Eye4Save – visuelle 3D-Überwachung für die sichere Mensch-Maschine-Kooperation

Daniel Kapusi[1], Rainer Jahn[1], Rico Nestler[2], Karl-Heinz Franke[1]

[1]Zentrum für Bild- und Signalverarbeitung e. V.
Werner-von-Siemens-Str. 10, D-98693 Ilmenau
eMail: daniel.kapusi@zbs-ilmenau.de
eMail: rainer.jahn@zbs-ilmenau.de
eMail: karl-heinz.franke@zbs-ilmenau.de
URL: http://www.zbs-ilmenau.de
[2]Technische Universität Ilmenau, Fak. IA, FG GDV,
Postfach 100 565, D-98684 Ilmenau
eMail: rico.nestler@tu-ilmenau.de
URL: http://www.tu-ilmenau.de

Zusammenfassung Motiviert durch den steigenden Bedarf an barrierefreien Sicherheitslösungen für moderne Mensch-Maschine-Arbeitsplätze ist das visuelle Überwachungssystem „Eye4Save“ entwickelt worden. Die Beobachtung der abzusichernden Umgebung erfolgt mit Hilfe mehrerer Stereo-Farbkameras aus verschiedenen Perspektiven. Die gewonnenen Tiefenbilder ermöglichen in Kombination mit den lokalen Farbinformationen eine robuste Extraktion von Vordergrundsegmenten bezüglich eines Hintergrundmodells. Die Fusion von vordergründigen Bild- und Tiefendaten aus den unterschiedlichen Ansichten erlaubt schließlich die vollständige 3D-Beschreibung der Vordergrundobjekte, die durch hierarchisch strukturierte Voxelmodelle repräsentiert werden. Durch die relationale Auswertung von Objekten in der Szene werden potentiell gefährliche Situationen erkannt und ggf. durch einen aktiven Eingriff in die Maschinensteuerung bereits im Ansatz entschärft. An Konzepten zur temporalen Objektverfolgung, insbesondere im Hinblick auf die Bewältigung von Ausnahmesituationen, wird gearbeitet. Solche Ausnahmesituationen entstehen z. B. durch „Verschmelzen“ oder „Auftrennen“ von Modellobjekten.

1 Einleitung und Stand der Technik

Unter dem Begriff „Industrie 4.0“ sollen sich bis zum Jahr 2025 intelligente Fabriken entwickeln, die u. a. durch starke Individualisierung der Produkte unter den Bedingungen einer hoch flexiblen Massenfertigung gekennzeichnet sind. Gerade im Hinblick auf Flexibilität werden in der Produktion von morgen zunehmend

menschliche Erfahrungen, Kreativität und Fertigkeiten in den Produktionsprozess zu integrieren sein. Andererseits wird der Mensch zukünftig bei einer breiten Palette weiterer Tätigkeiten von cyber-physischen Systemen assistiert [20]. In beiden Fällen rücken die Arbeitsbereiche von Menschen und Maschinen immer näher zusammen, wodurch die Ansprüche an die Sicherheitsanforderungen stark ansteigen. Um den hohen Sicherheitsanforderungen gerecht zu werden, werden bisher vorwiegend physisch trennende Schutzeinrichtungen, wie z. B. Schutzzäune, eingesetzt. Der Zugang in den Arbeitsbereich ist in diesem Fall räumlich sehr beschränkt und im Störfall wird die Reduzierung der Maschinenbewegung auf eine „sichere“ Geschwindigkeit bis hin zum Stillstand ganzer Fertigungslinien vorausgesetzt. Ein paralleles bzw. kooperatives Arbeiten von Mensch und Maschine ist unter diesen Bedingungen nicht möglich. Der gleichen Sicherheitsstrategie folgend sind berührungslos wirkende Schutzeinrichtungen wie Lichtvorhänge, Laserscanner und Kameras auf dem Markt, die ausgewiesene Zugangsbereiche überwachen. Ergänzend werden akustische und taktile Systeme eingesetzt. Beispiele sind Trittmatten und die druckempfindliche, künstliche Roboterhaut.

Modernere Entwicklungen im Bereich der kamerabasierten Überwachung setzen auf eine tiefenbildbasierte Beobachtung aus der Vogelperspektive [5, 13]. Hier wird ein virtueller Überwachungskegel über den Arbeitsbereich gestülpt. Durch den Verzicht auf weitere trennende Schutzeinrichtungen kann u. a. eine erweiterte und relativ flexible Zugänglichkeit zur Maschine erreicht werden, was die Akzeptanz und die Effektivität von Mensch-Maschine-Arbeitsplätzen erheblich steigert. Bei diesem Ansatz werden keine Objekte detektiert, sondern im virtuellen Überwachungskegel werden definierbare Schutzzonen aktiviert. Sie werden auf eine Mindestmenge von 3D-Punkten überprüft, die vom Hintergrund abweichen, und es wird entsprechend in die Maschinensteuerung eingegriffen. Nach dem Verlassen des Gefahrenbereiches muss der Wiederanlauf der Maschine allerdings manuell erfolgen, da der unmittelbare Wirkungsbereich des Roboters wegen seiner Eigenbewegung prinzipbedingt von der Überwachung ausgeschlossen werden muss.

