

Auswahl von Kameraaktionen zur wissensbasierten Szenenexploration

U. Ahlrichs, D. Paulus, H. Niemann

Lehrstuhl für Mustererkennung (LME, Informatik 5)
Martensstr. 3, Universität Erlangen–Nürnberg, 91058 Erlangen
Tel.: +49 (9131) 85–27826 — Fax: +49 (9131) 303811
ahlrichs@informatik.uni-erlangen.de
<http://www5.informatik.uni-erlangen.de/Persons/ah>

eingereicht zu

21. DAGM Symposium

Mustererkennung 1999
September 1999
Bonn

Rev.: 1.3
L^AT_EX'ed 9. September 1999

Auswahl von Kameraaktionen zur wissensbasierten Szenenexploration

U. Ahlrichs, D. Paulus, H. Niemann

Lehrstuhl für Mustererkennung (LME, Informatik 5)
Martensstr. 3, Universität Erlangen–Nürnberg, 91058 Erlangen
Tel.: +49 (9131) 85–27826 — Fax: +49 (9131) 303811
ahlrichs@informatik.uni-erlangen.de
<http://www5.informatik.uni-erlangen.de/Persons/ah>

Zusammenfassung Verfahren zur Szenenexploration basieren unter anderem auf der Auswahl von Kameraparametern für eine Bildaufnahme, die — wie von der Strategie des aktiven Sehens gefordert — für die weitere Verarbeitung optimal geeignet sein müssen. In diesem Artikel wird die Auswahl der Kameraparameter auf eine Auswahl von Kameraaktionen zurückgeführt, die in einer Wissensbasis als Konzepte eines semantischen Netzes repräsentiert werden. Zusätzlich zu diesem Wissen enthält die Wissensbasis Information über die Szene und die in die Aufgabe involvierten Objekte. Um aus mehreren Kameraaktionen, die zu einem Analysezeitpunkt ausführbar sind, die optimale auswählen zu können, wird ein Bewertungskalkül benötigt. In dem Artikel wird ein neues, auf Ideen der Entscheidungstheorie beruhendes Bewertungskalkül vorgestellt. Die Kameraaktionen werden anhand von Nützlichkeiten bewertet, die auf der Bewertung der Instanzen für die zu suchenden Objekte in der Szene basieren. Die Tragfähigkeit des Ansatzes wird am Beispiel der Exploration einer Büroszene belegt.

Schlüsselwörter: Aktives Sehen, visuelle Exploration, semantische Netze

Auswahl von Kameraaktionen zur wissensbasierten Szenenexploration

U. Ahlrichs, D. Paulus, H. Niemann

Lehrstuhl für Mustererkennung (LME, Informatik 5)
Martensstr. 3, Universität Erlangen–Nürnberg, 91058 Erlangen
Tel.: +49 (9131) 85–27826 — Fax: +49 (9131) 303811
ahlrichs@informatik.uni-erlangen.de
<http://www5.informatik.uni-erlangen.de/Persons/ah>

Zusammenfassung Verfahren zur Szenenexploration basieren unter anderem auf der Auswahl von Kameraparametern für eine Bildaufnahme, die — wie von der Strategie des aktiven Sehens gefordert — für die weitere Verarbeitung optimal geeignet sein müssen. In diesem Artikel wird die Auswahl der Kameraparameter auf eine Auswahl von Kameraaktionen zurückgeführt, die in einer Wissensbasis als Konzepte eines semantischen Netzes repräsentiert werden. Zusätzlich zu diesem Wissen enthält die Wissensbasis Information über die Szene und die in die Aufgabe involvierten Objekte. Um aus mehreren Kameraaktionen, die zu einem Analysezeitpunkt ausführbar sind, die optimale auswählen zu können, wird ein Bewertungskalkül benötigt. In dem Artikel wird ein neues, auf Ideen der Entscheidungstheorie beruhendes Bewertungskalkül vorgestellt. Die Kameraaktionen werden anhand von Nützlichkeiten bewertet, die auf der Bewertung der Instanzen für die zu suchenden Objekte in der Szene basieren. Die Tragfähigkeit des Ansatzes wird am Beispiel der Exploration einer Büroszene belegt.

