Kooperative Bildverarbeitung und Lokalisierung in einem Team von Robotern

Norbert Oswald, Reinhard Lafrenz und Paul Levi
Universität Stuttgart,
Institut für parallele und verteilte Höchstleistungsrechner,
Praktische Informatik - Bildverstehen

1 Motivation und Einleitung

In der Bildverarbeitung lag der Schwerpunkt in den vergangenen Jahren auf der Anwendung des Paradigmas des aktiven Sehens [Baj88] [AWB87]. Das aktive Sehen umfaßt die Analyse von Modellierungs- und Steuerungsstrategien für die Wahrnehmung, d.h. Modellierung der Sensoren, Objekte, Umgebung und Interaktion für einen vorgegebenen Zweck, z.B. zur Manipulation, Bewegung oder Erkennung. Dieses Paradigma beschreibt jedoch ausnahmslos die Fähigkeiten eines einzelnen Beobachters. Gerade in natürlichen Umgebungen hängt die Qualität der Bildverarbeitungsergebnisse nicht zuletzt von der Aufnahmeposition ab, die auch mit dem aktiven Ansatz nicht beliebig „wählbar“ ist, insbesondere dann nicht, wenn sich das beobachtete Objekt bewegt. Durch die Verteilung der Sichtaufgabe auf mehrere räumlich getrennte Beobachter eines Teams kann die Zuverlässigkeitsrate bei der Erkennung gesteigert werden. Die Vorteile solch eines verteilten Sehens sind:

- Nutzung von potentiellem Wissen des Teams
- Simultane Analyse von Szenen mit statischen oder dynamischen Objekten
- Generierung robuster Hypothesen

Verteilt berechnete Daten lassen sich jedoch im allgemeinen nur dann kombinieren, wenn sie in Bezug zueinander gesetzt werden können. Das bedeutet, daß für eine kooperative Bildverarbeitung jedem Teammitglied seine absolute Position bekannt sein muß. Die Fähigkeit zur Selbstlokalisierung bildet damit eine Voraussetzung zur visuellen Kooperation. In der vorliegenden Arbeit wird zunächst untersucht, wie und in welcher Form visuelle Kooperation in einem Team von Beobachtern, z.B. Robotern, stattfinden kann. Nach einer kurzen Einführung in die Problemstellung der Selbstlokalisierung werden Anwendungsbeispiele mit visueller Kooperation aus der high-level Bildverarbeitung vorgestellt.

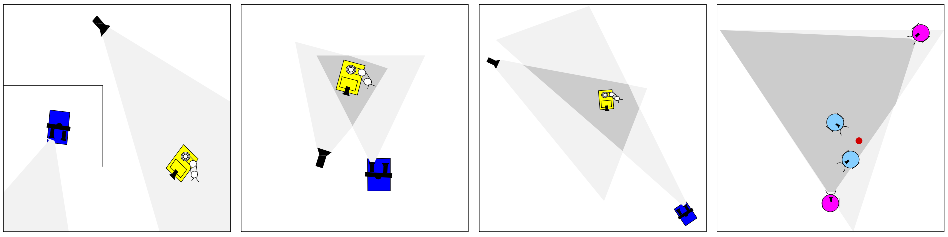
2 Kooperationsebenen für die Bildverarbeitung

Wie schon in der Einleitung erwähnt, sind an einer visuellen Kooperation in der Regel mehrere Beobachter beteiligt. Diese Art der Kooperation bezeichnen wir als *externe* Kooperation. Charakteristisches Merkmal der externen Kooperation ist das verteilte Sehen, also die simultane visuelle Wahrnehmung aus räumlich unterschiedlichen Positionen. Kooperation kann aber auch ausschließlich innerhalb eines Beobachters auftreten. Eine solche Form der Kooperation bezeichnen wir als *interne* Kooperation. Hier kooperieren an Stelle mehrerer Beobachter mehrere verschiedene Verfahren miteinander. Unabhängig von der Art der visuellen Kooperation wird mit Kooperation stets das Ziel verfolgt, eine Steigerung der Qualität der Ergebnisse zu erzielen oder diese überhaupt zu ermöglichen. Für visuelle Kooperation bzw. kooperative Bildverarbeitung müssen folgende Voraussetzungen gegeben sein:

- an der Ergebnisfindung sind mehrere Beobachter (verteiltes Sehen/ extern) oder verschiedene Verfahren eines Beobachters (intern) beteiligt,
- zwischen den Beteiligten besteht die Möglichkeit zur Kommunikation,
- es existieren Mechanismen zur Kooperation und
- Methoden zur Fusionierung

Die Fusionsmethoden hängen davon ab, ob bei der Kooperation zwischen Beobachtern eine *gemeinsame Sicht* auf ein Objekt oder eine Szene existiert oder nicht (Abb. 1). Gibt es keine gemeinsame Sicht, handelt es sich um komplementäre Daten. Dann wird lediglich beliebige Information adaptiert, die von einem anderen Beobachter stammt. Das ist die einfachste Form der Fusion. Im Falle einer gemeinsamen Sicht sind die Daten partiell redundant und sollen zu einer kombinierten Sicht verwendet werden. Hier sind komplexere Fusionsmethoden erforderlich. Die Wahl der geeigneten Fusionsmethode hängt jedoch stark vom Abstraktionsgrad der ausgetauschten Daten ab. Es liegt daher nahe, einen Bezug zwischen der Kooperation und dem Schichtenmodell in der Bildverarbeitung [Rad93] herzustellen.

