Probabilistische Modellierung bei der Objekterkennung
Seminar Wintersemester 2006 - 2007

Seminararbeit im Studiengang Informatik von

Emre Özyurt
Matr. Nr. 236 695

Theoretical Computer Science Dept. of Computer Science
Prof. Peter Rossmanith,
Rheinisch-Westfälische Technische Hochschule Aachen

Betreuer:

Inhaltsverzeichnis

1 Einleitung

Die Wissenschaftler und Bildverarbeitungsexperten haben sehr oft das Problem, bestimmte Objekte aus den vorhandenen Information in einem Bild zu erkennen und zu rekonstruieren. Oft kommt es vor, dass aufgrund der Unschärfe des Bildes oder der nicht optimalen Belichtung eine eindeutige Konstruktion nahezu unmöglich ist. Bei solch einem Fall können die probabilistischen Algorithmen verwendet werden, bei denen aus einem großen Vorrat an Objekten eines zufällig ausgewählt wird.

In dieser Arbeit wird beschrieben, wie die Objekte und ihre Positionen, mit Hilfe probabilistischer Methoden, aus einem eingegebenen 2 dimensionalen Bild erkannt werden. Das heißt, diese Methoden ermöglichen, 3 bzw. 2 dimensionale Objekte aus 2 dimensionalen Bildern zu identifizieren. Für die probabilistische Modellierung der Objekte werden die wichtigsten Eigenschaften (Ecke-Kante Beziehungen) von diesen benutzt. Mithilfe der probabilistischen Methoden, werden also die Zusammenhänge (Ähnlichkeiten) der Kanten- und Eckeigenschaften zwischen der projezierten Version des Objekts und dem eigentlichen Objekt geschätzt. Die Schätzung der Position des Objekts korrespondiert mit einem nicht linearen maximalen Wahrscheinlichkeitschätzungsproblem, welches von einem globalen Optimierungsverfahren gelöst wird. Für die Klassifizierung wird der Bayes'sche Klassifikator benutzt. Ein randomisiertes Verfahren sucht anschließend an die Modellierung, mit Hilfe der erhaltenen Objektparameter, die Position des Objektes.

2 Vorbereitung für die Arbeit

In diesem Abschnitt werden die Grundlagen, die nötig sind, um die im Laufe der Arbeit auftretenden Berechnungen zu verstehen, erklärt.

- Bedingte Wahrscheinlichkeit: $P(X|Y) = \frac{P(X \bigcap Y)}{P(Y)}$

  Bedingte Wahrscheinlichkeit beschreibt die Wahrscheinlichkeit des Auftretens des Ereignisses X unter der Annahme, dass das Ereignis Y schon aufgetreten ist. Um die Zugehörigkeit eines Objektmerkmales zu einer Objektklasse zu bestimmen, werden in dieser Arbeit diese Art von Berechnungen angewendet.

- Normalverteilung: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$

  Die Normalverteilung (engl.: normal distribution) ist ein Verteilungsmodell für kontinuierliche Zufallsvariablen. Die Normalverteilung ist das wichtigste Verteilungsmodell der Statistik[6]. In dieser Arbeit wurde die multivariante Normalverteilung benutzt, um die Wahrscheinlichkeit des Auftretens eines Merkmales zu berechnen.

- Bayes'scher Klassifikator: $P(C|x_1,...,x_n) = \frac{P(c)P(x_1,...,x_n|C)}{P(x_1,...,x_n)}$

  Der Bayes'sche Klassifikator dient dazu, die Zugehörigkeit der Zufallsvariablen zu den jeweiligen Klassen zu berechnen. Dadurch wird die Entscheidung getroffen, welcher Klasse das Objekt zugeordnet werden muss. Dabei ist $p(C|x_1,...,x_n)$ die Wahrscheinlichkeit, dass ein Objekt mit den Eigenschaften $x_1,...,x_n$ zur Klasse C gehört.

- Merkmalsextraktion und Merkmalstypen

  Der Begriff der Merkmalsextraktion (feature extraction) ist einer der wichtigsten Teilprozesse der Objekterkennung. Darunter versteht man die Bestimmung oder Herausstellung der Objekteigenschaften (Merkmale), die auf einem Bild zu erkennen sind.

