Maschinelles Sehen für mobile Roboter: Virtuell-aktive visuelle Odometrie

Machine Vision for Mobile Robots: Virtually-Active Visual Odometry

Volker Willert*, Technische Universität Darmstadt

* Korrespondenzautor: vwillert@rtr.tu-darmstadt.de

Zusammenfassung Der Beitrag befasst sich mit der Navigation von autonomen mobilen Robotern anhand von Videodaten. Es wird ein Ansatz zur Verbesserung visueller Odometriealgorithmen vorgestellt, der die Idee der Einflussnahme auf die Bilddaten durch ein aktives Kamerasystem aufgreift, ohne tatsächlich eine Aktorik zu benutzen. Es wird gezeigt, dass durch virtuelle Fixationen das räumliche Muster des Bildflusses so verändert werden kann, dass die daraus berechneten Kamerabewegungen genauer werden. Zur Berechnung der Kamerabewegung wird der aktive 7-Punkt-Algorithmus eingeführt. Anhand von statistischen Auswertungen wird abgeleitet, welche Fixationspunkte dazu besonders geeignet sind und ein Verfahren beschrieben, wie geeignete Fixationspunkte ausgewählt werden können. Desweiteren wird ein Vergleich zwischen dem virtuell-aktiven Ansatz und einem genauen visuellen Odometrieverfahren auf aktuellem Stand der Technik gegeben. ►►► **Summary** This paper deals with vision-based navigation of autonomous mobile robots. An approach to improve visual odometry algorithms is presented that picks up the idea of actively influencing the image data by actuating elements of an active camera system without using an active camera system effectively. It is shown that via virtual fixations the spatial optical flow pattern can be influenced in such a way that the camera motion computed based on this optical flow is getting more precise. For computation of the camera motion, the active 7-point algorithm is introduced. Based on statistical evaluations it is derived which fixation points are especially suited and a strategy how to choose such suited fixation points is described. A comparison between the presented virtual-active approach and a state of the art visual odometry implementation is given.

Schlagwörter Bildverarbeitung, Robotersehen, mobile Robotik, visuelle Odometrie ►►► **Keywords** Computer vision, robot vision, mobile robotics, visual odometry

1 Einleitung

Laut der Zeitschrift *Pictures of the Future* [1] ist für die künstliche Intelligenz das maschinelle Sehen nach über fünfzig Jahren Forschung weiterhin die größte Herausforderung. Als Beispiel wird das Forschungsprojekt *Fly & Inspect* genannt [2], welches eine mobile Plattform entwickelt, die in Zukunft mithilfe von Videokameras völlig autonom navigieren soll. Dazu müssen nur auf Basis von visuellen Daten Hindernisse erkannt, ihnen ausgewichen und bewegte Objekte detektiert, verfolgt und ihnen bei Bedarf auch gefolgt werden. Zur Lösung dieser Aufgabe spielt die Auswertung der visuellen Daten eine zentrale Rolle, denn diese Daten beinhalten eine Fülle von relevanter Information über die Umwelt, anhand derer die momentane Bewegung bzw. ganze Bewegungsabläufe des mobilen Roboters geplant werden können. Je präziser

und fehlerfreier die Umweltinformationen auf Basis der visuellen Daten, wie z. B. das Szenenprofil sowie die Art und Position von Hindernissen, extrahiert werden kann, desto genauer kann der Roboter in seiner Umgebung navigieren.

Grundlage für alle visuell getriebenen Navigationsaufgaben ist die Posenbestimmung der sich bewegenden Kamera. Die Pose einer Kamera besteht aus der Position und der Orientierung des bewegten Kamerakoordinatensystems bezüglich eines festen Weltkoordinatensystems. Die Kamerapose kann rein visuell über die sogenannte *visuelle Odometrie* [3] bestimmt werden. In den letzten dreißig Jahren wurde eine Menge Verfahren für unterschiedlichste Anwendungen entwickelt. Einen umfassenden Überblick liefern die aktuellen Artikel von Scaramuzza und Fraundorfer [4; 5].

Die Bildauswertung besteht bei der visuellen Odometrie im Wesentlichen aus zwei Schritten: 1) der Extraktion der 2D *Bildkoordinaten* von projizierten *raumfesten* 3D Punkten, die eine eindeutige Intensitätsstruktur im Bereich um die Bildkoordinate aufweisen (sogenannte visuelle Merkmale), damit 2) der *Bildfluss* dieser projizierten 3D Punkte möglichst eindeutig und genau bestimmt werden kann. Je präziser Bildkoordinaten und Bildfluss dieser 3D Punkte vorliegen, desto genauer kann die Pose und die Posenänderung der Kamera berechnet werden.

Bei fast allen Ansätzen (siehe z. B. [4; 5]) wird davon ausgegangen, dass sich die Kameras *passiv* bewegen. Das bedeutet, dass anhand der Bildauswertung kein Einfluss auf die Bewegung der Kamera genommen werden kann. Damit ist die Bildauswertung allen Schwierigkeiten ausgesetzt, die bei einer Bildaufnahme von einer bewegten Plattform aus entstehen. Angefangen von Bewegungsartefakten, wie Bewegungsunschärfe, die eine Verschlechterung der Bildqualität zur Folge haben, oder der Kameradekalibrierung aufgrund von Vibrationen und Erschütterungen, ergeben sich eine Reihe weiterer, grundlegender Probleme. Eine bewegte Kamera sieht in ihrem Sichtfeld nur für einen kurzen Augenblick die gleiche Szeneninformation. Selbst die Objekte, die für eine Weile im Sichtfeld zu sehen sind, ändern ihren Abstand zur Kamera und damit die Projektionsgröße und die abgebildete Auflösung. Zudem wechselt die Objektansicht, wodurch unterschiedliche Bereiche der Objekte im Raum abgebildet bzw. verdeckt werden. Schließlich ändern sich durch die Bewegung auch die Beleuchtungsverhältnisse, was zu Helligkeits- und Kontrastschwankungen führt. Daraus könnte man schlussfolgern: Bewegung verschlechtert die visuelle Perzeption und damit die visuell gestützte Roboternavigation.

Diesen Schwierigkeiten wirkt man bei der Bildauswertung entgegen, indem man visuelle Merkmale sucht, die möglichst robust gegenüber Bildrauschen und Bewegungsartefakten sind und möglichst invariant gegenüber Beleuchtungs- und Ansichtsänderungen erscheinen. Eine Übersicht über geeignete visuelle Merkmale für die visuelle Odometrie findet sich beispielsweise in [5].

Mit der Fähigkeit des Roboters sich selbst *aktiv* – und damit auch das Bildverarbeitungssystem – zu bewegen, ist die Möglichkeit für den Roboter eröffnet, zu entscheiden, welchen Ausschnitt aus der Umwelt er sich aus welcher Perspektive anschaut. Die Bildverarbeitung übernimmt damit nicht nur das Wahrnehmen, um zu agieren, sondern aus dem Wahrgenommenen kann auch abgeleitet werden, wie agiert werden soll, um besser wahrzunehmen [6]. Diese Idee wurde Anfang der neunziger Jahre formuliert und wird heute unter dem Begriff *Aktives Sehen* (active vision) zusammengefasst [7–9]. Dazu gehört auch die aktive Beeinflussung von Kameraparametern, wie dem Fokus, der Tiefenschärfe oder der Belichtungszeit.

