreuzkorrelation (CC) und unter Verwendung des MLP mit den in Abschnitt 8.1 ermittelten Parametern.

Für das Testset DS1 ergibt sich die höchste Erkennungsrate bei Verwendung von 23 Merkmalen und unter Nutzung der Kreuzkorrelation. Die Ergebnisse differieren dabei je nach verwendetem Klassifikator und können in Tabelle 8.6 für die höchsten Erkennungsraten nachgelesen werden. Die genauen Kurvenverläufe dazu sind in Abbildung

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1	78.34% (23 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, Normierung 1	68.34% (4 Merkmale)
MLP mit Normierung 1	66.09% (26 Merkmale)

Tabelle 8.6: Klassifikationsergebnisse für DS1 mit Tsirikolias-Mertzios-Momenten

8.7(a) zu sehen. Es wird deutlich, daß bereits mit der geringen Menge von 5 Merkmalen eine Erkennungsrate von 70% erreicht werden kann und die Erkennungsrate danach eher stagniert.

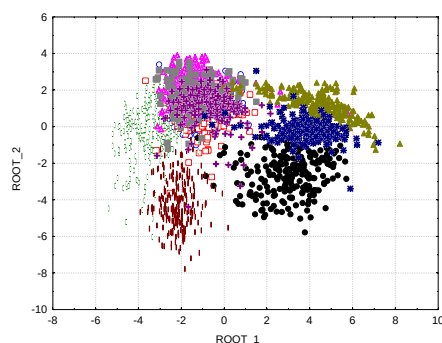
Für das zweite Datenset DS2 sind infolge der auch bezüglich Rotation gestörten Daten erhebliche Fehlklassifikationen zu erwarten, da die TM-Momente keine Rotationsinvarianz aufweisen. In der Tat fällt die Erkennungsrate auch deutlich auf etwa 60% ab und die 70% Korrektheit konnten nicht mehr erreicht werden. Die genauen Werte sind in der Tabelle 8.7 und in Abbildung 8.7(b) angegeben.

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1	61.15% (28 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, Normierung 1	49.44% (32 Merkmale)
MLP mit Normierung 1	55.36% (26 Merkmale)

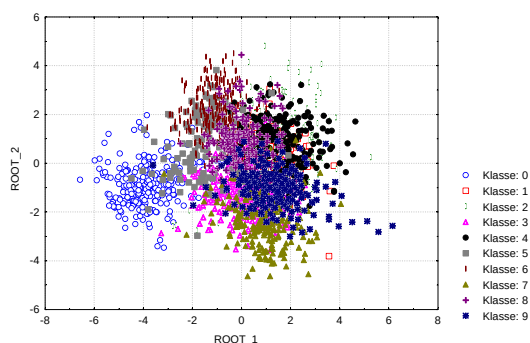
Tabelle 8.7: Klassifikationsergebnisse für DS2 mit Tsirikolias-Mertzios-Momenten

Validität der korrekten Klassifikationen für k NN bei TM-Momenten Für den Fall des k NN-Klassifikators wurde noch der Grad der Validität entsprechend des Konzepts aus Abschnitt 7.4.1 untersucht, dabei ergaben sich bei der korrekten Klassifikation (ermittelt über dem Set mit maximaler Erkennungsrate) die in Tabelle 8.8 gezeigten Werte. Man erhält für den ersten Datensatz eine akzeptable Validität unter den korrekten Klassifikationen, da im Mittel über 70% der analysierten nächsten Nachbarn (11/15) korrekt klassifiziert wurden. Auch die minimale Validität ist mit einem Viertel der nächsten Nachbarn noch relativ hoch. Somit können die Klassifikationsentscheidungen für korrekte Entscheidungen hier als relativ sicher angesehen werden. Im Falle des zweiten, gestörten Datensatzes ergibt sich für die mittlere Validität ein Abfall von 10%.

Über den besten Sets aus der Klassifikation wurden nun noch die maximal zu berechnenden Polynomordnungen bestimmt. Die Ergebnisse dazu sind in Tabelle 8.9 zu

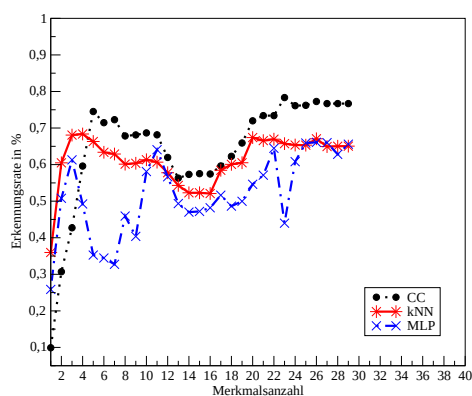


(a) DS1

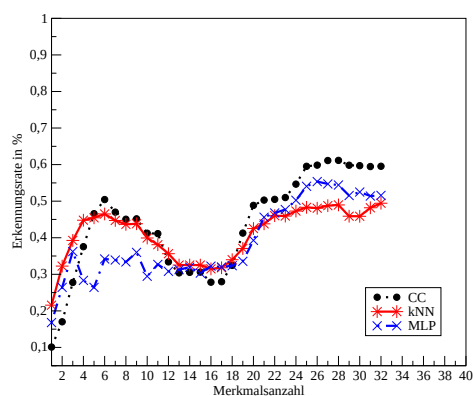


(b) DS2

Abbildung 8.6: Kanonische Analyse der 40 diskriminantesten TM-Momente für DS1 und DS2



(a) DS1



(b) DS2

Abbildung 8.7: Klassifikationsrate mit TM-Momenten für DS1 und DS2

Datensatz	Prüfgröße	Wert
DS1 (Setgröße= 4)	MIN	26.67 %
	MAX	100 %
	AVG	72.57 %
DS2 (Setgröße= 32)	MIN	20%
	MAX	100%
	AVG	62.40%

Tabelle 8.8: Validität der k NN Klassifikation bei Nutzung von Tsirikolias-Mertzios-Momenten

finden. Somit müssen also keine monomialen Polynome größer 20 berechnet werden.

Datensatz	maximale Ordnung $p + q$	$\max p$	$\max q$
DS1	34	14	10
DS2	19	20	19

Tabelle 8.9: Maximale Ordnung der Parameter p und q für Tsirikolias-Mertzios-Momente

Bei Berücksichtigung der späteren Ergebnisse ist diese Ordnung allerdings schon relativ hoch.

8.2.2 Untersuchungen zu normalisierten Momenten

Allgemeine statistische Daten

Im Anhang auf Seite 111 ist zur Information wieder die deskriptive Statistik gegeben. Wie an den in Tabelle 8.10 angegebenen exemplarischen Werten ersichtlich, ist die Translationsinvarianz vollständig und auch die Skalierungsinvarianz ist weitestgehend gegeben. Rotationsinvarianz konnte erwartungsgemäß nicht festgestellt werden. Die folgenden Analysen wurden wieder für beide Datensätze DS1 und DS2 durchgeführt.

Merkmalsselektion und Diskriminanzanalyse

Auch bei den normalisierten Momenten wurden für die verschiedenen Analysen Momente bis ($p \leq 20$) berechnet und mittels SFS die 40 diskriminantesten Merkmale ermittelt. Die in Abbildung 8.8(a) durchgeführte kanonische Analyse zeigt bereits eine gute Trennung der 10 Klassen und auch die Wilks-Lambda-Werte, die in Tabelle 8.11 dargestellt wurden, sind relativ niedrig. Somit weisen die NM-Momente gute Trennfähigkeiten auf. Allerdings wird beim Vergleich mit den TM-Momenten deutlich, daß die NM-Momente bei sehr wenigen Merkmalen (hier 5) eine schlechtere Trennfähigkeit als TM-Momente aufweisen und im weiteren Vergleich mit den orthogonalen Momenten ebenfalls sehr schlecht abschneiden. Für das ebenfalls untersuchte

Grad der Transformation	Parameter	Translation	Skaliert	Rotiert
0	(1, 1)	-0.1536	-0.1536	-0.1536
1	(1, 1)	-0.1536	-0.1636	0.1536
2	(1, 1)	-0.1536	-0.1478	0.0299
0	(3, 0)	-0.1291	-0.1291	-0.1291
1	(3, 0)	-0.1291	0.0282	0.2442
2	(3, 0)	-0.1291	-0.0797	0.0158
0	(0, 3)	0.2442	0.2442	0.2442
1	(0, 3)	0.2442	0.1925	0.1291
2	(0, 3)	0.2442	0.0401	1.9565

Tabelle 8.10: Einfluß der Translation, Skalierung oder Rotation auf normalisierte Momente

Merkmal	Wilks λ DS1	Merkmal	Wilks λ DS2
NM1	0.121	NM1	0.457
NM2	0.098	NM1	0.497
NM3	0.094	NM1	0.429
NM4	0.094	NM1	0.399
NM5	0.086	NM1	0.377

Tabelle 8.11: Diskriminanzwerte zu normalisierten Momenten

zweite Datenset DS2 verschlechtern sich die Wilks-Lambda-Werte erheblich und auch die Trennungsfähigkeit nimmt sichtbar ab.

Klassifikationsergebnisse

Die Klassifikationen wurden mit 40 diskriminanten Sets der Größe $\{1, \dots, 40\}$ für beide Datensätze unter Verwendung von Kreuzkorrelation, k NN-Analyse und MLP durchgeführt.

Für das Testset DS1, ergibt sich die höchste Erkennungsrate bei Verwendung des k NN mit 11 Merkmalen. Die höchsten Erkennungsraten pro verwendetem Klassifikator sind in Tabelle 8.12 gegeben. Dabei sind die genauen Kurvenverläufe in Abbildung

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1 (12 - Merkmale)	68.45%
k NN mit $k = 15$, $eps = 2.0$, Normierung 1 (11 - Merkmale)	71.39%
MLP mit Normierung 1 (11 - Merkmale)	63.31%

Tabelle 8.12: Klassifikationsergebnisse für DS1 mit normalisierten Momenten

8.9(a) zu sehen. Es wird deutlich, daß bereits mit der geringen Menge von 5 Merkmalen eine Erkennungsrate von 70% erreicht werden kann und danach nur noch eine

langsame Steigerung der Korrektheit erfolgt.

Für das zweite Datenset DS2 sind infolge der auch bezüglich Rotation gestörten Daten erhebliche Fehlklassifikationen zu erwarten, da die NM-Momente keine Rotationsinvarianz aufweisen. In der Tat fällt die Erkennungsrate auch deutlich auf etwa 47% ab. Die genauen Werte sind wieder in der folgenden Tabelle 8.13 und in Abbildung 8.9(b) dargestellt. Über den besten Sets aus der Klassifikation mit normalisierten

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1	45.11% (12 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, Normierung 1	46.60% (13 Merkmale)
MLP mit Normierung 1	42.36% (8 Merkmale)

Tabelle 8.13: Klassifikationsergebnisse für DS2 mit normalisierten Momenten

Momenten wurden noch die maximal zu berechnenden Polynomordnungen bestimmt. Die Ergebnisse sind in der folgenden Tabelle 8.14 dargestellt und zeigen, daß keine monomialen Polynome größer 15 berechnet werden müssen.

Datensatz	maximale Ordnung $p + q$	$\max p$	$\max q$
DS1	8	7	3
DS2	15	15	7

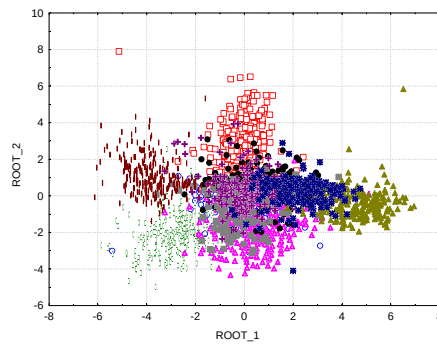
Tabelle 8.14: Maximale Ordnung der Parameter p und q für normalisierte Momente

Validität der korrekten Klassifikationen für k NN bei NM-Momenten Mit den Ergebnissen des k NN Klassifikators wurde wieder der Grad der Gültigkeit bestimmt, dabei ergaben sich für den Fall der korrekten Klassifikation (maximale Erkennungsrate) die in Tabelle 8.15 angegebenen Prozentwerte. Für den ersten Datensatz wurde eine

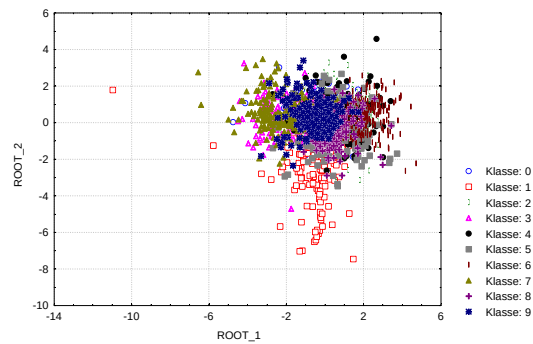
Datensatz	Prüfgröße	Wert
DS1 (Setgröße= 11)	MIN	26.67%
	MAX	100 %
	AVG	73.77 %
DS2 (Setgröße= 13)	MIN	20%
	MAX	100%
	AVG	53.48%

Tabelle 8.15: Validität der k NN-Klassifikation bei Nutzung von normalisierten Momenten

relativ hohe Validität unter den korrekten Klassifikationen erreicht, da im Mittel über 70% der analysierten nächsten Nachbarn (11/15) die selbe Klassifikation durchgeführt haben. Auch die minimale Validität ist mit einem Viertel der nächsten Nachbarn noch

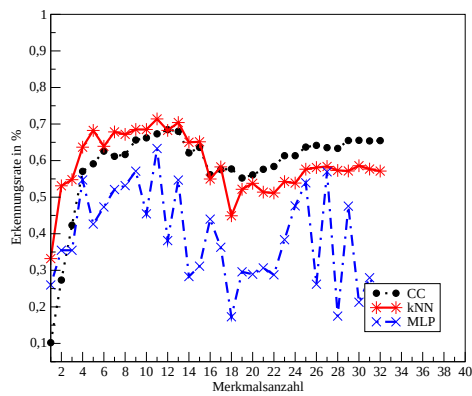


(a) DS1

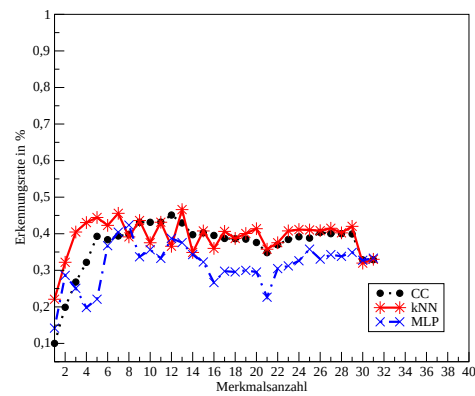


(b) DS2

Abbildung 8.8: Kanonische Analyse der 40 diskriminantesten NM-Momente für DS1 und DS2



(a) DS1



(b) DS2

Abbildung 8.9: Klassifikationsrate mit NM-Momenten für DS1 und DS2

relativ hoch. Somit können die Klassifikationsentscheidungen für korrekte Entscheidungen auch hier als relativ sicher angesehen werden. Im Falle des zweiten, gestörten Datensatzes nehmen sowohl die minimale als auch die mittlere Validität deutlich ab und für die mittlere Validität sinkt die Güte der Klassifikation sogar um über 20%.

Damit und mit den schlechten Klassifikationsraten für den gestörten Datensatz DS2 wird deutlich, daß die normalisierten geometrischen Momente auf die Rotation im Datensatz DS2 sehr sensibel reagieren.

8.3 Untersuchungen der orthogonalen Momente

In diesem Abschnitt werden die Ergebnisse der Analysen der orthogonalen Momente aus dem Abschnitt 5 vorgestellt. Dabei wurden die Merkmale für alle Verfahren bis $p \leq 20$ berechnet und darüber die ersten 40 diskriminanten Merkmalskombinationen mit SFS gebildet. Die Klassifikationen erfolgten jeweils unter Verwendung der Kreuzkorrelation (in den Grafiken mit CC bezeichnet), des k NN Klassifikators und des MLPs. Die Merkmalsvektoren bestehen dabei sowohl aus rotationsinvarianten Merkmalen, ermittelt durch Normbildung, als auch aus reellen Merkmalen (ohne Betragsbildung).