  In dieser Arbeit gibt es im wesentlichen zwei wichtige Merkmalstypen.

  Bildmerkmale Ein Bildmerkmal o ist eine Eigenschaft eines Objekts, die sich auf der projezierten Darstellung des Objekts befindet. Es wird gesagt, dass zwei Objekte genau dann zur selben Objektklasse gehören, wenn sie die gleiche 3-dimensionale Struktur aufweisen beziehungsweise, anders gesagt, die gleichen Merkmale besitzen. Dabei spielt die Farbe keine Rolle. Eine Menge von Bildmerkmalen kann mit $O = \{o_1, ..., o_m\}$ repräsentiert werden und benannt werden als *image space* des Objektes.

  Modellmerkmale Ein Modellmerkmal c_K ist eine Eigenschaft, die nach der Transformation eines Bildmerkmales erhalten wird. Sie dient dazu, die 2-dimensionalen Bilder in einem 3-dimensionalen Raum zu visualisieren. Eine Menge von Modellmerkmalen kann mit $\Omega_K = \{c_{K,1}, c_{K,2}, ..., c_{K,3}\}$ beschrieben werden und benannt werden als *model space* des Objektes.

  Ein Bildmerkmal kann mit der Zuordnungsfunktion $\varsigma : O \to \Omega$ aus dem *image space* O in den *model space* Ω abgebildet werden.

3 Methodenbeschreibung

3.1 Objekterkennung mit der statistischen Vorgehensweise

In diesem Teil wird erfasst, wie die Objektmodelle beschrieben werden und wie sie in den Bayesischen Klassifikator eingesetzt werden.

Damit die Objekterkennung effektiver und beständiger läuft, werden aus den Musterbildern Objektmerkmale berechnet, die uns erlauben, an dem beobachteten Objekt einen Vergleich und anschliessend eine Klassifikation durchzuführen. Die Aufgabe eines Klassifikators ist die Erkennung der verschiedenen (vorhandenen) Prototypen, den sogenannten Objektmodellen, in beliebigen Bildern, bei unterschiedlicher Objektlage und Position.

Bei der modellbasierten Objekterkennung wird die Ähnlichkeit zwischen den Objektmodellen und den Merkmalen der beobachteten Objekte gemessen. Die Klasse, deren Objektmodell dem beobachteten Muster am ähnlichsten ist, wird als die Modellklasse des Objektes ausgewählt [1]. Fast alle Klassifikationsalgorithmen basieren auf dem Bayes'schen Klassifikator. Der Bayes'sche Klassifikator kann mit dem Bayes'schen Satz bewiesen werden.

$$p(\Omega_K|c) = \frac{p(c \cap \Omega_K)}{p(c)},\ p(c|\Omega_K) = \frac{p(c \cap \Omega_K)}{p(\Omega_K)},\ p(c) > 0,\ p(\Omega_K) > 0$$

Dieseselbe Gleichung kann mit der folgenden Formel dargestellt werden;

$$p(c \cap \Omega_K) = p(\Omega_K|c)p(c),\ p(c \cap \Omega_K) = p(c|\Omega_K)p(\Omega_K)$$

Daraus folgt;

$$p(\Omega_K|c) = \frac{p(c|\Omega_K)p(\Omega_K)}{p(c)}$$

Hierbei kann der Term $p(c|\Omega_K)$ durch den Term $p(c; a_k)$ ersetzt werden, da $p(c; a_k)$ die Wahrscheinlichkeit angibt, dass ein Modellmerkmal c aus der Modellklasse Ω_k auftritt.

$$\lambda = \arg\max_K\ p(\Omega_K|c) = \arg\max_K\ \frac{p(\Omega_K)\, p(c; a_k)}{p(c)} \tag{1}$$