Schlüsselwörter: Aktives Sehen, visuelle Exploration, semantische Netze

1 Motivation

Typische Aufgaben von autonomen mobilen Systemen, die visuelle Information zur Lösung einer Aufgabe benutzen, erfordern im allgemeinen eine Exploration unterschiedlichster Umgebungen. Hierzu sind variable Strategien zur Exploration oder auch zur Erkennung von Objekten zwingend notwendig. Die visuelle Exploration einer Szene erfordert eine gezielte Einstellung der mechanischen und optischen Kameraparameter, wie sie z.B. Bestandteil der Strategie des aktiven Sehens ist [Alo88].

In diesem Artikel wird die Auswahl von Strategien zur Exploration bzw. die Auswahl der Kameraparameter auf eine Auswahl von Kameraaktionen zurückgeführt. Die Kameraaktionen werden als Konzepte in einem wissensbasierten Formalismus repräsentiert, wobei Abhängigkeiten zur Szenenrepräsentation berücksichtigt werden. Die Wissensbasis enthält somit Kameraaktionen und Objekte, die *einheitlich* als Konzepte eines semantischen Netzes repräsentiert werden. Dies erlaubt es, während der Analyse

¹ Diese Arbeit wurde unterstützt durch die *Deutsche Forschungsgemeinschaft* im Rahmen des Sonderforschungsbereichs 182 und des DFG-Projekts NI 191/12-1.

aufgrund der über eine Szene gewonnenen Information aus mehreren möglichen Kameraaktionen diejenige auszuwählen, die mit geringstmöglichem Aufwand die optimale Fortsetzung der Szenenexploration erlaubt. Hierzu stellen wir ein neues, auf den Grundlagen der Entscheidungstheorie [Jen96] basierendes Bewertungskalkül vor.

Semantische Netze erlauben eine intuitive Repräsentation der Objekte, die sich in einer Szene befinden: Die Positionen der Objekte können zum Beispiel unmittelbar als Attribute des entsprechenden Konzepts modelliert werden, was die Handhabung bei der Modellierung gegenüber Bayesnetzen, die in vergleichbaren Arbeiten verwendet wurden [Kre98,Rim93,Lev89], erleichtert.

Zur Nutzung des in der Wissensbasis repräsentierten Wissens während der Exploration wird, wie in vielen bisherigen Anwendungen, eine auf dem A*-Algorithmus basierende Kontrolle [Kum97] verwendet.

Nach einer Beschreibung der Wissensbasis in Abs. 2 wird in Abs. 3 die Bewertung für die Kameraaktionen vorgestellt. Abs. 4 erläutert die Bewertungsfunktionen der Konzepte, die der Szenenbeschreibung zuzuschreiben sind. Schließlich werden in Abs. 5 Experimente dargestellt, die die Tragfähigkeit des Ansatzes am Beispiel der Exploration einer Büroszene demonstrieren.

2 Die Wissensbasis

Zur *einheitlichen* Repräsentation des Wissens über die in die Aufgabe involvierten Objekte und die Kameraaktionen wird ein semantisches Netz verwendet, das auf dem ERNEST-Formalismus basiert [Sag85]. Als Anwendungsgebiet wird hier die Exploration einer statischen Büroszene gewählt. Momentan besteht die Aufgabe darin, drei Objekte, einen Locher, einen Klebestift und einen Abroller, in solch einer heterogenen Szene zu finden, wobei die Objekte anfänglich nur klein oder überhaupt nicht im Bild sind. Die Wissensbasis für das Anwendungsgebiet ist in Bild 1 dargestellt. Die Ovale bilden die Konzepte der Wissensbasis, die jeweils Attribute und Kanten enthalten. Die grauen Ovale stellen den Teil der Wissensbasis dar, der in den meisten konventionellen wissensbasierten Systemen [Han78,Mat90] zu finden ist und das Wissen über die Szene repräsentiert. Mit Hilfe der Konkretisierungskante werden Konzepte auf verschiedenen Abstraktionsebenen verbunden. Die Konzepte „Locher“ und „Farbregion“ bilden hierfür ein Beispiel. Die Konzepte der zu suchenden Objekte, z.B. „Locher“, werden als Bestandteile der Büroszene repräsentiert.

