

# Differenzbasierte Repräsentation räumlicher Relationen zur probabilistischen Szenenerkennung mittels hierarchischen Constellation Models

Bachelorarbeit  
von

Joshua Enrico Link

An der Fakultät für Informatik  
Institut für Anthropomatik und Robotik  
Lehrstuhl Prof. Dr.-Ing. R. Dillmann

Erstgutachter:	Prof. Dr.-Ing. R. Dillmann
Zweitgutachter:	???
Betreuender Mitarbeiter:	Dipl.-Inform. Pascal Meißner

Bearbeitungszeit: 11. Juni 2017 – 10. September 2017

Hiermit erkläre ich an Eides statt, dass ich die von mir vorgelegte Arbeit selbstständig verfasst habe, dass ich die verwendeten Quellen, Internet-Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen – die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Karlsruhe, den **(Datum)**

**ToDo**

---

Joshua Enrico Link

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>iv</b>
<b>1 Einführung</b>	<b>1</b>
<b>2 Motivation und Problemstellung</b>	<b>2</b>
2.1 Motivation . . . . .	2
2.2 Fokus der Arbeit . . . . .	3
<b>3 Grundlagen</b>	<b>5</b>
3.1 PSM(Probabilistic Scene Model) . . . . .	5
3.2 Datengetriebene Entwicklung . . . . .	5
<b>4 Konzept</b>	<b>6</b>
4.1 Ansatz . . . . .	6
4.2 Erkennungsalgorithmus . . . . .	7
4.2.1 Ausformuliert . . . . .	7
4.3 Wahrscheinlichkeitsabschätzung . . . . .	7
<b>5 Implementierung</b>	<b>9</b>
5.1 Umbau PSM . . . . .	9
5.1.1 Klassenaustausch . . . . .	9
5.1.2 Ersatzfunktionen . . . . .	9
5.1.3 Datenbankeinbindung . . . . .	9
5.2 Differenzbasierter Erkennungsalgorithmus . . . . .	10
5.2.1 Algorithmus . . . . .	10
5.2.2 Einbindung PSM . . . . .	10
<b>6 Evaluation</b>	<b>11</b>
6.1 Experiment 1: Frühstück . . . . .	11
6.2 Experiment 2: Office . . . . .	11
<b>7 Zusammenfassung und Ausblick</b>	<b>12</b>
<b>Literaturverzeichnis</b>	<b>13</b>

# Abbildungsverzeichnis

2.1	Beispiel: Relative Position eines Objekts zu einem anderen . . . . .	3
3.1	eine Grafik ohne Sinn und Verstand . . . . .	5
4.1	Beispiel: Kaffeetasse - Ausreißer in den Daten . . . . .	7
4.2	Algorithmus als True False Diagramm . . . . .	8

# 1. Einführung

In der Robotik ist die Servicerobotik wohl der Forschungsbereich, welcher den größten Alltagsbezug für den Menschen hat, da er sich mit der Entwicklung und Weiterentwicklung von autonomen Robotern beschäftigt, welche dem Menschen im Alltag assistieren. Man findet mittlerweile Roboter im Privaten, die das Putzen, Staubsaugen oder Rasenmähen übernehmen, in der Industrie, bei Montage und Fertigung, sowie auch in der Medizin, als Pflegehilfe, Botengänger oder Assistent.

Allerdings müssen die Roboter ihre Umwelt für komplexere Aufgaben so präzise wie möglich wahrnehmen und verstehen. Sie könne Aufgaben übernehmen bei denen sie gezielt Objekte umfahren, suchen und auch aufnehmen und benutzen. Dieser Funktionsumfang kann mit dem Prizip Programmieren durch Vormachen (PdV) ermöglicht werden, bei dem die Roboter Objekte und Tätigkeiten ihrer Umgebung kennen lernen, wieder erkennen und nachahmen können. So lässt sich die hohe Komplexität umgehen, die die manuelle Programmierung vieler Aufgaben mit sich bringen würde.

Um tatsächlich selbstständige Serviceroboter zu schaffen muss man aber noch zu einer Objekterkennung ein Kontextverständnis hinzufügen. Die Roboter müssen erkannte Objekte in einen Zusammenhang bringen, um die dadurch resultierenden Aufgaben zu verstehen. Zum Beispiel hat ein Teelöffel, welcher neben einer Tasse Tee liegt eine andere Aufgabe zu verrichten, als wenn er neben einem Becher Joghurt platziert ist. Nur am Kontext lässt sich dort entscheiden warum im einen Fall umgerührt und im anderen gelöffelt wird. Ebenso wäre ein Stück Butter verschieden zu verwenden, wenn es auf einem Frühstückstisch steht als wenn es mit anderen Zutaten neben einer Rührschüssel vorkommt.

