

# Farbhistogrammgestützte 3D-Objektverfolgung in einem Mensch-Roboter- Überwachungsszenario

Daniel Kapusi<sup>1</sup>, Rico Nestler<sup>2</sup>, Karl-Heinz Franke<sup>1</sup>

<sup>1</sup>Zentrum für Bild- und Signalverarbeitung e. V.

Werner-von-Siemens-Str. 10, D-98693 Ilmenau

eMail: daniel.kapusi@zbs-ilmenau.de

eMail: karl-heinz.franke@zbs-ilmenau.de

URL: <http://www.zbs-ilmenau.de>

<sup>2</sup>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. Bestandteil des Systems sind u. a. mehrere Stereo-Farbkameras, welche die abzusichernde Umgebung aus verschiedenen Perspektiven beobachten und in Analogie zum menschlichen räumlichen Sehen zusätzliche Tiefeninformationen liefern. Die Fusion der Bild- und Tiefendaten aus den unterschiedlichen Ansichten erlaubt die Detektion von Objekten im Raum und deren vollständige 3D-Erfassung. Allerdings werden dicht beieinanderliegende bzw. sich berührende Objekte nur als ein einzelnes verschmolzenes Objekt registriert, wodurch sämtliche Objektrelationen verloren gehen. Neben den charakterisierenden Merkmalen, wie Raumposition und Volumen ermöglicht insbesondere die Einbeziehung von Farbhistogrammen in die Überwachung eine robuste Verfolgung der Objekte über den zeitlichen Verlauf und in dem Zusammenhang die Erkennung von Vorkommnissen wie der Verschmelzung und dem Wiederzerfall von 3D-Objekten.

# 1 Einleitung

Auf der Hannover-Messe 2013 ist die Entwicklung zur vierten Stufe der industriellen Revolution eingeleitet worden. Unter dem Begriff „Industrie 4.0“ sollen sich bis zum Jahr 2025 intelligente Fabriken entwickeln, die sich durch weitgehende Integration von Kunden und Geschäftspartnern in Geschäfts- und Wertschöpfungsprozesse, die Verkopplung von Produktion und hochwertigen Dienstleistungen und durch eine starke Individualisierung der Produkte unter den Bedingungen einer hoch flexibilisierten (Großserien-)Produktion auszeichnen. Auch in der Produktion von morgen werden menschliche Erfahrungen, Kreativität und Fertigkeiten während der Produktionsprozesse benötigt. Der Mensch wird zukünftig von Cyber-physischen Systemen assistiert [21]. Dabei rücken die Arbeitsbereiche von Menschen und Maschinen immer näher zusammen, wodurch die Ansprüche an die Sicherheitsanforderungen an Mensch-Maschine-Umgebungen entsprechend steigen.

Um den hohen Sicherheitsanforderungen gerecht zu werden, werden derzeit vorwiegend physisch trennende Schutzeinrichtungen eingesetzt. Durch die räumliche Trennung zwischen sich bewegenden Maschinenteilen und Menschen durch entsprechende Schutzgitter werden Situationen, die zu Unfällen führen können, weitestgehend vermieden. Der Zugang zum Arbeitsbereich der Maschine erfolgt in der Regel über Schutztüren, die sich erst entriegeln lassen, wenn die Maschine sich mit einer als sicher definierten Geschwindigkeit bewegt bzw. gestoppt wurde. Stopp und Wiederanlauf der Maschine müssen in dem Fall manuell betätigt werden, was wiederum insbesondere bei häufigem Zutritt sich zu sehr langen zusätzlichen Standzeiten aufsummieren kann. Da die Phase des Stillstandes in den meisten Fällen negative Auswirkungen auf die gesamte Produktionskette hat, werden, wenn häufige Zugänge zur Maschine notwendig sind, berührungslos wirkende Schutzeinrichtungen, wie Lichtvorhänge, Laserscanner und Kameras, die ausgewiesene Zugangsbereiche überwachen und ggf. einen Maschinenstopp auslösen, bevorzugt. Darüber hinaus sind auch noch taktile, wie zum Beispiel druckempfindliche Trittmatten und künstliche Roboterhaut, oder akustische Systeme für die Gewinnung zusätzlicher Redundanz im Einsatz.

