Interessengetriebene audiovisuelle Szenenexploration

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

von der KIT-Fakultät für Informatik
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

Dipl.-Inf. Benjamin Kühn

aus Ludwigshafen am Rhein

Tag der mündlichen Prüfung:	06. Februar 2015
Erster Gutachter:	Prof. Dr.-Ing. J. Beyerer
Zweiter Gutachter:	Prof. Dr.-Ing. K. Kroschel

Danksagung

Die vorliegende Arbeit entstand im Rahmen des *Sonderforschungsbereichs 588 – humanoide Roboter* und wurde von der *Deutschen Forschungsgemeinschaft (DFG)* gefördert. Im diesem Kontext haben mich eine Vielzahl an Menschen begleitet, welche mich sowohl gefördert als auch gefordert haben, um die ehrgeizigen Ziele zu erreichen.

Als Erstes möchte ich mich bei Herrn Prof. Dr.-Ing. Jürgen Beyerer und Herrn Prof. Dr.-Ing. Kristian Kroschel bedanken für die Möglichkeit zur Promotion sowie dem stets konstruktiven Austausch, für die Freiheiten, die Möglichkeiten und das in mich gesetzte Vertrauen. Herrn Prof. Dr.-Ing. Jürgen Beyerer möchte ich insbesondere auch für die reibungslose Übernahme des Hauptreferats danken. Die intensiven Gespräche mit Herrn Prof. Dr.-Ing. Kristian Kroschel werde ich vermissen.

Außerdem bedanke ich mich bei den Kollegen/-innen sowohl am Karlsruher Institut für Technologie (KIT) als auch am Fraunhofer IOSB, welche für eine angenehme Arbeitsatmosphäre sorgten, immer ein offenes Ohr hatten sowie ein konstruktives Arbeiten ermöglicht haben.

Zwei Personen möchte ich in diesem Zusammenhang besonders hervorheben: Timo Machmer und Alexej Swerdlow begleiten mich schon seit vielen Jahren, sowohl vor, während, als auch nach der Promotion. Die Zusammenarbeit war und ist immer freundschaftlich, konstruktiv, kollegial, zielstrebig und angenehm. Vielen Dank dafür.

Bei Frau Eva-Maria Schubart möchte ich mich für die kurze, jedoch sehr angenehme Zeit am Institut für Nachrichtentechnik (INT) und spätere Treffen danken. Ihr sowie Isabelle Wappler, Marion Staub, Angelika Schreiber und Gaby Gross möchte ich danken für die stetige Hilfsbereitschaft in Verwaltungsangelegenheiten und eine reibungslose, herzliche Zusammenarbeit.

Bei meinen Freunden möchte ich mich bedanken, die auch in Zeiten, in denen ich mich etwas rar gemacht habe, stets zu mir gehalten haben und mich im Bestreben der Promotion stets unterstützt haben. Sie haben einen großen Betrag für eine gesunde Work-Life-Balance, viel Spaß und Humor geleistet. Danke.

Meinen Eltern Heidemarie und Wolfram Kühn sowie meinem Bruder Manuel möchte ich danken für ihre Stütze auf dem teils stressigen und zeitraubenden Weg zur Promotion. Sie haben mich bei meinem Vorhaben, das Abitur zu machen und später auch zu studieren in jeglicher Hinsicht unterstützt, bei der Entscheidung zur Promotion Mut zugesprochen und Recht behalten.

VIELEN DANK!

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungen, Akronyme, Notation und Formelzeichen

Abkürzungen und Akronyme

ANT	Ameisenalgorithmus
ARMAR	AnthRopomorphic Multi-Arm-Robot
CC	Cumulated Curiosity
CJAD	Cumulated Joint Angle Distances
CS	Cumulated Saliency
DB	DynamicBrain Robot
DCT	Discrete Cosine Transform
DFG	Deutsche Forschungsgemeinschaft
EP	Explorationspfad
FFT	Fast Fourier Transform
GA	Genetischer Algorithmus
GCC	Generalized Cross Correlation
GMM	Gaussian Mixture Model
HMI	Mensch-Maschine-Schnittstelle
HMM	Hidden-Markov-Modell
HSV	Farbmodell mit den Komponenten: Farbton (H), Farbsättigung (S) und Hellwert (V)
ICC	Integrated Cumulated Curiosity
ICS	Integrated Cumulated Saliency
ICM	Indices-based Correlation Measure
IFFT	Inverse Fast Fourier Transform
IOR	Inhibition of Return

IQDCT	Inverse Quaternion Discrete Cosine Transform
IOSB	(Fraunhofer-)Institut für Optronik, Systemtechnik und Bildauswertung
JAD	Joint Angle Distances
KIT	Karlsruher Institut für Technologie
KLD	Kullback-Leibler Divergenz
MATLAB	MATrix LABoratory
MCT	Modifizierte Census-Transformation
MFCC	Mel Frequency Cepstral Coefficients
MMI	Mensch-Maschine-Interaktion
NCJAD	Normalized Cumulated Joint Angle Distances; dt. Maß zur Bewertung der Gesamtbewegung auf einem Explorationspfad
NICC	Normalized Integrated Cumulated Curiosity; dt. Bewertungsmaß für die Abnahme der Neugier entlang eines Explorationspfads
NICS	Normalized Integrated Cumulated Saliency; dt. Bewertungsmaß für die Abnahme der Salienz entlang eines Explorationspfads
NN	Nearest Neighbor algorithm, dt. Nächster-Nachbar-Heuristik
OCSVM	One-Class Support Vector Machine
OPASCA	OPto-Acoustic SCene Analysis
PHAT	PHAse Transform
PSS	Parameter Search Space
PTU	Pan-Tilt-Unit, dt. Schwenk-Neige-Einheit
PTZ	Pan-Tilt-Zoom, dt. schwenken-neigen-zoomen
QDCT	Quaternion Discrete Cosine Transform
RGB-D	Kameras von diesem Typ besitzen neben den Grundfarben Rot (R), Grün (G) und Blau (B) noch Tiefeninformationen (D)
RGB	Farbmodell mit Grundfarben Rot (R), Grün (G) und Blau (B)
SFB	Sonderforschungsbereich
SRP	Steered Responce Power
STFT	Short-Time Fourier Transform
TDOA	Time Difference Of Arrival
TSP	Traveling Salesman Problem, dt. Rundreiseproblem oder Problem eines Handlungsreisenden
UBM	Universal Background Model

Notationen

$\circ$	Hadamard-Produkt (elementweises Produkt) zweier Matrizen
x oder X	skalare Größe
$\boldsymbol{x}$	Vektor
x_k	k-tes Element des Vektors $\boldsymbol{x}$
$\boldsymbol{X}$	Matrix
$\boldsymbol{X}^{-1}$	Inverse der Matrix $\boldsymbol{X}$
$\mathcal{X}$	Menge
$\lvert x \rvert$	Betrag von x
$\lVert \boldsymbol{x} \rVert$	Norm des Vektors $\boldsymbol{x}$
$\lvert \mathcal{X} \rvert$	Kardinalität von $\mathcal{X}$
$f(x)$	skalare Funktion von x
$f(\boldsymbol{x})$	skalare Funktion von Vektor $\boldsymbol{x}$
$\mathrm{sgn}(x)$	Signumfunktion
$\mathrm{erf}(x)$	Fehlerfunktion
$\mathcal{N}(\mu, \Sigma)$	Normalverteilung
$\mathrm{Cov}(\mathcal{X})$	Kovarianzmatrix
$\mathrm{Clust}(\mathcal{X}, d)$	Clusterfunktion zur Aufteilung der Menge $\mathcal{X}$ in disjunkte Teilmengen anhand des Kriteriums d
$\textsc{Histogramm}(\mathcal{X}, \boldsymbol{p})$	Histogramm über die Werte der Menge $\mathcal{X}$ an den Positionen $\boldsymbol{p}$
$\max(x, y)$	Maximum von x und y
$\min(x, y)$	Minimum von x und y
$\arg\max_x f(x)$	Argument, für welches $f(x)$ maximal ist
$\arg\min_x f(x)$	Argument, für welches $f(x)$ minimal ist
$\boldsymbol{x} \succ \hat{\boldsymbol{x}}$	$\hat{\boldsymbol{x}}$ wird strikt dominiert durch $\boldsymbol{x}$
Gegenstand	Klasse eines Objekts (hier: Gegenstand)
istGegenstand	Klassenrelation eines Objekts (hier: Relation auf die Klasse Gegenstand)

Die wichtigsten Formelzeichen

Aufmerksamkeit und Salienz

$a(t)$	ursprünglich ausgesendetes Audiosignal
β	Gewichtungsfaktor für PHAT-Gewichtungsfilter
c_i	unimodaler Salienzcluster
$c_j^{\text{akustisch}}$	akustischer Salienzcluster
c_k^{visuell}	visueller Salienzcluster
$c_{\mathcal{C}_i}$	räumlich-zeitlich fusionierter multimodaler Salienzcluster, generiert aus der Teilmenge $\mathcal{C}_i$
$\mathcal{C}$	Menge aller unimodaler Salienzcluster
$\mathcal{C}^{\text{akustisch}}$	akustische Teilmenge der Salienzclustermenge $\mathcal{C}$
$\mathcal{C}^{\text{visuell}}$	visuelle Teilmenge der Salienzclustermenge $\mathcal{C}$
$\mathcal{C}_i$	Teilmenge an räumlich zusammengehörigen Clustern
$\mathcal{C}_i^{\text{akustisch}}$	akustischer Anteil der Teilmenge $\mathcal{C}_i$
$\mathcal{C}_i^{\text{visuell}}$	visueller Anteil der Teilmenge $\mathcal{C}_i$
$\boldsymbol{d}$	Verschiebungsvektor zur Bestimmung des Zentrums $\boldsymbol{u}$ bei der Isophotenberechnung
$D_{\text{KL}}(f\|\|g)$	Kullback-Leibler-Divergenz
$E(t, \boldsymbol{x})$	Ergebnisakkumulator beim SRP-PHAT-Verfahren
E_{min}	minimale Schwelle im Ergebnisakkumulator beim SRP-PHAT-Verfahren
$f_{\text{3D}}\{\cdot\}$	Transformationsfunktion für die Rückprojektion von 2D-Punkten nach 3D
$f_{c_i}^G(\boldsymbol{x})$	gewichtete Gauß-Notation eines unimodalen Salienzclusters c_i
$h(t)$	Raumimpulsantwort
$\mathbb{H}$	Algebra für Quaternionen
$\boldsymbol{I}_q(t)$	vierdimensionales Bild zum Zeitpunkt t
$\boldsymbol{I}_I$	Intensität eines Bilds
$\boldsymbol{I}_{\text{RG}}$ bzw. $\boldsymbol{I}_{\text{BY}}$	rot-grüne bzw. blau-gelbe Farbopponenten
$\boldsymbol{I}_{\text{M}}$	Bewegungsbild
$\kappa(x, y)$	Isophotenkrümmung im Punkt (x, y)
$\boldsymbol{\mu}_{c_i}$	räumliche Position des unimodalen Salienzclusters c_i
$\boldsymbol{\mu}_{c_j}^{\text{akustisch}}$	räumliche Position des akustischen Salienzclusters $c_j^{\text{akustisch}}$
$\boldsymbol{\mu}_{c_k}^{\text{visuell}}$	räumliche Position des visuellen Salienzclusters c_k^{visuell}

$\boldsymbol{\mu}_{\mathcal{C}_i}$	räumliche Position des multimodalen Salienzclusters $c_{\mathcal{C}_i}$
$\mu_{\text{Prior}}^{l,k}$	Erwartungswert der a-priori-Wahrscheinlichkeitsdichte für den diskreten Zeitpunkt l und die Frequenz k
$\mu_{\text{Post}}^{l,k}$	Erwartungswert der a-posteriori-Wahrscheinlichkeitsdichte für den diskreten Zeitpunkt l und die Frequenz k
$\mathcal{M}$	Menge aller Mikrofonpaare
$n(t)$	additiver Rauschterm
ψ^{PHAT}	PHAT-Gewichtungsfilter
$\psi^{\text{PHAT-}\beta}$	PHAT-β-Gewichtungsfilter
$P_{\text{Prior}}^{l,k}$	a-priori-Wahrscheinlichkeitsdichte zum diskreten Zeitpunkt l und für die Frequenz k
$P_{\text{Post}}^{l,k}$	a-posteriori-Wahrscheinlichkeitsdichte zum diskreten Zeitpunkt l und für die Frequenz k
q	Quaternion (hyperkomplexe Zahl)
R_{ij}	Kreuzkorrelation
R_{ij}^{GCC}	generalisierte Kreuzkorrelation (engl.: Generalized Cross Correlation, kurz: GGC)
s_{c_i}	Salienz des unimodalen Salienzclusters c_i
$s_{c_j}^{\text{akustisch}}$	Salienz des akustischen Salienzclusters $c_j^{\text{akustisch}}$
$s_{c_k}^{\text{visuell}}$	Salienz des visuellen Salienzclusters c_k^{visuell}
$s_{\mathcal{C}_i}$	Salienz des multimodalen Salienzclusters $c_{\mathcal{C}_i}$
$S_{\text{akustisch}}(l)$	akustische Salienz zum diskreten Zeitpunkt l gemittelt über alle Frequenzen
$S_{\text{akustisch}}(l,k)$	akustische Salienz zum diskreten Zeitpunkt l für die Frequenz k
$\boldsymbol{S}_{\text{visuell}}(t)$	visuelle Salienzkarte zum Zeitpunkt t
$\boldsymbol{\Sigma}_{c_i}$	Positionsunsicherheit des unimodalen Salienzclusters c_i
$\boldsymbol{\Sigma}_{c_j}^{\text{akustisch}}$	Positionsunsicherheit des akustischen Salienzclusters $c_j^{\text{akustisch}}$
$\boldsymbol{\Sigma}_{c_k}^{\text{visuell}}$	Positionsunsicherheit des visuellen Salienzclusters c_k^{visuell}
$\boldsymbol{\Sigma}_{\mathcal{C}_i}$	Positionsunsicherheit des multimodalen Salienzclusters $c_{\mathcal{C}_i}$
$\mathcal{S}$	kartesischer Suchraum zur Bestimmung des Ursprungs der Schallquelle
$\tilde{\mathcal{S}}$	sphärischer Suchraum zur Bestimmung des Ursprungs der Schallquelle

$r(x, y)$	Radius von Punkt (x, y) ausgehend zum Zentrum der kreisförmigen Isophote
τ_{ij}	theoretischer Zeitversatz zwischen dem i-ten und j-ten Mikrofon
$\boldsymbol{u}$	Zentrum eines Peaks bei der Isophotenberechnung
$\mathrm{var}_{\mathrm{Prior}}^{l,k}$	Varianz der Normalverteilung der a-priori-Wahrscheinlichkeitsdichte für den diskreten Zeitpunkt l und die Frequenz k
$\mathrm{var}_{\mathrm{Post}}^{l,k}$	Varianz der Normalverteilung der a-posteriori-Wahrscheinlichkeitsdichte für den diskreten Zeitpunkt l und die Frequenz k
$\boldsymbol{V}_l$	l-te Kovarianzmatrix bei der Bestimmung der Gesamtkovarianzmatrix eines Clusters
$\boldsymbol{x} = (x, y, z)$	kartesischen Koordinaten
$\tilde{\boldsymbol{x}} = (r, \varphi, \vartheta)$	sphärische Koordinaten
$\hat{\boldsymbol{x}}(t)$	Position einer einzelnen ermittelten Schallquelle zum Zeitpunkt t
$\hat{\mathcal{X}}(t)$	Menge aller selektierten Raumpositionen zur Bestimmung von Schallquellen zum Zeitpunkt t
$\mathcal{X}_{c_k}^{\mathrm{2D}}$	visuelle 2D-Positionen des Clusters c_k
$\hat{\mathcal{X}}_{c_j} \subseteq \hat{\mathcal{X}}(t)$	räumlicher Cluster c_j von akustischen Positionen zur Bestimmung des Ursprungs der Schallquelle
$y(n)$	abgetastetes Audiosignal zum diskreten Zeitpunkt n
$y_i(t)$	Audiosignal von Mikrofon i
$\boldsymbol{Y}_{\mathrm{Prior}}(l, k)$	Ausgangsdaten zur Bestimmung der a-priori-Wahrscheinlichkeitsdichte zum diskreten Zeitpunkt l und für die Frequenz k
$\boldsymbol{Y}_{\mathrm{Post}}(l, k)$	Ausgangsdaten zur Bestimmung der a-posteriori-Wahrscheinlichkeitsdichte zum diskreten Zeitpunkt l und für die Frequenz k
$Y_{\mathrm{Spek}}(n, k)$	Spektrogramm eines Audiosignals zum Zeitpunkt n für die Frequenz k
$Y_i(t, \omega)$	Kurzzeit-Fourier-Transformierte des Audiosignals $y_i(t)$ von Mikrofon i

Objektzentrierte Umwelterfassung

a_i^o bzw. a_i	i-tes Attribut des Objektrepräsentanten o
$a_i^{o_p}$	i-tes Attribut des a-priori-Objekts o_p
$a^o_{<Name>}$	Attribut vom Typ <*Name*> des Objektrepräsentanten o
$\mathcal{A}_o$	Menge aller Attribute des Objektrepräsentanten o
$\mathcal{A}_{o_p}$	Attributwissen des a-priori-Objektrepräsentanten o_p
$\boldsymbol{d}_{a,i}$ bzw. $\boldsymbol{d}_{r,i}$	relevante Sensordaten für das Attribut a bzw. die Relation r zum Zeitpunkt i
$\mathcal{D}_a$ bzw. $\mathcal{D}_r$	relevante Sensordaten für das Attribut a bzw. die Relation r von mehreren Zeitpunkten
$f_a^{\mathrm{AT}}(\cdot)$	Analysetransformation zur Generierung eines Merkmalsvektors für das Attribut a
$f_r^{\mathrm{AT}}(\cdot)$	Analysetransformation zur Generierung eines Merkmalsvektors für die Relation r
$f_a^{\mathrm{SAT}}(\cdot)$	Attribut-spezifische Signaturakualisierungstransformation für das Attribut a
$f_r^{\mathrm{SAT}}(\cdot)$	Relation-spezifische Signaturakualisierungstransformation für die Relation r
$f_a^{\mathrm{SGT}}(\cdot)$	Attribut-spezifische Signaturgenerierungstransformation für das Attribut a
$f_r^{\mathrm{SGT}}(\cdot)$	Relation-spezifische Signaturgenerierungstransformation für die Relation r
$f_a^{\mathrm{dSGT}}(\cdot)$	Attribut-spezifische direkte Signaturgenerierungstransformation für das Attribut a
$f_r^{\mathrm{dSGT}}(\cdot)$	Relation-spezifische direkte Signaturgenerierungstransformation für die Relation r
$k^o_{a_i}$ bzw. k_{a_i}	Konfidenzwert des Attributs a_i des Objektrepräsentanten o
$k^o_{a_{<Name>}}$	Konfidenzwert des Attributs $a_{<Name>}$ des Repräsentanten o
$k^o_{r_j}$ bzw. k_{r_j}	Konfidenzwert der Relation r_j des Objektrepräsentanten o
$k^o_{r_{<Name>}}$	Konfidenzwert der Relation $r_{<Name>}$ des Repräsentanten o
o	Objektrepräsentant im Umweltmodell
o_p	A-priori Objektwissen für ein Objekt o_p
$\mathcal{O}$	Menge aller Objektrepräsentanten im Umweltmodell
$\mathcal{O}_{\mathrm{Signaturen}}$	Menge aller Objektsignaturen zur Wiedererkennung
$\mathcal{O}_{\mathrm{Vorwissen}}$	Menge des gesamten a-priori-Objektwissen
r_j^o bzw. r_j	j-te Relation des Objektrepräsentanten o

$r_j^{o_p}$	j-te Relation des a-priori-Objekts o_p
$r^o_{<Name>}$	Relation vom Typ <*Name*> des Objektrepräsentanten o
$\mathcal{R}_o$	Menge aller Relationen des Objektrepräsentanten o
$\mathcal{R}_{o_p}$	Relationwissen des a-priori-Objekts o_p
$\mathcal{S}_o$	Menge aller Signaturen für den Objektrepräsentanten o
$\mathcal{S}_{a_i \mid o}$	attributsbezogene Signatur für das Attribut a_i des Objektrepräsentanten o
$\mathcal{S}_{r_j \mid o}$	relationsbezogene Signatur für die Relation r_j des Objektrepräsentanten o
$\rho^o_{a_i}$ bzw. ρ_{a_i}	Priorität des Attributs a_i des Objektrepräsentanten o
$\rho^o_{a_{<Name>}}$	Priorität des Attributs $a_{<Name>}$ des Objektrepräsentanten o
$\rho^o_{r_j}$ bzw. ρ_{r_j}	Priorität der Relation r_j des Objektrepräsentanten o
$\rho^o_{r_{<Name>}}$	Priorität der Relation $r_{<Name>}$ des Objektrepräsentanten o
$t^o_{a_i}$ bzw. t_{a_i}	Attributtyp des Attributs a_i des Objektrepräsentanten o
$t^{o_p}_{a_i}$	Attributtyp des Attributs $a_i^{o_p}$ des a-priori-Objekts o_p
$t^o_{a_{<Name>}}$	Attributtyp des Attributs $a_{<Name>}$ des Objektrepräsentanten o
$t^o_{r_j}$ bzw. t_{r_j}	Relationstyp der Relation r_j des Objektrepräsentanten o
$t^{o_p}_{r_j}$	Relationstyp der Relation $r_j^{o_p}$ des a-priori-Objekts o_p
$t^o_{r_{<Name>}}$	Relationstyp der Relation $r_{<Name>}$ des Objektrepräsentanten o
t^o_z bzw. t_z	Zeitpunkt der Wahrnehmung des Objektrepräsentanten o
$\mathcal{T}_\mathcal{A}$	Menge aller im Umweltmodell vorhandener Attributtypen
$\mathcal{T}_\mathcal{R}$	Menge aller im Umweltmodell vorhandener Relationstypen
$w^o_{a_i}$ bzw. w_{a_i}	Attributwert des Attributs a_i des Objektrepräsentanten o
$w^{o_p}_{a_i}$ bzw. $\mathcal{W}^{o_p}_{a_i}$	gültiger Wertebereich bzw. Menge an validen Werten für das Attribut $a_i^{o_p}$
$w^o_{a_{<Name>}}$	Attributwert des Attributs $a_{<Name>}$ des Objektrepräsentanten o
$w^o_{r_j}$ bzw. w_{r_j}	Relationswert der Relation r_j des Objektrepräsentanten o
$w^{o_p}_{r_j}$ bzw. $\mathcal{W}^{o_p}_{r_j}$	gültiger Wertebereich bzw. Menge an validen Werten für die Relation $r_j^{o_p}$ des a-priori-Objekts o_p
$w^o_{r_{<Name>}}$	Relationswert der Relation $r_{<Name>}$ des Objektrepräsentanten o
$\mathcal{W}$	aktuell verfügbares Wissen im Umweltmodell
$\boldsymbol{x}_{a,i}$ bzw. $\boldsymbol{x}_{r,i}$	attributs- bzw. relationsbezogener Merkmalsvektor für das Attribut a bzw. die Relation r zum Zeitpunkt i

$\mathcal{X}_a$ bzw. $\mathcal{X}_r$	Menge aller Merkmalsvektoren aus verschiedenen Zeitschritten für das Attribut a bzw. die Relation r
$\mathcal{Z}_o$	Menge aller Zeitpunkte der Wahrnehmung eines Objektrepräsentanten

Wissensbasierte Neugier

e_{a_i}	Indikator für die Existenz des Attributs a_i
e_{r_j}	Indikator für die Existenz der Relation r_j
η_o	Induzierte Neugier für das im Umweltmodell repräsentierte Objekt o
d_{Cluster}	Abstandsmaß für die Clusterbildung
$f_{<Strategie>}(\cdot)$	verschiedene Fusionsstrategie zur Bestimmung der Aspekte und Teilaspekte
$f_\eta(\cdot)$	Fusionsstrategie für die Fusion der Aspekte der Neugier
$f_\kappa(\cdot)$	Fusionsstrategie für die Fusion der Teilaspekte der Komplexität
$f_\pi(\cdot)$	Fusionsstrategie für die Fusion der Teilaspekte der Neuartigkeit
$f_\nu(\cdot)$	Fusionsstrategie für die Fusion der Teilaspekte der Unsicherheit
$f_\varsigma(\cdot)$	Fusionsstrategie für die Fusion der Teilaspekte des Konflikts
$g_o^{<Name>}(\cdot)$	diverse Hilfsfunktionen zur Bestimmung eines Teilaspekts der Neugier für das Objekt o
κ_o	Aspekt: Komplexität des im Umweltmodell repräsentierten Objekts o
κ_o^{Objekt}	Teilaspekt: Komplexität eines Objekts
κ_o^{Szene}	Teilaspekt: Lokale Szenenkomplexität
π_o	Aspekt: Neuartigkeit des im Umweltmodell repräsentierten Objekts o
π_o^{Grad}	Teilaspekt: Grad der Neuartigkeit
$\pi_o^{\text{Häufigkeit}}$	Teilaspekt: Häufigkeit der Wahrnehmung
$\pi_o^{\text{Zeitpunkt}}$	Teilaspekt: Zeitpunkt der letzten Wahrnehmung
ν_o	Aspekt: Unsicherheit bzgl. des im Umweltmodell repräsentierten Objekts o
ν_o^{Klasse}	Teilaspekt: Typ- bzw. Identitätsunsicherheit
$\nu_o^{\text{Eigenschaften}}$	Teilaspekt: attributs- und relationsbezogene Unsicherheit

$\mathcal{O}_{\text{Gegenstände}}$	Untermenge der Objektrepräsentanten im Umweltmodell, die Gegenstände repräsentieren
$\mathcal{O}_{\text{Personen}}$	Untermenge der Objektrepräsentanten im Umweltmodell, die Personen repräsentieren
ς_o	Aspekt: Konflikte bzgl. des im Umweltmodell repräsentierten Objekts o
$\varsigma_o^{\text{Identität}}$	Teilaspekt: Konflikt bei der Identitätsbestimmung
$\varsigma_o^{\text{Multimodal}}$	Teilaspekt: Konflikt bei der multimodalen Wahrnehmung
$\varsigma_o^{\text{Vorwissen}}$	Teilaspekt: Konflikt mit dem a-priori-Wissen
$\Delta t_o^{\text{Zeitpunkt}}$	Zeitraum seit der letzten Wahrnehmung des im Umweltmodell repräsentierten Objekts o
Δt_{T}	Zeitraum bis zur vollständigen Neuartigkeit eines Objekts bei fehlender Observation
ω_i	Faktoren für die gewichtete Fusionsstrategie
$\Delta \mathcal{Z}_o$	Menge an Wahrnehmungen des im Umweltmodell repräsentierten Objekts o mit relativem Zeitbezug

Interessengetriebene Exploration einer Szene

$\gamma_s, \gamma_\eta, \gamma_d$	Einflussparameter für den interessengetriebenen Explorationspfad
ξ	bijektive Abbildung zur Generierung aller Permutationen der Objektmenge $\mathcal{P}_\mathcal{O}$
$\text{EP}_{\text{Allgemein}}$	allgemeine Definition eines Explorationspfads
$\text{EP}_{\text{Bewegung}}$	bewegungsoptimierter Explorationspfad
$\text{EP}_{\text{Neugier}}$	Neugier-basierter Explorationspfad
$\text{EP}_{\text{Referenz}}$	Referenzexplorationspfad
$\text{EP}_{\text{Salienz}}$	Salienz-basierter Explorationspfad
EP_{SNB}	interessengetriebener Explorationspfad
η_o	induzierte Neugier für das im Umweltmodell repräsentierte Objekt o
o	Objektrepräsentant im Umweltmodell
$\mathcal{P}_\mathcal{O}$	Menge aller Permutationen der Objektmenge $\mathcal{O}$
$\boldsymbol{q}_o$	benötigte Ausrichtung der Sensoren zur Fokussierung des im Umweltmodell repräsentierten Objekts o
Q	Anzahl an Gelenkwinkeln bzw. Freiheitsgraden
s_o	Salienz des im Umweltmodell repräsentierten Objekts o

Das OPASCA-System

$f_{\text{RGB}\rightarrow\text{HSV}}(\cdot)$	Transformationsfunktion zur Konvertierung von Bildpunkten aus dem RGB-Farbraum in den HSV-Farbraum
$H(\boldsymbol{x}); S(\boldsymbol{x}); V(\boldsymbol{x})$	Funktionen zur Selektion von Farbton (H), Sättigung (S) und Hellwert (V) eines Bildpunkts $\boldsymbol{x}$ im HSV-Farbraum
$\boldsymbol{I}_{\text{RGB}}$	Darstellung des gesamten Bilds im RGB-Farbraum
$\boldsymbol{I}_o^{\text{Grau}}$	Grauwert-basierter Bildausschnitt eines im Umweltmodell repräsentierten Objekts o
$\boldsymbol{I}_o^{\text{Kanten}}$	binäres Kantenbild eines im Umweltmodell repräsentierten Objekts o
λ^{Kanten}	Faktor, der den maximal benötigten Kantenanteil eines Objekts für eine vollständige Texturiertheit festlegt
$\lambda^{\text{Präsenz}}$	Faktor, der den maximal benötigten Anteil eines Objekts im Bild für eine vollständige Präsenz festlegt
φ_i	Winkel des Gradienten in Punkt i der Objektkontur
$\boldsymbol{x}^{\text{RGB}}$	Bildpunkt im RGB-Farbraum
$\boldsymbol{x}^{\text{HSV}}$	Bildpunkt im HSV-Farbraum
$\mathcal{X}_o^{\text{RGB}}$	Bildpunkte eines im Umweltmodell repräsentierten Objekts o im RGB-Farbraum
$\mathcal{X}_o^{\text{HSV}}$	Bildpunkte eines im Umweltmodell repräsentierten Objekts o im HSV-Farbraum

Experimentelle Evaluation

$\gamma_s^*, \gamma_\eta^*, \gamma_d^*$	optimale Einflussparameter für den interessengetriebenen Explorationspfad EP_{SNB}
$f : \mathbb{R}^m \rightarrow \mathbb{R}^n$	Abbildungsfunktion, welche die m Eingangsgrößen anhand von n Kriterien bewertet (Basis für die Pareto-Optimierung)
ι_N	Normalisierungsfaktor gibt den maximalen Positionsunterschied aller Objekte von zwei beliebigen Explorationspfaden mit denselben Objekten an
$I_{\text{EP}}(o)$	Funktion liefert die aktuelle Position des Objekts o auf dem Explorationspfad EP zurück
$\text{ICM}(\text{EP}_1, \text{EP}_2)$	Indizes-basiertes Korrelationsmaß für den Vergleich von zwei Explorationspfaden
$\lambda_{\text{NICS}}, \lambda_{\text{NICC}}, \lambda_{\text{NCJAD}}$	Skalarisierungsfaktoren zur Berücksichtigung des anteiligen Einflusses von Salienz, Neugier und Bewegung
$\Lambda_1, \ldots, \Lambda_{10}$	Zusammenstellungen von Skalarisierungsfaktoren zur Evaluation

η_{EP_i}	induzierte Neugier für das i-te Objekt auf dem Explorationspfad
$\mathcal{P}_{\mathrm{Front}}$	Pareto-Front
$\mathcal{P}_{\mathrm{Menge}}$	Pareto-Menge
$\mathrm{PSS}_{\mathrm{SNB}}$	Parameterauswahlbereich zur Optimierung der Einflussparameter: Salienz, Neugier und Bewegung
$\boldsymbol{q}_{\mathrm{EP}_i}$	Gelenkwinkelstellung um das i-te Objekt auf dem Explorationspfad EP zu fokussieren
Q	Anzahl an Gelenkwinkeln bzw. Freiheitsgraden
s_{EP_i}	Salienzwert des i-ten Objekts auf dem Explorationspfad EP
$\boldsymbol{x} \in \mathbb{R}^m$	Eingangsgrößen der Pareto-Optimierung (m-Tupel)
$\mathcal{X}$	Menge aller Kombinationen von Eingangsgrößen
$\mathcal{Y}$	bewertete Menge aller Kombinationen von Eingangsgrößen

Teil I

Grundlagen

1

Einführung

Dieses Kapitel beschäftigt sich mit den grundlegenden Überlegungen, der Motivation und dem Kontext sowie der Zielsetzung der vorliegenden Arbeit. Dabei werden die wissenschaftlichen Beiträge sowie die Anknüpfungspunkte zu anderen Teilbereichen der Wissenschaft hervorgehoben, bevor in den nachfolgenden Kapiteln detailliert auf die Voraussetzungen und Grundlagen eingegangen wird. Anschließend werden nach und nach die einzelnen Aspekte der interessengetriebenen Exploration näher beschrieben, in ein Gesamtkonzept überführt sowie abschließend evaluiert.

1.1 Motivation und Kontext der Arbeit

In den letzten Jahren hat eine immer stärker werdende Technisierung des alltäglichen Lebens stattgefunden. Dabei ist auch das Interesse an verschiedenen Arten von Robotern, z. B. für den Haushalt, immer weiter angestiegen. Insbesondere die Entwicklung und die praktische Einsatzmöglichkeit von speziellen Robotern mit einem beschränkten Aufgabengebiet, wie beispielsweise Staubsaugrobotern, sind heutzutage schon weit vorangeschritten. Auch in weiteren Bereichen des Haushalts sind immer mehr technologische Verbesserungen vorhanden. Diese sind allerdings meist nicht auf den ersten Blick sichtbar, sondern arbeiten z. B. in einem intelligenten Backofen oder einer modernen Waschmaschine, fast unsichtbar im Hintergrund.

Forscher arbeiten weltweit bereits seit vielen Jahren an Robotern und anderen autonomen System, die den Menschen bei der Arbeit – u. a. im Haushalt – unterstützen bzw. die Arbeit übernehmen. Dies ist nicht nur in einer immer älter werdenden Gesellschaft sinnvoll, sondern kommt auch beruflich stark eingespannten Personen zugute. Es gibt hierbei zwei grundlegend verschiedene Ansätze: Zum

einen werden Einzellösungen entwickelt, welche bestimmte Bereiche vereinfachen und dabei z. B. vorhandene alltägliche Geräte verbessern und aufwerten. Diese Technologien sind bereits heute für die breite Masse verfügbar, haben allerdings den Nachteil, dass sie jeweils nur einen relativ kleinen Teil abdecken und keine ganzheitlich unterstützende Funktionalität bieten. Zum anderen werden Ansätze verfolgt, bei denen ein mobiles und flexibles System mehrere Aufgaben bewältigen und sich an die jeweilige Situation anpassen kann.

Der durch die Deutsche Forschungsgemeinschaft (DFG) geförderte Sonderforschungsbereich (SFB) 588 „Humanoide Roboter – Lernende und kooperierende multimodale Roboter“ (vgl. [Son12]) hat sich zur Aufgabe gemacht, einen humanoiden Roboter zu entwickeln, der die zuvor angesprochenen Aufgaben des Alltags meistern und den Menschen vielfältig unterstützen kann. Dieser soll sich dabei an neue Situationen anpassen und neue Fertigkeiten erlernen können. Als Umgebung wurde die Küche ausgewählt, da diese eine Vielzahl an verschiedenen Herausforderungen und Aufgaben beinhaltet. Bei einem komplexen System wie einem humanoiden Roboter existieren neben den mechatronischen Herausforderungen, einschließlich der Regelung, vor allem auch anspruchsvolle Aufgaben im Bereich der Perzeption und Kognition. Die hier vorliegende Arbeit fand im Kontext des Sonderforschungsbereichs 588 statt.

1.2 Zielsetzung der Arbeit

Für einen humanoiden Roboter – oder allgemein: ein autonomes System – ist es notwendig, seine aktuelle Umgebung (auch: *Szene*) mit den darin befindenden Gegenständen und Personen – fortan auch zusammengefasst als *Objekte* bezeichnet – zu erfassen. Dazu ist es notwendig, bekannte Objekte jederzeit wiederzuerkennen, unbekannte Objektinformationen zu akquirieren und neue Objekte zu registrieren. Dies lässt sich im Kontext einer (vollständigen) Erfassung der aktuellen Umgebung als *Szenenexploration* beschreiben.

Das Ziel der vorliegenden Arbeit ist es, eine priorisierte und zugleich zielgerichtete Exploration der Objekte in einer Szene durchzuführen, bei der *interessante Objekte* bevorzugt untersucht werden. Dabei dienen Aspekte der Aufmerksamkeit und der Neugier als Grundlage für eine *interessengetriebene, audiovisuelle Szenenexploration*. In der Arbeit wurden folgende Themengebiete als Schwerpunkte identifiziert und näher untersucht:

- Definition und Realisierung von *wissensbasierter Neugier* für ein autonomes System durch Adaptation der situativen Bedingungen für Neugier beim Menschen,

- Entwurf und Realisierung eines *interessengetriebenen audiovisuellen Explorationsansatzes* zur bedarfsorientierten Erfassung von Objekten in einer Szene für die Bewältigung von Aufgaben und die Interaktion,
- Evaluation von mehreren Strategien zur Exploration einer Szene unter Berücksichtigung verschiedener *Priorisierungskriterien* (u. a. wissensbasierter Neugier).

Diese Arbeit verwendet die vorhandenen Erkenntnisse zur wissensbasierten und objektzentrierten Umwelterfassung (vgl. [Mac10b], [Küh10]) als Ausgangsbasis und erweitert diese. Dabei werden sowohl die Ansätze zur Generierung und Fusion von Umweltwissen (vgl. [Mac10a]) als auch die audiovisuellen Signaturen zur Objektwiedererkennung (vgl. [Swe09]) berücksichtigt. Beides kann für eine tiefgehendere Analyse von Objekten eingesetzt werden, d. h. eine detaillierte schrittweise Erfassung von Objektinformationen mit verschiedenartigen Sensortypen und unter Zuhilfenahme von weiteren Informationsquellen. Das im Rahmen der zuvor genannten Arbeiten entwickelte System zur audiovisuellen Szenenanalyse (kurz: OPASCA; vgl. [Mac10b]) wurde im Hinblick auf die interessengetriebene Exploration im Rahmen dieser Arbeit stark erweitert und angepasst. Dabei wurden u. a. die Aspekte *wissensbasierte Neugier* und *multimodale Salienz* (d. h. Auffälligkeit) aufgenommen. Darüber hinaus wurde die reine Perzeption in eine interessengetriebene Exploration der Szene überführt.

Neben den zuvor genannten Grundlagen existieren für die hier vorliegende Arbeit eine Reihe an Rahmenbedingungen und Annahmen. Diese werden im Folgenden kurz zusammengefasst:

- Für die Arbeit wird eine zentrale Erfassung der Sensordaten durch einen (mobilen) Roboter oder mit einem vergleichbaren Sensoraufbau vorausgesetzt. Dies begründet sich durch den Kontext der Arbeit, dem Sonderforschungsbereich 588, in dem ein humanoider Roboter mit onboard-Sensorik im Mittelpunkt steht. Dadurch ergeben sich auch die berücksichtigten Sensorarten: Es werden zwei Stereokameras für den Nah- bzw. Fernbereich zur gleichzeitigen Erfassung der Szene im Detail und im Überblick verwendet, sowie ein Mikrofonarray für die weitere Informationsakquise. Insgesamt ist somit eine Erfassung der aktuellen Szene mit mehreren Modalitäten möglich.
- Für die gleichzeitige Analyse aller in der Szene erfassten Objekte sind nicht genügend Rechenressourcen vorhanden. Außerdem kann mit den vorhandenen Sensoren immer nur ein Teil der gesamten Szene zu einem Zeitpunkt detailliert erfasst werden. Infolgedessen müssen sowohl Sensoren als auch Rechenressourcen gezielt eingesetzt werden.
- Des Weiteren wird die Annahme getroffen, dass die in der Szene vorhandenen Objekte nur teilweise als bekannt vorausgesetzt werden können, d. h. es können a-priori unbekannte Personen und Gegenstände vorhanden sein. Zu-

sätzlich existiert a-priori-Wissen über den aktuellen Raum, d. h. die Raumgeometrie ist bekannt einschließlich der im Raum befindlichen Möbel. Außerdem wird vorausgesetzt, dass eine Selbstlokalisierung mit großer Genauigkeit vorhanden ist und somit jederzeit eine absolute Position des Roboters und der wahrgenommenen Objekte zu einem Referenzpunkt im Raum zur Verfügung steht (vgl. [Gon08]).

- Im Rahmen der vorliegenden Arbeit sollen Ansätze, Verfahren und Vorgehensweisen entworfen bzw. verwendet werden, welche nicht nur für einen humanoiden Roboter geeignet sind, sondern sich auch auf andere autonome Systeme mit verteilten und/oder stationären Sensoreinheiten übertragen lassen.

Aus der eingangs beschriebenen *interessengetriebenen audiovisuellen Exploration* können folgende zusätzliche Bedingungen direkt oder indirekt abgeleitet werden:

- Im Rahmen der interessengetriebenen Exploration findet eine erweiterte objektzentrierte Erfassung statt, bei welcher die Objekte in einer Szene gezielt anhand von Priorisierungszielen nach und nach detailliert wahrgenommen werden.
- Die Priorisierungsziele der Exploration leiten sich aus der Definition des Begriffs *Interesse* im Rahmen der Exploration ab. Beim Menschen lässt sich Interesse als Kombination von Neugier und Salienz (Auffälligkeit) definieren. Diese und weitere Kriterien sollen vereint werden, um eine Priorisierung bei der Objektselektion im Rahmen der Exploration zu realisieren.
- Bei der Bestimmung sowohl der wissensbasierten Neugier als auch der Salienz sollen audiovisuelle Informationen über die Objekte berücksichtigt werden.
- Das vorhandene a-priori-Wissen ist eine wichtige Grundlage für die wissensbasierte Neugier. Dieses beeinflusst insbesondere den Teilaspekt *Neuartigkeit* und ist somit eine wichtige Wissensquelle.
- Während der Exploration ist es wichtig, auch sogenannte externe Faktoren zu berücksichtigen. Dies können beispielsweise anstehende Aufgaben, die Interaktion mit dem Menschen oder eine Veränderung der Umgebung sein. In diesem Zusammenhang kann die Exploration frühzeitig beendet oder aber unterbrochen werden und zu einem späteren Zeitpunkt gegebenenfalls fortgeführt werden. Im letzteren Fall sollten die zwischenzeitlichen Änderungen in der Szene berücksichtigt werden und keine vollständige Exploration vollzogen werden.

1.3 Wissenschaftliche Beiträge

Im Teilprojekt P2 „multimodale Exploration“ des Sonderforschungsbereichs 588 wurden die Grundlagen für die Bestimmung der Aufmerksamkeit eines Roboters geschaffen (vgl. [Sch11a], [Küh12b]). Dabei wurde die multimodale Salienz als ein Ansatz zur Bestimmung der Aufmerksamkeit genutzt. Im Rahmen der vorliegenden Arbeit wird aus den aktuellen audiovisuellen Sensordaten die Salienz bestimmt und in Form multimodaler Salienzcluster repräsentiert (vgl. Kapitel 2). Die dabei gewonnenen Informationen dienen als Grundlage für die Priorisierung der Objekte bei der gezielten Exploration einer Szene (vgl. Kapitel 5).

Ein wichtiger Beitrag der vorliegenden Arbeit ist die Übertragung der situativen Bedingungen für die Neugier beim Menschen aus der Psychologie auf ein autonomes System. Die einzelnen Bedingungen werden dabei auf Basis von Informationen aus einem objektzentrierten Umweltmodell (vgl. Kapitel 3) modelliert und anschließend zu einem Gesamtmaß für die wissensbasierte Neugier bzgl. eines Objekts vereint (vgl. Kapitel 4).

Für die Exploration einer Szene werden in der vorliegenden Arbeit neue Strategien zur Priorisierung von Objekten vorgestellt (vgl. Kapitel 5). Eine solche Priorisierung ist notwendig, da meist nur beschränkte Ressourcen (z. B. Rechenressourcen, Ausschnitt der Szene oder verfügbare Zeit) zur Verfügung stehen. Zur Bestimmung der Priorität und somit einer interessengetriebenen Explorationsstrategie wird eine Kombination von verschiedenen interessenbasierten Priorisierungskriterien (u. a. Neugier und Salienz) eingeführt. Dies stellt in dieser Kombination und Herangehensweise einen neuen Ansatz bei der Exploration einer Szene dar (vgl. Kapitel 5).

Die Exploration einer Szene ist ein komplexer Prozess, bei dem viele einzelne Komponenten (u. a. Bild- und Signalverarbeitung zur Objektdetektion und -klassifikation sowie übergeordnete Bestimmung und Ausführung einer Explorationsstrategie) ihren Beitrag leisten. Nur durch die gezielte Verknüpfung von einzelnen Ansätzen in einem ganzheitlichen System zur audiovisuellen Szenenanalyse und -exploration (vgl. Kapitel 6) ist es letztendlich möglich, eine Szene in der notwendigen Tiefe und mit der gesetzten Priorisierung effektiv zu explorieren.

Des Weiteren wird im Rahmen der Evaluation die Bestimmung der optimalen Parameter für die interessengetriebene Explorationsstrategie untersucht. Die Priorisierungsziele der einzelnen Beiträge zur interessengetrieben Exploration sind dabei grundsätzlich gegensätzlich. Durch die Vorgabe von Präferenzen können die Priorisierungsziele in Form von Einflussfaktoren gegeneinander in Relation gesetzt werden. Dabei kann gezeigt werden, dass bei einer geringfügigen Anpassung der Ausgangsparameter eine insgesamt viel bessere Erfüllung der einzelnen konträren Priorisierungsziele erreicht wird (vgl. Kapitel 7).

Zuletzt wird im Rahmen der vorliegenden Arbeit ein neues Maß für die Evaluation eingeführt, welches zum einen die Priorisierung der Objekte bei zwei unterschiedlichen Explorationsstrategien bewertet und zum anderen indirekt beurteilt, wie sehr ein gegebener Explorationspfad Aspekte wie Neugier oder Salienz berücksichtigt (vgl. Kapitel 7).

1.4 Gliederung der Arbeit

Die vorliegende Arbeit gliedert sich in fünf Teile. Im ersten Teil werden die Grundlagen der Arbeit vorgestellt. In *Kapitel 1* sind die Motivation und die Einordnung und der Kontext der Arbeit, die Zielsetzung sowie die wissenschaftlichen Beiträge enthalten. In *Kapitel 2* wird die multimodale Salienz beschrieben, welche visuell und akustisch herausstechende Wahrnehmungen repräsentiert und sowohl bei der Modellierung der Aufmerksamkeit als auch bei der interessengetriebenen Exploration ihre Anwendung findet. Zunächst werden dazu unimodale Salienzcluster definiert, welche die visuelle oder akustische Salienz darstellen. Anschließend werden diese zu multimodalen Salienzclustern fusioniert und repräsentieren somit räumlich lokalisierte herausstechende Wahrnehmungen von mehreren Modalitäten. In *Kapitel 3* wird die objektzentrierte Umwelterfassung vorgestellt, welche die formale Definition eines Umweltmodells für die Repräsentation von Personen, Gegenständen und a-priori-Wissen umfasst sowie die Themen Abstraktionsebenen, Klassenhierarchie, Systemarchitektur, Objektsignaturen und Lebenszyklus eines Objekts abdeckt.

Der zweite Teil der Arbeit stellt die interessengetriebene Szenenexploration vor. In *Kapitel 4* wird dazu die wissensbasierte Neugier vorgestellt, welche auf Erkenntnissen aus der Psychologie aufbaut und auf Informationen aus einem objektzentrierten Umweltmodell zurückgreift. Die Neugier bildet zusammen mit weiteren Interessenaspekten die Grundlage für die Exploration. In *Kapitel 5* wird die interessengetriebene Exploration beschrieben, welche die Salienz und die Neugier mit weiteren Parametern kombiniert. Diese Kriterien drücken das berücksichtigte Interesse für einzelne Objekte im Rahmen der audiovisuellen Exploration aus. Für die spätere Evaluation werden in diesem Kapitel weitere Explorationsstrategien definiert und dabei die Komplexität bei der Bestimmung des sogenannten Explorationspfads untersucht.

Der dritte Teil der Arbeit beschäftigt sich mit der Umsetzung und Evaluation der zuvor vorgestellten Ansätze und Methoden für ein reales System. In *Kapitel 6* wird ein ganzheitlicher Systemansatz für die Szenenexploration vorgestellt. Eine Beschreibung des generellen Aufbaus, der einzelnen Module sowie der Realisierung der interessengetriebenen Exploration erfolgt in diesem Zusammenhang. Die Grundlage dafür bildet das zuvor genannte OPASCA-System. In

Kapitel 7 findet die experimentelle Evaluation statt. Dabei werden zunächst die verwendeten Hardwareplattformen vorgestellt. Anschließend erfolgt eine Übersicht der Evaluationsdatensätze für unterschiedliche Szenarien. Danach werden Maße für die Bewertung und den Vergleich von Pfaden definiert, bevor abschließend die eigentliche Evaluation der Explorationspfade bzw. der Explorationsstrategien selbst erfolgt.

Der vierte Teil mit *Kapitel 8* fasst den Inhalt der Arbeit zusammen und gibt einen Ausblick auf weiterführende Ideen und Ansätze für zukünftige Arbeiten.

Den letzten Teil der Arbeit bildet der *Anhang*, welcher ausführliche Beispiele für die wissensbasierte Neugier inklusive der vorhandenen Daten im Umweltmodell zeigt. Hierbei werden auch die Ergebnisse der gewählten Fusionsstrategie zur Bestimmung der Neugier dargestellt.

2

Aufmerksamkeit und Salienz

In diesem Kapitel wird zunächst auf die Aufmerksamkeit näher eingegangen, da diese in vielen Bereichen, u. a. für einen humanoiden Roboter und andere autonome Systeme von hoher Bedeutung ist. Grundvoraussetzungen für die Aufmerksamkeit ist die Auffälligkeit (auch: Salienz) von einzelnen Objekten, Geräuschen oder Abläufen in einer Szene. Es wird im Folgenden auf die sogenannte multimodale Salienz näher eingegangen, welche die Grundlage der Aufmerksamkeit bildet und für verschiedene Modalitäten (z. B. visuell und akustisch) modelliert werden kann. Im weiteren Verlauf der vorliegenden Arbeit findet die Salienz Anwendung als ein Kriterium für die Priorisierung einzelner Objekte während der Szenenexploration.

2.1 Motivation

Im Laufe der letzten Jahrzehnte entstanden verschiedene Theorien, wie der Prozess der Aufmerksamkeitsselektion beim Menschen abläuft. Zum einen wurde die These aufgestellt, dass die Selektion der Aufmerksamkeit bereits in einem frühen Stadium der Wahrnehmung geschieht. So wurde in der Filtertheorie von Broadbent (vgl. [Bro58]) und Dämpfungstheorie von Treisman (vgl. [Tre64]) von einem Vorverarbeitungsschritt ausgegangen, welcher die sensorischen Reize reduziert und früh eine Selektion vornimmt, noch bevor eine semantische Erkennung erfolgt. Zum anderen wurde in der Theorie von Deutsch & Deutsch (vgl. [Deu63]) beschrieben, dass zunächst die Reize inhaltlich analysiert werden, bevor eine Selektion stattfindet. Neuere Theorien (vgl. [Bun90]) gehen von einem zweistufigen Prozess aus, welcher zunächst die sensorischen Informationen als Merkmale repräsentiert und gewichtet, bevor diese anschließend kategorisiert werden. Dabei kann nur eine bestimmte Anzahl an Objekten gleichzeitig kategorisiert und im

Kurzeitgedächtnis repräsentiert werden. Alle anderen Objekte werden nicht bewusst wahrgenommen und somit ausgeblendet. Aktuelle Forschungsergebnisse gehen von einem orts- oder objektbasierten Zusammenhang aus, welche in Studien von Brefczynski (vgl. [Bre99]) bestätigt werden konnten.

Diese Erkenntnisse über die Wahrnehmung beim Menschen können als Grundlage für die Modellierung der Aufmerksamkeit bei einem humanoiden Roboter oder einem anderen autonomen System mit entsprechenden Fähigkeiten Anwendung finden. Diese Systeme müssen in der Lage sein, neben expliziten Aufgaben und Tätigkeiten, stets auf die aktuelle Szene – d. h. auf Inhalte und Ereignisse in einem gewissen Umfeld – reagieren zu können. Dazu ist es notwendig, relevante Bereiche des sensorischen Inputs zu detektieren und irrelevante oder weniger wichtige Bereiche (partiell) zu unterdrücken. Die Selektion des Relevanten hängt von einer Vielzahl an Faktoren ab. Während die Aufmerksamkeit einen Bottom-up-Informationsfluss darstellt, ist die Selektion und Unterdrückung einzelner Stimuli ein Top-down-Prozess, welcher in der Regel von einer Vielzahl an zusätzlichen Informationsquellen (z. B. den aktuellen Aufgaben oder dem Verhalten bei bestimmten Ereignissen) beeinflusst wird. Ein ort- bzw. objektzentrierter Ansatz abstrahiert dabei die Szene insoweit, dass die wahrgenommenen sensorischen Informationen für die Darstellung in ihrer Komplexität reduziert werden und somit erst ein anschließender Verarbeitungsansatz auf einem abstrakteren Niveau (d. h. Objektebene) ermöglicht wird.

Die selektive Aufmerksamkeit kann dahingehend unterschieden werden, ob bei gleichbleibendem sensorischem Ausschnitt der Umwelt eine aktive Bewegung (z. B. Kopfdrehung oder Augenbewegung) in Richtung eines Objekts zur Fokussierung unternommen (engl.: overt attention) oder nur der aktuelle Fokus verschoben wird (engl.: covert attention). Letzteres kann auch eine virtuelle oder gedankliche Verschiebung des Aufmerksamkeitsfokus auf zuvor wahrgenommene Objekte sein. Zusätzlich ist die Fähigkeit, die Aufmerksamkeit nach einer Weile auf einen neuen Ort bzw. ein neues Objekt zu lenken, wichtig. Durch die sogenannte *Hemmung der Rückkehr* (engl.: inhibition of return; kurz: IOR) zu bereits fokussierten Objekten wird sichergestellt, dass kürzlich fokussierte Orte bzw. Objekte nicht unmittelbar wieder selektiert werden.

Die Aufmerksamkeit ist allgemein eine wichtige Fähigkeit bei der Exploration einer Szene – d. h. eine tiefgehendere Erfassung von beispielsweise Personen und Gegenständen in der aktuellen Umgebung –, da die Aufmerksamkeit dafür verantwortlich ist, den Fokus gezielt auf besonders wichtige Objekte und Ereignisse zu lenken (vgl. [Itt98]). Die Berücksichtigung von mehreren Modalitäten (bspw. visuell und akustisch) lässt eine vielfältige Modellierung der Aufmerksamkeit zu (vgl. [Sch11a], [Küh12b]). Dabei kann die aktuelle Aufmerksamkeit anhand der *Salienz* (auch: Auffälligkeit) für jedes Objekt bestimmt werden. Hierbei ist die Sa-

lienz als ein Reiz oder eine Wahrnehmung zu verstehen, welche(r) sich vom Rest der Szene abhebt und dadurch deutlicher wahrgenommen werden kann.

Die Salienz kann allgemein genutzt werden, um den Aufmerksamkeitsfokus eines Roboters festzulegen – oder wie im Rahmen der vorliegenden Arbeit: um speziell die Priorität von Objekten während der Exploration festzulegen. Dabei werden Objekte, die eine bedeutende Präsenz[1] in der Szene aufweisen, früher exploriert als Objekte mit einer unbedeutenden Präsenz (vgl. Kapitel 5; [Küh12b]).

2.2 Stand der Forschung

Im Bereich der Aufmerksamkeitsforschung wurden in der letzten Dekade eine Vielzahl an Artikeln (vgl. [Itt00], [Ora05], [Meg07], [But08], [Rue08], [Xu09], [Beg10]) publiziert, welche insbesondere Roboter oder andere autonome Systeme im Fokus haben. Die Veröffentlichungen setzen dabei auf eine Vielzahl an unterschiedlichen Herangehensweisen, um saliente Ausschnitte von Eingangssignalen (z. B. Bilder oder Geräusche) zu modellieren. Als *salient* sind hierbei, wie schon im vorigen Abschnitt erwähnt, beispielsweise Bereiche innerhalb eines Audiosignals oder eines Bildes zu bezeichnen, welche besonders hervorstechen und damit die Aufmerksamkeit des Menschen oder des Roboters auf sich ziehen.

Die sogenannten *Salienzkarten* (engl.: saliency maps) wurden von Koch und Ullman (vgl. [Koc85]) zur Bestimmung aktueller und künftiger visueller Schwerpunkte der Aufmerksamkeit auf Basis von grundlegenden Merkmalen (u. a. Orientierung, Farbe und Bewegung) eingeführt. Dies ist der klassische Ansatz, der meist eine Punkt-zu-Punkt-Referenz zwischen dem aktuellen Kamerabild und der Salienzkarte darstellt. Dabei korrespondieren i. d. R. saliente Regionen mit höheren Werten und weniger saliente bzw. nicht-saliente Regionen mit entsprechend geringeren Werten. Es entsteht eine Darstellung ähnlich dem Höhenprofil einer Landkarte: Die Region mit dem höchsten Wert erhält den aktuellen Fokus. Wird dieser unterdrückt, so erhält die Region mit dem nächsthöheren Wert den Fokus. Diese Art der Darstellung ist heute immer noch stark verbreitet (vgl. [Itt00], [Meg07] et al.) und dient auch als Ausgangsbasis für andere Ansätze wie Egosphären und Proto-Objekte (vgl. [Rue08], [Sch11a] et al.).

Salienzkarten, welche zunächst nur den aktuellen Kameraausschnitt repräsentierten, wurden im Laufe der Zeit immer mehr verfeinert. So wurden die Salienzkarten auf ein 360-Grad-*Panorama* (u. a. [Bur06], [Nic07], [Sar09]) erweitert oder auch für eine komplette *Egosphäre* (vgl. [Pet01], [Fle06], [Rue08], [Wel11])

[1] Objekte, welche sich von ihrer Umgebung stärker abheben als andere Objekte (d. h. herausstechende Merkmale besitzen) und somit vom Betrachter deutlicher wahrgenommen werden.

definiert. Diese Repräsentationen gehen meist von einem Sensoraufbau mit zwei oder mehr Freiheitsgraden bei einer festen Position aus, welche als Ursprung des Koordinatensystems dient. Eine Translation der Sensorposition führt meist zu einer aufwendigen Nachführung oder sogar zur Ungültigkeit der zuvor akquirierten Salienzinformation.

Losgelöst von einer festen Sensorposition kann die Salienz auch in Form von sogenannten *Proto-Objekten* (vgl. [Wal06], [Ora07], [Sch11a], [Küh12b] et al.) dargestellt werden. Dabei werden relevante Regionen als abstrakte Objekte im Raum repräsentiert und es findet somit keine kontinuierliche Darstellung der Salienz über den kompletten Raum mehr statt. Durch die Referenzierung der Proto-Objekte in einem globalen Koordinatensystem sowie die kompakte Darstellung und die damit verbundene Datenreduktion wird die Verwendung auch für mobile Roboter erleichtert.

Die Umsetzung von Salienzkarten für die *akustische Salienz* wurde beispielsweise von Kayser und Kalinli (vgl. [Kay05], [Kal07]) realisiert. Dabei werden in den Audiosignalen saliente Elemente gesucht und durch deren Lokalisation in eine akustische Karte analog zu einer visuellen Salienzkarte eingetragen.

In der Literatur findet die visuelle Salienz im Gegensatz zur akustischen eine höhere Beachtung. Insbesondere der Einbeziehung mehrerer Modalitäten kam erst in den letzten Jahren eine größere Bedeutung zu (siehe [Kay05], [Kal09], [Sch11a]). Dabei hat die Kombination visueller und akustischer Salienzen den Vorteil, dass sowohl sichtbare Bereiche als auch hörbare Ereignisse – die möglicherweise außerhalb des aktuellen Blickfeldes stattfinden – gleichzeitig wahrgenommen werden können.

Im folgenden Abschnitt wird die multimodale Salienz und deren Definition – wie diese in der Arbeit nachfolgend verwendet wird – näher beschrieben. Dabei werden saliente Regionen ähnlich wie bei Proto-Objekten als sogenannte *Salienzcluster* repräsentiert.

2.3 Multimodale Salienz

Die *multimodale Salienz* besteht in der vorliegenden Arbeit aus der *akustischen Salienz*, welche den hörbaren Anteil der Salienz repräsentiert, und der *visuellen Salienz*, welche den sichtbaren Anteil darstellt. Für beide Modalitäten werden zunächst getrennt *akustische* und *visuelle Salienzcluster* bestimmt, die räumliche Salienzrepräsentationen für die jeweilige Modalität bilden. Anschließend werden diese unimodalen Salienzcluster durch eine räumlich-zeitliche Fusion zu *multimodalen Salienzclustern* vereint.

2.3.1 Akustische Salienz

Die *akustische Salienz* beschreibt hörbare Ereignisse im näheren Umfeld, welche sich aufgrund von bestimmten Merkmalen von den umgebenden Geräuschen abheben. Die dadurch erzeugte Aufmerksamkeit lenkt den Fokus auf das Ereignis. Dies ist beispielsweise bei starken und abrupten Änderungen im akustischen Spektrum der Fall, welche beispielsweise durch das Herunterfallen eines Objekts ausgelöst werden.

Bestimmung der akustischen Salienz

In Anlehnung an ein Verfahren aus der Bildverarbeitung, der sogenannten *Bayesian Surprise* (vgl. [Itt06]), wurde der Ansatz für akustische Signale adaptiert (vgl. [Sch11a], [Küh12b]), um saliente hörbare Ereignisse zu detektieren. Als Grundlage zur Bestimmung der akustischen Salienz dient die diskrete Kurzzeit-Fouriertransformation (engl.: short-time Fourier transform; kurz: STFT).

Zunächst wird hierbei das empfangene Audiosignal $y(t)$ abgetastet und in einzelne sich überlappende Fenster der Länge N unterteilt (vgl. Abb. 2.1). Der Versatz der Fenster wird dabei über den Parameter L eingestellt und sollte die Korrelation der einzelnen Fenster berücksichtigen. Die diskreten Zeitpunkte werden durch den Parameter n dargestellt und das Ende eines jeden Fensters wird in Abhängigkeit des Parameters l definiert.

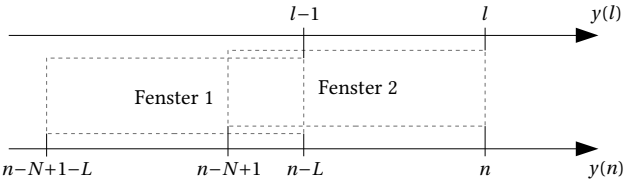

Abb. 2.1: Segmentierung des Audiosignals $y(n)$ in sich überlappende Fenster der Länge N mit Versatz L.

Anschließend werden die zuvor definierten Fenster eines Audiosignals mit einer Fensterfunktion (z.B. Hamming oder Blackman; vgl. [Kam12]) gewichtet, um unerwünschte Nebeneffekte (u. a. Leckeffekt) bei der Fouriertransformation zu verringern, da nur ein begrenzter Beobachtungszeitraum für die Bestimmung zur Verfügung steht. Im nächsten Schritt wird das Spektrogramm $Y_{\text{Spek}}(n,k)$ für jeden Zeitpunkt n mittels der diskreten STFT ermittelt über

$$Y_{\text{Spek}}(n,k) = |Y(n,k)|^2 = \left| \sum_{i=n-N+1}^{n} y(i) \cdot e^{-j\frac{2\pi}{N}ik} \right|^2 \quad \text{mit } k = 0,\ldots,N-1\,. \tag{2.1}$$

Hierbei stellen n und k diskrete Zeit- bzw. Frequenzparameter dar. Die Länge des Fensters ist mit dem Parameter N gegeben.

Für die Bestimmung der akustischen Salienz wird, wie eingangs erwähnt, die Bayesian Surprise verwendet, welche durch den Vergleich von a-priori- und a-posteriori-Wahrscheinlichkeitsdichten bestimmt werden kann. Hierfür wird die Bayes'sche Statistik verwendet. In dieser korrespondieren Wahrscheinlichkeiten mit dem Grad-des-Dafürhaltens eines jeweiligen Modelles, welches nach dem Satz von Bayes aktualisiert werden kann, wenn neue Daten akquiriert werden. Zu diesem Zweck wird in jedem Zeitschritt l die a-priori-Wahrscheinlichkeitsdichte $P_{\text{Prior}}^{l,k}$ mit den neuen Daten des aktuellen Spektrogramms $Y_{\text{Spek}}(l,k)$ für jede Frequenz separat aktualisiert und damit die a-posteriori-Wahrscheinlichkeitsdichte $P_{\text{Post}}^{l,k}$ bestimmt. Die Anzahl $M \in \mathbb{N}^+$ der genutzten Spektrogramme (d. h. die Historie) und damit das Zeitfenster, in welchem die beiden Wahrscheinlichkeitsdichtefunktionen bestimmt werden können, ist dabei für reale Anwendungen beschränkt. Somit stehen für die Bestimmung der a-priori- und a-posteriori-Dichten folgende Daten zur Verfügung:

$$\boldsymbol{Y}_{\text{Prior}}(l,k) = (Y_{\text{Spek}}(l-1,k), Y_{\text{Spek}}(l-2,k),\ldots, Y_{\text{Spek}}(l-M,k))^T \tag{2.2}$$

$$\boldsymbol{Y}_{\text{Post}}(l,k) = (Y_{\text{Spek}}(l,k), Y_{\text{Spek}}(l-1,k),\ldots, Y_{\text{Spek}}(l-M,k))^T\,. \tag{2.3}$$

Dieser Zusammenhang ist in Abb. 2.2 noch einmal dargestellt. Des Weiteren ist auch die zeitliche Beziehung zwischen dem ursprünglichen Audiosignal in Abb. 2.1 und den Spektrogrammen $\boldsymbol{Y}_{\text{Spek}}$ visualisiert.

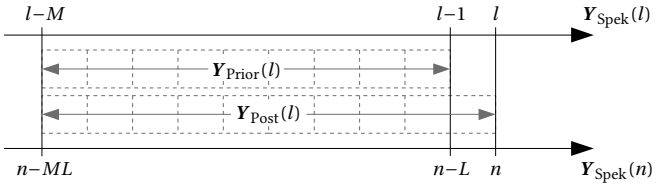

Abb. 2.2: Übersicht der vorhandenen Spektrogramme zur Bestimmung der a-priori- und a-posteriori-Wahrscheinlichkeitsdichten.

Die akustische Salienz kann durch den Vergleich der zuvor erwähnten Dichten auf signifikante Änderungen bestimmt werden. Dies kann mit Hilfe der Kullback-Leibler-Divergenz (engl.: Kullback-Leibler divergence; kurz: KLD) geschehen,

welche die *relative Entropie* zwischen zwei Wahrscheinlichkeitsdichtefunktionen f und g definiert:

$$D_{\mathrm{KL}}(f||g) := \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx. \tag{2.4}$$

Für d-dimensionale normalverteilte Dichten $\hat{f}$ und $\hat{g}$ kann dies in einer geschlossenen Form (vgl. [Her07]) berechnet werden mit

$$D_{\mathrm{KL}}(\hat{f}||\hat{g}) = \frac{1}{2}\left[\log \frac{|\boldsymbol{\Sigma}_{\hat{g}}|}{|\boldsymbol{\Sigma}_{\hat{f}}|} + \mathrm{tr}\left[\boldsymbol{\Sigma}_{\hat{g}}^{-1}\boldsymbol{\Sigma}_{\hat{f}}\right] - d + \left(\boldsymbol{\mu}_{\hat{f}} - \boldsymbol{\mu}_{\hat{g}}\right)^{\top} \boldsymbol{\Sigma}_{\hat{g}}^{-1}\left(\boldsymbol{\mu}_{\hat{f}} - \boldsymbol{\mu}_{\hat{g}}\right)\right]. \tag{2.5}$$

Die Bestimmung von mehrdimensionalen Normalverteilungen wird mit zunehmender Dimension ein immer aufwendigerer Prozess. In Hinblick auf die spätere Anwendung wird deshalb die akustische Salienz $S_{\mathrm{akustisch}}(l,k)$ mit Hilfe der KLD in jedem Blockschritt l für jede Frequenz k im Spektrogramm separat bestimmt. Infolgedessen werden nur eindimensionale Normalverteilungen benötigt, welche effizient bestimmt werden können. Somit vereinfacht sich Gl. 2.5 zu

$$\begin{aligned} S_{\mathrm{akustisch}}(l,k) &:= D_{\mathrm{KL}}\left(P_{\mathrm{Post}}^{l,k} \,\middle\|\, P_{\mathrm{Prior}}^{l,k}\right) \\ &= \frac{1}{2}\left[\log \frac{\mathrm{var}_{\mathrm{Prior}}^{l,k}}{\mathrm{var}_{\mathrm{Post}}^{l,k}} + \frac{\mathrm{var}_{\mathrm{Post}}^{l,k}}{\mathrm{var}_{\mathrm{Prior}}^{l,k}} - 1 + \frac{\left(\mu_{\mathrm{Post}}^{l,k} - \mu_{\mathrm{Prior}}^{l,k}\right)^2}{\mathrm{var}_{\mathrm{Prior}}^{l,k}}\right]. \end{aligned} \tag{2.6}$$

Die a-priori- und a-posteriori-Dichten für die Observationen $Y_{\mathrm{Spek}}(l,k)$ sind definiert als

$$P_{\mathrm{Prior}}^{l,k} := \mathcal{N}\left(\mu_{\mathrm{Prior}}^{l,k}, \mathrm{var}_{\mathrm{Prior}}^{l,k}\right) \quad \text{bzw.} \quad P_{\mathrm{Post}}^{l,k} := \mathcal{N}\left(\mu_{\mathrm{Post}}^{l,k}, \mathrm{var}_{\mathrm{Post}}^{l,k}\right) \tag{2.7}$$

und werden durch die erwartungstreuen Schätzwerte (vgl. [Kro11]) für die Mittelwerte $\mu_{\mathrm{Prior}}^{l,k}$ bzw. $\mu_{\mathrm{Post}}^{l,k}$ und die Varianzen $\mathrm{var}_{\mathrm{Prior}}^{l,k}$ bzw. $\mathrm{var}_{\mathrm{Post}}^{l,k}$ bestimmt:

$$\hat{\mu}_{\mathrm{Prior}}^{l,k} := \frac{1}{M}\sum_{m=1}^{M} Y_{\mathrm{Spek}}(l-m,k) \tag{2.8}$$

$$\widehat{\mathrm{var}}_{\mathrm{Prior}}^{l,k} := \frac{1}{M-1}\sum_{m=1}^{M}\left[Y_{\mathrm{Spek}}(l-m,k) - \hat{\mu}_{\mathrm{Prior}}^{l,k}\right]^2 \tag{2.9}$$

bzw.

$$\hat{\mu}_{\mathrm{Post}}^{l,k} := \frac{1}{M+1}\sum_{m=0}^{M} Y_{\mathrm{Spek}}(l-m,k) \tag{2.10}$$

$$\widehat{\mathrm{var}}_{\mathrm{Post}}^{l,k} := \frac{1}{M}\sum_{m=0}^{M}\left[Y_{\mathrm{Spek}}(l-m,k) - \hat{\mu}_{\mathrm{Post}}^{l,k}\right]^2. \tag{2.11}$$

Hierbei ist M, wie zuvor erwähnt, die Anzahl an verwendeten Spektrogrammen (Historie), über welche die Erwartungswerte und Varianzen bestimmt werden.

Abschließend lässt sich die akustische Salienz für den aktuellen Zeitschritt l durch Mittelung über alle Frequenzen des Spektrogramms unter Berücksichtigung der Symmetrie des Spektrogramms bestimmen als

$$S_{\text{akustisch}}(l) = \frac{1}{N/2} \sum_{k=0}^{N/2-1} S_{\text{akustisch}}(l,k)\,. \tag{2.12}$$

Dabei kann der Grad der Überlappung zwischen den einzelnen Frames, wie eingangs beschrieben, durch den Parameter L festgelegt werden. Neben der Mittelung über alle Frequenzen sind auch andere Verfahren denkbar, welche z. B. je nach Anwendung bestimmte Bereiche des Spektrogramms höher gewichten als andere.

Akustische Lokalisation mit SRP-PHAT-β

Nach der Definition der akustischen Salienz ist es notwendig, für die Bestimmung eines Salienzclusters den Ursprung der Schallquelle zu ermitteln. Dies kann mit verschiedenen Verfahren realisiert werden, welche je nach Einsatzzweck und den zur Verfügung stehenden Rechenressourcen (vgl. [Pap11], [Bec06]) ausgewählt werden können. In der Literatur und bei praktischen Anwendungen (z. B. [Swe08b], [Mac10a]) hat sich das sogenannte „*Steered Response Power with Phase Transform*"-Verfahren (kurz: SRP-PHAT; vgl. [DiB01]) bewährt. Für die Lokalisation der Schallquelle werden dabei i. d. R. mehrere Mikrofonpaare eingesetzt und die Tatsache ausgenutzt, dass das Audiosignal der Schallquelle eine unterschiedlich lange Laufzeit zu den verschiedenen Mikrofonen besitzt. Durch die Korrelation der einzelnen Signale der Mikrofonpaare kann auf den Ort der Schallquelle geschlossen werden.

Allgemein wird das empfangene Audiosignal $y(t)$ als Faltung der Raumimpulsantwort $h(t)$ mit dem ursprünglichen Signal $a(t)$ und einem additiven Rauschterm $n(t)$ angenommen:

$$y(t) = h(t) * a(t) + n(t)\,. \tag{2.13}$$

Die Bestimmung des Zeitversatzes zwischen zwei Mikrofonen i und j erfolgt mit Hilfe der Kreuzkorrelation. Hierzu werden zunächst die Kurzzeit-Fourier-Transformierten der Eingangssignale $y_i(t)$ und $y_j(t)$ der beiden Mikrofone bestimmt mit

$$Y_i(t,\omega) = \text{STFT}\{y_i(t)\} \quad \text{und} \quad Y_j(t,\omega) = \text{STFT}\{y_j(t)\}\,. \tag{2.14}$$

Die STFT wird hierbei mit einem gefensterten Signalausschnitt – z. B. mittels Hanning-Fensters mit fester Breite T – berechnet. Im Vergleich zur akustischen Salienz hängt die Breite des Fensters unmittelbar vom maximalen Zeitversatz der Mikrofonsignale ab und weist somit u. U. eine andere Breite auf. In jedem Zeitpunkt t kann nun $Y_i(t,\omega)$ für ein Mikrofon i und die zu untersuchende Frequenz ω bestimmt werden. Die Kreuzkorrelation lässt sich mit dem Signal eines zweiten Mikrofons j somit wie folgt berechnen

$$R_{ij}(t,\tau) = \int_{-\infty}^{\infty} Y_i(t,\omega) Y_j^*(t,\omega) e^{j\omega\tau} d\omega . \tag{2.15}$$

Um die Schätzung zu verbessern, kann im Frequenzbereich eine Vorfilterung eingesetzt werden, welche alle Frequenzkomponenten gleich gewichtet. Dieses auch als *Prewhitening* bezeichnete Verfahren ergänzt die Kreuzkorrelation um einen PHAT-Term (*Phase Transform*) $\psi_{ij}(t,\omega)$. Dadurch erhält man die generalisierte Kreuzkorrelation (kurz: GCC; vgl. [Kna76], [Mac09])

$$R_{ij}^{\mathrm{GCC}}(t,\tau) = \int_{-\infty}^{\infty} \psi_{ij}^{\mathrm{PHAT}}(t,\omega) Y_i(t,\omega) Y_j^*(t,\omega) e^{j\omega\tau} d\omega , \tag{2.16}$$

mit

$$\psi_{ij}^{\mathrm{PHAT}}(t,\omega) = \frac{1}{|Y_i(t,\omega) Y_j^*(t,\omega)|} . \tag{2.17}$$

Neuere Untersuchungen (vgl. vgl. [Don07]) haben gezeigt, dass eine partielle Filterung (engl.: partial whitening) speziell bei verrauschten Umgebungen zu besseren Ergebnissen führt. In [Don07] wurde die partielle Filterung durch Einführung eines zusätzlichen Parameters β realisiert

$$\psi_{ij}^{\text{PHAT-}\beta}(t,\omega) = \frac{1}{(|Y_i(t,\omega) Y_j^*(t,\omega)|)^{\beta}} \quad \text{mit } \beta \in [0,1] \tag{2.18}$$

und erfolgreich bei verrauschten Umgebungen für $\beta \in [0{,}5; 0{,}7]$ verifiziert.

Für den Spezialfall $\beta = 0$ ist $\psi_{ij}^{\text{PHAT-}\beta}(t,\omega) = 1$, was der normalen Kreuzkorrelation nach Gl. 2.15 entspricht, d.h. die Spektren gehen ungefiltert in die Berechnung ein. Ist hingegen $\beta = 1$, so sind Gl. 2.17 und Gl. 2.18 identisch und es liegt eine klassische PHAT-Filterung vor, bei welcher die kompletten Informationen über die Amplitude des Spektrums unberücksichtigt bleiben. Alle anderen Werte für den Parameter β führen zu einer partiellen Filterung.

Die Bestimmung der eigentlichen Position der Schallquelle erfolgt über die Auswertung der kumulierten Zeitversätze über einen Suchraum. Dieser stellt dabei

beispielsweise die Umgebung eines Roboters dar. Für die Berechnung werden die möglichen Positionen im Suchraum quantisiert. Dabei muss abgewogen werden zwischen den benötigten Rechenressourcen und der Genauigkeit, die eine Quantisierung aufweisen soll. Der Suchraum $\mathcal{S}$ kann dabei über eine Menge aus kartesischen Koordinaten $\boldsymbol{x} = (x, y, z) \in \mathcal{S}$ definiert oder alternativ über eine Menge festgelegt werden, bestehend aus sphärischen Koordinaten $\tilde{\boldsymbol{x}} = (r, \varphi, \vartheta) \in \tilde{\mathcal{S}}$.

Beim SRP-PHAT-Verfahren wird ein Ergebnisakkumulator E verwendet, welcher in Abhängigkeit der Positionen $\boldsymbol{x}$ des Suchraumes $\mathcal{S}$ die Kreuzkorrelationswerte bei gegebenen Laufzeitunterschieden für alle Mikrofonpaare $\mathcal{M}$ in jedem Punkt summiert

$$E(t, \boldsymbol{x}) = \frac{1}{|\mathcal{M}|} \sum_{(i,j) \in \mathcal{M}} R_{ij}^{\mathrm{GCC}}(t, \tau_{ij}(\boldsymbol{x})) \,. \tag{2.19}$$

$\tau_{ij}(\boldsymbol{x})$ ist der Laufzeitunterschied zwischen dem Eintreffen eines Signals bei Mikrofon i und j bei gegebenen Mikrofonpositionen sowie dem Ort $\boldsymbol{x}$ einer Schallquelle (engl.: time difference of arrival; kurz: TDOA). $\mathcal{M}$ ist die Menge aller Mikrofonpaare und $|\mathcal{M}|$ die dazugehörige Kardinalität.

Im Folgenden kann entweder nur eine aktive Schallquelle bestimmt werden oder gegebenenfalls eine Vielzahl von gleichzeitig aktiven Quellen. Der erste Fall kann durch eine Maximumsuche im Ergebnisakkumulator realisiert werden

$$\hat{\boldsymbol{x}}(t) = \arg\max_{\boldsymbol{x} \in \boldsymbol{S}} E(t, \boldsymbol{x}) \,. \tag{2.20}$$

Alternativ können auch lokale Maxima bestimmt werden, indem alle Punkte im Ergebnisakkumulator, die eine Schwelle überschreiten, als Kandidaten für eine Positionsbestimmung genutzt werden

$$\hat{\mathcal{X}}(t) = \{\boldsymbol{x} \in \mathcal{S} | E(t, \boldsymbol{x}) > E_{\min}\} \,. \tag{2.21}$$

$E_{\min}$ repräsentiert dabei die minimale Schwelle, die überschritten werden muss, damit die aktuelle Position $\boldsymbol{x}$ eine Schallquelle darstellt und in die Menge der akustisch aktiven Raumpositionen $\hat{\mathcal{X}}(t)$ aufgenommen wird. Die minimale Schwelle kann je nach Anwendung absolut oder in Abhängigkeit des Maximalwertes im Ergebnisakkumulator definiert werden, wobei bei Letzterem sichergestellt werden muss, dass eine Schallquelle aktiv ist. Mit Hilfe von typischen anwendungsspezifischen Audioaufnahmen kann der absolute Schwellwert im Vorfeld explizit bestimmt werden.

Die Bestimmung der akustischen Salienzcluster erfolgt auf Basis der akustischen Salienz und den soeben ermittelten Raumpositionen der Schallquellen. Die akustischen Salienzcluster sind zusammen mit den visuellen Salienzclustern und der multimodalen Fusion in Abschnitt 2.3.3 erläutert.

2.3.2 Visuelle Salienz

Bei der Bestimmung der visuellen Salienz werden gezielt auffällige (d. h. saliente) Regionen in Bildern hervorgehoben und in Form von Salienzkarten repräsentiert. Dazu wird u. a. die diskrete Cosinus-Transformation auf Grauwert-Bildern genutzt (vgl. [Hou12]). Dabei bleiben jedoch die im Ursprungsbild vorhandenen Farbinformationen unberücksichtigt. In der Literatur wird die Quaternion-basierte diskrete Cosinus-Transformation (kurz: QDCT) typischerweise bei der Verarbeitung von Mehrkanalbildern mit vier Kanälen eingesetzt. Ein Beispiel hierfür ist das Template-basierte Farbmatching (vgl. [Fen08]). Durch die Kombination beider Ansätze (vgl. [Sch12a]) können nun Mehrkanalbilder mit z. B. Farbe und Bewegung auf saliente Bereiche hin untersucht werden.

