UNIVERSITÄT KOBLENZ · LANDAU
Fachbereich 4: Informatik

AGAS
Arbeitsgruppe Aktives Sehen

# Bildbasierte Bewegungsschätzung aus Kamerafahrten anhand prägnanter Merkmale

**Diplomarbeit**

**zur Erlangung des Grades eines/r**
**Diplom-Informatikers / Diplom-Informatikerin**
**im Studiengang Computervisualistik**

vorgelegt von

Peter Decker

| | |
|---|---|
| Betreuer: | Prof. Dr.-Ing. Dietrich Paulus, Institut für Computervisualistik, Fachbereich Informatik |
| Erstgutachter: | Prof. Dr.-Ing. Dietrich Paulus, Institut für Computervisualistik, Fachbereich Informatik |
| Zweitgutachter: | Dipl.-Inf. Tobias Feldmann, Institut für Computervisualistik, Fachbereich Informatik |

Koblenz, im September 2007

## Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Richtlinien der Arbeitsgruppe für Studien- und Diplomarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden. ja ☐ nein ☐

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu. ja ☐ nein ☐

Koblenz, den ..... Unterschrift

# Danksagung

Ich möchte meinen Betreuern Prof. Dr.-Ing. Dietrich Paulus und Dipl.-Inf. Tobias Feldmann für ihre Unterstützung beim Erstellen dieser Arbeit danken. Außerdem gilt mein Dank allen Mitarbeitern und Studenten der Arbeitsgruppe Aktives Sehen, für ihre Hilfe und fachlichen Beistand. Ebenso danke ich meiner Familie dafür, dass sie mir das Studium ermöglicht hat.

# Inhaltsverzeichnis

# Kapitel 1

# Einleitung

In vielen Bereichen ist Wissen über Eigenbewegung oder die Lage im Raum Voraussetzung für angemessenes Handeln oder eine korrekte Bewertung anderer Daten. So sind z. B. in der Robotik Odometrieinformationen zur Lösung der SLAM-Problematik (simultaneous localization and mapping) nötig. Außerdem können Informationen über die Rotation zusätzliche Hinweise auf eine ungewollte Schräglage des Roboters geben. Eine weitere Anwendungsmöglichkeit besteht in AR (augmented reality) Szenarien. Hier ist die Bestimmung der Pose des Anwenders zu einem Objekt notwendig, um Interaktion im Sinne der AR und eine korrekte Augmentierung der realen Welt zu ermöglichen.

Kameras sind heutzutage kostengünstig, weit verbreitet und werden in viele Bereichen eingesetzt. Einzelne Bilder tragen potentiell große Mengen an Information, die genutzt werden können, um die beschriebenen Probleme zu lösen. Eine Möglichkeit bietet hier also die bildbasierte Bewegungsschätzung. Dabei wird aus einer Bildfolge die Bewegung der Kamera rekonstruiert. Dieses Verfahren ist auch als Visuelle Odometrie bekannt.

In dieser Arbeit stelle ich einen Ansatz zur Visuellen Odometrie vor, der eine Poseschätzung zwischen aufeinanderfolgenden, monokularen Frames ermöglicht und diese anschließend in eine globale Trajektorie einreiht.

Der Aufbau der Arbeit ist wie folgt:

In Kapitel 2 werden bekannte Verfahren zur Visuellen Odometrie und verwandte Metho-

den vorgestellt. In Kapitel 3 werden die verwendeten Algorithmen beschrieben und Veränderungen sowie Ergänzungen im Vergleich zum Stand der Technik dargelegt. Ein Hauptaugenmerk liegt hierbei auf der Menge der auf RANSAC (random sample consensus) aufbauenden Verfahren bis hin zu BEEM (balanced exploration and exploitation model search for efficient epipolar estimation). Zusätzlich wird das verwendete Kriterium für die Zuordnung von SIFT-Punktkorrespondenzen formalisiert und motiviert. Schließlich wird die Notwendigkeit des Prüfens auf reduzierte Bewegungsmodelle im Fall von auf dem 8-Punkte-Algorithmus basierenden Verfahren zur Schätzung der Epipolargeometrie herausgestellt. Im Anschluss findet sich eine Evaluation der interessantesten Aspekte des verwendeten Verfahrens, und dessen Grenzen werden aufgezeigt. Es wurden dazu sowohl reale wie auch synthetische Kamerafahrten zur Bewertung herangezogen. In Kapitel 5 werden die Ergebnisse zusammengefasst. Im Anhang sind die verwendeten mathematischen Verfahren detailliert beschrieben und es werden Hinweise zur Implementation gegeben.

# Kapitel 2

# Visuelle Odometrie

Viele Ansätze zur SLAM-Problematik unterscheiden sich von der Visuellen Odometrie in Zahl und Art der genutzten Sensorik. Strelow und Singh kombinieren z. B. inertiale Sensoren des Roboters sowohl mit omnidirektionalen als auch mit konventionellen Kameras, um stabilere Ergebnisse zu erzielen [SS03]. Auf dem Gebiet der reinen Visuellen Odometrie sind grundsätzlich zwei unterschiedliche Herangehensweisen zu unterscheiden:
Der monokulare und der Stereoaufbau. Beim Stereoaufbau ist eine genaue metrische Rekonstruktion der Bewegung möglich. Im Fall der monokularen Visuellen Odometrie bleibt die Translation stets mit einem unbekannten skalaren Faktor behaftet. Nister beschreibt beide Methoden in der Anwendung bei Bodenfahrzeugen [NBN06] sowie die monokulare Herangehensweise für eine fliegende Plattform [NNB04]. Davison beschäftigt sich mit den Herausforderungen unter Echtzeitbedingungen [Dav03] und nutzt als Anwendungsfall einen tragbaren Roboter mit Optik [DMM03].

Ein gemeinsames Teilproblem all dieser Ansätze ist die Schätzung der Bewegung zwischen zwei Frames. Im Fall eines Stereoaufbaus kann die Pose mit lediglich 3 rekonstruierten 3-D-Punktkorrespondenzen geschätzt werden [HLON91]. Bei der monokularen Visuellen Odometrie ist das Problem äquivalent zur Schätzung der Epipolargeometrie. Es existieren neben bekannten Verfahren wie dem normalisierten 8-Punkte-Algorithmus [Har97] Methoden, die mit weniger Punktkorrespondenzen auskommen, um eine Fundamental-

Matrix zu instanziieren [Nis03].

Voraussetzung für die korrekte Schätzung der Epipolargeometrie sind korrekte Zuordnungen zwischen Punkten der verschiedenen Frames sowie die Detektion von Outliern. Der Random Sample Consensus (RANSAC [FB81]) war der erste Algorithmus, mit dem auch bei einer großen Anzahl von Outliern zuverlässige Ergebnisse erzielt werden konnten, was mit anderen Ansätzen wie der least square Optimierung nicht möglich war. In den folgenden Jahren wurden zahlreiche Verbesserungsvorschläge für den ursprünglichen Algorithmus gemacht:
Torr und Zisserman änderten die Bewertungsfunktion der Güte des geschätzten Modells von der reinen Zahl der Inlier zu einer Maximum Likelihood Funktion [TZ00]. In [MC02] schlägt J. Matas ein verfrühtes Abbruchkriterium bei der Verifikation der Modelle zur Beschleunigung des Algorithmus vor. Chum erweitert RANSAC zu LO-RANSAC [CMK03] durch das Hinzufügen eines weiteren Schritts der lokalen Optimierung. Dieser wurde immer dann ausgeführt, wenn RANSAC ein Modell mit neuer maximaler Inlierzahl generiert hatte. Wie durch eine geschickte Wahl der verwendeten Punktkorrespondenzen, die zur Erzeugung des Modells genutzt werden, der Rechenaufwand reduziert werden kann, verdeutlicht er mit PROSAC [CM05]. Ilan Shimshoni kombinierte und erweiterte diese Ansätze im letzten Jahr zu BEEM [Shi06], einem Verfahren, das neben zufälliger globaler Suche, sowohl eine lokale Suche in der Nähe der gefundenen besten Lösung als auch ein iteratives Verfahren zur Verbesserung des aktuellen Modells in sich vereint. Es beschreibt außerdem die Möglichkeit, aus nur 2 SIFT-Merkmalskorrespondenzen eine Fundamental-Matrix zu instanziieren und aus degenerierten Konfigurationen zu entkommen. Mit dem Erkennen und Behandeln von degenerierten Konfigurationen, beispielsweise dominanten Ebenen im Bild, beschäftigten sich auch schon Chum, Werner und Matas [CWM05]. Torr, Zisserman und Maxbank lieferten eine komplette Übersicht über mögliche Bewegungsmodelle der Kamera, welche zu degenerierten Schätzung der Epipolargeometrie führen können. [TZM95]

# Kapitel 3

# Bildbasierte Bewegungsschätzung aus Kamerafahrten anhand prägnanter Merkmale

Zur Lösung des Problems wurde ein mehrstufiger Algorithmus entwickelt, der sich aus folgenden Teilen zusammensetzt:
Die einzige Vorarbeit besteht in einer initialen Kalibrierung der verwendeten Kamera. Anschließend können Bilder sukzessiv eingespeist werden, und die Kamerapose wird sowohl in Relation zu dem vorangehenden Frame als auch zu einem globalen Weltkoordinatensystem ermittelt. Dazu wird zuerst das Bild entzerrt, anschließend werden SIFT-Merkmale extrahiert. Mit Hilfe dieser Merkmale wird versucht, unter Verwendung eines angepassten BEEM-Algorithmus eine F-Matrix mit maximal möglichem Support Set zu instanziieren. Im Anschluss werden verschiedene reduzierte Bewegungsmodelle ebenfalls getestet, um Fehler aufgrund von degenerierten Konfigurationen auszuschließen. Aus den Informationen über die gewonnene Epipolargeometrie lässt sich die Bewegung zwischen den zwei Frames rekonstruieren und schließlich in die globale Trajektorie integrieren. Das Aktivitätsdiagramm 3.1 verdeutlicht diesen Ablauf.

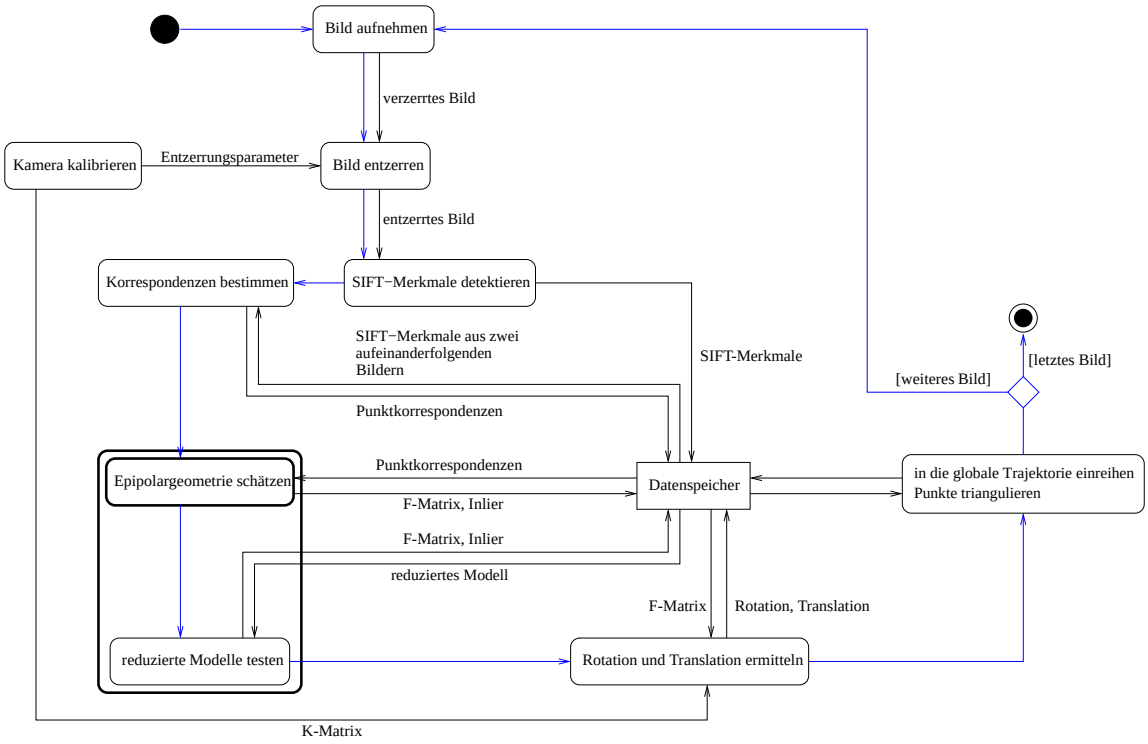


Bild 3.1: Ablauf der Verfahrens. Zur besseren Übersichtlichkeit ist der Kontrollfluss in blau, der Datenfluss in schwarz dargestellt. Die fett umrandeten Aktivitäten weichen vom Stand der Technik ab und werden in den entsprechenden Kapiteln genauer erläutert.

## 3.1 Überblick

Im Folgenden werden die verwendeten Algorithmen beschrieben, ihre Verwendung motiviert und Abweichungen vom Stand der Technik erläutert und begründet. Die vorverarbeitenden Schritte entsprechen hierbei bekannten und bewährten Verfahren. Zur Schätzung der Epipolargeometrie wurde der noch sehr junge BEEM-Algorithmus erstmals nach *C++* portiert und dem Anwendungsszenario angepasst. Zusätzlich wird er mit der Schätzung von reduzierten Modellen verknüpft, um eine robustere Rekonstruktion der Trajektorie aus Bildfolgen zu ermöglichen.

## 3.2 Kamerakalibrierung

Ziel der Kamerakalibrierung ist die Bestimmung der Kalibriermatrix $\boldsymbol{K}$, welche die intrinsischen Kameraparameter enthält. Des weiteren werden die Entzerrungsparameter, mit deren Hilfe die Linsenverzerrung ausgeglichen werden kann, bestimmt. Die Kalibrierung erfolgt im Vorfeld der eigentlichen Anwendung mit Hilfe eines Kalibriermusters nach der Methode von Zhang [Zha00]. Da die Ergebnisse jeweils für eine ganze Bildserie relevant sind und an mehreren Stellen (3.7.1, B.4, vgl. Bild 3.1) gebraucht werden, werden die Daten stets persistent in Dateien abgelegt.

## 3.3 Extraktion prägnanter Merkmale

Die Schätzung der Bewegung zwischen aufeinanderfolgenden Frames geschieht über eine Schätzung der Epipolargeometrie, welche die nötigen Informationen enthält. Die gängigen Verfahren zur Berechnung von Epipolargeometrien sind merkmalsbasiert, die überwältigende Mehrheit beruht hierbei auf Punktmerkmalen. In dieser Arbeit werden SIFT-Merkmale, wie von David Lowe in [Low04] beschrieben, verwendet, um Korrespondenzen zu finden und schließlich die F-Matrix zu schätzen. SIFT-Merkmale haben gegenüber anderen Punktmerkmalen mehrere Vorteile. Sie sind subpixelgenau und verfügen über einen charakteristischen Merkmalsdeskriptor, der die Zuordnungssuche (3.4) erleichtert und sogar die Grundlage für weitergehende Optimierungen bietet (3.5.3, 3.5.8). Außerdem sind die Merkmale invariant gegenüber Rotationen, welche im vorliegenden Anwendungsfall dringend mit berücksichtigt werden müssen.

## 3.4 Korrespondenzsuche zwischen Merkmalen

Ziel ist es, zwei Merkmalen aus (aufeinanderfolgenden) Bildern, welche Abbildungen des selben Punktes in der Welt sind, einander zuzuordnen. Man sagt auch, die Merkmale korrespondieren.
Sei $\mathrm{arise}(f, \boldsymbol{p}^{\mathrm{w}})$ die Relation, die Auskunft darüber gibt, ob das Merkmal $f$ einer Ab-

bildung des Weltpunktes $\boldsymbol{p}^{\mathrm{w}}$ entspringt. Daraus ergibt sich die gewünschte Relation für Korrespondenzen:

$$\mathrm{corr}(f_1, f_2) \Leftrightarrow \mathrm{arise}(f_1, \boldsymbol{p}^{\mathrm{w}}) \wedge \mathrm{arise}(f_2, \boldsymbol{p}^{\mathrm{w}}) \tag{3.1}$$

Da $\mathrm{arise}$ nicht direkt geprüft werden kann, wird die Ähnlickeit der Deskriptoren der Merkmale als Indiz dafür genommen, dass zwei Merkmale Abbildungen des selben Weltpunktes sind. Im Falle von SIFT-Merkmalen ist der Deskriptor $\boldsymbol{d}_f$ des Merkmals $f$ ein Vektor der Länge 128: $\boldsymbol{d}_f \in \mathbb{Z}^{128}$.
Die Funktion $\mathrm{dist} : \mathbb{Z}^{128} \times \mathbb{Z}^{128} \mapsto \mathbb{R}$ für den Abstand zweier Merkmalsdeskriptoren in Bezug auf deren Ähnlichkeit kann über die Länge des Differenzvektors definiert werden:

$$\mathrm{dist}(\boldsymbol{d}_{f_1}, \boldsymbol{d}_{f_2}) = \|\boldsymbol{d}_{f_1} - \boldsymbol{d}_{f_2}\| \tag{3.2}$$

Unklar ist an dieser Stelle noch die Norm, die der Betragsfunktion zugrunde gelegt wird. In [GIM99] wird auf mehrere Quellen verwiesen, die zu dem Schluss kommen, dass der Unterschied zwischen $l_1$ und $l_2$ Norm in hochdimensionalen Räumen für die Nachbarsuche zu vernachlässigen ist. Wir verwenden also als Abstandsmaß von zwei Merkmalsdeskriptoren die $l_1$-Norm:

$$\mathrm{dist}(\boldsymbol{d}_{f_1}, \boldsymbol{d}_{f_2}) = \sum_{i=1}^{124} |\boldsymbol{d}_{f_1 i} - \boldsymbol{d}_{f_2 i}| \tag{3.3}$$

So entsteht die symmetrische Relation $\mathrm{corr}(f_1, f_2)$, die festlegt, ob die Merkmale $f_1 \in F_1$ und $f_2 \in F_2$ korrespondieren. 3.4 fordert einen wechselseitig minimalen Abstand der Deskriptorverktoren:

$$\begin{aligned} \mathrm{corr}(f_1, f_2) \Leftrightarrow (\nexists f_i \in F_1 \mid \mathrm{dist}(f_i, f_2) < \mathrm{dist}(f_1, f_2)) \\ \wedge (\nexists f_j \in F_2 \mid \mathrm{dist}(f_1, f_j) < \mathrm{dist}(f_1, f_2)) \end{aligned} \tag{3.4}$$

Weitere Bedingungen können aus anwendungsabhängigen Vorgaben entstehen. Beispielsweise lässt sich eine maximale zugelassene Verschiebung in Pixelkoordinaten annehmen, wenn bekannt ist, dass die Bewegung der Kamera entsprechend gering, beziehungsweise die Framerate genügend hoch ist. Die Gefahr hierbei ist jedoch, bei schnellen Bewegungen

die Korrespondenzinformationen von Punkten, die am nächsten am Optischen Zentrum der Kamera liegen und somit die größte Translation erfahren, zu verlieren. Gerade diese Punkte bieten jedoch die meiste Information für die Schätzung der Epipolargeometrie, da die starke Bewegung, die sie zwischen den Frames erfahren, etwaige Ungenauigkeiten in der Lokalisation aufgrund von Rauschen bei weitem überwiegt. Da keine so gearteten Randbedingungen an den Algorithmus gestellt werden, wird hier von einer Verwendung einer Begrenzung der maximalen Bewegung von Pixeln zwischen aufeinanderfolgenden Frames abgesehen.