Andere Entwicklungen der kamerabasierten Überwachung setzen auf eine tiefenbildbasierte Beobachtung aus der Vogelperspektive [7, 14]. In diesem Ansatz wird ein virtueller Überwachungskegel über den zu überwachenden Bereich gestülpt, wodurch auf weitere trennende Schutzeinrichtungen weitestgehend verzichtet werden kann. Im Arbeitsraum wird praktisch eine maximale Zugänglichkeit zur Maschine erreicht, was die Akzeptanz und die Effektivität des Mensch-Maschine-Arbeitsplatzes steigert. Der virtuelle Überwachungskegel selbst wird dabei in frei definierbare Schutzzonen unterteilt. Es werden keine Objekte an sich detektiert, sondern lediglich Schutzzonen aktiviert, bei denen das Vorhandensein einer Mindestmenge von 3D-Punkten registriert wurde. In diesem Fall wird dann ggf. in die Maschinensteuerung eingegriffen. Der Wiederanlauf der Maschine nach dem Verlassen des Gefahrenbereiches muss allerdings auch hier manuell erfolgen, da der Wirkungsbereich des Roboters, in die der Mensch theoretisch eindringen kann, von der Überwachung bewusst ausgelassen werden muss, da sonst bereits seine Eigenbewegungen einen Maschinenstopp auslösen würden. Ein anderer Ansatz ist die kamerabasierte Überwachung aus mehreren unterschiedlichen

Perspektiven. Im Gegensatz zur bereits erwähnten monoperspektivischen Überwachung aus der Vogelperspektive mittels Tiefenkamera, wobei der Raum nur z-eindeutig bezüglich der Kameraperspektive erfasst wird, kann durch den Ansatz mit verschiedenen Ansichten die Szene prinzipiell vollständig dreidimensional erfasst werden. Dadurch lassen sich grundsätzlich auch Objekte detektieren, die aus einer Perspektive durch andere statische oder dynamische Objekte verdeckt sind. An der Universität Bayreuth wird seit 2001 an einem solchen multiperspektivischen Ansatz für die sichere Mensch-Roboter-Kooperation geforscht [2]. Es werden mehrere Grauwert- [6] bzw. Farbkameras oder auch Tiefenkameras [4] eingesetzt, die simultan Bilder aufzeichnen. In den Bildern werden die Objekte zunächst mittels einer grauwert-, farb- bzw. 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.

Seit 2009 wird auch am ZBS e. V. an einem bildbasierten Überwachungssystem für die Bewegungssteuerung von Robotern im sicheren Arbeitsprozess mit Menschen gearbeitet. Gestartet wurden die Forschungs- und Entwicklungsarbeiten im Rahmen eines vom Europäischen Fond für regionale Entwicklung geförderten zweijährigen Verbundprojektes (2008FE9063) zusammen mit Thüringer Unternehmen [10]. Unter dem Namen „Eye4Save“ [3, 11] wurde das Überwachungssystem erstmals 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-Farbkameras beobachtet. Diese sind dabei so angeordnet, dass die Schnittmenge der einzelnen Sichtkegel aus den verschiedenen Blickwinkeln im zu überwachenden Bereich möglichst groß ist, die Verdeckungen durch dynamische und statische Objekten minimiert werden und alle Bereiche in der Überwachungszone von einer vorgegebenen Mindestanzahl von Kameras einsehbar sind. Im Vorfeld der Installation, empfiehlt es sich, die optimalen Kamerapositionen und -ausrichtungen z. B. interaktiv mittels Simulation zu ermitteln. Dazu kann das vom Fraunhofer Institut für Digitale Medien (IDMT) entwickelte „Sim4Save“-Tool [22] eingesetzt werden. In [20] wird u. a. auch ein automatisches Optimierungsverfahren zur Ermittlung der optimalen Kamerapositionen und -ausrichtungen vorgestellt. Für ein Testszenario mit einem im Arbeitsraum befindlichen Industrieroboter wurde eine Konfiguration mit vier Stereo-Farbkameras gewählt, die in die oberen Raumecken montiert sind und den Arbeitsraum aus einer schrägen Aufsicht heraus beobachten.