Hauptaugenmerk des aktiven Sehens im Bereich der visuellen Navigation liegt auf der Fixation von Raumpunkten oder Punkten und Linien während der Bewegung, um die Bewegungsgleichungen zu reduzieren [10; 11]. Diese Idee wurde erstmals von Bandyopadhyay [12] aufgeworfen und von ihm gezeigt, dass sich durch Fixation eines Raumpunktes die freien Parameter der Kamerabewegung um eins reduzieren. Im Bereich des *visual SLAM* (Self Localization and Mapping) wurde vorgeschlagen die Positionierung zu schlecht identifizierten visuellen Landmarken aktiv zu verändern, um die Identifikationssicherheit zu steigern [13]. Dazu müssen Messung und Zustand als stochastische Variablen angenommen werden. Als Informationsmaße werden unter anderem der mittlere Transinformationsgehalt zwischen der Vorhersage der nächsten Merkmalskorrespondenz und des erwarteten Zustandes der Kamerapose herangezogen. Damit wird die aktive Kamerabewegung so gewählt, dass die Unsicherheit des Zustandes der Kamerapose unter der Bedingung aller bisher getätigten Messungen größtmöglich reduziert wird [14; 15].

Die Frage, ob durch die Fixation eines Raumpunktes auch die Genauigkeit der Schätzung der Kamerabewegung verbessert werden kann, wenn man davon ausgeht, dass sowohl die Bildkoordinaten, als auch der Bildfluss nicht exakt bestimmbar ist, wurde – nach bestem Wissen des Authors – bis jetzt noch nicht beantwortet. In diesem Artikel wird anhand von statistischen Auswertungen gezeigt, dass sich die Schätzgüte verbessern läßt, wenn der Fixationspunkt nach bestimmten Kriterien in Abhängigkeit von der momentanen Kamerabewegung ausgewählt wird. Daraus kann man schlussfolgern: Eine aktive Einflussnahme auf die Bewegung kann die visuelle Perzeption und damit die visuell gestützte Roboternavigation verbessern.

Im verbleibenden Artikel werden zuerst in den Abschnitten 2.1 bis 2.3 die Grundlagen der visuellen Odometrie zusammengefasst, die zum Entwurf der virtuell-aktiven visuellen Odometrie nötig sind. Danach werden in den Abschnitten 2.4 bis 2.7 die Teilschritte des virtuell-aktiven visuellen Odometrieverfahrens vorgestellt. Abschnitt 3 zeigt einen Vergleich dieses Verfahrens

mit einem sehr genauen visuellen Odometrieverfahren auf aktuellem Stand der Technik [16] anhand von realen Videosequenzen. Abschnitt 4 fasst die Ergebnisse zusammen.

2 Virtuell-aktive visuelle Odometrie

Im Folgenden wird die visuelle Odometrie als Beispiel für eine Verbesserung der visuellen Perzeption durch von der Perzeption getriebene Kamerabewegungen herangezogen. Dazu wird die Idee der sogenannten *aktiven Navigation* nach [12] zur Vereinfachung der Berechnung der Eigenbewegung einer mobilen Kamera aufgegriffen und die Auswirkungen auf den kontinuierlichen 8-Punkt-Algorithmus [17] hergeleitet. Danach wird eine Strategie zur Auswahl von virtuell-aktiven Kamerabewegungen vorgestellt, welche die Schätzung der Eigenbewegung verbessert.

2.1 Visuelle Odometrie

Als visuelle Odometrie bezeichnet man in der mobilen Robotik die Posenbestimmung einer bewegten Kamera anhand der Kameradaten. Sie ist Grundlage jeder visuellen Navigations- sowie Szenenrekonstruktionsaufgabe. Der große Vorteil der visuellen Odometrie gegenüber der Odometrie auf Basis von Radumdrehungen eines mobilen Roboters ist das Wegfallen von Odometriefehlern aufgrund von Radschlupf oder Unebenheiten [4]. Die Pose $g^t = (\mathbf{R}^t, \mathbf{T}^t) \in SE(3)$ zum Zeitpunkt t umfasst die Position $\mathbf{T}^t \in \mathbb{R}^3$ und die Orientierung $\mathbf{R}^t \in SO(3)$ des Kamerakoordinatensystems $\mathcal{C}$ einer Kamera im Raum bezüglich eines Weltkoordinatensystems $\mathcal{W}$. Hierbei entspricht $SE(3)$ der Menge aller Euklidischen Transformationen und $SO(3)$ der Menge aller Rotationsmatrizen im Raum. Die zeitlich veränderlichen Kamerakoordinaten $\mathbf{X}_\mathcal{C}^t = [X_\mathcal{C}^t, Y_\mathcal{C}^t, Z_\mathcal{C}^t]^\top$ jedes zeitlich festen Raumpunktes $p \in \mathbb{R}^3$ stehen über seine Weltkoordinaten $\mathbf{X}_\mathcal{W} = [X_\mathcal{W}, Y_\mathcal{W}, Z_\mathcal{W}]^\top$ und die zeitlich veränderliche Kamerapose g^t der bewegten Kamera in folgendem Zusammenhang:

$$\mathbf{X}_\mathcal{C}^t = \mathbf{R}^t \mathbf{X}_\mathcal{W} + \mathbf{T}^t \,. \tag{1}$$

Die aktuelle Pose wird über Integration von Relativposen $g^{t,t-\Delta t} = (\mathbf{R}^{t,t-\Delta t}, \mathbf{T}^{t,t-\Delta t}) \in SE(3)$ zwischen zwei aufeinanderfolgenden Kameraposen zu den Zeitpunkten t und $t-\Delta t$ berechnet (siehe Bild 1). Ausgehend von einer bekannten Startpose $g^0 = (\mathbf{R}^0, \mathbf{T}^0)$ kann die aktuelle Pose rekursiv berechnet werden:

$$\mathbf{R}^t = \left(\mathbf{R}^{t,t-\Delta t} + \eta_R^t\right)\mathbf{R}^{t-\Delta t} \,, \tag{2a}$$

$$\mathbf{T}^t = \left(\mathbf{R}^{t,t-\Delta t} + \eta_R^t\right)\mathbf{T}^{t-\Delta t} + \left(\mathbf{T}^{t,t-\Delta t} + \eta_T^t\right) . \tag{2b}$$

Damit werden auch die Messfehler η_R^t und η_T^t aufintegriert, was ein zunehmendes Abweichen der geschätzten von der realen Robotertrajektorie zur Folge hat. Deswegen ist es wichtig, die einzelnen Relativposen $g^{t,t-\Delta t}$ so genau wie möglich zu berechnen, um die Fehlerfortpflanzung so gering wie möglich zu halten [5]. Die

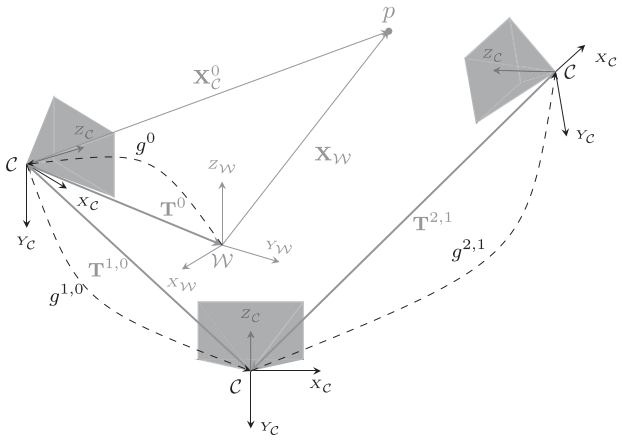

Bild 1 Prinzip der visuellen Odometrie.

Relativposen können über die Epipolareinschränkung direkt aus den Bildkoordinaten und den Bildpunktbewegungen von projizierten festen Raumpunkten berechnet werden.