8.3.1 Untersuchungen zu Legendre-Momenten

Allgemeine statistische Daten

Zur Information ist im Anhang auf Seite 112 die deskriptive Statistik dargestellt. Die Werte in Tabelle 8.16 zeigen, daß die Translationsinvarianz vollständig gegeben ist und auch bezüglich der Skalierungsinvarianz ähnliche Werte auftreten. Die Werte für die hinsichtlich Rotation transformierten Bilder weichen hingegen deutlich ab.

Grad der Transformation	Parameter	Translation	Skaliert	Rotiert
0	(1, 1)	−0.6277	−0.6277	−0.6277
1	(1, 1)	−0.6277	−0.8215	0.6440
2	(1, 1)	−0.6277	−0.6561	−0.0605
0	(3, 0)	−0.1766	−0.1766	−0.1766
1	(3, 0)	−0.1766	−0.0514	0.1002
2	(3, 0)	−0.1766	−0.1900	0.1111
0	(0, 3)	0.0763	0.0763	0.0763
1	(0, 3)	0.0763	0.1608	0.2028
2	(0, 3)	0.0763	0.1126	0.5759

Tabelle 8.16: Einfluß der Translation, Skalierung oder Rotation auf Legendre-Momente

Merkmalsselektion und Diskriminanzanalyse

Der in Abbildung 8.10(a) dargestellte Scatterplot zur kanonischen Analyse zeigt bereits eine außerordentlich gute Trennung der 10 Klassen und auch die Wilks-Lambda-Werte, die in Tabelle 8.17 notiert wurden, sind sehr niedrig. Somit weisen die Legendre-Momente sehr gute Trennfähigkeiten auf und erzeugen diskriminante Merkmale. Für das ebenfalls untersuchte zweite Datenset DS2 verschlechtern sich sowohl die Wilks-Lambda-Werte als auch die Trennungsfähigkeit. Dies ist vor allem auf die fehlende Rotationsinvarianz zurückzuführen.

Merkmal	Wilks λ DS1	Merkmal	Wilks λ DS2
LM1	0.038	LM1	0.163
LM2	0.024	LM1	0.191
LM3	0.021	LM1	0.174
LM4	0.020	LM1	0.158
LM5	0.018	LM1	0.150

Tabelle 8.17: Diskriminanzwerte zu Legendre-Momenten

Klassifikationsergebnisse

Für das Testset DS1 ergibt sich die höchste Erkennungsrate mit 96% bei Verwendung von 39 Merkmalen. Die Ergebnisse differieren dabei je nach verwendetem Klassifikator und sind in Tabelle 8.18 für die besten Klassifikationsraten mit den verwendeten drei Klassifikatoren angegeben. Dabei sind die genauen Kurvenverläufe in Abbildung

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1	96.32% (39 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, Normierung 1	93.95% (32 Merkmale)
MLP	94.47% (38 Merkmale)

Tabelle 8.18: Klassifikationsergebnisse für DS1 mit Legendre-Momenten

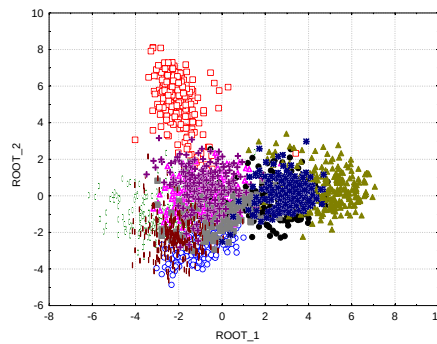
8.11(a) zu sehen. Es wird deutlich, daß bereits mit der geringen Menge von 7 Merkmalen eine Erkennungsrate von 80% erreicht werden kann und danach nur noch eine langsame Steigerung der Erkennungsrate erfolgt.

Für das zweite Datenset DS2 sind infolge der auch bezüglich Rotation gestörten Daten stärkere Fehlklassifikationen zu erwarten, da die Legendre-Momente keine Rotationsinvarianz aufweisen. In der Tat fällt die Erkennungsrate auf etwa 86% für die Kreuzkorrelation mit Normierung 1 ab. Insgesamt ist der Einfluß der Transformationen des DS2 auf die Legendre-Momente aber relativ gering. Die genauen Werte für die höchsten Erkennungsraten sind in der Tabelle 8.19 und in Abbildung 8.11(b) dargestellt. Über den besten Sets aus der Klassifikation mit Legendre-Momenten wurden

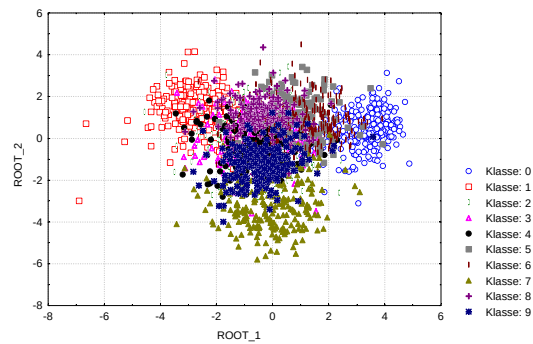
Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1	86.21% (35 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, Normierung 1	81.77% (32 Merkmale)
MLP	83.16% (39 Merkmale)

Tabelle 8.19: Klassifikationsergebnisse für DS2 mit Legendre-Momenten

noch die maximal zu berechnenden Polynomordnungen bestimmt. Die Ergebnisse sind in der folgenden Tabelle 8.20 dargestellt. Somit müssen also keine Legendre Polynome größer 13 berechnet werden, was sich auf die Komplexität außerordentlich günstig auswirkt.

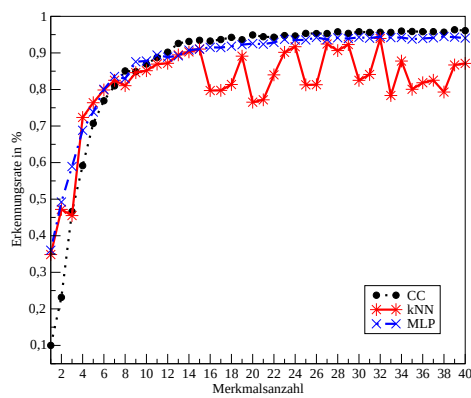


(a) DS1

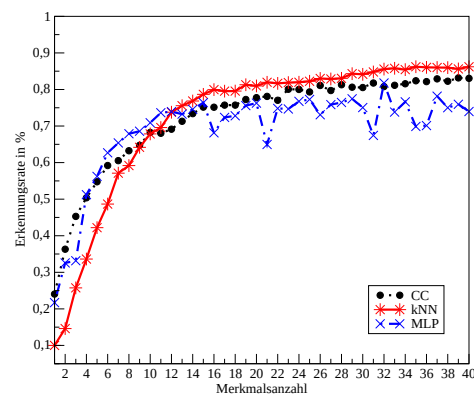


(b) DS2

Abbildung 8.10: Kanonische Analyse der 40 diskriminantesten Legendre-Momente für DS1 und DS2



(a) DS1



(b) DS2

Abbildung 8.11: Klassifikationsrate mit Legendre-Momenten für DS1 und DS2

Datensatz	maximale Ordnung $p + q$	$\max p$	$\max q$
DS1	19	9	13
DS2	19	11	10

Tabelle 8.20: Maximale Ordnung der Parameter p und q für Legendre-Momente

Validität der korrekten Klassifikationen für k NN bei Legendre-Momenten Für den Fall des k NN Klassifikators wurde noch der Grad der Validität der korrekten Klassifikationen untersucht, die erhaltenen Werte sind in Tabelle 8.21 dargestellt. Für den

Datensatz	Prüfgröße	Wert
DS1 (Setgröße= 32)	MIN	26.67%
	MAX	100%
	AVG	90.67%
DS2 (Setgröße= 32)	MIN	26.67%
	MAX	100%
	AVG	78.21%

Tabelle 8.21: Validität der k NN Klassifikation bei Nutzung von Legendre-Momenten

ersten Datensatz ergibt sich eine sehr gute Validität unter den korrekten Klassifikationen, da im Mittel über 90% der analysierten nächsten Nachbarn (14/15) die selbe Klassifikation durchgeführt haben. Auch die minimale Validität ist mit einem Viertel der nächsten Nachbarn noch relativ hoch. Somit können die Klassifikationsentscheidungen für korrekte Entscheidungen hier als sehr sicher angesehen werden. Im Falle des zweiten, gestörten Datensatzes nehmen sowohl die minimale als auch die mittlere Validität deutlich ab, sind aber im Vergleich mit anderen Verfahren noch immer hoch.

8.3.2 Untersuchungen zu Orthogonal Fourier-Mellin-Momenten

Allgemeine statistische Daten

Die deskriptive Statistik kann auf Seite 115 eingesehen werden. In der nachfolgenden Tabelle 8.22 wurden wieder einige OFM-Momente exemplarisch berechnet, um deren Invarianzeigenschaften zu prüfen. Dabei weisen die OFM-Momente in der hier verwendeten Form sowohl Translations-, Skalierungs- und Rotationsinvarianz auf.

Merkmalsselektion und Diskriminanzanalyse

Die in Abbildung 8.12(a) durchgeführte kanonische Analyse zeigt bereits eine gute Trennung der 10 Klassen und auch die Wilks-Lambda-Werte, die in Tabelle 8.23 dargestellt wurden, sind sehr niedrig. Deshalb besitzen die OFM-Momente gute Trennfähigkeiten. Für das ebenfalls untersuchte zweite Datenset DS2 verschlechtern sich die

Grad der Transformation	Parameter	Translation	Skaliert	Rotiert
0	(1, 1)	0.0207	0.0207	0.0207
1	(1, 1)	0.0207	0.0417	0.0255
2	(1, 1)	0.0207	0.0453	0.0196
0	(3, 0)	3.8764	3.8764	3.8764
1	(3, 0)	3.8764	4.2155	3.2226
2	(3, 0)	3.8764	2.3815	3.8041
0	(0, 3)	0.0020	0.0020	0.0020
1	(0, 3)	0.0020	0.0082	0.0012
2	(0, 3)	0.0020	0.0013	0.0012

Tabelle 8.22: Einfluß der Translation, Skalierung oder Rotation auf OFM-Momente

Merkmal	Wilks λ DS1	Merkmal	Wilks λ DS2
OFM1	0.026	OFM1	0.078
OFM2	0.020	OFM1	0.072
OFM3	0.013	OFM1	0.076
OFM4	0.012	OFM1	0.073
OFM5	0.012	OFM1	0.058

Tabelle 8.23: Diskriminanzwerte zu OFM-Momenten

Wilks-Lambda-Werte und auch die Trennungsfähigkeit (vgl. Abbildung 8.12(b)) nur gering, so daß die Invarianzeigenschaften der OFM-Momente deutlich werden.

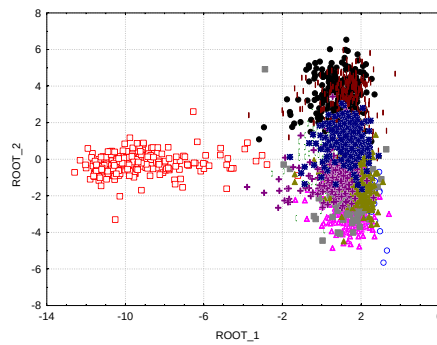
Klassifikationsergebnisse

Für das Testset DS1 ergibt sich die höchste Erkennungsrate bei Verwendung des MLP mit Normierung 1 und 32 Merkmalen. Die höchsten Erkennungsraten pro verwendetem Klassifikator sind für DS1 in Tabelle 8.24 und DS2 in Tabelle 8.25 angegeben.

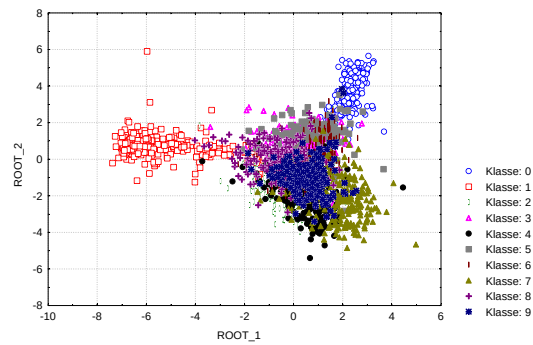
Klassifikator	Erkennungsrate
Kreuzkor. mit Normierung 1	92.47% (37 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, norm=1	87.28% (32 Merkmale)
MLP	92.84% (32 Merkmale)

Tabelle 8.24: Klassifikationsergebnisse für DS1 mit OFM-Momenten

Über den besten Sets aus der Klassifikation mit OFM-Momenten wurden noch die maximal zu berechnenden Polynomordnungen bestimmt. Die dabei erhaltenen Ergebnisse sind in Tabelle 8.26 aufgeführt. Somit müssen also keine OFM Polynomordnungen größer 11 berechnet werden.

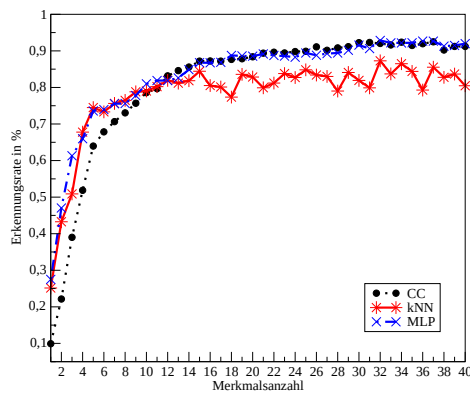


(a) DS1

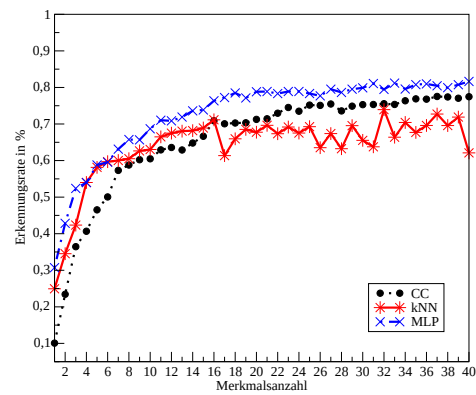


(b) DS2

Abbildung 8.12: Kanonische Analyse der 40 diskriminantesten OFM-Momente für DS1 und DS2



(a) DS1



(b) DS2

Abbildung 8.13: Klassifikationsrate mit Orthogonal-Fourier-Mellin-Momenten für DS1 und DS2

Klassifikator	Erkennungsrate
Kreuzkorr. mit Normierung 1	77.53% (37 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, $norm=1$	73.93% (32 Merkmale)
MLP	81.64 % (40 Merkmale)

Tabelle 8.25: Klassifikationsergebnisse für DS2 mit OFM-Momenten

Datensatz	maximale Ordnung $p + q$	$\max p$	$\max q$
DS1	11	11	6
DS2	08	7	7

Tabelle 8.26: Maximale Ordnung der Parameter p und q für OFM-Momente

Validität der korrekten Klassifikationen für k NN bei OFM-Momenten Für den Fall der Gültigkeitsprüfung sind die in Tabelle 8.27 angegebenen Werte vergleichbar mit denen der Legendre-Momente und allgemein recht gut. Die Klassifikationen sind somit

Datensatz	Prüfgröße	Wert
DS1 (Setgröße= 32)	MIN	20%
	MAX	100%
	AVG	89.10%
DS2 (Setgröße= 32)	MIN	20%
	MAX	100%
	AVG	78.40%

Tabelle 8.27: Validität der k NN-Klassifikation bei Nutzung von OFM-Momente

als sehr sicher einzustufen, da mit fast 90% Prozent für den ersten Datensatz und fast 80% für den gestörten Datensatz hohe Validitätswerte erzielt wurden.

8.3.3 Untersuchungen zu Pseudo-Zernike-Momenten

Allgemeine statistische Daten

Im Anhang auf Seite 110 ist die deskriptive Statistik der Pseudo-Zernike-Momente zu finden. Wie an den exemplarischen Werten aus Tabelle 8.27 ersichtlich, ist die Translationsinvarianz vollständig gegeben und auch bezüglich Skalierungsinvarianz können die Werte als nahezu identisch angesehen werden. Auch für die Rotation sind die Werte relativ ähnlich.