Der Kameraverbund wird so gesteuert, dass die Bildaufnahme einer Kamera immer absolut zeitgleich mit den anderen erfolgen kann. Als Verarbeitungseinheit wird ein mehrkerniger Industrie-PC verwendet.

In der Arbeitsphase wird aus den ermittelten Objektrelationen in Verbindung mit der gemessenen Laufzeit von der Bildaufnahme bis zum Ergebnis jeweils eine neue

Zeitvorgabe unter Einbeziehung der DIN EN ISO 13855 [19] ermittelt. Innerhalb des berechneten Zeitlimits muss die nächste Aktualisierung des Ergebnisses erfolgt sein. Ein unabhängiger Watchdog stoppt das System ggf. bei einer Überschreitung. Das System erfüllt nach diesem Prinzip harte Echtzeitanforderungen mit flexiblen Zeitschranken. Im Rahmen des Einrichtungsprozesses werden die Kameras bezüglich der inneren und äußeren Kameraparameter sowie auch colorimetrisch mit Hilfe eines Schachbretttargets kalibriert [12, 13]. Im weiteren Verlauf des Einrichtungsprozesses werden das Anlernen bzw. die interaktive Definition des Hintergrundbereiches durchgeführt.

Aus den aufgenommenen, verzeichnungskorrigierten und rektifizierten Bildern [8] der Stereo-Farbkameras wird pro Kamera paar ein Tiefenbild [15] berechnet. Mit Hilfe der lokalen Tiefen- und Farbinformationen werden 2D-Vordergrundobjekte separat für jede Stereo-Farbkamera bezüglich des Hintergrundes getrennt [9].

Die gefundenen 2D-Objekte aller Perspektiven werden in Form ihrer visuellen Hüllen mittels des Volumenschnittverfahrens [24] 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.



