

# Statistische Modellierung, Klassifikation und Lokalisation von Objekten

Joachim Hornegger

Lehrstuhl für Mustererkennung (Informatik 5),  
Martensstraße 3, 91058 Erlangen,  
email: [hornegger@informatik.uni-erlangen.de](mailto:hornegger@informatik.uni-erlangen.de)

**Zusammenfassung** In dieser Arbeit wird ein neuer Ansatz zur modellbasierten Erkennung und Lokalisierung dreidimensionaler Objekte in Grauwertbildern vorgestellt. Entgegen den üblicherweise verwendeten geometrischen Beschreibungen sind die hier eingesetzten Objektmodelle Wahrscheinlichkeitsdichten. Diese erlauben die Berechnung eines Dichtewertes für ein beobachtetes Bild und liefern damit ein wahrscheinlichkeitstheoretisches Maß, das die Grundlage für die Realisierung statistischer Klassifikatoren darstellt. Neben der allgemeinen Struktur von Modelldichten werden Verfahren zur automatischen Modellgenerierung aus Beispielaufnahmen sowie Algorithmen zur Lokalisierung und zur Klassifikation erläutert. Die benötigten mathematischen Hilfsmittel sind Methoden zur Parameterschätzung wie beispielsweise die Maximum-Likelihood-Schätzung und der Expectation-Maximization-Algorithmus und Verfahren zur effizienten Optimierung kontinuierlicher Funktionen. Die experimentelle Evaluierung des entwickelten Ansatzes und die Anwendung auf andere Problemstellungen der Mustererkennung unterstreichen die Vorteile und Tragweite statistischer Klassifikatoren.

## 1 Einführung

Die Mustererkennung beschäftigt sich mit der Entwicklung von Verfahren zur Analyse und Interpretation von Sensordaten und findet damit in den unterschiedlichsten Bereichen der Industrie und der Forschung potentielle Anwender. So werden Algorithmen der Mustererkennung zur Erkennung von Sprache oder auch zur Analyse von Bilddaten eingesetzt. Mit zunehmender Rechenleistung von Standardhardware gibt es immer mehr Anwendungen, die Mustererkennungsalgorithmen einsetzen. Beispielsweise gelten Belegleser in Banken oder Scanner an Supermarktkassen bereits heute vielerorts als selbstverständlich. Wesentliche Voraussetzungen für die Produktreife von Algorithmen der Mustererkennung sind kostengünstige, effiziente und robuste Lösungen, die die Akzeptanzschwelle der Benutzer überschreiten. Letztere ist ziemlich hoch anzusetzen, da viele Aufgaben der Mustererkennung — wie zum Beispiel das Lesen handgeschriebener Schriftzeichen — vom Menschen problemlos und mit großer Zuverlässigkeit gelöst werden.

Das Rechnersehen, mit dem sich der folgende Beitrag befaßt, zählt zu einem Teilgebiet der Mustererkennung, das alle Aktivitäten umfaßt, die mit der

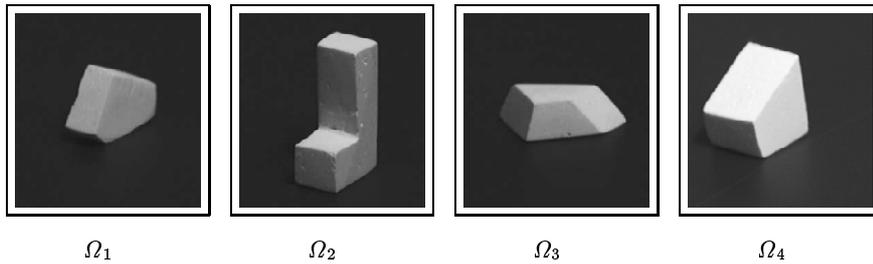
Verarbeitung visueller Daten im Zusammenhang stehen: Die rechnergestützte Qualitätskontrolle, die Detektion und Verfolgung sich bewegender Objekte, die Analyse medizinischer Bilddaten oder die bereits erwähnte Handschriftenerkennung sind typische Anwendungsbeispiele des Rechnersehens. Zu den schwierigsten Problemen des Rechnersehens zählt die Erkennung und Lokalisierung dreidimensionaler Objekte unter Verwendung von Grauwert- oder Farbbildern, die mit einer Standard-Videokamera aufgenommen werden. Bis heute gilt dieses Problem nur für Spezialfälle als gelöst, wobei in vielen Anwendungen die verfügbaren Algorithmen hinter den Erwartungen und Anforderungen der Industrie liegen. Ein Verfahren, das die Klassifikation und Lokalisierung beliebiger Objekte ermöglicht, gibt es bisher nicht. Das Problem der 3D-Objekterkennung ist somit immer noch von großem wissenschaftlichem Interesse, und es besteht in diesem Teilbereich der Mustererkennung noch ein erheblicher Forschungsbedarf.

Als langfristiges Ziel der 3D-Bildverarbeitung wird die Entwicklung von Verfahren betrachtet, die es dem Benutzer erlauben, ein Objekt vor eine robotergeführte Kamera zu legen. Der Rechner erlernt dann selbständig das Objekt und besorgt sich die Information, die für die Erkennung sowie für die Lokalisierung benötigt wird. Die 2D-Ansichten des 3D-Objekts sowie das statistische Verhalten beobachteter Bildmerkmale akquiriert der Computer selbständig. Eine manuelle Unterstützung oder die Bereitstellung eines CAD-Modells soll entfallen. Die folgenden Abschnitte führen ein theoretisches Konzept ein, das sich an dieser Zielsetzung orientiert und diesbezüglich einen wesentlichen Beitrag leistet.

Die Arbeit untergliedert sich insgesamt in acht Teile: Nach den einleitenden Bemerkungen führt der nächste Abschnitt die allgemeinen Grundkonzepte statistischer Klassifikatoren für die Bildverarbeitung ein. Aufbauend darauf werden unterschiedliche Möglichkeiten zur stochastischen Modellierung von Objekten vorgestellt. Die zum automatischen Training und zur Lokalisierung sowie Klassifikation erforderlichen Algorithmen sind Gegenstand der Abschnitte 4 und 5. Einen Eindruck der Leistungsfähigkeit der entwickelten Verfahren vermitteln die experimentellen Ergebnisse in Abschnitt 6. Ein kurzer Einblick in weitere Einsatzmöglichkeiten der vorgestellten statistischen Modellierung sowie einige Schlußbemerkungen runden die Arbeit ab.

## 2 Statistische Objekterkennung

Sensoren liefern im Kontext der Mustererkennung sogenannte *Muster*. Aus Gründen der Datenreduktion und der Effizienzsteigerung werden aus Mustern *Merkmale* berechnet, die eine *Klassifikation* der beobachteten Muster zulassen. Bei der Klassifikation werden Merkmale eines Musters einer von  $K$  Musterklassen zugeordnet. Ein typisches Klassifikationsproblem ist die Erkennung von Objekten. Abbildung 1 zeigt vier unterschiedliche Polyeder. Jedes dieser Objekte definiert eine Musterklasse, die im weiteren Verlauf mit  $\Omega_1$ ,  $\Omega_2$ ,  $\Omega_3$  und  $\Omega_4$  bezeichnet werden. Die Klassifikationsaufgabe besteht beispielsweise in der Erkennung dieser vier Klassen in beliebigen Grauwertbildern bei variierender Objektlage, bei unterschiedlicher Beleuchtung oder auch bei heterogenem Hin-



**Abbildung 1.** Polyedrische 3D-Objekte

tergrund. Abbildung 2 (links) zeigt ein Beispiel für ein Muster, das klassifiziert werden soll. Als Merkmale werden hierfür Eckpunkte und Vertices herangezogen, die ein Segmentierungsverfahren liefert [3]. Vom Klassifikator wird erwartet, daß er sich ausgehend von den in Abbildung 2, rechts, dargestellten Segmentierungsdaten für die Musterklasse  $\Omega_2$  entscheidet.

