

Hierarchische Modelle für das visuelle Erkennen und Lernen von Objekten, Szenen und Aktivitäten*

Jens Spehr

Institut für Robotik und Prozessinformatik
Technische Universität Braunschweig
jensspehr@gmx.de

Abstract: In zahlreichen Computer Vision Anwendungen müssen Objekte in einzelnen Bildern oder Bildsequenzen erlernt und erkannt werden. Viele dieser Objekte sind hierarchisch aufgebaut. In dieser Arbeit werden neue probabilistische hierarchische Modelle vorgestellt, die es ermöglichen mehrere Objekte verschiedener Kategorien, Skalierungen, Rotationen und aus verschiedenen Blickrichtungen effizient zu repräsentieren. Eine Idee ist hierbei, Ähnlichkeiten unter Objekten, Objektteilen oder auch Aktionen und Bewegungen zu nutzen, um redundante Informationen und Mehrfachberechnungen zu vermeiden. Das einheitliche Framework ermöglicht die Anwendung des vorgestellten Modells für die klassische Objekterkennung, die Erkennung von menschlichen Posen, die Aktivitätenerkennung und die Umfeldwahrnehmung von intelligenten Fahrzeugen. Eine detaillierte Ausführung dieser Arbeit ist in [Spe13] zu finden.

1 Einleitung

Das visuelle Erkennen und Lernen von Objekten, Szenen und Aktivitäten in bzw. aus Bildern oder Bildsequenzen ist eine sehr relevante, aber gleichzeitig auch sehr schwierige Aufgabe, die Anwendungen in vielen Bereichen wie Robotik, Industrie, Unterhaltungselektronik, Luft- und Raumfahrt, Transportsysteme oder Ambient Assisted Living hat. In industriellen Szenarien kann die Standard Computer Vision Aufgabe des Erfassens und Erkennens von Objekten aus Bilddaten meist mit einem einfachen Schwellwertverfahren und geeigneten Formdeskriptoren gelöst werden. In der realen Welt müssen hingegen mehrere Ansichten, verschiedene Artikulationen und Skalen beachtet und außerdem verschiedene Erscheinungsmerkmale für eine robuste Erkennung kombiniert werden. Darüber hinaus sollte die Repräsentation eine Invarianz gegenüber der typischen Bildverarbeitungsprobleme wie Beleuchtungsänderungen, Hintergrundstörungen oder Verdeckungen bieten. Die erkannten Objekte können zum Beispiel für die inhaltsbasierte Bildsuche oder in einer Augmented-Reality-Anwendung verwendet werden, bei der die Ansicht einer realen Umgebung durch computergenerierte Grafiken ergänzt wird. Andere derzeit sehr interessante

*Englischer Titel der Dissertation: „On Hierarchical Models for Visual Recognition and Learning of Objects, Scenes, and Activities“

Beispiele sind Ambient Assisted Living Anwendungen. Sie sind besonders im Hinblick auf die Alterung der Gesellschaft, die die Unterstützung der älteren Menschen in ihrer häuslichen Umgebung mehr und mehr notwendig macht, interessant. Auf der einen Seite steigt die Zahl der älteren Menschen, auf der anderen Seite nimmt die Zahl der Menschen, die Hilfe anbieten können, ab. Durch den damit verbundenen Rückgang an Pflegekräften werden freie Plätze in Altenheimen knapp. Ein vielversprechender Weg, dieses Dilemma zu lösen, ist es, älteren Menschen zu ermöglichen, so lange wie möglich in ihrer gewohnten häuslichen Umgebung zu bleiben, wobei gleichzeitig die Gesundheit der Person mit Hilfe eines Überwachungssystems sichergestellt wird. Solch ein Überwachungssystem extrahiert Schätzungen der Vitalparameter wie Gehgeschwindigkeit und erkennt kritische Werte und Veränderungen. Visuelle Sensoren wie Kameras haben den Vorteil, unaufdringlich und unsichtbar für den Benutzer zu sein, denn sie benötigen keine explizite Interaktion zwischen Benutzer und System. Solche visuellen Sensorsysteme können verwendet werden, um plötzlich eintretende kritische Ereignisse wie Stürze zu erfassen. Hierfür benötigt der visuelle Sensor jedoch ein effizientes Bildverarbeitungsframework, das Merkmale wie Körpergröße oder Körperachse aus den Bilddaten extrahiert. Darüber hinaus können vision-basierte Ansätze auch zur Sturzprävention verwendet werden. Durch eine Ganganalyse können Merkmale wie Schrittweite, Schritthöhe, Geschwindigkeit und Geschwindigkeitsvarianz, Gangharmonie, Ausgleichsbewegung und Körperschwankung extrahiert und zur Einschätzung, wie kritisch die aktuellen Gangparameter sind, verwendet werden. Eine noch größere Herausforderung ist die Repräsentation von menschlichen Verhaltensmustern und die Erkennung von Anomalien, die häufig auf gesundheitliche Probleme hindeuten.