Merkmalsselektion und Diskriminanzanalyse

Die in Abbildung 8.14(a) angegebene kanonische Analyse zeigt bereits eine gute Trennung der 10 Klassen und auch die Wilks-Lambda-Werte, die in Tabelle 8.29 dargestellt wurden, sind relativ niedrig. Somit weisen die PZM-Momente gute Trennfähigkeiten

Grad der Transformation	Parameter	Translation	Skaliert	Rotiert
0	(4, 0)	3.2064	3.2064	3.2064
1	(4, 0)	3.2064	4.3199	2.6052
2	(4, 0)	3.2064	1.9438	2.9075
0	(4, 2)	0.2664	0.2664	0.2664
1	(4, 2)	0.2664	0.1558	0.3310
2	(4, 2)	0.2664	0.3764	0.3390
0	(2, 2)	0.0910	0.0910	0.0910
1	(2, 2)	0.0910	0.1866	0.1082
2	(2, 2)	0.0910	0.1102	0.1139

Tabelle 8.28: Einfluß der Translation, Skalierung oder Rotation auf Pseudo-Zernike-Momente

auf. Für das ebenfalls untersuchte zweite Datenset DS2 verschlechtern sich die Wilks-

Merkmal	Wilks λ DS1	Merkmal	Wilks λ DS2
PZM1	0.018	PZM1	0.106
PZM2	0.014	PZM1	0.099
PZM3	0.017	PZM1	0.097
PZM4	0.015	PZM1	0.084
PZM5	0.014	PZM1	0.084

Tabelle 8.29: Diskriminanzwerte zu Pseudo-Zernike-Momenten

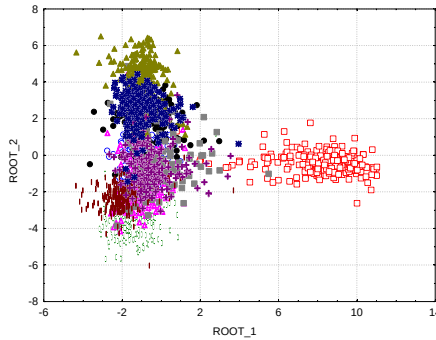
Lambda-Werte erheblich und auch die Trennungsfähigkeit nimmt sichtbar ab.

Klassifikationsergebnisse

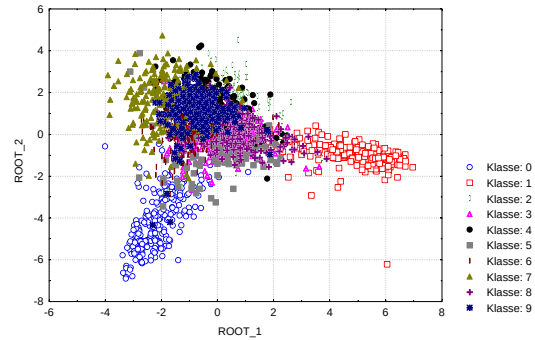
Für das Testset DS1 ergibt sich die höchste Erkennungsrate bei Verwendung des MLP mit Normierung 1 und 40 Merkmalen. In Tabelle 8.30 sind die höchsten Erkennungs-raten pro verwendetem Klassifikator angegeben. Der zweite Datensatz wird deutlich schlechter klassifiziert und es können nur noch Erkennungs-raten von knapp 80% erreicht werden. Die Werte dazu sind in Tabelle 8.31 angegeben.

Klassifikator	Erkennungsrate
Kreuzkorr. mit Normierung 1	91.22% (34 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, norm=2	86.78% (39 Merkmale)
MLP	91.67% (40 Merkmale)

Tabelle 8.30: Klassifikationsergebnisse für DS1 mit Pseudo-Zernike-Momenten

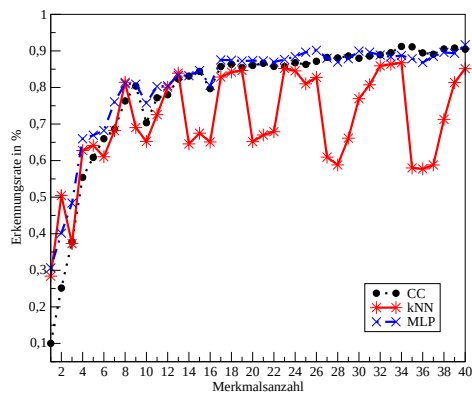


(a) DS1

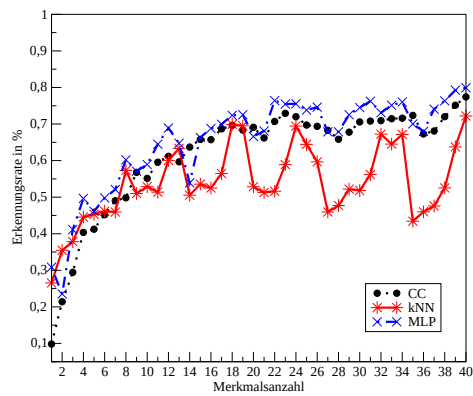


(b) DS2

Abbildung 8.14: Kanonische Analyse der 40 diskriminantesten Pseudo-Zernike-Momente für DS1 und DS2



(a) DS1



(b) DS2

Abbildung 8.15: Klassifikationsrate mit Pseudo-Zernike-Momenten für DS1 und DS2

Klassifikator	Erkennungsrate
Kreuzkorr. mit Normierung 1	77.42% (40 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, norm=2	72.20% (40 Merkmale)
MLP	79.91% (40 Merkmale)

Tabelle 8.31: Klassifikationsergebnisse für DS2 mit Pseudo-Zernike-Momenten

Validität der korrekten Klassifikationen für k NN bei PZ-Momenten Auch für die Pseudo-Zernike-Momente wurde der Grad der Gültigkeit der Klassifikation untersucht. Dabei traten für den Fall der korrekten Klassifikation (ermittelt über dem Set mit der maximalen Erkennungsrate) die in Tabelle 8.32 gegebenen Werte auf. Die Validität der Klassifikation ist somit auch für die Pseudo-Zernike-Momente relativ gut und kann als hoch angesehen werden. Die maximal zu berechnenden Polynomordnungen

Datensatz	Prüfgröße	Wert
DS1 (Setgröße= 34)	MIN	33.33%
	MAX	100%
	AVG	89.15%
DS2 (Setgröße= 40)	MIN	20%
	MAX	100%
	AVG	75.17%

Tabelle 8.32: Validität der k NN-Klassifikation bei Nutzung von Pseudo-Zernike-Momenten

für die besten Sets sind in Tabelle 8.33 gegeben und zeigen, daß keine Pseudo-Zernike Polynome der Ordnung größer 16 berechnet werden.

Datensatz	maximale Ordnung $p + q$	max p	max q
DS1	21	14	7
DS2	25	16	9

Tabelle 8.33: Maximale Ordnung der Parameter p und q für Pseudo-Zernike-Momente

8.3.4 Untersuchungen zu Zernike-Momenten

Allgemeine statistische Daten

Für die Tabelle 8.34 wurden einige Zernike-Momente unter verschiedenen Transformationen berechnet und es ist offensichtlich, daß sowohl die Translationsinvarianz als auch die Skalierungsinvarianz gegeben ist. Auch für die Rotation sind ähnlich Werte zu erkennen. Alle weiteren Analysen wurden für beide Datensätze DS1 und DS2 durchgeführt. Die deskriptive Statistik ist im Anhang auf Seite 109 dargestellt.

Grad der Transformation	Parameter	Translation	Skaliert	Rotiert
0	(4, 0)	0.0040	0.0040	0.0040
1	(4, 0)	0.0040	0.0098	0.0001
2	(4, 0)	0.0040	0.0281	0.0039
0	(4, 2)	0.1017	0.1017	0.1017
1	(4, 2)	0.1017	0.0456	0.1211
2	(4, 2)	0.1017	0.1219	0.1175
0	(2, 2)	0.1767	0.1767	0.1767
1	(2, 2)	0.1767	0.3168	0.1854
2	(2, 2)	0.1767	0.1917	0.1860

Tabelle 8.34: Einfluß der Translation, Skalierung oder Rotation auf Zernike-Momente

Merkmalsselektion und Diskriminanzanalyse

Die in Abbildung 8.16(a) durchgeführte kanonische Analyse zeigt bereits eine gute Trennung der 10 Klassen und auch die Wilks-Lambda-Werte, die in Tabelle 8.35 dargestellt wurden, sind relativ niedrig. Somit weisen die Zernike-Momente gute Trennfähigkeiten auf, wenn gleich im Fall des zweiten Datensatzes eine deutliche Verschlechterung auftritt.

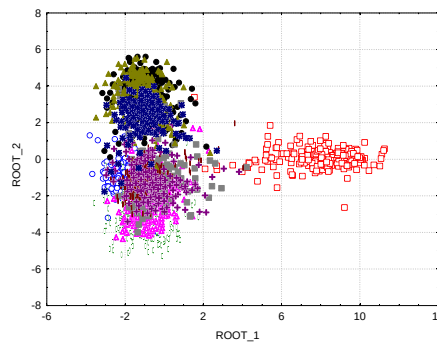
Merkmal	Wilks λ DS1	Merkmal	Wilks λ DS2
ZM1	0.023	ZM1	0.103
ZM2	0.027	ZM1	0.092
ZM3	0.019	ZM1	0.083
ZM4	0.016	ZM1	0.088
ZM5	0.014	ZM1	0.078

Tabelle 8.35: Diskriminanzwerte zu Zernike-Momenten

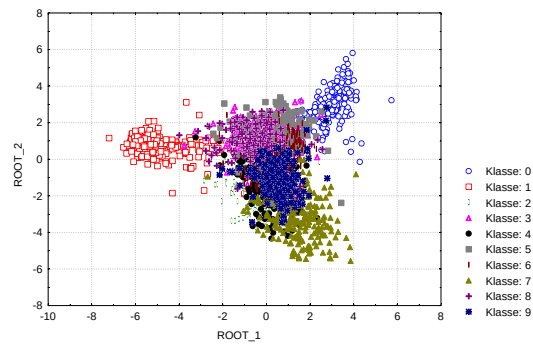
Klassifikationsergebnisse

Für das Testset DS1 ergibt sich die höchste Erkennungsrate von fast 94% bei Verwendung der Kreuzkorrelation mit Normierung 1 und 27 Merkmalen. Die höchsten Erkennungsrate pro verwendetem Klassifikator für beide Datensätze sind in den Tabellen 8.36 und 8.37 angegeben. Dabei ist die Erkennungsrate für den zweiten Datensatz mit 82.60% ebenfalls noch als hoch anzusehen.

Validität der korrekten Klassifikationen für k NN bei Zernike-Momenten Der Grad der Gültigkeit einer Klassifikation wurde unter Verwendung des k NN-Klassifikators für den Fall der korrekten Klassifikation untersucht und es ergaben sich die in Tabelle 8.38 gegebenen Werte. Man erkennt für den ersten Datensatz eine relativ hohe Validität

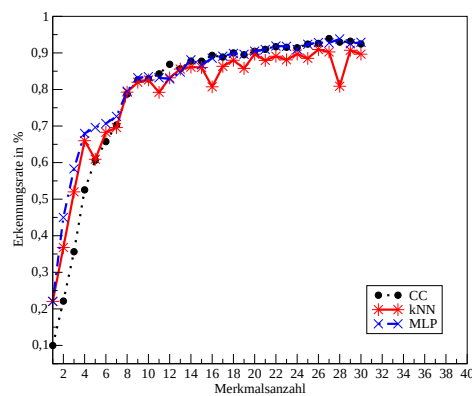


(a) DS1

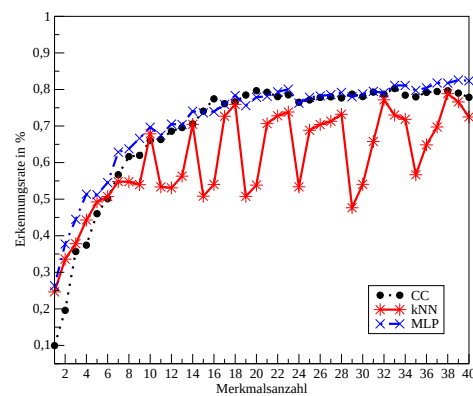


(b) DS2

Abbildung 8.16: Kanonische Analyse der 40 diskriminantesten Zernike-Momente für DS1 und DS2



(a) DS1



(b) DS2

Abbildung 8.17: Klassifikationsrate mit Zernike-Momenten für DS1 und DS2

Klassifikator	Erkennungsrate
Kreuzkorr. mit Normierung 1	93.93% (27 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, $norm=2$	90.94% (26 Merkmale)
MLP	93.83% (28 Merkmale)

Tabelle 8.36: Klassifikationsergebnisse für DS1 mit Zernike-Momenten

Klassifikator	Erkennungsrate
Kreuzkorr. mit Normierung 1	80.26% (33 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, $norm=2$	78.86% (38 Merkmale)
MLP	82.60% (39 Merkmale)

Tabelle 8.37: Klassifikationsergebnisse für DS2 mit Zernike-Momenten

unter den korrekten Klassifikationen, da im Mittel über 90% der analysierten nächsten Nachbarn (14/15) die selbe Klassifikation durchgeführt haben. Auch die minimale Validität ist mit einem Viertel der nächsten Nachbarn noch relativ hoch. Somit können die Klassifikationsentscheidungen für korrekte Entscheidungen auch hier als sehr sicher angesehen werden. Im Falle des zweiten, gestörten Datensatzes nehmen sowohl die minimale als auch die mittlere Validität deutlich ab und für die mittlere Validität sinkt die Güte der Klassifikation um über 10%.

Über den besten Sets aus der Klassifikation mit Zernike-Momenten wurden die maximal zu berechnenden Polynomordnungen bestimmt und deren Ergebnisse in Tabelle 8.39 notiert. Dabei wird deutlich, daß man bei den Zernike-Momenten mit Polynomordnungen ≤ 15 auskommt. Dies ist etwas mehr als bei den OFM-Momenten, so daß sich die Aussagen von Sheng et al. [58] zu bestätigen scheinen.

Datensatz	Prüfgröße	Wert
DS1 (Setgröße= 26)	MIN	26.67%
	MAX	100%
	AVG	90.14%
DS2 (Setgröße= 38)	MIN	20%
	MAX	100%
	AVG	78.76%

Tabelle 8.38: Validität der k NN-Klassifikation bei Nutzung von Zernike-Momenten

Datensatz	maximale Ordnung $p + q$	$\max p$	$\max q$
DS1	16	10	7
DS2	22	15	9

Tabelle 8.39: Maximale Ordnung der Parameter p und q für Zernike-Momente

8.3.5 Untersuchungen zu Wavelet-Momenten

Allgemeine statistische Daten

Im Anhang auf Seite 113 ist wieder die deskriptive Statistik für DS1 gegeben. Für die nachfolgenden Analysen wurden die Wavelet-Momente mit

$$m = 0, \dots, 3 \quad n = 0, 1, 2 \quad q = 0, \dots, 10$$

berechnet und mittels SFS die 40 diskriminantesten Merkmale ermittelt.

Merkmalsselektion und Diskriminanzanalyse

Die in Abbildung 8.18(a) durchgeführte kanonische Analyse zeigt bereits eine gute Trennung der 10 Klassen und auch die Wilks-Lambda-Werte, die in Tabelle 8.40 dargestellt wurden, sind relativ niedrig. Für das ebenfalls untersuchte zweite Datenset DS2

Merkmal	Wilks λ DS1	Merkmal	Wilks λ DS2
WM1	0.034	WM1	0.102
WM2	0.021	WM1	0.106
WM3	0.025	WM1	0.104
WM4	0.022	WM1	0.106
WM5	0.020	WM1	0.097

Tabelle 8.40: Diskriminanzwerte zu Wavelet-Momenten

verschlechtern sich die Wilks-Lambda-Werte deutlich und auch die Trennungsfähigkeit nimmt sichtbar ab.

Auch für die Wavelet-Momente wurden exemplarisch einige Werte berechnet und in Tabelle 8.41 angegeben. Dabei wird das Vorhandensein der drei Invarianzeigenschaften hinsichtlich Translation, Skalierung und Rotation deutlich.

Klassifikationsergebnisse

Die Klassifikationen wurden mit 40 diskriminanten Sets der Größe $\{1 \dots 40\}$ für beide Datensätze unter Verwendung von Kreuzkorrelation, k NN-Analyse und MLP durchgeführt.