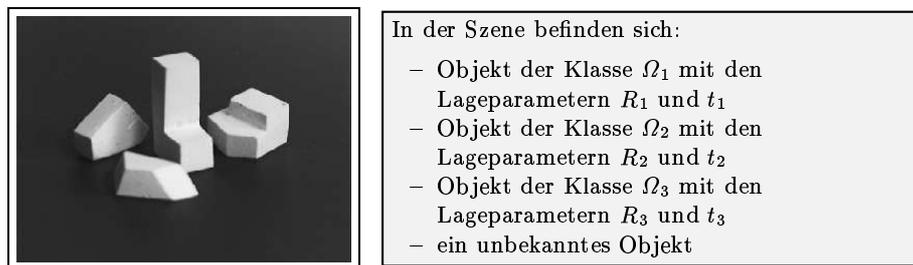
Grundsätzlich gibt es unterschiedliche Möglichkeiten, die Klassifikation vorzunehmen. Bei modellbasierten Ansätzen zur Objekterkennung wird die Ähnlichkeit zwischen Prototypen, den sogenannten *Modellen*, und den beobachteten Merkmalen gemessen. Die Entscheidung fällt dann für diejenige Musterklasse, deren Modell die größte Ähnlichkeit zum beobachteten Muster aufweist. Hierfür können ein Abstandsmaß minimiert oder Wahrscheinlichkeiten maximiert werden. Statistische Klassifikatoren setzen die Bayes-Entscheidungsregel ein und entscheiden sich für diejenige Klasse  $\Omega_\lambda$ , die für eine gegebene Beobachtung  $\mathbf{O}$  die *a-posteriori* Wahrscheinlichkeit  $p(\Omega_\kappa | \mathbf{O})$  maximiert. Bayes-Klassifikatoren zeichnen ihre Optimalität aus. Unter bestimmten Voraussetzungen minimiert die Bayes-Entscheidungsregel die Wahrscheinlichkeit für eine Fehlklassifikation. Gelingt also die Lösung einer Klassifikationsaufgabe durch die Realisierung eines Bayes-Klassifikators, so stellt ein Ergebnis der Entscheidungstheorie sicher, daß es keinen anderen Klassifikator geben kann, der eine geringere Fehlerrate aufweist. Wesentlich für den Bayes-Klassifikator ist die Kenntnis der *a-posteriori* Wahrscheinlichkeiten der einzelnen Musterklassen.

Gesetzt den Fall, ein erkanntes Objekt soll mit einem Robotergreifer aufgenommen werden, so ist zusätzlich zur Klassifikation auch die *Lokalisierung* des Objektes erforderlich. Neben der Klassennummer muß auch die Position und Orientierung des Objektes bezüglich eines geeignet zu definierenden Referenzkoordinatensystems bestimmt werden. Hinzu kommt, daß in den meisten Anwendungen von Mehrobjektszenen ausgegangen werden kann. In aller Regel wird sich nicht ein einzelnes Objekt vor homogenem Hintergrund in der Szene befinden, sondern mehrere Objekte bei heterogenem Hintergrund.

Aus formaler Sicht beschreibt ein Verfahren zur Erkennung und Lokalisierung von Objekten somit die Transformation einer Bildmatrix in eine symbolische Beschreibung. Abbildung 3 zeigt ein einfaches Beispiel für eine Szene mit mehreren Objekten und der gewünschten symbolischen Beschreibung, die neben



**Abbildung 2.** Ein Grauwertbild und das korrespondierende Ergebnis eines Eckendetektors



**Abbildung 3.** Transformation eines Grauwertbildes in eine symbolische Beschreibung

den Klassennummern auch die Lageparameter der einzelnen Objekte beinhaltet. Das Ziel statistischer Verfahren zur Objekterkennung besteht nun darin, diese Transformation mittels probabilistischer Methoden durchzuführen. Von den Algorithmen, die die Berechnung der symbolischen Beschreibung übernehmen, wird erwartet, daß sie robust, schnell, zuverlässig und flexibel sind und nach Möglichkeit keine Spezialhardware erfordern.

Im Gegensatz zu Verfahren für die Objekterkennung und für die Szenenanalyse haben Algorithmen zur automatischen Spracherkennung bereits Produktreife erreicht und werden kommerziell vertrieben. Der Grund hierfür ist zweifellos dem Einsatz statistischer Verfahren und der Realisierung von Bayes-Klassifikatoren zuzuschreiben. Auch in der Bildverarbeitung zählen statistische Verfahren zu etablierten Methoden, jedoch blieb der konsequente Einsatz probabilistischer Verfahren zur Objekterkennung bislang aus. Statistische Methoden beschränken sich größtenteils auf Problemstellungen der Bildvorverarbeitung und die statistische Modellierung möglicher Störgrößen bei der Bildaufnahme [6, 7, 8], obgleich es zahlreiche Argumente gibt, die den Einsatz von Statistik zur Erkennung von Objekten sinnvoll erscheinen lassen [4]:

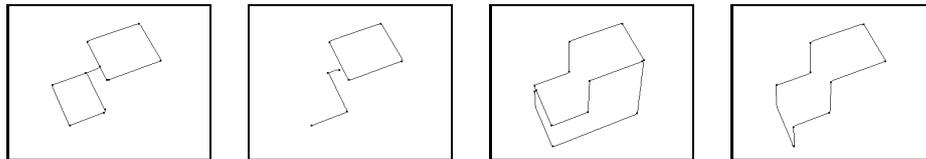
- Der ungebrochene Siegeszug statistischer Verfahren in der automatischen Spracherkennung motiviert, ähnliche Ansätze auch für die Bildverarbeitung zu erarbeiten.

- Die Entscheidungstheorie gewährleistet die Optimalität des Bayes-Klassifikators.
- Störungen, die beispielsweise durch variierende Beleuchtung oder durch Sensorrauschen hervorgerufen werden, erfordern eine adäquate mathematische Modellierung.
- Verfügbare theoretische Ergebnisse der mathematischen Statistik erleichtern möglicherweise die Modellierung und den Entwurf von Algorithmen zur Lokalisation und Klassifikation. Die mathematische Statistik verfügt über zahlreiche Forschungsergebnisse zur Modellierung und zur Parameterschätzung. Beispiele für Modelle sind Markovketten oder Markov-Zufallsfelder [6]. Zur Berechnung der freien Parameter der Dichtefunktionen können zum Beispiel die Maximum-Likelihood (ML) oder die Maximum-a-posteriori (MAP) Schätzung angewendet und theoretische Aussagen über die Robustheit von Schätzwerten eingesetzt werden.