2.2 Zeitkontinuierliche Epipolareinschränkung

Die zeitkontinuierliche Epipolareinschränkung nach [18] beschreibt den differenziellen Zusammenhang zwischen Bildpunktbewegungen $\mathbf{u}^t$ (auch *optischer Fluss* genannt), Bildpunktkoordinaten $\mathbf{x}^t$ von projizierten statischen Raumpunkten p und den Kamerageschwindigkeiten $V^t = (\boldsymbol{\omega}^t, \mathbf{v}^t)$.[1]

Die unterschiedlichen Kamerakoordinaten $\mathbf{X}_\mathcal{C}^t$ und $\mathbf{X}_\mathcal{C}^{t-\Delta t}$ eines festen Raumpunktes p zu zwei unterschiedlichen Zeitpunkten $(t, t-\Delta t)$ stehen über die Relativpose $g^{t,t-\Delta t}$ in folgendem Zusammenhang:

$$\mathbf{X}_\mathcal{C}^t = \mathbf{R}^{t,t-\Delta t}\mathbf{X}_\mathcal{C}^{t-\Delta t} + \mathbf{T}^{t,t-\Delta t} \,. \tag{3}$$

Ist die Kamerabewegung langsam im Vergleich zur Bildwiederholrate und damit der Zeitversatz Δt sehr klein (was bei mobilen Robotern meistens zutrifft), dann ist die zeitliche Änderung $\dot{\mathbf{X}}_\mathcal{C}^t$ der Koordinate über die Kamerageschwindigkeiten $V^t = (\boldsymbol{\omega}^t, \mathbf{v}^t)$ beschrieben:

$$\dot{\mathbf{X}}_\mathcal{C}^t = \widehat{\boldsymbol{\omega}}^t \mathbf{X}_\mathcal{C}^t + \mathbf{v}^t \,, \tag{4}$$

wobei $\widehat{\boldsymbol{\omega}}^t \in \mathbb{R}^{3\times 3}$ der schiefsymmetrischen Matrix des Vektors der Rotationsgeschwindigkeit $\boldsymbol{\omega}^t = [\omega_1^t, \omega_2^t, \omega_3^t]^\top$ und $\mathbf{v}^t = [v_1^t, v_2^t, v_3^t]^\top$ dem Vektor der Translationsgeschwindigkeit entspricht.[2] Die Projektion des Punktes p auf die Bildebene entsteht an der homogenen Bildkoordinate $\mathbf{x} = [x, y, 1]^\top$ im normierten Bildkoordinatensystem.[3] Damit stehen Raum- $\mathbf{X}_\mathcal{C}$ und Bildpunktkoordinaten $\mathbf{x}$, sowie deren Ableitungen $\dot{\mathbf{X}}_\mathcal{C}$

[1] Die Kamerageschwindigkeiten beschreiben die Änderung der Pose des Weltkoordinatensystems relativ zum Kamerakoordinatensystem aus Sicht der Kamera, also bezüglich der Kamerakoordinaten.
[2] Zur besseren Übersicht werden die Zeitindizes ab jetzt vernachlässigt.
[3] Es wird davon ausgegangen, dass die intrinsischen Parameter der Kamera bekannt sind.

und $\dot{\mathbf{x}} = [\dot{x}, \dot{y}, 0]^\top$ über die projektive Abbildung $\mathbf{X}_\mathcal{C} = Z_\mathcal{C}\mathbf{x}$ in folgendem Zusammenhang:

$$\dot{\mathbf{X}}_\mathcal{C} = \dot{Z}_\mathcal{C}\mathbf{x} + Z_\mathcal{C}\dot{\mathbf{x}}\,. \tag{5}$$

Setzt man (5) in (4) ein

$$\dot{\mathbf{x}} = \widehat{\boldsymbol{\omega}}\mathbf{x} + \frac{1}{Z_\mathcal{C}}\mathbf{v} - \frac{\dot{Z}_\mathcal{C}}{Z_\mathcal{C}}\mathbf{x} \tag{6}$$

und wendet das Skalarprodukt auf beide Vektoren der Gleichung (6) mit dem Vektor $\widehat{\mathbf{v}}\mathbf{x}$ an, dann erhält man die $Z_\mathcal{C}$-bereinigte, sogenannte *zeitkontinuierliche Epipolareinschränkung*

$$\mathbf{u}^\top\widehat{\mathbf{v}}\mathbf{x} + \mathbf{x}^\top\widehat{\boldsymbol{\omega}}\widehat{\mathbf{v}}\mathbf{x} = 0\,, \tag{7}$$

wobei $\mathbf{u} = [u_1, u_2, 0]^\top := \dot{\mathbf{x}}$. Diese Gleichung hängt nur noch von den Bildkoordinaten $\mathbf{x}$ und deren Bewegung $\mathbf{u}$, sowie der gesuchten Kamerageschwindigkeiten $V = (\boldsymbol{\omega}, \mathbf{v})$ ab.

2.3 Zeitkontinuierlicher 8-Punkt-Algorithmus

Der zeitkontinuierliche 8-Punkt-Algorithmus ist in [17] ausführlich beschrieben. Er überführt die zeitkontinuierliche Epipolareinschränkung in ein homogenes, lineares, (überbestimmtes) Gleichungssystem zur Bestimmung der Kamerageschwindigkeit, welches zur Lösung die Koordinaten $\{\mathbf{x}_i\}_{i=1}^N$ und optische-Fluss-Vektoren $\{\mathbf{u}_i\}_{i=1}^N$ von mindestens acht $N \geq 8$ projizierten festen Raumpunkten benötigt. Hier werden kurz die Zusammenhänge dargelegt, welche zum Verständnis des in Abschnitt 2.5 vorgestellten aktiven 7-Punkt-Algorithmus benötigt werden.

In [18] wird gezeigt, dass die Beziehung $\mathbf{x}^\top\mathbf{S}\mathbf{x} = \mathbf{x}^\top\widehat{\boldsymbol{\omega}}\widehat{\mathbf{v}}\mathbf{x}$ gilt und man aus der kontinuierlichen Epipolareinschränkung (7) nur die symmetrische Epipolar-Komponente $\mathbf{S} = (\widehat{\boldsymbol{\omega}}\widehat{\mathbf{v}} + \widehat{\mathbf{v}}\widehat{\boldsymbol{\omega}})/2$ der Matrix $\widehat{\boldsymbol{\omega}}\widehat{\mathbf{v}}$ erhalten kann. Damit muss die Gleichung

$$\mathbf{u}^\top\widehat{\mathbf{v}}\mathbf{x} + \mathbf{x}^\top\mathbf{S}\mathbf{x} = 0 \tag{8}$$

nach den Komponenten der Matrizen

$$\widehat{\mathbf{v}} = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} \quad \text{und} \quad \mathbf{S} = \begin{pmatrix} s_1 & s_2 & s_3 \\ s_2 & s_4 & s_5 \\ s_3 & s_5 & s_6 \end{pmatrix}$$

aufgelöst werden. Die rechte Seite der Gleichung (8) kann durch Umsortieren der Elemente als Skalarprodukt $\mathbf{a}^\top\mathbf{s}$ der beiden Vektoren

$$\mathbf{a} := [-u_2, u_1, u_2x - u_1y, x^2, 2xy, 2x, y^2, 2y, 1]^\top \in \mathbb{R}^9, \tag{9a}$$

$$\mathbf{s} := [v_1, v_2, v_3, s_1, s_2, s_3, s_4, s_5, s_6]^\top \in \mathbb{R}^9 \tag{9b}$$

geschrieben werden. Damit ist die kontinuierliche Epipolareinschränkung (8) in eine homogene Gleichung $\mathbf{a}^\top\mathbf{s} = 0$, die linear in den gesuchten Variablen $\mathbf{s}$ ist, umgeformt. Jedes Paar $(\mathbf{x}_i, \mathbf{u}_i)$ bildet einen Vektor $\mathbf{a}_i$ und eine Gleichung $\mathbf{a}_i^\top\mathbf{s} = 0$. Die Matrix $\mathbf{A} := [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]^\top \in \mathbb{R}^{N\times 9}$ bildet zusammen mit dem Vektor $\mathbf{s}$ ein homogenes lineares Gleichungssystem

$$\mathbf{A}\mathbf{s} = 0\,. \tag{10}$$

Falls die Matrix $\mathbf{A}$ den Rang acht hat, dann gibt es eine homogene Lösung $\widehat{\mathbf{s}}$, die eindeutig bis auf einen beliebigen Skalierungsfaktor λ ist. Das bedeutet, es kann nur die translatorische Bewegungsrichtung $\lambda\mathbf{v}$, nicht aber der Betrag $\|\mathbf{v}\|$ ermittelt werden. Da die Größen $(\mathbf{x}_i, \mathbf{u}_i)$ nicht exakt bestimmbar sind, werden üblicherweise mehr als die minimal benötigten acht Gleichungen $N \geq 8$ aufgestellt und das überbestimmte homogene Gleichungssystem über die Methode der kleinsten Quadrate $\widehat{\mathbf{s}} = \operatorname{argmin}_\mathbf{s}\|\mathbf{A}\mathbf{s}\|^2$ mittels einer Singulärwertzerlegung [19] gelöst.

