Kapitel 4

Kognitive Systeme

In den beiden vorangegangenen Jahren hat die Gruppe „Mustererkennung“ des ZPR die Aktivitäten auf dem Gebiet der Handschriftenerkennung um die Entwicklung einer angegliederten assoziativen Datenbank erweitert. Im Bereich Bildverarbeitung galt das Interesse vermehrt neuen anpassungsfähigen Algorithmen.

Durch die rasante Entwicklung der Hardware in den letzten Jahren sind hohe Rechenleistungen auch in PC-basierten Systemen verfügbar. Daher ist es mittlerweile möglich, Verfahren der Mustererkennung in der industriellen Praxis einzusetzen, die vor kurzem noch der reinen Forschung vorbehalten waren. Problematisch ist z. Zt. noch die Anpassung bestehender Software an Veränderungen im Produktionsprozeß. Solche Parameteranpassungen verlangen zumeist algorithmisches Wissen und sind deshalb ein großer Kostenfaktor. Um die Einsetzbarkeit der Mustererkennungsverfahren zu erhöhen, muß das Ziel heute die Entwicklung von selbstlernenden Algorithmen und die Erschließung einer Wissensbasis sein.

Im Bereich Handschriftenerkennung wurde der Schwerpunkt auf die Entwicklung der assoziativen Datenbank „DACCORD“ gelegt, da es sich als essentiell herausgestellt hat, die Ergebnisse von maschinellen Schrifterkennungssystemen durch eine solche Wissensbasis zu verbessern. Die hohe Durchsatzrate von „DACCORD“ selbst bei Datenbankgrößen von über 100 MByte hat schnell zum erfolgreichen Einsatz in einer deutschen Großbank bei der Verarbeitung von Überweisungsformularen geführt. Die maschinelle Schrifterkennung selbst konnte durch die Entwicklung dynamischer Segmentierungsverfahren und neuer hierarchischer Clusterungsverfahren weiter verbessert werden. Im Bereich der industriellen Qualitätskontrolle wurden die existierenden Anwendungen um Algorithmen zum automatischen Lernen und um *Region-of-Interest*-Verfahren erweitert. Letztere suchen in einem größeren Bild einen relevanten Ausschnitt wie z. B. Adressen auf Briefen.

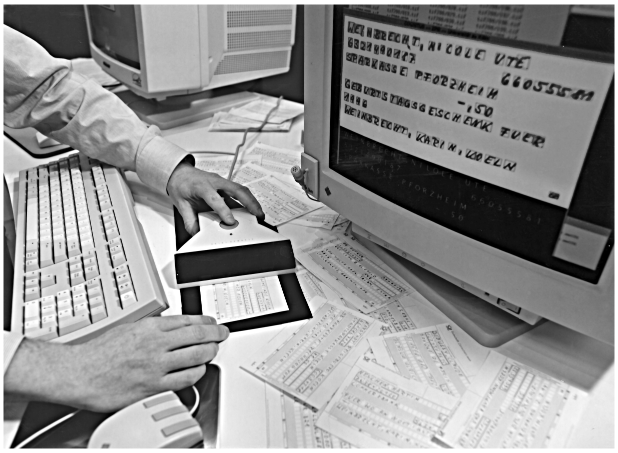

Abbildung 4.1: *Präsentation der Handschriftenerkennung auf der CeBIT ’96*

4.1 Assoziative Datenbanken

Bei der Klassifikation handgeschriebener Zeichen treten häufig mehrdeutige Fälle auf, d. h. ein einzelnes Zeichen kann mehrere verschiedene Interpretationen haben. Beispielsweise sind die Schreibweisen der Ziffern „1“ und „7“ bei manchen Schreibern sehr ähnlich. Solche Mehrdeutigkeiten können aber oft aufgelöst werden, wenn Kontextinformationen zur Verfügung stehen. So kann im obigen Beispiel eine Entscheidung getroffen werden, wenn die betrachtete Ziffer Teil einer Bankleitzahl ist und nur eine der beiden Möglichkeiten eine existierende Bankleitzahl ergibt.