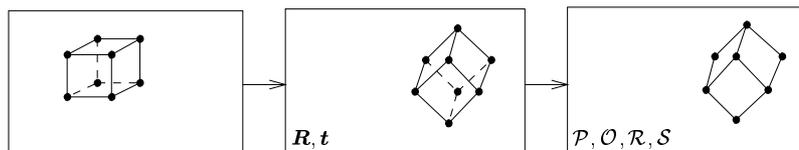
Das zentrale Problem bei der statistischen Klassifikation und Lokalisation von Objekten ist die Definition geeigneter Wahrscheinlichkeitsdichten, die die Berechnung von *a-posteriori* Wahrscheinlichkeiten erlauben. Im Gegensatz dazu können die *a-priori* Wahrscheinlichkeiten der Musterklassen aus einer klassifizierten Stichprobe über relative Häufigkeiten geschätzt werden. Da nicht offensichtlich ist, wie die statistischen Modelle — im folgenden auch *Modelldichten* genannt — zu konstruieren sind, ist es hilfreich, zunächst den Prozeß der Bildentstehung aus geometrischer sowie statistischer Sicht zu betrachten.

Ein 3D-Objekt kann durch eine Menge von Merkmalen repräsentiert werden. Beispielsweise eignen sich Linien oder Eckpunkte zur Charakterisierung polyedrischer Objekte. Durch Rotation und Translation im Raum werden diese geometrischen Merkmale transformiert, d.h., eine affine Abbildung gegeben durch eine Rotationsmatrix  $\mathbf{R}$  und einem Translationsvektor  $\mathbf{t}$  transformiert die Merkmale im *Modellraum*. Da in der Bildebene projizierte Modellmerkmale beobachtet werden, müssen die dreidimensionalen Modellmerkmale ins Zweidimensionale projiziert werden (Projektion  $\mathcal{P}$ ). Aufgrund der Tatsache, daß geometrische Merkmale wie zum Beispiel Eckpunkte eines Polyeders aus verrauschten Grauwertbildern bei variierender Beleuchtung berechnet werden, sind die beobachteten Merkmale der Bildebene fehlerhaft und weisen ein probabilistisches Verhalten auf (Rauschoperator  $\mathcal{R}$ ). Die Merkmale befinden sich nicht exakt an der Stelle, an der sie bei einer fehlerfreien Aufnahme und Segmentierung wären. Segmentierungsfehler führen dazu, daß einige Merkmale fehlen oder weitere Merkmale hinzukommen, für die es keine geometrische Entsprechung im Modell gibt (Segmentierungsoperator  $\mathcal{S}$ ). Abbildung 4 zeigt mehrere Segmentierungsergebnisse von Bildern, die bei konstanter Objektlage und variierender Beleuchtung aufgenommen wurden. Die sogenannte Selbstverdeckung  $\mathcal{O}$  bewirkt hier, daß nicht alle Modellmerkmale korrespondierende Merkmale im Bild haben.

Zusammenfassend stellt sich somit heraus, daß die einzelnen Schritte von der ursprünglichen 3D-Geometrie bis zur Transformation in den Bildraum innerhalb des statistischen Objektmodells repräsentiert werden müssen (vgl. Abbildung 5). Als Grundlage für die folgenden Abschnitte dient die Hypothese, daß



**Abbildung4.** Segmentierungsergebnisse bei variierender Beleuchtung



**Abbildung5.** Vom 3D-Modell zur 2D-Aufnahme

bereits die 3D-Merkmale als Zufallsgrößen interpretiert werden und die beobachteten Bildmerkmale aus einer Transformation von Zufallsgrößen hervorgehen. Stochastische Objektmodelle müssen diesen Überlegungen zufolge folgende Komponenten umfassen:

- die statistische Charakterisierung einzelner Merkmale,
- die Abhängigkeit der beobachteten Merkmale von der Objektlage,
- das verwendete Kameramodell, das die Projektion der 3D-Welt in die Bildebene festlegt, und
- die Zuordnung von Bild- und Modellmerkmalen.

Die schrittweise Entwicklung statistischer Modelle, die diesen Anforderungen gerecht werden, skizziert der folgende Abschnitt.

### 3 Statistische Objektmodellierung

Der Schwerpunkt bei der Modellierung von Objekten zu Zwecken der Erkennung und Lokalisierung liegt nicht notwendigerweise auf einer detailgetreuen Rekonstruktion der Objektgeometrie, sondern vielmehr auf der Entwicklung trennscharfer Objektbeschreibungen, die eine möglichst zuverlässige Klassifikation und Lokalisierung sicherstellen. Erwartungen, die mit einem Objektmodell verbunden werden, sind eine für die Erkennung ausreichende Modellierung von Verdeckungserscheinungen und Segmentierungsfehlern sowie die Möglichkeit, Hintergrundmerkmale adäquat beschreiben zu können. In einer Aufnahme können sich Objekte befinden, die keine Elemente der Modellbasis sind. Trotzdem sollte der Klassifikator in der Lage sein, die bekannten Objekte zu erkennen und

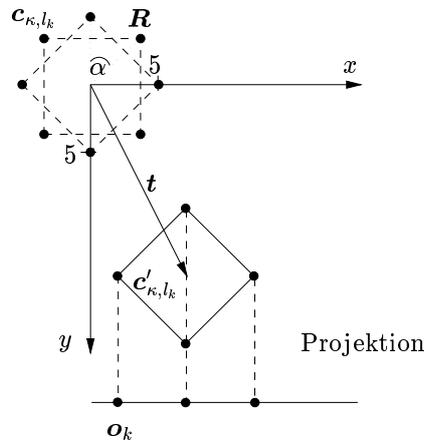
zu lokalisieren. Neben diesen Anforderungen, die aus Sicht der Bildverarbeitung unerlässlich sind, muß die Modelldichte zudem die Schätzung der Modellparameter aus Beispielaufnahmen sowie die Verwendung der Bayes-Entscheidungsregel erlauben.

### 3.1 Statistische Modellierung einzelner Merkmale

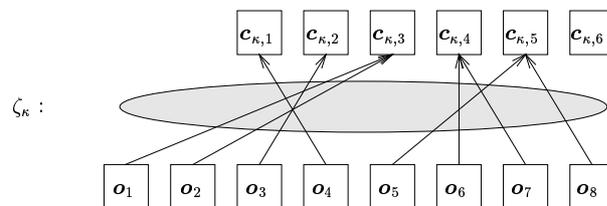
Die statistische Modellierung einzelner Merkmale kann unter Verwendung einzelner Dichtefunktionen erfolgen, was anhand eines einfachen Beispiels anschaulich erläutert wird. Um das Grundprinzip zu erklären, beschränkt man sich auf den Fall, nur 1D-Projektionen zweidimensionaler Objekte zu betrachten.