Dieser Klassifikator entscheidet sich für die Klasse Ω_K, die für ein Modellmerkmal eines beobachteten Objekts, die *a-posteriori* Zugehörigkeits-Wahrscheinlichkeit $p(\Omega_K|c)$ maximiert. Um die *a-posteriori* Wahrscheinlichkeit feststellen zu können, muss die *a-Priori* Wahrscheinlichkeit $p(\Omega_K)$ jeder Objektklasse Ω_K bekannt sein (siehe Kap.1). Da es schwer ist die a-posteriori-Wahrscheinlichkeit auf direktem Wege zu ermitteln, wird diese mittels der Bayes'schen Regel auf die a-priori Wahrscheinlichkeit und die klassenbedingte Wahrscheinlichkeit berechnet [3]. In der Formel bezeichnet die parametrische Dichtefunktion $p(c; a_k)$ des Modellmerkmals c; das probabilistische Verhalten der Zufallsgrösse mit dem klassespezifischen Parameter $a_K = \{\mu_K, \sum_k\}\quad \mu_K \in R^3, \sum_k \in R^{3\times3}$, wobei μ_K den Erwartungswertvektor und $\sum_k$ die Kovarianzmatrix beschreibt [1].

Die Tatsache, dass die Objekte auf den Bildern nicht optimal wie bei einem Musterbild positioniert sind, erschwert deren Klassifikation. Dieses Problem kann gelöst werden, indem die multivariante Normalverteilung benutzt wird. Der Vorteil der multivarianten Normalverteilung ist, dass die einzelnen Komponenten der multivarianten Normalverteilung auch bei Rotationen stochastisch unabhängig bleiben[5]. Also aus der Definiton von Normalverteilung $\mathcal{N}$ der Stetige Zufallsvariablen:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \Rightarrow X \sim \mathcal{N}(\mu, \sigma^2)$$

erhalten wir die Gleichung, die durch Ersetzen der klassespezifischen Parameter wie folgt aussieht.

$$\begin{aligned} p(c; a_k) = X \sim \mathcal{N}(\mu, \sigma^2) &\overset{(\mu \Leftarrow c; \mu_k, \sigma^2 \Leftarrow \Sigma_k)}{=} \mathcal{N}(c; \mu_k, \Sigma_k) \\ \overset{siehe\ kap.1}{=} &(2\pi)^{-\frac{1}{2}} |\Sigma_k|^{-\frac{1}{2}} exp(-\frac{1}{2}(c-\mu_k)^T \Sigma_k^{-1}(c-\mu_k)) \end{aligned} \tag{2}$$

Dadurch kann bestimmt werden, wie hoch die Wahrscheinlichkeit ist, dass ein Modellmerkmal c auftritt.

3.2 Merkmal-Transformation

Um ein 3-dimensionales Objekt statistisch modellieren zu können, braucht man eine Menge von Merkmalen, die das Objekt repräsentieren. Dafür eignen sich die Linien oder Eckpunkten des Objekts, also die sogenannten primitiven Merkmale. Da die Erkennung in der 2-dimensionalen Bildebene durchgeführt wird, müssen die 3-dimensionalen Modellmerkmale ins 2-dimensionale projeziert werden.Wie oben erwähnt wurde, ist es wichtig zu betrachten, dass die Objekte nicht so positioniert sind wie auf dem Muster. Diese Positionierung kann beschrieben werden, indem die primitiven Merkmale durch die Rotationsmatrix $R \in R^{D_m \times D_m}$ und den Translationsvektor $t \in R^{D_m}$ im Modellraum transformiert werden, wobei D_m die Dimension des Modellraumes ist. Beispielweise würde ein Eckpunkt c mit drei Achsengrößen c_1, c_2, c_3 beschrieben und nach der Rotation und Transformation mit der Gleichung $c' = (Rc_1 + t, Rc_2 + t, Rc_3 + t)$ repräsentiert werden.

