

# HMM-basierte Online Handschrifterkennung — ein integrierter Ansatz zur Text- und Formelerkennung

Vom Fachbereich Elektrotechnik  
der Gerhard-Mercator-Universität - Gesamthochschule Duisburg

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

genehmigte Dissertation

von

Andreas Kosmala

aus Herne

Referent: Prof. Dr.-Ing. habil. G. Rigoll

Korreferent: Prof. B. Hosticka Ph.D.

Tag der mündlichen Prüfung: 11. Dezember 2000

# Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter im Fachgebiet Technische Informatik des Fachbereichs Elektrotechnik an der Gerhard-Mercator-Universität Duisburg.

Dem Leiter des Fachgebietes Herrn Prof. Dr.-Ing. habil. Gerhard Rigoll gilt mein besonderer Dank für die Anregung zu dieser sehr interessanten Themenstellung, für die wertvollen Diskussionen und für die gewährten wissenschaftlichen Entfaltungsmöglichkeiten.

Prof. Ph.D. Bedrich Hosticka (Fraunhofer-Institut IMS, Duisburg) danke ich für die Übernahme des Korreferats und die kooperative Zusammenarbeit.

Danken möchte ich auch meinen Kollegen für die fundierten Diskussionen und die vielfältige Mithilfe. Insbesondere möchte ich mich bei Frau Dipl.-Ing. Anja Brakensiek für die Durchsicht dieser Arbeit und die zahlreichen Anmerkungen bedanken.

Ferner danke ich all den Versuchspersonen, die im Rahmen der Datenakquisition zu einer beträchtlichen Sammlung von Handschriftdaten beitrugen und damit viele Experimente und Entwicklungen ermöglicht haben.

Schließlich möchte ich mich bei meiner lieben Frau Antje und meinen Kindern Sophia, Jan Philipp und Niklas für deren Verständnis und deren stete Unterstützung - auch in sehr arbeitsintensiven Phasen - bedanken.

Ihnen sei diese Arbeit gewidmet.

Duisburg, im Dezember 2000

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Vorverarbeitung</b>	<b>4</b>
2.1	Neuabtastung . . . . .	5
2.2	Normalisierung . . . . .	7
2.2.1	Zeilenneigung - Skew . . . . .	8
2.2.2	Schriftneigung - Slant . . . . .	10
2.2.3	Größennormalisierung . . . . .	14
2.3	Ergebnisse . . . . .	16
2.4	Kapitelzusammenfassung . . . . .	18
<b>3</b>	<b>Merkmalsextraktion</b>	<b>20</b>
3.1	Trajektorienmerkmale . . . . .	21
3.1.1	Kettenkodierung . . . . .	21
3.1.2	Diskrete Cosinus Transformation der Trajektorie . . . . .	23
3.1.3	Bézierkurve . . . . .	25
3.1.4	B-Splines . . . . .	30
3.1.5	Hauptachsentransformation . . . . .	32
3.2	Bitmap-Merkmale . . . . .	33
3.2.1	Räumliche Unterabtastung der Bitmap . . . . .	34
3.2.2	Diskrete Cosinus Transformation (DCT) der Bitmap . . . . .	35
3.2.3	Walsh Transformation der Bitmap . . . . .	36
3.3	Ergebnisse . . . . .	36
3.3.1	Trajektorienmerkmale . . . . .	36
3.3.2	Bitmap-Merkmale . . . . .	38
3.3.3	Merkmalskombination . . . . .	39