Der hier verfolgte Ansatz ist die kamerabasierte Überwachung aus mehreren unterschiedlichen Perspektiven. Im Gegensatz zur zuvor geschilderten, monoperspektivischen Überwachung, die nur z-eindeutig bezüglich einer Kameraperspektive ist, kann durch den Ansatz mit verschiedenen Ansichten die Szene im Wesentlichen vollständig dreidimensional erfasst werden. Dadurch lassen sich grundsätzlich auch Objekte detektieren, die aus einer Einzelperspektive durch andere statische oder dynamische Objekte verdeckt sind. Unter anderem wird an der Universität Bayreuth seit 2001 an einem solchen multiperspektivischen Ansatz für die sichere Mensch-Roboter-Kooperation geforscht [1]. Mehrere Einzelkameras [4] oder auch Tiefenkameras [2] zeichnen simultan Bilder auf. Darin werden potentielle Objekte mit Hilfe einer grauwert-, farb- und distanzbasierten Vordergrundklassifikation segmentiert und über den Volumenschnitt in ein 3D-Voxelmodell [17] überführt. Anhand der 3D-Daten im Voxelraum werden Gefahrensituationen erkannt und Maschinen ggf. angehalten bzw. Ausweichmanöver gefahren.

Unter starker Beteiligung der Thüringer Industrie wurde 2009 ein im Rahmen des Europäischen Fond für regionale Entwicklung gefördertes, zweijähriges Verbundprojekt (2008FE9063) gestartet, aus dem Folgearbeiten und Entwicklungen des ZBS

e. V. zur Thematik „Bildbasiertes Überwachungssystem für die Bewegungssteuerung von Robotern im sicheren Arbeitsprozess mit Menschen“ hervorgegangen sind. Erste Ergebnisse wurden unter dem Namen „Eye4Save“ [8] auf der Hannover-Messe 2013 vorgestellt.

2 Prinzipieller Aufbau und Funktionsweise des kamerabasierten 3D-Überwachungssystems

Der Mensch-Maschine-Arbeitsraum wird aus unterschiedlichen Perspektiven mit Hilfe mehrerer Stereo-Farbkamera-Paare beobachtet. Diese sind dabei so angeordnet, dass die Schnittmenge der einzelnen Sichtkegel möglichst groß ist, die Verdeckungen durch dynamische und statische Objekten minimiert werden und alle erforderlichen Bereiche in der Überwachungszone von einer vorgegebenen Mindestanzahl von Kameras einsehbar sind. Die Nutzung von geeigneten Simulationstools, z. B. „Sim4Save“ vom Fraunhofer Institut für Digitale Medien (IDMT) [21] kann beim interaktiven Finden optimaler Kamerapositionen und –ausrichtungen sehr dienlich sein. Ober-Gecks et. al. stellen zu dieser Thematik in [19] ein automatisiertes Optimierungsverfahren vor. Für das diskutierte einfache Testszenario mit Industrieroboter wurden vier in den oberen Raumecken der Halle montierte Stereo-Farbkameras gewählt, die den Arbeitsraum aus einer schrägen Aufsicht heraus beobachten (siehe Abbildung 1). Alle Aufnahmen des Kameraverbundes sind zeitsynchronisiert. Als Verarbeitungseinheit wird ein Multi-Prozessor-Industrie-PC verwendet.
In der Arbeitsphase wird aus den ermittelten Objektrelationen und der gemessenen Laufzeit von der Bildaufnahme bis zum Ergebnis gemäß DIN EN ISO 13855 [18] ein Zeitlimit berechnet, in dem die nächste Aktualisierung des Ergebnisses erfolgt sein muss. Bei Überschreitung stoppt ein unabhängiger Watchdog das System. Das System erfüllt nach diesem Prinzip harte Echtzeitanforderungen mit flexiblen Zeitschranken. Beim Einrichten werden die Kameras mittels Schachbretttarget (siehe Abbildung 2) [10, 11] colorimetrisch und bezüglich der inneren und äußeren Kameraparameter kalibriert. Das Anlernen bzw. die interaktive Definition des Hintergrundbereiches gehören ebenfalls zum Einrichtungsprozess. Aus den aufgenommenen, verzeichnungskorrigier-

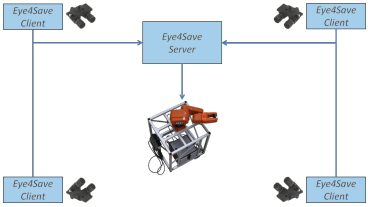

Abbildung 1: Kameraanordnung

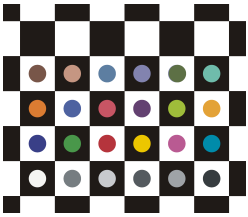
Abbildung 2: Kalibriertarget

ten und rektifizierten Bildern [6, 24] der Stereo-Farbkameras wird pro Kamerapaar ein Tiefenbild [15] berechnet. Die Trennung von 2D-Vordergrundsegmenten und Hintergrund erfolgt unter Verwendung der lokalen Tiefen- und Farbinformationen ebenfalls für jede Stereo-Farbkamera separat (siehe Abschnitt 3).