Im Schichtenmodell bestehend aus Rohbildebene, Merkmalsebene, Objektebene und



(a) (b) (c) (d)

Abbildung 1: Kooperation ohne gemeinsame Sicht (a) und mit gemeinsamer Sicht auf Merkmalsebene (b), Objektebene (c) und Szenenebene (d)

Szenenebene existieren unterschiedliche Voraussetzungen, Aufgaben und Ziele für jede Ebene. Diese werden in den folgenden Abschnitten näher beschrieben, wobei der Schwerpunkt auf der Bearbeitung redundanter Daten liegt.

2.1 Kooperation auf Rohbildebene

Auf der untersten Ebene des Schichtenmodells nach [Rad93] erfolgt die Bildvorverarbeitung. Typische Aufgaben sind Bildentzerrung, Bildverbesserung, Restauration oder Kodierung. Kooperation auf dieser Ebene erfolgt meist in Form von externer Kooperation ohne gemeinsame Sicht. Sie besteht darin, daß Bildmaterial (s. Abb. 1 (a)) zwischen den räumlich verteilten Beobachtern ausgetauscht wird, die an einer Aufgabenlösung beteiligt sind.

2.2 Kooperation auf Merkmalsebene

Bei der Kooperation auf Merkmalsebene mit überlappendem Sichtfeld liegen zumindest partiell redundante Daten (vgl. Abb. 1 (b)) vor. Ziel und Aufgabe der

Kooperation auf dieser Ebene ist die Rekonstruktion von Objekt oder Szene in ein gemeinsames Koordinatensystem, wobei die Zuordnung der Daten ohne Objektmodellwissen erfolgt.

Die meisten Anwendungen in der Literatur, die sich mit visueller Kooperation befassen, sind dieser Ebene zuzuordnen. In der Vergangenheit wurden häufig Anwendungen beschrieben, die der internen Kooperation zuzuordnen sind. Krotkov beispielsweise beschreibt in [Kro89] ein kooperatives Verfahren zur 3D-Rekonstruktion von Bildszenen mit Stereo und Fokus. Das Verfahren geht von einer unbekannten, statischen Szene mit beliebig strukturierten Objekten aus. Mithilfe einer Stereokamera wird für verschiedene Szenenpunkte sowohl mit einem Fokussierungs- als auch mit einem Stereoverfahren der Abstand von der Kamera berechnet. Dabei verifizieren die beiden Techniken ihre Ergebnisse gegenseitig; mit der Fokussierung werden falsche Stereokorrespondenzen, durch die Stereotechnik werden falsche Fokussierungspunkte identifiziert. Dies erhöht die Zuverlässigkeit im Vergleich zur alleinigen Anwendung eines der beiden Verfahren.

Ähnliches gilt für das Verfahren von [GST89], daß per interner Kooperation die Rekonstruktion von 3D-Objekten oder ganzer 3D-Szenen beschreibt. Es setzt ebenfalls eine statische Bildszene voraus. Mithilfe zweier verschiedener Algorithmen werden aus einer Folge von Stereoaufnahmen Tiefenkarten berechnet und in eine Gesamtdarstellung integriert. Dabei handelt es sich um ein Stereoverfahren und ein Verfahren, das auf der Berechnung des optischen Flusses basiert. Zur Aufnahme der Stereobilder wird die Kamera kreisförmig mit konstantem Radius um einen Fixationspunkt bewegt, normalerweise die Mitte der Bildszene. Dabei sind die Kamerabewegungen bekannt und im voraus festgelegt.

2.3 Kooperation auf Objektebene

Auf dieser Ebene lassen sich Interpretationen zu einem Objekt von Beobachtern mit unterschiedlichen Beobachtungspositionen kombinieren (vgl. Abb. 1 (c)). Dazu ist eine Transformation auf ein einheitliches Koordinatensystem erforderlich. Liegt eine gemeinsame Sicht vor, können die Daten redundant oder komplementär sein. Es müssen jedoch alle Beobachter dasselbe Objekt beobachten, damit Kooperation auf dieser Ebene möglich ist.

Beispiele in der Literatur für Kooperation auf dieser Ebene sind in Anwendungen im Zusammenhang mit *Multi-Perspective Interactive Video* (kurz: MPI-Video) zu finden [JKS96], [KKK+95], [CJK+94] [1]. MPI-Video-Systeme liefern interaktive und realistische Darstellungen realer Ereignisse, die gleichzeitig von mehreren Kameras an verschiedenen Orten aufgenommen werden. Sie finden Anwendung in verschiedenen Aufgaben wie interaktiven Fernsehübertragungen, Teletourismus, interaktive Filme, virtuelle Welten oder Leitsysteme. Im Unterschied zu herkömmlichem Video bietet ein MPI-Video-System viele Interaktionsmöglichkeiten, über die der Inhalt der gezeigten Videosequenzen und die Art der Darstellung gesteuert werden können wie z.B. in einer Sportübertragung die Wahl einer speziellen Perspektive oder die Auswahl eines bestimmten Objekts.