Für das Testset DS1 ergibt sich die höchste Erkennungsrate bei Verwendung des MLP mit Normierung 1 und 22 Merkmalen. Die höchsten Erkennungsraten pro Klassifikator sind dabei in Tabelle 8.42 gegeben. Dabei sind die genauen Kurvenverläufe

Grad der Transformation	Parameter	Translation	Skaliert	Rotiert
0	(0, 0, 1)	0.0035	0.0035	0.0035
1	(0, 0, 1)	0.0035	0.0022	0.0024
2	(0, 0, 1)	0.0035	0.0001	0.0003
0	(1, 0, 0)	0.0036	0.0036	0.0036
1	(1, 0, 0)	0.0036	0.0019	0.0029
2	(1, 0, 0)	0.0036	0.0028	0.0028
0	(2, 0, 0)	0.0021	0.0021	0.0021
1	(2, 0, 0)	0.0021	0.0008	0.0015
2	(2, 0, 0)	0.0021	0.0012	0.0014

Tabelle 8.41: Einfluß der Translation, Skalierung oder Rotation auf Wavelet-Momente

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1 (36 - Merkmale)	92.63%
k NN mit $k = 15$, $eps = 2.0$, Normierung 1 (34 - Merkmale)	89.49%
MLP mit Normierung 1 (22 - Merkmale)	94.00%

Tabelle 8.42: Klassifikationsergebnisse für DS1 mit Wavelet-Momenten

in Abbildung 8.19(a) zu sehen. Es wird deutlich, daß bereits mit der geringen Menge von 8 Merkmalen eine Erkennungsrate von 80% erreicht werden kann und danach nur noch eine langsame Steigerung der Korrektheit erfolgt.

Für das zweite Datenset DS2 fällt die Erkennungsrate auf etwa 86% ab. Die genauen Werte sind wieder in der folgenden Tabelle 8.43 und in Abbildung 8.19(b) dargestellt.

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1	79.02% (40 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, Normierung 1	73.97% (37 Merkmale)
MLP mit Normierung 1	85.57% (37 Merkmale)

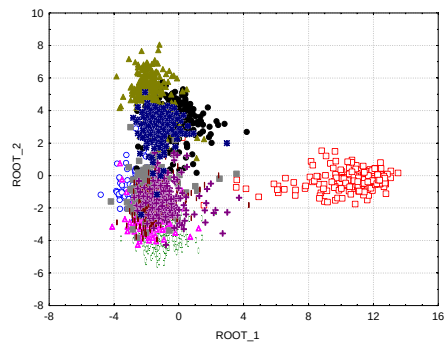
Tabelle 8.43: Klassifikationsergebnisse für DS2 mit Wavelet-Momenten

Validität der korrekten Klassifikationen für k NN bei Wavelet-Momenten Für den Fall des k NN-Klassifikators wurde wieder der Grad der Gültigkeit entsprechend des Konzepts aus Abschnitt 7.4.1 untersucht. Es ergaben sich für den Fall der korrekten Klassifikation (maximale Erkennungsrate) die folgenden, in Tabelle 8.43 gezeigten Werte. Man erhält für den ersten Datensatz eine relativ hohe Validität unter den korrekten Klassifikationen, da im Mittel fast 90% der analysierten nächsten Nachbarn (13/15) die selbe Klassifikation durchgeführt haben. Auch die minimale Validität ist mit einem Viertel der nächsten Nachbarn noch relativ hoch. Somit können die Klassifikationsentscheidungen für korrekte Entscheidungen auch hier als sicher angesehen werden. Im

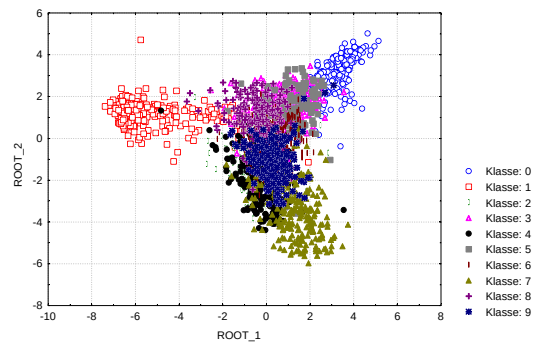
Datensatz	Prüfgröße	Wert
DS1 (Setgröße= 34)	MIN	26.67%
	MAX	100%
	AVG	87.52%
DS2 (Setgröße= 37)	MIN	20%
	MAX	100%
	AVG	72.90%

Tabelle 8.44: Validität der k NN-Klassifikation bei Nutzung von Wavelet-Momenten

Fälle des zweiten, gestörten Datensatzes nehmen sowohl die minimale als auch die mittlere Validität ab und für die mittlere Validität sinkt die Güte der Klassifikation um 15%.

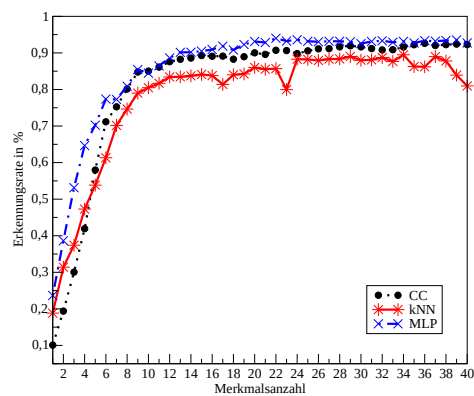


(a) DS1

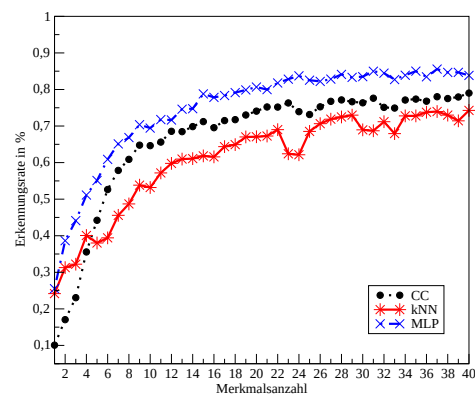


(b) DS2

Abbildung 8.18: Kanonische Analyse der 40 diskriminantesten Wavelet-Momente für DS1 und DS2



(a) DS1



(b) DS2

Abbildung 8.19: Klassifikationsrate mit Wavelet-Momenten für DS1 und DS2

8.4 Klassifikation mit kombinierten Momenten

Ähnlich wie bei Chim et al. [71] wurde auch hier zusätzlich zur Klassifikation mit einzelnen Merkmalsarten auch die Klassifikation mit kombinierten Merkmalen durchgeführt.

Dabei wurden die besten Merkmalskombinationen aus den vorigen Abschnitten kombiniert und darüber mittels SFS diskriminante Sets der Größe 1 bis 40 für Datensets DS1 und DS2 ermittelt. Im folgenden sind wieder die Plots zur kanonischen Analyse (Abbildung 8.20) und die Wilks-Lambda Werte in Tabelle 8.45 dargestellt.

Merkmal	Wilks λ DS1	Merkmal	Wilks λ DS2
TM1	0.014	TM1	0.071
TM2	0.009	TM1	0.073
TM3	0.008	TM1	0.063
TM4	0.005	TM1	0.066
TM5	0.004	TM1	0.058

Tabelle 8.45: Diskrimanzwerte zu kombinierten Momenten

Klassifikationsergebnisse

Die Klassifikation erfolgt wieder mittels Kreuzkorrelation, k NN-Analyse und unter Verwendung des MLP.

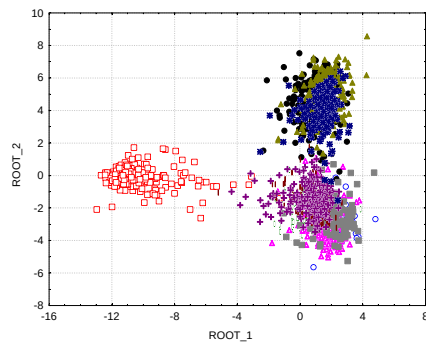
Für das Testset DS1 ergibt sich die höchste Erkennungsrate mit 96.51% (MLP) bei Verwendung von 36 Merkmalen. Die Ergebnisse differieren dabei je nach verwendetem Klassifikator und sind in Tabelle 8.46 gegeben. Dabei sind die genauen Kurven-

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1	94.82% (35 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, Normierung 1	92.87% (40 Merkmale)
MLP	96.51% (36 Merkmale)

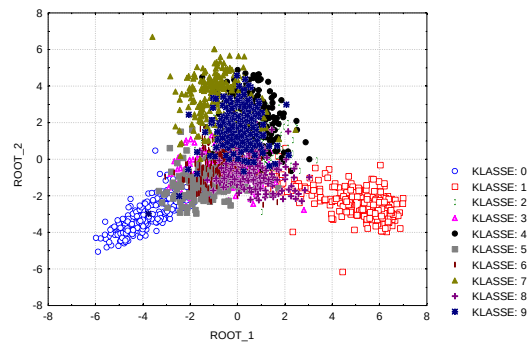
Tabelle 8.46: Klassifikationsergebnisse für DS1 bei Verwendung kombinierter Momente

verläufe in Abbildung 8.21(a) zu sehen.

Für das zweite Datenset DS2 verschlechtert sich die Klassifikationsrate und die Erkennungsrate fällt auf 86.3% für das MLP mit Normierung 1 ab. Insgesamt ist der Einfluß der Transformationen des DS2 auf den Klassifikationsprozeß aber relativ gering. Die genauen Werte sind in der folgenden Tabelle 8.47 und in Abbildung 8.21(b) dargestellt.

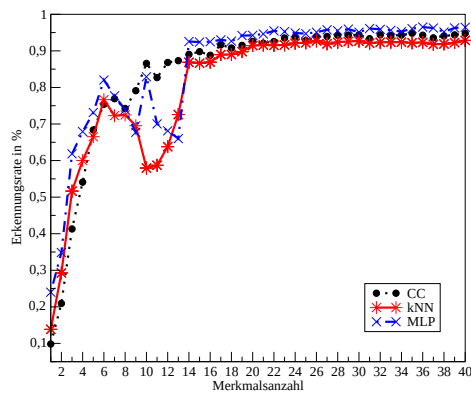


(a) DS1

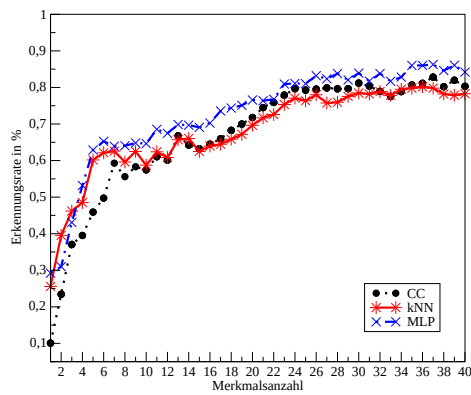


(b) DS2

Abbildung 8.20: Kanonische Analyse der 40 diskriminantesten kombinierten moment-basierten Merkmale für DS1 und DS2



(a) DS1



(b) DS2

Abbildung 8.21: Klassifikationsrate bei Verwendung von kombinierten Merkmalen für DS1 und DS2

Klassifikator	Erkennungsrate
Kreuzkorrelation mit Normierung 1	82.82% (37 Merkmale)
k NN mit $k = 15$, $eps = 2.0$, Normierung 1	80.17% (36 Merkmale)
MLP	86.28% (37 Merkmale)

Tabelle 8.47: Klassifikationsergebnisse für DS2 bei Verwendung kombinierter Momente

Validität der korrekten Klassifikationen für k NN bei kombinierten Momenten

Die Gültigkeit der Klassifikation wurde mit den Werten des k NN-Klassifikators ermittelt und ist in Tabelle 8.48 dargestellt. Es ergibt sich für den ersten Datensatz die

Datensatz	Prüfgröße	Wert
DS1	MIN	26.67%
	MAX	100%
	AVG	91.66%
DS2	MIN	26.67%
	MAX	100%
	AVG	79.63%

Tabelle 8.48: Validität der k NN-Klassifikation bei kombinierten Momenten

höchste bisher festgestellte Validität unter den korrekten Klassifikationen, da im Mittel deutlich über 90% der analysierten nächsten Nachbarn (14/15) die selbe Klassifikation durchgeführt haben. Auch die minimale Validität ist mit einem Viertel der nächsten Nachbarn noch relativ hoch. Somit können die Klassifikationsentscheidungen für korrekte Entscheidungen hier als sehr sicher angesehen werden. Im Falle des zweiten, gestörten Datensatzes nahm lediglich die mittlere Validität um etwa 10% ab, ist damit aber noch immer als sehr hoch anzusehen.

Wie bereits aus den Klassifikationsgraphen ersichtlich, sind die Erkennungsraten nur geringfügig höher als bei ausschließlicher Verwendung von Legendre-Momenten. Insgesamt ergab sich bei Verwendung kombinierter Merkmale keine deutliche Verbesserung der Klassifikationsrate. Allerdings wurde eine etwas bessere Klassifikationsleistung als bei den Legendre-Momenten erzielt, ohne daß in dem kombinierten Datensatz sehr viele Legendre-Momente berechnet werden mußten. Somit konnte durch Nutzung einfacher zu berechnender Momententypen sogar eine höhere Klassifikationsrate erzielt werden.

8.5 Zusammenfassung

In den vorigen Abschnitten wurden 7 momentgenerierende Verfahren analysiert und die erzeugten Merkmale im Klassifikationsprozeß untersucht. Dabei wurde deutlich, daß die orthogonalen Momente gegenüber den geometrischen Momenten (TM, NM) zu favorisieren sind, obwohl alle Verfahren diskriminante Merkmale mit hoher Zwischenklassenvarianz und niedriger Innerklassenvarianz erzeugen.

Da in den Analysen das Hauptaugenmerk auf dem Vergleich der Merkmalsarten und nicht primär auf der Klassifikation lag, kann angenommen werden, daß im einzelnen noch leicht verbesserte Klassifikationsergebnisse durch Variation der Parameter (Normierung, Gewichtung, . . .) oder Verwendung anderer Klassifikatoren möglich sind.

Im allgemeinen und für den Vergleich im besonderen wurden aber durch entsprechende Voranalysen Parametrisierungen für die Klassifikatoren gewählt, die einen gültigen Vergleich erlauben.

8.5.1 Normalisierte vs. Tsirikolias-Mertzios-Momente

Klassifikationsrate Die geometrischen Momente von Tsirikolias-Mertzios und die normalisierten geometrischen Momente zeigen in etwa gleiche Ergebnisse, sind aber mit etwa 80% Erkennungsrate beim ungestörten Datensatz 2 und ca. 60% Erkennungsrate für den Datensatz DS2 deutlich schlechter als die orthogonalen Momente. Für die normalisierten Momente läßt sich anhand der Analysen ein deutlich stabileres Klassifikationsverhalten bei Variation der Merkmalsanzahl feststellen. Allerdings waren die normalisierten Momente weit anfälliger gegenüber den Störungen im zweiten Datensatz.

Berechnungsaufwand Beide Momenttypen basieren auf relativ einfachen Berechnungen und können z.B. durch das Verfahren von Mertzios in Abschnitt 6.3 effizient und schnell berechnet werden.

8.5.2 Orthogonale Momente

Es wurden verschiedene Arten von orthogonalen Momenten in die Analysen einbezogen. Jedes dieser Verfahren ist zur Erzeugung translations-, rotations- und skalierungs-invarianter, diskriminanter Merkmale geeignet und über orthogonalen Polynomen definiert.

Klassifikationsrate Bei der Verwendung von Legendre-Momenten konnte die höchste Erkennungsrate (DS1 bei der Verwendung von maximal 40 Merkmalen) mit 96% erzielt werden, dicht gefolgt von den Zernike-Momenten mit etwa 94%. Dabei sind die Legendre-Momente und auch die Zernike-Momente sowohl in der Erkennungsrate als

auch bei Variation der Anzahl der verwendeten diskriminanten Merkmale als nahezu äquivalent zu betrachten und zeigen in der Klassifikation ein sehr stabiles Verhalten.

Die orthogonalen Fourier-Mellin-Momente (OFM) weisen ebenfalls ein relativ stabiles Klassifikationsverhalten auf und sind mit einer Erkennungsrate von 93% ebenfalls recht gut. Für die Pseudo-Zernike-Momente läßt sich feststellen, daß die Erkennungsrate noch einmal etwas schlechter als bei der Verwendung von OFM-Momenten sind, so daß die freiere Parametrisierung der Pseudo-Zernike-Momente im Vergleich zu den Zernike-Momenten keine Vorteile erbrachte.