Die 2D-Vordergrundsegmente aller Perspektiven werden in Form ihrer visuellen Hüllen mittels des Volumenschnittverfahrens (siehe Abschnitt 4) in einem gemeinsamen kartesischen Koordinatensystem fusioniert. Das Ergebnis sind zunächst zusammenhanglose 3D-Voxelmengen in einem Octree-Voxelraum, die nach einer anschließenden Gruppierung nach 3D-Nachbarschaften und unter Einbeziehung der Stereo-Tiefeninformationen zu 3D-Voxelobjekten vereint werden. Die segmentierten 3D-Voxelobjekte korrespondieren mit Objekten im realen Mensch-Maschine-Arbeitsraum und liefern somit ein modellhaftes Abbild der Szene zum Zeitpunkt der Bildaufnahme. Nach der Klassifikation der Objekte und der Auswertung ihrer Eigenschaften und Relationen können für die Überwachungsaufgabe notwendige Rückschlüsse auf kritische Situationen erfolgen.

3 Vordergrundsegmentierung in den einzelnen Stereo-Bildpaaren

Die Vordergrundsegmentierung (VS) basiert auf einem statistischen Hintergrundmodell (HM) für jedes Pixel. HM und VS sind in Jahn et. al. [7] u. a. ausführlich erläutert und werden nachfolgend kurz beschrieben. Unter Verzicht auf laterale Zusammenhänge wird von stochastisch unabhängigen Hintergrundmerkmalen ausgegangen. Diese Einschränkung ist nicht unumstritten, hat sich aber in der Praxis oft als hinreichend erwiesen. Das HM wird in einem Belehrungsprozess vor der Inbetriebnahme des Überwachungssystems erstellt, kann aber bei Erfordernis während des Betriebs adaptiert werden. Ausgangspunkt sind primäre Hintergrund-Pixelmerkmale

$$\vec{V}_{i,j} = (Y_{i,j}, U_{i,j}, V_{i,j}, D_{i,j}) \tag{1}$$

mit den transformierten Farbwerten YUV, dem Objektabstand D und den Pixelindizes i, j. Das HM eines Pixels wird oft durch eine multimodale Gaußverteilung mit fester Modenzahl dargestellt, deren Parameter aus der Belehrungsbildfolge iterativ berechnet werden [22, 3]. Wir verwenden als alternativen Ansatz mit gewissen Vorteilen bezüglich Flexibilität und mittlerem Speicherbedarf ein sogenanntes *Codebook* [12]. Es beschreibt allgemein eine multimodale Gleichverteilung, in unserem Fall eine 3D-Verteilung für YUV und eine davon unabhängige 1D-Verteilung für den Objektabstand D. Die Parameter einer Verteilungsmode heißen *Codebook-Eintrag*. Das Codebook eines Pixels setzt sich aus einer variablen Anzahl von Codebook-Einträgen zusammen. Ein solcher Eintrag lautet speziell für die k-te Verteilungsmode des Objektabstandes:

$$\vec{D}_k = (D_l^{(k)}, D_{min}^{(k)}, D_{max}^{(k)}, D_h^{(k)}, n_D^{(k)}). \tag{2}$$

Neben den Intervallgrenzen $D_{min}^{(k)}$, $D_{max}^{(k)}$ der Verteilungsmode ist ein umgebender Fangbereich definiert, der von $D_l^{(k)}$ bis $D_h^{(k)}$ reicht; $n_D^{(k)}$ ist die absolute Häufigkeit der Stichprobenwerte, die während der Belehrungsphase auf diese Verteilungsmode entfallen. Analog zu (2), jedoch mit 3D-Intervallgrenzen, ist der Codebook-Eintrag für einen YUV-Farbwert definiert.

Der erwähnte Fangbereich ist in der Belehrungsphase von Bedeutung: Wenn der aktuelle Merkmalswert D des Pixels in den Fangbereich eines schon vorhandenen k-ten Codebook-Eintrags fällt, wird D zu dieser Verteilungsmode gezählt (Inkrementierung

von $n_D^{(k)}$), und gegebenenfalls werden die Intervallgrenzen in (2) erweitert. Liegt D außerhalb aller Fangbereiche, wird ein neuer Eintrag der Gestalt $(D-\Delta, D, D, D+\Delta, 1)$ mit einem festen Wert Δ zum Codebook hinzugefügt. Ausgewählte Aspekte des HM lassen sich als Farb- oder Grauwertbild visualisieren. Ein Beispiel zur Visualisierung der YUV-Erwartungswerte ist in Abbildung 3 dargestellt. Die VS wird während der

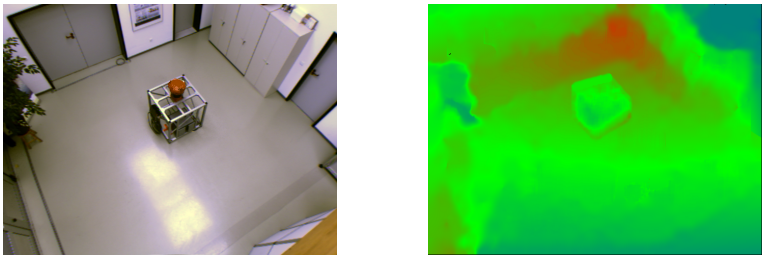

Abbildung 3: Visualisierung des Hintergrundmodells. *links* YUV-Erwartungswerte, nach RGB transformiert; *rechts* Distanzmodell (farbdistanzcodiert)

