

Kapitel 1

Einleitung

Gegenstand dieser Arbeit ist die Anwendung und Evaluierung von statistischen Verfahren für die rechnergestützte Erkennung von Mustern in Bildern und Bildsequenzen. Die verwendeten statistischen Verfahren basieren auf sogenannten Hidden-Markov-Modellen (HMM), die eine Erweiterung der Markov-Quellen darstellen. Die Hidden-Markov-Modelle sind, seit ihrer Einführung in dieses Gebiet Mitte der 70er Jahre, die dominierende Modellierungstechnik im Bereich der automatischen Spracherkennung geworden. Diese Dominanz ist vor allem dadurch zu erklären, daß die Markov-Modellierung auf die unsegmentierte Merkmalsequenz eines fließend gesprochenen Satzes angewendet werden kann. Anders formuliert, ist keine Vorsegmentierung in Einzelworte erforderlich. Eine solche Vorsegmentierung würde ihrerseits segmentierungstypische Fehler in das Gesamtsystem einbringen. Hidden-Markov-Modelle bieten also die Möglichkeit, einen fließend gesprochenen Satz in einem Schritt zu segmentieren und die Einzelkomponenten des Satzes (Worte, bzw. Phoneme) zu erkennen. Ein wichtiges Ziel dieser Arbeit ist es, dieses in der Spracherkennung entwickelte Prinzip auf Bilder und Bildsequenzen zu übertragen. Auf Bilder übertragen bedeutet dies etwa, ein Objekt aus einer Menge von definierten Klassen in einem gegebenen Bild aufzufinden und *simultan* das Objekt zu klassifizieren. In der vorliegenden Arbeit wird der integrierte Segmentierungs- und Erkennungsansatz beispielsweise auf handschriftliche Skizzen angewendet, in denen vordefinierte Symbole in einem ebenfalls von Hand schraffierten Hintergrund eingebettet sind (siehe auch Abb. 1.1). Die Erkenntnisse, die durch die Experimente auf diesen handschriftlich erstellten Mustern gewonnen wurden, konnten ferner genutzt werden, um Personen in Bildern von realen Szenen aufzufinden und zu identifizieren.

Ein offensichtliches Problem, welches bei der Anwendung von Hidden-Markov-Modellen auf Bildern auftritt, ist die inhärente eindimensionale Struktur dieser Modelle. In dieser Arbeit werden jedoch eine Vielzahl von effizienten Algorithmen für mehrdimensionale Mustererkennungsaufgaben, die auf den Hidden-Markov-Modellen basieren, vorgestellt. Dabei kommen vor allem Modellierungsmethoden zum Einsatz, die unter teilweiser Vernachlässigung von Nachbarschaftsbeziehungen höherdimensionale Daten modellieren. Dies

sind die sogenannten *pseudo* zwei- und dreidimensionalen HMMs, die erfolgreich auf Mustererkennungsaufgaben in Bildern und Bildsequenzen angewendet werden können.

1.1 Mustererkennung

Wie Eingangs erwähnt wurde, befaßt sich diese Arbeit mit statistischen Verfahren für die automatische Mustererkennung. Es ist somit sinnvoll zunächst festzulegen, was unter dem Begriff *Muster* verstanden werden soll und welche Realisierungen dieses Begriffs in dieser Arbeit verwendet werden. In [Jai00] findet sich die Definition eines Musters als *das Gegenteil von Chaos; es ist eine Entität, vage definiert, der ein Namen geben werden kann*. Die hier verwendeten Muster entstammen den Anwendungsgebieten Mensch-Maschine-Kommunikation sowie den Multimedia-Anwendungen. Im Speziellen wurden die folgenden Muster verwendet: handschriftlich eingegebene Piktogramme und Skizzen, Abbildungen von Werkzeugen, technische Zeichnungen, Personen in natürlichen Szenen, sowie Gesten. Insbesondere werden diese Muster in der Kombination mit Störeffekten, wie beispielsweise einer Teilüberdeckung oder einer Einbettung in eine komplexe Szene betrachtet und untersucht. Ein repräsentatives Beispiel für die hier untersuchten Muster ist in Abb. 1.1 dargestellt.



Abbildung 1.1: Beispiel für ein Muster, das in eine komplexe Szene integriert ist

Die Abbildung zeigt ein von Hand skizziertes Piktogramm, das Mensch oder Person bedeutet und welches sich in einer Umgebung befindet, die aus ähnlichen Konstruktionselementen, nämlich Strichen, zusammengesetzt ist, wie das Piktogramm selbst. Wenn nun die Aufgabe darin besteht, solche Bilder dahingehend zu analysieren, welches Piktogramm sie enthalten, dann muß der verwendete Algorithmus nicht nur das Piktogramm klassifizieren, sondern es auch innerhalb des Bildes auffinden. Dies kann mit Bezug auf die angegebene Definition auch beschrieben werden als die Aufgabe, das *Muster* im *Chaos* aufzufinden.