Korrekt einander zugeordnete Merkmale heißen Inlier. Merkmalspaare, die als korrespondierend angenommen werden, obwohl sie nicht dem selben Weltpunkt entspringen, nennt man Fehlzuordnungen oder auch Outlier. Die Gründe für solche Fehlzuordnungen sind verschieden.
Neben optischen Phänomenen wie sich bewegenden Glanzpunkten oder Spiegeln, welche im Bildentstehungsmodell nicht berücksichtigt werden, gehören sich wiederholende Strukturen in einem Bild und einfache Zufallskorrespondenzen zu den Hauptverursachern von Outliern. Eine naive Herangehensweise zur Eindämmung der zufälligen Fehlzuordnungen wäre es, einen Grenzwert für den maximalen Abstand der Deskriptorvektoren einzuführen. Leider ist diese Herangehensweise für solch hochdimensionale Räume wie den Merkmalsraum von SIFT-Merkmalen nicht gut geeignet. David Lowe empfiehlt als Maß für die Zuverlässigkeit einer Zuordnung die Distance Ratio $r$ [Low04]. Die Distance Ratio ist der Quotient aus dem Abstand zum potentiellen korrespondierenden Merkmal und dem Abstand zum zweitbesten. Ist dieser Quotient minimal, so ähnelt das zugeordnete Merkmal seinem Partner weitaus mehr als dem nächstbesten. Durch die Nutzung dieses Wissens wird sowohl die Wahrscheinlichkeit der Fehlzuordnung aufgrund eines sich wiederholenden Musters minimiert als auch die Anzahl der zufälligen falschen Zuordnungen reduziert. Geht der Quotient jedoch auf $1$ zu, so existieren offensichtlich weitere ähnliche Merkmale in dem Bild, was auf eine hohe Wahrscheinlichkeit einer möglichen Fehlzuordnung hindeutet. Es könnte sich um ein wiederkehrendes Merkmal handeln oder das passende Merkmal war nur zufällig das beste, hat jedoch keine stärker ausgeprägte Ähnlichkeit als ein beliebiges anderes. Als Schwellwert für gute Zuordnungen wird allgemein ein Wert von $r = 0.8$ angenommen. Alle Korrespondenzen mit einer schlechteren Distan-

ce Ratio werden verworfen.
Die Funktion für die Distance Ration $r$ einer Korrespondenz der Merkmale $f_1 \in F_1$ und $f_2 \in F_2$ ergibt sich wie folgt:

$$
\begin{aligned}
& r(\mathrm{corr}(f_1, f_2)) = \\
& \max\left(\frac{\mathrm{dist}(f_1, f_2)}{\mathrm{dist}(f_1, f_i)} \mid f_i \in F_2 \wedge (\nexists f_j \in F_2 \setminus f_2 \mid \mathrm{dist}(f_1, f_j) < \mathrm{dist}(f_1, f_i)),\right. \\
& \left.\frac{\mathrm{dist}(f_1, f_2)}{\mathrm{dist}(f_k, f_2)} \mid f_k \in F_1 \wedge (\nexists f_l \in F_1 \setminus f_1 \mid \mathrm{dist}(f_l, f_2) < \mathrm{dist}(f_k, f_2))\right)
\end{aligned}
\tag{3.5}
$$

Wie aus Gleichung 3.5 ersichtlich, wird das Maximum der zwei möglichen Distance Ratios angenommen. Dies hilft beispielsweise in Fällen, in denen eine Struktur in einem Bild lediglich einmal, im anderen Bild jedoch häufiger auftritt. Des weiteren unterstützt es den Anspruch, dass die Korrelation von Merkmalen und die damit verbundenen Relationen symmetrisch, also unabhängig von der Reihenfolge der Frames ist.

Eine gängige Technik der Bildverarbeitung ist die hierarchische Herangehensweise an komplexe Probleme. Die Idee ist, durch Reduktion der Information die Berechnung zu beschleunigen und gegebenenfalls Teilbereiche im Anschluss genauer zu untersuchen. Im Prinzip lässt sich ein ähnliches Verfahren auch im Fall der Merkmalsextraktion und Zuordnung anwenden, indem vorab Bilder mit geringerer Auflösung verwendet werden. Der zu erwartende Informationsverlust ist aufgrund der Subpixelgenauigkeit der SIFT-Merkmale sehr gering, auch wenn ausschließlich mit den geringeren Auflösungen gearbeitet wird. Dies birgt jedoch eine nicht gleich offensichtliche, inhärente Gefahr.
Betrachtet man die Wahrscheinlichkeit für eine zufällige Fehlzuordnung unter Miteinbeziehung des Distance Ratio Kriteriums (3.5), so ist offensichtlich, dass die Wahrscheinlichkeit für eine zufällige Fehlzuordnung unter anderem invers mit der Anzahl (nicht zugeordneter) Merkmale korreliert. Durch eine geringe Zahl an Merkmalen steigt die Wahrscheinlichkeit, dass zwei Merkmale sich zufällig gegenseitig am ähnlichsten sind. Gleichzeitig sinkt die Aussagekraft der Distance Ratio aufgrund der geringen Zahl an möglichen Vergleichspartnern, und damit die Wahrscheinlichkeit, auf diese Weise die Fehlzuordnung aufzudecken. In Bild 3.2 ist dies verdeutlicht.
Durch eine geringere Auflösung wird also offensichtlich nicht nur die Zahl der gefundenen

Merkmale abnehmen, sondern gleichzeitig der prozentuale Anteil an zufälligen Fehlzuordnungen steigen.

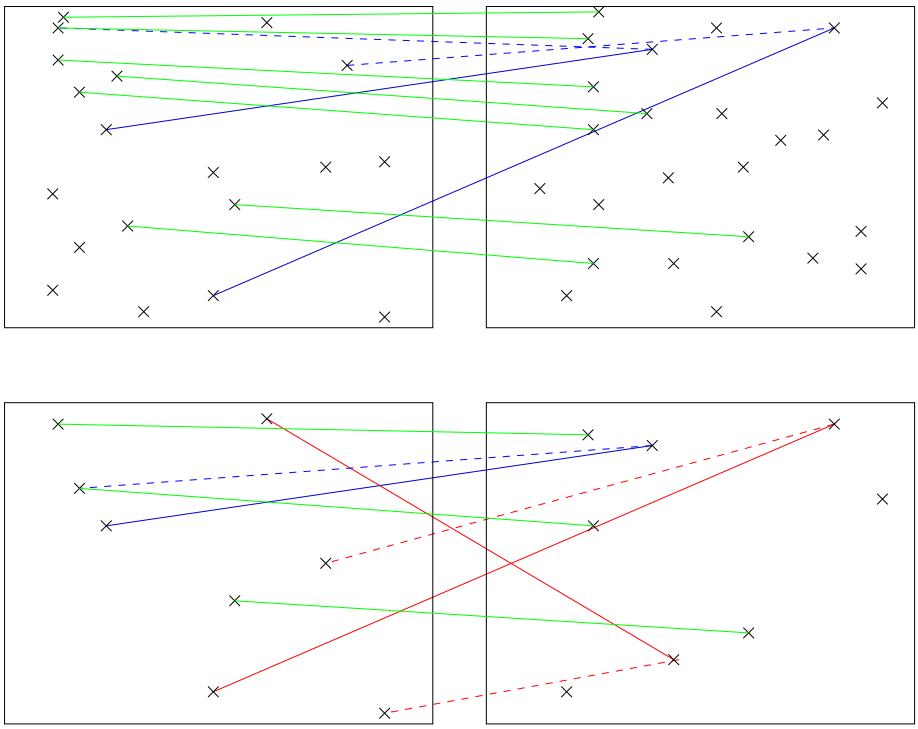

Bild 3.2: Grün: korrekte Zuordnungen; Blau: falsche Zuordnunge mit $r > 0.8$; Rot: falsche Zuordnungen mit $r \leq 0.8$; Gestrichelt: die zweitbeste Zuordnung - aus Gründen der Übersichlichkeit nur in eine Richtung dargestellt.

## 3.5 Schätzen der Epipolargeometrie

Der nächste Schritt ist die Schätzung der Fundamental-Matrix zwischen zwei aufeinanderfolgenden Bildern mit Hilfe der gefundenen Korrelationen. Zwei grundsätzliche Herangehensweisen können hier unterschieden werden: Iterative Algorithmen und ein linearer Ansatz. Aufgrund der Einfachheit und weil er sich weit verbreitet und bewährt hat, wird der Normalisierte 8-Punkte-Algorithmus verwendet [Har97].

### 3.5.1 Normalisierter 8-Punkte-Algorithmus

Der 8-Punkte-Algorithmus versucht die Epipolargeometrie in Form einer Fundamental-Matrix zu schätzen. Hierzu wird mit mindestens acht korrespondierenden Punktepaaren eine Messwertmatrix aufgestellt. Diese ist so aufgebaut, dass die Komponenten der Fundamental-Matrix in Vektorform im rechten Nullraum der Matrix stehen. Die wichtige Idee des Normalisierten 8-Punkte-Algorithmus ist eine Normalisierung, welche vor dem Aufstellen der Messwertmatrix durchgeführt wird. Die Koordinaten der korrespondierenden Punkte jedes Bildes werden so verschoben und skaliert, dass ihr Zentrum im Ursprung des Koordinatensystems liegt und der Mittelwert des Abstands dazu $\sqrt{2}$ beträgt. Dies garantiert eine kleinere Konditionszahl und eine stabilere Schätzung der F-Matrix, da numerische Instabilitäten vermieden werden. Nach der Lösung des aus der Messwertmatrix entspringenden linearen Gleichungssystems mit Hilfe der Singulärwertzerlegung wird durch Erzwingen des Ranges der F-Matrix (siehe Anhang B.2) die ähnlichste valide F-Matrix gefunden, welche anschließend noch in die ursprünglichen Koordinatensysteme zurücktransformiert werden muss.

### 3.5.2 RANSAC

Der 8-Punkte-Algorithmus kann nur ein korrektes Modell erstellen, wenn alle Punktkorrespondenzen, die zur Berechnung herangezogen werden, Inlier sind. Wie in 3.4 dargelegt, lassen sich allerdings Outlier unter den Zuordnungen nicht völlig vermeiden. Um also eine korrekte Schätzung zu erstellen und zu erkennen, wird ein Verfahren benötigt, welches die

Möglichkeit hat, aus outlierbehafteten Daten ein möglichst gutes Modell zu erstellen. Gut bedeutet in dem Zusammenhang, möglichst nahe der Realität, und somit ausschließlich konsistent mit allen Inliern.
Der Random Sample Consensus (RANSAC [FB81]) ist ein Paradigma, welches es ermöglicht, gute Modelle in fehlerbehaftete experimentelle Daten einzupassen. Es handelt sich um ein iteratives Verfahren, welches kontinuierlich neue Modelle aus einem minimalen, zufällig gewählten Datensatz erzeugt. Anhand der Menge der Daten, die das Modell unterstützen, misst es dessen Qualität. Diese Daten nennt man das *Support Set*, in diesem Fall sind es alle korrespondierenden Punktepaare $(\tilde{\boldsymbol{q}}^{\mathrm{p}}, \tilde{\boldsymbol{p}}^{\mathrm{p}})$, die die Epipolare Bedingung 3.6 für eine Fundamental-Matrix $\boldsymbol{F}$ erfüllen. Als Approximation wird für gewöhnlich der Schwellwert $\theta$ verwendet.

$$\begin{aligned} \tilde{\boldsymbol{q}}^{\mathrm{p}\mathrm{T}}\, \boldsymbol{F}\, \tilde{\boldsymbol{p}}^{\mathrm{p}} &= 0 \\ |\tilde{\boldsymbol{q}}^{\mathrm{p}\mathrm{T}}\, \boldsymbol{F}\, \tilde{\boldsymbol{p}}^{\mathrm{p}}| &\leq \theta \end{aligned} \tag{3.6}$$

Hauptkritikpunkt an RANSAC ist die hohe Anzahl an Iterationen $I$, die benötigt werden, um mit genügend großer Wahrscheinlichkeit $p$ ein gutes Modell zu finden (siehe Bild 3.3). Diese hängt sowohl von dem Anteil der Inlier $\alpha$ in den Messdaten ab als auch von der Größe $s$ der benötigteten Stichprobe zur Instanziierung eines Modells. Zusätzlich ist die Wahrscheinlichkeit, mit der mindestens ein Modell nur aus Inliern instanziiert werden soll zwar theoretisch variabel, sollte jedoch in der Praxis mit $p = 0.99$ angenommen werden. Dadurch soll verhindert werden, dass der Algorithmus terminiert, ohne ein gutes Modell gefunden zu haben.
Aus den erwähnten Größen ergibt sich folgender Zusammenhang:

$$(I - \alpha^s)^I = 1 - p \tag{3.7}$$

Gleichung 3.7 beschreibt näherungsweise die Wahrscheinlichkeit, bei $I$ Iterationen kein outlierfreies Modell zu generieren. Näherungsweise deshalb, weil $\alpha^s$ ein Ziehen mit Zurücklegen darstellt, was nicht der Realität entspricht, da Korrespondenzen nur einfach in die Modellerstellung einfließen. Gleichung 3.7 lässt sich umformen und nach $I$ auflösen, um ein Maß für die Zahl der benötigten Iterationen zu erhalten.

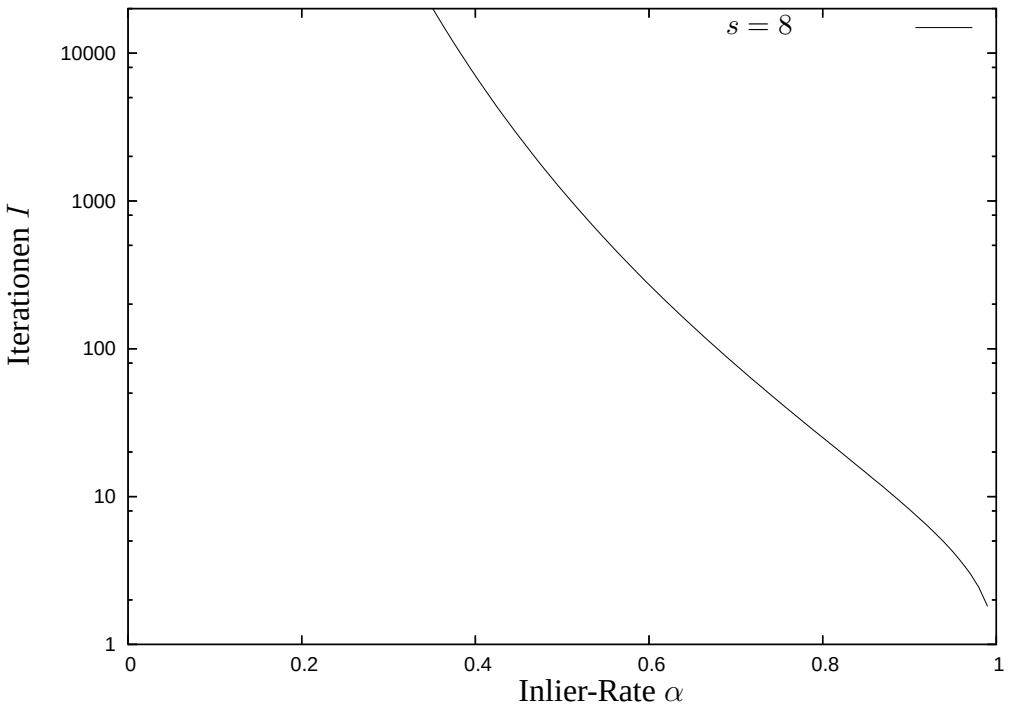


Bild 3.3: Aufwandsbetrachtung RANSAC für $I = \frac{log(1-0{,}99)}{log(1-\alpha^s)}$. Um bei einer Inlier-Rate $\alpha$ mit einer festen Wahrscheinlichkeit von $99\%$ mindestens ein Inliermodell mit dem Algorithmus zu generieren, werden $I$ Iterationen benötigt.

$$I = \frac{log(1-p)}{log(1-\alpha^s)} \tag{3.8}$$

Die einzige Möglichkeit, die Anzahl der benötigten Iterationen bei gleich bleibendem $p$ zu reduzieren, liegt offenbar darin, die Inlierrate $\alpha$ zu erhöhen oder die Zahl der benötigten Messwerte $s$ zu verringern. Im Folgenden werden mehrere Varianten von RANSAC vorgestellt, die dies mehr oder weniger direkt erreichen.

### 3.5.3 PROSAC

Der Progressive Sample Consensus (PROSAC [CM05]) ist in seiner Anwendung beschränkter als RANSAC. Er versucht durch zusätzliches Wissen, welches nur im Spezialfall der Suche nach Modellen, die aus korrespondierenden SIFT-Merkmalen entstehen, vorhanden ist, das Verfahren zu verbessern. Während RANSAC die Stichprobe zur Instanziierung jedes Modells vollkommen zufällig wählt, werden bei PROSAC bevorzugt Korrespondenzen mit guter Distance Ratio (siehe 3.4, 3.5) herangezogen. Unter der Annahme, dass die Distance Ratio als Ähnlichkeitsmaß der Merkmalsvektoren besser geeignet ist als zufälliges Ziehen, ist hier eine Beschleunigung des Algorithmus zu erwarten, da die Wahrscheinlichkeit, einen Inlier zu wählen, auf diese Weise größer ist als deren prozentualer Anteil $\alpha$.

### 3.5.4 LO-RANSAC

Der Locally Optimized RANSAC (LO-RANSAC [CMK03]) ergänzt den Standardansatz um einen zusätzlichen Schritt der lokalen Optimierung, der immer, wenn ein neues bestes Modell gefunden wurde, ausgeführt wird. In diesem Schritt wird aus dem Support Set des aktuellen Modells ein neues Modell erstellt. Die Rechtfertigung für diesen zusätzlichen Schritt ist die Feststellung, dass in der Praxis ein durch den 8-Punkte-Algorithmus berechnetes Modell aufgrund des Erzwingens des Rangs der Matrix oft zu ungenau wird, um alle tatsächlichen Inlier als solche zu erkennen. Werden jedoch mehr Punkte als die minimal notwendige Menge von acht verwendet, so stabilisiert sich das Ergebnis. Aufgrund des Aufwands von RANSAC (Gleichung 3.8) kann nicht von vornherein eine größere Menge an Punktepaaren $s$ zur Berechnung herangezogen werden, ohne die Zahl der Iterationen $I$ unnötig zu erhöhen.

In [CMK03] wurden neben dem Standard-RANSAC vier unterschiedliche Verfahren der lokalen Optimierung untersucht und bewertet:

**Einfach:** Bei dem einfachen Ansatz wird mit allen Punkten, die die Epipolare Bedingung 3.6 für einen Schwellwert $\theta$ erfüllen, linear ein neues Modell generiert.

**Iterativ:** Im Fall des Iterativen Ansatzes wählt man einen Schwellwert $k \cdot \theta$, berechnet wieder mit allen Inliern ein neues Modell, verringert den Schwellwert und wiederholt das Vorgehen solange, bis der Schwellwert $\theta$ erreicht ist.

**Innerer RANSAC:** Beim Inneren RANSAC wird erneut eine Stichprobe zur Modellschätzung gezogen. Die Punkte dazu werden ausschließlich aus dem Support Set $S_k$ des Modells der aktuellen $k$-ten RANSAC-Schätzung entnommen. Da es sich um Inlier handelt, muss die Stichprobe nicht minimal sein. Es wird eine Größe von $\min(\frac{S_k}{2}, 14)$ im Fall der F-Matrix Schätzung vorgeschlagen. Die Zahl der Inneren RANSAC Iterationen wird auf 10 gesetzt.

**Innerer RANSAC mit Iterationen:** Die Methode des Inneren RANSAC mit Iterationen kombiniert die beiden vorhergehenden, indem sie bei jeder der zehn inneren Iterationen die iterative Verfeinerung durchführt.

### 3.5.5 R-RANSAC

In [MC02] wird ein neues Abbruchkriterium für den Evaluationsschritt definiert, in welchem eine durch RANSAC generierte Hypothese getestet wird. Die Idee ist ein vorangehender Test auf einem zufälligen Datensatz aus den Testdaten, welcher aber deutlich kleiner ist. Nur wenn dieser Test erfolgreich absolviert wird, werden die restlichen Daten getestet. So können sehr schlechte Modelle, die nur ein verschwindend geringes Support Set haben, sehr schnell erkannt werden, und die Zeit für eine komplette Evaluierung der Qualität des Modells kann gespart werden.

### 3.5.6 Degenerierte Fundamental-Matrizen

Degenerierte Modelle entstehen, wenn die Eingabedaten nicht ausreichend sind, um eine eindeutige Lösung zu generieren. Dies führt im vorliegenden Fall zu mehreren Problemen. In realen Bildern ist es nicht auszuschließen, dass ein Großteil der gut zuzuordnenden Merkmale eine degenerierte Konfiguration bilden, also zur Schätzung einer degenerierten F-Matrix führen. Ein konkretes Beispiel wäre eine dominante Ebene im Bild, wie in

[CWM05] beschrieben. Aber auch Merkmalskorrespondenzen, die alle aus einem kleinen Bildausschnitt stammen, können zu falschen Schätzungen der F-Matrix führen. Die Gefahr hierbei liegt darin, dass die so generierten Modelle dadurch, dass alle zur Degeneration gehörenden Korrespondenzen in ihrem Support Set liegen, zwar von RANSAC aufgrund des großen Support Sets als sehr gutes Modell eingeschätzt werden, jedoch in Wirklichkeit nicht die komplette Bildgeometrie beschreiben. Der BEEM Ansatz ([Shi06], 3.5.7) versucht durch seinen Schritt der lokalen Optimierung, aus solch degenerierten Konfigurationen zu entkommen. Es werden absichtlich Korrespondenzen zur Berechnung eines neuen Modells herangezogen, welche nicht Teil des Support Sets des aktuellen Modells sind, in der Hoffnung, dass das aktuelle Modell einer degenerierten Konfiguration von Merkmalskorrespondenzen entsprungen ist, der so entkommen werden kann. Dieser Ansatz erweist sich jedoch als gefährlich, sollte die Kamerabewegung Grund für die Degeneration sein und nicht etwa die Wahl der Korrespondenzen. Im Fall einer reinen Translation beispielsweise werden für die Beschreibung des Modells der Kamerabewegung weniger Parameter benötigt als durch den 8-Punkte-Algorithmus geschätzt werden. Das geschätztes Modell entspricht also einer höherdimensionalen Bewegung als sie in der Realität erfolgt ist. Dies führt dazu, dass eine aus Inliern geschätzte F-Matrix ihr Support Set erweitern kann, wenn zusätzlich ein Outlier in die Schätzung miteinbezogen wird. Zur geometrischen Veranschaulichung ist in [TZM95] folgendes Beispiel gegeben:

Es soll eine Linie in eine Menge von 2-D-Punkten eingepasst werden (siehe Bild 3.4-Bild 3.6). Die soliden Geraden repräsentieren mögliche Lösungen, die gestrichelten Linien zeigen die Grenzen, in denen Punkte als konsistent mit dem Modell angesehen werden. In Bild 3.4 ist eine Menge nicht degenerierter Daten gegeben, welche zu einem eindeutigen Modell führen.