Somit braucht man eine Szenenerkennung, welche zuverlässig die Objekte erkennen und ihren jeweiligen Kontext verstehen und einschätzen kann. Diese Erkennung ist nicht immer eindeutig, da der eben genannte Löffel ebenso zwischen einem Becher Joghurt und einer Tasse Tee liegen könnte, deshalb bietet es sich an mit Wahrscheinlichkeitsabschätzungen des vorliegenden Kontexts zu arbeiten.

## 2. Motivation und Problemstellung

Das Kapitel geht in Motivation auf die Relevanz des Themas und der vorliegenden Arbeit ein und in Fokus der Arbeit auf die Problemstellung die bearbeitet wird, sowie diverse Einschränkungen und Annahmen die für die Arbeit festgelegt sind.

### 2.1 Motivation

Wie schon in der Einführung erwähnt ist es elementar wichtig, dass man eine zuverlässige Szenenerkennung und ein gutes Kontextverständnis schafft um den Robotern die Möglichkeit zu bieten, sinnvoll mit ihrer Umwelt zu interagieren. Mit einer präzisen Wahrscheinlichkeitseinschätzung wie die momentane Umgebung beschaffen ist, lässt sich abschätzen welche Aufgaben es zu bewältigen und welche Probleme zu lösen gilt. Um dies zu gewährleisten muss man in das System möglichst viele Referenzdaten einspeisen, damit es jeden vorhanden Kontext erkennen kann. Szenen werden zu diesem Zweck aufgebaut um der Erkennung als neue Szene vorgestellt. Jede Szene die die Szenenerkennung auf diese Weise lernt, hilft das Chaos der sie umgebenen Objekte mehr und mehr zu interpretieren und einzuordnen. Natürlich könnte man das System auch dynamisch mit jedem auftreten eines Objekts neue Kontexte beziehungsweise Szenen lernen lassen, das würde allerdings die Erkennung an Präzision einbüßen lassen.

Die Szenenerkennung nutzt also die Daten die sie zur Verfügung gestellt bekommt, bereits gelernte Szenen wiederzuerkennen. In dem bereits vorhandenen PSM(Probabilistic Scene Model)-System werden die Daten pro Szene zu einem Modell zusammengefasst, bei dem Auffällige Zusammenhänge berücksichtigt werden und scheinbar nicht miteinander in Verbindung stehende Objekte voneinander gelöst betrachtet werden. Zum Beispiel findet man eine Computermouse signifikant häufig vor dem Computerbildschirm und nie dahinter, allerdings kann dabei das räumliche Verhältnis der Maus zur Tastatur stark variieren. Die Maus wäre in diesem Beispiel manchmal direkt neben der Tastatur, manchmal dichter beim Bildschirm eben so oft weiter entfernt. In diesem Fall würde möglicherweise die Relation zwischen Maus und Tastatur wegfallen und nur jeweils das