Ein weiteres vielversprechendes Anwendungsgebiet ist das autonome Fahren eines intelligenten Fahrzeugs, welches eine robuste sensorbasierte Wahrnehmung der Umwelt erfordert. Durch 3D-Rekonstruktionstechniken wie 'structure from motion' stellt eine Kamera einen sehr leistungsfähigen 3D-Sensor dar. Allerdings ist die Interpretation der gesammelten Daten immer noch sehr anspruchsvoll und erfordert aufwendige Ansätze, die ein 'high-level' Verständnis der Szene bezüglich der Erkennung der Straße, anderen Fahrzeugen, Fußgängern, Parkplätzen und so weiter ermöglichen.

In dieser Arbeit werden hierarchische Modelle vorgestellt, die ideal für das Lösen der zuvor beschriebenen Herausforderungen geeignet sind. Objekte können in Teile zerlegt werden, diese Teile in visuelle Grundelemente und schließlich die Grundelemente in lokale Gradienten und Farbwerte. Statt Elemente verschiedener Ansichten oder Konfigurationen unabhängig voneinander zu modellieren, können Elemente untereinander geteilt werden, um so die Gesamtkomplexität der Modelle und die Anzahl der Teile zu reduzieren. Dies führt zu besseren Generalisierungseigenschaften und zu einer effizienten Darstellung, die den Rechenaufwand während des Lern- und Erkennungsprozesses reduziert. Es ermöglicht auch die vorgeschlagenen hierarchischen Modelle in realen Anwendungen in Echtzeit anzuwenden. Durch das Teilen von Informationen werden redundante Informationen vermieden und dadurch unnötige Mehrfachberechnungen reduziert. Dieses Teilen von Informationen ist auch besonders für Gelenkstrukturen, wie dem menschlichen Körper, geeignet, da hier der Grad des Teilens für verschiedene Konfigurationen sehr hoch ist. Ähnliche Ansätze zur hierarchischen Zerlegung und dem Teilen von Informationen können auch für Aktivitäten verwendet werden. Diese sind in Aktionen zerlegbar und Aktionen wie-

derum in Aktionsprimitive. Auch Verkehrsszenen lassen sich in Objekte wie Autos und Fußgänger zerlegen und diese wiederum in einfache Grundelemente.

2 Überblick über die Beiträge dieser Arbeit

Die wichtigsten Beiträge der neuen vorgeschlagenen hierarchischen Methoden und Modelle sind:

- Wir stellen eine hierarchische Repräsentation von blickpunktabhängigen Instanzen vor und bilden die Skalierung und Rotation der Instanzen direkt in der Hierarchie ab. Unsere Hierarchie ist weniger eingeschränkt als andere Darstellungen in der Literatur, d.h. unsere Hierarchien haben bis zu 20 Ebenen und flexible Abhängigkeiten zwischen den Schichten. Diese Flexibilität erlaubt es, die Wiederverwendbarkeit von Teilen zu maximieren und führt damit zu einer kompakten und effizienten Darstellung. Abweichend von früheren hierarchischen Modellen, in denen die höchste Ebene explizit die Objektebene darstellt, wird in der vorgestellten Repräsentation die Hierarchieebene direkt mit der Größe und Komplexität eines Objekts gekoppelt. So wird ein komplexes Objekt mit Textur auf einer höheren Ebene dargestellt als ein einfaches Objekte ohne Textur.
- Wir kombinieren die hierarchische Repräsentation mit einer „Grob-zu-Fein“-Suche mittels einer Ähnlichkeitshierarchie. Obwohl die Idee der „Grob-zu-Fein“-Suche bekannt ist, wird sie in dieser Arbeit erstmals direkt in die Hierarchie auf verschiedenen Abstraktionsebenen integriert und mit einer 'Scale-Space' Repräsentation kombiniert. Wie wir sehen werden, erlaubt diese Kombination das effiziente Generieren und Verifizieren von Teil- und Objekthypothesen. Diese werden in der kompositorischen Hierarchie auf einer groben Auflösungsstufe generiert, in einer „Grob-zu-Fein“-Ähnlichkeitshierarchie verfeinert, und schließlich in der kompositorischen Hierarchie auf feiner Auflösungsstufe bewertet.
- Wir schlagen eine unbeaufsichtigte 'top-down' Lernmethode vor, die die Wiederverwendbarkeit von Teilen maximiert und das Lernen von effizienten Hierarchien sowohl offline als auch online ermöglicht.
- Die Ansätze werden anhand verschiedener Anwendungen demonstriert: Objekterkennung, menschliche Posenschätzung, Aktivitätenerkennung und Szeneninterpretation für intelligente Fahrzeuge.
- Zu weiteren Beiträgen zählen die entwickelten Inferenztechniken: Kombination von 'bottom-up' Message Passing zur Hypothesengenerierung mit 'top-down' Message Passing zur Hypothesenverifikation, Ausführen des 'bottom-up' Message Passings in mehreren sequentiellen 'Sweeps' und beobachtbare 'high-level' Knoten mittels Importance Sampling.

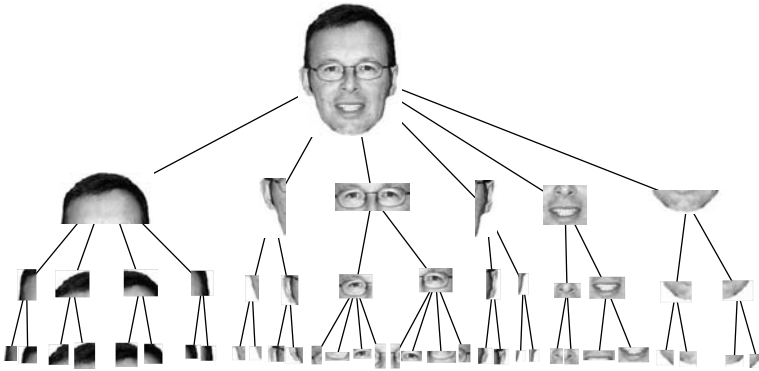


Abbildung 1: Einfaches Beispiel für eine hierarchische Dekomposition. Das Gesicht wird in seine Bestandteile zerlegt, diese Teile werden weiter zerlegt in kleinere Teile und so weiter. Die Kanten definieren die räumlichen Beziehungen zwischen den Elementen.

3 Ein einfaches Beispiel

Wir beginnen mit der Darstellung des Kompositions-konzeptes unter Verwendung der einfachen Hierarchie in Abb. 1. Das Gesicht wird in seine Bestandteile zerlegt, wie die Stirn, die Seiten mit den Ohren, die Augen, das Kinn, die Nase und den Mund. Diese Teile können wieder in kleinere Teile wie Haaransatz, ein einzelnes Auge oder ein Teil des Kinns zerlegt werden. Und schließlich werden diese kleinen Teile in visuelle Primitive wie Kanten oder Ecken zerlegt. Die kompositorische Hierarchie besteht aus zwei Komponenten: (1) die 'low-level' Merkmale und (2) die räumlichen Beziehungen zwischen den Merkmalen, Teilen und Primitiven. Die 'low-level' Features stellen die Beobachtungen dar und sind direkt aus dem Eingangsbild gewonnen. Alle höheren Ebenen sind nicht direkt beobachtbar, so dass Informationen von den unteren Ebenen über die räumlichen Relationen zu den höheren gesendet werden müssen. Diese räumlichen Relationen bestimmen die relativen Positionen zwischen dem Objekt und seinen Teilen.

Da ein Objekt auf verschiedene Weisen zerlegt werden kann, kann es auch durch verschiedene Hierarchien repräsentiert werden. Dies ist eine wichtige Eigenschaft, die dazu genutzt werden kann, unter allen Hierarchien diejenige auszuwählen, die besonders aus rechnerischer Sicht attraktiv ist. Die Idee hierbei ist redundante Berechnungen zu reduzieren, indem Informationen über ähnliche Teile wie z.B. das linke und rechte Auge nur einmal berechnet werden.