Durch schlechte Leseergebnisse bei schwierigen Handschriften oder durch die Ausgabe von mehreren Alternativen nach der Klassifikation wird der Suchraum stark vergrößert. Datenbanken in der Größenordnung von bis zu 100 MByte können dann nur mit Hilfe schneller Hash-Algorithmen in akzeptabler Zeit durchsucht werden. Diese

3<17>0100<50>0		37010050
	$\Longrightarrow$	
POSTB<A>NK <IL>T<SÖ><LI>N		POSTBANK KOELN

Abbildung 4.2: *Leseergebnis und Resultat des Datenbankabgleichs für die Bankleitzahl und den Banknamen. Zeichen in spitzen Klammern stellen mehrere Alternativen der Lese-Software dar.*

erlauben es, schon mit wenigen Operationen festzustellen, ob ein Eintrag in der Datenbank vorkommt. Im Sommer 1995 haben wir daher mit der Entwicklung der assoziativen Datenbank DACCORD begonnen. Das Ziel war dabei, eine hocheffiziente, portable Anwendung zu entwickeln, die auch auf Standard-PCs eine hohe Verarbeitungsgeschwindigkeit bei großen Datenbankvolumina erzielt.

Der Vergleich von Datenbankeinträgen mit den Leseergebnissen geschieht auf Basis der *Levenshtein-Distanz*, die ein Maß für die Abweichung zweier Zeichenketten ist. Dabei wird die erste Zeichenkette durch aufeinanderfolgende Einfüge- und Löschoperationen in die Zielzeichenkette überführt. Werden dann die verschiedenen Operationen mit Kosten belastet, so ergeben die aufsummierten Kosten die gesuchte Distanz. In unserer Arbeit wurde die Berechnung der Levenshtein-Distanz zu einem Einzelwortabgleich erweitert, bei dem zu jedem Wort des Quelltextes ein möglichst ähnliches Wort im Zieltext gesucht wird. Dieser wortweise Abgleich führte zu einer Halbierung der Anzahl der nicht erkannten Zeichenketten (Rejectrate) bei gleicher Fehlerquote.

Die aktuelle Version von DACCORD besitzt folgende Merkmale:

Bearbeitungsgeschwindigkeit: Auf einer Sparc-10 mit 60 MHz werden etwa 80 Formulare pro Sekunde korrigiert, auf einem Pentium-PC mit 133 MHz unter Windows etwa 70 Formulare pro Sekunde.

Speicherbedarf: Die Daten liegen komprimiert im Hauptspeicher. Beispielsweise können bei einem 64 Mbyte großen Hauptspeicher unter Windows NT über 110 MByte an Auftraggeber- und Empfängerdaten bearbeitet werden.

Zuverlässigkeit: Durch entsprechende Einstellung in einer ASCII-Konfigurationsdatei kann die Fehlerquote immer auf unter 0,1 % bei einer Rejectrate von weniger als 25 % gedrückt werden.

Portabilität: Alle Programmteile sind in ANSI-C geschrieben, so daß die Programme praktisch für jede Rechnerplattform zur Verfügung gestellt werden können. Zur Zeit ist das System auf Solaris, SunOS, Win32s und Windows NT implementiert.

Konfigurierbarkeit: Alle wesentlichen Parameter des Systems werden in ASCII-Konfigurationsdateien definiert und können problemlos den individuellen Anforderungen des Endkunden angepaßt werden.

DACCORD wird bereits von einer deutschen Großbank auf einem Parsytec-CC-Rechner unter Windows NT eingesetzt. Auf diesem Rechner werden pro Sekunde bis zu 10 Überweisungsformulare basierend auf Datenbanken mit ca. 1,8 Millionen Konten verarbeitet.

4.2 Handschriftenerkennung

Der bereits in den Vorjahren entwickelte Handschriftenklassifikator „COGNITUS“ wurde 1995 zu einem vollständigen System zur Verarbeitung hand- oder maschinengeschriebener Formulare ausgebaut. Die große Zuverlässigkeit und die hohe Verarbeitungsgeschwindigkeit von ca. drei Banküberweisungsformularen pro Sekunde auf einem Pentium-133 PC resultieren aus hochentwickelten Clusterungsverfahren, für deren Entwicklung unser Mitarbeiter Dr. Martin Neschen auf der Jahrestagung der Deutschen Arbeitsgemeinschaft für Mustererkennung 1995 mit dem DAGM-Preis ausgezeichnet wurde.

Die bereits verwendeten hierarchischen Clusterungsverfahren wurden weiter verbessert. So wird jetzt auf den ersten Verarbeitungsstufen das zu klassifizierende Zeichen nicht mehr bestimmten Zeichenklassen zugeordnet, sondern abstrakten Unterräumen zugeteilt. Dies führt, neben einer weiteren Beschleunigung des Verfahrens, zu besseren Resultaten, insbesondere bei mehrdeutigen Zeichen, da die Klassifikation erst auf tieferen Hierarchiestufen erfolgt.