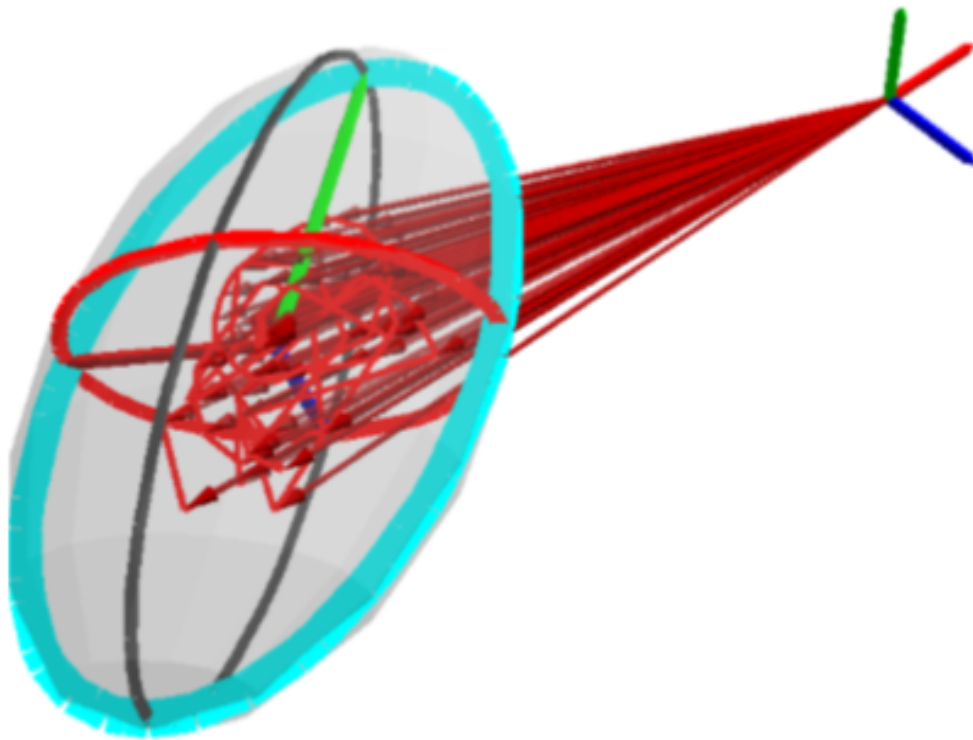


Abbildung 2.1: Beispiel: Relative Position eines Objekts zu einem anderen

räumliche Verhältnis zum Bildschirm betrachtet werden. Dadurch gibt es Vorteile in der Laufzeit und möglicherweise auch eine signifikantere Erkennung, allerdings findet natürlich auch ein Informationsverlust statt, der zu Fehlern führen kann. In dieser Arbeit wird ein Ansatz getestet, der dichter an den erhaltenen Daten arbeitet.

In Abbildung 2.1 sieht man wie das Modell eine Relation zwischen zwei Objekten aufgrund der erhaltenen Daten erstellt. Die roten Pfeile beschreiben hier die jeweiligen Relationen, die aus den Daten berechnet wurden und die Ellipsen zeigen, welchen Bereich das Modell als Basis für die Wahrscheinlichkeitsabschätzung benutzt. [?]

## 2.2 Fokus der Arbeit

Ziel der vorliegenden Arbeit ist es das PSM-System zu überarbeiten und einen neuen Modus der Szenenerkennung im PSM-System zu entwickeln. Dieser Modus soll alternativ zu den bereits vorhandenen Modi auswählbar sein und das System erweitern. Außerdem soll sowohl die Positionierung als auch die Rotation der erkannten Objekten mit den bekannten Szenen verglichen werden um eine Wahrscheinlichkeitsabschätzung auszugeben, welche die Wahrscheinlichkeit aller möglichen Szenen ausgibt sowie auch die Wahrscheinlichkeit, dass es sich um keine der Szenen handelt.

Um eine interaktive Szenenerkennung zu schaffen ist es sinnvoll das System datengetrieben zu programmieren. Datengetriebene Entwicklung zeichnet sich dadurch aus, dass sich der Programmfluss ändert aufgrund der Daten die das System während der Laufzeit als Eingabe erhält. Das PSM-System wurde bereits datengetrieben programmiert und in der zu dieser Arbeit gehörigen Implementierung wurde auch am datengetriebenen Entwicklungsmodell festgehalten.

Da die Erkennung innerhalb des PSM-Systems nutzbar sein soll sind Eingabe- und Ausgabeschnittstellen sowie die Visualisierung vorgeschrieben und sollen für gute Vergleichbarkeit denen des vorhandenen Systems entsprechen.

Die Arbeit ist wie folgend strukturiert. In 3 Grundlagen wird auf die Grundkenntnisse eingegangen die man braucht um den Rest der Arbeit gut zu verstehen. Vorallem wird in diesem Kapitel das bestehende PSM-System erklärt. In 4 Konzept wird das theoretische Konzept thematisiert, welches zum Algorithmus geführt hat, dass den neuen Modus vom alten System unterscheidet. Außerdem wird der Algorithmus selbst erläutert. In 5 Implementierung wird die Software beschrieben und dokumentiert die im Zuge dieser Arbeit entstanden ist sowie die Änderungen die am bestehenden PSM System vorgenommen wurden. 6 Evaluation beschreibt die durchgeführten Experimente und interpretiert sie. In 7 Zusammenfassung und Ausblick wird die Arbeit noch einmal zusammen gefasst und ein Ausblick darauf gegeben in welche Richtung man das bestehende System weiter entwickeln und auf welche Weise man möglicherweise die Szenenerkennung noch weiter verbessern kann. Am Schluss stehen die Quellen welche zur Recherche für diese Arbeit genutzt wurden.