Die Wavelet-Momente sind das neueste hier betrachtete Verfahren und ermöglichen auch die Erfassung von sehr lokalen Merkmalen. Nach entsprechender Merkmalsselektion, die die irrelevanten Merkmale entfernt, sind die verbleibenden Merkmale sehr diskriminant und führen zu fast, ebenso guten Erkennungsraten wie bei der Verwendung von Legendre-Momenten. Die vielfältige Parametrisierbarkeit gestattet es zugleich, die Merkmalsselektion in einem stärkeren Maße an die Problemstellung anzupassen.

Berechnungsaufwand Für alle orthogonalen Verfahren existieren effiziente Algorithmen vgl. Abschnitt 6.2, die eine Berechnung z.B. mittels Greenstheorem ermöglichen⁵. Auch ist es oft schon ausreichend, einfach die verwendeten Polynome bis zu einer bestimmten Ordnung vorzuberechnen um damit die zeitliche Komplexität zu reduzieren. Die Legendre-Momente sind als rekursives Verfahren relativ aufwendig, können aber ebenfalls durch Vorberechnung der Polynome in etwa ähnlich schnell wie z.B. die Zernike-Momente ermittelt werden, so daß dies kein Hinderungsgrund ist. Für die Legendre-Momente wurden von Shu et al. sowie Zhou [59, 60, 73] verschiedene neue effiziente Berechnungsmethoden vorgeschlagen.

8.5.3 Wie viele Merkmale sind nötig ?

Wie an den Grafiken zur Erkennungsrate ersichtlich ist, genügen meist die besten 20 Merkmale, um sowohl im DS1, als auch für den Datensatz DS2 das „Optimum“ bzgl. der Erkennungsrate des betrachteten Verfahrens zu erzielen. Lediglich bei den geometrischen Momenten genügen schon unter 10 Merkmale und die Legendre und Wavelet-Momente benötigen etwa 15 Merkmale.

Allgemein zeigen alle Kurven meist schon recht früh, ab einer bestimmten Merkmalsanzahl ein stagnatives bis sogar abfallendes Verhalten in der Erkennungsrate. Die Berechnung von Merkmalen und damit Polynomen höherer Ordnung kann deshalb oft entfallen, da die Erkennungsrate durch weitere Hinzunahme von Merkmalen nur noch sehr gering beeinflusbar ist. Dies führt ebenfalls dazu, daß die Berechnungen relativ schnell erfolgen können.

⁵Durch Umschreiben der Verfahren mit Bezug zu geometrischen Momenten.

8.5.4 Vergleichbarkeit mit anderen Arbeiten

Zur Erkennung von Schriftzeichen wurde bereits eine Vielzahl von Untersuchungen durchgeführt. Dabei treten bzgl. einer Vergleichbarkeit stets die selben Probleme auf. Zum einen sind die verwendeten Datensätze oft unterschiedlich und zum Teil nicht frei zugänglich, zum anderen sind durch die vielen in der OCR auftretenden Zwischenschritte (vgl. Abschnitt 1) vielfältigste Parametrisierungsunterschiede ein eben solches Problem.

Im folgenden sind die dargestellten Vergleiche deshalb nur bedingt als gültig anzusehen und können letztlich nur als Hinweis verstanden werden, ohne Anspruch auf Vollständigkeit. Neben den hier favorisierten statistischen Momenten sind, wie bereits erwähnt, auch andere Merkmale für die Schriftzeichenerkennung möglich, so wurden in einer Arbeit von Cai [9] Fourierdeskriptoren eingesetzt und diese in Kombination mit Hidden Markov Modellen zur Erkennung numerischer Zeichen genutzt. Eine andere Möglichkeit besteht wie im Abschnitt 3 angegeben darin, ein neuronales Netz nicht nur als Klassifikator, sondern auch als Merkmalsextraktor einzusetzen. Ein entsprechender Ansatz wurde von Ping et al. [49] verfolgt. Ein Ansatz unter Nutzung eines modularen Netzes wurde durch Oh et al. in [63] gegeben. Auch zu momentbasierter OCR gibt es, meist bereits in den entsprechenden Abschnitten erwähnte Arbeiten, bei denen unter anderem Zernike-Momente, Orthogonal-Fourier-Mellin Momente [32, 6, 33], Tsirikolias-Mertzios-Momente [68, 71] und Wavelet-Momente [57] genutzt wurden.

Neben der Nutzung von numerischen Merkmalen können auch symbolische Merkmale erfaßt und zur Klassifikation herangezogen werden. Ein entsprechendes System wurde von Ahmed et al. in [2] vorgestellt. Von Kim et al. [34] wurde ein Ansatz unter Nutzung der PCA-Methodik vorgestellt.

In Tabelle 8.5.4 werden die Ergebnisse anderer Autoren unter Angaben der verwendeten Datenbasis und der verwendeten Verfahren zum Vergleich mit denen in dieser Arbeit erhaltenen Ergebnissen darstellt.

Wie deutlich zu sehen ist, sind die Parametrisierungen und Datenbasen in den Arbeiten stets sehr unterschiedlich, was einen Vergleich über mehrere Arbeiten hinweg nahezu unmöglich macht. Allgemein läßt sich aber auch hier ableiten, daß größere Datensets und damit auch eine größere Trainingsdatenmenge zu besseren Ergebnissen führen. Die künstlich generierten Datensets mit Druckbuchstaben schneiden selbst bei älteren Arbeiten relativ gut ab, was einen Idealisierungseffekt vermuten läßt, der so bei handschriftlichen Daten meist nicht zu erkennen ist. Nur in wenigen Arbeiten wurden Datensätze im normalen und gestörten Zustand verglichen, wenn dies allerdings durchgeführt wurde (vgl. [33]), zeigten sich deutlich schlechtere Ergebnisse und es konnten meist nur Erkennungsraten $\leq 85\%$ erzielt werden.

Bezogen auf die in Tabelle 8.5.4 dargestellten Resultate, sind die, in dieser Arbeit ermittelten Klassifikationsraten im oberen Drittel angesiedelt. Dabei weisen die hier

Autoren	Datenbasis	Verfahren	Erkennungsrate
Cai et al. [9]	CEDAR CDROM 1 ca. 20000 numerische Symbole	Fourierdeskriptoren + HMM	93.3 – 96.7%
Ping et al. [49]	NIST Special DB3 5000 numerische Symbole) + eigene DB ca. 10000 numerische Symbole	Merkmale per NN extrahiert + andere	95.50 – 98.10%
Oh et al. [63]	CENPARMI ca. 6000 Symbole	Distanzbasierte Merkmale + modulares MLP	ca. 97.5%
Ahmed et al. [2]	CEDAR 5726 numerische Symbole	symbolische Merkmale	ca. 79.70%
Kim et al. [34]	UCI 5620 numerische Symbole	PCA	98.55
Shen et al. [57]	künstlicher Datensatz, 26 Großbuchstaben 30 pro Klasse	Wavelet-Momente + minimum distance classifier	$\approx 100\%$
Kan et al. [32]	NIST ca. 120000 numerische Symbole	Zernike-Momente, OFM-Momente + MLP	$\leq 82.1\%$
Belkasim et al. [6]	eigene Erhebung mit 300 numerischen Symbolen	Hu, Zernike und Pseudo-Zernike-Momente + kNN	ca. 90%
Khotanzad et al. [33]	eigene Erhebung 314 alphabetischer Druckbuchstaben	Zernike-Momente + NN, MMD	90 – 99%
Tsirikolias et al. [68]	eigene Erhebung 2496 alphabetischer Druckbuchstaben	Tsirikolias-Mertzios-Momente + Discrimination Cost vgl. Seite 57	94%
Chim et al. [71]	eigene Erhebung 2160 alphabetischer Druckbuchstaben	TM-Momente, Affine-Momente von Hu, Reiss + Distanzmessungen	96 – 98%

Tabelle 8.49: Auflistung der Ergebnisse anderer Arbeiten

verwendeten handschriftlichen Daten aus der NIST-Datenbank [45] eine hohe Schreibstilvarianz auf, so daß die Ergebnisse als relativ allgemeingültig angesehen werden können, wenn gleich die Größe des genutzten Datensatzes für die hier betrachtete Problemstellung des Vergleichs verschiedener momentbasierter Verfahren zur Schriftzeicherkennung lediglich als ausreichend anzusehen ist.

Generell muß, wie auch bei den Arbeiten anderer Autoren ersichtlich, davon ausgegangen werden, daß bei der Verwendung anderer Datensätze Varianzen in den hier ermittelten Ergebnissen auftreten⁶. Dies ist allerdings für den Vergleich der Eignung der momentbasierten Verfahren und deren unterschiedliche Diskriminanzfähigkeit eher nicht zu erwarten, so daß diese Aussagen auch bei anderen Datensätzen Bestand haben. Die Ergebnisse bezüglich der maximal aufgetretenen Merkmalsordnungen sind ebenfalls im allgemeinen als gültig anzusehen, da auch schon in anderen Artikeln z.B. bei Trier [67] gewisse Obergrenzen angedeutet werden.

Für die statistischen Momente im speziellen läßt sich aus den Ergebnissen dieser Arbeit und im Vergleich mit den Ergebnissen in Tabelle 8.5.4, die keine statistischen Momente nutzen, keine endgültige Aussage ableiten. Verallgemeinert sind die statistischen Momente in ihrer Anwendung für die OCR aber sehr verbreitet und zeigen mindestens genauso gute Ergebnisse wie bei der Verwendung anders artiger Merkmale. Allerdings fehlen bisher ausführliche Vergleiche verschiedener Merkmalsarten (die nicht nur momentbasiert sind) unter Nutzung verschiedenster Klassifikatoren, so daß sich diesbezüglich keine generelle Aussage ableiten läßt. Zumindest für die statistischen Momente wurde mit dieser Arbeit der, bisher umfassendste Vergleich zur Anwendung statistischer Momente in der OCR durchgeführt.

8.5.5 Aufwand allgemein - Empfehlungen

Wenn man den Aufwand für die Generierung der Prototypdaten (Lerndaten) mit dem eigentlichen Klassifikationsprozeß vergleicht, wird deutlich spürbar, daß vor allem die Klassifikation Zeit in Anspruch nimmt. Dies wird besonders bei der Kreuzkorrelation deutlich, die, wenn gleich mit sehr guten Klassifikationsraten glänzend, erheblich länger dauert, als z.B. bei der Verwendung des k NN-Klassifikators. Dies ist vor allem auf den vollständigen Vergleich des zu klassifizierenden Vektors mit allen Trainingsvektoren zurückzuführen. Da im allgemeinen die Klassifikation mit zunehmender Trainingsmenge besser wird, ergeben sich ab einer bestimmten Trainingssetgröße auch Probleme hinsichtlich des Speicheraufwandes, so daß auch unter diesem Gesichtspunkt Klassifikationsverfahren wie das MLP zu favorisieren sind.

Die Dauer der Klassifikation kann somit zum einen durch Beschränkung der Berechnung auf eine geringe Anzahl von Merkmalen reduziert werden und zum anderen durch Nutzung eines Klassifikators, bei dem konstante Berechnungskosten einmalig im Vorfeld der Klassifikation realisierbar sind. Somit ist ein neuronales Netz sowohl bezüglich der Geschwindigkeit, als auch mit Hinblick auf eine sehr gute Erkennungs-

⁶insbesondere mit Blick auf die Erkennungsrate

rate und Speicherkomplexität zu empfehlen, da durch Modellierung einer Gewichtsmatrix im Vorfeld der Klassifikation bereits ein erheblicher Teil des Klassifikationsaufwandes umgesetzt werden kann.

Erfolgt die Berechnung der Momente zudem mittels Greenstheorem oder bei der Verwendung vorausberechneter Polynome, sinkt der Berechnungsaufwand deutlich.

Kapitel 9

Zusammenfassung

In dieser Arbeit wurden verschiedene momentbasierte Verfahren zur Gewinnung von diskriminanten Merkmalen zur Schriftzeichenerkennung untersucht. Es wurde gezeigt, daß diese Form von Merkmalen zur Verwendung im Bereich der OCR geeignet ist und auch bei handschriftlichen Daten erfolgreich eingesetzt werden kann. Dabei wurden mit diesem Vergleich erstmalig verschiedenste, auch neuere momentbasierte Verfahren, über einer gleichartigen, größeren handschriftlichen Datenbasis unter Nutzung verschiedener Klassifikatoren, mit Bezug auf ihre Eignung im OCR-Bereich, analysiert. Die Analysen wurden dabei nicht nur über ungestörten handschriftlichen Daten durchgeführt, sondern zusätzlich über einem gestörten Datensatz, um die Invarianzeigenschaften zu prüfen.

Basierend auf den durchgeführten Analysen wurden Empfehlungen für die Art der zu verwendenden Verfahren und zur Anzahl und Ordnung der dabei zu ermittelten Merkmale gegeben.

Im ersten Teil der Arbeit wurde die Theorie zu den verschiedenen Arten statistischer Momente angegeben und Methoden zur effizienten Berechnung statistischer Momente zusammengetragen und implementiert. Dabei wurden die verschiedenen Verfahren kombiniert und ein System zur flexiblen Berechnung von Momenten entwickelt.

Die Analyse der verschiedenen Momenttypen und deren Nutzung im Bereich der OCR, wurde im zweiten Teil der Arbeit durch Nutzung verschiedener Klassifikatoren untersucht. Dabei wurden die Analysen über einem gestörten und ungestörten Datensatz durchgeführt und es konnten Erkennungsraten von über 96 % (DS1) bzw. über 86 % (DS2) erzielt werden. Die verschiedenen Momenttypen wurden verglichen und es konnte gezeigt werden, daß die Legendre¹- und Wavelet-Momente die beste Eignung für dieses Anwendungsfeld aufweisen.

Abschließend wurde die Verwendung von kombinierten Momenten untersucht, die Ergebnisse zeigen, daß bei Beschränkung auf maximal 40 Merkmale keine signifikante Steigerung der Erkennungsrate im Vergleich zu den Legendre-Momenten auftritt. Insgesamt läßt sich feststellen, daß alle hier untersuchten Verfahren zur Ermittlung von

¹Insbesondere die Legendre-Momente wurden bis jetzt in der Klassifikation kaum eingesetzt, da die ursprüngliche, rekursive Definition sehr rechenaufwendig ist.

momentbasierten Merkmalen diskriminante Merkmale erzeugen, die zur Anwendung in der OCR geeignet sind. Dabei verbleiben jedoch Probleme bei sehr stark gestörten Symboldaten, die so, bis jetzt noch nicht vollständig gelöst sind.

Als ein besonderes Problem beim Vergleich mit anderen Arbeiten ergaben sich die sehr unterschiedlichen Datenbasen, die in den verschiedenen anderen Artikeln genutzt wurden und daß diese zum Teil nicht verfügbar oder nur kommerziell zugänglich sind. Weitere Analysen erscheinen unter Nutzung auch kommerzieller Datenbanken deshalb als sinnvoll.

Die momentbasierten Verfahren zeigen teilweise eine sehr gute Eignung im Bereich der OCR handschriftlicher Daten. Entsprechend der Analysen bzgl. des zweiten Datensatzes DS2 und der Analysen verschiedener Autoren (vgl. z.B. [33, 57]) existieren allerdings noch immer Defizite in der Stabilität der erfaßten Merkmale bzgl. affiner Transformationen oder Rauschen, die die Erkennungsrate negativ beeinflussen.

Untersuchungen unter Einbeziehung anderer Merkmale, möglicherweise auch struktureller Natur, und weiterer Klassifikatoren, wie modularer neuronaler Netze oder kombinierter Klassifikation, wären, wenn gleich auch sehr aufwendig, ebenfalls vielversprechend.

Allgemein kann aber vermutet werden, daß damit in erster Linie der meist unerwünschte Einfluß affiner Transformationen oder der Einfluß von Rauschen abgemildert werden kann. Eine Erkennungsrate von 100% kann bei Einzelsymbolerkennung im Realfall meist sowieso nicht erreicht werden, so daß auch der Einfluß des Kontextes und damit weiterer, auf der reinen Einzelsymbolklassifikation aufbauender Verfahren, berücksichtigt werden muß.

Somit ergeben sich weitere, interessante Möglichkeiten zur Fortsetzung OCRbezogener Analysen und Arbeiten. Wobei sich diese mehr in Richtung einer intelligenten Klassifikation unter starker Einbeziehung des Kontextes und komplexer Modellierungen, wie z.B. mit Hidden-Markov-Modellen, modularen neuronalen Netze oder kombinierten Klassifikationen, abzeichnen.

