

Schätzung der Oberflächenorientierung aus Glanzlichtverfolgung

Diplomarbeit

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

vorgelegt von

Conrad Emde

Betreuer: Prof. Dr.-Ing. Dietrich Paulus, Institut für Computervisualistik,
Fachbereich Informatik
Erstgutachter: Prof. Dr.-Ing. Dietrich Paulus, Institut für Computervisualistik,
Fachbereich Informatik
Zweitgutachter:

Koblenz, im November 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

Zusammenfassung

Die Erstellung räumlicher Abbilder aus planaren Ansichten gewinnt immer mehr Bedeutung in der modernen Medizintechnik. 3D-Rekonstruktionen haben wesentlich zur besseren Detektion, wie auch zu Optimierung und Innovation in der Diagnostik und Behandlungsmethodik bestimmter Krankheitsbilder beigetragen. Durch die Verfahren der Bildverarbeitung ist es möglich, aus Bildsequenzen eine 3D-Abbildung der gefilmten Szene zu erstellen. Ziel dieser Diplomarbeit soll es sein, zu untersuchen, inwieweit sich aus der Aufnahmetechnik aus einer Reihe unkalibrierter Endoskopiebilder weitere Rückschlüsse über die Oberflächenbeschaffenheit des betrachteten Gewebes ziehen lassen.

Hierbei wird das Phänomen zugrundegelegt, daß bei der Aufnahme der Bilder Glanzlichter auftreten, wenn die Beleuchtung am Kamerakopf orthogonal zur Gewebeoberfläche auftrifft. Diese Glanzlichter geben daher implizit Aufschluss über die Oberflächenorientierung des Gewebes. Aufgabe ist es nun, diese Glanzlichter in einer Reihe von unkalibrierten Endoskopieaufnahmen zu finden, die Bilder aus der Sequenz einander zuzuordnen, also Korrespondenzen zwischen den Bildern zu finden, und unter Einbeziehung der Kamerageometrie Rückschlüsse auf die Gewebeoberfläche zu ziehen.

Zuerst müssen hierfür die Glanzlichter in den Einzelbildern der Sequenz gefunden werden. Dazu wird ein Verfahren verwendet, welches die Glanzlichter durch eine Zerlegung des HSV-Farbraums detektiert und deren Mittelpunkt errechnet. Um die Kamerageometrie zu schätzen, werden mithilfe eines Punktverfolgers Punktkorrespondenzen zwischen den Einzelbildern erstellt, anhand derer sich die Fundamentalmatrix durch RANSAC errechnen läßt. Unter Anwendung eines Autokalibrierungsverfahrens werden aus den geschätzten Fundamentalmatrizen dann in einem abschließenden Schritt die internen Kameraparameter ermittelt. So sollte möglich sein, die Glanzlichter durch eine Sequenz von Bildern zu verfolgen und die Oberflächennormalen einem Referenzbild zuzuordnen.

Inhaltsverzeichnis

- 1 Einführung 1**
 - 1.1 Endoskopie 2
 - 1.2 Aufgabenstellung 5
 - 1.3 Methode 6
 - 1.4 Aufteilung der Arbeit 6

- 2 Grundlagen 9**
 - 2.1 Homogene Koordinaten 11
 - 2.2 Kamerakoordinatensystem → Bildkoordinatensystem 11
 - 2.3 Bildkoordinatensystem → Pixelkoordinatensystem 13
 - 2.4 Epipolargeometrie 15
 - 2.5 Essential Matrix 16
 - 2.6 Fundamental Matrix 17
 - 2.7 Glanzlichtentstehung 18

- 3 Stand der Technik 21**
 - 3.1 Punktverfolger 21
 - 3.2 Autokalibrierung 22

3.3	Glanzlichtdetektion	23
3.4	Normalenbestimmung	23
4	Eigener Ansatz	25
4.1	Punktverfolger	25
4.1.1	Lucas-Kanade mit Pyramiden	27
4.1.2	SIFT	28
4.2	Fundamentalmatrixschätzung	28
4.2.1	Berechnung der Fundamentalmatrix	29
4.2.2	RANSAC	30
4.3	Autokalibrierung	31
4.3.1	Verfahren - zwei Bilder	31
4.3.2	Verfahren - Sequenz	33
4.3.3	Dynamic Hill Climbing	34
4.4	Essentialmatrix Dekomposition	36
4.4.1	Dekomposition - Zwei Bilder	36
4.4.2	Dekomposition - Sequenz	39
4.5	Normalenbestimmung	39
4.6	Glanzlichtdetektion	40
4.6.1	HSV-Schwellwertanalyse	40
4.6.2	Bestimmen des Zentroids	42
5	Experimente und Ergebnisse	45
5.1	Implemetation	45
5.2	Versuchsaufbau und Durchführung	46
5.2.1	Testbilder	46

