Kollisionsvermeidung durch raum-zeitliche Bildanalyse

Collision Avoidance based on Space-Time Image Analysis

Uwe Franke, Clemens Rabe, Stefan Gehrig, DaimlerChrysler AG, Sindelfingen

Zusammenfassung Mehr als 1/3 aller Unfälle mit Personenschäden passieren im städtischen Bereich, primär an Kreuzungen. Eine Unterstützung des Fahrers durch geeignete Assistenzsysteme erfordert das Verstehen dieser sehr komplexen Situationen, insbesondere das sichere Erkennen anderer bewegter Verkehrsteilnehmer. Der Beitrag zeigt, wie man durch eine geschickte Fusion von Stereosehen und Bewegungswahrnehmung zu einer robusten und schnellen Detektion relevanter bewegter Objekte kommt. Dabei schätzt das als 6D-Vision bezeichnete Verfahren simultan Ort und Bewegung einzelner Bildpunkte und erlaubt somit eine Detektion bewegter Objekte bereits auf Pixelebene. Unter Verwendung eines Kalman-Filters propagiert der Algorithmus die aktuelle Interpretation ins nächste Bild, sodass er sich in Echtzeit darstellen lässt. Beispiele kritischer Situationen im Innenstadtbereich verdeutlichen die Leistungsfähigkeit des 6D-Vision-Prinzips, das auch im Bereich der mobilen Roboter wertvolle Beiträge leisten kann. ►►► **Summary** More than one third of all traffic accidents with injuries occur in urban areas, especially at intersections. A suitable driver assistance system for such complex situations requires the understanding of the scene, in particular a reliable detection of other moving traffic participants. This contribution shows how a robust and fast detection of relevant moving objects is obtained by a smart combination of stereo vision and motion analysis. This approach, called 6D Vision, estimates location and motion of pixels simultaneously which enables a detection of moving objects on a pixel level. Using a Kalman Filter, the algorithm propagates the current interpretation to the next image. Hence a real-time implementation is achieved. Examples of critical situations in urban areas exhibit the potential of the 6D Vision concept which can also be extended to robotics applications.

KEYWORDS I.2.9 [Robotics] Autonomous Vehicles, I.2.10 [Vision and Scene Understanding] 3D/Stereo Scene Analysis, Motion, I.4.8 [Scene Analysis] Depth Cues, Motion, Object Recognition, Stereo, I.5.4 [Applications] Computer Vision

1 Einführung

Ein Blick auf die Unfallstatistik in Bild 1 zeigt deutlich drei Hauptursachen für Unfälle mit Personenschaden. Es sind dies das Abkommen von der Fahrbahn, Auffahrunfälle und Kollisionen an Kreuzungen. In der Fachwelt besteht Einigkeit, dass die trotz aller Erfolge der passiven Sicherheit immer noch große Zahl schwerer Unfälle nur durch umgebungserfassende Fahrerassistenzsystem nachhaltig reduziert werden kann.

Für die beiden ersten Situationsklassen gibt es bereits heute wirksame Unterstützungssysteme. Für das Abkommen von der Fahrbahn – meist aus Unaufmerksamkeit des Fahrers – hat sich das videobasierte Lane-Departure-Warning im Nutzfahrzeugbereich etabliert. Seit 2000 wird für den Mercedes-Benz Actros (schwere Klasse NFZ) der „Spurassistent“ angeboten, der den Fahrer im Fall einer Gefahr akustisch seitenrichtig warnt.

Für das Erkennen potenzieller Auffahrsituationen im fließenden Verkehr eignet sich besonders die Radarsensorik. Moderne ACC-Systeme in Fahrzeugen der Oberklasse warnen nicht nur den Fahrer, sondern greifen auch wirkungsvoll in die Bremse ein. Seit Juni 2006 verfügt wiederum der Mercedes-Benz Actros als weltweit erstes Fahrzeug über ein automatisches Notbremssystem, das nach akustischer und haptischer Fahrerwarnung sogar eine Vollbremsung ausführt.

Der dritte Unfallschwerpunkt ist eindeutig der Kreuzungsbereich. Es mag auf den ersten Blick überraschen, dass sich hier 35% aller Unfälle mit Personenschaden ereignen und 17% aller im Straßenverkehr getöteten Personen zu beklagen

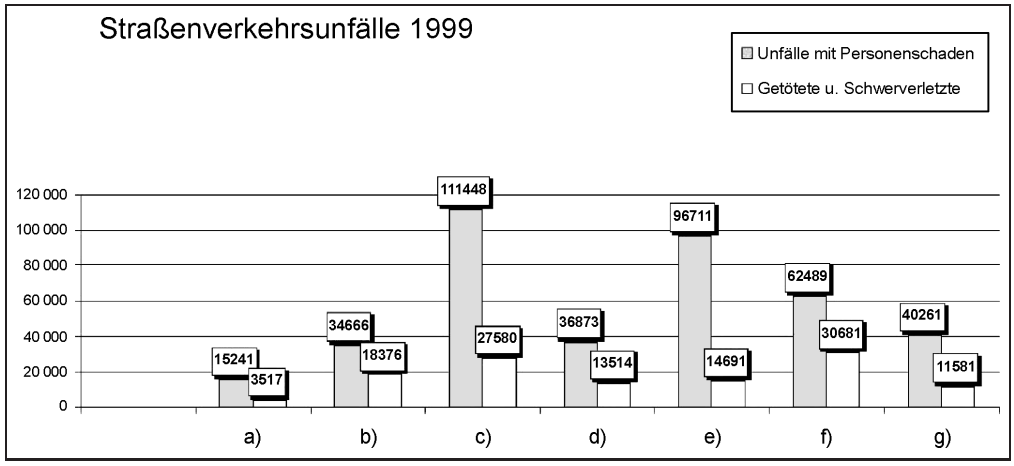

Bild 1 Unfallstatistik des statistischen Bundesamtes für Deutschland 1999 [3]. Dargestellt sind die Anzahl der Unfälle mit Personenschaden sowie die Anzahl an Getöteten und Schwerverletzten für einen Zusammenstoß mit einem anderen Fahrzeug, das seitlich in gleicher Richtung fährt (a), entgegenkommt (b), einbiegt oder kreuzt (c). Daneben sind die Zahlen für den Fall eines Zusammenstoßes zwischen Fahrzeug und Fußgänger (d), Auffahrunfälle (e), Abkommen von der Fahrbahn (f) sowie Unfälle anderer Art (g) angegeben.