In Bild 3.5 hingegen liegt eine degenerierte Konfiguration der Eingangsdaten vor. Die Punkte beschreiben eher einen Punkt als eine Linie (das Äquivalent zu einer reinen Translation, wenn nach einer F-Matrix gesucht wird). Es sind offensichtlich mehrere Lösungen für eine Linie möglich. Besser wäre an dieser Stelle, statt einer Linie einen Punkt als reduziertes Modell zu verwenden, da dieses den Eingabedaten viel eher entspricht. Übertragen auf das Problem der Schätzung der Epipolargeometrie heißt das: Liegt eine reine Translation vor, so sind viele unterschiedliche F-Matrizen mit einem großen Support Set möglich.

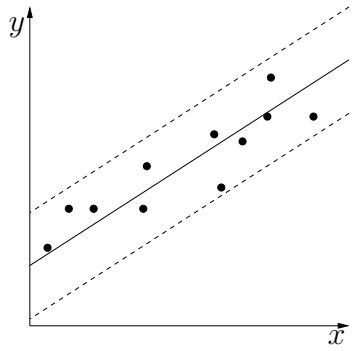


Bild 3.4: Ein nicht degeneriertes Datenset. Die Lösung ist eindeutig.

Es wäre jedoch besser, ein Modell für eine reine Translation aufzustellen, da es der Realität eher entspricht.

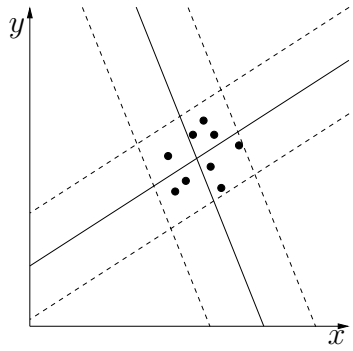


Bild 3.5: Degenerierte Daten, mehrere Lösungen sind möglich.

Ein zusätzliches Problem bei degenerierten Lösungen ist, dass ein einzelner Outlier, der in die Berechnung mit einbezogen wird, eine degenerierte Konfiguration als nicht degeneriert erscheinen lassen kann (siehe Bild 3.6). Das so entstehende Modell ist zwar mathematisch korrekt und hat das größtmögliche Support Set, entspricht jedoch nicht der Wirklichkeit. In diesem Fall wäre es angemessener, nach einem Modell mit weniger Parametern (also einem Punkt respektive einer reinen Translation) zu suchen, und so den Outlier als solchen zu erkennen.

Durch das Testen auf reduzierten Modellen kann also verhindert werden, dass Outlier zur

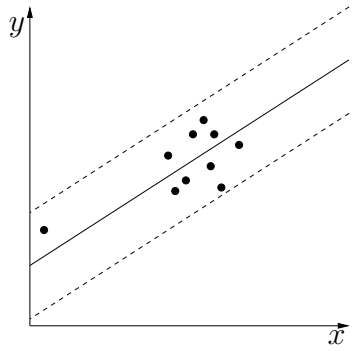


Bild 3.6: Ein einzelner Outlier lässt das degenerierte Datenset als nicht degeneriert erscheinen.

Schätzung der Epipolargeometrie im Falle einer degenerierten Konfiguration miteinbezogen und nicht als solche erkannt werden. Dazu müssen mögliche reduzierte Modelle identifiziert und untersucht werden. Im Fall von Visueller Odometrie sind vor allem die reine Translation der Kamera (z. B. bei der Vorwärtsbewegung eines Roboters), eine Drehung um das Optische Zentrum und die Identität, also Bewegungslosigkeit, von Relevanz. Zur Schätzung des Zusammenhangs zwischen zwei aufeinanderfolgenden Bildern sollte also neben der Suche nach einer F-Matrix mit größtmöglichem Support Set noch geprüft werden, ob ein reduziertes Modell existiert, dessen Support Set nur unwesentlich kleiner ist. In diesem Fall liegt nahe, dass die eigentliche Bewegung der des reduzierten Modells entspricht, und das größere Support Set der durch den BEEM oder ähnliche Algorithmen geschätzten F-Matrix auf die Miteinbeziehung von Outliern zurückzuführen ist. Hierzu wird ein Schwellwert für die Größe des Support Sets des reduzierten Modells verwendet. In [TZM95] wird als Grenze $t_S = 0.95$ vorgeschlagen. Liefert also ein reduziertes Modell ein Support Set, dessen Größe $95\%$ des Support Sets der F-Matrix Schätzung nicht unterschreitet, so wird das reduzierte Modell als korrekt angenommen.

### 3.5.7 BEEM

Der originale BEEM-Algorithmus [Shi06] vereint viele der bereits erläuterten Methoden, sowie das Locality-Sensitive Hashing [GIM99] zur Beschleunigung der Korrespondenzsu-

che. Außerdem beschreibt er eine Methode zum Schätzen der Epipolargeometrie aus nur zwei SIFT-Merkmalen. Die grundlegende Idee hinter dem Algorithmus ist ein Zusammenspiel von einem vorangehenden, und vier Hauptzuständen, die abhängig vom momentan aktuellen Modell und dem Vorgängerzustand versuchen, eine Lösung mit maximalem Support Set für die F-Matrix zu finden.
Die Zustände des originalen BEEM-Algorithmus sind:

- Prior Estimation
- Global Exploration
- Exploitation
- Local Exploration
- Model Quality Estimation

**Prior Estimation:** Ein erster Zustand, der nur einmal angenommen wird, ist die Prior Estimation. Hier wird unter Verwendung von empirischen Daten aus den Distance Ratios der Merkmalskorrespondenzen versucht, den Anteil $\alpha$ an Inliern in den angenommenen Korrespondenzen zu schätzen. Mit Hilfe dieser Daten lässt sich für jede Korrespondenz eine Wahrscheinlichkeit $P_{in}$ ermitteln, ob es sich um einen Inlier handelt. Der Algorithmus wechselt anschließend in den Zustand der Global Exploration.

**Global Exploration:** In diesem Zustand versucht der Algorithmus, durch globale Exploration eine initiale Schätzung der Epipolargeometrie zu erreichen. Um die Anzahl der Iterationen zu verringern, die benötigt werden, um eine F-Matrix ausschließlich aus Inliern zu instanziieren, wird an dieser Stelle das 2-SIFT-Verfahren verwendet (3.5.8). Hiermit ist es möglich, aus nur zwei SIFT-Korrespondenzen eine Epipolargeometrie zu schätzen. Besonders bei einem geringen Anteil $\alpha$ an Inliern reduziert dies den zu erwartenden Aufwand des Verfahrens erheblich (siehe Bild 3.7). Zusätzlich wird das Wissen aus der Prior Estimation verwendet, um im Sinne von PROSAC bevorzugt Korrespondenzen mit einer hohen Inlierwahrscheinlichkeit $P_{in}$

zu wählen. Sollte die so geschätzte F-Matrix ein für diesen Zustand neues maximales Support Set aufweisen, schließt sich der Exploitation Zustand an, ansonsten die Model Quality Estimation.

**Exploitation:** Im Exploitation Zustand wird eine lokale Optimierung auf dem aktuellen Modell durchgeführt. Dies geschieht durch eine iterative Verfeinerung wie in LO-RANSAC beschrieben, allerdings mit einem kleinen Unterschied. Jedes mal, wenn ein neues lokales Optimum gefunden wurde, startet der Prozess erneut. Nach zehn Iterationen ohne Verbesserung schließt sich der Zustand der Model Quality Estimation an.

**Local Exploration:** In diesem Schritt werden bewusst Korrespondenzen, die nicht im Support Set liegen, für die nächste Schätzung der F-Matrix miteinbezogen. Die Menge an Zuordnungen, aus der das neue Modell geschätzt wird, richtet sich nach der Größe des Support Sets $|S_{best}|$ des aktuell besten Modells. Es werden $min(|\frac{S_{best}}{2}|, 13)$ Korrespondenzen aus dem Support Set der aktuell besten Lösung, sowie eine Korrespondenz, die nicht im Support Set liegt, verwendet. Letztere liefert hierbei vier Merkmale, wie im 2-SIFT-Verfahren beschrieben. Der Algorithmus hat so bessere Chancen, aus degenerierten Konfigurationen zu entkommen, wie sie z. B. bei einer merkmalsreichen dominanten Ebene im Bild auftreten. Sobald die Wahrscheinlichkeit $P_q$ (siehe Model Quality Estimation) mit $1.0$ geschätzt wird, werden in diesem Schritt nacheinander alle Merkmale, die nicht im Support Set liegen, nach ihrer Inlierwahrscheinlichkeit $P_{in}$ geordnet und getestet. Wird auf diese Weise ein neues bestes Support Set für diesen Zustand gefunden, folgt der Exploitation Schritt, ansonsten die Model Quality Estimation.

**Model Quality Estimation:** Ziel des Zustandes ist es, die Variable $P_q$ zu schätzen. $P_q$ beschreibt die Wahrscheinlichkeit, dass das Support Set des aktuellen Modells nicht deshalb erreicht wurde, weil zufällig Outlier konsistent mit dem Modell sind. Sollte $P_q$ auf $1.0$ geschätzt werden, also die Wahrscheinlichkeit, dass das Support Set nur zufällig seine Größe erreicht hat, minimal sein, so kann der Algorithmus terminieren. Eine zusätzliche Bedingung für den Abbruch ist noch, dass im Local Exploration Schritt bereits alle Korrespondenzen, die nicht im Support Set liegen, getestet

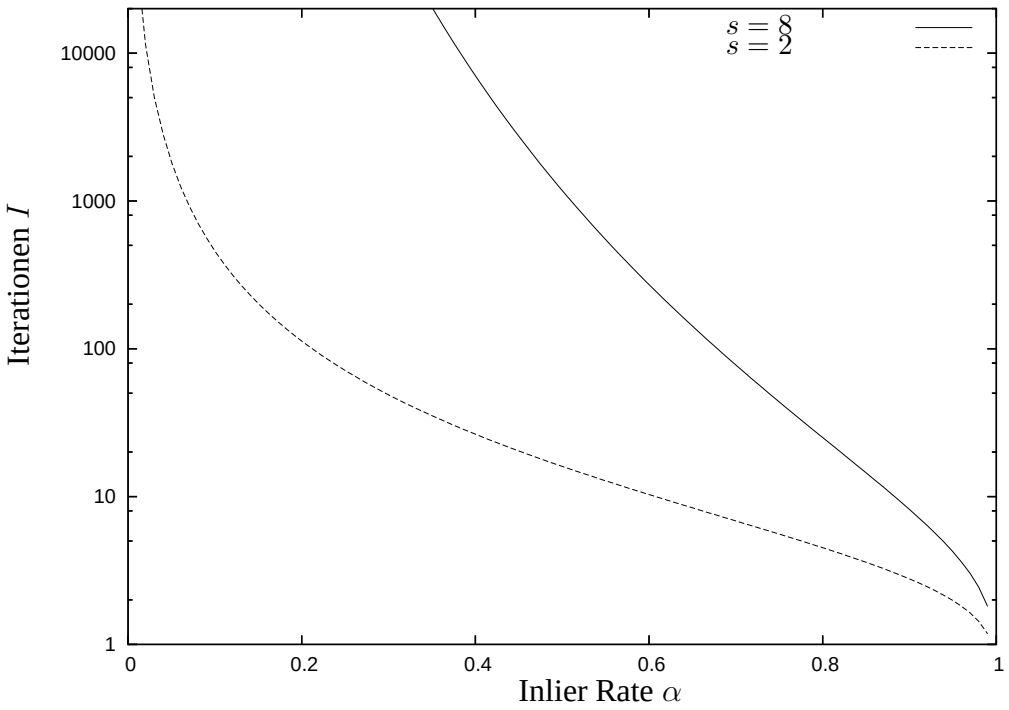


Bild 3.7: Vergleich des zu erwartenden Aufwands von 8-Punkt RANSAC mit 2-SIFT RANSAC für $p = 0,99$. Um mit einer Festen Wahrscheinlichkeit ein von $99\%$ ein Inliermodell zu generieren, werden bei nur zwei benötigten Punktkorrespondenzen wesentlich weniger Iterationen benötigt.

wurden. Sollten diese Terminierungskriterien nicht erfüllt sein, kehrt der Algorithmus mit einer Wahrscheinlichkeit von $P_q$ in den Local Exploration Schritt zurück, mit $1 - P_q$ in den Global Exploration Schritt.

Bild 3.8 verdeutlicht den Ablauf des Algorithmus.

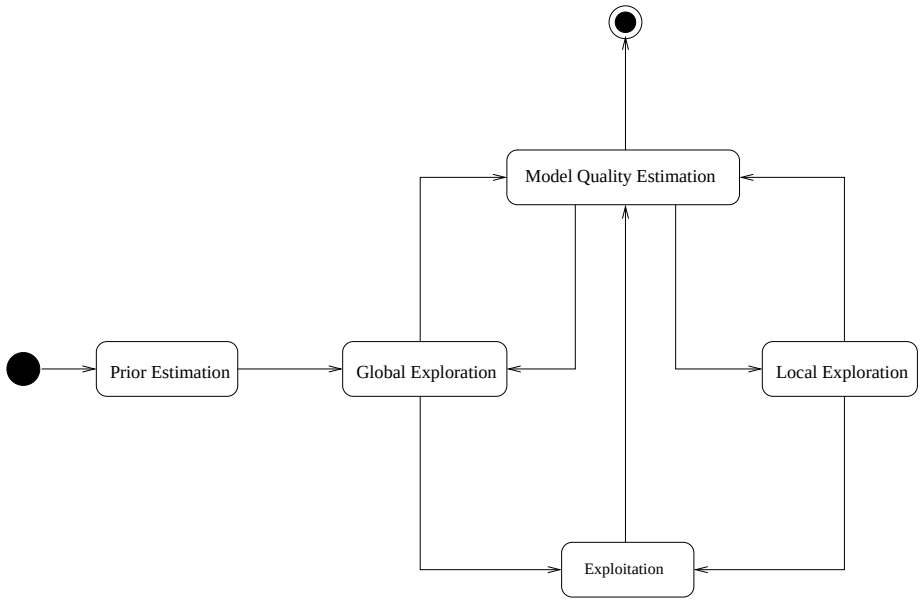


Bild 3.8: Ablauf des BEEM-Algorithmus, vgl. [Shi06].

### 3.5.8 2-SIFT-Verfahren

Das in [Shi06] beschriebene Verfahren zur Bestimmung einer Epipolargeometrie aus nur 2 SIFT-Zuordnungen nutzt die zusätzlichen Informationen der Orientierung und Skala, in welcher das SIFT-Merkmal gefunden wurde, um daraus 4 Punktmerkmale zu generieren. So werden aus 2 SIFT-Korrespondenzen 8 Punktkorrespondenzen, was ausreichend ist, um mit Hilfe des 8-Punkte-Algorithmus eine F-Matrix zu schätzen. Der sonst beim 8-Punkte-Algorithmus übliche Schritt des Erzwingens des Rangs der F-Matrix wird in diesem Fall nicht ausgeführt. Dies resultiert zwar darin, dass die entstehende Matrix keine valide F-Matrix ist, jedoch liefert das 2-SIFT-NSC (no singularity constraint) bessere Ergebnisse als das 2-SIFT-Verfahren mit anschließender Rangerzwingung. Da die geschätzte F-Matrix nur den einen Zweck hat, ein initiales Support Set an Inlier zu erkennen, ist dieser Nebeneffekt vernachlässigbar.
Die 4 Punkte werden folgendermaßen berechnet:
Es sei $S$ die Skala, in der das Merkmal detektiert wurde, $\Theta$ die Orientierung und $l = \frac{7}{8}\frac{w}{2}$, wobei $w$ die Größe des Deskriptorfensters in Pixel ist (gewöhnlicherweise $16$). Es resul-

tieren folgende Merkmalspositionen aus einem SIFT-Merkmal an der Stelle $(x, y)$:

- $(x, y)$
- $(x + l{\cdot}S{\cdot}\cos(\Theta), y + l{\cdot}S\sin(\Theta))$
- $(x + l{\cdot}S\cos(\Theta + \frac{2\pi}{3}), y + l{\cdot}S\sin(\Theta + \frac{2\pi}{3}))$
- $(x + l{\cdot}S\cos(\Theta + \frac{4\pi}{3}), y + l{\cdot}S\sin(\Theta + \frac{4\pi}{3}))$

Drei der Punkte liegen offensichtlich in regelmäßigen Abständen auf einem Kreis, der in die Größe des Deskriptorfensters eingepasst ist, der vierte liegt im Zentrum, also auf dem gleichen Punkt wie das ursprüngliche Merkmal. Bild 3.5.8 zeigt, wie aus 2 SIFT-Merkmalen 8 Punktkorrespondenzen entstehen. Die Ungenauigkeit ist bei den grünen Punkten deutlich zu sehen.

## 3.6 Kombination der Ansätze zur Schätzung der Epipolargeometrie

Die in Kapitel 3.5 vorgestellten Ansätze zur Schätzung der Epipolargeometrie werden im Folgenden zum endgültigen Algorithmus zusammengeführt. Dieser setzt sich aus zwei Hauptverfahren zusammen:

1. Der Berechnung einer Fundamental-Matrix mit maximalem Support Set,
2. Der Prüfung auf mögliche reduzierte Modelle.

Die Notwendigkeit des Betrachtens von reduzierten Modellen zusätzlich zur Schätzung der Epipolargeometrie wurde in 3.5.6 bereits motiviert.

### 3.6.1 Berechnung einer Fundamental-Matrix

Die Berechnung einer Fundamental-Matrix wird durch den leicht veränderten BEEM-Algorithmus geleistet, jedoch mit einigen kleinen Änderungen. Der originale

Bild 3.9: 8 Punktkorrespondenzen, generiert aus 2 SIFT-Merkmalen (grün und cyan), die in verschiedenen Skalen liegen.

BEEM-Algorithmus wurde unter anderem darauf hin entwickelt, mit einer sehr großen Stereobasis zurecht zu kommen. Da im Fall der Visuellen Odometrie keine generell gültige Aussage über die Stereobasis in aufeinanderfolgenden Frames gemacht werden kann, scheint vor allem der Ansatz der auf empirischen Daten beruhenden Schätzung der Inlierwahrscheinlichkeit aufgrund der Distance Ratio als nicht passend. Folgerichtig wird zwar an der inhaltlichen Aussage der Distance Ratio als Hinweis auf die Verlässlichkeit einer Zuordnung festgehalten, es wird jedoch nicht weiter zwischen dieser und der Wahrscheinlichkeit, nach der mit PROSAC im Fall der globalen Exploration Merkmalskorrespondenzen ausgewählt werden, unterschieden. Die in dem Paper [Shi06] gezeigte Graphik Fig.5(b), welche die Wahrscheinlichkeit eines Inliermodells $P_q$ in Abhängigkeit

der Anzahl der Iterationen und der Größe des Support Sets zeigt, beruht ebenfalls auf empirischen Daten. In Anbetracht der geringen Varianz der Kurven (sie unterscheiden sich in Bezug auf die Größe des Support Sets lediglich im Bereich von $1\%$-$3\%$ Inlierrate) und der generellen Kritik an der Verwendung explizit empirischer Daten ohne die Garantie vergleichbarer Szenarien, lässt sich diese Funktion auch leicht durch einen Schwellwertverfahren approximieren.