Betriebsphase mit jeder Einzelaufnahme durchgeführt. Ein Pixel wird genau dann als Vordergrund klassifiziert, wenn sein Merkmalsvektor nicht mit ausreichender Sicherheit einer Verteilungsmode zugeordnet werden kann. Der Entscheidungsalgorithmus ist mehrstufig und vom Nutzer modifizierbar. Es ist möglich, allein die Partialverteilung der Farbvalenz oder des Objektabstandes zur Klassifizierung heranzuziehen. Wenn die Codebook-Einträge des Objektabstandes genügend signifikant sind, wird ihnen üblicherweise eine hohe Priorität zugewiesen. Dadurch wird die Tatsache, dass der Hintergrund mit den entferntesten Objekten identisch ist, optimal genutzt. Bei ungünstigen Bedingungen, wenn die Objektabstände nicht zuverlässig bestimmbar sind, wird nur anhand der YUV-Partialverteilung klassifiziert. Um auch laterale Abhängigkeiten im Bild zu berücksichtigen, wird das segmentierte Binärbild in folgenden Schritten morphologisch nachbearbeitet: Closing-Operation, Zusammenhangsanalyse mittels RLE-Codierung, Löschen von flächenmäßig kleinen Segmenten.
In Abbildung 4 ist ein Beispiel zur VS dargestellt. Als Vordergrund sind im Idealfall Personen, Maschinen und Gegenstände zu betrachten, die im Hintergrundmodell „unwahrscheinlich“, also in der Visualisierung nach Abbildung 3 unsichtbar sind. Die Verbindung der Merkmale Farbe und Objektabstand mit der Codebook-Methode ist als Verfahren zur VS nach unserer Kenntnis neuartig.

4 Fusion von 2.5D Daten zu 3D-Daten

Beim Volumenschnitt (SfS – Shape from Silhouette) werden die Silhouetten der Vordergrundobjekte aus allen verfügbaren Ansichten in das gemeinsame Weltkoordinatensystem projiziert (siehe Abbildung 6). Die Schnittmenge aller Volumenkegel der Silhouetten aus den verschiedenen Perspektiven bildet die visuellen Hüllen der jeweiligen Vordergrundobjekte im Raum. Das heißt, dass sich die eigentlichen realen Objekte innerhalb dieser Hüllen befinden. Konkave Formen können durch die visuelle Hülle nicht abgebildet werden. Je mehr Perspektiven verwendet werden,

Abbildung 4: Beispiel zur Vordergrundsegmentierung. *links* Kamerabild. *rechts* Segmentierung mit Farbe (weiße Füllflächen an strukturarmen Bildregionen) und Abstand (farbdistanzcodiert)

desto mehr nähern sich die visuellen Hüllen den tatsächlichen Objektformen an und desto stärker verkürzen sich dadurch die erforderlichen Sicherheitsabstände zwischen Mensch und Maschine. Das bedeutet im Umkehrschluss, dass selbst durch den ausfallbedingten Wegfall einer Perspektive die Überwachungsaufgabe des Systems grundsätzlich weiterhin durchgeführt werden kann, wenn der Ausfall registriert wird. Die visuellen Hüllen um die Objekte wären dann entsprechend größer, die gemessenen relativen Objektdistanzen kleiner und ein Maschinenstopp würde eher ausgelöst. Die SfS-Implementierung kann durch inverses Vorgehen wesentlich vereinfacht werden. Ausgehend vom vorgegebenen diskreten 3D-Suchraum wird jedes Voxel in die Bildkoordinaten aller Kameras projiziert [14]. Nur wenn in allen „gültigen“ Ansichten an den korrespondierenden Stellen Pixel als Vordergrundpixel markiert sind, wird das betrachtete Voxel ebenfalls als gültig gesetzt und bildet ein Element der visuellen Hülle. Das Verfahren wird noch performanter, indem im Voxelraum hierarchisch im Top-Down-Prinzip vorgegangen wird [23]. Dazu werden Octree-Datenstrukturen verwendet. Die Verarbeitung in den Octree-Datenstrukturen wird parallelisiert auf mehreren Prozessorkernen durchgeführt. Durch die adaptive Quantisierung im Voxelraum werden sowohl der Speicherverbrauch als auch der Rechenaufwand drastisch reduziert. Im Anschluss an die Ermittlung der Hüllvoxel im 3D-Weltkoordinatensystem werden sich berührende Voxel zu 3D-Objekten zusammengefasst. Durch die hierarchische Struktur des Octree-Voxelraumes erfolgt die 3D-Zusammenhangsprüfung der Voxel sehr recheneffizient. Dabei können problemangepasst unterschiedliche Nachbarschaften (6er für Würfelflächen, 18er für Würfelkanten und 26er für Würfelecken) verwendet werden. Die 6er Nachbarschaft ist oft physikalisch sinnvoll, ausreichend und effizient.

Beim Schnitt von mehreren rückprojizierten Volumenkegeln aus mehreren Silhouetten pro Perspektive können auch „Hüllen“ im Voxelraum entstehen, die kein reales 3D-Objekt enthalten. Entstehen derartige Bereiche durch Okklusion, so müssen sie bei einer Sensoranordnung, die nicht von allen Seiten beobachtet, sicherheitshalber als Pseudo-Objekte bestehen bleiben (siehe Abbildung 7). Allerdings können reale Objekte von solchen Pseudoobjekten durch Projektion der 2.5D-Tiefeninformationen der Stereopaare in den Voxelraum und Vergleich der 3D-Punktdichte innerhalb der visuellen Hüllen unterschieden werden. Diejenigen Pseudo-Objekte, die sich aus