sind. Die meisten Unfälle passieren hier beim Ein- und Abbiegen, beim Kreuzen sowie durch Missachtung einer roten Ampel oder eines Stoppschilds. Schuld daran sind nach im Rahmen öffentlich geförderter Projekte (z. B. INVENT, PREVENT) durchgeführten Unfallanalysen in zwei von drei Fällen die Faktoren Ablenkung, Unaufmerksamkeit und Fehleinschätzung. Dahinter steht die Tatsache, dass Kreuzungssituationen – verglichen mit Autobahnen – extrem komplex und ungleich schwerer prädizierbar sind.

Dementsprechend hoch sind die Anforderungen an ein wirkungsvolles Assistenzsystem. Es beginnt mit dem Verstehen der Infrastruktur, also Ampeln, Verkehrszeichen, Zebrastreifen, Richtungspfeile, Kreuzungsgeometrie usw. Des Weiteren müssen alle relevanten Verkehrsteilnehmer „gesehen“ und ihre Bewegungen abgeschätzt werden, um potenzielle Kollisionen voraussagen zu können. In einem letzten Schritt muss die aktuelle Verkehrssituation „verstanden“ werden, um die wahrscheinlichen Absichten der beteiligten Verkehrsteilnehmer abschätzen, Fehlverhalten antizipieren und wirkungsvoll drohende Kollisionen vermeiden zu können.

Da angesichts der Vielfalt der genannten Aufgaben der Kamera eine zentrale Bedeutung beim Erkennen gefährlicher Situationen zukommt, beschäftigt sich die DaimlerChrysler Forschung seit nunmehr 10 Jahren mit dem videobasierten Verstehen des innerstädtischen Verkehrs [8]. Auf dem Gebiet der Infrastrukturerkennung aus dem Fahrzeug heraus sind die Arbeiten weit fortgeschritten. Beispielsweise beschreibt Lindner [12] ein bildbasiertes System für die Erkennung von Ampeln, das in Kombination mit einer digitalen Karte Erkennungsraten von über 95% und Fehldetektionsraten kleiner 1/h erreicht. Für die gut untersuchte Aufgabe der Verkehrszeichenerkennung wird eine noch höhere Performanz erzielt.

Diese Module können dazu beitragen, den Fahrer vor dem fehlerhaften Einfahren in entsprechend geregelte Kreuzungen zu bewahren. Die größte Gefahr geht jedoch von anderen sich im Kreuzungsbereich bewegenden Verkehrsteilnehmern aus. Diese müssen von der eingesetzten umgebungserfassenden Sensorik schnell und sicher erkannt, ihre Position und Größe ermittelt und ihre aktuelle Bewegung nach Betrag und Richtung analysiert werden. Angesichts der Vielfalt möglicher Objekte wird ein Verfahren angestrebt, das alle potenziell gefährlichen oder gefährdeten Objekte unabhängig vom Aussehen und Blickwinkel erkennen kann.

Der vorliegende Beitrag stellt ein neues, leistungsfähiges Verfahren der Bildverarbeitung vor, das in der Lage ist, bewegte Objekte durch eine Kombination von räumlicher (Stereosehen) und zeitlicher (optischer Fluss) Bildanalyse in Echtzeit zu detektieren und zu vermessen. Nach einer Diskussion möglicher bekannter videobasierter Verfahren der Objektdetektion im nächsten Abschnitt beschreibt Abschnitt 3 das als 6D-Vision bezeichnete Prinzip. Die Leistungsfähigkeit dieses Ansatzes wird im Abschnitt 4 anhand zweier typischer Situationen aus dem städtischen Bereich gezeigt.

2 Detektion bewegter Objekte

2.1 Sehen mit einem oder zwei Augen?

Nicht ohne Grund hat die Evolution die Fähigkeit des stereoskopischen Sehens entwickelt. Der wesentliche Vorteil liegt in der Tatsache, dass bereits ein kurzer Blick vor unseren Augen ein dreidimensionales Abbild der Umgebung entstehen lässt. In vielen Fällen reicht bereits diese allein durch Stereo erhaltene 3D-Information für die Detektion von Hindernissen und ihre Vermessung aus.

Möchte man in technischen Sehsystemen aus Aufwandsgründen auf die notwendige zweite Kamera verzichten, gibt es eine Reihe von

Merkmalen wie z. B. Textur, Schattierung und Bewegung, die dem „Einäugigen“ Hinweise auf die Dreidimensionalität der Szene geben. Am erfolgsversprechensten ist es, durch Auswertung mehrerer aufeinander folgender Bilder mit Hilfe von Verfahren des „depth-from-motion“ die bei der Abbildung verloren gegangene Tiefeninformation zu rekonstruieren. Das setzt eine Bewegung des Beobachters voraus und gelingt zuverlässig nur bei stationären Objekten, die sich zudem nicht zu nahe an der optischen Achse befinden dürfen.