An dieser Stelle sollen die wichtigsten Erkenntnisse aus den zuvor genannten Quellen zusammengefasst dargestellt werden, welche für die Bestimmung der visuellen Salienz notwendig sind.

Quaternion-basierte diskrete Cosinus-Transformation (QDCT)

Durch die Erweiterung der komplexen Zahlen auf vier Dimensionen lässt sich ein Quaternion q (vgl. [Fen08], [Sch12a]) als hyperkomplexe Zahl wie folgt definieren

$$q = a + bi + cj + dk \in \mathbb{H} \qquad \text{mit } a, b, c, d \in \mathbb{R}. \tag{2.22}$$

Dabei besteht dieses aus einem Realteil (mit 1 als Basis) und drei Imaginärteilen (mit i, j und k als Basis). Die Algebra $\mathbb{H}$ definiert die folgenden Regeln

$$i^2 = j^2 = k^2 = ijk = -1 \tag{2.23}$$
$$ij = -ji = k \tag{2.24}$$
$$jk = -kj = i \tag{2.25}$$
$$ki = -ik = j \tag{2.26}$$

und es gilt für $x, y \in \mathbb{H}$

$$x + y = (x_0 + y_0) + (x_1 + y_1)i + (x_2 + y_2)j + (x_3 + y_3)k \tag{2.27}$$

$$\begin{aligned} x \cdot y = {} & (x_0 y_0 - x_1 y_1 - x_2 y_2 - x_3 y_3) \\ & + (x_0 y_1 + x_1 y_0 + x_2 y_3 - x_3 y_2)i \\ & + (x_0 y_2 - x_1 y_3 + x_2 y_0 + x_3 y_1)j \\ & + (x_0 y_3 + x_1 y_2 - x_2 y_1 + x_3 y_0)k \end{aligned} \tag{2.28}$$

$$\overline{x} = \overline{x_0 + x_1 i + x_2 j + x_3 k} = x_0 - x_1 i - x_2 j - x_3 k \tag{2.29}$$

$$x \cdot \overline{x} = x_0^2 + x_1^2 + x_2^2 + x_3^2 \tag{2.30}$$

$$|x| = \sqrt{x_0^2 + x_1^2 + x_2^2 + x_3^2}\,. \tag{2.31}$$

Die links-/rechtsseitige 2D-Quaternion DCT lässt sich nach Feng (vgl. [Fen08]) wie folgt bestimmen

$$\mathrm{QDCT}_L(u,v) = \alpha_u^M \cdot \alpha_v^N \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} u_q \cdot I_q(m,n) \cdot \beta_{u,m}^M \cdot \beta_{v,n}^N \tag{2.32}$$

$$\mathrm{QDCT}_R(u,v) = \alpha_u^M \cdot \alpha_v^N \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_q(m,n) \cdot \beta_{u,m}^M \cdot \beta_{v,n}^N \cdot u_q\,, \tag{2.33}$$

wobei $\boldsymbol{I}_q \in \mathbb{H}^{M\times N}$ eine $M \times N$ Quaternionmatrix ist, u_q ein Einheitsquaternion ist (d.h. $u_q^2 = -1$) und α_u^M, α_v^N sowie $\beta_{u,m}^M, \beta_{v,n}^N$ wie folgt definiert sind

$$\alpha_u^M = \begin{cases} \sqrt{1/M}, & \text{für } u = 0 \\ \sqrt{2/M}, & \text{für } u \neq 0 \end{cases} \quad \text{bzw.} \quad \alpha_v^N = \begin{cases} \sqrt{1/N}, & \text{für } v = 0 \\ \sqrt{2/N}, & \text{für } v \neq 0 \end{cases} \tag{2.34}$$

und

$$\beta_{u,m}^M = \cos\left(\frac{\pi(2m+1)u}{2M}\right) \quad \text{bzw.} \quad \beta_{v,n}^N = \cos\left(\frac{\pi(2n+1)v}{2N}\right). \tag{2.35}$$

Die inverse Quaternion-basierte diskrete Cosinus-Transformation (kurz: IQDCT) wird analog zur QDCT links- und rechtsseitig definiert als

$$\mathrm{IQDCT}_L(m,n) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha_u^M \cdot \alpha_v^N \cdot u_q \cdot C_q(u,v) \cdot \beta_{u,m}^M \cdot \beta_{v,n}^N \quad \text{und} \tag{2.36}$$

$$\mathrm{IQDCT}_R(m,n) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha_u^M \cdot \alpha_v^N \cdot C_q(u,v) \cdot \beta_{u,m}^M \cdot \beta_{v,n}^N \cdot u_q\,, \tag{2.37}$$

wobei u_q, α_u^M, α_v^N sowie $\beta_{u,m}^M, \beta_{v,n}^N$ entsprechend der QDCT definiert sind. $\boldsymbol{C}_q \in \mathbb{H}^{M\times N}$ ist die Quaternionmatrix für die Rücktransformation.

Bestimmung der Salienzkarte

In [Hou12] konnte gezeigt werden, dass durch die Kombination von diskreter Cosinus-Transformation und Signumfunktion (kurz: sgn) in einem Bild der Hintergrund größtenteils unterdrückt und Merkmale, welche saliente Bereiche und

Objekte darstellen, hervorgehoben werden können. Für ein Mehrkanal-Farbbild lässt sich somit die visuelle Salienzkarte wie folgt bestimmen (vgl. [Hou12]):

$$\boldsymbol{S}_{\text{visuell}}^{\text{DCT}}(\boldsymbol{I}) = g * \sum_{c} \left[T(\boldsymbol{I}_c) \circ \overline{T(\boldsymbol{I}_c)} \right] \tag{2.38}$$

mit

$$T(\boldsymbol{I}_c) = \text{IDCT}\left\{ \text{sgn}\left(\text{DCT}\{\boldsymbol{I}_c\}\right) \right\}, \tag{2.39}$$

wobei $\boldsymbol{I}_c$ der c-te Farbkanal des Bildes $\boldsymbol{I}$ ist, g die Impulsantwort eines Gauß-Filters darstellt und $\circ$ das Hadamard-Produkt[2] repräsentiert. Diese Erkenntnisse lassen sich auf eine QDCT-basierte Salienzkarte unter Zuhilfenahme der Signumfunktion und der DCT für Quaternionen übertragen (vgl. [Sch12a]). In diesem Zusammenhang kann die Signumfunktion für Quaternionen als „Ausrichtung" eines Quaternion $x \in \mathbb{H}$ angesehen werden und ist wie folgt definiert (vgl. [Sch12a])

$$\text{sgn}(x) = \begin{cases} \frac{x_0}{|x|} + \frac{x_1}{|x|}\,i + \frac{x_2}{|x|}\,j + \frac{x_3}{|x|}\,k & \text{für } |x| \neq 0 \\ 0 & \text{für } |x| = 0 \end{cases}. \tag{2.40}$$

Die visuelle Salienzkarte $\boldsymbol{S}_{\text{visuell}}(t)$ zum Zeitpunkt t kann analog zu Gl. 2.38 wie folgt bestimmt werden (vgl. [Sch12a], [Hou12]):

$$\boldsymbol{S}_{\text{visuell}}(t) = g * \left[T(\boldsymbol{I}_q(t)) \circ \overline{T(\boldsymbol{I}_q(t))} \right] \tag{2.41}$$

mit

$$T(\boldsymbol{I}_q(t)) = \text{IQDCT}_L\left\{ \text{sgn}\left(\text{QDCT}_L\left\{\boldsymbol{I}_q(t)\right\}\right) \right\} \tag{2.42}$$

und

$$\boldsymbol{I}_q(t) = \boldsymbol{I}_{\text{I}}(t) + \boldsymbol{I}_{\text{RG}}(t)\,i + \boldsymbol{I}_{\text{BY}}(t)\,j + \boldsymbol{I}_{\text{M}}(t)\,k, \tag{2.43}$$

wobei $\boldsymbol{I}_q(t)$ ein vierdimensionales Bild zum Zeitpunkt t ist. Dieses besteht aus den vier Komponenten: Intensität ($\boldsymbol{I}_I$), rot-grünen bzw. blau-gelben Farbopponenten[3] ($\boldsymbol{I}_{\text{RG}}$ bzw. $\boldsymbol{I}_{\text{BY}}$) und Bewegung ($\boldsymbol{I}_{\text{M}}$). Im Ergebnisbild werden durch die Signumfunktion saliente Regionen durch hohe Werte repräsentiert und weniger saliente mit niedrigeren Werten.

[2] elementweises Produkt
[3] Gegenfarbpaare, vgl. [Hur57]

Segmentierung der Salienzkarte mit Hilfe der Isophotenkrümmung

Die Segmentierung der zuvor gewonnenen Salienzkarte $\boldsymbol{S}_{\text{visuell}}(t)$ geschieht anhand sogenannter Isophoten (vgl. [Lic05]). Diese repräsentieren im Allgemeinen Linien mit gleichen Intensitäten, vergleichbar mit den Höhenlinien in einer Landkarte. Übertragen auf Salienzkarten sind Isophoten geschlossene Linien mit gleicher Salienz, welche saliente Regionen umschließen. Unter der Annahme, dass die salienten Regionen annähernd kreisförmig sind und einen Peak im Inneren besitzen, lässt sich das Zentrum als Ort bestimmen, zu dem die Gradienten der Pixel der umliegenden Region zeigen. Auf diese Art und Weise können effizient die Zentren bestimmt werden, auch wenn die Regionen in ihren Salienzwerten und/oder ihrer Größe stark variieren oder sich teilweise überlappende Peaks besitzen.

Zur Bestimmung der salienten Peaks wird nun die lokale Isophotenkrümmung $\kappa(x, y)$ in allen Punkten der Salienzkarte $\boldsymbol{S}_{\text{visuell}}$ betrachtet (vgl. [Val08])

$$\kappa(x, y) = -\frac{S_{cc}(x, y)}{S_g(x, y)}, \tag{2.44}$$

wobei $\boldsymbol{S}_{cc}$ die zweite Ableitung der Salienzkarte in allen Punkten in Richtung der Senkrechten zur Gradientenrichtung ist und $\boldsymbol{S}_g$ die erste Ableitung in Gradientenrichtung ist. Dies lässt sich auch direkt durch Ableitungen in kartesischen Basiskoordinaten (x, y) darstellen als

$$\kappa(x, y) = -\frac{S_{cc}}{S_g} = -\frac{S_y^2 S_{xx} - 2S_x S_y S_{xy} + S_x^2 S_{yy}}{(S_x^2 + S_y^2)^{3/2}}, \tag{2.45}$$

und in Matrixschreibweise mit elementweiser Multiplikation ($\circ$) und Division (/)

$$\boldsymbol{\kappa} = -\boldsymbol{S}_{cc} / \boldsymbol{S}_g = -\left(\boldsymbol{S}_y^2 \circ \boldsymbol{S}_{xx} - 2\boldsymbol{S}_x \circ \boldsymbol{S}_y \circ \boldsymbol{S}_{xy} + \boldsymbol{S}_x^2 \circ \boldsymbol{S}_{yy}\right) / \left(\boldsymbol{S}_x^2 + \boldsymbol{S}_y^2\right)^{3/2}, \tag{2.46}$$

wobei $\boldsymbol{S}_x$ und $\boldsymbol{S}_y$ die ersten Ableitungen der Salienzkarte $\boldsymbol{S}_{\text{visuell}}$ in allen Punkten in x- bzw. y-Richtung, $\boldsymbol{S}_{xx}$ und $\boldsymbol{S}_{yy}$ die zweiten Ableitungen in die jeweilige Richtung sind. Bei $\boldsymbol{S}_{xy}$ wurde die Salienzkarte in allen Punkten zuerst in x- und dann in y-Richtung abgeleitet.

Durch Ausnutzung der Tatsache, dass die lokale Krümmung $\kappa(x, y)$ reziprok zum Radius r des Kreises in jedem Pixel ist, d.h. $r(x, y) = 1/\kappa(x, y)$, kann durch Kombination mit der lokalen Ausrichtung grob die Position des Zentrums jedes Peaks abgeschätzt werden (vgl. Abb. 2.3; links). Dazu werden die Verschiebungsvektoren $\boldsymbol{d}_x$, $\boldsymbol{d}_y$ mit

$$d_x(x, y) = S_x(x, y) \cdot S_d(x, y) \quad \text{bzw.} \quad d_y(x, y) = S_y(x, y) \cdot S_d(x, y) \tag{2.47}$$

und

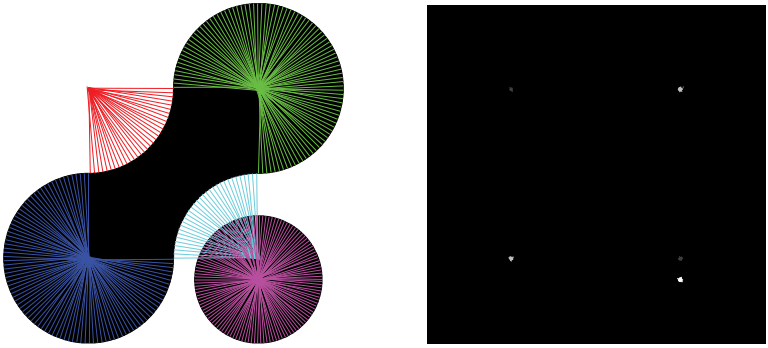

Abb. 2.3: Die Verschiebungsvektoren, die auf das Zentrum eines Peaks zeigen (links) und die Akkumulatorkarte, welche die Position und Ausprägung der Peaks anzeigt (rechts) für ein ideales Beispiel.

$$\boldsymbol{S}_d = -\left(\boldsymbol{S}_x^2 + \boldsymbol{S}_y^2\right) / \left(\boldsymbol{S}_y^2 \circ \boldsymbol{S}_{xx} - 2\boldsymbol{S}_x \circ \boldsymbol{S}_y \circ \boldsymbol{S}_{xy} + \boldsymbol{S}_x^2 \circ \boldsymbol{S}_{yy}\right) \tag{2.48}$$

berechnet (vgl. [Val08]), um das Zentrum des Peaks $\boldsymbol{u}(x,y) = (u_x(x,y), u_y(x,y))$ für das aktuelle Pixel (x,y) zu ermitteln mit

$$u_x(x,y) = p_x(x,y) + d_x(x,y) \quad \text{und} \quad u_y(x,y) = p_y(x,y) + d_y(x,y)\,. \tag{2.49}$$

Hierbei stellt $\boldsymbol{p}(x,y) = (p_x(x,y), p_y(x,y)) = (x,y)$ die Koordinaten der Pixel in x- bzw. y-Richtung dar, d.h. den Index in Ordinaten- und Abszissenrichtung. In einer Akkumulatorkarte (vgl. Abb. 2.3; rechts), welche dieselbe Größe hat wie die ursprüngliche Salienzkarte, wird für jeden Punkt der Salienzkarte das dazugehörige Zentrum des Peaks bestimmt und der Wert an dieser Stelle um Eins inkrementiert (Voting). Alternativ kann auch der Salienzwert des aktuellen Pixels beim Voting für ein Zentrum genutzt werden, sodass vom Zentrum weiter entfernte Punkte in der Regel ein geringeres Gewicht haben. Zusammengefasst lässt sich sagen, dass saliente Regionen in der Ursprungskarte auf Punkte mit hohen Werten in der Akkumulatorkarte abgebildet werden, welche die Basis für die visuelle Salienz bilden.

Um das Ergebnis zu verbessern, können verschiedene Verfahren eingesetzt werden. Beispielsweise können nur Punkte mit einem maximalen Abstand bei dem Voting für ein Zentrum berücksichtigt werden und so Ausreißer unterdrückt werden. Weitere Ausreißer, welche z. B. durch Rauschen entstanden sind, können mit Hilfe von *Convex peeling* (vgl. [Hod04]) entfernt werden. Des Weiteren können Regionen, die prozentual zu viele Ausreißer beherbergen, aus der Betrachtung entfernt werden.

Zur Bestimmung der visuellen Salienzcluster wird nun die bereinigte Akkumulatorkarte untersucht. Dabei wird ein Verfahren zur Vermeidung einer erneuten Untersuchung einer Region (engl.: location-based inhibition of return;

vgl. [Cho08]) angewendet, welches auch bei Salienzkarten (siehe bspw. [Sch10], [Rue08], [Itt98]) eingesetzt werden kann. Hierbei wird in der Akkumulatorkarte der Punkt ausgewählt, der die Region mit der höchsten Salienz repräsentiert. Die dazugehörigen Pixel in der Salienzkarte können dabei ebenfalls bestimmt werden. Im nächsten Schritt werden diese in der Akkumulatorkarte zurückgesetzt, um diese bei der Wahl der nächsten Salienzcluster nicht mehr zu berücksichtigen (Inhibition). Dieser Vorgang wird solange wiederholt, bis keine weiteren ausschlaggebenden Peaks in der Akkumulatorkarte mehr vorhanden sind.

2.3.3 Salienzcluster

Salienzcluster dienen der Modellierung von wichtigen Regionen und Ereignissen, die im aktuellen Sensorbereich wahrgenommen werden können. Nachdem in den vorherigen Abschnitten die Grundlagen für die akustische und visuelle Salienz gelegt wurden, erfolgt nun die Bestimmung der jeweiligen *unimodalen Salienzcluster*. Diese repräsentieren, wie eingangs des Kapitels erwähnt, die akustisch relevanten Ereignisse bzw. die visuell hervorstechenden Bildregionen. Anschließend erfolgt eine Fusion zu *multimodalen Salienzclustern*. Diese können beispielsweise für die Festlegung des aktuellen Aufmerksamkeitsfokus eines autonomen Systems oder auch während der Analyse einer Szene für die Priorität einzelner Bereiche genutzt werden.

Unimodale Salienzcluster

Sowohl visuell saliente Regionen als auch akustisch saliente Ereignisse werden im Folgenden mit einem einheitlichen Konzept, dem Salienzcluster, repräsentiert. Die Menge aller unimodalen Salienzcluster $\mathcal{C}$ besteht im Rahmen der vorliegenden Arbeit aus einer visuellen $\mathcal{C}^{\text{visuell}}$ sowie einer akustischen $\mathcal{C}^{\text{akustisch}}$ Teilmenge und lässt sich wie folgt definieren

$$\mathcal{C} := \mathcal{C}^{\text{akustisch}} \cup \mathcal{C}^{\text{visuell}} = \{c_1, \ldots, c_N\} \quad \text{mit } N \in \mathbb{N}, \tag{2.50}$$

wobei jeder unimodale Cluster c_i aus einem 3-Tupel mit Salienzwert s_{c_i}, mittlerer räumlicher Position $\boldsymbol{\mu}_{c_i}$ sowie einer räumlichen Positionsunsicherheit $\boldsymbol{\Sigma}_{c_i}$ dargestellt wird und zusammengefasst in Gl. 2.51 zu sehen ist:

$$c_i = (s_{c_i}, \boldsymbol{\mu}_{c_i}, \boldsymbol{\Sigma}_{c_i}) \in \mathcal{C}\,. \tag{2.51}$$

Hiermit lässt sich jeder unimodale Salienzcluster mit Hilfe einer gewichteten Gauß-Notation beschreiben als

$$f_{c_i}^{\text{G}}(\boldsymbol{x}) = \frac{s_{c_i}}{\sqrt{(2\pi)^3 \det(\boldsymbol{\Sigma}_{c_i})}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{c_i})^T \boldsymbol{\Sigma}_{c_i}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_{c_i})}\,. \tag{2.52}$$

Erzeugung der akustischen Salienzcluster

Die akustischen Salienzcluster werden durch räumliches Clustern (Mean-Shift-Verfahren; vgl. [Com02]) der Geräuschpositionen $\hat{\mathcal{X}}(t)$ (siehe Gl. 2.21) in Kombination mit den zuvor berechneten Salienzwerten $S_{\text{akustisch}}(t)$ ermittelt. Jedes räumliche Cluster $\hat{\mathcal{X}}_{c_j} \subseteq \hat{\mathcal{X}}(t)$ bildet dabei die Grundlage für die Bestimmung eines akustischen Salienzclusters

$$c_j^{\text{akustisch}} := \left(s_{c_j}^{\text{akustisch}}, \boldsymbol{\mu}_{c_j}^{\text{akustisch}}, \boldsymbol{\Sigma}_{c_j}^{\text{akustisch}}\right) \in \mathcal{C}^{\text{akustisch}} \tag{2.53}$$

mit

$$s_{c_j}^{\text{akustisch}} = S_{\text{akustisch}}(t), \quad \hat{\boldsymbol{\mu}}_{c_j}^{\text{akustisch}} = \mathbb{E}[\hat{\mathcal{X}}_{c_j}] \quad \text{und} \quad \hat{\boldsymbol{\Sigma}}_{c_j}^{\text{akustisch}} = \text{Cov}(\hat{\mathcal{X}}_{c_j}).$$

Die Clusterbildung bietet den Vorteil, dass rauschbehaftete Positionen bei einer gewissen Anzahl an Punkten durch die Mittelung verbessert werden können. Dennoch wird mit Hilfe der Kovarianzmatrix der räumlichen Verteilung einzelner Teilpositionen Rechnung getragen.

Erzeugung der visuellen Salienzcluster

Die visuellen Salienzcluster können aus den Ergebnissen der Isophoten-basierten Segmentierung (vgl. Abschnitt 2.3.2) direkt bestimmt werden. Dabei formieren die einzelnen Regionen in der Salienzkarte, welche nach und nach ausgewertet werden die räumlichen Cluster c_k^{2D} zur Bildung der visuellen Salienzcluster. Ein solcher Cluster lässt sich zunächst in 2D definieren als

$$c_k^{\text{2D}} := (s_{c_k}^{\text{visuell}}, \boldsymbol{\mu}_{c_k}^{\text{2D}}, \boldsymbol{\Sigma}_{c_k}^{\text{2D}}). \tag{2.54}$$

Hierbei ist $s_{c_k}^{\text{visuell}}$ der Salienzwert des Peaks der aktuellen Region, $\boldsymbol{\mu}_{c_k}^{\text{2D}}$ die Position des Peaks selbst und $\boldsymbol{\Sigma}_{c_k}^{\text{2D}}$ die Kovarianzmatrix, welche die räumliche Ausdehnung beschreibt. Mit Hilfe von Tiefeninformationen (beispielsweise von einer Stereokamera oder einem RGB-D-Sensor) werden nun die 2D-Informationen in eine 3D-Repräsentation überführt. Hierbei kann aufgrund von verdeckten Ansichten die 3D-Form des Clusters nicht einwandfrei bestimmt werden, was sich auf die Kovarianzmatrix auswirkt. Ein visueller Salienzcluster c_k^{visuell} wird durch die Transformation aller 2D-Punkte einer Region $\mathcal{X}_{c_k}^{\text{2D}}$ nach 3D mit anschließender Mittelwert- und Kovarianzmatrixbestimmung ermittelt. Dabei wird das Lochkameramodell (vgl. [Har03]) verwendet, um die Transformation der 2D-Informationen durchzuführen ($f_{\text{3D}}\{\cdot\}$). Der zuvor bestimmte Salienzwert $s_{c_k}^{\text{visuell}}$ wird hingegen direkt übernommen. Ein visueller Salienzcluster in 3D ist somit definiert als

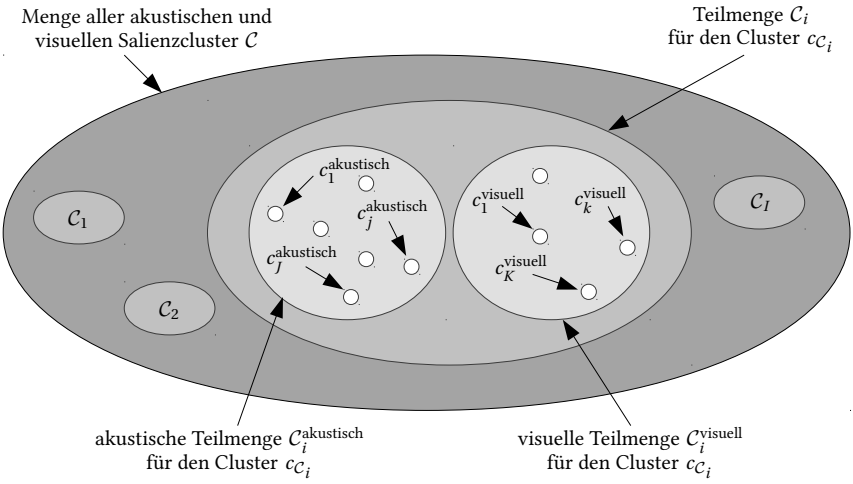

Abb. 2.4: Die Zusammenhänge der einzelnen Salienzcluster und der Mengen von verschiedenen Salienzclustern.

$$c_k^{\text{visuell}} := \left(s_{c_k}^{\text{visuell}}, \boldsymbol{\mu}_{c_k}^{\text{visuell}}, \boldsymbol{\Sigma}_{c_k}^{\text{visuell}}\right) \in \mathcal{C}^{\text{visuell}}. \tag{2.55}$$

mit

$$\hat{\boldsymbol{\mu}}_{c_k}^{\text{visuell}} = \mathbb{E}\left[f_{3\text{D}}\left\{\mathcal{X}_{c_k}^{2\text{D}}\right\}\right] \quad \text{und} \quad \hat{\boldsymbol{\Sigma}}_{c_k}^{\text{visuell}} = \text{Cov}\left(f_{3\text{D}}\left\{\mathcal{X}_{c_k}^{2\text{D}}\right\}\right). \tag{2.56}$$

Fusion von unimodalen Salienzclustern

Die akustischen und visuellen Salienzcluster werden durch räumliche und zeitliche Gruppierung zu multimodalen Salienzclustern fusioniert. Die räumliche Gruppierung geschieht mit Hilfe des Mean-Shift-Verfahrens (vgl. [Com02]) und über mehrere Zeitschritte hinweg. Dies hat den Vorteil, dass Rauscheinflüsse reduziert werden und mehr Informationen für die Fusion zur Verfügung stehen.

Mit dem Mean-Shift-Verfahren wird die Menge an akustischen und visuellen Salienzclustern in disjunkte Teilmengen unterteilt (vgl. Abb. 2.4). Eine Teilmenge an räumlich zusammengehörigen Clustern $\mathcal{C}_i$ ist definiert als

$$\mathcal{C}_i = \left\{c_1^{\text{akustisch}}, \ldots, c_J^{\text{akustisch}}; c_1^{\text{visuell}}, \ldots, c_K^{\text{visuell}}\right\} \subseteq \mathcal{C}, \quad J, K \in \mathbb{N} \tag{2.57}$$

mit

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \ldots \cup \mathcal{C}_i \cup \ldots \cup \mathcal{C}_I, \quad I \in \mathbb{N}. \tag{2.58}$$

Des Weiteren kann die Teilmenge $\mathcal{C}_i$ in einen akustischen $\mathcal{C}_i^{\text{akustisch}}$ und einen visuellen Anteil $\mathcal{C}_i^{\text{visuell}}$ unterteilt werden (vgl. Abb. 2.4). Dies lässt sich formal beschreiben als

$$\mathcal{C}_i^{\text{akustisch}} = \mathcal{C}_i \cap \mathcal{C}^{\text{akustisch}} \quad \text{bzw.} \quad \mathcal{C}_i^{\text{visuell}} = \mathcal{C}_i \cap \mathcal{C}^{\text{visuell}}. \tag{2.59}$$

Bei der Fusion der akustischen und visuellen Salienzcluster der Teilmenge $\mathcal{C}_i$ kann nun die Zuverlässigkeit einzelner Sensoren berücksichtigt und die Cluster einer Modalität unterschiedlich stark gewichtet werden. Die Gewichte können beispielsweise für eine konkrete Sensorkombination und mit Hilfe einer Vielzahl an Testobjekten bestimmt werden. Im Falle von idealen Sensoren oder falls die Gewichtung aufgrund fehlender Sensorinformationen nicht objektiv bestimmbar ist, bieten sich gleiche Gewichte für die Linearkombination bei der Fusion der Salienzen an. Diese können gegebenenfalls zur Laufzeit dynamisch angepasst werden. Alternative Ansätze, welche den Zeitpunkt und die Art der Fusion anders wählen (vgl. [Ona07]), lassen sich anstelle der hier vorgestellten Vorgehensweise ebenfalls nutzen.

Die Fusion der akustischen und visuellen Salienzen zu einer multimodalen Salienz für einen Cluster $c_{\mathcal{C}_i}$ lässt sich somit realisieren als

$$s_{\mathcal{C}_i} = \frac{1}{2} \sum_{c_j \in \mathcal{C}_i^{\text{akustisch}}} w_{\mathcal{C}_i,c_j}^{\text{akustisch}} \cdot f_{c_j}^G(\boldsymbol{\mu}_{\mathcal{C}_i}) + \frac{1}{2} \sum_{c_k \in \mathcal{C}_i^{\text{visuell}}} w_{\mathcal{C}_i,c_k}^{\text{visuell}} \cdot f_{c_k}^G(\boldsymbol{\mu}_{\mathcal{C}_i}), \tag{2.60}$$

wobei die Salienz jedes akustischen bzw. visuellen Salienzclusters durch eine gewichtete Gauß'sche Dichtefunktion (siehe Gl. 2.52) im Clusterzentrum repräsentiert wird. Die Gewichte der einzelnen Modalitäten werden bestimmt durch

$$w_{\mathcal{C}_i,c_j}^{\text{akustisch}} = \frac{s_{c_j}}{\sum_{c_m \in \mathcal{C}_i^{\text{akustisch}}} s_{c_m}} \quad \text{und} \quad w_{\mathcal{C}_i,c_k}^{\text{visuell}} = \frac{s_{c_k}}{\sum_{c_n \in \mathcal{C}_i^{\text{visuell}}} s_{c_n}}. \tag{2.61}$$

Zur Bestimmung der Position wird der gewichtete räumliche Mittelwert aller sich im Cluster $c_{\mathcal{C}_i}$ befindenden unimodalen Salienzcluster verwendet

$$\boldsymbol{\mu}_{\mathcal{C}_i} = \mathbb{E}[\mathcal{C}_i] = \sum_{c_l \in \mathcal{C}_i} \boldsymbol{\mu}_{c_l} w_{\mathcal{C}_i,c_l} \qquad \text{mit} \quad w_{\mathcal{C}_i,c_l} = \frac{s_{c_l}}{\sum_{c_o \in \mathcal{C}_i} s_{c_o}}. \tag{2.62}$$

Anschließend lässt sich die räumliche Positionsunsicherheit $\boldsymbol{\Sigma}_{\mathcal{C}_i}$ durch iterative Fusion der Kovarianzmatrizen (vgl. [Smi86]) der einzelnen unimodalen Salienzcluster c_l des Clusters $c_{\mathcal{C}_i}$ bestimmen mit

$$\boldsymbol{V}_l = \boldsymbol{V}_{l-1} - \boldsymbol{V}_{l-1}\left(\boldsymbol{V}_{l-1} + \boldsymbol{\Sigma}_{c_l}\right)^{-1} \boldsymbol{V}_{l-1}, \quad \forall\, l = 2, \ldots, L_{\mathcal{C}_i}, \tag{2.63}$$

wobei die Startbedingung mit $\boldsymbol{V}_1 = \boldsymbol{\Sigma}_{c_1}$ definiert ist und die gesuchte Kovarianzmatrix $\boldsymbol{\Sigma}_{\mathcal{C}_i} = \boldsymbol{V}_{L_{\mathcal{C}_i}}$ ist. Mit $L_{\mathcal{C}_i} = |\mathcal{C}_i|$ wird die Anzahl an unimodalen Clustern im aktuellen Cluster $\mathcal{C}_i$ bezeichnet.

Der räumlich-zeitlich fusionierte multimodale Salienzcluster $c_{\mathcal{C}_i}$ kann abschließend beschrieben werden durch das 3-Tupel, bestehend aus der Salienz $s_{\mathcal{C}_i}$, der Position $\boldsymbol{\mu}_{\mathcal{C}_i}$ und der Kovarianzmatrix $\boldsymbol{\Sigma}_{\mathcal{C}_i}$:

$$c_{\mathcal{C}_i} = \left(s_{\mathcal{C}_i}, \boldsymbol{\mu}_{\mathcal{C}_i}, \boldsymbol{\Sigma}_{\mathcal{C}_i}\right). \tag{2.64}$$

Um die Zuordnung von salienten Regionen und Ereignissen zu multimodalen Salienzclustern auch über die Zeit hinweg eindeutig gewährleisten zu können, ist es notwendig, den unimodalen Salienzclustern auch bereits vorhandene multimodale Salienzcluster aus einem früheren Zeitschritt zuzuordnen und nicht nur neue multimodale Salienzcluster zu erzeugen. Dies geschieht über einen kombinierten Assoziationsfilter- und Trackingansatz (beispielsweise Partikelfilter; vgl. [Ris04]), welcher die Zuordnung vornimmt und sich bewegende multimodale Salienzcluster nachverfolgt. Die Fusion von multimodalen Salienzclustern mit zugeordneten unimodalen Salienzclustern wird dabei analog zur Bestimmung der multimodalen Salienzcluster realisiert.

2.4 Ergebnisse

An dieser Stelle sollen Ergebnisse der visuellen und akustischen Salienz kurz zusammengefasst werden, welche im Rahmen der Explorationsstrategien im späteren Verlauf der Arbeit verwendet werden.

2.4.1 Visuelle Salienz

In Abb. 2.5 sind zwei Szenarien (Büro und Frühstück) mit einer Reihe von verschiedenen Gegenständen zu sehen. Die dazu korrespondierenden Salienzkarten sind jeweils darunter zu sehen und zeigen die visuelle Salienz. In den Salienzkarten bedeuten rötliche Farbtöne eine hohe Salienz und bläuliche eine niedrige Salienz. Wie zu sehen ist, besteht eine hohe Korrelation zwischen Objekten in einer Szene und Bereichen in den Salienzkarten mit erhöhten Werten. Eine ausführliche Evaluation des in der vorliegenden Arbeit zusammengefassten QDCT-Ansatzes zur Bestimmung der Salienzkarte ist in den Veröffentlichungen von Schauerte (vgl. [Sch12a], [Sch12b]) enthalten.

2.4.2 Akustische Salienz

Im Gegensatz zur visuellen Salienz ist die akustische Salienz nicht nur auf den aktuellen Kameraausschnitt begrenzt, vielmehr können durch den Einsatz

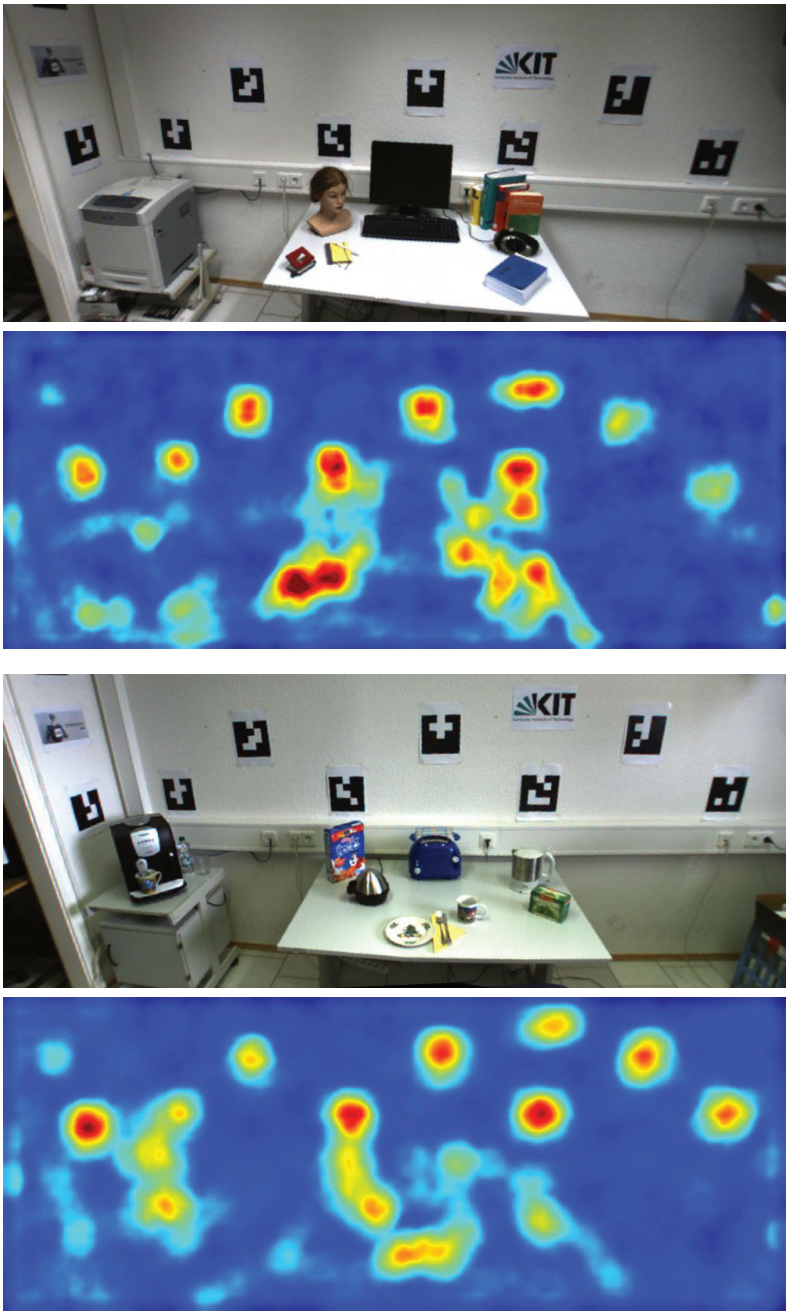

Abb. 2.5: Bild einer Büroszene und die dazugehörige visuelle Salienz in Form einer Salienzkarte (beide oberen Bilder) und ein Frühstücksszenario mit korrespondierender Salienzkarte (beide unteren Bilder).

von omni-direktionalen Mikrofonen meist die Geräuschquellen eines gesamten Raumes wahrgenommen werden.

Für die nachfolgenden Beispiele wurde eine Abtastrate von 48 kHz gewählt. Die Länge eines Fensters N beträgt 512 Samples bei einer Überlappung von 256 Samples. Es wurde für die Fensterung der Frames ein Hamming-Fenster gewählt. Die Anzahl an Spektrogrammen M zur Bestimmung des aktuellen akustischen Salienzwertes wurde auf 10 festgelegt.

Das Beispiel in Abb. 2.6 zeigt ein circa 7 s langes Audiosignal einer sprechenden Person (a), die zu zwei Zeitpunkten (in Magenta markiert) je einen Gegenstand auf einen Tisch fallen lässt. Vor der Analyse wird zunächst dem Audiosignal weißes Rauschen mit niedriger Amplitude addiert (b), d. h. SNR = 3 dB. Dabei ist das ursprüngliche Sprachsignal nicht mehr deutlich zu sehen, akustisch jedoch noch wahrnehmbar. Das dazugehörige Spektrogramm (c) zeigt sowohl die Sprache als auch die Geräusche, wobei Letzteres deutlicher zu sehen ist. Die akustische Salienz (d) zeigt deutliche Ausschläge zu den Zeitpunkten, an denen die Gegenstände fallen gelassen wurden, und geringere Ausschläge an Stellen, an denen die Person (nach kurzen Pausen) zu sprechen beginnt. Zu beachten ist die logarithmische Skalierung der Salienz zur besseren Visualisierung.

In Abb. 2.7 ist dasselbe Audiosignal zu sehen (a), allerdings diesmal mit einer deutlich höheren Rauschamplitude (b), d. h. SNR = −12 dB. Das ursprüngliche Signal ist dabei nicht mehr zu sehen. Im Spektrogramm (c) ist das Rauschen deutlicher zu sehen, und sowohl die Sprache als auch die Geräusche sind nur noch ansatzweise zu erkennen. Die darunter aufgetragene Salienz (d) zeigt jedoch noch deutliche Ausschläge an den Stellen, an denen die Objekte auf die Tischoberfläche aufgetroffen sind und Geräusche erzeugt haben. Die akustische Salienz lässt sich somit auch bei starken Rauscheinflüssen noch sicher bestimmen.

2.5 Schlussbetrachtungen

In diesem Kapitel wurden die akustische und visuelle Salienz modelliert, welche in Form von akustischen bzw. visuellen Salienzclustern auf einer abstrakten Ebene repräsentiert werden. Jeder unimodale Salienzcluster stellt dabei eine visuell prägnante Region oder ein akustisch auffälliges Geräusch dar. Durch die sukzessive Fusion von unimodalen Salienzclustern werden anschließend multimodale Salienzcluster generiert.

Als Grundlage für die akustische Salienz dient die relative Entropie, welche mittels der Kullback-Leibler-Divergenz auf Basis von Spektrogrammen bestimmt wird. Durch die akustische Lokalisation mittels SRP-PHAT-β können die Posi-

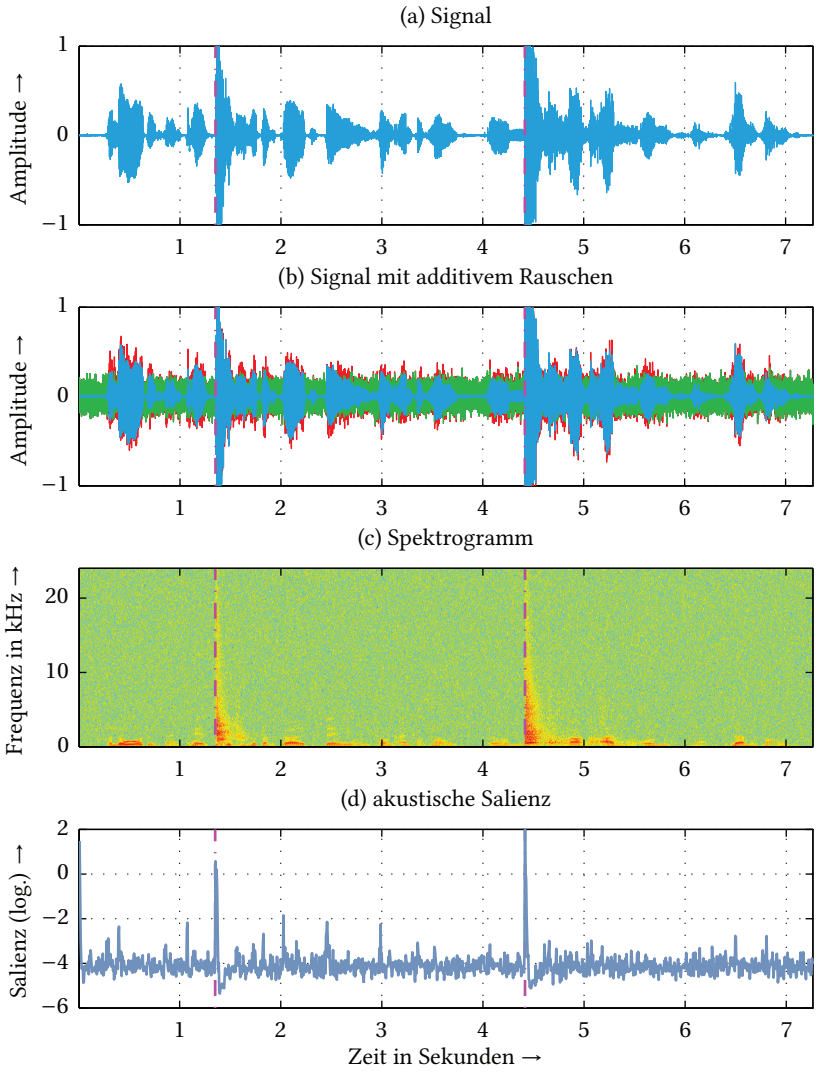

Abb. 2.6: Eine Aufnahme, in der eine Person spricht und Gegenstände an den farbig markierten Stellen auf einen Tisch fallen lässt (a). Das Signal mit additivem Rauschen (b; in Rot), das Rauschen selbst (b; in Grün) und das ursprüngliche Signal (b; in Blau) sind zum Vergleich dargestellt (SNR = 3 dB). Das Spektrogramm (c) und die akustische Salienz (d) für das rauschbehaftete Signal sind darunter abgebildet.

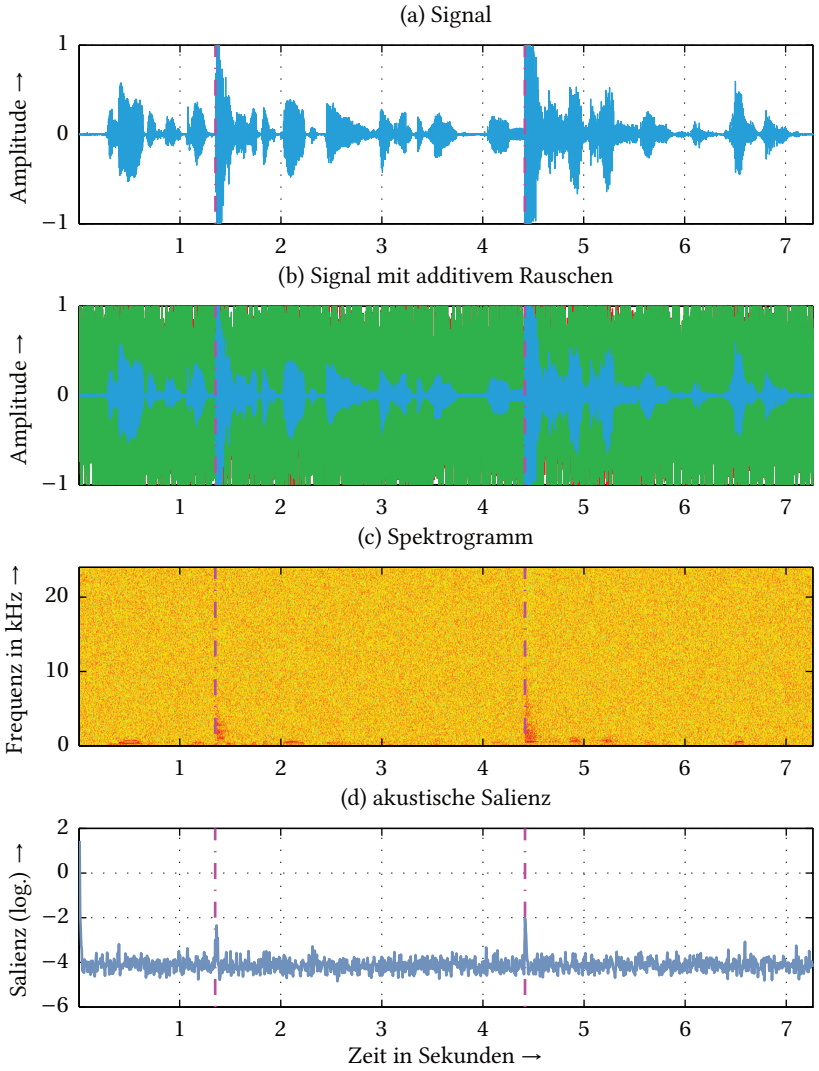

Abb. 2.7: Dasselbe Eingangssignal wie in Abb. 2.6 (a), jedoch mit einem wesentlich stärkeren additiven Rauschen (b), d. h. SNR = −12 dB. Im Spektrogramm darunter sind die Sprache und die Objektgeräusche nicht mehr so deutlich zu sehen (c). Die akustische Salienz (d) zeigt dennoch deutlich Ausschläge an den gekennzeichneten Stellen (in Magenta), welche durch Geräusche der Gegenstände verursacht wurden.

tionen der Schallquellen bestimmt werden und so akustische Salienzcluster generiert werden.

Die visuelle Salienz wird mit Hilfe der Quaternion-basierten diskreten Cosinus-Transformation ermittelt und als Salienzkarte repräsentiert. Eine anschließende Segmentierung unter Ausnutzung der Isophotenkrümmung und die Verwendung von 3D-Informationen einer Stereokamera ermöglichen die Erstellung von visuellen Salienzclustern.

Durch ein räumliches und zeitliches Clustern der unimodalen Salienzcluster können multimodale Salienzcluster generiert werden, welche saliente Regionen im Kamerabild bzw. markante Geräusche in der Umgebung repräsentieren. Anhand einiger Beispiele konnte gezeigt werden, dass sowohl die visuelle Salienz als auch die akustische Salienz die in einer Szene relevanten Objekte und Ereignisse darstellen.

Zusammenfassend lässt sich sagen, dass ein humanoider Roboter oder ein anderes autonomes System in der Lage sein muss, wichtige Ereignisse bzw. wichtige Objekte wahrzunehmen, um auf diese reagieren zu können. Durch die Bestimmung der Salienz von Objekten und Geräuschen lässt sich allgemein die Aufmerksamkeit eines Roboters steuern. Dabei ist insbesondere eine schnelle Reaktion auf die Geschehnisse im Umfeld eines Roboters wichtig. Im Rahmen der vorliegenden Arbeit wird im späteren Verlauf die Salienz als ein Priorisierungskriterium bei der Exploration einer Szene verwendet.

3

Objektzentrierte Umwelterfassung

In diesem Kapitel wird die Wissensakquise und -repräsentation in einem autonomen System detailliert vorgestellt. Für den alltäglichen Betrieb solcher Systeme ist es notwendig, verschiedene Informationen über die aktuelle Umgebung zu erfassen und in aufbereiteter Form bereitzustellen. Für die Bewältigung von Aufgaben und die Interaktion mit Menschen werden beispielsweise die Personen und Gegenstände im näheren Umfeld erfasst und auf einheitliche Art und Weise im System repräsentiert.

3.1 Umweltmodell

In der Literatur findet sich eine Vielzahl teils unterschiedlicher, aber auch sehr ähnlicher Ansätze für die Modellierung der Umwelt und der Repräsentation der gewonnenen Informationen. Dabei steht meist die Anwendung stark im Vordergrund. Für einen mobilen Roboter zur Kartierung großer Flächen, der seine Umwelt als Ganzes erfasst, werden oft komplexe 3D-Daten als Repräsentation der Umwelt gewählt. Dies hat jedoch den Nachteil, dass die Datenmenge sehr groß ist und die Daten meist sehr abstrakt sind. Ein dazu gegensätzlicher Ansatz ist es, die aktuelle Szene (d. h. die Umgebung) in Form von Objekten zu repräsentieren (vgl. [Ghe08], [Bau10], [Mac10a], [Bel12], et al.). Ein sogenanntes Umweltmodell dient humanoiden Robotern und anderen autonomen Systemen dabei als Gedächtnisstruktur und repräsentiert die wahrgenommen Informationen, insbesondere Personen und Gegenstände, in einheitlicher Form als Objekte. Zusätzlich kann im Umweltmodell auch a-priori vorhandenes Wissen (z. B. die Raumgeometrie) als Objekte modelliert werden. Bei dieser objektzentrierten Darstellung der Umwelt werden alle vorhandenen Informationen über ein Objekt in Form von Objekteigenschaften und -beziehungen abstrahiert. Diese Informationen können

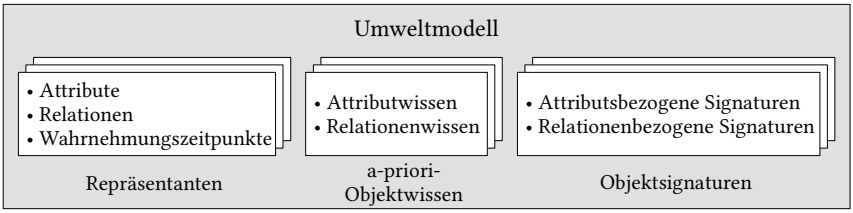

Abb. 3.1: Aufbau des Umweltmodells mit den wahrgenommenen Objekten (als Repräsentanten), dem a-priori-Objektwissen und den Objektsignaturen.

jederzeit beispielsweise für die Bewältigung von anstehenden Aufgaben oder die Interaktion mit dem Menschen verwendet werden.

Im Rahmen der vorliegenden Arbeit wird die grundlegende Definition eines Umweltmodells aus [Mac10a] fortgeführt und weiterentwickelt. Dabei werden die Anforderungen an ein Umweltmodell wie beispielsweise *dynamische Erweiterbarkeit*, *Speicherung von heterogenen Daten* und *Formalisierbarkeit* übernommen. Zusätzlich werden neue Anforderungen definiert, um u. a. der wissensbasierten Neugier (vgl. Kapitel 4) Rechnung zu tragen. Anschließend wird die Definition und Realisierung von Signaturen aus [Swe09] fortgeführt, um die Identifizierung und Wiedererkennung von Gegenständen und Personen zu ermöglichen.

3.1.1 Formale Definition des Umweltmodells

Das *Umweltmodell* stellt eine Repräsentation der realen Welt dar. Es ist die Wissensgrundlage für einen humanoiden Roboter oder ein anderes autonomes System und stellt somit eine Gedächtnisstruktur dar. Das Umweltmodell (vgl. Abb. 3.1) umfasst die aktuell wahrgenommenen Objekte $\mathcal{O}$, das vorhandene a-priori-Wissen (u. a. das vordefinierte Objektwissen $\mathcal{O}_{\text{Vorwissen}}$) und die Signaturen $\mathcal{O}_{\text{Signaturen}}$ zur Wiedererkennung von Objekten (vgl. [Swe09]). Das gesamte aktuell verfügbare Wissen $\mathcal{W}$ lässt sich somit formal als 3-Tupel definieren.

$$\mathcal{W} = \{\mathcal{O}, \mathcal{O}_{\text{Vorwissen}}, \mathcal{O}_{\text{Signaturen}}\} \tag{3.1}$$

Die Gegenstände und Personen der realen Welt werden dabei als *Objekte* zusammengefasst. Die dazugehörigen Repräsentationen der Objekte im Umweltmodell werden als *Objektrepräsentanten* (kurz: Repräsentanten) bezeichnet. Die über die Zeit veränderlichen Repräsentanten werden über *Attribute*, *Relationen* und *Zeitpunkte der Wahrnehmung* durch ein 3-Tupel definiert:

$$o = (\mathcal{A}_o, \mathcal{R}_o, \mathcal{Z}_o) \in \mathcal{O} \tag{3.2}$$

Im Gegensatz zu [Mac10a] werden die dort definierten Attribute in der vorliegenden Arbeit unterteilt in *Attribute* und *Relationen*. Relationen stellen Eigenschaften mit externen Bezügen dar. Diese können räumliche Verknüpfungen („steht auf *Tisch*“) oder auch eine Beziehung zu einem anderen Objekt („Bruder von *Benjamin*“) darstellen. Des Weiteren werden auch sogenannte Klassenrelationen damit modelliert (vgl. Abschnitt 3.2.2). Alle sonstigen Eigenschaften eines Objekts werden als Attribute realisiert. Die Attribute und Relationen stellen die kleinsten Informationseinheiten eines Objektrepräsentanten dar.

Attribute

Die Menge aller Attribute $\mathcal{A}_o$ eines Objektrepräsentanten o beinhaltet die objektbeschreibenden Eigenschaften, welche entweder durch Sensoren akquiriert wurden oder aus anderen Informationsquellen, wie beispielsweise dem a-priori-Wissen, stammen. Die Menge an Attributen ist definiert über

$$\mathcal{A}_o = \{a_1, \ldots, a_I\} \quad \text{mit } a_i := \left(t_{a_i}, w_{a_i}, k_{a_i}, \rho_{a_i}\right) \in \mathcal{A}_o\,, \quad I \in \mathbb{N}. \tag{3.3}$$

Hierbei ist das objektspezifische Attribut a_i definiert über ein 4-Tupel bestehend aus Attributtyp t_{a_i}, Attributwert w_{a_i}, Konfidenzwert $k_{a_i} \in [0,1]$ und Priorität $\rho_{a_i} \in [0,1]$.

Der *Attributtyp* t_{a_i} ist ein eindeutiger Bezeichner für ein Attribut und legt gleichzeitig auch den Datentyp und gegebenenfalls die Maßeinheit für das Attribut fest. Die Menge aller im Umweltmodell existierender Attributtypen ist definiert als $\mathcal{T}_\mathcal{A} = \{t_{a_1}, \ldots, t_{a_N}\}$. Ein Beispiel für einen Attributtyp ist die Größe einer Person. Hierbei werden als Datentyp eine Fließkommazahl und als Einheit beispielsweise Meter festgelegt. Neben einfachen Datentypen, wie ganzen Zahlen oder Fließkommazahlen, können auch komplexere Datentypen, wie Vektoren oder Matrizen, hinterlegt werden, welche z. B. die 3D-Position oder die Oberfläche eines Objekts repräsentieren.

Der *Attributwert* w_{a_i} stellt die zum Attribut a_i gehörenden Daten dar. Der Datentyp des Attributwertes ist aufgrund des Attributtyps festgelegt. Durch neu akquirierte Sensorinformationen oder erworbenes Wissen ist der Attributwert über die Zeit veränderlich. Ein gutes Beispiel hierfür ist die 3D-Position einer Person: Diese besteht aus einem Vektor mit x-, y-, z-Komponente und ist in Meter angegeben. Durch die Erfassung der Bewegungen einer Person ändert sich entsprechend das Attribut *Position* im Laufe der Zeit.

Der *Konfidenzwert* $k_{a_i} \in [0,1]$ repräsentiert die Sicherheit der im Attributwert w_{a_i} hinterlegten Daten. Die Bestimmung des Konfidenzwertes ist eng mit der im Attributwert gespeicherten Art an Daten verknüpft (vgl. [Mac10a]).

Die *Priorität* $\rho_{a_i} \in [0,1]$ stellt die Wichtigkeit des Attributes in Bezug auf die Gesamteigenschaft eines Objekts dar. Diese kann vorab durch Expertenwissen oder auch durch Zugriffsstatistiken eines Attributs, z. B. durch ein autonomes System zur Bewältigung von Aufgaben, bestimmt werden. Sollte beides nicht zur Verfügung stehen, wird die Priorität für alle Attribute gleich bewertet. Die Priorität wurde im Vergleich zu [Mac10a] ergänzt, um für die Attribute eine unterschiedliche Gewichtung zu ermöglichen, da manche Eigenschaften eine wichtigere Rolle spielen können. Ein Beispiel hierfür ist die Priorisierung des Attributs *Identität* zum schnelleren Auffinden einer bestimmten Person.

Relationen

Die *Objektrelationen* stellen die Beziehungen zwischen dem aktuellen Objekt und seiner Umwelt, oder auch zu anderen Objekten, dar. Die Relationen eines Objektrepräsentanten $\mathcal{R}_o$ lassen sich wie folgt definieren:

$$\mathcal{R}_o = \{r_1, \ldots, r_J\} \quad \text{mit } r_j := \left(t_{r_j}, w_{r_j}, k_{r_j}, \rho_{r_j}\right) \in \mathcal{R}_o\,, \quad J \in \mathbb{N}\,. \tag{3.4}$$

In diesem Zusammenhang besteht eine spezifische Relation r_j aus dem 4-Tupel Relationstyp t_{r_j}, Relationswert w_{r_j}, Konfidenzwert k_{r_j} und Priorität ρ_{r_j}. Die letzten beiden sind analog zu den Attributen definiert.

Der *Relationstyp* t_{r_j} ist ein eindeutiger Bezeichner für eine Relation r_j und legt gleichzeitig den Datentyp für den Relationswert w_{r_j} fest. Die Menge aller im Umweltmodell vorhandener Relationstypen ist definiert als $\mathcal{T}_\mathcal{R} = \{t_{r_1}, \ldots, t_{r_M}\}$. Ein Beispiel für den Relationstyp ist der Besitzer eines Gegenstandes: Der Relationstyp *Besitzer* legt den Datentyp des Relationswertes fest. Dabei handelt es sich nicht um einen einfachen Datentyp, wie eine Fließkommazahl oder Matrix, sondern um einen komplexen Datentyp, der einen Verweis auf die entsprechende Person darstellt.

Der *Relationswert* w_{r_j} stellt den eigentlichen Verweis auf ein Ziel dar. Dies kann ein konkreter Bezeichner, z. B. *Benjamin* oder *Alexej*, sein oder auch eine symbolische Verknüpfung über einen Objektrepräsentanten (z. B. o_3). Des Weiteren kann der Relationswert auch die Zugehörigkeit des aktuellen Objektrepräsentanten zur einer Gruppe von Objekten abbilden.

Es existiert eine Vielzahl an Relationen, die sich in verschiedene Kategorien einteilen lassen. Ortsbezogene und objektbezogene Relationen sind hierbei die wichtigsten Vertreter. So werden der Wohnort einer Person über einen Ortsbezug und die Verwandtschaft über einen Objektbezug dargestellt. Beides wird über Relationen im Umweltmodell modelliert. Des Weiteren können Relationen, wie in Abschnitt 3.2.2 beschrieben, genutzt werden, um die Klassenhierarchie von Objekten zu modellieren.

Zeitpunkte der Wahrnehmungen

Für die spätere Bestimmung der wissensbasierten Neugier (insbesondere für den Teilaspekt Neuartigkeit; vgl. Abschnitt 4.3.3) ist es wichtig, die *Zeitpunkte der Wahrnehmung* eines Objekts zu kennen. Der aktuelle Zeitpunkt der Wahrnehmung eines Gegenstands oder einer Person – auch während einer längeren Phase – wird stets mit t_Z bezeichnet. Die Menge aller Zeitpunkte $\mathcal{Z}_o$ lässt sich darstellen als

$$\mathcal{Z}_o = \{t_1, \ldots, t_Z\} \subseteq \mathbb{R} \quad \text{mit } Z \in \mathbb{N}. \tag{3.5}$$

Für eine einheitliche Darstellung sollten die Zeitpunkte $t_z \in \mathcal{Z}_o$ einen absoluten zeitlichen Bezug haben sowie eine angemessene zeitliche Auflösung besitzen. Für die vorliegende Arbeit werden als Referenzzeitpunkt der Beginn des Jahres 1 n. Chr. gewählt und als Einheit Tage. Der aktuelle Zeitpunkt wird relativ zum Referenzzeitpunkt angeben und ist dabei anteilig in Tagen als Fließkommazahl dargestellt.

a-priori-Objektwissen

Unter *a-priori-Wissen* werden alle Informationen zusammengefasst, welche im Vorhinein definiert wurden. So umfasst dieses beispielsweise Informationen über die aktuelle Umgebung in Form von Raummodellen. Das a-priori vorhandene Objektwissen wird durch a-priori-Objekte definiert. Jedes a-priori-Objekt o_p besteht dabei aus Attributen und Relationen:

$$o_p = \left(\mathcal{A}_{o_p}, \mathcal{R}_{o_p}\right) \in \mathcal{O}_{\text{Vorwissen}}. \tag{3.6}$$

Die Menge aller a-priori-Objekte wird mit $\mathcal{O}_{\text{Vorwissen}}$ bezeichnet. Die Attribute sind definiert als ein Tupel bestehend aus Attributtyp $t_{a_i}^{o_p}$ und gültigem Wertebereich $w_{a_i}^{o_p}$ bzw. Menge an validen Werten $\mathcal{W}_{a_i}^{o_p}$:

$$\begin{aligned} \mathcal{A}_{o_p} = \left\{a_1^{o_p}, \ldots, a_I^{o_p}\right\} \quad & \text{mit } a_i^{o_p} := \left(t_{a_i}^{o_p}, w_{a_i}^{o_p}\right) \in \mathcal{A}_{o_p} \\ & \text{bzw. } a_i^{o_p} := \left(t_{a_i}^{o_p}, \mathcal{W}_{a_i}^{o_p}\right) \in \mathcal{A}_{o_p}, \quad I \in \mathbb{N}. \end{aligned} \tag{3.7}$$

Die Relationen sind dazu analog ebenfalls über ein Tupel definiert, welches aus dem Relationstyp $t_{r_j}^{o_p}$ und den gültigen Werten $\mathcal{W}_{r_j}^{o_p}$ bzw. Wertebereichen $w_{r_j}^{o_p}$ besteht:

$$\begin{aligned} \mathcal{R}_{o_p} = \left\{r_1^{o_p}, \ldots, r_J^{o_p}\right\} \quad & \text{mit } r_j^{o_p} := \left(t_{r_j}^{o_p}, w_{r_j}^{o_p}\right) \in \mathcal{R}_{o_p} \\ & \text{bzw. } r_j^{o_p} := \left(t_{r_j}^{o_p}, \mathcal{W}_{r_j}^{o_p}\right) \in \mathcal{R}_{o_p}, \quad J \in \mathbb{N}. \end{aligned} \tag{3.8}$$

Ein Beispiel für die Attribute ist die Farbe eines Apfels. Das dazugehörige a-priori-Objekt definiert existierende Farben und Farbkombinationen von Äpfeln auf der Welt. Eine dazugehörige Relation kann mögliche Herkunftsländer von Äpfeln definieren.

3.1.2 Formale Kurzformen für Attribute und Relationen

Im Folgenden werden noch einige *Kurzformen* für Attribute und Relationen des aktuellen Objektrepräsentanten o eingeführt. So lässt sich ein Attribut oder eine Relation mit einem bestimmten Typ direkt schreiben als:

$$a_{\text{Identität}} := a_i \quad \text{mit } t_{a_i} = \text{„Identität“ und } a_i \in \mathcal{A}_o \tag{3.9}$$

bzw.

$$r_{\text{Wohnort}} := r_j \quad \text{mit } t_{r_j} = \text{„Wohnort“ und } r_j \in \mathcal{R}_o\,. \tag{3.10}$$

Analog dazu lassen sich auch der Attributtyp, der Attributwert, der Konfidenzwert und die Priorität für ein spezifisches Attribut definieren als:

$$t_{a_{\text{Identität}}} := t_{a_i} \quad \text{mit } t_{a_i} = \text{„Identität“ und } a_i \in \mathcal{A}_o \tag{3.11}$$

$$w_{a_{\text{Identität}}} := w_{a_i} \quad \text{mit } t_{a_i} = \text{„Identität“ und } a_i \in \mathcal{A}_o \tag{3.12}$$

$$k_{a_{\text{Identität}}} := k_{a_i} \quad \text{mit } t_{a_i} = \text{„Identität“ und } a_i \in \mathcal{A}_o \tag{3.13}$$

$$\rho_{a_{\text{Identität}}} := \rho_{a_i} \quad \text{mit } t_{a_i} = \text{„Identität“ und } a_i \in \mathcal{A}_o\,. \tag{3.14}$$

Für den Relationstyp, den Relationswert mit dem dazu gehörenden Konfidenzwert und die Priorität für eine bestimmte Relation gelten analoge Kurzformen wie bei den Attributen. Dabei beziehen sich alle Informationen stets auf den aktuellen Objektrepräsentanten o.

Sind Attribute oder Relationen von mehreren Repräsentanten zu unterscheiden, so wird dies entsprechend gekennzeichnet. $a_i^{o_1}$ stellt beispielsweise das i-te Attribut des Objektrepräsentanten o_1 dar, wohingegen $a_i^{o_2}$ das i-te Attribut des Objektrepräsentanten o_2 bezeichnet. Dies gilt analog auch für Relationen und kann mit den zuvor eingeführten Kurzformen kombiniert werden: $k_{a_{\text{Identität}}}^{o_3}$ bezeichnet beispielsweise den Konfidenzwert des Attributs *Identität* für den Objektrepräsentanten o_3.

3.2 Abstraktionsebenen und Klassenhierarchie

Im Abschnitt 3.1.1 wurden die Grundlagen für die Repräsentation von Objekten im Umweltmodell gelegt. Werden nun die vielfältigen Anforderungen an ein autonomes System betrachtet, so lässt sich feststellen, dass die notwendige Wahrnehmung der einzelnen Objekte nicht immer in gleicher Detailliertheit erfolgen

muss. Die Navigation eines mobilen Roboters ist hierfür ein gutes Beispiel, da hierfür nicht alle Details über die Objekte in seiner Umgebung zu jedem Zeitpunkt bekannt sein müssen. Die Position und Ausdehnungen von Hindernissen sowie die Raumgeometrie reichen hierbei vollkommen aus. Soll ein Roboter hingegen eine bestimmte Tasse von einem Tisch holen, so sind wesentlich detailliertere Informationen über die Umgebung notwendig. Daher ist es sinnvoll, das Wissen über ein Objekt auf verschiedenen Abstraktionsebenen zu repräsentieren (vgl. [Mac10a]).

3.2.1 Abstraktionsgrad

Das Wissen über Objekte wird im hier vorgestellten Umweltmodell durch das Hinzufügen von neuen Attributen und Relationen immer größer. Infolgedessen wird der *Abstraktionsgrad* immer geringer. Im Rahmen der vorliegenden Arbeit werden nur physikalische Objekte im Umweltmodell repräsentiert, sodass diese auch immer eine Position aufweisen, welche durch das entsprechende Attribut *Position* dargestellt wird. Dies ist die abstrakteste Darstellung eines Objekts. Im Nachfolgenden kann sich durch neue Sensorinformationen der Abstraktionsgrad verändern. Werden mit Hilfe der gewonnenen Informationen neue Attribute und/oder Relationen hinzugefügt, so sinkt der Abstraktionsgrad. Im Gegensatz dazu können durch neue Sensorinformationen auch vorhandene Attribute und/oder Relationen angepasst werden, wodurch der Abstraktionsgrad nicht beeinflusst wird. Ein Beispiel hierfür ist die aktuelle Position. Führen neue Informationen zu der Erkenntnis, dass die bisher im Attribut und/oder in der Relation gespeicherten Daten nicht der Wahrheit entsprechen, d. h. der Konfidenzwert unter eine minimale Schwelle sinkt, so werden diese invalidiert, was infolgedessen zu einer Erhöhung des Abstraktionsgrads führen kann.

3.2.2 Klassenhierarchie

Die Klasse eines Objektrepräsentanten hängt eng mit dem Abstraktionsgrad zusammen. Alle Repräsentanten, die bestimmte Attribute und/oder Relationen besitzen, gehören zu einer spezifischen Klasse. Im einfachsten Fall hat ein Repräsentant das Attribut *Position* und gehört damit zur Klasse Punkt. Diese Zugehörigkeiten werden in der vorliegenden Arbeit durch das Vorhandensein sogenannter *Klassenrelationen* realisiert. Dabei handelt es sich um spezielle Relationen, welche einen Verweis auf eine Objektklasse herstellen. Sie werden durch das Prefix „ist“ gekennzeichnet. Ein Repräsentant, welcher einen Gegenstand abbildet, besitzt eine Relation istGegenstand mit einem Verweis auf die Objektklasse (Gegenstand) als Relationswert. Der dazugehörige Konfidenzwert gibt

die Sicherheit für die Klassenzugehörigkeit an. Durch neue Informationen können auch weitere Klassenrelationen zu einem Repräsentanten hinzugefügt werden. Ein Repräsentant kann dabei auch mehrere Klassenrelationen gleichzeitig aufweisen. Jede neue Klassenrelation ermöglicht in der Regel auch das Hinzufügen weiterer Attribute und/oder Relationen. Ist die neue Klasse beispielsweise GEGENSTAND, so kann das Attribut *Farbe* für diesen Repräsentanten hinzugefügt und bestimmt werden. Durch den Abstraktionsgrad und die im Laufe der Wahrnehmung hinzukommenden Klassenrelationen lässt sich eine *Klassenhierarchie* $\mathcal{H}$ definieren. In der vorliegenden Arbeit beginnt diese, wie bereits zuvor erwähnt, stets mit der Klasse PUNKT, da alle hier behandelten Objekte stets eine Position besitzen, und lässt sich je nach Detailgrad der Informationen über ein Objekt mit jeder neuen Klassenrelation weiter fortführen. Sollten in Zukunft noch abstraktere Informationen (insbesondere ohne Position) im Umweltmodell repräsentiert werden, so kann die Klasse ENTITÄT und die dazugehörige Klassenrelation istENTITÄT als Basis in der Klassenhierarchie ergänzt werden.

In Abb. 3.2 sind zwei Beispiele zu sehen, welche den Zusammenhang von Abstraktionsgrad und Klassenhierarchie darstellen. Zu Beginn weisen beide Objektrepräsentanten nur eine Zugehörigkeit zur Klasse PUNKT auf. Durch weitere Sensorinformationen kann festgestellt werden, dass es sich bei Objekt #6 um eine *Person* und bei Objekt #9 um einen *Gegenstand* handelt. Die farblich gekennzeichneten Attribute und Relationen kommen mit jeder neuen Klasse hinzu und somit sinkt auch der Abstraktionsgrad immer weiter. Letztendlich kann festgestellt werden, dass es sich bei Objekt #9 um einen *Mixer* handelt, welcher eine bestimmte Farbe, einen definierten Zustand und einen Besitzer hat. Bei Objekt #6 ist bekannt, dass es sich um die bestimmte Person (*Benjamin*) handelt, welche eine gewisse Größe hat und in einem bestimmten Ort wohnt. Bei Letzterem wurde der Wohnort aus dem vorhandenen a-priori-Wissen abgeleitet. Die aktuelle Ausprägung der Klassenhierarchie für den Objektrepräsentanten o_6 lautet somit $\mathcal{H}_{o_6} = \{\text{PUNKT}, \text{PERSON}, \text{BENJAMIN}\}$ und entsprechend für den Objektrepräsentanten o_9 ist diese $\mathcal{H}_{o_9} = \{\text{PUNKT}, \text{GEGENSTAND}, \text{MIXER}\}$.

3.2.3 Abstraktionsebenen und Wissensabhängigkeiten

Die schrittweise Erweiterung der Repräsentanten um neue Attribute und/oder Relationen ist ein wichtiger Teil der Umwelterfassung. Dabei lässt sich der Abstraktionsgrad mittels der zuvor definierten Klassen in *Abstraktionsebenen* unterteilen. Für das Hinzufügen neuer Attribute und/oder Relationen müssen bestimmte Voraussetzungen erfüllt sein, bevor ein Repräsentant erweitert werden kann. Zum einen müssen bestimmte Attribute bzw. Relationen vorhanden sein und zum anderen muss der Konfidenzwert eine zuvor festgelegte Schwelle überschreiten. So setzt beispielsweise die Erzeugung des Attributs *Farbe* die

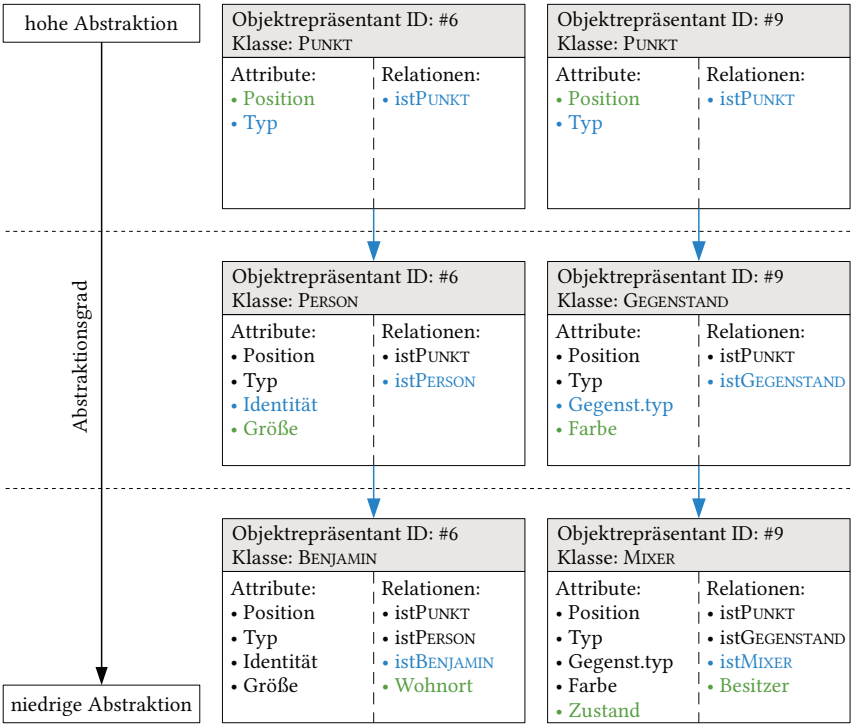

Abb. 3.2: Beispiel für verschiedene Abstraktionsgrade und Klassenhierarchien anhand von zwei Objekten.

Klassenrelation istGEGENSTAND voraus, da dieses Attribut spezifisch für einen Gegenstand ist. Zusätzlich muss die Sicherheit in Bezug auf die Klasse GEGENSTAND hoch sein, was sich in einem hohen Konfidenzwert der Klassenrelation istGEGENSTAND widerspiegelt. Diese Art von Abhängigkeiten zwischen den Attributen und Relationen im Umweltmodell wird fortan als *Wissensabhängigkeiten* bezeichnet (vgl. [Mac10a]).

Mit den in Abschnitt 3.2.2 beschriebenen Klassenrelationen lassen sich Repräsentanten anhand ihrer Attribute und Relationen zu bestimmten Klassen zuordnen. So steht beispielsweise die Relation istPUNKT für die Klasse aller realen Objekte. Entsprechendes gilt bei den Klassenrelationen istPERSON und istGEGENSTAND, welche die Klasse aller Personen bzw. Gegenstände repräsentieren.

Es wäre theoretisch möglich, alle im Umweltmodell vorhandenen Attribut- und Relationstypen beliebig den Repräsentanten zuzuordnen. Durch die Verwendung der zuvor definierten Wissensabhängigkeiten wird die Zuordnung auf die sinn-

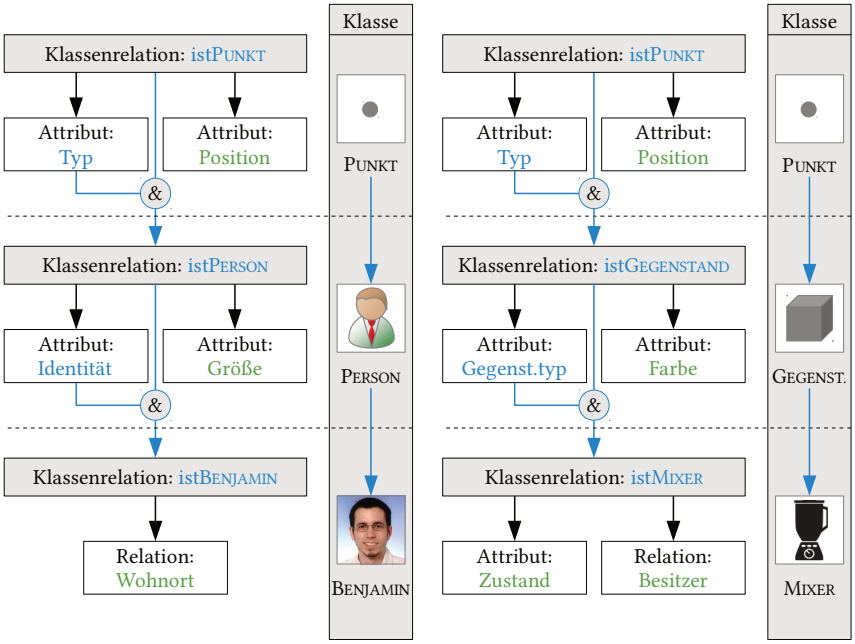

Abb. 3.3: Die Klassenhierarchie für eine Person (links) und einen Gegenstand (rechts) ist durch die Wissensabhängigkeiten von einzelnen Attribut- und Relationstypen bestimmt.

vollen Fälle reduziert. Die in Abschnitt 3.2.2 definierte Klassenhierarchie lässt sich nun durch die Wissensabhängigkeiten erzeugen.

In Abb. 3.3 ist der Zusammenhang zwischen den Abstraktionsebenen, den Wissensabhängigkeiten der Attribute und Relationen sowie der Klassenhierarchie anhand von zwei Beispielen dargestellt. Mit Hilfe der Klassenrelationen (istPUNKT, istPERSON, istGEGENSTAND, usw.) können über die Wissensabhängigkeiten schrittweise neue Attribute und Relationen hinzugefügt werden (schwarze Pfeile). Anhand von speziellen Attributen und Klassenrelationen (blau) können unter Berücksichtigung der dazugehörigen Konfidenzwerte neue Klassenrelationen erzeugt werden (blaue Pfeile). Dadurch kann eine neue Klasse in die Hierarchie eingefügt werden und somit das Objekt detaillierter repräsentiert werden. Es können auch zusätzliche Attribute und Relationen (grün) hinzugefügt werden, die keine weitere Ebene in der Klassenhierarchie erzeugen (beispielsweise die *Größe* oder die *Farbe*) und dennoch ein Objekt näher beschreiben. Die kumulativen hierarchischen Abhängigkeiten von allen Attributen und Relationen und die Zuordnung zu den einzelnen Objektklassen sind in Abb. 3.3 ebenfalls zu sehen.

3.3 Multimodale objektbeschreibende Signaturen

Die sogenannten *objektbeschreibenden Signaturen* (kurz: Objektsignaturen) werden für die multimodale Beschreibung und spätere Wiedererkennung von Gegenständen und Personen in einer Szene benötigt (vgl. [Swe09]). Objektbeschreibende Signaturen sind dabei eine Zusammenstellung aus attributs- und relationsbezogenen Signaturen, welche ein konkretes Objekt oder eine Gruppe von Objekten eindeutig charakterisieren. Dafür werden aus den akquirierten Sensordaten eines Objekts repräsentative Merkmale extrahiert und durch sogenannte Signaturgenerierungstransformationen (vgl. [Swe09]) in beschreibende mathematische Modelle für eine spätere Klassifizierung überführt. In den folgenden Unterabschnitten werden die Erkenntnisse in Bezug auf Signaturen aus [Swe09] zusammenfassend dargestellt und in Hinblick auf Relationen ergänzt.