Zusätzlich zur Repräsentation der Szenenobjekte werden Konzepte für die Kameraaktionen in die Wissensbasis integriert (weiße Ovale). Auf der höchsten Abstraktionsebene entsprechen die Kameraaktionen „direkteSuche“ und „indirekteSuche“ Explorationsstrategien, um Objekte in der Szene zu suchen. So wird z.B. bei jeder Instanziierung des Konzepts „direkteSuche“ ein neuer Schwenkwinkel und eine neue Zoomeinstellung so berechnet, dass man Überblicksbilder für die Szene erhält. Durch Aneinanderreihen dieser Bilder erhält man ein Überblicksbild der kompletten Szene. Bei der Berechnung des Konzepts „indirekteSuche“ wird zunächst nach einem „Zwischenobjekt“ gesucht [Wix94], das sich in der Nähe des Zielobjekts befindet. Meistens werden hier mit einer kleinen Brennweite lokalisierbare Objekte, z.B. Tische, verwendet. Die momentan integrierte indirekte Suche repräsentiert diese Vorgehensweise beispielhaft an Hand der

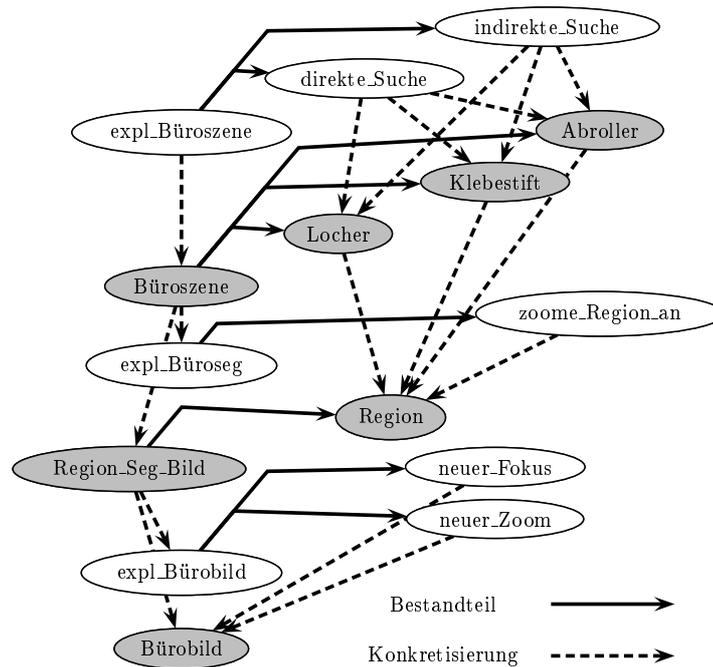


Bild1. Semantisches Netz, das die Repräsentation von Kameraaktionen (weiße Ovale) mit der Szenenrepräsentation (graue Ovale) kombiniert. Generalisierungen zu einigen Konzepten wurden aus Gründen der Übersichtlichkeit weggelassen.

in der Szenenwissensbasis enthaltenen Objekte. Die Kameraaktion „*zoomRegionAn*“ auf der mittleren Abstraktionsebene hat den Effekt, dass Regionen angezoomt werden, die potentielle Hypothesen für Objekte sind und für eine Verifikation zu klein sind. Auf der untersten Abstraktionsebene befinden sich datengetriebene Kameraaktionen wie z.B. die Einstellung des Fokus. Die Konzepte „*direkteSuche*“ und „*indirekteSuche*“ repräsentieren *konkurrierende* Kameraaktionen, d.h. Kameraaktionen, die nicht gleichzeitig in einem Analyseschritt ausgeführt werden sollen. Bei einer zeitgleichen Ausführung würde die Kamera unnötig hin- und herbewegt. Zur Repräsentation solcher konkurrierender Kameraaktionen werden Modalitätsmengen verwendet, die bei der Konzeptdefinition angegeben werden und eine kompakte Beschreibung von Varianten zur Bildung einer Instanz des entsprechenden Konzepts erlauben. So enthält das Konzept „*explBüroSzene*“ z.B. für seine Bestandteile „*direkteSuche*“ und „*indirekteSuche*“ je eine Modalitätsmenge. Gleiches gilt für die Konzepte „*Büroszene*“ und „*Reg-SegBild*“ mit entsprechenden Modalitätsmengen für ihre Konkretisierungen.