4 Hierarchische Graphische Modelle

Das Ziel der vorgeschlagenen hierarchischen graphischen Modelle ist es, verschiedene Instanzen unterschiedlicher Objektklassen in Bildern, Bildsequenzen oder andere Szenen-repräsentationen zu erkennen. Der Begriff „Objekt“ wird in diesem Zusammenhang als

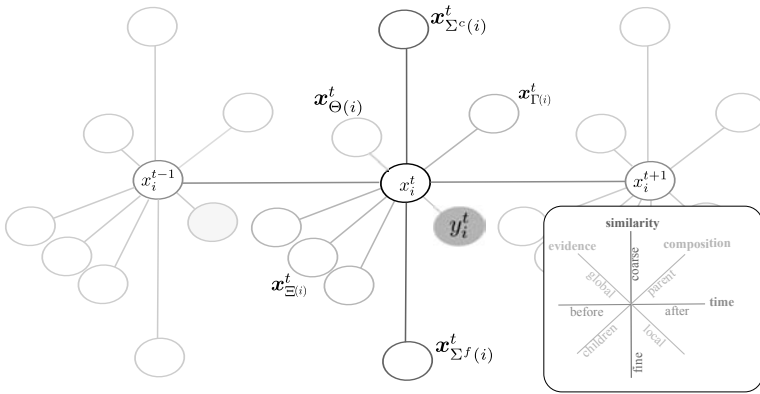


Abbildung 2: Informationsquellen von Knoten x_i (Evidenzen, Komposition, Ähnlichkeit und Zeit).

ein allgemeiner Begriff für visuelle Objekte, visuelle Teile, visuelle Merkmale, visuelle Grundtypen, aber auch Aktivitäten, Aktionen oder Bewegungsprimitive verwendet. Im Folgenden werden zwei verschiedene Arten von Hierarchien betrachtet, deren Unterscheidung für das Verstehen, der in dieser Arbeit vorgeschlagenen Ansätze sehr wichtig ist: kompositorische Hierarchien und Ähnlichkeitshierarchien. In kompositorischen Hierarchien wird die Zerlegung eines Objektes durch Knoten und Kanten repräsentiert, wobei Kanten räumliche Beziehung zwischen den Eltern und den Kindern beschreiben. Auf diese Weise können komplexe 'high-level' Knoten rekursiv in einfache 'low-level' Merkmale zerlegt werden. Zu solchen Hierarchien gehört auch die in Abb. 1 dargestellte Hierarchie. Ähnlichkeitshierarchien beschreiben hingegen Ähnlichkeiten zwischen Objekten und zwischen Teilen. In dieser Arbeit werden sie verwendet, um eine „Grob-zu-Fein“-Suche durchzuführen. In der Realität sind alle Messungen mit Unsicherheiten behaftet. Die Knoten der Hierarchie sind daher mit Zufallsvariablen assoziiert, die die Unsicherheiten in Form von Wahrscheinlichkeitsdichtefunktionen beschreiben. Die ungerichtete Graphenrepräsentation entspricht einem 'Markov Random Field'.

Es werden nun die Informationsquellen diskutieren, auf die ein Knoten in der Hierarchie zugreifen kann: Wir unterscheiden vier verschiedene Arten von Informationsquellen: evidenzbasierte, kompositorische, ähnlichkeitsbasierte und zeitliche Informationsquellen (dargestellt in Abb.2 durch verschiedene Farben). Der Knoten i kann seine a-posteriori Marginalverteilung $b_i(x_i) = p(x_i|\mathbf{y})$ durch das Kombinieren der Informationen seiner Nachbarn wie folgt berechnen:

$$b_i(x_i) \propto p(x_i|\mathbf{y}_{obs})p(x_i|\mathbf{y}_{comp})p(x_i|\mathbf{y}_{sim})p(x_i|\mathbf{y}_{temp}) \tag{1}$$

Über die Verteilung $p(x_i|\mathbf{y}_{obs})$ kommen die Sensorinformationen in das graphische Modell. Hierbei wird zwischen lokaler und globaler Evidenz unterschieden. Lokale Evidenzen beschreiben lokale Merkmale des Eingangsbildes oder -sequenz, die mit Merkmalsextraktionsverfahren wie beispielsweise Kantendetektoren aus den Bilddaten bestimmt werden. Globale Evidenzen lassen sich aus den Eingangsdaten als Ganzes ableiten und beschreiben so globale Merkmale wie z.B. Farbhistogramme des gesamten Bildes. Informationen, wie

