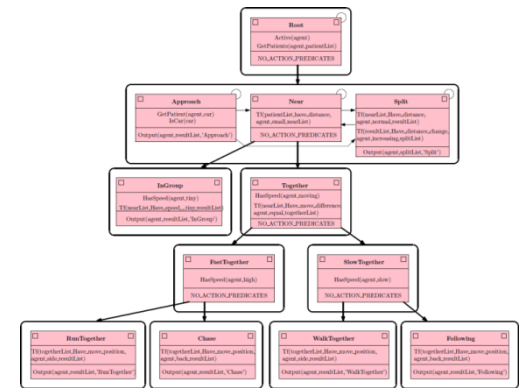
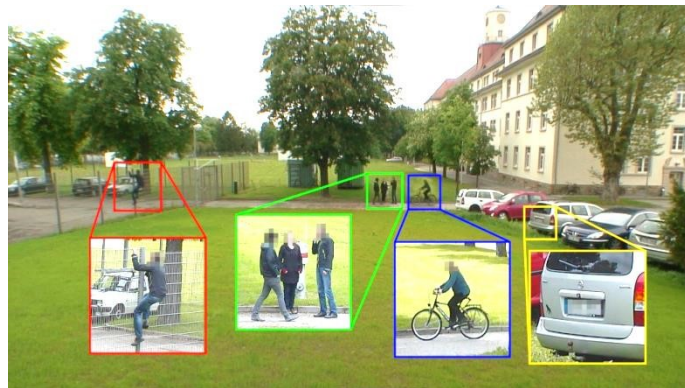
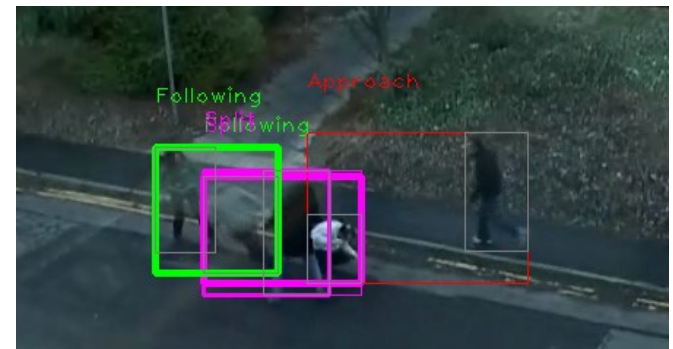


BEGRIFFLICHE SITUATIONSANALYSE BEI VIDEOÜBERWACHUNG IM PERSONENKONTEXT

David Münch, Fraunhofer IOSB

14.07.2015



13. Vorlesung: Verhaltensbasierte Bewegungsprädiktion

“Drei Nutzungsformen für dasselbe Verhaltenswissen:

1. Begriffliche Bildfolgenauswertung
2. Bildfolgenernerierung
3. Rückkopplung”

“Ein Formalismus für Auswertung/Generierung/Rückkopplung.”

“Quantitative Bildauswertung wird durch qualitatives Wissen stabilisiert.”

“Traversierung eines Situationsgraphen ergibt eine Situation für jeden Zeitpunkt und jedes Fahrzeug.”

Motivation

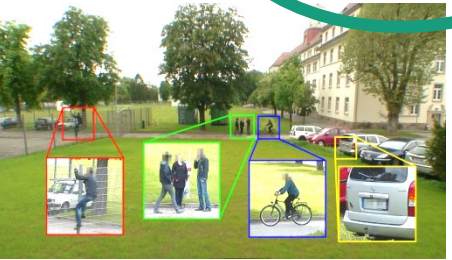
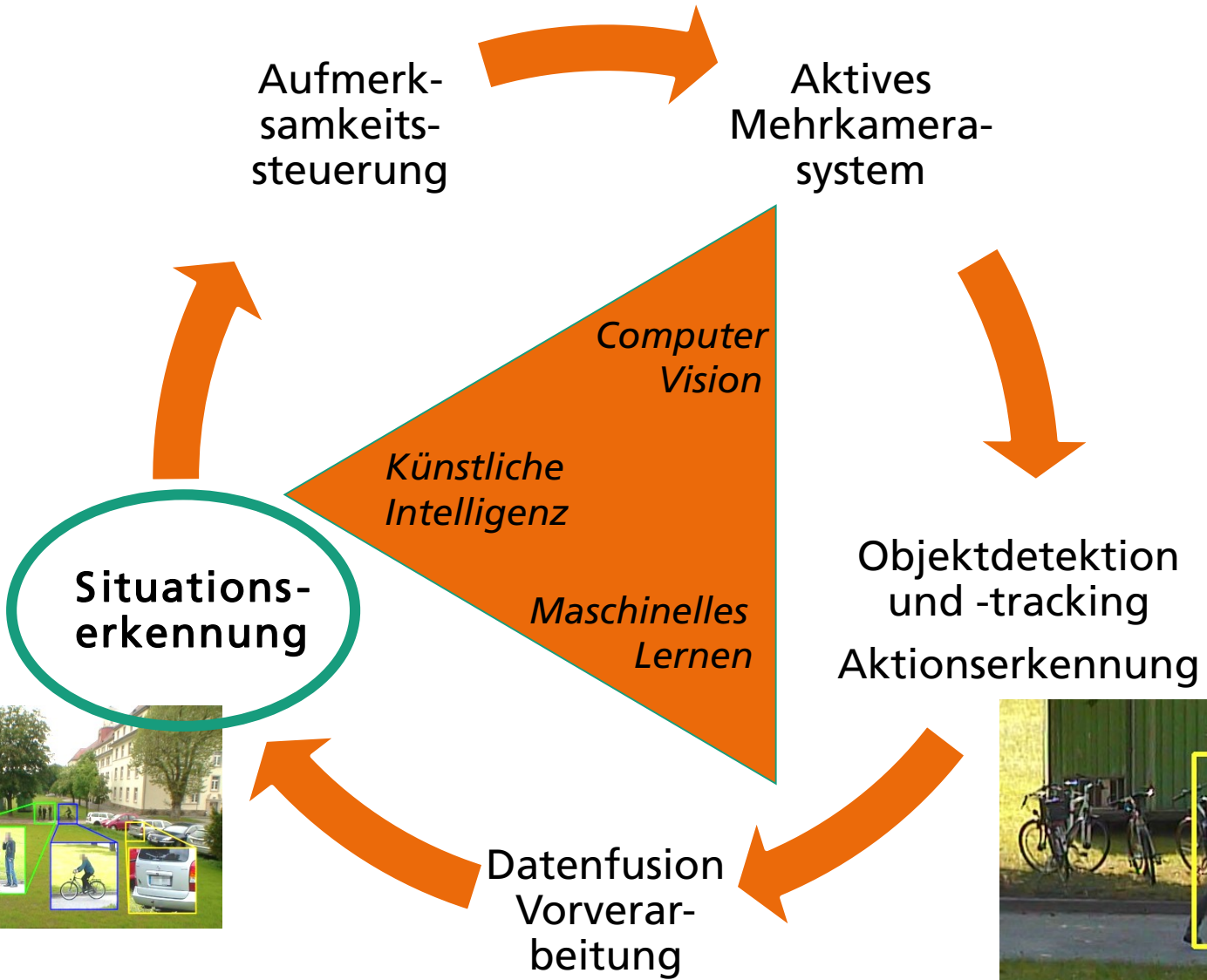
Automatische Video- basierte Situations- erkennung

- Videoüberwachung
 - Detektion von Gefahrensituationen
 - Bedienerunterstützung
 - Ereignisprotokoll
- Semantische Suche in Videos