5.2.2	Glanzlichtdetektion	50
5.2.3	Punktverfolger	50
5.2.4	Fundamentalmatrixschätzung	51
5.2.5	Autokalibrierung und Essentialmatrixdekomposition	51
5.3	Ergebnisse	52
5.3.1	Glanzlichtdetektion	52
5.3.2	Punkttracking und Fundamentalmatrixschätzung	53
5.3.3	Autokalibration	56
5.3.4	Geometrieschätzung	57
6	Fazit	59
6.1	Ergebnisanalyse	59
6.2	Zukunft	60
A	Mathematische Bezeichner und Symbole	61
B	Implementationsdetails	65
B.1	Benötigte Bibliotheken	65
B.2	Doxygen-Dokumentation	65

Kapitel 1

Einführung

Bildgebende Verfahren halten zunehmend Einzug in die moderne Medizintechnik. In der Bildverarbeitung entwickelte Methoden haben einen maßgeblichen Einfluss auf die Entwicklung neuer Möglichkeiten zur Diagnose und Therapie bestimmter Krankheitsbilder genommen und auch zur verbesserten Veranschaulichung von Problematiken in der Ausbildung angehender Mediziner geführt.

Durch Aufnahmen aus dem Inneren des Körpers, mithilfe von Videokameras, können in der Endoskopie Spiegelungen vorgenommen werden, um Krankheiten zu diagnostizieren und den Arzt intraoperativ zu leiten. Endoskopie gehört zu den *minimal invasiven* Eingriffen in der Chirurgie, die darauf abzielen, den Patienten einer möglichst geringen Belastung auszusetzen und eine schwerwiegende Operation zu umgehen. Im schlimmsten Falle wird eine endoskopische Untersuchung als unangenehm empfunden, was aber infolge der gegebenen Schonung des Patienten im Hinblick auf Eingriff und Nachbehandlung zu vernachlässigen ist.

Eine bildbasierte Behandlung macht allerdings aufgrund ihrer Natur eine visuelle Unterstützung des Arztes nötig. Dadurch, dass der Eingriff nicht unmittelbar am erkrankten Organ erfolgt, sondern durch ein Videobild in den Operationsaal übertragen wird, wird die Koordination der Instrumente und die Klassifikation des Gewebes ungleich schwieriger als bei einer herkömmlichen Operation. Die Bildverarbeitung versucht, dem Arzt bestmögliche Bedingungen durch ergänzende Informationen aus dem Bildmaterial zu schaf-

fen. Eine wichtige Rolle spielen hier 3D-Daten, die eine Hilfestellung bei der räumlichen Interpretation der Bilder leisten können. So wird im Rahmen der Endoskopie an 3D-Rekonstruktionen gearbeitet, die ein Abbild der Situation im Inneren des Körpers erstellen. Verfahren zur Rekonstruktion von räumlichen Szenen aus planaren Bildsequenzen wurden im Laufe der letzten Jahrzehnte in der Bildverarbeitung entwickelt und finden nun in der medizintechnischen Praxis ihre Anwendung.

Um die vorhandenen Möglichkeiten zu ergänzen, soll in dieser Arbeit geprüft werden, inwieweit sich aus dem Aufnahmeverfahren der Endoskopiebilder selbst ein neuer Erkenntnisgewinn ziehen lässt, indem man die häufig auftretenden Glanzlichter in den Aufnahmen verwendet, um die Normale der Gewebeoberfläche zu bestimmen. Ausgehend von der Annahme, dass die am Endoskop angebrachte Lichtquelle im Falle eines Glanzlichtes orthogonal zur Gewebeoberfläche reflektiert, wird die Position und Orientierung der Kamera ermittelt und damit die Normale bestimmt. Dies kann dazu beitragen, eine robustere Rekonstruktion des Gewebes anzufertigen.

1.1 Endoskopie

Bei der Endoskopie (gr.: $\epsilon\nu\delta\omicron$ in, hinein $\sigma\kappa\omicron\pi\epsilon\iota\nu$ sehen, schauen) handelt es sich um eine 1850/1851 von Hermann von Helmholtz entwickelte Diagnosetechnik in der Medizin. Als *minimal invasive* Behandlungstechnik bietet sie neben den ökonomischen Vorteilen die Möglichkeit einer patientenschonenden Therapie, die sowohl während des operativen Eingriffs als auch bei der Nachbehandlung der Krankheit möglichst geringfügige Komplikationen nach sich ziehen sollte.

Bei einer endoskopischen Behandlung wird eine Kamera mittels einer starren oder beweglichen Röhre durch einen sogenannten *Trokar* in den Körper des Patienten eingeführt. Der *Trokar* ist ein am Ende angespitzter Leitstift mit einem 0,5 bis 12 mm umfassenden Innendurchmesser und wird wahlweise über eine natürliche Körperöffnung oder einen kleinen chirurgischen Schnitt in den Körper geleitet. Durch die entstandene Öffnung wird dann das Endoskop geschleust. Unter Zuhilfenahme von mikrochirurgischen Instrumenten, die direkt an den Endoskopkopf gekoppelt sind, ermöglicht die Endoskopie die *minimal invasive Chirurgie*, wobei die aufgenommenen Bilder in Echtzeit in den Operationssaal übertragen



Bild 1.1: Flexibles und starres Endoskop

werden und den Arzt bei der Behandlung leiten.