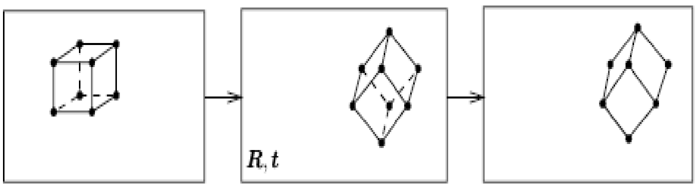

Abbildung 1: Die Projezierung vom 3-dimensionalen Raum in den 2-dimensionalen für ein beobachtetes Objekt

Mit Hilfe dieser Darstellung kann der Bayes'sche Klassifikator (aus Formel 1) mit den Bildmerkmalen des beobachteten Objektes benutzt werden.

$$\lambda = \arg\max_K \ p(\Omega_K|o) \overset{(aus\ 1)}{=} \arg\max_K \frac{p(\Omega_K)\, p(o; a_k, R, t)}{p(o)} \tag{3}$$

Wie vorher gesagt wurde, kann ein Objekt selbstverständlich nicht mit einem einzigen Bildmerkmal beschrieben werden. Für ein Objekt mit seinen beobachteten Bildmerkmalen $O = \{o_1, o_2, ..., o_m\}$ kann dann die Dichtefunktion mit $p(O; B_K, R, t)$ bezeichnet werden. Dabei präsentiert B_K alle objektspezifischen Parameter a_k.

3.3 Objektmodellierung mit der statistischen Vorgehensweise

Wie oben erwähnt wurde, ist die Detaillierung des Objekts bei der Objekterkennung wichtig. Dabei ist auch wichtig, wie sinnvoll diese detaillierte Beschreibung ist. Der Schwerpunkt bei der Objektmodellierung ist es nämlich, die Merkmale so zu entwickeln, dass diese Merkmale das entsprechende Objekt von den anderen Objekten differenziert [2]. In diesem Teil wird auf die Objektmodellierung eingegangen, welche die probabilistische Beschreibung der beobachteten Bildmerkmale ermöglicht.

Um feststellen zu können, ob ein aus einem Bild extrahiertes Objekt O, einer der Modellklassen Ω_K angehört, müssen alle Bildmerkmale o_k, $k \in \{1, ..., K\}$, in Modellmerkmale umgewandelt werden. Auch wenn es in der Realität anders ist und die Merkmale untereinander eine Korrelation aufweisen, wird hier dennoch angenommen, dass alle stochastisch unabhängig sind. Das ermöglicht, wie im Laufe der Arbeit erklärt werden wird, eine effizientere Berechnung und eine reduzierte Laufzeitkomplexität.

In diesem Abschnitt wird beschrieben, wie die aus dem Bild erhaltenen Merkmale, mit Hilfe der Zuordnungsfunktion, in Modellmerkmale umgewandelt werden und wie bei den gestörten Bildern das Ausschalten der Hintergrundmerkmale funktioniert.

3.3.1 Die diskrete Zuordnungsfunktion

Die Zuordnungsfunktion dient dazu, die Bildmerkmale eines Objektes in Modellmerkmale eines Musters abzubilden. Die in dem Grundlagen-Kapitel beschriebene Darstellung kann wie folgt erweitert werden:

$$\varsigma_K : \begin{cases} O \rightarrow 1, ..., n_K \\ o_k \mapsto l_k \;\; k = 1, ..., m \end{cases} \tag{4}$$

Am Ende wird die Bildmerkmalmenge auf einen diskreten Vektor $\varsigma_K = (\varsigma_k(o_1), ..., \varsigma_k(o_m))^T$ abgebildet. Da die Zuordnung ein stochastisches Modell hat, muss die Bedingung $\sum_{\varsigma_k} p(\varsigma_k) = 1$ erfüllt werden. Zusammen mit Hilfe der diskreten Zuordnungsfunktion und dem oben erwähnten Hinweis, dass die Merkmale stoschastisch unabhängig sind, lässt sich die folgende Formel erstellen.

$$p(o_1, o_2, ...o_m) = p(o_1) \cdot p(o_2) \cdot p(o_3)$$

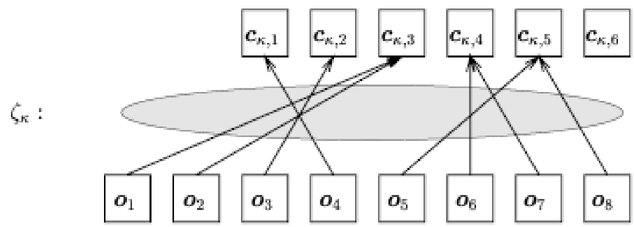

Abbildung 2: Die Zuordnungsfunktion bildet die Bildmerkmale randomisiert auf die Modellmerkmale ab