3.3.1 Anforderungen an objektbeschreibende Signaturen

Für eine zuverlässige Klassifizierung von Objekten mittels objektbeschreibender Signaturen werden folgende Anforderungen definiert (vgl. [Swe09]):

- *Eindeutigkeit:* Alle Signaturen müssen die korrespondierenden Objekte eindeutig beschreiben. Dabei kann eine Signatur nicht mehrere Objekte beschreiben, wenn diese durch den aktuellen Abstraktionsgrad unterschieden werden können, d. h., die Personen *A* und *B* besitzen verschiedene Signaturen, wenn diese auf individueller Ebene betrachtet werden, wohingegen die beiden Personen bei der Betrachtung auf einer höheren Ebene nur gemeinsame personenbezogene Signaturen besitzen. Diese beschreiben dabei allgemeine Attribute und Relationen von Personen.
- *Speicher- und Recheneffizienz:* Die Signaturen erfordern eine kompakte Repräsentation der Objekte anhand von effektiven Merkmalen und nicht in Form von Sensorrohdaten, d. h. unverarbeiteten Audio- bzw. Videodaten. Die Signaturgenerierung erfolgt anhand von objektbeschreibenden Merkmalen, welche die relevanten Informationen aus Sensorsignalen repräsentieren. Die Merkmalsgenerierung selbst erfolgt mittels Analysetransformation (vgl. [Hof98]) und reduziert i. d. R. den Datenumfang.
- *Aktualisierungseigenschaft:* Die Aktualisierung bzw. Anpassung bereits generierter Signaturen ist notwendig, wenn initial wenige oder unsichere Informationen über ein Objekt vorhanden sind. Des Weiteren ist auch eine Verfeinerung von bereits vorhandenem Wissen notwendig, um eine Verbesserung der Signaturen zu erreichen. Dabei erfolgt die Anpassung bereits vorhandener Signaturen durch neue Informationen mit einer geringeren Gewichtung, um eine Überanpassung zu vermeiden.

- *Vervollständigungseigenschaft:* Objekte werden u. U. bei ihrer ersten Erfassung und der anschließenden Signaturgenerierung nicht vollständig erfasst, sodass zu einem späteren Zeitpunkt neue Modalitäten und weitere Eigenschaften hinzukommen. Infolgedessen wird eine bereits vorhandene Objektsignatur erweitert, um eine vollständigere Erfassung zu ermöglichen. Sowohl die Aktualisierung als auch die Vervollständigung sind kontinuierliche Prozesse und für eine ganzheitliche Wahrnehmung dringend erforderlich.

3.3.2 Formale Definitionen der Signaturen

Im Nachfolgenden wird die Generierung von attributs- und relationsbezogenen Signaturen als Basis für objektbeschreibende Signaturen formal beschrieben.

Attributs- und relationsbezogene Signaturen

Die Erfassung und Repräsentation der aktuellen Szene und der darin befindlichen Gegenstände und Personen erfolgt, wie bereits zuvor thematisiert, auf Basis von Objektrepräsentanten in einem Umweltmodell. Dabei werden die Objekte der realen Welt in Form von Attributen und Relationen dargestellt (vgl. Abschnitt 3.1.1). Sogenannte attributsbezogene Signaturen charakterisieren in [Swe09] bestimmte Eigenschaften eines spezifischen Objekts oder Objekttyps eindeutig. Im Rahmen der vorliegenden Arbeit werden die rein attributsbezogenen Signaturen analog durch relationsbezogene Signaturen ergänzt.

Aus den Sensordaten für das aktuell betrachtete Objekt o werden in einem ersten Schritt die relevanten Informationen extrahiert und in Form von beschreibenden Merkmalen repräsentiert. Dabei werden mit Hilfe einer für das Attribut bzw. für die Relation spezifischen Analysetransformation (engl.: analysis transform; kurz: AT; vgl. [Hof98]) die relevanten Informationen für eine spätere Wiedererkennung aus dem Sensorsignal extrahiert und in Form von speichereffizienten Merkmalen kompakt repräsentiert. Formal lässt sich dies beschreiben als

$$\boldsymbol{x}_{a,i} := f_a^{\mathrm{AT}}\left(\boldsymbol{d}_{a,i}\right) \quad \text{bzw.} \quad \boldsymbol{x}_{r,i} := f_r^{\mathrm{AT}}\left(\boldsymbol{d}_{r,i}\right), \tag{3.15}$$

wobei $\boldsymbol{d}_{a,i}$ bzw. $\boldsymbol{d}_{r,i}$ die relevanten Sensordaten für das Attribut a bzw. die Relation r zum Zeitpunkt i sind. Mit Hilfe der korrespondierenden Analysetransformationen $f_a^{\mathrm{AT}}(\cdot)$ bzw. $f_r^{\mathrm{AT}}(\cdot)$ werden die attributs- bzw. die relationsbezogenen Merkmalsvektoren $\boldsymbol{x}_{a,i}$ und $\boldsymbol{x}_{r,i}$ extrahiert.

Die Menge aller Merkmalsvektoren aus verschiedenen Zeitschritten für das Attribut a bzw. die Relation r bildet die Grundlage für die Generierung der jeweiligen Signatur und ist definiert als

$$\mathcal{X}_a := \{\boldsymbol{x}_{a,1}, \ldots, \boldsymbol{x}_{a,k}\} \quad \text{bzw.} \quad \mathcal{X}_r := \{\boldsymbol{x}_{r,1}, \ldots, \boldsymbol{x}_{r,l}\} \quad \text{mit } k, l \in \mathbb{N}^+ . \tag{3.16}$$

Mit Hilfe der attributs- bzw. relationsspezifischen Signaturgenerierungstransformation (engl.: signature generation transform; kurz: SGT; vgl. [Swe09]) werden die Merkmalsvektoren in eine für das Objekt o spezifische Signatur überführt:

$$\mathcal{S}_{a|o} = f_a^{\mathrm{SGT}}(\mathcal{X}_a) \quad \text{bzw.} \quad \mathcal{S}_{r|o} = f_r^{\mathrm{SGT}}(\mathcal{X}_r) . \tag{3.17}$$

Anstatt die Merkmale direkt durch eine Signaturgenerierungstransformation in eine attributs- bzw. relationsbezogene Signatur zu überführen, kann zunächst alternativ auch eine Merkmalsanalysetransformation (engl.: feature analysis transform; kurz: FAT; vgl. [Swe09]) verwendet werden, um neue bzw. angepasste Merkmale zu generieren:

$$\tilde{\mathcal{X}}_a := f_a^{\mathrm{FAT}}(\mathcal{X}_a) = \tilde{\boldsymbol{x}}_{a,1}, \ldots, \tilde{\boldsymbol{x}}_{a,m} \quad \text{bzw.} \quad \tilde{\mathcal{X}}_r := f_r^{\mathrm{FAT}}(\mathcal{X}_r) \quad \text{mit } m \in \mathbb{N}^+ \tag{3.18}$$

Ein einfaches Beispiel hierfür ist die Transformation der Ausdehnung eines Objekts, repräsentiert als Vektor $\boldsymbol{x}$, in ein Volumen $\tilde{x}$ mit $\boldsymbol{x}_{a,i} \in \mathbb{R}^3$ und $\tilde{x}_{a,i} \in \mathbb{R}^1$.

Neben den zuvor beschriebenen Möglichkeiten, kann durch die Verwendung einer direkten Signaturgenerierungstransformation (kurz: dSGT) aus den relevanten Daten

$$\mathcal{D}_a := \{\boldsymbol{d}_{a,1}, \ldots, \boldsymbol{d}_{a,k}\} \quad \text{bzw.} \quad \mathcal{D}_r := \{\boldsymbol{d}_{r,1}, \ldots, \boldsymbol{d}_{r,l}\} \quad \text{mit } k, l \in \mathbb{N}^+ , \tag{3.19}$$

eine attributs- bzw. relationsbezogene Signatur erzeugt werden:

$$\tilde{\mathcal{S}}_{a|o} = f_a^{\mathrm{dSGT}}(\mathcal{D}_a) \quad \text{bzw.} \quad \tilde{\mathcal{S}}_{r|o} = f_r^{\mathrm{dSGT}}(\mathcal{D}_r) . \tag{3.20}$$

Diese Art der Signaturgenerierung kann bei Aktualisierung der attributs- bzw. relationsbezogenen Signatur (vgl. Abschnitt 3.3.3) einen entscheidenden Nachteil aufweisen, da u. U. die originalen Sensordaten benötigt werden, welche in den meisten Fällen um ein Vielfaches größer sind als die kompakten Merkmale aus Gl. 3.16.

Objektbeschreibende Signaturen

Eine objektbeschreibende Signatur ist eine Zusammenstellung aus korrespondierenden attributs- und relationsbezogenen Signaturen für ein spezifisches Objekt o. Die Signatur repräsentiert dabei die aktuelle Ausprägung des Objekts und dient als Grundlage für eine Identifizierung sowie spätere Wiedererkennung. Formal lässt sich eine objektbeschreibende Signatur definieren als

$$\mathcal{S}_o = \bigcup_a \mathcal{S}_{a|o} \bigcup_r \mathcal{S}_{r|o} \in \mathcal{O}_{\text{Signaturen}} , \tag{3.21}$$

wobei $\mathcal{O}_{\text{Signaturen}}$ die Menge aller Objektsignaturen im Umweltmodell ist (vgl. Abschnitt 3.1.1). Die Generierung von einer oder mehreren attributs- bzw. relationsbezogenen Signaturen erfolgt nach Gl. 3.17.

3.3.3 Aktualisierung und Vervollständigung von Signaturen

Bei der Signaturgenerierung ist es wichtig ausreichend viele Merkmalsvektoren zu akquirieren, welche gleichzeitig ein Objekt oder eine Gruppe von Objekten möglichst gut beschreiben und die Gefahr einer Verwechslung reduzieren. Dies ist notwendig, damit eine sichere Wiedererkennung von Objekten gewährleistet werden kann. Dazu ist es erforderlich, während der Phase der Informationsakquise das Objekt kontinuierlich nachzuverfolgen. In einem realen Szenario ist es nicht immer möglich, ausreichend viele Informationen gleichzeitig zu akquirieren bzw. alle Aspekte eines Objekts (insbesondere akustische Merkmale) zu erfassen. Deshalb müssen zuvor generierte Signaturen auf Basis von neuen Sensordaten aktualisiert bzw. um zuvor nicht erfasste Aspekte ergänzt werden.

Aktualisierung von attributs- und relationsbezogenen Signaturen

Die Aktualisierung einer attributs- bzw. relationsbezogenen Signatur kann prinzipiell auf zwei unterschiedliche Weisen geschehen: Zum einen werden im Laufe der Zeit weitere Merkmalsvektoren generiert, um mit Hilfe der zuvor extrahierten Vektoren sowie einer attributs- bzw. relationsspezifischen Signaturgenerierungstransformation eine neue Signatur zu erzeugen. Zum anderen erfolgt eine Adaptation der zuvor generierten Signatur anhand neuer Merkmalsvektoren.

Formal lässt sich die Signaturaktualisierung auf Merkmalsebene mit Hilfe der ursprünglichen Signaturgenerierungstransformation wie folgt definieren:

$$\mathcal{S}_{a|o}^{\text{neu}} = f_a^{\text{SGT}}\{\underbrace{\boldsymbol{x}_{a,1},\ldots,\boldsymbol{x}_{a,k}}_{\text{alt}},\underbrace{\boldsymbol{x}_{a,k+n},\ldots,\boldsymbol{x}_{a,k+n+m}}_{\text{neu}}\} \quad \text{mit } k,n,m \in \mathbb{N}^+ \tag{3.22}$$

Dabei werden sowohl die ursprünglichen Merkmalsvektoren $\{\boldsymbol{x}_{a,1},\ldots,\boldsymbol{x}_{a,k}\}$ als auch weitere Merkmalsvektoren $\{\boldsymbol{x}_{a,k+n},\ldots,\boldsymbol{x}_{a,k+n+m}\}$, welche zu späteren Zeitpunkten extrahiert wurden, bei der Signaturgenerierungstransformation und somit der Signaturaktualisierung berücksichtigt. Für relationsbezogene Signaturen gilt ein analoges Vorgehen.

Bei der Signaturaktualisierung auf Merkmalsebene liegt der Vorteil in der einfachen und analogen Herangehensweise wie bei der Signaturgenerierung. Dies erfordert jedoch die permanente Verfügbarkeit aller vorheriger Merkmalsvektoren, was zugleich ein erheblicher Nachteil ist, da die Datenmenge und somit auch die benötigte Zeit zur Signaturgenerierung immer größer wird. Eine direkte Adaptation von Signaturen durch neue Merkmalsvektoren ist dazu ein alternativer Ansatz, welcher jedoch nicht immer möglich ist. Diese Vorgehensweise erfordert eine spezielle Repräsentation der Signatur selbst, welche eine nachträgliche Anpassung ermöglicht. Wird die Signatur beispielsweise über ein parametrisches

Modell repräsentiert, so kann dieses durch neue Merkmalsvektoren und durch eine entsprechende Signaturaktualisierungstransformation (engl.: signature adaptation transform; kurz: SAT) adaptiert werden:

$$\mathcal{S}_{a|o}^{\text{neu}} = f_a^{\text{SAT}}\left\{\mathcal{S}_{a|o}^{\text{alt}}; \boldsymbol{x}_{a,k+n}, \ldots, \boldsymbol{x}_{a,k+n+m}\right\} \quad \text{mit } k, n, m \in \mathbb{N}^+ \tag{3.23}$$

Ein Beispiel hierfür sind Gaußsche Mischverteilungsmodelle (kurz: GMM), welche durch die iterative Nutzung des EM-Algorithmus eine Anpassung an die neuen Objektinformationen ermöglichen (vgl. [Swe09], [Bis06]) und somit die Aktualisierungseigenschaft von Signaturen realisieren.

Signaturvervollständigung

Die Vervollständigung einer Signatur kann auf zwei verschiedenen Ebenen erfolgen: zum einen die Ergänzung fehlender Modalitäten bei einer attributs- bzw. relationsbezogenen Signatur und zum anderen das Hinzufügen neuer attributs- bzw. relationsbezogenen Signaturen.

Die Ergänzung fehlender Modalitäten sei beispielhaft anhand einer attributsbezogenen Signatur dargestellt und lässt sich einfach auf relationsbezogene Signaturen übertragen: Jedes Attribut a kann potentiell sowohl durch visuelle als auch akustische Sensordaten in Form von korrespondierenden Merkmalsvektoren repräsentiert werden:

$$\boldsymbol{x}_{a,i}^{\text{akustisch}} := f_{a^{\text{akustisch}}}^{\text{AT}}\left(\boldsymbol{d}_{a,i}^{\text{akustisch}}\right) \quad \text{bzw} \quad \boldsymbol{x}_{a,i}^{\text{visuell}} := f_{a^{\text{visuell}}}^{\text{AT}}\left(\boldsymbol{d}_{a,i}^{\text{visuell}}\right). \tag{3.24}$$

Durch entsprechende Signaturgenerierungstransformationen können diese in für jede Modalität separate attributsbezogene Signaturen überführt werden:

$$\mathcal{S}_{a^{\text{akustisch}}|o} = f_{a^{\text{akustisch}}}^{\text{SGT}}\left\{\boldsymbol{x}_{a,1}^{\text{akustisch}}, \ldots, \boldsymbol{x}_{a,k}^{\text{akustisch}}\right\} \tag{3.25}$$

bzw.

$$\mathcal{S}_{a^{\text{visuell}}|o} = f_{a^{\text{visuell}}}^{\text{SGT}}\left\{\boldsymbol{x}_{a,1}^{\text{visuell}}, \ldots, \boldsymbol{x}_{a,k}^{\text{visuell}}\right\}. \tag{3.26}$$

Eine gemeinsame attributsbezogene Signatur für die aktuelle Ausprägung des Attributs a des Objekts o ist definiert als

$$\mathcal{S}_{a|o} = \mathcal{S}_{a^{\text{akustisch}}|o} \cup \mathcal{S}_{a^{\text{visuell}}|o}. \tag{3.27}$$

Eine Aktualisierung der modalen Einzelsignaturen erfolgt analog zu der im vorherigen Abschnitt beschriebenen Vorgehensweise.

Die Vervollständigungseigenschaft für objektbeschreibende Signaturen lässt sich für komplett neue attributs- bzw. relationsbezogene Signaturen für das Objekt o wie folgt beschreiben:

$$\mathcal{S}_o^{\text{neu}} = \mathcal{S}_o^{\text{alt}} \cup \mathcal{S}_{a^{\text{neu}}|o} \quad \text{bzw.} \quad \mathcal{S}_o^{\text{neu}} = \mathcal{S}_o^{\text{alt}} \cup \mathcal{S}_{r^{\text{neu}}|o}\,. \tag{3.28}$$

Diese Vorgehensweise ist analog zu der Definition von objektbeschreibenden Signaturen in Gl. 3.21. Ein Beispiel für eine komplett neue Relation für eine konkrete Person ist der Wohnort, welcher aus Information im Rahmen eines Dialogs extrahiert werden konnte.

3.3.4 Generierung und Zuordnung von objektbeschreibenden Signaturen

Für eine dauerhafte Speicherung und spätere Wiederverwendung von objektbeschreibenden Signaturen ist es notwendig, eine sichere Zuordnung der realen Objekte zu ihren Objektrepräsentationen im Umweltmodell zu gewährleisten (vgl. [Swe09]). Bei Gegenständen können mit Hilfe einer Objektklasse mehrere gleichartige Objekte identifiziert werden, wohingegen bei Personen jede Klasse genau eine Person eindeutig beschreibt. Die Zuordnung selbst kann auf verschiedene Weisen geschehen, wobei der Mensch in diesem Kontext als objektiver Guide ein sehr großes Vertrauen genießt (vgl. [Swe09]).

In Abb. 3.4 sind zwei Beispiele für den Prozess der Signaturgenerierung und Zuordnung dargestellt. Auf der linken Seite ist das Objekt o_3 dargestellt, welches die Person *Timo* repräsentiert und auf der rechten Seite das Objekt o_6, welches die Person *Benjamin* repräsentiert. Es werden in beiden Fällen sowohl visuelle als auch akustische Sensordaten akquiriert und daraus die entsprechenden Merkmale extrahiert (vgl. Abschnitt 3.3.2). Aufgrund bereits vorhandener Signaturen ist es möglich, festzustellen, dass es sich bei den Objekten o_3 und o_6 um Personen handelt. Jedoch wurden diese zuvor noch nicht erfasst und sind dementsprechend unbekannt.

Während des Prozesses der Signaturgenerierung steht bei Objekt o_3 ein Guide zur Verfügung, welcher mit Hilfe von Mensch-Maschine-Interaktion eine Zuordnung des Objekts zur Klasse Timo vornimmt. Im Beispiel auf der rechten Seite ist dies nicht der Fall. Infolgedessen kann für das Objekt o_3 bei der Generierung der attributsbezogenen Signatur $\mathcal{S}_{a_{\text{Identität}}|o_3}$ eine Zuordnung zur Klasse Timo direkt erfolgen, wohingegen bei Objekt o_6 die Zuordnung zur Klasse Benjamin für die attributsbezogene Signatur $\mathcal{S}_{a_{\text{Identität}}|o_6}$ nicht direkt erfolgen kann und erst zu einem späteren Zeitpunkt nachgeholt wird. Durch die erneute Einbeziehung eines Guides erfolgt dabei eine sichere Zuordnung der zuvor erzeugten, zunächst

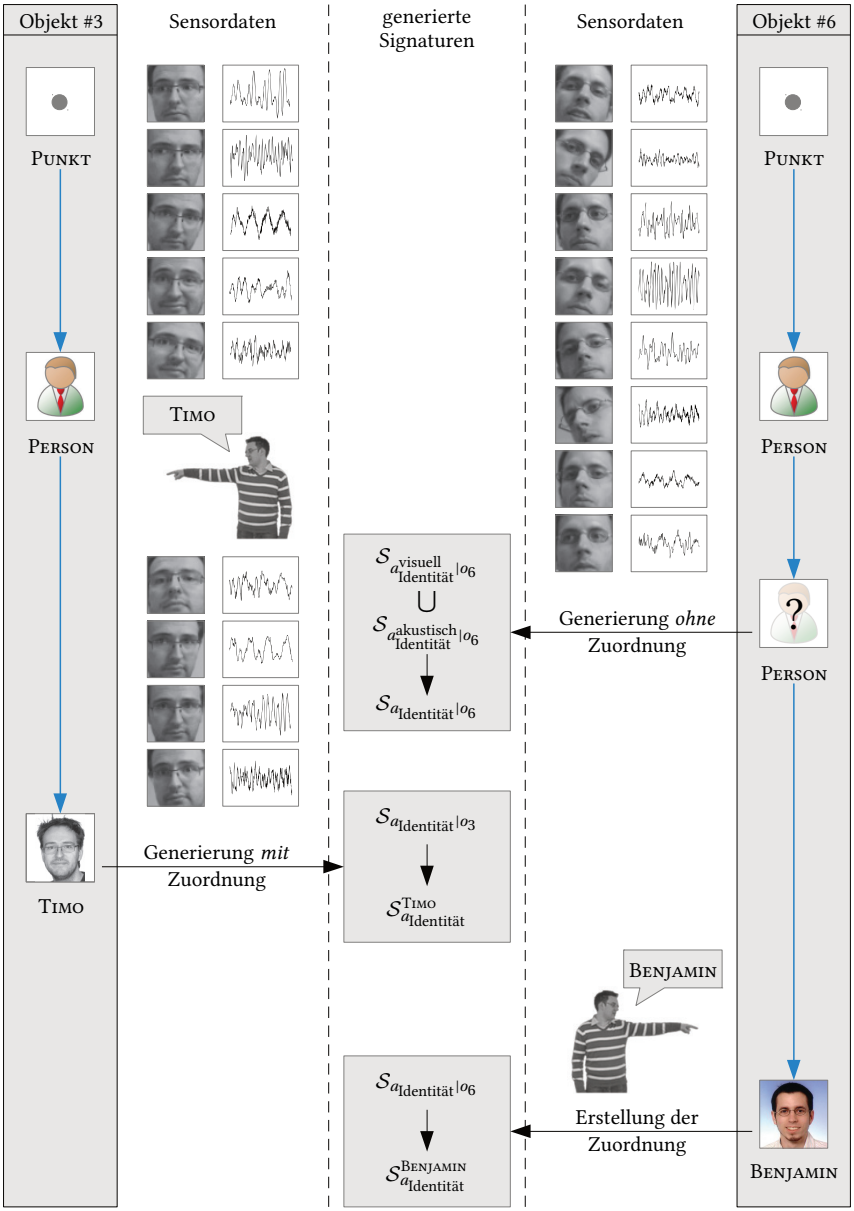

Abb. 3.4: Beispiel für den Prozess der Generierung und Zuordnung von attributsbeschreibenden Signaturen: Jeweils werden sowohl visuelle als auch akustische Merkmale für die Generierung der attributsbeschreibenden Signatur $\mathcal{S}_{a_{\text{Identität}}|o}$ verwendet. Die direkte Zuordnung bei der Generierung (links) und die spätere Zuordnung (rechts) erfolgt dabei jeweils durch einen menschlichen Guide.

temporären attributsbezogenen Signatur $\mathcal{S}_{a_{\text{Identität}}|o_6}$ zu der entsprechenden Klasse BENJAMIN. Die attributsbezogenen Signaturen werden, wie in Gl. 3.25, Gl. 3.26 und Gl. 3.27 beschrieben, aus den Merkmalsvektoren (vgl. Gl. 3.24) bestimmt.

Neben der Zuordnung einzelner attributs- und/oder relationsbezogener Signaturen zu den korrespondieren Objektklassen kann auch eine Zuordnung von kompletten objektbeschreibenden Signaturen zu entsprechenden Klassen erfolgen. Dies ermöglicht die Wiedererkennung von Personen und Gegenständen zu einem späteren Zeitpunkt, unabhängig von den zuvor vorhandenen Objektrepräsentationen im Umweltmodell. Der Übergang von Informationen aus dem Arbeitsgedächtnis des Menschen in das Langzeitgedächtnis stellt einen vergleichbaren Prozess dar.

Eine gewisse Anzahl an Grundsignaturen sollte bei der Realisierung eines solchen Ansatzes (z. B. für einen humanoiden Roboter) immer vorhanden sein. Diese können mit Hilfe von neuen Informationen dynamisch erweitert oder verfeinert werden. Neue Signaturen können ebenfalls im laufenden Betrieb – u. a. aus den gesammelten Sensorinformationen – erstellt werden (vgl. [Swe09]). Weitere Quellen für neue Signaturen, die insbesondere Alltagsgegenstände repräsentieren, können beispielsweise über das Internet bereitgestellt und verteilt werden. Bei humanoiden Robotern kann der Austausch von Signaturen z. B. auch direkt mittels Roboter-Roboter-Interaktion erfolgen (vgl. [Swe09]).

Zusammenfassend lässt sich sagen, dass Signaturen eine wichtige Grundlage für die Umwelterfassung bilden, da diese zur Identifizierung und Wiedererkennung von realen Objekten dienen und somit auch ein Teil des nachfolgend beschriebenen Lebenszyklus eines Objektrepräsentanten sind.

3.4 Lebenszyklus eines Objektrepräsentanten

Der *Lebenszyklus eines Objektrepräsentanten* ist analog zu [Mac10a] bzw. [Swe09] modelliert und beginnt mit der Wahrnehmung eines Objekts durch die Sensoren und der anschließenden Erstellung eines Repräsentanten im Umweltmodell. Im Laufe der Zeit werden mehr Informationen über das Objekt gesammelt und in Form von neuen Attributen und Relationen dargestellt. Der gesamte Lebenszyklus eines Repräsentanten lässt sich in drei Phasen gliedern und wird im Folgenden beschrieben:

1. *Detektionsphase*: In der ersten Phase werden aus den Detektionen (Wahrnehmungen), die bisher keinem im Umweltmodell existierenden Repräsentant zugeordnet werden können, neue Repräsentanten erzeugt. Die Repräsentanten besitzen zunächst nur ein Attribut *Position* sowie die initiale Zugehörigkeit zur Klasse PUNKT. Zu diesem Zeitpunkt ist wenig über das Objekt be-

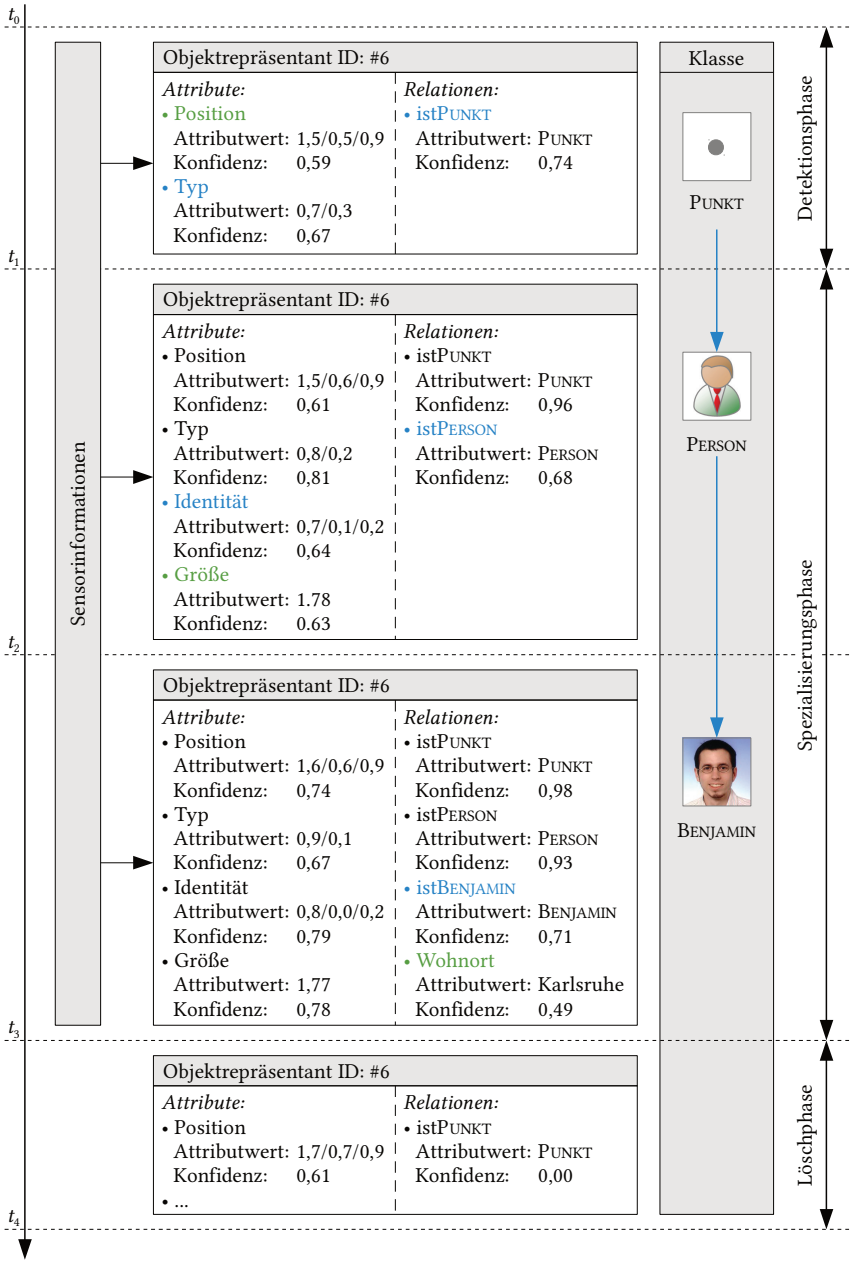

Abb. 3.5: Der Lebenszyklus eines Repräsentanten im Umweltmodell anhand eines Beispiels mit den drei Phasen Detektion, Spezialisierung und Löschung.

kannt und somit wird dieses nur über seine Position repräsentiert. Die Existenz des gesamten Objekts wird durch den Konfidenzwert der Klassenrelation istPunkt modelliert. Durch neue Informationen steigt i. d. R. der Konfidenzwert und bei Überschreiten eines bestimmten Wertes tritt ein Repräsentant in die nächste Phase ein. Sollte hingegen ein Objekt nicht mehr erfasst werden, so sinkt die Sicherheit für dessen Existenz.

2. *Spezialisierungsphase:* Diese Phase ist in der Regel die längste Phase und dient u. a. dazu, mit neuen Sensorinformationen das vorhandene Wissen über ein Objekt zu verbessern. Dabei können zwei Fälle unterschieden werden: Zum einen können die neuen Sensorinformationen mit bereits vorhandenem Wissen in den Attributen und Relationen fusioniert werden und zum anderen können neue Attribute und Relationen erzeugt werden. Dies kann eine Reduzierung des Abstraktionsgrades und/oder das Hinzufügen einer neuen Klassenzugehörigkeit zur Folge haben. Sollten über einen längeren Zeitraum keine Informationen über das Objekt mehr wahrgenommen werden, so tritt der Objektrepräsentant in die Löschphase ein.

3. *Löschphase*: Die letzte Phase des Lebenszyklus eines Objektrepräsentanten ist die Löschung. Aufgrund von ausbleibenden Informationen sinkt der Konfidenzwert für die Klassenrelation istPunkt. Fällt dieser unter eine bestimmte Schwelle und ist zuvor eine erneute Bestätigung der Existenz nicht möglich, so wird infolgedessen der Repräsentant aus dem Umweltmodell gelöscht.

In Abb. 3.5 ist ein Beispiel für den Lebenszyklus eines Objektrepräsentanten mit den drei Phasen dargestellt. Im Mittelpunkt steht die schrittweise Wahrnehmung einer bestimmten Person zu den verschiedenen Zeitpunkten:

t_0: Durch die Sensoren wird ein neues Objekt wahrgenommen und im Umweltmodell wird ein entsprechender Repräsentant mit der Identifikationsnummer (kurz: ID) 6 angelegt. Ein neuer Repräsentant gehört stets zur Klasse Punkt und besitzt, wie zuvor bereits mehrmals angesprochen, das Attribut *Position* und die Klassenrelation istPunkt. Aus den Sensorinformationen und mit Hilfe vorhandener Signaturen lassen sich im Folgenden bereits erste Erkenntnisse über den Objekttyp gewinnen, welche im Attribut *Typ* abgelegt werden. Der Attributwert ist ein Vektor, der in diesem Beispiel die Wahrscheinlichkeiten für die Zugehörigkeit zu den beiden Objekttypen Personen und Gegenstände darstellt.

t_1: Neue Sensorinformationen führen zu einem Anstieg der Konfidenzwerte für die Klassenrelation istPunkt und das Attribut *Typ*. Beim Überschreiten eines Schwellenwertes (beispielsweise $k^{o_6}_{a_{\text{Typ}}} \geq 0{,}75$) kann durch die definierten Wissensabhängigkeiten eine neue Klassenrelation erzeugt werden. Aufgrund der Informationen des Attributs *Typ* kann die Relation istPerson hinzugefügt und somit die Klasse Person in der Klassenhierarchie ergänzt werden.

Des Weiteren kommen die Attribute *Größe* und *Identität* hinzu. Das Attribut *Identität* ist wie das Attribut *Typ* ein Vektor, der im Gegensatz dazu die Übereinstimmung mit bekannten Personen angibt.

t_2: Werden weitere Sensorinformationen akquiriert, so steigen in diesem Beispiel wiederum die Konfidenzwerte für die Klassenrelation istPERSON und das Attribut *Identität*. Ist ein vordefinierter Schwellenwert überschritten, so können neue Attribute und Relationen über die Wissensabhängigkeiten hinzugefügt werden. Die neue Klassenrelation istBENJAMIN wird dabei aufgrund der Informationen im Attribut *Identität* und einer entsprechenden Signatur erzeugt. Zusätzlich wird die Relation *Wohnort* aus den a-priori-Informationen ergänzt.

t_3: Die Person wird im Folgenden eine Zeit lang wahrgenommen, anschließend bleiben neue Sensorinformationen jedoch aus. Infolgedessen sinkt der Konfidenzwert der Klassenrelation istPUNKT. Eine erneute Wahrnehmung des Objekts ist jedoch nicht möglich – die erfasste Person hat z. B. das Haus verlassen –, sodass der Konfidenzwert sinkt und beim Unterschreiten einer zuvor definierten Schwelle der Repräsentant aus dem Umweltmodell gelöscht wird.

t_4: Im Umweltmodell ist der Objektrepräsentant (ID: #6) nicht mehr vorhanden.

3.5 Systemarchitektur

Nach der Einführung eines Umweltmodells, von Abstraktionsebenen, Wissensabhängigkeiten und der Klassenhierarchie, Objektsignaturen sowie dem Lebenszyklus eines Objektrepräsentanten wird in diesem Abschnitt auf die Systemarchitektur näher eingegangen. Diese repräsentiert die übergeordneten Zusammenhänge und den Informationsaustausch in einem autonomen System (vgl. [Mac10a]). In Abb. 3.6 ist eine schematische Darstellung der Systemarchitektur zu sehen, welche die nachfolgend beschriebenen Komponenten zeigt.

3.5.1 Multimodale Sensoren

Für die Erfassung der Umwelt sind Sensoren in einem autonomen System essentiell. Durch die Kombination einer Vielzahl an Sensoren kann die Menge an wahrgenommener Informationen erhöht werden. Dabei können Sensoren unterschiedlichen Typs eingesetzt werden, welche die Umgebung mit verschiedenen Modalitäten wahrnehmen. Beispiele hierfür sind visuelle und akustische Sensoren, welche durch ihre Kombination eine Szene auf mehrere sich ergänzende Arten erfassen können. Auch verschiedene bildgebende Sensoren, wie beispielsweise Farbkameras, Infrarotkameras und Kameras mit Tiefeninformationen (Stereokameras, RGB-D-Sensoren, …), können miteinander kombiniert werden. Die

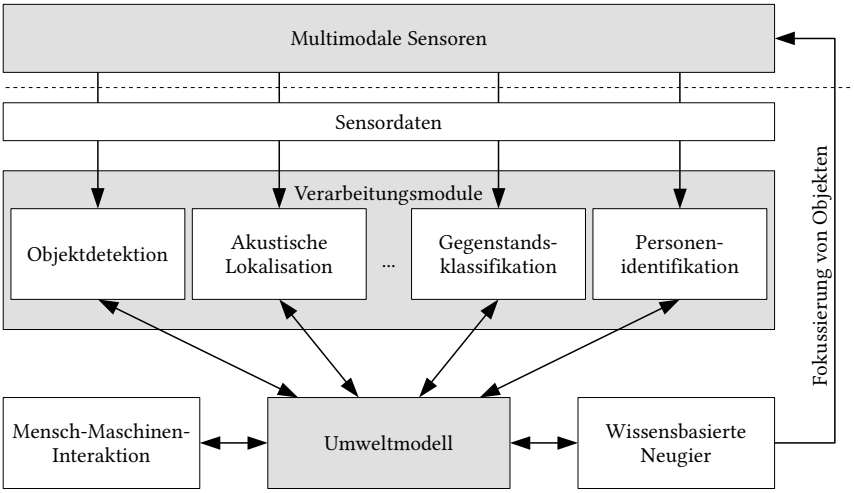

Abb. 3.6: Die Systemarchitektur für die objektzentrierte Umwelterfassung zeigt die Sensoren, die Verarbeitungsmodule, welche aus den Sensordaten das Objektwissen erzeugen, und das Umweltmodell als zentrale Wissensbasis. Zusätzlich interagieren weitere Systembestandteile mit dem Umweltmodell, um spezifische Funktionalitäten zu ermöglichen.

akquirierten Sensordaten werden in der Systemarchitektur an verschiedene spezialisierte Algorithmen für die Weiterverarbeitung übergeben (vgl. Abb. 3.6).

3.5.2 Module

Für die Erfassung der Umgebung ist eine Vielzahl an spezialisierten Algorithmen (z. B. für die Gesichtsdetektion oder die akustische Lokalisation von Schallquellen) notwendig. Jeder einzelne Algorithmus besitzt dabei meist nur die Fähigkeit, einen Teilbeitrag zur Gesamtwahrnehmung (beispielsweise Personenidentifikation) zu liefern. Die Kombination mehrerer Algorithmen gewährleistet i. d. R. eine sich ergänzende und somit zuverlässige Wahrnehmung der kompletten Szene.

Die Algorithmen werden in sogenannten *Modulen* gekapselt, welche für die Erzeugung von Objektwissen verantwortlich sind, d. h. die Generierung eines bestimmten Attribut- oder Relationstyps (z. B. die Größe einer Person). Durch die Modularisierung ergeben sich drei entscheidende Vorteile: Erstes ist durch die Erweiterung des Systems mit neuen Modulen eine Einbeziehung von neuen Verfahren und somit neuen Attributen und/oder Relationen möglich. Der Austausch bereits existierender Module durch neue Module mit verbesserten Algorithmen ermöglicht zweitens eine genauere und/oder schnellere Erfassung der Umwelt. Und drittens kann durch die Verwendung von Modulen eine Unabhängigkeit von

den eingesetzten Sensortypen erreicht werden, da das Umweltmodell das Wissen unabhängig von einem bestimmten Sensortyp repräsentiert. Durch den Austausch von sensorspezifischen Modulen zur Akquise der Sensordaten kann somit jederzeit der Sensor eines anderen Herstellers oder sogar ein anderer Sensortyp verwendet werden. Die im Umweltmodell vorhandenen Attribute und Relationen müssen dafür nicht angepasst werden.

In der Systemarchitektur werden sogenannte *Verarbeitungsmodule* (vgl. Abb. 3.6) genutzt, um aus den akquirierten Sensorinformationen Wissen über die erfassten Objekte zu generieren sowie dieses im Umweltmodell zu hinterlegen. Auch bereits vorhandenes Wissen im Umweltmodell kann bei der Generierung von neuem Objektwissen berücksichtigt werden. Ein Beispiel hierfür ist die Nachverfolgung der Position einer Person. Dabei wird sowohl die aktuelle Position, als auch eine gewisse Historie an Positionen aus der näheren Vergangenheit verwendet. Eine konkrete Ausprägung der Module ist in Abschnitt 6.3 näher erläutert.

In Abschnitt 3.2.3 wurden die Wissensabhängigkeiten definiert, welche nun im Folgenden genutzt werden können, um zu überprüfen, ob die Voraussetzungen für Nutzung eines Moduls für ein bestimmtes Objekt erfüllt sind. Ist dies der Fall, kann das Modul entsprechendes Wissen über ein Objekt erzeugen. Andernfalls sind die notwendigen Informationen nicht vorhanden und das Modul muss bei Bestimmung der Attribute und Relationen nicht berücksichtigt werden, d. h., nur wenn ein Objekt beispielsweise ein Gegenstand ist, wird auch der Gegenstandstyp ermittelt. Diese Vorgehensweise reduziert insgesamt den Umfang der benötigten Rechenressourcen erheblich.

3.5.3 Interaktion mit dem Umweltmodell

Das Umweltmodell stellt die zentrale Wissensbasis in der Systemarchitektur dar (vgl. Abb. 3.6). Die Verarbeitungsmodule und andere Systemkomponenten generieren und nutzen das dort vorhandene Objektwissen sowie die hinterlegten Objektsignaturen. Die Verarbeitungsmodule verwenden dabei Sensorinformationen sowie Objektwissen aus dem Umweltmodell, um neue und/oder detailliertere Objektinformationen zu generieren und im Umweltmodell zu hinterlegen. Über externe Schnittstellen kann das aktuelle Wissen auch außerhalb des Umweltmodells zur Verfügung gestellt werden.

Allgemein kann bei einem autonomen System das vorhandene Wissen über eine *Mensch-Maschine-Schnittstelle* (engl.: human-machine interface; HMI) dem Nutzer, z. B. über eine grafische Benutzerschnittstelle, zur Verfügung gestellt werden. Außerdem können auch neue Informationen, Anweisungen oder Aufgaben über Eingabegeräte wie beispielsweise Maus oder Tastatur in das System eingebracht werden.

Ein humanoider Roboter tritt mit seinem Gegenüber meist in Form einer direkten *Mensch-Maschine-Interaktion* (engl.: man-machine interaction; MMI) in Kontakt. Typische Formen der Interaktion sind hierbei Dialoge und/oder Gesten. Dabei wird bereits akquiriertes Wissen genutzt und ausgetauscht. Zusätzliche Informationen können beispielsweise zum Lösen eines Konflikts oder einer Aufgabe im Zuge dessen eingeholt werden.

Die in Abb. 3.6 dargestellte *wissensbasierte Neugier* (vgl. Kapitel 4) für ein Objekt wird in erster Linie nur durch direkte Interaktion mit dem Umweltmodell bestimmt. Dabei wird auf die darin enthaltenen Informationen (insbesondere die Attribute und Relationen eines Objektrepräsentanten) zurückgegriffen, um daraus Aussagen in Bezug auf die induzierte Neugier zu generieren. Im Rahmen der Exploration einer Szene (vgl. Kapitel 5) wird im späteren Verlauf der Arbeit auch die Neugier für ein Objekt berücksichtigt und es erfolgt somit auch eine Rückkopplung zu den Sensoren im System, da während der Exploration eines bestimmten Objekts dieses fokussiert wird, um mehr Informationen zu erfassen und somit die Neugier zu reduzieren.

3.6 Schlussbetrachtungen

In diesem Kapitel wurden die Grundlagen für die Erfassung von Personen und Gegenständen in einer Szene gelegt. Durch die Definition eines objektzentrierten Umweltmodells als Gedächtnisstruktur für ein autonomes System können Personen und Gegenstände, welche mit verschiedenen Sensoren erfasst werden, einheitlich repräsentiert werden. Die Eigenschaften und Beziehungen dieser Objekte werden in Form von Attributen und Relationen dargestellt. Mit Hilfe von a-priori bekanntem Objektwissen kann sowohl das akquirierte Wissen auf seine Konsistenz geprüft, als auch fehlende Informationen ergänzt werden. Die Generierung von multimodalen Objektsignaturen ermöglicht eine spätere Wiedererkennung von zuvor wahrgenommenen Objekten sowie deren Identifikation.

Der Lebenszyklus eines Objektrepräsentanten ist eine weitere wichtige Komponente der objektzentrierten Umwelterfassung und ist durch die drei Phasen Detektion, Spezialisierung und Löschung definiert. Dabei durchläuft ein Repräsentant i. d. R. mehrere Stufen in der Klassenhierarchie. Ausgehend von einer sehr abstrakten Repräsentation eines Objekts zu Beginn der Erfassung (Klasse: Punkt) werden durch weitere Sensorinformationen immer mehr Eigenschaften erfasst. Durch die Nutzung von Wissensabhängigkeiten und die Einführung von Abstraktionsebenen können Attribute und Relationen gezielt immer detaillierter erfasst werden, sodass die Abstraktion abnimmt und immer weitere Klassenzugehörigkeiten, wie beispielsweise Gegenstand und Mixer, hinzukommen.

Eine modulare Systemarchitektur dient der objektzentrierten Umwelterfassung als Grundlage für die Informationsakquise. Mit Hilfe von multimodalen Sensoren und einer Vielzahl an unterschiedlichen Verarbeitungsmodulen wird aus den Sensorinformationen das Wissen über die Objekte in der Szene abgeleitet und im Umweltmodell hinterlegt. Dieses Wissen kann beispielsweise für die Bewältigung von Aufgaben oder die Mensch-Maschine-Interaktion gezielt genutzt werden. Die objektzentrierte Umwelterfassung dient auch als Grundlage für die nachfolgend in Kapitel 4 beschriebene wissensbasierte Neugier, welche u. a. aus den im Umweltmodell vorhandenen bzw. fehlenden Informationen die Neugier für ein Objekt bestimmt.

Teil II

Interessengetriebene Szenenexploration

4

Wissensbasierte Neugier

In diesem Kapitel wird die wissensbasierte Neugier, welche einem autonomen System die Fähigkeit verleiht, eine spezielle Art von Neugier bei der Wahrnehmung von Objekten zu entwickeln, eingeführt und erörtert. Wichtige Grundlagen hierfür sind neben dem bereits vorhandenen Objektwissen insbesondere auch die noch nicht erfassten Eigenschaften bzw. Informationen über ein Objekt.

Als Ausgangsbasis für die wissensbasierte Neugier dienen Erkenntnisse aus der Psychologie über die Neugier des Menschen sowie das im vorherigen Kapitel beschriebene objektzentrierte Umweltmodell.

4.1 Motivation

Ähnlich wie es auch beim Menschen der Fall ist, muss ein autonomes System in der Lage sein, eine Vielzahl an verschiedenen Aufgaben zu erfüllen. Die konkrete Art der Aufgabenstellung und deren Bewerkstelligung mögen sich durchaus unterscheiden, dennoch sind grundlegende Gemeinsamkeiten und Voraussetzungen vorhanden. Eine davon ist die Wahrnehmung der Umgebung: Hierbei wird unter Zuhilfenahme der Salienz (vgl. Kapitel 2) die Möglichkeit geschaffen, den Fokus auf bestimmte Elemente und Ereignisse in einer Szene zu lenken. Neben dieser sehr instinktiven Orientierungsreaktion besitzt der Mensch ein Streben nach Neuem und Verbogenem, welches allgemein als *Neugier* definiert ist (vgl. [Hof55]).

Im Rahmen der vorliegenden Arbeit wird in Kapitel 5 ein interessengetriebener Explorationsansatz für die Erfassung der aktuellen Umgebung eingeführt und beschrieben. Die wissensbasierte Neugier hat dabei eine sehr wichtige Funktion, da

aufgrund von beschränkten Ressourcen (z. B. aktueller Sensorbereich, Rechenressourcen) während der Erfassung von Gegenständen und Personen in einer Szene eine Selektion bzw. Priorisierung vorgenommen werden muss. Dabei ist das bereits vorhandene Wissen über ein Objekt eine wichtige Einflussgröße, welche zusammen mit dem festgelegten Grad der Detailliertheit bei der Erfassung maßgeblich die Dauer des gesamten Explorationsprozesses beeinflusst.

Die Neugier bietet die Möglichkeit der Priorisierung von Objekten während der Exploration abseits von rein visuellen oder akustischen Merkmalen, wie es bei der Salienz (vgl. Kapitel 2) der Fall ist. Eine ganzheitliche Betrachtung von Objekten ist das Ziel, wobei in der Vergangenheit erworbenes Wissen (vgl. Abschnitt 3.3) mit berücksichtigt wird. So werden bei der Betrachtung eines Objekts beispielsweise Widersprüche mit zuvor wahrgenommenen Informationen untersucht. Des Weiteren werden auch neue Aspekte oder Veränderungen wahrgenommen, sodass infolgedessen vorhandene als auch fehlende Informationen über ein Objekt inkludiert werden. Außerdem werden häufig wahrgenommene Objekte i. d. R. durch den Menschen nicht mehr so intensiv betrachtet wie neue Objekte in einem Raum oder teilweise bzw. gänzlich unbekannte Objekte.

Die zuvor beschriebenen Aspekte der Neugier beim Menschen haben alle den Erwerb oder die Verfeinerung von Wissen zum Ziel. Dieses deckt sich mit den Zielen der Exploration einer Szene. Im Gegensatz dazu gibt es noch andere Formen der Neugier, welche in Abschnitt 4.2 detailliert beschrieben sind. Einige Neugierarten spielen aufgrund ihrer eher negativen Eigenschaften für ein autonomes System eine untergeordnete Rolle bzw. besitzen keine Relevanz. Hierzu zählen u. a. die Sensationslust oder die Neugier des Menschen in Bezug auf Privates.

4.2 Bezüge zur Psychologie

In den nächsten Abschnitten wird aus psychologischer Sicht Bezug auf verschiedene Aspekte der Neugier genommen. Zunächst erfolgt eine historische Einordnung des Begriffs Neugier, bevor anschließend verschiedene Arten von Neugier vorgestellt werden. Abschließend werden die situativen Bedingungen, welche die Entstehung von Neugier fördern, näher beschrieben.

4.2.1 Historische Prägung des Begriffs „Neugier“

Um die Vielschichtigkeit der Neugier und die historische Prägung bzw. Interpretation besser zu verstehen, erfolgt an dieser Stelle eine chronologische Übersicht:

Der Begriff „Neugier“ wurde schon früh geprägt (vgl. [Loe94]). So wird in der Antike von Aristoteles Neugier als „[einen] angeborenen Wunsch nach Information“ (vgl. [Luk06]) definiert und somit als erstrebenswerte Tugend. Später wurde von Marcus Tullius Cicero, einem römischen Politiker, Schriftsteller und Philosoph, die Neugier beschrieben als „innere Liebe zu lernen und Wissen zu erwerben[,] ohne an irgendeinen Profit zu denken“ (vgl. [Luk06]).

Im Mittelalter hingegen traten auch Aspekte der Neugier in den Vordergrund, welche einen deutlichen Bezug zur Sensationslust aufwiesen und diese zusätzlich als menschlichen Trieb beschrieben, ähnlich dem Sexualtrieb. So schriebt Augustinus von Hippo, Kirchenlehrer und Philosoph, in seinen Bekenntnissen (Confessiones), dass die Neugierde ein „eitles und merkwürdiges Begehren nach Wissen“ (vgl. [Luk06]) sei. In diesem Kontext wird auch der Begriff Augenlust geprägt, welcher u. a. deutliche Bezüge zur Sensationslust aufweist. Später, zu Zeiten von Galileo Galilei, in denen sich die Neugier in der Wissenschaft widerspiegelte und im Streben nach neuen Erkenntnissen über die Natur und das Universum mündete, stand dieses Streben in Konflikt mit der Kirche. In diesen wie auch späteren Zeiten lässt sich feststellen, dass „Neugier … unerwünscht sein [kann], da sie bestehende Machtstrukturen in Frage stellt“ (vgl. [Kri76]).

In Zeiten der Aufklärung wurde der Begriff Neugier stärker differenziert. So unterscheidet der Philosoph und Historiker David Hume die Neugier in eine *gute*, bei der die „Liebe zum Wissen“ (vgl. [Luk06]) eine entscheidende Rolle spielt und eine *schlechte*, bei der z. B. die Lust nach Sensationen und Informationen über private Ereignisse von Mitmenschen im Mittelpunkt stehen. Immanuel Kant beschreibt in Anlehnung an körperliche Bedürfnisse die wissensgerichtete Neugier als „Appetit nach Wissen“ (vgl. [Luk06]). Der Philosoph und Germanist Johannes Hoffmeister schreibt Mitte des 20. Jahrhunderts in seinem Wörterbuch der philosophischen Begriffe: „Neugier ist das als ein Reiz auftretende Verlangen, Neues zu erfahren und insbesondere Verborgenes kennenzulernen.“ (vgl. [Hof55]).

Eine weitere Differenzierung der Neugier anhand der verschiedenen Arten erfolgt im nächsten Abschnitt.

4.2.2 Arten von Neugier

In der Neugierforschung wurden im Laufe der Zeit verschiedene Theorien entwickelt (vgl. [Sch08]). So wurde beispielsweise von McDougall oder Murray Neugier als Instinkt, Trieb, Bedürfnis oder Motiv identifiziert (vgl. [McD26], [Mur38]). Cattell beschreibt hingegen die Neugier als Teil der Persönlichkeitseigenschaft *Offenheit für Erfahrungen* (vgl. [Cat50]). Berlyne charakterisiert und differenziert wiederum die Neugier anhand der beiden Faktoren Ausrichtung und Gegenstandsbezug (vgl. [Ber60]).

Der Begriff Neugier subsumiert in den verschiedenen Theorien eine Reihe von Aspekten (vgl. [Loe94]): So ist einerseits die Neugier die treibende Kraft hinter Lernprozessen und bei der Suche nach neuen Erkenntnissen in der Wissenschaft (vgl. [Sch03]). Anderseits ist die Neugier auch bei der Exploration von unbekannten Umgebungen und neuen Situationen ein essentieller Faktor. In anderen Teilbereichen wie der Selbstmotivation oder der Kreativität spielt die Neugier ebenfalls eine wichtige Rolle. Weitere, eher negativ besetzte Ausprägungen wie beispielsweise die Sensationslust seien der Vollständigkeit halber auch noch genannt. Je nach Ausprägung und Aspekt der Neugier kann die Ursache dafür sowohl intrinsisch als auch extrinsisch motiviert sein (vgl. [Day81], [Loe94]).

Im Rahmen der vorliegenden Arbeit wird auf die Herangehensweise von Berlyne (vgl. [Ber60]) zur Unterscheidung der Neugier näher eingegangen. Dieser verknüpft die beiden Aspekte *Ausrichtung* (diversiv oder spezifisch) und *Gegenstandsbezug* (perzeptuell oder epistemisch), sodass sich vier differenzierende Kombinationen und somit Arten der Neugier erzeugen lassen, welche im Nachfolgenden näher erläutert werden:

- *diversiv perzeptuell*: Diese Art der Neugier entsteht bei der ungerichteten Suche bzw. Exploration nach neuen Eindrücken. Dabei steht als Antrieb die Langeweile im Vordergrund und diese tritt meist in Umgebungen mit wenig visuellen Reizen auf.
- *diversiv epistemisch*: Hierbei steht der Wissenserwerb im Vordergrund, welcher wie zuvor aus der Langeweile heraus motiviert ist. Diese Art von Neugier besitzt ebenfalls einen explorierenden und ungerichteten Charakter, die durch den Erwerb von Wissen gekennzeichnet ist.
- *spezifisch perzeptuell*: Diese Ausprägung der Neugier wird durch bestimmte Reizmuster bzw. Sinneseindrücke hervorgerufen, welche im Rahmen einer näheren Betrachtung detailliert analysiert werden.
- *spezifisch epistemisch*: Hierbei wird die Neugier anhand einer konkreten Fragestellung ausgelöst und durch deren Beantwortung aufgelöst. Die Fragestellung kann dabei sowohl intrinsisch als auch extrinsisch motiviert sein.

Berlyne (vgl. [Ber60], [Ber66]) verwendet für die ersten beiden Ausprägungen auch den Begriff der Exploration anstatt der Neugier, da die Ausrichtung diversiv ist und mit der Suche nach Neuem verbunden ist. Hierbei wird durch Perzeption von neuen Reizen die auftretende Langeweile reduziert bzw. abgebaut. Ein mittleres Reizniveau, bei dem Reize vorhanden sind, jedoch nicht durch ihre Anzahl überfordern, wird dabei als angenehm empfunden (vgl. [Rot11]). Die diversiven Ausrichtungen der Neugier werden größtenteils dem Verhalten von Tieren zugeordnet, wohingegen die Varianten der spezifischen Neugier in weiten Teilen dem Menschen zugeordnet sind.

Im Rahmen der vorliegenden Arbeit wird die *wissensbasierte Neugier* für autonome Systeme definiert und eingeführt, welche die Varianten der spezifischen Neugier vereint. Dabei wird die Aufgabenstellung der Wissensakquisition im Rahmen der Exploration einer Szene mit der Fragestellung der Wichtigkeit einzelner Objekte kombiniert. Die Neugier dient dabei als ein wichtiger Faktor bei der Priorisierung der zu explorierenden Objekte. Die im nachfolgenden Abschnitt definierten situativen Bedingungen der Neugier fungieren in diesem Zusammenhang als Ausgangsbasis für die wissensbasierte Neugier.

4.2.3 Situative Bedingungen

Die sogenannten situativen Bedingungen sind die fundamentalen Größen für die Bestimmung der wissensbasierten Neugier und somit auch der Fokussierung auf einzelne Fragestellungen, Sachverhalte oder Objekte. In zahlreichen experimentellen Studien mit äußeren und objektivierbaren Gegebenheiten wurden die situativen Bedingungen von Berlyne bestimmt (vgl. [Ber74]). Diese sind *Neuartigkeit*, *Komplexität*, *Ungewissheit* sowie *Konflikt* und stehen untereinander in Beziehung. Im Folgenden werden die einzelnen Bedingungen wie bei Lukesch (vgl. [Luk06]) zusammenfassend vorgestellt:

- Die *Neuartigkeit* beschreibt die Konstellation, bei der eine Gegebenheit von einer vertrauten abweicht. Dabei ist zum einen ein Reiz oder auch eine Situation neuartig, wenn dieser bzw. diese zuvor noch nicht wahrgenommen wurde (*völlige Neuartigkeit*). Zum anderen besitzt ein Reiz eine unterschiedliche Neuartigkeit, je nachdem, ob dieser in der jüngsten Vergangenheit (*kurzfristige Neuartigkeit*) oder vor ein paar Tagen oder sogar Monaten (*langfristige Neuartigkeit*) zuletzt erfasst wurde. Eine weitere Unterscheidung wird zwischen *absoluter* und *relativer Neuartigkeit* getroffen. Dabei wurde bei Erstem (wie zuvor schon angesprochen) etwas noch nie wahrgenommen, wohingegen sich bei Letzterem die Wahrnehmung aus bereits bekannten Aspekten zusammensetzt bzw. eine Variante von bereits Bekanntem ist.
- Die *Komplexität* stellt den „Grad an Vielfalt oder Verschiedenheit in einem Reizmuster“ dar (vgl. [Luk06]). So nimmt die Komplexität mit einer steigenden Anzahl an verschiedenen Elementen in einem Reizmuster zu. Die Komplexität sinkt hingegen, wenn einzelne Elemente zu einer Einheit zusammengefasst werden. Ein Beispiel hierfür wären verschiedenartige Linien, die kombiniert einen Buchstaben oder ein Objekt ergeben. Dies ist jedoch abhängig vom vorhandenen Wissen und den Erfahrungen einer Person. Mit sogenannten Gestaltgesetzen (vgl. [Met53]) kann die Zusammengehörigkeit bzw. Unterscheidung einzelner Elemente bestimmt werden. Hierzu zählen unter anderem die Prägnanz, Nähe, Ähnlichkeit und Kontinuität. Dabei wird

versucht, möglichst einfache, geschlossene, symmetrische und gleichartige Formen entstehen zu lassen.

- Die *Ungewissheit* entsteht durch den Vergleich von Reizen mit Erwartungen, welche tendenziell unterschiedlich sind. Des Weiteren können Wahrnehmungen auch mehrere Erwartungen gleichzeitig aktivieren und somit zur Ungewissheit führen. Berlyne (vgl. [Ber74]) bezieht sich dabei auf zwei unterschiedliche Aspekte: zum einen ein Vergleich des Wertebereichs für ein Merkmal und zum anderen die Wahrscheinlichkeit für das Eintreten bzw. Nichteintreten eines Ereignisses. Die Ungewissheit ist maximal, wenn die Wahrnehmung mit dem Wertebereich eines Merkmals nicht übereinstimmt sowie das Eintreten und das Nichteintreten gleichwahrscheinlich sind.
- Der *Konflikt* beschreibt Reaktionen, die ein Objekt oder eine Sache hervorruft, welche nicht miteinander vereinbar sind (z. B. die Stimme und das Erscheinungsbild einer Person). Der Konflikt steigt, je ähnlicher die konkurrierenden Reaktionstendenzen sind, je stärker die konkurrierenden Reaktionstendenzen sind und je höher die Anzahl an konkurrierenden Reaktionstendenzen ist. Die Reaktionstendenz lässt sich dabei beschreiben als eine Tendenz für die Interpretation einer Wahrnehmung oder eines Sachverhaltes aufgrund von bestimmten Merkmalen. Der Mensch besitzt neben den zuvor angesprochenen Aspekten auch eine Reihe an gedanklichen Konflikten wie beispielsweise Zweifel, Perplexität, Widerspruch, gedankliche Inkongruenz, Verwirrung und Irrelevanz (vgl. [Ber74]).

Nach Untersuchungen von Wentworth und Witryol (vgl. [Wen90]) stehen die ersten drei Bedingungen in Beziehung zueinander. Die Ungewissheit nimmt dabei den höchsten Stellenwert ein, gefolgt von der Neuartigkeit und der Komplexität. Der Konflikt wurde im Rahmen der experimentellen Studien nicht explizit betrachtet.

4.3 Systematische Realisierung der wissensbasierten Neugier

Zunächst werden die Anforderungen an die wissensbasierte Neugier definiert sowie die grundlegende Vorgehensweise bei der Umsetzung erläutert, bevor die systematische Realisierung für ein autonomes System vorgestellt wird. Anschließend wird die Modellierung der Teilaspekte der Neugier sowie abschließend die Bestimmung der Gesamtneugier für einzelne Objekte detailliert beschrieben.

4.3.1 Anforderungen an die wissensbasierte Neugier

Im Folgenden werden die Anforderungen an die Realisierung der wissensbasierten Neugier für ein autonomes System, wie es beispielsweise ein humanoider Roboter ist, definiert:

- *Übertragung der Erkenntnisse aus der Psychologie:* Die zuvor beschriebenen situativen Bedingungen für die Neugier, d. h. Neuartigkeit, Komplexität, Ungewissheit und Konflikt (vgl. Abschnitt 4.2), sollen bei der Realisierung berücksichtigt und systematisch in einem autonomen System nachgebildet werden.
- *Verwendung eines objektzentrierten Umweltmodells:* In einem Umweltmodell (vgl. Abschnitt 3.1) sind die aktuell erfassten Informationen über die Objekte in der aktuellen Szene enthalten. Dieses Wissen soll als Grundlage zur Bestimmung der wissensbasierten Neugier dienen.
- *Nutzung von a-priori-Wissen:* Neben den im Umweltmodell aktuell vorhandenen Objektinformationen soll auch das dort abgelegte a-priori-Objektwissen bei der Realisierung der wissensbasierten Neugier berücksichtigt werden.
- *Kriterium zur Objektpriorisierung:* Die wissensbasierte Neugier soll die Wichtigkeit der Objekte widerspiegeln und kann somit als ein Kriterium für die Priorisierung der Objekte während der Exploration dienen (vgl. Kapitel 5). Dabei sind Objekte, welche eine geringe Neugier hervorrufen, dem autonomen System bereits hinreichend bekannt und weisen keinen Konflikt auf, wohingegen eine hohe Neugier u. a. auf Objektaspekte hinweist, welche teilweise bzw. völlig unbekannt sind oder einen Konflikt induzieren.

4.3.2 Umsetzung der wissensbasierten Neugier

Für die Realisierung der wissensbasierten Neugier sollen die in Abschnitt 4.2 beschriebenen vier Teilaspekte berücksichtigt werden:

Zur Bestimmung der *Neuartigkeit* werden Informationen aus dem Umweltmodell – in Bezug auf die Zeitpunkte und die Häufigkeit der Wahrnehmung – verwendet. Die Neuartigkeit soll, wie beim Menschen auch, bei häufigerer Wahrnehmung eines Objekts immer weiter abnehmen. Ein zweiter Punkt ist die Vollständigkeit der Wahrnehmung: Wurden einige Eigenschaften eines Objekts noch nicht erfasst, so ist entsprechend die Neuartigkeit höher als bei anderen Objekten.

Die *Komplexität* eines Objekts setzt sich aus den visuellen Eigenschaften zusammen. Hierbei werden die Textur und die Form des Objekts berücksichtigt. Des Weiteren sind auch die Präsenz im Sichtfeld sowie eventuell vorhandene Signalfarben Komponenten zur Bestimmung der Komplexität. In Anlehnung an die Ge-

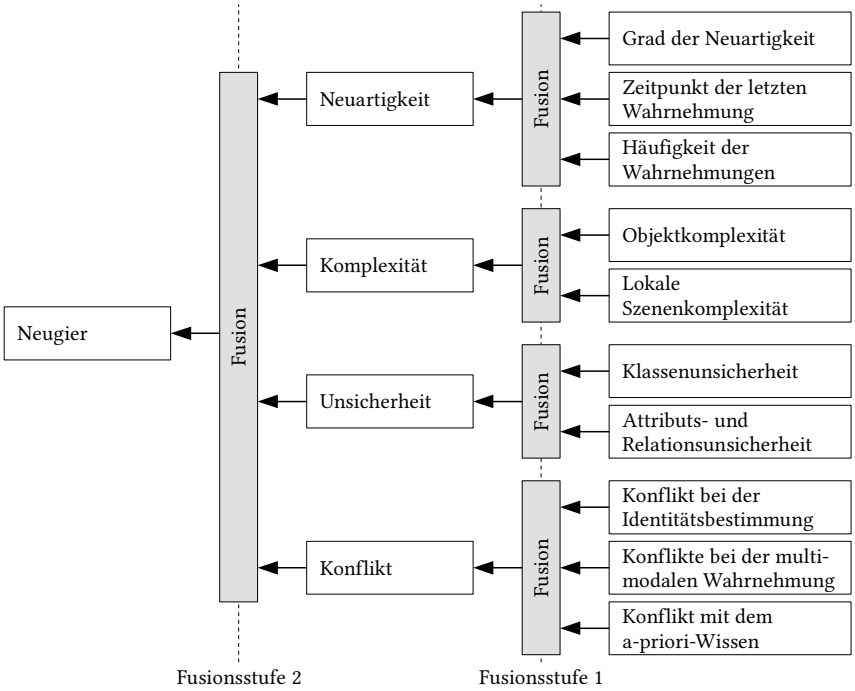

Abb. 4.1: Zusammenhang der Aspekte und Teilaspekte der Neugier sowie die hierarchische Fusion zur Bestimmung der wissensbasierten Neugier.

staltgesetze (vgl. [Met53]) wird die räumliche Anordnung der in der Szene vorhandenen Objekte berücksichtigt.

Im Rahmen der Arbeit wird die Ungewissheit beim Menschen durch die Unsicherheit in Bezug auf verschiedene Objekteigenschaften modelliert. Zur Bestimmung der *Unsicherheit* werden die Konfidenzwerte einzelner Attribute und Relationen der Objekte betrachtet. Des Weiteren werden auch der Konfidenzwert des Attributs *Identität* bzw. *Typ* eines Objekts berücksichtigt. Hierbei drückt der im Umweltmodell definierte Konfidenzwert ein Sicherheitsmaß aus und stellt somit eine entgegengesetzte Aussage zur Unsicherheit dar.

Der letzte Aspekte der Neugier ist der *Konflikt* und wird durch drei sich ergänzende Faktoren definiert: Ein Konflikt liegt beispielsweise vor, wenn ein und dieselbe Person mehr als einmal in einer Szene vorhanden ist, d. h., mehreren Personen wird die gleiche Identität zugewiesen. Des Weiteren ergibt sich ein Konflikt, falls die Wahrnehmungen bzw. Informationen aus unterschiedlichen Quellen stammen und eine sehr starke Differenz bzgl. der Aussage aufweisen, d. h., wenn z. B. die akustische Wahrnehmung nicht mit der visuellen überein-

stimmt. Eine weitere Ursache für einen Konflikt liegt vor, falls die akquirierten Objektinformationen sich von dem als fix angenommenem Wissen unterscheiden, d. h., entweder sind die a-priori definierten Informationen möglicherweise falsch bzw. lückenhaft oder die aktuelle Wahrnehmung ist fehlerhaft. A-priori-Informationen werden i. d. R. gezielt zusammengestellt und für ein autonomes System bereitgestellt. Die Informationen müssen vorab auf Konsistenz und Fehlerfreiheit überprüft werden.

Alle Aspekte der Neugier bestehen wie zuvor beschrieben aus verschiedenen Teilkomponenten, die mit Hilfe von geeigneten Fusionsfunktionen hierarchisch zur wissensbasierten Neugier fusioniert werden (vgl. Abb. 4.1). Dies geschieht für jedes erfasste Objekt im Umweltmodell separat.

4.3.3 Neuartigkeit

Für die Neuartigkeit, die ein Objekt beim Betrachter hervorruft, sind im Wesentlichen drei Einflussfaktoren entscheidend: Als Erstes ist die Vollständigkeit der Wahrnehmung ausschlaggebend. Ein Objekt kann beispielsweise mit mehreren Modalitäten wahrgenommen werden, sodass bei einer fehlenden akustischen Wahrnehmung eine nur unvollständige Beschreibung möglich ist. Fehlende Attribute und Relationen sind somit typisch für die Neuartigkeit eines Objekts. Als Zweites spielt der Zeitpunkt der letzten Wahrnehmung eine wichtige Rolle, da Objekte, die über Monate oder Jahre hinweg nicht wahrgenommen wurden, eine priorisierte Wahrnehmung verursachen. Nur Objekte, welche zuvor noch nie erfasst wurden, besitzen eine noch höhere Priorität. Als Letztes ist die Häufigkeit, mit der Objekte wahrgenommen werden, entscheidend, da vielfach wahrgenommene Objekte im Alltag i. d. R. eine geringere Auffälligkeit besitzen als Objekte, welche nur selten wahrgenommen werden.

Die Neuartigkeit besteht im Rahmen der vorliegenden Arbeit aus dem Grad der Neuartigkeit π_o^{Grad}, dem Zeitpunkt der letzten Wahrnehmung $\pi_o^{\text{Zeitpunkt}}$ sowie der Häufigkeit der Wahrnehmung eines Objekts $\pi_o^{\text{Häufigkeit}}$ und ist definiert als

$$\pi_o := f_\pi\left(\pi_o^{\text{Grad}}, \pi_o^{\text{Zeitpunkt}}, \pi_o^{\text{Häufigkeit}}\right) \quad \text{mit } o \in \mathcal{O}. \tag{4.1}$$

Mit Hilfe einer geeigneten Fusionsfunktion f_π kann aus diesen Teilaspekten die Gesamtneuartigkeit bestimmt werden. Eine Diskussion über verschiedene Fusionsfunktionen erfolgt in Abschnitt 4.3.7.

Grad der Neuartigkeit

Die Neuartigkeit hängt nicht nur davon ab, ob ein Objekt bekannt oder unbekannt ist, sondern auch davon, wie viel Wissen gesammelt wurde. Das Wissen

über ein Objekt ist in dessen Attributen $\mathcal{A}_o$ und Relationen $\mathcal{R}_o$ repräsentiert. Somit lässt sich der Grad der Neuartigkeit in deren Abhängigkeit definieren:

$$\pi_o^{\text{Grad}} := f_o^{\text{Grad}}(\mathcal{A}_o, \mathcal{R}_o)\,. \tag{4.2}$$

Dabei werden mögliche Attribute und Relationen, welche für ein Objekt noch nicht erfasst wurden, dennoch generiert und entsprechend gekennzeichnet (*„nicht erfasst“*). Aufgrund der unterschiedenen Bedeutsamkeit von einzelnen Attributen bzw. Relationen ist es notwendig, diese bei der nachfolgenden Bestimmung des Neuartigkeitsgrades unterschiedlich zu gewichten. Dazu wurde bereits in Gl. 3.3 bzw. Gl. 3.4 die Priorität von Attributen und Relationen definiert. Die Wahl der Priorität ist abhängig von der jeweiligen Eigenschaft und lässt sich über Statistiken oder auch anhand von Anwendungsszenarien im Vorfeld festlegen. Auch eine Anpassung zur Laufzeit ist denkbar, falls bestimmte Attribute oder Relationen auf jeden Fall und besonders sicher erfasst werden sollen, z. B. um einen Auftrag zu erfüllen. Sollte dies nicht möglich sein, so ist eine Gleichgewichtung aller Prioritäten eine sinnvolle Alternative.

Der Grad der Neuartigkeit (engl.: level of novelty; kurz: LN) kann unter Berücksichtigung der Existenz eines Attributes oder einer Relation (e_{a_i} bzw. e_{r_j}) in Kombination mit der zuvor angesprochenen Priorität (ρ_{a_i} bzw. ρ_{r_j}) wie folgt definiert werden:

$$\overline{g_o^{\text{LN}}}(\mathcal{A}_o, \mathcal{R}_o) = \frac{\sum_{i=1}^{I} e_{a_i}\rho_{a_i} + \sum_{j=1}^{J} e_{r_j}\rho_{r_j}}{\sum_{i=1}^{I} \rho_{a_i} + \sum_{j=1}^{J} \rho_{r_j}} \quad \text{mit } a_i \in \mathcal{A}_o; r_j \in \mathcal{R}_o; I, J \in \mathbb{N} \tag{4.3}$$

$$g_o^{\text{LN}}(\mathcal{A}_o, \mathcal{R}_o) = 1 - \overline{g_o^{\text{LN}}}(\mathcal{A}_o, \mathcal{R}_o) \tag{4.4}$$

Die Indikatorfunktionen für die Existenz eines Attributs e_{a_i} und einer Relation e_{r_j} sind unter Zuhilfenahme des jeweiligen Attributwerts w_{a_i} bzw. Relationswerts w_{r_j} definiert als:

$$e_{a_i} = \begin{cases} 0, & \text{falls } w_{a_i} = \text{„nicht erfasst“} \\ 1, & \text{sonst} \end{cases} \tag{4.5}$$

$$e_{r_j} = \begin{cases} 0, & \text{falls } w_{r_j} = \text{„nicht erfasst“} \\ 1, & \text{sonst} \end{cases}\,. \tag{4.6}$$

In Gl. 4.3 wird dabei für die Vergleichbarkeit verschiedener Objektrepräsentanten die Summe aller Attributs-/Relationsprioritäten als ein Normierungsfaktor eingeführt, sodass der Funktionswert im abgeschlossenen Intervall $[0, 1]$ liegt.

Unter Berücksichtigung des Zusammenhangs zwischen den Wissensabhängigkeiten, den Abstraktionsebenen und der Klassenhierarchie für ein Objektrepräsentanten

sentant im Umweltmodell (vgl. Abschnitt 3.2.3) lässt sich der Grad der Neuartigkeit aus den Attributen und Relationen der jeweiligen Ebene h in der Klassenhierarchie bestimmen (engl.: hierarchical level of novelty; kurz: HLN):

$$g_o^{\overline{\text{HLN}}}(\mathcal{A}_o, \mathcal{R}_o) = \min\left\{ \bigcup_{h=1,\ldots,H} \left\{ \frac{\sum_{i=1}^{I} e_{a_i}^h \rho_{a_i}^h + \sum_{i=1}^{I} e_{r_i}^h \rho_{r_i}^h}{\sum_{i=1}^{I} \rho_{a_i}^h + \sum_{i=1}^{I} \rho_{r_i}^h} \right\} \right\} \tag{4.7}$$

$$g_o^{\text{HLN}}(\mathcal{A}_o, \mathcal{R}_o) = 1 - g_o^{\overline{\text{HLN}}}(\mathcal{A}_o, \mathcal{R}_o)\,. \tag{4.8}$$

Der Grad der Neuartigkeit wird mittels der Existenz und Priorität der jeweiligen Attribute bzw. Relationen eines Objektrepräsentanten für jede Ebene h in der Hierarchie getrennt bestimmt. $e_{a_i}^h$ bzw. $e_{r_j}^h$ stellen die Existenz für das i-te bzw. j-te Element der Attributs-/Relationsmenge auf der h-ten Ebene dar. Analoges gilt für die jeweiligen Prioritäten $\rho_{a_i}^h$ und $\rho_{r_j}^h$.

Abschließend sollte bei der Definition der Gesamtfunktion noch der Spezialfall berücksichtigt werden, dass das Objektrepräsentant als „Unbekannt“ klassifiziert wird. Ein entsprechender Eintrag ist im Attributwert des virtuellen Attributs a_{Klasse} zu finden und sollte deshalb zu einer vollkommenen Neuartigkeit führen. Die Funktion zur Bestimmung des Neuartigkeitsgrads lässt sich somit unter Zuhilfenahme des Attributwerts $w_{a_{\text{Klasse}}}$ und von Gl. 4.4 bzw. Gl. 4.8 definieren als

$$f_o^{\text{Grad}}(\mathcal{A}_o, \mathcal{R}_o) = \begin{cases} 1, & \text{falls } w_{a_{\text{Klasse}}} = \text{„Unbekannt“} \\ g_o^{\text{LN}}(\mathcal{A}_o, \mathcal{R}_o), & \text{sonst} \end{cases} \tag{4.9}$$

bzw.

$$f_o^{\text{Grad}}(\mathcal{A}_o, \mathcal{R}_o) = \begin{cases} 1, & \text{falls } w_{a_{\text{Klasse}}} = \text{„Unbekannt“} \\ g_o^{\text{HLN}}(\mathcal{A}_o, \mathcal{R}_o), & \text{sonst} \end{cases}\,. \tag{4.10}$$

Zeitpunkt der letzten Wahrnehmung

Ein weiterer wichtiger Teilaspekt der Neuartigkeit ist der Zeitpunkt der letzten Wahrnehmung eines Objekts. Dieser lässt sich bestimmen über

$$\pi_o^{\text{Zeitpunkt}} := f_o^{\text{Zeitpunkt}}\left(\Delta t_o^{\text{Zeitpunkt}}\right) \tag{4.11}$$

mit

$$\Delta t_o^{\text{Zeitpunkt}} = \begin{cases} \infty, & \text{falls } \mathcal{Z}_o \setminus t = \emptyset \\ t - \max(\mathcal{Z}_o \setminus t), & \text{sonst} \end{cases}\,, \tag{4.12}$$

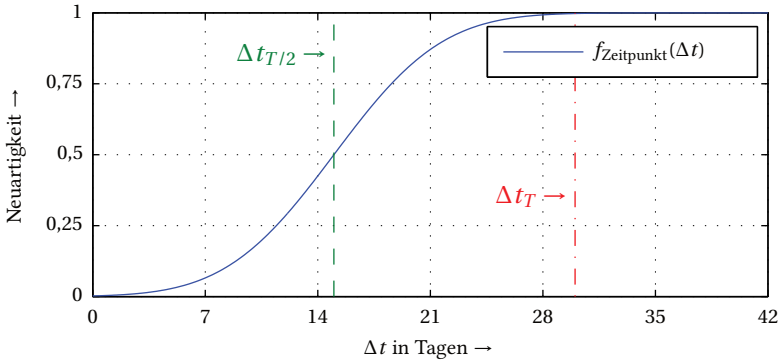

Abb. 4.2: Zusammenhang zwischen dem Zeitpunkt der letzten Wahrnehmung und der Neuartigkeit eines Objekts mit $\Delta t_{T/2}$ als Zeitpunkt, an dem die Neuartigkeit auf 0,5 ansteigt.

wobei $\Delta t_o^{\text{Zeitpunkt}}$ den Zeitraum seit der letzten Wahrnehmung und t den aktuellen Zeitpunkt darstellt. $\mathcal{Z}_o$ sind die Zeitpunkte der vorherigen Wahrnehmung des im Umweltmodell repräsentierten Objekts o (vgl. Abschnitt 3.1.1). Sollte das Objekt zuvor noch nicht wahrgenommen sein, so wird per Definition ein unendlich langer Zeitraum festgelegt. Um eine normierte Aussage zu erhalten, wird der Zeitraum seit der letzten Wahrnehmung durch eine Hilfsfunktion $f_o^{\text{Zeitpunkt}}(\cdot)$ in ein geeignetes Maß überführt. Dazu ist es notwendig, den Verlauf der Neuartigkeit in Bezug auf den Zeitraum festzulegen. Dies geschieht in der vorliegenden Arbeit mit Hilfe der Fehlerfunktion erf(x) (in Anlehnung an nicht-exponentielle Lernkurven; vgl. [Gal04]) sowie der zeitlichen Konstante $\Delta t_{\text{T}/2}$:

$$f_o^{\text{Zeitpunkt}}\left(\Delta t_o^{\text{Zeitpunkt}}\right) = \frac{1}{2}\left(1 - \text{erf}\left(2 - 2 \cdot \frac{\Delta t_o^{\text{Zeitpunkt}}}{\Delta t_{\text{T}/2}}\right)\right) \tag{4.13}$$

mit

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, \text{d}t \,. \tag{4.14}$$

Der Verlauf der Kurve ist beispielhaft in Abb. 4.2 dargestellt und zeigt den Zusammenhang mit der Konstanten $\Delta t_{\text{T}/2}$. Diese gibt den Zeitraum in Tagen an, bis die Neuartigkeit auf 0,5 angestiegen ist. Nach einer fehlenden Observation (d. h. Wahrnehmung) von Δt_{T} Tagen tritt eine nahezu vollständige Neuartigkeit auf. Die Kurve wurde so gewählt, dass in den ersten Tagen die Neuartigkeit nur sehr langsam ansteigt und erst später deutlich zunimmt. Dies fördert die erneute Analyse von länger nicht mehr wahrgenommen Objekten deutlich und reduziert die Neuartigkeit bei häufiger wahrgenommenen Objekten.