### 3.6.2 Testen mit reduzierten Modellen

Unter der Annahme, dass mit dem vorhergehenden Schritt erfolgreich eine F-Matrix mit einem maximalen Support Set $N$ gefunden wurde, können nun, wie in 3.5.6 beschrieben, Modelle für die Fälle der reinen Translation, Rotation um das optische Zentrum und Bewegungslosigkeit geprüft werden. In Anhang B.5 werden die Verfahren genauer beschrieben. Die Modelle werden mit RANSAC generiert, wobei die Zahl der Iterationen auf das Doppelte der von Gleichung 3.8 vorhergesagten Zahl gesetzt wird. Dies lässt sich darüber erklären, dass der Anteil der Inlier $\alpha$ durch die Größe des Support Sets der F-Matrix-Schätzung ermittelt wird. Er ist somit in den Fällen, in denen das reduzierte Modell relevant wäre, immer größer als der tatsächliche. Sollte mit einem der reduzierten Modelle ein Support Set von mindestens $95\%$ der Größe von $N$ erreicht werden, so wird dieses Modell als korrekt angenommen. Da die Grenze der benötigten Inlier im Voraus bekannt ist, könnte beim Testen der reduzierten Modelle ein verfrühtes Abbruchkriterium, wie in R-RANSAC beschrieben, verwendet werden. Alternativ wird solange getestet, bis der Schwellwert unterschritten wird. Für den Fall, dass mehrere reduzierte Modelle in Frage kommen, weicht das Verfahren von dem in [TZM95] beschriebenen Vorgehen ab. Dort würde aus den reduzierten Modellen dasjenige gewählt, welches das größte Support Set hat. Diese Herangehensweise provoziert allerdings den gleichen Fehler, der durch die Untersuchung der reduzierten Modellen in Bezug auf die F-Matrix vermieden werden soll. So kann bei Bewegungslosigkeit ein Modell, welches eine reine Translation schätzt, ein größeres Support Set erreichen, indem es Outlier in die Berechnung mit einfließen lässt. Dies ist besonders wahrscheinlich, wenn zur Instanziierung des Modells mehr als die minimal benötigte Menge an Korrespondenzen benutzt wird. Wir bewerten also bei mehreren reduzierten Modellen, die ein Support Set von mehr als $95\%$ der F-Matrix habenn dasje-

nige höher, welches die geringstdimensionale Bewegung beschreibt. Die Reihenfolge ist dann wie folgt:
Keine Bewegung, reine Translation, reine Rotation.

Je nachdem, um welches Modell es sich handelt, entfällt so der nächste Schritt der Extraktion der Rotation und Translation aus der F-Matrix. Im Fall von Bewegungslosigkeit wird der aktuelle Frame verworfen.

## 3.7 Zerlegung der Essential-Matrix

Um eine komplette Trajektorie der Kamerabewegung zu erzeugen, ist es notwendig, die Rotation und Translation zwischen aufeinanderfolgenden Frames zu berechnen. Diese Informationen können aus der E-Matrix gewonnen werden. Hierbei ist zu beachten, dass die Translation lediglich eine Richtung, nicht aber einen absoluten Wert liefern kann. Es muss also außerdem sichergestellt werden, dass sich der Skalierungsfaktor der Translation zwischen den unterschiedlichen Framepaaren nicht ändert, da die Trajektorie anderenfalls nicht mehr in sich konsistent wäre.

### 3.7.1 Extraktion der Essential-Matrix aus der Fundamental-Matrix

Das Ergebnis der Schätzung der Epipolargeometrie (3.5) ist im Allgemeinen die Fundamental-Matrix $\boldsymbol{F}$. Aus ihr lässt sich mit Hilfe der Kalibriermatrix $\boldsymbol{K}$ die Essential-Matrix $\boldsymbol{E}$ extrahieren:

$$\boldsymbol{E} = \boldsymbol{K}^{\mathrm{T}}\, \boldsymbol{F}\, \boldsymbol{K} \tag{3.9}$$

### 3.7.2 Aufbau der Essential-Matrix

Es sei im Folgenden $\boldsymbol{P_1}$ die Projektionsmatrix der ersten Kamera, die einen Weltpunkt in die Bildebene der ersten Kamera projiziert durch $\tilde{\boldsymbol{p}}^{\mathrm{i}} = \boldsymbol{P_1}\, \tilde{\boldsymbol{p}}^{\mathrm{w}}$. Es sei entsprechend $\boldsymbol{P_2}$ die Projektionsmatrix der zweiten Kamera.

Wir wählen $\boldsymbol{P_1}$ so, dass das Weltkoordinatensystem seinen Ursprung im Optischen Zentrum der ersten Kamera hat. Die Optische Achse liegt auf der z-Achse:

$$\boldsymbol{P_1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3.10}$$

$\boldsymbol{P_2}$ definiert sich über eine Rotation $\boldsymbol{R}$ und Translation $\boldsymbol{t}$, die auf alle Punkte angewendet werden, bevor die Projektion in die Bildebene erfolgt:

$$\boldsymbol{P_2} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \end{pmatrix} \tag{3.11}$$

Eine Multiplikation mit dieser Matrix $\tilde{\boldsymbol{p}}^{\mathrm{i}} = \boldsymbol{P_2}\,\tilde{\boldsymbol{p}}^{\mathrm{w}}$ entspricht

$$\tilde{\boldsymbol{p}}^{\mathrm{i}} = \boldsymbol{R}\,\tilde{\boldsymbol{p}}^{\mathrm{w}} + \boldsymbol{t} \tag{3.12}$$

Es wird also zuerst die Rotation, dann die Translation auf einen Weltpunkt angewendet, um ihn in das Kamerakoordinatensystem zu überführen, bevor er auf die Bildebene projiziert wird.

Die Essential-Matrix $\boldsymbol{E}$ wird im Allgemeinen als Produkt der Kreuzproduktmatrix $[\boldsymbol{t}]_\times$ (siehe Anhang B.3) des Translationsvektors $\boldsymbol{t}$ und der Rotationsmatrix $\boldsymbol{R}$ verstanden:

$$\boldsymbol{E} = [\boldsymbol{t}]_\times \boldsymbol{R} \tag{3.13}$$

Zusätzlich gilt für jedes Punktepaar $\tilde{\boldsymbol{p}}^{\mathrm{i}}, \tilde{\boldsymbol{q}}^{\mathrm{i}}$ in homogenen Koordinaten, welches eine Abbildung des selben Weltpunktes in die beiden Bildebenen ist, die Epipolare Bedingung (vgl. 3.6):

$$\tilde{\boldsymbol{q}}^{\mathrm{iT}}\,\boldsymbol{E}\,\tilde{\boldsymbol{p}}^{\mathrm{i}} = 0 \tag{3.14}$$

In praktischen Anwendungen wird lediglich ein Schwellwert $\theta$ statt $0$ gefordert:

$$|\tilde{\boldsymbol{q}}^{\mathrm{iT}}\,\boldsymbol{E}\,\tilde{\boldsymbol{p}}^{\mathrm{i}}| \leq \theta \tag{3.15}$$

Aus der Epipolaren Bedingung 3.14 folgt

$$\tilde{\boldsymbol{q}}^{\mathrm{iT}}\,a{\cdot}\boldsymbol{E}\,\tilde{\boldsymbol{p}}^{\mathrm{i}} = 0 \tag{3.16}$$

und somit nach 3.13 auch

$$\tilde{\boldsymbol{q}}^{\mathrm{iT}}\, a{\cdot}([\boldsymbol{t}]_\times\, \boldsymbol{R})\, \tilde{\boldsymbol{p}}^{\mathrm{i}} = 0 \tag{3.17}$$

Der Faktor $a$ kann nicht auf die Rotationsmatrix angewendet werden, ohne deren Eigenschaften zu verletzen ($\det(\boldsymbol{R}) = 1$). Somit muss er auf $\boldsymbol{T}$ angewendet werden. Also kann die in $\boldsymbol{E}$ enthaltene Translation $\boldsymbol{t}$ nur eine Richtung, jedoch keinen absoluten Betrag angeben. Bild 3.10 verdeutlicht dies anschaulich.

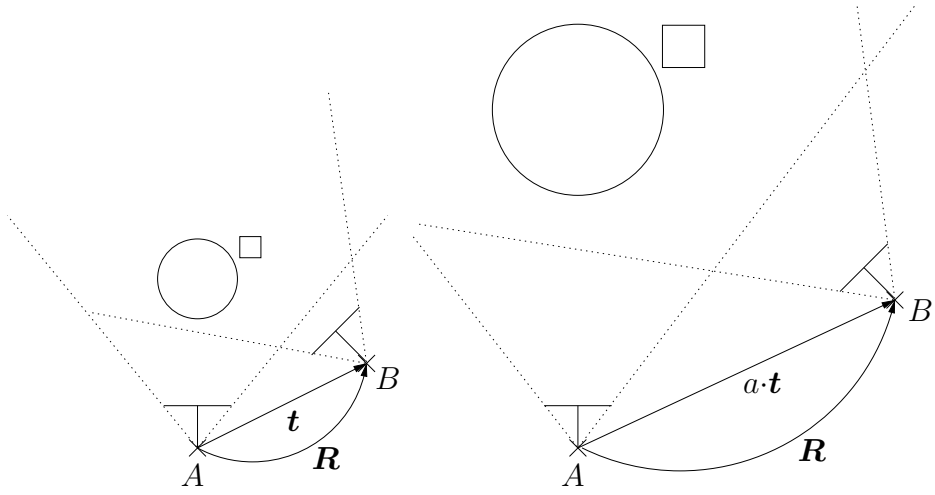


Bild 3.10: Bei bekannter Rotation und Translationsrichtung zwischen den Kameras kann ohne Wissen über die Welt keine Aussage über den absoluten Betrag der Translation gemacht werden. In beiden Beispielen (links und rechts) liefern die Kameras $A$ und $B$ jeweils identische Bilder, obwohl die Translation sich in ihrem Betrag um einen Faktor $a$ unterscheidet.

### 3.7.3 Schätzen der Rotation und Translation

Ziel ist es, eine gegebene Essential-Matrix in ihre Rotation und Translation zu zerlegen. Ein mögliches Verfahren wird in [HZ03] beschrieben. Hierzu sei $\boldsymbol{S}$ eine schiefsymmetrische Matrix, $\boldsymbol{R}$ weiterhin eine Rotationsmatrix:

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{S}\,\boldsymbol{R} \\ [\boldsymbol{t}]_\times\,\boldsymbol{R} &= \boldsymbol{S}\,\boldsymbol{R} \end{aligned} \tag{3.18}$$

Als Hilfsmatrizen definieren wir

$$\boldsymbol{W} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\boldsymbol{Z} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Sei außerdem

$$\boldsymbol{U} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \boldsymbol{V}^{\mathrm{T}} = svd(\boldsymbol{E}) \tag{3.19}$$

die Singulärwertzerlegung der normierten Essential-Matrix $\boldsymbol{E}$. Es ergeben sich folgende mögliche Zerlegungen für $\boldsymbol{E} = \boldsymbol{S}\,\boldsymbol{R}$:

$$\begin{aligned} \boldsymbol{S} &= \boldsymbol{U}\,\boldsymbol{Z}\,\boldsymbol{U}^{\mathrm{T}} \\ \boldsymbol{R_1} &= \boldsymbol{U}\,\boldsymbol{W}\,\boldsymbol{V}^{\mathrm{T}} \\ \boldsymbol{R_2} &= \boldsymbol{U}\,\boldsymbol{W}^{\mathrm{T}}\,\boldsymbol{V}^{\mathrm{T}} \end{aligned} \tag{3.20}$$

Aus Gleichung 3.13 folgt

$$\boldsymbol{t}^{\mathrm{T}}\,\boldsymbol{E} = \boldsymbol{t}^{\mathrm{T}}[\boldsymbol{t}]_\times\,\boldsymbol{R} = \boldsymbol{0} \tag{3.21}$$

Die Translation $\boldsymbol{t}$ liegt also im linken Nullraum von $\boldsymbol{E}$ und lässt sich somit alternativ zu $\boldsymbol{S}$ in 3.20 als letzter Spaltenvektor von $\boldsymbol{U}$ bestimmen.

Offensichtlich folgt aus 3.18 auch

$$\boldsymbol{E} = -\boldsymbol{S}\,-\,\boldsymbol{R} \tag{3.22}$$

als eine korrekte Zerlegung. Die Vorzeichen beider Matrizen $\boldsymbol{S}$ und $\boldsymbol{R}$ sind also durch die Zerlegung alleine nicht eindeutig bestimmbar.

Wie in 3.17 gezeigt, ist $\boldsymbol{t}$ zusätzlich nur eindeutig bis auf einen skalaren Faktor $a$, welcher auch negativ sein kann. Dies ist ein weiterer Grund, warum für die Rekonstruktion der korrekten Translationsrichtung sowohl $\boldsymbol{t}$ als auch $-\boldsymbol{t}$ in Betracht gezogen werden müssen.

### 3.7.4 Auflösen der Mehrdeutigkeiten

Zusammenfassend sind folgende Zerlegungsergebnisse als korrekte Lösungen für $\boldsymbol{R}$ und $\boldsymbol{t}$ denkbar:

$$\begin{aligned} \boldsymbol{R} &= \pm \boldsymbol{R_1} \\ \boldsymbol{R} &= \pm \boldsymbol{R_2} \\ \boldsymbol{t} &= \pm \boldsymbol{t} \end{aligned} \tag{3.23}$$

Um diese Mehrdeutigkeit aufzulösen, werden zwei Schritte durchgeführt.
Die Entscheidung über das Vorzeichen der Rotationsmatrizen $\boldsymbol{R_1}$ und $\boldsymbol{R_2}$ kann mit Hilfe der Bedingungen, die an Rotationsmatrizen gestellt werden, bewältigt werden. Nur eine der jeweils beiden Möglichkeiten $\boldsymbol{R}$ oder $-\boldsymbol{R}$ wird die für Rotationsmatrizen geforderte Determinante $\det(\boldsymbol{R}) = 1$ haben; die andere wird die Determinante $-1$ besitzen.

Die anschießende Entscheidung zwischen den 4 Kombinationsmöglichkeiten aus $\boldsymbol{R_1}$, $\boldsymbol{R_2}$ sowie $\boldsymbol{t}$ und $-\boldsymbol{t}$ kann aufgrund geometrischer Überlegungen gefällt werden (vgl. [HZ03]). Der Unterschied zwischen $\boldsymbol{t}$ und $-\boldsymbol{t}$ betrifft offensichtlich die Richtung, in welche sich die Kamera bewegt hat. Der Unterschied zwischen den möglichen Rotationen liegt in einer $180°$ Drehung der zweiten Kamera um die Grundlinie zwischen den optischen Zentren der Kameras. Bild 3.11 verdeutlicht die vier möglichen Positionen der zweiten Kamera $B$ in Bezug zur ersten Kamera $A$. Von links nach rechts ist das Vorzeichen der Translation $\boldsymbol{t}$ invertiert, von oben nach unten wird zwischen den zwei Rotationen $\boldsymbol{R_1}$ und $\boldsymbol{R_2}$ alterniert, also die 2. Kamera um $180°$ um die Grundlinie gedreht.

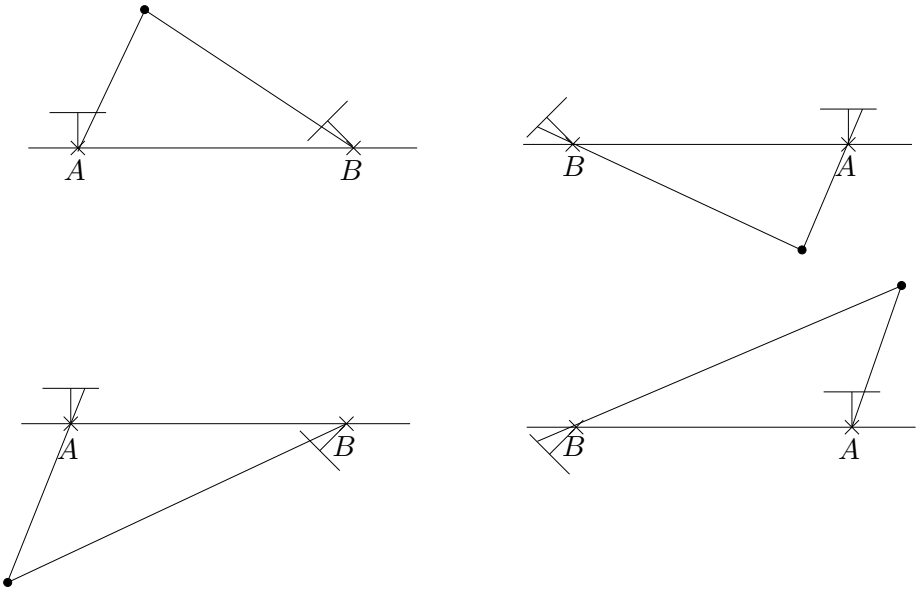


Bild 3.11: Geometrische Veranschaulichung der vier validen Faktorisierungsmöglichkeiten der Essential-Matrix.
Von links nach rechts ist das Vorzeichen der Translation $\boldsymbol{t}$ invertiert, von oben nach unten wird zwischen den zwei Rotationen $\boldsymbol{R_1}$ und $\boldsymbol{R_2}$ alterniert.

Rückprojiziert man einen Punkt aus beiden Bildern mit minimalem quadratischen Abstand zurück in die Welt (siehe Anhang B.4), so liegt er genau in einem der 4 Fälle vor beiden Kameras. In den anderen Fällen liegt er entweder hinter beiden, oder abwechselnd hinter einer und vor der anderen Kamera. So kann man durch Prüfen der z-Koordinate des rückprojizierten Punktes in beiden Kamerasystemen die einzig korrekte, weil einzig physikalisch mögliche Lösung, ermitteln. Sollte die Fundamental-Matrix falsch geschätzt worden sein, so kann es vorkommen, dass sich nicht diese vier Lösungen ergeben, oder dass andere korrespondierende Punkte anschließend hinter den Kameras rekonstruiert werden.

Das Ergebnis ist die Projektionsmatrix $\boldsymbol{P_2} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \end{pmatrix}$ der Kamera des zweiten Frames, unter der Annahme, dass $\boldsymbol{P_1} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \end{pmatrix}$ ist, also im Bezug auf das Koordinatensystem des vorangehenden Frames. Zur Einreihung in die globale Trajektorie ist jetzt noch lediglich eine Umrechnung in Weltkoordinaten und eine Anpassung der Skalierung nötig.

### 3.7.5 Einreihung in eine globale Trajektorie

Gegeben ${}^{\mathrm{i+1}}\boldsymbol{R}_{\mathrm{i}}$ und ${}^{\mathrm{i+1}}\boldsymbol{t}_{\mathrm{i}}$, die Rotation und Translation aus der Zerlegung der Essential-Matrix des Frames $i+1$. Es handelt sich also um die Transformationen aus der Projektionsmatrix von Frame $i+1$ in Bezug auf auf Frame $i$. Weiterhin gegeben sind die Projektionsmatrizen $\boldsymbol{P}_{\mathrm{wcs}}..\boldsymbol{P}_{\mathrm{i}}$ aller vorhergehender Frames. Gesucht sind ${}^{\mathrm{i+1}}\boldsymbol{R}_{\mathrm{wcs}}$ und ${}^{\mathrm{i+1}}\boldsymbol{t}_{\mathrm{wcs}}$, welche durch $\boldsymbol{P}_{\mathrm{i+1}} = \begin{pmatrix} {}^{\mathrm{i+1}}\boldsymbol{R}_{\mathrm{wcs}} & {}^{\mathrm{i+1}}\boldsymbol{t}_{\mathrm{wcs}} \end{pmatrix}$ die Projektionsmatrix des Frames $i+1$ in Bezug auf das Weltkoordinatensystem liefern. Es gilt folgender Zusammenhang:

$$\begin{aligned} {}^{\mathrm{i+1}}\boldsymbol{R}_{\mathrm{wcs}} &= {}^{\mathrm{i+1}}\boldsymbol{R}_{\mathrm{i}}\, {}^{\mathrm{i}}\boldsymbol{R}_{\mathrm{wcs}} \\ {}^{\mathrm{i+1}}\boldsymbol{t}_{\mathrm{wcs}} &= {}^{\mathrm{i+1}}\boldsymbol{t}_{\mathrm{i}} + {}^{\mathrm{i+1}}\boldsymbol{R}_{\mathrm{i}}\, {}^{\mathrm{i}}\boldsymbol{t}_{\mathrm{wcs}} \end{aligned} \tag{3.24}$$

Um die Position und Orientierung einer Kamera in Weltkoordinaten, also in Relation zum ersten Frame, zu erhalten genügt es, die Transponierte der Rotation und das Negative der Translation ihrer Projektionsmatrix zu betrachten.

### 3.7.6 Korrektes Skalieren

Da die Translation zwischen zwei aufeinanderfolgenden Frames nur bis auf einen Faktor genau bestimmt werden kann (3.17), muss Sorge dafür getragen werden, dass dieser Faktor zwischen allen Framepaaren zu äquivalenten Rückprojektionen führt. So wird gewährleistet, dass die rekonstruierte Bewegung sich immer im gleichen Maßstab befinden. Das bedeutet unter anderem, dass der Abstand der rückprojizierten Weltpunkte $\boldsymbol{p}_{\boldsymbol{i}}^{\mathrm{w}}$ und $\boldsymbol{p}_{\boldsymbol{i+1}}^{\mathrm{w}}$ aus den Framepaaren $(F^{t-2}, F^{t-1})$ sowie $(F^{t-1}, F^{t})$ gleich sein muss. Zwischen der Translation der Kamera von Frame $F^{t-1}$ nach $F^{t}$, und dem Abstand der aus Merkmalen dieser beiden Frames triangulierten Weltpunkte, besteht ein direkter linearer Zusammenhang. Um also die korrekte Skalierung der Translation in den aktuellen Frame $F^{t}$ zu gewährleisten, muss sichergestellt werden, dass der Abstand der rückprojizierten Weltpunkte zu dem des vorhergehenden Framepaares identisch ist. Um ein stabiles Maß für die Skalierung zu erreichen, wird der Median aller Abstände $|\boldsymbol{p}_{\boldsymbol{i}}^{\mathrm{w}} - \boldsymbol{p}_{\boldsymbol{i+1}}^{\mathrm{w}}|$ als Maß für die Skalierung der Translation gewählt.