Daraus folgt;

$$p(O|\varsigma_K; B_K, R, t) \stackrel{Unabhängigkeit}{=} \prod_{k=1}^{m} p(o_k; a_{K,l_k}, R, t) \tag{5}$$

Dieselbe Funktion lässt sich für eine mögliche Zuordnung so darstellen;

$$p(O; B_K, R, t) \stackrel{(aus\ 5)}{=} p(\varsigma_K) \prod_{k=1}^{m} p(o_k; a_{K,l_k}, R, t) \tag{6}$$

Wobei der Parameter a_{K,l_k} zusätzlich die Information enthält, dass dieser Parameter der Klasse Ω_K angehört und gemäß der Zuordnungsfunktion die ausgewählte Modellmerkmalsnummer l_k ist.

Durch den Vektor ς_K kann die Zufallsgröße der Zuordnung bestimmt werden. Für eine bestimmte Zuordnung würde die zusammmengesetzte Dichte Funktion wie folgt aussehen;

$$\begin{aligned} p(O; B_K, R, t) & \\ &= p(O, \varsigma_K | B_K, R, t) \\ &= p(\varsigma_K) p(O|\varsigma_K, B_K, R, t) \\ &= p(\varsigma_K) p(O|\varsigma_K, a_{K,1}, a_{K,2}, ..., a_{K,n_K}, R, t) \end{aligned} \tag{7}$$

Selbstverständlich erzeugt ς_K einen zufälligen Vektor und es ist vorher nicht erkennbar, ob dieser Vektor einer der Objektklassen angehört.

Gesucht wird aber ein Vektor, der zu einer der Objektklasse gehört. Dazu wird Marginalisierung benutzt, bei der alle Zuordnungen für jede Variable der Bildmerkmalmenge summiert werden. Dadurch wird die Randdichte berechnet[2]. Das ergibt die Modelldichte einer beobachteten Bildmerkmalmenge O. Also:

$$
\begin{aligned}
p(O; B_K, R, t) &= \sum_{\varsigma_K} p(O, \varsigma_K | B_K, R, t) \\
&\overset{(aus\ 7)}{=} \sum_{\varsigma_K} p(\varsigma_K) p(O|\varsigma_K, a_{K,1}, a_{K,2}, ..., a_{K,n_K}, R, t) \\
&\overset{(aus\ 5)}{=} \sum_{\varsigma_K} p(\varsigma_K) \prod_{k=1}^{m} p(o_k; a_{K,\varsigma_K(o_k)}, R, t)
\end{aligned} \tag{8}
$$

Durch diese Marginalisierung kann alle mögliche Merkmale überprüft werden. Das sorgt für eine demokratische Messung [1].

3.3.2 Das Hintergrundproblem

Die Methode, die hier vorgestellt wird, funktioniert einwandfrei bei den Trainingsbildern, wo sich Objekte mit variierenden Positionen befinden. Aber um die Methode in der Realität benutzen zu können, muss die Hintergrundfläche ebenfalls betrachtet werden, um Konflikte zu vermeiden.

Dafür wird die Dichtefunktion so erweitert, dass ein gewonnnenes Merkmal entweder einem Hintergrund oder einer Modellklasse zugewiesen wird.

$$
\varsigma_{H,K}(o_k) = \begin{cases} 0 & \text{, falls } O_k \text{ dem Hintergrund zugeordnet wird} \\ K & \text{, falls } O_k \text{ zur Objektklasse } \Omega_K \text{ gehört} \end{cases} \tag{9}
$$

Zusammen mit der Formel aus 8;

$$
p(O; B_H, B_K, R, t) \overset{(aus\ 8)}{=} \sum_{\varsigma_{H,K}, \varsigma_K} p(\varsigma_{H,K}, \varsigma_K) \left(\prod_{\substack{k=1 \\ \varsigma_{H,K(o_k)}=K}}^{m} p(o_k; a_{K,\varsigma_K(o_k)}, R, t) \right) \cdot \left(\prod_{\substack{k=1 \\ \varsigma_{H,K(o_k)}=0}}^{m} p(o_k; a_H) \right) \tag{10}
$$

Wobei B_H alle hintergrundspezifischen Parameter a_H und die Wahrscheinlichkeit $p(\varsigma_{H,K}) \in [0, 1]$ des randomisierten Vektors $\varsigma_{H,K}$ präsentiert.