Statistik über die Häufigkeit der Detektion

Neben dem Zeitraum seit der letzten Wahrnehmung eines Objekts spielt auch die Häufigkeit eine wichtige Rolle, da diese ein Indikator für das vorhandene Wissen über ein Objekt sein kann. Über Objekte, die in der Vergangenheit in einer gewissen Häufigkeit wahrgenommen wurden, liegt i. d. R. eine größere Menge an Daten vor, als bei anderen Objekten die seltener erfasst wurden. Im Folgenden wird die Verteilung des Auftretens näher untersucht und mit Hilfe von

$$\pi_o^{\text{Häufigkeit}} := f_o^{\text{Häufigkeit}}(\Delta\mathcal{Z}_o), \quad \text{mit } \Delta\mathcal{Z}_o = \bigcup_{t_k \in \mathcal{Z}_o} \{t_k - t\} \tag{4.15}$$

in eine konkrete Aussage überführt. $\Delta\mathcal{Z}_o$ stellt dabei die Menge an Wahrnehmungen eines Objekts mit relativem Zeitbezug t dar. Die Zeitpunkte in der Vergangenheit werden mit einem negativen Vorzeichen gekennzeichnet. Für die nachfolgende Bestimmung der Häufigkeit wird über die Menge an relativen Zeitpunkten $\Delta\mathcal{Z}_o$ ein Histogramm gebildet:

$$\boldsymbol{h}^{\text{Häufigkeit}}(\Delta\mathcal{Z}_o) = \text{Histogramm}(\Delta\mathcal{Z}_o, \boldsymbol{\tau}), \quad \text{mit } \boldsymbol{\tau} = (-\Delta t_\text{T}, \ldots, -1, 0), \tag{4.16}$$

wobei die Werte zeitlich diskretisiert, d. h. auf feste Zeitpunkte $\boldsymbol{\tau}$ beschränkt werden. Das entsprechende Intervall ist über die maximale Neuartigkeitsspanne Δt_T festgelegt. Für die Auswertung werden zwei Maße definiert, welche zum einen das mittlere Auftrittsverhalten und zum anderen die zeitliche Struktur der Wahrnehmung widerspiegeln.

Die mittlere Anzahl an Wahrnehmungen (engl: average number of detections; AD) in einem Beobachtungszeitraum wird als Maß für die generelle Wahrnehmung eines Objekts verwendet und lässt sich wie folgt definieren:

$$g_o^{\text{AD}}(\Delta\mathcal{Z}_o) = \frac{1}{\Delta t_\text{T}} \sum_{i=-\Delta t_\text{T}}^{0} h_i^{\text{Häufigkeit}}(\Delta\mathcal{Z}_o). \tag{4.17}$$

Im Folgenden ist es notwendig, eine Gewichtungsfunktion für die mittlere Anzahl an Wahrnehmungen einzuführen, da zunächst wenige Wahrnehmungen nur einen geringen Einfluss haben sollen und erst im weiteren zeitlichen Verlauf mit einer immer größer werdenden Anzahl an Wahrnehmungen (d. h. mehr Informationen) der Einfluss zunehmen soll, bis er schließlich gegen sein Maximum konvergiert. Die Fehlerfunktion (siehe Gl. 4.14) beschreibt, wie andere Sigmoidfunktionen auch, dieses Verhalten sehr gut. Die gewichtete mittlere Anzahl an Wahrnehmungen (engl.: weighted average number of detections; kurz: WAD) ist wie folgt definiert:

$$g_o^{\text{WAD}}(\Delta\mathcal{Z}_o) = \frac{1}{2}\left(1 - \text{erf}\left(2 - 4 \cdot \frac{g_o^{\text{AD}}(\Delta\mathcal{Z}_o)}{\zeta^{\text{WAD}}}\right)\right). \tag{4.18}$$

In diesem Zusammenhang wird ein zusätzlicher Parameter ζ^{WAD} eingeführt, welcher den Punkt definiert, an dem der Funktionswert sein Maximum nahezu erreicht hat. Um nun eine Aussage in Bezug auf die Neuartigkeit zu erhalten, muss die Bedeutung der Funktion umgekehrt werden

$$g_o^{\overline{\mathrm{WAD}}}(\Delta\mathcal{Z}_o) = 1 - g_o^{\mathrm{WAD}}(\Delta\mathcal{Z}_o)\,. \tag{4.19}$$

In Abb. 4.3 ist ein Beispiel mit einem Gegenstand und einer Person zu sehen. Die zeitlichen Wahrnehmungen sind dabei als Histogramm (vgl. Abb. 4.3; oben) dargestellt für $\Delta t_{\mathrm{T}} = 30\,\mathrm{Tage}$. Die Ergebnisse für die Gewichtung mit $\zeta^{\mathrm{WAD}} = 0{,}5$ sind darunter zu erkennen (vgl. Abb. 4.3; Mitte). Es lässt sich erkennen, dass die Person eher selten wahrgenommen wurde – und zudem vor mehr als zwei Wochen das letzte Mal. ζ^{WAD} wurde so gewählt, dass das Objekt im Schnitt jeden zweiten Tag erkannt werden müsste, um nicht „Neu“ zu sein. Im Falle von jedem vierten Tag steigt die Neuartigkeit auf 0,5 an. Dies ist in Fällen, wie z. B. für einem humanoiden Haushaltsroboter, sinnvoll, kann jedoch je nach Anwendungsdomäne angepasst werden (i. d. R. verlängert werden).

Eine weitere Möglichkeit, das zuvor gewonnene Histogramm (vgl. Gl. 4.16) auszuwerten, ist eine Untersuchung der zeitlichen Struktur der Wahrnehmung. Dies geschieht über die zeitliche Akkumulierung des Histogramms und beschreibt somit die zeitlich kumulierten Wahrnehmungen (engl.: cumulated number of detections; kurz: CD) und lässt sich bestimmen durch

$$g_o^{\mathrm{CD}}(\tau_i;\Delta\mathcal{Z}_o) = \frac{1}{\Delta t_{\mathrm{T}}} \sum_{j=0}^{\tau_i} h_j^{\mathrm{Häufigkeit}}(\Delta\mathcal{Z}_o)\,. \tag{4.20}$$

Hierbei sind der Parameter $\boldsymbol{\tau}$ und das Histogramm $\boldsymbol{h}^{\mathrm{Häufigkeit}}(\Delta\mathcal{Z}_o)$ wie in Gl. 4.16 definiert. Die Fläche unter der Kurve (engl.: integrated cumulated number of detections; kurz: ICD) definiert ein Maß für die zeitliche Struktur der Wahrnehmungen:

$$g_o^{\mathrm{ICD}}(\Delta\mathcal{Z}_o) = \int_{-\Delta t_T}^{0} g_o^{\mathrm{CD}}(p;\Delta\mathcal{Z}_o)\,\mathrm{d}p \quad \text{mit } p \in [0,1] \tag{4.21}$$

Diese lässt sich durch die Trapezregel oder Simpson-Regel (vgl. [Fre07]) nummerisch approximieren.

Die Gleichverteilung der Anzahl an aufgetretenen Wahrnehmungen mit mindestens einer Wahrnehmung pro Tag stellt ein ideales Ergebnis dar – mit Hinblick auf den zukünftigen Einsatz von humanoiden Robotern z. B. im Haushalt. Dieses liefert jedoch keinen Maximalwert für die zuvor definierte Funktion. Deshalb ist eine zusätzliche Normierung des Wertebereichs notwendig (engl.: normalized integrated cumulated number of detections; kurz: NICD):

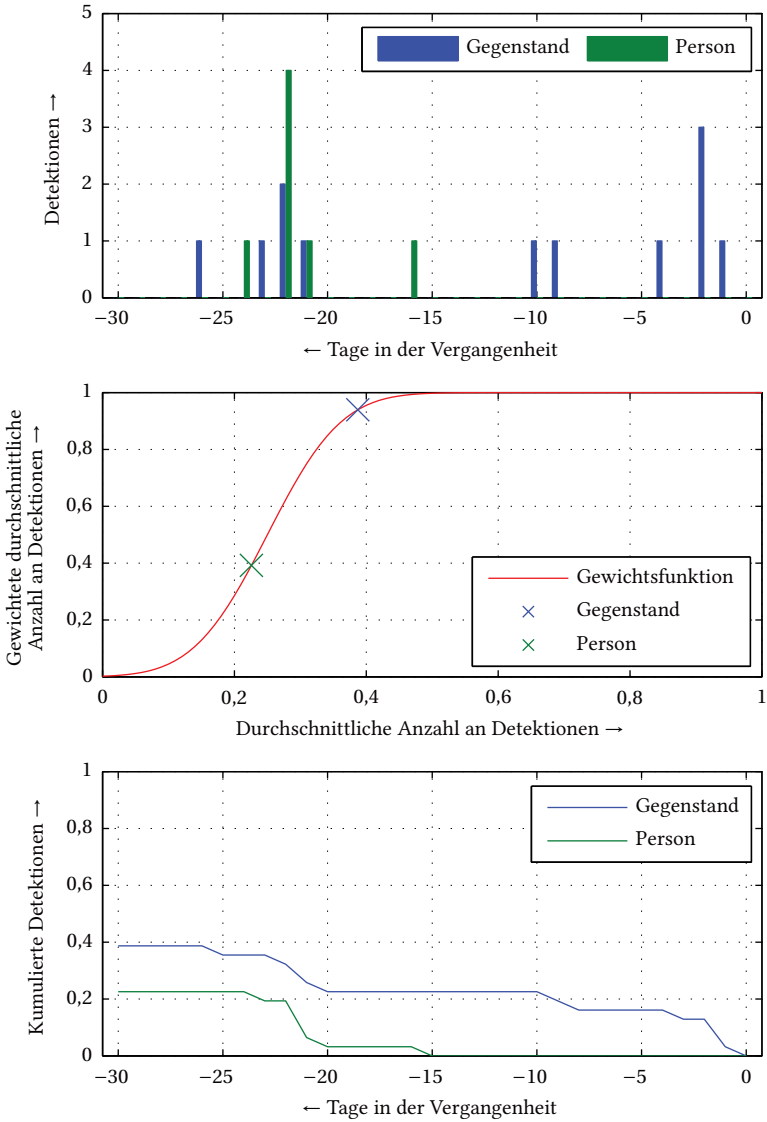

Abb. 4.3: Wahrnehmung eines Gegenstandes bzw. einer Person in den letzten 30 Tagen (oben), Gewichtsfunktion zur Bewertung der durchschnittlichen Anzahl an Wahrnehmungen im Zeitraum (Mitte), kumulierte und normierte Kurve über die Wahrnehmung in der Vergangenheit (unten).

$$g_o^{\mathrm{NICD}}(\Delta\mathcal{Z}_o) = \begin{cases} 1, & \text{falls } g_o^{\mathrm{ICD}}(\Delta\mathcal{Z}_o) \geq \zeta^{\mathrm{ICD1}} \\ g_o^{\mathrm{ICD}}(\Delta\mathcal{Z}_o) \cdot \zeta^{\mathrm{ICD2}}, & \text{sonst} \end{cases}. \tag{4.22}$$

Im Falle der Gleichverteilung lässt sich ein ideales $\zeta^{\mathrm{ICD1}} = 0{,}5$ bestimmen. Damit der Wertebereich der Funktion $g_o^{\mathrm{NICD}}(\Delta\mathcal{Z}_o) \in [0,1]$ kontinuierlich ist, wird die Gewichtung mit dem Skalierungsfaktor $\zeta^{\mathrm{ICD2}} = \frac{1}{\zeta^{\mathrm{ICD1}}} = 2$ durchgeführt. Um nun neuartige Objekte höher zu bewerten und bekannte niedriger, muss Gl. 4.22 bzgl. der Aussage umgekehrt werden. Dies kann wie folgt beschrieben werden:

$$g_o^{\overline{\mathrm{NICD}}}(\Delta\mathcal{Z}_o) = 1 - g_o^{\mathrm{NICD}}(\Delta\mathcal{Z}_o). \tag{4.23}$$

Die Ergebnisse für ein Beispiel mit einer Person und einem Gegenstand sind in Abb. 4.3 (unten) visualisiert.

Abschließend lässt sich durch Gewichtung der Häufigkeit mit der zeitlichen Struktur der Wahrnehmung die eingangs gesuchte Funktion zur Bestimmung der Neuartigkeit anhand des Auftretens in einem gewissen Zeitfenster wie folgt definieren:

$$f_o^{\mathrm{Häufigkeit}}(\Delta\mathcal{Z}_o) = g_o^{\overline{\mathrm{NICD}}}(\Delta\mathcal{Z}_o) \cdot g_o^{\overline{\mathrm{WAD}}}(\Delta\mathcal{Z}_o) \tag{4.24}$$

Zusammenfassung

Die Neuartigkeit eines Objekts kann durch den Grad der Neuartigkeit π_o^{Grad}, den Zeitpunkt der letzten Wahrnehmung $\pi_o^{\mathrm{Zeitpunkt}}$ sowie der Häufigkeit der Wahrnehmungen $\pi_o^{\mathrm{Häufigkeit}}$ bestimmt werden. In den voran gegangenen Abschnitten wurden die angesprochenen Kriterien definiert. Daraus lässt sich mittels einer geeigneten Fusionsfunktion f_π die Gesamtneuartigkeit für ein Objektrepräsentant o definieren als

$$\pi_o := f_\pi\left(\pi_o^{\mathrm{Grad}}, \pi_o^{\mathrm{Zeitpunkt}}, \pi_o^{\mathrm{Häufigkeit}}\right) \quad \text{mit } o \in \mathcal{O}. \tag{4.25}$$

In Abschnitt 4.3.7 werden verschiedene Fusionsfunktionen vorgestellt sowie deren Vor- und Nachteile aufgezeigt.

4.3.4 Komplexität

Bei der Modellierung der Komplexität spielen verschiedene Aspekte eine Rolle. Neben dem vorhandenen Wissen, auf dessen Basis z. B. aus einfachen Teilstrukturen (beispielsweise Linien) bereits bekannte komplexere Objekte gebildet werden können (z. B. Buchstaben), basiert die Komplexität auf verschiedenen Aspekten

wie (Un-)Regelmäßigkeit in Form und Textur, Heterogenität der Objekte und die Anzahl an verschiedenartigen bzw. gleichartigen Objekten (vgl. [Ber60]).

Im Rahmen der vorliegenden Arbeit werden die zuvor genannten Aspekte mit Hilfe von zwei Größen repräsentiert: als Erstes die Komplexität eines Objekts κ_o^{Objekt}, welche sich aus bestimmten Eigenschaften wie beispielsweise der Textur oder der Kontur zusammensetzt und als Zweites die räumliche Beziehung der Objekte untereinander κ_o^{Szene}. Die Gesamtkomplexität κ_o eines im Umweltmodell repräsentierten Objekts o lässt sich unter Verwendung der Fusionsfunktion f_κ definieren als

$$\kappa_o := f_\kappa\left(\kappa_o^{\text{Objekt}}, \kappa_o^{\text{Szene}}\right) \quad \text{mit } o \in \mathcal{O}. \tag{4.26}$$

Im Gegensatz zur Neuartigkeit macht die Definition von Komplexität in Bezug auf Personen wenig Sinn. Deshalb ist die Komplexität im Folgenden nur für Gegenstände definiert und beschrieben.

Komplexität eines Objekts

Die Objektkomplexität ist in Abhängigkeit der Menge an Attributen $\tilde{\mathcal{A}}_o$ definiert, welche zur Komplexität einen aktiven Beitrag leisten:

$$\kappa_o^{\text{Objekt}} := f_o^{\text{Objekt}}(\tilde{\mathcal{A}}_o) \tag{4.27}$$

Die verwendeten Attribute hängen von der Klasse des Objektrepräsentanten ab und lassen sich durch Hinzunahme neuer Verfahren, welche entsprechende Attribute erzeugen, erweitern. Im Rahmen der vorliegenden Arbeit werden die Texturiertheit, die Form der Objektkontur, die Farbgebung und die allgemeine Präsenz in einer Szene berücksichtigt (vgl. Abschnitt 6.3.3).

Es lässt sich nun eine Attributskomplexität (engl.: attribute complexity; kurz: AC) für jedes berücksichtigte Attribut bestimmen. Dies geschieht mit Hilfe einer attributspezifischen Transformationsfunktion $f_{a_i}^{\text{Komplexität}}(\cdot)$, welche den Attributwert w_{a_i} in ein Maß für die Komplexität überführt:

$$g_o^{\text{AC}}(a_i) = f_{a_i}^{\text{Komplexität}}(w_{a_i}) \quad \text{mit } a_i \in \tilde{\mathcal{A}}_o. \tag{4.28}$$

Dabei wird der Attributwert in das Intervall $[0, 1]$ überführt (z. B. in Kombination mit einer Min-Max-Normalisierung; vgl. [Kit98]) unter der Prämisse, dass eine höhere Komplexität einen größeren Funktionswert darstellt als eine geringere Komplexität.

Die Objektkomplexität (engl. object complexity; kurz: OC) lässt sich als gewichtete Attributskomplexität unter Zuhilfenahme des Konfidenzwertes k_{a_i} und der Priorität ρ_{a_i} des jeweiligen Attributs a_i wie folgt definieren:

$$g_o^{\mathrm{OC}}\left(\tilde{\mathcal{A}}_o\right) = \frac{\sum_{a_i \in \tilde{\mathcal{A}}_o} g_o^{\mathrm{AC}}(a_i) \cdot k_{a_i} \cdot \rho_{a_i}}{\sum_{a_i \in \tilde{\mathcal{A}}_o} \rho_{a_i}} \in [0,1]\,. \tag{4.29}$$

Die Gesamtfunktion für die Objektkomplexität berücksichtigt den Sonderfall, dass ein Objekt auch eine Person repräsentieren kann. In diesem Fall existiert die Klassenrelation $r_{\mathrm{istGegenstand}}$ nicht. Außerdem sind die benötigten Attribute $\tilde{\mathcal{A}}_o$ nicht vorhanden und somit ist die Objektkomplexität nicht bestimmbar. Infolgedessen wird bei der späteren Fusion dieser Teilaspekt nicht berücksichtigt. Andernfalls wird die in Gl. 4.29 bestimmte Objektkomplexität verwendet:

$$f_o^{\mathrm{Objekt}}(\tilde{\mathcal{A}}_o) = \begin{cases} g_o^{\mathrm{OC}}(\tilde{\mathcal{A}}_o), & \text{falls } \exists r_{\mathrm{istGegenstand}} \in \mathcal{R}_o \wedge \tilde{\mathcal{A}}_o \neq \emptyset \\ \text{„}\textit{nicht definiert}\text{“}, & \text{sonst} \end{cases} \tag{4.30}$$

Lokale Szenenkomplexität

Die räumliche Komplexität einer Szene ist durch die Anordnung und Anzahl der lokal vorhandenen Objekte beschrieben. Im Folgenden wird die lokale Komplexität der Szene den in einer Nachbarschaft vorhandenen Objekten selbst zugeordnet. Jedes Objekt ist dabei Mitglied einer lokalen Gruppe an Objekten (engl.: cluster). Die allgemeine Funktion zur Bestimmung der lokalen Szenenkomplexität für ein Objektrepräsentant o ist über die Menge aller Gegenstände $\mathcal{O}_{\mathrm{Gegenstände}}$ selbst definiert:

$$\kappa_o^{\mathrm{Szene}} := f_o^{\mathrm{Szene}}(\mathcal{O}_{\mathrm{Gegenstände}}) \tag{4.31}$$

mit

$$\mathcal{O}_{\mathrm{Gegenstände}} = \left\{ o_j \in \mathcal{O} \,\middle|\, \exists r_{\mathrm{istGegenstand}}^{o_j} \in \mathcal{R}_{o_j} \right\}. \tag{4.32}$$

Für die Bestimmung der lokalen Nachbarschaft werden nur die Positionen aller Gegenstände (engl.: object positions; kurz: OP) verwendet, da die entsprechenden Attributwerte bei Personen keine sinnvolle Repräsentation für die Komplexität darstellt.

$$g^{\mathrm{OP}}(\mathcal{O}_{\mathrm{Gegenstände}}) = \left\{ \bigcup_{o_j \in \mathcal{O}_{\mathrm{Gegenstände}}} \left\{ w_{a_{\mathrm{Position}}}^{o_j} \right\} \right\} \tag{4.33}$$

Die lokale Nachbarschaft wird durch Clusterbildung bestimmt über

$$\mathcal{C} = \mathrm{Clust}\left(g^{\mathrm{OP}}(\mathcal{O}_{\mathrm{Gegenstände}}), d_{\mathrm{Cluster}} \right) \quad \text{mit } \mathcal{C}_o \in \mathcal{C} \text{ und } w_{a_{\mathrm{Position}}}^{o} \in \mathcal{C}_o\,, \tag{4.34}$$

wobei $\mathcal{C}_o$ das Cluster repräsentiert, welches den Positionswert des Objekts o beinhaltet. *Clust* ist eine Clusterfunktion, welche die Menge der Positionswerte, die über die Funktion $g^{\mathrm{OP}}(\mathcal{O}_{\mathrm{Gegenstände}})$ bestimmt werden, in disjunkte Teilmengen unterteilt in Abhängigkeit des Distanzmaßes d_{Cluster}. Hierfür wird das Mean-Shift-Clusterverfahren (vgl. [Com02]) verwendet. Die Gesamtfunktion lässt sich nun unter Berücksichtigung einiger Sonderfälle schreiben als

$$f_o^{\mathrm{Szene}}(\mathcal{O}_{\mathrm{Gegenstände}}) = \begin{cases} f_{\mathrm{Anzahl}}(\mathcal{C}_o), & \text{falls } \exists r_{\mathrm{istGegenstand}} \in \mathcal{R}_o \wedge |\mathcal{C}_o| > 1 \\ \text{„}\textit{nicht definiert}\text{“}, & \text{sonst} \end{cases} \tag{4.35}$$

mit f_{Anzahl} als Bewertungsfunktion, sodass mit zunehmender Anzahl an Objekten der Funktionswert immer weiter gegen den Maximalwert 1 konvergiert. Diese ist beispielsweise definiert als

$$f_{\mathrm{Anzahl}}(\mathcal{C}_o) := \mathrm{erf}\left(2 \cdot \frac{|\mathcal{C}_o| - 1}{\zeta_{\mathrm{Anzahl}}}\right) \quad \text{mit } \zeta_{\mathrm{Anzahl}} = 4\,, \tag{4.36}$$

falls ab vier Objekten in einem Cluster (d. h. in einer lokalen Umgebung) die nahezu höchste lokale Szenenkomplexität erreicht werden soll. Dabei kann bei der Clusterbildung über den Parameter d_{Cluster} einen maximalen Abstand der Objekte in einem Cluster (vgl. Gl. 4.34) berücksichtigt werden. Ein Wert von $d_{\mathrm{Cluster}} = 0{,}25\,\mathrm{cm}$ stellt in Experimenten ein sinnvolles Distanzmaß dar. Die Fehlerfunktion wurde aufgrund ihrer guten Konvergenz für eine beliebige Anzahl an Objektrepräsentanten gewählt.

Zusammenfassung

Die Komplexität eines Objekts ist nur für Gegenstände definiert, da für Personen eine Definition von Komplexität nicht sinnvoll ist. Mit einer Auswahl an Attributen, welche eine Komplexitätseigenschaft repräsentieren, wird die Objektkomplexität κ^{Objekt} bestimmt. Dabei werden u. a. Eigenschaften wie die Texturiertheit und die Form berücksichtigt. Bei der Szenekomplexität κ^{Szene} ist hingegen die räumliche Anordnung der Objekte entscheidend. Die Gesamtkomplexität lässt sich nun mit Hilfe der Fusionsfunktion f_κ definieren als

$$\kappa_o := f_\kappa\left(\kappa_o^{\mathrm{Objekt}}, \kappa_o^{\mathrm{Szene}}\right) \quad \text{mit } o \in \mathcal{O}\,. \tag{4.37}$$

Für die Fusionsfunktion sei an dieser Stelle auf den Abschnitt 4.3.7 verwiesen, in dem mögliche Fusionsfunktionen vorgestellt und verglichen werden.

4.3.5 Unsicherheit

Die Unsicherheit ist ein wichtiger Bestandteil der Neugier, da der Mensch bestrebt ist, (neue) Informationen zu sammeln, um die Unsicherheit in Bezug auf ein Objekt oder einen Sachverhalt zu reduzieren.

Im Rahmen der vorliegenden Arbeit werden die folgenden zwei Arten der Unsicherheit realisiert: Zum einen ist dies die Unsicherheit in Bezug auf die Identität einer Person oder den Typ eines Gegenstands v_o^{Klasse} und zum anderen in Bezug auf die Eigenschaften eines Objekts $v_o^{\text{Eigenschaften}}$. Beide Arten der Unsicherheit entstehen durch eine unvollständige Erfassung von Objekten, sowohl in der Vergangenheit als auch zum aktuellen Zeitpunkt.

Die Gesamtunsicherheit für ein im Umweltmodell repräsentiertes Objekt o ist mit Hilfe der Fusionsfunktion f_v definiert als

$$v_o := f_v\left(v_o^{\text{Klasse}}, v_o^{\text{Eigenschaften}}\right) \quad \text{mit } o \in \mathcal{O}. \tag{4.38}$$

Typ- bzw. Identitätsunsicherheit

Einen wichtigen Bestandteil bei der Unsicherheit eines Objekts spielt die Typbzw. Identitätsunsicherheit (je nachdem ob es sich um einen Gegenstand oder eine Person handelt), d. h. der Grad, wie sehr der Aussage über die aktuelle Klasse eines Objekts in diesem Zusammenhang misstraut wird (kurz: Klassenungewissheit). Diese lässt sich auch direkt als Funktion des Konfidenzwertes der aktuellen Klasse (d. h. der letzten Stufe in der Klassenhierarchie) schreiben:

$$v_o^{\text{Klasse}} := f_o^{\text{Klasse}}\left(k_{r_{\text{istKlasse}}}\right). \tag{4.39}$$

Die Klassenunsicherheit stellt eine gegensätzliche Aussage zum Konfidenzwert (Maß der Sicherheit) für eine Objektklasse dar und lässt sich somit wie folgt formulieren:

$$f_o^{\text{Klasse}}\left(k_{r_{\text{istKlasse}}}\right) = 1 - k_{r_{\text{istKlasse}}} \quad \text{mit } k_{r_{\text{istKlasse}}} \in [0,1], \tag{4.40}$$

wobei $k_{r_{\text{istKlasse}}}$ der Konfidenzwert der aktuellen Klasse ist. In diesem Kontext stellen konkrete Personen und Gegenstände (z. B. $r_{\text{istBenjamin}}$ oder $r_{\text{istTeekanne}}$) ebenfalls eine eigene Klasse dar (vgl. Abschnitt 3.2.2).

Attributs- und relationsbezogene Unsicherheit

Wie eingangs im vorliegenden Kapitel erwähnt, ist der Grad der Unsicherheit von erfassten Attributen und Relationen (kurz: *Eigenschaften*) für die Neugier

ebenfalls ein wichtiger Beitrag. Die Unsicherheit bzgl. der Objekteigenschaften lässt sich in diesem Zusammenhang definieren über

$$v_o^{\text{Eigenschaften}} := f_o^{\text{Eigenschaften}}(\mathcal{A}_o, \mathcal{R}_o)\,. \tag{4.41}$$

Der Grad der attributs- und relationsbezogenen Unsicherheit (engl.: level of uncertainty; kurz: LU) lässt sich – in Anlehnung an den vorherigen Abschnitt – über den Konfidenzwert für ein Attribut bzw. eine Relation definieren. Hierbei wird die Wichtigkeit der einzelnen Eigenschaften, wie auch bei der Neuartigkeit zuvor, mitberücksichtigt. Hierzu wird der im Umweltmodell vorhandene Konfidenzwert k in Kombination mit der Priorität ρ verwendet und auf einen festen Wertebereich normiert. Die Unsicherheit wird somit definiert als

$$g_o^{\overline{\text{LU}}}(\mathcal{A}_o, \mathcal{R}_o) = \frac{\sum_{i=1}^{I} k_{a_i}\rho_{a_i} + \sum_{j=1}^{J} k_{r_j}\rho_{r_j}}{\sum_{i=1}^{I} \rho_{a_i} + \sum_{j=1}^{J} \rho_{r_j}} \in [0,1] \tag{4.42}$$

$$g_o^{\text{LU}}(\mathcal{A}_o, \mathcal{R}_o) = 1 - g_o^{\overline{\text{LU}}}(\mathcal{A}_o, \mathcal{R}_o)\,. \tag{4.43}$$

Wie auch bei der Neuartigkeit zuvor können die Wissensabhängigkeiten, die Abstraktionsebenen und die Klassenhierarchie eines Objektrepräsentanten im Umweltmodell (vgl. Abschnitt 3.2.3) berücksichtigt werden. Die attributs- und relationsbezogene Unsicherheit kann somit für die jeweilige Ebene in der Klassenhierarchie bestimmt werden (engl.: hierarchical level of uncertainty; kurz: HLU):

$$g_o^{\overline{\text{HLU}}}(\mathcal{A}_o, \mathcal{R}_o) = \min\left\{ \bigcup_{h=1,\dots,H} \left\{ \frac{\sum_{i=1}^{I} k_{a_i}^h \rho_{a_i}^h + \sum_{i=1}^{I} k_{r_i}^h \rho_{r_i}^h}{\sum_{i=1}^{I} \rho_{a_i}^h + \sum_{i=1}^{I} \rho_{r_i}^h} \right\} \right\} \tag{4.44}$$

$$g_o^{\text{HLU}}(\mathcal{A}_o, \mathcal{R}_o) = 1 - g_o^{\overline{\text{HLU}}}(\mathcal{A}_o, \mathcal{R}_o)\,. \tag{4.45}$$

Auf diese Art und Weise werden die Eigenschaften der jeweiligen Ebene in der Hierarchie h zugeordnet und eine Unsicherheit für jede Ebene separat bestimmt. Die Gesamtunsicherheit für ein Objekt bzgl. seiner Eigenschaften wird bei diesem Ansatz als Maximum der Unsicherheit (d. h. Minimum des priorisierten Konfidenzwerts) aller Ebenen bestimmt.

Somit kann die attributs- und relationsbezogene Unsicherheit nun für ein Objektrepräsentant über alle seine vorhandenen Eigenschaften bestimmt werden oder zunächst separat für alle Ebenen in der Hierarchie mit

$$f_o^{\text{Eigenschaften}}(\mathcal{A}_o, \mathcal{R}_o) = g_o^{\text{LU}}(\mathcal{A}_o, \mathcal{R}_o) \tag{4.46}$$

bzw.

$$f_o^{\text{Eigenschaften}}(\mathcal{A}_o, \mathcal{R}_o) = g_o^{\text{HLU}}(\mathcal{A}_o, \mathcal{R}_o)\,. \tag{4.47}$$

Zusammenfassung

Die Unsicherheit lässt sich zusammenfassend beschreiben als eine Unsicherheit bzgl. der Klasse eines Objektrepräsentanten ν_o^{Klasse} und als eine Unsicherheit in Bezug auf die Eigenschaften eines Objektrepräsentanten $\nu_o^{\text{Eigenschaften}}$. Bei Letzterem kann mit Hilfe der hierarchischen Vorgehensweise auch die Unsicherheit auf jeder Ebene der Objekthierarchie bestimmt werden. Die Gesamtunsicherheit lässt sich mit der Fusionsfunktion f_ν definieren als

$$\nu_o := f_\nu\left(\nu_o^{\text{Klasse}}, \nu_o^{\text{Eigenschaften}}\right) \quad \text{mit } o \in \mathcal{O}. \tag{4.48}$$

In Abschnitt 4.3.7 ist die Gesamtfusion aller Aspekte der Neugier dargestellt. Dabei werden auch verschiedene Fusionsfunktionen vorgeschlagen und bewertet, die für die Unsicherheit angewendet werden können.

4.3.6 Konflikt

Wahrgenommene Objekte erzeugen immer dann Konflikte, wenn gewisse Erwartungshaltungen nicht erfüllt werden. Dies lässt sich anhand von drei Einzelkonflikten beschreiben: Als Erstes lässt sich ein Konflikt in Bezug auf die Identität einer Person feststellen, falls in einer Szene mehrere Personen mit scheinbar derselben Identität existieren. Insbesondere bei eineiigen Zwillingen ist dies häufiger der Fall, da der optische Unterschied marginal ist. Als Zweites tritt ein Konflikt auf, falls Objekte mit mehreren Modalitäten wahrgenommen werden (z. B. visuell und akustisch) und die wahrgenommenen Informationen nicht miteinander vereinbar sind. Als Letztes ist ein Konflikt vorhanden, wenn als fest angenommene Voraussetzungen oder Fakten verletzt werden (z. B. Existenz von roten Bananen). Dies wird insbesondere durch fehlendes oder falsches Wissen bzw. durch eine schlechte Wahrnehmung verursacht und/oder verstärkt.

Im Rahmen der vorliegenden Arbeit wird der Gesamtkonflikt ς_o für ein Objektrepräsentant mit Hilfe von drei Einzelkonflikten und der Fusionsfunktion f_ς definiert:

$$\varsigma_o := f_\varsigma\left(\varsigma_o^{\text{Identität}}, \varsigma_o^{\text{Multimodal}}, \varsigma_o^{\text{Vorwissen}}\right), \quad o \in \mathcal{O}. \tag{4.49}$$

Hierbei repräsentiert $\varsigma_o^{\text{Identität}}$ den Konflikt bei mehreren Personen mit derselben Identität, $\varsigma_o^{\text{Multimodal}}$ den Konflikt bei gegensätzlicher multimodaler Wahrnehmung und $\varsigma_o^{\text{Vorwissen}}$ den Konflikt mit dem a-priori-Wissen. Der Konflikt bzgl. der Identität $\varsigma_o^{\text{Identität}}$ bleibt bei Gegenständen unberücksichtigt.

Konflikt bei der Identitätsbestimmung einer Person

Werden Personen und Gegenstände betrachtet, so ergibt sich ein genereller Unterschied bzgl. deren Anzahl. So existiert ein und die dieselbe Person nur ein einziges Mal auf der Welt, wohingegen Gegenstände desselben Typs im Allgemeinen mehrmals vorkommen können. Selbst bei Zwillingen existieren in der Regel sichtbare Unterschiede, wohingegen zwei gleiche Gegenstände optisch nicht unterscheidbar sein können.

Vor diesem Hintergrund wird speziell für Personen ein Konflikt definiert, der sich im mehrfachen Vorhandensein ein und derselben Person im Umweltmodell ausdrückt. Dieser Konflikt wird für den aktuellen Objektrepräsentant o in Abhängigkeit aller im Weltmodell vorhandener Personen $\mathcal{O}_{\text{Personen}}$ wie folgt definiert:

$$\varsigma_o^{\text{Identität}} := f_o^{\text{Identität}}(\mathcal{O}_{\text{Personen}}) \tag{4.50}$$

mit

$$\mathcal{O}_{\text{Personen}} = \left\{ o_j \in \mathcal{O} \,\middle|\, \exists r^{o_j}_{\text{istPERSON}} \in \mathcal{R}_{o_j} \right\}. \tag{4.51}$$

Für die Bestimmung des Konfliktausmaßes wird zunächst eine Indikatorfunktion definiert, welche prüft, ob zwei Personen dieselbe Identität besitzen (engl.: identity conflict; kurz: IDC), und anschließend wird der Grad des Konfliktes mit einer zweiten Bewertungsfunktion bestimmt. Die Indikatorfunktion

$$g^{\text{IDC}}(o, o_n) = \begin{cases} 1, & \text{falls } (t^o_{a_{\text{Identität}}}, w^o_{a_{\text{Identität}}}) = (t^{o_n}_{a_{\text{Identität}}}, w^{o_n}_{a_{\text{Identität}}}) \wedge o \neq o_n \\ 0, & \text{sonst} \end{cases} \tag{4.52}$$

überprüft, ob das Typ-Wert-Tupel $(t_{a_{\text{Identität}}}, w_{a_{\text{Identität}}})$ für das Attribut *Identität* zweier Objektrepräsentanten identisch ist. Hierbei ist o die aktuelle Person und o_n eine weitere Person aus dem Umweltmodell. Ein Konflikt mit sich selbst wird durch die Nebenbedingung $o \neq o_n$ ausgeschlossen.

Durch Hinzunahme eines Bewertungskriteriums wird die zuvor definierte Entscheidung über das Vorhandensein eines Konflikts relativiert. Das Minimum der Konfidenzwerte der Identität beider Objekte ist hierbei ein geeignetes Kriterium, da bei stark unterschiedlichen Werten der Konflikt niedrig ist und bei gleichen Werten der Grad des Konflikts vom Absolutwert abhängig ist. Das Bewertungskriterium ist wie folgt definiert:

$$\Delta k_{a_{\text{Identität}}}(o, o_n) = \min \left\{ k^{o_n}_{a_{\text{Identität}}}, k^{o}_{a_{\text{Identität}}} \right\} \in [0, 1] \tag{4.53}$$

Anschließend wird der Sonderfall berücksichtigt, dass das zu untersuchende Attribut $a_{\text{Identität}}$ und der dazugehörige Attributwert beim Objektrepräsentant o bzw. Vergleichsobjektrepräsentant o_n nicht vorhanden sind bzw. nicht erfasst wurden:

$$\Delta\tilde{k}_{a_{\text{Identität}}}(o,o_n) = \begin{cases} \Delta k_{a_{\text{Identität}}}(o,o_n), & \text{falls } \exists t^{o_n}_{a_{\text{Identität}}} \wedge w^{o_n}_{a_{\text{Identität}}} \neq \text{„nicht erfasst“} \\ & \quad \wedge \exists t^{o}_{a_{\text{Identität}}} \wedge w^{o}_{a_{\text{Identität}}} \neq \text{„nicht erfasst“} \\ 0, & \text{sonst} \end{cases} \tag{4.54}$$

Der Konflikt bzgl. der Identität zwischen zwei Objekten wird durch die Bewertung der Indikatorfunktion (vgl. Gl. 4.52) mit dem Minimum der korrespondierenden Konfidenzwerte (vgl. Gl. 4.54) beschrieben. Hierbei wird der Gesamtkonflikt für Objektrepräsentant o als Maximum der Einzelkonflikte mit den anderen Objekten $o_n \in \mathcal{O}_{\text{Personen}}$ im Umweltmodell definiert:

$$\tilde{f}_o^{\text{Identität}}(\mathcal{O}_{\text{Personen}}) = \max\left\{ \bigcup_{n=1}^{N} \left\{ \Delta\tilde{k}_{a_{\text{Identität}}}(o,o_n) \cdot g^{\text{IDC}}(o,o_n) \right\} \right\} \tag{4.55}$$

mit $N = |\mathcal{O}_{\text{Personen}}|$ und $o_n \in \mathcal{O}_{\text{Personen}}$.

Der Sonderfall, dass der zu untersuchende Objektrepräsentant o keine Person ist, muss noch berücksichtigt werden:

$$f_o^{\text{Identität}}(\mathcal{O}_{\text{Personen}}) = \begin{cases} \tilde{f}_o^{\text{Identität}}(\mathcal{O}_{\text{Personen}}), & \text{falls } o \in \mathcal{O}_{\text{Personen}} \\ \text{„nicht definiert“}, & \text{sonst} \end{cases} \tag{4.56}$$

Konflikt durch die multimodale Wahrnehmung

Weitere Konflikte können dadurch entstehen, dass einzelne Attribute oder Relationen unterschiedliche multimodale Aussagen haben. Hierbei ist multimodal nicht nur auf unterschiedliche Sensor(typ)en bezogen, sondern auch auf verschiedene Verfahren zur Bestimmung eines Attributs bzw. einer Relation. In der Regel wird in solchen Fällen stets eine Fusion durchgeführt, um ein gemeinsames Attribut bzw. eine gemeinsame Relation zu bestimmen.

Der multimodale Konflikt wird als Funktion in Anhängigkeit aller Attribute und Relationen definiert als:

$$\varsigma_o^{\text{Multimodal}} := f_o^{\text{Multimodal}}(\mathcal{A}_o, \mathcal{R}_o)\,. \tag{4.57}$$

Im Folgenden wird exemplarisch auf die *Identität* einer Person bzw. den *Typ* eines Gegenstandes eingegangen. Die nachfolgend beschriebenen Ansätze sind generisch und lassen sich auf beliebige Attribute und Relationen anwenden.

Zur Bestimmung des Konflikts wird eine Indikatorfunktion definiert, welche das Vorhandensein eines Konflikts zwischen dem Ergebnis der Modalität m und dem Gesamtergebnis (engl.: modality conflict; kurz: MC) für ein Attribut bzw. eine Relation darstellt:

$$g_o^{\mathrm{MC}}(w_{a_i^m}, w_{a_i}) = \begin{cases} 1, & \text{falls } w_{a_i^m} \neq w_{a_i} \\ 0, & \text{sonst} \end{cases} \tag{4.58}$$

bzw.

$$g_o^{\mathrm{MC}}(w_{r_j^m}, w_{r_j}) = \begin{cases} 1, & \text{falls } w_{r_j^m} \neq w_{r_j} \\ 0, & \text{sonst} \end{cases} \tag{4.59}$$

Hierbei repräsentieren $w_{a_i^m}$ bzw. $w_{r_j^m}$ das Ergebnis, wenn nur die Modalität m genutzt werden würde, und w_{a_i} bzw. w_{r_j} das Ergebnis bei der Berücksichtigung aller Modalitäten. Letzteres ist die Referenz bei dem Vergleich mit dem Ergebnis jeder einzelnen Modalität (z. B. akustisch und visuell). Für das Beispiel mit der Identität einer Person bedeutet dies, dass die Gesamtidentität $w_{a_{\text{Identität}}}$ mit den Identitäten $w_{a^m_{\text{Identität}}}$ der einzelnen Modalitäten m verglichen wird. Konkret werden die akustische und visuelle Identität einer Person jeweils mit der Gesamtidentität verglichen. Für kontinuierliche Attribut- bzw. Relationswerte kann zusätzlich eine Unschärfe beim Vergleich einführt werden. Dies lässt sich über den Parameter ε_{a_i} bzw. ε_{r_j} erzielen:

$$g_o^{\mathrm{MC}}(w_{a_i^m}, w_{a_i}) = \begin{cases} 1, & \text{falls } w_{a_i^m} \notin \left[w_{a_i} - \varepsilon_{a_i}, w_{a_i} + \varepsilon_{a_i} \right] \\ 0, & \text{sonst} \end{cases} \tag{4.60}$$

bzw.

$$g_o^{\mathrm{MC}}(w_{r_j^m}, w_{r_j}) = \begin{cases} 1, & \text{falls } w_{r_j^m} \notin \left[w_{r_j} - \varepsilon_{r_j}, w_{r_j} + \varepsilon_{r_j} \right] \\ 0, & \text{sonst} \end{cases} . \tag{4.61}$$

Der Gesamtkonflikt für ein Attribut bzw. eine Relation wird durch die Bewertung mit einem zusätzlichen Kriterium definiert. Hierbei wird das Minimum der Konfidenzwerte wie auch schon im Abschnitt zuvor genutzt. Dieses Mal wird jedoch das Minimum zwischen dem Konfidenzwert für die einzelne Modalität m (z. B. akustisch und visuell) und dem Gesamtkonfidenzwert bestimmt:

$$\Delta k(k_{a_i^m}, k_{a_i}) = \min\{k_{a_i^m}, k_{a_i}\} \quad \text{bzw.} \quad \Delta k(k_{r_j^m}, k_{r_j}) = \min\{k_{r_j^m}, k_{r_j}\} \tag{4.62}$$

Somit lässt sich z. B. für den Typ eines Gegenstands der Konflikt bestimmen aus den Konfidenzwerten $k_{a^m_{\text{Typ}}}$ für die einzelnen Modalitäten m und dem Gesamtkonfidenzwert $k_{a_{\text{Typ}}}$ des Attributes a_{Typ}.

Die Gesamtfunktion lässt sich beschreiben durch die Gewichtung der Indikatorfunktion für das Vorhandensein eines Konflikts mit der Bewertungsfunktion für die korrespondierenden Konfidenzwerte. Dies geschieht für jede Modalität unabhängig, sodass der Gesamtkonflikt bzgl. der Wahrnehmung als Maximum der Teilkonflikte aller Attribute und Relationen definiert werden kann:

$$\tilde{f}_o^{\text{Multimodal}}(\mathcal{A}_o) = \max\left\{\bigcup_{i=1}^{I}\bigcup_{m=1}^{M_i}\left\{\Delta k(k_{a_i^m}, k_{a_i}) \cdot g_o^{\text{MC}}(w_{a_i^m}, w_{a_i})\right\}\right\} \tag{4.63}$$

$$\tilde{f}_o^{\text{Multimodal}}(\mathcal{R}_o) = \max\left\{\bigcup_{j=1}^{J}\bigcup_{m=1}^{M_j}\left\{\Delta k(k_{r_j^m}, k_{r_j}) \cdot g_o^{\text{MC}}(w_{r_j^m}, w_{r_j})\right\}\right\} \tag{4.64}$$

$$f_o^{\text{Multimodal}}(\mathcal{A}_o, \mathcal{R}_o) = \max\left\{\tilde{f}_o^{\text{Multimodal}}(\mathcal{A}_o), \tilde{f}_o^{\text{Multimodal}}(\mathcal{R}_o)\right\} \tag{4.65}$$

Hierbei stellt M_i bzw. M_j die Anzahl an verwendeten Modalitäten für das jeweilige Attribut a_i bzw. die jeweilige Relation r_j dar. Der Spezialfall, dass nur eine Modalität vorhanden ist, muss hierbei nicht gesondert berücksichtigt werden, da in diesem Fall die Indikatorfunktion stets Null liefert und somit kein Konflikt vorhanden ist.

Konflikte in Bezug auf das a-priori-Wissen

Bereits im Vorfeld kann Wissen über Objekte bereitgestellt werden (a-priori-Wissen). Dabei können bestimmte Eigenschaften von Objekten, welche vorab bekannt sind und sich in i. d. R. nicht ändern, definiert werden. Des Weiteren können auch Wertebereiche oder Mengen vorgegeben werden, welche gültige Werte für die Eigenschaft darstellen. Ein Beispiel hierfür ist die Farbe von Äpfeln, da nur Variationen von Grün, Gelb und Rot existieren. Sollte dennoch z. B. ein blauer Apfel wahrgenommen werden, so liegt ein Konflikt mit dem a-priori-Wissen vor.

Allgemein lässt sich dieser Konflikt in Abhängigkeit der aktuellen Attribute und Relationen eines Objektrepräsentanten o sowie dem a-priori-Wissen für Objekte beschreiben als

$$\varsigma_o^{\text{Vorwissen}} := f_o^{\text{Vorwissen}}(\mathcal{O}_{\text{Vorwissen}}) \quad \text{mit } o \in \mathcal{O}. \tag{4.66}$$

Hierbei stellt $\mathcal{O}_{\text{Vorwissen}}$ die Menge aller a-priori-Objekte dar. Diese beschreiben die a-priori bekannten (Teil-)Informationen für eine konkrete Objektklasse (vgl. Abschnitt 3.1.1). Das Vorliegen eines Konflikts für ein konkretes Attribut a_i (engl.: a-priori attribute conflict; kurz: AAC) lässt sich beschreiben mit

$$c^o_{a_i} := g^{\text{AAC}}_o(\mathcal{O}_{\text{Vorwissen}}, a_i)$$
$$= \begin{cases} 1, & \text{falls } \exists p,k : r^o_{\text{istKlasse}} = r^{o_p}_{\text{istKlasse}} \wedge t^o_{a_i} = t^{o_p}_{a_k} \wedge w^o_{a_i} \notin \mathcal{W}^{o_p}_{a_k} \\ 0, & \text{sonst} \end{cases}, \quad (4.67)$$

wobei $o_p \in \mathcal{O}_{\text{Vorwissen}}$ ist und k ein Laufindex über die Attribute des aktuellen a-priori-Objekts o_p darstellt. Es wird zunächst überprüft, ob ein passendes a-priori-Objekt existiert, bei dem die aktuelle Klasse übereinstimmt (1. Bedingung), bei dem der aktuelle Attributstyp vorhanden ist (2. Bedingung) und der aktuelle Wert des Attributs nicht im vordefinierten Wertebereich liegt oder nicht in der Teilmenge der a-priori-Daten beinhaltet ist (3. Bedingung). Treffen alle drei Aussagen zu, so liegt ein Konflikt für das aktuelle Attribut a_i des Objektrepräsentanten o vor.

Analog dazu kann auch der Konflikt für die aktuelle Relation r_j eines Objektrepräsentanten o (engl.: a-priori relation conflict; kurz: ARC) beschrieben werden als

$$c^o_{r_j} := g^{\text{ARC}}_o(\mathcal{O}_{\text{Vorwissen}}, r_j)$$
$$= \begin{cases} 1, & \text{falls } \exists p,k : r^o_{\text{istKlasse}} = r^{o_p}_{\text{istKlasse}} \wedge t^o_{r_j} = t^{o_p}_{r_k} \wedge w^o_{r_j} \notin \mathcal{W}^{o_p}_{r_k} \\ 0, & \text{sonst} \end{cases}. \quad (4.68)$$

Im Falle von Relationen sind die drei Kriterien: die Übereinstimmung der Klasse, das Vorhandensein des Relationstyps und die fehlende Übereinstimmung des Wertebereichs bzw. das Nichtvorhandensein in der Teilmenge des Relationswerts.

Um nun den Gesamtkonflikt des Objekts bzgl. des Vorwissens zu bestimmen, werden sowohl der aktuelle Konfidenzwert als auch die Priorität der Attribute bzw. Relationen berücksichtigt. Hierzu werden die zuvor bestimmten Konfliktindikatoren aus Gl. 4.67 und Gl. 4.68 jeweils mit dem Konfidenzwert $k^o_{a_i}$ bzw. $k^o_{r_j}$ und der Priorität $\rho^o_{a_i}$ bzw. $\rho^o_{r_j}$ des Repräsentanten o im Umweltmodell gewichtet:

$$f^{\text{Vorwissen}}_o(\mathcal{O}_{\text{Vorwissen}}) = \frac{\sum_{i=1}^{I} c^o_{a_i} \rho^o_{a_i} k^o_{a_i} + \sum_{j=1}^{J} c^o_{r_j} \rho^o_{r_j} k^o_{r_j}}{\sum_{i=1}^{I} c^o_{a_i} \rho^o_{a_i} + \sum_{j=1}^{J} c^o_{r_j} \rho^o_{r_j}} \quad \in [0,1] \quad (4.69)$$

Der Grad des Konfliktes bzgl. des a-priori-Wissens wird bei der Bestimmung mit den Prioritäten und den Konflikten der Attribute bzw. Relationen gewichtet, sodass eine implizite Normalisierung auf das Intervall $[0,1]$ stattfindet.

Es lassen sich neben den zuvor definierten 1:1-Beziehungen der Eigenschaften zwischen dem aktuellen und dem a-priori-Wissen natürlich auch wesentlich kompliziertere Beziehungen definieren. So könnte z. B. eine bestimmte Apfelsorte eine spezifische Farbkombination haben und nur in einem Land angebaut werden. Diese Verknüpfung ermöglicht eine sehr detaillierte Definition von gültigen

Eigenschaften. Diese Art und Weise ist sehr aufwendig, da die Datenmenge sehr schnell sehr groß wird und in der Regel entweder manuell definiert wird oder zumindest bei automatischer Generierung aus Statistiken überprüft werden sollte. Des Weiteren erfordert die Definition von Wissen in solch einem Detailliertheitsgrad, dass dieses regelmäßig überprüft und angepasst wird, da sonst Fehler entstehen können, die keine sind. Für das Beispiel mit den Äpfeln muss deshalb die Liste der Herkunftsländer sowie der Farbvariationen für neue Sorten angepasst werden.

Zusammenfassung

Der Konflikt eines Objekts wurde in den vorherigen Abschnitten definiert als Konflikt bzgl. der Identität einer Person $\varsigma_o^{\text{Identität}}$, in Bezug auf die multimodale Wahrnehmung $\varsigma_o^{\text{Multimodal}}$ und mit dem vorhandenen a-priori-Wissen $\varsigma_o^{\text{Vorwissen}}$. Der Gesamtkonflikt lässt sich mit Hilfe einer geeigneten Fusionsfunktion f_ς allgemein schreiben als

$$\varsigma_o := f_\varsigma\left(\varsigma_o^{\text{Identität}}, \varsigma_o^{\text{Multimodal}}, \varsigma_o^{\text{Vorwissen}}\right) \quad \text{mit } o \in \mathcal{O}\,. \tag{4.70}$$

Eine Auswahl an möglichen Fusionsfunktionen sowie die Vor- und Nachteile werden im nächsten Abschnitt diskutiert. Dort ist auch die hierarchische Fusion aller Teilaspekte für die Neugier aufgeführt.

4.3.7 Fusion

Die Fusion der Einzelergebnisse der Teilaspekte sowie der Teilaspekte selbst zur wissensbasierten Neugier geschieht wie in Abb. 4.4 dargestellt in zwei Stufen. In jeder Stufe erfolgt eine Fusion der Teilergebnisse. Die Fusion in der untersten Stufe wurde bereits allgemein in Gl. 4.1, Gl. 4.26, Gl. 4.38 und Gl. 4.49 definiert. Die Gesamtfusion der Teilaspekte ist definiert als

$$\eta_o := f_\eta\left(\pi_o, \kappa_o, \nu_o, \varsigma_o\right)\,. \tag{4.71}$$

Bei Personen wird der Teilaspekte Komplexität κ_o nicht mitberücksichtigt, da dieser nicht sinnvoll bestimmbar ist (vgl. Abschnitt 4.3.4):

$$\eta_o^{\text{Person}} := f_\eta\left(\pi_o, \nu_o, \varsigma_o\right)\,. \tag{4.72}$$

Je nach Fusionsfunktion ist es u. U. notwendig, dass die zu berücksichtigenden Eingangsgrößen denselben Wertebereich besitzen. Dazu ist gegebenenfalls eine Normalisierung der Eingangsgrößen vor der eigentlichen Fusion notwendig. Im Folgenden werden verschiedene Normalisierungsverfahren beispielhaft aufgelistet, bevor anschließend die Fusionsfunktionen beschrieben werden.

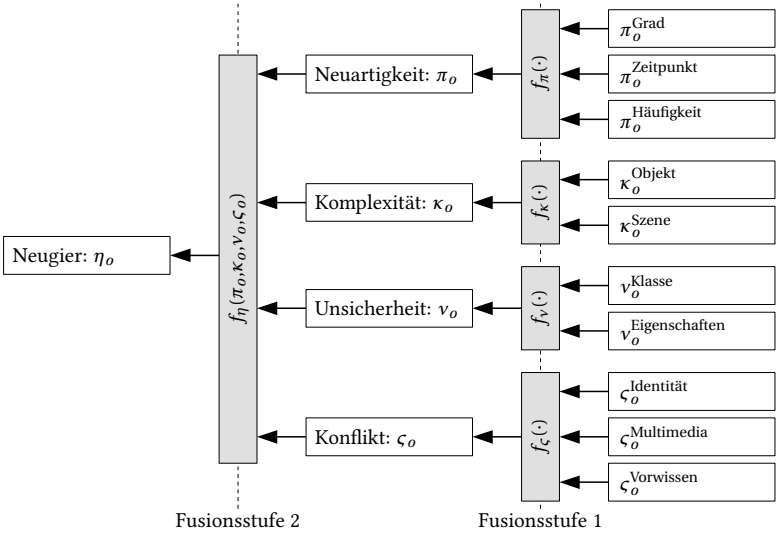

Abb. 4.4: Übersicht über die Fusionshierarchie zur Bestimmung der Neugier und zur Rückverfolgung der Quellen der Neugier.

Normalisierung

Für die Fusion unterschiedlicher Größen kann es je nach Fusionsfunktion notwendig sein, dass der Wertebereich der einzelnen Eingangsgrößen angepasst werden muss. Zu diesem Zweck können verschiedene Normalisierungsverfahren wie beispielsweise *Min-Max-Normalisierung*, *Decimal-Scaling*, *z-Score-Normalisierung* oder *tanh-Normalisierung* verwendet werden (vgl. [Jai05]).

Im Rahmen der Definition der Teilaspekte der Neugier wurde deren Wertebereich bereits auf ein gemeinsames Intervall $[0, 1]$ festgelegt. Die Generierung eines aufwendigen Trainingsdatensatzes für die Bestimmung einzelner Parameter, wie es bei manchen Normalisierungstechniken notwendig ist, entfällt somit.

Fusionsfunktionen

Nach der Festlegung der Teilaspekte Neuartigkeit, Komplexität, Unsicherheit und Konflikt werden nun abschließend für die Bestimmung der wissensbasierten Neugier verschiedene Fusionsfunktionen (ähnl. [Jai05], [Sne03], [Kit98], [Ver99]) kurz vorgestellt.

- *Minimum:* Diese Fusionsfunktion ist für eine gegebene Menge an Eingangsdaten $x_i \in X$ definiert als

$$x^{\text{Minimum}} := f_{\text{Minimum}}(x_1, x_2, \ldots, x_N) = \min\{x_1, x_2, \ldots, x_N\} \quad \text{mit } N \in \mathbb{N}. \tag{4.73}$$

Diese Funktion hat den Vorteil, dass ein großer Funktionswert mit insgesamt sehr hohen Einzelbeiträgen korreliert ist, d. h., alle Eingangsgrößen besitzen eine hohe Wichtigkeit. Aber genau dieser vermeintliche Vorteil macht die Fusionsfunktion auch sehr anfällig für eine einzelne große Abweichung vom Durchschnittswert, da nur ein einziger niedriger Wert letztendlich das Ergebnis bestimmt. In Bezug auf die wissensbasierte Neugier ist dieser Umstand von großem Nachteil, da einzelne Teilaspekte sehr wohl geringe Werte aufweisen können (z. B. Neuartigkeit).

- *Maximum:* Bei dieser Fusionsfunktion wird für eine gegebene Menge an Eingangsdaten $x_i \in X$ nur der größte Wert ausgewählt. Formal lässt sich dies definieren als

$$x^{\text{Maximum}} := f_{\text{Maximum}}(x_1, x_2, \ldots, x_N) = \max\{x_1, x_2, \ldots, x_N\} \quad \text{mit } N \in \mathbb{N}. \tag{4.74}$$

Diese Vorgehensweise deckt sich mit den frühen Theorien der selektiven Wahrnehmung beim Menschen (vgl. [Bro58]), in denen nur der stärkste Sinneseindruck weiterverarbeitet wird. Die Funktion beruht genau wie das Minimum letztendlich nur auf einem Wert der Eingangsdaten und kann damit ebenfalls die Leistungsfähigkeit des Verfahrens einschränken. Zusätzlich können mit dieser Funktion leicht mehrere Objekte denselben Wert (1) besitzen, da nur der stärkste Teilaspekt berücksichtigt wird.

- *Produkt:* Die multiplikative Verknüpfung der Teilergebnisse kann definiert werden über

$$x^{\text{Produkt}} := f_{\text{Produkt}}(x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} x_i \quad \text{mit } N \in \mathbb{N}. \tag{4.75}$$

Bei dieser Art der Fusion sorgt ein einziger geringer Wert eines Teilaspektes dafür, dass die Gesamtneugier stark einbricht. Ein solches Verhalten ist nicht sinnvoll, da die Beiträge aller durch nur einen Teilaspekt stark reduziert werden können. Im Extremfall bedeutet dies, dass nur durch einen Teilaspekt mit einem niedrigen Ergebnis die komplette Neugier unterdrückt werden kann.

- *Summe:* Bei der Summe trägt jeder Wert $x_i \in X$ kumulativ zum Gesamtergebnis bei. Dieser Zusammenhang lässt sich formal definieren als

$$x^{\text{Summe}} := f_{\text{Summe}}(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N} x_i \quad \text{mit } N \in \mathbb{N}. \tag{4.76}$$

Diese Fusionsfunktion hat den Vorteil, dass auch eine Vielzahl an kleinen Werten bei der Fusion berücksichtigt wird, sodass diese insgesamt einen moderaten Gesamtbeitrag leisten können.

- *Mittelwert:* Der Mittelwert ist vergleichbar zur Summe, mit dem Unterschied, dass eine Gewichtung mit der Anzahl der verwendeten Eingangsdaten erfolgt. Die Fusionsfunktion mit Mittelwert lässt sich definieren als

$$x^{\text{Mittelwert}} := f_{\text{Mittelwert}}(x_1, x_2, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \text{mit } N \in \mathbb{N}. \tag{4.77}$$

  Wird die wissensbasierte Neugier mehrerer Objekte verglichen, die mit $f_{\text{Mittelwert}}(\boldsymbol{x})$ bestimmt wurden, so ist das Verhältnis zwischen den Objekten vergleichbar mit dem Ergebnis der Fusionsfunktion $f_{\text{Summe}}(\boldsymbol{x})$, falls alle Teilaspekte dieselbe Anzahl an Eingangsdaten haben. Dies ist im Rahmen dieser Arbeit jedoch nicht der Fall.

- *Median:* Diese Fusionsfunktion ist ähnlich dem Mittelwert, hat jedoch bereits bei einer moderaten Menge an Daten den Vorteil, einzelne abweichende Werte sehr gut zu unterdrücken. Der Median wird bestimmt durch

$$x^{\text{Median}} := f_{\text{Median}}(x_1, \ldots, x_N) = \begin{cases} x_{(N+1)/2}, & \text{falls } N \text{ ungerade} \\ \frac{1}{2}\left(x_{N/2} + x_{N/2+1}\right), & \text{falls } N \text{ gerade} \end{cases}. \tag{4.78}$$

  Aufgrund der geringen Anzahl an Eingangswerten x_i (d. h. Teilaspekten) für die Fusion bei der Neugier liefert diese Fusionsfunktion keine guten Ergebnisse, da einzelne starke Aspekte dabei nicht berücksichtigt werden. Diese abweichenden Werte sollen jedoch nicht als „Ausreißer“ aufgefasst werden, sondern einen aktiven Beitrag zur Neugier leisten. Infolgedessen bleiben Teilaspekte der Neugier u. U. unberücksichtigt.

- *Gewichtete Fusion:* Bei dieser Fusion werden einzelne Teilaspekte anhand von Gewichten $\omega_i \in [0, 1]$ unterschiedlich stark priorisiert. Allgemein lässt sich die gewichtete Fusion definieren über

$$x^{\text{Gewichtet}} := f_{\text{Gewichtet}}(x_1, \ldots, x_N; \omega_1, \ldots, \omega_N) = \sum_{i=1}^{N} x_i \cdot \omega_i \quad \text{mit } N \in \mathbb{N}. \tag{4.79}$$

  Eine dynamische Anpassung der Gewichte in einem autonomen System ist in bestimmten Fällen sinnvoll und hängt von der konkreten Anwendung ab. Bei der Exploration einer Szene könnte beispielsweise mit der gewichteten Fusion am Anfang die Priorität auf neuartige Objekte gelegt werden. Im Laufe der Zeit sollte jedoch eine Verlagerung der Gewichtung auf Aspekte wie Unsicherheit und Konflikt erfolgen. Letztendlich ist auch dem Aspekt der Komplexität eine höhere Bedeutung zuzuordnen. Die Bestimmung der Gewichte muss dabei sorgfältig vorgenommen werden, da sonst ein Ungleichgewicht bei der Fusion entsteht und nicht die gewünschte Priorisierung für die Anwendung erfolgt.

Werden die zuvor vorgestellten Fusionsfunktionen genauer betrachtet, so lässt sich feststellen, dass $f_{\text{Summe}}(\boldsymbol{x})$ und $f_{\text{Mittelwert}}(\boldsymbol{x})$ Spezialfälle von $f_{\text{Gewichtet}}(\boldsymbol{x};\boldsymbol{\omega})$ sind. Bei entsprechender Wahl der Gewichte $\boldsymbol{\omega}$ lässt sich dies einfach zeigen:

$$f_{\text{Gewichtet}}(x_1,\ldots,x_N;\omega_1,\ldots,\omega_N) = f_{\text{Summe}}(x_1,\ldots,x_N), \text{ falls } \omega_i = 1 \tag{4.80}$$
$$f_{\text{Gewichtet}}(x_1,\ldots,x_N;\omega_1,\ldots,\omega_N) = f_{\text{Mittelwert}}(x_1,\ldots,x_N), \text{ falls } \omega_i = N^{-1}$$

Werden die Ergebnisse der Fusionsfunktionen bei einem anschließenden Verarbeitungsprozess nicht absolut betrachtet, sondern nur das Verhältnis untereinander, so liefern die Fusionsfunktionen $f_{\text{Summe}}(\boldsymbol{x})$ und $f_{\text{Mittelwert}}(\boldsymbol{x})$ gleiche Resultate, falls N konstant ist. In einem mehrstufigen Fusionsprozess, wie er in Abb. 4.4 dargestellt ist, ergeben sich jedoch Unterschiede, da die Anzahl der Eingangsgrößen bei den Teilaspekten in der Fusionsstufe 1 stets variiert und auch in Fusionsstufe 2 die Anzahl unterschiedlich sein kann, falls es sich um eine Person handelt, da dort der Teilaspekt Komplexität nicht definiert ist (vgl. Abschnitt 4.3.4). Somit ist die Fusionsfunktion $f_{\text{Mittelwert}}(\boldsymbol{x})$ der Funktion $f_{\text{Summe}}(\boldsymbol{x})$ für diese konkrete Anwendung vorzuziehen.

Abschließend wird die gewichtete Fusionsfunktion $f_{\text{Gewichtet}}(\boldsymbol{x};\boldsymbol{\omega})$ betrachtet, bei welcher der Einfluss der Teilaspekte durch die Gewichte ω_i vorgenommen werden kann. Die Bestimmung der Gewichte hängt von einer Vielzahl an Größen ab, welche i. d. R. über die Zeit veränderlich sind. Ist es für ein autonomes System a-priori oder zur Laufzeit nicht möglich die einzelnen Gewichte in jeder Fusionsstufe zu bestimmen (d. h. 14 Gewichte in zwei Fusionsstufen; vgl. Abb. 4.4), kann eine Gleichverteilung aller Teilaspekte $\omega_i = N^{-1}$ angenommen werden. Dies geschieht im Falle der Fusion durch $f_{\text{Mittelwert}}(\boldsymbol{x})$. Im Rahmen der vorliegenden Arbeit wird für die Bestimmung der Neugier die Fusionsfunktion $f_{\text{Mittelwert}}(\boldsymbol{x})$ gewählt, da zudem allen Teilaspekten derselbe Einfluss eingeräumt werden soll.

4.4 Zusammenhang mit der objektzentrierten Umwelterfassung

Bei der objektzentrierten Umwelterfassung (vgl. Kapitel 3) werden Objekte mit einer zunehmenden Anzahl an Sensorinformationen immer detaillierter erfasst. Dabei durchläuft ein Objektrepräsentant im Umweltmodell mehrere Spezialisierungsphasen. Die in Abschnitt 4.3 beschriebene wissensbasierte Neugier hängt direkt mit den Phasen der Spezialisierung zusammen.

4.4.1 Abhängigkeit der Neugier von den Spezialisierungsphasen

Mit jeder Spezialisierung gelangt ein Objektrepräsentant im Umweltmodell auf eine neue Ebene in der Klassenhierarchie. Dabei kommen neue Attribute

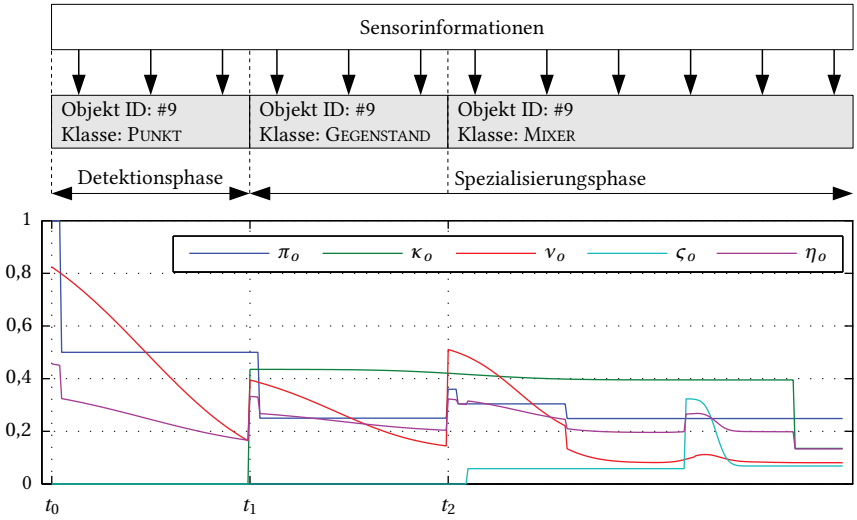

Abb. 4.5: Beispiel für den Verlauf der wissensbasierten Neugier η_o für ein Objekt in Zusammenhang mit der objektzentrierten Umwelterfassung (Detektions- und Spezialisierungsphase). Die Teilaspekte Neuartigkeit π_o, Komplexität κ_o, Unsicherheit ν_o und Konflikt ς_o zur Bestimmung der wissensbasierten Neugier sind ebenfalls aufgeführt. Hinweis: Die Zeitpunkte t_0, t_1 und t_2 liegen nur für Darstellungszwecke gleich weit auseinander. Die ersten beiden Phasen sind i. d. R. um ein Vielfaches kürzer als Phase 3.

und/oder Relationen hinzu, welche zuvor noch nicht oder unzureichend erfasst wurden. Diese Vorgehensweise hat einen direkten Einfluss auf die wissensbasierte Neugier und insbesondere auf die Teilaspekte Unsicherheit und Neuartigkeit. So kann durch eine Spezialisierung zunächst die Neugier aufgrund neuer Attribute und/oder Relationen ansteigen, da bei der Erfassung anfangs i. d. R. größere Unsicherheiten vorhanden sind, und erst mit weiteren Informationen (z. B. neuen Sensordaten) wiederum sinken.

In Abb. 4.5 ist dieser Zusammenhang beispielhaft veranschaulicht für einen Mixer, der neben weiteren Gegenständen auf einem Tisch steht. In der kurzen Detektionsphase sind zunächst die Neuartigkeit π_o und somit die Neugier groß, da insbesondere der Typ des Objekts unbekannt ist. Bei jeder Spezialisierung (t_1 und t_2) wird eine neue Stufe in der Klassenhierarchie (PUNKT → GEGENSTAND → MIXER) erreicht. Dabei werden neue Attribute und/oder Relationen hinzugefügt und infolgedessen steigt zunächst die Neugier aufgrund der Unsicherheit und Neuartigkeit an. Mit neuen Sensorinformationen sinkt im Laufe der Zeit die Neugier tendenziell wieder.

In Abb. 4.6 sind die einzelnen Teilaspekte der wissensbasierten Neugier für das vorherige Beispiel (vgl. Abb. 4.5) getrennt dargestellt für die einzelnen Phasen

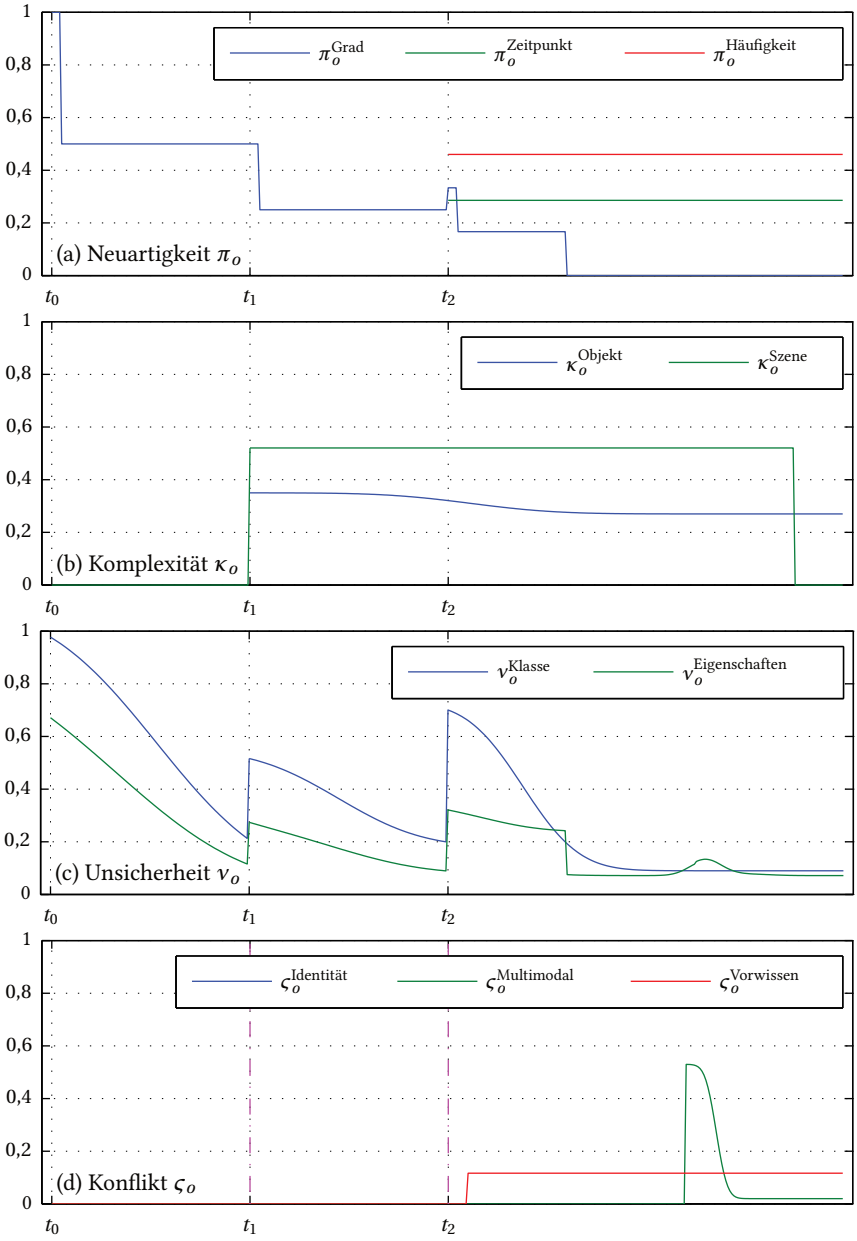

Abb. 4.6: Beispiel für den zeitlichen Verlauf der Teilaspekte Neuartigkeit (a), Komplexität (b), Unsicherheit (c) und Konflikt (d), welche als Grundlage zur Bestimmung der wissensbasierten Neugier (vgl. Abb. 4.5) dienen.

der Detektion und Spezialisierung. In den nachfolgenden Abschnitten werden die Teilaspekte anhand des Beispiels detailliert beschrieben:

Neuartigkeit

Der *Grad der Neuartigkeit* (π_o^{Grad}) hängt von den erfassten Objekteigenschaften ab. In den einzelnen Phasen besitzt der Objektrepräsentant o verschiedene Attribute und Relationen (Phase 1: *Position*, *Typ*; Phase 2: *Position*, *Typ*, *Gegenstandstyp*, *Farbe*; Phase 3: *Position*, *Typ*, *Gegenstandstyp*, *Farbe*, *Zustand*, *Besitzer*). Zu Beginn jeder Phase sind alle neuen Attribute zunächst unbekannt. Ausnahmen bilden die Attribute *Typ* und *Gegenstandstyp*, welche bis zum Ende der jeweiligen Phase unbekannt sind und danach den Wechsel in eine neue Stufe in der Klassenhierarchie veranlassen (vgl. Abschnitt 3.2). Die Relation Besitzer wird erst zu einem späteren Zeitpunkt mit Hilfe des a-priori-Wissens sicher bestimmt (vgl. Abb. 4.7), sodass erst damit der Grad der Neuartigkeit auf null sinkt (vgl. Abb. 4.6a).

Der *Zeitpunkt der Wahrnehmung* ($\pi_o^{\text{Zeitpunkt}}$) eines Objekts kann erst bestimmt werden, wenn eine konkrete Ausprägung der Klasse stattfindet. Dies ist erst ab Phase 3 der Fall (vgl. Abb. 4.6a), da dort der Gegenstandstyp Mixer ermittelt wird. Der Mixer wurde zuletzt vor 12 Tagen wahrgenommen, sodass sich ein entsprechender Wert (vgl. Gl. 4.13) ergibt.

Bei der *Häufigkeit der Wahrnehmung* ($\pi_o^{\text{Häufigkeit}}$) ist es wichtig, wie oft das Objekt in einem bestimmten Zeitraum wahrgenommen wurde. Liegt keine konkrete Ausprägung bzgl. der Klasse vor (d. h. Punkt, Gegenstand), so lässt sich dieser Aspekt nicht sinnvoll bestimmen. Mit Beginn von Phase 3 wird das Objekt als Mixer erkannt, welcher einen relativ hohen Wert für dieses Kriterium aufweist (vgl. Abb. 4.6a). Dies lässt sich darauf zurückführen, dass der Mixer nicht sehr häufig in den letzten 30 Tagen wahrgenommen wurde.

Komplexität

Die Komplexität ist, wie in Abschnitt 4.3.4 beschrieben, nur für Gegenstände definiert und infolgedessen lässt sich dieser Teilaspekt für das vorliegende Beispiel erst ab Phase 2 bestimmen (Punkt → Gegenstand → Mixer).

Die *Objektkomplexität* (κ_o^{Objekt}) ist relativ moderat, da die Beiträge zur Komplexität für einen Mixer nicht sehr groß sind. Seine äußere Form trägt maßgeblich, neben der leuchtend-blauen Farbe, zum Ergebnis bei (vgl. Abb. 4.6b).

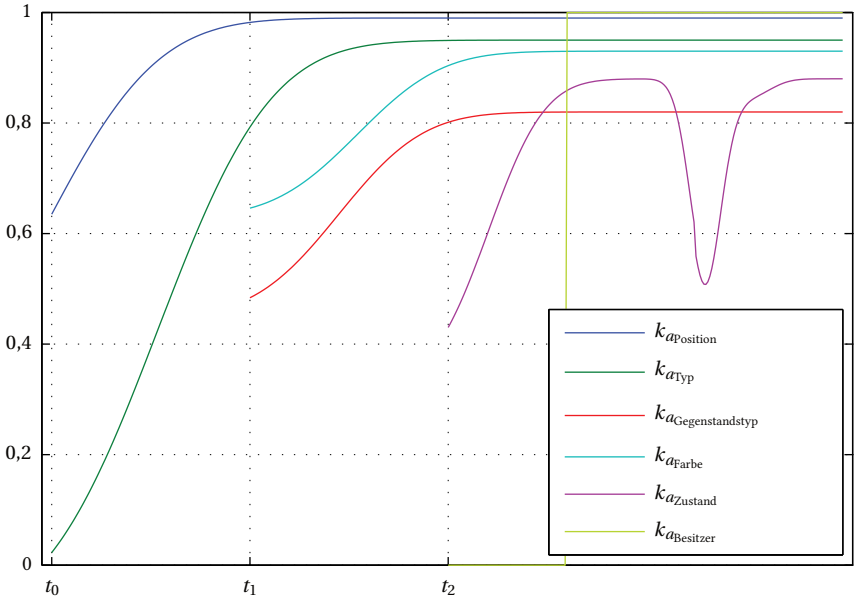

Abb. 4.7: Beispiel für den zeitlichen Verlauf der Konfidenzwerte zur Bestimmung der Unsicherheit (vgl. Abb. 4.6) im Kontext des Lebenszyklus eines Objekts. Der Konfidenzwert für den Zustand $k_{a_{\text{Zustand}}}$ sinkt aufgrund einer Zustandsänderung kurzzeitig ab, bevor dieser erneut ansteigt.

Die *lokale Szenenkomplexität* (κ_o^{Szene}) ist diesem Beispiel aufgrund der verschiedenen Gegenstände auf dem Tisch bis weit in Phase 3 recht hoch. Mit dem Einräumen der Spülmaschine werden alle Objekte bis auf den Mixer vom Tisch entfernt, sodass in dessen unmittelbarer Umgebung keine weiteren Gegenstände vorhanden sind. Somit sinkt die lokale Szenenkomplexität abrupt ab (vgl. Abb. 4.6b).

Unsicherheit

Die *Klassenunsicherheit* (v_o^{Klasse}) hängt direkt vom Konfidenzwert für die neueste Klasse in der Klassenhierarchie (Punkt ⟶ Gegenstand ⟶ Mixer) ab. Zu Beginn jeder Phase ist dieser recht niedrig und steigt im Verlauf der Zeit i. d. R. an. Somit sinkt im gleichen Maße die Klassenunsicherheit (vgl. Abb. 4.6c).

Die *Attributs- und Relationsunsicherheit* ($v_o^{\text{Eigenschaften}}$) hängt von den Konfidenzwerten der erfassten Attribute und Relationen (vgl. Abb. 4.7) ab. Zu Beginn jeder neuen Phase werden neue Eigenschaften erfasst, welche zunächst einen niedrigeren Konfidenzwert aufweisen. Dieser steigt mit weiteren Sensorinformati-

onen meist rasch an, sodass infolgedessen die Attributs- und Relationsunsicherheiten sinken (vgl. Abb. 4.6c). In diesem Beispiel ist in Phase 3 ein kurzfristiger Anstieg der Unsicherheit zu sehen. Zu diesem Zeitpunkt wird der Mixer eingeschaltet, dadurch sinkt zunächst kurzfristig der Konfidenzwert für das Attribut *Zustand* und infolgedessen steigt die Unsicherheit an. Mit neuen Sensorinformationen steigt der Konfidenzwert an und damit sinkt auch die Unsicherheit erneut (vgl. Abb. 4.6c).