Sieht man einmal von den MPI-Anwendungen ab, gibt es in der Literatur zu dieser Ebene so gut wie keine Anwendungen. Das ist um so erstaunlicher, als daß visuelle Kooperation auf dieser Ebene gerade in Multiroboteranwendungen von großem Interesse sein sollte. Beobachter können auf dieser Ebene bereits selbständig Identifizierungen von Objekten durchführen. Typischerweise ist dieser Vorgang in natürlichen Szenarien sehr anfällig gegenüber Einflüssen wie Helligkeitsschwankungen,

[1] Je nach Anwendung in MPI gibt es auch Aufgaben der Kooperation, die der Merkmalsebene zugeordnet werden

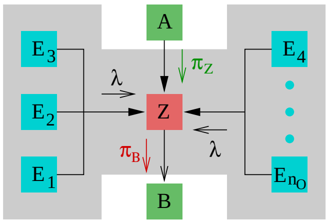
Verdeckungen oder Abschattungen. Daher bietet es sich geradezu an, durch Kooperation mit mehreren Beobachtern zu einer robusteren Klassifizierung zu gelangen. An unserem Institut wurde aus diesem Grunde der Einfluß der visuellen Kooperation auf dieser Ebene untersucht. Zwei Fragen standen dabei im Vordergrund: welche Methoden zur Informationsintegration sind geeignet und welchen Nutzen bringt eine Kooperation auf dieser Ebene.

Ein Identifizierungsverfahren liefert in der Regel eine Menge von möglichen Interpretationen für ein Objekt. Dabei kann es sich um vermutete Klassen- oder Ansichtszuordnungen handeln. Zur Fusionierung dieser Daten, die Hall [Hal92] als Identitätsfusion bezeichnet, eignen sich prinzipiell Techniken wie Kalmanfilter, Methode der kleinsten Quadrate, Bayes'sche Beliefnetze, Dempster-Shafer, Fuzzy-Mengen oder neuronale Netze, um nur ein paar zu nennen. Im Gegensatz zur Positionsfusion muß zwischen den aufgestellten Hypothesen der Objekterkennung jedoch keine Metrik existieren. In so einem Fall sind Techniken wie Kalmanfilter oder Methode der kleinsten Quadrate nicht anwendbar. Andere Verfahren dagegen, wie die Bayes'schen Ansätze, neuronale Netze und Fuzzymethoden, erscheinen dagegen für die anstehende Aufgabe geeignet zu sein. Sie setzen voraus, daß mit jedem Bearbeitungsschritt alle in Betracht kommenden Objekte und/ oder Ansichten zumindest implizit bewertet worden sind. Von den in frage kommenden Techniken hat sich gezeigt, daß Bayes'sche Beliefnetze [Pea87] relativ robust gegenüber Störungen sind und das sie einigermaßen schnell auf Veränderungen reagieren können. Das ist insbesondere im Falle von Objektbewegung und daraus resultierendem ständigen Wechsel der Objektorientierung von Interesse. Zudem sind Bayes'sche Beliefnetze bei der Integration von Information unabhängig von der Anzahl der beteiligten Verfahren bzw. Beobachter. Im folgenden wird der Einsatz von Beliefnetzen für die Kooperation nach [OL99] [OL97] kurz erläutert.

Beliefnetze schätzen die Identität und/ oder die aktuelle Orientierung eines Zielobjekts, indem sie die Ergebnisse der Objekterkennung von mehreren Beobachtern bzw. Verfahren miteinander verknüpfen. Besteht eine Modellhierarchie, kann für jede Hierarchieebene ein separates Beliefnetz benutzt werden. Beliefnetze sind azyklisch gerichtete Graphen bestehend aus Knoten und Kanten, in denen die Knoten Hypothesenwerte repräsentieren und die Kanten direkte Abhängigkeiten zwischen verbundenen Kanten beschreiben. Ein einzelner Knoten ist mit einem oder mehreren Vorgängern und Nachfolgern verknüpft. Er repräsentiert zu jedem Zeitpunkt den Zustand des Beliefnetzes. Erworbenes Wissen eines Knotens kann sowohl zum Vorgänger als auch zum Nachfolger propagiert werden, in der Regel aber ist für diese Aufgabe nur eine top-down Propagierung sinnvoll. Abb. 2 (a) zeigt das Design eines Beliefknotens. Beliefknoten besitzen zwei Arten von Eingabe und Ausgabe, den *causal support* und den *diagnostic support*. In einem Beliefknoten Z wird der causal support $\pi(Z_i)$ aus den propagierten Werten $\pi_Z(A_j)$ seines Vorgängerknotens A mit den bedingten Wahrscheinlichkeiten $P(Z_i|A_j)$ berechnet mit

$$\pi(Z_i) = \sum_{j=1}^{N} P(Z_i|A_j)\pi_Z(A_j), \tag{1}$$

Dabei bezeichnet N die Anzahl der Modelle. Bei den bedingten Wahrscheinlichkeiten $P(Z_i|A_j)$ handelt es sich um *à* priori Wissen, das Auskunft gibt über Schätzungen zeitlicher Abhängigkeiten zwischen Modellen. Diese Schätzungen werden in sogenannten Transitionsmatrizen zusammengefaßt. Transitionsmatrizen können für jeden Beliefknoten verschieden sein, ihre Dimension hängt von der Anzahl der Modelle ab. Jedes Element in der Transitionsmatrix zeigt mit einer Wahrscheinlichkeit $p_{i,j}$ den Übergang von Modell M_i zu Modell M_j in einem Zeitintervall Δt an. Die Wahrscheinlichkeiten $p_{i,j}$ in der Transitionsmatrix können beliebig festgelegt wer-