1.2 Integrierte Ansätze zur Segmentierung und Klassifizierung

Anhand von Abb. 1.1 wurde im letzten Abschnitt beschrieben, daß die Aufgabe, ein solches Bild dahingehend zu analysieren, welches Piktogramm es enthält, im wesentlichen aus zwei Unteraufgaben besteht: Zunächst ist das gegebene Bild in die Bestandteile *Piktogramm* und *Hintergrund* zu segmentieren und anschließend ist das gefundene Piktogramm zu klassifizieren.

Die Segmentierung ist ein erster, wichtiger Schritt in der Bildverarbeitung und kann definiert werden als ein Prozeß, der ein Bild in nicht überlappende Regionen unterteilt und dies derart, daß die einzelnen Regionen homogen sind und die Vereinigung zweier benachbarter Regionen keine homogene Region ergibt (aus [Pal93]). Ein Segmentierungsschritt wird zum Beispiel für die folgenden Anwendungen benötigt: beim Auffinden von Schlüsselworten in verrauschtem, eingescanntem Text, bei der Identifizierung von Personen in natürlichen Bildern, et cetera. Zusätzlich zu dem Segmentierungsschritt wird bei den genannten Anwendungen ein Klassifikationsschritt erforderlich sein. In vielen konventionellen Bildbearbeitungsansätzen werden der Segmentierungs- und der Klassifizierungsschritt weitgehend unabhängig voneinander entwickelt.

Diese Unterteilung in voneinander unabhängig entwickelte Verarbeitungsstufen, entsprechend des *Top-Down*-Ansatzes, führt in vielen Fällen zu suboptimalen Ergebnissen. Der Grund hierfür ist die Existenz eines informationstheoretischen Theorems, welches besagt, daß während der Informationsübertragung über verbundene Einzelstufen wichtige Informationen verloren gehen können. Formal ausgedrückt, gehorcht der Transinformationsverlust $\Delta I_{\Omega}(X, Y) = I(X; \Omega) - I(Y; \Omega)$ bezogen auf die Größe Ω über eine Kette von $N - 1$ Transformationsstufen der Gleichung

$$\Delta I_{\Omega}(X_1, X_N) = \sum_{i=1}^{N-1} [I(X_i; \Omega) - I(X_{i+1}; \Omega)] \quad (1.1)$$

Da die Größe $\Delta I_{\Omega}(X, Y)$ für jede Transformation $Y = f(X)$ stets nicht-negativ ist, kann der Transinformationsverlust einer Transformation in dieser Reihe nicht wiedergewonnen werden. Ein üblicher Ansatz um diesem Effekt zu begegnen, ist, die Größe ΔI_{Ω} für die einzelnen Transformationen individuell zu minimieren, was das Risiko birgt, daß in einer frühen Stufe der Kette die Information in eine Form transformiert wird, die es den folgenden Stufen erschwert, den eigenen Informationsverlust zu minimieren (siehe auch [Plu91]). Um diese Nachteile der voneinander unabhängig entwickelten Verarbeitungsstufen zu vermeiden, wurden alternative, integrierte Verfahren entwickelt (siehe z.B. [Fri97]). Im Bereich der automatischen Spracherkennung ist ein Ansatz überaus populär geworden, der im nächsten Abschnitt vorgestellt wird und der die integrierte Segmentierung und Klassifizierung von Worten in fließend gesprochenen Sätzen ermöglicht.

1.3 Statistische Mustererkennung mit Hidden-Markov-Modellen

Wie bereits erwähnt wurde, ist die populärste Anwendung der Hidden-Markov-Modelle der Bereich der Zeitreihen-Klassifikation und hier insbesondere die rechnergestützte Spracherkennung. In den letzten Jahren wurde diese Methode ebenfalls in der Online-Handschrifterkennung erfolgreich eingesetzt (siehe z.B. [Nat93]), welches ein Anwendungsgebiet ist, das große Ähnlichkeit mit der Spracherkennung aufweist, da ebenfalls Zeitreihen verarbeitet werden. Erkennungsaufgaben, bei denen das Muster sich nicht über die *Zeit*, sondern mit dem *Ort* verändert, sind ebenfalls mit Hidden-Markov-Modellen gelöst worden. Solche Aufgaben sind beispielsweise die Erkennung von Gesichtern ([Sam94b, Eic99b, Eic00]), die Abfrage von Bilddatenbanken ([Lin97, Mul98a]) oder handgezeichneten Piktogrammen ([Mul98b, Mul99f]). Die hierbei zu verarbeitenden Daten sind zweidimensional und somit reichen die aus der Spracherkennung bekannten Verfahren nicht mehr aus. Erst die Einführung sog. *pseudo* zweidimensionaler Hidden-Markov-Modelle ermöglichte die Bearbeitung solcher Mustererkennungsaufgaben.