3.4 Erlernen der Modellmerkmale

In diesem Abschnitt wird darauf eingegangen, wie die verschiedenen Modellparameter geschätzt werden.

Bei dem Training werden für ein Objektmodell Ω_K, N verschiedene Bilder jeweils aus einer anderen Sicht aufgenommen und dabei identifizierte Merkmale werden für jede Sicht registriert. Die Anzahl der Bildmerkmale bei dem ϱ-ten Bild wird mit ${}^{\varrho}m$ beschrieben. Die Parameterschätzung des

Modells sieht folgendermaßen aus:

$$\hat{B}_K = \arg\max_{B_K} \sum_{\varrho=1}^{N} \log p({}^{\varrho}O; B_K, {}^{\varrho}R, {}^{\varrho}t) \tag{11}$$

Hierbei summiert $\hat{B}_K$ alle Modellmerkmalparameter von dem Objekt K. Der Logarithmus wird hier benutzt, damit die Zahl sinkt. Da es um eine statistische Anwendung geht, ist hier nur wichtig, wie das Verhältnis zwischen der verschiedenen Winkel sind, und nicht die Größe der Zahl. Die Symbole ${}^{\varrho}O, {}^{\varrho}R, {}^{\varrho}t$ beschreiben (in der Reihenfolge): die Bildmerkmalmenge, die Rotation und die Translation des ϱ-ten Trainingbildes von dem Objekt K. Diese Maximierung von $\hat{B}_K$ wird mit Hilfe des *Expectation-Maximization-Algorithmus* gemacht.

Expectation-Maximization-Algorithmus

Dieser Algorithmus wird bei der Schätzung von B_K-Werten benutzt, um die beste Beschreibung von Trainingsdaten zu finden. Das heißt: der Schätzungswert eines Parameters muss das *Maximum-Likelihood-Kriterium* maximieren [3].

Problematisch dabei ist, dass alle Parameter geschätzt werden müssen, obwohl durch die randomisierte Zuordnung nicht alle Objektmerkmale zur Verfügung stehen (siehe *Abb.* 2). Die Lösungsidee besteht darin, dass die fehlenden Daten unter Berücksichtigung der aktuell beobachteten Daten und Modellparameter geschätzt werden [4]. Der EM-Algorithmus beruht also auf dem Prinzip "versteckter" Zufallsvariablen und basiert auf einem iterativen Verfahren, das in jedem Iterationsschritt versucht, die am besten geeigneten Parameter zu finden. Für die versteckten Variablen werden Erwartungswerte berechnet (Expectation), mit denen ein neuer Maximum-Likelihood-Parametersatz aus dem beobachtbaren vorhergehenden Parametersatz (Maximization) berechnet wird [3]. Nach der Identifikation sowohl existierender als auch fehlender Zufallsgrößen wird die *Kullback-Leibler-Statistik* berechnet, welche in [5] ausführlicher beschrieben wird.

Wichtig bei dem EM-Algorithmus ist die Initialisierung der Schätzwerte. Da es beim EM-Algorithmus um ein lokales Optimierungsverfahren geht, kann dieses nur benutzt werden, wenn der initialisierte Schätzwert in der Nähe eines Optimum liegt [2]. Das bedeutet, vor der Schätzung müssen Anzahl der Modellmerkmale und die ersten groben und passenden Initialisierungswerte für Modellmerkmalparameter, sprich die Erwartungswerte und Kovarianzmatrizen gegeben werden.