Bei bewegten Objekten können diese Ansätze zu nicht prädizierbaren Fehlern führen, solange keine weiteren Informationen vorhanden sind. Kommt beispielsweise ein Fahrzeug rechtwinklig mit konstanter Geschwindigkeit so auf uns zu, dass es zu einem Unfall kommen muss, wird im Bild keine Verschiebung wahrgenommen, was schnell als „unendlich weit entfernt“ und damit ungefährlich interpretiert wird.

Etwas günstiger ist die Situation bei der Detektion querbewegter Objekte, bei denen es nicht zu einer solchen „stehenden Peilung“ kommt. Bild 2 zeigt den optischen Fluss, d. h. die Verschiebung einzelner Bildpunkte zwischen zwei aufeinander folgenden Bildern. Deutlich hebt sich in diesem Bild das querende Auto ab. Gelingt es, den Fußpunkt des Objektes zu ermitteln, kann man bei bekannter Kamerageometrie und Winkel der optischen Achse zur Straße auf die Entfernung schließen. Flussbasierte Systeme für die Detektion vorausfahrender Fahrzeuge wurden im Rahmen von PROMETHEUS entwickelt und erprobt [7]. Leider erweist sich die Fußpunktbestimmung in der Praxis als problematisch. Die theoretischen Grenzen monokularer Verfahren werden in [11] aufgezeigt.

In Bereichen, in denen unser räumliches Sehen versagt, setzen wir deshalb Modellwissen ein, um aus der Größe gesehener Objekte auf die Entfernung zu schließen. Das gelingt uns Menschen scheinbar mühelos, impliziert aber die Fähigkeit eines technischen Systems, eine sehr große Zahl von Objekten verschiedenen Aussehens in einem großen Skalenbereich in der Bildfolge sicher zu detektieren. Im Extremfall muss dies auch bei Objekten geleistet werden, die der Rechner vorher noch nie „gesehen“ hat.

Aus den genannten Gründen sind die Autoren davon überzeugt, dass sich auch in den intelligenten Fahrzeugen der Zukunft das Prinzip des Stereosehens dauerhaft durchsetzen wird, da es die Detektion beliebiger Hindernisse in minimaler Zeit und mit einem Minimum an Annahmen ermöglicht.

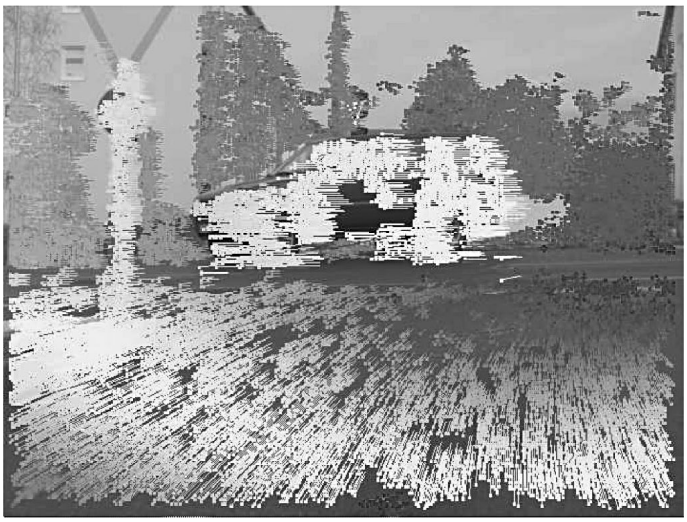

Bild 2 Darstellung des optischen Flusses. Deutlich erkennt man das durch die Eigenbewegung induzierte Expansionsfeld und das querende Fahrzeug. Das Flussfeld wurde mit Hilfe des in [14] beschriebenen echtzeitfähigen Verfahrens bestimmt.

2.2 Stereosehen

Das zentrale Problem des Stereosehens ist die so genannte Korrespondenzanalyse. Darunter versteht man das Finden korrespondierender Punkte eines Objektes in beiden Bildern des Stereokamerasystems. Sind die Parameter beider Kameras bekannt (neben den relativen Orientierungen zueinander u. a. auch die stets leicht unterschiedlichen Brennweiten und Verzeichnungsparameter der Optiken), reduziert sich die Aufgabe auf eine eindimensionale Suche entlang der so genannten Epipolarlinien. In modernen Systemen wird zur Vereinfachung stets eine so genannte Rektifizierung vorgenommen, die alle störenden Einflüsse eliminiert und die Suche in korrespondierenden Bildzeilen zulässt. Auch aus Rechenzeitgründen ist die Stereoanalyse also dem optischen Fluss überlegen, dessen Berechnung notwendigerweise eine ausgedehnte zweidimensionale Suche erfordert, vgl. Bild 2.

In der Praxis weit verbreitet sind korrelationsbasierte Verfahren der Korrespondenzanalyse, die meist an von einem Interestoperator gelieferten Stellen die Disparität, d. h. die Verschiebung zwischen rechtem und linkem Bild, ermitteln. Dies ist auf heutigen Standardprozessoren in Echtzeit möglich; parallel dazu existieren Hardwarelösungen. Die Entfernung L des betrachteten Bildpunktes ergibt sich bei rektifizierten Bildpaaren aus der Disparität d durch einfache Triangulation gemäß

$$L = \frac{fb}{d} \qquad (1)$$

Dabei bezeichnet b die Basisbreite, d. h. den Abstand der beiden Kameras und f ist die in Bildpunkten angegebene Brennweite.