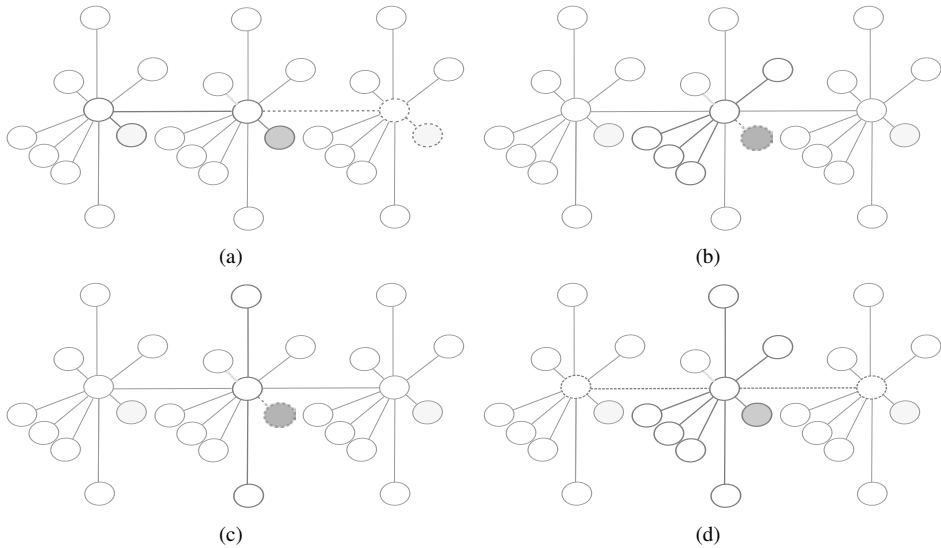


Abbildung 3: Beispiele für Informationsquellen (rot) in den folgenden Ansätzen: a) Zeitliches Tracking (z.B. Kalman- oder Partikelfilter). b) Kompositorische Hierarchie [SP05, ZLH⁺08, FL07]. c) Ähnlichkeitshierarchie (z.B. [Gav98]). d) Kombinierte kompositorische und Ähnlichkeitshierarchie (diese Arbeit).

sie im Beispiel aus Abschnitt 3 beschrieben wurden, werden durch $p(x_i | \mathbf{y}_{comp})$ berücksichtigt. Hierbei werden auf der einen Seite Informationen berücksichtigt, die von den Kinderknoten 'bottom-up' an x_i gesendet werden. Auf der anderen Seite werden Informationen vom Elternteil 'top-down' an x_i gesendet. Die Verteilung $p(x_i | \mathbf{y}_{sim})$ berücksichtigt Ähnlichkeit zwischen den Knoten. Hierbei wird zwischen Ähnlichkeiten zu Knoten auf einer groben und einer feinen Auflösung unterschieden. Genau diese Ähnlichkeitskanten sind es, aus denen die Ähnlichkeitshierarchien zusammengesetzt sind. Sie spielen eine entscheidende Rolle beim „Grob-zu-Fein“ Generieren von Objekthypothesen. Bisher wurde die Bestimmung von Objekthypothesen lediglich zu einem festen Zeitpunkt betrachtet. Oft werden Sensorinformationen jedoch über die Zeit gesammelt, so dass es neben den Sensorinformationen aus dem aktuellen Zeitschritt auch Informationen aus der Vergangenheit und offline auch aus der Zukunft gibt. Diese Informationen werden durch $p(x_i | \mathbf{y}_{temp})$ berücksichtigt.

Abb. 3 zeigt einige Beispiele für Ansätze aus der Literatur und ihren verwendeten Informationsquellen. Ansätze, die im wesentlichen auf zeitlichen Informationen beruhen, sind in Abb. 3 a) dargestellt, wobei in rot die verwendeten Kanten hervorgehoben wurden. Bekannte Beispiele dieser Ansätze sind der Kalman- oder Partikelfilter. Abb. 3 b) zeigt die Informationsquellen von kompositorischen Ansätzen, die Objekte in kleinere Teile zerlegen. Ähnlichkeitshierarchien verwenden die in Abb. 3 c) dargestellten Kanten. Die in dieser Arbeit vorgestellten Ansätze greifen auf alle zur Verfügung stehenden Informationen, wie in Abb. 3 d) dargestellt, zu.