## 3. Grundlagen

### 3.1 PSM(Probabilistic Scene Model)

### 3.2 Datengetriebene Entwicklung

Länge max. halb so lang wie Konzept + Implementierung

simpel beschreiben

konkret:

bestehendes System : PSM

Relevanz erklären?

Datengetriebene Entwicklung erklären

Viele Bilder benutzen, auch aus Joachims Arbeit

Auch aus Joachims Arbeit

[?]



Abbildung 3.1: eine Grafik ohne Sinn und Verstand

## 4. Konzept

Das folgende Kapitel thematisiert das Konzept, dass im Zuge der vorliegenden Arbeit entwickelt wurde, um die vorgestellte Problemstellung zu lösen. Zwischen vielen verschiedenen möglichen Ansätzen einen neuen Modus für das PSM System zu entwickeln, entschied ich mich für einen differenzbasierten Vergleich der Positions- und Rotationsrelationen. Dieser wird im ersten Abschnitt erläutert und anschließend wird der Algorithmus sprachlich, grafisch und in Pseudocode erklärt. Zum Schluss wird die Wahrscheinlichkeitsabschätzung für den Algorithmus erklärt und begründet.

### 4.1 Ansatz

Im vorhandenen PSM-System werden im Learner die erhaltenen Positions- und Rotationsdaten zu einem Modell zusammengefasst, dass die Vorkommen aller Objekte in Relation zueinander zusammenfasst. Dieser Vorgang führt dazu, dass teilweise Zusammenhänge in den Daten betont werden, aber auch zu einem Informationsverlust da Ausreißer und mutmaßliche Fehlmessungen dadurch verloren gehen. Deshalb kann es sinnvoll sein direkt auf den gemessenen Daten zu arbeiten, um ein Ergebnis zu erhalten welches alle Daten berücksichtigt.

Wir betrachten folgendes Szenario. Ein Roboter soll seine Aufgaben aufgrund von einer Szenenerkennung einschätzen und durchführen. In der Szene "Kaffee" gibt es eine volle Kaffeetasse und einen Teelöffel und seine Aufgabe ist es mit dem Löffel den Kaffee umzurühren. In seinen Referenzdaten zu der Szene war der Löffel meist direkt neben der Tasse und nur in einem Fall ein Stück weiter entfernt. Allerdings gilt jede einzelne aufgezeichnete Referenzszene auf äquivalente Weise als Beispiel für die Szene "Kaffee". Das Parametermodell würde diese Ausreißerdaten allerdings glätten und kaum berücksichtigen, sodass der Roboter die Szene selbst mit genau dem Aufbau aus den Referenzdaten möglicherweise nicht erkennen würde. Wenn man allerdings die Erkennung direkt auf den Referenzdaten basiert, erkennt die Szenenerkennung den Ausreißer auch, da sie ja eine Instanz der Szene mit diesem vergleicht, welche diesem entspricht.