# Kapitel 4

# Experimente und Ergebnisse

In diesem Kapitel wird zuerst eine Übersicht über die Implementation des Verfahrens gegeben. Im Anschluss werden ausgesuchte Testszenarien vorgestellt, und zur Evaluation des 2-SIFT Verfahrens, des Testens auf reduzierte Modelle und der erreichten Genauigkeit des Gesamtalgorithmus verwendet. Zum Schluss werden die Grenzen des Verfahrens aufgezeigt.

## 4.1 Implementation

### 4.1.1 Octave Prototyp

Im Vorfeld wurde ein Prototyp in *Octave* implementiert, der die Funktionalität vom Schätzen der Epipolargeometrie bei gegebenen Inliern über die Zerlegung der Essential-Matrix bis zur Einreihung in die globale Trajektorie umfasst. Hauptzweck war die Validierung mathematischer Methoden wie der Zerlegung der Essential-Matrix und die Möglichkeit, schnell neue Ideen testen zu können. In Anhang C.1 werden die implementierten Methoden genauer beschrieben.

### 4.1.2 Hauptprogramm

Die Hauptfunktionalität wurde in *C++* implementiert. Aufbauend auf eine in der Arbeitsgruppe vorhandenen Implementation wurde eine Bibliothek für den Kalibriervorgang erstellt. Eine weitere, davon unabhängige Bibliothek enthält die Methoden zur bildbasierten Bewegungsschätzung anhand prägnanter Merkmale sowie weiter für die Evaluation benötigte Algorithmen. Für den SIFT-Merkmalsdetektor wurde die Originalklasse von David G. Lowe verwendet. Mit Hilfe dieser Bibliotheken wurden einige Programme erstellt, die die Kamerakalibrierung, Entzerrung, Berechnung und Visualisierung einer Trajektorie sowie die Visualisierung von SIFT-Merkmalskorrespondenzen zwischen zwei Bildern leisten. Informationen über die Programme, Systemvoraussetzungen und verwendete Bibliotheken finden sich in Anhang C.2.

## 4.2 Versuchsaufbau und Durchführung

### 4.2.1 Einleitung

Zur Untersuchung der verwendeten Algorithmen wurden mehrere Kamerafahrten durchgeführt, um Testbildfolgen zu generieren (Tabelle 4.1). Alle verwendeten Kameras wurden im Vorfeld kalibriert (siehe Abschnitt 4.2.2).

Beschreibung der Szenen:

**Tasse:** Die Kamera bewegt sich im Uhrzeigersinn spiralförmig auf die Tasse zu. Am Ende bewegt sie sich fast linear von der Tasse weg. Die Szene ist vergleichbar mit Anwendungsfällen in der Augmented Reality, in denen ein Benutzer mit HMD o. ä. sich einem relevanten Objekt nähert und es von mehreren Seiten begutachtet. Eine Poseschätzung wäre in diesem Fall nötig, um dem Benutzer zusätzliche Informationen zu dem Objekt stabil in das Bild des HMD einblenden zu können. Eine Schwierigkeit in der Szene sind die Glanzpunkte auf der Oberfläche der Tasse, die zu Fehlzuordnungen führen können (Bild 4.1).

**Labor:** Für 3 Frames steht die Kamera still, anschließend vollführt sie eine hauptsächlich

| Szene | Kamera | Objektiv | Aufnahme Modus | Frames |
|---|---|---|---|---|
| Tasse | Imaging Source DFK-31BF03 | Computar $3.6mm, 1:1.6$ | $1024 \times 768$ mono | 8 |
| Labor | Imaging Source DFK-31BF03 | Pentax $8.5mm, 1:1.5$ | $1024 \times 768$ mono | 10 |
| Neon& Chrome | *Maya* Renderer | - | $1024 \times 768$ mono | 5 |

Tabelle 4.1: Übersicht über die Testbildfolgen.

nach vorne gerichtete Bewegung durch einen Raum. Am Ende dreht sie nach links ab. Die Szene ist vergleichbar mit den Kamerainformationen, die ein mobiler Roboter erhält, während er sich bewegt. Hier kann die Rekonstruktion der Trajektorie bei der SLAM-Problematik hilfreich sein. Eine Hauptschwierigkeit der Szene stellt die Aufnahmemodalität dar. Durch die große Brennweite des Objektivs ist der abgebildete Ausschnitt der Welt entsprechend klein, wodurch das Finden von nicht degenerierten Merkmalsmengen erschwert wird. Zusätzlich bewirken Drehungen um die x- und y-Achsen stärkere Änderungen im Bild (Bild 4.2).

**Neon&Chrome:** Die Szene[1] wurde mit *Maya* gerendert, daher sind exakte Ground Truth Daten verfügbar. Die Kamera bewegt sich für 3 Frames um das Auto, dann steigt sie senkrecht nach oben. Die Schwierigkeiten in der Szene sind sowohl die Glanzpunkte auf dem Auto als auch die häufig vorkommenden sich exakt wiederholenden Strukturen auf dem Gebäude und am Auto (die Lichterkette, Wandziegel, die Löcher im Auspuffrohr. Bild 4.3).

[1]Maya-Modell von Christophe Desse und Matthew Thain, im Rahmen einer 'Lighting Challenge' auf http://www.3drender.com/challenges/index.htm (Stand 29.08.2007)

Bild 4.1: Testbildfolge ’Tasse’.

### 4.2.2 Kamerakalibrierung und Entzerrung

Zur Bestimmung der intrinsischen Kameraparameter wurde ein Schachbrettmuster mit bekannter Geometrie aus verschiedenen Ansichten aufgenommen. Die Ecken zwischen den Schachbrettfeldern dienen dabei als leicht zu detektierende Punkte, deren Lage in der Welt zueinander bekannt ist. Bild 4.4 zeigt exemplarische Kalibrierbilder vor und nach der Entzerrung.

Bild 4.2: Testbildfolge ’Labor’.

### 4.2.3 Bildbasierte Bewegungsschätzung

Mit Hilfe der entstandenen entzerrten Bildfolgen wurde versucht, die Kamerabewegung zu rekonstruieren. Das Ergebnis sind Daten über die Rotation und Translation zwischen

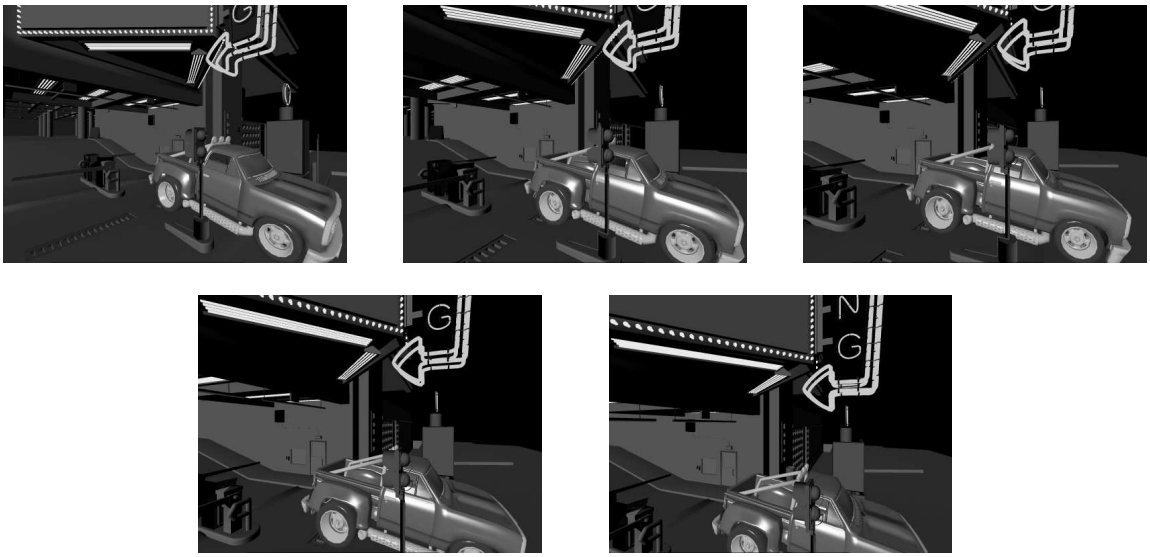

Bild 4.3: Testbildfolge ’Neon&Chrome’.

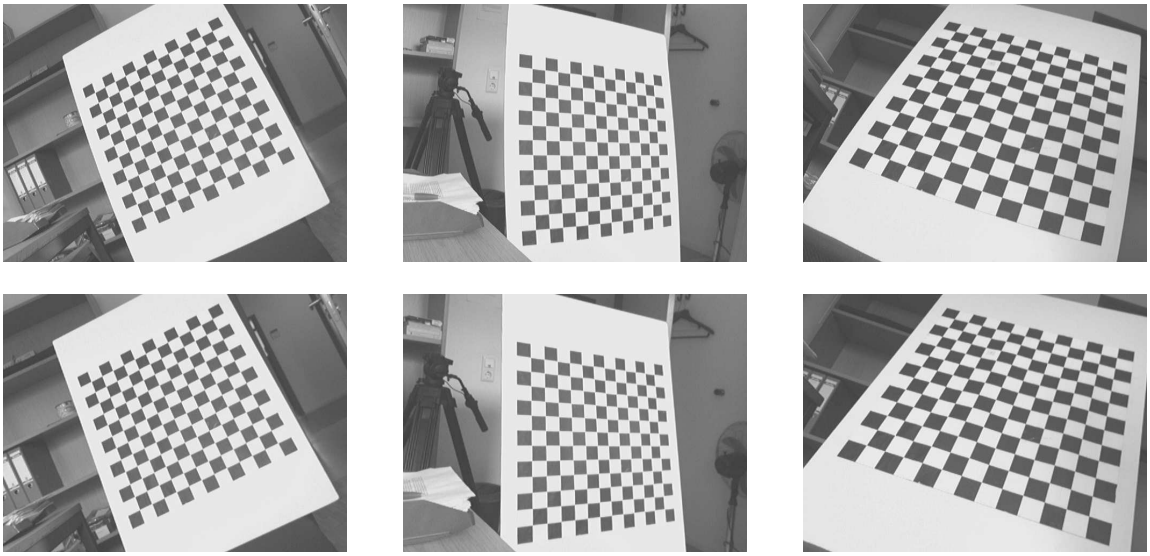

Bild 4.4: Aufnahmen des Kalibriermusters (oben) und Ergebnisse nach der Entzerrung (unten).

den aufeinanderfolgenden Frames, sowie relativ zu einem gemeinsamen Weltkoordinaten-

system. Diese wurden in *OpenGL* visualisiert (siehe Bild 4.5).

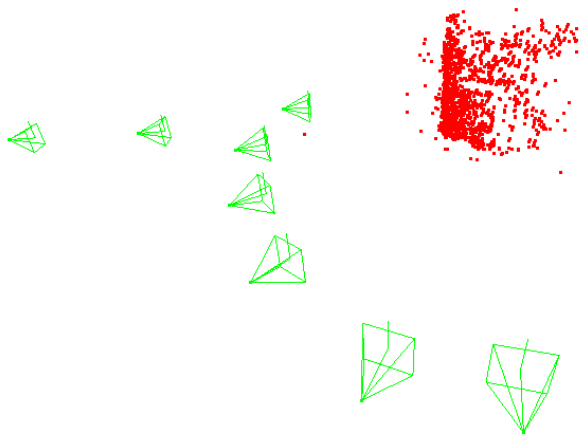

Bild 4.5: Rekonstruierte Trajektorie der Kamerafahrt ’Tasse’. Grün sind die Kamerapositionen der einzelnen Frames zu sehen, rot die rekonstruierten 3-D-Punkte.

## 4.3 Ergebnisse

### 4.3.1 2-SIFT Verfahren

Bei dem verwendeten 2-SIFT Verfahren ist bekannt, dass die Genauigkeit zu Gunsten des potentiellen Geschwindigkeitsgewinns stark vernachlässigt wird. Es stellt sich die Frage, ob das Verfahren trotzdem eine ausreichende Menge an Inliern als solche erkennt, und ob die auftretende Ungenauigkeit gegebenenfalls durch Maßnahmen wie die Lockerung des Schwellwertes $\theta$ der Epipolaren Bedingung ausgeglichen werden kann. Hierbei muss jedoch auch die Gefahr beachtet werden, Outlier als *falsch positiv* zu beurteilen. Folgerichtig wurde untersucht, in wie weit Modelle, die auf diese Weise erzeugt werden, Inlier und Outlier als passende Zuordnungen bewerten. Dazu wurde auf den Testbildfolgen ’Tasse’ und ’Labor’ für jedes Framepaar jeweils 200 mal mit dem 2-SIFT Verfahren

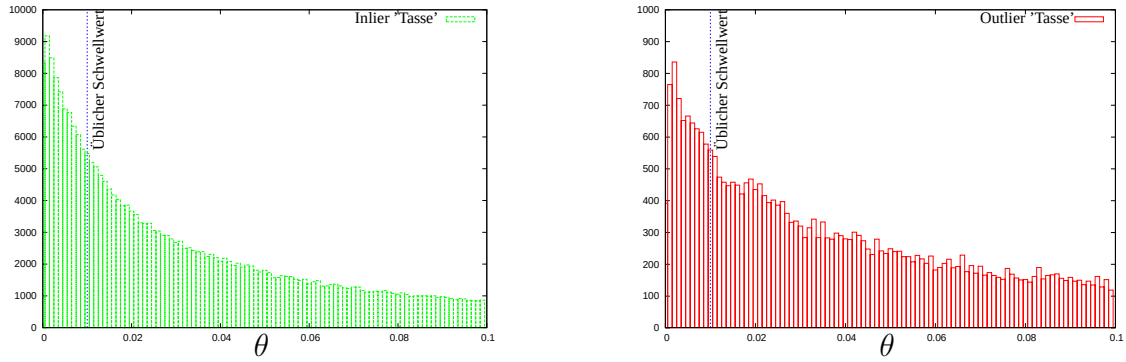


Bild 4.6: Histogramme für die Ergebnisse der Epipolaren Bedingung bei der Testbildfolge 'Tasse' unter Verwendung des 2-SIFT-Verfahrens.

eine F-Matrix erzeugt. Die Auswahl der dazu verwendeten Merkmale geschah zufällig aus der Menge der vom BEEM-Algorithmus entdeckten Inlier. Schließlich wurden alle potentiellen Punktezuordnungen (Inlier sowie Outlier) auf die Epipolare Bedingung (siehe 3.6) hin untersucht. Es wurden Histogramme für den Schwellwert $\theta$ erstellt (siehe Bild 4.6 und Bild 4.7).

Wie erhofft übersteigt die Zahl der Inlier mit einem guten Ergebnis $\theta$ die der Outlier. Allerdings scheint das Verhältnis stark szenenabhängig zu sein. So würden im Fall 'Labor' fast $17\%$ der Inlier und nur $3,5\%$ der Outlier die Epipolare Bedingung mit dem gängigen Schwellwert $\theta = 0,01$ erfüllen. In der Szene 'Tasse' hingegen nähern sich die Werte auf $17,7\%$ zu $8,3\%$ an. Hier besteht also eine erhöhte Gefahr, dass Outlier zufällig mit einem 2-SIFT-Modell konsistent sind. Betrachtet man die Möglichkeit einer Verbesserung durch eine Lockerung des Schwellwerts $\theta$, so wird leider klar, dass der Verlauf der Histogramme für Inlier und Outlier sich zu stark ähnelt, und somit das Verhältnis nicht verbessert werden kann.

### 4.3.2 Epipolare Schätzung bei degenerierten Konfigurationen

Das folgende Experiment zeigt mögliche Auswirkungen, wenn auf den Schritt des Testens der reduzierten Modelle verzichtet wird. Als Grundlage dient hier die Szene 'Labor' wie in

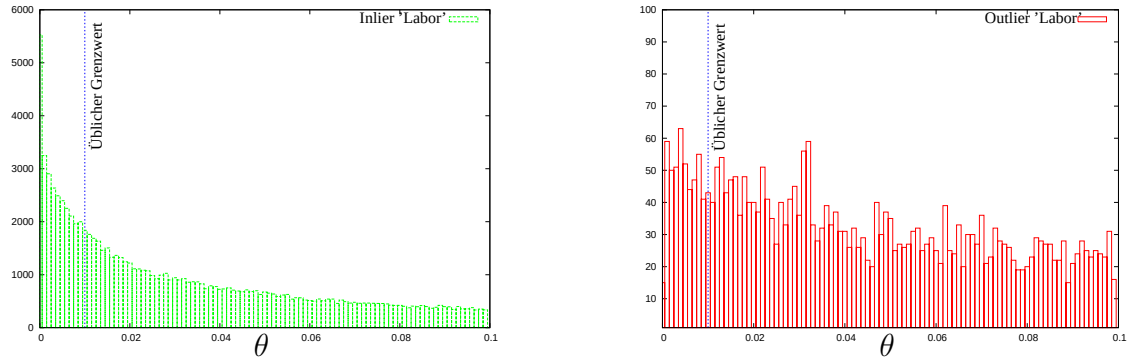


Bild 4.7: Histogramme für die Ergebnisse der Epipolaren Bedingung bei der Testbildfolge 'Labor' unter Verwendung des 2-SIFT-Verfahrens.

| Szene | $\theta \leq 0.01$ | $\theta \leq 0.02$ | $\theta \leq 0.03$ |
|---|---|---|---|
| Labor Inlier | 16,927% | 26,240% | 32,652% |
| Labor Outlier | 3,507% | 6,816% | 9,684% |
| Tasse Inlier | 17,777% | 28,837% | 36,577% |
| Tasse Outlier | 8,348% | 14,427% | 19,454% |

Tabelle 4.2: Ergebnisse der Epipolaren Bedingung beim 2-SIFT-Verfahren. Das Verhältnis von Inliern zu Outliern kann durch eine Lockerung des Schwellwertes $\theta$ nicht verbessert werden.

4.2.1 beschrieben. In der Szene gibt es zwei relevante Stellen, an welchen das Testen von reduzierten Modellen zu einer offensichtlichen Verbesserung der Ergebnisse führt. Wird beim annähernden Stillstand der Kamera während der ersten drei Frames der Kamerafahrt nicht auf Bewegungslosigkeit getestet, so besteht die Gefahr, eine ungeeignete Skala für die Repräsentation in Weltkoordinaten zu wählen. Dies führt im Endeffekt zu größeren Fehlern bei der Schätzung der Pose, sowie der Rekonstruktion von 3-D-Punkten (siehe Bild 4.8). Während der Vorwärtsbewegung der Kamera kann es ebenfalls zu großen Fehleinschätzungen bezüglich der Translation kommen, wenn das entsprechende reduzierte Modell für reine Translation nicht betrachtet wird. In Bild 4.9 sieht man ein mögliches Resultat. Werden hingegen die Unbewegtheit der Kamera sowie die reine Translation als solche erkannt, ist das Ergebnis weitaus plausibler. Die rekonstruierte Trajektorie beschreibt

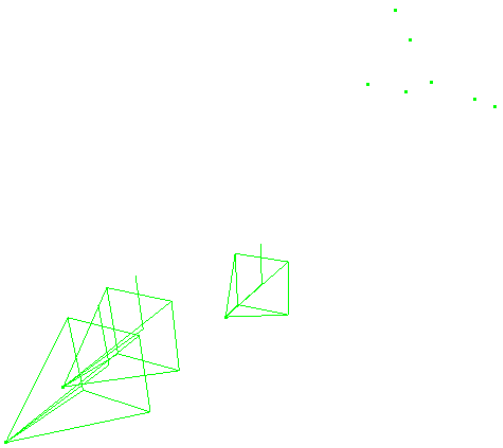

Bild 4.8: Ergebnistrajektorie der Szene ’Labor’ ohne Tests auf reduzierte Modelle.

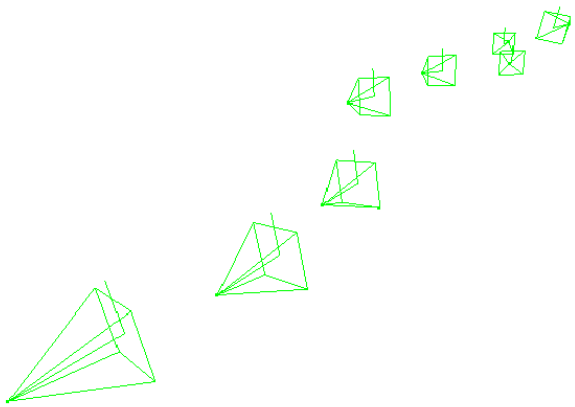

Bild 4.9: Ergebnistrajektorie der Szene ’Labor’ unter Auslassung des Tests auf reine Translation.

den Weg der Kamera: Bild 4.10.