3.4	Kapitelzusammenfassung . . . . .	42
<b>4</b>	<b>Modellierung</b>	<b>43</b>
4.1	Vergleich zwischen diskreten und kontinuierlichen Modellen . . . . .	44
4.2	Verwendung neuronaler Netze zur Merkmalsquantisierung . . . . .	47
4.2.1	Die Transinformation als Gütemaß für das Training neuronaler Netze	48
4.2.2	Automatische Handschrifterkennung - ein Nachrichtenübertragungssystem . . . . .	50
4.3	Hybrides NN/HMM Handschrifterkennungssystem . . . . .	53
4.3.1	MMI-Training einzelner Codebücher . . . . .	54
4.3.2	Simultanes MMI-Training multipler Codebücher . . . . .	55
4.4	Ergebnisse . . . . .	58
4.5	Kapitelzusammenfassung . . . . .	60
<b>5</b>	<b>Strukturoptimierung bei Verwendung kontextabhängiger Modelle</b>	<b>62</b>
5.1	Bi- und Trigrapheme . . . . .	63
5.2	Parameterreduktion . . . . .	65
5.2.1	Selektives Verfahren . . . . .	66
5.2.2	Parameter-Tying . . . . .	67
5.3	Datengetriebenes State-Clustering . . . . .	68
5.4	Entscheidungsbaum basiertes State-Clustering . . . . .	72
5.4.1	Einbeziehung typographischer Eigenschaften . . . . .	79
5.4.2	Clustering Markov'scher Quellen . . . . .	80
5.4.3	Tree-Sweeping . . . . .	84
5.5	Ergebnisse . . . . .	86
5.6	Kapitelzusammenfassung . . . . .	89
<b>6</b>	<b>Textmodellierung</b>	<b>91</b>
6.1	Verknüpfung von Graphem- und Textmodellen . . . . .	92
6.2	Modellebenen . . . . .	93
6.2.1	Graphemmodelle . . . . .	93
6.2.2	Textmodelle . . . . .	93
6.3	Dekodierung . . . . .	95
6.4	Ergebnisse und Kapitelzusammenfassung . . . . .	97

<b>7</b>	<b>Formelerkennung</b>	<b>98</b>
7.1	Grundlegende Funktionsweise . . . . .	99
7.2	Vorverarbeitung, Merkmalsextraktion und Dekodierung . . . . .	101
7.3	Strukturanalyse . . . . .	103
7.4	Parsing – LRTD-Ansatz . . . . .	104
7.4.1	Fehlerkorrektur . . . . .	105
7.4.2	Spezielle mathematische Operatoren . . . . .	106
7.4.3	Sub- und Superskript . . . . .	108
7.5	2D-Parsing mit kontextuellen Graph-Grammatiken . . . . .	110
7.6	Manuelle Korrekturfunktionen . . . . .	111
7.6.1	Funktionsweise manueller Korrekturfunktionen . . . . .	112
7.6.2	Löschen . . . . .	113
7.6.3	Ersetzen . . . . .	115
7.6.4	Einfügen . . . . .	116
7.6.5	Rückgängig und Wiederholen . . . . .	117
7.6.6	Neuzeichnen . . . . .	117
7.7	Ergebnisse . . . . .	119
7.8	Kapitelzusammenfassung . . . . .	121
<b>8</b>	<b>Zusammenfassung</b>	<b>122</b>
<b>A</b>	<b>Zusammenfassung grundlegender informationstheoretischer Zusammenhänge</b>	<b>126</b>
<b>B</b>	<b>Beschreibung der verwendeten Datenbasen</b>	<b>128</b>
B.1	Schreiberabhängige Textdatenbasis . . . . .	128
B.2	Schreiberunabhängige Textdatenbasis . . . . .	129
B.3	Schreiberabhängige Formeldatenbasis . . . . .	130
	<b>Literaturverzeichnis</b>	<b>132</b>