Konflikt

Der *Konflikt bei Identitätsbestimmung* ($\varsigma_o^{\text{Identität}}$) ist für alle Gegenstände nicht definiert und wird somit bei der Bestimmung des Teilaspekts nicht berücksichtigt (vgl. Abb. 4.6d).

Ein *Konflikt bei multimodaler Wahrnehmung* ($\varsigma_o^{\text{Multimodal}}$) tritt immer dann auf, wenn beispielsweise akustische und visuelle Sensorinformationen zu stark unterschiedlichen Ergebnissen führen. Im Beispiel (vgl. Abb. 4.6d) wird der Mixer in Phase 3 (Mitte) eingeschaltet. Zunächst stimmt die akustische Wahrnehmung nicht mit der visuellen überein. Im ersten Moment wird der Mixer akustisch als Kaffeemühle wahrgenommen. Der Konflikt nimmt jedoch mit der Zeit recht schnell wieder ab, da neue akustische Sensorinformationen die Erkennung verbessern und den Mixer akustisch korrekt klassifizieren.

Ein *Konflikt mit dem a-priori-Wissen* ($\varsigma_o^{\text{Vorwissen}}$) tritt immer dann auf, wenn akquirierte Informationen in Konflikt mit vorab definiertem Wissen stehen. In Abb. 4.6d ist dies für das Attribut *Farbe* zu sehen, da in diesem Beispiel zur Veranschaulichung das vorhandene a-priori-Wissen künstlich eingeschränkt wurde und der gültige Wertebereich nur aus weißen, grauen und schwarzen Mixern besteht.

Zusammenfassung

Mit Hilfe der in Abschnitt 4.3.7 gewählten Fusionsfunktion lassen sich aus den einzelnen Graphen in Abb. 4.6 die Teilaspekte und somit auch die wissensbasierte Neugier bestimmen (Abb. 4.5), welche einen deutlichen Bezug zu den Spezialsierungsphasen im Umweltmodell aufweist.

4.4.2 Zurückverfolgung der wesentlichen Ursachen der Neugier

Bevor eine gezielte Reduzierung der Neugier erfolgen kann, muss zunächst deren Ursache(n) bestimmt werden. Dies wird durch die stufenweise Zurückverfolgung

in der Fusionshierarchie (vgl. Abb. 4.4) ermöglicht: In Abhängigkeit der verwendeten Fusionsfunktion kann beginnend mit der letzten Fusionsstufe, in umgekehrter Reihenfolge, für jede Stufe der höchste Beitrag bestimmt werden. Zunächst kann eine Hauptrichtung (Neuartigkeit, Komplexität, Unsicherheit oder Konflikt) ermittelt werden und anschließend sukzessive (falls eindeutig möglich) der Ursprung mit dem zunächst höchsten Beitrag bestimmt und reduziert (vgl. Abschnitt 4.4.3) werden. Sind mehrere Ursachen für eine erhöhte Neugier verantwortlich, so muss der Prozess der Zurückverfolgung unter Berücksichtigung bereits untersuchter Ursachen iterativ durchgeführt werden, bis entweder die Neugier unter eine vordefinierte Schwelle gefallen ist, die maximale Anzahl an Iterationen erreicht ist oder die Zeit, die für das Betrachten eines Objekts eingeplant war, verstrichen ist.

4.4.3 Reduzierung der Neugier

Zur Reduzierung der Neugier können je nach Quelle verschiedene Strategien verfolgt werden. Dazu ist es jedoch notwendig, wie in Abschnitt 4.4.2 angesprochen, die Ursache der Neugier zu identifizieren.

Liegt beispielsweise ein Konflikt mit dem a-priori-Wissen vor, so ist zu überprüfen, ob durch nähere Betrachtung des Objekts der Konflikt reduziert werden kann, da beispielsweise eine höhere Auflösung des Objekts im Kamerabild vorliegt und somit mehr Details erfasst werden können. Des Weiteren kann auch das a-priori-Wissen unvollständig sein, da zum Zeitpunkt der Erstellung die konkrete Ausprägung des Objekts nicht existierte. Um diesem Umstand entgegenzuwirken, müssen weitere Informationsquellen hinzugenommen werden. Das können neben online verfügbaren Datenbeständen auch Personen sein, welche im Rahmen eines Dialogs (d. h. mittels Mensch-Maschine-Interaktion) zur Klärung des Sachverhaltes beitragen können.

Bei den meisten Aspekten der Neugier kann eine Reduzierung durch „genaueres Hinsehen und -hören“ sowie durch den Einsatz von weiteren Ressourcen bzw. weiteren Algorithmen erreicht werden. Diese Tatsache gilt insbesondere für die Unsicherheit, da durch weitere und neue Sensorinformationen der Konfidenzwert für einzelne Attribute und/oder Relationen erhöht und die Unsicherheit somit reduziert werden kann.

Die Neuartigkeit hingegen nimmt i. d. R. mit häufigerer Wahrnehmung eines Objekts ab. Dabei werden die Informationen über ein Objekt langfristig in Form von Signaturen (Abschnitt 3.3; vgl. [Swe09]) abgelegt, welche außerdem zu einem detaillierten Objektwissen führen. Die Neuartigkeit eines Objekts sinkt außerdem durch häufigere und zeitnahe Wahrnehmungen, bei denen Objekte immer seltener neue Attribute oder Relationen aufweisen.

Die Teilaspekte der Komplexität sind davon unabhängig und teilweise konstant. Die vorhandene Textur und Form eines Objekts, welche sich bei einer erneuten Wahrnehmung nicht verändert, verursacht dieselbe Komplexität. Daher müssen Verfahren zur Abschwächung (vgl. Abschnitt 2.1; inhibition of return) dieser Art von Komplexität für das jeweilige Objekt angewendet werden, sodass erst nach einem längeren Zeitraum die Komplexität wieder ihre Wirkung entfaltet. Die Beschaffenheit der aktuellen Szene, welche das räumliche Zusammenspiel mehrerer Objekte (vgl. Abschnitt 4.3.2) beschreibt, liefert hingegen eine Komplexität für ein einzelnes Objekt, die abhängig von dessen aktueller Umgebung und somit zeitlich variabel ist.

4.5 Schlussbetrachtungen

Die wissensbasierte Neugier ist ein wichtiger Aspekt für autonome Systeme, da diese u. a. die Relevanz für eine genauere Betrachtung einzelner Objekte in einer Szene festlegt. In diesem Kapitel wurden dazu verschiedene Anforderungen für die Realisierung der wissensbasierten Neugier definiert:

Die perzeptive Neugier des Menschen wurde anhand von Erkenntnissen aus der Psychologie definiert; auch die dafür notwendigen situativen Bedingungen wurden herausgestellt. Dabei konnten die Neuartigkeit, die Komplexität, die Ungewissheit und der Konflikt als Teilaspekte der wissensbasierten Neugier identifiziert werden. Mit Hilfe der Objektdefinition auf Basis von Attributen und Relationen, deren Repräsentation in einem Umweltmodell sowie dem vorhandenen apriori-Wissen und den Zeitpunkten der Wahrnehmungen (vgl. Kapitel 3) konnten die Teilaspekte der wissensbasierten Neugier einzeln modelliert werden. Durch die Wahl einer geeigneten Fusionsfunktion ist es möglich, die Teilaspekte der Neugier zu einem Gesamtmaß für die Neugier zu vereinen und somit indirekt eine Priorisierung der Objekte vorzunehmen.

Im Rahmen der objektzentrierten Umwelterfassung ändert sich die Neugier im Laufe der verschiedenen Spezialisierungsphasen eines Objekts. Zu Beginn jeder Phase steigen die Neuartigkeit und die Unsicherheit aufgrund von neuen Attributen und Relationen i. d. R. zunächst kurzzeitig an und sinken bei genauerer Betrachtung und Akquise zusätzlicher Informationen oftmals recht schnell wieder. Dies lässt sich leicht nachvollziehen, da die Konfidenzwerte der Attribute und Relationen bei einzelnen Teilaspekten wichtige Faktoren sind und in dem beschriebenen Fall i. d. R. ansteigend sind. Weitere Gründe für eine erhöhte Neugier können durch eine detaillierte Betrachtung der einzelnen Teilaspekte identifiziert werden und entsprechend durch gezielte Maßnahmen (z. B. zur Auflösung von Konflikten) reduziert werden.

Im Anhang A ist ein weiteres ausführliches Beispiel zur Bestimmung der wissensbasierten Neugier von vier Objekten (zwei Personen und zwei Gegenständen) angefügt. Dabei sind die aktuell im Umweltmodell vorhandenen Daten der Objekte sowie das a-priori-Wissen exemplarisch aufgeführt.

Die wissensbasierte Neugier wird in Kapitel 5 im Rahmen der interessengetriebenen Exploration berücksichtigt und erfüllt dabei u. a. die Anforderung der Priorisierung von Objekten.

5

Interessengetriebene Exploration einer Szene

In diesem Kapitel wird die interessengetriebene Exploration, welche Aspekte des Menschen bei der Wahrnehmung einer Szene berücksichtigt, vorgestellt. In diesem Zusammenhang werden verschiedene Kriterien vorgestellt, welche die Priorität der Objekte während der Exploration beeinflussen und somit verschiedene Schwerpunkte setzen. In diesem Kontext repräsentieren die wissensbasierte Neugier und die multimodale Salienz wichtige Interessenaspekte für eine interessengetriebene Exploration.

5.1 Motivation

Für ein autonomes System, wie es beispielsweise ein humanoider Roboter ist, hat die Wahrnehmung der Umgebung durch Sensoren eine hohe Bedeutung. Diese bildet die Grundlage für eine Vielzahl an Aktionen (z. B. das Erledigen von Aufgaben oder die Interaktion mit Personen). Neben der reinen Wahrnehmung des aktuellen Ausschnitts einer Szene ist die gezielte Suche nach bestimmten Informationen, wie beispielsweise Gegenständen und Personen in der gesamten Szene, sehr wichtig. Die gezielte tiefgehendere Wahrnehmung einer Szene wird in diesem Kontext als *Exploration* bezeichnet und bildet die Grundlage dafür.

Die Exploration kann je nach Anwendungsdomäne jedoch ganz unterschiedlich aufgefasst werden. Im Folgenden werden hierzu verschiedene Beispiele gegeben, die alle eine gezielte Wissensakquisition gemein haben:

- Ein Schwarm an einfachen, mobilen Robotern kann eine unbekannte Umgebung mit seinen visuellen Sensoren explorieren und nach und nach eine vollständige Karte, z. B. für die Navigation, erstellen (vgl. [Yog09]). Dies wird

u. a. in Katastrophenszenarien (z. B. nach einem radioaktiven Unfall) notwendig, um das Leben von Menschen zu schützen und um einen ersten Eindruck vom Ausmaß der Katastrophe zu erhalten.

- In einem intelligenten Raum (engl: smart room) ist hingegen die Umgebung bekannt und mit einer Vielzahl an Sensoren ausgestattet (vgl. [Ess00], [Mos07], [Sti08]). Die darin befindlichen Personen und Gegenstände können einschließlich ihrer Interaktion erfasst und nachvollzogen werden. Eine Situations- und Intentionserkennung hilft beim ganzheitlichen Verständnis der aktuellen Abläufe.
- Aufgaben wie beispielsweise die aktive 3D-Objektsuche mit und ohne beschränktem Zeithorizont (vgl. [And11], [Wel11]) oder die physikalische Interaktion (haptische Exploration) mit unbekannten Gegenständen zum Erstellen einer 3D-Objektrepräsentation (vgl. [Sch11b], [Bie07]) werden ebenfalls als Exploration aufgefasst.
- Durch die visuelle Aufmerksamkeit in Form von Salienzkarten (vgl. Abschnitt 2.2) können die wichtigsten Elemente einer Szene vorrangig exploriert werden (vgl. [Itt98]). Des Weiteren können aus den Salienzkarten auch Landmarken für die Selbstlokalisation und die Navigation extrahiert werden (vgl. [Fri08], [Bur06], [Sar09]). Diese Vorgehensweise ist besonders hilfreich in unbekannten Umgebungen und dient der Orientierung.
- Allgemein kann bei der Szenenexploration eine Kombination von mehreren visuellen Sensoren genutzt werden, um ähnlich wie beim peripheren und fovealen Sehen des Menschen eine Detektion, Verfolgung und Klassifizierung von Objekten (vgl. [Ude03]) zu ermöglichen. Im Rahmen dieses Prozesses können mit fovealen Sensoren Detailinformationen von einzelnen Personen und Gegenständen wahrgenommen und mit peripheren Sensoren ein Überblick über die Umgebung gewonnen werden. Das dabei generierte Wissen weist einen unterschiedlichen Abstraktionsgrad auf (vgl. Abschnitt 3.2).

Die Exploration von verschiedenartigen Umgebungen, welche sich auch in einer Form von Überwachung von bestimmten Bereichen widerspiegelt, hat ein breites Feld an Anwendungen. Im Folgenden sind hierfür einige Anwendungsbeispiele bzw. Forschungsprojekte aufgeführt:

- Die maritime Überwachung von Staatsgrenzen und die Detektion von Ereignissen (vgl. [Fis12], [Tar09]) hat eine zunehmende Bedeutung erfahren. Dies wurde beispielsweise im Rahmen des EU-Projekts WIMA^{2}S – *Wide Maritime Area Airborne Surveillance* – untersucht (vgl. [WIM09]).
- Die aufgabenorientierte Videoüberwachung und -auswertung von bestimmten Personen, einzelnen Firmenbereichen oder abgelegenen Liegenschaften (vgl. [Mos10], [Mon11]) ist eine immer wichtiger werdende Anwendung der

Sensorerfassung. Dabei wird gerade in jüngster Zeit immer häufiger auch die Frage nach der Privatsphäre in solch einem Umfeld gestellt (vgl. [Vag12]). Im Rahmen der Eigenforschung des Fraunhofer-Instituts IOSB wurde dazu das Projekt NEST – *Network Enabled Surveillance and Tracking* – konzipiert, welches sich u. a. mit diesen Aufgabenstellungen befasst (vgl. [NES13]).

- Bei verheerenden Katastrophen, wie es z. B. schwere Naturkatastrophen, Chemie- bzw. Nuklearunfälle sind, ist eine umfassende Exploration des Katastrophenumfeldes wichtig. Durch den Einsatz und die Koordination von mobilen Robotern und Drohen kann der Ort des Geschehens erkundet und gesichert werden. Im Fraunhofer-Übermorgenprojekt SENEKA – *Sensornetzwerk mit mobilen Robotern für das Katastrophenmanagement* – wird diese Aufgabenstellung untersucht (vgl. [SEN13]).
- In der Zukunft sollen humanoide Roboter (vgl. [Asf06]) den Menschen im Alltag unterstützen. Dazu ist es wichtig, dass diese ihre Umgebung erfassen können (vgl. [Swe09], [Mac10a]) und den Menschen gezielt bei der Arbeit unterstützen oder diese für ihn übernehmen. Der Sonderforschungsbereich (SFB) 588 Humanoide Roboter – *Lernende und kooperierende multimodale Roboter* – beschäftigt sich mit dieser Fragestellung im Kontext eines Küchenszenarios (vgl. [Son12]).

Die vorliegende Arbeit ist dem Kontext der zuletzt genannten Anwendungsdomäne zuzuordnen und entstand im Rahmen des Sonderforschungsbereichs 588 im Teilbereich der perzeptuellen Exploration einer Szene. Ein humanoider Roboter hat, wie der Mensch auch, nur eingeschränkte Möglichkeiten bei der Wahrnehmung einer ganzen Szene. Während der Exploration einer Szene können nicht alle darin befindlichen Objekte mit einem hohen Detailgrad gleichzeitig erfasst und die dabei gewonnen Informationen verarbeitet werden. Ursachen dafür sind zum einen der aktuelle Sensorbereich, der die komplette Szene nicht zu jedem Zeitpunkt detailliert erfasst, sondern stets nur einzelne Ausschnitte durch foveale Sensoren im hohen Detailgrad abdeckt. Zum anderen ist aufgrund der vorhandenen begrenzten Rechenressourcen eine vollständige sowie gleichzeitige detaillierte Erfassung aller Informationen über die Objekte in der kompletten Szene nicht möglich.

Deshalb erfolgt eine Exploration der Szene, bei der für den Roboter wichtige Objekte möglichst früh erfasst werden sollen. Im Nachfolgenden wird dazu die sogenannte *interessengetriebene Exploration* eingeführt, bei der u. a. menschliche Aspekte bei der Priorisierung einzelner Objekte eine Rolle spielen. Die dort vorgestellte Herangehensweise ist jedoch nicht nur auf einen humanoiden Roboter beschränkt, sondern kann auch auf andere autonome Systeme, z. B. zur Überwachung von Liegenschaften, übertragen werden.

5.2 Grundlegende Definitionen

In diesem Abschnitt wird zunächst durch die Definition einiger Begriffe die grundlegende Vorgehensweise bei der Exploration einer Szene beschrieben, bevor anschließend im nächsten Abschnitt auf die interessengetriebene Exploration – eine spezielle Form der Exploration – näher eingegangen wird.

5.2.1 Explorationsstrategie

Die *Explorationsstrategie* definiert, nach welchem Priorisierungskriterium bzw. nach welchen Priorisierungskriterien die Abfolge der zu untersuchenden Gegenstände und Personen in einer Szene festgelegt wird. Im einfachsten Fall legt die Reihenfolge der Wahrnehmung der Objekte auch die Abfolge bei der Analyse der Objekte (vgl. Abschnitt 3.4) fest. Es können auch andere bzw. weitere Kriterien, wie die zuvor definierte multimodale Salienz oder die wissensbasierte Neugier, berücksichtigt werden. Letztendlich kann sogar die Bewegung, die zur Fokussierung eines Objekts notwendig ist, miteinbezogen werden.

5.2.2 Explorationspfad

Der *Explorationspfad* (EP) wird im Rahmen dieser Arbeit als eine konkrete Ausprägung der Explorationsstrategie für eine gegebene Menge an Objekten definiert. Der Explorationspfad setzt dabei die Priorisierungsziele der Explorationsstrategie um (vgl. [Küh12b]). Formal lässt sich ein Explorationspfad $\mathrm{EP} \in \mathcal{P}_{\mathcal{O}}$ als Permutation $\mathcal{P}(\cdot)$ aller Objekte $\mathcal{O}$ definieren:

$$\mathcal{P}_{\mathcal{O}} := \mathcal{P}(\mathcal{O}) = \left\{\xi : \mathcal{O} \to \mathcal{O} \mid \xi \text{ ist bijektiv}\right\} \quad \text{mit } N = |\mathcal{O}|\,. \tag{5.1}$$

Zur Veranschaulichung sind in einem Beispiel vier Objekte $o_1, \ldots, o_4$ vorhanden, und der dazugehörige Explorationspfad wurde mit einer zuvor festgelegten Explorationsstrategie bestimmt als $\mathrm{EP}_{\mathrm{Beispiel}} = (o_4, o_2, o_1, o_3)$. Hierbei wird zunächst das Objekt o_4, gefolgt von o_2 untersucht, danach folgen die Objekte o_1 und o_3.

Die Menge an möglichen Explorationspfaden steigt mit der Anzahl der vorhandenen Objekte an und wird durch die Anzahl an Objektpermutationen $|\mathcal{P}_{\mathcal{O}}| = N!$ bestimmt. Mit Hilfe des Priorisierungskriteriums bzw. der Priorisierungskriterien der Explorationsstrategie lässt sich bei der Bestimmung des Explorationspfades die Anzahl an zielführenden Pfaden i. d. R. stark reduzieren, sodass nur ein oder wenige Kandidaten (in Bezug auf die Explorationsstrategie) am Ende vorhanden sind.

5.2.3 Exploration

Im Rahmen der *Exploration einer Szene* soll eine tiefgehendere Wahrnehmung von Personen und Gegenständen in der aktuellen Umgebung erfolgen und somit Wissen akquiriert werden, welches zu einem späteren Zeitpunkt als Grundlage für die Interaktion mit Personen oder Gegenständen als auch zur Bewältigung von Aufgaben genutzt werden kann. Dazu wird zunächst der *Explorationspfad* (vgl. Abschnitt 5.2.2) initial anhand der ersten abstrakten Objektinformationen, der bereits vorhandenen Informationen im Umweltmodell und der gewählten *Explorationsstrategie* (vgl. Abschnitt 5.2.1), erstellt. Dies geschieht beispielsweise, nachdem ein Roboter einen ersten visuellen Schwenk durch einen Raum gemacht hat und dabei einen Großteil der vorhandenen Objekte bereits initial, d. h. auf sehr hohem Abstraktionsniveau, erfasst hat. Anschließend können die gefundenen Objekte entsprechend des Explorationspfads nach und nach untersucht werden. Dabei sinkt mit neuen Sensorinformation i. d. R. der Abstraktionsgrad der Objektrepräsentanten zunehmend (vgl. Abschnitt 3.2). Während der Exploration können zusätzlich neue Objekte wahrgenommen werden, welche z. B. aufgrund von Verdeckungen, der Größe oder fehlender Präsenz zuvor nicht erfasst wurden. Aus diesem Grund ist eine dynamische Anpassung des Explorationspfads an neue Gegebenheiten während der Exploration notwendig.

5.3 Interessengetriebene Szenenexploration

Nach der allgemeinen Definition der verschiedenen Begriffe der Exploration einer Szene im vorherigen Abschnitt wird nun darauf aufbauend und auf Basis der in Kapitel 3 beschriebenen objektzentrierten Umwelterfassung sowie unter Einbeziehung verschiedener Interessenaspekte eine spezielle Form der Szenenexploration – die interessengetriebene Szenenexploration – nachfolgend definiert und beschrieben.

5.3.1 Von der objektzentrierten Umwelterfassung zur interessengetriebenen Exploration

Bei der objektzentrierten Umwelterfassung (vgl. Kapitel 3; [Swe09], [Mac10a]) wird nach und nach aus den aktuellen Sensorinformationen immer mehr Wissen über Objekte im momentanen Erfassungsbereich der Sensoren akquiriert. Dies erfolgt auf mehreren Abstraktionsebenen und mit Hilfe einer Klassenhierarchie sowie unter Ausnutzung von Wissensabhängigkeiten (vgl. Abschnitt 3.2). Die Repräsentation der gewonnenen Objektinformationen sowie der Objekte selbst erfolgt dabei in einem Umweltmodell (vgl. Abschnitt 3.1).

Eine Erweiterung der objektzentrierten Umwelterfassung ist die in Abschnitt 5.2 definierte allgemeine Szenenexploration. Diese ergänzt die objektzentrierte Umwelterfassung um eine gezielte Exploration der Szene, welche eine Fokussierung von Objekten für eine ganzheitliche Erfassung aufweist. Somit kann beispielsweise ein humanoider Roboter nach und nach seine komplette Umgebung erfassen, um damit anfallende Aufgaben in Zukunft schneller bewältigen zu können und stets ein Bild über die aktuelle Szene zu besitzen. Insbesondere in vollständig neuen und somit unbekannten Umgebungen kann bei vollständiger Erfassung ein solcher Vorgang eine lange Zeit in Anspruch nehmen, was i. d. R. nicht optimal ist.

Bei der interessengetriebenen Exploration einer Szene wird dieser Problematik entgegengetreten, indem den Objekten in einer Szene unterschiedliche Bedeutungen zugeordnet werden. D. h., während dieser Form der Exploration werden die Objekte zuerst und intensiver erfasst, welche eine höhere Wichtigkeit besitzen als andere. Bei dieser Vorgehensweise werden die wichtigsten Objekte bereits früh erfasst. Dies ist besonders vorteilhaft, falls aus zeitlichen oder anderen Gründen die Exploration vorzeitig unterbrochen und somit nicht bis zum Ende durchgeführt wird. Dazu ist es jedoch notwendig, Kriterien für die Priorisierung der Objekte einzuführen. Die interessengetriebene Exploration beschäftigt sich im Nachfolgenden mit genau dieser Fragestellung.

Die Perzeption einzelner Objekte erfolgt dabei durch eine Fokussierung im Rahmen der Exploration und der in Abschnitt 3.2 beschriebenen Vorgehensweise zur sukzessiven Erfassung und Repräsentation von Objekten in einem Umweltmodell. Die notwendigen Anforderungen und Voraussetzungen an die interessengetriebene Exploration sind in den folgenden Unterabschnitten beschrieben.

Voraussetzungen für die interessengetriebene Exploration

Im Rahmen der vorliegenden Arbeit werden einige zusätzliche Voraussetzungen für die Exploration festgelegt, welche im Nachfolgenden aufgeführt sind:

- *vorhandenes Grundwissen*: Es wird ein gewisses Grundwissen über die Szene, wie z. B. die Raumgeometrie und die Möbel, als vorhanden vorausgesetzt, welches im Rahmen der Exploration nicht weiter untersucht werden soll. Die Akquisition dieses Wissens ist entweder a-priori gegeben oder wurde bereits zuvor erfasst.
- *Wahrnehmung von Objekten*: Während der Exploration werden Gegenstände auf Oberflächen und an Wänden erfasst und die gewonnenen Informationen werden in einem Umweltmodell repräsentiert (vgl. Kapitel 3). Des Weiteren werden auch Personen, die sich ebenfalls aktuell im Raum aufhalten, berücksichtigt und wahrgenommen.

Anforderungen an die interessengetriebene Exploration

Die interessengetriebene Exploration besitzt im Vergleich zu der zuvor beschriebenen allgemeinen Exploration (vgl. Abschnitt 5.2.3) zusätzliche Anforderungen, welche im Folgenden aufgeführt sind:

- *Erweiterung der objektzentrierten Umwelterfassung*: Die in Kapitel 3 vorgestellte objektzentrierte Umwelterfassung wird u. a. durch die gezielte Fokussierung der Sensoren auf einzelne Objekte zur ganzheitlichen Exploration der Szene erweitert. Dazu muss eine passende Explorationsstrategie festgelegt werden und ein Explorationspfad bestimmt werden, um die Objekte sukzessiv zu fokussieren und detaillierter zu erfassen.
- *priorisierte Erfassung der Umgebung*: Personen und Gegenstände, die bestimmte Kriterien erfüllen (z. B. „unbekannt", „lange nicht gesehen" oder „salient") werden priorisiert und somit früher als andere Objekte untersucht, die diese Kriterien nicht oder weniger erfüllen.
- *Anlehnung an den Menschen*: Erkenntnisse über die Salienz und die Neugier des Menschen werden genutzt, um die Exploration einer Szene zu verbessern. Bei der Bestimmung der Salienz werden im Rahmen der vorliegenden Arbeit visuelle und/oder akustische Sensorinformation genutzt, um besonders hervorstechende Objekte und Geräusche in der zu erfassenden Szene zu bestimmen (vgl. Kapitel 2). Die Neugier hingegen wird nicht nur anhand von Sensorinformationen bestimmt, sondern hauptsächlich mit Hilfe von bereits gewonnenem Wissen im Umweltmodell und noch fehlender Objektinformationen (vgl. Kapitel 4). Die Neugier und die Salienz ermöglichen somit eine Fokussierung bzw. Priorisierung von Objekten im Rahmen der Exploration.
- *Einfluss von externen Faktoren auf den Ablauf der Exploration*: Während der Exploration können natürlich jederzeit neue Objekte hinzukommen, sodass der Explorationspfad dynamisch angepasst werden muss. Dies gilt auch für besonders prägnante akustische Ereignisse – außerhalb des aktuellen Sichtfeldes –, welche eine hohe Priorisierung hervorrufen.

  Neben der laufenden Anpassung des Explorationspfads ist es außerdem wichtig, dass ein humanoider Roboter oder ein anderes autonomes System jederzeit den regulären Explorationsprozess für andere Dinge unterbrechen kann, um beispielsweise eine wichtige Aufgabe zu erledigen oder mit einem Menschen zu interagieren. Diese sogenannten „High-Level"-Prozesse, welche die aktuelle Exploration unterbrechen können, sind essentiell für einen humanoiden Roboter (vgl. [Küh12a]). Noch nicht explorierte Objekte können anschließend untersucht werden. Bereits untersuchte Objekte werden durch entsprechende Unterdrückungsmechanismen i. d. R. nicht noch ein zweites Mal für die Exploration ausgewählt (engl.: *inhibition of return*; vgl. [Sch11a]).

In den folgenden Abschnitten wird beschrieben, wie die einzelnen Anforderungen an die interessengetriebene Exploration erfüllt werden können.

5.3.2 Interessengetriebene Perzeption

Die interessengetriebene Perzeption bildet die Grundlage für die interessengetriebene Exploration einer Szene und erfüllt dabei die Anforderungen an eine priorisierte Erfassung der Umwelt in Anlehnung an den Menschen.

Bei der interessengetriebenen Perzeption wird eine Szene unter verschiedenen Gesichtspunkten wahrgenommen, welche alle eine Form von Interesse implizieren. Die sogenannten *Interessenaspekte* beschreiben Eigenschaften, die das Interesse für ein Objekt erhöhen und sind definiert – im Rahmen der vorliegenden Arbeit – durch eine Kombination aus Salienz und Neugier, welche durch den aktuell wahrnehmbaren Ausschnitt der Umwelt ergänzt werden. Diese Aspekte stellen eine priorisierte Sichtweise auf die Umwelt dar, bei der einzelne Objekte besonders hervorgehoben und andere abgeschwächt bzw. ausgeblendet werden.

Multimodale Salienz

Ein wichtiger Aspekt der interessengetriebenen Perzeption ist die multimodale Salienz (vgl. Kapitel 2). Diese hebt Objekte, welche sich aufgrund von besonderen audiovisuellen Merkmalen von ihrer Umgebung abheben, besonders hervor. In diesem Zusammenhang werden Objekte mit einer höheren Salienz als interessanter von einem Menschen bzw. durch ein autonomes System wahrgenommen als Objekte mit einer geringeren Salienz. Im Rahmen der vorliegenden Arbeit wurde in Kapitel 2 die audiovisuelle Salienz in Form von multimodalen Salienzclustern modelliert, welche eine Bewertung einzelner Bereiche der Szene vornimmt.

Wissensbasierte Neugier

Ein weiterer wichtiger Aspekt der interessengetriebenen Perzeption ist die wissensbasierte Neugier (vgl. Kapitel 4). Diese ist maßgeblich für das Interesse an einem Objekt verantwortlich. Die wissensbasierte Neugier umfasst die Teilaspekte Neuartigkeit, Komplexität, Unsicherheit und Konflikt, welche durch die Merkmale eines Objekts und das vorhandene bzw. fehlende Wissen bestimmt werden (vgl. Abschnitt 4.3). Objekte, welche eine größere Neugier induzieren als andere, erwecken auch ein höheres Maß an Interesse und treten bei der Betrachtung mehrerer Objekte infolgedessen in den Vordergrund.

Aktueller Umweltausschnitt

Der letzte Aspekt der interessengetriebenen Perzeption ist der aktuelle Ausschnitt der Umwelt, welcher durch die Sensoren momentan erfasst werden kann. Analog dazu ist beim Menschen das aktuelle Blickfeld, in dem Objekte in Abhängigkeit ihrer Position ohne weitere Bewegung entweder foveal oder nur peripher wahrgenommen werden. Bei vielen Aufgaben im Alltag nimmt der Mensch eine gezielte Fokussierung vor, da Objekte nur im Sehzentrum detailliert wahrgenommen werden können (foveale Wahrnehmung). Eine gleichzeitige Erfassung von Objektkonturen und insbesondere Bewegungen in den Randbereichen des Sichtfelds ist ebenfalls möglich (periphere Wahrnehmung), sodass der Mensch Objekte im gesamten Blickfeld ohne Fokussierung verfolgen kann. Zusätzlich können durch bereits geringe Bewegungen der Augen diese Objekte schnell fokussiert und somit detailliert erfasst werden. Übertragen auf ein autonomes System kann durch den Einsatz von mehreren Kameras mit unterschiedlicher Brennweite ein analoges Verhalten abgebildet werden. So können Objekte im Zentrum detailliert erfasst werden, während Objekte im Randbereich weiterhin nachverfolgt werden können.

Durch die Berücksichtigung des aktuell wahrnehmbaren Umweltausschnitts lässt sich somit allgemein ein Verhalten ableiten, bei dem Objekte bevorzugt werden, die sich direkt im Zentrum des Blickfelds befinden und somit detailliert wahrgenommen werden, im Gegensatz zu anderen Objekten, die eine immer größer werdende Bewegung für die Fokussierung erfordern.

Erweiterung des Umweltmodells

Für die interessengetriebene Perzeption wird die Definition des Umweltmodells (vgl. Abschnitt 3.1) erweitert, um die angesprochenen Aspekte aufzunehmen. Hierfür wird zunächst die Definition eines Objektrepräsentanten (vgl. Gl. 3.2) zu einem 4-Tupel erweitert:

$$o = (\mathcal{A}_o, \mathcal{R}_o, \mathcal{Z}_o, s_o) \in \mathcal{O} \tag{5.2}$$

Hierbei besteht jeder Repräsentant weiterhin aus der Menge an Attributen $\mathcal{A}_o$, Relationen $\mathcal{R}_o$ und Zeitpunkten der Wahrnehmung $\mathcal{Z}_o$. Der Salienzwert s_o (vgl. Kapitel 2) kann durch die räumliche Zuordnung der vorhandenen Salienzcluster (vgl. Gl. 2.64) zu den im Umweltmodell repräsentierten Objekten bestimmt werden.

Die letzten beiden Aspekte der interessengetriebenen Perzeption werden in Form von Relationen zwischen einem Beobachter – beispielsweise einem humanoiden

Roboter – und den Objektrepräsentanten im Umweltmodell modelliert. Die Relationen zwischen dem Roboter und den Objektrepräsentanten $\mathcal{O}$ sind wie folgt definiert:

$$\mathcal{R}_{\text{Roboter}\leftrightarrow\mathcal{O}} := \left\{ r_1^{o_1}, r_2^{o_1}, \ldots, r_j^{o_i}, r_{j+1}^{o_i}, \ldots, r_{J-1}^{o_N}, r_J^{o_N} \right\} \quad \text{mit } N, J \in \mathbb{N}^+ \tag{5.3}$$

und

$$r_j^{o_i} := \left(t_{r_j}^{o_i}, w_{r_j}^{o_i}, v_{r_j}^{o_i} \right). \tag{5.4}$$

In diesem Zusammenhang beschreibt $r_j^{o_i}$ die j-te Relation zwischen dem Beobachter und dem Objektrepräsentanten o_i. Jede Relation r_j besteht dabei aus einem Relationstyp t_{r_j}, einem Relationswert w_{r_j} und zusätzlich einem Verweis v_{r_j} auf einen Repräsentanten im Umweltmodell (ähnl. Gl. 3.4).

Die wissensbasierte Neugier η_{o_i}, welche durch das i-te Objekt hervorgerufen wird (vgl. Kapitel 4; Gl. 4.71), ist formal über die Relation $r_{\text{Neugier}}^{o_i}$ mit dem Objektrepräsentanten o_i verknüpft:

$$r_{\text{Neugier}}^{o_i} := \left(t_{r_{\text{Neugier}}}^{o_i}, w_{r_{\text{Neugier}}}^{o_i}, v_{r_{\text{Neugier}}}^{o_i} \right) = \left(\text{Neugier}, \eta_{o_i}, o_i \right) \in \mathcal{R}_{\text{Roboter}\leftrightarrow\mathcal{O}}\,. \tag{5.5}$$

Die benötigte Ausrichtung der Sensoren (vgl. Abschnitt 5.4.1) zur Fokussierung des Objekts o_i wird über den Parameter $\boldsymbol{q}_{o_i}$ festgelegt. Diese kann für eine Anpassung des aktuell erfassbaren Umweltausschnitts genutzt werden, um beispielsweise mit fovealen Sensoren eine deutlich verbesserte Erfassung eines Objekts (d. h. höhere Auflösung und verbessertes Signal-zu-Rausch-Verhältnis) zu ermöglichen. Formal wird dieser Zusammenhang über die Relation $r_{\text{Fokus}}^{o_i}$ beschrieben, die wie folgt definiert ist:

$$r_{\text{Fokus}}^{o_i} := \left(t_{r_{\text{Fokus}}}^{o_i}, w_{r_{\text{Fokus}}}^{o_i}, v_{r_{\text{Fokus}}}^{o_i} \right) = \left(\text{Fokus}, \boldsymbol{q}_{o_i}, o_i \right) \in \mathcal{R}_{\text{Roboter}\leftrightarrow\mathcal{O}}\,. \tag{5.6}$$

5.3.3 Interessengetriebene Explorationsstrategie

Die interessengetriebene Explorationsstrategie bestimmt den Explorationspfad und bildet in Kombination mit der Erweiterung der objektzentrierten Erfassung der Umwelt eine Grundlage für die interessengetriebene Exploration. Diese kann jederzeit durch externe Einflussfaktoren beeinflusst werden, sodass Mechanismen zur Unterbrechung und Fortführung der Exploration notwendig sind.

Externe Einflussfaktoren

Bei allen Einflüssen, die während der Exploration von außen auf ein autonomes System einwirken (z. B. Ereignisse, Aufgaben oder Interaktion), ist es wichtig,

dass wenn diese die Exploration unterbrechen, eine anschließende Fortsetzung unter Berücksichtigung der bereits explorierten Objekte möglich ist. Dazu ist es notwendig, dass möglichst viele Objekte währenddessen nachverfolgt und die restlichen schnell – beispielsweise über ihre Position oder durch markante Merkmale – wiedergefunden werden.

Bei stationären autonomen Systemen kann diese Problematik gelöst werden, indem neben den verteilten Kameras beispielsweise eine Pan-Tilt-Zoom-Kamera (kurz: PTZ-Kamera) für die Exploration eingesetzt wird. Die verteilten Kameras können dabei die Gegenstände und Personen in der Umgebung nachverfolgen, während mit einer PTZ-Kamera einzelne Objekte fokussiert und näher untersucht werden können.

Im Fall von humanoiden Robotern, wie beispielsweise ARMAR-III (vgl. [Asf08]) oder DynamicBrain DB (vgl. [Atk00]), welche über keine externen Sensoren verfügen, werden mehrere Kameras mit unterschiedlicher Brennweite für die Erfassung der Umwelt eingesetzt. In Anlehnung an das periphere Sehen beim Menschen wird mit einer Kamera mit geringer Brennweite ein Großteil der Szene im Überblick wahrgenommen und mit Hilfe einer zweiten Kamera mit großer Brennweite wird – analog zum fovealen Sehen – ein kleinerer Ausschnitt der Szene wesentlich genauer betrachtet (vgl. [Ude03]).

Allgemein können nach einer Unterbrechung der Exploration, ausgehend von den letzten bekannten Positionen, die noch ausstehenden Objekte, aber auch neue Objekte bei der Fortsetzung berücksichtigt werden. Dabei wird der noch fehlende Abschnitt des Explorationspfads an die neue Situation angepasst: Nicht mehr vorhandene Objekte werden entfernt, neue Objekte werden ergänzt und die noch nicht explorierten Objekte werden aktualisiert. Die nun auf dem noch fehlenden angepassten Abschnitt des Explorationspfads befindlichen Objekte werden sukzessive mit Hilfe der in Kapitel 3 beschriebenen Vorgehensweise detailliert erfasst. Eine erneute vollständige Exploration der Szene ist mit diesem Ansatz nicht notwendig, da die bereits explorierten Objekte nachverfolgt und bei der Fortsetzung i. d. R. nicht zum wiederholten Male betrachtet werden.

Die Anforderung an die interessengetriebene Exploration, dass der Einfluss von externen Faktoren beim Ablauf der Exploration berücksichtigt wird, ist mit dieser Herangehensweise erfüllt.

Erweiterung der objektzentrierten Erfassung

Der interessengetriebene Explorationsansatz erweitert den Gedanken der allgemeinen Exploration in Abschnitt 5.2 um den Faktor Interesse und berücksichtigt dabei die objektzentrierte Vorgehensweise bei der Erfassung der Umwelt (vgl. Kapitel 3; [Swe09], [Mac10a]).

Mit Hilfe der interessengetriebenen Explorationsstrategie lässt sich die Priorität einzelner Objekte während der Exploration bestimmen. Die Wichtigkeit einzelner Objekte leitet sich aus den Interessensaspekten aus Abschnitt 5.3.2 ab. Durch eine gezielte Gewichtung der drei Aspekte *wissensbasierte Neugier*, *multimodale Salienz* und *aktueller Umweltausschnitt* kann Einfluss auf die aktuelle Explorationspriorität der einzelnen Objekte genommen und somit auch ein Schwerpunkt bei der Exploration gesetzt werden. Die formale Definition der interessengetriebenen Explorationsstrategie zur Bestimmung eines interessengetriebenen Explorationspfads erfolgt in Abschnitt 5.4.2.

Der interessengetriebene Explorationspfad bestimmt die Priorisierung und somit letztendlich die Reihenfolge der zu untersuchenden Objekte während der Exploration. Dazu wird im Rahmen der Exploration das aktuelle Objekt auf dem Pfad ausgewählt und mit den Sensoren fokussiert. Anschließend wird mit Hilfe der objektzentrierten Erfassung das Objekt nach und nach erfasst (vgl. Kapitel 3). Dabei sind zu Beginn zunächst wenige Informationen vorhanden und der Objektrepräsentant umfasst noch sehr abstrakt Informationen. Im Laufe einer tiefgehenderen Wahrnehmung und mit der Zeit können immer mehr und sichere Informationen generiert werden, sodass das Objekt immer detaillierter im Umweltmodell repräsentiert werden kann. Dabei werden gezielt die Aspekte der Neugier berücksichtigt, sodass die Neugier im Verlauf der Zeit insgesamt sinkt. Ist die Neugier hinreichend gering, so wird anschließend mit dem nächsten Objekt auf dem Pfad fortgefahren (vgl. Abschnitt 6.4.2).

5.4 Verschiedene Explorationsstrategien

Im Nachfolgenden werden neben der zuvor vorgestellten interessengetriebenen Explorationsstrategie noch weitere Explorationsstrategien formal definiert, welche im Rahmen der späteren Evaluation (vgl. Kapitel 7) gegenübergestellt werden. Dabei werden die Explorationsstrategien unterteilt vorgestellt nach der Anzahl der berücksichtigten Priorisierungskriterien. Abschließend erfolgt für jede einzelne Strategie eine Abschätzung der Komplexität bei der Bestimmung des Explorationspfads in Abhängigkeit von der Anzahl an berücksichtigten Objekten.

5.4.1 Explorationsstrategien mit einem Priorisierungskriterium

Zunächst werden Explorationspfade vorgestellt, bei denen Explorationsstrategien mit einem Priorisierungskriterium als Grundlage dienen. Neben der Referenzstrategie, bei der die Priorität der Objekte anhand der Zeitpunkte der Wahrnehmung bestimmt wird, werden weitere Explorationsstrategien mit Salienz,

Neugier und Bewegung als einzelne Priorisierungskriterien für die spätere Evaluation getrennt vorgestellt.

Referenzexplorationsstrategie

Der Referenzexplorationspfad $\mathrm{EP}_{\mathrm{Referenz}}$ ist über eine Explorationsstrategie definiert, bei der die Priorität der Objekte während der Exploration durch die Reihenfolge der Wahrnehmung bestimmt wird, d. h., Objekte, die früher als andere erfasst wurden, werden auch entsprechend früher vollständig exploriert. Der Referenzexplorationspfad ist definiert als

$$\mathrm{EP}_{\mathrm{Referenz}} = \left(o_{i_1}, o_{i_2}, \ldots, o_{i_N}\right) \quad \text{mit } t_{o_{i_1}} \leq t_{o_{i_2}} \leq \ldots \leq t_{o_{i_N}}\,. \tag{5.7}$$

Hierbei werden die Objekte den Zeitpunkten der Wahrnehmung $t_{o_{i_1}}, \ldots, t_{o_{i_N}}$ entsprechend aufsteigend sortiert. Frühere Zeitpunkte sind dabei per Definition kleiner als spätere Zeitpunkte. Die Objekte $o_1, \ldots, o_N$ werden dabei durch die Indices $i_1, \ldots, i_N$ auf die Reihenfolge des Referenzexplorationspfads abgebildet.

Salienzbasierte Explorationsstrategie

Bei dieser Strategie werden, wie zuvor auch, alle erfassten und noch nicht untersuchten Objekte berücksichtigt. Diese werden bei der salienzbasierten Explorationsstrategie entsprechend ihrer Salienz $\boldsymbol{s}$ absteigend sortiert und anschließend analysiert. Hierbei werden, wie bei den anderen Strategien auch, ebenfalls Objekte miteinbezogen, die sich zurzeit nicht im Blickfeld des Roboters befinden, jedoch im Umweltmodell repräsentiert sind. Der salienzbasierte Explorationspfad $\mathrm{EP}_{\mathrm{Salienz}}$ lässt sich definieren als

$$\mathrm{EP}_{\mathrm{Salienz}} = \left(o_{i_1}, o_{i_2}, \ldots, o_{i_N}\right) \quad \text{mit } s_{o_{i_1}} \geq s_{o_{i_2}} \geq \ldots \geq s_{o_{i_N}}\,. \tag{5.8}$$

Die räumlichen Zusammenhänge zwischen den einzelnen Objekten werden bei dieser Explorationsstrategie ähnlich der neugierbasierten Explorationsstrategie nicht berücksichtigt. Dies führt in der Regel zu einer hohen Aktivität bzw. hohen Eigenbewegungen zur Fokussierung der Objekte während der Exploration.

Neugierbasierte Explorationsstrategie

Die neugierbasierte Explorationsstrategie berücksichtigt bei der Priorisierung die aktuelle wissensbasierte Neugier $\boldsymbol{\eta}$ für die Objekte. Dabei werden auf dem Explorationspfad Objekte, die eine höhere Neugier hervorrufen, früher ausgewählt als andere. Der neugierbasierte Pfad $\mathrm{EP}_{\mathrm{Neugier}}$ ist definiert als

$$\mathrm{EP}_{\mathrm{Neugier}} = \left(o_{i_1}, o_{i_2}, \ldots, o_{i_N}\right) \quad \text{mit } \eta_{o_{i_1}} \geq \eta_{o_{i_2}} \geq \ldots \geq \eta_{o_{i_N}}. \tag{5.9}$$

Entsprechend werden die Objekte absteigend nach dem Grad der induzierten Neugier selektiert. Diese Vorgehensweise hat den Vorteil, dass Objekte entsprechend ihrer hervorgerufenen Neugier möglichst schnell ausgewählt werden. Andere Einflussfaktoren, wie z. B. die benötigte Bewegung, die Salienz usw., werden dabei nicht explizit berücksichtigt.

Bewegungsoptimierte Explorationsstrategie

Die bewegungsoptimierte Explorationsstrategie berücksichtigt bei der Bestimmung des Explorationspfads implizit den aktuellen Umweltausschnitt. Dies geschieht, indem die Bewegung für die Fokussierung aller Objekte minimiert wird. Räumlich benachbarte Objekte werden dabei auch in zeitlich enger Reihenfolge betrachtet. Die konkrete Optimierung des Explorationspfads hängt von der eingesetzten Hardware bzw. dem Roboter ab. Nimmt man vereinfachend an, dass keine translatorischen Bewegungen für die Fokussierung eines Objekts notwendig sind, so lässt sich das Problem über die Minimierung von Gelenkwinkelbewegungen lösen. Ein bewegungsoptimaler Explorationspfad $\mathrm{EP}_{\mathrm{Bewegung}}$ für eine vorhandene Menge an Objekten lässt sich bestimmen durch Berücksichtigung der Bewegung in Form von sogenannten *Gelenkwinkeldistanzen*. Formal lässt sich dies beschreiben durch

$$\mathrm{EP}_{\mathrm{Bewegung}} = \underset{\mathrm{EP} \in \mathcal{P}_{\mathcal{O}}}{\arg\min} \left\{ \sum_{k=1}^{N} d_{o_{i_k}} \right\} \quad \text{mit } d_{o_{i_k}} = \left\| \boldsymbol{q}_{o_{i_k}} - \boldsymbol{q}_{o_{i_{k-1}}} \right\|, \tag{5.10}$$

wobei i_k das k-te Objekt auf dem aktuellen Explorationspfad $\mathrm{EP} \in \mathcal{P}_{\mathcal{O}}$ ist. $\boldsymbol{q}_{o_{i_k}}$ stellt die Gelenkwinkelstellung dar, die notwendig ist, um das k-te Objekt zu fokussieren. Dementsprechend ist $\boldsymbol{q}_{o_{i_{k-1}}}$ die Gelenkwinkelstellung des vorherigen Objekts auf dem Explorationspfad. Die initiale Stellung der Gelenkwinkel $\boldsymbol{q}_{o_{i_0}}$ ist dabei die Ausgangskonfiguration der Winkel bei der ersten Berechnung des Explorationspfads. Die L1-Norm der Gelenkwinkeländerungen $d_{m,n} = \left\| \boldsymbol{q}_m - \boldsymbol{q}_n \right\|$ dient hierbei als Maß für die Veränderung zwischen zwei Gelenkwinkelstellungen (kurz: Gelenkwinkeldistanz; engl.: joint angle distances; JAD) und damit als Maß für die notwendige Bewegung. Durch die Minimierung der Gelenkwinkeldistanzen über die Menge an möglichen Objektpermutationen $\mathcal{P}_{\mathcal{O}}$, d. h. Explorationspfaden, wird abschließend der optimale Pfad bestimmt.

Allgemein führt die bewegungsoptimierte Explorationsstrategie zu einer Verringerung der Eigenbewegung und somit auch zu einem zeit- und energieeffizienteren Pfad. Der Nachteil ist – wie bei allen zuvor vorgestellten Explorationsstrategien –, dass keine weiteren Aspekte berücksichtigt werden und somit die bewegungsoptimierte Explorationsstrategie weder den Zeitpunkt der Objektwahrnehmung noch die Salienz oder die induzierte Neugier der Objekte miteinbezieht.

5.4.2 Explorationsstrategien mit mehreren Priorisierungskriterien

Die zuvor vorgestellten Explorationsstrategien berücksichtigen immer nur ein Priorisierungskriterium gleichzeitig. Dies soll nachfolgend durch die Einbeziehung von weiteren Priorisierungskriterien und einer gesamtheitlichen Optimierung zu einem generellen Ansatz erweitert werden.

Allgemeine Definition

Im Rahmen einer Szenenexploration sollen verschiedene Einflussfaktoren bei der Selektion von Objekten berücksichtigt werden. Dies lässt sich allgemein definieren über

$$\mathrm{EP}_{\mathrm{Allgemein}} := \underset{\mathrm{EP} \in \mathcal{P}_{\mathcal{O}}}{\arg\max} \left\{ \sum_{k=1}^{N} \gamma_1 \cdot c_1\left(o_{i_k}\right) + \gamma_2 \cdot c_2\left(o_{i_k}\right) + \ldots + \gamma_J \cdot c_J\left(o_{i_k}\right) \right\} \tag{5.11}$$

mit

$$c_j\left(o_{i_k}\right) \in [0,1], \quad \gamma_1 + \gamma_2 + \ldots + \gamma_J = 1, \quad \gamma_j \in [0,1] \quad \text{und} \quad J \in \mathbb{N}. \tag{5.12}$$

Hierbei liefert die Funktion $c_j\left(o_{i_k}\right)$ den normierten Wert des j-ten Priorisierungskriteriums für das Objekt o_{i_k} und γ_j stellt einen dazu korrespondierenden Einflussparameter dar. Wichtig hierbei ist, dass der Wertebereich der Priorisierungskriterien entsprechend so definiert ist, dass hohe Werte das Kriterium sehr gut erfüllen und niedrigere Werte entsprechend weniger gut. Dieser Ansatz lässt sich leicht erweitern durch die Hinzunahme neuer Priorisierungskriterien.

Interessengetriebene Explorationsstrategie

In der vorliegenden Arbeit sind die aktuelle Salienz eines Objekts und die wissensbasierte Neugier die beiden Haupteinflussfaktoren für die interessengetriebene Exploration. Neben diesen ist die Bewegung zum Fokussieren eines Objekts eine weitere wichtige Größe. Dabei wird der aktuelle Umweltausschnitt berücksichtigt, sodass sichtbare Objekte im Vergleich zu Objekten außerhalb des Blickfelds begünstigt werden. Bei einem humanoiden Roboter beeinflusst die Bewegung auch indirekt Faktoren, wie beispielsweise den Energieverbrauch, und der aktuelle Umweltausschnitt die benötigte Zeit zum Fokussieren eines bestimmten Objekts.

Somit sind die Priorisierungskriterien für die interessengetriebene Explorationsstrategie, in Anlehnung an die interessengetriebene Perzeption (vgl. Abschnitt 5.3.2), die Salienz, die Neugier und die Bewegung (kurz: SNB). Der daraus resultierende Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}$ ist definiert als

$$\mathrm{EP}_{\mathrm{SNB}}(\gamma_s, \gamma_\eta, \gamma_d) = \underset{\mathrm{EP} \in \mathcal{P}_\mathcal{O}}{\arg\max} \left\{ \sum_{k=1}^{N} \gamma_s \cdot f_{\mathrm{Salienz}}\left(s_{o_{i_k}}\right) + \gamma_\eta \cdot f_{\mathrm{Neugier}}\left(\eta_{o_{i_k}}\right) + \gamma_d \cdot f_{\mathrm{Bewegung}}\left(d_{o_{i_k}}\right) \right\}. \tag{5.13}$$

mit

$$\gamma_s + \gamma_\eta + \gamma_d = 1, \quad \gamma_s, \gamma_\eta, \gamma_d, f(\cdot) \in [0,1]. \tag{5.14}$$

Um nun gezielt den Einfluss der Salienz $s_{o_{i_k}}$, der Neugier $\eta_{o_{i_k}}$ und der Bewegung $d_{o_{i_k}}$ auf den Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}$ berücksichtigen zu können, werden die drei Parameter γ_s, γ_η und γ_d eingeführt (analog zu Gl. 5.11). Die drei Funktionen zur Einbeziehung der Priorisierungskriterien Salienz, Neugier und Bewegung ergänzen das Optimierungsproblem und sind im Folgenden näher beschrieben.

Die Salienz $s_{o_{i_k}}$ wird im Folgenden berücksichtigt durch

$$f_{\mathrm{Salienz}}\left(s_{o_{i_k}}\right) := \tilde{s}_{o_{i_k}} = s_{o_{i_k}}^{\mathrm{MinMax}} \cdot \omega_{N,k}^{s} \tag{5.15}$$

mit

$$s_{o_{i_k}}^{\mathrm{MinMax}} := \frac{s_{o_{i_k}} - s_{\min}}{s_{\max} - s_{\min}} \in [0,1], \quad \omega_{N,k}^{s} := \frac{N-k+1}{N} \quad \text{und} \quad N = |\mathcal{O}|. \tag{5.16}$$

Hierbei wird der ursprüngliche Salienzwert mit Hilfe einer Min-Max-Normalisierung (vgl. Abschnitt 4.3.7) in das Intervall [0,1] transformiert. Die dafür notwendigen Minimal- und Maximalwerte ($s_{\min}$ bzw. $s_{\max}$) werden anhand von realen Testdaten ermittelt. Zusätzlich wird eine weitere Gewichtung $\omega_{N,k}^{s}$ eingeführt, welche eine Bestrafung darstellt, die umso höher ist, je später ein Objekt für die Exploration ausgewählt wird. Hohe Salienzwerte sind davon mehr betroffen als niedrigere Werte, sodass infolgedessen bei der Optimierung Objekte mit höherer Salienz zuerst ausgewählt werden müssen, um dieses Kriterium gut zu erfüllen.

Die Einbeziehung der Neugier $\eta_{o_{i_k}}$ erfolgt durch

$$f_{\mathrm{Neugier}}\left(\eta_{o_{i_k}}\right) := \tilde{\eta}_{o_{i_k}} = \eta_{o_{i_k}} \cdot \omega_{N,k}^{\eta} \tag{5.17}$$

mit

$$\eta_{o_{i_k}} \in [0,1], \quad \omega^{\eta}_{N,k} := \frac{N-k+1}{N} \quad \text{und} \quad N = |\mathcal{O}|. \tag{5.18}$$

Der Wert für die wissensbasierte Neugier für ein Objekt $\eta_{o_{i_k}}$ liegt bereits im Intervall [0,1] und daher muss keine weitere Anpassung vorgenommen werden (vgl. Abschnitt 4.3.7). Die zusätzliche Gewichtung $\omega^{\eta}_{N,k}$ erfolgt analog zu Gl. 5.15.

Die Gelenkwinkeldistanz $d_{o_{i_k}}$ beschreibt die benötigte Bewegung zur Fokussierung des Objekts o_{i_k}. Die Funktion zu deren Bestimmung lässt sich über die L1-Norm für den Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}$ definieren als

$$f_{\text{Bewegung}}\left(d_{o_{i_k}}\right) := \tilde{d}_{o_{i_k}} = 1 - d_{o_{i_k}} \cdot \underbrace{\frac{1}{Q\pi}}_{\text{Normierungsfaktor für die maximale Gesamtbewegung}} = 1 - \left\| \boldsymbol{q}_{o_{i_k}} - \boldsymbol{q}_{o_{i_{k-1}}} \right\| \cdot \frac{1}{Q\pi}, \tag{5.19}$$

wobei $\boldsymbol{q}_{o_{i_k}}$ die Gelenkwinkelstellung zur Fokussierung des Objekts o_{i_k} und $\boldsymbol{q}_{o_{i_{k-1}}}$ die Gelenkwinkelstellung des vorherigen Objekts ist. Formal sei $\boldsymbol{q}_{o_{i_0}}$ als die Ausgangsgelenkwinkelstellung zu Beginn der Exploration definiert. Zusätzlich wird mit der Anzahl an Gelenkwinkeln Q und der maximalen Bewegung pro Gelenkwinkel π normiert. Damit die Bedingung an eine einheitliche Bedeutung der Funktionswerte gegeben ist – d. h., hohe Werte sind optimal und niedrigere Werte suboptimal –, wird das Produkt aus Gelenkwinkeldistanz und Normierungsfaktor, welches im Intervall $[0,1]$ liegt, vom Wert 1 subtrahiert.

Der Wertebereich für die zuvor beschriebenen Funktionen der Salienz, Neugier und Bewegung (vgl. Gl. 5.15, Gl. 5.17 bzw. Gl. 5.19) ist aufgrund der jeweiligen Funktionsdefinition beschränkt auf das Intervall $[0,1]$.

Zusammengefasst ist der interessengetriebene Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}$ wie folgt definiert

$$\mathrm{EP}_{\mathrm{SNB}}(\gamma_s, \gamma_\eta, \gamma_d) = \underset{\mathrm{EP} \in \mathcal{P}_{\mathcal{O}}}{\arg\max} \left\{ \sum_{k=1}^{N} \gamma_s \cdot \tilde{s}_{o_{i_k}} + \gamma_\eta \cdot \tilde{\eta}_{o_{i_k}} + \gamma_d \cdot \tilde{d}_{o_{i_k}} \right\} \tag{5.20}$$

mit $N = |\mathcal{O}|$ und ermöglicht unterschiedlich starke Einflüsse von Salienz, Neugier und Bewegung je nach Wahl der Parameter γ_s, γ_η und γ_d. Die Bestimmung der Einflussparameter γ_s, γ_η und γ_d wird im Rahmen der Evaluation in Kapitel 7 beschrieben und kann dabei dynamisch an die aktuelle Anwendung bzw. Aufgabe angepasst werden.

Mit dieser Definition des interessengetriebenen Explorationspfads werden die Anforderungen an die interessengetriebene Exploration (vgl. Abschnitt 5.3.1) für eine priorisierte Erfassung der Umgebung erfüllt. Des Weiteren wurde durch Berücksichtigung der Optimierungskriterien Salienz und Neugier eine Anlehnung an das Verhalten des Menschen bei der Exploration einer Szene ermöglicht.

Strategie	Komplexität	Bemerkung
$EP_{Referenz}$	$O(n \log n)$	Schnelle Sortierung z. B. mit Heapsort oder Mergesort (vgl. [Ott12]); Falls die Objekte entsprechend der Wahrnehmung indiziert sind nur $O(1)$.
$EP_{Salienz}$	$O(n \log n)$	Sortierung z. B. mit Heapsort oder Mergesort
$EP_{Neugier}$	$O(n \log n)$	Sortierung z. B. mit Heapsort oder Mergesort
$EP_{Bewegung}$	$O(n!/2)$	Exakte Lösung des symmetrischen TSP
EP_{SNB}	$O(n!)$	Exakte Lösung des asymmetrischen TSP

Tabelle 5.1: Ein Vergleich der Komplexität bei der Bestimmung der verschiedenen Explorationspfade mit n als Anzahl an wahrgenommenen Objekten. Für die letzten beiden Strategien muss das Problem eines Handlungsreisenden (engl.: traveling salesman problem; kurz: TSP) gelöst werden, welches eine deutlich höhere Komplexität aufweist.

5.4.3 Abschätzung der Komplexität bei der Pfadbestimmung

Im Folgenden wird die Komplexität bei der Bestimmung der Explorationspfade für die zuvor vorgestellten Explorationsstrategien gegenübergestellt. Dabei werden Heuristiken zur schnelleren Berechnung von besonders aufwendigen Explorationspfaden zusammenfassend vorgestellt.

Allgemein

Die in den vorherigen Abschnitten vorgestellten Explorationsstrategien weisen ein stark unterschiedliches Verhalten in Bezug auf deren effiziente Berechenbarkeit auf. In Tabelle 5.1 ist ein allgemeiner Vergleich der Komplexität bei der Bestimmung der zuvor vorgestellten Explorationsstrategien zusammengefasst.

Betrachtet man zunächst die Referenz-, die salienzbasierten und die neugierbasierten Explorationsstrategien, so hängt der resultierende Explorationspfad jeweils nur von einer Größe ab. Durch eine effektive Sortierung (vgl. [Ott12]) der Objekte anhand eines einfachen Kriteriums (Zeitpunkt der Wahrnehmung, Salienz bzw. Neugier) lässt sich der Pfad sehr effektiv bestimmt. Die Komplexität beträgt $O(n \log n)$, wobei n die Anzahl der berücksichtigten Objekte darstellt.

Bei der bewegungsoptimierten und interessengetriebenen Explorationsstrategie ist die Bestimmung des Explorationspfads schwieriger, da nicht (nur) aufgrund von Objekteigenschaften der Pfad bestimmt wird, sondern unter Berücksichtigung der Beziehungen der Objekte untereinander. Dies lässt sich als Variante des Rundreiseproblems definieren, bei dem eine Liste mit Städten besucht werden muss und dabei die Gesamtstrecke minimiert wird. Im Nachfolgenden wird

zunächst das Rundreiseproblem kurz vorgestellt, bevor die Lösungen zur Bestimmung der Pfade für die beiden Explorationsstrategien Bewegung und Interesse einzeln diskutiert werden.

Definition des Rundreiseproblems

Das *Rundreiseproblem* oder *Problem eines Handlungsreisenden* (engl.: traveling salesman problem; kurz: TSP) beschreibt klassischerweise die Suche nach einer minimalen Tour für eine gegebene Liste mit Städten. Dabei wird die Entfernung verwendet, um die Kosten für die Reise von einer zur anderen Stadt zu beschreiben. Das TSP lässt sich bei einer ganzen Klasse von ähnlichen Optimierungsproblemen anwenden (beispielsweise beim „Layout" von integrierten Schaltkreisen oder bei der DNA-Sequenzierung). Die Problemstellung wird meist als ein Graph dargestellt, bei dem im klassischen TSP die Knoten die Städte darstellen und die Kanten die Entfernung zwischen den Städten. Je nach Anwendung lassen sie verschiedene Arten des Rundreiseproblems definieren:

- Ein *asymmetrisches TSP* (kurz: ATSP) kann durch einen gerichteten Graphen abgebildet werden. Dabei stellen die asymmetrischen Kanten zwischen zwei Knoten die Kosten dar, wobei der Graph nicht zwangsweise vollständig sein muss. Ein Beispiel hierfür ist die Einbahnstraße, welche unterschiedlich lange und somit nicht mehr symmetrische Wege zwischen zwei Orten erzeugt.
- Das *symmetrische TSP* (kurz: STSP) ist eine Vereinfachung des ATSP, da die Kanten des Graphen ungerichtet sind, d. h., die Kosten sind in beiden Richtungen gleich hoch. In Anlehnung an das Beispiel mit den Städten ist der Weg von Stadt A nach Stadt B gleich lang wie von Stadt B nach Stadt A.
- Ein *metrisches TSP* (kurz: MTSP) beschreibt ein TSP, das die Dreiecksungleichung für die in den Kanten hinterlegten Kosten zwischen allen Knoten erfüllt. Dies ist implizit erfüllt, falls bei der Bestimmung der Kosten eine Metrik wie beispielsweise der Euklidische Abstand oder die Manhattan-Metrik verwendet wird. Werden die Kosten jedoch nicht als (reine) Abstände interpretiert, sondern mit einer Kostenfunktion bestimmt, welche die Dreiecksungleichung verletzt (z. B. durch eine zusätzliche Gewichtung), so handelt es sich um ein *nicht-metrisches TSP* (kurz: NTSP).

In Abb. 5.1 sind Beispiele für die vorgestellten Varianten des Rundreiseproblems zu sehen. Dabei wird deutlich, dass auch ein asymmetrisches TSP ein metrisches TSP sein kann und auch ein symmetrisches TSP nicht immer metrisch sein muss.

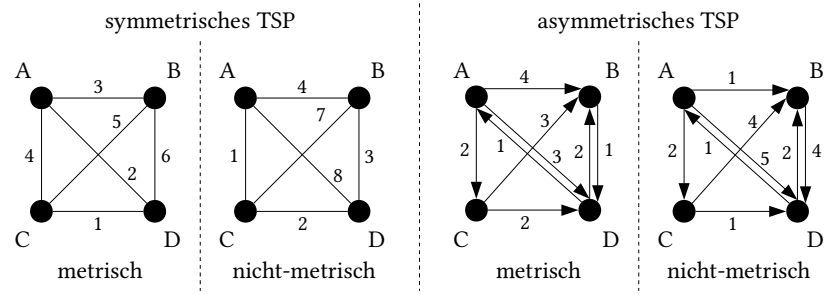

Abb. 5.1: Beispiele für verschiedene Varianten des Rundreiseproblems. Dabei wird deutlich, dass symmetrische TSP nicht immer metrisch sein müssen und asymmetrische TSP auch metrisch sein können.

Lösung des Rundreiseproblems

Die exakte Lösung des Rundreiseproblems ist NP-vollständig (vgl. [Joh97]) und daher nur für Probleme mit einer geringen Anzahl an Knoten (typischerweise bis zu 10-12 Knoten, je nach eingesetzter Hardware) in wenigen Sekunden berechenbar. Für größere Probleme muss entsprechend wesentlich mehr Zeit investiert werden, d. h. Minuten, Stunden oder Tage, da aufgrund der Komplexität des Problems das Laufzeitverhalten zur Bestimmung der Lösung einen exponentiellen Charakter besitzt. Alternativ kann eine Heuristik verwendet werden, die möglicherweise einen suboptimalen Pfad zurückliefert, der jedoch mit einem wesentlich geringeren zeitlichen Aufwand (z. B. polynomiell) bestimmt werden kann.

Die Beschreibungen sowie ein Vergleich der im Nachfolgenden kurz aufgeführten Heuristiken sind beispielsweise in [Gol80], [Joh97], [Glo01], [Hah07] und [App07] zu finden. Dabei können verschiedene Arten von Heuristiken unterschieden werden:

- Die *Eröffnungsheuristiken* liefern sehr schnell einen Pfad, der als Ausgangsbasis für weitere Heuristiken verwendet werden kann. Bekannte Vertreter der Eröffnungsheuristiken sind: Nearest-/Farthest-Neighbor-Heuristik, Nearest-/Farthest-Insertion-Heuristik, minimale Spannbaumheuristik (engl.: Minimum-Spanning-Tree-Heuristik) und Christofides-Heuristik, wobei die letzten beiden Heuristiken eine obere Grenze für die Länge des Pfads einhalten, falls ein metrisches STSP vorliegt. Die Pfadlänge der Minimum-Spanning-Tree-Heuristik ist dabei maximal doppelt so lang wie der optimale Pfad. Bei der Christofides-Heuristik kann sogar als Maximallänge das 1,5-fache des optimalen Pfades garantiert werden.
- Bei *Verbesserungsheuristiken* wird ausgehend von einem suboptimalen Pfad versucht, das Ergebnis zu verbessern, indem beispielsweise Teile des Pfades

neu angeordnet werden. Bekannte Vertreter sind u. a. die 2-opt-, 3-opt-, k-opt-Heuristik (Lin-Kernighan-Heuristik) und v-opt-Heuristik. Dabei wird bei den ersten drei Heuristiken jeweils eine feste Anzahl an Kanten vertauscht und somit versucht, einen kürzeren Pfad zu finden. Bei der v-opt-Heuristik wird hingegen versucht, durch Austausch einer variablen Anzahl an Kanten das Ergebnis zu verbessern.

- Die *metaheuristische Verfahren* beschreiben Ansätze, bei denen meist durch die Kombination von lokalen und globalen Optimierungskriterien versucht wird, lokale Minima zu vermeiden und somit ein besseres Gesamtergebnis zu erzielen. Beispiele hierfür sind *Tabu-Suche* oder *simulierte Abkühlung* (engl.: simulated annealing). Andere Heuristiken haben hingegen Prozesse in der Natur zum Vorbild und versuchen hierdurch gute Ergebnisse bei der Lösung des TSP zu erreichen. Bekannte Vertreter sind Ameisenalgorithmen (engl.: ant colony optimization), künstliche neuronale Netze (engl.: neural network algorithms) oder genetische Algorithmen (engl.: genetic algorithms).

Die zuvor vorgestellten Verfahren lassen sich teilweise nicht für die Lösungen von allen Varianten des TSP nutzen oder liefern teils schlechte Ergebnisse für einzelne TSP-Varianten. Deshalb ist es wichtig, im Einzelfall dasjenige Verfahren auszuwählen, welches für ein konkretes Problem erwartet die besten Ergebnisse liefert. In den nächsten beiden Abschnitten werden dazu die bewegungsoptimierte und die interessengetriebene Explorationsstrategie untersucht in Bezug auf die vorliegende Variante des Rundreiseproblems.

Bewegungsoptimierte Explorationsstrategie

Wird die bewegungsoptimierte Explorationsstrategie nach Gl. 5.10 betrachtet, so kann festgestellt werden, dass die Minimierung der kumulierten Gelenkwinkeldistanzen ein zum Rundreiseproblem äquivalentes Problem darstellt. Die Tatsache, dass am Ende die Startposition nicht erneut aufgesucht werden muss, ändert nichts an der NP-Vollständigkeit des Problems. Stellt man das gesamte Problem als Graph dar, so sind die Knoten die Objekte und die Kanten die benötigte Bewegung zwischen den Objekten. Letztere ist in beide Richtungen (zwischen zwei Objekten) gleich groß und wird wie zuvor schon beschrieben durch die L1-Norm dargestellt. Es handelt sich um ein MTSP, da die Kanten durch eine Metrik (L1-Norm) definiert sind. Des Weiteren sind die Kanten, wie zuvor beschrieben, ungerichtet, sodass es sich auch um ein symmetrisches TSP handelt. Zwei Besonderheiten müssen bei der Lösung des TSP berücksichtigt werden: Zum einen ist die Startposition fest, d. h. die aktuelle Sicht auf eine Szene. Dies kann als ein zusätzlicher Knoten im TSP dargestellt werden, der als Ausgangspunkt der Optimierung dient; zum anderen muss, wie zuvor erwähnt, die Startposition am Ende

nicht erneut aufgesucht werden. Letzteres bietet eine zusätzliche Möglichkeit für die Optimierung des Pfads.

Werden nun die möglichen Verfahren zur Lösung des Problems betrachtet, so lässt sich feststellen, dass bis zu einer gewissen Anzahl an Objekten der optimale Explorationspfad direkt bestimmt werden kann, indem alle möglichen Objektpermutationen berechnet werden und der kürzeste Pfad ausgewählt wird. Ab einer kritischen Anzahl an Objekten, die sich je nach vorhandenen Rechenressourcen vorab bestimmen lässt, ist es notwendig, Heuristiken einzusetzen, welche das Problem suboptimal lösen. In Kapitel 7 werden im Rahmen der Evaluation eine Auswahl an passenden Heuristiken für dieses TSP miteinander verglichen.

Interessengetriebene Explorationsstrategie

Die interessengetriebene Explorationsstrategie, wie sie in Gl. 5.20 definiert ist, berücksichtigt neben der in Gl. 5.19 verwendeten Bewegungsdistanz noch zwei weitere Priorisierungskriterien: die wissensbasierte Neugier und die multimodale Salienz (vgl. Gl. 5.15 und Gl. 5.17). Wird dieser Zusammenhang als ein ungerichteter Graph dargestellt (vgl. Abb. 5.2), so ist die Distanz die Kantenlänge, und die wissensbasierte Neugier und die Salienz werden den jeweiligen Knoten zugeordnet, welche die Objekte repräsentieren. Die aktuelle Ausrichtung und Position zu Beginn der Exploration werden durch den Startknoten realisiert. Werden nun die in den Knoten kodierten Priorisierungskriterien direkt auf die Kanten übertragen (analog zu Gl. 5.20), so wird aus dem ungerichteten Graphen ein gerichteter Graph. Das Beispiel in Abb. 5.2 veranschaulicht diese Vorgehensweise. Dabei wurden die Kantengewichte analog zu Gl. 5.20 berechnet für $\gamma_s = 0{,}25$, $\gamma_\eta = 0{,}5$ und $\gamma_d = 0{,}25$. Aufgrund der inversen Definition in Gl. 5.20 (Maximierungsproblem) muss für eine optimale Lösung nun der Pfad mit der maximalen Länge gefunden.

Die Bestimmung eines maximalen Pfads ist, wie bei der bewegungsoptimierten Explorationsstrategie, eine zum Rundreiseproblem analoge Fragestellung. Es gibt für die interessengetriebene Explorationsstrategie jedoch zwei wichtige Unterschiede: Zum einen liegt ein gerichteter Graph vor, welcher zu einem asymmetrischen TSP (ATSP) führt, und zum anderen werden die Kosten der Kanten nicht über eine Metrik definiert, sodass die Dreiecksungleichung verletzt wird und somit ein nicht-metrisches ATSP vorliegt. Die Vorgaben an den Pfad für eine feste Startposition und eine offene Endposition gelten hingegen für die interessengetriebene Explorationsstrategie ebenfalls.

In Abb. 5.2 sind sowohl der Startknoten als auch die Tatsache, dass der Graph nicht metrisch ist, dargestellt. Der Umweg von C nach A über B ist dabei im Sinne der Optimierung günstiger als der direkte Weg von C nach A. Somit muss

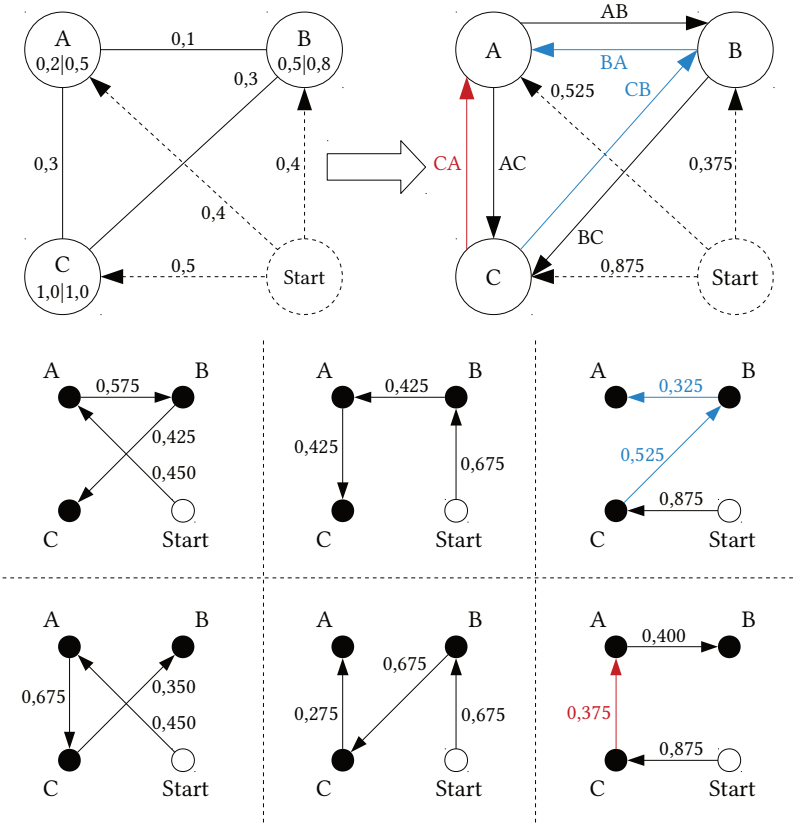

Abb. 5.2: Definition der Objektzusammenhänge mit Hilfe eines ungerichteten Graphen und von Objektgewichten (oben links); Übertragung der Objektgewichte (Neugier und Salienz) mit $\gamma_s = 0{,}25$, $\gamma_\eta = 0{,}5$ und $\gamma_d = 0{,}25$ auf die Pfade (analog zu Gl. 5.20) und somit die Definition eines gerichteten Graphen (oben rechts; unten). Die Ausgangsposition ist mittels Startknoten gekennzeichnet und besitzt gerichtete Kanten zu allen Objekten. Dies ist für die erste Fokussierung eines Objekts notwendig.

für eine größere Anzahl an Knoten eine Heuristik verwendet werden, welche für ein nicht-metrisches ATSP eine nahezu optimale Lösung liefert. Im Rahmen der vorliegenden Arbeit heißt das, dass die Summe der Kantengewichte auf dem zurückgelegten Weg nur unwesentlich geringer sein sollte als das Optimum bei exakter Berechnung. Alternativ besteht auch die Möglichkeit, jedes ATSP in ein symmetrisches TSP (STSP) zu überführen. Dabei werden Knoten mit mindestens einer asymmetrischen Kante verdoppelt (vgl. [Jon83]). Das resultierende STSP ist weiterhin nicht-metrisch und es erhöht sich der Aufwand zur Lösung des Problems aufgrund einer größeren Anzahl an Knoten. Dies ist nur dann zu rechtferti-

gen, falls die eingesetzte Heuristik für das STSP wesentlich effizienter ist als eine Heuristik für das ATSP. In Kapitel 7 wird eine Auswahl an möglichen Verfahren zur direkten Lösung dieses nicht-metrischen ATSP im Rahmen der Evaluation gegenübergestellt.

5.5 Schlussbetrachtungen

Die interessengetriebene Exploration bildet eine wichtige Grundlage für einen humanoiden Roboter, um die Umgebung mit den darin enthaltenen Gegenständen und Personen gezielt und tiefgehend wahrzunehmen. Dies ist eine wichtige Voraussetzung für die Bewältigung von Aufgaben, die Interaktion mit der Umwelt u. v. m. Der Explorationsansatz ist dabei generisch, sodass dieser für eine Vielzahl von anderen autonomen Systemen genutzt werden kann.

Für die interessengetriebene Exploration wurden folgende Anforderungen identifiziert und realisiert: Erweiterung der objektzentrierten Umwelterfassung zu einem ganzheitlichen Explorationsprozess; priorisierte Erfassung der Umgebung; Anlehnung an den Menschen sowie die Berücksichtigung des Einflusses von externen Faktoren während der Exploration.

Neben der Exploration, welche einen kompletten Prozess einschließlich der Koordination der Wahrnehmung einer Szene darstellt, muss zwischen der Explorationsstrategie zur Festlegung der Priorisierung einzelner Objekte und dem daraus resultierenden Explorationspfad mit einer konkreten Objektabfolge unterschieden werden. Neben der interessengetriebenen Explorationsstrategie mit mehreren Priorisierungskriterien (Salienz, Neugier und Bewegung) wurden auch weitere Explorationsstrategien wie beispielsweise die Referenz-, neugierbasierte, salienzbasierte oder bewegungsoptimierte Explorationsstrategie definiert, welche nur ein Priorisierungskriterium besitzen und im Rahmen einer späteren Evaluation zum Vergleich herangezogen werden. Je nach Wahl und Parametrierung der einzelnen Strategien ist eine unterschiedliche Priorisierung der Objekte während der Exploration möglich.

Des Weiteren besitzen die Explorationsstrategien auch eine unterschiedliche Komplexität bei der Bestimmung des Explorationspfads. Manche Pfade lassen sich durch eine schnelle Sortierung der Objekte anhand einfacher Kriterien bestimmen (beispielsweise der Zeitpunkt der Wahrnehmung oder die Salienz), andere setzen hingegen das Lösen einer komplexen Problemstellung (z. B. des Rundreiseproblems) voraus.

Auf die interessengetriebene Exploration wird in den folgenden Kapiteln im Rahmen der Umsetzung in einem ganzheitlichen System als auch bei der Evaluation in Kapitel 7 nochmals eingegangen.

Teil III

Umsetzung und Evaluation

6

Das OPASCA-System

In diesem Kapitel wird ein ganzheitliches System vorgestellt, welches die zuvor teils separat beschriebenen Elemente und Grundlagen der interessengetriebenen Exploration vereinigt. Dabei wird zunächst auf den strukturellen Aufbau des Systems näher eingegangen; anschließend werden die einzelnen Bestandteile (Module) vorgestellt. Anhand eines ausführlichen Beispiels wird die generelle Arbeitsweise des Systems näher erläutert und somit die grundlegende Informationsakquise für die interessengetriebene Exploration vorgestellt.