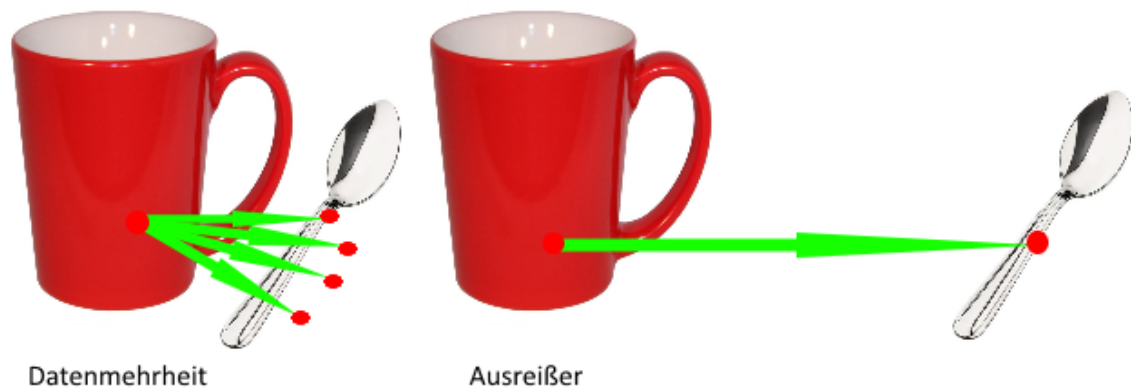


Abbildung 4.1: Beispiel: Kaffeetasse - Ausreißer in den Daten

Abbildung 4.1 verdeutlicht das genannte Szenario. Die roten Punkte stehen für die gemessenen Positionen und die grünen Pfeile stehen für die räumliche Relation zwischen den Objekten. Links sieht man, dass die meisten Messungen einen kleinen Abstand zwischen Löffel und Tasse haben, rechts ist der Ausreißer dargestellt, der möglicherweise vom alten System nicht als die gelernte Szene erkannt wird.

Im differenzbasierten Modus sollen also gemessene Objekte direkt mit den Referenzdaten verglichen werden, die das System bereits gelernt hat. Der Algorithmus betrachtet alle Objekte vollvermascht, sodass er die Szenenreferenz findet, die die maximale Ähnlichkeit zu den gemessenen Objekten hat. Darauf basierend wird die Wahrscheinlichkeit abgeschätzt, dass die gemessenen Objekte die Referenzszenen enthalten oder repräsentieren. Dabei stören zusätzliche Objekte die Erkennung nicht und eine Unvollständigkeit der Szene führt zu einer kleineren Wahrscheinlichkeit aber nicht zu direkter Ablehnung, da die Szene ja noch durch weitere Objekterkennungen vervollständigt werden könnte.

## 4.2 Erkennungsalgorithmus

### 4.2.1 Ausformuliert

## 4.3 Wahrscheinlichkeitsabschätzung

komplexer mathematischer formulieren  
 Vergleichsbasierte Erkennung erklären  
 Stochastische Richtigkeit beweisen

[DSS93]