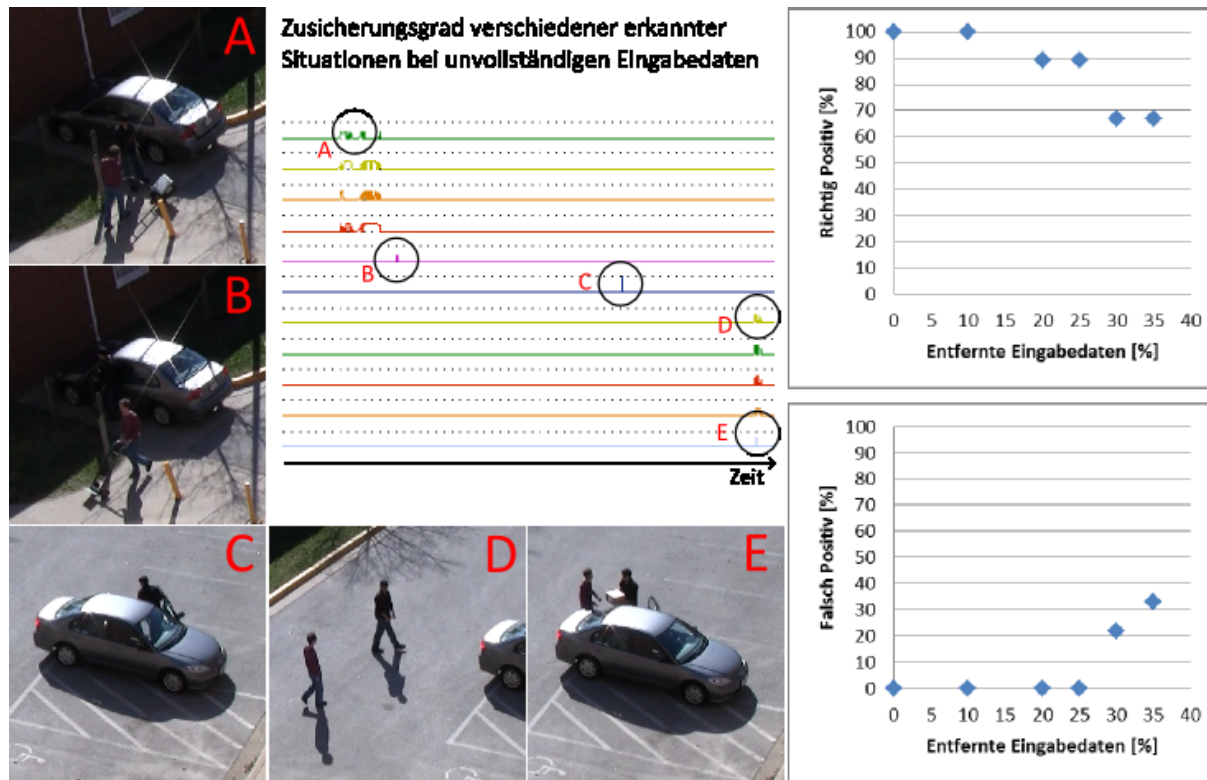
Weitere Entwicklungen des Kognitiven Sichtsystems

- Übertragung der Situationserkennung auf den Diskursbereich Videoüberwachung im Personenkontext.
- Erweiterung des Formalismus um
 - mehrere konkurrierende Hypothesen,
 - Propagierung der Unschärfe durch den gesamten Inferenzprozess.
- Umgang mit unvollständigen und fehlerhaften Daten.
- Komplexitätsreduktion.
- Anwendungsbereich Videoüberwachung:
 - Detektion von Gefahrensituationen
 - Bedienerunterstützung
 - Ereignisprotokoll



Übertragung auf den Diskursbereich Videoüberwachung

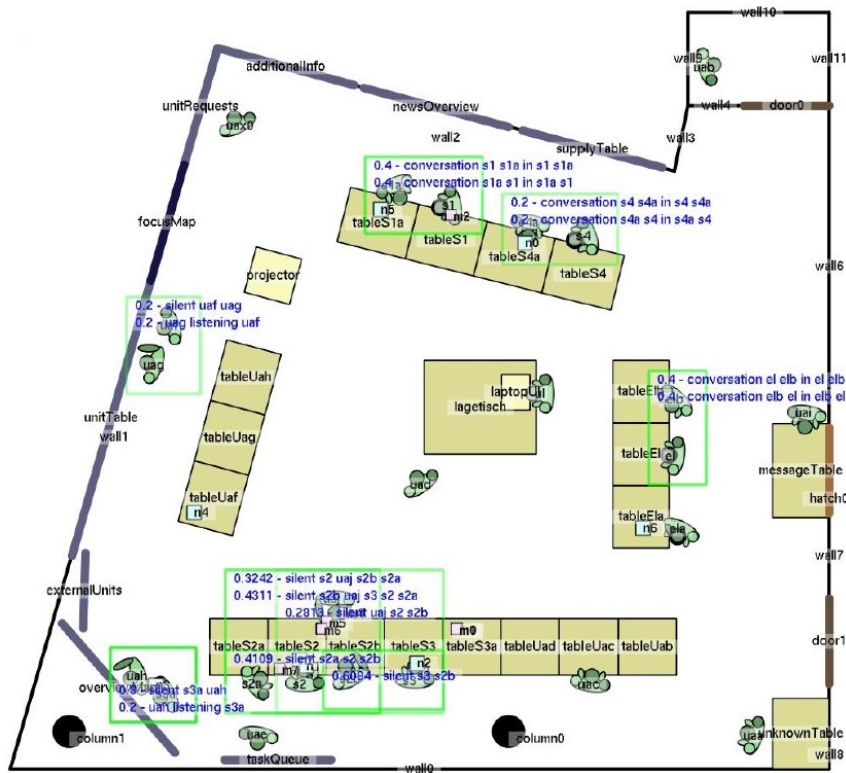
Untersuchungen auf künstlichen (perfekten) Daten oder manuell annotierten realen Daten waren erfolgreich und vielversprechend.



D. Münch, J. IJsselmuide, A.-K. Grosselfinger, M. Arens, R. Stiefelhagen. **Rule-Based High-Level Situation Recognition from Incomplete Tracking Data.** Proc. of the 6th International Symposium on Rules: Research Based and Industry Focused, Montpellier, France, 2012.

Übertragung auf den Diskursbereich Videoüberwachung

Untersuchungen auf künstlichen (perfekten) Daten oder manuell annotierten realen Daten waren erfolgreich und vielversprechend.



J. IJsselmuiden, A.-K. Grosselfinger, D. Münch, M. Arens, R. Stiefelhagen. **Automatic Behavior Understanding in Crisis Response Control Rooms**. Proc. of International Joint Conference on Ambient Intelligence, Pisa, Italy, 2012.

Probleme

Reale Daten sind:

- fehlerbehaftet,
- unvollständig und
- verrauscht.

Komplexität/
Gruppen von
Agenten.

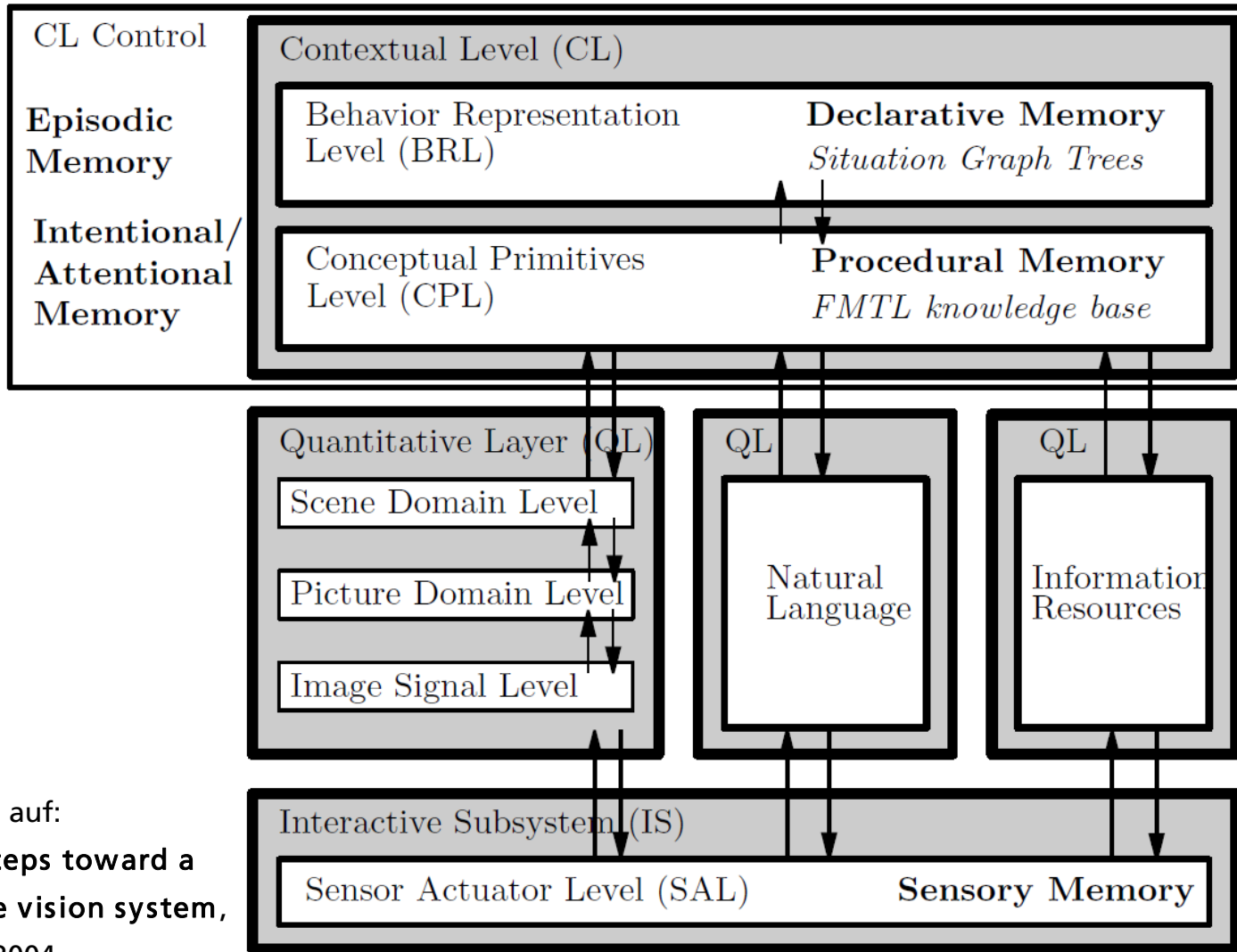
Selektive Informations-
beschaffung.