(a)

t_{k-1} \ t_k	M_1	M_2	$\dots$	M_{n-1}	M_n
M_1	$p_{1,1}$	$p_{1,2}$	$\dots$	$p_{1,n-1}$	$p_{1,n}$
M_2	$p_{2,1}$	$p_{2,2}$	$\dots$	$p_{2,n-1}$	$p_{2,n}$
M_3	$p_{3,1}$	$p_{3,2}$	$\dots$	$p_{3,n-1}$	$p_{3,n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
M_n	$p_{n,1}$	$p_{n,2}$	$\dots$	$p_{n,n-1}$	$p_{n,n}$

(b)

Abbildung 2: : Ein einzelner Beliefknoten Z *(a) Transitionsmatrix mit Übergängen von* M_i *nach* M_j *im Zeitintervall* $\Delta t = t_k - t_{k-1}$ *(b)*

den, als Randbedingung muß lediglich erfüllt sein, daß sie zeilenweise aufsummiert den Wert 1 ergeben. Existiert kein Zusammenhang zwischen den Modellen, wird als Transitionsmatrix die Einheitsmatrix verwendet. Besteht ein Zusammenhang zwischen Modellen, z.B. wenn verschiedene Ansichten pro Modell betrachtet werden, bieten sich Verteilungsfunktionen wie die Gauß- oder die Gleichverteilung an, um die Werte $p_{i,j}$ zu bestimmen.

Zu jedem Zeitschritt gelangen die Hypothesen der bis zu n_O Beobachter (bzw. Verfahren) als diagnostic support $\lambda(E_i)$ in den Beliefknoten Z. Aus diesen Hypothesen berechnet sich der totale diagostic support $\lambda(Z_i)$ mit

$$\lambda(Z_i) = \prod_l \lambda(E_l) \tag{2}$$

Die Beliefwerte eines Beliefknoten berechnen sich aus causal support und diagnostic support sowie einer Normalisierungskonstante α mit

$$BEL(Z) = \alpha\lambda(Z_i)\pi(Z_i) \tag{3}$$

Aus den Beliefwerten erhält man die Werte $\pi_B(Z)$ zur top-down Propagierung mit $\pi_B(Z) = BEL(Z)$. Bei der Objekterkennung wird also zu jedem Zeitschritt für jeden Knoten im Netz ein Vektor von Beliefwerten errechnet. Dieser Vektor gibt Auskunft über die Modell- oder Ansichtszuordnung eines Objekts und stellt damit das durch die Teammitglieder fusionierte Ergebnis des Erkennungsvorgangs dar.

2.4 Kooperation auf Szenenebene

Während es bei der Kooperation auf der dritten Ebene darum geht, ein einzelnes Objekt zu verifizieren, sollen in der vierten Ebene alle beteiligten Objekte und deren Beziehung zueinander erfaßt werden. Typische Werkzeuge dafür sind relationale Strukturen, semantische Netze oder Grammatiken. Die so entstandene Beschreibung der Szene bildet die Grundlage für situationsbezogenes Handeln. Ziele der Kooperation auf dieser Ebene sind die Validierung der lokalen Sicht auf die Szene sowie der Aufbau einer konsistenten globalen Szenenkarte (vgl. Abb. 1 (d)), vorausgesetzt, es liegt eine gemeinsame Sicht vor.

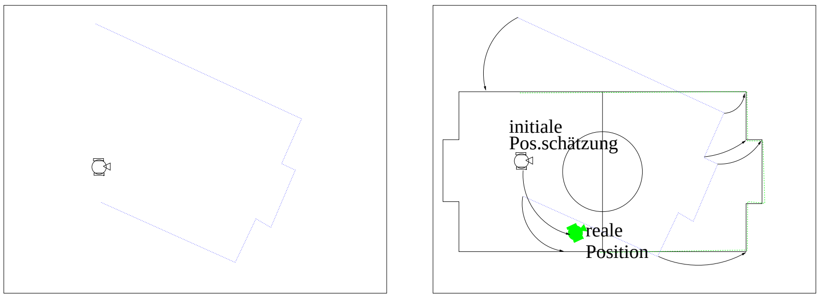
3 Selbstlokalisierung

Bei der externen Kooperation mit gemeinsamer Sicht tauschen Beobachter eines Teams bereits ausgewertete Informationen untereinander aus. Damit diese Informationen verknüpft werden können, muß die absolute Position der Beobachter bekannt sein[2]. Die Bestimmung der absoluten Position ermittelt jeder einzelne Beobachter durch Selbstlokalisation. Typische Sensoren für die Selbstlokalisierung sind:

- Odometrie, also die Positionsbestimmung aus der Radbewegung
- Ultraschall
- Laser-Abstandssensoren (Laserscanner)
- Kamera

Die Odometrie liefert Informationen über die Bewegung der Räder, die meistens mittels Microcontroller ausgewertet und direkt in Positionsinformation umgerechnet wird. Schwäche dieses Sensors ist die sich mit der Zeit erhöhende Ungenauigkeit der Positionsschätzung, die außerdem nur relativ zu einer initalen Position ist. Absolute Positionsschätzungen können stets mit Ultraschall, Laser oder Kamera ermittelt werden. Diese Sensoren liefern mit jeder Messung gemäß ihrem Sensorsichtfeld einen Datensatz mit einem oder mehreren Abstandswerten.