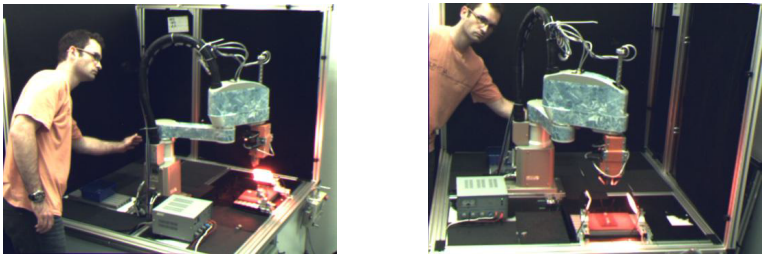

Abbildung 5: Mensch-Roboter-Szene 1 (nur 2 Kamerapaare, linke und rechte Perspektive)

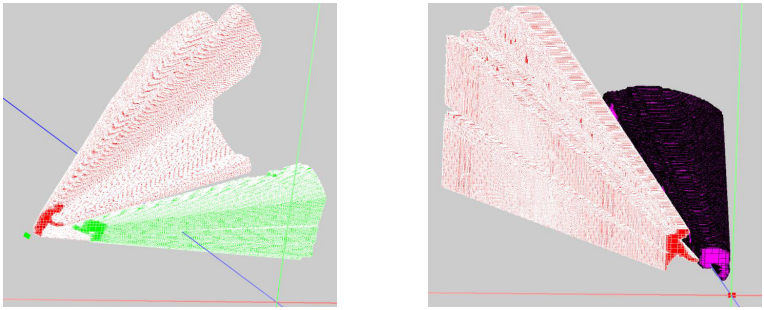

Abbildung 6: Volumenkegel (linke und rechte Perspektive)

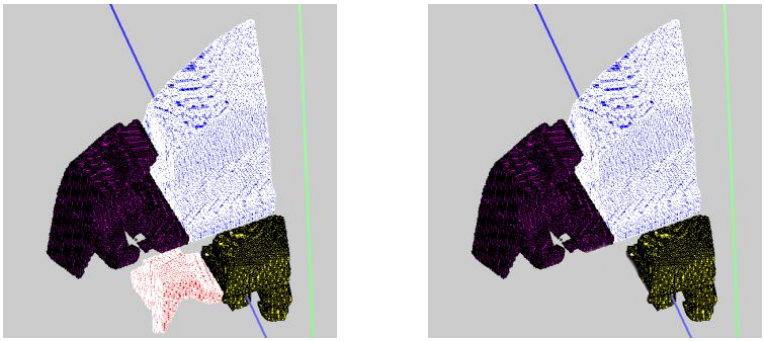

Abbildung 7: *links* Volumenschnitt im Voxelraum. *rechts* Entfernung von Pseudo-Objekten (Blaues Pseudo-Objekt muss bestehen bleiben, da dieser Bereich durch die realen Objekte im Vordergrund für beide Kameras nicht einsehbar ist.)

mindestens einer der jeweiligen Kameraperspektiven vor einem tatsächlichen Objekt befinden, werden auf diese Weise entfernt. Da die Kameras den Arbeitsbereich von allen Seiten beobachten, ist das für nahezu alle Pseudo-Objekte der Fall (siehe Abbildung 8).

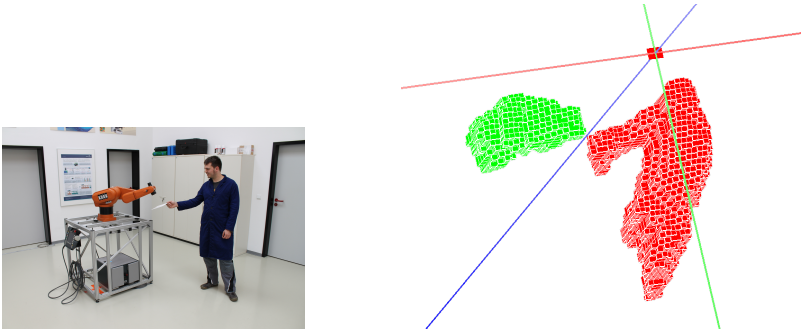

Abbildung 8: *links* Mensch-Roboter-Szene 2. *rechts* Voxeldarstellung (Aufgrund allseitiger Kameraperspektiven konnten alle Pseudo-Objekte entfernt werden.)

5 Konzept zur temporalen Objektverfolgung im Voxelraum

Sich berührende Objekte können nach dem bisher geschilderten Ansatz bei ausschließlicher Betrachtung einer Momentaufnahme nicht separiert werden. Das betrifft natürlich auch solche Objekte, deren Objektrelationen sicherheitsrelevant sind, etwa die zwischen Mensch und Maschine. In der Regel überdecken Bestandteile von ansonsten beweglichen Maschinen statische Raumpositionen, so dass sie über dieses Kriterium als Maschinen-Objekt klassifizierbar sind. Eine schwellwertbasierte Volumenüberwachung ermöglicht eine einfache Berührungsdetektion ohne das Vorwissen über die Szene mit einbezogen werden muss. Die Verfolgung der Objekte über die Zeit und die Feststellung des Verschmelzens in jedem Einzelfall ist allerdings als sicherere Alternative der Volumenüberwachung vorzuziehen. Da die Objekte bereits durch die vorausgegangenen Verarbeitungsschritte formerhaltend detektiert wurden, konzentriert sich die Aufgabe der Objektverfolgung auf die Wiederzuordnung von Objekten aus der Momentaufnahme und bereits verfolgten Objekten der Szenenvergangenheit. Dabei sind Verschmelzungen von Modellobjekten sowie deren Zerfall zu erkennen.
Gegeben sei eine Menge $\check{A}_{t-1}$ von m zeitlich verfolgten Objekten zum Zeitpunkt $t-1$ und eine Menge B_t von n zuletzt detektierten und noch unbekannten Objekten zum Zeitpunkt t. Das Ziel ist die Zuordnung der Objekte $o \in B_t$ zu den Objekten $\check{o} \in \check{A}_{t-1}$. Folgende Zuordnungsfälle können auftreten ($i, j, k, l, m, n \in \mathbb{N}^*$):