Die sich daraus ergebende Matrix $\widehat{\mathbf{S}}$ ist zwar symmetrisch, jedoch nicht zwangsweise eine symmetrische Epipolar-Komponente. Dies wird durch eine Projektion dieser Matrix auf den symmetrischen epipolaren Raum gewährleistet. Danach kann die Kamerabewegung V, wie in [17] beschrieben, extrahiert werden.

2.4 Aktiver Navigationsansatz

Der aktive Navigationsansatz wurde bereits in [12] vorgestellt. Dort wird gezeigt, dass das Fixieren des projizierten Punktes im Bildhauptpunkt durch eine, der passiven Kamerabewegung überlagerte, aktive Rotationsbewegung der Kamera die freien Parameter der Kamerabewegung von fünf auf vier reduziert, wenn die aktive Rotationsbewegung genau bekannt ist.

Differenziert man Gleichung (1) nach der Zeit, erhält man die Herleitung von Gleichung (4) und sieht den Zusammenhang zwischen Kamerageschwindigkeiten V und Posenänderung $\dot{g}$ ausgehend von einer beliebigen Kamerapose g:

$$\dot{\mathbf{X}}_\mathcal{C} = \underbrace{\dot{\mathbf{R}}\mathbf{R}^\top}_{\widehat{\boldsymbol{\omega}}}\mathbf{X}_\mathcal{C} + \underbrace{\dot{\mathbf{T}} - \dot{\mathbf{R}}\mathbf{R}^\top\mathbf{T}}_{\mathbf{v}} = \widehat{\boldsymbol{\omega}}\mathbf{X}_\mathcal{C} + \mathbf{v}\,. \tag{11}$$

Es zeigt sich, dass die zeitliche Änderung der Koordinaten $\mathbf{X}_\mathcal{C}^\mathcal{O} = [0, 0, 0]^\top$ des Kameraursprungs $\mathcal{O}$ der translatorischen Kamerabewegung $\mathbf{v} = \dot{\mathbf{X}}_\mathcal{C}^\mathcal{O}$ entspricht. Realisiert man nun eine der passiven Kamerabewegung $\boldsymbol{\omega}$ additiv überlagerte aktive Rotationsbewegung $\boldsymbol{\omega}_a = [\omega_{1,a}, \omega_{2,a}, 0]^\top$, so dass ein Raumpunkt $\mathcal{F}$ (Fixationspunkt) immer im Bildhauptpunkt $\mathbf{x}_o = [0, 0, 1]^\top$ abgebildet (fixiert) wird, die optische Achse also immer durch den Fixationspunkt verläuft, dann gilt: $\mathbf{X}_\mathcal{F} = \mathbf{T} = [0, 0, T_3]^\top$, $\dot{\mathbf{X}}_\mathcal{F} = \dot{\mathbf{T}} = [0, 0, \dot{T}_3]^\top$ und $\widehat{\boldsymbol{\omega}} + \widehat{\boldsymbol{\omega}}_a = \dot{\mathbf{R}}$. Hierbei besitzt der Fixationspunkt immer die Kamerakoordinaten $\mathbf{X}_\mathcal{F} = [0, 0, Z_\mathcal{F}]^\top$ und die entsprechende translatorische Bewegung $\dot{\mathbf{X}}_\mathcal{F} = [0, 0, \dot{Z}_\mathcal{F}]^\top$. Außerdem wird die Kamerapose auf ein Weltkoordinatensystem mit diesem Fixationspunkt als Ursprung bezogen, welches parallel $\mathbf{R}^\top = \mathbf{I}$ zum aktuellen Kamerakoordinatensystem ausgerichtet ist.

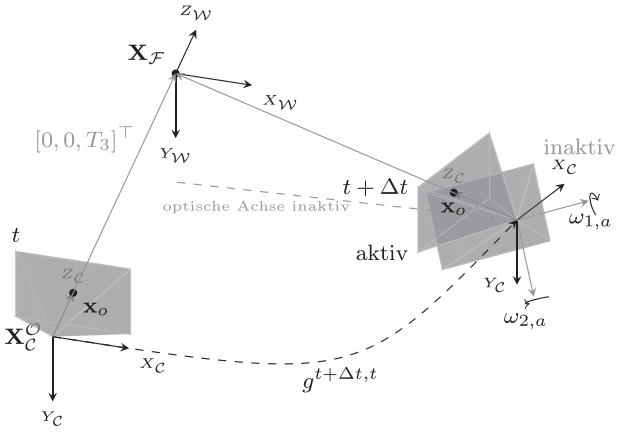

Bild 2 Aktiver Navigationsansatz.

Diese Zusammenhänge sind nochmals in Bild 2 skizziert. Setzt man diese Bedingungen in Gleichung (11) ein

$$\mathbf{v} = \dot{\mathbf{X}}_\mathcal{F} - (\widehat{\boldsymbol{\omega}} + \widehat{\boldsymbol{\omega}}_a)\mathbf{X}_\mathcal{F}\,, \tag{12}$$

dann erhält man eine Abhängigkeit zwischen Komponenten der translatorischen Kamerabewegung $\mathbf{v}$ und Komponenten der passiven rotatorischen Kamerabewegung $\boldsymbol{\omega}$ bei bekannter aktiver Rotationsbewegung $\boldsymbol{\omega}_a$:

$$v_1 = -(\omega_2 + \omega_{2,a})T_3\,, \tag{13a}$$
$$v_2 = (\omega_1 + \omega_{1,a})T_3\,, \tag{13b}$$
$$v_3 = \dot{T}_3\,. \tag{13c}$$

Wenn die Entfernung des Fixationspunktes T_3 unbekannt ist, dann kann wieder nur die translatorische Bewegungsrichtung[4] $\lambda\mathbf{v}$, nicht aber der Betrag $\|\mathbf{v}\|$ ermittelt werden.