Abbildung 1: Mensch-Roboter-Szene

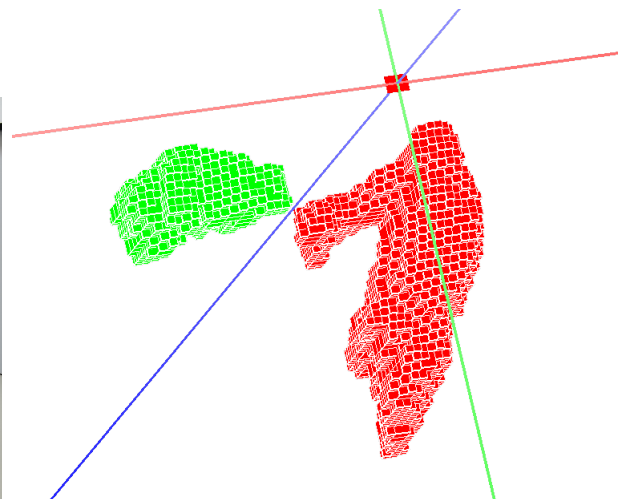


Abbildung 2: Voxelrepräsentation

## 3 Objektklassifikation und -verfolgung

### 3.1 Konzept

Um eine Szene im Mensch-Maschine-Arbeitsraum hinsichtlich der Einhaltung von Sicherheitsanforderungen beurteilen zu können, ist eine Klassifikation der Objekte, von denen eine potentielle Gefahr ausgeht und von Objekten für die Ausnahmeregelungen gelten sollen, da sie ebenfalls zum Inventar gehören, notwendig. Im einfachen Fall stationärer Objekte (z. B. Industrieroboter mit Schwenkarm) kann dies bereits unter Einbeziehung der Position im Arbeitsraum erfolgen, die vom jeweiligen stationären Objekt in jedem Fall exklusiv eingenommen wird. Befinden sich jedoch Objekte sehr nahe bei einander oder berühren sie sich gar, dann werden sie vom Überwachungssystem nur als einzelnes Objekt mit entsprechend vergrößertem Volumen wahrgenommen. Eine zusätzliche schwellwertbasierte Volumenüberwachung würde an dieser Stelle bereits eine einfache Erkennung von Maschinenberührungen ohne Einbeziehung eines zeitlichen Kontextes erlauben. Diese Variante sollte allerdings nur eine Option sein, die zusätzliche Redundanz in das Überwachungssystem bringt. Mehr Sicherheit verspricht die Erkennung und die Behandlung von verschmolzen bzw. zerfallenen Objekten, die eine temporale Auswertung der Szene erforderlich macht.

Die Klassifikation der Inventarobjekte muss aufgrund des sicherheitskritischen Aspektes der Überwachungsanwendung absolut zuverlässig sein. Eine Fehlklassifikation könnte fatale Folgen haben, da in dem Fall ggf. die Sicherheitsabstände zwischen den falschen Objekten überwacht würden. Daher sollen diese Objekte im Arbeitsraum nicht vollständig automatisch klassifiziert, sondern von einem menschlichen Bediener einmalig manuell zugeordnet werden. Die einmal so gewonnene sichere initiale Zuordnung soll im weiteren zeitlichen Verlauf durch die Verfolgung der Objekte aufrechterhalten werden. Das bedingt, dass die initial klassifizierten Objekte den Überwachungsbereich der Kameras nicht verlassen, was bei den stationären Maschinenobjekten auch nicht zu erwarten ist. Die Klassifikation aller weiteren gegenständlichen Inventarobjekte im Raum zum Zeitpunkt des Einlernvorganges erlaubt die Anwendung von Ausnahmeregelungen, etwa für Objekte, die zum Produktionsprozess gehören und sich daher zeitweilig sehr nahe an den Maschinen befinden dürfen. Für alle weiteren unbekannt Objekte (Menschen), die während der Arbeitsphase von außen in den Arbeitsbereich eindringen und die somit keine interaktive Klassenzuordnung aufweisen, sind keine Ausnahmeregelungen zulässig. Für sie gelten alle Sicherheitsanforderungen ohne Einschränkungen.

### 3.2 Möglichkeiten der Objektzuordnung

In diesem Abschnitt werden die Varianten aufgezeigt, die bei der Zuordnung eines Objektes im aktuellen Zeitpunkt zu einem Objekt im vorherigen Zeitpunkt auftreten können. 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 unbekannt 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  $\check{o}_{j\dots l} \in \check{A}_{t-1}$  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  $o_{i\dots k} \in B_t$ .

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

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 eine direkte Objektzuordnung,
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  $(m(2^n - n - 1) + n(2^m - 1))$  verschiedene Kombinationsmöglichkeiten für die Zuordnungen. Aus der Menge der zugeordneten Objekte  $\forall \check{o} \in \check{A}_{t-1}$  werden  $(2^m - m - 1)$  Kandidatenobjekte

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

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

$$\check{\check{B}}_t := \mathcal{P}(B_t) \setminus B_t \quad (2)$$

für alle möglichen Objektzerfälle. Neue Kandidatenobjekte werden durch Kombination mit allen Einzelobjekten innerhalb der jeweiligen Mengen  $\check{\check{A}}_{t-1}$  bzw.  $\check{\check{B}}_t$  durch Vereinigung der korrespondierenden Voxelmengen und der damit zusammenhängenden Objekteigenschaften gebildet. Ein Kandidatenobjekt für die Verschmelzung bzw. den Zerfall kann aus zwei und mehr Einzelobjekten bestehen. Alle zum Kandidatenobjekt 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 diese Stelle bereits verworfen werden. Das senkt die Anzahl der möglichen Kombinationen. Die Anzahl der verbliebenen plausiblen Verschmelzungskandidatenobjekte  $\check{\check{A}}_{t-1}$  ist  $\check{m} \in \mathbb{N} \mid 0 \leq \check{m} \leq 2^m - m - 1$ . Die Anzahl der verbliebenen plausiblen Zerfallskandidatenobjekte  $\check{\check{B}}_t$  ist  $\check{n} \in \mathbb{N} \mid 0 \leq \check{n} \leq 2^n - n - 1$ .