Wie vorher erwähnt wurde, werden für diese Schätzung die Hintergrundmerkmale alle mitsummiert. Zusammen mit diesen Überlegungen und der Formel 7 lässt sich die Wahrscheinlichkeitsverteilung vom Auftreten des k-

ten Bildmerkmals ${}^{\varrho}o_k$ von der ϱ-ten Sicht mit der folgenden Formel darstellen.

$$
\begin{aligned}
p(\varrho, k, l) &= p({}^{\varrho}o_k | \varsigma_K({}^{\varrho}o_k) = l; \hat{B}_K^{(i)}, {}^{\varrho}R, {}^{\varrho}t) \\
&\overset{(aus\ kap\ 1)}{=} \frac{\hat{p}_{K,l}^{(i)} p({}^{\varrho}o_k; \hat{a}_{K,l}^{(i)}, {}^{\varrho}R, {}^{\varrho}t)}{p({}^{\varrho}o_k; \hat{B}_K^{(i)}, {}^{\varrho}R, {}^{\varrho}t)} \\
&\overset{(aus\ 8)}{=} \frac{\hat{p}_{K,l}^{(i)} p({}^{\varrho}o_k; \hat{a}_{K,l}^{(i)}, {}^{\varrho}R, {}^{\varrho}t)}{\sum_{l=1}^{n_K} \hat{p}_{K,l}^{(i)} p({}^{\varrho}o_k; \hat{a}_{K,l}^{(i)}, {}^{\varrho}R, {}^{\varrho}t)}
\end{aligned}
\tag{12}
$$

Wobei die ${}^{k}\hat{p}_{K,l}$ die Wahrscheinlichkeitsverteilung einer Zuordnungfunktion $p(\varsigma_K(o_k))$ bezüglich des Bildmerkmals o_k beschreibt. Nach diesen Überlegungen muss, um den EM-Algorithmus anzuwenden, nach den optimalen Wahrscheinlichkeitverteilungen der Zuordnungfunktionen sowie den objektspezifischen Parametern gesucht werden. Benötigt werden also die Zuordnungsfunktionen, deren Wahrscheinlichkeitsverteilungen beim Aufsummieren von jedem einzelnen Trainingsbild das Maximum liefern[3,5]. Dafür wird die Q-Funktion, die alle Zuordnungen von jedem Bildmerkmal, aus jeder Sicht summiert, benutzt. Ausgangspunkt ist die Differenz zweier *Log-Likelihood-Funktionen* (siehe Formel unten). Es sei $\hat{B}_K^{(i)}$ ein gegebener Parametersatz, dann wird $\hat{B}_K^{(i+1)}$ neu geschätzt, so dass die neuen Parameter die Verteilung besser als die alten beschreiben. Mit Hilfe der Formel 11 lässt sich die Q-Funktion wie folgt definieren [1].

$$
\begin{aligned}
Q(\hat{B}_K^{(i+1)}; \hat{B}_K^{(i)}) = \sum_{\varrho=1}^{N} \sum_{k=1}^{{}^{\varrho}m} \sum_{l=1}^{n_K} p(\varrho, k, l) \log \hat{p}_{K,l}^{(i+1)} + \\
\sum_{\varrho=1}^{N} \sum_{k=1}^{{}^{\varrho}m} \sum_{l=1}^{n_K} p(\varrho, k, l) \log p({}^{\varrho}o_k; \hat{a}_{K,l}^{(i+1)}, {}^{\varrho}R, {}^{\varrho}t)
\end{aligned}
\tag{13}
$$

Hierbei beschreibt N Anzahl der Sichten, ${}^{\varrho}m$ anzahl der Bildmerkmale für die jeweilige Sicht und n_K Anzahl der Modellmerkmale. Der neue Parametersatz wird als der gegebene Parametersatz verwendet, und $\hat{B}_K^{(i+1)}$ wird wiederum neu geschätzt. Dieser iterative Vorgang wiederholt sich solange, bis kein besserer Parametersatz gefunden wird.

3.4.1 Position des Objekts

Nachdem eine Menge von Bildmerkmalen durch Zuordnungsfunktion auf eine Modellklasse abgebildet wird und die Modellmerkmale durch den Bayes'schen Klassifikator in eine Objektklasse zugeordnet werden, bleibt die Aufgabe zu schätzen, wie und wo das Objekt auf dem Bild positioniert ist.