Bei der Selbstlokalisierung wird versucht, eine Überdeckung von Sensordaten und Modell zu finden. Dies kann einerseits auf den Abstandsdaten direkt erfolgen, oder aber auf Linien, die aus den Abstandsdaten extrahiert wurden ([SC94]). Im folgenden soll ein iteratives Verfahren näher erläutert werden ([Cox90]). Der Messdatensatz wird solange gedreht und verschoben, bis die Summe der Abstandsquadrate zwischen den Messdaten und den zugeordneten Linien eine Schwelle unterschreitet. Bei diesem Verfahren ist eine Initialschätzung der Position erforderlich. Im ersten Schritt wird der Messdatensatz um die geschätzte Position verdreht und verschoben. Dann wird zu jedem Messpunkt die nächste Linie des Modells bestimmt und der Abstand berechnet. Über ein Gradientenverfahren lässt sich die Abweichung des Scans vom Modell berechnen. Die verbesserte Schätzung ergibt sich aus der ursprünglichen Schätzung und der Abweichung. Liegt die Abweichung unter einer vorgegebenen Schwelle, terminiert das Verfahren.



(a) (b)

Abbildung 3: Selbstlokalisierung: (a) Laserscan, (b) Matching mit Linienmodell

Bei der Selbstlokalisierung wird häufig versucht, den Suchraum einzuschränken. Beispielsweise kann mittels Odometrie eine grobe Positionsschätzung erfolgen, die dann durch Laser- oder Ultraschalldaten verbessert wird. Dies ist besonders dann nötig, wenn mehrere Situationen mit ähnlichen Sensorwerten zu erwarten sind, wie dies in Büroumgebungen mit gleichartig eingerichteten Räumen der Fall ist. Dann kann zwar die Position innerhalb des Raums bestimmt werden, nicht jedoch um welchen der Räume es sich handelt. Um dies zu bestimmen, kann ausgehend von einer Initialposition der Weg verfolgt werden, der eine grobe Orientierung mittels Odometrie

[2]Es wird angenommen, daß die Kameraorientierung in der Selbstlokalisierung berücksichtigt ist

erlaubt.

Ein weiteres Problem ist die Symmetrie der Umgebung. So ist das Spielfeld beim Roboterfußball symmetrisch aufgebaut, so daß die Positionsschätzung immer zwei zum Mittelpunkt symmetrische Positionen liefert. Hier kann bei Berücksichtigung der Zeit der maximal mögliche Fahrweg und die maximale Drehung zwischen zwei Messungen errechnet werden. Folgen diese dicht genug aufeinander, können Fehllokalisierungen ausgeschlossen werden.

4 Kooperative Bildverarbeitung in der Anwendung

Die meisten Anwendungen mit kooperativer Bildverarbeitung benötigen Kenntnis über den Standort der Beobachter. Eine Ausnahme bildet dabei die interne Kooperation. Hier wird keine Kenntnis der absoluten Position in Weltkoordinaten benötigt, da sich alle an der Kooperation beteiligten Verfahren auf das gleiche Koordinatensystem beziehen. Bei der externen Kooperation geht es dagegen darum, Daten zu integrieren, die von einem anderen Blickwinkel erfaßt wurden. Liegt keine gemeinsame Sicht vor, so macht ein Informationsaustausch ohne Positionsangabe im allgemeinen keinen Sinn. Im Falle einer gemeinsamen Sicht auf eine Szene oder ein Objekt werden zwei grundsätzliche Kooperationsziele unterschieden, die *Verifikation* und die *Lokalisation*. Die Verifikation hat zum Ziel, Hypothesen mehrerer Beobachter zum Zwecke einer robusteren Erkennung zu kombinieren unter der Voraussetzung, daß die beteiligten Beobachter ihre absolute Position kennen. Bei der Lokalisation besteht - im Unterschied zur Selbstlokalisierung - die Aufgabe darin, die relative räumliche Position anderer Beobachter zu erkennen. Dies setzt voraus, daß alle beteiligten Beobachter in der Lage sind, selbständig robuste Hypothesen zu ermitteln. Verifikation und Lokalisation sind im Prinzip auf jeder Ebene der Kooperation möglich. Für ein Team von Beobachtern sind jedoch erst Anwendungen interessant, die höheren Ebenen zugeordnet werden können. Im folgenden werden daher drei Anwendungen auf Objekt- bzw. Szenenebene beschrieben.