Während der *Analyse* werden zu den Konzepten sogenannte *Instanzen* berechnet und bewertet, die für die „*Szenenkonzepte*“ eine Realisierung des Konzepts in den Sensordaten darstellen und für die „*Kameraaktionskonzepte*“ die Berechnung einer neuen Kameraeinstellung oder Ausführung einer Kameraaktion bedeuten. Eine Instanz zu einem „*Szenenkonzept*“, z.B. für das Konzept *Locher*, entspricht einer *Hypothese* darüber,

ob ein Objekt, z.B. der Locher, gefunden wurde oder nicht, wobei die Bewertung der Instanz etwas über die Güte der Hypothese aussagt. Die Instanzen werden in *Suchbaumknoten* verwaltet, deren Bewertung sich aus den Instanzbewertungen ergibt und die den Suchraum für die A*-basierte Kontrolle bilden. Innerhalb der Suchbaumknoten wird die Reihenfolge der Instanzberechnungen mit Hilfe von in den Konzepten definierten *Prioritäten* festgelegt, womit z.B. sichergestellt ist, dass zunächst Instanzen zu allen Objekten vorliegen, bevor über eine Kameraaktion entschieden wird. Durch die Verwendung von Prioritäten wird erreicht, dass diese Sequentialisierung nicht aufgrund von Instanzbewertungen durchgeführt werden muß, was eine Vermischung von Kontrolle und Wissensrepräsentation zur Folge hätte. Für konkurrierende Instanzen, die aufgrund von mehrdeutigen Segmentierungsergebnissen entstehen, und für die durch Modalitäten entstehenden Mehrdeutigkeiten werden konkurrierende Suchbaumknoten gebildet. Zur Bewertung der Suchbaumknoten werden Bewertungsfunktionen für die Konzepte benötigt, die angeben wie *gut* die berechneten Instanzen sind. Diese Bewertungsfunktionen werden in den folgenden Abschnitten vorgestellt.

3 Kameraaktionsbewertung

Das Ziel der Ausführung der Kamerabewegungen besteht in dem Gewinnen von mehr Information über die Szene, um die Unsicherheit über die Zwischenergebnisse zu reduzieren. Die Kontrolle muß entscheiden, ob zu einem Zeitpunkt weitere Information benötigt wird und welche Kameraaktion eine Einstellung der Kamera liefert, mit der man die Information mit geringst möglichen Kosten erhält. Sie verwendet dazu die innerhalb eines Suchbaumknotens vorliegenden Instanzen, um abzuschätzen, inwieweit das Analyseziel erreicht ist. In dem hier betrachteten Anwendungsbeispiel sind z.B. Instanzen für die zu suchenden Objekte entscheidend, um aus den Kameraaktionen „indirekteSuche“ (*iS*) und „direkteSuche“ (*dS*) die optimale herauszusuchen oder sich für das Ende der Analyse „keineAktion“ (*kA*) zu entscheiden. Eine indirekte Suche soll nur dann ausgeführt werden, wenn der Klebestift und der Abroller gefunden wurden, während der Locher noch fehlt. In allen anderen Fällen soll eine direkte Suche angestoßen werden. Die Information, ob die in den Instanzen enthaltenen Hypothesen dem jeweils gesuchten Objekt entsprechen, spiegelt sich in der Bewertung der betreffenden Instanzen wider. Das heißt, es liegt für jede Instanz eine Hypothese vor, mit den Zuständen *Objekt gefunden* und *Objekt nicht gefunden* und einer entsprechenden Wahrscheinlichkeit. Abhängig von den Zuständen dieser Hypothesen wird die optimale Kameraaktion bestimmt.