# Abbildungsverzeichnis

1.1	Systemübersicht . . . . .	3
2.1	Neuabtastung - prinzipielle Wirkungsweise . . . . .	6
2.2	Neuabtastung mit verschiedenen Vektorlängen . . . . .	7
2.3	Zeilenrotation . . . . .	9
2.4	Entropie der Integralprojektion als Funktion des Zeilenneigungswinkels . . . . .	10
2.5	Entropieverlauf als Funktion des Scherungswinkels . . . . .	11
2.6	Vorzeichenkonvention für die Schriftscherung . . . . .	12
2.7	Beispiele für nicht eindeutig schätzbare Slant-Winkel . . . . .	13
2.8	Bereichsunterteilung der Handschrift . . . . .	14
2.9	Iterative Bereichsdetektion . . . . .	15
3.1	Verlauf von $(x,y)(k)$ einer Schriftprobe . . . . .	23
3.2	Konstruktion einer Bézierkurve . . . . .	26
3.3	Bézierkurve - $\vec{B}_3(t)$ . . . . .	28
3.4	Abgetastete Trajektorienstücke des Wortes 'Go' mit approximierten Kontrollpunkten und rekonstruierten Bézier-Kurven . . . . .	30
3.5	Abgetastete Trajektorienstücke des Wortes 'Go' aus Abb. 3.4 mit approximierten Kontrollpunkten und rekonstruierten B-Splines . . . . .	31
3.6	Superposition relativer Fenster (Trainings-Set) . . . . .	33
3.7	Gleitende Bitmap . . . . .	34
3.8	Unterabtastung der Bitmap . . . . .	35
4.1	Prinzipielle Darstellung kontinuierlicher und diskreter HMM . . . . .	45
4.2	Alignment nach Initialisierung und nach vollständigem Training . . . . .	46
4.3	Informationstheoretisches Kanalmodell (Berger'sches Diagramm) . . . . .	50
4.4	Übersicht eines verteilten Erkennungssystems . . . . .	52

4.5	Verlauf der Transinformation für separates und simultanes MMI-Training . . . . .	59
5.1	Kontextbedingte Graphemvariationen . . . . .	63
5.2	Trigraphem-Häufigkeiten der Trainingsdatenbasis (1K-Lexikon) . . . . .	66
5.3	Parameterverknüpfung bei Trigraphemen . . . . .	68
5.4	Datengetriebenes State-Clusterings . . . . .	70
5.5	Cluster-Splitting . . . . .	73
5.6	Prinzip des Entscheidungsbaum basierten State-Clusterings . . . . .	78
5.7	Clustering Markov'scher Quellen . . . . .	81
5.8	Beispiele rekonstruierter Grapheme . . . . .	83
5.9	Übersicht zum Entscheidungsbaum basierten Clustering . . . . .	84
5.10	Split-and-Merge von Zustands-Clustern . . . . .	85
5.11	Entwicklung der Master-Knoten durch Tree-Sweeping . . . . .	87
6.1	Stackalgorithmus mit zeitabhängigen Stacks und zeitsynchroner Stackbear- beitung . . . . .	95
6.2	Anordnung der Stacks und Zuordnung der Hypothesen . . . . .	96
7.1	Benutzungsregeln zur zeitlichen Abfolge handschriftlicher Eingaben . . . . .	100
7.2	Systemübersicht des Formeleditors . . . . .	102
7.3	Bestimmung der räumlichen Segmentierung aus zeitlicher Segmentierung . . . . .	104
7.4	Beispiel zur Fehlerkorrektur . . . . .	105
7.5	Detektion von Unter- und Obergrenze am Beispiel eines Integrals . . . . .	107
7.6	Gruppierung von Teilausdrücken . . . . .	108
7.7	Implementierte Editierbefehle . . . . .	113
7.8	Löschen von Teilausdrücken . . . . .	114
7.9	Löschen mehrerer Teilausdrücke . . . . .	115
7.10	Ersetzen von Teilausdrücken . . . . .	115
7.11	Einfügen von Teilausdrücken . . . . .	117
7.12	Aktualisierung des Schriftbildes . . . . .	118
7.13	Korrekt erkannte Beispiele . . . . .	120
8.1	Oberfläche des Demonstrators . . . . .	125
B.1	Vorsegmentierte Einzelzeichen . . . . .	129

B.2	Textpassagen für das Training . . . . .	130
B.3	Beispiele aus der Formeldatenbasis (Test und Training) . . . . .	131