Anhang A

Deskriptive Statistik zu verschiedenen Momenttypen

Tabelle A.1: Deskriptive Statistik zu Momenten von Hu

Klasse	Merkmal	Mittelwert	Minimum	Maximum	σ^1
0	HUM1	0.379401	0.1696	0.754356	0.094877
	HUM2	0.027683	0.0001	0.215405	0.028550
	HUM3	0.606680	0.0002	7.918150	0.923554
	HUM4	0.178705	0.0003	4.400940	0.426639
	HUM5	-0.024339	-6.3759	4.288800	0.660127
	HUM6	0.013804	-0.2407	1.048850	0.078501
	HUM7	-0.126576	-25.9434	1.163830	1.654529
1	HUM1	0.55464	0.00635	0.991	0.1567
	HUM2	0.34509	0.02639	3.937	0.3373
	HUM3	1.14131	0.00643	42.616	3.9578
	HUM4	0.73099	0.00131	41.398	3.5957
	HUM5	13.85927	-0.02941	1738.790	142.0525
	HUM6	0.77408	-0.40359	82.130	6.5021
	HUM7	0.02628	-3.68236	11.856	0.8092
2	HUM1	0.413115	0.1799	0.75418	0.098206
	HUM2	0.044945	0.0002	0.35416	0.045582
	HUM3	7.068509	0.4797	49.46520	6.295657
	HUM4	0.562790	0.0017	3.91964	0.625107
	HUM5	0.959548	-10.9811	28.3329	3.782159
	HUM6	0.031781	-0.4018	1.68512	0.194300
	HUM7	-0.745983	-17.1759	9.75312	2.594867
3	HUM1	0.437742	0.2379	0.9929	0.10932
	HUM2	0.062634	0.0006	0.4551	0.05482
	HUM3	2.950950	0.0491	29.0649	3.17627
	HUM4	1.070652	0.0014	12.8603	1.31905
	HUM5	2.806497	-7.5247	248.0840	17.29514
	HUM6	0.090496	-0.4956	5.9414	0.55723
	HUM7	1.067503	-16.5254	34.7419	4.12721
4	HUM1	0.381851	0.1918	0.67136	0.079626
	HUM2	0.027398	0.0003	0.15273	0.028435
	HUM3	5.465908	0.0646	25.30660	4.368992
	HUM4	0.583440	0.0027	6.99576	0.937662
	HUM5	2.163513	-2.1018	82.72890	8.336224
	HUM6	0.108606	-0.0581	2.51187	0.304177
	HUM7	-0.601351	-29.7037	3.36319	2.658549
5	HUM1	0.47973	0.2583	0.868	0.1156
	HUM2	0.08949	0.0022	1.200	0.1238
	HUM3	5.17244	0.0020	70.447	8.6815

Tabelle A.1: Deskriptive Statistik zu Momenten von Hu

Klasse	Merkmal	Mittelwert	Minimum	Maximum	σ^1
	HUM4	1.45269	0.0014	39.269	3.6687
	HUM5	19.64738	-44.7022	2045.870	179.6174
	HUM6	0.47957	-2.4436	39.292	3.4714
	HUM7	4.68202	-14.7576	283.418	27.0138
6	HUM1	0.359860	0.17419	0.6284	0.08561
	HUM2	0.031984	0.00019	0.2915	0.03895
	HUM3	4.662499	0.24528	17.6677	3.25084
	HUM4	1.593221	0.02139	9.8311	1.67885
	HUM5	7.225873	-1.19511	126.8310	16.55927
	HUM6	0.334308	-0.05102	3.5697	0.60643
	HUM7	2.470551	-1.40569	27.0706	4.50216
7	HUM1	0.49199	0.2498	0.9840	0.12236
	HUM2	0.07741	0.0000	0.4296	0.06792
	HUM3	15.46143	1.9703	55.1260	9.65356
	HUM4	3.07221	0.1474	20.1333	2.59345
	HUM5	19.51337	-61.1114	571.4090	57.82644
	HUM6	0.72039	-1.2198	10.3124	1.23844
	HUM7	16.57833	-15.5362	126.8540	24.18510
8	HUM1	0.386806	0.21296	0.73275	0.093233
	HUM2	0.066783	0.00085	0.33637	0.057478
	HUM3	1.514968	0.00660	18.25680	2.251891
	HUM4	0.275416	0.00003	6.49568	0.712926
	HUM5	0.776172	-0.76839	40.79580	4.536015
	HUM6	0.047376	-0.16783	2.21364	0.242650
	HUM7	0.149604	-1.24457	19.36170	1.449197
9	HUM1	0.386620	0.211815	0.60919	0.087563
	HUM2	0.049373	0.000355	0.21242	0.043177
	HUM3	4.637298	0.374812	17.77990	3.163744
	HUM4	1.041653	0.002906	6.93842	1.179413
	HUM5	4.039207	-0.060745	58.63760	8.262906
	HUM6	0.270106	-0.006691	2.50376	0.411308
	HUM7	1.205877	-0.903387	13.55770	2.268547

Klasse	Merkmal	Minimum	Maximum	Mittelwert	σ
0	ZM(1, 1)	0	0,19347	$7,21E-03$	$1,71E-02$
	ZM(2, 0)	0,00001	2,70317	0,1563191	0,2557451
	ZM(2, 2)	0,00158	0,76839	0,2201414	0,1483114
	ZM(3, 1)	0,00028	1,3189	$8,66E-02$	0,132858
	ZM(3, 3)	0,00016	0,26951	$3,25E-02$	$4,09E-02$
1	ZM(1, 1)	0	0,09934	$3,79E-03$	$7,81E-03$
	ZM(2, 0)	0,26614	3,31882	1,0933278	0,3962652
	ZM(2, 2)	0,28557	1,20475	0,8451691	0,1513891
	ZM(3, 1)	0,0001	0,76728	$2,72E-02$	$6,63E-02$
	ZM(3, 3)	0,00002	0,24216	$7,33E-03$	$1,73E-02$
2	ZM(1, 1)	0,00001	0,13774	$1,37E-02$	$2,00E-02$
	ZM(2, 0)	0,00015	1,53309	0,2211546	0,2480385
	ZM(2, 2)	0,00428	1,00602	0,2900681	0,1879314
	ZM(3, 1)	0,00253	1,01246	0,249102	0,191906
	ZM(3, 3)	0,00041	0,86117	0,2538903	0,17213
3	ZM(1, 1)	0,00001	0,13971	$2,99E-02$	$3,09E-02$
	ZM(2, 0)	0	1,03684	0,1481816	0,1765849
	ZM(2, 2)	0,01302	1,07298	0,4637745	0,1906973
	ZM(3, 1)	0,00054	0,74609	0,2213743	0,1495638
	ZM(3, 3)	0,00009	0,42153	$6,13E-02$	$5,97E-02$
4	ZM(1, 1)	0	0,06789	$8,11E-03$	$1,02E-02$
	ZM(2, 0)	0,00383	2,20533	1,0378109	0,4812057
	ZM(2, 2)	0,00059	0,53569	0,1000123	$8,00E-02$
	ZM(3, 1)	0,00044	0,48192	$7,54E-02$	$7,58E-02$
	ZM(3, 3)	0,00031	0,82008	0,1596625	0,150375
5	ZM(1, 1)	0,00001	0,12633	$2,54E-02$	$2,70E-02$
	ZM(2, 0)	0,00012	1,36264	0,3899996	0,3727666
	ZM(2, 2)	0,02171	0,99365	0,3700465	0,1956782
	ZM(3, 1)	0,00191	0,76739	0,1538519	0,1383649
	ZM(3, 3)	0,00042	0,35616	$6,61E-02$	$7,28E-02$
6	ZM(1, 1)	0,00009	0,12709	$2,12E-02$	$1,63E-02$
	ZM(2, 0)	0,0017	2,10862	0,9421108	0,5071887
	ZM(2, 2)	0,0002	0,63536	0,1266846	0,1052023
	ZM(3, 1)	0,00006	0,82577	0,2426946	0,1382246
	ZM(3, 3)	0,00604	0,54317	0,1158689	$7,77E-02$
7	ZM(1, 1)	0,0022	0,1063	$3,82E-02$	$2,14E-02$
	ZM(2, 0)	0,00396	1,9811	0,5677892	0,369701
	ZM(2, 2)	0,00155	0,73253	0,2310001	0,1534799
	ZM(3, 1)	0,05256	1,05755	0,3084457	0,1539539
	ZM(3, 3)	0,01991	0,89442	0,2669415	0,173762
8	ZM(1, 1)	0,00002	0,11567	$8,06E-03$	$1,46E-02$
	ZM(2, 0)	0,00169	1,66295	0,3598767	0,2600057
	ZM(2, 2)	0,00036	1,09145	0,4535005	0,2080848
	ZM(3, 1)	0,00015	0,48934	$5,02E-02$	$5,62E-02$
	ZM(3, 3)	0,00002	0,44207	$5,81E-02$	$7,54E-02$
9	ZM(1, 1)	0,00068	0,07712	$1,61E-02$	$1,21E-02$
	ZM(2, 0)	0,00862	3,54853	1,1348922	0,483892
	ZM(2, 2)	0,00454	0,63771	0,1739505	0,1073102
	ZM(3, 1)	0,00073	0,49189	$8,90E-02$	$8,47E-02$
	ZM(3, 3)	0,00401	0,35704	$8,63E-02$	$5,72E-02$

Tabelle A.2: Deskriptive Statistik zu den Zernike - Momenten

Klasse	Merkmal	Minimum	Maximum	Mittelwert	σ
0	PZM(1, 0)	0,00048	6,99143	1,070927	0,8100839
	PZM(1, 1)	0	0,07695	$4,30E-03$	$8,25E-03$
	PZM(2, 0)	0	8,49827	0,3603781	0,6865031
	PZM(2, 1)	0,00001	0,81866	$5,40E-02$	$9,79E-02$
	PZM(2, 2)	0,00111	0,43255	0,15984	$9,12E-02$
1	PZM(1, 0)	2,2979	7,96235	4,1574061	0,7309203
	PZM(1, 1)	0	0,02731	$1,68E-03$	$2,90E-03$
	PZM(2, 0)	3,19831	18,1315	8,3811701	1,6913096
	PZM(2, 1)	0,00003	0,38748	$1,69E-02$	$3,73E-02$
	PZM(2, 2)	0,05003	0,50449	0,3028846	$7,40E-02$
2	PZM(1, 0)	0,33187	5,00944	1,9843879	0,7776978
	PZM(1, 1)	0,0001	0,07045	$6,58E-03$	$9,08E-03$
	PZM(2, 0)	0,00056	9,3578	2,0833058	1,3819387
	PZM(2, 1)	0,00133	0,59755	0,118814	0,1095426
	PZM(2, 2)	0,00102	0,46413	0,1408817	$8,93E-02$
3	PZM(1, 0)	0,45752	4,07336	1,7459711	0,6482297
	PZM(1, 1)	0	0,07194	$1,31E-02$	$1,54E-02$
	PZM(2, 0)	0,31879	6,98806	2,1352842	1,1008359
	PZM(2, 1)	0,00162	0,58357	0,1614822	0,1035504
	PZM(2, 2)	0,01324	0,63241	0,263407	0,1059741
4	PZM(1, 0)	1,12767	6,1754	4,0062524	1,0035712
	PZM(1, 1)	0,00001	0,02164	$3,44E-03$	$4,16E-03$
	PZM(2, 0)	0,33456	12,6044	4,7889497	2,4492867
	PZM(2, 1)	0,00139	0,26209	$4,80E-02$	$4,18E-02$
	PZM(2, 2)	0,0008	0,19563	$4,94E-02$	$3,18E-02$
5	PZM(1, 0)	0,47122	4,69727	2,4048643	1,0879864
	PZM(1, 1)	0,00004	0,05657	$1,37E-02$	$1,30E-02$
	PZM(2, 0)	0,06473	7,8568	2,5701441	1,7635946
	PZM(2, 1)	0,00052	0,39521	$9,26E-02$	$9,54E-02$
	PZM(2, 2)	0,00644	0,54941	0,2085248	0,1059223
6	PZM(1, 0)	1,08413	6,01284	3,7872716	1,1050033
	PZM(1, 1)	0,00002	0,04928	$5,42E-03$	$6,81E-03$
	PZM(2, 0)	0,00536	9,31744	3,0234084	2,1287094
	PZM(2, 1)	0,0131	0,47017	0,1737017	$8,04E-02$
	PZM(2, 2)	0,00035	8,11877	$8,59E-02$	0,5074232
7	PZM(1, 0)	0,8781	5,79613	2,9895616	0,8640133
	PZM(1, 1)	0,00006	0,05786	$1,42E-02$	$1,15E-02$
	PZM(2, 0)	0,00004	8,3074	1,885659	1,4495397
	PZM(2, 1)	0,03918	0,69773	0,2428438	0,1040939
	PZM(2, 2)	0,00042	0,26574	$8,62E-02$	$5,65E-02$
8	PZM(1, 0)	0,49133	5,24207	2,4572801	0,7311872
	PZM(1, 1)	0	0,08784	$4,71E-03$	$9,25E-03$
	PZM(2, 0)	0,13468	8,67037	3,3096592	1,4006561
	PZM(2, 1)	0,00005	0,36776	$3,62E-02$	$4,32E-02$
	PZM(2, 2)	0,0057	0,52078	0,2311042	$9,53E-02$
9	PZM(1, 0)	0,49302	8,31603	4,2000554	0,9968747
	PZM(1, 1)	0,00014	0,04554	$7,43E-03$	$7,31E-03$
	PZM(2, 0)	0,00167	16,5654	4,4877375	2,4251911
	PZM(2, 1)	0,00599	0,30878	0,1161068	$6,53E-02$
	PZM(2, 2)	0,00284	0,23308	$7,88E-02$	$4,19E-02$

Tabelle A.3: Deskriptive Statistik zu Pseudo-Zernike Momenten

Klasse	Merkmal	Minimum	Maximum	Mittelwert	σ
0	NM(0, 3)	-1,40656	0,55367	$2,29E-03$	0,1986973
	NM(0, 4)	0,0092035	0,307775	$5,70E-02$	$3,62E-02$
	NM(0, 5)	-2,42461	0,56061	$-4,49E-03$	0,204105
	NM(0, 6)	0,00155	0,30395	$2,41E-02$	$2,70E-02$
	NM(0, 7)	-3,47894	0,47294	$-9,89E-03$	0,2391005
1	NM(0, 3)	-1,20119	0,62848	$-1,97E-02$	0,1257579
	NM(0, 4)	0,000052	0,236192	$2,87E-02$	$3,53E-02$
	NM(0, 5)	-3,00673	0,7843	$-2,87E-02$	0,255648
	NM(0, 6)	0	9,70061	0,33337	1,3078788
	NM(0, 7)	-6,95955	0,97326	$-0,1337908$	0,8103723
2	NM(0, 3)	-1,17097	1,53866	-0,181035	0,3942838
	NM(0, 4)	0,0122947	0,271089	$6,20E-02$	$4,13E-02$
	NM(0, 5)	-1,24583	2,16967	-0,1063033	0,4407676
	NM(0, 6)	0,00255	0,28616	$3,46E-02$	$3,91E-02$
	NM(0, 7)	-1,29383	2,76637	$-3,82E-02$	0,4731376
3	NM(0, 3)	-2,09904	0,14459	-0,3933963	0,3340169
	NM(0, 4)	0,0057581	0,275655	$4,56E-02$	$3,52E-02$
	NM(0, 5)	-3,34762	0,12019	-0,3352277	0,4204042
	NM(0, 6)	0,00079	0,31672	$2,50E-02$	$3,47E-02$
	NM(0, 7)	-4,7825	0,08143	-0,2814564	0,5090744
4	NM(0, 3)	-1,28843	1,06499	$-7,33E-02$	0,3599888
	NM(0, 4)	0,0083436	0,200421	$6,17E-02$	$3,56E-02$
	NM(0, 5)	-1,3407	1,47057	$-3,34E-02$	0,3107742
	NM(0, 6)	0,00135	0,1655	$3,02E-02$	$2,72E-02$
	NM(0, 7)	-1,24803	1,82118	$-4,77E-03$	0,2702326
5	NM(0, 3)	-0,85383	6,31953	0,4210217	0,7860121
	NM(0, 4)	0,0108607	2,17924	0,124255	0,204411
	NM(0, 5)	-1,10137	25,6639	0,7584164	2,430428
	NM(0, 6)	0,00224	6,35576	0,144771	0,5660547
	NM(0, 7)	-1,27514	94,8311	1,5540177	8,4623765
6	NM(0, 3)	-0,22232	0,6672	0,1343711	0,1531116
	NM(0, 4)	0,0051999	0,319598	$3,25E-02$	$2,59E-02$
	NM(0, 5)	-0,14879	0,99874	$8,97E-02$	0,1239072
	NM(0, 6)	0,00067	0,30828	$1,23E-02$	$2,13E-02$
	NM(0, 7)	-0,08349	1,50003	$5,61E-02$	0,1186411
7	NM(0, 3)	-1,35876	0,40018	-0,2543673	0,3378295
	NM(0, 4)	0,0078253	0,233207	$5,06E-02$	$3,39E-02$
	NM(0, 5)	-1,50814	0,6107	-0,217996	0,3331395
	NM(0, 6)	0,00124	0,20035	$2,54E-02$	$2,77E-02$
	NM(0, 7)	-1,87439	0,78731	-0,1765235	0,3277251
8	NM(0, 3)	-0,21877	1,44479	0,1525066	0,2571214
	NM(0, 4)	0,0027812	0,19494	$3,29E-02$	$2,70E-02$
	NM(0, 5)	-0,15749	1,5687	0,1211531	0,2439919
	NM(0, 6)	0,00023	0,18251	$1,43E-02$	$2,02E-02$
	NM(0, 7)	-7,59152	2,07685	$4,21E-02$	0,6209857
9	NM(0, 3)	-0,57935	0,35725	-0,1038996	0,1709682
	NM(0, 4)	0,0048685	0,0792519	$2,55E-02$	$1,30E-02$
	NM(0, 5)	-0,3818	0,24081	$-5,04E-02$	$9,67E-02$
	NM(0, 6)	0,00061	0,03495	$7,95E-03$	$5,87E-03$
	NM(0, 7)	-0,23017	0,16808	$-2,20E-02$	$5,07E-02$