6.1 Einleitung

Das Akronym OPASCA steht für **op**to-**a**coustic **sc**ene **a**nalysis und ist der Name für ein System zur audiovisuellen Szenenexploration und -analyse. Der Fokus des Systems liegt insbesondere in der multimodalen Erfassung von Personen und Gegenständen in einer Szene. Das System wurde im Rahmen des Teilprojekts P2 (*interaktive multimodale Exploration*) des Sonderforschungsbereichs 588 (*Humanoide Roboter – Lernende und kooperierende multimodale Roboter*) entwickelt.

In den Arbeiten von Machmer und Swerdlow (vgl. [Mac10a], [Swe09]) wurden bereits erste theoretische Grundlagen für das OPASCA-System gelegt. Bei der ersten Arbeit wird insbesondere auf die Generierung und die Fusion von Umweltwissen eingegangen, wohingegen sich die zweite Arbeit mit den audiovisuellen Signaturen (vgl. Abschnitt 3.3) für die Wiedererkennung im Rahmen der Umwelterfassung befasst. Neben den theoretischen Grundlagen wurde auch ein Live-System für einen Hardwareaufbau – bestehend aus einer Stereokamera, einem Mikrofonarray und einer Schwenk-Neige-Einheit – realisiert.

Für die Entwicklung des OPASCA-Systems wurde die Programmierumgebung MATLAB der Firma *Mathworks* (vgl. [The12]) als Grundlage verwendet. Durch

die Bereitstellung der sogenannten Toolboxen für die digitale Signal-/Bildverarbeitung konnten Standardkomponenten wie beispielsweise FFT und 2D-DCT direkt in MATLAB verwendet werden. Des Weiteren ist eine flexible Verwendung von MATLAB-Code, C/C++-, C#[1]- und Java-Code, möglich, welcher beispielsweise auch die Anbindung von externen Schnittstellen zu Sensoren und Aktoren ermöglicht.

Aufbauend auf diesen Grundlagen wurde im Rahmen der vorliegenden Arbeit das OPASCA-System wesentlich weiterentwickelt und erweitert. Die wichtigsten Beiträge sind hierbei die multimodale Salienz bzw. Aufmerksamkeit (vgl. Kapitel 2), die wissensbasierte Neugier (vgl. Kapitel 4) und die interessengetriebene Explorationsstrategie (vgl. Kapitel 5). Letzteres ermöglicht den Übergang von einer reinen Umwelterfassung des aktuellen Sensorbereichs (vgl. [Swe09], [Mac10a]) hin zu einer aktiven perzeptuellen Exploration einer kompletten Szene (vgl. Kapitel 5). Darüber hinaus wurden die vorhandenen Möglichkeiten signifikant erweitert, a-priori-Wissen vorab zu hinterlegen und zusätzliche Objektinformationen im Rahmen eines Dialogs zu akquirieren.

6.2 Überblick und Aufbau

Der grundsätzliche Aufbau des OPASCA-Systems ist in Abb. 6.1 zu sehen und zeigt den Zusammenhang der einzelnen Systemkomponenten untereinander. Der Systemaufbau erfolgt auf Basis der in Abschnitt 3.5 beschriebenen Systemarchitektur.

Das OPASCA-System ist durch seine modulare Struktur leicht an neue Sensorplattformen anpassbar, d. h. es existieren Systemmodule für die Ansteuerung der Sensoren und Aktoren, welche einheitliche Schnittstellen für die Datenakquisition und die Ansteuerung der Aktoren zur Verfügung stellen. Somit ist das eigentliche System mit seinen Verarbeitungsmodulen (vgl. Abschnitt 3.5.2) nicht direkt an bestimmte Sensoren eines Herstellers gebunden.

Das Modul zur Objektdetektion erstellt neue Objektrepräsentanten im Umweltmodell, falls die aus den Sensordaten extrahierten Informationen keinem bereits existierenden Repräsentant zugeordnet werden können. Die Verarbeitungsmodule generieren Informationen über die im Umweltmodell vorhandenen Repräsentanten und ermöglichen somit eine immer detailliertere Repräsentation der Objekte (vgl. Abschnitt 3.2).

Das Umweltmodell selbst enthält Repräsentanten für die aktuell erfassten Objekte in der Umgebung, das vorhandene a-priori-Wissen sowie die objektbe-

[1] C#-Unterstützung nur unter Windows vorhanden.

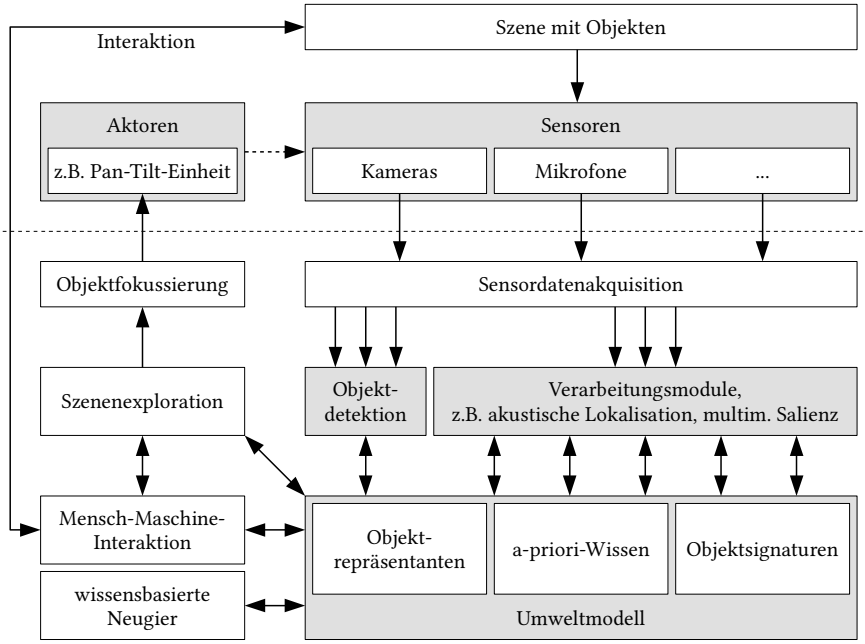

Abb. 6.1: Schematischer Aufbau des OPASCA-Systems mit den Schnittstellen zu den Sensoren und Aktoren, den Akquise- und Verarbeitungsmodulen, dem Umweltmodell sowie weiteren Bestandteilen zur Bestimmung der Neugier, zur Interaktion und zur Szenenexploration.

schreibenden Signaturen (vgl. Abschnitt 3.1). Letztere werden von den Verarbeitungsmodulen genutzt, um zuvor erfasste Objekte wiederzuerkennen.

Mit den im Umweltmodell vorhandenen Informationen kann nun die wissensbasierte Neugier für die einzelnen Objekte bestimmt werden (vgl. Abschnitt 4.3). Des Weiteren bilden diese Informationen auch die Grundlage für die Mensch-Maschine-Interaktion (vgl. Abschnitt 3.5.3) und die Exploration einer Szene (vgl. Kapitel 5).

Sowohl die Interaktion als auch die Exploration nehmen Einfluss auf den aktuellen Ausschnitt der Szene und die Akquise von Informationen. Durch die Mensch-Maschine-Interaktion können beispielsweise direkt Informationen über bestimmte Objekte, die im Umweltmodell repräsentiert sind, eingeholt werden, wohingegen bei der Exploration gezielt der Fokus der Sensoren verschoben wird (durch Aktoren wie beispielsweise eine Schwenk-Neige-Einheit – engl.: Pan-Tilt-Einheit) und somit durch die Verarbeitungsmodule neue Informationen aus den Sensordaten extrahiert werden können. Im Folgenden werden dazu die im Rah-

men der vorliegenden Arbeit verwendeten Verarbeitungsmodule kategorisiert und näher erläutert.

6.3 Module

In Abschnitt 3.5.2 wurde das Modulkonzept bei der Umwelterfassung bereits eingeführt. Dort wurden allgemein Verarbeitungsmodule beschrieben, welche aus den Sensordaten und/oder den im Umweltmodell erfassten Informationen neues Wissen über Gegenstände und Personen im aktuellen Erfassungsbereich der Sensoren generieren. Im Folgenden werden die Module in verschiedene Kategorien eingeteilt, um deren Bedeutung für das System herauszustellen. Im Anschluss daran werden die wichtigsten Module und die dazugehörigen Algorithmen für die Wahrnehmung von Gegenständen und Personen in einer Szene vorgestellt.

6.3.1 Kategorien

Die Module des Systems werden aufgrund ihrer Funktionalität in vier verschiedene Kategorien unterteilt (vgl. [Mac10a]):

- *Initiierungsmodule* sind für die Akquise von grundlegenden Informationen aus den aktuellen Sensordaten verantwortlich. Diese können zur späteren Erzeugung von neuen Objektrepräsentanten im Umweltmodell genutzt werden. Ebenfalls werden diese Informationen den bereits vorhandenen Repräsentanten zugeordnet (u. a. über die aktuelle Position eines Objekts).
- *Spezialisierungsmodule* ergänzen und aktualisieren das Wissen über die im Umweltmodell vorhandenen Objektrepräsentanten. Dabei werden aus den aktuellen Sensorinformationen und den im Umweltmodell bereits vorhandenen Daten neue Attribute und Relationen erzeugt bzw. vorhandene konkretisiert. Infolgedessen nimmt i. d. R. der Abstraktionsgrad eines Repräsentanten mit der Zeit ab, d. h., es findet eine „Spezialisierung“ des Repräsentanten statt.
- *Fusionsmodule* werden benötigt, wenn ein Attribut und/oder eine Relation durch Daten aus unterschiedlichen Quellen bestimmt wird. Ein Beispiel hierfür ist die Identität einer Person, die mit akustischen und visuellen Sensoren wahrgenommen wurde. Durch die Fusion wird das Wissen durch ein einziges Attribut bzw. durch eine einzelne Relation beschrieben. Im vorherigen Beispiel ist dies das Attribut Identität. Die Fusion kann nicht nur mit Daten aus unterschiedlichen Informationsquellen, sondern auch von verschiedenen Zeitpunkten erfolgen, d. h. bei der Fusion wird die Historie miteinbezogen.

- *Dienstleistungsmodule* fassen alle Module zusammen, welche für generelle Funktionen des Systems verantwortlich sind. Beispiele hierfür sind Module zur Ansteuerung von Sensoren oder Aktoren. Aber auch Funktionen wie beispielsweise das Nachverfolgen der momentanen Objektposition oder die Zuordnung der aktuellen Sensordaten zu den im Umweltmodell vorhandenen Repräsentanten werden in Modulen dieses Typs abgebildet.

Neben den zuvor vorgestellten Modulkategorien wird im Rahmen der vorliegenden Arbeit eine weitere Kategorie eingeführt:

- *Analysemodule* nutzen ausschließlich Informationen aus dem Umweltmodell, um neue Objektaspekte bzw. neues Wissen zu generieren. Dies kann als eine Fusion von Objektinformationen auf einer höheren Ebene aufgefasst werden. Beispiele hierfür sind Situationserkennung, Intentionserkennung oder auch die in der vorliegenden Arbeit vorgestellte wissensbasierte Neugier.

6.3.2 Die wichtigsten Module zur Umwelterfassung

Für die Umwelterfassung werden Algorithmen und somit auch Systemmodule mit einer Vielzahl an unterschiedlichen Fähigkeiten benötigt. Beginnend mit der Sensordatenakquisition und der Regelung der Aktoren werden Algorithmen zur Lokalisation, Klassifikation bzw. Identifikation, Nachverfolgung und Fusion benötigt. Die wichtigsten Module zur Erfassung von Personen und Gegenständen in einer Szene werden in den folgenden Abschnitten zusammengefasst.

Sensordatenakquisition und Regelung der Aktoren

Die Ausgangsbasis für die Erfassung der Umwelt liefern die akquirierten Daten von verschiedenen Sensoren. Im OPASCA-System stehen sensorspezifische Dienstleistungsmodule zur Verfügung, welche die verwendete Sensorhardware abstrahieren. Somit ist das System nicht an die Hardware eines bestimmten Herstellers gebunden und es kann von den nachfolgenden Modulen der Umwelterfassung direkt auf die akquirierten Daten, d. h. Kamerabilder, Audiosignale usw. zugegriffen werden.

Neben den Sensoren sind Aktoren, welche u. a. eine Ausrichtung der Sensoren und somit eine Änderung des Erfassungsbereichs ermöglichen, eine sehr wichtige Komponente für die Szenenexploration. Für das OPASCA-System werden im Rahmen der vorliegenden Arbeit zwei verschiedene Sensoraufbauten verwendet (vgl. Abschnitt 7.1), jedoch ist das Systemkonzept nicht nur auf diese beiden beschränkt. Der eine Sensoraufbau ist eine Kombination aus zwei Stereokameras und einem Mikrofonarray mit einer Schwenk-Neige-Einheit (vgl. [Küh12a]),

der andere ist der Kopfaufbau des Roboters ARMAR-III (vgl. [Asf08]), welcher vergleichbare Sensoren besitzt. Beide unterscheiden sich in der Anzahl der Freiheitsgrade und in der Möglichkeit der Regelung. Der Sensoraufbau mit Schwenk-Neige-Einheit lässt nur einfache Bewegungen in zwei Ebenen zu, wohingegen der ARMAR-III Kopf mit seinen sieben Freiheitsgraden sehr flexibel geregelt werden kann (vgl. [Mil11]). Für jeden Sensoraufbau existiert im System ein separates Dienstleistungsmodul für die Ansteuerung.

Akustische Lokalisation von Schallquellen

Für die Bestimmung der Position einer Schallquelle (d. h. Sprache oder Geräusch eines Gegenstands) wird das in Abschnitt 2.3.1 beschriebene korrelationsbasierte Lokalisationsverfahren mit mehreren Mikrofonpaaren verwendet und in Form eines Initiierungsmoduls realisiert:

Durch die Kombination einzelner Mikrofonpaare eines Mikrofonarrays können die Positionen von Schallquellen in Form einer akustischen Landkarte bestimmt werden. Als Algorithmus zur Bestimmung der Position kommt das im Englischen als *Steered Response Power with Phase Transform* (SRP-PHAT) bezeichnete Verfahren zum Einsatz (vgl. [Kna76]). Wird dieser Ansatz um einen zusätzlichen Parameter zur Gewichtung des Prewhitening-Terms erweitert, so entsteht das sogenannte SRP-PHAT-β-Verfahren (vgl. [Don07]; Abschnitt 2.3.1), welches zu einer durchschnittlich höheren Lokalisationsgenauigkeit führt (vgl. [Pap11]).

Akustische Klassifikation von Schallquellen

Die Klassifikation von Schallquellen wird im OPASCA-System durch mehrere Spezialisierungsmodule umgesetzt. Hierbei handelt es sich um Module zur (Vor-)Klassifikation von Schallquellen, zur Klassifikation von Geräuschen und zur akustischen Identifikation von Personen. Des Weiteren wird in einem weiteren Spezialisierungsmodul der aktuelle Betriebszustand eines Geräts anhand der emittierten Geräusche bestimmt:

Die akustische (Vor-)Klassifikation von Schallquellen, d. h. die Entscheidung, ob es sich bei dem Objekt um einen Gegenstand oder eine Person handelt, erfolgt auf Grundlage von spektralen Merkmalen. Sogenannte *Mel-Frequenz-Cepstrum-Koeffizienten* (engl.: mel-frequency cepstral coefficients; kurz: MFCC) sind aus der Spracherkennung bekannt und können auch bei der Klassifikation eingesetzt werden, um die relevanten Merkmale in einem Audiosignal bzgl. der Sprache und der Geräusche zu extrahieren (vgl. [O'S00]). Für die Klassifikation von Geräuschen werden *Gaußsche Mischverteilungsmodelle* (engl.: Gaussian mixture

models; kurz: GMM) eingesetzt, welche für jede Geräuschklasse ein unabhängiges Modell bilden. Bei der Identifikation von Personen wird ein sogenanntes *universelles Hintergrundmodell* (engl.: universal background model; kurz: UBM) als Basis verwendet (vgl. [Rey00]), welches ein deutlich komplexeres Modell für menschliche Sprache darstellt. Die einzelnen Modelle für jede Person werden aus dem UBM-Modell mit Hilfe von Trainingsdaten abgeleitet. Beide Verfahren haben sich inner- und außerhalb des OPASCA-Systems in Evaluationen bewährt (vgl. [Swe08a], [Swe09], [Mac10a], [Rey00]).

Insbesondere bei Geräten kann neben dem Typ bei einigen Objekten, wie beispielsweise einem Kaffeeautomaten, zusätzlich auch noch der Betriebszustand akustisch bestimmt werden. Dazu werden mit Hilfe von *Hidden-Markov-Modellen* (kurz: HMM) Zustandsfolgen für eine spätere Klassifikation modelliert. Dies wurde im OPASCA-System beispielhaft für einen Kaffeevollautomaten mit seinen verschiedenen Betriebszuständen während der Zubereitung realisiert (vgl. [Wol12]).

Bei der akustischen Klassifikation ist es wichtig, zu erkennen, ob das aktuelle Geräusch zu einer bekannten Klasse gehört oder nicht. Dies wurde exemplarisch am Beispiel von Küchengeräten und mit Hilfe von *Ein-Klassen-Supportvektormaschinen* (engl.: one-class support vector machine; kurz: OCSVM) untersucht (vgl. [Lóp11]).

Visuelle Personendetektion

Zur Detektion von Personen in einer Szene kann eine Vielzahl an Algorithmen eingesetzt werden. Dabei kann sowohl der ganze Körper als auch nur das Gesicht als repräsentatives Merkmal verwendet werden. Im Rahmen der vorliegenden Arbeit wurden mehrere Verfahren zur Gesichtsdetektion betrachtet: zum einen das weit verbreitete Verfahren von Viola & Jones (vgl. [Vio04]), zum anderen ein Verfahren, welches auf der *modifizierten Census-Transformation* (engl: modified census transform; kurz: MCT) basiert (vgl. [Frö04]). Aufgrund der sehr hohen Detektionsrate und sehr geringen Rate an Fehldetektionen (vgl. [Frö04]) sowie einem Konfidenzmaß für die Sicherheit einer Gesichtsdetektion, wird im Rahmen der vorliegenden Arbeit die MCT-basierte Gesichtsdetektion verwendet und in Form eines Initiierungsmoduls im OPASCA-System realisiert.

Visuelle Personenidentifikation

Für die visuelle Identifikation von Personen wird das Gesicht als primäres Merkmal verwendet. Dabei wird ein Ansatz verfolgt, welcher auf der Verwendung der *diskreten Kosinustransformation* (engl.: discrete cosine transform; kurz: DCT) für

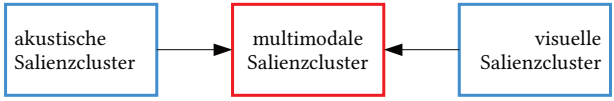

Abb. 6.2: Die multimodalen Salienzcluster und deren Teilkomponenten wurden als Module realisiert; Initiierungsmodule (blau); Fusionsmodul (rot).

lokale Gesichtsbereiche beruht (vgl. [Eke05]). Die Gesichtsidentifikation wird im System durch ein Spezialisierungsmodul umgesetzt. Das Verfahren wurde in realen Anwendungen inner- und außerhalb des OPASCA-Systems erfolgreich evaluiert (vgl. [Eke09], [Wan09], [Tsa09], [Mac10a], [Lie11]).

Visuelle Gegenstandsdetektion und -klassifikation

Für die Detektion und Klassifikation von Gegenständen wird im OPASCA-System eine Kombination von verschiedenen Ansätzen verfolgt (vgl. [Wei08a], [Wei08b]). Dazu zählen u. a. texturbasierte Verfahren, aber auch normalisierte Farbhistogramme (vgl. [Jua09], [Swe09]). Für die Segmentierung von Objekten auf a-priori bekannten Oberflächen kommt ein Regionwachstumsverfahren zum Einsatz (vgl. [Mac10a]). Je nach verwendeter Hardware (Stereokamera, RGB-D-Sensor) und vorhandenem a-priori-Wissen können auch 3D-Modelle verwendet werden (vgl. [Kas12]). Im Rahmen der vorliegenden Arbeit wurde auf die Verwendung von 3D-Modellen, aufgrund der insgesamt hohen Rechenanforderungen der bisher vorgestellten Verfahren, verzichtet. Eine Ergänzung von 3D-Modellen ist jedoch als Erweiterung in Form von weiteren Initiierungs- und Spezialisierungsmodulen im System jederzeit möglich.

Multimodale Salienzcluster

Die Verfahren zur Bestimmung der *multimodalen Salienzcluster* sind in Kapitel 2 beschrieben. Mit Hilfe der akustischen Salienz, welche mittels Bayesian Surprise bestimmt wird (vgl. [Sch11a], [Küh12b]), und der visuellen Salienz, welche auf Quaternion-basierten Salienzkarten beruht (vgl. [Sch12a]), wird durch die Fusion eine gemeinsame multimodale Salienz bestimmt. Durch die zusätzliche örtliche Bestimmung der Positionen der Salienzen können die in Abschnitt 2.3.3 beschriebenen multimodalen Salienzcluster erstellt werden. Die Umsetzung im OPASCA-System erfolgt durch zwei Initiierungsmodule zur Bestimmung der unimodalen Salienzcluster und einem Fusionsmodul zur Erzeugung der multimodalen Salienzcluster (vgl. Abb. 6.2).

Tracking von Personen und Gegenständen

Die *kontinuierliche Nachverfolgung* (engl.: Tracking) von bewegten Objekten in einer Szene ist ein essentieller Bestandteil für das System und wird in Form eines Dienstleistungsmoduls realisiert.

Es ist notwendig, für die Nachverfolgung den bereits im Umweltmodell vorhandenen Repräsentanten aus dem letzten Zeitschritt die aktuelle Detektion, d. h. die Positionen der realen Objekte in der Szene, zuzuordnen. Dabei müssen für neue Gegenstände und Personen entsprechende Objektrepräsentanten im Umweltmodell angelegt werden. Mit den zugeordneten neuen Positionsdaten und den Werten aus der Vergangenheit kann eine Position für den nächsten Zeitschritt prädiziert werden. Dieses Vorgehen kann auf verschiedene Weisen realisiert werden: Die verbreitetsten Verfahren sind der Nächste-Nachbar-Filter (vgl. [Ron96]), der Kalmanfilter (vgl. [Kal60]) und der Partikelfilter (vgl. [Aru02]). Es existiert eine Vielzahl an Varianten, die je nach Anforderung ihre Stärken und Schwächen aufweisen. Für das OPASCA-System wird ein sogenannter *Interacting Markov Chain Monte Carlo particle filter* (kurz: IMPF) verwendet, welcher erfolgreich evaluiert wurde (vgl. [Pal11]).

Fusion von multimodalen Informationen

Die Fusion von Informationen im OPASCA-System geschieht zu zwei unterschiedlichen Zeitpunkten. Zum einen werden früh im Explorationsprozess bei der *räumlich-zeitlichen Fusion* (engl.: spatio-temporal fusion) – mittels der aktuellen Position eines Objekts sowie einer gewissen Historie – die aktuelle Aussagen über verschiedene Aspekte eines Objekts verbessert. Ein Beispiel hierfür sind die visuellen Informationen zur Bestimmung der Identität einer Person. Zum anderen findet eine *Fusion von multimodalen Informationen* eines Attributs oder einer Relation zu einem späteren Zeitpunkt (engl.: late fusion) statt. Dabei werden Daten aus unterschiedlichen Informationsquellen (meist von mehreren Sensortypen) zu einer gemeinsamen Aussage fusioniert. Die visuellen Aussagen über die Identität einer Person werden hierbei beispielsweise mit den akustischen Ergebnissen zu einer Gesamtidentität vereinigt.

Mit dem Stellenwert der Fusion im Rahmen der Umwelterfassung sowie den unterschiedlichen Fusionsansätzen für ein System zur audiovisuellen Szenenanalyse beschäftigt sich die Arbeit von Machmer (vgl. [Mac10a]) ausführlich. Dort wird auch die Fusion im OPASCA-System im Rahmen der Umsetzung und Evaluation der theoretischen Grundlagen näher beschrieben.

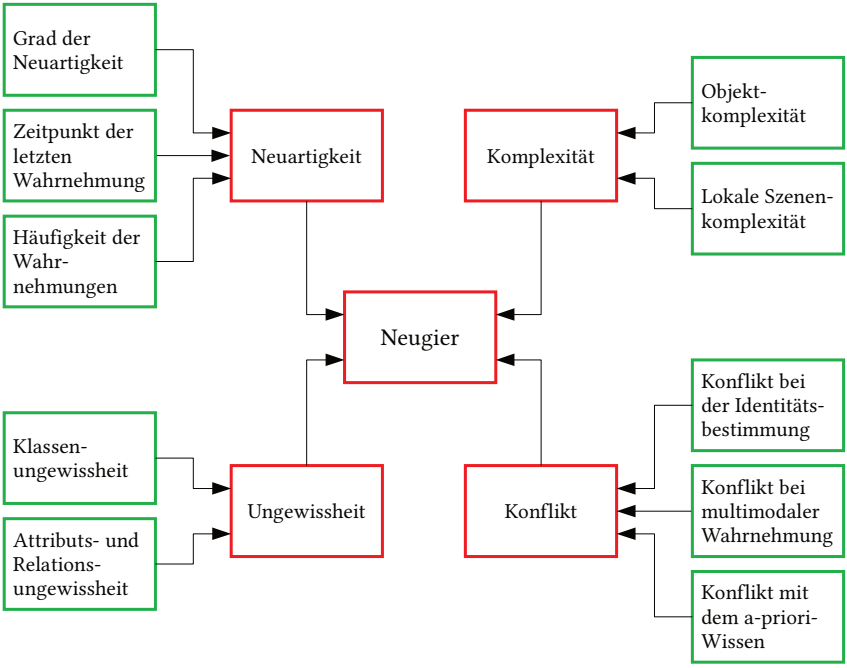

Abb. 6.3: Die Teilkomponenten der wissensbasierten Neugier (vgl. Abb. 4.1) werden mit Hilfe von Analysemodulen (grün) und Fusionsmodulen (rot) realisiert.

6.3.3 Module zur Bestimmung der wissensbasierten Neugier

Die theoretischen Grundlagen zur Bestimmung der *wissensbasierten Neugier* wurden in Kapitel 4 beschrieben und basieren u. a. auf Erkenntnissen aus der Psychologie. Die verwendeten situativen Bedingungen zur Bestimmung der Neugier sind hierbei die Neuartigkeit, die Komplexität, die Unsicherheit und der Konflikt. Durch die Kombination aller Teilaspekte wird ein gemeinsames Maß bestimmt, welches die wissensbasierte Neugier für ein Objekt beschreibt. Die Neugier basiert dabei auf den aktuellen Objektinformationen im Umweltmodell, d. h. Attributen, Relationen sowie Zeitpunkten der Wahrnehmung.

Die wissensbasierte Neugier wird im Rahmen der vorliegenden Arbeit mit einer Vielzahl an Analyse- und Fusionsmodulen realisiert, welche in Abb. 6.3 dargestellt sind. Durch eine hierarchische Vorgehensweise bei der Fusion werden zunächst die einzelnen Teilaspekte (vgl. Abschnitt 4.3.7) bestimmt und anschließend zur wissensbasierten Neugier fusioniert. Dies geschieht für jeden Objektrepräsentant im Umweltmodell separat.

Module zur Bestimmung der Neuartigkeit

Die *Neuartigkeit* umfasst die drei Bestandteile: Grad der Neuartigkeit, Zeitpunkt der letzten Wahrnehmung und Häufigkeit der Wahrnehmungen (vgl. Abschnitt 4.3.3; Abb. 6.3). Diese sind jeweils als ein Analysemodul realisiert. In Kombination mit einem Fusionsmodul wird das Gesamtergebnis des Teilaspekts Neuartigkeit bestimmt (vgl. Abschnitt 4.3.7).

Module zur Bestimmung der Komplexität

Zur Bestimmung der *Komplexität* werden zwei Aspekte herangezogen (vgl. Abschnitt 4.3.4; Abb. 6.3): zum einen die Objektkomplexität und zum anderen die lokale Szenenkomplexität. Für die Objektkomplexität wird nur eine Teilmenge an Attributen berücksichtigt, welche im Folgenden näher betrachtet werden:

- Die *Texturiertheit eines Objekts* lässt sich mit Hilfe eines Kantendetektors – angewendet auf den grauwertbasierten Bildausschnitt eines Objekts $\boldsymbol{I}_o^{\text{Grau}}$ – bestimmen. Im Rahmen der vorliegenden Arbeit wurden verschiedene Kantendetektoren untersucht und der Canny-Kantendetektor ausgewählt, da dieser u. a. ein rauscharmes Kantenbild generiert (vgl. [Gon02]). Das binärwertige Kantenbild wird unter Berücksichtigung der beiden Parameter Standardabweichung σ, welche für den Gauß'schen Weichzeichnungsfilter im Vorverarbeitungsschritt verwendet wird, und Schwellenwerte $\boldsymbol{t}$, welche zur Festlegung von schwachen und starken Kanten genutzt werden, bestimmt:

$$\boldsymbol{I}_o^{\text{Kanten}} := \text{Canny}\left(\boldsymbol{I}_o^{\text{Grau}}, \sigma, \boldsymbol{t}\right) \quad \text{mit } \sigma = 1 \text{ und } \boldsymbol{t} = (0{,}019; 0{,}047)\,. \tag{6.1}$$

  Die Werte für die beiden Parameter wurden aus [Gon02] übernommen. Abschließend wird der Kantenanteil eines Objekts als Maß für die Texturiertheit eines Objekts definiert:

$$w_{a_{\text{Textur}}} := \min\left\{1; f_{\text{Anteil}}\left(\boldsymbol{I}_o^{\text{Kanten}}, \lambda^{\text{Kanten}}\right)\right\} \quad \text{mit } \lambda^{\text{Kanten}} \in [0, 1] \tag{6.2}$$

  und

$$f_{\text{Anteil}}\left(\boldsymbol{I}_o^{\text{Kanten}}, \lambda^{\text{Kanten}}\right) := \frac{1}{\lambda^{\text{Kanten}} \cdot \left|\boldsymbol{I}_o^{\text{Kanten}}\right|} \sum_{x,y} \boldsymbol{I}_o^{\text{Kanten}}(x, y)\,. \tag{6.3}$$

  Dabei ist λ^{Kanten} ein Kantenfaktor, der den maximal benötigten Kantenanteil eines Objekts für eine vollständige Texturiertheit festlegt. Der Wertebereich von $w_{a_{\text{Textur}}}$ wurde aufgrund des Kantenfaktors λ^{Kanten} in Gl. 6.2 nach oben auf 1 beschränkt.

- Die *Kontur eines Objekts* dient als weitere Größe für die Bestimmung der Komplexität. Alle Objekte werden dafür zunächst auf eine Standardgröße skaliert, um eine Invarianz vom Abstand zur Kamera zu gewährleisten. Anschließend wird der Winkel des Gradienten φ in jedem Punkt i der Objektkontur bestimmt und die kumulierte Änderung der Ausrichtung als Maß für die Komplexität interpretiert. Abschließend erfolgt eine Bewertung mit der Länge der Objektkontur L selbst, um so ein normalisiertes Maß $w_{a_{\text{Kontur}}}$ zu bestimmen:

$$w_{a_{\text{Kontur}}} := \frac{1}{L} \sum_{i=1}^{L} |\varphi_i - \varphi_{i-1}| \quad \text{mit } \varphi_0 = \varphi_L \text{ und } w_{a_{\text{Kontur}}} \in [0, 2\pi) \,. \tag{6.4}$$

- Bei der *Farbgebung eines Objekts* werden Objekte, welche farblich hervorstechen, höher bewertet als andere. Hierzu wird der Bildausschnitt $\mathcal{X}_o^{\text{RGB}}$ eines Objekts im HSV-Farbraum (vgl. [Gon02]) betrachtet: $\mathcal{X}_o^{\text{HSV}}$. Dieser Farbraum besitzt die drei Komponenten Farbton (H), Sättigung (S) und Hellwert (V). Dabei ist der Farbton auf einem Farbkreis beschrieben als Winkel von Null bis 360 Grad, während die Sättigung und der Hellwert in einem Intervall von Null bis Eins bestimmt werden. Dominante Farben lassen sich in diesem Farbraum bei einer hohen Sättigung und einem hohen Hellwert abbilden. Der Farbton ist dabei beliebig. Ein Bildpunkt ist im HSV- und RGB-Farbraum beschrieben über seinen Ort und den dazugehörigen Farbwerten, als Vektor $\boldsymbol{x}^{\text{HSV}} = (x, y, H, S, V)$ bzw. $\boldsymbol{x}^{\text{RGB}} = (x, y, R, G, B)$. Die relative Farbgebung eines Objekts lässt sich wie folgt aus der Bildpunktmenge des Objekts $\mathcal{X}_o^{\text{RGB}}$ nach Transformation in den HSV-Farbraum (d. h. $\mathcal{X}_o^{\text{HSV}} := f_{\text{RGB}\to\text{HSV}}(\mathcal{X}_o^{\text{RGB}})$; vgl. [Gon02]) bestimmen:

$$w_{a_{\text{Farbgebung}}} := \frac{1}{|\mathcal{X}_o^{\text{HSV}}|} \sum_{\boldsymbol{x}^{\text{HSV}} \in \mathcal{X}_o^{\text{HSV}}} S\left(\boldsymbol{x}^{\text{HSV}}\right) \cdot V\left(\boldsymbol{x}^{\text{HSV}}\right) \tag{6.5}$$

  mit

$$S\left(\boldsymbol{x}^{\text{HSV}}\right), V\left(\boldsymbol{x}^{\text{HSV}}\right) \in [0, 1] \,. \tag{6.6}$$

  Hierbei liefern die Funktionen $S\left(\boldsymbol{x}^{\text{HSV}}\right)$ und $V\left(\boldsymbol{x}^{\text{HSV}}\right)$ die normierten Werte $[0, 1]$ für die Sättigung und die Helligkeit für den transformierten Bildpunkt $\boldsymbol{x}^{\text{RGB}}$ im HSV-Farbraum, d. h. $\boldsymbol{x}^{\text{HSV}}$. In Gl. 6.5 wird somit über das Produkt aus Sättigungs- und Helligkeitswert aller Bildpunkte des Objekts im HSV-Farbraum summiert und anschließend mit der Anzahl an Bildpunkten normiert, um eine von der Größe des Objekts unabhängige Aussage zu erhalten.

- Die *allgemeine Präsenz* eines Objekts ist in der vorliegenden Arbeit durch dessen Größe im Bild beschrieben. Ist der Anteil der Bildpunkte eines Objekts $|\mathcal{X}_o^{\text{RGB}}|$ im aktuellen Kamerabild $\boldsymbol{I}_{\text{RGB}}$ hoch, so ist auch dessen Präsenz hoch. Dieser Zusammenhang ist wie folgt definiert:

$$w_{a_{\text{Präsenz}}} := \begin{cases} 1, & \text{falls } \frac{|\mathcal{X}_o^{\text{RGB}}|}{\lambda^{\text{Präsenz}} \cdot |\boldsymbol{I}_{\text{RGB}}|} \geq 1 \\ \frac{|\mathcal{X}_o^{\text{RGB}}|}{\lambda^{\text{Präsenz}} \cdot |\boldsymbol{I}_{\text{RGB}}|}, & \text{sonst} \end{cases} \quad \text{mit } \lambda^{\text{Präsenz}} \in (0, 1]\,. \tag{6.7}$$

Dabei ist $\lambda^{\text{Präsenz}}$ ein Präsenzfaktor, der den maximal benötigten Anteil eines Objekts im Bild für eine vollständige Präsenz festlegt.

Die Objektkomplexität wird aus den Werten der zuvor beschriebenen Attribute (Texturiertheit eines Objekts, Kontur eines Objekts, Farbgebung eines Objekts und allgemeine Präsenz) ermittelt (vgl. Abschnitt 4.3.4) und in Form eines Analysemoduls im OPASCA-System realisiert. Mit Hilfe der Ergebnisse des zweiten Analysemoduls zur Bestimmung der lokalen Szenenkomplexität (vgl. Abschnitt 4.3.4) lässt sich ein Gesamtergebnis für die Komplexität durch die Verwendung eines entsprechenden Fusionsmoduls bestimmen (vgl. Abschnitt 4.3.7).

Module zur Bestimmung der Unsicherheit

Der Teilaspekt *Unsicherheit* der wissensbasierten Neugier gliedert sich in die Klassenunsicherheit und die Attributs- bzw. die Relationsunsicherheit (vgl. Abschnitt 4.3.5; Abb. 6.3). Die einzelnen Unsicherheiten werden im OPASCA-System in Analysemodulen getrennt bestimmt und anschließend durch ein Fusionsmodul zur Gesamtunsicherheit für jeden Objektrepräsentanten im Umweltmodell separat fusioniert (vgl. Abschnitt 4.3.7).

Module zur Bestimmung des Konflikts

Der letzte Teilaspekt der wissensbasierten Neugier ist der *Konflikt* und besteht aus drei Bestandteilen: Konflikt bei der Identitätsbestimmung, Konflikt bei der multimodalen Wahrnehmung und Konflikt mit dem a-priori-Wissen (vgl. Abschnitt 4.3.6; Abb. 6.3). Wie bei den anderen Teilaspekten zuvor auch werden die Teilkonflikte zunächst einzeln in entsprechenden Analysemodulen bestimmt, bevor anschließend der Gesamtkonflikt für jeden Repräsentanten im Umweltmodell separat durch ein Fusionsmodul bestimmt wird (vgl. Abschnitt 4.3.7).

Modul zur Bestimmung der Neugier

Alle vier Teilaspekte der Neugier werden durch eine Kombination von verschiedenen Analyse- und Fusionsmodulen realisiert. Die Neugier selbst wird, wie eingangs dargestellt (vgl. Abb. 6.3), durch ein weiteres Fusionsmodul für jeden Objektrepräsentanten im Umweltmodell bestimmt.

6.4 Interessengetriebene Szenenexploration

Die interessengetriebene Exploration einer Szene ist ein komplexer Prozess (vgl. Abb. 6.1), bei dem potentiell alle Komponenten des Systems (vgl. Abschnitt 6.2) beteiligt sind: Von der Sensordatenakquisition über die verschiedenen Verarbeitungsmodule, welche Wissen über Objekte liefern, bis hin zu Systembestandteilen wie die wissensbasierte Neugier oder auch die Mensch-Maschine-Interaktion sowie die Ausrichtung der Sensoren durch Aktoren sind alle Systemkomponenten involviert.

6.4.1 Allgemeine Vorgehensweise

In einem ersten Schritt wird eine initiale Menge an Objekten bestimmt, welche im Rahmen der Exploration tiefgehender erfasst werden soll (vgl. Abschnitt 3.4). Dazu können – beispielsweise durch einen ersten Blick über die Szene zu Beginn des Explorationsprozesses – Objekte auf einem relativ hohen Abstraktionsniveau im Umweltmodell erfasst werden. Diese Menge an Objekten dient als Ausgangsbasis für die Exploration und kann im weiteren Verlauf – auch während der Exploration – durch neue Objekte ergänzt werden.

Anschließend wird anhand der ausgewählten Explorationsstrategie ein Explorationspfad bestimmt (vgl. Abschnitt 5.2). Die darauf befindlichen Objekte werden im Folgenden nach und nach analysiert. Dies geschieht durch eine sukzessive Fokussierung der Objekte entlang des Pfads unter Verwendung entsprechender Aktoren zur Ausrichtung der Sensoren, da eine gleichzeitige detaillierte Erfassung aller Objekte in realen Anwendungen aufgrund von beschränkten Sensorbereichen und Rechenressourcen i. d. R. nicht möglich ist. Das aktuelle Objekt wird dabei durch Sensoren mit verschiedenen Modalitäten im Rahmen der objektzentrierten Erfassung detaillierter wahrgenommen. Dazu werden mit Hilfe der Verarbeitungsmodule (vgl. Abschnitt 6.3) die Objektinformationen aus den Sensordaten gewonnen und dem zu untersuchenden Objektrepräsentant im Umweltmodell zugeordnet. Im Laufe der Zeit werden weitere Informationen akquiriert und infolgedessen sinkt i. d. R. das Abstraktionsniveau. Dabei können neben den reinen Sensorinformationen auch andere Informationsquellen (z. B. externe Wissensdatenbanken, Interaktionen mit den in der Umgebung befindlichen Menschen) hinzugezogen werden, um Konflikte zu lösen und um zusätzliches Wissen einzubringen, welches u. U. nicht direkt durch Sensoren erfasst werden kann (z. B. Informationen über den Wohnort oder Geschwister einer Person).

Während der Exploration wird entweder jedem Objekt eine gewisse Zeit für die Erfassung zugewiesen oder es werden Kriterien definiert, die erfüllt werden müssen, bevor das nächste Objekt fokussiert werden kann. Diese Kriterien können

beispielsweise eine gewisse Stufe in der Klassenhierarchie oder ein bestimmter Detailgrad bei der Objekterfassung sein (vgl. Abschnitt 3.2). Die Exploration einer Szene ist abgeschlossen, sobald alle Objekte entsprechend den definierten Kriterien hinreichend erfasst wurden oder die geplante Zeit für die gesamte Exploration einer Szene überschritten wurde.

Aufgrund der Komplexität einer Szene macht es Sinn, bei der Bestimmung des Explorationspfads verschiedene Priorisierungskriterien anzugeben (vgl. Abschnitt 5.4), sodass bei einer Unterbrechung bzw. einem vorzeitigen Beenden der Exploration ein Großteil der wichtigsten Objekte bereits erfasst wurden, da diese die zuvor festgelegten Kriterien überdurchschnittlich erfüllt haben. Dazu wurde im Rahmen der vorliegenden Arbeit die interessengetriebene Explorationsstrategie in Kapitel 5 eingeführt.

6.4.2 Besonderheiten der interessengetriebenen Szenenexploration

Bei der interessengetriebenen Explorationsstrategie werden aus den aktuellen Objektinformationen im Umweltmodell und den Priorisierungskriterien der Explorationsstrategie die Priorität eines jeden Objekts bestimmt, welche infolgedessen die Generierung eines Explorationspfads ermöglicht. Die Priorität der Objekte für die Exploration wird sowohl durch die wissensbasierte Neugier als auch durch die multimodale Salienz bestimmt. Im OPASCA-System existiert dafür eine Vielzahl an Modulen zu deren Bestimmung (vgl. Abschnitte 6.3.2 und 6.3.3). Des Weiteren wird der jeweils aktuelle Erfassungsbereich der Sensoren bei der Festlegung des Explorationspfads mit berücksichtigt und in Form einer Bewegungsminimierung als drittes Kriterium für die interessengetriebene Explorationsstrategie festgelegt (vgl. Abschnitt 5.4.2).

Während der Exploration müssen Kriterien festgelegt werden, welche entscheiden, wann ein Objekt ausreichend erfasst wurde und infolgedessen mit dem nächsten Objekt auf dem Explorationspfad fortgefahren werden kann. Im Falle der interessengetriebenen Exploration werden zwei Kriterien definiert: zum einen die aktuelle Salienz des Objekts und zum anderen die Neugier für das aktuelle Objekt.

Die multimodale Salienz definiert in der vorliegenden Arbeit, wie sehr sich sowohl die visuellen als auch akustischen Merkmale eines Objekts von denen seiner Umgebung unterscheiden (vgl. Abschnitt 2.3). Die Salienz wird daher häufig für die Bestimmung der Aufmerksamkeit genutzt. Im Rahmen der Exploration wird durch die Betrachtung eines Objekts ein Abschwächungsprozess initiiert, der im Laufe der Zeit zu einer vollständigen Unterdrückung der Salienz für das aktuelle Objekt führt, sodass das Objekt infolgedessen aufgrund seiner Salienz nicht ein zweites Mal untersucht wird (engl.: inhibition of return; vgl. [Küh12b]).

Die wissensbasierte Neugier ist ein wichtiger Faktor für das Interesse an einem Objekt. Im Laufe einer tiefgehenderen Objekterfassung und der damit verbundenen Wissensakquise sinkt i. d. R. die Neugier, sodass bei Unterschreiten einer zuvor definierten Schwelle ein Objekt als hinreichend erfasst gilt. Dazu muss allerdings möglichst detailliertes und vollständiges Wissen über ein Objekt vorhanden sein, es dürfen keine Konflikte mehr vorhanden sein und die Unsicherheit für die einzelnen Attribute und Relationen muss auf einem niedrigen Niveau sein. Dies kann erreicht werden, indem ein Objekt intensiver und über einen längeren Zeitraum betrachtet wird und dabei beispielsweise Zoomkameras oder foveale Kameras mit entsprechender Brennweite für eine detaillierte sensorische Erfassung eingesetzt werden. Des Weiteren kann auch der Blickwinkel verändert werden, um zusätzliche Informationen aus dem anderen Betrachtungswinkel zu erhalten. Sind einzelne Aspekte eines Objekts nicht alleine anhand der Sensorinformationen zu erfassen, so können weitere Informationsquellen – wie beispielsweise externe Datenbestände – hinzugezogen werden. Außerdem kann, falls gerade verfügbar und das Problem auf eine andere Weise nicht gelöst werden kann, auch der Mensch im Rahmen eines Dialogs zur Lösung des Problems beitragen. Ist dies nicht möglich, so kann das Objekt entsprechend gekennzeichnet werden und zu einem späteren Zeitpunkt erneut betrachtet werden (z. B. nach dem Ende der eigentlichen Exploration). Die Definition einer Schwelle für die Neugier legt indirekt auch die Detailliertheit der Erfassung und die benötigte Zeit für ein Objekt während der Exploration fest. Die Festlegung der Schwelle hängt daher von der konkreten Anwendung und den äußeren Bedingungen ab, z. B.: *Sind Personen oder andere zusätzliche Informationsquellen vorhanden? Wurde die Umgebung bereits zuvor einmal erfasst?* u. v. m.

Diesen zuvor beschriebenen komplexen Ablauf steuert ein Dienstleistungsmodul zur Szenenexploration (vgl. Abb. 6.1), welches direkten Zugriff auf die Informationen aus dem Umweltmodell, die Aktoren für die Objektfokussierung und die Mensch-Maschine-Interaktion zur Akquise von zusätzlichen Objektinformationen sowie zum Lösen von Konflikten hat. Indirekt stehen über das Umweltmodell auch Informationen wie die wissensbasierte Neugier, die multimodale Salienz und weitere Daten aus den Verarbeitungsmodulen zur Verfügung.

6.4.3 Tiefgehendere Objekterfassung am Beispiel einer Person

Für eine tiefgehendere Erfassung eines Objekts sind verschiedene Systemmodule verantwortlich. Im nachfolgenden Beispiel wird eine Person detektiert, identifiziert und nachverfolgt. Die dafür benötigten Module sowie die strukturellen Zusammenhänge und Abhängigkeiten sind in Abb. 6.4 visualisiert. Die Module sind entsprechend ihrer Kategorie (vgl. Abschnitt 6.3.1) farblich gekennzeichnet. Das komplette OPASCA-System enthält über 50 verschiedene Module, welche

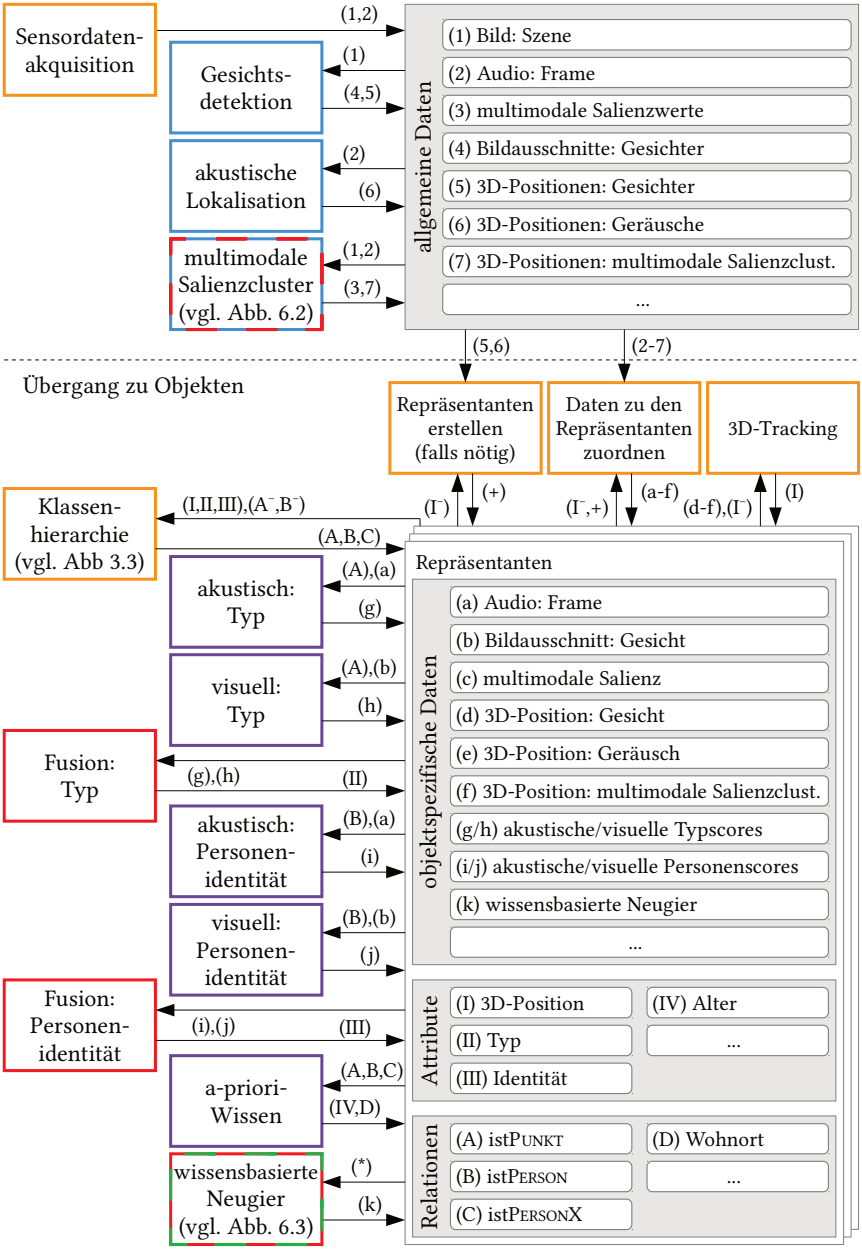

Abb. 6.4: Der Ausschnitt des OPASCA-Systems zur multimodalen Erfassung von Personen zeigt die Interaktion der Module mit dem Umweltmodell. Initiierungsmodule (blau); Spezialisierungsmodule (lila); Fusionsmodule (rot); Dienstleistungsmodule (orange); Analysemodule (grün); (+) = neuer Repräsentant; (*) = alle objektspezifischen Daten, Attribute und Relationen des Repräsentants; (X^-) = spezifische Daten, Attribute bzw. Relationen aus dem vorherigen Zeitschritt.

u. a. weitere Aspekte von Personen – beispielsweise die Körpergröße – erfassen können sowie die multimodale Wahrnehmung von Gegenständen ermöglichen.

Im oberen Bereich (Abb. 6.4) sind die Module dargestellt, welche *allgemeine Daten* generieren und/oder benötigen (1-7). Hierzu zählen u. a. sowohl die Akquisition von Sensordaten durch ein Dienstleistungsmodul als auch die initiale Detektion von grundlegenden Merkmalen (z. B. *Gesichtsdetektion* oder *akustische Geräuschlokalisation*) durch Initiierungsmodule sowie die Bestimmung der multimodalen Salienzcluster durch eine Kombination von Initiierungs- und Fusionsmodulen.

Im mittleren Bereich (Abb. 6.4) erfolgt der Übergang von allgemeinen zu *objektspezifischen Daten*. Dafür verantwortlich sind das Dienstleistungsmodul zur Datenassoziation sowie das Dienstleistungsmodul für die Erstellung von neuen Repräsentanten. Beides geschieht über die *3D-Positionen* der Sensordaten bzw. der Objekte. Dabei wird die Objektposition mit Hilfe eines weiteren Dienstleistungsmoduls nachgeführt. In diesem Beispiel werden über die 3D-Positionen der Gesichter sowie die akustischen 3D-Positionen der Sprache neue Objektrepräsentanten erzeugt, welche Personen darstellen, sowie allgemeine Daten zu den bereits vorhandenen Repräsentanten zugeordnet.

Im unteren Bereich (Abb. 6.4) sind die Repräsentanten selbst dargestellt, mit den *objektspezifischen Daten* (a-k), den *Attributen* (I-IV) und den *Relationen* (A-D), sowie die Module, welche aus den objektspezifischen Daten das Objektwissen (d. h. Attribute und Relationen) erzeugen. Dazu wird eine Kombination aus Spezialisierungs- und Fusionsmodulen verwendet, welche die objektspezifischen Daten wie beispielsweise *Audioframes*, *Bildausschnitte* oder die *multimodale Salienz* der zugeordneten Salienzcluster (vgl. Abschnitt 5.3.2) sowie das bereits vorhandene Wissen für eine sukzessive Wissensgenerierung nutzen (vgl. Abschnitt 3.2). Dabei wird zunächst ausgehend von der *Position* als erstes Attribut der *Typ* und anschließend bei Personen auch die *Identität* bestimmt. Ein weiteres Dienstleitungsmodul bestimmt die aktuelle Klassenhierarchie eines Objekts (vgl. Abschnitt 3.2.2). Für Personen werden beispielsweise die Klassenrelationen istPunkt, istPerson und istPersonX (eine spezifische Person) ermittelt. Mit Hilfe von a-priori-Wissen oder durch weitere Systemmodule, welche die Sensordaten auswerten, kann das *Alter* einer Person bestimmt werden bzw. der aktuelle *Wohnort* aus einer Wissensdatenbank abgerufen werden.

Das Modul für die Bestimmung der *wissensbasierten Neugier* benötigt die vorhandenen objektspezifischen Daten und das generierte Wissen (Attribute und Relationen). Zusammen mit der multimodalen Salienz dienen diese Informationen als Grundlage für die Bestimmung eines Explorationspfads im Rahmen der interessengetriebenen Szenenexploration (vgl. Abschnitt 5.4.2).

Im System sind neben den hier dargestellten Modulen (vgl. Abb. 6.4) noch weitere Module zur Wahrnehmung von Gegenständen und weiteren Merkmalen von

Personen vorhanden. Dabei wird aufgrund der Wissensabhängigkeiten (vgl. Abschnitt 3.2.3) nicht jedes Modul für alle Repräsentanten eingesetzt, sondern nur die passenden Module für den jeweiligen Objekttyp anhand der Klassenhierarchie dynamisch verwendet. Dieser Ansatz ermöglicht eine flexible Erweiterung durch neue Module und garantiert eine individuelle tiefgehende Erfassung aller Objekte in einer Szene.

6.5 Schlussbetrachtungen

In diesem Kapitel wurde ein modulares System vorgestellt, welches die in den vorherigen Kapiteln beschriebenen theoretischen Grundlagen in einem realen System vereint. Durch seinen modularen Aufbau ist eine künftige Erweiterung bzw. Verbesserung mit neuen Verfahren und Ansätzen flexibel möglich.

Die Anforderungen an die interessengetriebene Szenenexploration (vgl. Abschnitt 5.3.1) wurden beim Entwurf des zuvor beschriebenen OPASCA-Systems und der Interaktion der Teilkomponenten (insbesondere der Module) berücksichtigt. Die priorisierte Erfassung der Umgebung im Rahmen der interessengetriebenen Exploration (vgl. Abschnitt 6.4.2) gewährleistet eine bevorzugte Exploration von Objekten, welche bestimmte Kriterien besonders gut erfüllen. Die Kriterien sind die wissensbasierte Neugier, die multimodale Salienz und die Berücksichtigung des aktuellen Wahrnehmungsbereichs der Sensoren. Alle drei Kriterien lehnen sich an menschliches Verhalten bei der Exploration an.

Des Weiteren wurde auch der Einfluss von externen Faktoren auf den Ablauf der Exploration berücksichtigt und in Form von Unterbrechungen und Fortsetzen sowie vorzeitigem Beenden der Exploration realisiert. Darüber hinaus können durch externe Informationsquellen, wie beispielsweise Wissensdatenbanken oder durch die Interaktion mit Menschen, weitere Objektinformationen bezogen werden.

Die erweiterte objektzentrierte Umwelterfassung wurde abschließend anhand eines Beispiels für die Wahrnehmung von Personen in einer Szene in Abschnitt 6.4.3 erläutert. Dabei wurde die Beziehung der Module untereinander, die Interaktion mit dem Umweltmodell, der Bezug zu den Klassenhierarchien sowie die Generierung von Salienz und Neugier als auch die Erzeugung von Attributen und Relationen im Zusammenhang erläutert.

7

Experimentelle Evaluation

In diesem Kapitel wird die interessengetriebene Explorationsstrategie im Vergleich zu anderen Explorationsstrategien evaluiert. Dies erfolgt sowohl mit Simulations- als auch mit realen Evaluationsdaten und unter Zuhilfenahme des zuvor vorgestellten OPASCA-Systems. In diesem Kontext wird für eine objektive Evaluation ein neues Maß für den Vergleich von Explorationspfaden eingeführt sowie zusätzlich verschiedene Hardwareplattformen verwendet. Im Rahmen der Evaluation kann gezeigt werden, dass mit Hilfe der drei Priorisierungskriterien wissensbasierte Neugier, multimodale Salienz und minimale Bewegung gezielt Schwerpunkte bei der interessengetriebenen Exploration gesetzt werden können.

7.1 Demonstratoren

Im Folgenden werden sogenannte *Demonstratoren* vorgestellt, welche für die im späteren Verlauf beschriebene Evaluation genutzt werden. Dabei ist ein Demonstrator immer eine spezifische Hardwareplattform (z. B. ein bestimmter humanoider Roboter), welche im Rahmen der Evaluation genutzt bzw. für die öffentliche Demonstration von Funktionalitäten und Fähigkeiten verwendet wird. Es werden im Rahmen der Evaluation zwei verschiedene Hardwareplattformen verwendet: zum einen der im Sonderforschungsbereich 588 entwickelte ARMAR-III-Roboterkopf (vgl. [Asf08]) und zum anderen ein mechanisch vereinfachter Sensoraufbau, welcher über ähnliche Sensoren verfügt wie der Roboterkopf. Beide Hardwareplattformen verwenden für die Evaluation das in Kapitel 6 beschriebene OPASCA-System.

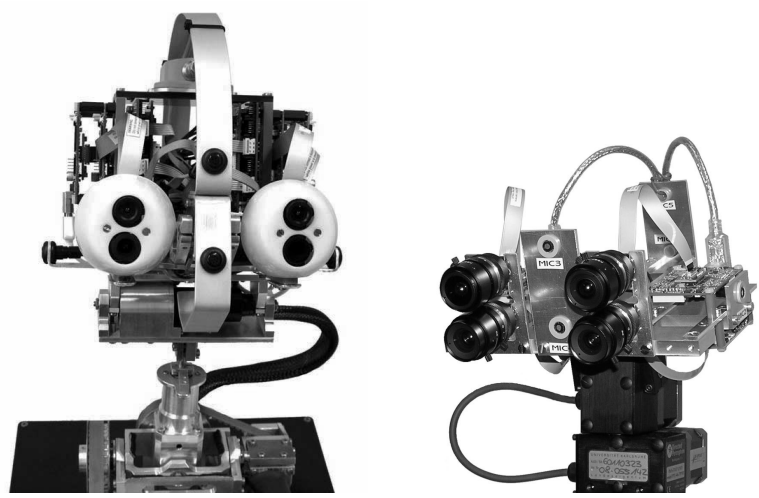

Abb. 7.1: Der ARMAR-III-Roboterkopf (links) und der PTU-Sensoraufbau (rechts) besitzen einen ähnlichen Sensorbau mit zwei Stereokamerapaaren und einem Mikrofonarray.

7.1.1 ARMAR-III-Roboterkopf

Im Rahmen des Sonderforschungsbereichs 588 wurden mehrere Generationen des humanoiden Roboters ARMAR entwickelt (vgl. [Son12], [Asf06], [Asf13]). Dieser soll den Menschen bei seiner täglichen Arbeit, insbesondere in der Küche, unterstützen. Neben dem kompletten Roboter stehen auch einzelne ARMAR-Köpfe (vgl. [Asf08]) – in größerer Stückzahl – zur Verfügung, welche baugleich mit ARMAR-III sind und somit dieselben Bewegungs- und Perzeptionseigenschaften wie das Original besitzen. In der vorliegenden Arbeit wurde ein solcher Kopf (vgl. Abb. 7.1; links) verwendet, um die nachfolgenden Evaluationen durchzuführen.

7.1.2 PTU-Sensoraufbau

Im Rahmen der vorliegenden Arbeit wurde darüber hinaus ein vereinfachter multimodaler Sensoraufbau entwickelt (vgl. Abb. 7.1; rechts), welcher ähnliche perzeptive Eigenschaften wie der ARMAR-III-Roboterkopf besitzt, jedoch in seiner Herstellung kostengünstiger ist. Für viele Perzeptionsaufgaben (z. B. die Erfassung der Umwelt) bildet die Kombination aus einer Schwenk-Neige-Einheit (engl.: pan-tilt unit; kurz: PTU), mehreren Kameras und einem Mikrofonarray eine sehr gute Ausgangsbasis. Im Gegensatz zum ARMAR-III-Roboterkopf können die Anzahl und die Positionen der Kameras und Mikrofone variabel verändert werden, sodass beispielsweise die Basis der Stereokameras angepasst werden

	ARMAR-III-Kopf	PTU-Sensoraufbau
Stereokamerapaare	2	2
- Brennweite	3,5 mm / 6,0 mm	4,0 mm / 12,0 mm
- max. Auflösung	640 × 480 Pixel	640 × 480 Pixel
- max. Bildwiederholrate	30 fps	30 fps
Mikrofonarray	1	1
- Mikrofone	6 (omnidirektional)	6 (omnidirektional)
- Abtastrate	48 kHz	48 kHz
Sonstiges		
- Freiheitsgrade	7	2 (Schwenk-Neige-Einheit)
- inertiale Messeinheit	1	–

Tabelle 7.1: Vergleich der Sensoren und Aktoren der beiden verwendeten Demonstratoren

kann. Des Weiteren kann durch das Hinzufügen zusätzlicher Mikrofone eine höhere Lokalisationsgenauigkeit erzielt werden. Der Austausch einzelner Sensoren ist durch die Bauweise einfach und flexibel möglich.

7.1.3 Vergleich der Hardwarekomponenten

In Tabelle 7.1 sind die wichtigsten Eckdaten der beiden Demonstratoren gegenübergestellt. Die Demonstratoren besitzen je vier Farbkameras des Typs *DragonFly2* der kanadischen Firma *Point Grey* (vgl. [Poi14]). Die Kameras lassen sich paarweise zu zwei Stereokameras mit unterschiedlicher Brennweite kombinieren. Damit soll sichergestellt werden, dass sowohl ein großer Teil der Szene im Überblick als auch ein detaillierter Ausschnitt im Zentrum in 3D erfasst werden kann. Die beiden Hardwareplattformen befinden sich während der Evaluation in zwei unterschiedlichen Räumen. Durch die verschiedenen Raumgrößen werden Brennweiten von 6 mm bzw. 12 mm zur Wahrnehmung von Details in der jeweiligen Szene und 3,5 mm bzw. 4 mm zur Perzeption der Szene im Überblick verwendet. Die Anordnung der Kameras und Mikrofone auf den Demonstratoren ist bei beiden Hardwareplattformen vergleichbar. Der ARMAR-III-Kopf besitzt Mikrofone vom Typ *ECM-C115.CE7* von *Sony* (vgl. [Son14]), während der PTU-Sensoraufbau über Mikrofone vom Typ *MCE 60.18* der Firma *beyerdynamic* (vgl. [bey13]) verfügt, welche beide eine omnidirektionale Charakteristik aufweisen. Der ARMAR-III-Kopf hat insgesamt sieben Freiheitsgrade, verteilt in Hals, Kopf und Augen, wohingegen die Schwenk-Neige-Einheit des Sensoraufbaus nur zwei Freiheitsgrade aufweist. Der ARMAR-III besitzt zusätzlich Inertialsensoren zur Messung von Beschleunigungen und Drehraten. Die verbaute inertiale Messeinheit (engl.: inertial measurement unit; kurz: IMU) der Firma *Xsens*

(vgl. [Xse14]) ermöglicht eine kombinierte 6D-Messung von Orientierung und Beschleunigung.

7.2 Datensätze

Für die nachfolgende Evaluation wurden im Rahmen der Arbeit zwei Datensätze erstellt: zum einen ein Simulationsdatensatz und zum anderen ein Realdatensatz mit Daten der zwei zuvor vorgestellten Demonstratoren.

7.2.1 Simulationsdatensatz

Der Datensatz für die Simulation besteht aus 100 Szenen, welche jeweils zwischen 10 und 20 Objekte beinhalten, die zufällig in einem virtuellen Raum verteilt sind (vgl. Abb. 7.2). Die multimodalen Salienzwerte s_o (vgl. Abb. 7.3; links) und die wissensbasierte Neugier η_o (vgl. Abb. 7.3; rechts) der Objekte $o \in \mathcal{O}$, welche für die Evaluation benötigt werden, sind entsprechend einer beschränkten Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ mit $\mu = 0{,}5$ und $\sigma = 1/6$ im Intervall $(0, 1]$ verteilt. Diese Anordnung spiegelt die Situation wider, in der nur wenige Objekte eine sehr hohe bzw. sehr niedrige Salienz besitzen, d. h., sehr dominant aus der Szene hervorstechende bzw. nicht weiter auffallende Objekte sind eher unwahrscheinlich. Des Weiteren induzieren die meisten Objekte eine mittlere Neugier, da diese beispielsweise weder völlig neuartig sind noch mit maximaler Sicherheit vollständig erfasst wurden. Die Anzahl an Objekten pro Szene entspricht über alle Szenen gemittelt näherungsweise einer Gleichverteilung für den Wertebereich $\{10, \ldots, 20\}$.

Die Fokussierung der Objekte wird mit Hilfe einer simulierten Schwenk-Neige-Einheit und einer darauf montierten virtuellen Kamera realisiert. Der Aufbau befindet sich an einem festen Ort im virtuellen Raum und kann durch seine zwei Freiheitsgrade alle Objekte im Raum durch gezielte Bewegungen erfassen. Zur Exploration der Objekte auf dem Explorationspfad werden diese durch Neigen und Schwenken der virtuellen Kamera stets in das Zentrum der Kamera gerückt.

7.2.2 Realdatensatz

Für die Evaluation werden Aufnahmen von Szenen mit verschiedenem Inhalt erstellt. Hierzu zählt ein Küchenszenario (vgl. Abb. 7.4; oben), das Geschirr, Geräte, wie beispielsweise Toaster oder Eierkocher, aber auch diverse Frühstückszutaten wie beispielsweise Cornflakes beinhaltet. Des Weiteren ist ein Büroszenario

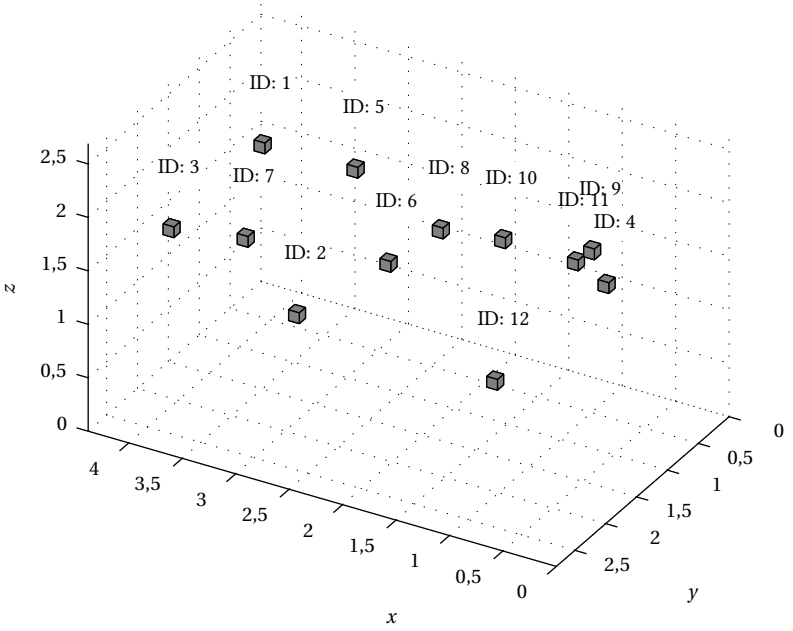

Abb. 7.2: Beispiel für die Verteilung der Objekte in einem virtuellen Raum

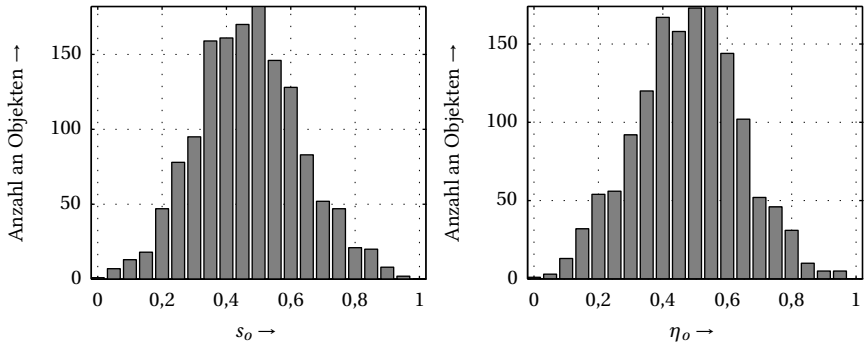

Abb. 7.3: Verteilung der Salienzwerte (links) und der wissensbasierten Neugier (rechts) im Simulationsdatensatz

(vgl. Abb. 7.4; unten) mit Büchern, Computer und Arbeitsmaterialien vorhanden. Es existiert außerdem eine gewisse Variation in den Szenarien, da die Elemente in einer Szene bzgl. ihrer Anzahl und Position über die einzelnen Aufnahmen hinweg variiert werden. Zusätzlich werden die Aufnahmen mit unterschiedlichen Demonstratoren (vgl. Abschnitt 7.1) durchgeführt, welche je in einem anderen

Abb. 7.4: Zusammengesetzte Bildausschnitte der verschiedenen Szenarien durch einen Schwenk über die aktuelle Szene. In Raum 1 wurde der ARMAR-III-Kopf verwendet (oben) und in Raum 2 wurde der Sensoraufbau mit Schwenk-Neige-Einheit und zwei Stereokameras verwendet (unten).

Raum stehen. Diese Vorgehensweise soll zeigen, dass die interessengetriebene Szenenexploration unabhängig von der Umgebung und den verwendeten Demonstratoren funktioniert. Es wurden in jedem Raum bzw. mit jedem Demonstrator je 25 Szenarien aufgenommen, welche eine Länge von 30 s besitzen und einen Schwenk über die Szene beinhalten. Die Aufnahmen umfassen sowohl die aktuellen audiovisuellen Sensordaten als auch die aktuelle Ausrichtung und Bewegung der Sensoren. Für jede Aufnahme wurden zusätzlich vier verschiedene Sätze an a-priori-Informationen hinterlegt, welche einen unterschiedlichen Umfang sowie eine variable Granularität an vorhandenem Objektwissen (vgl. Abschnitt 3.1.1) aufweisen. In Tabelle 7.2 ist eine Übersicht über die Zusammensetzung des Realdatensatzes zu sehen.

7.3 Bewertungsmaße

Die Bewertung der in Abschnitt 5.4 vorgestellten Explorationsstrategien erfolgt mit Hilfe von drei Maßen. Diese bewerten einen beliebigen Explorationspfad anhand der drei Kriterien: multimodale Salienz, wissensbasierte Neugier und minimale Bewegung. Hierbei wird der salienzbasierte Explorationspfad als Optimum

Szenario	*Anzahl an Videosequenzen* ARMAR-III-Kopf	PTU-Sensoraufbau	Summe
Frühstück	15	15	30
Büro	10	10	20
Insgesamt	25	25	50

Tabelle 7.2: Aufbau des Realdatensatzes für die Evaluation der verschiedenen Explorationsstrategien. Für jede Aufnahme wurden jeweils vier verschiedene Sätze an a-priori-Wissen in unterschiedlicher Granularität hinterlegt.

für die rein salienzgetriebene Exploration festgelegt, der neugierbasierte Explorationspfad für eine rein neugiergetriebene Exploration und der bewegungsoptimierte Pfad für eine Exploration mit minimaler Bewegung, d. h. unter Berücksichtigung des aktuellen Umweltausschnitts.

7.3.1 Abnahme der Salienz

Für einen gegebenen Explorationspfad lässt sich anhand der Salienz der einzelnen Objekte ein Maß bestimmen, das angibt, wie sehr die Salienz bei der Erstellung des Pfads, durch eine entsprechende Explorationsstrategie, berücksichtigt wurde. Dazu wird eine monoton ansteigende Funktion definiert, welche in Abhängigkeit des Explorationspfads die abschnittsweise kumulierte Salienz (engl.: cumulated saliency; kurz: CS) beschreibt durch

$$\mathrm{CS}(i;\mathrm{EP}) = \mathrm{CS}(i-1;\mathrm{EP}) + \frac{s_{\mathrm{EP}_i}}{\mathrm{CS}_{\max}(\mathrm{EP})} \quad \text{mit } \mathrm{CS}(0;\mathrm{EP}) = 0\,. \tag{7.1}$$

Hierbei ist s_{EP_i} der Salienzwert des i-ten Objekts auf dem Explorationspfad EP. Für eine Vergleichbarkeit zwischen verschiedenen Messungen bzw. Explorationspfaden wird der aktuelle Salienzwert mit der kumulierten Salienz des gesamten Pfads

$$\mathrm{CS}_{\max}(\mathrm{EP}) = \sum_{i=1}^{N} s_{\mathrm{EP}_i} \tag{7.2}$$

normiert, wobei N die Anzahl der Objekte auf dem Pfad repräsentiert. Im Idealfall werden Objekte mit einem größeren Salienzwert zuerst untersucht, sodass der Anstieg der abschnittsweise definierten Kurve mit der Anzahl der untersuchten Objekte abnimmt. Die Fläche unter der Kurve $\mathrm{CS}(i;\mathrm{EP})$ definiert ein Maß (engl.: integrated cumulated saliency; kurz: ICS), welches eine Aussage über die Berücksichtigung der Salienz bei der Bestimmung des Explorationspfads EP macht:

$$\mathrm{ICS}(\mathrm{EP}) = \int_0^N \mathrm{CS}(p;\mathrm{EP})\,\mathrm{d}p\,. \tag{7.3}$$

Dieses Maß lässt sich durch die Trapez- oder Simpson-Regel (vgl. [Fre07]) numerisch approximieren. Dabei ist der Abstand zwischen den einzelnen Punkten der Kurve bei der Approximation umgekehrt proportional zu der Anzahl an Objekten auf dem Explorationspfad. Diese Betrachtung ist notwendig, da im Rahmen der späteren Evaluation Pfade mit unterschiedlicher Länge verglichen werden.

Abschließend lässt sich feststellen, dass bei Pfaden, die salientere Objekte zuerst untersuchen, die Steigung der approximierten CS-Kurve im Verlauf stets abnimmt und somit im Vergleich zu anderen Pfaden immer eine größere Fläche entsteht. Das Maß bewertet somit die Abnahme der Salienzwerte entlang des Explorationspfads. Der Explorationspfad $\mathrm{EP}_{\mathrm{Salienz}}$ liefert für dieses Maß aufgrund seiner Definition (vgl. Abschnitt 5.4.1) den höchsten Wert.

7.3.2 Abnahme der Neugier

Analog zum Maß für die Abnahme der Salienz lässt sich ein Maß definieren, welches beurteilt, wie sehr ein gegebener Explorationspfad die wissensbasierte Neugier, die ein Objekt induziert, priorisiert. Dazu lässt sich zunächst die kumulierte Neugier (engl.: cumulated curiosity; kurz: CC) analog zur kumulierten Salienz definieren:

$$\mathrm{CC}(i;\mathrm{EP}) = \mathrm{CC}(i-1;\mathrm{EP}) + \frac{\eta_{\mathrm{EP}_i}}{\mathrm{CC}_{\max}(\mathrm{EP})} \quad \text{mit } \mathrm{CC}(0;\mathrm{EP}) = 0 \tag{7.4}$$

und

$$\mathrm{CC}_{\max}(\mathrm{EP}) = \sum_{i=1}^{N} \eta_{\mathrm{EP}_i}\,. \tag{7.5}$$

Hierbei ist η_{EP_i} die induzierte Neugier des i-ten Objekts auf dem Explorationspfad EP und N die Anzahl an Objekten auf dem Pfad. Die kumulierte Neugier aller Objekte $\mathrm{CC}_{\max}(\mathrm{EP})$ des Explorationspfads EP wird in Gl. 7.4 zur Normierung verwendet. Analog zur Abnahme der Salienz wird ein Maß definiert, welches die Fläche unter der approximierten Kurve beschreibt. Die integrierte kumulierte Neugier (engl.: integrated cumulated curiosity; kurz: ICC) bewertet dabei Explorationspfade anhand der Reihenfolge der berücksichtigten Objekte und deren induzierter Neugier. Die Berechnung erfolgt durch Integration der kumulierten Neugier entlang des Explorationspfads:

$$\mathrm{ICC}(\mathrm{EP}) = \int_0^N \mathrm{CC}(p;\mathrm{EP})\,\mathrm{d}p\,. \tag{7.6}$$

Die Fläche lässt sich wiederum, analog zur Salienz, mit Hilfe der Trapez- oder Simpson-Regel (vgl. [Fre07]) numerisch approximieren. Das Maß bewertet insgesamt die Abnahme der induzierten Neugier der Objekte entlang eines Explorationspfads. Der Explorationspfad $\mathrm{EP}_{\mathrm{Neugier}}$ stellt aufgrund seiner Definition (vgl. Abschnitt 5.4.1) das Optimum für dieses Maß dar.

7.3.3 Minimale Bewegung

Der momentane Ausschnitt der Umwelt, welcher mit den Sensoren aktuell erfasst werden kann, ist ein wichtiger Aspekt der interessengetriebenen Perzeption. Daraus abgeleitet wird ein Maß für die benötigte Bewegung während der Exploration definiert. Die Reduzierung der Bewegung über den kompletten Pfad optimiert gleichzeitig mehrere Parameter: Es wird die Zeit zum Fokussieren von Objekten, die benötigte Energie und gleichzeitig auch indirekt der Verschleiß an Bauteilen reduziert. Für einen kompletten Explorationspfad lässt sich über die kumulierten Gelenkwinkeldistanzen (engl.: cumulated joint angle distances; kurz: CJAD) ein Maß festlegen, welches die Bewegung (des Kopfs) repräsentiert:

$$\mathrm{CJAD}(\mathrm{EP}) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{Q}\left\|\boldsymbol{q}_{\mathrm{EP}_i} - \boldsymbol{q}_{\mathrm{EP}_{i\text{-}1}}\right\|\right). \tag{7.7}$$

Hierbei ist $\boldsymbol{q}_{\mathrm{EP}_i}$ die erforderliche Gelenkwinkelstellung, um das i-te Objekt auf dem Explorationspfad EP zu fokussieren, $\boldsymbol{q}_{\mathrm{EP}_{i\text{-}1}}$ hingegen die Gelenkwinkelstellung des zuvor untersuchten Objekts. $\boldsymbol{q}_{\mathrm{EP}_0}$ ist die Ausgangsgelenkwinkelstellung zu Beginn des Explorationsvorgangs. Um eine Unabhängigkeit von der Anzahl der zu explorierenden Objekte N und der vorhandenen Gelenkwinkel Q zu gewährleisten und um mehrere Explorationspfade vergleichen zu können, wird das Maß zur Bestimmung der Bewegung mit beiden Größen normiert. Kleine CJAD-Werte stellen eine geringe Bewegung dar und somit einen Pfad, der das Kriterium der (Kopf-)Bewegung sehr gut erfüllt. Der Explorationspfad $\mathrm{EP}_{\mathrm{Bewegung}}$ liefert aufgrund seiner Definition (vgl. Abschnitt 5.4.1) für das Maß der minimalen Bewegung das beste Ergebnis.

7.3.4 Normalisierung der Maße

Für eine bessere Vergleichbarkeit der Maße wird eine Normalisierung auf einen einheitlichen Wertebereich $[0, 1]$ eingeführt. Dabei kann die Tatsache ausgenutzt werden, dass die minimalen und maximalen Werte der zuvor definierten Maße für eine Menge an Objekten bekannt sind. Die optimalen Werte können durch die Explorationspfade $\mathrm{EP}_{\mathrm{Salienz}}$, $\mathrm{EP}_{\mathrm{Neugier}}$ bzw. $\mathrm{EP}_{\mathrm{Bewegung}}$ bestimmt werden. Die pessimalen Werte (d. h. schlechtesten Werte) können anhand der entgegengesetzten Explorationsstrategien und somit der Explorationspfade $\mathrm{EP}_{!\mathrm{Salienz}}$, $\mathrm{EP}_{!\mathrm{Neugier}}$ bzw. $\mathrm{EP}_{!\mathrm{Bewegung}}$ bestimmt werden, welche im nächsten Abschnitt beschrieben sind.