1. Die Objektzuordnung ist eindeutig.

$$o_i \mapsto \check{o}_j$$

2. Mehrere Objekte $\check{o}_{j\ldots l} \in \check{A}_{t-1}$ verschmelzen zu einem Objekt $o_i \in B_t$.

$$o_i \mapsto \dot{\bigcup}_j^l \check{o}_j \ , \ j \neq l$$

3. Ein Objekt $\check{o}_j \in \check{A}_{t-1}$ zerfällt in mehrere Objekte $o_{i\ldots k} \in B_t$.

$$\dot{\bigcup}_i^k o_i \mapsto \check{o}_j \ , \ i \neq k$$

4. Ein neues Objekt $o_i \in B_t$ kommt in den Überwachungsbereich.

$$o_i \not\mapsto \forall \check{o} \in \check{A}_{t-1}$$

5. Ein Objekt $\check{o}_j \in \check{A}_{t-1}$ verlässt den Überwachungsbereich.

$$\forall o \in B_t \not\mapsto \check{o}_j$$

Zunächst sollen nur die Fälle 1-3 betrachtet werden. Es ergeben sich

1. mn Möglichkeiten für die direkte Bildung von Objektpaaren,
2. $n(2^m - m - 1)$ Möglichkeiten für eine Verschmelzung der Objekte und
3. $m(2^n - n - 1)$ Möglichkeiten für einen Zerfall der Objekte.

Zusammengenommen bestehen $(m(2^n - n - 1) + n(2^m - 1))$ zu bewertende Varianten für die Zuordnungen. Aus der Menge der zugeordneten Objekte $\forall \check{o} \in \check{A}_{t-1}$ werden $(2^m - m - 1)$ Kandidaten

$$\check{A}_{t-1} := \mathcal{P}(\check{A}_{t-1}) \setminus \check{A}_{t-1} \tag{3}$$

für alle möglichen Objektverschmelzungen gebildet und analog dazu aus der Menge der noch nicht zugeordneten Objekte $\forall o \in B_t$ $(2^n - n - 1)$ Kandidaten

$$\breve{B}_{t-1} := \mathcal{P}(B_t) \setminus B_t \tag{4}$$

für alle möglichen Objektzerfälle. Neue Kandidaten werden durch Kombination aller Einzelobjekte innerhalb der jeweiligen Mengen $\check{A}_{t-1}$ bzw. $\breve{B}_t$ gebildet. Dazu werden die korrespondierenden Voxelmengen und die damit zusammenhängenden Objekteigenschaften vereinigt. Ein Kandidat für die Verschmelzung bzw. den Zerfall kann aus zwei und mehr Einzelobjekten bestehen. Alle zum Kandidat zugehörigen Einzelobjekte werden über die jeweils kürzeste Objekt-zu-Objekt-Distanz direkt oder indirekt miteinander verbunden. Als Distanzmaß wird die kürzeste euklidische Entfernung zwischen den Objekten im Octree-Voxelraum verwendet. Durch die Prüfung auf Einhaltung einer Maximaldistanz, die in Abhängigkeit von der Zeitdauer t ermittelt wird, können nicht plausible Verschmelzungs- bzw. Zerfallskandidaten an dieser Stelle bereits verworfen werden. Das senkt die Anzahl der möglichen Kombinationen. Die Anzahl der verbliebenen plausiblen Verschmelzungskandidaten $\check{A}_{t-1}$ ist $\check{m} \in \mathbb{N} \mid 0 \leq \check{m} \leq 2^m - m - 1$. Die Anzahl der verbliebenen plausiblen Zerfallskandidaten $\breve{B}_t$ ist $\breve{n} \in \mathbb{N} \mid 0 \leq \breve{n} \leq 2^n - n - 1$.

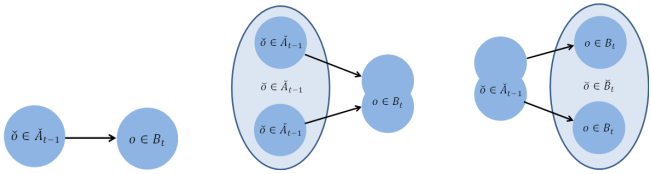

Abbildung 9: Varianten der Objektzuordnung (Direktpaarung, Verschmelzung, Zerfall)

Zwischen allen möglichen mn Objekteinzelpaarungen $B_t \mapsto \check{A}_{t-1}$ und plausiblen $\check{m}n$ Objektverschmelzungspaarungen $B_t \mapsto \check{A}_{t-1}$ sowie $m\check{n}$ Objektzerfallspaarungen $\check{B}_t \mapsto \check{A}_{t-1}$ werden jeweils Distanzmaße zwischen den Merkmalen Schwerpunkt $\underline{s_{xyz}}$, Volumen V und Farbhistogramm H berechnet. Der Ausdruck $feat$ in (5) substituiert jeweils die Merkmale xyz, V, H.