2.5 Aktiver linearer 7-Punkt-Algorithmus

Wir gehen davon aus, dass die Entfernung T_3 nicht bekannt ist und definieren aufgrund des Zusammenhanges in den Gleichungen (13) als Schätzvektoren die mit λ normierte Translationsgeschwindigkeit

$$\begin{aligned}\mathbf{v}' &= [v_1', v_2', v_3']^\top = \lambda[v_1, v_2, v_3]^\top \\ &= [-(\omega_2 + \omega_{2,a}), \omega_1 + \omega_{1,a}, v_3']^\top = [-\omega_2', \omega_1', v_3']^\top\end{aligned}$$

und die Überlagerung von passiver $\boldsymbol{\omega}$ und aktiver $\boldsymbol{\omega}_a$ Rotationsgeschwindigkeit

$$\boldsymbol{\omega}' = [\omega_1', \omega_2', \omega_3']^\top = [(\omega_1 + \omega_{1,a}), \omega_2 + \omega_{2,a}, \omega_3]^\top\,.$$

Damit reduziert sich die Zahl der freien Parameter um eins, weil das Element $s_6 = -(\omega_1' v_1' + \omega_2' v_2') = -(\omega_1'\omega_2' - \omega_1'\omega_2') = 0$ der symmetrischen Epipolarkomponente $\mathbf{S}$ Null wird. Die Dimension des Gleichungssystems (10) kann um eins reduziert werden

$$\mathbf{a}' := [-u_2, u_1, u_2 x - u_1 y, x^2, 2xy, 2x, y^2, 2y]^\top \in \mathbb{R}^8, \tag{14a}$$
$$\mathbf{s}' := [v_1', v_2', v_3', s_1, s_2, s_3, s_4, s_5]^\top \in \mathbb{R}^8 \tag{14b}$$

[4] $\lambda = 1/T_3$ entspricht dem unbekannten Skalierungsfaktor.

und man benötigt nur noch mindestens $N \geq 7$ Punktpaare $\{(\mathbf{x}_i, \mathbf{u}_i)\}_{i=1}^N$ um das lineare Gleichungssystem $\mathbf{A}'\mathbf{s}' = 0$ mit $\mathbf{A}' \in \mathbb{R}^{N\times 8}$ zu lösen. Wegen der Reduktion des Gleichungssystems auf mindestens 7 lineare Gleichungen durch eine zusätzliche bekannte aktive Rotationsbewegung $\boldsymbol{\omega}_a$, wird der Algorithmus im Folgenden als *aktiver linearer 7-Punkt-Algorithmus* benannt[5]. Weitere algorithmische Vereinfachungen ergeben sich durch den aktiven Navigationsansatz nicht. Auch hier muss eine Projektion der geschätzten Matrix $\widehat{\mathbf{S}}$ in den symmetrischen Epipolarraum erfolgen. Danach kann die Kamerabewegung V analog zum kontinuierlichen linearen 8-Punkt-Algorithmus aus Abschnitt 2.3 extrahiert werden.

2.6 Virtuell-aktive Kamera

Für rein rotatorische Kamerabewegungen ist die Verschiebung der projizierten Raumpunkte auf dem Kamerabild unabhängig von den Entfernungen der Raumpunkte. Zudem werden keine Raumpunkte von anderen verdeckt bzw. kommen keine neuen Raumpunkte hinter Verdeckungen hervor. Deswegen kann die Ansicht, die nach (vor) der Rotation zu sehen ist für den Bereich des Sichtfeldes, der auch vor (nach) der Rotation zu sehen war, ebenso ohne die Rotationsbewegung aktiv auszuführen, erzeugt werden. Diese Idee kann durch die geometrische Transformation der Rückwärtsabbildung [20] mit dem durch die Rotation erzeugten optischen Fluss realisiert werden und wird im Folgenden als virtuell-aktive Kamera bezeichnet. Im Bereich der Videovorverarbeitung von handgefilmten Aufnahmen werden ähnliche Verfahren zur Kompensation von Handbewegungen herangezogen und mit Videostabilisierung bezeichnet [21].

Die der Roboterbewegung überlagerte rotatorische Kamerabewegung $\boldsymbol{\omega}_a$, die das Fixieren des projizierten Punktes im Bildhauptpunkt gewährleistet, erzeugt einen bestimmten optischen Fluss $\mathbf{u}_a$. Die zu diesem Fluss gehörigen Flussgleichungen erhält man aus Gleichung (6), wenn man für die Rotationsbewegung die aktive Rotationsbewegung $\boldsymbol{\omega} = \boldsymbol{\omega}_a$ einsetzt und die Translationsbewegung gleich Null $\mathbf{v} = \mathbf{0}$ setzt:

$$\mathbf{u}_a = \widehat{\boldsymbol{\omega}}_a\mathbf{x} - \frac{\dot{Z}_\mathcal{C}}{Z_\mathcal{C}}\mathbf{x}\,, \quad \text{mit} \quad \frac{\dot{Z}_\mathcal{C}}{Z_\mathcal{C}} = \omega_{1,a}y - \omega_{2,a}x\,. \tag{15}$$

Daraus ergeben sich die von den Entfernungen $Z_\mathcal{C}$ unabhängigen Flusskomponenten

$$u_{1,a} = (1+x^2)\omega_{2,a} - xy\omega_{1,a}\,, \tag{16a}$$
$$u_{2,a} = -(1+y^2)\omega_{1,a} + xy\omega_{2,a} \tag{16b}$$

und die geometrische Vorwärtstransformation [20], die die Positionen der projizierten Bildpunkte in den Bildern

[5] Nicht zu verwechseln mit dem nichtlinearen 7-Punkt-Algorithmus aus [23]. Dort wird die *Fundamentalmatrix* der *diskreten* Epipolareinschränkung über ein *nichtlineares* Gleichungssystem, das aus 7 Punktkorrespondenzen erzeugt wird, berechnet.

vor $\mathbf{x}^v$ und nach $\mathbf{x}^n$ der Rotation verknüpft:

$$x^n = x^v + u_{1,a} = x^v + (1 + (x^v)^2)\omega_{2,a} - x^v y^v \omega_{1,a}\,, \tag{17a}$$

$$y^n = y^v + u_{2,a} = y^v - (1 + (y^v)^2)\omega_{1,a} + x^v y^v \omega_{2,a}\,. \tag{17b}$$

Aus dem Fluss $\mathbf{u}_o$ des Bildhauptpunktes $\mathbf{x}_o = [0,0]^\top$ erhält man durch Einsetzen in Gleichung (16) die benötigte Rotation $\boldsymbol{\omega}_a$, welche den Bildhauptpunkt fixiert:

$$\omega_{1,a} = -u_{2,o}\,, \quad \omega_{2,a} = u_{1,o}\,. \tag{18}$$

Den Fluss des Bildhauptpunktes $\mathbf{u}_o$ erhält man über ein beliebiges Trackingverfahren, z. B. über eine normierte Kreuzkorrelation oder die Methode von Lucas und Kanade [18]. Die gesuchte Rückwärtsabbildung, welche die entstehenden Bildpunktkorrespondenzen bei einer gegebenen virtuell-aktiven Rotation beschreibt, ergeben sich dann aus dem negativen Fluss des Bildhauptpunktes zu

$$x^v = x^n - (1 + (x^n)^2)u_{1,o} - x^n y^n u_{2,o}\,, \tag{19a}$$

$$y^v = y^n - (1 + (y^n)^2)u_{2,o} - x^n y^n u_{1,o}\,. \tag{19b}$$

Damit auch Punkte fixiert werden können, die nicht im Bildhauptpunkt liegen, kann durch eine zusätzliche Rückwärtsabbildung des Ausgangsbildes jeder beliebige Fixationspunkt in den Bildhauptpunkt verschoben werden. Das Ergebnis ist in Bild 3 zu sehen. Bild 3a zeigt den Bildfluss (schwarze Striche) von strukturstarken Bildpunkten (weiße Kreise) zwischen dem Bild in 3a und dem darauf folgenden Bild in 3b ohne eine Fixation des Fixationspunktes (schwarzer Kreis) im Bildhauptpunkt (weißes Quadrat). Durch die Fixation des Fixationspunktes im Bildhauptpunkt wie in Bild 3c und d zu sehen (schwarzer Kreis liegt nun in weißem Quadrat) ergibt sich der Bildfluss relativ zur kompensierten Fixationspunktbewegung.

Dieser Relativfluss führt zu genaueren Abschätzungen der Kamerabewegung. Der Hauptgrund dafür ist der folgende: Der Relativfluss besitzt durch die Fixation betragsmäßig kleinere Flussvektoren. Bei kleineren Beträgen können aktuelle Flussschätzer den Fluss genauer bestimmen [22]. Die Beträge und Richtungen der Vektoren verteilen sich relativ zum Bildhauptpunkt gleichmäßiger über den Bildort als ohne eine Fixation des Bildhauptpunktes. Der Fluss im Bildhauptpunkt beträgt jetzt Null. Da die Beträge und Richtungen der Vektoren gleichmäßiger über den Bildort verteilt sind, sind auch die Fehler des verwendeten Flussschätzers räumlich gleichmäßiger verteilt. Dadurch liefert das lineare Ausgleichsproblem des 7- bzw. 8-Punkt-Algorithmus bessere Ergebnisse, wie in Abschnitt 3 gezeigt wird.