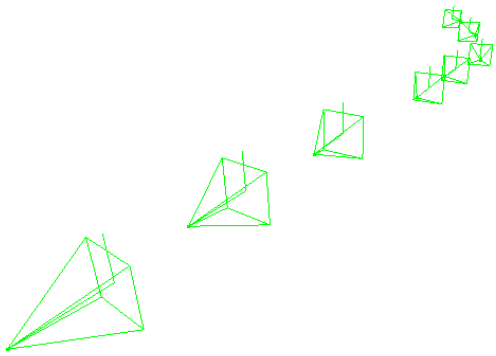

Bild 4.10: Ergebnistrajektorie der Szene ’Labor’ unter Miteinbeziehung von reduzierten Modellen.

### 4.3.3 Genauigkeit des Verfahrens

Um eine Aussage über die Genauigkeit des Verfahrens machen zu können, wurde der Algorithmus auf der Szene ’Neon&Chrome’ evaluiert, da nur hier exakte Ground Truth Daten vorliegen. Auch hier wurde im Vorfeld eine Kamerakalibrierung auf synthetisch erzeugten Bilddaten durchgeführt. Für eine Auflösung von $1024 \times 768$ ergab sich folgende $\boldsymbol{K}$-Matrix:

$$\boldsymbol{K} = \begin{pmatrix} 512.1085815430 & 0.0000000000 & 511.5093688965 \\ 0.0000000000 & -512.0780639648 & 383.3175354004 \\ 0.0000000000 & 0.0000000000 & 1.0000000000 \end{pmatrix} \tag{4.1}$$

Gleichung 4.1 zeigt deutlich einen gewissen Kalibrierungsfehler, der selbst bei perfekten Kameras wie der des *Maya* Renderers nicht auszuschließen ist. Die Position des Hauptpunktes ($\boldsymbol{K}_{1:2,3}$) weicht fast um einen halben Pixel vom korrekten Wert $(512, 384)$ ab, und auch die Relation der Brennweite zur Sensorgröße ($\boldsymbol{K}_{1,1}$, $\boldsymbol{K}_{2,2}$) zeigt eine kleine Abwei-

chung vom erwarteten Wert $(512, -512)$.

Für das eigentliche Experiment wurde die Trajektorie der Szene je 100 mal mit verschiedenen Verfahren rekonstruiert. Dabei wurden der BEEM, LO-RANSAC sowie RANSAC Algorithmus verwendet, jeweils mit und ohne Testen auf reduzierte Modelle. Die Anzahl der Iterationen von LO-RANSAC und RANSAC richtete sich nach der Menge, die der BEEM-Algorithmus gebraucht hatte. Als Iteration wurde hier jede Erstellung eines Modells gewertet. Dies hat unter anderem zur Folge, dass der BEEM-Algorithmus im Exploitation-Schritt minimal zehn Iterationen benötigt. Als Gütemaß für die Genauigkeit der rekonstruierten Trajektorie wurden die Abweichungen zur Ground Truth in allen Frames (außer dem ersten, da dieser lediglich den Startpunkt vorgibt und somit nie eine Abweichung hat) betrachtet. Einen absoluten Wert wie den Abstand der geschätzten Kameraposition zur korrekten Kameraposition anzugeben, macht hier keinen Sinn. Dieser ist nämlich direkt vom beliebig wählbaren Skalierungsfaktor, der zwischen den Frames der Welt bzw. der Kamera herrscht, abhängig. Daher wurden hierbei stattdessen die drei folgende Maße als Indiz für die Genauigkeit der Verfahren verwendet:

**Translationsrichtung ($\measuredangle_t$):** Die Translationsrichtung gibt den Winkel an, in welchem sich die Kamera zum nächsten Frame hin bewegt hat. Hier wurde der Winkel in Grad zwischen der korrekten und der geschätzten Bewegungsrichtung als Gütemaß verwendet (vgl. Bild 4.11).

**Rotationsachse ($\measuredangle_{\text{R Axis}}$):** Die Rotation, der die Kamera zwischen zwei aufeinanderfolgenden Frames unterliegt, lässt sich in der Axis-Angle-Repräsentation als Rotationsachse und Drehwinkel darstellen (siehe Anhang B.1). Der Winkel zwischen den Rotationsachsen der korrekten und der geschätzten Rotation dient als weiteres Gütemaß.

**Rotationswinkel ($\measuredangle_{\text{R Angle}}$):** Der Rotationswinkel beschreibt, wieweit die Kamera um die Rotationsachse gedreht wurde. Die Differenz des korrekten und des geschätzten Winkels in Grad bildet das letzte Gütemaß (vgl. Bild 4.12).

Von allen drei Gütemaßen wurden jeweils der Mittelwert $\mu$ und die Standardabweichung $\sigma$ der 100 Experimente pro Verfahren ermittelt. Die Ergebnisse sind in Tabelle 4.3 zu sehen.

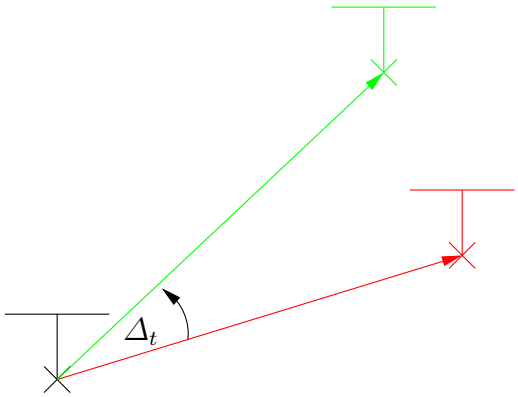


Bild 4.11: Fehler beim Schätzen der Translationsrichtung.

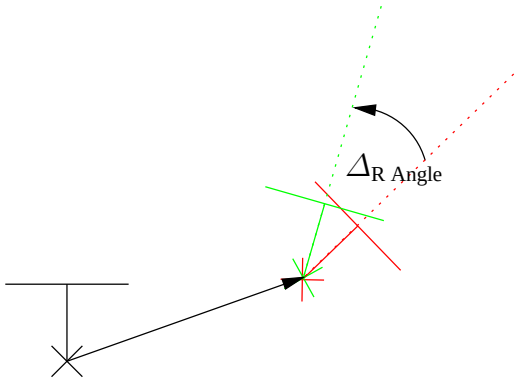


Bild 4.12: Fehler beim Schätzen des Rotationswinkels.

Wie aus Tabelle 4.3 ersichtlich, ist die Schätzung der Translationsrichtung durch den Algorithmus mit bis zu $6^\circ$ durchschnittlichem Fehler der ungenaueste Teil. Die Bestimmung der Rotationsachse und besonders des Rotationswinkels hingegen liefern weitaus genauere Ergebnisse. Dies entspricht auch den subjektiven Beobachtungen, die während zahlreicher Experimente gemacht wurden: während es bei der Schätzung der Translationsrichtung hin und wieder zu Ungenauigkeiten und Ausreißern kommt, ist die Schätzung der Rotation unabhängig davon sehr stabil. Ein Grund für die zunehmenden Fehler bei der Translationsrichtung kann auch aus der Wahl der Gütemaße resultieren: Während Fehler in der Schätzung der Rotation und Translation sich auf die Translationen

| Verfahren | $\mu(\Delta_t)$ | $\sigma(\Delta_t)$ | $\mu(\Delta_{\text{R Axis}})$ | $\sigma(\Delta_{\text{R Axis}})$ | $\mu(\Delta_{\text{R Angle}})$ | $\sigma(\Delta_{\text{R Angle}})$ |
|---|---|---|---|---|---|---|
| BEEM | $6,1351$ | $\mathbf{3,5150}$ | $\mathbf{2,2093}$ | $\mathbf{1,7568}$ | $0,68897$ | $\mathbf{0,65182}$ |
| LO-RANSAC | $5,6275$ | $5,7025$ | $2,2296$ | $3,0119$ | $0,84074$ | $0,99635$ |
| RANSAC | $\mathbf{5,6125}$ | $5,5929$ | $2,2432$ | $3,0316$ | $\mathbf{0,64953}$ | $0,78330$ |
| BEEM | $19,334$ | $15,378$ | $12,102$ | $11,307$ | $1,6129$ | $1,7704$ |
| LO-RANSAC | $19,429$ | $16,036$ | $12,171$ | $11,546$ | $1,7336$ | $1,9657$ |
| RANSAC | $19,343$ | $15,649$ | $12,073$ | $11,402$ | $1,6457$ | $1,8367$ |

Tabelle 4.3: Abweichungen von der Ground Truth bei der Szene 'Neon&Chrome'. Bei dem oberen Tripel wurde im Anschluss an das Schätzen der Epipolargeometrie ein Testen auf reduzierte Modelle durchgeführt, beim unteren Tripel wurde dieser Schritt ausgelassen.

der folgenden Frames relativ zum Ursprung auswirken (Bild 4.13), bewirken Fehler in der Translation keine Folgefehler für die Rotationsschätzung (Bild 4.14).

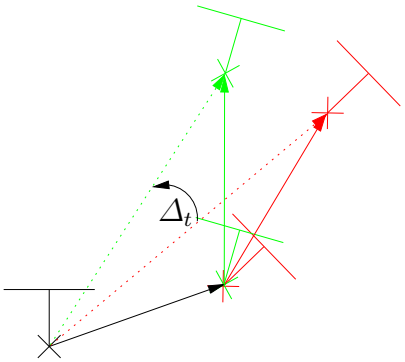


Bild 4.13: Fehler in der Rotationsschätzung beeinflussen Translationsfehler für die folgenden Frames. Obwohl die Translation stehts korrekt geschätzt wurde, verursacht der Fehler in der Rotationsschätzung einen Fehler bei der Bewertung der Translation in den nachfolgenden Frames ($\Delta_t$).

Vergleicht man die Ergebnisse des BEEM-Algorithmus mit denen der RANSAC-Derivate

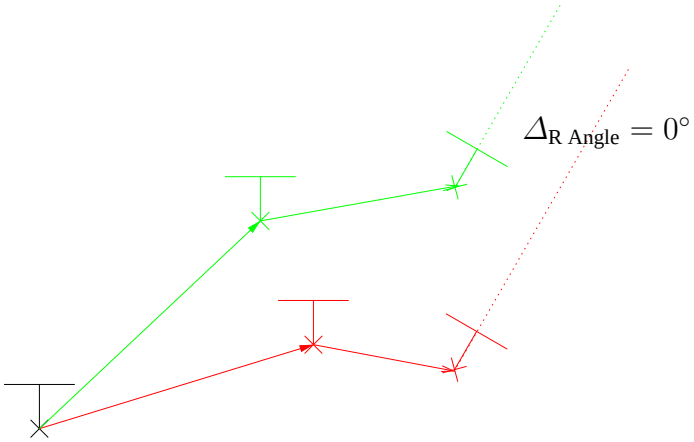


Bild 4.14: Fehler in der Translationsschätzung wirken sich nicht auf die Rotationsschätzung für folgende Frames aus. Trotz wiederholt falscher Translationsschätzungen (rot) ist die Blickrichtung der Kamera unbeeinflusst, es gibt also keine Auswirkungen auf den Rotationsfehler ($\Delta_{\text{R Angle}}$).

so fällt auf, dass die Ergebnisse von BEEM nicht deutlich besser in Bezug auf die Mittelwerte der Fehler sind. Die Unterschiede hier sind im Gegenteil eher gering. Dies kann mit der teils überdurchschnittlichen Zahl an Iterationen zu tun haben, die der BEEM-Algorithmus für die vorliegenden Frames benötigte. Ein Hauptgrund hierfür ist die in Abschnitt 4.3.1 untersuchte Ungenauigkeit der 2-SIFT-Methode. Teils verbringt der BEEM-Algorithmus bis zu 1000 Iterationen mit der globalen Suche, welche sich nicht direkt auf die Qualität des resultierenden Modells auswirkt. Durch diese hohe Zahl an Iterationen wird die Wahrscheinlichkeit, durch die RANSAC-Derivate ein gutes Modell zu finden, stark erhöht. Zusätzlich ist aufgrund der strengen Restriktionen beim Zuordnen der Merkmale nur eine sehr geringer Anteil an Outliern vorhanden, was den zusätzlichen Aufwand von BEEM nicht rechtfertigt.
Es ist jedoch ebenso offensichtlich, dass die von BEEM erzielten Ergebnisse, die eine konsequent deutlich geringeren Standardabweichung aufweisen, weitaus stabiler sind. Der Algorithmus kann somit als robuster bezeichnet werden. Außerdem besitzt er gegenüber den RANSAC-Derivaten den inhärenten Vorteil der intelligenten Abbruchbedingung. Somit ist er besser geeignet für unbekannte Szenen und passt den Aufwand besser an die

Umstände an.
In den unteren drei Zeilen von Tabelle 4.3 sieht man die Ergebnisse von Abschnitt 4.3.2 nochmals bestätigt: Ohne eine anschließende Prüfung auf reduzierte Modelle sind die Ergebnisse weitaus schlechter und im Bereich des Unbrauchbaren anzusiedeln. Der Schritt des Testens auf reduzierte Modelle ist also nach erfolgter Schätzung der Epipolargeometrie auf Basis des Normalisierten 8-Punkte-Algorithmus unabdingbar. Besonders in Anwendungsfällen aus dem Bereich der Robotik, in denen es oft zu Stillständen oder reinen Translationen zwischen aufeinanderfolgenden Frames kommen kann, wäre eine Auslassung dieses Schrittes fatal.

### 4.3.4 Laufzeit

Im Folgenden soll an zwei verschiedenen Beispielen die Laufzeit des Verfahrens demonstriert werden. Es handelt sich um die Zeit für das Hinzufügen des zweiten und dritten Frames aus der Testbildfolge ’Tasse’.

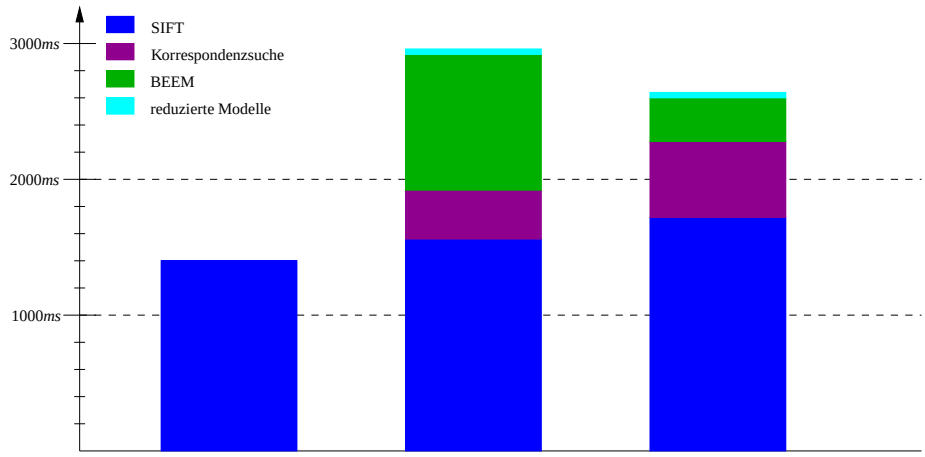


Bild 4.15: Laufzeitbetrachtung der ersten drei Frames der Testbildfolge ’Tasse’.

In Bild 4.15 sind die Anteile der unterschiedlichen Schritte des Gesamtverfahrens aufgezeigt. Die Laufzeit der Mekmalsdetektion durch den SIFT-Algorithmus steigt mit der Zahl der gefundenen Merkmale. Erwartungsgemäß führt dies ebenfalls zu einer Steigerung der

Zeit, die für die Korrespondenzsuche benötigt wird. Die Schätzung der Epipolargeometrie durch BEEM ist jedoch weniger abhängig davon, wie viele potentielle Zuordnungen als Eingabe vorliegen. Sie hängt hauptsächlich von der Zahl der benötigten Iterationen ab. Diese richtet sich danach, wie schnell ein passendes Modell in der Global Exploration gefunden wurde und wie oft dieses anschließend verfeinert wird. Die abschließende Suche nach reduzierten Modellen benötigt nur einen verschwindent geringen Bruchteil der gesamten Zeit.

### 4.3.5 Grenzen des Verfahrens

Viele Grenzen des Verfahrens liegen in den Grenzen der verwendeten Algorithmen begründet. So stellt der Schritt der Schätzung der Epipolargeometrie auf Basis des 8-Punkte-Algorithmus gewisse Anforderungen an die Szene. Optische Phänomene, die im Bildentstehungsmodell nicht berücksichtigt wurden, können zu einer kompletten Fehlschätzung der Epipolargeometrie führen. Beispiele hierfür sind Glanzpunkte und reflektierende Flächen. Eine weitere Beschränkung, die der 8-Punkte-Algorithmus mit sich bringt, ist die benötigte Verteilung der Merkmale im Bild. Liegen diese auf einer Ebene in der Welt oder sammeln sich ausschließlich in einem kleinen Bildausschnitt, so kann die Rekonstruktion der Kamerapositionen ebenfalls große Fehler aufweisen. Zusätzlich kann es vorkommen, dass Bilder mit geringer Auflösung szenenabhängig einfach zu wenige Merkmale enthalten, die korrekt zugeordnet werden können.
Das verwendete 2-SIFT Verfahren erreicht zwar meist sehr schnell ein ausreichendes Support Set, um die Suche lokal weiter zu führen, in einigen Fällen jedoch konnten immense Probleme beobachtet werden. So wurden teilweise um den Faktor $10$ bis $1000$ mal mehr Iterationen benötigt als gewöhnlich, um im Global Exploration Schritt ein passendes Modell zu finden. Grund hierfür ist wahrscheinlich die Ungenauigkeit der auf diese Weise erzeugten Modelle, die zu einer zu geringen Zahl korrekt erkannter Inlier führt.

# Kapitel 5

# Zusammenfassung und Ausblick

Im Folgenden werden der Verlauf und die Ergebnisse der Arbeit nochmals zusammengefasst und die wichtigsten Erkenntnisse herausgestellt. Im Anschluss findet sich ein Ausblick auf mögliche, sich anschließende Arbeiten.

## 5.1 Zusammenfassung

Im Rahmen der Arbeit ’Bildbasierte Bewegungsschätzung aus Kamerafahrten anhand prägnanter Merkmale’ wurde ein mehrstufiger Algorithmus entwickelt, der es ermöglicht, aus Bildfolgen eine Trajektorie der Kamerabewegung zu rekonstruieren. Die Kalibrierung der Kamera beruht auf dem Verfahren von Zhang und ermöglicht den Ausgleich der durch das Objektiv entstehenden radialen Verzerrung der Bilder. Die sich anschließende Detektion prägnanter Merkmale wird durch den SIFT-Operator geleistet, welcher neben subpixelgenauer Lokalisation der Merkmale zusätzlich einen stark markanten Deskriptor zu deren Beschreibung liefert. Außerdem sind die Merkmale invariant gegenüber Rotationen, was für einige mögliche Anwendungsfälle sehr relevant ist. Die Suche nach Korrespondenzen wurde auf Basis der Distance Ratio ausgeführt. Hier wurde eine komplette Formalisierung der Korrelationsbeziehung zwischen Merkmalsvektoren präsentiert, welche eindeutig eine symmetrische Beziehung zwischen SIFT-Merkmalsvektoren definiert, die den an eine Korrespondenz gestellten Ansprüchen gerecht wird. Zusätzlich wurde moti-

viert, warum die sonst in der Bildverarbeitung gängige Methode der Hierarchisierung zur Reduktion des Aufwands in diesem speziellen Fall zu schlechteren Inlier-Raten in den gefundenen Korrespondenzen führen kann. Anschließend wurde ein genereller Überblick über den RANSAC-Algorithmus und die aus ihm entspringenden Derivate gegeben. Hierbei diente der zu erwartende Aufwand als Hauptkriterium für die Verbesserungen. Diese Entwicklung gipfelte im BEEM-Algorithmus, der neusten Entwicklung zur Schätzung von Epipolargeometrien auf Basis des 8-Punkte-Algorithmus, welcher viele Ansätze seiner Vorgänger in sich vereint. Dieser BEEM-Algorithmus wurde erstmals in *C++* implementiert, vereinfacht und an das gegeben Szenario angepasst. Zusätzlich wurde die Schätzung der Epipolargeometrie durch den anschließenden Schritt des Testens auf reduzierte Modelle erweitert, was für die korrekte Schätzung einer Trajektorie einen wichtigen Schritt darstellt. So wird nicht nur ein mathematisch korrektes Modell, sondern das der Realität entsprechende Modell für die Kamerabewegung berechnet. Die Bewertung bei der Wahl zwischen unterschiedlichen reduzierten Modellen wurde ebenfalls neu motiviert und angepasst. Die aus der Kombination dieser beiden Verfahren resultierende Fundamental-Matrix wurde im Anschluss faktorisiert, um die Rotation und Translationsrichtung der Kamerabewegung zu ermitteln. Diese wurde schließlich in die globale Trajektorie eingereiht und durch einen Vergleich der rückprojizierten Punkte wurde die korrekte Skalierung gewährleistet. Zur Evaluation des Algorithmus wurden drei Kamerafahrten herangezogen. Zwei reale Kamerafahrten ähneln solchen, wie sie bei den Hauptanwendungsszenarien wie der Robotik oder AR-Anwendungen entstehen könnten. Die dritte ist eine synthetische Kamerafahrt, welche so exakte Ground Truth liefert. Alle Szenen enthalten besondere Schwierigkeiten, wie Glanzpunkte auf Objektoberflächen oder kleine Öffnungswinkel der Kamera, welche auch bei echten Anwendungen vorkommen können. Unter Verwendung dieser Kamerafahrten wurde das 2-SIFT-Verfahren in Bezug auf die Anzahl der erkannten Inlier und als *falsch positiv* klassifizierten Outlier untersucht und begründeter Abstand davon genommen, den Schwellwert der Epipolaren Bedingung bei der Bewertung von 2-SIFT-Modellen zu lockern. Die Notwendigkeit des Prüfens auf reduzierte Bewegungsmodelle wurde exemplarisch an einer Kamerafahrt demonstriert, an welcher die Rekonstruktion der Trajektorie bei Auslassung dieses Schrittes vollkommen scheitert. Schließlich wurde die Genauigkeit auf Basis von synthetischen Daten getestet, welche sich bei der geschätzten Rotation im Bereich von $2^\circ$ bewegt. Die Abweichung der geschätzten Translationsrich-

tung war größer. Sie ist allerdings nicht direkt mit dem Wert der Rotation vergleichbar, da hier Folgefehler einen größeren Einfluss haben. Schließlich wurden die Grenzen des Verfahrens aufgezeigt und begründet.