Je nach Anwendungszweck gibt es zwei Arten von Endoskopen: starre und flexible. Flexible Endoskope, sogenannte *Flexoskope*, finden Verwendung in der Spiegelung bestimmter Organe und sind somit eher der Diagnostik zuzuordnen. Sie werden meist in natürliche Körperöffnungen eingeführt. Starre Endoskope, auch *Boroskope*, hingegen erlauben minimal invasive chirurgische Eingriffe und erfordern meist einen chirurgischen Schnitt in der Gegend um das Organ.

Die folgende Aufzählung soll ein paar Beispiele geben, welche Körperbereiche sich durch die Endoskopie untersuchen lassen:

- Lunge, Atemwege
- Speiseröhre
- Magen-Darm-Trakt
- Harnwege, weibliches Geschlechtsorgan
- Gelenke
- sonstige Organe wie Leber, Galle, Niere etc.

Ein Endoskop besteht im Wesentlichen aus drei Komponenten: der Lichtquelle, der Kamera und optional, mikrochirurgischen Instrumenten am Endoskopende. Um das Organ

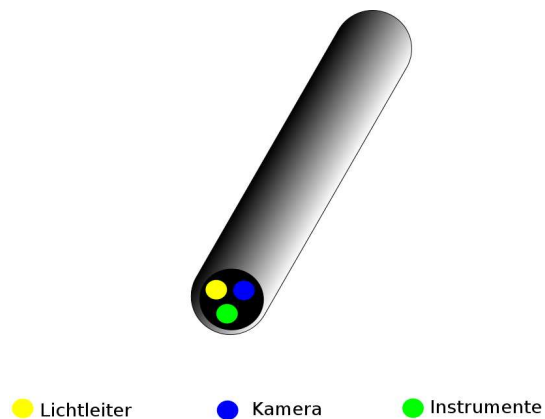


Bild 1.2: Schematischer Querschnitt durch ein Endoskop

zu beleuchten, werden in der Endoskopie Xenon-Lampen eingesetzt, deren ausgestrahltes Licht über Glasfasern zum Endoskopkopf geleitet wird. Dadurch sendet man Licht auf das Gewebe, welches reflektiert und vom CCD-Chip der Kamera aufgenommen wird, der ein Abbild der Szene im Inneren erstellt. Eine direkte Anbringung der Lichtquelle am Endoskopkopf ist nicht möglich, da sonst das sensible Gewebe durch deren Hitze verbrennen könnte. Bild 1.2 zeigt einen schematischen Querschnitt durch das Endoskop.

Im vorgegebenen Fall wurde mit unkalibrierten Bildern aus einem Teilbereich der Endoskopie gearbeitet, der Laparoskopie. Hierbei handelt es sich um Bilder aus dem Bauchbereich des Patienten. Das Endoskop wird durch einen kleinen Schnitt in der Bauchgegend eingeführt, um die Aufnahmen zu machen. Gleichzeitig nimmt der Arzt unter Verwendung spezieller Instrumente den Eingriff vor. Hierbei entstehen allerdings oft Probleme, die sich durch die unnatürliche Sicht auf das zu behandelnde Gewebe ergeben. Durch die Einschränkung des Sichtfeldes, die zum Teil schlechte Bildqualität und die erschwerten Bedingungen im Umgang mit den Geräten wird die gewohnte Arbeitsweise des Arztes stark beeinträchtigt. Demnach werden höchste Anforderungen an den Arzt gestellt, was die Erfahrung mit der Darstellungsweise und das Feingefühl im Umgang mit den Instrumenten betrifft. Zudem können Glanzlichter bei der Zuordnung der Gewebeflächen irritieren.