Tabelle A.4: Deskriptive Statistik zu normalisierten Momenten

Klasse	Merkmal	Minimum	Maximum	Mittelwert	σ
0	LM(1, 1)	-1, 23272	0, 32905	-0, 5383598	0, 2824658
	LM(1, 2)	-0, 36188	1, 88072	0, 2054938	0, 2298531
	LM(1, 3)	-0, 81056	2, 24242	0, 5926719	0, 5082581
	LM(1, 4)	-1, 5577	0, 89721	-0, 3367682	0, 3537077
	LM(1, 5)	-1, 61778	1, 10595	0, 1384283	0, 4459868
1	LM(1, 1)	-1, 52045	0, 45717	-0, 8759565	0, 4089963
	LM(1, 2)	-1, 32249	0, 69434	-1, 27E - 02	0, 1833137
	LM(1, 3)	-1, 35936	3, 54295	2, 3660323	0, 9911847
	LM(1, 4)	-1, 46333	2, 10197	-4, 43E - 02	0, 2642663
	LM(1, 5)	-4, 62884	2, 04613	-2, 8793303	1, 2398663
2	LM(1, 1)	-1, 31838	0, 68659	-0, 4462989	0, 3783545
	LM(1, 2)	-0, 45733	1, 19514	0, 3269744	0, 3405783
	LM(1, 3)	-0, 79941	2, 67243	0, 7059405	0, 5939198
	LM(1, 4)	-1, 9869	1, 30763	-0, 5800353	0, 5320953
	LM(1, 5)	-2, 03852	1, 61075	-0, 1615016	0, 6156529
3	LM(1, 1)	-1, 1811	0, 874	-0, 2498105	0, 3924817
	LM(1, 2)	-0, 33149	1, 12688	0, 2942292	0, 2210072
	LM(1, 3)	-1, 37093	1, 92056	0, 4114131	0, 659625
	LM(1, 4)	-1, 65175	1, 55427	-0, 2909279	0, 5460991
	LM(1, 5)	-1, 57477	0, 98674	-0, 1327597	0, 4428143
4	LM(1, 1)	-0, 85786	0, 62881	-0, 2061102	0, 3042024
	LM(1, 2)	-1, 45816	0, 32886	-0, 4535567	0, 3099189
	LM(1, 3)	-1, 70284	1, 661	0, 2946718	0, 6648308
	LM(1, 4)	-0, 86275	2, 63788	1, 1647849	0, 6233206
	LM(1, 5)	-1, 8222	2, 7472	7, 23E - 02	0, 8374187
5	LM(1, 1)	-1, 42231	0, 65389	-0, 4108061	0, 4708375
	LM(1, 2)	-1, 03177	0, 88068	-3, 04E - 02	0, 3487892
	LM(1, 3)	-1, 05524	1, 92236	0, 4586114	0, 7157027
	LM(1, 4)	-1, 53654	1, 62829	-2, 54E - 02	0, 6430109
	LM(1, 5)	-1, 13996	1, 48532	0, 2142304	0, 4886152
6	LM(1, 1)	-0, 81187	0, 77663	-3, 49E - 02	0, 3187588
	LM(1, 2)	-7, 63384	0, 86563	-9, 07E - 02	0, 5194369
	LM(1, 3)	-1, 69815	2, 17002	0, 1118623	0, 6715027
	LM(1, 4)	-1, 55944	0, 76468	-0, 4835801	0, 4127007
	LM(1, 5)	-2, 54403	1, 57346	-9, 39E - 02	0, 6791993
7	LM(1, 1)	-0, 9047	1, 03064	-8, 19E - 02	0, 3910839
	LM(1, 2)	-1, 09157	0, 72372	-0, 140435	0, 3225832
	LM(1, 3)	-2, 2969	2, 20192	0, 3066003	0, 8452363
	LM(1, 4)	-0, 37918	2, 18561	1, 143585	0, 5397087
	LM(1, 5)	-2, 46622	1, 97718	-0, 3488293	0, 8545025
8	LM(1, 1)	-1, 39816	0, 34446	-0, 5118707	0, 3672328
	LM(1, 2)	-0, 96605	0, 70515	-0, 1153935	0, 2534276
	LM(1, 3)	-0, 71785	2, 5482	0, 9189409	0, 6853421
	LM(1, 4)	-1, 33997	1, 70442	0, 2648232	0, 5453413
	LM(1, 5)	-2, 46061	1, 22522	-0, 3555666	0, 6204881
9	LM(1, 1)	-0, 82403	0, 8296	-0, 1113211	0, 3194067
	LM(1, 2)	-0, 50636	0, 49233	1, 72E - 02	0, 1964277
	LM(1, 3)	-2, 00587	1, 87217	0, 2346915	0, 7032957
	LM(1, 4)	-0, 7167	1, 64064	0, 6684085	0, 4698034
	LM(1, 5)	-2, 00702	1, 88464	-0, 1495536	0, 703068

Tabelle A.5: Deskriptive Statistik zu Legendre Momenten

Klasse	Merkmal	Minimum	Maximum	Mittelwert	σ
0	WM(0, 0, 1)	0	0,00242	$1,65E-04$	$2,48E-04$
	WM(0, 0, 2)	0	0,0015	$4,32E-04$	$3,38E-04$
	WM(0, 0, 3)	0	0,00058	$7,23E-05$	$8,34E-05$
	WM(0, 0, 4)	0	0,00072	$1,47E-04$	$1,52E-04$
	WM(0, 0, 5)	0	0,00027	$2,28E-05$	$3,23E-05$
1	WM(0, 0, 1)	0	0,00092	$4,06E-05$	$1,05E-04$
	WM(0, 0, 2)	0,00021	0,00319	$1,50E-03$	$4,54E-04$
	WM(0, 0, 3)	0	0,00057	$3,08E-05$	$6,29E-05$
	WM(0, 0, 4)	0	0,0027	$7,43E-04$	$4,36E-04$
	WM(0, 0, 5)	0	0,00051	$2,40E-05$	$5,15E-05$
2	WM(0, 0, 1)	0	0,00133	$3,36E-04$	$2,65E-04$
	WM(0, 0, 2)	0	0,00121	$2,00E-04$	$1,88E-04$
	WM(0, 0, 3)	0	0,00158	$2,01E-04$	$2,34E-04$
	WM(0, 0, 4)	0	0,00107	$1,54E-04$	$1,50E-04$
	WM(0, 0, 5)	0	0,00068	$1,24E-04$	$1,27E-04$
3	WM(0, 0, 1)	0	0,0013	$3,61E-04$	$2,40E-04$
	WM(0, 0, 2)	0	0,00137	$1,98E-04$	$1,65E-04$
	WM(0, 0, 3)	0	0,00105	$2,47E-04$	$2,03E-04$
	WM(0, 0, 4)	0	0,001	$1,61E-04$	$1,62E-04$
	WM(0, 0, 5)	0	0,00051	$8,72E-05$	$8,66E-05$
4	WM(0, 0, 1)	0	0,00088	$1,46E-04$	$1,37E-04$
	WM(0, 0, 2)	0	0,00192	$3,19E-04$	$2,68E-04$
	WM(0, 0, 3)	0	0,00167	$2,88E-04$	$2,90E-04$
	WM(0, 0, 4)	0	0,00097	$1,97E-04$	$1,80E-04$
	WM(0, 0, 5)	0	0,00137	$2,08E-04$	$2,31E-04$
5	WM(0, 0, 1)	0	0,00134	$2,43E-04$	$2,32E-04$
	WM(0, 0, 2)	0,00002	0,00212	$7,22E-04$	$4,21E-04$
	WM(0, 0, 3)	0	0,0008	$1,87E-04$	$1,91E-04$
	WM(0, 0, 4)	0	0,00078	$1,81E-04$	$1,68E-04$
	WM(0, 0, 5)	0	0,00047	$9,39E-05$	$9,84E-05$
6	WM(0, 0, 1)	0,00001	0,00155	$4,36E-04$	$2,43E-04$
	WM(0, 0, 2)	0	0,00194	$2,11E-04$	$2,57E-04$
	WM(0, 0, 3)	0	0,00076	$2,32E-04$	$1,72E-04$
	WM(0, 0, 4)	0	0,00061	$1,13E-04$	$1,23E-04$
	WM(0, 0, 5)	0	0,00044	$6,61E-05$	$7,47E-05$
7	WM(0, 0, 1)	0,00008	0,00205	$5,82E-04$	$2,87E-04$
	WM(0, 0, 2)	0	0,00151	$3,86E-04$	$2,84E-04$
	WM(0, 0, 3)	0	0,00079	$2,27E-04$	$1,78E-04$
	WM(0, 0, 4)	0	0,00097	$1,19E-04$	$1,31E-04$
	WM(0, 0, 5)	0	0,00065	$1,02E-04$	$1,14E-04$
8	WM(0, 0, 1)	0	0,00075	$8,60E-05$	$1,03E-04$
	WM(0, 0, 2)	0	0,00083	$1,79E-04$	$1,72E-04$
	WM(0, 0, 3)	0	0,00071	$1,54E-04$	$1,47E-04$
	WM(0, 0, 4)	0	0,00138	$3,61E-04$	$3,06E-04$
	WM(0, 0, 5)	0	0,0011	$9,35E-05$	$1,21E-04$
9	WM(0, 0, 1)	0	0,00094	$2,15E-04$	$1,69E-04$
	WM(0, 0, 2)	0	0,0011	$1,75E-04$	$1,85E-04$
	WM(0, 0, 3)	0	0,00111	$2,96E-04$	$2,25E-04$
	WM(0, 0, 4)	0	0,00055	$1,62E-04$	$1,30E-04$
	WM(0, 0, 5)	0	0,00059	$6,64E-05$	$8,59E-05$

Tabelle A.6: Deskriptive Statistik zu Wavelet Momenten

Klasse	Merkmal	Minimum	Maximum	Mittelwert	σ
0	TM(0, 3)	-0,38	0,35	-2,36E-03	0,1254
	TM(0, 4)	1,5	2,02	1,7311	9,56E-02
	TM(0, 5)	-1,62	1,37	-2,75E-02	0,4746
	TM(0, 6)	2,58	5,18	3,6088	0,4718
	TM(0, 7)	-5,8	4,68	-0,135	1,5504
1	TM(0, 3)	-1,5	0,69	-6,97E-02	0,2214
	TM(0, 4)	1,6	3,25	2,085	0,2494
	TM(0, 5)	-6,38	4,83	-0,3325	1,191
	TM(0, 6)	3,17	14,95	5,8059	1,8737
	TM(0, 7)	-29,22	28,16	-1,4161	6,2888
2	TM(0, 3)	-0,73	0,72	-0,1562	0,2799
	TM(0, 4)	1,69	2,94	2,1447	0,248
	TM(0, 5)	-4,12	4,8	-0,5807	1,5221
	TM(0, 6)	3,35	13,52	6,2148	1,7269
	TM(0, 7)	-21,77	29,18	-1,719	7,4015
3	TM(0, 3)	-1,12	0,2	-0,4075	0,2229
	TM(0, 4)	1,79	4,02	2,2532	0,2977
	TM(0, 5)	-9,1	1,03	-2,1458	1,3255
	TM(0, 6)	3,95	27,64	7,2144	2,5367
	TM(0, 7)	-76,08	4,48	-10,2851	7,988
4	TM(0, 3)	-0,73	0,64	-6,55E-02	0,2666
	TM(0, 4)	1,37	2,89	1,9464	0,2397
	TM(0, 5)	-3,63	3,79	-0,2197	1,0997
	TM(0, 6)	2,22	11,6	4,9204	1,4116
	TM(0, 7)	-18,77	20,13	-0,6467	4,4827
5	TM(0, 3)	-0,51	0,97	0,2157	0,3182
	TM(0, 4)	1,57	3,39	2,3864	0,3859
	TM(0, 5)	-2,74	6,39	1,2712	1,8452
	TM(0, 6)	2,87	17,43	7,9843	2,7783
	TM(0, 7)	-13,4	41,66	6,8518	9,9704
6	TM(0, 3)	-0,35	0,79	0,171	0,1761
	TM(0, 4)	1,72	3,31	2,0379	0,1687
	TM(0, 5)	-1,8	5,34	0,8422	0,8714
	TM(0, 6)	3,55	17,79	5,3643	1,277
	TM(0, 7)	-7,81	40,22	3,6222	4,2863
7	TM(0, 3)	-1,16	0,41	-0,2574	0,3123
	TM(0, 4)	1,38	3,81	2,1703	0,3837
	TM(0, 5)	-6,81	2,17	-1,3656	1,6754
	TM(0, 6)	2,21	21,79	6,5235	2,9066
	TM(0, 7)	-47,86	10,31	-6,6953	9,0163
8	TM(0, 3)	-0,36	1,14	0,1725	0,2411
	TM(0, 4)	1,62	3,98	2,1556	0,3216
	TM(0, 5)	-2,11	8,94	0,9679	1,408
	TM(0, 6)	3,13	25,73	6,328	2,6077
	TM(0, 7)	-10,28	68,94	4,9799	8,5288
9	TM(0, 3)	-0,75	0,44	-0,1693	0,2535
	TM(0, 4)	1,56	2,8	2,0382	0,1649
	TM(0, 5)	-3,24	2,13	-0,7022	1,0939
	TM(0, 6)	2,85	10,16	5,3491	0,9775
	TM(0, 7)	-13,12	8,85	-2,6156	4,3377