4.1 Verifikation und Lokalisation auf Objektebene

In [OL99] wurde am Beispiel einer Referenzfahrt des mobilen Roboters *Aramis* der Einfluß der Kooperation auf die Qualität der Erkennungsergebnisse gezeigt. Abb. 4 zeigt Ausschnitte der Referenzfahrt, die von den beiden statischen Beobachtern E_1 und E_2 verfolgt wurden. Der tatsächliche Orientierungswinkel, also der Winkel zwischen den optischen Achsen der beiden Beobachter, der zur Transformation der Hypothesen auf ein einheitliches Koordinatensystem benötigt wird, stammt im Falle der Verifikation von den errechneten Selbstlokalisationen von E_1 und E_2. Bei der Lokalisation wird dieser Winkel dagegen ermittelt. Der tatsächliche Orientierungswinkel beträgt in diesem Szenario ca. 190°. Beide Beobachter sind in der Lage, neben der Identität des Objekts auch dessen Orientierung bzw. Ansicht zu bestimmen. Dazu vergleichen sie die Inhalte der konvexen Hüllen aus der Bildsequenz von Abb. 4 mit 5 Referenzobjekten einer Modelldatenbank, die aus jeweils 24 Ansichten in 15°-Schritten von 0° bis 360° bestehen.

Dieser Erkennungsprozeß liefert für jede Ansicht eine Bewertung und damit implizit auch eine Modellzuordnung. Die individuellen Erkennungsergebnisse bilden den Ausgangspunkt zur Untersuchung des Einflußes der Kooperation. Abb. 5 (a) zeigt jeweils die maximal bewertete Hypothese für die 74 Bilder der Bildfolge. Die Erkennungsrate hängt stark von der Qualität der Eingabedaten ab, sie liegt bei E_1 bei ca. 76%, bei E_2 nur bei 56%.

Durch die Verifikation kann diese Rate deutlich gesteigert werden, wie man an der

(a)

(b)

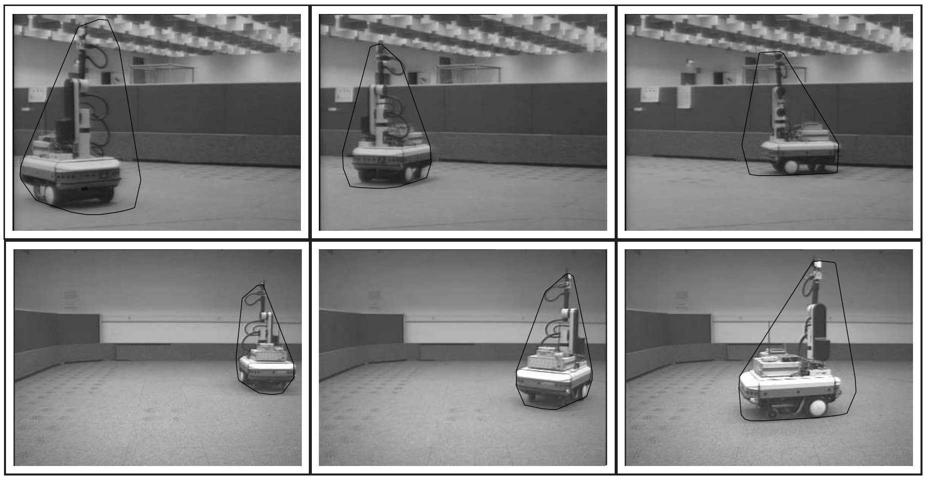


Abbildung 4: Die Bilder #1, #30 und #60 aus der Sicht der beiden Beobachter E_1 (a) und E_2 (b)

Verify-Kurve aus Abb. 5 (b) erkennt. Als Fusionmethode wurden die in Kap. 2.3 vorgestellten Beliefnetze verwendet. Zu jedem Zeitschritt gelangen die Hypothesen der beiden Beobachter simultan als diagnostic support in das Netz. Man erkennt deutlich, daß im kooperativen Fall die Zuordnung zum Modell *Aramis* stets korrekt ist und die geschätzte Ansicht gar nicht oder nur minimal von der tatsächlichen abweicht.

Für die Lokalisation hat jeder Beobachter lokal die Qualität seiner Hypothesen durch Verwendung eines eigenen Beliefnetzes verbessert. Mit dieser kontinuierlichen Betrachtung der Hypothesen konnte die Erkennungsrate immerhin auf 80% bei E_1 und 71% bei E_2 verbessert werden. Der Orientierungswinkel zwischen den Beobachtern wurde in der *Locate*-Kurve aus Abb. 5 (b) durch einfache Mittelung der maximalen Hypothesen von E_1 und E_2 bestimmt. Wenn das Ergebnis der Lokalisation so robust wie in diesem Fall ist, kann es sogar zur Verifizierung des Orientierungswinkels aus der Selbstlokalisierung verwendet werden.