$$D_{feat} := \left\{ \begin{array}{l} d_{feat}(\check{o}, o),\ \forall \check{o} \in \check{A}_{t-1} \forall o \in B_t \\ d_{feat}(\check{o}, o),\ \forall \check{o} \in \check{A}_{t-1} \forall o \in B_t \\ d_{feat}(\check{o}, o),\ \forall \check{o} \in \check{A}_{t-1} \forall o \in \check{B}_t \end{array} \right\} \begin{array}{l} \textit{Direktpaarung} \\ \textit{Verschmelzung} \\ \textit{Zerfall} \end{array} \tag{5}$$

Die euklidische Distanz zwischen den Schwerpunkten $\underline{s_{xyz}}(\check{o}_i)$ und $\underline{s_{xyz}}(o_j)$ ist:

$$d_{xyz}(\check{o}_i, o_j) = \left\| \underline{s_{xyz}}(\check{o}_i) - \underline{s_{xyz}}(o_j) \right\|_2 \tag{6}$$

$$d_{xyz}(\check{o}_i, o_j) \in D_{xyz}$$

Die Volumendifferenz zwischen $V(\check{o}_i)$ und $V(o_j)$ ist:

$$d_V(\check{o}_i, o_j) = |V(\check{o}_i) - V(o_j)| \tag{7}$$

$$d_V(\check{o}_i, o_j) \in D_V$$

Auf dem 20. Farbworkshop 2014 [9] wurden verschiedene Verfahren für das histogrammbasierte Farbmatching diskutiert sowie entsprechende Distanzmaße untersucht und bewertet. In experimentellen Untersuchungen [16] wies das Volumenmerkmal V die größten Schwankungen gegenüber den Merkmalen Schwerpunktposition $\underline{s_{xyz}}$ und Farbhistogramm H auf. Die Merkmale werden zur Ermittlung eines Gesamtdistanzmaßes d_{total} für jede Objektpaarung entsprechend ihrer Signifikanz gewichtet fusioniert.

Die berechneten Distanzmaße $\forall d_{total} \in D_{total}$ aller Objektpaarungen werden in einer Liste der Größe nach aufsteigend sortiert. Der niedrigste Distanzwert korrespondiert mit der ähnlichsten Objektpaarung und steht damit am Anfang der Liste. Die Liste wird nun von oben nach unten bis zum einem vordefinierten Distanzschwellwert abgearbeitet und Objektpaarungen werden dann als gültig gesetzt, wenn keines der beteiligten Einzelobjekte $\check{o} \in \check{A}_{t-1}$ bzw. $o \in B_t$ an eine übergeordnete gültige Objektpaarung vergeben wurde.

Die übrigen Objekte $\check{o} \in \check{A}_{t-1}$, die aufgrund der zu geringen Ähnlichkeit noch nicht einer Objektpaarung $B_t \mapsto \check{A}_{t-1}$, $B_t \mapsto \check{A}_{t-1}$ oder $\check{B}_t \mapsto \check{A}_{t-1}$ zugeordnet wurden, müssen demnach Objekte sein, die den Überwachungsbereich wieder verlassen haben.

Diese Schlussfolgerung muss allerdings über einen ausreichend großen Abstand zu den Maschinenobjekten bzw. eine ausreichende Nähe zum Rand des Überwachungsbereiches auf Plausibilität geprüft werden. Genauso muss für die verbliebenen Objekten $o \in B_t$, die als neue Objekte in den Überwachungsbereich treten, eine hinreichende Nähe zum Randbereich vorhanden sein.

6 Zusammenfassung und Ausblick

In diesem Beitrag wurde das visuelle „Eye4Save"-Überwachungssystem für die barrierefreie Absicherung von Mensch-Maschine-Arbeitsplätzen vorgestellt. Mittels mehrerer Stereo-Farbkameras und einer dafür entwickelten Verarbeitungslösung werden Vordergrundobjekte bezüglich eines Hintergrundmodells nach den Kriterien Farbe und Tiefe detektiert. Die extrahierten Silhouetten der Vordergrundobjekte werden mittels Schnitt ihrer Volumenkegel in einen 3D-Voxelraum überführt. Durch Analyse der 3D-Objektrelationen werden kritische Situationen erkannt und durch den Eingriff in die Maschinensteuerung stoppen gefährliche Bewegungen für die Dauer des Aufenthaltes in Maschinennähe. Die nächsten Arbeiten werden der Umsetzung und Verbesserung des Tracking-Ansatzes, der Kombination der passiven 3D-Sensorik mit weiteren Prinzipien zur Erfassung von Tiefeninformationen sowie der Erhöhung von Sicherheit und Verfügbarkeit des Systems gewidmet sein.

Literatur

[1] Ebert, Dirk und Dominik Henrich: *Safe Human-Robot-Cooperation: Problem Analysis, System Concept and Fast Sensor Fusion.* In: *IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems, Baden-Baden, Germany*, August 2001.

[2] Fischer, Markus und Dominik Henrich: *Surveillance of Robots using Multiple Colour or Depth Cameras with Distributed Processing.* In: *Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2009), Aug 30*, 2009.

[3] Gassmann, F. und S. Müller: *Hintergrundschätzung für Aufgaben der Verkehrsflussanalyse mittels digitaler Bildverarbeitung.* In: *48. Internationales Wissenschaftliches Kolloquium, TU Ilmenau*, September 2003.