Schritthaltende Verarbeitung
bei komplexen Situationen.

Mehrere Situationen pro
Agent pro Zeiteinheit.

→ Die Situationserkennung muss explizit mit
diesen Effekten umgehen können.

Cognitive Vision System



Basierend auf:
**Nagel, Steps toward a
 cognitive vision system,**
 AI Mag., 2004.

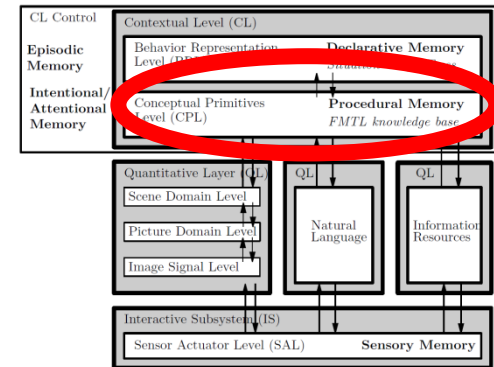
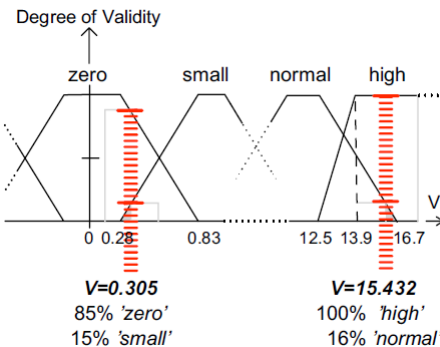
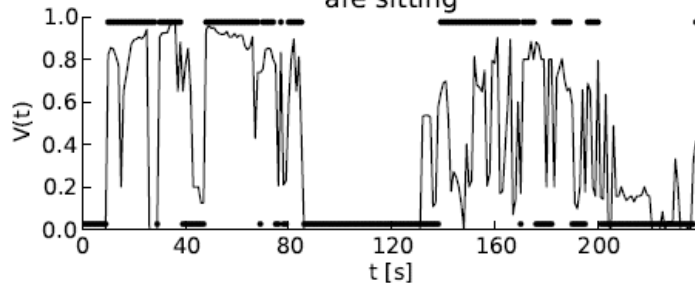
Begriffliches Teilsystem – Begrifflich-Primitive-Ebene

- Quantitative Information → Begriffliches Basiswissen (FMTL Prädikate)
- Ein Satz von Basisregeln
- Bereich: räumliche und zeitliche Relationen in kurzen Zeitintervallen
- Weitestgehend Diskursbereichs-unabhängig
- Unterstützung von Unsicherheit
- Inferenzmaschine: F-LIMETTE

$$\square \{ [\diamond_{-2} \textit{Distance}(p, q, d_{-2}) \wedge \diamond_{-1} \textit{Distance}(p, q, d_{-1}) \wedge \textit{Distance}(p, q, d_0) \wedge \diamond_1 \textit{Distance}(p, q, d_1) \wedge \diamond_2 \textit{Distance}(p, q, d_2) \wedge \textit{Derivative}(d_{-2}, d_{-1}, d_0, d_1, d_2, d') \wedge \textit{DistanceChangeCategory}(d', c)] \rightarrow \textit{DistanceChange}(p, q, c) \}$$

$$\square \{ \forall p [\textit{HasType}(p, \textit{agent}) \rightarrow \exists q [\textit{HasType}(q, \textit{chair}) \wedge \textit{AtSeat}(p, q) \rightarrow \textit{EverybodyAtSeat}()]] \}$$

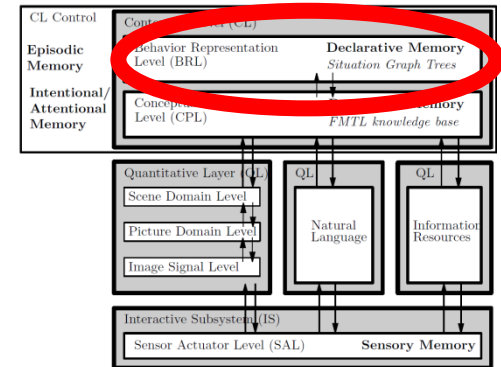
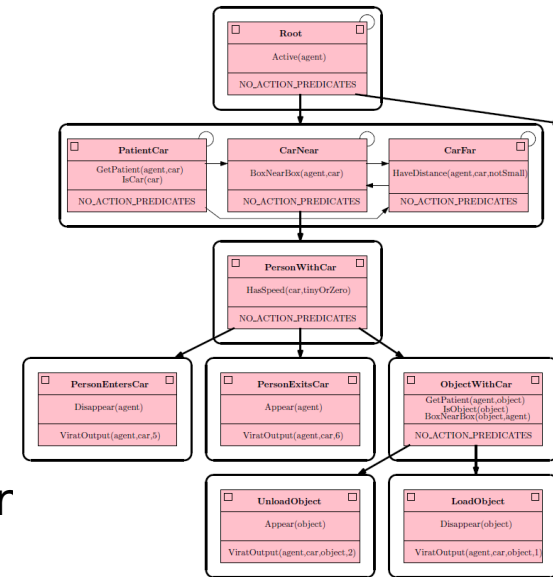
s4 s4a from Group s4 s4a are sitting



Begriffliches Teilsystem – Verhaltens-Repräsentations-Ebene

Wie werden Situationen repräsentiert?

- Wissensrepräsentation in Situationsgraphenbäumen
 - Graphisch editierbar
 - Einfach zu erweitern und wiederzuverwender
 - Interpretierbar (va. Blackbox)
 - Neuer Editor: SGTyEditor
- Erschöpfende parallele Situationserkennung:
 - Ein Thread pro Agent
 - Wissensaustausch zwischen den Threads
 - Watchdog beendet zu lange Anfragen



Begriffliches Teilsystem – Verhaltens-Repräsentations-Ebene

Wie man einen SGT benutzt: 7 Gestaltungskriterien (GK)

GK1: reine Verhaltensbeschreibung

GK2: klar strukturiert

GK3: temporale Kanten für zeitliche Nachfolge

GK4: Untergraphen modellieren den beobachtbaren Übergang von Start- zu Zielsituation

GK5: sich gegenseitig ausschließende Situationen in unterschiedlichen Untergraphen

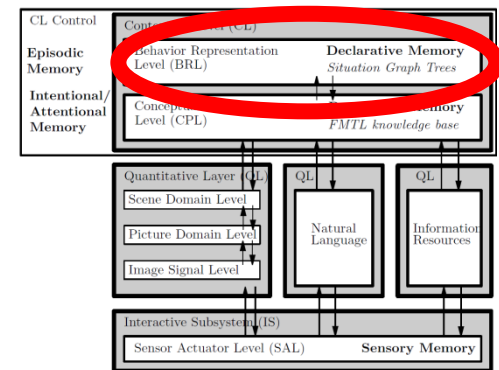
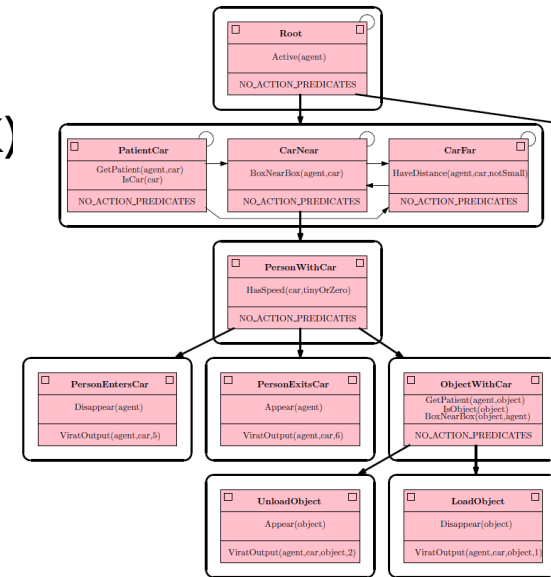
GK6: Situationen beschreiben Verhalten von Agent

GK7: Unabhängig von konkreter Szene

Basierend auf:

Harlad, Nutzung logikbasierter Verhaltensrepräsentation zur natürlichsprachlichen Beschreibung von Videos,

Dissertation, KIT, 2011.