Abbildung 6 zeigt ein Quadrat, das die Musterklasse  $\Omega_\kappa$  definiert. Als Merkmale  $\mathbf{c}_{\kappa, l_k}$  mit  $l_k \in \{1, 2, 3, 4\}$  werden die Eckpunkte betrachtet. Jeder Eckpunkt  $\mathbf{c}_{\kappa, l_k}$  wird nun als eine zweidimensionale Zufallsgröße interpretiert. Das probabilistische Verhalten der Zufallsgrößen wird jeweils durch eine Dichtefunktion  $p(\mathbf{c}_{\kappa, l_k} | \mathbf{a}_{\kappa, l_k})$  mit den Parametern  $\mathbf{a}_{\kappa, l_k}$  beschrieben. Bei normalverteilten Punktmerkmalen bedeutet das beispielsweise, daß mit jedem Punkt eine zweidimensionale Normalverteilung assoziiert wird, die wiederum durch Mittelwertvektoren und Kovarianzmatrizen definiert werden. Die Rotation  $\mathbf{R}$  um den angegebenen Winkel  $\alpha$  und die Translation um den Vektor  $\mathbf{t}$  induziert eine Transformation der Eckpunkte und damit der Zufallsgrößen. Die Dichtefunktionen der transformierten Zufallsgrößen können unter Verwendung des Transformationssatzes für Dichten berechnet werden. Damit verfügen die Dichtefunktionen  $p(\mathbf{c}_{\kappa, l_k} | \mathbf{a}_{\kappa, l_k}, \mathbf{R}, \mathbf{t})$  der einzelnen Merkmale über zwei Arten von Parametern: einerseits sind die Parameter  $\mathbf{a}_{\kappa, l_k}$  enthalten, die das probabilistische Verhalten in der Geometrie widerspiegeln, andererseits liegen die Parameter  $\mathbf{R}$  und  $\mathbf{t}$  vor, welche die Transformation der Merkmale beschreiben und die einzelnen Dichtefunktionen koppeln. Bei der Projektion der 2D-Punkte in die Bildebene geht eine Dimension der ursprünglichen 2D-Zufallsvektoren verloren. Beobachtet werden lediglich 1D-Zufallsgrößen  $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_m$ , deren probabilistische Eigenschaften sich aus den Dichtefunktionen der 2D-Modellmerkmale ableiten lassen. In der Wahrscheinlichkeitsrechnung erfolgt die Elimination einer Dimension (in diesem Beispiel die  $y$ -Koordinate) durch Marginalisierung, d.h., die Berechnung der Randdichten liefert  $p(\mathbf{o}_k | \mathbf{a}_{\kappa, l_k}, \mathbf{R}, \mathbf{t})$  zu den einzelnen projizierten Merkmalen  $\mathbf{o}_k, k = 1, 2, \dots, m$ .

Die skizzierten Schritte werden für jedes Merkmal durchgeführt, und folglich steht ein probabilistisches Maß für das Auftreten bestimmter 1D-Punkte unter Vorgabe einer Rotation und Translation zur Verfügung. Für eine Menge paarweise statistisch unabhängiger Beobachtungen  $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_m\}$  hat das zur Folge, daß sich die Dichte der beobachteten Merkmale aus dem Produkt einzelner Wahrscheinlichkeitsdichten berechnen läßt. Wichtig hierbei ist die Kenntnis, welches beobachtete Merkmal zu welchem Modellmerkmal korrespondiert.



**Abbildung6.** Statistische Modellierung transformierter Merkmale



**Abbildung7.** Diskrete Zuordnung von Bild- und Modellmerkmalen

### 3.2 Statistische Modellierung der Zuordnung

Die diskrete Zuordnungsfunktion  $\zeta_{\kappa}$  bildet Bild- auf Modellmerkmale der Musterklasse  $\Omega_{\kappa}$  ab, die mit  $c_{\kappa, 1}, c_{\kappa, 2}, \dots, c_{\kappa, n_{\kappa}}$  bezeichnet werden. Abbildung 7 veranschaulicht ein Beispiel für eine mögliche Zuordnung bei acht beobachteten Merkmalen und sechs Modellmerkmalen. Die Zuordnung der Bild- und Modellmerkmale ist in aller Regel unbekannt und wird in den meisten Ansätzen zur Objekterkennung mit geometrisch basierten Verfahren berechnet. Die Realisierung eines statistischen Klassifikators schließt den Einsatz von derartigen Verfahren zur Berechnung der Zuordnung aus, da ein durchgehender statistischer Ansatz angestrebt wird. Die statistische Modellierung der Zuordnung ist somit erforderlich. Hierzu wird mit jeder Zuordnung  $\zeta_{\kappa}$  eine diskrete Funktion assoziiert, die die beobachteten Merkmale auf die Indizes der korrespondierenden Modellmerkmale abbildet. Folglich induziert jede Zuordnung  $\zeta_{\kappa}$  einen diskreten Vektor, der als Zufallsgröße aufgefaßt werden kann. Diesem Vektor kann dann eine diskrete Wahrscheinlichkeit zugeordnet werden. Der Zufallsvektor für das obige Beispiel

ist somit  $\zeta_\kappa = (3, 3, 2, 1, 5, 4, 4, 5)^T$  mit der diskreten Wahrscheinlichkeit  $p(\zeta_\kappa)$ . Aufgrund des stochastischen Modells müssen die einzelnen Wahrscheinlichkeiten die Bedingung  $\sum_{\zeta_\kappa} p(\zeta_\kappa) = 1$  erfüllen.

### 3.3 Konstruktion statistischer Objektmodelle

Bei bekannter Zuordnung  $\zeta_\kappa$  läßt sich die Dichtefunktion für eine Menge beobachteter Merkmale in der Form

$$p(\mathbf{O}, \zeta_\kappa | \mathbf{B}_\kappa, \mathbf{R}, \mathbf{t}) = p(\zeta_\kappa) p(\mathbf{O} | \zeta_\kappa, \{\mathbf{a}_{\kappa,1}, \mathbf{a}_{\kappa,2}, \dots, \mathbf{a}_{\kappa,n_\kappa}\}, \mathbf{R}, \mathbf{t}) \quad (1)$$

schreiben. Da die Zuordnungsfunktion in der Regel unbekannt ist, wird hier — wie im Falle der Projektion von Modellmerkmalen in die Bildebene — über die nicht beobachtete Zufallsgröße summiert und die Randdichte berechnet. Die Modelldichte für eine beobachtete Merkmalmenge  $\mathbf{O}$  ist somit

$$p(\mathbf{O} | \mathbf{B}_\kappa, \mathbf{R}, \mathbf{t}) = \sum_{\zeta_\kappa} p(\zeta_\kappa) p(\mathbf{O} | \zeta_\kappa, \{\mathbf{a}_{\kappa,1}, \mathbf{a}_{\kappa,2}, \dots, \mathbf{a}_{\kappa,n_\kappa}\}, \mathbf{R}, \mathbf{t}) \quad (2)$$

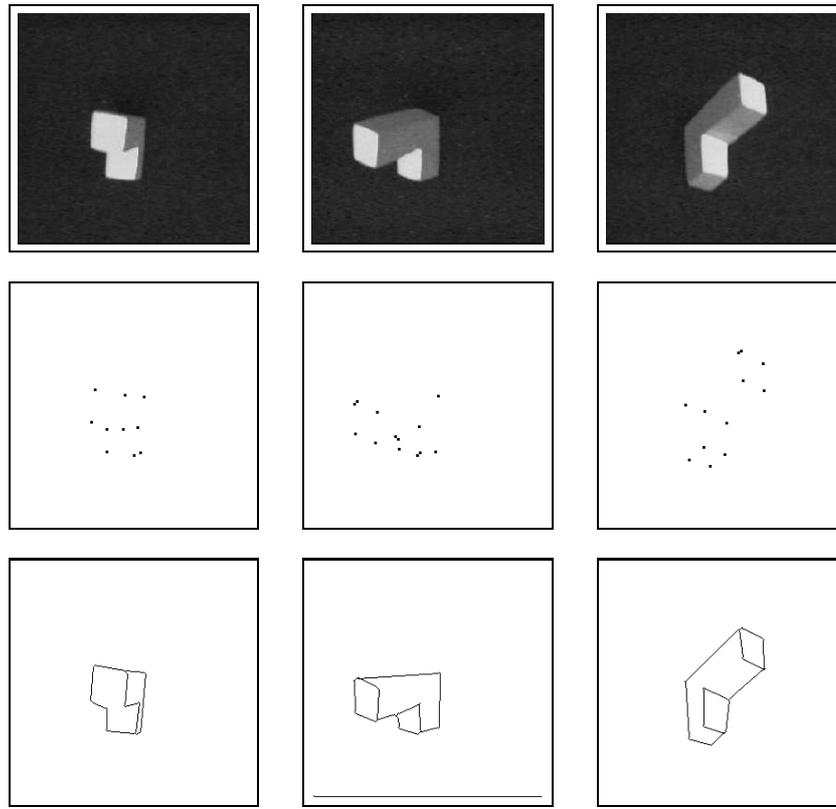
Die Anzahl der Multiplikationen und Additionen zur Auswertung der Modelldichte wird — unabhängig von den gewählten Merkmalen — beschränkt durch  $\mathcal{O}(m n_\kappa^m)$ . Durch die Einführung statistischer Abhängigkeiten beschränkter Ordnung läßt sich die Berechnungskomplexität jedoch deutlich reduzieren. In [4] wird gezeigt, daß die Komplexität für statistisch abhängige Zuordnungen  $g$ -ter Ordnung die Anzahl der Rechenschritte durch  $\mathcal{O}(m n_\kappa^{g+1})$  beschränkt wird. Des weiteren belegt eine genauere Analyse von (2), daß sich aus obiger Modellierung mit  $g = 0$  Mischungsverteilungen und mit  $g = 1$  Hidden-Markov-Modelle ergeben.