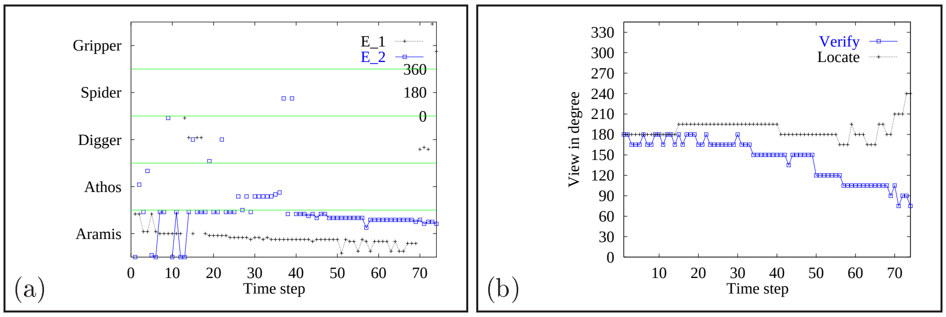


Abbildung 5: (a) Ausgangshypothesen für die Verifikation durch die beiden Beobachter E_1 und E_2 (b) Ergebnisse der Verifikation und Lokalisation

4.2 Selbstlokalisierung in einem Team von Robotern

Ein typisches Anwendungsbeispiel für Kooperation auf der dritten Ebene ohne gemeinsame Sicht auf ein Objekt ist die Selbstlokalisierung mit Hilfe eines oder mehrerer anderer Beobachter. In einer Büroumgebung mit gleichartigen Räumen wie z.B. in Abb. 6 (a) kann es vorkommen, daß ein Beobachter sich nicht eindeutig lokalisieren kann und mehrere gleichwahrscheinliche Hypothesen erhält. Durch visuelle Kooperation in einem Team von Beobachtern kann diese Mehrdeutigkeit aufgelöst werden.

Dazu wird ein zweiter Beobachter, der seine absolute Position eindeutig bestimmen kann, beauftragt, Beobachter 1 zu identifizieren. Der Identifizierungsvorgang, der individuell ausgeführt wird, liefert eine Objekthypothese die keine Objektentfernung beinhalten muß. Objekthypothese und absolute Position des Beobachter 2 werden Beobachter 1 mitgeteilt. Mit Hilfe dieser Information ist Beobachter 1 nun in der Lage, alle Hypothesen außerhalb dieses Raumes zu verwerfen und eine korrekte Selbstlokalisation durchzuführen.

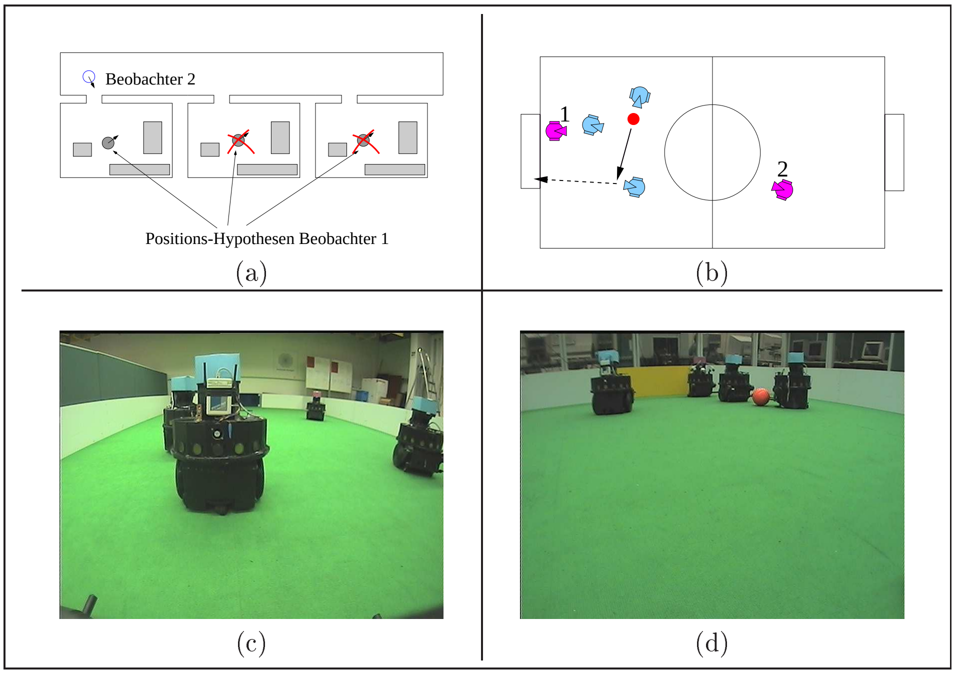


Abbildung 6: Selbstlokalisierung im Team (a), Kooperation auf Szenenebene (b), Sichten von Beobachter 1 (c) und Beobachter 2 (d)

4.3 Verifikation auf Szenenebene

Aufgrund der Komplexität sind Realisierungen visueller Kooperation auf der Szenenebene bisher nur in relativ einfachen Szenarien möglich. Ein typisches Anwendungsgebiet sind fußballspielende Roboter. Wegen des relativ einfachen Umfelds[3] kann jeder Roboter für sich eine visuelle Szenenerkennung durchführen. Der Einfluß der Kooperation auf die Szeneninterpretation im Roboterfußball sei am Beispiel einer Spielsituation beschrieben, wie sie bei den CoPS[4] Stuttgart realisiert wurde.