Leitet man diese Beziehung nach der Disparität ab, zeigt sich, dass die Genauigkeit der Entfernungsschätzung quadratisch mit der Entfer-

Bild 3 Stereoskopische Analyse einer einfachen Kreuzungssituation unter der Verwendung des in [9] beschriebenen korrelationsbasierten Verfahrens. Der querende Fahrradfahrer befindet sich in einer Entfernung von 12 m zum Beobachter. In dieser Entfernung kann er gerade noch vom dahinter stehenden Auto getrennt werden.

nung abnimmt. In der Praxis erzielt man bei der Disparitätsschätzung Genauigkeiten von 0,2 pixel (häufig werden bessere Werte genannt, die aber einer genauen Überprüfung unter realen Bedingungen nicht Stand halten). Bei üblichen Basisbreiten von 25 cm und Standardobjektiven mit einem Öffnungswinkel von 40 Grad ergibt sich bei Verwendung von $1/3''$ Bildsensoren eine Genauigkeit der Entfernungsschätzung im interessanten Abstand von 30 m (entsprechend 2 Sekunden Fahrt bei 50 km/h) von $\pm 0{,}9$ m. Bei einem Abstand von 45 m steigt die Unsicherheit bereits auf ± 2 m.

Die Probleme, die die begrenzte Auflösung im Kreuzungsbereich impliziert, zeigt Bild 3. Der Fahrradfahrer befindet sich 12 m vor uns. In dieser Entfernung kann er gerade noch vom dahinter stehenden Fahrzeug getrennt werden, was bei größerem Abstand nicht mehr möglich ist. Eine Hindernisdetektion, die ausschließlich auf den 3D-Daten der Stereoanalyse agiert, würde in diesem Fall nur ein großes Objekt detektieren.

2.3 Raum-zeitliche Auswertung

Will man bewegte Objekte sicher von stationären Hindernissen trennen, erfordert dies eine zusätzliche Bewegungsanalyse. In der Literatur finden sich hierzu verschiedene Ansätze: Waxman und Duncan untersuchten schon 1986 die Beziehungen der optischen Flussfelder der linken und rechten Kamera [16]. Der daraus abgeleitete Relativfluss stellt eine Beziehung zwischen der Disparität und ihrer zeitlichen Änderung auf und ermöglicht die direkte Bestimmung der relativen Geschwindigkeit zu einem Objekt. Argyros et al. vergleichen in [1] den Normalen-Fluss einer Kamera mit dem Normalen-Fluss zwischen den beiden Stereokameras. Widersprüchliche Daten deuten auf bewegte Objekte hin und werden detektiert, jedoch nicht vermessen.

Als nachteilig erweist sich die Tatsache, dass nur direkt aufeinander folgende Bildpaare betrachten werden. Das impliziert eine große Empfindlichkeit gegenüber dem immer vorhandenen Messrauschen und Unsicherheiten in der Fahrzeugbewegung. Die Messgenauigkeit kann durch Vergrößerung des zeitlichen Abstandes der betrachteten Bildpaare erhöht werden, was aber nicht im Sinne einer schnellen Reaktion bei plötzlich auftauchenden Objekten ist.

Wünschenswert ist vielmehr ein Ansatz, der es erlaubt, durch „längeres Hinsehen“ die Situation immer genauer und zuverlässiger zu analysieren. Das beinhaltet nicht nur die Unterscheidung stationär/bewegt, sondern auch die möglichst präzise Schätzung von Ort und Bewegungsvektor kritischer Objekte. Dang et al. kombinieren in [5] die Bildverschiebung und Disparitätsmessung aller auf einem Objekt liegenden Punkte in einem Kalman Filter und schätzen dessen 3D-Position und Geschwindigkeit. Voraussetzung dafür ist jedoch eine korrekte Objektsegmentierung. Der im Folgenden beschriebene 6D-Vision-Ansatz umgeht dies durch die Bestimmung von Orts- und Bewegungsvektor auf Bildpunktebene.

3 Das 6D-Vision Prinzip

Versucht man die Bewegung eines weit entfernten Punktes anhand zweier zeitlich aufeinander folgend gemessener 3D-Positionen zu schätzen, ist die Unsicherheit wegen des in Abschnitt 2.2 erwähnten Stereo-Messrauschens zu groß, um vertrauenswürdige Aussagen tätigen zu können. Dies gilt insbesondere dann, wenn die Messungen in kurzen Abständen von 40–80 ms durchgeführt werden.

Die zentrale Idee des 6D-Vision besteht darin, im Sinne des „länger Hinsehens“ interessierende Bildpunkte über mehrere Bilder zu verfolgen und die beschriebene Unsicherheit durch zeitliche Integration kontinuierlich zu verringern, gleichzeitig aber zu jedem Zeitpunkt eine optimale Schätzung bereitzustellen. Hierzu wird angenommen, dass sich die Punkte als Teile massebehafteter Körper kurzfristig geradlinig im Raum bewegen.

Mathematisch elegant lässt sich diese Aufgabenstellung mit Hilfe des Kalman-Filters lösen. Dabei handelt es sich ein rekursives Verfahren zur Schätzung des Zustands eines Systems, das mit jeder Messung seinen geschätzten Zustandsvektor so korrigiert, dass die Varianz des Schätzfehlers minimiert wird (vgl. [17]). Dieser Zustandsvektor besteht aus sechs Komponenten: den drei Ortskoordinaten $(X, Y, Z)^T$ und den

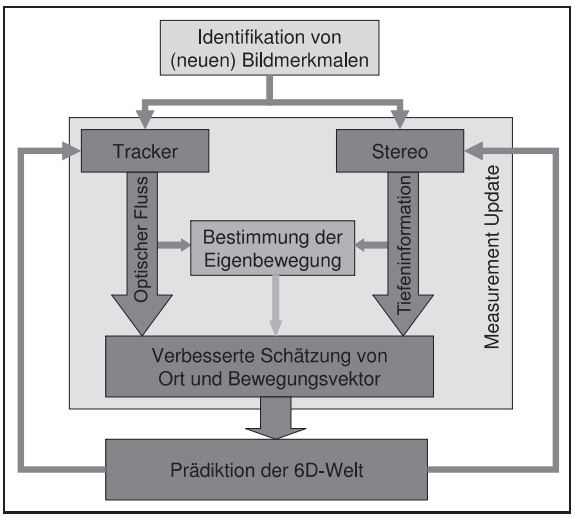

Bild 4 Das 6D-Vision Konzept.

drei Geschwindigkeitskomponenten $(\dot{X}, \dot{Y}, \dot{Z})^T$ des betrachteten Punktes, die vom Verfahren simultan geschätzt werden und ihm den Namen gegeben haben.