## 4 Erlernen von Objekten

Die Modelldichte wird charakterisiert durch die statistische Modellierung der Zuordnungsfunktion und der einzelnen Merkmale. Die Struktur der Modelldichte variiert von Objekt zu Objekt. Die Anzahl der modellierten Merkmale hängt von der Anzahl der Merkmale des Objektes ab. Stehen die einzelnen Komponenten der Modelldichten fest, so reduziert sich die automatische Modellgenerierung auf die Berechnung der *Modellstruktur* und der *Modellparameter*. Zur automatischen Berechnung der Modellstruktur sind bis heute lediglich heuristische Algorithmen bekannt, weshalb sich dieser Abschnitt lediglich auf die Schätzung der Modellparameter konzentriert.

### 4.1 Parameterschätzung aus unvollständigen Daten

Das zur Parameterschätzung verfügbare Trainingsmaterial bestehe aus  $N$  Ansichten des zu erlernenden Objekts und die  $\varrho$ -te Ansicht enthalte  $\ell m$  Merkmale. Abbildung 8 zeigt exemplarisch einige Ansichten und die daraus errechneten



**Abbildung 8.** Beispiele für Trainingsansichten

Punkt- und Linien-Merkmale wie sie für die Modellgenerierung eingesetzt werden.

Die für das Training verfügbaren Aufnahmen weisen einige Eigenschaften auf, die für die Wahl des Parameterschätzverfahrens von entscheidender Wichtigkeit sind:

1. Die beobachteten Merkmale des Bildraums sind Projektionen der Merkmale des höherdimensionalen Modells. Die Tiefeninformation geht durch die Projektion verloren.
2. Die Zuordnungen der beobachteten Merkmale und der Modellmerkmale sind kein Bestandteil der beobachteten Trainingsdaten.

Die Schätzung der Modellparameter umfaßt einerseits die Berechnung der Parameter, die die einzelnen Merkmale charakterisieren, und andererseits die Bestimmung der Wahrscheinlichkeiten einzelner Zuordnungen. Alle Parameter müssen geschätzt werden, obwohl nur projizierte Merkmale bei fehlender Zuordnung

zur Verfügung stehen. Demzufolge liegt ein Parameterschätzproblem aus unvollständigen Daten vor, für das die Anwendung des Prinzips der fehlenden Information und der Einsatz des *Expectation–Maximization–Algorithmus* (EM–Algorithmus, [1]) naheliegen.

## 4.2 Expectation–Maximization–Algorithmus

Beim EM–Algorithmus handelt es sich um ein iteratives Parameterschätzverfahren, das ausgehend von einem initialen Schätzwert für die gesuchten Parameter die Likelihood–Funktion sukzessive maximiert. Trotz unvollständiger Trainingsdaten können so die Modellparameter geschätzt werden. Die Grundidee besteht hierbei im wesentlichen darin, die unbeobachteten Größen durch Schätzwerte zu ersetzen und iterativ sowohl die gesuchten Parameter der Dichtefunktion als auch die Schätzwerte der verborgenen Zufallsgrößen zu berechnen. Die iterative Verfahren terminiert, wenn sich die Neuschätzungen nicht mehr ändern.

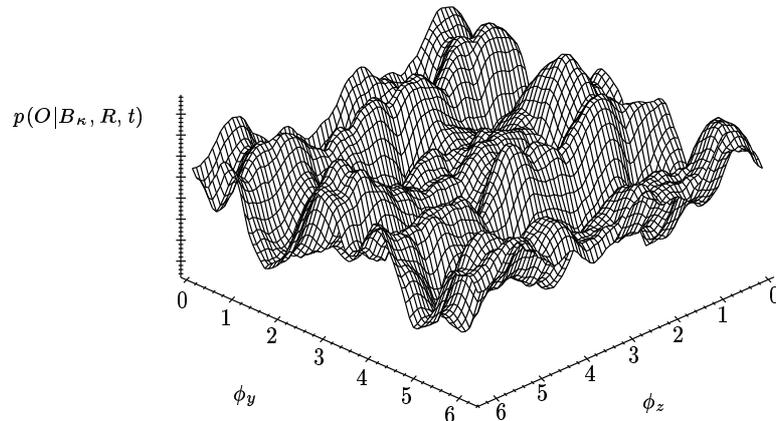
Die Berechnung der Schätzformeln läuft dabei nach einem fest vorgegebenen Schema ab [4]: Nach der Identifikation der beobachtbaren und nicht beobachtbaren Zufallsgrößen kann die sogenannte *Kullback–Leibler–Statistik* berechnet werden, die bezüglich der freien Parameter zu maximieren ist. Als kritisch für die Anwendung des EM–Algorithmus stellt sich die Initialisierung heraus. Da es sich beim EM–Algorithmus um ein lokales Optimierungsverfahren handelt, führt der EM–Algorithmus nur zum Erfolg, wenn der initiale Schätzwert für die Parameter in der *Nähe* des globalen Optimums liegt. Die entscheidenden Vorteile des EM–Algorithmus sind, daß der Suchraum häufig in kleinere, voneinander unabhängige Teile zerfällt, in vielen Fällen geschlossene Iterationsformeln angegeben werden können und die abgeleiteten Schätzverfahren nicht nur sehr wenig Speicher benötigen, sondern auch numerisch stabil sind.

Der EM–Algorithmus kann zur Berechnung der Parameter unterschiedlichster Modelle eingesetzt werden. Die Baum–Welch–Formeln für Hidden–Markov–Modelle lassen sich damit ebenso berechnen, wie die Schätzformeln zur Generierung statistischer Objektmodelle. Beispielsweise sind in [4] Schätzverfahren angegeben wie aus 2D–Punktmerkmalen 3D–Mittelwertvektoren geschätzt werden können, ohne das Korrespondenzproblem lösen zu müssen. Eine ausführliche Darstellung der Schätzalgorithmen und deren Herleitungen sind in [4] dargestellt.