Stehen x_i alle Informationsquellen zur Verfügung, stellt sich die schwierige Frage, wie das

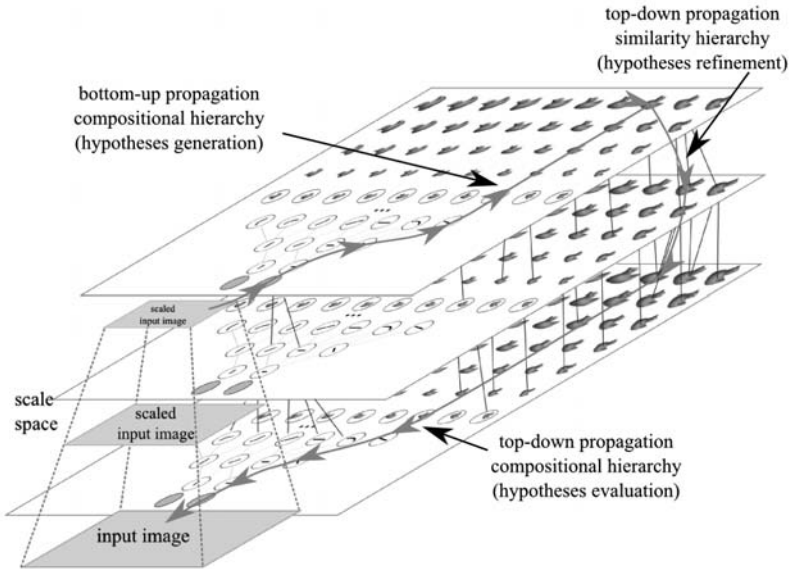


Abbildung 4: Das Versenden von Nachrichten in einer kompositorischen Hierarchie und Ähnlichkeitshierarchie für das effiziente Lösen des Inferenzproblems.

Inferenzproblem zu lösen ist. In dieser Arbeit wird hierfür Belief Propagation verwendet. Da in unserem Fall die Zufallsvariablen kontinuierlich und multimodal sind, wird nicht-parametrisches Belief Propagation angewendet. Dieser Ansatz ist ähnlich dem bekannten Partikelfilter, bei dem die Wahrscheinlichkeitsdichtefunktion durch eine Menge von Partikeln repräsentiert wird. In dieser Arbeit werden verschiedene Ansätze vorgestellt, mit denen sich das Inferenzproblem effizient und in Echtzeit lösen lässt. Abb. 4 gibt ein Beispiel für das Lösen des Inferenzproblem durch intelligentes Nachrichtenversenden. Um Objekte im Eingangsbild erkennen zu können, wird das Bild in Form einer Bildpyramide an die Kombination aus kompositorischer Hierarchie und Ähnlichkeitshierarchie angelegt. Die Bildpyramide enthält neben dem Eingangsbild in Originalauflösung auch verschiedene skalierte Versionen. Die grundlegende Idee des effizienten Nachrichtenversendens ist (1) Objekthypothesen auf einer groben Auflösungsstufe durch das 'bottom-up' Senden von Nachrichten in der kompositorischen Hierarchie zu generieren, (2) diese anschließend in der Ähnlichkeitshierarchie von der groben Auflösung hin zur detaillierten zu verfeinern und (3) schließlich in der kompositorischen Hierarchie auf der feinen Auflösung 'top-down' zu evaluieren (siehe roten Pfad in Abb. 4).

Objekterkennung: Die Erkennung von Objekten aus Bildern ist ein klassisches Problem aus dem Bereich Computer Vision. Eine besondere Herausforderung ist hierbei verschiedene Objekte aus verschiedenen Blickrichtungen und Skalierungen zu erkennen. Die Ansätze aus dieser Arbeit lösen dieses Problem effizient, indem Ähnlichkeiten unter den Objekten ausgenutzt werden und so eine effiziente Lösung des Inferenzproblems ermöglicht wird. Abb. 5 zeigt ein qualitatives Beispielergebnis.



Abbildung 5: Beispielresultat für die rotations- und skalierungsunabhängige Erkennung von verschiedenen Objekten: (a) Eingangsbild (b) Erkannte Objekte in rot/blau hervorgehoben.