Dafür werden die Parameter R und t so ausgewählt, dass durch die Dichtefunktion der Maximalwert geschätzt wird. Also muss die Dichtefunktion

$$\arg\max_{R,t}\ p(O; B_K, R, t) \tag{14}$$

den maximalen Wert liefern. Um die Objektlage zu bestimmen muss das globale Maximum dieser Funktion berechnet werden. Für diesen Zweck eignet sich am besten die adaptive Zufallssuche.

/* Adaptive Zufallssuche */
Eingabe: Dichtefunktion mit der beobachteten Merkmalen
a zufällige Punkte mit gleicher Wahrscheinlichkeit generieren und die besten b dieser Punkte nach dem Funktionswert der Dichtefunktion sortiert speichern
solange das Abbruchkriterium nicht erfüllt wird
die Information von gespeicherten Punkten ausnutzen und weiterer Punkte generieren, danach die Dichtefunktion an diesen Punkte auswerten.
die neue Punkte in die Liste einfügen, und die schlechtesten wieder entfernen
die Parameter mit einem Faktor $0 < x < 1$ multiplizieren, damit den Suchbereich für die weitere Punkte eingeschränkt wird
bestimme das globale Optimum unter Verwendung der in der Liste gespeicherten Punkte
Ausgabe: Koordinaten des globalen Maximums

Abbildung 3: Die Arbeitsweise der adaptiven Zufallssuche [2]

Es ist auch möglich den Suchraum durch eine Transformation einzuschränken, indem die 2-dimensionale Darstellung des Objekts auf die x-Achse projeziert wird. Dadurch wird eine 1-dimensionale Information über ein Merkmal des Objekts, welches bei einer Rotation immer den gleichen Wert aufweist, erhalten. Dieser Vorgang reduziert den Suchaufwand.

4 Zusammenfassung

In dieser Seminararbeit wurde ein statistisches Verfahren als alternative Lösung von Problemen in der Bildverarbeitung vorgestellt.

Zunächst werden von den erhaltenen Bildern die Bildmerkmale extrahiert. Die extrahierten Bildmerkmale werden zufällig auf Objektmerkmale abgebildet. Dann werden die Wahrscheinlichkeitsverteilungen der dadurch entstandenen Objekte berechnet. Hierbei wurden ebenfalls die Rotation und Translation des Objektes betrachtet. Mittels des Bayes'schen Klassifikators wird dann entschieden, welcher Objektklasse die Objektmerkmale am besten angehören. Nachdem die Entscheidung getroffen wurde, blieb nur die Aufgabe übrig, die Lage des Objekts zu bestimmen. Zur Lösung dieser Aufgabe wurde dann das Verfahren der adaptiven Zufallssuche vorgestellt.

Es wurde gesagt, dass eins der wichtigsten Bestandteile dieses Verfahrens, das Erlernen des Objektmodells ist. Es werden für ein einziges Objekt viele Trainingsbilder gemacht, deren gesamten Parameter durch eine spezielle Marginalisierung aufsummiert werden. Dadurch wurden die Objektparameter identifiziert und die sogenannten Objektklassen und dazu gehörigen Modellmerkmale gebildet.

5 Literatur

1. *J. Hornegger and H. Niemann. Probabilistic Modeling and Recognition of 3-D Objects. International Journal of Computer Vision, Volume 39, Number 3, September 2000*

2. *J. Hornegger. Statistische Modellierung, Klassifikation und Lokalisation von Objekten. Lehrstuhl für Mustererkennung. 1999.*

3. *J. Hornegger. Statistische Modellierung, Klassifikation und Lokalisation von Objekten. Shaker. Aachen (1996).*

4. *M. Motter. Statistische Modellierung von Bildinhalten* für die Bilderkennung. Thesis, RWTH Aachen. 2001.

5. *S. Dabek. Erkennung von 3D-Objekten mit probabilistischen Methoden.Institut für Informatik WWU Münster. 20.6.2005.*