Tabelle A.7: Deskriptive Statistik zu Tsirikolias - Momenten

Klasse	Merkmal	Minimum	Maximum	Mittelwert	σ
0	OFM(0, 1)	0	0,13014	$6,31E-03$	$1,19E-02$
	OFM(0, 2)	0	0,16335	$2,27E-02$	$2,40E-02$
	OFM(0, 3)	0	0,08802	$7,52E-03$	$9,50E-03$
	OFM(0, 4)	0	0,02456	$4,40E-03$	$4,52E-03$
	OFM(0, 5)	0	0,04868	$2,79E-03$	$4,56E-03$
1	OFM(0, 1)	0	0,1545	$4,34E-03$	$1,09E-02$
	OFM(0, 2)	0,13717	0,89185	0,5548198	0,1082976
	OFM(0, 3)	0	0,16819	$2,75E-03$	$1,18E-02$
	OFM(0, 4)	0,01275	0,81988	0,2963059	0,1168479
	OFM(0, 5)	0	0,05593	$2,71E-03$	$4,82E-03$
2	OFM(0, 1)	0,00002	0,13018	$2,23E-02$	$2,03E-02$
	OFM(0, 2)	0,00411	0,39468	$9,71E-02$	$7,13E-02$
	OFM(0, 3)	0,00001	0,31424	$5,27E-02$	$5,24E-02$
	OFM(0, 4)	0,00037	0,17917	$3,46E-02$	$3,31E-02$
	OFM(0, 5)	0,00006	0,11971	$1,81E-02$	$1,99E-02$
3	OFM(0, 1)	0,00022	0,11774	$2,88E-02$	$2,29E-02$
	OFM(0, 2)	0,00021	0,34103	$8,95E-02$	$5,95E-02$
	OFM(0, 3)	0,00001	0,13005	$2,16E-02$	$2,16E-02$
	OFM(0, 4)	0,00006	0,08307	$1,64E-02$	$1,54E-02$
	OFM(0, 5)	0,00004	0,07291	$1,29E-02$	$1,34E-02$
4	OFM(0, 1)	0,00001	0,0685	$1,14E-02$	$1,16E-02$
	OFM(0, 2)	0,00024	0,28207	$3,93E-02$	$4,02E-02$
	OFM(0, 3)	0,00101	0,34116	$6,91E-02$	$5,58E-02$
	OFM(0, 4)	0,00003	0,12235	$2,85E-02$	$2,59E-02$
	OFM(0, 5)	0,00065	0,20699	$6,60E-02$	$4,56E-02$
5	OFM(0, 1)	0,00012	0,1255	$2,33E-02$	$2,41E-02$
	OFM(0, 2)	0,00491	0,40792	$8,83E-02$	$7,25E-02$
	OFM(0, 3)	0,00011	0,13504	$2,43E-02$	$2,54E-02$
	OFM(0, 4)	0,00011	0,13906	$2,02E-02$	$2,47E-02$
	OFM(0, 5)	0	0,074	$1,28E-02$	$1,51E-02$
6	OFM(0, 1)	0,001	0,1108	$2,73E-02$	$1,71E-02$
	OFM(0, 2)	0,00014	0,35342	$4,43E-02$	$5,38E-02$
	OFM(0, 3)	0,001	0,17481	$4,13E-02$	$2,31E-02$
	OFM(0, 4)	0,00026	0,10742	$2,10E-02$	$1,72E-02$
	OFM(0, 5)	0,00003	0,07585	$1,00E-02$	$1,05E-02$
7	OFM(0, 1)	0,00608	0,12412	$3,65E-02$	$1,59E-02$
	OFM(0, 2)	0,00295	0,33985	$9,11E-02$	$6,66E-02$
	OFM(0, 3)	0,00949	0,2623	$8,72E-02$	$4,00E-02$
	OFM(0, 4)	0,00009	0,13034	$1,38E-02$	$1,41E-02$
	OFM(0, 5)	0,00024	0,14158	$3,58E-02$	$2,89E-02$
8	OFM(0, 1)	0,00002	0,05277	$6,22E-03$	$9,14E-03$
	OFM(0, 2)	0,00012	0,45431	0,1344259	$9,36E-02$
	OFM(0, 3)	0,00006	0,16287	$3,25E-02$	$3,08E-02$
	OFM(0, 4)	0,00004	0,12427	$2,20E-02$	$2,16E-02$
	OFM(0, 5)	0,00001	0,12239	$1,30E-02$	$1,70E-02$
9	OFM(0, 1)	0,00005	0,05376	$9,99E-03$	$1,09E-02$
	OFM(0, 2)	0,00027	0,26159	$3,95E-02$	$4,17E-02$
	OFM(0, 3)	0,00342	0,1454	$4,70E-02$	$2,64E-02$
	OFM(0, 4)	0,0004	0,07168	$1,77E-02$	$1,24E-02$
	OFM(0, 5)	0,00029	0,0955	$1,73E-02$	$1,33E-02$

Tabelle A.8: Deskriptive Statistik zu Orthogonal Fourier-Mellin Momenten

Literaturverzeichnis

- [1] Saranli Afsar and Mubeccel Demirekler. A statistical unified framework for rank-based multiple classifier decision combination. *Pattern Recognition*, 34:865–884, 2001.
- [2] Maher Ahmed and Rabab Kreidieh Ward. An expert system for general symbol recognition. *Pattern Recognition*, 33:1975–1988, 2000.
- [3] H. C. Andrews. Multidimensional rotations in feature selection. *IEEE Trans. Computers*, 20:1045–1051, 1971.
- [4] Nafiz Arica and Fatos T. Yarman-Vural. An overview of character recognition focused on off-line handwriting. *IEEE Transactions on systems, man and cybernetics Part C: Applications and reviews*, 31:216–233, may 2001.
- [5] Ivar Balslev, Kasper Doring, and Rene Decker Eriksen. Weighted central moments in pattern recognition. *Pattern Recognition Letters*, 21:381–384, 2001.
- [6] S. O. Belkasim, M. Shridilar, and M. Ahmadi. Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition*, 24:1117–1138, 1991.
- [7] A.B. Bhatia and E. Wolf. On the circle polynomials of zernike and related orthogonal sets. *Proceedings Cambridge Philosophical Society*, 50:40–48, 1954.
- [8] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.
- [9] Jinhai Cai and Zhi-Qiang Liu. Hidden markov models with spectral features for 2d shape recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 23 (12):1454–1459, 2001.
- [10] R. Courant and D. Hilbert. *Methods of Mathematics in Physics, Vol I*. Interscience, New York, 1953.
- [11] Luciano da Fontoura Costa & Roberto M. Cesar Jr. *Shape Analysis & Classification*. CRC Press, 2001.

- [12] J.P.Marques de Sa. *Pattern Recognition - Concepts, Methods and Applications*. Springer, Berlin, Heidelberg, New York, 2001.
- [13] P. A. Devijer and J. Kittler. *Pattern Recognition: a statistical approach*. Prentice Hall, London, 1982.
- [14] R. O. Duda and P. E. Hart. *Pattern classification and scene analysis*. John Wiley & Sons, New York, 1973.
- [15] Adolf Ebeling. Kein x für ein u. *C't*, 19:176–179, 2001.
- [16] M. Egmont-Petersen, D. de Ridder, and H. Handels. Image processing with neural networks – a review. *Pattern Recognition*, 35(10):2279–2301, 2002.
- [17] E.K. Etienne and Mike (eds) Nachtegaal. *Fuzzy Techniques in image processing*. Springer, Physica-Verlag, New York, 2000.
- [18] Jan Flusser. Fast calculation of geometric moments of binary images. *Pattern Recognition and Medical Computer Vision*, pages 265–274, 1998.
- [19] Jan Flusser. On the independence of rotation moment invariants. *Pattern Recognition*, 33:1405–1410, 2000.
- [20] Jan Flusser. Refined moment calculation using image block representation. *IEEE Transactions on Image Processing*, 9:1977–1978, 2000.
- [21] Jan Flusser and Tomas Suk. Degraded image analysis: an invariant approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:590–603, june 1998.
- [22] M. D. Garris and J. L. Blue. *NIST Form-Based Handprint Recognition System*. National Institute of Standards and Technology, 1995.
- [23] R. C. Gonzalez and R. E. Woods. *Digital Image Processing, 2nd ed.* Prentice Hall, Upper Saddle River, NJ., 2002.
- [24] Simon Haykin. *Neural Networks - A comprehensive foundation*. Prentice Hall, New Jersey, 1999.
- [25] L. Heutte, T. Paquet, J.V.Moreau, Y. Lecourtier, and C. Olivier. A structural/statistical feature based vector for handwritten character recognition. *Pattern Recognition Letters*, 19:629–641, 1998.
- [26] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Transactions on information theory*, 8:179–187, 1962.
- [27] Davis P. J. Plane regions determined by complex moments. *Journal of Approximation theory*, 19:148–153, 1977.

- [28] A. Jain and D. Zongker. Feature selection: Evaluation, application and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:153–158, 1997.
- [29] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31:264 – 323, september 1999.
- [30] Anil K. Jain, Robert P.W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 22:4–37, 2000.
- [31] X. Y. Jiang and H. Bunke. Simple and fast computation of moments. *Pattern Recognition*, 24:801–806, 1991.
- [32] Chao Kan and Mandyam D. Srinath. Invariant character recognition with zernike and orthogonal fourier-mellin moments. *Pattern Recognition*, 35:143–154, 2002.
- [33] Alireza Khotanzad and Yaw Hua Hong. Invariant image recognition by zernike moments. *IEEE transaction on pattern analysis and machine intelligence*, 12:489–497, 1990.
- [34] Hyun-Chul Kim, Daijin Kim, and Sung Yang Bang. A numerak character recognition using the pca mixture model. *Pattern Recognition Letters*, 23:103–111, 2002.
- [35] Mineichi Kudo and Jack Sklansky. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33:25–41, 2000.
- [36] Y LeChun, B. Boser, and J.S. Denker. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1 (4):541 – 551, 1989.
- [37] Thomas Lehmann, Walter Oberschelp, Erich Pelikan, and Rudolf Repges. *Bildverarbeitung für die Medizin*. Springer, Berlin, Heidelberg, New York, 1997.
- [38] Jia-Guu Leu. Computing a shapes moments from its boundary. *Pattern Recognition*, 24:949–957, 1991.
- [39] Yajun Li. Reforming the theory of invariant moments for pattern recognition. *Pattern Recognition*, 25:723–730, 1992.
- [40] Simon X. Liao and Miroslaw Pawlak. On the accuracy of zernike moments for image analysis. *IEEE transactions on pattern analysis and machine intelligence*, 20:1358–1364, 1998.
- [41] W. G. Lin and S. Wang. A note on the calculation of moments. *Pattern Recognition Letters*, 15:1065–1070, 1994.

- [42] Alexander G. Mamistvalov. n -dimensional moment invariants and conceptual mathematical theory of recognition n -dimensional solids. *IEEE Transactions on pattern analysis and machine intelligence*, 20:819–831, august 1998.
- [43] Basil G. Mertzios and Dimitris A. Karras. *On applying fast and efficient methods in pattern recognition*. IOS Press, J. S. Byrnes (Ed.) / NATO ASI, 1999.
- [44] R. Mukundan and K. R. Ramakrishnan. Fast computation of legendre and zernike moments. *Pattern Recognition*, 28:1433–1442, 1995.
- [45] National Institut of Standards and Technology. *FL3 Handwritten Symbol Database Subset of the NIST Special Database 1*. NIST, <ftp://sequoyah.ncsl.nist.gov/pub/databases/data>.
- [46] Jaehwa Park and Venu Govindaraju. Use of adaptive segmentation in handwritten phrase recognition. *Pattern Recognition*, 35:245–252, 2002.
- [47] J.R. Parker. *Algorithms for image processing and computer vision*. Wiley, New York, Chichester, Brisbane, Toronto, Singapore, Weinheim, 1997.
- [48] Wilfried Philips. A new fast algorithm for moment computation. *Pattern Recognition*, 26:1619–1621, 1993.
- [49] Zhang Ping and Chen Lihui. A novel feature extraction method and hybrid tree classification for handwritten numeral recognition. *Pattern Recognition Letters*, 23:45–56, 2002.
- [50] Rejean Plamondon and Sargur N. Srihari. On-line and off-line handwriting recognition a comprehensive survey. *IEEE Transactions on Pattern Analysis AND Machine Intelligence*, 22:63–84, 2000.
- [51] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.
- [52] T.H. Reiss. The revised fundamental theorem of moment invariants. *IEEE Transactions on pattern analysis and machine intelligence*, 13:830–834, august 1991.
- [53] Thomas H. Reiss. *Recognizing Planar Objects Using Invariant Image Features*. Springer, Berlin, Heidelberg, New York, 1993.
- [54] Brian D. Ripley. *Pattern recognition and neural networks*. Cambridge University Press, Cambridge, UK, 1996.
- [55] Irene Rothe, Herbert Süsse, and Klaus Voss. The method of normalization to determine invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:366–376, 1996.

- [56] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.
- [57] Dinggang Shen and Horace H.S. Ip. Discriminative wavelet shape descriptors for recognition of 2-d patterns. *Pattern Recognition*, 32:151–165, 1999.
- [58] Y. Sheng and L. Shen. Orthogonal fourier-mellin moments for invariant pattern recognition. *Journal of Optical Society of America*, 11:1748–1757, 1994.
- [59] Huazhong Shu, Limin Luo, Xudong Bao, and Wenxue Yu. An efficient method for computation of legendre moments. *Graphical Models*, 62:237–262, 2000.
- [60] H.Z. Shu, L.M. Luo, W.X. Yu, and Y. Fu. A new fast method for computing legendre moments. *Pattern Recognition*, 33:341–348, 2000.
- [61] Mark H. Singer. A general approach to moment calculation for polygons and line segments. *Pattern Recognition*, 26:1019–1028, 1993.
- [62] P. Somol, P. Pudil, J. Novovicova, and P. Paclik. Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, 20:1157–1163, 1999.
- [63] Ching Y. Suen and Il-Seok Oh. A class-modular feedforward neural network for handwritten recognition. *Pattern Recognition*, 35:229–244, 2002.
- [64] Yuan Y. Tang, Ching Y. Suen, and Seong-Whan Lee. Automatic document processing: a survey. *Pattern Recognition*, 29:1931–1952, 1996.
- [65] Michael Reed Teague. Image analysis via the general theory of moments. *Journal of Optical Society of America*, 70:920–930, 1980.
- [66] Cho-Huak Teh and Roland T. Chin. On image analysis by the method of moments. *IEEE Transactions on pattern analysis and machine intelligence*, 10:496–513, 1988.
- [67] Oivind Due Trier, Anil K. Jain, and Torfinn Taxt. Feature extraction methods for character recognition - a survey. *Pattern Recognition*, 29:641–662, 1996.
- [68] K. Tsirikolias and B.G.Mertzios. Statistical pattern recognition using efficient two-dimensional moments with applications to character recognition. *Pattern Recognition*, 26:877–883, 1993.
- [69] Ake Wallin and Olaf Kübler. Complete sets of complex zernike moment invariants and the role of the pseudoinvariants. *IEEE Transactions on pattern analysis and machine intelligence*, 17:1106–1110, november 1995.
- [70] Luren Yang and Fritz Albregtsen. Fast and exact computation of cartesian geometric moments using discrete green’s theorem. *Pattern Recognition*, 29:1061–1073, 1996.

-
- [71] Y.C.Chim, A.A.Kassim, and Y.Ibrahim. Character recognition using statistical moments. *Image and Vision Computing*, 17:299–307, 1999.
 - [72] M. F. Zakaria, L. J. Vroomen, P. Zsombor-Murray, and J.M. van Kessel. Fast algorithm for the computation of moment invariants. *Pattern Recognition*, 20:639–643, 1987.
 - [73] J.D. Zhou, H.Z. Shu, L.M. Luo, and W.X. Yu. Two new algorithms for efficient computation of legendre moments. *Pattern Recognition*, 35:1143–1152, 2002.

Danksagung

Ich möchte mich an erster Stelle bei Dr. Raouf Hamzaoui für die Auswahl des interessanten Themas und die stetige Betreuung herzlich bedanken. Für die ständige geistige und moralische Unterstützung möchte ich mich ebenso herzlich bei Katharina Buchta, Tobias Dörfler, Michael Köhler, Stephanie Schmidt und Andrea Simmel bedanken. Auch möchte ich mich bei meinen Studienfreunden Ingmar Bauer und Heiko Stamer für die vielen konstruktiven Gespräche bedanken, die ebenso zum Gelingen der Arbeit beitrugen. In statistischen Fragen beriet mich Dr. Frank Piontek - auch ihm ein herzliches Danke. Ebenso möchte ich mich bei Dr. Thomas Villmann bedanken, der mich bei Fragen der Datenreduktion konstruktiv unterstützte. Ein besonderer Dank bei der Bereitstellung literarischer Quellen gilt dem Max-Planck-Institut für neuropsychologische Forschung, dessen Bibliothek mir eine große Hilfe war. Ganz besonders möchte ich mich auch bei meinen Eltern bedanken, die mir zu jeder Zeit mit Rat und Tat zur Seite standen.

Selbständigkeitserklärung

Hiermit erkläre ich, daß ich diese Diplomarbeit eigenständig und unter ausschließlicher Nutzung der angegebenen Hilfsmittel angefertigt habe.

Leipzig, den 26. Juli 2002

Frank-Michael Schleif