Es gibt noch weitere Vorteile dieser virtuell-aktiven Kamera. Die Kamerabewegung kann beliebig schnell in alle Richtungen simuliert werden, da Massenträgheit und Stellgrößenbegrenzungen wie bei einer realen aktiven Kamera nicht vorhanden sind. Das genaue Fixieren eines Raumpunktes mit einer realen aktiven Kamera erfordert eine präzise Mechanik, wohingegen die virtuell-aktive Kamera so genau fixieren kann, wie es die Genauigkeit des verwendeten Trackingverfahrens erlaubt. Je nachdem, wie eindeutig und stabil die Struktur um den ausgewählten Fixationspunkt im Bild ist, kann beispielsweise mit dem Tomasi-Kanade-Verfahren [25] Subpixelgenauigkeit erreicht werden. Da die Fixation nur virtuell geschieht können in einem Bild gleichzeitig mehrere Fixationspunkte in parallelen Verarbeitungsprozessen virtuell fixiert werden. So kann zum Beispiel eine virtuelle Fixation mit günstigen Flussvektorverteilungen für die Berechnung der Translationsgeschwindigkeit und eine weitere virtuelle Fixation mit günstigen Flussvektorverteilungen für die Berechnung der Rotationsbewegung realisiert werden.

Der entscheidende Nachteil der virtuell-aktiven Kamera ist in Bild 3c und d zu sehen. Es werden keine neuen Szenenteile sichtbar (schwarzer Bereich am Bildrand), die nicht im vorhandenen Sichtfeld vor der Fixation zu sehen sind. Deswegen wird der Sichtbereich bei der virtuellen

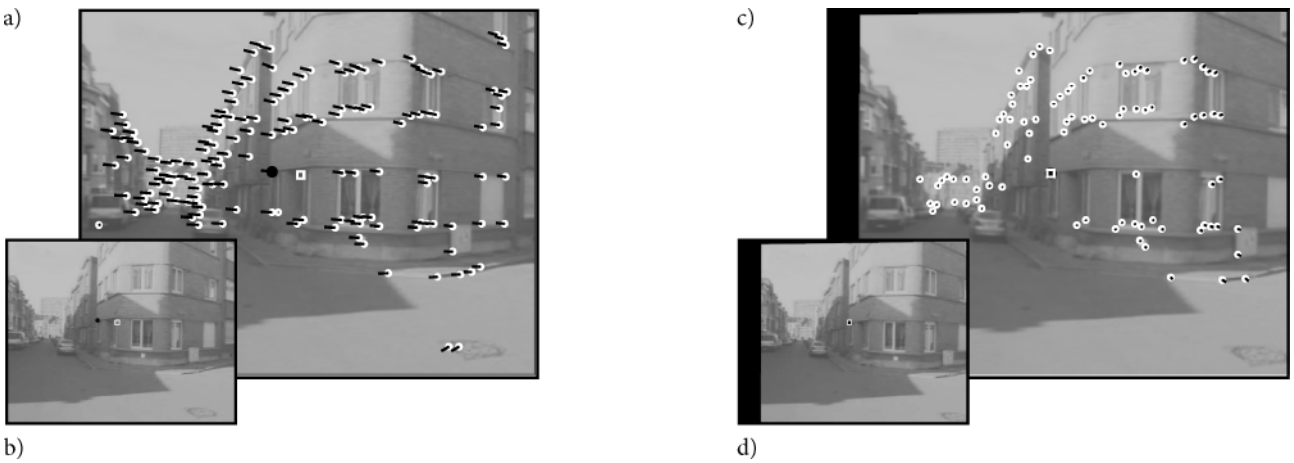

Bild 3 Unterschiede im Bildfluss (schwarze Striche) von kontrastreichen Bildpunkten (weiße Kreise) **(a)** ohne und **(c)** mit Fixation im Bildhauptpunkt (weißes Quadrat). Ohne virtuell-aktive Kamera bewegt sich der ausgewählte Fixationspunkt (schwarzer Kreis) zwischen zwei Bildern **(a)** und **(b)** der Bildfolge. Bei einer Fixation des Fixationspunktes durch eine virtuell-aktive Kamera verharrt der Fixationspunkt (schwarzer Kreis) zwischen zwei Bildern **(c)** und **(d)** der Bildfolge an der Koordinate des Bildhauptpunktes (weißes Quadrat).

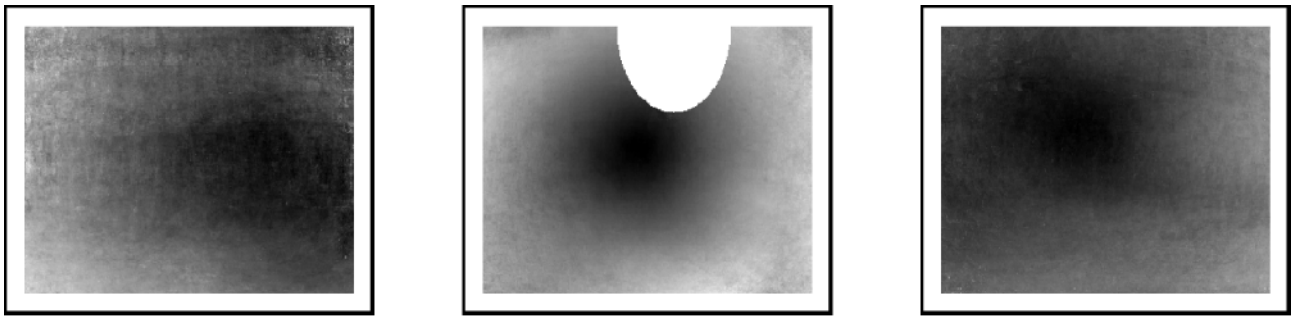

Bild 4 Fixationsabhängige mittlere Schätzfehler für verschiedene Bewegungskategorien – (**a**) Linkskurvenfahrt (**b**) Geradeausfahrt und (**c**) Rechtskurvenfahrt. Je heller der Grauton, desto größer ist der Schätzfehler. Schwarz entspricht einem Schätzfehler von Null. Weiß sind Bereiche, für die keine Schätzung möglich war bzw. zu wenig Schätzungen wegen fehlender Grauwertstruktur (z. B. Himmel) vorhanden sind.

Fixation umso kleiner, je weiter der Fixationspunkt sich vom Bildhauptpunkt entfernt.

Es stellt sich nun die Frage, was für Vorteile die Fixation im Bildhauptpunkt für die Schätzung der Eigenbewegung bringt, außer dass sich die Gleichungen des 8-Punkt-Algorithmus zu denen des aktiven 7-Punkt-Algorithmus reduzieren. Beziehungsweise, ob durch die Fixation eines bestimmten Bildpunktes die Schätzgenauigkeit verbessert werden kann.

2.7 Adaptive Fixationspunktauswahl

Um die Frage zu beantworten, ob eine Fixation eine Auswirkung auf die Schätzgüte von visuellen Odometrialgorithmen haben kann, wurden in dieser Arbeit Videosequenzen[6] ausgewertet, für die ziemlich genaue visuelle Odometriedaten mittels des Algorithmus aus [16] zur Verfügung stehen. Diese werden hier als Referenzbewegungen herangezogen.