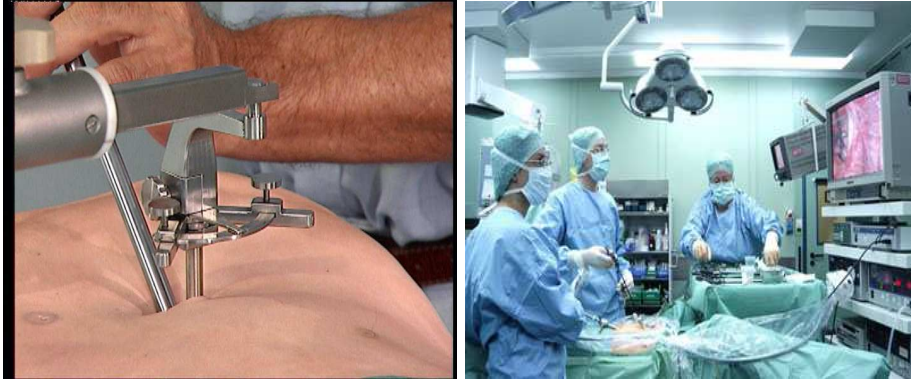


Bild 1.3: Aufnahmen aus dem OP bei einer Laparoskopie

1.2 Aufgabenstellung

In Endoskopiebildern treten häufig Glanzlichter an den Stellen auf, an denen das Licht orthogonal zur Geweboberfläche reflektiert wurde. Diese sind meist störend bei der Behandlung und erschweren die Koordination der Instrumente und den Ablauf zusätzlich, da das darunter liegende Gewebe nicht sichtbar ist. Mittlerweile gibt es Techniken, um Glanzlichter zu eliminieren und das Gewebe zu rekonstruieren, siehe [Vog06] und [Hei03]. Diese Glanzlichter geben jedoch andererseits auch implizit Aufschluss über die Oberflächennormale des betrachteten Gewebes, wenn sich die ursprüngliche Kameraposition und -orientierung ermitteln lässt. Steht sie fest, lässt sich die Oberflächennormale im Mittelpunkt abtragen. Diese kann dann z.B. als ergänzende Information für eine 3D-Rekonstruktion verwendet werden, was zusätzlichen Nutzen bei der Orientierung während dieses 'Ferneingriffs' in den Körper verspricht.

Ziel der vorliegenden Arbeit soll nun sein, weitere 3D-Informationen direkt aus dem verwendeten Aufnahmeverfahren der Endoskopie zu ziehen. Die in der Bildsequenz auftretenden Glanzlichter sollen extrahiert und ihr Mittelpunkt bestimmt werden. Hiernach schätzt man für jedes Bildpaar die *Fundamentalmatrix*, um eine Grundlage für die Positions- und Orientierungsschätzung der Kamera zu erhalten. Anhand der berechneten Fundamentalmatrizen können die internen Kameraparameter approximiert werden. In einem letzten Arbeitsschritt wird dann aus den entstandenen Daten die *Kamerapose* berechnet, wodurch sich die Oberflächennormale des Gewebes rückrechnen lässt.

1.3 Methode

Zur Detektion der Glanzlichter nutzt man ein Schwellwertverfahren. Da sich Glanzlichter durch ihre Helligkeit stark vom zugrundeliegenden Gewebe abheben, lassen sie sich leicht aus ihrem Umfeld separieren. Im Folgenden wird das Bild hierfür in den *HSV-Farbraum* transferiert und in Einzelkanäle zerlegt. Durch das Setzen von Schwellwerten auf Sättigungs- und Dunkelstufenkanal lassen sich die Glanzlichter dann aus der Bildumgebung herausfiltern. Danach werden die gefundenen Glanzlichter in zusammenhängende Einzelflächen aufgeteilt und deren Mittelpunkt errechnet.

Im folgenden Arbeitsschritt sollen Punktkorrespondenzen zwischen den Teilbildern der Sequenz ermittelt werden. In diesem Fall verwendet man zwei Verfahren, *Scale Invariant Feature Transform* und den *Lucas-Kanade Optical Flow* Algorithmus. Die beiden Tracker haben sich in der Praxis für die Punktverfolgung bewährt. Stehen die Punktkorrespondenzen fest, kann man mittels des *RANSAC* die Fundamentalmatrizen der Bilder berechnen. Zuletzt wird aus den Fundamentalmatrizen die Kamerapose geschätzt. Dies geschieht durch eine *Dynamic Hill Climbing* genannte Methode. Hierbei approximiert man unter Annahme bestimmter Grundvoraussetzungen die internen Kameraparameter, die dann durch Einbezug der Fundamentalmatrix die Berechnung der *Essentialmatrix* erlauben. Aus dieser wiederum lassen sich Informationen über die Kamerapose, also deren Rotation und Translation gewinnen. Steht die Kamerapose fest, ergibt sich aus der Normalen zur Bildebene gleichzeitig die Normale des Gewebes und somit dessen geometrische Orientierung an der Stelle des Glanzlichtes.

Die Implementation des geschilderten Verfahrens stützt sich auf Intel's OpenCV Bibliothek, die einen immensen Funktionsumfang für die Bildverarbeitung zur Verfügung stellt.

1.4 Aufteilung der Arbeit

Der Aufbau der Arbeit ist wie folgt: direkt im Anschluss findet sich ein Überblick über die Grundlagen der Entstehung von Glanzlichtern, der Epipolargeometrie und der Autokalibration. Im darauffolgenden Kapitel möchte ich auf bereits vorhandene Lösungsansätze zur Detektion von Glanzlichtern, zur Berechnung der Epipolargeometrie und der Schätzung

der internen Kameraparameter eingehen. In Kapitel 4 werde ich den von mir gewählten Ansatz im Detail veranschaulichen und dessen Funktionsweise näher beschreiben. Kapitel 5 umfasst die Implementation und die experimentell ermittelten Ergebnisse meines Verfahrens. Eine Zusammenfassung und Bewertung der Ergebnisse findet sich in Kapitel 6.