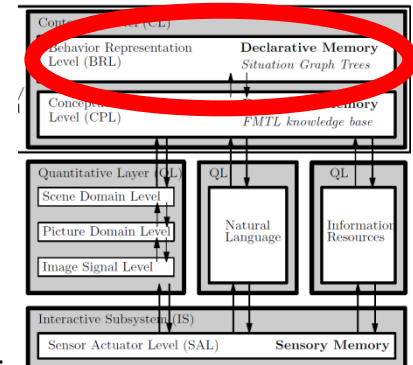


Erschöpfende parallele Situationserkennung

Algorithm 1: Fuzzy Situation Graph Tree Traversal

Input: SGT, *object*

```
1 if object occurs for the first time then
2    $G \leftarrow$  SGT root graph;
3   forall the  $s | s \in G \wedge s$  is start situation do
4     if  $s$  can be instantiated then
5       forall the  $spec | spec \in s \wedge spec$  is specialization do
6          $s := spec, G :=$  graph containing  $spec$ ;
7         start recursion goto line 3;
8       evaluate actions of  $s$ ;
9   else
10    forall the  $predSit | predSit$  is prediction situation of  $s$  (all the last situations
11    of the already known object) do
12      if instantiate  $predSit$  successful then
13         $s := predSit, G :=$  graph containing  $predSit$ ;
14        start recursion goto line 5;
15      else
16        if  $predSit$  is instantiated end situation  $\wedge predSit \in G$  then
17          instantiation successful;
18        else
19          instantiation failed;
```



Eigenschaften

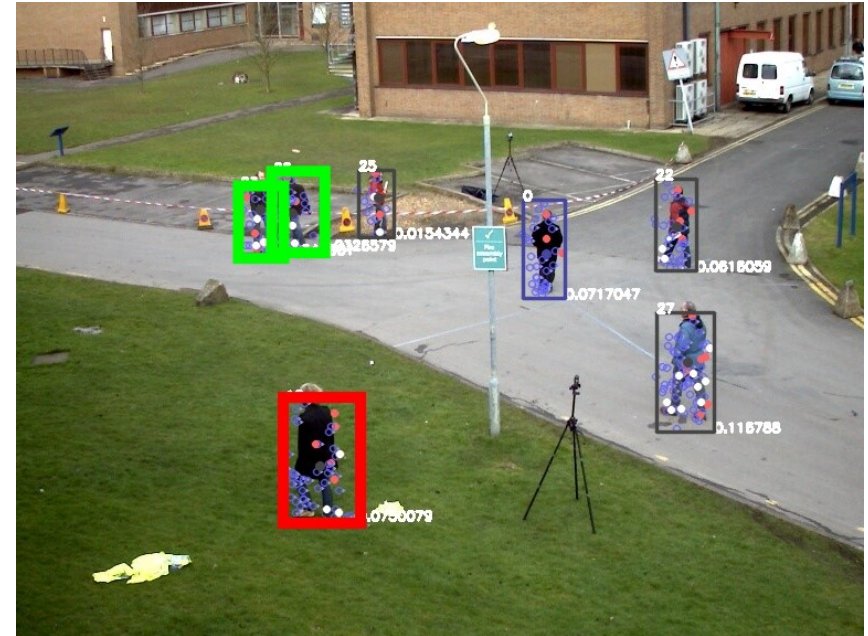
- Trennung des Wissens auf verschiedene Ebenen
- Diskursbereichsunabhängig
- Situationsmodellierung von Hand
- Gut geeignet für komplexe Situationen mit zeitlichen Abfolgen, zu denen es schwierig ist Trainingsdaten zu beschaffen
- Sprache ist einfach für Menschen verständlich
- Warum wird eine Situation erkannt? Beweisbaum!
- Semantische Information durch das ganze System hindurch
- Parallelisierbar
- Unsicherheit wird durch das ganze System propagiert

Evaluation

Evaluation von verschiedenen Situationen auf dem PETS2009 Datensatz.

Ausgabe für den 669. Frame:

- 1 | 669 ! have_speed(obj_0, normal).
- 1 | 669 ! have_speed(obj_19, normal).
- 1 | 669 ! in_illegal_zone(obj_19).
- 1 | 669 ! have_speed(obj_22, normal).
- 1 | 669 ! have_speed(obj_23, normal).
- 1 | 669 ! close_to(obj_23,obj_26).
- 1 | 669 ! pass_by(obj_23,obj_26).
- 0.673 | 669 ! have_speed(obj_25, normal).
- 0.327 | 669 ! have_speed(obj_25,high).
- 1 | 669 ! have_speed(obj_26, normal).
- 1 | 669 ! close_to(obj_26,obj_23).
- 1 | 669 ! pass_by(obj_26,obj_23).
- 1 | 669 ! have_speed(obj_27, normal).



PETS2009, Frame 669. Zwei Situation erkannt: *in_illegal_zone(obj_19)* und *pass_by(obj_23, obj_26)*.

Evaluation – CAVIAR 1st

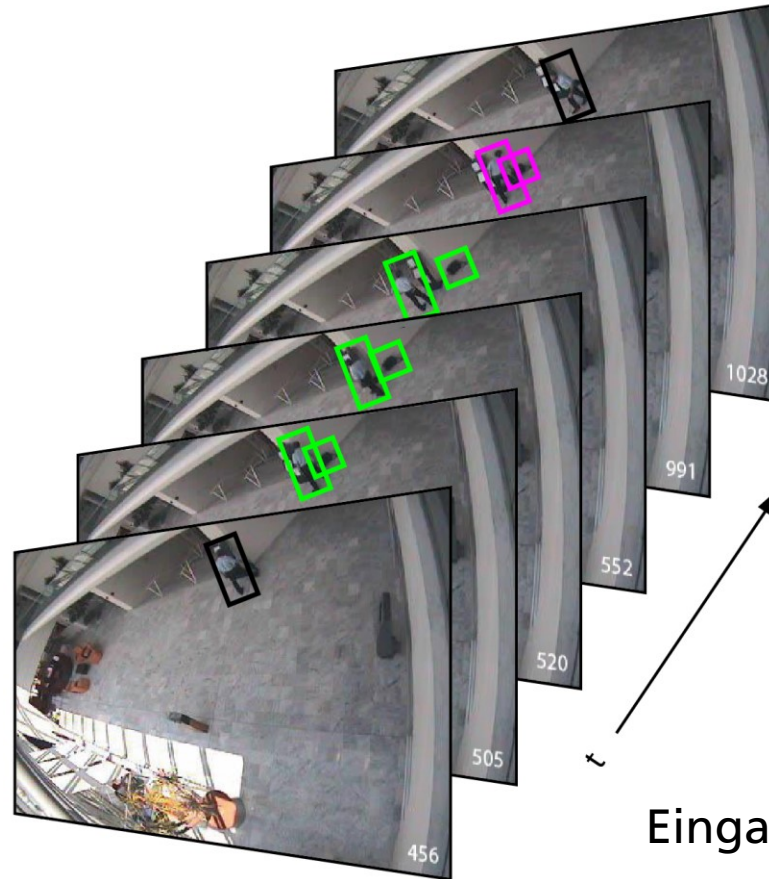


Zeit	Truth value	Ausgabe
350	0.550	walkTogether(a1,a2)
390	0.552	standTogether(a1,a2)
402	1.000	splitUp(a1,a2)
410	1.000	standTogether(a1,a3)
441	1.000	walkTogether(a1,a3)



Eingabe: Grundwahrheit.

Evaluation – CAVIAR 1st



Eingabe: Grundwahrheit.