## 5 Klassifikation und Lokalisation

Die Klassifikation von Objekten unter Verwendung der Bayes–Entscheidungsregel erfordert die Berechnung der *a–posteriori* Wahrscheinlichkeiten. Für eine Menge von Merkmalen  $\mathbf{O}$  wird die bedingte Wahrscheinlichkeit  $p(\Omega_\kappa|\mathbf{O})$  berechnet, mit der eine Musterklasse  $\Omega_\kappa$  vorliegt unter der Bedingung, daß die beobachteten Merkmale vorliegen. Der Klassifikator entscheidet sich für die Klasse mit der maximalen *a–posteriori* Wahrscheinlichkeit, d.h.,

$$\lambda = \operatorname{argmax}_\kappa p(\Omega_\kappa|\mathbf{O}) = \operatorname{argmax}_\kappa \frac{p(\Omega_\kappa)p(\mathbf{O}|\mathbf{B}_\kappa, \mathbf{R}, \mathbf{t})}{p(\mathbf{O})} \quad (3)$$



**Abbildung 9.** Zweidimensionaler Schnitt durch die Funktion, deren globales Maximum die Objektlage charakterisiert

Da die numerische Auswertung der *a-posteriori* Wahrscheinlichkeiten die Berechnung von  $p(\mathbf{O}|\mathbf{B}_\kappa, \mathbf{R}, \mathbf{t})$  erfordert, müssen für die Klassifikation die Rotations- und Translationsparameter bekannt sein und somit vorab berechnet werden. Bei der gesuchten Matrix  $\mathbf{R}$  und dem Vektor  $\mathbf{t}$  handelt es sich um Parameter der Wahrscheinlichkeitsdichte, weshalb Parameterschätzverfahren — wie beispielsweise die Maximum-Likelihood Schätzung — zur Lokalisation eingesetzt werden können. Die Parameterschätzung entspricht der globalen Optimierung einer multivariaten Funktion, deren Parameter die Freiheitsgrade der Rotation und Translation sind. Abbildung 9 zeigt einen zweidimensionalen Schnitt durch eine sechsdimensionale Funktion, wie sie bei der Lokalisierung von 3D-Objekten in 2D-Aufnahmen vorkommt. Um die Objektlage zu berechnen, muß das globale Maximum dieser Funktion berechnet werden. Zur Lösung dieser Optimierungsaufgabe scheiden deterministische Gittersuchverfahren aus Komplexitätsgründen aus, wie ein einfaches Zahlenbeispiel belegt: Bei einer Diskretisierung der drei Rotationswinkel in  $10^\circ$ -Schritte, einer Reduktion der drei Komponenten des Verschiebungsvektors auf 30 Quantisierungsstufen und einer Auswertzeit von 7 ms für einen Parametersatz werden für die Funktionsauswertung an allen Gitterpunkten etwa 2450 Stunden benötigt.

Die Praxis hat gezeigt, daß sich probabilistische Suchverfahren zur Lösung der globalen Optimierungsaufgabe eignen. Bei der hier gewählten adaptiven Zufallssuche paßt sich der Zufallsgenerator, der Punkte im Suchraum generiert, mit zunehmender Anzahl von Funktionsauswertungen an den Funktionsverlauf insofern an, als Bereiche, in denen hohe Funktionswerte beobachtet werden, wahrscheinlicher besucht werden als Gebiete des Suchraums, in denen bisher nur niedrige Funktionswerte beobachtet wurden. Einen allgemeinen Überblick über die Funktionsweise der adaptiven Zufallssuche gibt Abbildung 10.

/* Adaptive Zufallssuche */				
<b>Eingabe:</b> Modelldichte, beobachtete Merkmale				
werte die Modelldichte an $a$ zufällig erzeugten Punkten im Suchraum aus und speichere die besten $b$ dieser Punkte in einer nach dem Funktionswert sortierten Liste				
solange das Abbruchkriterium nicht erfüllt wird				
<table border="1"> <tr> <td>generiere neue Punkte unter Ausnutzung der in der Liste enthaltenen Information und werte die Modelldichte an diesen Punkten aus</td> </tr> <tr> <td>füge die neuen Punkte in die sortierte Liste ein</td> </tr> <tr> <td>entferne einige der schlechtesten Punkte aus der Liste</td> </tr> <tr> <td>modifiziere die Parameter, welche die Erzeugung von neuen Versuchspunkten steuern</td> </tr> </table>	generiere neue Punkte unter Ausnutzung der in der Liste enthaltenen Information und werte die Modelldichte an diesen Punkten aus	füge die neuen Punkte in die sortierte Liste ein	entferne einige der schlechtesten Punkte aus der Liste	modifiziere die Parameter, welche die Erzeugung von neuen Versuchspunkten steuern
generiere neue Punkte unter Ausnutzung der in der Liste enthaltenen Information und werte die Modelldichte an diesen Punkten aus				
füge die neuen Punkte in die sortierte Liste ein				
entferne einige der schlechtesten Punkte aus der Liste				
modifiziere die Parameter, welche die Erzeugung von neuen Versuchspunkten steuern				
bestimme das globale Optimum unter Verwendung der in der Liste gespeicherten Punkte				
<b>Ausgabe:</b> Koordinaten des globalen Maximums				

**Abbildung10.** Prinzip der adaptiven Zufallssuche

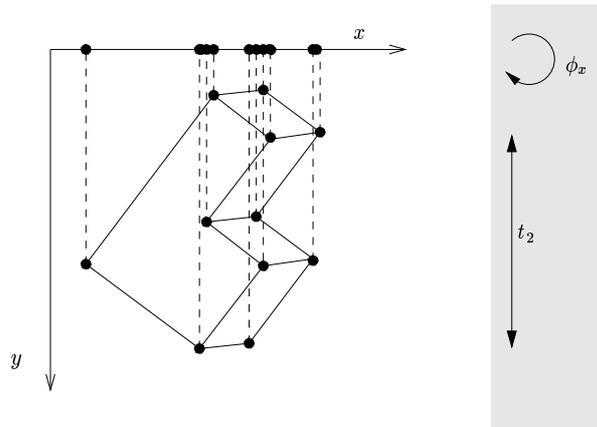
Neben dem Einsatz probabilistischer Suchverfahren bieten sich für die Objektlokalisierung noch weitere Möglichkeiten zur Reduktion des Suchaufwandes:

1. Durch geschickte Transformation des Suchproblems ist es häufig möglich, Abhängigkeiten zwischen einzelnen Dimensionen des Suchraums zu eliminieren. Beispielsweise sind 1D-Projektionen von 2D-Punkten auf die  $x$ -Achse des Bildkoordinatensystems unabhängig von Translationen parallel zur  $y$ -Achse und Rotationen um die  $x$ -Achse (vgl. Abbildung 11).
2. Geometrische Eigenschaften von Objekten können bei der Schätzung der Lageparameter dazu führen, daß bestimmte Bereiche des Suchraums *a-priori* ausgeschlossen werden können. Beispielsweise gibt es Objekte, die aus physikalischen Gründen nur sehr spezielle stabile Lagen einnehmen können. Die unmöglichen Positionen des Objektes können somit im Rahmen der Parameteroptimierung von vorneherein ausgeschlossen werden bzw. in Form einer Regularisierung innerhalb der Optimierung berücksichtigt werden.

Damit ist der Einsatz statistischer Verfahren zur Klassifikation und Lokalisation zumindest aus theoretischer Sicht geklärt. Im folgenden wird anhand einiger Experimente der praktische Einsatz dieser Verfahren erläutert.