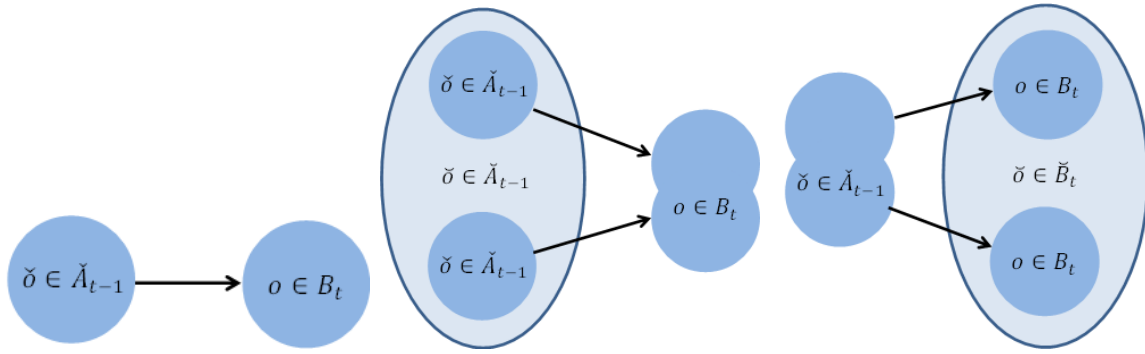


Abbildung 3: Varianten der Objektzuordnung (Direktpaarung, Verschmelzung, Zerfall)

### 3.3 Korrespondenzsuche

Zwischen allen möglichen  $mn$  Objekteinzelpaarungen  $B_t \mapsto \check{A}_{t-1}$  und plausiblen  $\check{m}\check{n}$  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 (3) substituiert jeweils die Merkmale  $xyz, V, H$ .

$$D_{feat} := \left\{ \begin{array}{l} d_{feat}(\check{o}, o), \forall \check{o} \in \check{A}_{t-1} \forall o \in B_t \\ d_{feat}(\check{o}, o), \forall \check{o} \in \check{A}_{t-1} \forall o \in B_t \\ d_{feat}(\check{o}, o), \forall \check{o} \in \check{A}_{t-1} \forall o \in \check{B}_t \end{array} \right\} \begin{array}{l} \text{Direktpaarung} \\ \text{Verschmelzung} \\ \text{Zerfall} \end{array} \quad (3)$$

Euklidische Distanz zwischen den Schwerpunkten  $\underline{s}_{xyz}(\check{o}_i)$  und  $\underline{s}_{xyz}(o_j)$  ist:

$$d_{xyz}(\check{o}_i, o_j) = \|\underline{s}_{xyz}(\check{o}_i) - \underline{s}_{xyz}(o_j)\|_2 \quad (4)$$

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

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

$$d_V(\check{o}_i, o_j) = |V(\check{o}_i) - V(o_j)| \quad (5)$$

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

In der Arbeit von Krawetzke[16] wurden das sehr schnelle Color-Indexing Verfahren [23] zur Berechnung der Farbhistogramme mit acht Bins verwendet und Distanzmaße für das histogrammbasierte Farbmatching experimentell untersucht. Im Vergleich mit den Distanzmaßen *Sum of Squared Differences (SSD)*, *Bhattacharyya* und *Histogrammschnitt* konnten mit dem *Chi-Quadrat-Test* im Mensch-Maschine-Testumfeld für die 8-Bin-Histogramme die signifikantesten Ergebnisse erzielt werden.