Die Entscheidung über die optimale Kameraaktion wird anhand von Nützlichkeiten getroffen. Diese *Nützlichkeit* wird in der Entscheidungstheorie mit Hilfe von numerischen Werten gemessen, die in einer *Nützlichkeitstabelle* angegeben werden. Abgebildet auf den Bayesnetzformalismus ergeben sich die sogenannten Entscheidungsnetze, die Aktions- und Nützlichkeitsknoten enthalten [Jen96] und eine effiziente Berechnung der optimalen Aktion erlauben. Die Nützlichkeitstabelle enthält als Einträge die Werte der Funktion $U(a, h)$, die die Nützlichkeit einer Aktion a bei Vorliegen des Zustands h der Zufallsvariablen H angibt, wobei $a \in \mathcal{A}$ und \mathcal{A} die Menge der möglichen ausführbaren Aktionen bezeichnet. Außerdem ist im allgemeinen nur eine Wahrscheinlichkeits-

$I(L)$	$I(A)$	$I(K)$	dS	iS	kA
0	0	0	1	0	0
0	0	1	1	0	0
0	1	0	1	0	0
0	1	1	0	1	0
1	0	0	1	0	0
1	0	1	1	0	0
1	1	0	1	0	0
1	1	1	0	0	1

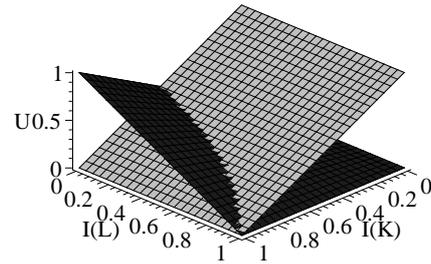


Bild2. Links: Nützlichkeitstabelle für die Kameraaktionen „indirekteSuche“ (iS), „direkteSuche“ (dS) und „keine Aktion“ (kA). Bei den Instanzen gibt der Eintrag die Hypothese an, ob ein Objekt gefunden wurde (Eintrag 1) oder nicht (Eintrag 0); Rechts: Gütefunktionen für „indirekteSuche“ (grauer Graph) und „direkteSuche“ (schwarzer Graph) in Abhängigkeit von der Bewertung für Instanzen von Locher und Klebestift, wobei die Bewertung für eine Instanz zu Abroller gleich Eins gesetzt wurde.

verteilung über H bekannt, die sich aus den Wahrscheinlichkeiten für das Vorliegen der Objekte in den Instanzbewertungen ergibt, jedoch nicht h selbst. Im Folgenden wird der Ansatz anhand der Kameraaktionen „indirekteSuche“ und „direkteSuche“ erläutert. So wird die Menge \mathcal{A} für dieses Beispiel von den beiden genannten Kameraaktionen gebildet. Als Hypothese H wird ein Vektor $(I(L), I(A), I(K))$ definiert, so dass man als Zustände alle Konfigurationen über das Vorliegen von entsprechenden Objekten in Form der Instanzen Locher $I(L)$, Abroller $I(A)$ und Klebestift $I(K)$ erhält. Es ergibt sich damit die in Bild 2 dargestellte Nützlichkeitstabelle. Die Zustände $h_1 \dots h_8$ der Zufallsvariablen H werden durch Angabe einer 1 für die Hypothese, dass ein Objekt gefunden wurde, oder 0, dass kein Objekt gefunden wurde, beschrieben. Zum Beispiel wurden für $(I(L), I(A), I(K)) = (0, 0, 0)$ weder Locher, noch Klebestift und Abroller gefunden. Als Nützlichkeiten werden hier momentan 0 und 1 verwendet, je nachdem ob eine Kameraaktion für ein h_i ausgeführt werden soll oder nicht. So wäre zum Beispiel in dem Zustand $(I(L), I(A), I(K)) = (0, 1, 1)$ eine Ausführung der direkten Suche zeitaufwendiger als die Ausführung der indirekte Suche, wobei beide die gleiche Information, Regionen, von denen eine als Locher zu interpretieren ist, liefern.

Da der Zustand der Hypothese h im allgemeinen nicht bekannt ist, läßt sich nur die *mittlere Nützlichkeit* EU einer Aktion a durch Summation über alle möglichen Zustände von H beim Vorliegen der Evidenz e berechnen, so dass sich

$$EU(a|e) = \sum_{h \in H} U(a, h)p(h|e) \quad (1)$$

ergibt. Die Evidenz e ist durch im entsprechenden Suchbaumknoten vorliegende Instanzen für Locher, Abroller und Klebestift bekannt. Es ergibt sich die Wahrscheinlichkeit $p(h|e)$ für den Zustand der Hypothese aus den Bewertungen der Instanzen (vgl. Abs. 4).