Evaluation – CAVIAR 1st

Zeit	Truth value	Ausgabe
456	1.000	agent(A)
504	0.735	hasType(C,bag)
505	0.242	distanceChange(A,C,increasing)
505	0.242	distance(A,C,notSmall)
519	0.154	hasType(C,bag)
520	1.000	distanceChange(A,C,increasing)
520	0.564	distance(A,C,notSmall)
552	1.000	distanceChange(A,C,increasing)
552	1.000	distance(A,C,notSmall)
991	0.627	hasType(C,bag)
1028	1.000	agent(A)

Erweiterung der Basisregeln und des Inferenzprozesses

- Bisher waren Relationen zwischen Agenten binär
- Exponentielle Explosion der Relation, wenn viele Agenten



Einführung von
listenbasierten Regeln

- Erweiterte Regeln und Inferenzprozess durch Filter ergänzt, die Prädikate auf eine ganze Liste von Agenten etc. anwenden

Erweiterung der Basisregeln und des Inferenzprozesses

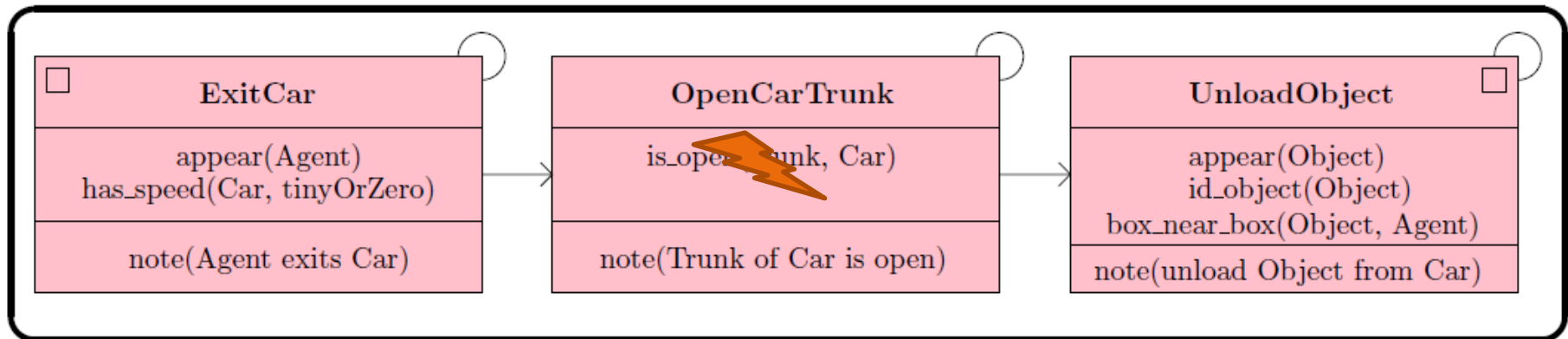
$\square \{ \mathbf{Truefilter}(in, Fun, agent, parameter, res) \leftarrow$
 $\mathbf{Truefilter_}(in, Fun, agent, parameter, res) \wedge res \langle \rangle [] \}$

$\square \{ \mathbf{Truefilter_}([elem|in], Fun, agent, parameter, res) \leftarrow$
 $functor = ..[Fun, agent, elem, parameter]$
 $\wedge [(call(functor) \wedge res = [elem|new] \wedge !) \vee res = new]$
 $\wedge \mathbf{Truefilter_}(in, Fun, agent, parameter, new) \}$

$\square \mathbf{Truefilter_}([], -, -, -, [])$

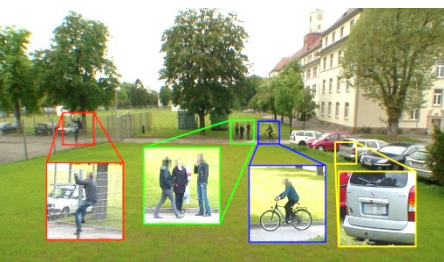
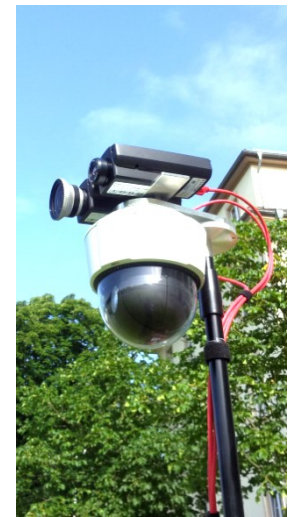
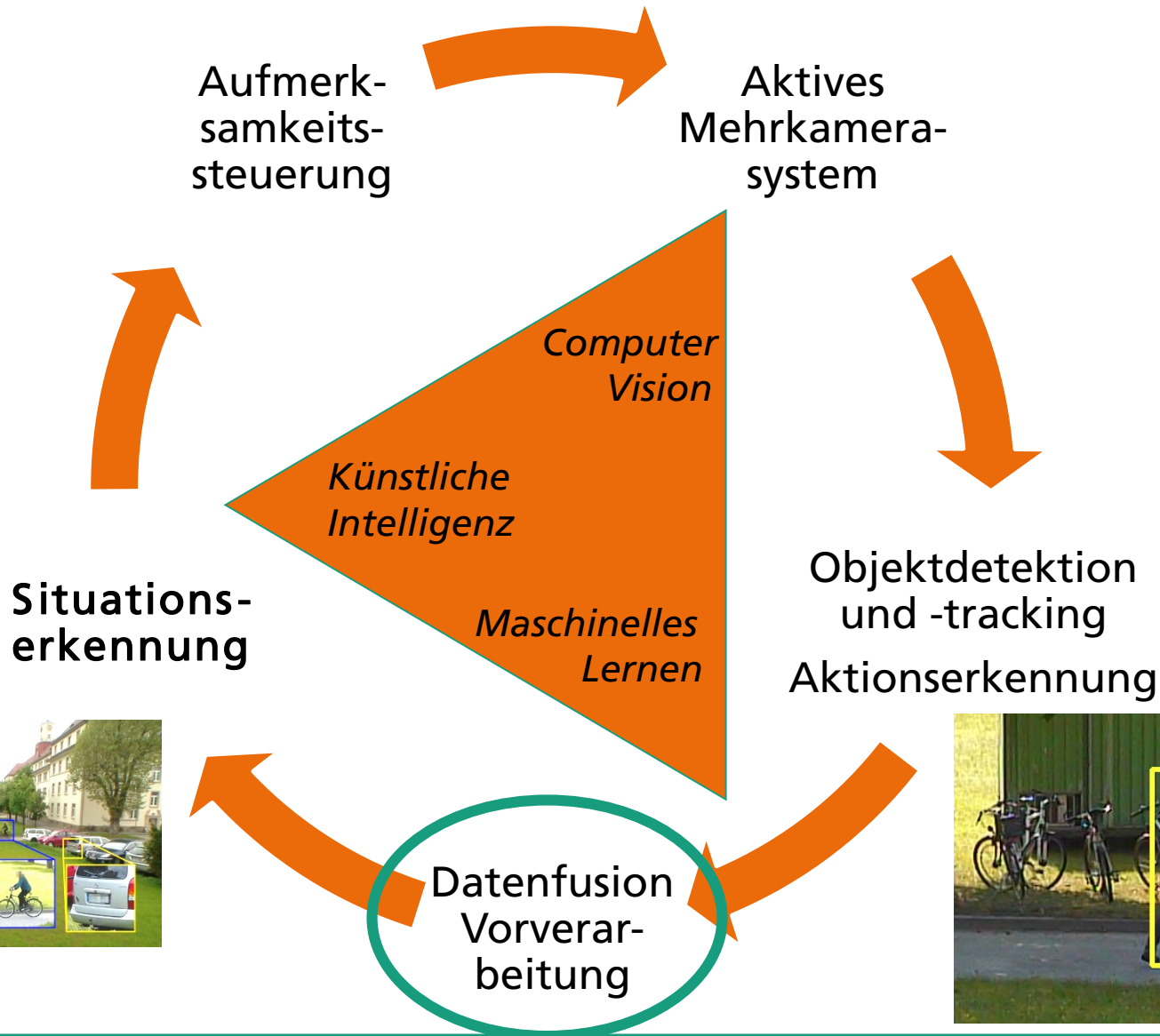
Behandlung unvollständiger Daten - Verhaltens-Repräsentations-Ebene

- Halluzination (Abduktion) fehlender Information.



Was passiert wenn `is_open(trunk, Car)` nicht abgeleitet werden kann?

→ Halluziniere `is_open(trunk, Car)` und mache weiter!



Zweigeteilter Inferenzprozess zur Komplexitätsreduktion

Unterstützung der Basis-Begrifflichen Ebene durch nicht-deklarative Methoden um die kombinatorische Komplexität und Redundanz zu reduzieren.

1. Clustering von einfachen Basisprädikaten.
2. Anwendung der Situationserkennung auf den geclusterten Ergebnissen.

Basisregel: Zeitlicher und räumlicher Abstand.