Die Distanz zwischen Farbhistogrammen mit  $nb \in \mathbb{N}^*$  Bins nach dem *Chi-Quadrat-Test* ist:

$$d_H(\check{o}_i, o_j) = d_{H_{\chi^2}}(\check{o}_i, o_j) = \sum_{bin=1}^{nb} \frac{H_{\chi^2, bin}(o_j) - H_{\chi^2, bin}(\check{o}_i)}{H_{\chi^2, bin}(\check{o}_i)} \quad (6)$$

$$d_H(\check{o}_i, o_j) \in D_H$$

Beim einfachen uniformen Binning treten prinzipbedingt Quantisierungsfehler auf, die umso größer sind, je weniger Bins verwendet werden. Die willkürlich gesetzten Klassengrenzen interferieren mit den Farbvalenzen der Pixel und sprunghafte Umschläge von einem Bin zum benachbarten sind schon bei kleinsten Farbveränderungen möglich. Um den entgegen zu wirken, werden in [5] die Verwendung von unscharfen Histogrammen [18] bzw. graphenbasierte Ansätze [1] zum histogrammbasierten Farbmatching empfohlen.

Aufgrund der unterschiedlichen Wertebereiche werden alle Distanzmaße auf Werte zwischen 0 und 1 normiert, damit ein allgemeines Distanzmaß ermittelt werden kann, welches alle Merkmale vereint. Der Ausdruck  $feat$  in (7) steht auch hier wie schon in (3) stellvertretend für die verwendeten Merkmale  $xyz, V, H$ .

$$d_{norm}(D_{feat}, i, j) = \begin{cases} \frac{d_{feat}(\check{o}_i, o_j) - \min(D_{feat})}{\max(D_{feat}) - \min(D_{feat})}, & \max(D_{feat}) \neq \min(D_{feat}) \\ 0, & \text{sonst} \end{cases} \quad (7)$$

$$d_{xyz,norm}(\check{o}_j, o_i) = d_{norm}(D_{xyz}, i, j) \quad (8)$$

$$d_{V,norm}(\check{o}_j, o_i) = d_{norm}(D_V, i, j) \quad (9)$$

$$d_{H,norm}(\check{o}_j, o_i) = d_{norm}(D_H, i, j) \quad (10)$$

Die normierten Distanzmerkmale werden mit  $\alpha_{xyz}, \alpha_V, \alpha_H \in \mathbb{R}_+ | \alpha_{xyz} + \alpha_V + \alpha_H > 0$  gewichtet. In experimentellen Untersuchungen [16] wies das Volumenmerkmal  $V$  die größten Schwankungen gegenüber den Merkmalen Schwerpunktposition  $s_{xyz}$  und Farbhistogramm  $H$  auf. Es wird empfohlen, den Volumenwichtungsfaktor gegenüber den anderen Wichtungsfaktoren zu verkleinern, z. B.  $\alpha_V = 0.5\alpha_{xyz} = 0.5\alpha_H = 0.5$ . Das Distanzmaß  $d_{total}(\check{o}_i, o_j)$  über alle Merkmale wird dann wie folgt berechnet:

$$d_{total}(\check{o}_i, o_j) = \sqrt{\frac{\alpha_{xyz}d_{xyz,norm}(\check{o}_j, o_i)^2 + \alpha_V d_{V,norm}(\check{o}_j, o_i)^2 + \alpha_H d_{H,norm}(\check{o}_j, o_i)^2}{\alpha_{xyz} + \alpha_V + \alpha_H}} \quad (11)$$

$$d_{total}(\check{o}_i, o_j) \in \mathbb{R}_+ | 0 \leq d_{total}(\check{o}_i, o_j) \leq 1$$

Die berechneten Distanzmaße  $\forall d_{total} \in D_{total}$  aller Objektpaarungen werden in einer Liste der Größe nach aufsteigend sortiert, so dass der niedrigste Distanzwert, der demzufolge mit der ähnlichsten Objektpaarung korrespondiert, an der Spitze der Liste steht. 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 bereits vergeben wurde. Die übrigen Objekte  $\check{o} \in \check{A}_{t-1}$ , die aufgrund der zu geringen Ähnlichkeit noch nicht an eine 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 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. Ansonsten liegt ein Fehlerfall vor. Genauso verhält es sich mit den verbliebenen Objekten  $o \in B_t$ , die als neue Objekte in den Überwachungsbereich treten. Auch hier muss eine Nähe zum Randbereich vorhanden sein, andernfalls ist ein plötzliches Erscheinen mitten im Raum nicht plausibel.