Pessimale Explorationsstrategien

Für das Maß der abnehmenden Salienz (ICS) ist der Maximalwert über den Explorationspfad $\text{EP}_{\text{Salienz}}$ definiert. Eine untere Grenze kann bestimmt werden, indem eine Strategie verfolgt wird, bei der stets das Objekt ausgewählt wird, welches die nächstgrößere Salienz besitzt, d. h., der Explorationspfad $\text{EP}_{!\text{Salienz}}$ folgt den aufsteigenden Salienzwerten der Objekte und lässt sich somit definieren als

$$\text{EP}_{!\text{Salienz}} = (o_{i_1}, o_{i_2}, \ldots, o_{i_N}) \quad \text{mit } s_{o_{i_1}} \leq, \ldots, \leq s_{o_{i_N}}. \tag{7.8}$$

Die Objekte $o_1, \ldots, o_N$ werden über die Indices $i_1, \ldots, i_N$ durch eine entsprechende Sortierung der Salienz s_{o_i} auf die gewünschte Reihenfolge für den Explorationspfad abgebildet. N ist dabei die Anzahl an Objekten.

Analoges gilt für das Maß der abnehmenden Neugier (ICC). Der Maximalwert ist über den Explorationspfad $\text{EP}_{\text{Neugier}}$ definiert. Die pessimale Explorationsstrategie, die stets das Objekt mit der niedrigsten induzierten Neugier als Nächstes auswählt, stellt eine untere Schranke dar und ist somit die schlechtmöglichste Strategie. Der dazugehörige Explorationspfad $\text{EP}_{!\text{Neugier}}$ ist definiert als

$$\text{EP}_{!\text{Neugier}} = (o_{i_1}, o_{i_2}, \ldots, o_{i_N}) \quad \text{mit } \eta_{o_{i_1}} \leq, \ldots, \leq \eta_{o_{i_N}}. \tag{7.9}$$

Hierbei ist η_{o_i} die wissensbasierte Neugier des Objekts o_i und N die Anzahl an Objekten. Die Objekte werden über die zusätzliche Indizierung $i_1, \ldots, i_N$ sortiert.

Das Maß für die Bewegung (CJAD) hat sein Minimum für den Explorationspfad $\text{EP}_{\text{Bewegung}}$, da bei dieser Explorationsstrategie aus allen möglichen Pfaden der Pfad ausgewählt wird, welcher die geringste Gelenkwinkeldistanz besitzt. Eine obere Grenze kann erreicht werden, indem eine entgegengesetzte Strategie verfolgt wird. Der pessimale Explorationspfad kann bestimmt werden durch

$$\text{EP}_{!\text{Bewegung}} = \underset{\text{EP} \in \mathcal{P}_{\mathcal{O}}}{\arg\max} \left\{ \sum_{k=1}^{N} d_{o_{i_k}} \right\} \quad \text{mit } d_{o_{i_k}} = \left\| \boldsymbol{q}_{o_{i_k}} - \boldsymbol{q}_{o_{i_{k-1}}} \right\|. \tag{7.10}$$

Die Bewegungsdistanz ist hierbei maximal und kein anderer Pfad besitzt ein insgesamt höheres und somit schlechteres Bewegungsverhalten.

Min-Max-Normalisierung

Werden die zuvor definierten Explorationsstrategien und Bewertungsmaße betrachtet, so lässt sich folgender Zusammenhang zwischen diesen herstellen:

$$\text{ICS}(\text{EP}_{!\text{Salienz}}) \leq \text{ICS}(\text{EP}) \leq \text{ICS}(\text{EP}_{\text{Salienz}}) \tag{7.11}$$

$$\text{ICC}(\text{EP}_{!\text{Neugier}}) \leq \text{ICC}(\text{EP}) \leq \text{ICC}(\text{EP}_{\text{Neugier}}) \tag{7.12}$$

$$\text{CJAD}(\text{EP}_{\text{Bewegung}}) \leq \text{CJAD}(\text{EP}) \leq \text{CJAD}(\text{EP}_{!\text{Bewegung}}) \tag{7.13}$$

Es existiert für jedes Maß eine obere und untere Grenze, die ein beliebiger Explorationspfad EP nicht über- bzw. unterschreitet. Durch diese Tatsache kann der Wertebereich für eine bekannte Menge an Objekten normiert werden. Dazu können verschiedene Normalisierungsverfahren eingesetzt werden, welche bereits in Abschnitt 4.3.7 vorgestellt wurden. Die *Min-Max-Normalisierung* behält die Ursprungsverteilung bei und transformiert diese in den Wertebereich $[0, 1]$. Das Ergebnis davon ist auch das gewünschte Ziel der hier angestrebten Normalisierung der Maße. Die Ober- und Untergrenzen zur Normalisierung lassen sich explizit anhand der Objekte bestimmen. Die Min-Max-Normalisierung ist allgemein wie folgt definiert:

$$v' = \frac{v - \min \mathcal{V}}{\max \mathcal{V} - \min \mathcal{V}} \quad \text{mit } v \in \mathcal{V} \tag{7.14}$$

Damit lassen sich neue normalisierte Maße für die Abnahme der Salienz (NICS), die Abnahme der Neugier (NICC) und die minimale Bewegung (NCJAD) wie folgt definieren:

$$\text{NICS}(\text{EP}) = \frac{\text{ICS}(\text{EP}) - \text{ICS}(\text{EP}_{!\text{Salienz}})}{\text{ICS}(\text{EP}_{\text{Salienz}}) - \text{ICS}(\text{EP}_{!\text{Salienz}})} \tag{7.15}$$

$$\text{NICC}(\text{EP}) = \frac{\text{ICC}(\text{EP}) - \text{ICC}(\text{EP}_{!\text{Neugier}})}{\text{ICC}(\text{EP}_{\text{Neugier}}) - \text{ICC}(\text{EP}_{!\text{Neugier}})} \tag{7.16}$$

$$\text{NCJAD}(\text{EP}) = \frac{\text{CJAD}(\text{EP}) - \text{CJAD}(\text{EP}_{!\text{Bewegung}})}{\text{CJAD}(\text{EP}_{\text{Bewegung}}) - \text{CJAD}(\text{EP}_{!\text{Bewegung}})}. \tag{7.17}$$

Die Maße sind jeweils so definiert, dass ein hoher Wert eine optimale Übereinstimmung und ein niedriger Wert eine suboptimale Übereinstimmung mit den jeweiligen Kriterien darstellt. Für NCJAD(EP) bedeutet dies, dass Pfade die eine geringe Bewegung verursachen, einen hohen Wert besitzen und Pfade mit viel Bewegung einen entsprechend niedrigeren Wert aufweisen. Bei NICS(EP) bzw. NICC(EP) führt die Berücksichtigung der Salienz bzw. der Neugier der einzelnen Objekte – in absteigender Reihenfolge – stets auch zu einem hohen Wert und umgekehrt bei Nichteinhaltung zu einem niedrigeren Wert.

7.4 Evaluation der Pfadoptimierung

Im Rahmen der Evaluation wird mit Hilfe der zuvor definierten normierten Maße ein optimaler Explorationspfad für eine gegebene Menge an Objekten bestimmt. Dazu wird zunächst untersucht, ob dies auch durch eine unabhängige Optimierung der drei Kriterien wissensbasierte Neugier, multimodale Salienz und Bewegung möglich ist. Die Möglichkeit dieser multikriteriellen Optimierung wird mittels Pareto-Optimierung untersucht.

7.4.1 Pareto-Optimierung

Die zuvor definierten Bewertungsmaße (vgl. Abschnitt 7.3) stellen verschiedene Optimierungsziele für einen Explorationspfad dar. Diese sind jedoch sehr unterschiedlich und lassen sich u. U. unabhängig voneinander optimieren. Um dies zu überprüfen, wird in den nachfolgenden Abschnitten die sogenannte *Pareto-Front* bestimmt und untersucht.

Die Idee dahinter stammt aus dem 19. Jahrhundert von Vilfredo Pareto und ist u. a. in den Wirtschaftswissenschaften (vgl. [Brü07]) weit verbreitet, z. B. bei der Bestimmung der Grenzproduktivität, d. h. der Optimierung der Produktionsauslastung. Die Erkenntnisse von Pareto finden Verwendung sowohl allgemein bei der multikriteriellen Optimierung von Problemen als auch speziell in der Spieltheorie (vgl. [Chi08]).

Definition

Die Pareto-Optimierung versucht, mehrere Eingangsgrößen gleichzeitig zu optimieren, sodass ein ideales Ergebnis erzielt wird. Dazu werden zunächst verschiedene Bewertungskriterien für die Eingangsgrößen definiert, die im Rahmen einer Optimierung minimiert oder maximiert werden. Dabei kann die Optimierung einer Eingangsgröße eine Verschlechterung einer anderen Eingangsgröße zufolge haben. Die *Pareto-Menge* – auch Pareto-Optima genannt – stellt eine Menge aller Kombinationen von Eingangsgrößen dar, bei denen keine Eingangsgröße anhand der Bewertungskriterien weiter optimiert werden kann, ohne eine andere zu verschlechtern. Die *Pareto-Front* stellt die Bildmenge (in Bezug auf die Bewertungskriterien) der Pareto-Menge dar und bildet die Grundlage der Optimierung.

Die Eingangsgrößen werden als m-Tupel $\boldsymbol{x} = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$ beschrieben und dienen zusammen mit den n Bewertungskriterien $f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})$, welche maximiert werden sollen, als Grundlage für die Optimierung. Dies kann als eine Abbildung $f : \mathbb{R}^m \to \mathbb{R}^n$ der Eingangsgrößen mit den Bewertungskriterien aufgefasst werden und ist wie folgt formuliert:

$$\boldsymbol{y} := (y_1, y_2, \ldots, y_n) = \big(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})\big) \text{ mit } \boldsymbol{x} \in \mathbb{R}^m,\ \boldsymbol{y} \in \mathbb{R}^n. \tag{7.18}$$

Sind nur bestimmte Kombinationen von Eingangsgrößen möglich, so lassen sich diese auch als eine Menge $\mathcal{X}$ beschreiben. Durch Anwendung der verschiedenen Bewertungskriterien ergibt sich somit eine Bildmenge $\mathcal{Y} := f(\mathcal{X})$, welche zur Pareto-Optimierung genutzt wird. Dabei ist die Pareto-Front $\mathcal{P}_{\text{Front}} \subseteq \mathcal{Y}$ und die Pareto-Menge $\mathcal{P}_{\text{Menge}} \subseteq \mathcal{X}$ die dazu korrespondierende Urbildmenge.

Eine bestimmte Kombination von Eingangsgrößen $\hat{\boldsymbol{x}} \in \mathcal{X}$ ist Element der Pareto-Menge, falls folgende Zusammenhänge gelten:

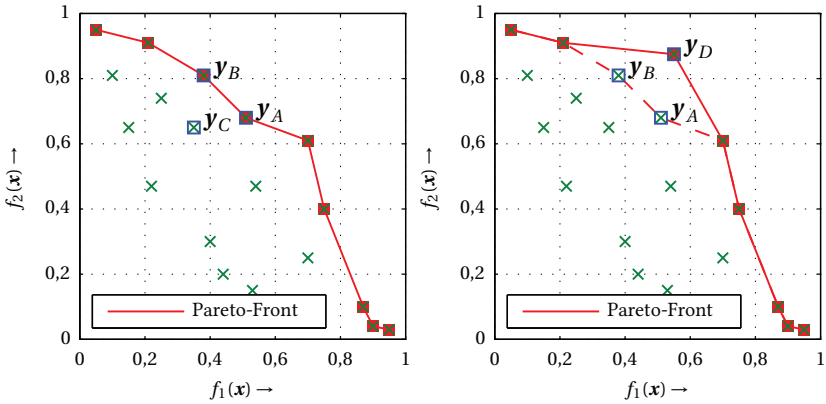

Abb. 7.5: Die Pareto-Front repräsentiert die Menge an Punkten, welche für die Bewertungskriterien $f_1(\boldsymbol{x})$ und $f_2(\boldsymbol{x})$ ein Optimum bilden, d. h., eine Verbesserung in Bezug auf ein Kriterium führt zu einer Verschlechterung bei mindestens einem anderen Kriterium. Der Punkt $\boldsymbol{y}_C$ wird durch die Punkte $\boldsymbol{y}_A$ bzw. $\boldsymbol{y}_B$ strikt dominiert (links), und falls der Punkt $\boldsymbol{y}_D$ existieren würde, so dominiert dieser die Punkte $\boldsymbol{y}_A$ bzw. $\boldsymbol{y}_B$ strikt (rechts). Die jeweils dominierten Punkte sind nicht Pareto-optimal und bilden somit keinen Teil der Pareto-Front. Die optimalen Eingangsgrößen in diesen Beispielen sind u. a. $\boldsymbol{x}_A$ und $\boldsymbol{x}_B$ (links) bzw. $\boldsymbol{x}_D$ (rechts).

$$\hat{\boldsymbol{x}} \in \mathcal{P}_{\text{Menge}} \iff \hat{\boldsymbol{y}} \in \mathcal{P}_{\text{Front}} \iff \nexists \boldsymbol{y} \in \mathcal{Y} \setminus \hat{\boldsymbol{y}} : \boldsymbol{y} \succ \hat{\boldsymbol{y}}\,. \tag{7.19}$$

Die Notation $\boldsymbol{y} \succ \hat{\boldsymbol{y}}$ bedeutet in diesem Fall, dass $\hat{\boldsymbol{y}}$ strikt dominiert wird durch $\boldsymbol{y}$. Die Existenz von strikt dominierenden Kombinationen ist ein notwendiges Kriterium für die Bestimmung der Pareto-Menge und ist wie folgt definiert:

$$\boldsymbol{y} \succ \hat{\boldsymbol{y}} : (\forall i \in \{1,\ldots,n\} : y_i \geq \hat{y}_i \wedge \exists j \in \{1,\ldots,n\} : y_j > \hat{y}_j)\,. \tag{7.20}$$

In Abb. 7.5 sind zwei Beispiele für die Pareto-Optimierung zu sehen. Dort sind die Bewertungskriterien $f_1(\boldsymbol{x})$ und $f_2(\boldsymbol{x})$ für die konkreten Kombinationen von Eingangsgrößen $\boldsymbol{x}_A, \boldsymbol{x}_B, \boldsymbol{x}_C, \boldsymbol{x}_D, \ldots, \boldsymbol{x}_Z \in \mathbb{R}^m$ dargestellt sowie die Pareto-Front mit den Pareto-optimalen Kombinationen. Im linken Bild ist zu sehen, dass der Punkt $\boldsymbol{y}_C \in \mathbb{R}^2$ nicht Teil der Pareto-Front ist, wohingegen die Punkte $\boldsymbol{y}_A \in \mathbb{R}^2$ und $\boldsymbol{y}_B \in \mathbb{R}^2$ es sind. Dies lässt sich einfach nachvollziehen, indem die strikte Dominanz der einzelnen Punkte überprüft wird. Für Punkt $\boldsymbol{y}_C$ gilt $\boldsymbol{y}_A \succ \boldsymbol{y}_C$ und $\boldsymbol{y}_B \succ \boldsymbol{y}_C$, weshalb dieser nicht Teil der Pareto-Front ist. Die Punkte $\boldsymbol{y}_A$ und $\boldsymbol{y}_B$ werden hingegen von keinem anderen Punkt strikt dominiert. Im rechten Bild (vgl. Abb. 7.5) ist dies jedoch nicht länger der Fall: Durch das Hinzufügen einer neuen Kombination von Eingangsgrößen ist der Punkt $\boldsymbol{y}_D \in \mathbb{R}^2$ entstanden, welcher u. a. die beiden Punkte $\boldsymbol{y}_A$ und $\boldsymbol{y}_B$ strikt dominiert, welche infolgedessen nicht länger Teil der Pareto-Front sind.

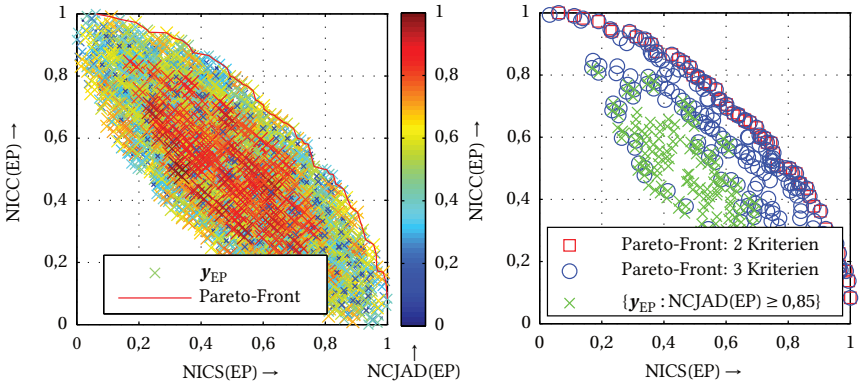

Abb. 7.6: Das Ergebnis der Pareto-Optimierung für eine gegebene Menge an Objekten zeigt die Pareto-optimalen Kombinationen (Pareto-Front). Die Bewertungskriterien sind die Maße NICS und NICC (links) bzw. NICS, NICC und NCJAD (rechts) für den jeweils zu untersuchenden Explorationspfad EP. Im linken Bild sind alle möglichen Explorationspfade dargestellt und zusätzlich ist der jeweilige NCJAD-Wert entsprechend der nebenstehenden Farbskala kodiert. Im rechten Bild hingegen sind nur die optimalen Explorationspfade bzgl. der gewählten Kriterien dargestellt.

Realisierung für die Optimierung von Explorationspfaden

Der zuvor beschriebene Ansatz kann auch für die Optimierung von Explorationspfaden herangezogen werden. Dabei wird ein möglicher Explorationspfad EP mit den drei zuvor definierten Maßen *Abnahme der Salienz* NICS(EP), *Abnahme der Neugier* NICC(EP) und *minimale Bewegung* NCJAD(EP) als Kriterien für die Optimierung bewertet. Für jeden potentiellen Explorationspfad $\text{EP} \in \mathcal{P}_{\mathcal{O}}$ (Menge aller möglichen Explorationspfade; vgl. Gl. 5.1) existiert somit ein Punkt $\boldsymbol{y}_{\text{EP}} \in \mathcal{Y}$, welcher zur anschließenden Bestimmung der Pareto-Front genutzt wird. Die Pareto-optimalen Explorationspfade bzgl. der zuvor genannten Kriterien lassen sich aus den Punkten der Pareto-Front ableiten (vgl. Gl. 7.19).

Der letztendlich verwendete Pfad für die Exploration wird aus der Pareto-Menge der Pfade ausgewählt. Dabei kann, falls mehrere Pfade vorhanden sind, noch eine gewisse Priorisierung vorgenommen werden, sodass beispielsweise alle Kriterien nahezu denselben Einfluss haben oder einzelne Kriterien, wie beispielsweise die Abnahme der Neugier bzw. der Salienz, bevorzugt werden. Die Pareto-Optimierung übernimmt diese Entscheidung nicht, diese liefert nur eine Menge von Pareto-optimalen Explorationspfaden.

In Abb. 7.6 sind die Ergebnisse der Pareto-Optimierung für eine Szene mit sieben Objekten dargestellt. Als Kriterien für die Optimierung werden die zuvor vorgestellten Bewertungsmaße (vgl. Abschnitt 7.3.4) gewählt. Im linken Bild werden

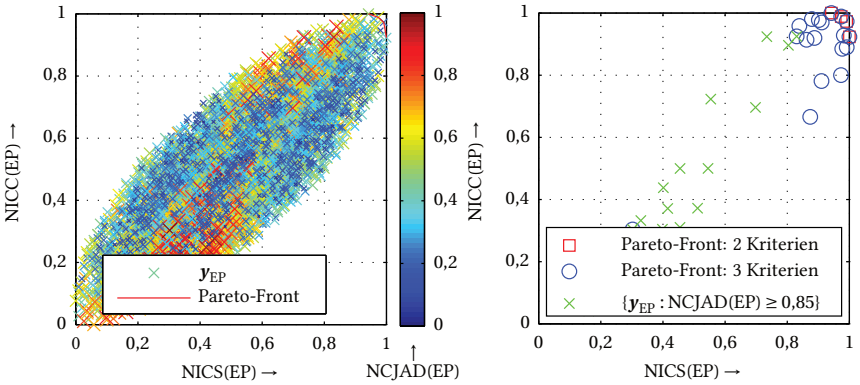

Abb. 7.7: Das Ergebnis der Pareto-Optimierung für ein weiteres Beispiel mit anderen Objekten zeigt eine andere Verteilung. Die Anzahl an möglichen Explorationspfaden ist ebenfalls sehr hoch (links). Hierbei sind jedoch wesentlich weniger Pareto-optimale Explorationspfade vorhanden, welche die drei Kriterien sehr gut erfüllen (rechts). Im linken Bild ist der NCJAD-Wert des jeweiligen Explorationspfads entsprechend der nebenstehenden Farbskala kodiert.

nur die beiden Maße NICS und NICC verwendet. Die Ergebnisse der Bewertung mit den Maßen für alle möglichen Explorationspfade einer festen Menge an gegebenen Objekten sind farbig mit „×“ gekennzeichnet. Die Farbe bzw. Größe stellt im linken Bild den Wert des Bewertungsmaßes NCJAD dar. Dabei bedeuten größere und rötlichere „×“, dass es sich um einen Explorationspfad handelt, welcher das Kriterium Bewegung gut bis sehr gut berücksichtigt, d. h. $\text{NCJAD(EP)} \in (0{,}66; 1]$. Mittelgroße und grünliche „×“ beachten entsprechend die Bewegung weniger gut, d. h. $\text{NCJAD(EP)} \in (0{,}33; 0{,}66]$. Kleine und bläuliche „×“ berücksichtigen die Bewegung kaum oder gar nicht, d. h. $\text{NCJAD(EP)} \in [0; 0{,}33]$. Die Menge aller Punkte $\mathcal{Y}$ wird durch die Pareto-Front $\mathcal{P}_{\text{Front}}$ (rote Linie) begrenzt. In diesem Beispiel sind Pfade, welche die Bewegung besser berücksichtigen, eher im Zentrum untergebracht. Dies lässt sich in Abb. 7.6 (rechts) noch deutlicher sehen. Dort sind für dieselben Objekte nur die Pareto-optimalen Explorationspfade in Bezug auf die zwei Kriterien Abnahme der Neugier und Salienz (rote Quadrate) und für alle drei Kriterien (blaue Kreise) zu sehen. Durch die Einbeziehung des Bewertungskriteriums Bewegung (NCJAD) wird dabei die Menge an Pareto-optimalen Pfaden größer. Zusätzlich sind alle Explorationspfade dargestellt, welche mindestens 85 % des Optimums bzgl. der Bewegung erfüllen (grüne Kreuze). Dies unterstreicht die Aussage bzgl. der Bewegung im linken Bild.

In Abb. 7.7 ist ein weiteres Beispiel für eine andere Szene mit ebenfalls sieben Objekten dargestellt. Dabei ist die Verteilung der Pareto-optimalen Kombination recht unterschiedlich zum ersten Beispiel. Hier existieren Pfade, bei denen in

einem sehr hohen Maß alle drei Kriterien berücksichtigt werden können. Dies lässt sich in Abb. 7.7 (rechts) direkt ablesen durch die Existenz von Explorationspfaden, welche sowohl hohe Werte für NICS und NICC besitzen und zusätzlich das Kriterium NCJAD(EP) ≥ 0,85 erfüllen. Des Weiteren ist die Pareto-Menge im Vergleich zum ersten Beispiel wesentlich kleiner, was sich auf die Form der Verteilung von $\mathcal{Y}$ zurückführen lässt.

Die Ergebnisse der beiden Beispiele verdeutlichen, dass die Wahl eines optimalen Explorationspfads stark von den Objekten in einer Szene abhängt. So existieren einerseits Konstellationen wie in Abb. 7.6, bei denen eine ideale Kombination aller Kriterien nicht möglich ist und nur einzelne Kriterien zuungunsten anderer weiter optimiert werden können. Andererseits können auch Explorationspfade wie in Abb. 7.7 existieren, welche alle drei Kriterien nahezu optimal erfüllen. Generell kann eine unabhängige Optimierung aller Kriterien für die Bestimmung eines Explorationspfads nicht vorgenommen werden.

Vor- und Nachteile

Der Vorteil der Pareto-Optimierung liegt in der Bestimmung von idealen Kombinationen der Eingangsgrößen bzw. der Auswahl von Pareto-optimalen Explorationspfaden. In den meisten Fällen besteht die Pareto-Menge aus mehreren Elementen, sodass eine Variation der Präferenzen bzgl. einer Einflussgröße weiterhin möglich ist. Dies erfordert jedoch ein weiteres Verfahren, mit dem die Wahl einer optimalen Lösung, d. h. eines konkreten Explorationspfads, vorgenommen wird.

Die Menge aller möglichen Explorationspfade $\mathcal{P}_\mathcal{O}$ dient bei der Pareto-Optimierung als Eingangsgröße. Diese steigt jedoch mit der Anzahl an vorhandenen Objekten N in einer Szene sehr stark (d. h. $|\mathcal{P}_\mathcal{O}| = N!$) an. Jeder Explorationspfad besitzt potentiell eine unterschiedliche Optimalität bzgl. der Kriterien Salienz, Neugier und Bewegung, sodass eine scheinbar nur leichte Variation zwischen beispielsweise EP_A und EP_B zu einem völlig anderen Ergebnis führen kann. Die Pareto-Optimierung erfordert daher die Berücksichtigung aller Eingangsgrößen zur Bestimmung der Pareto-Front bzw. Pareto-Menge. Diese lässt sich theoretisch komplett bestimmen, jedoch ist dies aufgrund der hohen Anzahl an möglichen Explorationspfaden selbst bei einer moderaten Zahl an Objekten innerhalb eines kurzen Zeitfensters (d. h. binnen weniger Sekunden) nicht möglich.

Lösungsansatz für die interessengetriebene Explorationsstrategie

Ein anderer Lösungsansatz für die zuvor beschriebene Problematik ist die sogenannte Skalarisierung, bei der die Bewertungskriterien zu einer einzigen Größe

für die Optimierung zusammengefasst werden. Für die interessengetriebene Explorationsstrategie (vgl. Abschnitt 5.4.2) in Kombination mit den Bewertungskriterien NICS, NICC und NCJAD bedeutet das, dass die Kriterien einzeln gewichtet und anschließend als gemeinsame Größe optimiert werden (vgl. Abschnitt 7.4.3). Dabei können die optimalen Einflussparameter γ_s^*, γ_η^* und γ_d^* für den letztendlich ausgewählten interessengetriebenen Explorationspfad bestimmt werden. Dazu ist es notwendig, eine Vielzahl an Explorationspfaden effektiv zu bestimmen.

Die Bestimmung des interessengetriebenen Explorationspfads ist ein zum nichtmetrischen asymmetrischen TSP analoges Problem (vgl. Abschnitt 5.4.3), welches NP-vollständig ist und eine Komplexität von $O(N!)$ bei der exakten Bestimmung besitzt. In Abschnitt 5.4.3 wurden verschiedene Heuristiken vorgestellt, welche im nachfolgenden Abschnitt evaluiert werden und eine wesentlich effektivere Bestimmung ermöglichen.

7.4.2 Lösung des Rundreiseproblems

Durch die Lösung des Rundreiseproblems kann neben dem interessengetriebenen auch der für den späteren Vergleich benötigte bewegungsorientierte Explorationspfad bestimmt werden. Aufgrund der hohen Komplexität ist dabei eine exakte Bestimmung eines optimalen Pfads meist nicht möglich. Es existiert jedoch eine Vielzahl an Heuristiken zur Approximation des Problems (vgl. Abschnitt 5.4.3). Im Rahmen der vorliegenden Arbeit werden fünf verschiedene Verfahren untersucht und dabei speziell an die Bedingungen der Exploration mit einer festen Ausgangsposition und einer beliebigen Endposition angepasst, da die meisten Realisierungen dies nicht explizit ermöglichen. Dadurch kann der Pfad nochmals optimiert werden, da die Rückkehr zur Ausgangsposition nicht notwendig ist. Des Weiteren wird die benötigte Zeit zur Bestimmung des Pfads berücksichtigt, sodass die Explorationspfade in einem autonomen System sehr schnell zur Laufzeit berechnet werden können.

Für die Evaluation wurde die Nearest-Neighbor-Heuristik (kurz: NN), die 2-opt-Heuristik mit NN (vgl. [Joh97]), eine Variante des Ameisenalgorithmus (vgl. [Dor96]) sowie ein angepasster genetischer Algorithmus (vgl. [Kir11]) untersucht. Des Weiteren wurde die exakte Lösung (Brute-Force-Methode) berechnet, falls die Anzahl an Objekten klein genug war, um den Explorationspfad in wenigen Sekunden berechnen zu können.

In Abb. 7.8 sind die Ergebnisse für die zuvor vorgestellten Heuristiken zusammengefasst. Dabei wurde der Simulations- und Realdatensatz als eine gemeinsame Datenbasis verwendet und in zwei Gruppen unterteilt. In der ersten Gruppe kann aufgrund der geringen Anzahl an Objekten das Problem innerhalb eines kurzen Zeitfensters (d. h. innerhalb von 10 Sekunden) exakt gelöst werden. In

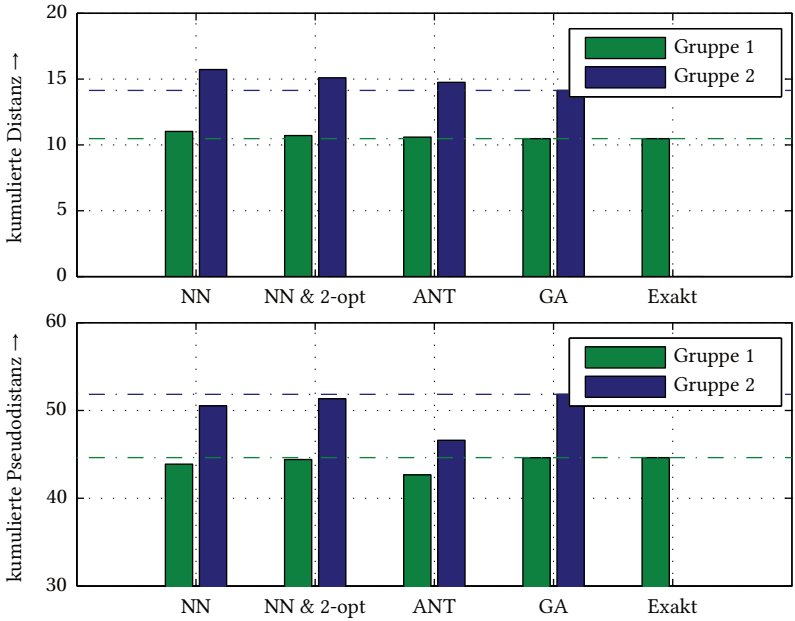

Abb. 7.8: Die Diagramme zeigen die Ergebnisse für das STSP (oben) bzw. das nicht-metrische ATSP (unten). Diese sind jeweils unterteilt in den Fall, dass das Problem exakt berechnet werden kann, innerhalb eines vorgegebenen Zeitfensters (Gruppe 1) und falls aufgrund der höheren Anzahl an Objekten nur eine Approximation möglich ist (Gruppe 2), kombiniert für den Simulations- und Realdatensatz. Die optimalen Distanzen für die jeweilige Gruppe sind durch die entsprechend farblich gekennzeichneten Linien dargestellt. Beim ATSP wurden die Kanten durch zusätzliche Gewichte der Knoten bewertet (vgl. Abschnitt 5.4.3), sodass keine Vergleichbarkeit mit dem STSP gegeben ist. NN: Nearest-Neighbor-Heuristik; NN & 2-opt: 2-opt-Heuristik auf Basis von NN; ANT: Ameisenalgorithmus; GA: genetischer Algorithmus; Exakt: Brute-Force-Methode.

der zweiten Gruppe werden alle Fälle mit mehr als 10 Objekten betrachtet, bei denen die Verwendung einer Heuristik zwingend erforderlich ist. Im oberen Diagramm (vgl. Abb. 7.8; oben) sind die Ergebnisse für das metrische STSP, wie es bei der bewegungsoptimierten Explorationsstrategie der Fall ist, dargestellt. Die kumulierte zurückgelegte Distanz dient als Vergleich. Es ist zu sehen, dass für Gruppe 1 bei allen Verfahren die Bestimmung eines guten Pfads etwas einfacher ist als bei Gruppe 2, d. h., der Abstand zur unteren Linie ist jeweils vergleichsweise geringer als zur korrespondierenden oberen Linie. Das untere Diagramm (vgl. Abb. 7.8; unten) zeigt die Ergebnisse für das nicht-metrische ATSP, welches für die interessengetriebene Explorationsstrategie Anwendung findet. Mit Ausnahme des Ameisenalgorithmus (ANT) sind vergleichbare Aussagen wie beim oberen Diagramm zu treffen. Bei der interessengetriebenen Explorationsstrate-

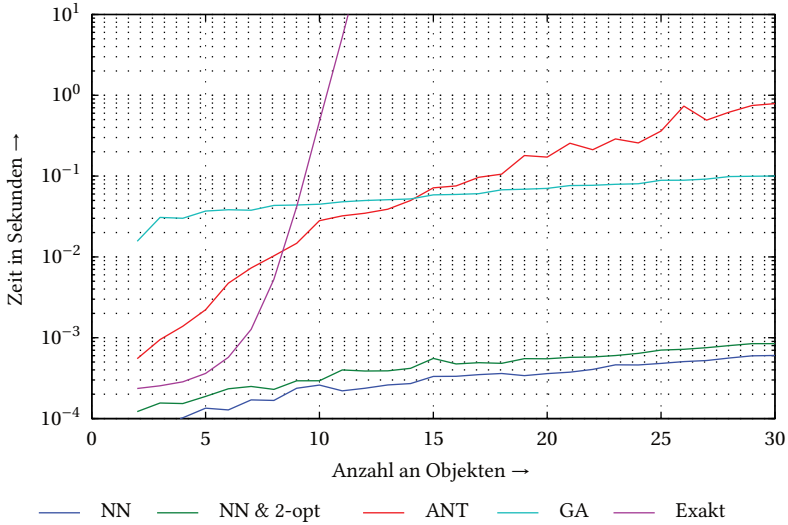

Abb. 7.9: Laufzeitvergleich der Algorithmen in Abhängigkeit von der Anzahl an Objekten. NN: Nearest-Neighbor-Heuristik; NN & 2-opt: 2-opt-Heuristik auf Basis von NN; ANT: Ameisenalgorithmus; GA: genetischer Algorithmus; Exakt: Brute-Force-Methode.

gie entspricht die dargestellte Pseudodistanz keiner realen Distanz mehr, da die Kantengewichte durch Einbeziehung der Salienz und der Neugier verändert werden (vgl. Abschnitt 5.4.3). Außerdem ist nun die maximale Distanz aufgrund der Definition des konkreten Optimierungsproblems optimal.

Der genetische Algorithmus und die 2-opt-Heuristik liefern in allen Fällen sehr gute Ergebnisse für dieses modifizierte Rundreiseproblem (d. h. feste Ausgangsposition und beliebige Endposition), wobei der genetische Algorithmus etwas bessere Ergebnisse hervorbringt, dafür jedoch mehr Zeit bei der Berechnung benötigt (vgl. Abb. 7.9). Der Ameisenalgorithmus liefert im Falle des nichtmetrischen ATSP im Rahmen der vorliegenden Arbeit schlechtere Ergebnisse. Dies ist begründet durch die Einführung von zusätzlichen objektreihenfolgeabhängigen Gewichtungen für die Salienz und die Neugier (vgl. Gl. 5.15, Gl. 5.17) beim interessengetriebenen Explorationspfad. Insgesamt ist die benötigte Rechenzeit für alle Verfahren mit Ausnahme der Brute-Force-Methode (ab einer Anzahl von 10 Objekten) vergleichsweise gering und steigt mit Ausnahme der Brute-Force-Methode und des Ameisenalgorithmus auch nur moderat mit der Objektanzahl an.

7.4.3 Interessengetriebener Explorationspfad

Nachdem im vorherigen Abschnitt die Heuristiken zur effektiven Lösung des Rundreiseproblems gegenüber gestellt wurden, erfolgt nun die Betrachtung des interessengetriebenen Explorationspfads. Dabei wird der genetische Algorithmus (vgl. [Kir11]) zur effektiven Bestimmung eines Pfads verwendet. Die Skalarisierung des Optimierungsproblems erfolgt mit Hilfe der normierten Bewertungsmaße (vgl. Abschnitt 7.3) und führt zur Bestimmung der optimalen Einflussparameter γ_s^*, γ_η^* und γ_d^* für den interessengetriebenen Explorationspfad. Die Datensätze aus Abschnitt 7.2 dienen als Grundlage für die Evaluation.

Bewertungsmaße

In Abb. 7.10 sind die Ergebnisse der normierten Bewertungsmaße (vgl. Abschnitt 7.3.4) in Abhängigkeit von den Einflussparametern γ_s, γ_η und γ_d für die interessengetriebene Explorationsstrategie (vgl. Abschnitt 5.4.2) dargestellt. Aufgrund des Wertebereichs von γ_s, γ_η und γ_d (d. h. $[0,1]$) sowie der Nebenbedingung $\gamma_s + \gamma_\eta + \gamma_d = 1$, ergibt sich Einschränkung auf eine Ebene in Form eines Dreiecks. Dabei ist das gemittelte Ergebnis für den Simulationsdatensatz (links) und den gesamten Realdatensatz (rechts) angegeben. Wie zu erkennen ist, liegt das Optimum bzgl. der Salienz (oben) bei einem sehr hohen γ_s-Wert und niedrigen γ_η- bzw. γ_d-Werten. Für die wissensbasierte Neugier verhält sich dies für hohe γ_η-Werte und niedrige γ_s- bzw. γ_d-Werte (Mitte). Die Bewegung nähert sich ihrem Optimum an, je geringer die γ_s- und γ_η-Werte werden und desto höher die γ_d-Werte werden (unten). Dieses Verhalten lässt sich auch aus Gl. 5.20 und der Definition der normierten Bewertungsmaße direkt ableiten. In Abb. 7.11 sind die dazu korrespondierenden Standardabweichungen für die jeweiligen Datensätze zu sehen. Die Abweichungen sind für alle Bewertungsmaße und Parameterkombinationen recht gering und zeigen eine deutliche Abnahme in Richtung der optimalen Kombinationen der Einflussparameter.

Bestimmung der optimalen Einflussparameter

Der Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}$ der interessengetriebenen Explorationsstrategie besitzt die drei Parameter γ_s, γ_η und γ_d, welche indirekt Einfluss auf die Berücksichtigung der Salienz und der Neugier und auch die benötigte Bewegung zur Fokussierung von Objekten nehmen. Wie in Abschnitt 7.4.1 im Rahmen der Pareto-Optimierung angesprochen wurde, kann ein Optimierungsproblem mit mehreren abhängigen Eingangsgrößen mit Hilfe einer Skalarisierungsfunktion gelöst werden.

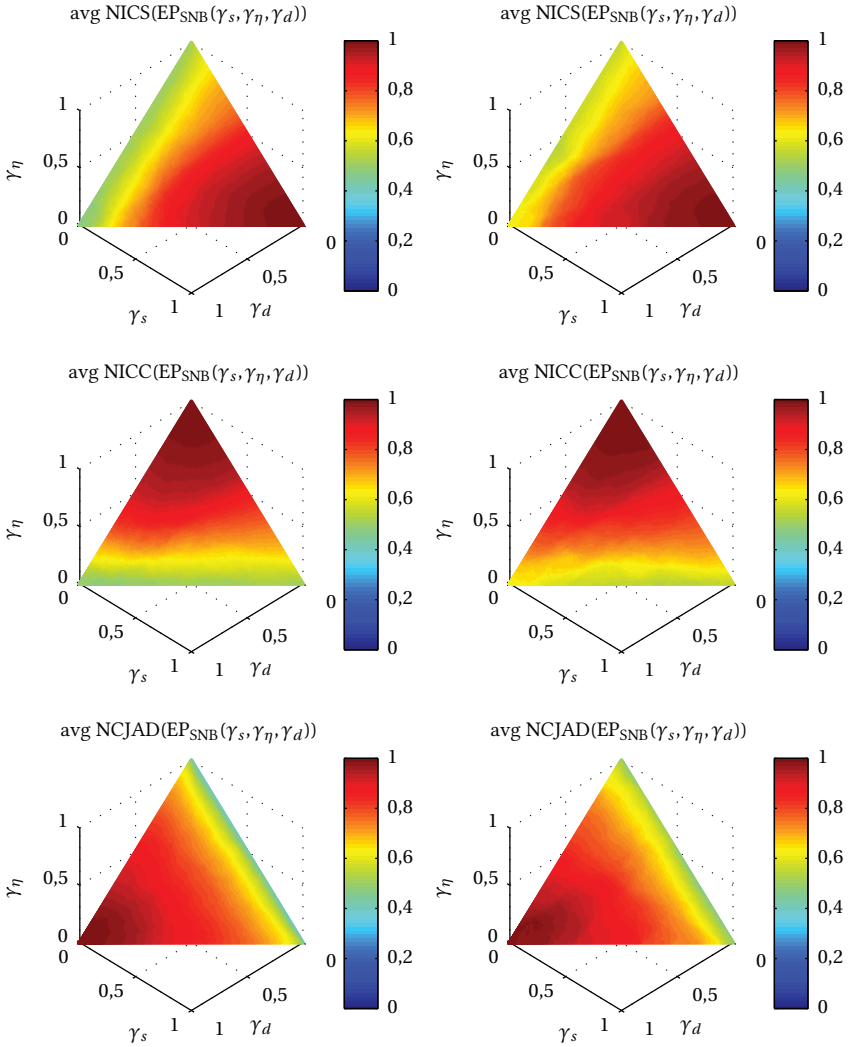

Abb. 7.10: Ergebnisse der normierten Bewertungsmaße für den interessengetriebenen Explorationspfad EP$_{SNB}$ und den Simulationsdatensatz (links) bzw. den Realdatensatz (rechts) – gemittelt über alle Szenen.

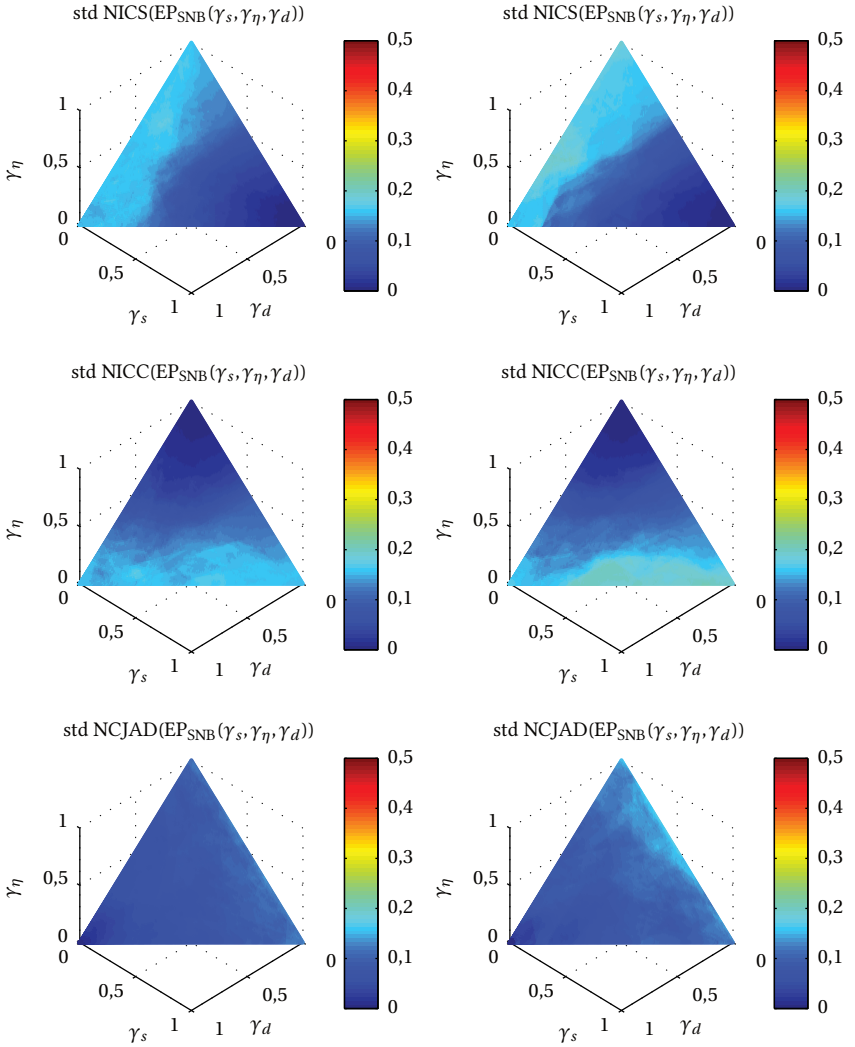

Abb. 7.11: Standardabweichung der Ergebnisse der normierten Bewertungsmaße für den interessengetriebenen Explorationspfad EP_{SNB} und den Simulationsdatensatz (links) bzw. den Realdatensatz (rechts).

Die Skalarisierungsfunktion wird in der vorliegenden Arbeit als sogenannter Parameterauswahlbereich (engl.: parameter selection space; kurz: PSS) bezeichnet und dient im weiteren Verlauf zur Bestimmung der idealen Einflussparameter γ_s^*, γ_η^* und γ_d^* für den interessengetriebenen Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}$. Der Auswahlbereich ist unter Zuhilfenahme der Skalarisierungsfaktoren λ_{NICS}, λ_{NICC} und λ_{NCJAD} wie folgt definiert:

$$\begin{aligned}\mathrm{PSS}_{\mathrm{SNB}}(\gamma_s,\gamma_\eta,\gamma_d;\lambda_{\mathrm{NICS}},\lambda_{\mathrm{NICC}},\lambda_{\mathrm{NCJAD}}) = {}& \lambda_{\mathrm{NICS}}\cdot\mathrm{NICS}(\mathrm{EP}_{\mathrm{SNB}}(\gamma_s,\gamma_\eta,\gamma_d))+\\ & \lambda_{\mathrm{NICC}}\cdot\mathrm{NICC}(\mathrm{EP}_{\mathrm{SNB}}(\gamma_s,\gamma_\eta,\gamma_d))+\\ & \lambda_{\mathrm{NCJAD}}\cdot\mathrm{NCJAD}(\mathrm{EP}_{\mathrm{SNB}}(\gamma_s,\gamma_\eta,\gamma_d))\end{aligned} \tag{7.21}$$

mit

$$\lambda_{\mathrm{NICS}},\lambda_{\mathrm{NICC}},\lambda_{\mathrm{NCJAD}}\in[0,1] \quad \text{und} \quad \lambda_{\mathrm{NICS}}+\lambda_{\mathrm{NICC}}+\lambda_{\mathrm{NCJAD}}=1\,. \tag{7.22}$$

Die optimalen Parameter von $\mathrm{EP}_{\mathrm{SNB}}(\gamma_s^*,\gamma_\eta^*,\gamma_d^*)$ werden mit den Skalarisierungsfaktoren λ_{NICS}, λ_{NICC} und λ_{NCJAD} zur direkten Berücksichtigung des anteiligen Einflusses von Salienz, Neugier und Bewegung und durch eine Maximumsuche im Parameterauswahlbereich $\mathrm{PSS}_{\mathrm{SNB}}$ bestimmt:

$$(\gamma_s^*,\gamma_\eta^*,\gamma_d^*) = \underset{\gamma_s,\gamma_\eta,\gamma_d}{\arg\max}\left\{\mathrm{PSS}_{\mathrm{SNB}}(\gamma_s,\gamma_\eta,\gamma_d;\lambda_{\mathrm{NICS}},\lambda_{\mathrm{NICC}},\lambda_{\mathrm{NCJAD}})\right\}. \tag{7.23}$$

Die verwendeten Bewertungsmaße NICS bzw. NICC konvergieren für den interessengetriebenen Explorationspfad (vgl. Gl. 5.20) bei $\gamma_s\to 1$, $\gamma_\eta\to 0$, $\gamma_d\to 0$ bzw. $\gamma_s\to 0$, $\gamma_\eta\to 1$, $\gamma_d\to 0$ gegen ihr Optimum – für NCJAD hingegen liegt das Optimum bei $\gamma_s\to 0$, $\gamma_\eta\to 0$, $\gamma_d\to 1$. Aufgrund dieser Eigenschaften kann der interessengetriebene Explorationspfad an die unterschiedlichen Gegebenheiten während der Exploration ideal angepasst werden und gegebenenfalls einzelne Priorisierungskriterien bei entsprechender Parametrierung ausgeblendet werden.

Zur Veranschaulichung der Optimierung sind in Abb. 7.12 die Ergebnisse für verschiedene Kombinationen von λ_{NICS}, λ_{NICC} und λ_{NCJAD} für den Simulationsdatensatz (links) und den Realdatensatz (rechts) dargestellt. Die einzelnen Wertebereiche $\Lambda_1,\ldots,\Lambda_{10}$ der Skalarisierungsfaktoren sind in Tabelle 7.3 zusammengefasst. Bei der Variation der Faktoren wird eine gleichmäßige Zu- bzw. Abnahme der Werte vorgenommen, sodass entsprechend Gl. 7.22 gilt. In Abb. 7.12 ist deutlich zu sehen, dass die optimalen Einflussparameter γ_s^*, γ_η^* und γ_d^* entsprechend des jeweiligen Wertebereichs Λ der Skalarisierungsfaktoren variieren und somit das Ziel der Optimierung erfüllt ist.

In Abb. 7.13 sind die gemittelten Ergebnisse bei einer Gleichgewichtung der Skalarisierungsfaktoren ($\lambda_{\mathrm{NICS}}=\lambda_{\mathrm{NICC}}=\lambda_{\mathrm{NCJAD}}=1/3$) sowohl für den Simulationsdatensatz (links) als auch für den Realdatensatz (rechts) dargestellt. Die optimalen

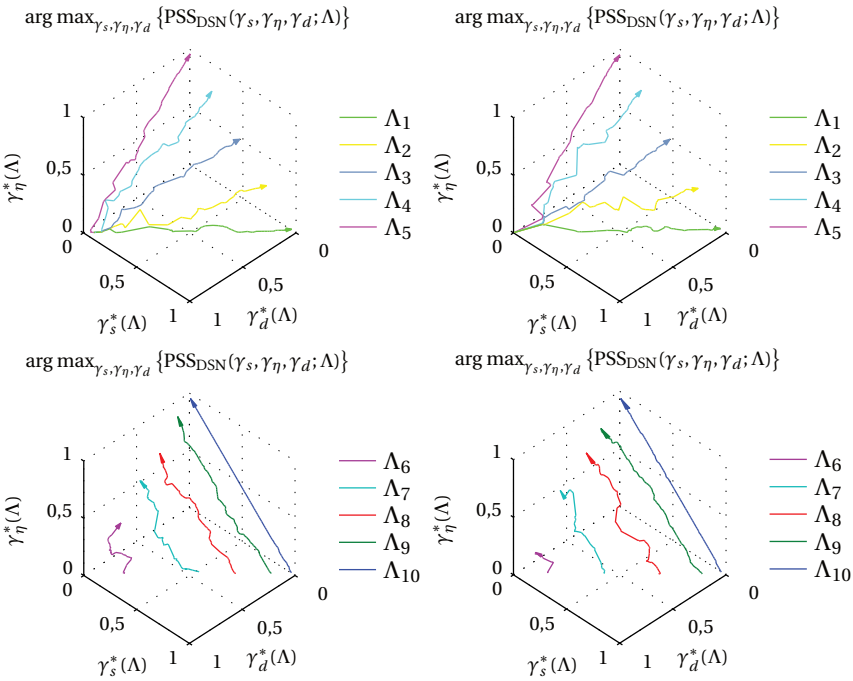

Abb. 7.12: Ergebnisse der Variation der Skalarisierungsfaktoren zur Bestimmung der idealen Einflussparameter γ_s^*, γ_η^* und γ_d^* für den interessengetriebenen Explorationspfad EP_{SNB}, getrennt nach Simulationsdatensatz (links) und Realdatensatz (rechts). Die Wertebereiche $\Lambda_1, \ldots, \Lambda_{10}$ für die Variation sind in Tabelle 7.3 angeben. Diese werden in den Grafiken in Richtung der Pfeilspitze durchlaufen.

	→ Λ_1	→ Λ_2	→ Λ_3	→ Λ_4	→ Λ_5
λ_{NICS}	0,00…1,00	0,00…0,75	0,00…0,50	0,00…0,25	0,00
λ_{NICC}	0,00	0,00…0,25	0,00…0,50	0,00…0,75	0,00…1,00
λ_{NCJAD}	1,00…0,00	1,00…0,00	1,00…0,00	1,00…0,00	1,00…0,00
	→ Λ_6	→ Λ_7	→ Λ_8	→ Λ_9	→ Λ_{10}
λ_{NICS}	0,25…0,00	0,50…0,00	0,75…0,00	0,90…0,00	1,00…0,00
λ_{NICC}	0,00…0,25	0,00…0,50	0,00…0,75	0,00…0,90	0,00…1,00
λ_{NCJAD}	0,75	0,50	0,25	0,10	0,00

Tabelle 7.3: Übersicht der Wertebereiche $\Lambda_1, \ldots, \Lambda_{10}$ für die Variation der Skalarisierungsfaktoren λ_{NICS}, λ_{NICC} und λ_{NCJAD} zur Bestimmung der idealen Einflussparameter γ_s^*, γ_η^* und γ_d^*.

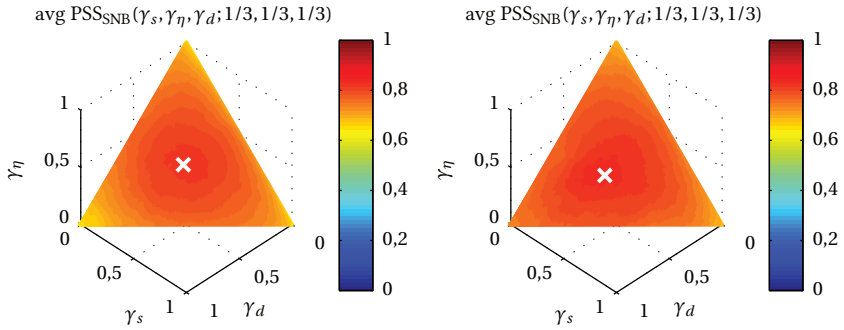

Abb. 7.13: Bestimmung der idealen Einflussparameter $(\gamma_s^*, \gamma_\eta^*, \gamma_d^*)$ = (0,32, 0,33, 0,35) für den Simulationsdatensatz (links) und $(\gamma_s^*, \gamma_\eta^*, \gamma_d^*)$ = (0,31, 0,27, 0,42) für den Realdatensatz (rechts) im Parameterauswahlbereich PSS$_{SNB}$.

EP	NICS(EP)	NICC(EP)	NCJAD(EP)
EP$_{Referenz}$	0,593 (−0,258)	0,495 (−0,296)	0,530 (−0,333)
EP$_{Salienz}$	1,000 (0,149)	0,536 (−0,256)	0,487 (−0,376)
EP$_{Neugier}$	0,557 (−0,293)	1,000 (0,209)	0,460 (−0,403)
EP$_{Bewegung}$	0,615 (−0,236)	0,619 (−0,172)	1,000 (0,137)
EP$_{SNB}(\gamma_s^*, \gamma_\eta^*, \gamma_d^*)$	0,851	0,791	0,863

Tabelle 7.4: Vergleich der Explorationsstrategien mit Hilfe der normierten Bewertungsmaße für den Realdatensatz. Die Abweichung der Maße zwischen den einfachen Pfaden und dem interessengetriebenen Explorationspfad ist in Klammern angegeben.

Einflussparameter γ_s^*, γ_η^* und γ_d^* lassen sich wie zuvor über eine Maximumsuche im Parameterauswahlbereich (vgl. Gl. 7.23) bestimmen. Für den Realdatensatz ergeben sich vergleichbare Ergebnisse bzgl. der Parametrierung von γ_s^*, γ_η^* und γ_d^* wie beim Simulationsdatensatz.

In Tabelle 7.4 sind die Ergebnisse der Bewertungsmaße für alle Explorationsstrategien (vgl. Abschnitt 5.4) und den Realdatensatz im Vergleich zusammengefasst. Dabei werden die zuvor bestimmten Werte für γ_s^*, γ_η^* und γ_d^* (Gleichgewichtung des Einflusses von Salienz, Neugier und Bewegung) für den interessengetriebenen Explorationspfad verwendet. Dabei zeigt sich gerade für diesen Explorationspfad, dass alle Ergebnisse der Bewertungsmaße im Vergleich zu den anderen Strategien insgesamt höher sind. Alle anderen Pfade erfüllen maximal eines der Kriterien sehr gut, wohingegen die anderen Kriterien vergleichsweise schlecht erfüllt werden (vgl. Tabelle 7.4). Für den salienzbasierten Explorationspfad sind dies die Bewertungsmaße NICC bzw. NCJAD, für den neugierbasierten Pfad NICS bzw. NCJAD und für den bewegungsoptimierten Pfad NICS bzw. NICC. Der Referenzpfad weist insgesamt geringere Werte für alle Bewertungsmaße auf.

7.5 Evaluation der Explorationsstrategien

Im Rahmen der Evaluation wird überprüft, ob die zuvor definierten Bewertungsmaße (vgl. Abschnitt 7.3) den Einfluss auf die Priorität der Objekte während der Exploration korrekt widerspiegeln und somit die Ergebnisse aus den vorangegangenen Abschnitten bestätigen. Hierzu ist es notwendig, ein Maß zu definieren, welches zwei Explorationspfade unabhängig von den verwendeten Strategien der einzelnen Pfade vergleicht. Ziel ist es, zu zeigen, dass die interessengetriebene Explorationsstrategie abhängig von der Parametrierung der Einflussparameter γ_s, γ_η und γ_d, ein adäquates Ergebnis für den Einfluss von Salienz, Neugier und Bewegung liefert. Abschließend werden anhand eines Beispiels die im Rahmen der vorliegenden Arbeit vorgestellten Explorationsstrategien für eine konkrete Szene überprüft.

7.5.1 Indizesbasiertes Korrelationsmaß für Explorationspfade

Die interessengetriebene Explorationsstrategie (vgl. Abschnitt 5.4.2) kann mit Hilfe der Einflussparameter γ_s, γ_η und γ_d angepasst werden. Dabei wird der Einfluss von Salienz und Neugier sowie der benötigten Bewegung zur Fokussierung von Objekten variiert. Im Rahmen der Evaluation wird die interessengetriebene Explorationsstrategie mit weiteren zuvor definierten Explorationsstrategien (vgl. Abschnitt 5.4.1) durch ein unabhängiges Maß verglichen. Die Priorität der Objekte während der Exploration, d. h. die Reihenfolge, in der die Objekte selektiert und analysiert werden, dient als Maß für die Übereinstimmung zwischen zwei Explorationspfaden. Formal kann dieses sogenannte indizesbasierte Korrelationsmaß (engl.: indices-based correlation measure; kurz: ICM) über die kumulierten Unterschiedene der auftretenden Objektindizes der zu vergleichenden Pfade definiert werden:

$$\mathrm{ICM}(\mathrm{EP}_1, \mathrm{EP}_2) = 1 - \frac{1}{\iota_N} \sum_{n=1}^{N} \left| I_{\mathrm{EP}_1}(o_n) - I_{\mathrm{EP}_2}(o_n) \right| \quad \text{mit } N \in \mathbb{N}. \tag{7.24}$$

Dabei stellt $I_{\mathrm{EP}}(\cdot)$ eine Funktion dar, welche die Position eines Objekts auf dem aktuellen Explorationspfad EP zurückliefert. o_n ist das n-te Objekt aller bei der Exploration berücksichtigten Objekte N. Die Basis beider Explorationspfade EP_1 und EP_2 bilden jeweils dieselben Objektrepräsentanten im Umweltmodell.

Das indizesbasierte Korrelationsmaß $\mathrm{ICM}(\mathrm{EP}_1, \mathrm{EP}_2)$ bewertet die unterschiedlichen Prioritäten der Objekte innerhalb zweier Explorationspfade. Falls die Objekte in nahezu derselben Reihenfolge untersucht werden, liefert das Maß einen hohen Wert und somit eine große Übereinstimmung, andernfalls einen erheblich kleineren Wert und somit eine geringere Übereinstimmung bzw. eine höhere Abweichung. Für die Vergleichbarkeit der Ergebnisse verschiedener Szenen

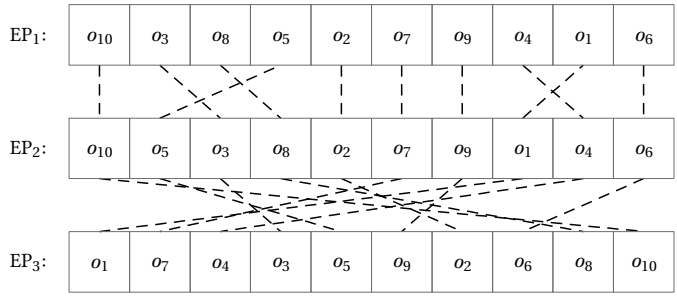

Abb. 7.14: Beispiel für den indizesbasierten Vergleich von verschiedenen Explorationspfaden

mit möglicherweise einer unterschiedlichen Anzahl an Objekten wird der Normalisierungsfaktor $\iota_N = \lfloor 0{,}5 \cdot N^2 \rfloor$ (vgl. [Wil14]) definiert, welcher der maximalen kumulierten Abweichung zwischen zwei Explorationspfaden mit N Objekten entspricht. Somit definiert das indizesbasierte Korrelationsmaß aus Gl. 7.24 die normierte kumulierte Übereinstimmung der Prioritäten der Objekte eines Pfads im Vergleich zu einem anderen Pfad.

In Abb. 7.14 ist ein Beispiel für drei verschiedene Explorationspfade mit denselben Objekten $o_1, \ldots, o_{10}$ angegeben, welche durch unterschiedliche Explorationsstrategien erzeugt werden. Ein Vergleich von EP_1 mit EP_2 zeigt eine relativ große Übereinstimmung bzgl. der Priorität der Objekte:

$$\mathrm{ICM}(\mathrm{EP}_1, \mathrm{EP}_2) = 1 - \frac{1}{50} \cdot (1+0+1+1+2+0+0+1+0+0) = 0{,}88\,.$$

Wird hingegen der zweite mit dem dritten Explorationspfad verglichen, so sind die Prioritäten beider Explorationsstrategien sehr unterschiedlich. Entsprechend bewertet das indizesbasierte Korrelationsmaß die Diskrepanz zwischen den beiden Explorationspfaden mit einem vielfach geringeren Wert:

$$\mathrm{ICM}(\mathrm{EP}_2, \mathrm{EP}_3) = 1 - \frac{1}{50} \cdot (7+2+1+6+3+2+4+5+1+9) = 0{,}20\,.$$

Im Nachfolgenden wird der interessengetriebene Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}$ in Abhängigkeit von den Einflussparametern γ_s, γ_η und γ_d mit den Explorationspfaden für Salienz, Neugier und Bewegung, d. h. $\mathrm{EP}_{\mathrm{Salienz}}$, $\mathrm{EP}_{\mathrm{Neugier}}$ und $\mathrm{EP}_{\mathrm{Bewegung}}$, verglichen. Diese Gegenüberstellung wird sowohl für den Simulationsdatensatz als auch für den Realdatensatz getrennt durchgeführt. Die gemittelten Ergebnisse für die beiden Datensätze sind in Abb. 7.15 dargestellt.

Wie deutlich zu sehen ist, nimmt die durchschnittliche Übereinstimmung zwischen den Explorationspfaden zu, je höher die Gewichtung des jeweiligen Ein-

flusses beim interessengetriebenen Explorationspfad ist: Der Vergleich des interessengetriebenen Explorationspfads $\mathrm{EP}_{\mathrm{SNB}}(\gamma_s, \gamma_\eta, \gamma_d)$ mit dem rein salienzbasierten Pfad $\mathrm{EP}_{\mathrm{Salienz}}$ (vgl. Abb. 7.15; oben) zeigt einen optimalen Bereich bei hohen γ_s- und niedrigen γ_η- bzw. γ_d-Werten. Für den Vergleich mit dem neugierbasierten Pfad $\mathrm{EP}_{\mathrm{Neugier}}$ gilt dies entsprechend für niedrige γ_s- bzw. γ_s- und hohe γ_η-Werte (vgl. Abb. 7.15; Mitte). Der bewegungsoptimierte Ansatz $\mathrm{EP}_{\mathrm{Bewegung}}$ weist optimale Übereinstimmungen mit dem interessengetriebenen Explorationspfad für hohe γ_d- und niedrige γ_s- bzw. γ_η-Werte auf (vgl. Abb. 7.15; unten). Ein Vergleich der Ergebnisse des Simulations- (links) mit dem Realdatensatz (rechts) zeigt bei den oberen vier Grafiken einen geringen und bei den unteren beiden einen etwas größeren Unterschied. Dies lässt sich u. a. durch die größere Möglichkeit der Positionsvariation im Simulationsdatensatz erklären: Im Realdatensatz sind die Objekte nur auf Tischen und Ablageflächen sowie an Wänden platziert, wohingegen bei der Simulation die Objekte frei im Raum positioniert sind.

Werden die Ergebnisse aus Abb. 7.15 den Resultaten für die zuvor definierten Bewertungsmaße und den interessengetriebenen Explorationspfad (siehe Abb. 7.10; Seite 171) gegenübergestellt, so zeigt sich ein vergleichbares Ergebnis für die Parametrierung von γ_s, γ_η und γ_d bzgl. des Optimums. Durch die unabhängige Auswertung mit dem indizesbasierten Korrelationsmaß (Gl. 7.24) kann somit gezeigt werden, dass die interessengetriebene Explorationsstrategie für jede Parametrierung von γ_s, γ_η und γ_d den zuvor bestimmten Einfluss von Salienz, Neugier und Bewegung besitzt. Dies kann zusätzlich gestützt werden durch die Tatsache, dass die Standardabweichung für das indizesbasierte Korrelationsmaß über alle Szenen des Simulations- und Realdatensatz gering ist – für die jeweiligen optimalen Parameter γ_s, γ_η und γ_d. Beim Vergleich mit der bewegungsoptimierten Exploration (vgl. Abb. 7.15, Abb. 7.16; unten) liefert das indizesbasierte Korrelationsmaß nur bei $\gamma_s = 0$, $\gamma_\eta = 0$ und $\gamma_d = 1$ ein optimales Ergebnis der Gegenüberstellung zweier Pfade. Alle anderen Kombinationen von γ_s, γ_η und γ_d können tendenziell eine geringere Übereinstimmung mit $\mathrm{EP}_{\mathrm{Bewegung}}$ aufweisen, ohne gleich eine wesentlich höhere Bewegung zu erzeugen. Diese Tatsache lässt sich durch das Vertauschen oder das Durchlaufen in umgekehrter Reihenfolge von einzelnen Abschnitten des Pfads erklären.

In Tabelle 7.5 sind die Ergebnisse für das indizesbasierte Korrelationsmaß und den Fall von ausgeglichenem Einfluss von Salienz, Neugier und Bewegung ($\lambda_{\mathrm{NICS}} = \lambda_{\mathrm{NICC}} = \lambda_{\mathrm{NCJAD}} = 1/3$) für den interessengetriebenen Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}(\gamma_s^*, \gamma_\eta^*, \gamma_d^*)$ dargestellt. Wie deutlich zu sehen ist, besitzt der Referenzexplorationspfad nur eine relativ geringe Übereinstimmung mit anderen Pfaden und somit eine andere Objektpriorisierung. Die Salienz-, Neugier- und bewegungsbasierten Explorationspfade weisen untereinander ebenfalls eine geringe Übereinstimmung auf. Dies lässt sich auch aufgrund der gegensätzlichen Optimierungsziele begründen. Dennoch ist es möglich, wie der interessengetriebene Explorationspfad zeigt, auf die einzelnen Aspekte einzugehen. Eine vollkommene

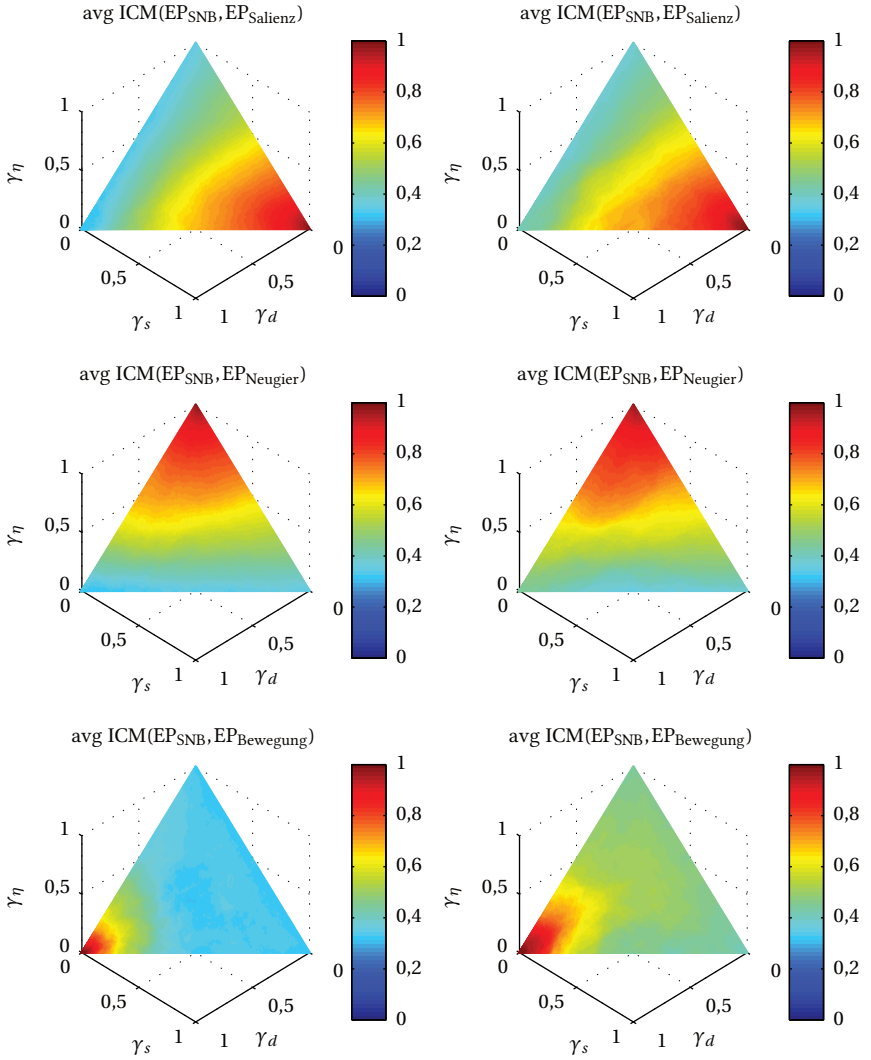

Abb. 7.15: Die Übereinstimmung der Objektreihenfolge für EP_{SNB} im Vergleich zu $EP_{Salienz}$, $EP_{Neugier}$ bzw. $EP_{Bewegung}$ in Abhängigkeit von den Parametern γ_s, γ_η und γ_d über den Simulationsdatensatz (links) und den Realdatensatz (rechts) gemittelt.

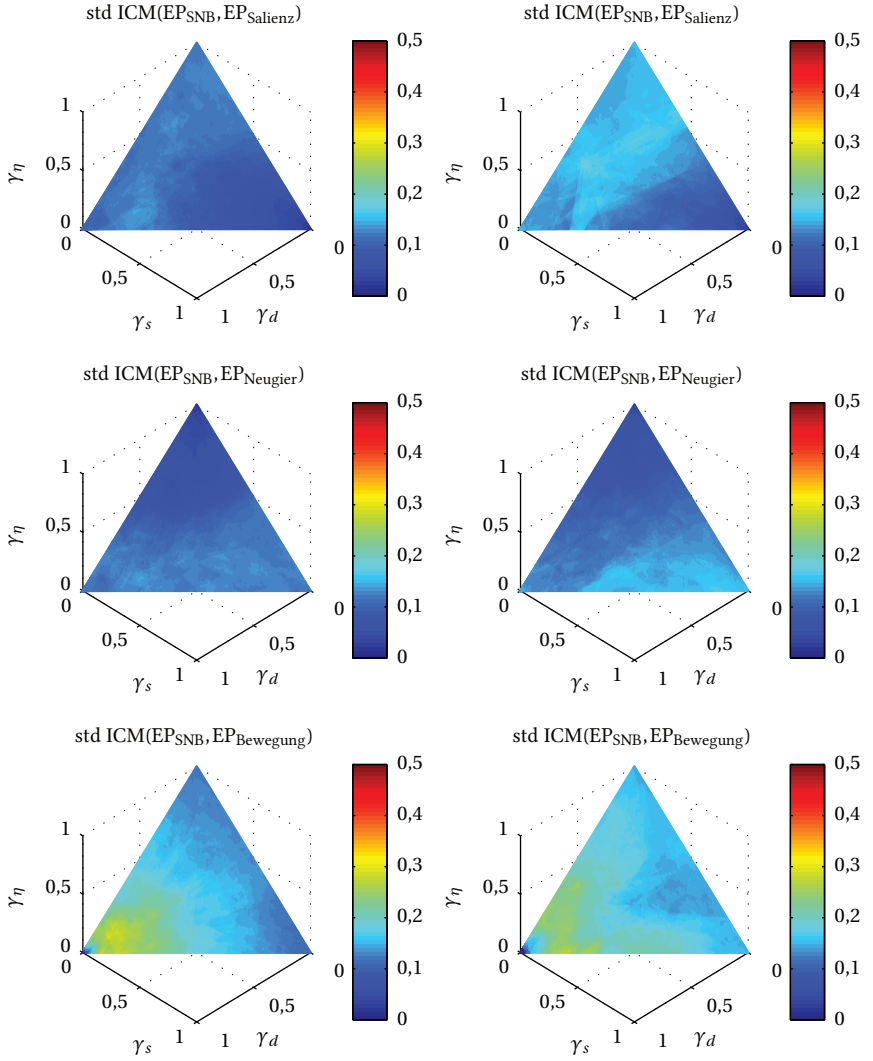

Abb. 7.16: Die Standardabweichung der mittleren Übereinstimmung in der Objektreihenfolge für EP_{SNB} im Vergleich zu $EP_{Salienz}$, $EP_{Neugier}$ bzw. $EP_{Bewegung}$ in Abhängigkeit von den Parametern γ_s, γ_η und γ_d für den Simulationsdatensatz (links) und den Realdatensatz (rechts).

$\mathrm{ICM}(\mathrm{EP}_1, \mathrm{EP}_2)$	$\mathrm{EP}_{\mathrm{Referenz}}$	$\mathrm{EP}_{\mathrm{Salienz}}$	$\mathrm{EP}_{\mathrm{Neugier}}$	$\mathrm{EP}_{\mathrm{Bewegung}}$	$\mathrm{EP}_{\mathrm{SNB}}$
$\mathrm{EP}_{\mathrm{Referenz}}$	–	0,396	0,356	0,410	0,373
$\mathrm{EP}_{\mathrm{Salienz}}$	0,396	–	0,375	0,409	0,606
$\mathrm{EP}_{\mathrm{Neugier}}$	0,356	0,375	–	0,443	0,566
$\mathrm{EP}_{\mathrm{Bewegung}}$	0,410	0,409	0,443	–	0,523
$\mathrm{EP}_{\mathrm{SNB}}(\gamma_s^*, \gamma_\eta^*, \gamma_d^*)$	0,373	0,606	0,566	0,523	–

Tabelle 7.5: Vergleich der Explorationsstrategien mit Hilfe des indizesbasierten Korrelationsmaßes für den Realdatensatz bei gleichem Einfluss von Salienz, Neugier und Bewegung für den interessengetriebenen Explorationspfad $\mathrm{EP}_{\mathrm{SNB}}(\gamma_s^*, \gamma_\eta^*, \gamma_d^*)$.

Übereinstimmung ist nur zu erreichen, wenn die Priorisierung bzgl. der Salienz und Neugier für alle Objekte gleich ist, d. h., dieselbe Reihenfolge wie der jeweils korrespondiere Vergleichspfad generiert wird. Zusätzlich muss genau diese in Bezug auf die Bewegung ebenfalls optimal sein. Dies wird in einem realen System nahezu nie der Fall sein. In diesem Zusammenhang lässt sich auch feststellen, dass das indizesbasierte Korrelationsmaß ein sehr strenges Maß ist, welches jede Abweichungen in der Objektpriorisierung bestraft, selbst wenn der Unterschied zwischen einzelnen Objekten in Bezug auf Salienz, Neugier und Bewegung teils marginal sein kann. Vor diesem Hintergrund erzielen die interessengetriebene Explorationspfade sehr gute Ergebnisse.

7.5.2 Beispiel für die unterschiedlichen Explorationsstrategien

Abschließend werden nun anhand eines konkreten Beispiels die verschiedenen Explorationsstrategien mit den dazugehörigen Pfaden $\mathrm{EP}_{\mathrm{Referenz}}$, $\mathrm{EP}_{\mathrm{Salienz}}$, $\mathrm{EP}_{\mathrm{Neugier}}$, $\mathrm{EP}_{\mathrm{Bewegung}}$ und $\mathrm{EP}_{\mathrm{SNB}}$ erläutert.

In Abb. 7.17 ist ein zusammengesetztes Bild – ein Kameraschwenk von links nach rechts – eines Frühstücksszenarios mit verschiedenen Gegenständen zu sehen. Dieses Szenario dient als Grundlage für die weiteren Erläuterungen: Mit Hilfe der in Kapitel 2 beschriebenen Verfahren lässt sich zunächst eine Salienzkarte (vgl. Abb. 7.17; Mitte) und im Anschluss daran die multimodalen Salienzcluster bestimmen. Des Weiteren werden durch das OPASCA-System (vgl. Kapitel 6) die in der Szene vorhandenen Objekte, durch die objektzentrierte Umwelterfassung, wie in Kapitel 3 beschrieben, zunächst initial erfasst und dabei die multimodalen Salienzcluster den Objektrepräsentanten zugeordnet (vgl. Abb. 7.17; unten). Die wissensbasierte Neugier wird, wie in Kapitel 4 beschrieben, für jeden Repräsentant im Umweltmodell bestimmt. Mit diesen Informationen können die Explorationspfade der verschiedenen Explorationsstrategien bestimmt werden.

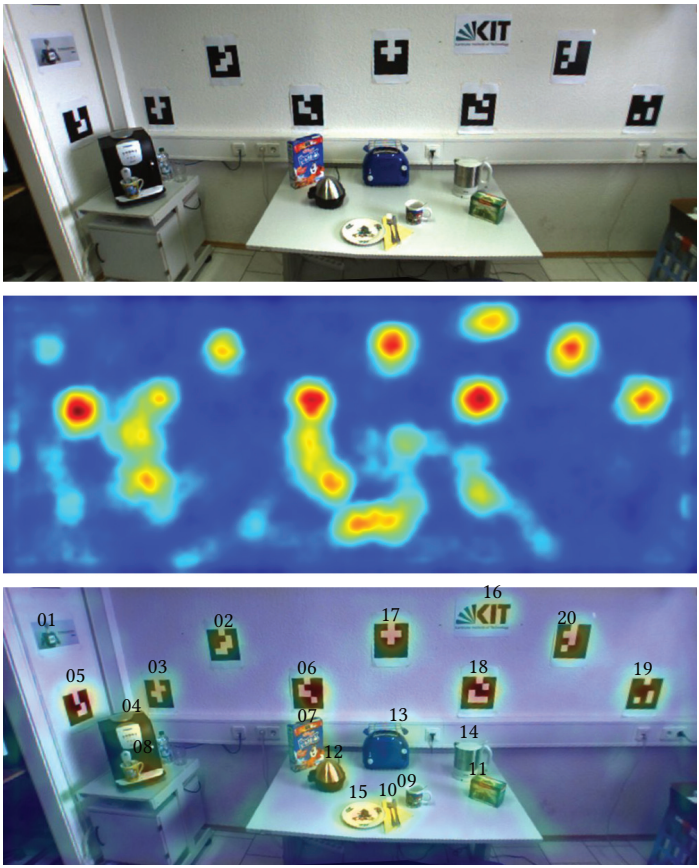

Abb. 7.17: Beispiel für die Exploration einer Szene, welches das zusammengesetzte Bild einer kompletten Szene (oben), die dazugehörige Salienzkarte (Mitte) sowie die aktuellen Objekte in der Szene (unten) darstellt

Referenzexplorationspfad

In Abb. 7.18 ist der Referenzexplorationspfad (vgl. Abschnitt 5.4.1) für die im Umweltmodell erfassten Objekte zu sehen. Die Nummern geben dabei die Priorität während der Exploration an. Als Ausgangspunkt dient die aktuelle Position der Kamera. Die Reihenfolge entspricht den Zeitpunkten der Erfassung der einzelnen Objekte, sodass ein gewisser Fluss bei der Exploration von links nach rechts entsteht. Es wird keine zusätzliche Priorisierung der Objekte bzgl. der Salienz, Neugier oder Bewegung vorgenommen.

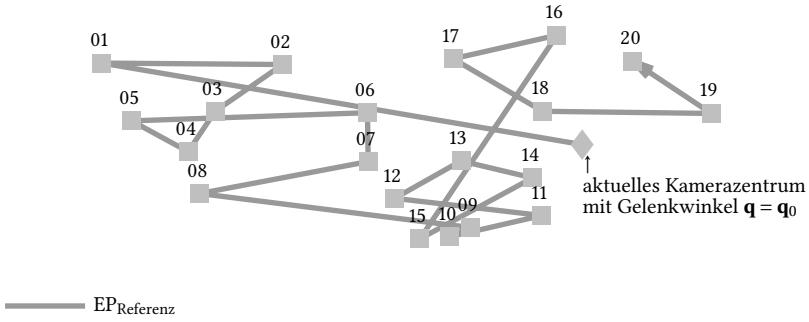

Abb. 7.18: Der Referenzexplorationspfad für die Exploration der Szene aus Abb. 7.17 zeigt die ursprüngliche Reihenfolge der Wahrnehmungen.