Kapitel 2

Grundlagen

Um einen tieferen Einblick in die Problematik zu erhalten, ist es nötig einige Grundlagen des Bildgebungsprozesses und deren mathematische Modellierung zu erläutern. Dies soll im nun folgenden Kapitel geschehen. Hierbei lehne ich mich an die Ausführungen in [TV98] und [HZ03] an.

Bei der Abbildung einer Szene der realen Welt durch eine Kamera werden mehrere aufeinander aufbauende Transformationen in verschiedenen Koordinatensystemen vorgenommen. Diese Koordinatensysteme, die im Verlauf betrachtet werden, sind folgende:

- das *Weltkoordinatensystem* WCS
- das *Kamerakoordinatensystem* CCS
- das *Bildkoordinatensystem* ICS
- das *Pixelkoordinatensystem* PCS

Beim *Weltkoordinatensystem* handelt es sich um ein dreidimensionales Koordinatensystem, welches als Referenzkoordinatensystem für die restlichen Koordinatensysteme zu verstehen ist. Im *Weltkoordinatensystem* werden die Objektpunkte der dargestellten Szene durch $\mathbf{p}^w = (p_x^w, p_y^w, p_z^w)^T$ dargestellt. Ebenso definiert sich die Position der Kamera anhand des *Weltkoordinatensystems*.

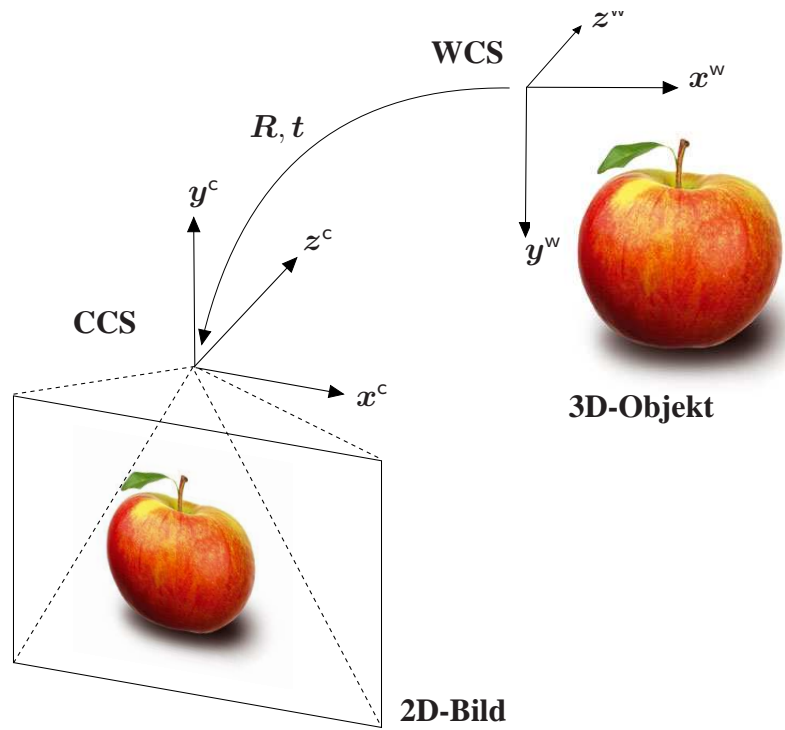


Bild 2.1: Welt- und Kamerakoordinatensystem

Das dreidimensionale *Kamerakoordinatensystem* legt die Position und Orientierung der Kamera fest. Der Ursprung des *Kamerakoordinatensystems* wird durch das optische Zentrum der Kamera vorgegeben, die Blickrichtung verläuft entlang der z-Achse, wobei x- und y-Achse kollinear zu den Bildspalten und -zeilen verlaufen. Punkte im *Kamerakoordinatensystem* lassen sich durch $\mathbf{p}^c = (p_x^c, p_y^c, p_z^c)^T$ repräsentieren.

Der Ursprung des *Bildkoordinatensystems* liegt im sogenannten *Hauptpunkt* $\mathbf{H} = (H_x, H_y)^T$ dem Mittelpunkt des Bildes. x- und y-Achse verlaufen parallel zu den Achsen des Kamerakoordinatensystems. Einen zweidimensionalen Punkt im *Bildkoordinatensystem* bezeichnet man mit $\mathbf{p}^i = (p_x^i, p_y^i)^T$.

Das *Pixelkoordinatensystem* ist ebenfalls zweidimensional und sein Ursprung liegt üblicherweise in der linken oberen Ecke des Bildes. Die Achsen des *Pixelkoordinatensystems* sind parallel zu denen des Bildkoordinatensystems ausgerichtet. Ein Punkt im Bildkoordinatensystem wird durch $\mathbf{p}^p = (p_x^p, p_y^p)^T$ gekennzeichnet.