Angenommen, ein Spieler (in Abb. 6 (b) der Spieler mit der Nummer 1) soll situationsbezogen handeln: a) sofort stören, wenn Ball und Gegner in der eigenen Spielhälfte erkannt werden und der Gegner einen Torschuß beabsichtigt, b) einen

[3] In diesem Szenario existieren die folgenden Objekte: roter Ball, gelbes Tor, blaues Tor, grüner Boden, Gegner oder Mitspieler farblich markiert entweder in hellblau oder violett

[4] CoPS = *C*ooperative S*o*ccer *P*laying Robot*s*, Roboterfußballteam am IPVR

gezielten Paß auf einen Mitspieler spielen, falls nur der Ball, aber kein Gegner sichtbar ist ansonsten c) abwarten und nichts tun. In Abb. 6 (c) erkennt Spieler 1 in seiner Umgebung die gegnerischen Roboter, aber keinen Ball, weil dieser verdeckt ist. Folglich würde er nichts tun. Sein Mitspieler, der Spieler mit der Nummer 2, der die gleiche Szene aus einem anderen Blickwinkel beobachtet, ist in der Lage, Gegner und Ball zu erkennen (Abb. 6 (d)). Kooperation auf dieser Ebene bedeutet nun, daß Spieler 2 sein Wissen Spieler 1 zur Verfügung stellt. Spieler 1 erhält durch Fusion seines lokalen Wissens mit dem übermittelten eine zusätzliche Information, die er nicht selbst ermitteln konnte. Mit der neuen Information ändert sich der Kontext aus seiner Sicht, so daß Spieler 1 die Gefahrensituation erkennen und entsprechend handeln kann.

5 Zusammenfassung

In der vorliegenden Arbeit wurde der Begriff der visuellen Kooperation bzw. kooperativen Bildverarbeitung definiert, analysiert und in Bezug gesetzt zum Schichtenmodell der Bildverarbeitung. Dabei hat sich gezeigt, daß durch visuelle Kooperation auf allen Ebenen der Bildverarbeitung eine Steigerung der Qualität der Ergebnisse erreicht werden kann. Visuelle Kooperation ist einerseits zwischen verschiedenen Verfahren eines Beobachters möglich, andererseits durch Verteilung der Sichtaufgabe an mehrere Beobachter. Die Wahl einer geeigneten Fusionsmethode hängt von der Art der Kooperation ab, der Abstraktionsebene und davon, ob die Sichtweisen der Beobachter auf ein Objekt oder eine Szene redundant oder komplementär sind. Das Ziel der externen Kooperation besteht im allgemeinen darin, Hypothesen mehrerer Beobachter zum Zwecke einer robusteren Erkennung zu kombinieren. Eine solche Verifikation ist nur dann möglich, wenn jeder Beobachter in der Lage ist, eine Selbstlokalisierung durchzuführen. Die Fähigkeit zur Selbstlokalisation bildet somit die Basis zur visuellen Kooperation in einem Team von Robotern.

Literatur

[AWB87] J. Aloimonos, I. Weiss, and A. Bandopadhay. Active vision. *International Journal on Computer Vision*, pages 333–356, 1987.

[Baj88] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8), August 1988.

[CJK+94] S. Chatterjee, R. Jain, A. Katkere, P. Kelly, D.Y. Kuramura, and S. Moezzi. Modeling and interactivity in mpi-video. Technical Report VCL-94-103, Visual Computing Laboratory, University of California, San Diego, December 1994.

[Cox90] I.J. Cox. Blanche: Position estimation for an autonomous robot vehicle. In I.J. Cox and G.T. Wilfong, editors, *Autonomous robot vehicles*, pages 221–228. Academic Press, New York, 1990.

[GST89] E. Grosso, G. Sandini, and M. Tistarelli. 3d object reconstruction using stereo and motion. *IEEE Transaction on Systems, Man and Cybernetics*, 19(6):1465–1488, 1989.

[Hal92] D. Hall. *Mathematical Techniques in Multi-sensor Fusion.* Artech House Inc., 1992.

[JKS96] R. Jain, A. Katkere, and J. Schlenzig. Mpi video: Content-centric interactivity and gestalt video. Technical report, Visual Computing Laboratory, University of California, San Diego, La Jolla, CA, 1996.

[KKK+95] P.H. Kelly, A. Katkere, D.Y. Kuramura, S. Moezzi, S. Chatterjee, and R. Jain. An architecture for multiple perspective interactive video. Technical report, Visual Computing Laboratory, La Jolla, CA, 1995.

[Kro89] E.P. Krotkov. *Active Computer Vision by Cooperative Focus and Stereo*. Springer, 1989.

[OL97] N. Oswald and P. Levi. Cooperative vision in a multi-agent architecture. *Lecture Notes in Computer Science*, 1310:709–716, 1997.

[OL99] N. Oswald and P. Levi. A framework for cooperative object recognition. Technical Report 1999/12, University of Stuttgart, 1999.

[Pea87] J. Pearl. Distributed revision of composite beliefs. *Artificial Intelligence*, 33:173–215, 1987.

[Rad93] B. Radig, editor. *Verarbeiten und Verstehen von Bildern*. Oldenbourg, 1993.

[SC94] B. Schiele and J.L. Crowley. Comparison of position estimation techniques using occupancy grids. In *Proceedings of the 1994 IEEE International Conference on Robotics and Automation, ICRA'94*, S. Diego (CA), USA, May 8–13 1994. IEEE Computer Society Press.