[4] Gecks, Thorsten und Dominik Henrich: *SIMERO: Camera Supervised Workspace for Service Robots.* In: *ASER 2004, 2nd Workshop on Advances in Service Robotics, Feldafing, Germany*, Mai 2004.

[5] Hader, S. und J. Haussmann: *Verfahren und Vorrichtung zum Überwachen eines dreidimensionalen Raumbereichs*, Juni 5 2008. DE Patent App. DE200,610,057,605.

[6] Hartley, Richard und Andrew Zisserman: *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521540518, Second Auflage, 2004.

[7] Jahn, Rainer, Rico Nestler, Karl-Heinz Franke und Dorit Merhof: *Color-Codebooks zur Vorder- und Hintergrundsegmentierung.* In: *Tagungsband 20. Workshop Farbbildverarbeitung.* German Color Group, September 2014. ISBN: 978-3-00-047016-5.

[8] Kapusi, Daniel, Martin Klewer und Peter Pharow: *Mehr Sicherheit für die industrielle Robotik.* SPS-Magazin, Seiten 96–97, Mai 2013.

[9] Kapusi, Daniel, Rico Nestler und Karl-Heinz Franke: *Farbhistogrammgestützte 3D-Objektverfolgung in einem Mensch-Roboter-Überwachungsszenario.* In: *Tagungsband*

20. Workshop Farbbildverarbeitung. German Color Group, September 2014. ISBN: 978-3-00-047016-5.

[10] Kapusi, Daniel, Philipp Prinke, Rainer Jahn, Karl-Heinz Franke und Rico Nestler: *Geometric and Colorimetric Calibration For Multiview Camera Adjustments.* In: *IWK - 55th. International Scientific Colloquium.* TU Ilmenau, September 2010. ISBN: 978-3-938843-53-6.

[11] Kapusi, Daniel, Philipp Prinke, Darko Vehar, Rainer Jahn, Karl-Heinz Franke und Rico Nestler: *Simultaneous Geometric and Colorimetric Cameracalibration.* In: *Tagungsband 16. Workshop Farbbildverarbeitung.* German Color Group, Oktober 2010. ISBN: 978-3-00-032504-5.

[12] Kim, K., T. H. Chalidabhongse, D. Harwood und L. Davis: *Real-time foreground-background segmentation using codebook model.* Real-Time Imaging, 11(3):172–185, 2005.

[13] Klumpp, W., M. Reichenbach, M. Schreiber, V. Zipter und M. Zürn: *Verfahren zum Überwachen eines Arbeitsbereichs und Überwachungsanordnung für einen Arbeitsbereich*, März 7 2013. DE Patent App. DE201,210,014,057.

[14] Koch, Torsten und Karl-Heinz Franke: *3D-Aspekte bei der Bildfusion am Beispiel von IR- und VIS-Daten.* In: *Proceedings zum Workshop "Automation durch integrierte Bildverarbeitung".* AT&M Technologie und Marketing GmbH, Europäisches Technologie Cluster "Integrated Machine Vision (EUTIST-IMV)", 2004.

[15] Konolige, Kurt: *Small vision system: Hardware and implementation.* In: *Proc. of the Intl. Symp. of Robotics Research (ISRR)*, Seiten 111–116, 1997.

[16] Krawetzke, Martin: *Relat. Objektbeschreibungen für die 3D-Szenenanalyse am Bsp. einer kamerabas. Überwachung von Roboterarbeitsplätzen.* Masterarbeit, TU Ilmenau, März 2014. bearbeitet beim ZBS e. V.

[17] Kuhn, Stefan und Dominik Henrich: *Multi-View Reconstruction of Unknown Objects in the Presence of Known Occlusions.* In: *5th International Symposium on Visual Computing (ISVC 2009)*, 2009.

[18] Normenausschuss Sicherheitstechnische Grundsätze und Normenausschuss Maschinenbau: *Sicherheit von Maschinen - Anordnung von Schutzeinrichtungen im Hinblick auf Annäherungsgeschwindigkeiten von Körperteilen (ISO 13855:2010); DeutscheFassung EN ISO 13855:2010.* Technischer Bericht, DIN Deutsches Institut für Normung e. V., Oktober 2010.

[19] Ober-Gecks, Antje, Maria Hänel, Dominik Henrich und Tobias Werner: *Fast multi-camera reconstruction and surveillance with human tracking and optimized camera configurations.* In: *8th German Conference on Robotics*, 2014.

[20] Referat IT-Systeme: *Zukunftsbild Industrie 4.0.* Technischer Bericht, Bundesministerium für Bildung und Forschung, November 2013.

[21] *Sim4Save - Intelligente Raumüberwachung für Industriearbeitsplätze*, 2013. Fraunhofer IDMT.

[22] Stauffer, C. und W. E. L. Grimson: *Adaptive background mixture models for real-time tracking.* IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2:2246–2252, 1999.

[23] Szeliski, Richard: *Rapid Octree Construction from Image Sequences.* CVGIP: Image Understanding, 58(1):23–32, July 1993.

[24] Vehar, Darko: *Metrische 3D-Rekonstruktion von natürlichen Szenen durch Bildanalyse binokularer Kameras.* Diplomarbeit, Technische Universität Ilmenau, 2011. bearbeitet beim ZBS e. V.