Zeitlich sich verändernde Sequenzen von Bildern stellen ein dreidimensionales Mustererkennungsproblem dar, und daher werden in der vorliegenden Arbeit die pseudo zweidimensionalen Hidden-Markov-Modelle erweitert auf pseudo dreidimensionale HMMs. Mustererkennungsaufgaben, die mit Sequenzen von Bildern arbeiten, sind z.B. die Erkennung von Gesten ([Rig96]), sowie die Indexierung von Videoaufnahmen ([Eic99a]). Die Verwendung von HMMs in den oben genannten Anwendungen hat die folgenden Vorteile: Die Modelle werden mit Beispieldaten trainiert und können somit leicht auf eine gegebene Mustererkennungsaufgabe adaptiert werden. Es ist weiterhin möglich, heterogene Merkmale mit einem einzelnen Modell zu verarbeiten. So besteht der Merkmalvektor in der Spracherkennung oftmals aus cepstralen Merkmalen und zusätzlich z.B. aus der Energie eines Sprachsignals. Die Eigenschaft der HMMs, die eine zentrale Rolle in dieser Arbeit einnimmt, ist, auf integrierte Weise Segmentierungen zu finden und ein Klassifizierungsergebnis zu erzeugen. Dieser Ansatz wurde in der vorliegenden Arbeit erfolgreich auf Bildern und Bildsequenzen eingesetzt. Dabei werden neben den pseudo zwei- und dreidimensionalen Hidden-Markov-Modellen auch die aus der Spracherkennung bekannten eindimensionalen Modelle verwendet, die beispielsweise auf elegante Weise die Umrißkurve einer geometrischen Form beschreiben können. Eingesetzt werden die HMMs dabei bei Problemen der handschriftlichen Piktogrammerkennung, beim Auffinden von Bildern in Bilddatenbasen, der Personenerkennung und -detektion und in der Gestikerkennung.

1.4 Gliederung der Arbeit

Die vorliegende Arbeit ist wie folgt gegliedert: Im Anschluß an diese Einführung führt Kapitel 2 in die Theorie der eindimensionalen Hidden-Markov-Modelle ein. Hidden-Markov-Modelle stellen einen doppelt stochastischen Prozeß dar und werden aufbauend auf den zugrundeliegenden sog. Markov-Quellen beschrieben. Anschließend wird in Kapitel 3 eine neuartige rotationsinvariante Modellierung von Piktogrammen und Objektformen mit eindimensionalen Hidden-Markov-Modellen vorgestellt. Die integrierten Segmentierungs- und Klassifizierungseigenschaften der HMMs werden dabei verwendet, um die Orientierung eines gedrehten Objektes herauszufinden und das Objekt zu erkennen.

Kapitel 4 führt in die Theorie der statistischen, zweidimensionalen Modellierung ein und beschreibt zunächst die Markovschen Wahrscheinlichkeitsfelder (engl. Markov-Random-Fields). Aus diesen gehen durch Einführung einer kausalen Abhängigkeit und eines zweiten stochastischen Prozesses die zweidimensionalen Hidden-Markov-Modelle hervor. Die Vereinfachungen, die gemacht werden müssen, um von diesen Modellen zu den pseudo zweidimensionalen HMMs zu gelangen, sind ebenfalls Gegenstand des Kapitels 4. Diese P2DHMMs werden im darauf folgenden Kapitel 5 verwendet, um zweidimensionale Muster in komplexen Umgebungen aufzufinden und zu klassifizieren. Dies erfolgt durch ein neuartiges Verfahren, das P2DHMMs in Kombination mit an den Bildkontext adaptierten Umgebungszuständen verwendet. Auch bei diesem Verfahren erfolgt die Klassifikation und die Segmentierung auf integrierte Weise in einem algorithmischen Schritt.

Kapitel 6 beschreibt die neuartigen pseudo dreidimensionalen Hidden-Markov-Modelle, die für die Erkennung von Bildsequenzen eingesetzt werden können. Das Kapitel gibt eine theoretische Einführung in diese Modelle und beschreibt Experimente, die auf einer Gestendatenbasis durchgeführt wurden. Dabei werden die P3DHMMs eingesetzt, um 12 verschiedene Gesten zu klassifizieren. Es wird ferner ein Ausblick auf einen integrierten Ansatz gegeben, der die gemeinsame Segmentierung und Klassifikation von dreidimensionalen Mustern ermöglicht. Den Abschluß der Arbeit bildet eine Zusammenfassung der erzielten Ergebnisse.