Salienzbasierter Explorationspfad

Die salienzbasierte Explorationsstrategie (vgl. Abschnitt 5.4.1) berücksichtigt bei der Priorisierung der Objekte ausschließlich die Salienz. Bewegungen sowie die wissensbasierte Neugier werden bei der Bestimmung des Explorationspfads nicht miteinbezogen. Wie erwartet folgt der salienzbasierte Explorationspfad (vgl. Abb. 7.19a; roter Pfad) strikt der Salienz der Objekte (vgl. Abb. 7.17; Mitte) und berücksichtigt keine anderen Priorisierungskriterien. Daher ist auch eine deutlich größere Bewegung als bei anderen Strategien zu sehen (beispielsweise Abb. 7.19c; grüner Pfad). Damit verlängert sich die Zeit für die Exploration alleine durch die größeren Bewegungen, da Objektnachbarschaften nicht ausgenutzt werden. Jedoch werden die Objekte entsprechend ihrer Salienz zeitnah und priorisiert analysiert.

Neugierbasierter Explorationspfad

Die neugierbasierte Explorationsstrategie (vgl. Abschnitt 5.4.1) nutzt für die Bestimmung des aktuellen Explorationspfads die wissensbasierte Neugier für die Objekte mit den Teilaspekten Neuartigkeit, Komplexität, Unsicherheit und Konflikt. Entsprechend der Definition (siehe Gl. 5.9) wird für den Explorationspfad immer das Objekt ausgewählt, welches die nächstkleinere Neugier induziert. In Abb. 7.19b (gelber Pfad) ist der Explorationspfad für die Objekte der Szene zu sehen. Die Priorität und somit die Reihenfolge bei der Exploration wurde aus den aktuellen Informationen im Umweltmodell bestimmt und entsprechend mit Nummern über den Objekten gekennzeichnet. Wie auch beim salienzbasierten Explorationspfad ergeben sich für den neugierbasierten Pfad dieselben Nachteile in Bezug auf die Bewegung: längere Exploration, höherer Energieverbrauch und

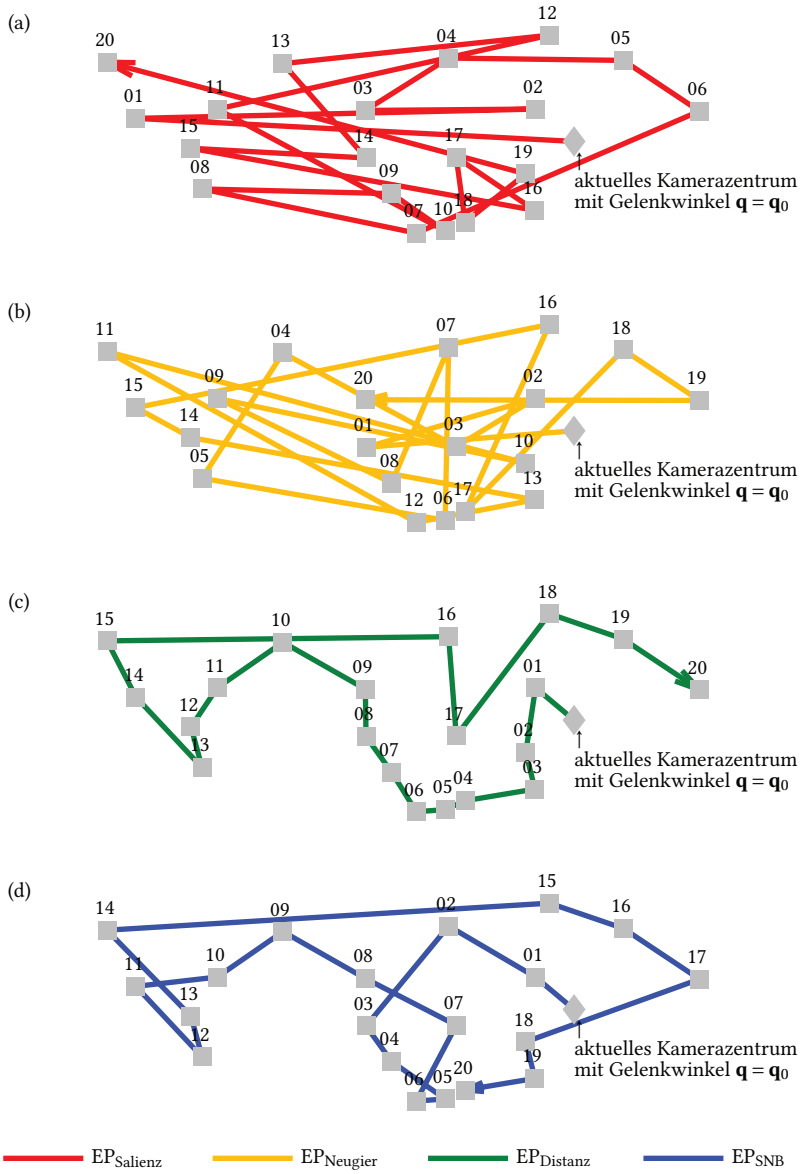

Abb. 7.19: Der salienzbasierte (a), der neugierbasierte (b), der bewegungsoptimierte (c) und der interessengetriebene Explorationspfad (d) im Vergleich für die detektierten Objekte des Explorationsbeispiels aus Abb. 7.17.

gegebenenfalls mehr Verschleiß. Zusätzlich wird auch die Salienz der Objekte nicht berücksichtigt. Entscheidende Vorteile sind jedoch die Berücksichtigung von Konflikten und Unsicherheiten sowie die Priorisierung von neuen und unbekannten Objekten durch die Neugier.

Bewegungsoptimierter Explorationspfad

Bei der bewegungsoptimierten Explorationsstrategie steht die Minimierung der Gelenkwinkelbewegungen einzig im Vordergrund (vgl. Abschnitt 5.4.1). Im Vergleich zu den meisten anderen vorgestellten Strategien lässt sich erkennen, dass die lokalen Beziehungen zwischen den Objekten sehr gut ausgenutzt werden (vgl. Abb. 7.19c; grüner Pfad), um einen optimalen Explorationspfad bzgl. der Bewegungen zu garantieren. Dabei kann die Bewegung deutlich reduziert werden, wenngleich auch die Salienz (vgl. Abb. 7.19a; roter Pfad) und die Neugier (vgl. Abb. 7.19b; gelber Pfad) bei diesem Ansatz nicht explizit berücksichtigt werden. Dies bedeutet damit auch, dass die Vorteile der salienz- und neugierbasierten Explorationsstrategie auch gleichzeitig die Nachteile dieser Strategie sind. Zusätzlich ist die benötigte Rechenzeit zur Bestimmung des optimalen Pfads – bei exakter Berechnung – wesentlich höher als bei der Referenzexplorationsstrategie oder der salienz-/neugierbasierten Explorationsstrategie, da das Problem NP-vollständig ist (vgl. Abschnitt 5.4.1). Durch eine gute Heuristik lässt sich dieser Nachteil weitestgehend ausräumen. Der Vorteil der bewegungsoptimierten Explorationsstrategie ist u. a. die durchschnittlich kürzere Zeit zum Fokussieren eines Objekts.

Interessengetriebener Explorationspfad

Die interessengetriebene Explorationsstrategie (vgl. Abschnitt 5.4.2) verfolgt im Gegensatz zu den zuvor genannten Strategien mehrere Optimierungsziele gleichzeitig. Mit Hilfe der Parameter γ_s, γ_η und γ_d (vgl. Gl. 5.20) kann der Einfluss der Salienz und der Neugier sowie die Bewegung auf den Explorationspfad explizit beeinflusst werden. Wie in Abschnitt 7.4.1 gezeigt, ist eine gleichzeitige Optimierung aller Parameter nicht immer möglich. Deshalb wurden die Optimierungsparameter λ_{NICS}, λ_{NICC} und λ_{NCJAD} zur Skalarisierung des Problems in Kombination mit den Bewertungsmaßen definiert, mit deren Hilfe ein Explorationspfad $\text{EP}_{\text{SNB}}(\gamma_s^*, \gamma_\eta^*, \gamma_d)$ bestimmt werden kann, der den gewünschten anteiligen Einfluss von Salienz, Neugier und Bewegung berücksichtigt. In Abb. 7.19d (blauer Pfad) ist der interessengetriebene Explorationspfad für $\lambda_{\text{NICS}} = \lambda_{\text{NICC}} = \lambda_{\text{NCJAD}} = 1/3$ dargestellt. Im Vergleich zum salienz- und neugierbasierten Explorationspfad ist eine deutliche Verringerung der Bewegung zu sehen. Des Weiteren werden bei der Priorisierung auch die Salienz und die Neugier berücksichtigt.

ICM(EP_1, EP_2)	$EP_{Referenz}$	$EP_{Salienz}$	$EP_{Neugier}$	$EP_{Bewegung}$	EP_{SNB}
$EP_{Referenz}$	–	0,210	0,380	0,370	0,290
$EP_{Salienz}$	0,210	–	0,310	0,300	0,510
$EP_{Neugier}$	0,380	0,310	–	0,460	0,610
$EP_{Bewegung}$	0,370	0,300	0,460	–	0,510
$EP_{SNB}(\gamma_s^*, \gamma_\eta^*, \gamma_d^*)$	0,290	0,510	0,610	0,510	–

Tabelle 7.6: Vergleich des salienz-, neugier- und bewegungsbasierten Pfads sowie des interessengetriebenen Explorationspfads untereinander mit dem indizesbasierten Korrelationsmaß für das Beispiel aus Abb. 7.19. Der Referenzexplorationspfad ist zusätzlich als Vergleich angegeben.

EP	NICS(EP)	NICC(EP)	NCJAD(EP)
$EP_{Referenz}$	0,362 (−0,377)	0,567 (−0,258)	0,742 (−0,139)
$EP_{Salienz}$	1,000 (+0,261)	0,417 (−0,408)	0,399 (−0,482)
$EP_{Neugier}$	0,436 (−0,303)	1,000 (+0,175)	0,318 (−0,563)
$EP_{Bewegung}$	0,430 (−0,309)	0,650 (−0,175)	1,000 (+0,119)
$EP_{SNB}(\gamma_s^*, \gamma_\eta^*, \gamma_d^*)$	0,739	0,825	0,881

Tabelle 7.7: Vergleich der Explorationspfade anhand der Bewertungsmaße für das Beispiel in Abb. 7.19. Die Abweichung der Maße zwischen den einfachen Pfaden und dem interessengetriebenen Explorationspfad ist in Klammern angegeben.

Vergleich der Explorationspfade

In Tabelle 7.6 ist ein Vergleich aller Explorationspfade aus Abb. 7.18 und Abb. 7.19 dargestellt. Dabei wird das indizesbasierte Korrelationsmaß ICM (vgl. Gl. 7.24) zum Vergleich der Pfade verwendet. Die Abweichung des interessengetriebenen Explorationspfads zu den anderen Pfaden ($EP_{Salienz}$, $EP_{Neugier}$ bzw. $EP_{Bewegung}$) ist bei einer Gleichgewichtung der Optimierungsziele ($\gamma_s^*, \gamma_\eta^*, \gamma_d^*$) jeweils relativ moderat. Die paarweise Übereinstimmungen zwischen $EP_{Referenz}$, $EP_{Salienz}$, $EP_{Neugier}$ und $EP_{Bewegung}$ zeigen, im Gegensatz dazu, eine viel geringere Korrelation. Wie bereits in Abschnitt 7.5.1 gesehen, ergibt sich eine insgesamt große Übereinstimmung mit den Optimierungszielen für alle Parametrierungen der interessengetriebenen Explorationsstrategie.

Eine Betrachtung der einzelnen Werte der normierten Bewertungsmaße NICS, NICC und NCJAD für alle Pfade aus Abb. 7.18 und Abb. 7.19 erfolgt in Tabelle 7.7: Der interessengetriebene Explorationspfad weist hohe Werte für alle drei Bewertungsmaße auf, wohingegen die anderen Explorationsstrategien, bis auf das jeweils optimale Bewertungsmaß (Wert: 1,0), keine höheren Werte erreichen. Insbesondere beim Maß für die minimale Bewegung (NCJAD) sind Steigerungen von über 100 % – im Vergleich zum Salienz- und Neugier-basierten Explorati-

onspfad – möglich. Durch die Verwendung des interessengetriebenen Explorationspfads mit einer moderaten Abweichung vom jeweiligen Optimum der Priorisierungskriterien lassen sich somit insgesamt deutlich höhere Werte bei allen Bewertungsmaßen erreichen.

7.6 Schlussbetrachtungen

In diesem Kapitel wurden zunächst zwei Demonstratoren vorgestellt, welche mit den vorhandenen On-Board-Sensoren eine multimodale Erfassung von Personen und Gegenständen ermöglichen. Des Weiteren wurden sowohl ein Simulationsdatensatz als auch ein Realdatensatz vorgestellt, welche die Grundlage für die Evaluation bilden und im Falle der realen Daten verschiedene Frühstücks- und Büroszenarien beinhalten. Der Realdatensatz wurde unter Zuhilfenahme der Demonstratoren in verschiedenen Räumen aufgenommen.

Für die Bewertung von Explorationspfaden wurden neue allgemeingültige Maße vorgestellt, welche ausdrücken, inwieweit ein Pfad eines der Priorisierungskriterien Salienz, Neugier bzw. Bewegung berücksichtigt. Die Maße sind dabei unabhängig von der Länge des Explorationspfads sowie den absoluten Werten der einzelnen Kriterien definiert.

Ausgehend davon wurde die generelle Abhängigkeit der Priorisierungskriterien untereinander mit Hilfe der Pareto-Optimierung untersucht und festgestellt, dass die einzelnen Kriterien sich generell nicht unabhängig voneinander optimieren lassen. Als Lösung für eine effektive Bestimmung von interessengetriebenen Explorationspfaden unter Ausnutzung der zuvor genannten Kriterien wurde eine Skalarisierung des Problems, in Kombination mit einer sehr guten Heuristik zur Lösung des nicht-metrischen asymmetrischen Rundreiseproblems, vorgestellt. Mit Hilfe der Skalarisierungsfaktoren λ_{NICS}, λ_{NICC} und λ_{NCJAD} ist die Festlegung des anteiligen Einflusses der Priorisierungskriterien über die Bewertungsmaße direkt möglich.

Durch die Einführung des indizesbasierten Korrelationsmaßes (ICM) zum Vergleich zweier Explorationspfade untereinander, unabhängig von der zugrundeliegenden Strategie, konnte die Priorisierung einzelner sowie aller Objekte auf dem Pfad gegenübergestellt werden. Dabei wurde deutlich, dass durch die Verfolgung einer interessengetriebenen Explorationsstrategie eine wesentlich bessere Priorisierung der Objekte bei der Exploration erfolgen kann.

Abschließend wurde anhand eines Beispiels veranschaulicht, dass die Optimierung des Explorationspfads mit Hilfe der interessengetriebenen Explorationsstrategie – entsprechend der Priorisierung – sehr gute Ergebnisse liefert.

Teil IV

Abschließende Betrachtungen

8

Zusammenfassung und Ausblick

In diesem Kapitel erfolgt die abschließende Zusammenfassung der Arbeit, welche die erreichten Ziele sowie die wissenschaftlichen Beiträge herausstellt. Des Weiteren erfolgt ein Ausblick auf zukünftige Forschungsmöglichkeiten, welche auf den gewonnenen Erkenntnissen aus der vorliegenden Arbeit beruhen und diese fortführen.

8.1 Zusammenfassung

Die vorliegende Arbeit wurde durch den Sonderforschungsbereich 588 – Humanoide Roboter – gefördert, der sich zur Aufgabe gemacht hat, neue Verfahren, Konzepte, Algorithmen und mechatronische Komponenten zu entwickeln, welche einem humanoiden Roboter die Fähigkeit verleihen, gemeinsam mit dem Menschen in einem Umfeld, wie beispielsweise der Küche, zu agieren.

Für humanoide Roboter, wie auch andere autonome Systeme, ist es essentiell, die aktuelle Umgebung im benötigten Detailgrad wahrzunehmen (z. B. zur Bewältigung von Aufgaben) und mit ihr zu interagieren. Die Szenenexploration, welche eine ganzheitliche Erfassung weiter Teile einer Umgebung ermöglicht, wurde im Rahmen der vorliegenden Arbeit auf Grundlage von Salienz (auch: Auffälligkeit) und Neugier in eine *interessengetriebene Exploration* überführt, bei der die Priorisierung der Analyse einzelner Objekte aufgrund der zuvor genannten *Interessensaspekte* erfolgt. Zusätzlich wird der *aktuell erfassbare Ausschnitt* der Umwelt durch die Sensoren – in Anlehnung an das Blickfeld des Menschen – berücksichtigt.

Die *Salienz* findet eine weite Verbreitung im Bereich der Robotik und wird insbesondere für die Modellierung der Aufmerksamkeit verwendet (vgl. [Itt98],

[Kay05], [Rue08], [Sch10]). Im Rahmen der Arbeit wurde sowohl die akustische als auch die visuelle Salienz detailliert betrachtet und in Form von unimodalen Salienzclustern modelliert, welche sowohl optisch auffällige Regionen in einem Bild als auch herausstechende Bereiche von Audiosignalen beschreiben (vgl. Kapitel 2). Durch eine räumlich-zeitliche Fusion einzelner unimodaler Salienzcluster zu *multimodalen Salienzclustern* ist eine kombinierte Repräsentation entstanden, welche als Grundlage für die Salienz von Objekten dient (vgl. [Küh12b]).

Die interessengetriebene Exploration berücksichtigt neben der Salienz auch die *wissensbasierte Neugier* (vgl. Kapitel 4): Im Rahmen der vorliegenden Arbeit wurden mit Hilfe von Erkenntnissen aus der Psychologie die *situativen Bedingungen* für die Neugier des Menschen (vgl. [Ber74], [Luk06]) auf ein autonomes System mit einem Umweltmodell (vgl. Kapitel 3, [Mac10b], [Mac10a], [Küh12a]) übertragen. Die einzelnen Bestandteile der wissensbasierten Neugier sind *Neuartigkeit, Komplexität, Unsicherheit und Konflikt*. Die zuvor genannten Aspekte der Neugier wurden im Rahmen der vorliegenden Arbeit durch verschiedene Teilaspekte modelliert, welche auf den aktuellen Informationen über ein Objekt im Umweltmodell basieren. Hierzu zählt das vorhandene Wissen, welches sich u. a. aus dem a-priori definierten und dem aktuell durch Sensoren wahrgenommenen Wissen über Personen und Gegenstände zusammensetzt. Im Wesentlichen sind dies die *Attribute* und *Relationen* eines Objekts sowie die *Zeitpunkte der Wahrnehmung*. Gerade Letzteres hat einen signifikanten Einfluss auf die Neuartigkeit eines Objekts.

Im Rahmen der vorliegenden Arbeit wurden verschiedene *Anforderungen an die interessengetriebene Exploration* (vgl. Kapitel 5) formuliert: Hierzu zählen die Erweiterung der objektzentrierten Umwelterfassung (vgl. [Swe09], [Mac10a]) zu einem *ganzheitlichen Explorationsansatz*, die *priorisierte Erfassung* der Umgebung durch Berücksichtigung von *Interessenaspekten*, die *Anlehnung an den Menschen* sowie die Berücksichtigung von *externen Faktoren*, welche den Ablauf der Exploration beeinflussen. Insbesondere der zuletzt genannte Aspekt ist von entscheidender Bedeutung, da eine vollständige Exploration der aktuellen Umgebung sehr viel Zeit beanspruchen kann. Dringende Aufgaben oder andere „High-Level"-Prozesse (z. B. die Interaktion mit dem Menschen) können die aktuelle Exploration unterbrechen. Durch den Einsatz verschiedener Sensoren wird währenddessen versucht, einen Großteil der Szene im Blick zu behalten. Bei der anschließenden Fortführung der Exploration wird ein Unterdrückungsmechanismus (*inhibition of return*; vgl. [Küh12b]) verwendet, welcher die Anzahl an noch zu explorierenden Objekten stark reduziert, indem die bereits gesammelten Objektinformationen berücksichtigt und die Objekte in der Szene möglichst lange nachverfolgt bzw. schnell wiedererkannt werden. Mit Hilfe der *interessengetriebenen Perzeption* sowie der daraus abgeleiteten interessengetriebenen Szenenexploration lassen sich die zuvor genannten Ziele unter Berücksichtigung der Anforderungen realisieren. Hierfür wird unter Zuhilfenahme der *interessenge-*

triebenen Explorationsstrategie für die zu betrachtenden Objekte ein *Explorationspfad* bestimmt, welcher die Interessenaspekte Salienz, Neugier und aktueller Umweltaussicht berücksichtigt. Unter Verwendung der *Einflussparameter* γ_s und γ_η kann dabei u. a. der Einfluss von Salienz und Neugier auf den Explorationspfad festgelegt werden.

Die Grundlage für die Evaluation bildet das OPASCA-System (vgl. Kapitel 6, [Mac10b], [Küh10], [Küh12a]), welches im Rahmen der Arbeiten von Machmer und Swerdlow (vgl. [Mac10a], [Swe09]) zur Generierung und Fusion von Umweltwissen sowie audiovisuellen Signaturen entstand. Im Rahmen der vorliegenden Arbeit wurde das Grundsystem zur objektzentrierten Umwelterfassung um weitere Aspekte – u. a. multimodale Salienz und wissensbasierte Neugier – ergänzt und mit Hilfe der interessengetriebenen Exploration zu einem *ganzheitlichen System zur Szenenexploration* weiterentwickelt. Des Weiteren wurde auch die Berücksichtigung von a-priori-Wissen erweitert sowie die Akquise von Zusatzinformationen im Rahmen eines Dialogs mit dem Menschen ergänzt. Das OPASCA-System besitzt ein *Umweltmodell als zentrale Gedächtnisstruktur*, in welchem u. a. das a-priori-Wissen sowie das aktuell vorhandene Objektwissen abgelegt sind. Die Algorithmen und Verfahren zur Szenenexploration wurden in Form von Systemmodulen realisiert, welche mit dem Umweltmodell interagieren und sukzessiv das Wissen über die wahrgenommenen Objekte erweitern (vgl. Kapitel 3; Abstraktionsebenen, Klassenhierarchie, Wissensabhängigkeiten, Objektsignaturen, Lebenszyklus).

Im Rahmen der Evaluation (vgl. Kapitel 7) wurde für zwei unterschiedliche Demonstratoren (ARMAR-III-Roboterkopf und PTU-Sensoraufbau; vgl. [Asf08], [Küh12a]) ein Realdatensatz erstellt, welcher alltägliche Szenen aus den Bereichen Frühstück und Büro beinhaltet. Der Realdatensatz enthält sowohl visuelle Sensordaten von zwei Stereokamerapaaren mit unterschiedlicher Brennweite als auch akustische Daten von einem Mikrofonarray. Die Sensoren sind dabei als Einheit beweglich, sodass jeder Punkt im Raum fokussiert werden kann. Neben dem Realdatensatz wurde ein Simulationsdatensatz für den Vergleich angelegt. Im Rahmen der vorliegenden Arbeit wurden drei *Bewertungsmaße* für Explorationspfade eingeführt, welche die *Abnahme der Salienz* und die *Abnahme der Neugier* entlang des Explorationspfads sowie die *Gesamtbewegung während der Exploration* bewerten. Eine Analyse der interessengetriebenen Explorationsstrategie mit den Interessenaspekten Salienz, Neugier und Bewegung hat dabei gezeigt, dass eine unabhängige Optimierung nicht möglich ist. Vielmehr führt die Priorisierung einer Größe i. d. R. zu einer reduzierten Berücksichtigung der anderen beiden Größen. Durch die Einführung von Skalarisierungsfaktoren zum anteiligen Einfluss von Salienz, Neugier und Bewegung konnte das Optimierungsproblem für den interessengetriebenen Explorationspfad gelöst werden. Im Rahmen der Evaluation konnte gezeigt werden, dass der interessengetriebene Explorationspfad im Vergleich zu anderen Explorationspfaden – wie salienz- oder neugierba-

sierten Pfaden oder auch dem bewegungsoptimierten Explorationspfad – sowie einem Referenzpfad, welcher der Reihenfolge der Objekterfassung entspricht, eine überdurchschnittlich gute Anpassungsmöglichkeit bietet. Ein Vergleich mit dem *indizesbasierten Korrelationsmaß* sowie ein abschließendes Beispiel haben die zuvor getroffenen Aussagen untermauert.

Zusammenfassend sind die *wichtigsten Beiträge* der vorliegenden Arbeit die Verwendung von multimodalen Salienzclustern zur Modellierung von Auffälligkeiten in einer Szene, die wissensbasierte Neugier auf Basis der situativen Bedingungen für die Neugier beim Menschen, die interessengetriebene Szenenexploration zur ganzheitlichen Erfassung einer Szene unter Berücksichtigung verschiedener Interessenaspekte sowie die Einführung von Bewertungs- und Evaluationsmaßen für Explorationspfade.

8.2 Ausblick

Im Rahmen der vorliegenden Arbeit sind eine Reihe an Erkenntnissen, Ansätzen, Vorgehensweisen und Methoden entstanden, welche entweder separat oder auch im Ganzen fortgeführt werden können:

- Die verwendete *Systemarchitektur* kann in zukünftigen Arbeiten um weitere Aspekte und Module ergänzt werden, welche mit Hilfe neuer Verfahren und Algorithmen eine noch *detailliertere Wahrnehmung der Umwelt* ermöglichen und *weitere Sensoren* sowie *Informationsquellen* einbinden. Beispiele hierfür sind verteilte Sensoren (z. B. Kameranetzwerke), haptische Sensoren (z. B. Kraftsensoren) oder Biosignale vom Menschen (z. B. über eine SmartWatch), welche eine vielfältige multimodale Wahrnehmung der Umwelt ermöglichen. Des Weiteren können auch Verfahren, welche beispielsweise die menschlichen Emotionen (vgl. [Gri07]) interpretieren, ergänzt werden.
- Durch den *Austausch von vorhandenem Wissen* zwischen Robotern und/oder anderen autonomen Systemen in unterschiedlichen Umgebungen (z. B. in einem anderen Haushalt) kann die Menge an vorhandenem Wissen stetig vergrößert werden. Hierfür sind eine *Generalisierung und die Konsolidierung* von teils heterogenem Wissen erforderlich. Außerdem können durch den Einsatz und die *Kooperation von mehreren Robotern* bzw. anderen autonomen Systemen verschiedene Aufgaben, wie beispielsweise die Exploration eines größeren Bereichs, schneller und zuverlässiger erledigt werden.
- Die Einbeziehung der *Salienz* bietet bei verschiedenen Prozessen der Bild- und Signalverarbeitung einen Mehrwert in Bezug auf auffällige Bild- und Audiobereiche. Anwendungen, wie beispielsweise die Überwachung von Liegenschaften, können aufgewertet werden, indem die Salienz als zusätzliche

Information einfließt und somit z. B. Vorkommnisse, wie ein Einbruchsversuch, effektiver detektiert werden können.

- Die *wissensbasierte Neugier* dient allgemein als Antrieb für die Wissensakquise, indem beispielsweise Konflikte aufgezeigt werden, Neuartiges priorisiert behandelt wird oder auch Unsicherheit betrachtet wird. Diese Aspekte eröffnen ein weites Feld an Anwendungen, welche insbesondere der humanoiden Robotik zugutekommen. Aber auch andere autonome Systeme zur Informationsgewinnung und -konsolidierung können von der wissensbasierten Neugier profitieren.
- Die *interessengetriebene Exploration* berücksichtigt bei der Generierung von Explorationspfaden eine spezielle Strategie, bei der verschiedene gewichtete Interessenaspekte – multimodale Salienz, wissensbasierte Neugier und aktueller Umweltausschnitt – zur Priorisierung der Objekte herangezogen werden. Neben den zuvor genannten Interessenaspekten können in weiteren Arbeiten auch individuelle Aspekte, wie beispielsweise Vorlieben oder persönliche Ziele, berücksichtigt werden.
- Die in der vorliegenden Arbeit eingeführten *Bewertungsmaße* können auch auf andere Strategien zur Exploration (z. B. anwendungs- oder aufgabenorientierte Explorationsstrategien) übertragen werden. Das *indizesbasierte Korrelationsmaß* ermöglicht bereits jetzt einen von der Strategie unabhängigen Vergleich von zwei Explorationspfaden.
- Die *aufgabenorientierte Exploration* ist eine Fortführung der vorliegenden Arbeit mit Hinblick auf verschiedene Aufgaben, die humanoide Roboter oder andere autonome Systeme im täglichen Betrieb bewältigen müssen. Dabei werden nicht mehr alle in der Szene vorhandenen Objekte detailliert exploriert, sondern nur auf der für die Aufgabe notwendigen Abstraktionsebene. Die Bewältigung von Aufgaben ist in diesem Zusammenhang eine besondere Form von Interesse und kann ebenfalls als Interessenaspekt bei der Exploration modelliert werden. Wird das Küchenszenario „Tisch decken“ als Beispiel betrachtet, so werden das benötigte Geschirr und der Tisch im Detail erfasst, andere Objekte in der Küche (z. B. Stühle) hingegen mit einem höheren Abstraktionsgrad (z. B. als Hindernisse). Die Aufgabe bildet somit ein eigenes Kriterium für die Priorisierung der Objekte während der Exploration.

Teil V

Anhang

A

Beispiel zur Bestimmung der Neugier

In diesem Kapitel wird anhand von konkreten Beispielen die Bestimmung der wissensbasierten Neugier durch das vorhandene a-priori-Wissen und die im Umweltmodell erfassten Informationen dargestellt. Dafür werden exemplarisch die Attribute, die Relationen und die erfassten Zeitpunkte der Wahrnehmung der im Umweltmodell vorhandenen Objektrepräsentanten aufgeführt sowie das hinterlegte a-priori-Wissen definiert. Anschließend werden für die zuvor beschriebenen Objektrepräsentanten die Teilaspekte und Aspekte sowie die wissensbasierte Neugier selbst bestimmt.

A.1 a-priori-Wissen

In Tabelle A.1 und A.2 ist exemplarisch das a-priori-Wissen für die Gegenstandtypen *Stabmixer* und *Äpfel* aufgeführt. Für die einzelnen Attribute und Relationen werden entsprechend gültige Wertebereiche oder eine Auswahl an validen Werten definiert (vgl. Abschnitt 3.1.1). Beispiele hierfür sind die Länge eines Stabmixers oder die Farben eines Apfels. An dieser Stelle sind nur einige Attribute und Relationen beispielhaft aufgeführt. Das OPASCA-System (vgl. Kapitel 6) enthält hingegen einen wesentlich größeren Satz an a-priori-Wissen.

$\mathcal{A}_{o_p}$	$\mathcal{R}_{o_p}$	Typ	Link	Werte
a_1	–	Gegenst.typ	–	Stabmixer
a_2	–	Farbe	–	{Weiß, Grau, Schwarz, Silber}
a_3	–	Länge	–	25 cm bis 40 cm
–	r_1	–	Nutzung	{Küche}

Tabelle A.1: Vorhandenes a-priori-Wissen über den Gegenstand *Stabmixer*

$\mathcal{A}_{o_p}$	$\mathcal{R}_{o_p}$	Typ	Link	Werte
a_1	–	Gegenst.typ	–	Apfel
a_2	–	Größe	–	6 cm bis 11 cm
a_3	–	Farbe	–	{Grün, Gelb, Rot}
–	r_1	–	Herkunft	{Deutschland, Frankreich}

Tabelle A.2: Vorhandenes a-priori-Wissen über den Gegenstand *Apfel*

A.2 Objektrepräsentanten im Umweltmodell

Im Umweltmodell werden die im Rahmen der Perzeption gewonnenen Informationen über Gegenstände oder Personen in Form von Objekten repräsentiert. Diese besitzen unter anderem, wie in Gl. 3.2 definiert, Attribute, Relationen und die Zeitpunkte der letzten Wahrnehmungen. In den folgenden Abschnitten werden die akquirierten Informationen für vier Objektrepräsentanten exemplarisch aufgeführt.

A.2.1 Attribute und Relationen

In Tabelle A.3 sind die erfassten Informationen des Objektrepräsentanten o_1 aufgeführt. Es handelt sich bei dem Objekt um eine *Person*, welche sowohl akustisch als auch visuell erfasst wurde. Die Identität der Person ist *Benjamin*. Des Weiteren wurden die Größe und die Haarfarbe bestimmt. Das Alter und der Wohnort sind hingegen unbekannt.

Bei dem zweiten Objektrepräsentant (o_2) handelt es sich um einen Gegenstand vom Typ *Stabmixer*. Die erfassten Informationen sind in Tabelle A.4 dargestellt. Der Mixer wurde ebenfalls multimodal erfasst und besitzt eine Reihe von Attributen und Relationen, die allesamt gut bis sehr gut erfasst wurden.

In Tabelle A.5 sind die wahrgenommenen Informationen des dritten Objektrepräsentanten (o_3) zusammengefasst. Das Objekt repräsentiert eine Person (*Peter*). Diese wurde zwar audiovisuell erfasst, die Klassifikationsergebnisse sind aufgrund von äußeren Einflüssen jedoch schlecht, sodass eine falsche Identität bestimmt wird. Die restlichen Attribute sind recht gut erfasst.

Der vierte Repräsentant (o_4) ist ein *Apfel*, welcher nur visuell wahrgenommen wurde. Die im Umweltmodell vorhandenen Informationen sind in Tabelle A.6 dargestellt. Die Eigenschaften *Herkunft* und *Farbe* wurden fehlerhaft erfasst.

Hinweis: Bei den Attributen *Identität* und *Gegenstandtyp* ist in den jeweiligen Tabellen (A.3, A.4, A.5 und A.6) nur der Gewinner in der Spalte Wert aufgeführt.

$\mathcal{A}_o$	$\mathcal{R}_o$	Typ	Link	Wert	Konfidenz	Priorität
a_1	–	Identität	–	Benjamin	0,85	1,00
		↪ akustisch	–	Peter	0,31	–
		↪ visuell	–	Benjamin	0,92	–
a_2	–	Größe	–	1,80 m	0,71	0,25
a_3	–	Alter	–	*nicht erfasst*	1,00	0,40
a_4	–	Haarfarbe	–	Braun	0,37	0,30
–	r_1	–	Wohnort	*nicht erfasst*	1,00	0,50

Tabelle A.3: Attribute und Relationen eines wahrgenommenen Objekts (o_1), welches die Person *Benjamin* repräsentiert.

$\mathcal{A}_o$	$\mathcal{R}_o$	Typ	Link	Wert	Konfidenz	Priorität
a_1	–	Gegenst.typ	–	Stabmixer	0,72	1,00
		↪ akustisch	–	Stabmixer	0,78	–
		↪ visuell	–	Stabmixer	0,69	–
a_2	–	Farbe	–	Schwarz	0,95	0,60
a_3	–	Länge	–	34,7 cm	0,88	0,45
a_4	–	Zustand	–	aktiv	0,98	0,80
a_5	–	Textur	–	0,02	1,00	1,00
a_6	–	Kontur	–	0,16	1,00	1,00
a_7	–	Farbgebung	–	0,06	1,00	1,00
–	r_1	–	Position	auf Tisch	1,00	0,75

Tabelle A.4: Attribute und Relationen eines wahrgenommenen Objekts (o_2), welches den Gegenstand *Stabmixer* repräsentiert.

$\mathcal{A}_o$	$\mathcal{R}_o$	Typ	Link	Wert	Konfidenz	Priorität
a_1	–	Identität	–	Benjamin	0,23	1,00
		↪ akustisch	–	Peter	0,08	–
		↪ visuell	–	Benjamin	0,44	–
a_2	–	Größe	–	1,71 m	0,89	0,25
a_3	–	Alter	–	25 Jahre	0,64	0,40
a_4	–	Haarfarbe	–	Schwarz	0,91	0,30
–	r_1	–	Wohnort	*nicht erfasst*	1,00	0,50

Tabelle A.5: Attribute und Relationen eines wahrgenommenen Objekts (o_3), welches die Person *Peter* repräsentiert.

$\mathcal{A}_o$	$\mathcal{R}_o$	Typ	Link	Wert	Konfidenz	Priorität
a_1	–	Gegenst.typ	–	Apfel	0,81	1,00
		↪ visuell	–	Apfel	0,81	–
a_2	–	Farbe	–	Blau	0,73	0,60
a_3	–	Durchmesser	–	9,2 cm	0,76	0,55
a_4	–	Sorte	–	*nicht erfasst*	1,00	0,50
a_5	–	Textur	–	0,16	1,00	1,00
a_6	–	Kontur	–	0,18	1,00	1,00
a_7	–	Farbgebung	–	0,37	1,00	1,00
–	r_1	–	Herkunft	Alaska	0,56	0,70

Tabelle A.6: Attribute und Relationen eines wahrgenommenen Objekts (o_4), welches den Gegenstand *Apfel* repräsentiert.

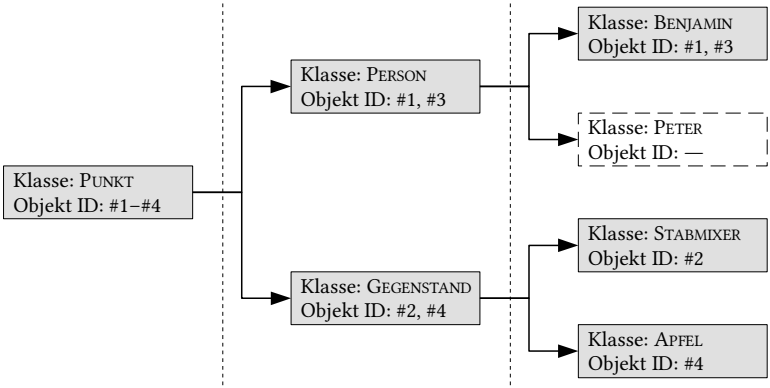

Abb. A.1: Klassenhierarchie der einzelnen Objektrepräsentanten $o_1, \ldots, o_4$ aus den Tabellen A.3, A.4, A.5 und A.6

Die Klassenrelationen und die Klassenhierarchie (vgl. Abschnitt 3.2.2) der zuvor beschriebenen Objektrepräsentanten sind in den korrespondierenden Tabellen nicht ausdrücklich aufgeführt, da diese bei der Bestimmung der wissensbasierten Neugier nicht explizit berücksichtigt werden. Zur Vollständigkeit werden diese im Folgenden angegeben: Für die beiden Personen in Tabelle A.3 bzw. Tabelle A.5 sind die Relationen istPUNKT, istPERSON und istBENJAMIN vorhanden. Für die Gegenstände in Tabelle A.4 bzw. Tabelle A.6 existieren die Klassenrelationen istPUNKT, istGEGENSTAND und istSTABMIXER bzw. istAPFEL. Die dazugehörige Klassenhierarchie ist für alle Repräsentanten in Abb. A.1 dargestellt. Aufgrund von fehlerhaften Informationen ist bei Repräsentant o_3 die Identität der Person falsch (vgl. Tabelle A.5) und somit auch die Zuordnung in der Klassenhierarchie, d. h. Klasse BENJAMIN anstatt PETER.

A.2.2 Zeitpunkte der Wahrnehmung

In Tabelle A.7 sind die Zeitpunkte der letzten Wahrnehmungen der zuvor aufgelisteten Objektrepräsentanten $o_1, o_2, \ldots, o_4$ aufgeführt. Die Zeitdifferenz ist in Tagen relativ zum aktuellen Zeitpunkt angegeben. Das negative Vorzeichen deutet auf den Bezug zur Vergangenheit hin.

Objekte $\mathcal{O}$	$\Delta\mathcal{Z}_o$ (in Tagen)						
o_1: Benjamin	−15,99	−20,21	−21,13	−21,16	−21,32	−21,71	−23,24
o_2: Stabmixer	−0,26	−1,35	−1,75	−1,44	−3,57	−8,51	−9,82
	−20,19	−21,04	−21,16	−22,73	−25,20	−30,85	−34,30
o_3: Peter	−3,65	−4,22	−4,88	−5,79	−6,18	−7,31	−10,68
	−10,77	−11,05	−11,07	−13,17	−14,10	−14,30	−14,72
	−15,93	−16,15	−16,81	−19,93	−20,32	−20,56	−20,71
	−20,73	−22,09	−22,09	−25,85	−26,05	−26,18	−26,22
	−26,43	−26,43	−26,74	−27,25	−27,41	−28,20	−28,87
	−29,22	−30,28	−30,53	−31,85	−32,55	−33,48	−33,70
o_4: Apfel	−5,95	−6,35	−6,67	−7,89	−9,86	−10,91	−13,98
	−14,32	−15,20	−15,76	−15,93	−16,88	−16,93	−20,90
	−21,48	−22,98	−23,04	−23,30	−24,14	−26,27	−30,34
	−31,85	−32,76	−32,95	−33,21	−34,03	−35,38	−35,40

Tabelle A.7: Zeitpunkte der Wahrnehmung der Objekte in der Vergangenheit (relativ in Tagen)

A.3 Detaillierte Ergebnisse

A.3.1 Einzelergebnisse der Teilaspekte

In den folgenden Abschnitten werden die Einzelergebnisse der wissensbasierten Neugier – die Neuartigkeit, die Komplexität, die Unsicherheit und der Konflikt – für die zuvor definierten Objektrepräsentanten $o_1, \ldots, o_4$ näher erläutert.

Neuartigkeit

Die Neuartigkeit eines Objekts kann, wie in Abschnitt 4.3.3 beschrieben, aus den vorhandenen Informationen im Umweltmodell bestimmt werden. Die Ergebnisse für die Neuartigkeit sind in Tabelle A.8 enthalten. Der Grad der Neuartigkeit π_o^{Grad} ist für Objektrepräsentant o_1 und o_3 etwas höher als bei den restlichen Repräsentanten, was sich auf die wenigen vorhandenen Informationen sowie nicht erfasste Attribute und Relationen zurückführen lässt (vgl. Tabelle A.3 und Tabelle A.5). Die Ergebnisse für den Zeitpunkt der letzten Wahrnehmung $\pi_o^{\text{Zeitpunkt}}$ können direkt aus den Werten in Tabelle A.7 abgeleitet werden. Es ist zu sehen, dass das Objekt, welche durch o_1 repräsentiert wird, vor knapp 16 Tagen das letzte Mal wahrgenommen wurde. Alle anderen Objekte wurden zwischenzeitlich mehrmals erfasst. Dementsprechend ist bei o_1 auch der Wert am höchsten. Die Häufigkeit der Wahrnehmung $\pi_o^{\text{Häufigkeit}}$ kann für die Objekte, wie zuvor schon, aus den Daten in Tabelle A.7 bestimmt werden. Die Objekte o_2, o_3 und o_4 wurden über die Zeit recht häufig wahrgenommen, wohingegen das Objekt o_1 nur in einem relativ kurzen Zeitraum exploriert wurde. Somit ist auch der Wert für das Objekt o_1 um ein Vielfaches höher als bei den restlichen Objekten.

Komplexität

Für die Objektrepräsentanten o_2 und o_4 lässt sich die Komplexität wie in Abschnitt 4.3.4 beschrieben bestimmen. Die beiden Personen (Repräsentanten o_1 und o_3) besitzen definitionsgemäß keine Komplexität. Die Objektkomplexität κ_o^{Objekt} lässt sich in diesem Beispiel durch die drei Attribute a_{Textur}, a_{Kontur} und $a_{\text{Farbgebung}}$ bestimmen. Die Ergebnisse sind in Tabelle A.9 für die Objektrepräsentanten o_2 und o_4 zu finden. Dort sind auch die Resultate für die Szenenkomplexität κ_o^{Szene} aufgeführt.

Unsicherheit

Die Bestimmung der Unsicherheit für die Objektrepräsentanten im Umweltmodell erfolgt wie in Abschnitt 4.3.5 beschrieben. Die Ergebnisse für die Typ- bzw. Identitätsunsicherheit ν_o^{Klasse} sind in Tabelle A.10 angegeben. Dabei fällt eine erhöhte Unsicherheit für den Objektrepräsentanten o_3 auf. Diese lässt sich durch einen geringen Konfidenzwert beim Attribut Identität erklären, d. h., die gesammelten Informationen über die Person lassen keine sichere Bestimmung der Identität zu. Die Resultate für attributs- und relationsbezogene Unsicherheit $\nu_o^{\text{Eigenschaften}}$ setzen sich aus den Konfidenzwerten und Gewichten aller Attribute und Relationen zusammen. Aufgrund der zuvor angesprochenen Unsicherheit der Identität bei Repräsentant o_3 ist auch diese Unsicherheit davon beeinflusst. Durch den geringen Konfidenzwert für die Haarfarbe bei Objektrepräsentant o_1 ist auch die Unsicherheit leicht erhöht. Die anderen Repräsentanten haben entsprechend ihrer Konfidenzwerte eine geringere Unsicherheit.

Konflikt

Der Konflikt für die Objektrepräsentanten des Umweltmodells lässt sich wie in Abschnitt 4.3.6 beschrieben bestimmen. Dabei werden drei Arten von Konflikten berücksichtigt (vgl. Tabelle A.11): Als erstes wird der Konflikt in Bezug auf die Identität einer Person $\varsigma_o^{\text{Identität}}$ betrachtet. Dieser ist für zwei Repräsentanten symmetrisch und hängt von den Konfidenzwerten der Identität der beiden Personen ab. Da der entsprechende Konfidenzwert für den Objektrepräsentanten o_3 wesentlich geringer ist als für den Objektrepräsentanten o_1, ist entsprechend der Konflikt für beide Personen nicht sehr hoch (vgl. Tabelle A.3 und Tabelle A.5). Für die Gegenstände ist analog kein Konflikt bzgl. des Gegenstandstyps möglich, da diese im Gegensatz zum Menschen mehrmals vorkommen können. Die zweite Konfliktart entsteht durch eine unterschiedliche multimodale Wahrnehmung $\varsigma_o^{\text{Multimodal}}$ eines Objekts. Der Konflikt lässt sich bestimmen, indem die Teilergebnisse für ein Attribut bzw. eine Relation jeder Modalität explizit ermittelt und diese mit den fusionierten Ergebnissen vergleichen werden. Für das Attribut *Identität* des Objektrepräsentanten o_3 in Tabelle A.5 liegt ein Konflikt vor, welcher sich aus dem Minimum der beiden Konfidenzwerte bestimmen lässt (vgl. Gl. 4.62). Als letzte Konfliktart wird der Konflikt mit dem a-priori-Wissen $\varsigma_o^{\text{Vorwissen}}$ betrachtet. Deutlich zu sehen ist, dass nur für den Objektrepräsentant o_4 ein solcher Konflikt vorliegt. Dies ist sowohl durch die wahrgenommene *Farbe* des Apfels als auch durch das erfasste *Herkunftsland* begründet (vgl. Tabelle A.6), da beide Informationen nicht mit denen im a-priori-Wissen definierten Vorgaben (vgl. Tabelle A.2) übereinstimmen.

A.3.2 Gesamtergebnisse für die wissensbasierte Neugier

Die Ergebnisse mit den Daten aus den Beispielen zuvor und der gewählten Fusionsfunktion $f_{\text{Mittelwert}}$ (vgl. Gl. 4.77) sind nach Teilaspekten getrennt in den Tabellen A.8 bis A.11 dargestellt. Das Gesamtergebnis für die wissensbasierte Neugier ist in Tabelle A.12 für alle Beispiele aufgeführt.

Objekte $\mathcal{O}$	π_o^{Grad}	$\pi_o^{\text{Zeitpunkt}}$	$\pi_o^{\text{Häufigkeit}}$	π_o
o_1: Benjamin	0,327	0,574	0,526	0,476
o_2: Stabmixer	0,000	0,003	0,033	0,012
o_3: Peter	0,204	0,016	0,000	0,073
o_4: Apfel	0,079	0,044	0,000	0,041

Tabelle A.8: Ergebnisse der Teilaspekte der Neuartigkeit

Objekte $\mathcal{O}$	κ_o^{Objekt}	κ_o^{Szene}	–	κ_o
o_1: Benjamin	–	–	–	–
o_2: Stabmixer	0,080	0,223	–	0,151
o_3: Peter	–	–	–	–
o_4: Apfel	0,237	0,223	–	0,230

Tabelle A.9: Ergebnisse der Teilaspekte der Komplexität

Objekte $\mathcal{O}$	ν_o^{Klasse}	$\nu_o^{\text{Eigenschaften}}$	–	ν_o
o_1: Benjamin	0,150	0,194	–	0,172
o_2: Stabmixer	0,280	0,058	–	0,169
o_3: Peter	0,770	0,414	–	0,592
o_4: Apfel	0,190	0,076	–	0,133

Tabelle A.10: Ergebnisse der Teilaspekte der Unsicherheit

Objekte $\mathcal{O}$	$\varsigma_o^{\text{Identität}}$	$\varsigma_o^{\text{Multimodal}}$	$\varsigma_o^{\text{Vorwissen}}$	ς_o
o_1: Benjamin	0,230	0,310	0,000	0,180
o_2: Stabmixer	–	0,000	0,000	0,000
o_3: Peter	0,230	0,080	0,000	0,103
o_4: Apfel	–	0,000	0,638	0,319

Tabelle A.11: Ergebnisse der Teilaspekte des Konflikts

Objekte $\mathcal{O}$	π_o	κ_o	ν_o	ς_o	η_o
o_1: Benjamin	0,476	–	0,172	0,180	0,276
o_2: Stabmixer	0,012	0,151	0,169	0,000	0,083
o_3: Peter	0,073	–	0,592	0,103	0,256
o_4: Apfel	0,041	0,230	0,133	0,319	0,181

Tabelle A.12: Ergebnisse aller Aspekte der wissensbasierten Neugier

Literaturverzeichnis

[And11] Andreopoulos, A.; Hasler, S.; Wersing, H.; Janssen, H.; Tsotsos, J. und Körner, E.: Active 3D Object Localization Using a Humanoid Robot. *IEEE Transaction on Robotics* (2011), Bd. 27(1):S. 47–64

[App07] Applegate, D. L.; Bixby, R. E.; Chvátal, V. und Cook, W. J.: *The Traveling Salesman Problem: A Computational Study*, Princeton University Press (2007)

[Aru02] Arulampalam, M. S.; Maskell, S.; Gordon, N. und Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* (2002), Bd. 50(2):S. 174–188

[Asf06] Asfour, T.; Regenstein, K.; Azad, P.; Schröder, J.; Bierbaum, A.; Vahrenkamp, N. und Dillmann, R.: ARMAR-III: An integrated humanoid platform for sensory-motor control, in: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Genua, Italien (2006), S. 169–175

[Asf08] Asfour, T.; Welke, K.; Azad, P.; Ude, A. und Dillmann, R.: The Karlsruhe Humanoid Head, in: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Daejeon, Südkorea (2008), S. 447–453

[Asf13] Asfour, T.; Schill, J.; Peters, H.; Klas, C.; Bücker, J.; Sander, C.; Schulz, S.; Kargov, A.; Werner, T. und Bartenbach, V.: ARMAR-4: A 63 DOF Torque Controlled Humanoid Robot, in: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Altanta, Georgia, USA (2013)

[Atk00] Atkeson, C. G.; Hale, J. G.; Pollick, F.; Riley, M.; Kotosaka, S.; Schaul, S.; Shibata, T.; Tevatia, G.; Ude, A.; Vijayakumar, S.; Kawato, E. und Kawato, M.: Using humanoid robots to study human behavior. *IEEE Journal on Intelligent Systems and their Applications* (2000), Bd. 15(4):S. 46–56

[Bau10] Baum, M.; Gheţa, I.; Belkin, A.; Beyerer, J. und Hanebeck, U. D.: Data Association in a World Model for Autonomous Systems, in: *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Salt Lake City, Utah, USA (2010), S. 187–192

[Bec06] Bechler, D.: *Akustische Sprecherlokalisation mit Hilfe eines Mikrofonarrays*, Dissertationsschrift, Universität Karlsruhe (TH), Fakultät für Elektrotechnik und Informationstechnik, Karlsruhe (2006)

[Beg10] Begum, M.; Karray, F.; Mann, G. K. I. und Gosine, R. G.: A Probabilistic Model of Overt Visual Attention for Cognitive Robots. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics* (2010), Bd. 40:S. 1305–1318

[Bel12] Belkin, A.; Kuwertz, A.; Fischer, Y. und Beyerer, J: World modeling for autonomous systems, in: *Innovative Information Systems Modelling Techniques*, InTech – Open Access Publisher (2012), S. 137–158

[Ber60] Berlyne, D. E.: *Conflict, arousal, and curiosity*, McGraw-Hill Book Company (1960)

[Ber66] Berlyne, D. E.: Curiosity and Exploration. *Science* (1966), Bd. 153:S. 25–33

[Ber74] Berlyne, D. E.: *Konflikt, Eregung, Neugier*, Klett, Stuttgart (1974)

[bey13] beyerdynamic: Mikrofon MCE 60, `http://www.beyerdynamic.de/shop/media/datenblaetter/DAT_MCE60_DE.pdf` (2013)

[Bie07] Bierbaum, A.; Welke, K.; Burger, D.; Asfour, T. und Dillmann, R.: Haptic Exploration for 3D Shape Reconstruction using Five-Finger Hands, in: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Pittsburgh, Pennsylvania, USA (2007), S. 616–621

[Bis06] Bishop, C. M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, New York, 1. Aufl. (2006)

[Bre99] Brefczynski, J. A. und DeYoe, E. A.: A physiological correlate of the spotlight of visual attention. *Nature Neuroscience* (1999), Bd. 2(4):S. 370–374

[Bro58] Broadbent, D. E.: *Perception and communication*, Pergamon Press (1958)

[Brü07] Brümmerhoff, D.: *Finanzwissenschaft*, Oldenbourg Verlag (2007)

[Bun90] Bundesen, C.: A theory of visual attention. *Psychological Review* (1990), Bd. 97(4):S. 523–547

[Bur06] Bur, A.; Tapus, A.; Ouerhani, N.; Siegwart, R. und Hugli, H.: Robot Navigation by Panoramic Vision and Attention Guided Features, in: *IEEE International Conference on Pattern Recognition (ICPR)*, Hong Kong, China (2006), S. 695–698

[But08] Butko, N. J.; Zhang, L.; Cottrell, G. W. und Movellan, J. R.: Visual saliency model for robot cameras, in: *IEEE International Conference on Robotics and Automation (ICRA)*, Pasadena, Kalifornien, USA (2008), S. 2398–2403

[Cat50] Cattell, R. P.: *Personality. A systematic theoretical and factual study*, New York: McGraw-Hill (1950)

[Chi08] Chinchuluun, A.; Pardalos, P. M.; Migdalas, A. und Pitsoulis, L.: *Pareto optimality, game theory and equilibria*, Bd. 17, Springer (2008)

[Cho08] Chou, W.-L. und Yeh, S.-L.: Location- and object-based inhibition of return are affected by different kinds of working memory. *The Quarterly Journal of Experimental Psychology* (2008), Bd. 61(12):S. 1761–1768

[Com02] Comaniciu, D. und Meer, P.: Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI)* (2002):S. 603–619

[Day81] Day, H. I.: *Neugierforschung: Grundlagen, Theorien, Anwendungen*, Kap. Neugier und Erziehung, H.-G. Voss und H. Keller (Hrsg.), Beltz, Weinheim (1981), S. 226–262

[Deu63] Deutsch, J. und Deutsch, D.: Attention: Some theoretical considerations. *Psychological Review* (1963), Bd. 70(1):S. 80–90

[DiB01] DiBiase, J. H.; Silverman, H. F. und Brandstein, M. S.: *Robust localization in reverberant rooms*, Kap. 8, Springer, Berlin (2001), S. 157–180

[Don07] Donohue, K. D.; Hannemann, J. und Dietz, H. G.: Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments. *Signal Processing* (2007), Bd. 87(7):S. 1677–1691

[Dor96] Dorigo, M.; Maniezzo, V. und Colorni, A.: Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* (1996), Bd. 26(1):S. 29–41

[Eke05] Ekenel, H. K. und Stiefelhagen, R.: Local appearance based face recognition using discrete cosine transform, in: *13th European Signal Processing Conference (EUSIPCO)*, Antalya, Türkei (2005), S. 2484–2487

[Eke09] Ekenel, H. K.: *A Robust Face Recognition Algorithm for Real-World Applications*, Dissertationsschrift, Universität Karlsruhe (TH), Fakultät für Informatik, Karlsruhe (2009)

[Ess00] Essa, I. A.: Ubiquitous sensing for smart and aware environments. *IEEE Personal Communications* (2000), Bd. 7(5):S. 47–49

[Fen08] Feng, W. und Hu, B.: Quaternion Discrete Cosine Transform and its Application in Color Template Matching, in: *Congress on Image and Signal Processing (CISP)*, Bd. 2, Sanya, Hainan, China (2008), S. 252–256

[Fis12] Fischer, Y. und Beyerer, J.: A Top-Down-View on Intelligent Surveillance Systems, in: *International Conference on Systems (ICONS)*, Saint Gilles, La Réunion, Frankreich (2012), S. 43–48

[Fle06] Fleming, K. A.; Peters II, R. A. und Bodenheimer, R. E.: Image mapping and visual attention on a sensory ego-sphere, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Peking, China (2006), S. 241–246

[Fre07] Freund, R. W. und Hoppe, R. W.: *Stoer/Bulirsch: Numerische Mathematik 1*, Springer, 10 Aufl. (2007)

[Fri08] Frintrop, S. und Jensfelt, P.: Attentional Landmarks and Active Gaze Control for Visual SLAM. *IEEE Transactions on Robotics* (2008), Bd. 24(5):S. 1054–1065

[Frö04] Fröba, B. und Ernst, A.: Face detection with the modified census transform, in: *IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, Seoul, Südkorea (2004), S. 91–96

[Gal04] Gallistel, C. R.; Fairhurst, S. und Balsam, P.: The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America* (2004), Bd. 101(36):S. 13124–13131

[Ghe08] Gheţa, I.; Heizmann, M. und Beyerer, J: Object oriented environment model for autonomous systems, in: *Skövde Workshop on Information Fusion Topics (SWIFT)*, Skövde, Schweden (2008), S. 9–12

[Glo01] Glover, F.; Gutin, G.; Yeo, A. und Zverovich, A.: Construction heuristics for the asymmetric TSP. *European Journal of Operational Research (EJOR)* (2001), Bd. 129(3):S. 555–568

[Gol80] Golden, B.; Boldin, T.; T., Doyle und Stewart Jr., W.: Approximate Traveling Salesman Algorithms. *Operations Research* (1980), Bd. 28(3):S. 694–711

[Gon02] González, R. C. und E., Woods R.: *Digital Image Processing*, Prentice Hall, 2. Aufl. (2002)

[Gon08] Gonzalez-Aguirre, D.; Asfour, T.; Bayro-Corrochano, E. und Dillmann, R.: Model-based visual self-localization using geometry and graphs, in: *International Conference on Pattern Recognition (ICPR)*, Tampa, Florida, USA (2008), S. 1–5

[Gri07] Grimm, M.: *Audiovisuelle Emotionserkennung für die Mensch-Maschinen-Interaktion*, Dissertationsschrift, Universität Karlsruhe (TH), Fakultät für Elektrotechnik und Informationstechnik, Karlsruhe (2007)

[Hah07] Hahsler, M. und Hornik, K.: TSP – Infrastructure for the Traveling Salesperson Problem. *Journal of Statistical Software (JSS)* (2007), Bd. 23(2):S. 1–21

[Har03] Hartley, R. und Zisserman, A.: *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, 2. Aufl. (2003)

[Her07] Hershey, J. R. und Olsen, P. A.: Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models, in: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Bd. 4, Honolulu, Hawaii, USA (2007), S. 317–320

[Hod04] Hodge, V. J. und Austin, J.: A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review* (2004), Bd. 22:S. 85–126

[Hof55] Hoffmeister, J.: *Wörterbuch der philosophischen Begriffe*, Meiner Verlag, 2. Aufl. (1955)

[Hof98] Hoffmann, R.: *Signalanalyse und -erkennung – Eine Einführung für Informationstechniker*, Springer-Verlag, Berlin Heidelberg, 4. Aufl. (1998)

[Hou12] Hou, X.; Harel, J. und Koch, C.: Image Signature: Highlighting Sparse Salient Regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2012), Bd. 34(1):S. 194–201

[Hur57] Hurvich, L. M. und Jameson, D.: An opponent-process theory of color vision. *Psychological Review* (1957), Bd. 64(6, Pt.1):S. 384–404

[Itt98] Itti, L.; Koch, C. und Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transaction on Pattern Analysis Machine Intelligence* (1998), Bd. 20(11):S. 1254–1259

[Itt00] Itti, L. und Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research* (2000), Bd. 40(10-12):S. 1489–1506

[Itt06] Itti, L. und Baldi, P. F.: Bayesian Surprise Attracts Human Attention, in: *Advances in Neural Information Processing Systems (NIPS)*, MIT Press, Cambridge, Massachusetts, USA (2006), S. 547–554

[Jai05] Jain, A.; Nandakumar, K. und Ross, A.: Score normalization in multimodal biometric systems. *Pattern Recognition* (2005), Bd. 38(12):S. 2270–2285

[Joh97] Johnson, D. S. und McGeoch, L. A.: *Local search in combinatorial optimization*, Kap. The Traveling Salesman Problem: A Case Study in Local Optimization, Princeton University Press (1997), S. 215–310

[Jon83] Jonker, R. und Volgenant, T.: Transforming asymmetric into symmetric traveling salesman problems. *Operations Research Letters* (1983), Bd. 2(4):S. 161–163

[Jua09] Juan, L. und Gwun, O.: A comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing (IJIP)* (2009), Bd. 3(4):S. 143–152

[Kal60] Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering* (1960), Bd. 82(Series D):S. 35–45

[Kal07] Kalinli, O. und Narayanan, S.: A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech, in: *Interspeech*, Antwerpen, Belgien (2007), S. 1941–1944

[Kal09] Kalinli, O. und Narayanan, S.: Prominence Detection Using Auditory Attention Cues and Task-Dependent High Level Information. *IEEE Transaction on Audio, Speech, and Language Processing* (2009), Bd. 17(5):S. 1009–1024

[Kam12] Kammeyer, K.-D. und Kroschel, K.: *Digitale Signalverarbeitung – Filterung und Spektralanalyse mit MATLAB™-Übungen*, Springer Vieweg Verlag, 8. Aufl. (2012)

[Kas12] Kasper, A.; Xue, Z. und Dillmann, R.: The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics. *The International Journal of Robotics Research* (2012), Bd. 31(8):S. 927–934

[Kay05] Kayser, C.; Petkov, C. I.; Lippert, M. und Logothetis, N. K.: Mechanisms for allocating auditory attention: an auditory saliency map. *Current Biology* (2005), Bd. 15(21):S. 1943–1947

[Kir11] Kirk, J.: Traveling Salesman Problem – Genetic Algorithm (Release: 2.3), `http://www.mathworks.de/matlabcentral/fileexchange/13680-traveling-salesman-problem-genetic-algorithm` (2011)

[Kit98] Kittler, J.: On combining classifiers . *IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI)* (1998):S. 226–239

[Kna76] Knapp, C. H. und Carter, G. C.: The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing* (1976), Bd. 24(4):S. 320–327

[Koc85] Koch, C. und Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology* (1985), Bd. 4:S. 219–227

[Kri76] Krieger, R.: *Determinanten der Wissbegier: Untersuchungen zur Theorie der intrinsischen Motivation*, Bern, Huber (1976)

[Kro11] Kroschel, K.; Rigoll, G. und Schuller, B.: *Statistische Informationstechnik*, Springer-Verlag Berlin Heidelberg, 5. Aufl. (2011)

[Küh10] Kühn, B.; Belkin, A.; Swerdlow, A.; Machmer, T.; Beyerer, J. und Kroschel, K.: Knowledge-Driven Opto-Acoustic Scene Analysis Based on an Object-Oriented World Modelling Approach for Humanoid Robots, in: *Joint 41st International Symposium on Robotics and the 6th German Conference on Robotics (ISR/ROBOTIK)*, München, Deutschland (2010), S. 1296–1303

[Küh12a] Kühn, B.; Schauerte, B.; Stiefelhagen, R. und Kroschel, K.: A Modular Audio-Visual Scene Analysis and Attention System for Humanoid Robots, in: *43rd International Symposium on Robotics (ISR)*, Taipei, Taiwan (R.O.C.) (2012), S. 1039–1044

[Küh12b] Kühn, B.; Schauerte, B.; Stiefelhagen, R. und Kroschel, K.: Multimodal Saliency-based Attention: A Lazy Robot's Approach, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vilamoura, Portugal (2012), S. 807–814

[Lic05] Lichtenauer, J.; Hendriks, E. und Reinders, M.: Isophote Properties as Features for Object Detection, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Bd. 2, San Diego, Kalifornien, USA (2005), S. 649–654

[Lie11] Liebscher, S.: *Entwurf und Umsetzung eines Systems zur Personenverifikation*, Diplomarbeit, Universität Mannheim, Lehrstuhl für Elektrotechnik, Mannheim (2011)

[Loe94] Loewenstein, G.: The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin* (1994), Bd. 116(1):S. 75–98

[Luk06] Lukesch, H.: *Einführung in die Pädagogische Psychologie*, Bd. 1 von *Psychologie in der Lehrerausbildung*, S.Roderer-Verlag (2006)

[Mac09] Machmer, T.; Swerdlow, A.; Kühn, B. und Kroschel, K.: Position Estimation of Car Occupants by Means of Speech Analysis, in: *International Conference on Acoustics including the 35th German Annual Conference on Acoustics (NAG/DAGA)*, Rotterdam, Niederlande (2009)

[Mac10a] Machmer, T.: *Generierung und Fusion von Umweltwissen für eine wissensbasierte Umwelterfassung*, Dissertationsschrift, Karlsruher Institut für Technologie (KIT), Fakultät für Elektrotechnik und Informationstechnik, Karlsruhe (2010)

[Mac10b] Machmer, T.; Swerdlow, A.; Kühn, B. und Kroschel, K.: Hierarchical, Knowledge-oriented Opto-acoustic Scene Analysis for Humanoid Robots and Man-machine Interaction, in: *IEEE International Conference on Robotics and Automation (ICRA)*, Anchorage, Alaska, USA (2010), S. 2389–2396

[McD26] McDougall, W.: *An Introduction to Social Psychology*, Boston: John W. Luce & Co. (1926)

[Meg07] Meger, D.; Forssén, P.; Lai, K.; Helmer, S.; McCann, S.; Southey, T.; Baumann, M.; Little, J. J. und Lowe, D. G.: Curious George: An Attentive Semantic Robot, in: *IROS Workshop: From sensors to human spatial concepts*, San Diego, Kalifornien, USA (2007), S. 503–511

[Met53] Metzger, W.: *Gesetze des Sehens*, Kramer, Frankfurt/Main (1953)

[Mil11] Milighetti, G.; De Luca, A. und Vallone, L.: Adaptive Predictive Gaze Control of a Redundant Humanoid Robot Head, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, Kalifornien, USA (2011), S. 3192–3198

[Mon11] Monari, E.: *Dynamische Sensorselektion zur auftragsorientierten Objektverfolgung in Kameranetzwerken*, Karlsruher Schriften zur Anthropomatik, KIT Scientific Publishing (2011)

[Mos07] Mostefa, D.; Moreau, N.; Choukri, K.; Potamianos, G.; Chu, S. M.; Casas, J. R.; Turmo, J.; Cristoferetti, L.; Tobia, F.; Pnevmatikakis, A.; Mylonakis, V.; Talantzis, F.; Burger, S.; Stiefelhagen, R.; Bernardin, K. und Rochet, C.: The CHIL Audiovisual Corpus for Lecture and Meeting Analysis inside Smart Rooms. *Language Resources and Evaluation* (2007), Bd. 41(3-4):S. 389–407

[Mos10] Mossgraber, J.; Reinert, F. und Vagts, H.: An Architecture for a Task-Oriented Surveillance System: A Service- and Event-Based Approach, in: *IEEE International Conference on Systems (ICONS)*, Les Menuires, Frankreich (2010), S. 146–151

[Mur38] Murray, H. A.: *Explorations in personality*, New York: Oxford University Press (1938)

[NES13] NEST: Network Enabled Surveillance and Tracking, `http://www.iosb.fraunhofer.de/servlet/is/4563/` (2013)

[Nic07] Nickel, K. und Stiefelhagen, R.: Fast audio-visual multi-person tracking for a humanoid stereo camera head, in: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Pittsburgh, Pennsylvania, USA (2007), S. 434–441

[Ona07] Onat, S.; Libertus, K. und König, P.: Integrating audiovisual information for the control of overt attention. *Journal of Vision* (2007), Bd. 7(10)

[Ora05] Orabona, F.; Metta, G. und Sandini, G.: Object-based Visual Attention: a Model for a Behaving Robot, in: *Workshop: Attention and Performance in Computational Vision, IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, Kalifornien, USA (2005), S. 89–96

[Ora07] Orabona, F.; Metta, G. und Sandini, G.: A Proto-object Based Visual Attention Model, in: *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint*, Bd. 4840, Springer Berlin Heidelberg (2007), S. 198–215

[O'S00] O'Shaughnessy, D.: *Speech Communications - Human and Machine*, Kap. Speech Analysis, IEEE Press (2000), S. 173–227

[Ott12] Ottmann, T. und Widmayer, P.: *Algorithmen und Datenstrukturen*, Spektrum Akademischer Verlag, 5. Aufl. (2012)

[Pet01] Peters, R. A. II; Hambuchen, K. E.; Kawamura, K. und Wilkes, D. M.: The Sensory Ego-Sphere as a Short-Term memory for Humanoids, in: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Tokyo, Japan (2001), S. 22–24

[Poi14] Point Grey Research, Inc.: Dragonfly2 CCD Color Camera, `http://ww2.ptgrey.com/firewire/dragonfly-2` (2014)

[Rey00] Reynolds, D. A.; Quatieri, T. F. und Dunn, R. B.: Speaker verification using adapted Gaussian mixture models. *Digital signal processing* (2000), Bd. 10(1):S. 19–41

[Ris04] Ristic, B.; Arulampalam, S. und Gordon, N.: *Beyond the Kalman filter: Particle filters for tracking applications*, Artech House Publishers (2004)

[Ron96] Rong Li, X. und Bar-Shalom, Y.: Tracking in clutter with nearest neighbor filters: Analysis and performance. *IEEE Transactions on Aerospace and Electronic Systems* (1996), Bd. 32(3):S. 995–1010

[Rot11] Rothermund, K. und Eder, A.: *Motivation und Emotion*, Springer (2011)

[Rue08] Ruesch, J.; Lopes, M.; Bernardino, A.; Hornstein, J.; Santos-Victor, J. und Pfeifer, R.: Multimodal Saliency-Based Bottom-Up Attention: A Framework for the Humanoid Robot iCub, in: *IEEE International Conference on Robotics and Automation (ICRA)*, Pasadena, Kalifornien, USA (2008), S. 962–967

[Sar09] Sarvadevabhatla, R. K. und Ng-Thow-Hing, V.: Panoramic attention for humanoid robots, in: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Paris, Frankreich (2009), S. 215–222

[Sch03] Schiefele, U.; Streblow, L.; Ermgassen, U. und Moschner, B.: Lernmotivation und Lernstrategien als Bedingungen der Studienleistung. *Zeitschrift für Pädagogische Psychologie* (2003), Bd. 17(3):S. 185–198

[Sch08] Schmid, S.: *Neugier und epistemisches Handeln*, Dissertationsschrift, Universität Mannheim, Fakultät für Sozialwissenschaften, Mannheim (2008)

[Sch10] Schauerte, B. und Fink, G. A.: Focusing Computational Visual Attention in Multi-Modal Human-Robot Interaction, in: *International Conference on Multimodal Interfaces (ICMI)*, Beijing, China (2010)

[Sch11a] Schauerte, B.; Kühn, B.; Kroschel, K. und Stiefelhagen, R.: Multimodal Saliency-based Attention for Object-based Scene Analysis, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, Kalifornien, USA (2011), S. 1173–1179

[Sch11b] Schiebener, D.; Ude, A.; Morimoto, J.; Asfour, T. und Dillmann, R.: Segmentation and learning of unknown objects through physical interaction, in: *IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (2011), S. 500–506

[Sch12a] Schauerte, B. und Stiefelhagen, R.: Predicting Human Gaze using Quaternion DCT Image Signature Saliency and Face Detection, in: *IEEE Workshop Applications of Computer Vision (WACV)*, Breckenridge, Colorado, USA (2012), S. 137–144

[Sch12b] Schauerte, B. und Stiefelhagen, R.: Quaternion-based Spectral Saliency Detection for Eye Fixation Prediction, in: *European Conference on Computer Vision (ECCV)*, Florenz, Italien (2012), S. 116–129

[SEN13] SENEKA: Sensornetzwerk mit mobilen Robotern für das Katastrophenmanagement, http://www.iosb.fraunhofer.de/servlet/is/28266/ (2013)

[Smi86] Smith, R. C. und Cheeseman, P.: On the Representation and Estimation of Spatial Uncertainty. *The International Journal of Robotics Research* (1986), Bd. 5(4):S. 56–68

[Sne03] Snelick, R.; Indovina, M.; Yen, J. und Mink, A.: Multimodal Biometrics: Issues in Design and Testing, in: *ACM International Conference on Multimodal Interfaces (ICMI)*, Vancouver, British Columbia, Kanada (2003), S. 68–72

[Son12] Sonderforschungsbereich (SFB) 588: Humanoide Roboter – Lernende und kooperierende multimodale Roboter, http://www.sfb588.uni-karlsruhe.de/ (2012)

[Son14] Sony Europe GmbH: Mikrofon Sony ECM-C115, http://www.sony.de/support/de/product/ecm-c115 (2014)

[Sti08] Stiefelhagen, R.; Bernardin, K.; Bowers, R.; Rose, R. T.; Michel, M. und Garofolo, J.: The CLEAR 2007 Evaluation, in: *Multimodal Technologies for Perception of Humans*, Bd. 4625 von *Lecture Notes in Computer Science*, Springer Berlin Heidelberg (2008), S. 3–34

[Swe08a] Swerdlow, A.; Machmer, T.; Kühn, B. und Kroschel, K.: Robust sound source identification for a humanoid robot, in: *Electronic Speech Signal Processing (ESSV)*, Frankfurt, Deutschland (2008)

[Swe08b] Swerdlow, A.; Machmer, T.; Kühn, B. und Kroschel, K.: Speaker Position Estimation in Vehicles by Means of Acoustic Analysis, in: *34. Deutsche Jahrestagung für Akustik (DAGA)*, Dresden, Deutschland (2008)

[Swe09] Swerdlow, A.: *Audiovisuelle Signaturen für eine objektzentrierte Umwelterfassung*, Dissertationsschrift, Karlsruher Institut für Technologie (KIT), Fakultät für Elektrotechnik und Informationstechnik, Karlsruhe (2009)

[Tar09] Tares, T.; Greidanus, H.; Jurquet, G. und Helie, P.: Wide Maritime Area Airborne Surveillance SoS (WiMAAS), in: *IEEE Systems Conference (SysCon)*, Vancouver, British Columbia, Kanada (2009), S. 123–126

[The12] The MathWorks, Inc.: MATrix LABoratory (MATLAB), Version 2012a, http://www.mathworks.com/ (2012)

[Tre64] Treisman, A. M.: Monitoring and storage of irrelevant messages and selective attention. *Journal of Verbal Learning and Verbal Behavior* (1964), Bd. 3:S. 449–459

[Ude03] Ude, A.; Atekeson, C. G. und Cheng, G.: Combining peripheral and foveal humanoid vision to detect, pursue, recognize and act, in: *IEEE International Conference on Intelligent Robots and Systems (IROS)*, Bd. 3, Las Vegas, Nevada, USA (2003), S. 2173–2178

[Vag12] Vagts, H. und Jakoby, A.: Privacy-aware access control for video data in intelligent surveillance systems, in: *SPIE Mobile Multimedia/Image Processing, Security, and Applications*, Baltimore, Maryland, USA (2012), S. 84060I–84060I–14

[Val08] Valenti, R. und Gevers, T.: Accurate eye center location and tracking using isophote curvature, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, Alaska, USA (2008), S. 1–8

[Ver99] Verlinde, P.; Druyts, P.; Chollet, G. und Acheroy, M.: Applying Bayes based classifiers for decision fusion in a multi-modal identity verification system, in: *International Symposium on Pattern Recognition: In Memoriam Pierre Devijver*, Brüssel, Belgien (1999)

[Vio04] Viola, P. und Jones, M.: Robust Real-time Face Detection. *International Journal of Computer Vision (IJCV)* (2004), Bd. 52(2):S. 137–154

[Wal06] Walther, D. und Koch, C.: Modeling attention to salient proto-objects. *Neural Networks* (2006), Bd. 19(9):S. 1395–1407

[Wei08a] Weinmann, Martin: *Entwicklung eines Verfahrens zur Extraktion und Klassifikation relevanter Objekte in Bildern unter Verwendung von Farb- und Tiefeninformationen*, Projektarbeit, Universität Karlsruhe (TH), Karlsruhe (2008)

[Wei08b] Weinmann, Michael: *Ansichtsbasierte Objektklassifikation und Lageschätzung*, Projektarbeit, Universität Karlsruhe (TH), Karlsruhe (2008)

[Wel11] Welke, K.: *Memory-Based Active Visual Search for Humanoid Robots*, Dissertationsschrift, Karlsruher Institut für Technologie (KIT), Fakultät für Elektrotechnik und Informationstechnik, Karlsruhe (2011)

[Wen90] Wentworth, N. und Witryol, S. L.: Information theory and collative motivation: incentive value of uncertainty, variety, and novelty for children. *Genetic, Social, and General Psychology Monographs* (1990), Bd. 116(3):S. 301–322

[Wil14] Wilson, D. W.: Maximum sum of displacements of elements in a permutation of (1..n), `http://oeis.org/A007590` (2014)

[WIM09] WIM^{2}S: Wide Maritime Area Airborne Surveillance, `http://www.wimaas.eu` (2009)

[Xse14] Xsens Technologies B.V.: MTi-10 IMU: Inertial Measurement Unit, `http://www.xsens.com/en/general/mti-10-series` (2014)

[Xu09] Xu, T.; Chenkov, N.; Kühnlenz, K. und Buss, M.: Autonomous Switching of Top-down and Bottom-up Attention Selection for Vision Guided Mobile Robots, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, St. Louis, Missouri, USA (2009), S. 4009–4014

[Yog09] Yogeswaran, M. und Ponnambalam, S. G.: An extensive review of research in swarm robotics, in: *IEEE World Congress on Nature & Biologically Inspired Computing (NaBic)*, Coimbatore, Indien (2009), S. 140–145

Betreute Abschlussarbeiten von Studierenden

[Gür12] Gür, M.: *Segmentierung von 3D-Punktwolken zur Detektion von a-priori unbekannten Gegenständen in einer Szene.* Diplomarbeit, Karlsruher Institut für Technologie (KIT), Karlsruhe, 2012.

[Kol10] Kolka, R.: *Integration einer mobilen Roboterplattform in ein System zur opto-aktustischen Szenenanalyse.* Studienarbeit, Karlsruher Institut für Technologie (KIT), Karlsruhe, 2010.

[Lóp11] López Armendáriz, J.: *Detection of unknown sound sources for learning new object classes.* Abschlussarbeit (Erasmus-Sokrates), Karlsruher Institut für Technologie (KIT), Karlsruhe, 2011.

[Pal11] Palacios Gómez, C.: *Knowledge-based localization and tracking of objects and persons in a scene.* Abschlussarbeit (Erasmus-Sokrates), Karlsruher Institut für Technologie (KIT), Karlsruhe, 2011.

[Pap11] Papantoni, V.: *Akustische Lokalisation von Schallereignissen mit Hilfe eines humanoiden Roboterkopfes.* Diplomarbeit, Karlsruher Institut für Technologie (KIT), Karlsruhe, 2011.

[Ram10] Sharma, R.: *Evaluation of face detection and recogition algorithms in a surveillance system.* Projektarbeit, Karlsruher Institut für Technologie (KIT), Karlsruhe, 2010.

[Tsa09] Tsanga, S.: *Gesichtsidentifikation für ein System zur opto-akustischen Szenenanalyse.* Diplomarbeit, Universität Karlsruhe (TH), Karlsruhe, 2009.

[Wan09] Wang, X.: *Face identification for an opto-acoustic scene analysis system.* Studienarbeit, Universität Karlsruhe (TH), Karlsruhe, 2009.

[Wol12] Wolf, D.: *Automatische Zustandserfassung von Küchengeräten mittels Mikrofonen für einen humanoiden Roboter*. Bachelorarbeit, Karlsruher Institut für Technologie (KIT), Karlsruhe, 2012.

Förderung

Die vorliegende Arbeit entstand im Rahmen des Teilprojekt P2 *interaktive multimodale Exploration* des Sonderforschungsbereichs (SFB) 588 *Humanoide Roboter – Lernende und kooperierende multimodale Roboter* und wurde von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Für die Unterstützung möchte ich mich stellvertretend bei dem Sprecher des Sonderforschungsbereichs, Herrn Prof. Dr.-Ing. Rüdiger Dillmann, und bei Herrn Prof. Dr.-Ing. Jürgen Beyerer, der dem Vorstand angehört und meine Arbeit mitbetreut hat, recht herzlich bedanken.