Das Filtern besteht aus den zwei Schritten: „Prädiktion“ und „Update“. Im Prädiktionsschritt wird der Zustandsvektor anhand eines zeitdiskreten Systemmodells vom aktuellen in den nächsten Zeitschritt überführt. Aus diesem prädizierten Zustand werden zu erwartende Messungen bestimmt und anschließend im Update-Schritt mit den tatsächlichen Messungen verglichen. Eventuelle Abweichungen werden genutzt, um den Zustandsvektor optimal zu korrigieren. Der Einsatz des Kalman Filters in der Bildfolgenanalyse geht auf E. D. Dickmanns zurück [6]. Der von ihm vorgeschlagene „4D“-Ansatz hat sich als extrem leistungsfähiges Verfahren für das Objekt-Tracking erwiesen und wird heute in vielen praktischen Applikationen eingesetzt.

Das präsentierte 6D-Vision-Verfahren überträgt dieses Prinzip auf die Ebene der Bildpunkte. Bild 4 zeigt schematisch den Ablauf. In dem von den Kameras aufgenommenen Bildpaar werden zunächst Features (Bildmerkmale) identifiziert, die sich zur weiteren Verarbeitung besonders eignen. Diese werden vom Tracker über die Zeit hinweg verfolgt; in der aktuellen Implementierung wird der Kanade–Lucas–Tomasi-Tracker verwendet [13; 15]. Die parallel im Stereoblock ermittelte 3D-Position wird genutzt, um die prädizierte Position und Geschwindigkeit des jeweils betrachteten Features zu verbessern (Update-Schritt des Kalman-Filters). Da die Bestimmung der Bewegungskomponenten relativ zum Fahrzeug durchgeführt wird, ist zudem die Kenntnis der Eigenbewegung notwendig. Diese kann entweder durch die im Auto vorhandene Inertialsensorik bestimmt oder ebenfalls aus den Bilddaten errechnet werden. Im Versuchsfahrzeug UTA kommt erfolgreich das Verfahren von Badino [2] zum Einsatz.

Im anschließenden Prädiktionsschritt liefert der Kalman-Filter den erwarteten Ort, der nur noch vom Messsystem verifiziert werden muss. Da der Suchbereich dafür drastisch begrenzt werden kann, verringert sich der Aufwand für die Berechnung von optischem Fluss und Stereo signifikant. Darüber hinaus können grobe Messfehler erkannt und eliminiert werden.

3.1 Mathematische Beschreibung

Sei $\vec{p} = (X, Y, Z)^T$ die 3D-Position eines beobachteten Punktes und $\vec{v} = (\dot{X}, \dot{Y}, \dot{Z})^T$ sein zugehöriger Geschwindigkeitsvektor. Nach einem Zeitintervall Δt lautet die Position zum Zeitschritt $k + 1$:

$$\vec{p}_{k+1} = R\vec{p}_k + \vec{T} + \Delta t R\vec{v}_k \quad (2)$$

wobei R die Rotation und $\vec{T}$ die Translation der Szene, d. h. die inverse Kamerabewegung darstellen. Der Geschwindigkeitsvektor $\vec{v}$ lautet entsprechend:

$$\vec{v}_{k+1} = R\vec{v}_k \quad (3)$$

Kombiniert man Position und Geschwindigkeitsvektor im 6D-Zustandsvektor $\vec{x} = (X, Y, Z, \dot{X}, \dot{Y}, \dot{Z})^T$, so ergibt sich das zeitdiskrete lineare Systemmodell des Kalman-Filters:

$$\vec{x}_k = A_k\vec{x}_{k-1} + B_k + \vec{\omega} \quad (4)$$

mit der Zustandstransitionsmatrix

$$A_k = \begin{bmatrix} R_k & \Delta t_k R_k \\ 0 & R_k \end{bmatrix}, \quad (5)$$

der Kontrollmatrix

$$B_k = \begin{bmatrix} \vec{T}_k \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (6)$$

und dem mittelwertfreien Gauß’schen Rauschterm $\vec{\omega}$.

Der Messvektor $\vec{z} = (u, v, d)^T$ setzt sich aus der vom Tracker bestimmten aktuellen Bildposition $(u, v)^T$ und der vom Stereosystem gemessenen Disparität d zusammen. Schaut die Kamera entlang der positiven Z-Achse, so lautet das nichtlineare Messmodell:

$$z = \begin{bmatrix} u \\ v \\ d \end{bmatrix} = \frac{1}{Z}\begin{bmatrix} Xf \\ Yf \\ bf \end{bmatrix} + \vec{v}. \quad (7)$$

Hierbei beschreiben f die fokalen Länge in Bildpunkten und b die Basisbreite des Stereosystems. Der Rauschterm $\vec{v}$ wird wiederum als mittelwertfreies Gauß’sches Rauschen angenommen.