## 4 Automatisches Wiederanfahren

Das automatische Wiederanfahren einer Maschine darf nur erfolgen, wenn absolut ausgeschlossen ist, dass sich keine Menschen mehr in ihrem Wirkungsbereich befinden. Im Fall, dass mehrere Personen sehr eng beisammen in den Überwachungsbereich hineintreten, d. h. sie werden vom Überwachungssystem wahrgenommen als ein verschmolzenes Objekt, und sich der inzwischen gestoppten Maschine bis zur Berührung nähern, darf die Maschine unter keinen Umständen wieder automatisch anfahren, solange sich nicht alle der Personen wieder von der Maschine entfernt haben.

Daher sollte für Maschinen- und sonstige Inventarobjekte, für die ja ggf. Ausnahmeregelungen für die Sicherheitsabstände gelten, die Rückkehr zu ihrem ursprünglichen Zustand (d. h. sie sind mit keinem anderen Objekt mehr verschmolzen) erst nach Verifikation der Ähnlichkeit mit ihrem gespeicherten Urzustand möglich sein. Für die Prüfung dieser Selbstähnlichkeit sollten die Merkmale entsprechend ihres diskriminativen Einflusses noch vor der Verschmelzung gewichtet werden, um beim Wiederzerfall eindeutig den ursprünglichen Zustand der Objekte wieder erkennen zu können.

Nähert sich beispielsweise eine Person im Blaumann einer ebenfalls blauen Maschine, dann ist das Merkmal Farbe wenig aussagekräftig. Der Einfluss des Farbhistogramms sollte dann entsprechend niedrig gewichtet sein, damit andere Merkmale, wie z. B. das Volumen, eine eindeutige Entscheidung erlauben. Sind keine der Merkmale ausreichend diskriminativ, dann darf nach dem Zerfall eines verschmolzenen Objektes das Rücksetzen auf den ursprüngliche Zustand nur durch einen manuellen Eingriff erfolgen, nach dem sichergestellt ist, dass der Wirkungsbereich der Maschine frei von Personen ist.

## 5 Zusammenfassung und Ausblick

Ausgangspunkt für das vorgestellte Konzept ist ein kamerabasiertes Überwachungssystem für Mensch-Maschine-Arbeitsplätze, welches Objekte im Raum in Form von Voxelmengen ohne jeden zeitlichen Bezug detektiert. Daran soll nun eine zeitliche Verfolgung von zum einen interaktiv markierten Inventarobjekten (Maschinen und sonstige Gegenstände in der Szene) sowie unbekanntem Objekten (Eintritt in die Szene erst während der Arbeitsphase, z. B. Personen) erfolgen. Die Objekte werden vom jeweils aktuellen Zeitschritt zu den bereits zeitlich verfolgten Objekten im vorherigen Zeitschritt zugeordnet, wobei aus den Merkmalen Schwerpunktposition der Voxelmenge, Volumen der Voxelmenge sowie Farbhistogramm ein gemeinsames Maß für die Objektähnlichkeit berechnet wird. Außerdem werden Hypothesen für mögliche Verschmelzungen von Objekten bzw. den Zerfall von Objekten aufgestellt und diese neben den direkten Objekt-zu-Objekt-Zuordnungen ebenfalls untersucht. In Form eines Rankings werden dann in der Reihenfolge der besten Ähnlichkeit die Objektpaarungen miteinander verknüpft, wobei eine vorgegebene Mindestähnlichkeit eingehalten werden muss.