## 6 Experimentelle Ergebnisse

Die experimentelle Evaluierung der entwickelten Verfahren erfolgt unter Verwendung der in Abbildung 1 dargestellten Objekte. Zum Training der Modelldichten werden je Objekt 400 Aufnahmen verwendet. Die Experimente zur Objektklassifikation beruhen auf insgesamt 1600 Beispielaufnahmen. Trainings- und Teststichprobe sind disjunkt, und der zur Auswertung verwendete Rechner ist eine



**Abbildung11.** Projektion der 2D-Merkmale auf eine Koordinatenachse

HP 9000/735 mit 99 MHz Taktrate, 124 MIPS, 147 SPECmark89 und 64 MB Hauptspeicher. Als Evaluierungskriterien dienen die erzielten Rechenzeiten und Erkennungsraten, und als Merkmale werden in allen Beispielen entweder 2D-Punkte und Linien verwendet.

## 6.1 Experimente zur Objektlokalisierung

In den experimentellen Vergleich globaler Optimierungsverfahren zur Maximierung der Wahrscheinlichkeitsdichte  $p(\mathbf{O}|\mathbf{B}_\kappa, \mathbf{R}, \mathbf{t})$  fließen folgende Algorithmen ein [4]:

- V1: die adaptive Zufallssuche,
- V2: die adaptiven Zufallssuche in Kombination mit dem Simplex-Verfahren zur lokalen Optimierung,
- V3: das Simulated Annealing für kontinuierliche Funktionen,
- V4: das Multistart-Verfahren,
- V5: das Grid-Simplex-Verfahren und
- V6: die rein probabilistische Suche.

Tabelle 1 zeigt eine Gegenüberstellung der gewählten globalen Optimierungsalgorithmen. Die Parameter der Suchalgorithmen werden hier derart bestimmt, daß in 20 zufällig gewählten Aufnahmen das globale Maximum sicher gefunden wird. Auf der Grundlage eines Rechenzeitvergleichs geht eindeutig die adaptive Zufallssuche, die mit lokalen Optimierungsverfahren gekoppelt wird, als Sieger hervor.

Unter Verwendung der so ermittelten Parametrierung führt ein Experiment mit insgesamt 400 2D-Ansichten eines 3D-Objektes, das aus zehn Modellpunkten besteht, zu einer Erfolgsrate von 82%. Nur in 18% der Fälle wird mit der

Optimierungsverfahren	Funktionsauswertungen	Rechenzeit [sec]
V1	10010	75
V2	8560	64
V3	41300	310
V4	585000	4380
V5	1820000	13600
V6	10000000	74500

**Tabelle 1.** Mittlere Anzahl der benötigten Auswertungen der Modelldichte sowie die Rechenzeiten bis zur Detektion des globalen Maximums

erweiterten adaptiven Zufallssuche (V2) das globale Maximum bezüglich der Lageparameter verfehlt.

Einige Lokalisierungsergebnisse sind in Abbildung 12 dargestellt. Zur Visualisierung der berechneten Lageparameter erfolgt eine Rückprojektion des Drahtgittermodells in die Szene. Diese erlaubt eine visuelle Beurteilung der erzielten Ergebnisse.

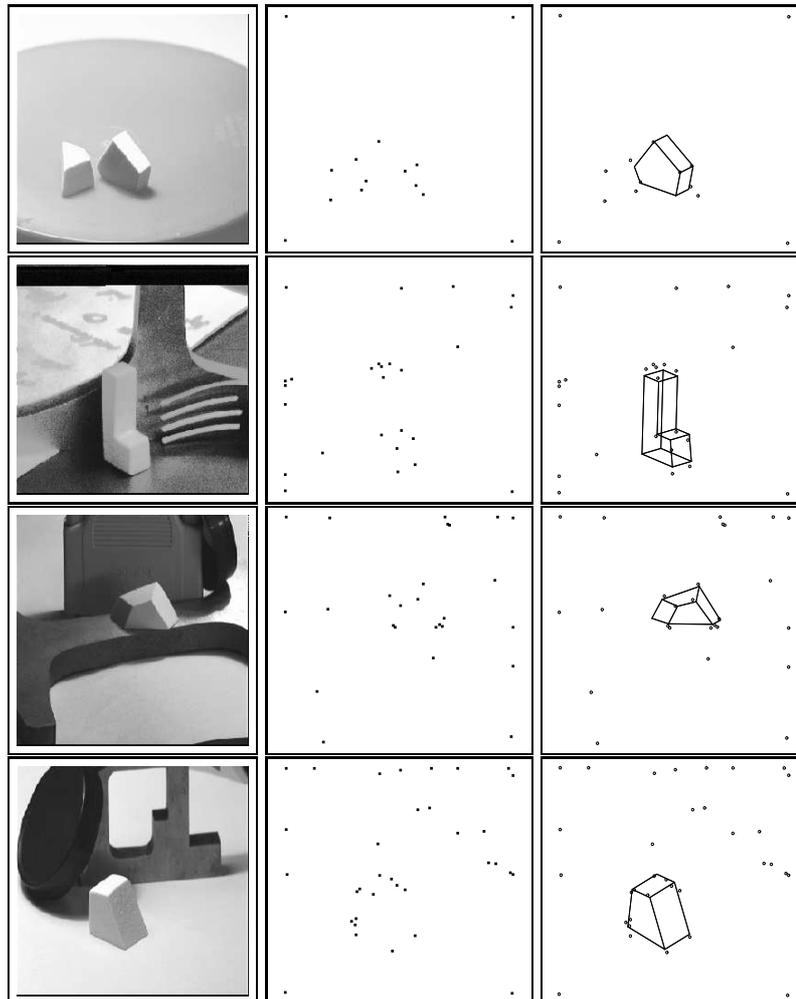
## 6.2 Experimente zur Objektklassifikation

Die verfügbaren Punkt und Linienmerkmale sind von der Objektlage abhängig, und somit ist die Berechnung der Lageparameter der Klassifikation voranzustellen. Fehler in der Lokalisation können ebenso für die Fehlklassifikation verantwortlich sein, wie fehlende Merkmale oder eine Übersegmentierung. Aufgrund dieser Überlegung darf wegen der bisher erzielten Lokalisationsergebnisse kaum eine Erkennungsrate von über 82% erwartet werden.

Für den Test des Klassifikators werden insgesamt 400 Aufnahmen pro Objektklasse verwendet. Da diese Bilder mit einer robotergeführten Kamera aufgenommen wurden, ist zu jeder Aufnahme die Objektlage bekannt. Unter Verwendung von Punktmerkmalen ergibt die Klassifikation für diese Stichprobe bei gegebener Objektlage eine Erkennungsrate von 96%. Bei unbekannter Objektlage steigt die Fehlerrate auf insgesamt 32% im Fall von Punktmerkmalen und auf 41% im Fall von Linienmerkmalen (vgl. Tabelle 2).

## 7 Weitere Anwendungen in der Mustererkennung

Die eingeführten statistischen Modelle können nicht nur zur Erkennung und Lokalisierung dreidimensionaler Objekte unter Verwendung segmentierter Punkt- oder Linienmerkmale eingesetzt werden, sondern sie finden inzwischen auch Verwendung zur Lösung weiterer Problemstellungen in der Mustererkennung. Varianten der in Abschnitt 3 eingeführten Modelle kommen sowohl in der Spracherkennung als auch in der Bildverarbeitung zum Einsatz.