Zu einer großen Verbesserung der Gesamtleistung hat auch die Einführung einer *dynamischen Segmentierung* geführt, da selbst auf Banküberweisungsformularen oft einzelne Zeichen überlappen bzw. in mehrere Fragmente zerfallen. Bei einer *dynamischen Segmentierung* wird der Text nicht

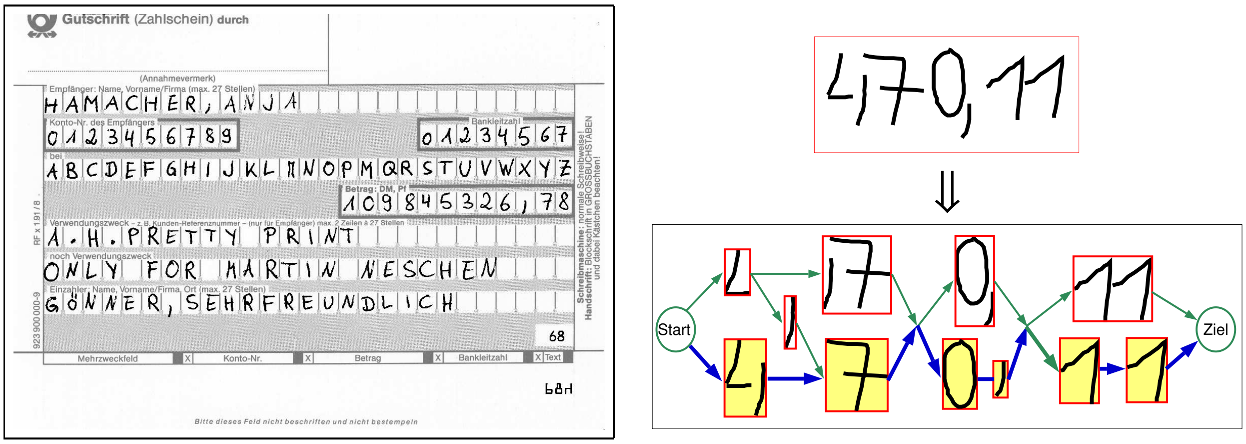

Abbildung 4.3: *Banküberweisungsformular und dynamische Segmentierung*

starr in einzelne Zeichen zertrennt, sondern es werden an jeder Stelle mehrere Trennvarianten vorgeschlagen. Jedes Zeichen wird dann einzeln klassifiziert, und anschließend wird mittels dynamischer Programmierung die beste Segmentierung ausgewählt. Hier zeigen sich die Vorteile unseres hierarchischen Klassifikators, da dieser schon nach wenigen Vergleichen sehr schnell eine Vorentscheidung treffen kann.

4.3 Bildverarbeitung

Qualitätskontrolle in der Lebensmittelindustrie

Die optische Kontrolle von Lebensmitteln stellt wegen der besonders hohen Variabilität des Aussehens der Waren ein schwieriges Problem in der Nahrungsmittelindustrie dar. Die Bedeutung anpassungsfähiger Algorithmen in der industriellen Qualitätskontrolle läßt sich an diesem Beispiel gut dokumentieren, da durch schnellen Produktwechsel ein auch von Laien handhabbares System notwendig ist. Hierzu ist bei der Einführung eines neuen Produktes immer eine „Einlern“-Phase mit dem Programm zu durchlaufen.

Bei einem bereits gelernten Menü verläuft das Klassifizierungsverfahren in drei Schritten. Im ersten Schritt werden durch das *Split-and-Merge-Verfahren* zusammenhängende Regionen ähnlicher Farbe gesucht. Diese werden dann mit den Farbprototypen der einzelnen Lebensmittel verglichen. Im letzten Schritt werden die Flächeninhalte und Positionen bewertet. Das Verfahren ist hochoptimiert und benötigt zur Verarbeitung eines einzelnen Bildes mit bis zu 1 MByte Daten auf einem Standard-PC lediglich eine Sekunde.

Da die Produktion jedoch oft umgestellt wird, ist ein einfach zu bedienendes Einlern-Programm in der Praxis von großer Bedeutung. Dieses sollte in Echtzeit Bilder verarbeiten und die notwendigen Benutzereingaben auf einfache „gut“ oder „schlecht“ Antworten beschränken. Dazu wurde ein mehrstufiges Verfahren entwickelt, das sich folgendermaßen zusammenfassen läßt:

Clusterung im Ortsraum: Genau wie in der Klassifizierung wird dazu ein *Split-and-Merge*-Verfahren eingesetzt, das die wesentlichen Bildinhalte der Bildfolge durch wenige Cluster beschreibt.