4 Resultate aus der Praxis

Bild 5 zeigt das Ergebnis der 6D-Vision Schätzung für den Fahrradfahrer aus Bild 3. Die eingezeichneten Pfeile zeigen vom jeweils untersuchten Bildpunkt auf seine erwartete

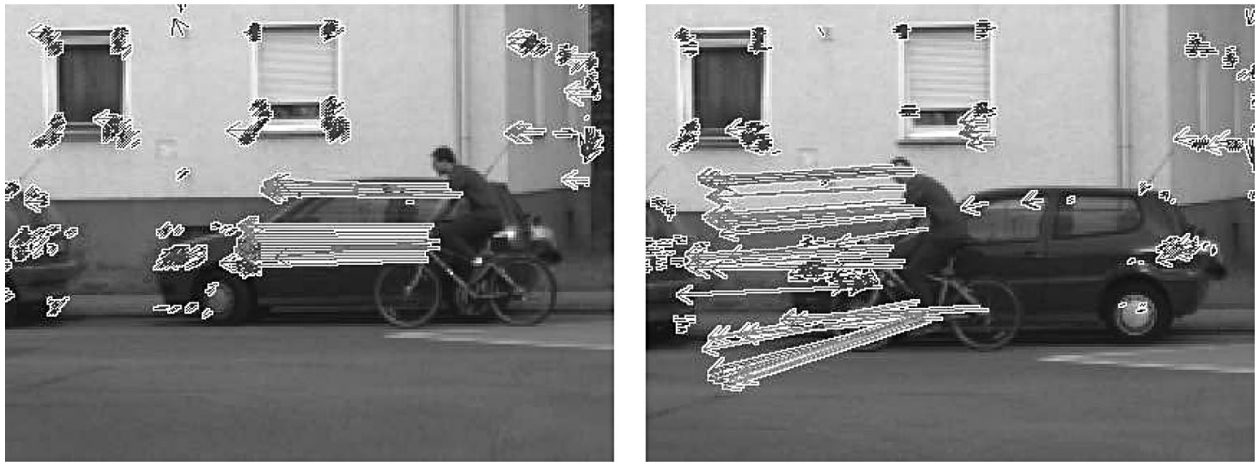

Bild 5 Ergebnis der Geschwindigkeitsschätzung für ein querendes Objekt ohne bildbasierte Eigenbewegungsschätzung. Die Pfeile zeigen die prädizierte Position des jeweils untersuchten Weltpunktes nach 0,5 s.

3D-Position nach 0,5 s. Im vorliegenden Beispiel fährt der Fahrradfahrer annähernd konstant mit 4 $\frac{m}{s}$. Das rechte Bild ist genau 0,5 s später aufgenommen. Wie man sieht, entspricht die Prädiktion sehr gut der tatsächlichen Bewegung. Im Gegensatz zur reinen Stereo-Analyse aus Bild 3 ist jetzt der Radfahrer sehr einfach vom Hintergrund zu trennen.

Für entgegenkommende Fahrzeuge liefert das System vergleichbar gute Ergebnisse. Bild 6 zeigt ein entgegenkommendes Fahrzeug wiederum im zeitlichen Abstand von 0,5 s. Die Eigengeschwindigkeit betrug 50 $\frac{km}{h}$, für das entgegenkommende Fahrzeug wurde ein Geschwindigkeit von 40 $\frac{km}{h}$ ermittelt. Die gestrichelte Linie gibt die erwartete bzw. tatsächliche Position der Unterkante des Fahrzeugs nach 0,5 s an.

In beiden Beispielen wurde die Eigenbewegung ausschließlich anhand der Inertialsensorik, bestehend aus Geschwindigkeits- und Gierratensensor, bestimmt. Nicht modellierte Nick- und Wankbewegungen des Fahrzeuges z. B. infolge schlechter Straßen werden deshalb als Bewegungen der Welt interpretiert. In Bild 7 (Ausschnittsvergößerung) ist deutlich der Effekt einer unkompensierten Nick-Bewegung zu sehen, die hier zu einer scheinbaren Abwärtsbewegung der stationären Punkte führt. Nach Aktivierung der bildbasierten Eigenbewegungsschätzung ist dieser Effekt, der genauso bei Wankbewegungen auftritt, verschwunden. Dank einer verbesserten Prädiktion kann zusätzlich eine größere Zahl von Punkten stabil verfolgt werden.

Die zeitliche Integration erlaubt nicht nur die Schätzung der Bewegung, sondern verbessert gleichzeitig die Genauigkeit der Ortsschätzung aller Punkte, da inhärent eine Mittelung vieler Messungen durchgeführt wird, die bei bewegtem Fahrzeug als annähernd unkorreliert aufgefasst werden können (vgl. [10]).

Das 6D-Vision Prinzip wurde 2005 auf der Abschlusspräsentation des BMBF-Projektes INVENT der Öffentlichkeit vorgestellt. Mit