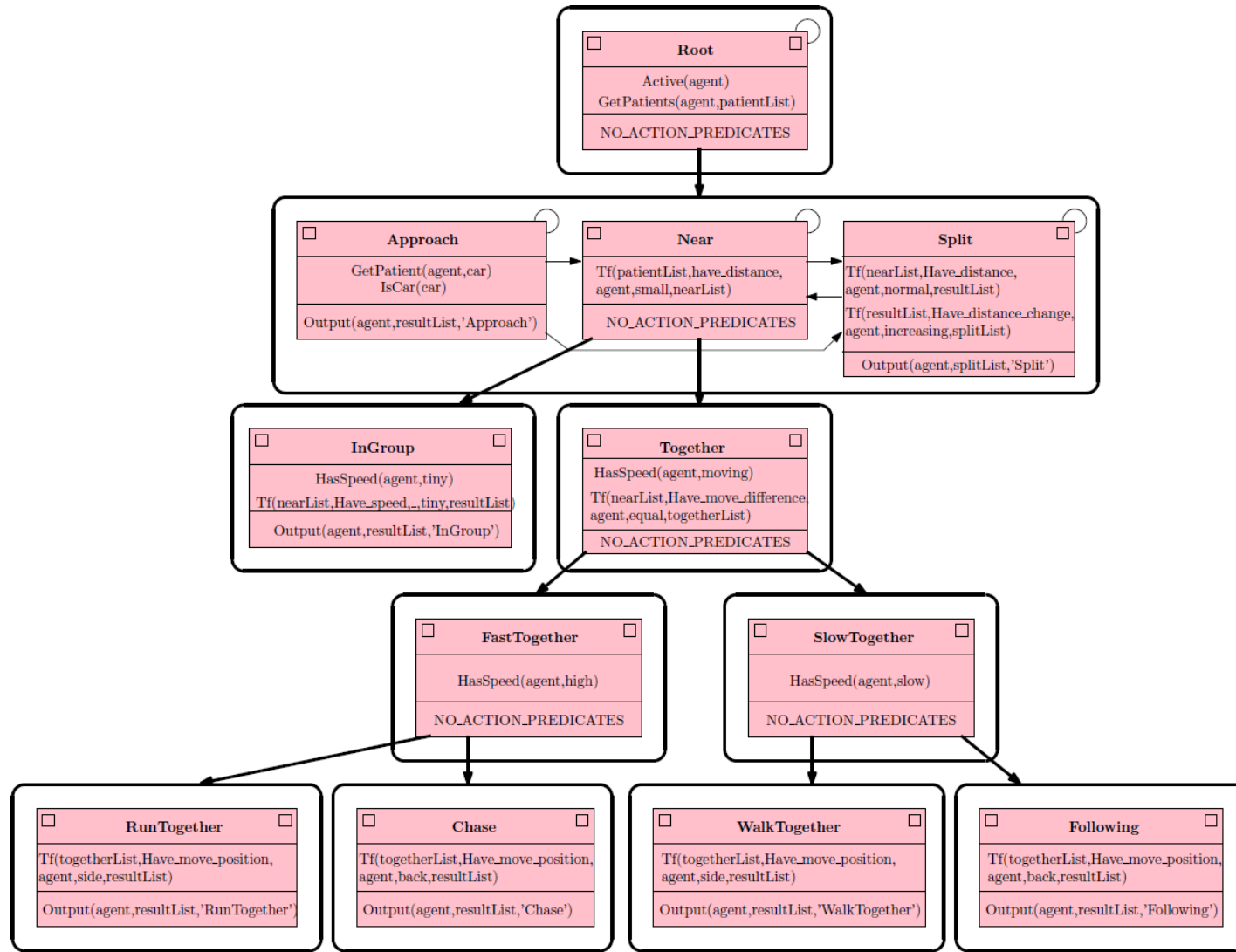
Evaluation mit dem BEHAVE Datensatz

Situationen von Interesse des BEHAVE Datensatzes:

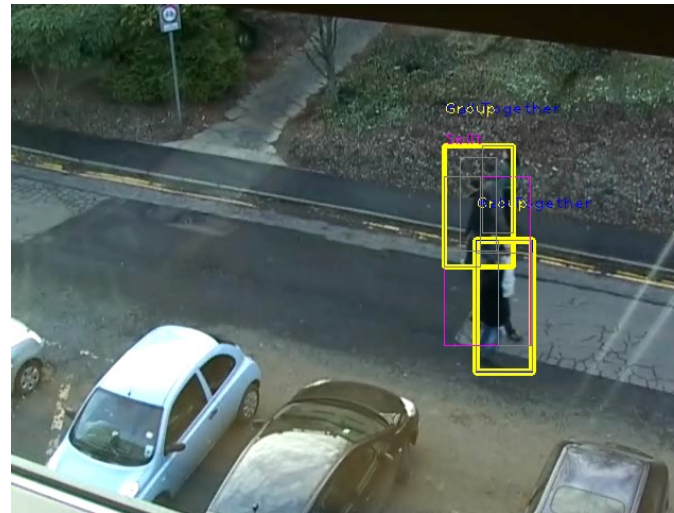
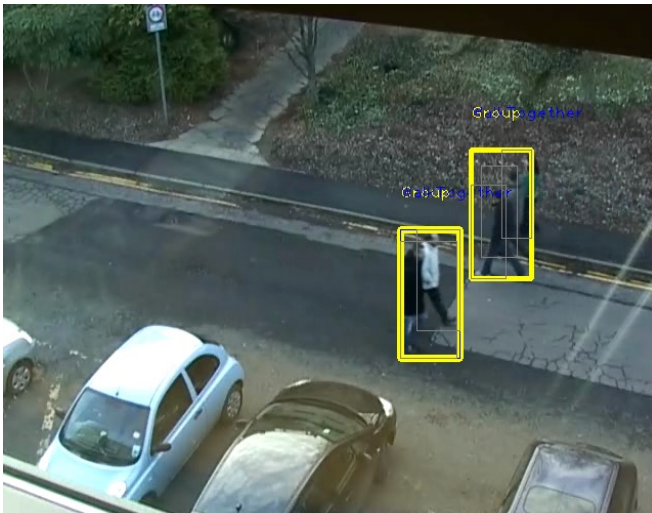
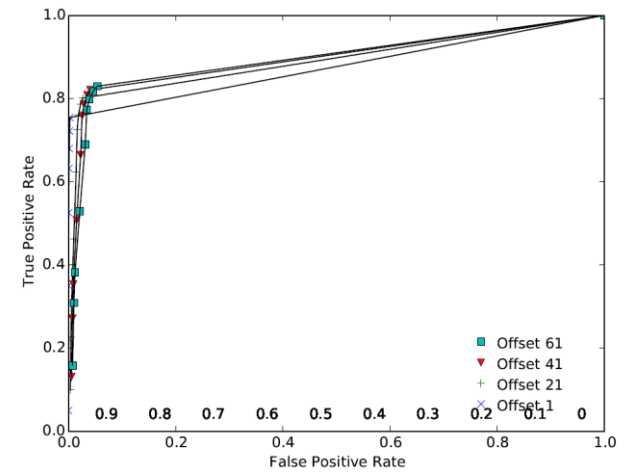
- In group
- Approach
- Walk together
- Meet
- Split
- Ignore
- Chase
- Fight
- Run together
- Following