2.1 Homogene Koordinaten

Homogene Koordinaten stellen eine Erweiterung des *reellen Raums* \mathbb{R}^n dar. Ein Punkt $\mathbf{p} \in \mathbb{R}^n$ wird in den *projektiven Raum* $\mathbb{P}^n \subset \mathbb{R}^{n+1}$ auf den homogenen Punkt $\tilde{\mathbf{p}}$ übertragen, indem man eine zusätzliche Koordinate w ungleich 0 ergänzt:

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix} \longrightarrow \tilde{\mathbf{p}} = \begin{pmatrix} wp_1 \\ wp_2 \\ \vdots \\ wp_n \\ w \end{pmatrix} = \begin{pmatrix} \tilde{p}_1 \\ \tilde{p}_2 \\ \vdots \\ \tilde{p}_n \\ w \end{pmatrix} \quad (2.1)$$

Die Umrechnung von *homogenen Koordinaten* in *karthesische Koordinaten* ergibt sich dann aus $\mathbf{p} = (\tilde{p}_1/w, \tilde{p}_2/w, \dots, \tilde{p}_n/w)^T$. Zwei homogene Punkte $\tilde{\mathbf{p}}$ und $\tilde{\mathbf{q}}$ werden *equivalent* genannt, wenn:

$$\tilde{\mathbf{p}} \cong \tilde{\mathbf{q}} \iff \tilde{\mathbf{p}} = \lambda \tilde{\mathbf{q}} \quad \text{mit} \quad \lambda \in \mathbb{R} \setminus \{0\} \quad (2.2)$$

Equivalenten homogenen Punkte entsprechen also demselben karthesischen Punkt.

Da *homogene Koordinaten* in verschiedenen Koordinatensystemen vorkommen, notieren wir dieses durch einen hochgestellten Index. Beispielsweise stellt $\tilde{\mathbf{p}}^w$ einen homogenen Punkt im Weltkoordinatensystem dar.

2.2 Kamerakoordinatensystem \rightarrow Bildkoordinatensystem

Zur mathematischen Beschreibung der perspektivischen Projektion der Kamerakoordinaten in Bildkoordinaten hat sich in der Bildverarbeitung und Computergraphik das *Lochkameraprinzip* durchgesetzt. Die Annahme des *Lochkameraprinzips* ist sehr gebräuchlich in der Bildverarbeitung, wie man z.B. [TV98], [HZ03], [Zha00] oder auch [Tsa87] entnehmen kann.

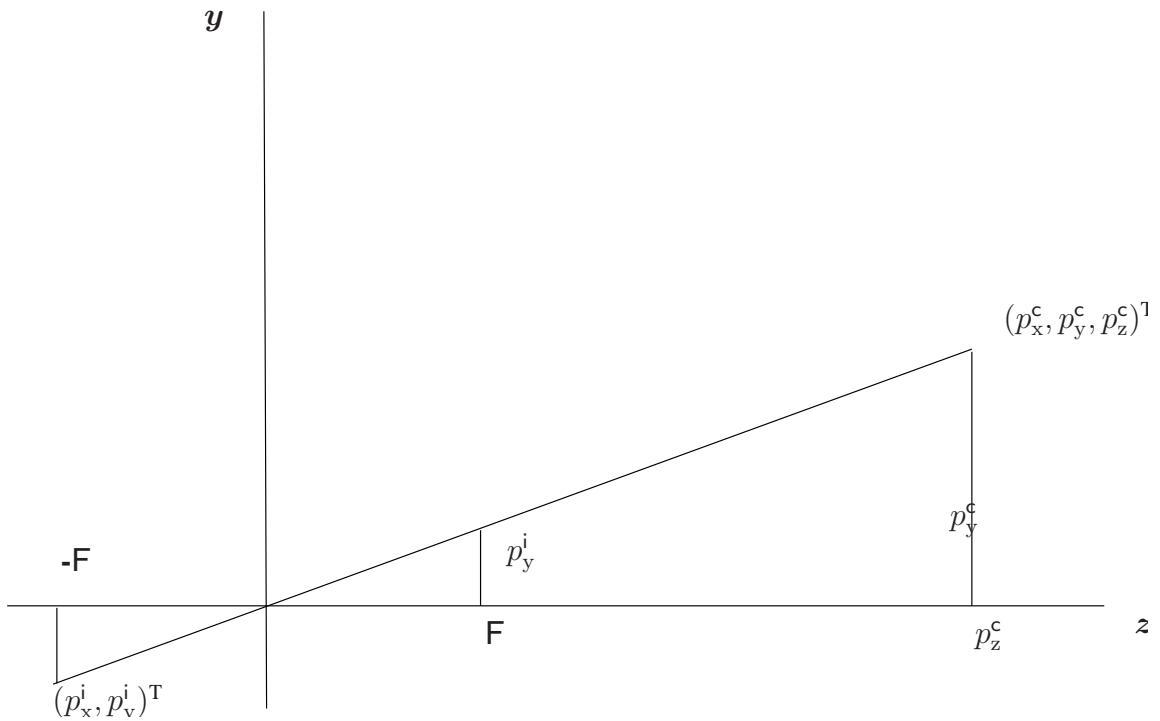


Bild 2.2: Lochkameraprinzip

Durch das *Lochkameraprinzip* wird der Punkt $\mathbf{p}^c = (p_x^c, p_y^c, p_z^c)^T$ in Kamerakoordinaten auf den Bildpunkt $\mathbf{p}^i = (p_x^i, p_y^i)^T$ nach folgender Formel abgebildet:

$$\begin{pmatrix} p_x^i \\ p_y^i \end{pmatrix} = \begin{pmatrix} F p_x^c / p_z^c \\ F p_y^c / p_z^c \end{pmatrix} \mapsto \begin{pmatrix} F p_x^c \\ F p_y^c \\ p_z^c \end{pmatrix} \quad (2.3)$$

F^1 bezeichnet hierbei die Brennweite der Kamera. Unter Verwendung von homogenen Koordinaten lässt sich die Gleichung zu einer linearen Abbildung umformen:

$$\tilde{p}^i = \begin{pmatrix} \tilde{p}_x^i \\ \tilde{p}_y^i \\ \tilde{p}_z^i \end{pmatrix} = \underbrace{\begin{pmatrix} F & 0 & 0 & 0 \\ 0 & F & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{P_{\text{per}}} \begin{pmatrix} \tilde{p}_x^c \\ \tilde{p}_y^c \\ \tilde{p}_z^c \\ 1 \end{pmatrix} \quad (2.4)$$

wobei

$$p^i = \begin{pmatrix} p_x^i \\ p_y^i \end{pmatrix} = \begin{pmatrix} \tilde{p}_x^i / \tilde{p}_z^i \\ \tilde{p}_y^i / \tilde{p}_z^i \end{pmatrix} \quad (2.5)$$

2.3 Bildkoordinatensystem \rightarrow Pixelkoordinatensystem

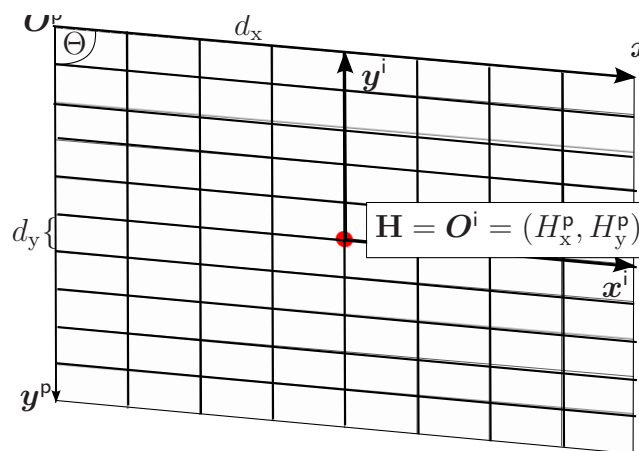


Bild 2.3: Pixelkoordinatensystem

Da die Daten digitaler Bilder im Pixelkoordinatensystem vorliegen, muss auch die Transformation von homogenen Bildkoordinaten \tilde{p}^i in homogene Pixelkoordinaten \tilde{p}^p modelliert werden. In diesem Fall kommt die sogenannte *Kalibrationsmatrix* \mathbf{K} zum Einsatz,

¹Man beachte, daß F hier für die Brennweite steht, während in den weiteren Kapiteln F zur Bezeichnung der Fundamentalmatrix verwendet wird.

die die *intrinsischen Kameraparameter* beinhaltet. *Intrinsische Kameraparameter* umfassen die Eigenschaften der Kamera, die sich bei der Bewegung nicht ändern, also konstant bleiben. Zu ihnen zählen:

- die Koordinaten des *Hauptpunkts*: H_x und H_y
- die *effektiven Brennweiten* in x- und y-Richtung: F_x und F_y
- die *Scherung*: s

Da der Ursprung des Pixelkoordinatensystems bei der Digitalisierung üblicherweise in der linken oberen Ecke des Bildes angesiedelt wird, während der Ursprung des Kamerakoor-
dinatensystems gewöhnlich in der Mitte des Bildes liegt, ist es nötig diese Translation zu modellieren. Diese Translation spiegeln die Koordinaten *Hauptpunkts* $H = (H_x, H_y)$ wieder.