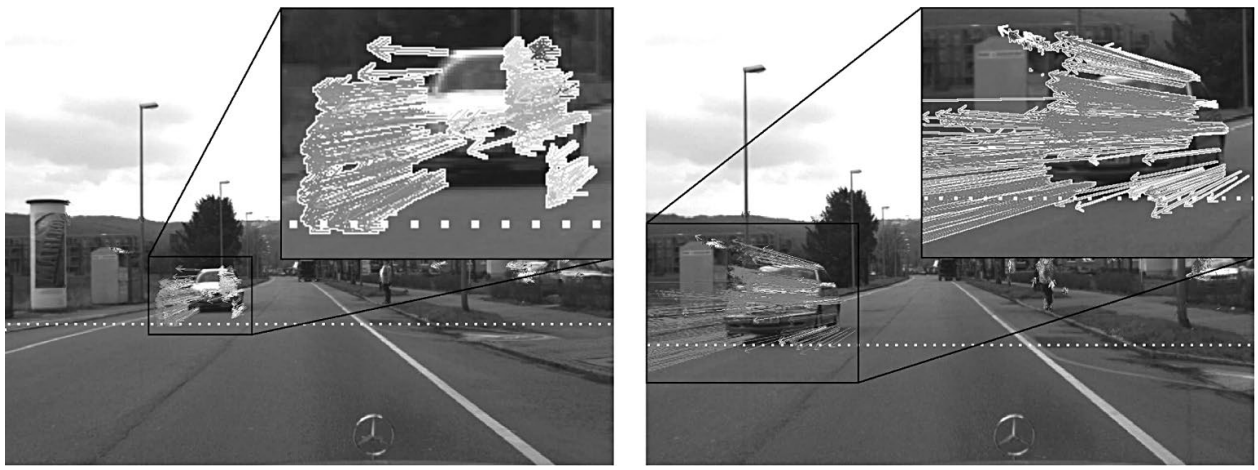

Bild 6 Ergebnis der Geschwindigkeitsschätzung für ein entgegenkommendes Fahrzeug. Die Eigengeschwindigkeit betrug 50 $\frac{km}{h}$, die des entgegenkommenden Fahrzeugs 40 $\frac{km}{h}$.

Bild 7 Ergebnis der Geschwindigkeitsschätzung ohne (links) und mit (rechts) bildbasierter Eigenbewegungsschätzung.

Schrittgeschwindigkeit querende Fußgänger wurden dabei zuverlässig aus Entfernungen von 30 m erkannt. Die Kollisionsgefahr mit einem sich schnell seitlich nähernden Fahrradfahrer wurde trotz stehender Peilung zwei bis drei Sekunden vor dem Zusammenstoß analysiert und bei ausbleibender Fahrerreaktion eine Notbremsung ausgelöst.

5 Zusammenfassung

Der Beitrag zeigt, wie durch zeitliche Integration von Stereo- und Bewegungsanalyse die für das Verstehen von Verkehrszenen zentrale Aufgabe der Detektion bewegter Objekte robust und echtzeitfähig gelöst werden kann. Die Bewegung plötzlich auftretender Objekte lässt sich typischerweise innerhalb von nur 4 Frames ermitteln. Bei unserer aktuellen Zykluszeit von 80 ms entspricht das einer Verzögerung von lediglich einer drittel Sekunde. Dabei schätzt der 6D-Vision-Ansatz simultan Ort und Bewegungsvektor einzelner Bildpunkte und ist so Ansätzen überlegen, die zunächst Objekte im Raum segmentieren und anschließend verfolgen. Der Schritt von einem 3D-Ortszustandsvektor zu dem 6D-Orts- und Bewegungsvektor erleichtert die Objektsegmentierung und die anschließende Objektverfolgung signifikant. Aktuell untersucht wird die Segmentierung mittels leistungsfähiger Verfahren wie „Level Set“ und „Graph Cut“. Dabei gefundene Segmente (bewegte Objekte) werden anschließend im Sinne des klassischen „4D“-Ansatzes verfolgt. Dadurch wird es möglich, zusätzlich zur Objektgeschwindigkeit auch die Beschleunigung zu schätzen. Eine deutliche Verzögerung signalisiert das im Kreuzungsbereich entscheidende „ich habe Sie gesehen“.

Dank der integrierten Schätzung der Eigenbewegung haben auch starke Nick- und Wankbewegungen keine negativen Auswirkungen auf die erzielten Resultate. Die durch den Kalman-Filter realisierte zeitliche Integration führt auch bei stationären Objekten zu einer signifikanten Steigerung der Schätzgenauigkeit.

Der beschriebene Ansatz ist nicht auf einen speziellen Feature-Tracker und das verwendete Stereoverfahren beschränkt. Dies ist bedeutsam, da aktuell vielerorts an preisgünstigen Hardwarelösungen (FPGA, ASICs) sowohl für die Stereoanalyse als auch für die Bewegungsanalyse gearbeitet wird, was der raum-zeitlichen Bildanalyse den Weg in die Praxis ebnen wird.

Schwächen der aktuellen Implementierung liegen in nicht erkannten fehlerhaften Stereokorrespondenzen, die stationären Punkten fälschlicherweise eine Bewegung zuordnen, sowie in dem Problem, dass in kontrastarmen Bildbereichen noch keine ausreichende Zahl von Messungen verfügbar ist bzw. die Ketten der verfolgten Punkte unzuverlässig sind, was wiederum bewegte Punkte vortäuscht. Solche Fehler müssen von den nachfolgenden Verarbeitungsstufen, insbesondere dem Objekttracking, beherrscht werden.

Die vorgestellte Fusion von Orts- und Bewegungsanalyse wird im BMWi-Projekt AKTIV weiterentwickelt. Unterstützung bei der Analyse von Kreuzungssituationen ist aus den USA zu erwarten: die DARPA hat das Projekt „Urban Challenge“ ausgeschrieben, bei dem im November 2007 vollautomatische Fahrzeuge komplexe Innenstadtsituationen beherrschen müssen [4]. Auch wenn dabei Sensoren zum Einsatz kommen werden, die absolut nicht fahrzeug- und serientauglich sind, wird dieses Projekt entscheidende Impulse für zukünftige Kreuzungsassistenzsysteme generieren.

Literatur

[1] A. Argyros, M. Lourakis, P. Trahanias, S. Orphanoudakis: Qualitative Detection of 3D Motion Discontinuities. In: *Proc. of the IEEE Int'l Conf. on Intelligent Robots and Systems* Vol. 3, Nov. 1996, pp. 1630–1637.