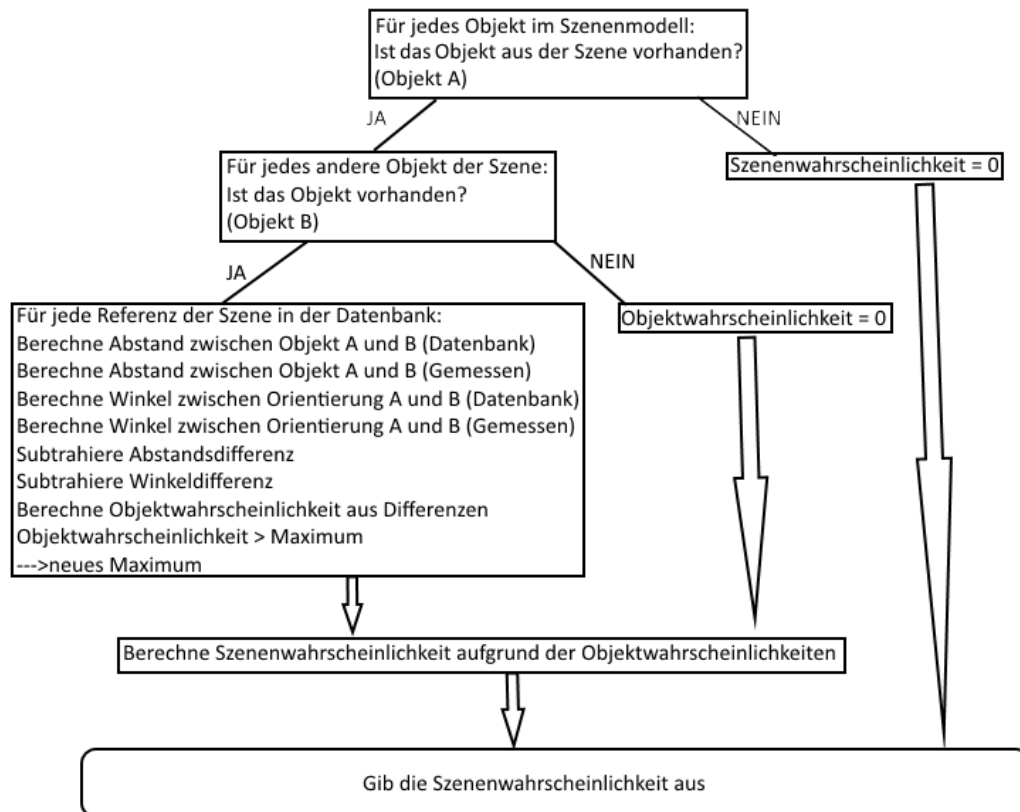


Abbildung 4.2: Algorithmus als True False Diagramm

## 5. Implementierung

Im Kapitel Implementierung wird alles beschrieben und erklärt, was am bestehenden PSM-Projekt verändert und hinzugefügt wurde. Außerdem werden ausgewählte hinzugefügte und veränderte Klassen sowie launch-Dateien dokumentiert, sodass das Kapitel das Verständnis und die Nutzung der Neuheiten im System vereinfacht. Nachdem der Umbauprozess beschrieben wird, bei dem eine Klasse komplett aus dem PSM-Projekt ausgetauscht wurde, widmet sich das Kapitel der Umsetzung des Algorithmuskonzepts und der Einbettung in das vorhandene System.

### 5.1 Umbau PSM

Da das Paket "pbd\_msgs" nicht kostenlos zur Verfügung gestellt wird, mussten alle Vorkommen der Klassen aus diesem Paket zu alternativen Ersatzklassen geändert werden. Teilweise konnte man dies durch simple Ersetzung erreichen, allerdings gab es nicht für jede Klasse eine Ersatzklasse mit dem selben Funktionsumfang. In den Fällen, in denen Funktionen fehlten oder geringfügig anders funktionierten, konnte man den Umbau durch kleine Anpassungen erreichen oder musste eigene Funktionen schreiben, welche die nötigen Operationen verrichten konnten. Alle auf diese Weise programmierten Funktionen wurden hinreichend auf Gleichheit mit ihren Ursprungsfunktionen in ihrer Funktionsweise getestet, indem die Ergebnisse bei gleichen Eingangsparametern abgeglichen wurden. Außerdem habe ich eine Datenbankschnittstelle für das PSM-System hinzugefügt.

#### 5.1.1 Klassenaustausch

#### 5.1.2 Ersatzfunktionen

#### 5.1.3 Datenbankeinbindung

Im vergleichbaren ISM Projekt, welches eine ähnliche Zielsetzung wie das ISM Projekt hat, allerdings einen nicht stochastischen Ansatz verfolgt, kann man die Szenen, die

das System lernen soll aus einer Datenbank auslesen. Aus Gründen der Vergleichbarkeit wie auch dem Komfort ist es sinnig, diese Datenbankschnittstelle für das PSM System nachzurüsten. Um dies zu erreichen wurde ein Datenbankpfad in die launch Datei des Learners(learner.launch) hinzugefügt und der veraltete rosbagfiles Parameter damit ersetzt. Die Datenbank wird innerhalb der SceneLearningEngine Klasse ausgelesen und die erhaltenen Daten, welche jeweils Vorkommen und von verschiedenen Objekten in diversen Szenen repräsentieren, werden anschließend konvertiert, sodass sie für das System sinnvolle AsrObject-Instanzen werden.

## 5.2 Differenzbasierter Erkennungsalgorithmus

Einerseits beschreibt dieses Kapitel, wie der im Konzept vorgestellte Algorithmus programmiertechnisch umgesetzt, andererseits wie dieser in das bestehende System eingebunden wurde.

### 5.2.1 Algorithmus

### 5.2.2 Einbindung PSM

Da davon abgesehen wird die Parametrisierung des Szenenmodells zu nutzen, braucht man den Learner und das Szenenmodell nur noch zu dem Zweck, dass überliefert wird, welche Szenen potentiell erkannt werden sollen. Die einzigen Informationen die der algorithmus nutzt sind der Name der Szene und Anzahl der enthaltenen Objekte. In der Launch Datei der Erkennung(inference.launch) wurde ein neuer Parameter hinzugefügt um den Pfad der Datenbank anzugeben. ...

alle Klassen die umgebaut wurden  
neuer differencebased modus

## 6. Evaluation

### 6.1 Experiment 1: Frühstück

### 6.2 Experiment 2: Office

Viele Bilder, beschreiben Daten  
Text interpretiert  
Fazit am Ende

[DSS93]

## 7. Zusammenfassung und Ausblick

Zwei Sätze zu jedem größeren Kapitel

[DSS93]



# Literaturverzeichnis

- [DSS93] Randall Davis, Howard Shrobe und Peter Szolovits: *What is a Knowledge Representation?* AI Magazine, 14(1):17–33, 1993. <http://www.aaai.org/ojs/index.php/aimagazine/article/view/1029>.