Posenerkennung: Ähnlich der Erkennung von Objekten können die vorgeschlagenen Ansätze auch zur menschlichen Posenschätzung verwendet werden. Dieses Problem stellt eine noch größere Herausforderung dar, da die Gelenke des menschlichen Körpers zusätzliche Freiheitsgrade darstellen, die mitgeschätzt werden müssen. Glücklicherweise ähneln sich die Erscheinungsbilder des menschlichen Körpers für verschiedene Konfigurationen sehr stark, so dass viele Informationen geteilt werden können. Abb. 6 zeigt Beispiele für erkannte menschliche Posen, die für eine Ganganalyse verwendet wurden. Bemerkenswert ist hierbei, dass der menschliche Körper sehr robust unter verschiedenen Skalierungen erkannt werden kann, dadurch dass die Skalierung des Modells direkt in der Repräsentation abgebildet ist.

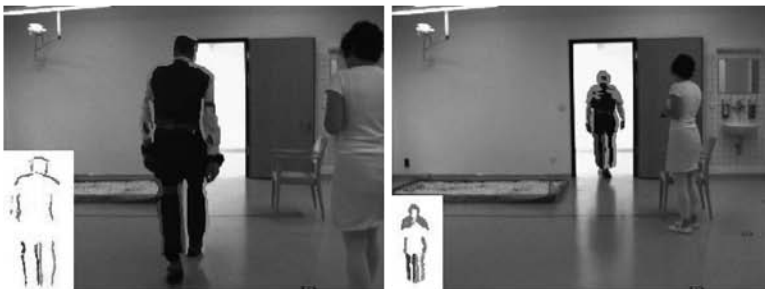


Abbildung 6: Beispiele für die Erkennung der menschlichen Pose: Die Pose ist farblich hervorgehoben in rot (Kopf), blau (linkes Bein) und grün (rechtes Bein).

Verhaltensanalyse: Werden die kompositorischen Hierarchien und ihre Kanten nicht nur räumlich sondern auch raum-zeitlich betrachtet, können die Ansätze auch für die Erkennung von Aktivitäten und somit zur Verhaltensanalyse verwendet werden. Entscheidend ist hierbei, dass neben den räumlichen Beziehungen auch zeitliche Abhängigkeiten modelliert werden. Abb. 7 zeigt ein Beispiel. Die Aktivität 'Kaffee trinken' wird durch eine Menge an Bewegungsvektoren beschrieben (links in rot hervorgehoben). Diese Bewegungen wurden aus der Bildsequenz mithilfe von optischen Fluss berechnet. Die Hierarchie

(rechts) beschreibt nun die Zerlegung der Aktivität 'Kaffee trinken' in einzelne Aktionen wie 'Greifen der Tasse' oder 'Führen der Tasse zum Mund' und schließlich die Zerlegung der Aktionen in einzelne Bewegungsprimitive.

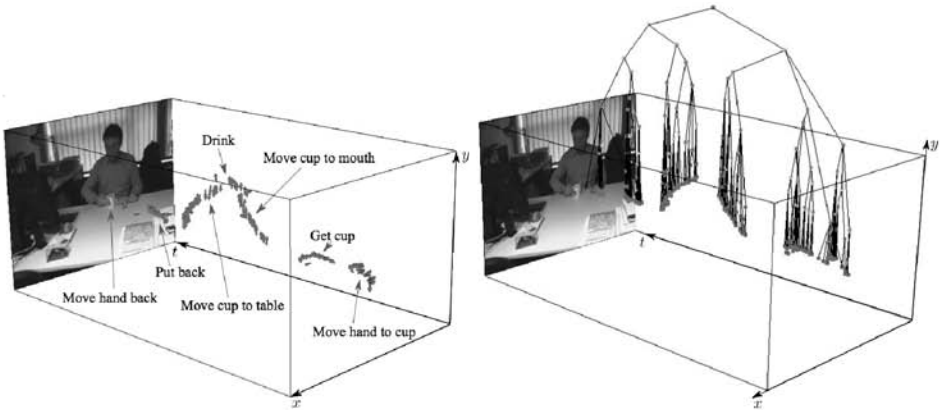


Abbildung 7: Beispiel für die Aktivitätenerkennung: Erkannte Bewegungsvektoren im Raum-Zeit Volumen (links). Zerlegung der gesamten Aktivität in einzelne Aktionen.