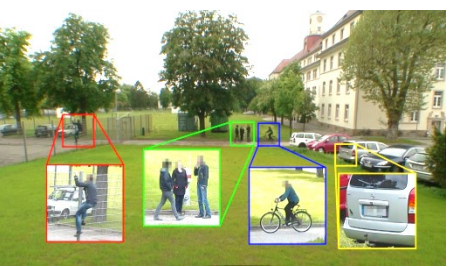
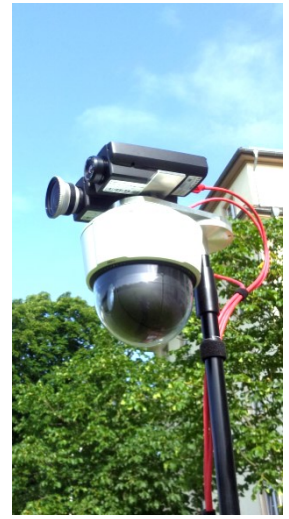
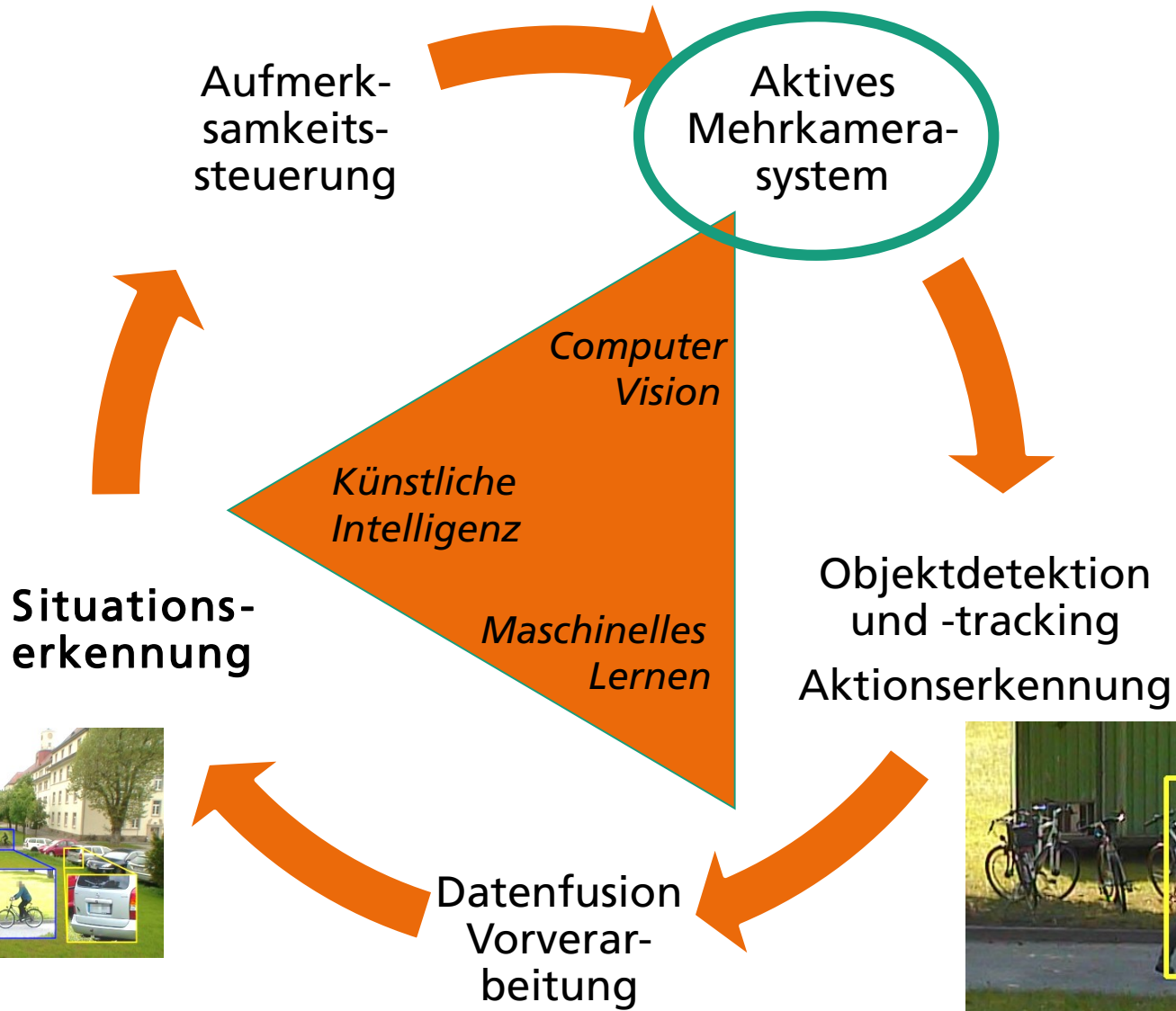
Blunsden, S., Fisher, R.: The behave video dataset: ground truthed video for multiperson behavior classification. *Annals of the BMVA* 2010(4), 1-12 (2010).

Situationsgraphenbaum für BEHAVE Datensatz

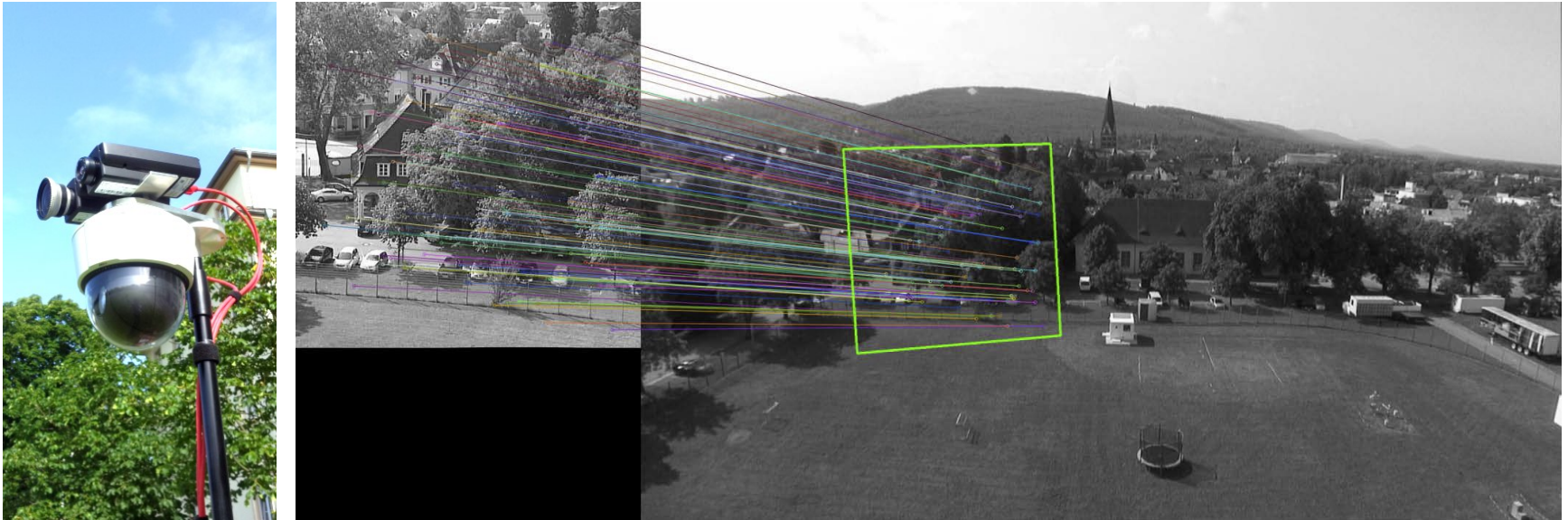


Auswertung auf BEHAVE Datensatz





Active multi-camera system

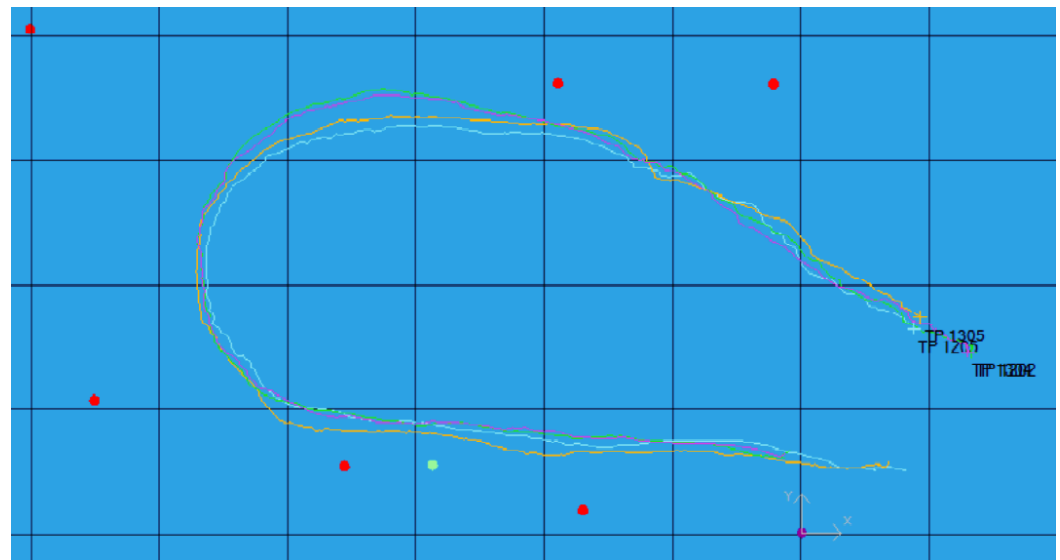


D. Münch, A.-K. Grosselfinger, W. Hübner, M. Arens.
Automatic Unconstrained Online Configuration of a Master-Slave Camera System.
Proc. of the 9th International Conference on Computer Vision Systems, St. Petersburg, Russia, 2013.

Selektive Informationsgewinnung

- “Semantische Rückkopplung”
- Es kann nicht die gesamte Welt im größten Detail beobachtet werden.
- Beispiel: Person kann auf Übersichtsbild erkannt werden. Um die Person dann aber zu identifizieren werden gezielt hochaufgelöste Ansichten der Person aufgenommen. (vgl. Cognitive Visual Tracking and Camera Control, Bellotto, 2012)
- Beispiel: Bei der Verfeinerung von Situationen stehen verschiedene Möglichkeiten/Verfahren/Sensoren/etc. mit unterschiedlicher Konfidenz zur Informationsbeschaffung zur Verfügung.
- Beispiel: Algorithmen werden mit verschiedenen Parametern initialisiert, je nach Anwendungsort in der aktuell beobachteten Szene.