Für die Bewertung von „indirekteSuche“ und „direkteSuche“ ergeben sich dabei die in Bild 2 dargestellten Graphen. Je nach Wahrscheinlichkeit für das Auftreten einer Instanz für Locher und Klebestift wird entweder eine indirekte oder eine direkte Suche durchgeführt.

Die optimale Aktion ist diejenige, die die mittlere Nützlichkeit maximiert:

$$\text{opt}(a) = \operatorname{argmax}_a EU(a|e). \quad (2)$$

Der hier am Beispiel der Konzepte „indirekteSuche“ und „direkteSuche“ erläuterte Ansatz lässt sich auf die Bewertung aller Aktionen in der kompletten Wissensbasis übertragen. Für das Konzept „Büroszene“ kann die Entscheidung über ein Fortfahren der Analyse mit der in der Instanz „RegSegBild“ berechneten Segmentierung oder ein Anzoomen der segmentierten Regionen zum Beispiel basierend auf der Nützlichkeit der segmentierten Farbregionen getroffen werden.

4 Szenenbewertung

Für die Realisierung des entscheidungstheoretischen Ansatzes zur Auswahl der Kameraaktionen benötigt man Wahrscheinlichkeitsverteilungen für die Instanzen zu den Objekten (vgl. Abs. 3), aus denen sich dann die Wahrscheinlichkeitsverteilung für die Hypothese H ergibt. Diese Verteilungen geben an, mit welcher Wahrscheinlichkeit ein bestimmter Sensordatenausschnitt zu dem in einem Konzept repräsentierten Wissen paßt. Zur Integration des Bewertungskalküls in das semantische Netz werden gerade diese Wahrscheinlichkeiten als Bewertungen für Instanzen berechnet, z.B. zu dem Konzept „Locher“. Die Bewertung einer Instanz $I(C_k)$ zu einem Konzept C_k setzt sich aus den Bewertungen der Attribute $A_h^{(k)}$, der Relationen $R_i^{(k)}$, der Bestandteile $B_j^{(k)}$ ¹ und der Konkretisierungen $K_q^{(k)}$ des Konzepts C_k zusammen. Es ist somit naheliegend auch diese Netzwerkelemente mit Hilfe von Wahrscheinlichkeiten zu bewerten. Daraus ergibt sich folgender Ansatz für die Bewertung einer Instanz:

$$p(I(C_k) | A_{1,\dots,n}^{(k)}, R_{1,\dots,m}^{(k)}, B_{1,\dots,p}^{(k)}, K_{1,\dots,z}^{(k)}) = \alpha p(I(C_k)) \cdot \prod_{h=1}^n p(A_h^{(k)} | I(C_k)) \prod_{i=1}^m p(R_i^{(k)} | I(C_k)) \prod_{j=1}^p p(B_j^{(k)} | I(C_k)) \prod_{q=1}^z p(K_q^{(k)} | I(C_k))$$

Mit $p(I(C_k) | A_{1,\dots,n}^{(k)}, R_{1,\dots,m}^{(k)}, B_{1,\dots,p}^{(k)}, K_{1,\dots,z}^{(k)})$ wird die Wahrscheinlichkeit für das Auftreten einer Instanz zum Konzept C_k bezeichnet, unter der Bedingung, dass die Attribute, Relationen, Bestandteile und Konkretisierungen des Konzepts bestimmte Werte annehmen. Besitzt ein Konzept C_k z.B. keine Relationen wird der entsprechende Term 1; gleiches gilt für die anderen Netzwerkelemente. α bezieht sich auf den durch Anwendung der Bayesregel entstehenden Nenner, der für eine Maximierung unerheblich ist. Aufgrund fehlenden a-priori Wissens wird für die Wahrscheinlichkeiten $p(I(C_k))$ die Gleichverteilung angenommen. Außerdem wird vereinfachend vorausgesetzt, dass für die einzelnen Verteilungsdichten die Unabhängigkeitsannahme gilt. Die Bewertung der Instanzen für die gesuchten Objekte basiert in dem hier gewählten Anwendungsgebiet auf der Bewertung der drei Attribute Höhe, Breite und Farbe. Es gilt somit hierfür geeignete Verteilungsdichten, z.B. $p(A_h^{(L)} | I(L))$, $h = 1, \dots, 3$ für den Locher, zu trainieren.