**Abbildung 12.** Beispiele für die Lokalisation von 3D-Objekten in Grauwertaufnahmen bei heterogenem Hintergrund

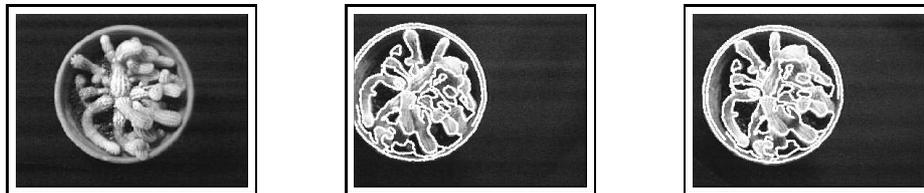
### 7.1 Probabilistische Modellierung der Grauwerte

Die Experimente zur 3D-Objekterkennung belegen eine Grundregel der Mustererkennung, die besagt, daß ein Klassifikator nur so leistungsfähig sein kann, wie es die gewählten Merkmale zulassen. Punktmerkmale sind zwar aus mathematischer Sicht einfach zu handhaben aber in aller Regel unzureichend für eine sichere Klassifikation beliebiger Objekte. Die darauf aufbauenden Klassifikatoren weisen daher auch relativ geringe Erkennungsraten auf. In jüngeren Arbeiten [5, 9] wird aus diesem Grund damit begonnen, nicht das statistische Verhalten segmentier-

3D-Objekt	Erkennungsrate [%]		Rechenzeit pro Bild [sec]	
	Punkte	Strecken	Punkte	Strecken
$\Omega_1$	47	44	466	1882
$\Omega_2$	78	82	485	2101
$\Omega_3$	58	36	465	1933
$\Omega_4$	89	76	471	1520
<b>Mittelwerte</b>	68	59	472	1859

**Tabelle2.** Erkennungsrate und Rechenzeit bei der Verwendung von Punkt- und Streckenmerkmalen

ter Merkmale zu modellieren, sondern direkt die Grauwerte unter Verwendung von Modelldichten probabilistisch zu charakterisieren. Erste Ergebnisse belegen, daß mit derartigen Ansätzen die Lokalisation von Objekten möglich wird, die mit Merkmalen aus einer Segmentierung und klassischen Lokalisationsverfahren bislang nicht möglich war. Abbildung 13 illustriert diesen Vorteil anhand eines Beispiels. Das linke Grauwertbild zeigt eine Pflanze, die beiden rechten Aufnahmen demonstrieren die erzielten Lokalisationsergebnisse unter Verwendung der Grauwertverteilungen. Zur Visualisierung der berechneten Rotations- und Translationsparameter werden die Konturlinien des Referenzobjektes und das Grauwertbild überlagert.

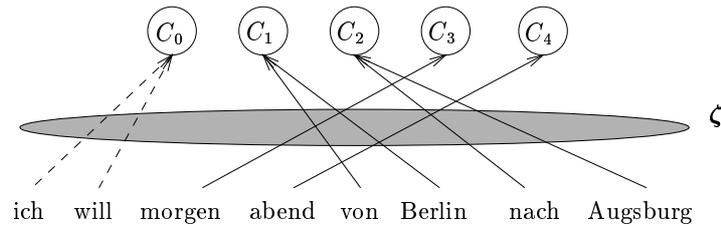


**Abbildung13.** Segmentierungsfreie Lokalisierung

## 7.2 Training und Erkennung semantischer Attribute

Ein wichtiges Problem im Bereich der Spracherkennung ist die semantische Analyse von Wortketten. In der Wortkette zu einer Anfrage an ein Zugauskunftssystem müssen beispielsweise der Abfahrtsort, der Zielort und die Abfahrtszeit detektiert werden. Die einzelnen Wörter sind somit semantischen Attributen zuzuordnen. Ein Beispiel für eine Zuordnung von Wörtern und Attributen  $C_0, C_1, \dots, C_4$  zeigt Abbildung 14. Wie im Falle der Bild- und Modellmerkmale

läßt sich die diskrete Zuordnungsfunktion auch hier probabilistisch modellieren (vgl. Abschnitt 3). Vorausgesetzt eine geeignete Initialisierung ist möglich, so lassen sich die statistischen Eigenschaften der Zuordnungsfunktion  $\zeta$  unter Verwendung des EM-Algorithmus aus einer Trainingsstichprobe trotz unbekannter Zuordnung für statistische Abhängigkeiten beliebiger Ordnung schätzen [2].



**Abbildung14.** Zuordnung semantischer Attribute

## 8 Schlußbemerkungen

Die eingeführte statistische Modellierung und die betrachteten Beispiele zeigen, daß sich statistische Verfahren durchaus zur Lösung von Problemen in der Bildverarbeitung erfolgreich einsetzen lassen. Die skizzierten Anwendungen, die über die segmentierungsbasierte Objekterkennung hinausgehen, belegen die Generalisierbarkeit und die Tragweite der erzielten Resultate innerhalb der Mustererkennung. Die Kopplung unterschiedlicher statistischer Größen in einer Modelldichte und der Einsatz des EM-Algorithmus zum automatischen Training eröffnen völlig neue Möglichkeiten für die Verwendung statistischer Verfahren. Die Komplexität der Trainingsalgorithmen und der Berechnungsaufwand für die Auswertung der Modelldichten werden direkt von den gewählten statistischen Abhängigkeiten beeinflusst.

Als ein noch offenes Problem wurde die automatische Generierung der Modellstruktur angesprochen. Während die Ordnung der statistischen Abhängigkeiten meist vom verfügbaren Trainingsmaterial abhängt, kann beispielsweise die Anzahl der Modellmerkmale weder analytisch noch mit effizienten heuristischen Schätzverfahren aus der Stichprobe berechnet werden. Ebenso kritisch für die Leistungsfähigkeit des Objekterkenners ist die Wahl der verwendeten Merkmale. Eine systematische Berechnung von Merkmalen, die in Kombination mit den statistischen Modellen zu einem optimalen Systemverhalten führen, ist wünschenswert. Mit den bisher verfügbaren Mitteln ist dies jedoch noch nicht realisierbar. Bleibt abschließend zu hoffen, daß die im Rahmen dieser Arbeit eingeführten und getesteten Methoden zumindest einen kleinen Beitrag zum Erreichen dieses langfristig angestrebten Ziels leisten können.

## Literatur

1. A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1):1–38, 1977.
2. J. Haas, J. Hornegger, R. Huber, and H. Niemann. Probabilistic semantic analysis of speech. In E. Paulus and F. Wahl, editors, *Mustererkennung 1997*. Springer, Berlin, erscheint September 1997.
3. M. Harbeck. *Objektorientierte linienbasierte Segmentierung von Bildern*. Shaker Verlag, Aachen, 1996.
4. J. Hornegger. *Statistische Modellierung, Klassifikation und Lokalisation von Objekten*. Shaker, Aachen, 1996.
5. J. Hornegger and D. Paulus. Bayesian vision: From intensity marginals to mutual information and entropic object recognition. Technical report, Lehrstuhl für Mustererkennung (Informatik 5), Universität Erlangen, März 1997.
6. S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer, Heidelberg, 1996.
7. K. V. Mardia. *Statistics and Images*, volume 2 of *Advances in Applied Statistics*. Carfax Publishing Company, Abingdon, 1994.
8. K. V. Mardia and G. K. Kanji. *Statistics and Images*, volume 1 of *Advances in Applied Statistics*. Carfax Publishing Company, Abingdon, 1993.
9. D. Paulus and J. Hornegger. *Pattern Recognition of Images and Speech in C++*. Advanced Studies in Computer Science. Vieweg, Braunschweig, 1997.