## 5.2 Ausblick

Der momentane Ablauf des Algorithmus beinhaltet noch keinen Informationsaustausch über mehr als 2 Frames hinweg, abgesehen von der Skalierung. Methoden, die die Epipolargeometrien zwischen mehr als zwei Kameras schätzen, oder Ansätze wie der Bündel-Ausgleich, könnten hier eine stabilere Rekonstruktion der Trajektorie unterstützen. Eine mögliche Änderung des Algorithmus zur Berechnung der Epipolargeometrie weg vom 8-Punkte-Algorithmus hin zu beispielsweise dem 5-Punkte-Algorithmus, oder eine Kombination beider, könnte Probleme wie planare Flächen im Bild noch besser lösen. Der SIFT-Algorithmus wäre in diesem Fall neu zu bewerten und anzupassen. Zur Beschleunigung der Verfahren könnte die Merkmalssuche auf die GPU ausgelagert werden, ein Thema, das im Moment viel Interesse auf sich zieht. Der Aufwand für die Suche nach ähnlichen Merkmalen kann, wie im BEEM-Paper beschrieben, durch mathematische Methoden wie das Locality-Sensitive Hashing reduziert werden. Durch Verwendung des 5-Punkte-Algorithmus würde das 2-SIFT-Verfahren weitestgehend obsolet, da nicht mehr die volle Menge von acht Punktkorrespondenzen benötigt wird. Jedoch wäre eine insgesamte Reduktion des Aufwands aufgrund der geringeren Zahl der benötigten Punkte trotzdem denkbar. Die Tauglichkeit des 2-SIFT-Verfahrens für den 5-Punkte-Algorithmus müsste untersucht werden.

# Anhang A

# Mathematische Bezeichner und Symbole

Im Folgenden werden die verwendeten Variablen und Bezeichner, sowie die vorhandenen Koordinatensysteme aufgelistet und kurz erläutert.

| Variable/Bezeichner | Bedeutung |
|---|---|
| $r_i$ | Distance Ratio der i-ten Korrespondenz |
| $P_q$ | Wahrscheinlichkeit, dass das aktuelle Modell kein Outliermodell ist |
| $t_S$ | Schwellwert für die Größe des Support Sets eines reduzierten Modells, um es als korrekt zu bewerten |
| *Fortgesetzt auf der nächsten Seite* | |

*Fortsetzung der vorherigen Seite*

| Variable/Bezeichner | Bedeutung |
|---|---|
| $[\boldsymbol{t}]_\times$ | Kreuzproduktmatrix des Vektors $\boldsymbol{t}$ |
| $\tilde{\boldsymbol{p}}^{\mathrm{w}}$ | Punkt $\boldsymbol{p} \in \mathbb{P}^3$ in homogenen Weltkoordinaten |
| $\tilde{\boldsymbol{p}}^{\mathrm{i}}$ | Punkt $\boldsymbol{p} \in \mathbb{P}^2$ in homogenen Bildkoordinaten |
| $\tilde{\boldsymbol{p}}^{\mathrm{p}}$ | Punkt $\boldsymbol{p} \in \mathbb{P}^2$ in homogenen Pixelkoordinaten |
| $\alpha$ | Prozentualer Anteil an guten Daten, Inlier |
| $\beta$ | Rotationswinkel einer Rotationsmatrix in Axis-Angle-Repräsentation |
| $\mu(A)$ | Mittelwert von $A$ |
| $\sigma(A)$ | Standardabweichung von $A$ |
| $\Delta_t$ | Abweichung der Translationsrichtung in Grad |
| $\Delta_{\text{R Axis}}$ | Abweichung der Rotationsachse in Grad |
| $\Delta_{\text{R Angle}}$ | Abweichung des Rotationswinkels in Grad |
| $\boldsymbol{v}$ | Rotationsachse einer Rotationsmatrix in Axis-Angle-Repräsentation |
| $svd(\boldsymbol{M})$ | Singulärwertzerlegung einer Matrix $svd(\boldsymbol{M}) = \boldsymbol{U}\,\Sigma\,\boldsymbol{V}^{\mathrm{T}}$ |
| $\mathrm{trace}(\boldsymbol{M})$ | Spur einer Matrix $\mathrm{trace}(\boldsymbol{M}) = \sum_i \boldsymbol{M}_{i,i}$ |
| $\theta$ | Schwellwert für die Epipolare Bedingung 3.15 |
| $\Theta$ | Orientierung eines SIFT-Features |
| $\boldsymbol{E}$ | Die Essential-Matrix |
| $\boldsymbol{F}$ | Die Fundamental-Matrix |
| $\boldsymbol{H}$ | Eine Homographie-Matrix |

*Fortgesetzt auf der nächsten Seite*

| *Fortsetzung der vorherigen Seite* | |
|---|---|
| Variable/Bezeichner | Bedeutung |
| $\boldsymbol{I}$ | Die Identitäts-Matrix |
| $\boldsymbol{K}$ | Die Intrinsische Matrix |
| $\boldsymbol{P}$ | Eine Projektionsmatrix. Sie bildet einen Punkt in homogenen Weltkoordinaten auf seinen Bildpunkt in Homogenen Bildkoordinaten ab. Projektionsmatrizen setzten sich aus einer Translation $\boldsymbol{t}$ und einer Rotation $\boldsymbol{R}$ zusammen: $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \end{pmatrix}$ |
| ${}^{\mathrm{i}}\boldsymbol{R}_{\mathrm{j}}$ | Rotation, die einen Punkt vom Koordinatensystem i in das Koordinatensystem j überführt |

Tabelle A.1: Übersicht über die verwendeten Variablen und Bezeichner.

| Koordinatensystem | Ursprung | Ausrichtung |
|---|---|---|
| Maya Szenenkoordinaten | Szenenabhängig | Rechtshand 3-D |
| *OpenGL* Weltkoordinaten | Optisches Zentrum des ersten Frames | Rechtshand 3-D |
| Weltkoordinaten | Optisches Zentrum des ersten Frames | Linkshand 3-D |
| Kamerakoordinaten | Optisches Zentrum der Kamera | Linkshand 3-D |
| Bildkoordinaten | Hauptpunkt auf der Bildebene | x-Achse nach rechts<br>y-Achse nach oben |
| Pixelkoordinaten | Obere linke Ecke des Bildes | x-Achse nach rechts (Spalten)<br>y-Achse nach unten (Zeilen) |

Tabelle A.2: Übersicht über die verwendeten Koordinatensysteme.

# Anhang B

# Mathematische Verfahren

## B.1 Extraktion der Axis-Angle-Repräsentation aus einer Rotationsmatrix

Gegeben sei eine Rotationsmatrix $\boldsymbol{R} \in \mathbb{R}^{(3,3)}$, gesucht der Vektor $\boldsymbol{v} \in \mathbb{R}^3$ und der Winkel $\beta$, welche zusammen die gleiche Rotation in Axis-Angle-Repräsentation beschreiben.

$$\begin{aligned} \beta &= \arccos\left(\frac{\mathrm{trace}(\boldsymbol{R}) - 1}{2}\right) \\ v_{\mathrm{x}} &= \frac{1}{2 \cdot \sin\beta} \cdot (R_{3,2} - R_{2,3}) \\ v_{\mathrm{y}} &= \frac{1}{2 \cdot \sin\beta} \cdot (R_{1,3} - R_{3,1}) \\ v_{\mathrm{z}} &= \frac{1}{2 \cdot \sin\beta} \cdot (R_{2,1} - R_{1,2}) \end{aligned} \tag{B.1}$$

## B.2 Erzwingen des Rangs einer Matrix mit Hilfe der Singulärwertzerlegung

Der Rang einer Matrix kann mit Hilfe der Singulärwertzerlegung erzwungen werden. Die Idee ist, eine Matrix $\boldsymbol{M}$ mittels Singulärwertzerlegung zu zerlegen, die Singulärwerte in $\Sigma$ so zu ändern, dass sie dem gewünschten Rang entsprechen, und die Matrix anschließend mit dem neuen $\Sigma'$ wieder zu $\boldsymbol{M}'$ zusammenzusetzen. Hierbei macht man sich zu Nutze, dass die Anzahl der Singulärwerte $\sigma \neq 0$ den Rang der Matrix bestimmt, und diese geordnet auf der Diagonalen von $\Sigma$ liegen ($\sigma_i \geq \sigma_{i+1}$). Im Folgenden wird exemplarisch der Rang $m < n$ einer Matrix $\boldsymbol{M}$ erzwungen.

$$\boldsymbol{U}\,\Sigma\,\boldsymbol{V}^{\mathrm{T}} = svd(\boldsymbol{M}) \tag{B.2}$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_n \end{pmatrix}$$

$$\Sigma' = \begin{pmatrix} \sigma_1 & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \sigma_m & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{M}' = \boldsymbol{U}\,\Sigma'\,\boldsymbol{V}^{\mathrm{T}}$$

## B.3 Die Kreuzproduktmatrix

Das Kreuzprodukt $\boldsymbol{v}\times\boldsymbol{u}$ zweier Vektoren lässt sich linear mit Hilfe der Kreuzproduktmatrix $[\boldsymbol{v}]_\times$ linearisiert darstellen: $\boldsymbol{v}\times\boldsymbol{u} = [\boldsymbol{v}]_\times\,\boldsymbol{u}$.

$$[\boldsymbol{v}]_\times = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_1 & 0 & -v_1 \\ -v_2 & v_3 & 1 \end{pmatrix} \tag{B.3}$$

## B.4 Triangulierung eines Weltpunktes

Seien $\boldsymbol{P}^1$ und $\boldsymbol{P}^2$ die Projektionsmatrizen der beiden Kameras, $\tilde{\boldsymbol{p}}^{\mathrm{i}}$ und $\tilde{\boldsymbol{q}}^{\mathrm{i}}$ die beobachteten Bildpunkte eines Weltpunktes $\tilde{\boldsymbol{w}}^{\mathrm{w}}$. Gesucht ist der Weltpunkt $\tilde{\boldsymbol{w}}^{\mathrm{w}}$, dessen Projektionen $\boldsymbol{P}^1\,\tilde{\boldsymbol{w}}^{\mathrm{w}}$ und $\boldsymbol{P}^2\,\tilde{\boldsymbol{w}}^{\mathrm{w}}$ den geringsten quadratischen Abstand zu den beobachteten Punkten haben.

$$\tilde{\boldsymbol{w}}^{\mathrm{w}} = \underset{\tilde{\boldsymbol{w}}^{\mathrm{w}}}{\operatorname{argmin}} \left\|\boldsymbol{P}^1\,\tilde{\boldsymbol{w}}^{\mathrm{w}} - \tilde{\boldsymbol{p}}^{\mathrm{i}}\right\|^2 \cdot \left\|\boldsymbol{P}^2\,\tilde{\boldsymbol{w}}^{\mathrm{w}} - \tilde{\boldsymbol{q}}^{\mathrm{i}}\right\|^2 \tag{B.4}$$

Sei $\begin{pmatrix} u_1 \\ v_1 \\ s_1 \end{pmatrix}$ der optimale Punkt der Abbildung im ersten Bild, so ergibt sich:

$$\boldsymbol{P^1}\,\tilde{\boldsymbol{w}}^{\mathrm{w}} = \begin{pmatrix} P_{1,1} & P_{1,2} & P_{1,3} & P_{1,4} \\ P_{2,1} & P_{2,2} & P_{2,3} & P_{2,4} \\ P_{3,1} & P_{3,2} & P_{3,3} & P_{3,4} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix} \tag{B.5}$$

$$\boldsymbol{P^1}\,\tilde{\boldsymbol{w}}^{\mathrm{w}} = \begin{pmatrix} P_{1,1}{\cdot}w_1 + P_{1,2}{\cdot}w_2 + P_{1,3}{\cdot}w_3 + P_{1,4}{\cdot}w_4 \\ P_{2,1}{\cdot}w_1 + P_{2,2}{\cdot}w_2 + P_{2,3}{\cdot}w_3 + P_{2,4}{\cdot}w_4 \\ P_{3,1}{\cdot}w_1 + P_{3,2}{\cdot}w_2 + P_{3,3}{\cdot}w_3 + P_{3,4}{\cdot}w_4 \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ s_1 \end{pmatrix}$$

Die Umformung aus dem $\mathbb{P}^2$ in den $\mathbb{R}^2$ liefert

$$\boldsymbol{p}^{\mathrm{i}} = \begin{pmatrix} \frac{u_1}{s_1} \\ \frac{v_1}{s_1} \end{pmatrix} \tag{B.6}$$

Mit $\boldsymbol{p}^{\mathrm{i}} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ und $\boldsymbol{q}^{\mathrm{i}} = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$ ergibt sich aus B.4

$$\left(\frac{u_1}{s_1} - p_1\right) + \left(\frac{v_1}{s_1} - p_2\right) + \left(\frac{u_2}{s_2} - q_1\right) + \left(\frac{u_2}{v_2} - q_1\right) \stackrel{!}{=} 0 \tag{B.7}$$

Die Summanden können einzeln betrachtet werden:

$$\begin{aligned} \left(\frac{u_1}{s_1} - p_1\right) &= 0 \\ u_1 - p_1 \cdot s_1 &= 0 \\ P_{1,1}w_1 + P_{1,2}w_2 + P_{1,3}w_3 + P_{1,4}w_4 & \\ -p_1 \cdot (P_{3,1}w_1 + P_{3,2}w_2 + P_{3,3}w_3 + P_{3,4}w_4) &= 0 \end{aligned} \tag{B.8}$$

Zusammengenommen liefern so alle vier Summanden insgesamt vier Gleichungen zum Aufstellen einer Messwertmatrix:

$$\begin{pmatrix} {P^1}_{1,1} - p_1 \cdot {P^1}_{3,1} & {P^1}_{1,2} - p_1 \cdot {P^1}_{3,2} & {P^1}_{1,3} - p_1 \cdot {P^1}_{3,3} & {P^1}_{1,4} - p_1 \cdot {P^1}_{3,4} \\ {P^1}_{2,1} - p_2 \cdot {P^1}_{3,1} & {P^1}_{2,2} - p_2 \cdot {P^1}_{3,2} & {P^1}_{2,3} - p_2 \cdot {P^1}_{3,3} & {P^1}_{2,4} - p_2 \cdot {P^1}_{3,4} \\ {P^2}_{1,1} - q_1 \cdot {P^2}_{3,1} & {P^2}_{1,2} - q_1 \cdot {P^2}_{3,2} & {P^2}_{1,3} - q_1 \cdot {P^2}_{3,3} & {P^2}_{1,4} - q_1 \cdot {P^2}_{3,4} \\ {P^2}_{2,1} - q_2 \cdot {P^2}_{3,1} & {P^2}_{2,2} - q_2 \cdot {P^2}_{3,2} & {P^2}_{2,3} - q_2 \cdot {P^2}_{3,3} & {P^2}_{2,4} - q_2 \cdot {P^2}_{3,4} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix} = 0 \tag{B.9}$$

Dieses Gleichungssystem kann mit Hilfe der Singulärwertzerlegung berechnet werden. Das Ergebnis ist der Weltpunkt $\tilde{\boldsymbol{w}}^{\mathrm{w}}$, der den kleinsten quadratischen Rückprojektionsfehler liefert.

## B.5 Messwertmatrizen für reduzierte Modelle

### B.5.1 Keine Bewegung

Das Modell für Bewegungslosigkeit zwischen zwei Frames ist eindeutig und hat somit keine Freiheitsgrade. Es wird getestet, ob korrespondierende Punkte innerhalb einer gewissen Toleranzgrenze identische Koordinaten haben.

### B.5.2 Reine Translation

Im Vorfeld wird hier der gleiche Schritt zur Normalisierung der Koordinaten wie im Fall des Acht-Punkte Algorithmus vorgenommen.
Zur Instanziierung eines Modells, das eine reine Translation schätzt, werden lediglich zwei korrespondierende Punktepaare benötigt. Um ein stabileres Ergebnis zu erhalten, ist es jedoch sinnvoll, mehr Punktepaare zu verwenden. Seien $\tilde{\boldsymbol{p}}_i^{\mathrm{i}}$ und $\tilde{\boldsymbol{q}}_i^{\mathrm{i}}$ die Korrespondierenden Punkte in Bildkoordinaten.

$$\tilde{\boldsymbol{q}}^{\mathrm{iT}} [\boldsymbol{t}]_\times \tilde{\boldsymbol{p}}^{\mathrm{i}} \stackrel{!}{=} 0 \tag{B.10}$$

$$(q_x, q_y, 1) \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ 1 \end{pmatrix} = 0$$

$$(q_y{\cdot}t_3 - t_2, -q_x{\cdot}t_3 + t_1, q_x{\cdot}t_2 - q_y{\cdot}t_1) \begin{pmatrix} p_x \\ p_y \\ 1 \end{pmatrix} = 0 \tag{B.11}$$

$$q_y{\cdot}t_3{\cdot}p_x - t_2{\cdot}p_x - q_x{\cdot}t_3{\cdot}p_y + t_1{\cdot}p_y + q_x{\cdot}t_2 - q_y{\cdot}t_1 = 0$$

$$(p_y - q_y){\cdot}t_1 + (q_x - p_x){\cdot}t_2 + (q_y{\cdot}p_x - q_x{\cdot}p_y){\cdot}t_3 = 0$$

Daraus folgt der Aufbau der Messwertmatrix:

$$\begin{pmatrix} p_y - q_y & q_x - p_x & q_y \cdot p_x - q_x \cdot p_y \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = 0 \tag{B.12}$$

### B.5.3 Rotation um das optische Zentrum

Eine Rotation um das optische Zentrum der Kamera ist äquivalent zu einer Homographie der Bildebene. Es genügt also, eine Homographiematrix zwischen den beiden Bildern zu schätzen. Im Vorfeld wird hier wieder der gleiche Schritt der Normalisierung der Punktkoordinaten wie im Fall des Acht-Punkte Algorithmus vorgenommen. Es sind mindestens 4 Punktkorrespondenzen nötig, um eine Homographie zu schätzen. Sei $\boldsymbol{H}$ die Homographiematrix, die einen Punkt $\tilde{\boldsymbol{p}}^{\mathrm{i}} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$ auf $\tilde{\boldsymbol{q}}^{\mathrm{i}} = \begin{pmatrix} x' \\ y' \\ k' \end{pmatrix}$ abbildet. Es gilt:

$$\tilde{\boldsymbol{q}}^{\mathrm{i}} = \boldsymbol{H}\, \tilde{\boldsymbol{p}}^{\mathrm{i}} \tag{B.13}$$
$$\begin{pmatrix} x' \\ y' \\ k' \end{pmatrix} = \begin{pmatrix} H_{1,1} & H_{1,2} & H_{1,3} \\ H_{2,1} & H_{2,2} & H_{2,3} \\ H_{3,1} & H_{3,2} & H_{3,3} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Da $k' = 1$, gilt:

$$\begin{aligned} x' &= \frac{H_{1,1} \cdot x + H_{1,2} \cdot y + H_{1,3}}{H_{3,1} \cdot x + H_{3,2} \cdot y + H_{3,3}} \\ y' &= \frac{H_{2,1} \cdot x + H_{2,2} \cdot y + H_{2,3}}{H_{3,1} \cdot x + H_{3,2} \cdot y + H_{3,3}} \end{aligned} \tag{B.14}$$

Das entstandene Gleichungssystem umformen und ausmultiplizieren:

$$\begin{aligned} H_{1,1} \cdot x + H_{1,2} \cdot y + H_{1,3} - x' \cdot (H_{3,1} \cdot x + H_{3,2} \cdot y + H_{3,3}) = 0 \\ H_{2,1} \cdot x + H_{2,2} \cdot y + H_{2,3} - y' \cdot (H_{3,1} \cdot x + H_{3,2} \cdot y + H_{3,3}) = 0 \end{aligned} \tag{B.15}$$

$$H_{1,1}{\cdot}x + H_{1,2}{\cdot}y + H_{1,3} - H_{3,1}{\cdot}x{\cdot}x' - H_{3,2}{\cdot}y{\cdot}x' - H_{3,3}{\cdot}x' = 0 \qquad \text{(B.16)}$$
$$H_{2,1}{\cdot}x + H_{2,2}{\cdot}y + H_{2,3} - H_{3,1}{\cdot}x{\cdot}y' - H_{3,2}{\cdot}y{\cdot}y' - H_{3,3}{\cdot}y' = 0$$

Sei $\boldsymbol{h} = (H_{1,1}..H_{3,3})^{\mathrm{T}}$ die Homographiematrix in Vektorform. Es ergibt sich folgender Aufbau für die Messwertmatrix:

$$\begin{pmatrix} x & y & 1 & 0 & 0 & 0 & -x'{\cdot}x & -x'{\cdot}y & -x' \\ 0 & 0 & 0 & x & y & 1 & -y'{\cdot}x & -y'{\cdot}y & -y' \\ \vdots & & \vdots & & \vdots & & & \vdots & \end{pmatrix} \boldsymbol{h} = 0 \qquad \text{(B.17)}$$

# Anhang C

# Implementationsdetails

## C.1 Octave Prototyp

Der in *Octave* implementierte Prototyp umfasst die folgende Funktionalität:

**8-Punkte-Algorithmus** `F = eightPoint(M,sing_cons)`

8-Punkte-Algorithmus aus der BEEM-Implementation von Ilan Shimshoni. Eingabeparameter sind die Messwertmatrix, und eine boolesche Variable, die Auskunft über die Anwendung des Singularity Constraint gibt. Rückgabewert ist die F-Matrix.