¹ Hier wird vereinfachend von der Bewertung der Bestandteile statt der Bewertung der Instanzen der Bestandteile gesprochen. Gleiches gilt für die Konkretisierungen.

Hierzu wird eine Normalverteilung für die auftretenden Attributwerte angenommen, die sich in Experimenten als tragfähig erwiesen hat. Für jedes Objekt wird der Mittelwert und die Varianz der Verteilung anhand einer klassifizierten Trainingsmenge von 40 Aufnahmen bestimmt. Während der Analyse werden anhand der trainierten Dichte die Wahrscheinlichkeiten für das Auftreten eines bestimmten Attributwertes x berechnet. Diese Wahrscheinlichkeiten werden vereinfachend durch

$$p(x) \approx \exp\left(\frac{(x - \mu)^2}{-2\sigma^2}\right) \quad (3)$$

approximiert. Vergleichbare Ansätze zur Bewertung von Segmentierungsergebnissen finden sich in [Kre98,Hor96].

5 Experimente

Zum gegenwärtigen Zeitpunkt der Entwicklung wurden Experimente mit dem Teil der Wissensbasis durchgeführt, der das Wissen über die Szene („Büroszene“) und über die Kameraaktion auf der obersten Abstraktionsebene („direkteSuche“ und „indirekteSuche“) enthält. Der untere Teil der Wissensbasis in Bild 1 wird als ein eigenständiges Modul zur Verfügung gestellt [Pau98]. Dieser Teil enthält die Konzepte „Bürobild“, „neuerZoom“, „neuerFokus“, „explBürobild“, „explBüroSeg“ und „zooomeRegionAn“. In diesem Modul werden die Hypothesen für rote Objekte mit Hilfe einer Histogramm-Rückprojektion berechnet [Swa91], die auf ein Überblicksbild angewendet wird, das mit minimaler Brennweite aufgenommen wurde. Um diese Hypothesen zu verifizieren, werden sie durch eine Veränderung der Schwenkposition und Variieren der Brennweite der Kamera fovealisiert, d.h. formatfüllend im Bild dargestellt. Dies ist genau die Aufgabe des unteren Teils der Wissensbasis aus Bild 1. Das primitive Konzept der Wissensbasis in der Experimentierumgebung ist somit das Konzept „Farbregion“. An dieses Konzept werden Farbregionen gebunden, die auf der Basis der Bilder mit den fovealisierten Objekten berechnet wurden. Bei den 15 Experimenten befanden sich durchschnittlich sieben verschiedene rote Objekte in der Büroszene, von denen nur drei in der Wissensbasis modelliert waren. Die Positionen der Objekte in der Szene waren in jedem Experiment unterschiedlich. Die drei für die Verifikation interessanten Objekte wurden von den datengetriebenen Modulen unter Verwendung der Histogramm-Rückprojektion in 40 von 45 möglichen Fällen gefunden. Durchschnittlich wurden sieben Bilder mit fovealisierten Objekten generiert, d.h. sieben Objekthypothesen werden in einem Übersichtsbild gefunden. In 85 % der Fälle wurde eine Farbregion dem korrekten Objekt zugeordnet. Durch Fehler in der Zuordnung wurde in drei Fällen die falsche Kameraaktion ausgelöst.

Ein Verarbeitungszyklus zur Analyse für eine Kameraeinstellung (d.h. von der datengetriebenen Generierung bis zur Berechnung der optimalen Instanz von „explOffice“) dauert ca. fünf Minuten. Die meiste Zeit wird für die Bewegung der Kamera und das Warten bis zum Erreichen der Zielposition, für die Median-Filterung bei der Rückprojektion und die Segmentierung der Farbregionen benötigt. Die Verifikation der hypothetisierten Farbregionen mit der A*-Kontrolle dauert abhängig von der gefundenen Zahl an Segmentierungshypothesen zwischen 0.14 und 0.66 sec.