[2] H. Badino, U. Franke, C. Rabe, S. Gehrig: Stereo Vision based Detection of moving Objects under strong Camera Motion. In: *Proce. of the First Int'l Conf. on Computer Vision Theory and Applications* Vol. 2, Feb. 2006, pp. 253–260.

[3] Bundesministerium für Verkehr, Bau- und Wohnungswesen: Verkehr in Zahlen 2000. Deutscher Verkehrs-Verlag, Hamburg, 2000.

[4] http://www.darpa.mil

[5] T. Dang, C. Hoffmann, C. Stiller: Fusing Optical Flow and Stereo Disparity for Object Tracking. In *Proc. of the 5th IEEE Int'l Conf. on Intelligent Transportation Systems*, 2002, pp. 112–117.

[6] E. D. Dickmanns: 4D-Szenenanalyse mit integralen raum-/zeitlichen Modellen. In: *Proc. of the 9th DAGM Symp.*, Vol. 149, 1987, pp. 257–271 – ISBN 3-540-18375-2.

[7] W. Enkelmann: Obstacle Detection by Evaluation of Optical Flow Fields from Image Sequences. In: *Proc. of the First European Conf. on Computer Vision*, Vol. 427, 1990, pp. 134–138.

[8] U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, C. Wöhler: Autonomous Driving Goes Downtown. In: *IEEE Intelligent Systems*, 1998, Vol. 13, No. 6, pp. 40–48.

[9] U. Franke, A. Joos: Real-time Stereo Vision for Urban Traffic Scene Understanding. In: *Proc. of the IEEE Intelligent Vehicles Symposium*, 2000, pp. 273–278.

[10] U. Franke, C. Rabe, H. Badino, S. Gehrig: 6D-Vision: Fusion of Stereo and Motion for Robust Environment Perception. In: *Proc. of the 27th DAGM Symp.*, 2005, pp. 216–223.

[11] J. Klappstein, F. Stein, U. Franke: Flussbasierte Eigenbewegungsschätzung und Detektion von fremdbewegten Objekten. In: *Proc. of the 4th Workshop Fahrerassistenzsysteme*, 2006, pp. 78–88 – ISBN 3-9809121-2-4.

[12] F. Lindner, U. Kressel, S. Kaelberer: Robust Recognition of Traffic Signals. In: *Proce. of the IEEE Intelligent Vehicles Symposium*, 2004, pp. 49–53.

[13] B. Lucas, T. Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision. In: *Proc. of the 7th Int'l Joint Conf. on Artificial Intelligence*, 1981, pp. 674–679.

[14] F. Stein: Efficient Computation of Optical Flow Using the Census Transform. In: *Proc. of the 26th DAGM Symp.*, 2004, pp. 79–86.

[15] C. Tomasi, T. Kanade: Detection and Tracking of Point Features. In: *Technical report CMU-CS-91-132*, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1991.

[16] A. M. Waxman, J. H. Duncan: Binocular image flows: Steps toward stereo-motion fusion. In: *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, 1986, pp. 715–729.

[17] G. Welch, G. Bishop: An Introduction to the Kalman Filter. In: *Technical report TR 95-041*, University of North Carolina at Chapel Hill, Department of Computer Science, 2004. TS[a]

1 Dr.-Ing. Uwe Franke wurde 1958 in Witzenhausen geboren. Er hat an der RWTH Aachen Elektrotechnik studiert und 1988 über regionenorientierte Bildcodierungsverfahren mit Auszeichnung promoviert. Seit 1989 ist er bei DaimlerChrysler im Bereich Forschung Fahrerassistenzsysteme tätig. Seit 2000 leitet er dort die Arbeitsgruppe „Videobasierte Umgebungserfassung". Im Jahre 2002 war er Program Chair der IEEE Intelligent Vehicles Conference. Sein spezielles Interesse gilt dem Verstehen komplexer Verkehrsszenen, insbesondere in der Innenstadt.
Adresse: DaimlerChrysler AG, HPC W50/G 024, 71059 Sindelfingen, Tel.: +49-7031-4389-873, Fax: +49-7031-4389-264, E-Mail: uwe.franke@daimlerchrysler.com

2 Dipl.-Inf. (FH) Clemens Rabe wurde 1979 in Gießen geboren. Sein Studium der Technischen Informatik schloss er 2005 ab (Dipl.-Inf. (FH)). Seit 2005 promoviert er an der Universität Kiel in Zusammenarbeit mit DaimlerChrysler.
Adresse: DaimlerChrysler AG, HPC W50/G 024, 71059 Sindelfingen, Tel.: +49-7031-4389-881, Fax: +49-7031-4389-264, E-Mail: clemens.rabe@daimlerchrysler.com

3 Dr. rer. nat. Stefan Gehrig wurde 1968 in Sindelfingen geboren. Sein Studium der Technischen Informatik an der Berufsakademie schloss er 1991 ab (Dipl. Ing. (BA)). Nachfolgend studierte er Physik an den Universitäten Stuttgart, Tübingen, San Jose und Berkeley, welches er 1997 mit Auszeichung abschloss (Dipl.-Phys.). Er promovierte an der Universität Tübingen in Zusammenarbeit mit DaimlerChrysler (Dr. rer. nat., 2000), wo er heute noch in der Forschung arbeitet. Sein Schwerpunkt ist Bildverarbeitung im automotiven Umfeld.
Adresse: DaimlerChrysler AG, HPC W50/G 024, 71059 Sindelfingen, Tel.: +49-7031-4389-874, Fax: +49-7031-4389-264, E-Mail: stefan.gehrig@daimlerchrysler.com