**Kreuzproduktmatrix** `Tx = crossProdMat(t)`

Gibt eine zum Vektor T korrespondierende Kreuzproduktmatrix zurück.

**Zufällige Rotationsmatrix** `R = randomRotationMatrix(limit)`

Liefert eine zufällig erzeugte Rotationsmatrix. Der Parameter `limit` dient zur Einschränkung des Rotationswinkels, wobei `1` totale Freiheit bedeutet, `0.5` einen maximalen Winkel von $\pm 90°$ und `0` keine Rotation.

**Zufällige Translation** `t = randomTranslationVector(limit)`

Liefert einen zufälligen Translationsvektor mit Werten im Bereich von `[-limit .. limit]`.

**E-Matrix Zerlegung** `[R, t] = decomposeE(E, p, q)`

Zerlegt eine Essential-Matrix, mit Hilfe einer gegeben Punktkorrespondenz in Bildkoordinaten, in eine Rotationsmatrix $\boldsymbol{R}$ und einen Translationsvektor $\boldsymbol{t}$, die zusammen die 2. Kameramatrix beschreiben.

**Triangulierung eines Weltpunktes** `pw = triangulate(P1, P2, p, q)`

Liefert den Weltpunkt der Bildpunkte `p` und `q`, die mit den Projektionsmatrizen `P1` und `P2` erzeugt wurden, der den geringsten quadratischen Rückprojektionsfehler hat (siehe auch Anhang B.4).

**Rekonstruktion einer Trajektorie** `trajectory.m`

Dieses Skript generiert zufällig eine Menge Kameras und Weltpunkte. Es projiziert die Punkte anschließend in die Bildebenen und berechnet daraus die Epipolargeometrien. Die resultierenden E-Matrizen werden wieder in eine Rotation und Translation zerlegt, und der Fehler zu den ursprünglichen Kameras gemessen. Die Möglichkeit des Miteinbeziehens von Rauschen besteht.

## C.2 Hauptimplementation

Die Hauptimplementation liegt in *C++* vor, und wurde als *Qt*-Projekt angelegt. Zentraler Bestandteil sind zwei Bibliotheken, die erzeugt werden. Sie beinhalten die Funktionalität der Kalibrierung (`libDecker07cal.a`) und der Visuellen Odometrie (`libDecker07vo.a`). Mehrere Programme, die diese Bibliotheken verwenden, wurden erstellt und werden später erläutert. Die Verzeichnisstruktur im Programmordner sieht folgendermaßen aus:

```
bin/
cal/
Calibration/
Eval/
libs/
OpenGLvis/
StereoGUI/
Undistortion/
vo/
```

Die Verzeichnisse `cal` und `vo` enthalten die Quelldateien für die Bibliotheken zur Kalibrierung und Visuellen Odometrie, welche im Verzeichnis `libs` erzeugt werden. In `bin` werden die Binaries abgelegt, welche aus den `main` Methoden in den Unterverzeichnissen, die mit einem Großbuchstaben beginnen, erzeugt werden. Sie linken alle gegen die jeweils benötigte Bibliothek.

### C.2.1 Kalibrierung

`./Calibration <Bildliste>`
z. B.: `./Calibration calibrationImage*`
Das Programm erwartet eine Liste von Bildern des Kalibriermusters im *pgm*-Format als Eingabe und versucht, daraus eine Kalibrierung zu erstellen. Im Erfolgsfall wird eine Datei

`cameraParameters.txt` erzeugt, in der alle intrinsischen Kameraparameter gespeichert sind. Es handelt sich hierbei um die Werte des *OpenCV*-Strukts `CvCamera`, welches neben der Bildgröße, der Intrinsischen Matrix und der Entzerrungsparameter auch die extrinsischen Parameter speichern kann.

### C.2.2 Entzerrung

`./Undistortion Kamerparameterdatei <Bildliste>`
z. B.: `./Undistortion cameraParameters.txt image*`
Das Programm erwartet als erstes Argument eine Datei mit den intrinsischen Kameraparametern, wie sie von `Calibration` erzeugt wird, gefolgt von einer Liste von Bildern im *pgm*-Format, welche es zu entzerren gilt. Die entzerrten Bilder werden mit dem Präfix `undistorted_` im *pgm*-Format abgespeichert.

### C.2.3 SIFT-Korrespondenzen

`./StereoGUI`
Ein Programm mit Graphischer Oberfläche in *Qt*. Es können zwei Bilder im *pgm*-Format geladen werden, in welchen anschließend SIFT-Merkmale detektiert und mit Hilfe des BEEM-Algorithmus Korrespondenzen ermittelt werden. In Bild C.1 ist die Oberfläche des Programms zu sehen.

### C.2.4 OpenGL Visualisierung

`./OpenGLvis Kameraparameterdatei <entzerrte Bildliste>`
z. B.: `./OpenGLvis cameraParameters.txt undistorted_image*`
Das Programm liest die intrinsischen Kameraparameter aus dem ersten Argument, und versucht dann die Liste der entzerrten Bilder sukzessive mit dem in Kapitel 3 vorgestellten Verfahren in eine Globale Trajektorie einzureihen. Diese wird am Schluss in *OpenGL* visualisiert. Die Betrachterposition lässt sich hierbei frei mit Hilfe der Maus und den Pfeiltasten wählen, die rekonstruierten Punkte können durch Drücken der `p`-Taste ein- und aus-

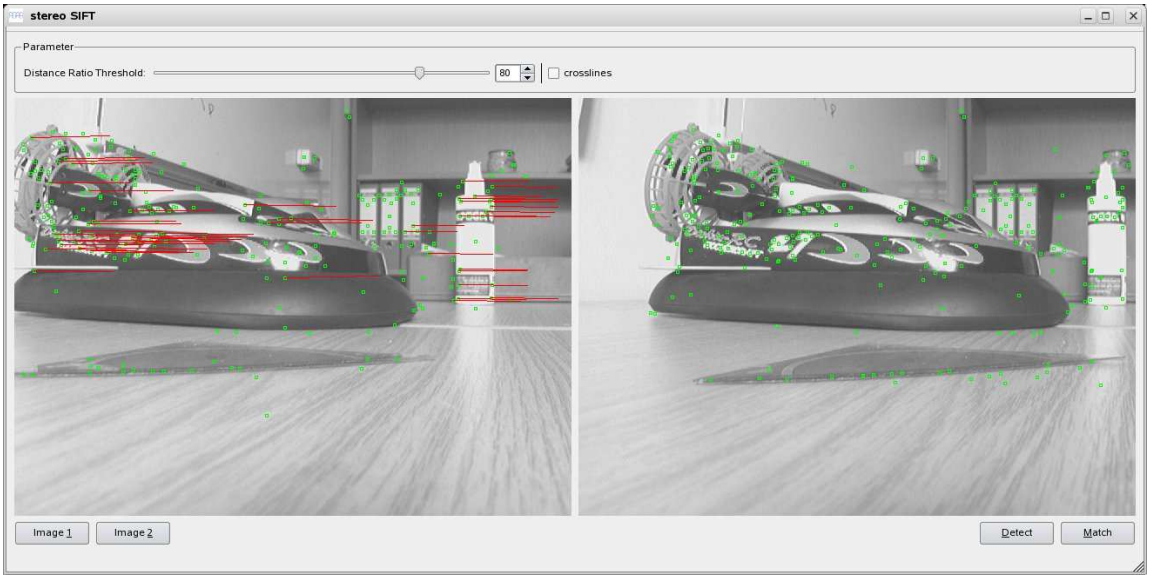


Bild C.1: Graphische Oberfläche zur Visualisierung von Korrespondierenden SIFT-Merkmalen zwischen zwei Bildern.

geschaltet werden. Die globale Skalierung lässt sich mit `+` und `-` anpassen. Bild C.2.4 zeigt eine solche Szene aus mehreren Positionen.

### C.2.5 Benötigte Bibliotheken und Abhängigkeiten

Als *build*-Tool wurde `QMake` in der Version `2.00a`, als Bestandteil von `Qt 4.1.0`, verwendet. Die Oberfläche wurde ebenfalls mit *Qt* entworfen. Als Bildverarbeitungsbibliothek wird `OpenCV 1.0.0` benutzt. Es wird sowohl für die Kalibrierung als auch für die bildbasierte Bewegungsschätzung benötigt.
Zusätzlich verwendet die Bibliothek `libDecker07vo.a` für mathematische Subroutinen das Paket `lapackpp-2.4.10`. Das Programm `Calibration` lässt sich allein unter Einbindung von `libDecker07cal.a` kompilieren. Alle weiteren Programme benötigen zusätzlich die Bibliothek zur Visuellen Odometrie und teilweise *OpenGL* sowie *glut*. Bild C.3 verdeutlicht diese Zusammenhänge.

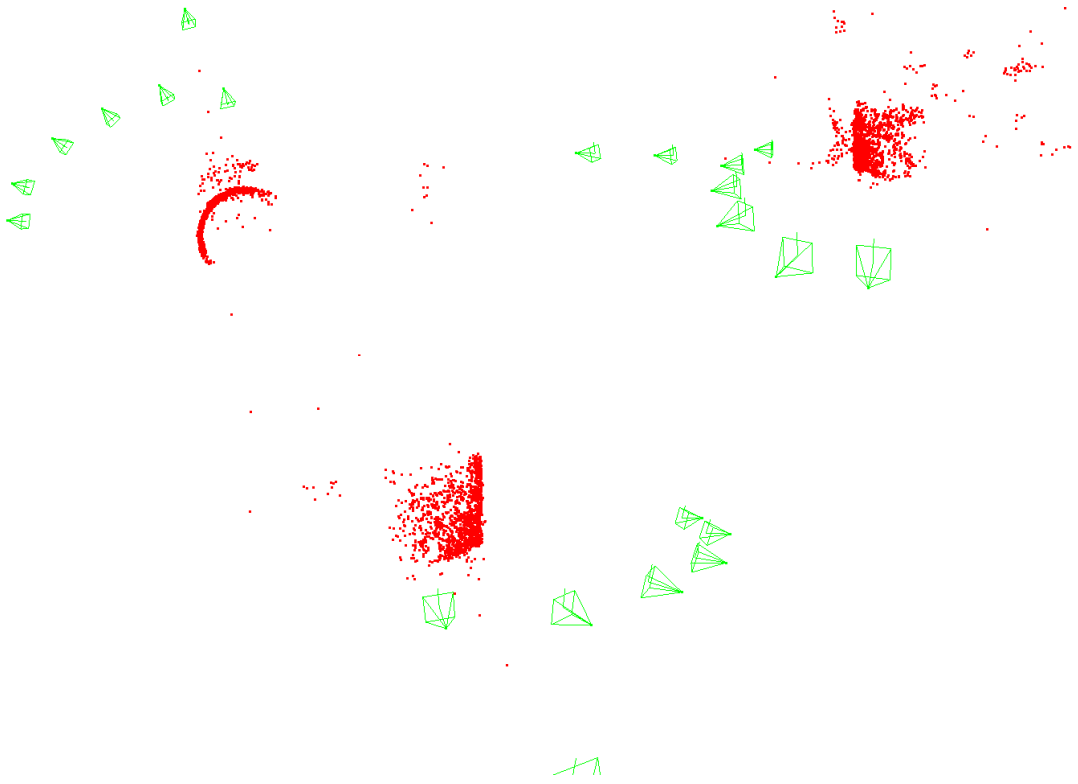

Bild C.2: Eine in *OpenGL* visualisierte Trajektorie aus verschiedenen Blickwinkeln.

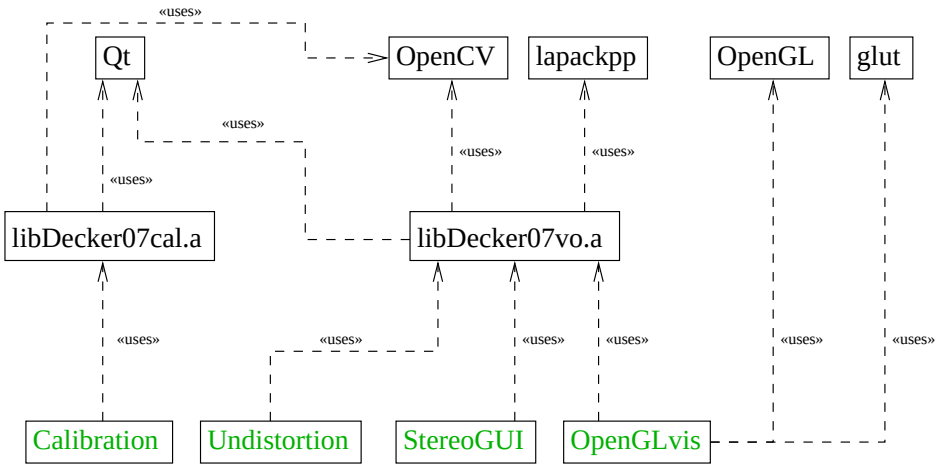


Bild C.3: Abhängigkeiten zwischen externen Bibliotheken (oben), selbst erzeugten Bibliotheken (Mitte) und Programmen (unten, grün). Die Beziehung «uses» bezeichnet die Abhängigkeit und ist in der Regel als transitiv zu verstehen.

# Anhang D

# Aufbau der CD

Der Ausarbeitung liegt eine CD bei, auf der die wichtigsten Quelldaten gespeichert sind. Der Aufbau ist wie folgt:

```
Ausarbeitung/
  Arbeit/
  VortragOberseminar/
  Decker2007.pdf

Literatur/

Programmcode/
  Hauptimplementation/
  OctavePrototyp/

Sonstiges/
  Labor/
  NeonChrome/
  Tasse/
```

Im Verzeichnis `Ausarbeitung` liegt dieses Dokument, in den Unterverzeichnissen die Quellen dazu sowie zum Vortrag im Oberseminar Aktives Sehen vom 09.05.2007. Im Ordner `Literatur` finden sich die zitierten und weitere relevante Papers. Sie sind ebenfalls in der Literaturdatenbank abgelegt. Der Programmcode findet sich im gleichnamigen Ordner, getrennt nach *C++*- und *Octave*-Implementation. Zur Erzeugung der Bibliotheken und Programme genügt bei gegebenen Systemvoraussetzungen ein `qmake && make` im obersten Ordner der `Hauptimplementation`. Unter `Sonstiges` finden sich die zur Evaluation verwendeten Kamerafahrten.

# Literaturverzeichnis


[CM05] CHUM, O. ; MATAS, J.: Matching with PROSAC - Progressive Sample Consensus. In: SCHMID, Cordelia (Hrsg.) ; SOATTO, Stefano (Hrsg.) ; TOMASI, Carlo (Hrsg.): *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* Bd. 1. Los Alamitos, USA : IEEE Computer Society, 6 2005, 220-226

[CMK03] CHUM, O. ; MATAS, J. ; KITTLER, J.: Locally optimized RANSAC. In: G. GOOS, J. van L. J. Hartmanis H. J. Hartmanis (Hrsg.): *DAGM 2003: Proceedings of the 25th DAGM Symposium*. Heidelberger Platz 3, 14197, Berlin, Germany : Springer-Verlag, 9 2003 (LNCS 2781), S. 236–243

[CWM05] CHUM, O. ; WERNER, Tomá ; MATAS, J.: Two-view Geometry Estimation Unaffected by a Dominant Plane. In: SCHMID, Cordelia (Hrsg.) ; SOATTO, Stefano (Hrsg.) ; TOMASI, Carlo (Hrsg.): *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)* Bd. 1. Los Alamitos, USA : IEEE Computer Society, 6 2005, S. 772–780

[Dav03] DAVISON, Andrew: Real-time simultaneous localisation and mapping with a single camera. In: *9th international Conference on Computer Vision (ICCV)*. Nice, France, 2003

[DMM03] DAVISON, Andrew ; MAYOL, Walterio ; MURRAY, David: Real-Time Localisation and Mapping with Wearable Active Vision. In: *IEEE International Symposium on Mixed and Augmented Reality*, IEEE Computer Society Press, 10 2003

[FB81] FISCHLER, Martin A. ; BOLLES, Robert C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: *Communications of the ACM* 24 (1981), Nr. 6, 381-395. `file:///lab/as/Docs/Articles/Fischler1981RSC.pdf`

[GIM99] GIONIS, Aristides ; INDYK, Piotr ; MOTWANI, Rajeev: Similarity Search in High Dimensions via Hashing. In: *VLDB '99: Proceedings of the 25th International Conference on Very Large Data Bases*. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1999, S. 518–529

[Har97] HARTLEY, Richard I.: In Defense of the Eight-Point Algorithm. In: *Pattern Analysis and Machine Intelligence* 19 (1997), 6, Nr. 6, S. 580–593

[HLON91] HARALICK, R. M. ; LEE, C. ; OTTENBERG, K. ; NÖLLE, M.: Analysis and Solutions of the Three Point Perspective Pose Estimation Problem. In: *Proceedings Computer Vision and Pattern Recognition 1991*. Lahaina, Maui, Hawaii, 6 1991, S. 592–598

[HZ03] HARTLEY, Richard I. ; ZISSERMAN, Andrew: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003

[Low04] LOWE, David G.: Distinctive Image Features from Scale-Invariant Keypoints. In: *International Journal of Computer Vision* 60 (2004), Nr. 2, 91-110. `http://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf`

[MC02] MATAS, J. ; CHUM, Ondrej: Randomized RANSAC. In: WILDENAUER, H. (Hrsg.) ; KROPATSCH, W. (Hrsg.): *Proceedings of the Computer Vision Winter Workshop '02*. Wien, Austria, 2 2002, 49–58

[NBN06] NISTÉR, David ; BERGEN, James R. ; NARODITSKY, Oleg: Visual Odometry for Ground Vehicle Applications. In: *Journal of Field Robotics* 23 (2006), Nr. 1. `http://www.vis.uky.edu/~dnister/Publications/2006/JFR/JFR1.pdf`. – inaugural issue

[Nis03] NISTÉR, David: An Efficient Solution to the Five-Point Relative Pose Problem. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2003, 195-202

[NNB04] NISTÉR, David ; NARODITSKY, Oleg ; BERGEN, James R.: Visual Odometry. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2004, 652-659

[Shi06] SHIMSHONI, Ilan: Balanced Exploration and Exploitation Model Search for Efficient Epipolar Geometry Estimations (BEEM). In: *9th European Conference on Computer Vision (ECCV 2006)* Bd. II, 2006, 151-164. – BEEM, Code-Demo available at Shimshonis Webpage

[SS03] STRELOW, D. ; SINGH, S.: Online motion estimation from image and inertial measurements. In: *Workshop on Integration of Vision and Inertial Sensors* (2003). `http://www.dennis-strelow.com/publications/`

[TZ00] TORR, P. H. S. ; ZISSERMAN, A.: MLESAC: a new robust estimator with application to estimating image geometry. In: *Computer Vision and Image Understanding* 78 (2000), Nr. 1, S. 138–156

[TZM95] TORR, P.H.S. ; ZISSERMAN, A. ; MAYBANK, S.J.: Robust detection of degenerate configurations for the fundamental matrix. In: *Fifth International Conference on Computer Vision* (1995), S. 1037

[Zha00] ZHANG, Zhengyou: A flexible new technique for camera calibration. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), Nr. 11, 1330-1334. `http://research.microsoft.com/~zhang/calib/`