6 Zusammenfassung und Ausblick

In diesem Artikel wurde eine *einheitliche* Repräsentation des Wissens über eine Szene und die Kameraaktionen zur Exploration einer Szene vorgestellt. Um aus Kameraaktionen die optimale auswählen zu können, wird ein auf Grundlagen der Entscheidungstheorie basierendes Bewertungskalkül verwendet. Dieses basiert auf der Berechnung von *Nützlichkeiten* für die verschiedenen Kameraaktionen, die von der Bewertung für Instanzen der Objekte aus der Szenenrepräsentation abhängen. Diese werden mit Hilfe von trainierten Normalverteilungen bewertet.

Weitere Arbeiten konzentrieren sich auf die Implementierung des unteren Teils der Wissensbasis für das präsentierte Anwendungsgebiet. Dies beinhaltet die Erweiterung des Bewertungskalküls auf alle, in dem semantischen Netz enthaltenen Aktionen. Weiterhin muß noch mit Hilfe eines statistischen Tests überprüft werden, inwiefern die Normalverteilungsannahme für die Verteilungsdichten zur Attributbewertung zutrifft. Außerdem wird eine Reduktion der Verarbeitungszeit angestrebt.

Literatur

- [Alo88] Aloimonos, J.; Weiss, I.; Bandyopadhyay, A.: *Active Vision*, *International Journal of Computer Vision*, Bd. 2, Nr. 3, 1988, S. 333–356.
- [Han78] Hanson, A.; Riseman, E.: *VISIONS: A Computer System for Interpreting Scenes*, in Hanson, A.; Riseman, E. (Hrsg.): *Computer Vision Systems*, Academic Press, Inc., New York, 1978, S. 303–333.
- [Hor96] Hornegger, J.; Nöth, E.; Fischer, V.; Niemann, H.: *Semantic Network Meet Bayesian Classifiers*, in Jähne, B.; Geißler, P.; Haußecker, H.; Hering, F. (Hrsg.): *Mustererkennung 1996*, Springer, Berlin, September 1996, S. 260–267.
- [Jen96] Jensen, F. V.: *An Introduction to Bayesian Networks*, UCL Press, London, 1996.
- [Kre98] Krebs, B.; Korn, B.; Wahl, F.: *A task driven 3d object recognition system using bayesian networks*, in *International Conference on Computer Vision*, Bombay, India, 1998, S. 527–532.
- [Kum97] Kummert, F.; Fink, G.; Sagerer, G.: *Schritthaltende hybride Objektdetektion*, in Paulus, E.; Wahl, F. (Hrsg.): *Mustererkennung 1997*, Braunschweig, 1997, S. 137–144.
- [Lev89] Levitt, T.; Binford, T.; Ettinger, G.; Gelband, P.: *Probability Based Control for Computer Vision*, in *Proc. of DARPA Image Understanding Workshop*, 1989, S. 355–369.
- [Mat90] Matsuyama, T.; Hwang, V.: *SIGMA. A Knowledge-Based Aerial Image Understanding System*, Bd. 12 von *Advances in Computer Vision and Machine Intelligence*, Plenum Press, New York and London, 1990.
- [Pau98] Paulus, D.; Ahlrichs, U.; Heigl, B.; Niemann, H.: *Wissensbasierte aktive Szenenanalyse*, in Levi, P. (Hrsg.): *Mustererkennung 1998*, Springer, Heidelberg, September 1998, S. 185–192.
- [Rim93] Rimey, R.: *Control of Selective Perception using Bayes Nets and Decision Theory*, Department of Computer Science, College of Arts and Science, University of Rochester, Rochester, New York, 1993.
- [Sag85] Sagerer, G.: *Darstellung und Nutzung von Expertenwissen für ein Bildanalyse-system*, Bd. 104 von *Informatik-Fachberichte*, Springer-Verlag, Berlin, 1985.
- [Swa91] Swain, M. J.; Ballard, D. H.: *Color Indexing*, *International Journal of Computer Vision*, Bd. 7, Nr. 1, November 1991, S. 11–32.
- [Wix94] Wixson, L.: *Gaze Selection for Visual Search*, Department of Computer Science, College of Arts and Science, University of Rochester, Rochester, New York, 1994.