Zuerst wurden die Videosequenzen in drei unterschiedliche Bewegungskategorien – Geradeausfahrt, Rechtskurve und Linkskurve – unterteilt. Dann wurde für jede Kategorie für jedes Bildpaar jeder Sequenz die Kamerabewegung für jeden möglichen Fixationspunkt mit dem 8-Punkt-Algorithmus berechnet. Das bedeutet, man bekommt für jeden virtuell fixierbaren Bildpunkt, also jeden Bildpunkt, der erfolgreich mittels Rückwärtsabbildung in den Bildhauptpunkt verschoben und mittels Kanade-Tomasi-Tracker im Bildhauptpunkt gehalten werden konnte, eine eigene Schätzung der Kamerabewegung. Dann wurde für jeden Fixationspunkt der Betrag der Differenz zwischen der geschätzten Kamerabewegung und der Referenzbewegung berechnet und über alle Bilder einer Kategorie für jede Fixationspunktkoordinate getrennt gemittelt. Diese fixationsabhängigen mittleren Schätzfehler sind in Bild 4 für Linkskurvenfahrt (a) Geradeausfahrt (b) und Rechtskurvenfahrt (c) abgebildet.

Man erkennt deutlich, dass die unterschiedlichen Kategorien, unterschiedliche Fehlerverteilungen zeigen. Bei allen drei Kategorien läßt sich eine radial verlaufende Fehlerzunahme mit unterschiedlichen Minimumspositionen feststellen. Während bei der Geradeausfahrt das Minimum im Bereich des Bildhauptpunktes liegt, verschiebt sich bei einer Links- bzw. Rechtskurve das Minimum nach rechts bzw. links. Anhand dieser Fehlerhistogramme wird ein Verfahren zur Auswahl eines Fixationspunktes in Abhängigkeit von der Kategorie und der Lage des Bildhauptpunktes abgeleitet: Um die aktuelle Kamerabewegung einer Kategorie zuzuordnen wird zuerst der Bildfluss von visuellen Merkmalen ohne Fixation berechnet. Der Punkt des geringsten Abstandes zu allen Geraden entlang der einzelnen Flussvektoren, wird dann als einfaches aber robustes Auswahlkriterium verwendet. Liegt dieser Punkt innerhalb des Bildes, dann wird die Kategorie Geradeausfahrt gewählt. Liegt er links oder rechts außerhalb des Bildes, dann wird entsprechend

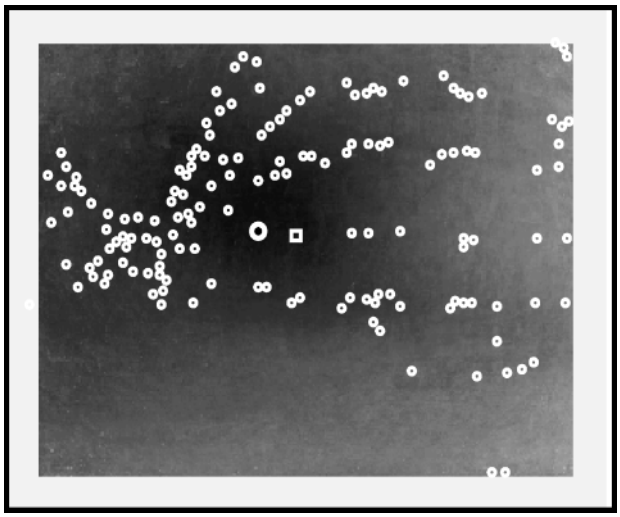

Bild 5 Fixationspunktauswahl (großer weißer Kreis) in Abhängigkeit von der Trackinggüte (kleine weiße Kreise), der Fehlerstatistik (Grauwertbild, schwarz entspricht geringstem Fehler) der Bewegungskategorie und dem Abstand zum Bildhauptpunkt (weißes Viereck).

[6] Die Videosequenzen [16] sind aus einem fahrenden Auto mit einer Bildwiederholrate von 25 Hz aufgenommen worden und von Nico und Kurt Cornelis von der KU Leuven veröffentlicht worden.

Kategorie Links- bzw. Rechtskurvenfahrt gewählt. Als Fixationspunkt wird dann aus allen visuellen Merkmalen (siehe Bild 5, kleine weiße Kreise), die einen bestimmten Wert im Fehlerhistogramm (siehe Bild 5, Grauwertbild) unterschreiten, dasjenige ausgesucht, das am nächsten zum Bildhauptpunkt (siehe Bild 5, weißes Viereck) liegt. In Bild 5 ist ein Beispiel einer Fixationspunktauswahl (großer weißer Kreis) visualisiert.

3 Auswertung

Der Vergleichsalgorithmus nach [16] besteht aus einer klassischen *Structure-from-Motion*-Verarbeitungskette [23]. Durch Hinzunahme von Triangulation [18] kombiniert mit einem *Sliding-Window*-Bündelblockausgleich [18] werden die Ergebnisse nochmals optimiert und die Fehleraufintegration reduziert [4]. Das Ergebnis der sich ergebenden Trajektorie (VO nach [16]) aus 1000 Bildern der Videosequenz (was einer Fahrtzeit bei 25 Hz Bildwiederholrate von 40 Sekunden entspricht) ist in Bild 6 zu sehen. Die Beträge der Translationsvektoren wurden auf eins normiert, da sie ohne zusätzliche Tiefeninformation nicht berechenbar sind. Deswegen sind die Achsen der x-y-Ebene einheitenlos.

Der kontinuierliche 8-Punkt-Algorithmus nach [17] (siehe Bild 6), der als Basisalgorithmus für ein lokales visuelles Odometrieverfahren angesehen werden kann, schneidet im Vergleich am schlechtesten ab. Benutzt man die virtuell-aktive visuelle Odometrie (VaO) in Kombination mit dem aktiven 7-Punkt-Algorithmus, ergibt sich eine Verbesserung zum 8-Punkt-Algorithmus ohne Fixation (siehe Bild 6). Es bleibt jedoch eine deutliche Abweichung zum Vergleichsalgorithmus nach [16] bestehen. Der Grund dafür ist die Ungenauigkeit bei der Fixation des Fixationspunktes mit dem Tomasi-Kanade-Tracker. Dadurch entstehen leichte Abweichungen in der aktiven Rotation $\boldsymbol{\omega}_a$ und Gleichung (13) ist nicht mehr exakt erfüllt. Damit ist Element s_6 der symmetrischen Epipolarkomponente nicht mehr exakt Null und die Bedingung zur Reduktion beim aktiven 7-Punkt-Algorithmus nicht mehr exakt gegeben. Verzichtet man auf diese Reduktion und wendet den kontinuierlichen 8-Punkt-Algorithmus *inklusive* der Fixation an, dann können die Abweichungen in der aktiven Rotation $\boldsymbol{\omega}_a$ ausgeglichen werden. Man erreicht mit diesem *lokalen*[7] Schätzverfahren *ohne globale* Nachoptimierung mittels Bündelblockausgleich (siehe Bild 6) nahezu die Schätzergebnisse des globalen Verfahrens nach [17], welches eine globale Nachoptimierung in jedem Zeitschritt beinhaltet. Der Rechenaufwand für die Rückwärtsabbildung um die Fixation zu realisieren ist wesentlich geringer als der Rechenaufwand zur Realisierung eines Bündelblockausgleichs. Damit kann das virtuell-aktive Odometrieverfahren als potentielle Alternative zu bestehenden Verfahren angesehen werden.