Clusterung im Farbraum: Mit Hilfe des *k-Means-Verfahrens* werden die Hauptfarben der Bildfolge bestimmt.

Generierung des Merkmalsraumes: Basierend auf den Nachbarschaftseigenschaften werden Merkmalsvektoren generiert und zusammen mit der durch den Benutzer erfolgten Klassifizierung abgespeichert.

Überprüfung der Konvergenz: Sind bereits einige Bilder klassifiziert, so werden weitere durch ein einfaches Nächste-Nachbar-Verfahren beurteilt. Stimmt die automatische Beurteilung mit den Benutzereingaben hinreichend oft überein, so ist das neue Menü erfolgreich eingelernt.

Das beschriebene Verfahren kann nach der Einlernphase sowohl Farb- als auch Anordnungsfehler selbständig erkennen. Zusammen mit dem Klassifizierungsmodul ist es auf einem Standard-PC unter Windows implementiert und leistet alle Berechnungen in Echtzeit. Somit sind auch Laien in der Lage, das System nach einem Produktwechsel vor Ort neu zu konfigurieren.

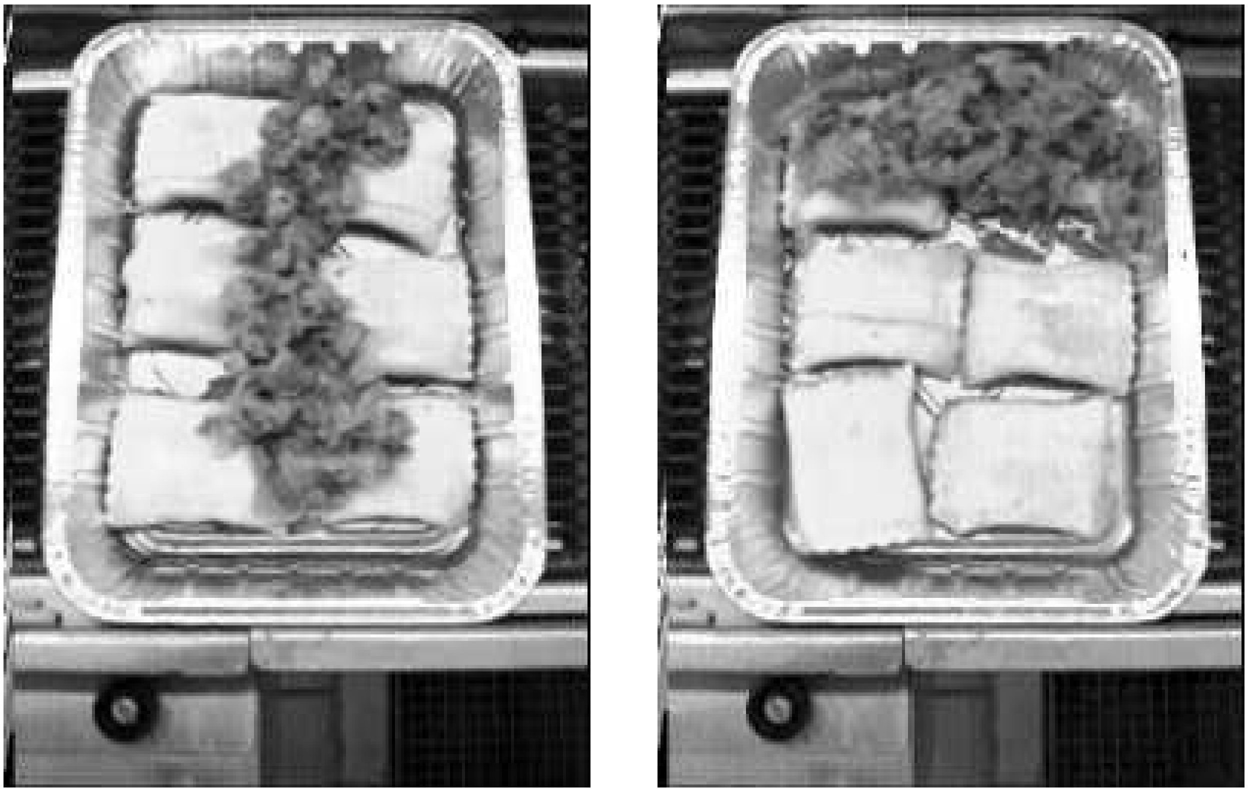

Abbildung 4.4: *Bild des Menüs „Maultaschen", rechts mit Anordnungsfehler*