Szenenverstehen für intelligente Fahrzeuge: Die vorgeschlagenen Ansätze wurden auch zur Interpretation des Fahrzeugumfelds in Echtzeit verwendend. Basierend auf einfachen Kamerasensoren, die unter den Seitenspiegeln angebracht wurden, wurden die Ansätze verwendet, um effizient Parklücken zu erkennen. Die Idee war hierbei, die Parkplatzszenen hierarchisch in einzelne Parklücken zu zerlegen. Die Parklücken wurden wiederum in Markierungen und ggf. einem geparkten Fahrzeug zerlegt. Die Zerlegung wurde über geometrische Grundkörper weitergeführt bis schließlich einfache Belegungswahrscheinlichkeiten einer kartenbasierten Umfeldrepräsentation erreicht wurden. Abb. 8 (links) zeigt den Versuchsträger Paul auf der Hannover Messe 2008. Ergebnisse sind rechts dargestellt, wobei oben die Eingangsdaten und unten die Ergebnisse gezeigt sind.

5 Fazit

In dieser Arbeit wurde eine neue hierarchische Repräsentation vorgestellt, die für ein breites Spektrum von Anwendungen geeignet ist. Unsere Ansätze nutzen drei grundlegende Konzepte: kompositorische Hierarchien, „Grob-zu-Fein“-Ähnlichkeitshierarchien, und das 'Sharing' von Teilen, um eine robuste und effiziente Darstellung von Multi-Objekten, Multi-Skalen- und Multi-Ansichten zu erhalten. Die in dieser Arbeit untersuchten Anwendungen spiegeln nur einen kleinen Teil der möglichen Einsatzgebiete wieder. Andere Anwendungsgebiete sehen wir überall dort, wo Modelle hierarchisch in Teile zerlegt werden können. Die vorgeschlagenen Verfahren sind besonders geeignet, wenn das Modell eine große Menge von verschiedenen Instanzen, Posen oder Konfigurationen in Echtzeit



Abbildung 8: Ergebnisse für das Szenenverstehen intelligenter Fahrzeuge: Versuchtsträger Paul (links), Eingangsdaten als Draufsicht (rechts oben), Beispielergebnisse (rechts unten).

repräsentieren muss. Für all diese Anwendungen liefert diese Arbeit wichtige Grundkonzepte, effiziente Methoden sowie praktische 'How-to'-Beispiele.

Literatur

- [FL07] S. Fidler und A. Leonardis. Towards Scalable Representations of Object Categories: Learning a Hierarchy of Parts. Minnesota, USA, June 2007.
- [Gav98] D M. Gavrila. Multi-Feature Hierarchical Template Matching Using Distance Transforms. In *Proceedings of the 14th International Conf. on Pattern Recognition-Volume 1*, Seiten 439–, Washington, DC, USA, 1998.
- [SP05] F. Scalzo und J. H. Piater. Statistical Learning of Visual Feature Hierarchies. Jgg. 3, Seite 44, 2005.
- [Spe13] Jens Spehr. *On Hierarchical Models for Visual Recognition and Learning of Objects, Scenes, and Activities*. Dissertation, Technische Universität Braunschweig, 2013.
- [ZLH⁺08] Long Zhu, Chenxi Lin, Haoda Huang, Yuanhao Chen und Alan L. Yuille. Unsupervised Structure Learning: Hierarchical Recursive Composition, Suspicious Coincidence and Competitive Exclusion. Jgg. 5303 of *Lecture Notes in Computer Science*, Seiten 759–773, 2008.



Jens Spehr studierte Elektrotechnik an der TU Braunschweig. Nach dem Abschluss des Studiums 2006 war er als wissenschaftlicher Mitarbeiter am Institut für Robotik und Prozessinformatik der TU Braunschweig beschäftigt, an dem er 2013 promovierte. Momentan arbeitet er bei der Konzernforschung der Volkswagen AG, wo er verantwortlich ist für die Umfeldwahrnehmung von automatisch fahrenden Fahrzeugen. Zu seinen Forschungsinteressen zählen hierarchische Modelle im Bereich Computer Vision aber auch grundlegende Konzepte im Bereich des maschinellen Lernens und der künstliche Intelligenz.