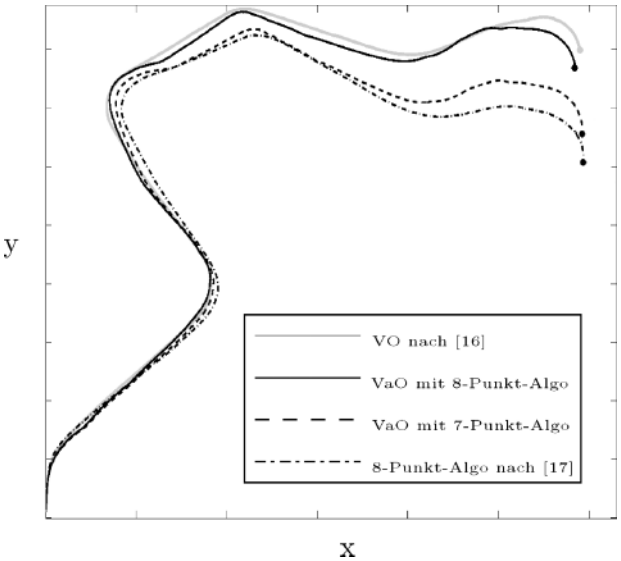

Bild 6 Vergleich der Trajektorienverläufe in der x-y-Ebene für verschiedene visuelle Odometrieverfahren.

4 Zusammenfassung und Ausblick

Dieser Artikel hat gezeigt, dass die Genauigkeit beim aktiven Navigationsansatz entscheidend von der Wahl des Fixationspunktes und der Genauigkeit der Fixation abhängt. Zur Wahl des Fixationspunktes wurde ein Verfahren vorgestellt, das zu einer Verbesserung der Schätzgüte von visuellen Odometrieverfahren führt. Da selbst durch die hier eingeführte virtuell-aktive Kamera die Fixation bisher nicht genau genug umgesetzt werden konnte, kann der Vorteil der Parameterreduktion in den Gleichungen der Epipolareinschränkung noch nicht zufriedenstellend ausgenutzt werden. Dies wurde durch die Herleitung des aktiven 7-Punkt-Algorithmus und dem Vergleich mit dem bekannten 8-Punkt-Algorithmus bei Fixation gezeigt. Es bleibt zu zeigen, ob ein genaueres Trackingverfahren, als der hier benutzte Tomasi-Kanade-Tracker, die Genauigkeit des aktiven 7-Punkt-Algorithmus noch erhöhen kann. Desweiteren sollte gezeigt werden, ob die Flussfeldkonfigurationen bei der gleichzeitigen virtuell-aktiven Fixation von einem Punkt und einer Linie noch günstiger ausfallen, und dadurch eine weitere Verbesserung der Schätzgüte der virtuell-aktiven visuellen Odometrie möglich ist.

Danksagung

Bei Harald Maier und Dong Gu Kim bedanke ich mich herzlich für die Umsetzung des beschriebenen Verfahrens zur virtuell-aktiven visuellen Navigation in Matlab und die Auswertung der Testdaten.

Literatur

[1] R. Froböse, „Maschinenauge, sei wachsam“, *Pictures of the Future, Die Zeitung für Forschung und Innovation, Siemens, Herbst 2011.*

[7] Jede Schätzung basiert nur auf zwei Bildern und es werden keine Annahmen an Zusammenhänge zwischen Kamerabewegungen oder Bilddaten in einem größeren Zeitfenster berücksichtigt.

[2] R. He, A. Bachrach, M. Achtelik, A. Geramifard, D. Gurdan, S. Prentice, J. Stumpf and N. Roy, „On the Design and Use of a Micro Air Vehicle to Track and Avoid Adversaries“.
[3] D. Nister, O. Naroditsky and J. Bergen „Visual Odometry for Ground Vehicle Applications“, *Journal of Field Robotics, 23(1): 3–20, 2006.*
[4] D. Scaramuzza and F. Fraundorfer „Visual Odometry, Part I: The First 30 Years and Fundamentals“, *IEEE Robotics & Automation Magazine, 18(4): 80–91, 2011.*
[5] D. Scaramuzza and F. Fraundorfer „Visual Odometry, Part II: Matching, Robustness, Optimization and Applications“, *IEEE Robotics & Automation Magazine, 18(4): 1–11, 2012.*
[6] D. Kragic and M. Vincze, „Vision for Robotics“, *Foundations and Trends in Robotics, 1(1): 1–78, 2010.*
[7] J. Aloimonos, I. Weiss, and A. Bandyopadhyay „Active Vision“, *International Journal of Computer Vision, 1(4): 333–356, 1988.*
[8] R. Bajcsy „Active Perception“, *Proceedings of the IEEE, 76(8): 996–1005, 1988.*
[9] B. Mertsching „Aktives Sehen – Eine kurze Einführung“, *Künstliche Intelligenz, Heft 1: 5–6, 1999.*
[10] D. Konstantinos „Fixation simplifies 3D motion estimation“, *Computer Vision and Image Understanding, 68(2): 158–169, 1996.*
[11] S. Soatto and P. Perona „Reducing Structure from Motion: A General Framework for Dynamic Vision, Part I: Modeling“, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(9): 933–942, 1998.*
[12] A. Bandyopadhyay „A Computational Study of Rigid Motion Perception“, *University of Rochester, Dissertation, 1986.*
[13] T. Vidal-Calleja, A. J. Davison, J. Andrade-Cetto and D. W. Murray „Active Control for Single Camera SLAM“, *IEEE International Conference on Robotics and Automation, S. 1930–1936, 2006.*
[14] A. J. Davison „Active Search for Real-Time Vision“, *IEEE International Conference on Computer Vision, S. 66–73, 2005.*
[15] A. J. Davison and D. W. Murray „Simultaneous Localisation and Map Building using Active Vision“, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7): 865–880, 2002.*
[16] B. Leibe, N. Cornelis, K. Cornelis, and L. Van Gool „Dynamic 3D Scene Analysis from a Moving Vehicle“, *IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, June 2007.*
[17] Y. Ma, J. Kosecka and S. Sastry „Linear differential algorithm for motion recovery: A geometric approach“, *International Journal of Computer Vision, 44(3): 219–249, 2000.*
[18] Y. Ma, S. Soatto, J. Kosecka and S.Sastry „An Invitation to 3-D Vision“, *Springer Verlag, 2004.*
[19] C. Voigt and J. Adamy „Formelsammlung der Matrizenrechnung“, *Oldenbourg Wissenschaftsverlag, 2007.*
[20] B. Jähne „Digitale Bildverarbeitung“, *6. Auflage, Springer-Verlag, 2005.*
[21] C. Kurz, T. Thormählen and H. P. Seidel „Scene-Aware Video Stabilization by Visual Fixation“, *Conference for Visaul Media Production, S. 1–6, 2009.*
[22] K. Pauwels, M. Lappe and M. van Hulle „Fixation as a Mechanism for Stabilization of Short Image Sequences“, *International Journal of Computer Vision, 72(1): 67–78, 2007.*
[23] R. Hartley and A. Zisserman „Multiple View Geometry in Computer Vision“, *Camebridge University Press, 2000.*
[24] J. Shi and C. Tomasi „Good features to track“, *IEEE Conference on Computer Vision and Pattern Recognition, S. 593–600, 1994.*
[25] C. Tomasi and T. Kanade „Shape and motion from image streams under orthography“, *International Journal of Computer Vision, 9(2): 137–154, 1992.*

Manuskripteingang: 1. Mai 2012

Dr.-Ing. Volker Willert ist Gruppenleiter der Forschergruppe *Maschinelles Sehen und Mobile Robotik* des Fachgebietes Regelungstheorie und Robotik unter der Leitung von Prof. Dr.-Ing. J. Adamy am Institut für Automatisierungstechnik und Mechatronik im Fachbereich Elektrotechnik und Informationstechnik der Technischen Universität Darmstadt. Hauptarbeitsgebiete: Probabilistische Inferenzmethoden in der Regelungstechnik, Bildverarbeitung für mobile Systeme, mobile Multi-Agenten-Systeme.

Adresse: Technische Universität Darmstadt, Fachbereich Elektrotechnik und Informationstechnik, Fachgebiet Regelungstheorie und Robotik, D-64283 Darmstadt,
E-Mail: vwillert@rtr.tu-darmstadt.de