Die genannten Überlegungen sollen in den „Eye4Save“-Demonstrator umgesetzt und evaluiert werden. Es ist zu erwarten, dass insbesondere durch die Einbeziehung der Informationen aus den Farbhistogrammen eine sehr robuste Verfolgung der Objekte über die Zeit gewährleistet sein wird. Durch die Erschließung des zeitlichen Zusammenhan-

ges sollen sicherheitskritische Mängel ausgeräumt werden, wie z. B. die Unwissenheit über eventuell miteinander verschmolzene Objekte, die sich aus der ausschließlichen Auswertung der Momentaufnahme ohne jeglichen zeitlichen Bezug ergeben. In der DIN EN ISO 13855 [19] wird eine konstante Schrittgeschwindigkeit definiert, deren Annahme maßgeblich den erforderlichen Sicherheitsabstand für Systeme ohne Zeitbezug beeinflusst. Durch die Auswertung des zeitlichen Verlaufes ist es möglich die realen relativen Objektgeschwindigkeiten zu schätzen. So könnten durch die Auswertung von Beschleunigungs- und Geschwindigkeitsvektoren auch nicht intendierte Gesten, wie z. B. Stolpern, die auf Gefahrensituationen hindeuten, vorhergesagt und die Sicherheit von Mensch-Maschine-Arbeitsplätzen über die Norm hinaus erhöht werden.

## Literatur

- [1] BALTHASAR, DIRK und VOLKER REHRMANN: *Robustes histogrammbasiertes Farbmatching*. In: *Tagungsband 5. Workshop Farbbildverarbeitung*. German Color Group, Oktober 1999.
- [2] 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*.
- [3] *Eye4Save - Bildbasiertes 3D-Überwachungssystem für die sichere Mensch-Maschine-Kooperation*, 2013. ZBS e. V.
- [4] 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*.
- [5] FRANKE, KARL-HEINZ: *Grundlagen der Farbbildverarbeitung: Histogrammbasiertes Farbmatching*, 2012. VL-Script, TU Ilmenau, Fak. IA, FG GDV, ISSN 1432-3346.
- [6] 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*.
- [7] HADER, S. und J. HAUSSMANN: *Verfahren und Vorrichtung zum Überwachen eines dreidimensionalen Raumbereichs*, Juni 5 2008. DE Patent App. DE200,610,057,605.
- [8] HARTLEY, RICHARD und ANDREW ZISSERMAN: *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, Second Auflage, 2004.

- [9] 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.
- [10] KAPUSI, DANIEL und RAINER JAHN: *Bildbasiertes Live-Überwachungssystem für die sichere Kooperation von Mensch und Maschine*. Technischer Bericht, ZBS e. V., August 2011.
- [11] KAPUSI, DANIEL, MARTIN KLEWER und PETER PHAROW: *Mehr Sicherheit für die industrielle Robotik*. SPS-Magazin, Seiten 96–97, Mai 2013.
- [12] 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.
- [13] 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.
- [14] 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.
- [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] LAMBERT, P., N. HERVEY und H. GRECU: *Image Retrieval using Spatial Chromatic Histograms*. CCGIV, 2004(1):343–347, 2004.
- [19] NORMENAUSSCHUSS SICHERHEITSTECHNISCHE GRUNDSÄTZE und NORMENAUSSCHUSS MASCHINENBAU: *Sicherheit von Maschinen - Anordnung von Schutzeinrichtungen im Hinblick auf Annäherungsgeschwindigkeiten von Körperteilen (ISO 13855:2010); Deutsche Fassung EN ISO 13855:2010*. Technischer Bericht, DIN Deutsches Institut für Normung e. V., Oktober 2010.
- [20] 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.

- [21] REFERAT IT-SYSTEME: *Zukunftsbild Industrie 4.0*. Technischer Bericht, Bundesministerium für Bildung und Forschung, November 2013.
- [22] *Sim4Save - Intelligente Raumüberwachung für Industriearbeitsplätze*, 2013. Fraunhofer IDMT.
- [23] SWAIN, MICHAELJ. und DANAH. BALLARD: *Color indexing*. International Journal of Computer Vision, 7(1):11–32, 1991.
- [24] SZELISKI, RICHARD: *Rapid Octree Construction from Image Sequences*. CVGIP: Image Understanding, 58(1):23–32, July 1993.