IOSB_VCA_Datensatz

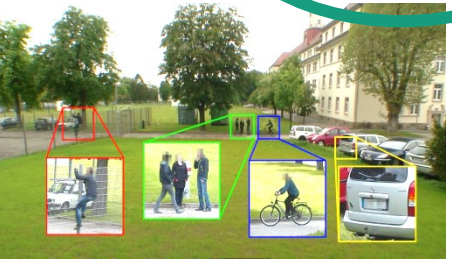
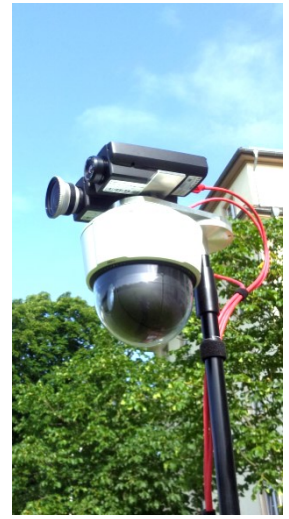
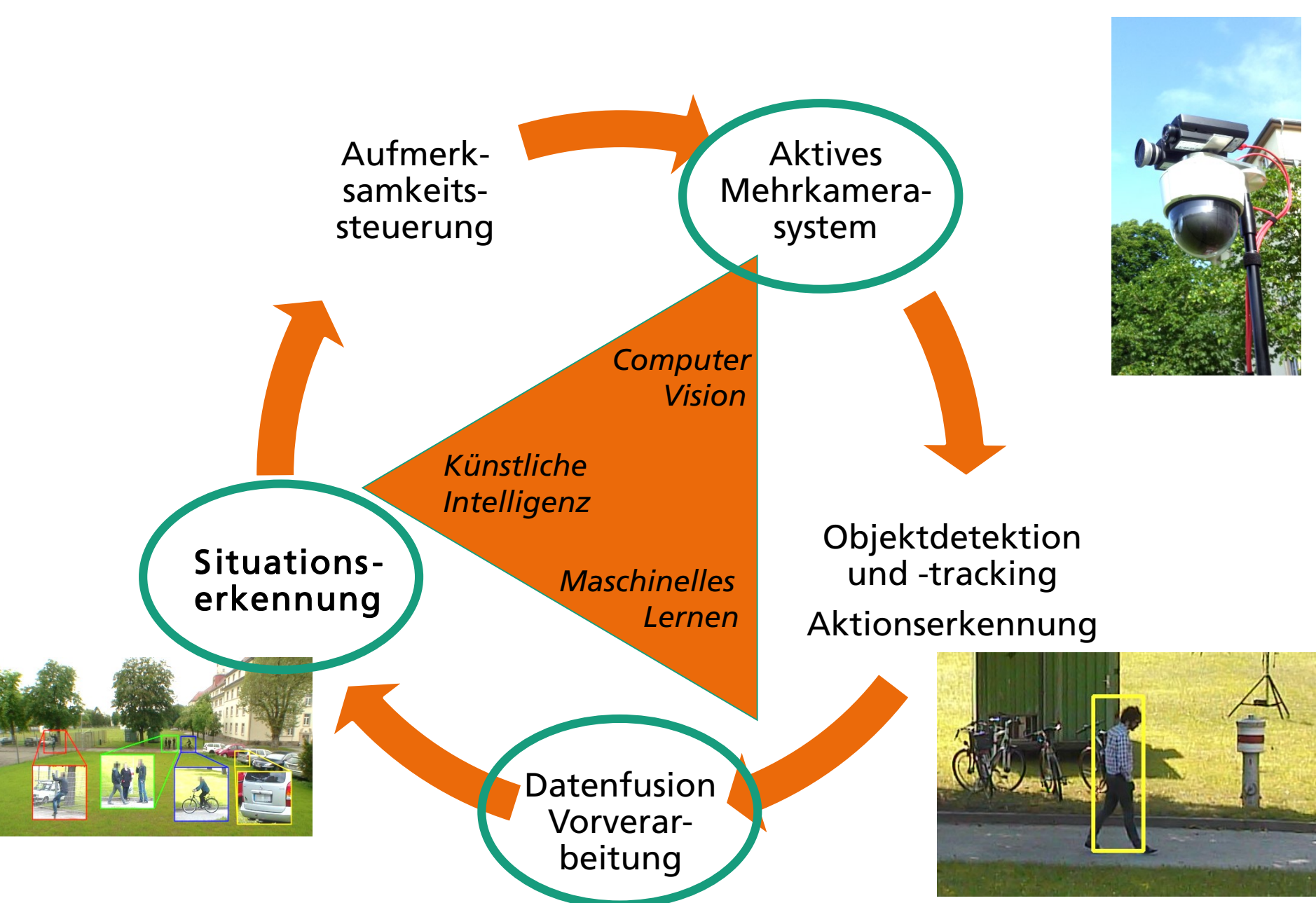


20 verschiedene Szenen:

- Einzel,
- Paarweise,
- Gruppen

Zentimetergenaue
Grundwahrheit der
Personen aufgezeichnet.

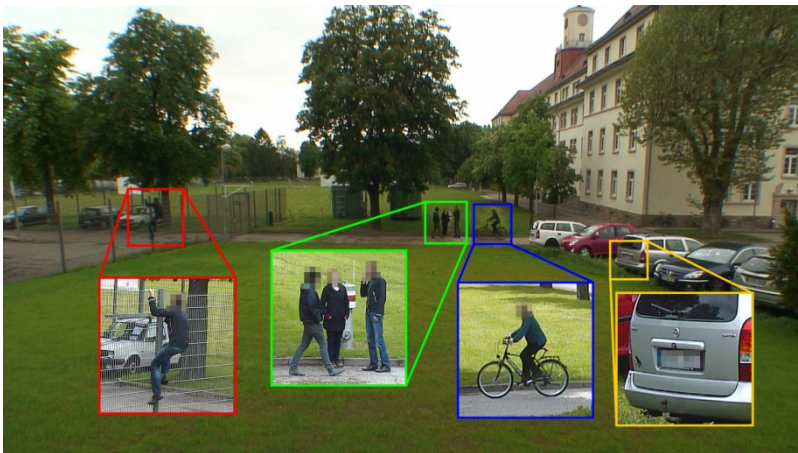
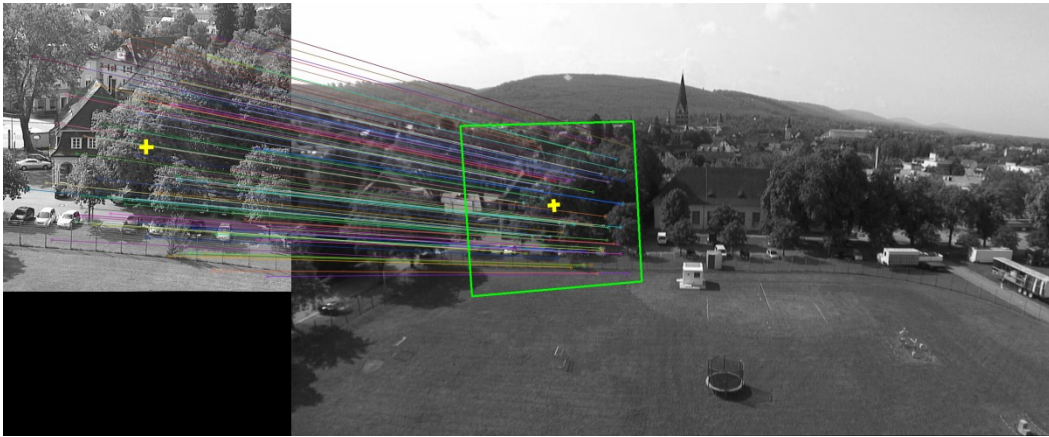




Weitere Entwicklungen des Kognitiven Sichtsystems

- Ist der unser Ansatz für Situationserkennung im Personenkontext geeignet? ✓
- Funktioniert unser Ansatz auf "perfekten" Daten? ✓
- Können mehrere mögliche Ableitungen gleichzeitig betrachtet werden? ✓
- Kann die Unsicherheit durch die Inferenz propagiert werden? ✓
- Kann mit mehreren Agenten umgegangen werden? ✓
- Ist eine schritthaltende Verarbeitung möglich? ✓
- Selektive Informationsbeschaffung? ✓

Masterarbeit beim Fraunhofer IOSB in Ettlingen: Multikamerasysteme im Bereich Objekterkennung – Inhaltsbezogene Videoanalyse



Kontakt:

Ann-Kristin Grosselfinger

ann-kristin.grosselfinger@iosb.fraunhofer.de

<https://recruiting.fraunhofer.de/Vacancies/22500/Description/1>