# Tabellenverzeichnis

2.1	Ergebnisse der Slantkorrektur . . . . .	17
2.2	Ergebnisse der Größenskalierung . . . . .	18
3.1	Erkennungsergebnis - <i>Sinus- und Cosinuswinkel</i> mit $F = 11$ und $N_m = 22$ .	37
3.2	Erkennungsergebnis - <i>DCT</i> mit $F = 11$ und $N_m = 6$ . . . . .	37
3.3	Erkennungsergebnis - <i>Bézierkurve</i> mit $F = 11$ und $N_m = 8$ . . . . .	37
3.4	Erkennungsergebnis - <i>B-Spline</i> mit $F = 11$ und $N_m = 8$ . . . . .	38
3.5	Erkennungsergebnis - <i>Hauptachsentransformation</i> mit $F = 16$ und $N_m = 8$	38
3.6	Erkennungsergebnis - <i>spatial unterabgetastete Bitmap</i> mit $F_{BM} = 30 \times 30$ und $N_m = 9$ . . . . .	38
3.7	Erkennungsergebnis - <i>DCT der Bitmap</i> mit $F_{BM} = 30$ und $N_m = 9$ . . . . .	39
3.8	Erkennungsergebnis - <i>Walsh-Transformation</i> mit $F_{BM} = 30$ und $N_m = 9$ .	39
3.9	MMI und log-Likelihood von Einzelmerkmalen . . . . .	40
3.10	Gesamt-Transinformation $I(S, Y^1, Y^2)$ in bit für Kombinationen aus Bitmap- und Trajektorienmerkmalen . . . . .	40
3.11	Gesamt-Transinformation $I(S, Y^1, Y^2)$ in bit für zwei Bitmap-Merkmale .	41
3.12	Gesamt-Transinformation $I(S, Y^1, Y^2)$ in bit für zwei Trajektorienmerkma- le ( $F=11$ ) . . . . .	41
3.13	Erkennungsraten kombinierter Merkmale . . . . .	41
4.1	Erkennungsergebnis - diskretes System, schreiberabhängig, 30K Lexikon .	58
4.2	Erkennungsergebnis - schreiberabhängig, 30K Lexikon . . . . .	60
4.3	Erkennungsergebnis - schreiberabhängig, 200K Lexikon . . . . .	61
5.1	Modellexpansion bei Verwendung von Allographemen am Beispiel des Wor- tes 'Kontext' . . . . .	64
5.2	Erkennungsergebnisse - Vergleich verschiedener Reduktionsansätze, schrei- berabhängig, 30K Lexikon . . . . .	87

5.3	Vergleich verschiedener Kontextbereiche, schreiberabhängig, 200k Lexikon	88
6.1	Erkennungsergebnisse - verschiedene Kontexttiefen, schreiberunabhängig, Lexikon-frei . . . . .	97
7.1	Erkennungsraten des schreiberabhängigen Formeleditors . . . . .	119

# Verzeichnis der verwendeten Formelzeichen

$\mathbf{a}$	Kontrollpunkt einer Bézierkurve oder eines B-Splines
$\vec{B}$	Bézierkurve
$b$	Ausgabeverteilung eines HMM-Zustandes
$C$	Buchstabenfolge
$c$	Buchstabe, Zeichen, Graphem
$d$	Abstandsmaß zwischen HMM-Zuständen
$d_C$	Abstandsmaß zwischen Clustern von HMM-Zuständen
$F$	Fenstergröße
$F_{BM}$	Bitmap-Fenster
$\mathcal{F}$	Zeitfenster einer Trajektorie
$f_i$	Aktivierungsfunktion des $i$ -ten Neurons
$\mathbf{f}$	Abtastpunkt
$g$	binäre Bitmap
$g'_k$	Ausschnitt aus $g$ um $k$ -ten Abtastpunkt
$h_i$	Hilfsfunktionen
$H$	Entropie
$h$	Histogramm
$I$	Transinformation
$\mathcal{I}$	Informationsgehalt
$i, j, J$	allgemeine Zählvariablen
$K$	Länge einer Schriftprobe in Abtastpunkten
$k$	Zeitdiskreter Index nach der Neuabtastung der Stiftrajektorie
$L$	Likelihood
$l$	allgemeine Zählvariable
$\hat{l}$	Vorgabewert für die Abtastvektorlänge
$\vec{m}$	Merkmalsvektor
$M$	Dimensionalität

$m$	(skalares) Merkmal
$m_x, m_y$	Mittelpunkt eines umschreibenden Rechtecks
$N$	Kontexttiefe
$N^-$	Rückweisungsschwelle für die Größennormalisierung
$N_{\Phi}^-$	Rückweisungsschwelle für die Slantkorrektur
$N_C$	Codebuchgröße
$N_c$	Zeichenvorrat
$N_m$	Anzahl der Komponenten eines Merkmalsvektors
$n$	allgemeine Zählvariable
$o(\cdot)$	Soft-Max Funktion
$p$	Wahrscheinlichkeitsdichtefunktion
$\tilde{p}$	Stiftdruck
$Q(\cdot)$	Kullback-Leibler Distanz
$\mathcal{Q}$	Menge von Fragen
$q$	Frage zur Aufspaltung eines Clusters
$R$	Matrix der Bernstein-Operatoren
$r_{\mathbf{v}}$	Radius des Abtastpunktes $\mathbf{v}$
$r$	Bernstein-Operator
$\vec{S}_3$	Kubischer B-Spline
$S$	Folge von HMM-Zuständen
$S^*$	optimale HMM-Zustandsfolge
$\mathcal{S}$	Menge von HMM-Zuständen (Cluster)
$s$	HMM-Zustand
$t$	Zeit
$U$	Eigenwertmatrix
$V$	Folge von Abtastvektoren
$\vec{v}$	Abtastvektor
$\mathbf{v}_0, \mathbf{v}_1$	Anfangs- bzw. Endpunkt eines Abtastvektors
$W$	Wortfolge, bestehend aus Worten $w$
$W^*$	optimale Wortfolge
$w$	Wort
$X$	beobachtete Folge von Merkmalsvektoren
$x_l, x_r$	linke und rechte Begrenzung des umschreibenden Rechtecks
$x$	Kartesische $x$ -Komponente der Stiftposition
$\vec{x}$	allgemeiner Merkmalsvektor der Beobachtungsfolge $X$
$Y$	Folge von VQ-Codewörtern
$y_{BL}$	Position der Basislinie
$y_{KL}$	Position der Kernlinie

$y_o, y_u$	obere und untere Begrenzung des umschreibenden Rechtecks
$y_n$	(diskretes) VQ-Codewort
$y$	Kartesische $y$ -Komponente der Stiftposition
$Z, z$	Indiziert Merkmalsstrom bzw. Codebuch
$\alpha(k)$	Winkel des $k$ -ten Abtastvektors
$\beta$	Lernparameter
$\gamma_s$	Frequenzierung des HMM-Zustandes $s$
$\Delta T$	Abtastintervall
$\eta$	normalisierter, kubischer B-Spline
$\theta$	Zeilenneigungswinkel
$\Lambda$	Parametersatz des Sprachmodells
$\lambda$	Parametersatz der HMM
$\lambda_{VQ}$	Parametersatz des Vektorquantisierers
$\vec{\lambda}_i$	$i$ -ter Codebuchvektor des Vektorquantisierers
$\Phi$	Schriftneigungswinkel
$\phi_{\mathbf{v}}$	Winkel des Radiusvektors von $\mathbf{v}$ zur $x$ -Achse
$\ \cdot\ _2$	Euklid'sche Norm

# Verzeichnis der verwendeten Abkürzungen

BM	Bitmap
CAGD	Computer Aided Geometric Design
DCT	Diskrete Cosinus Transformation
DFT	Diskrete Fourier Transformation
dpi	Geometrische Auflösung (dots per inch)
FB	Forward-Backward (Algorithmus)
HMM	Hidden Markov Modell
HSE	Handschrifterkennung
INRIA	Institut National de Recherche en Informatique et en Automatique
JPG/JPEG	Joint Photographic Experts Group
LRTD	Left-Right-Top-Down
ME	Merkmalsextraktion
ML	Maximum Likelihood
MLP	Mutli-Layer Perzeptron
MMI	Maximale Transinformation (Maximum Mutual Information)
MPEG	Moving Picture Experts Group
NA	Neuabtastung
NN	Neuronales Netz
nN	Nächster Nachbar
OCR	Optical Character Recognition
OFR	Optical Formula Recognition
PDA	Personal Digital Assistant
PIM	Personal Information Manager
RBF	Radiale Basisfunktionen
TB	Tree-based (Clustering)
VQ	Vektorquantisierung
WD	Schreiberabhängig
WDF	Wahrscheinlichkeitsdichtefunktion
WI	Schreiberunabhängig

WTA Winner-Takes-All

WYSIWYG What-you-see-is-what-you-get Textverarbeitungsprogramme