Außerdem sind die Elemente des CCD-Chips in der Regel nicht quadratisch, sondern weisen eine unterschiedliche Breite und Höhe auf. Bezeichnet man nun die Breite eines Sensorelements mit d_x und die Höhe mit d_y so ergeben sich die *effektiven Brennweiten* F_x in x- respektive F_y in y-Richtung anhand von:

$$F_x = F/d_x \text{ und } F_y = F/d_y \quad (2.6)$$

Technische Schwierigkeiten bei der Fabrikation von CCD-Chips können dazu führen, dass die Achsen des Bildkoordinatensystems nicht orthogonal zueinander verlaufen. Man spricht hierbei von einer *Scherung* s der Bildachsen. Sei Θ der Winkel zwischen x- und y-Achse des Bildkoordinatensystems, so erhält man die *Scherung* wie folgt:

$$s = -F_x \tan \Theta \quad (2.7)$$

Diese Eigenschaften werden in der sogenannten *Kalibrationsmatrix* \mathbf{K} zusammengefasst, die die Transformation zwischen homogenen Bildkoordinaten $\tilde{\mathbf{p}}^i = (\tilde{p}_x^i, \tilde{p}_y^i, 1)^T$ in homogene Pixelkoordinaten $\tilde{\mathbf{p}}^p = (\tilde{p}_x^p, \tilde{p}_y^p, 1)^T$ vornimmt.

$$\mathbf{K} = \begin{pmatrix} F_x & s & H_x \\ 0 & F_y & H_y \\ 0 & 0 & 1 \end{pmatrix}$$

$$\tilde{\mathbf{p}}^p = \mathbf{K} \tilde{\mathbf{p}}^i \quad (2.8)$$

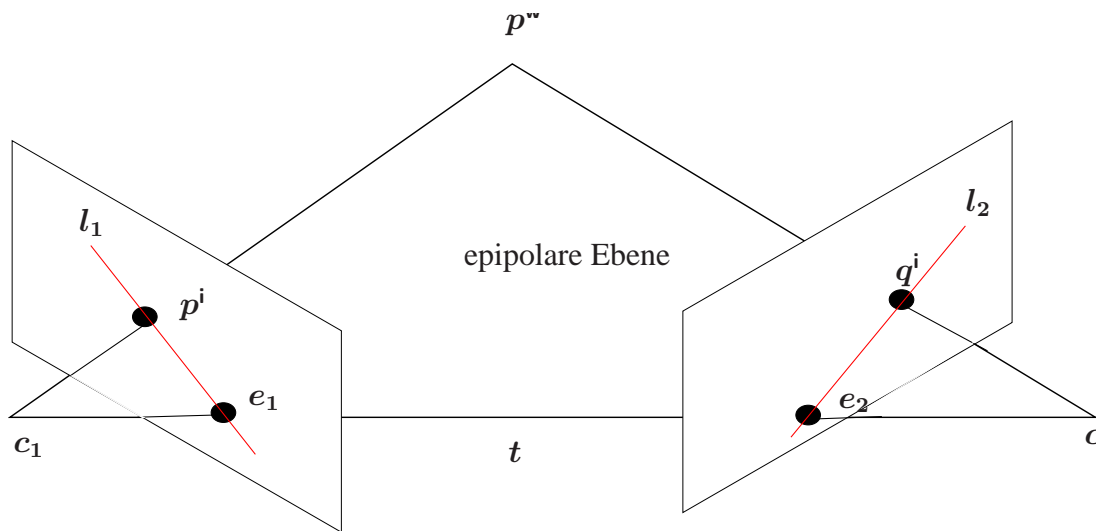


Bild 2.4: Epipolargeometrie

2.4 Epipolargeometrie

Die *Epipolargeometrie* beschreibt die mathematischen Zusammenhänge zwischen zwei Bildern derselben Szene aus verschiedenen Blickrichtungen. Anhand der *Epipolargeometrie* ist es möglich, die Kamerapose, also die relative Ausrichtung der Kameras zueinander, zu schätzen.

Angenommen zwei Kameras mit den Kamerazentren c_1 und c_2 betrachten aus zwei unterschiedlichen Positionen den Weltpunkt $p^w = (p_x^w, p_y^w, p_z^w)^T$ in einer statischen Szene. Dieser Punkt wird auf die Punkte $p^i = (p_x^i, p_y^i)^T$ im linken beziehungsweise $q^i = (q_x^i, q_y^i)^T$ im rechten Bildkoordinatensystem abgebildet. Ohne Beschränkung der Allgemeinheit sei der Ursprung des linken Kamerakoordinatensystems so gewählt, dass er mit dem Ursprung des Weltkamerakoordinatensystems zusammenfällt.

Dann kann man den Sichtstrahl der Kamera zum Objektpunkt, also die Verbindungslinie zwischen dem linken Kamerazentrum c_1 und dem Weltpunkt p^w , im rechten Bild ermitteln und umgekehrt. Die Projektionen dieser Verbindungslinien in den Bildern werden *Epipolarlinien* genannt. Diese Situation ist in Bild 2.4 dargestellt. l_2 ist die epipolare Linie zu $\overline{c_1 p^w}$.

Werden nun mehrere Objektpunkte von beiden Kameras erfasst, so schneiden sich alle

epipolaren Linien im *Epipol* e , dem Schnittpunkt von $\overline{c_1 c_2}$ mit der Bildebene. Der *Epipol* ist somit die Projektion des jeweils anderen Kamerazentrums ins Bild.

Die Abbildung der Weltpunkte vor der Projektion in die Bildebene geschieht mithilfe der sogenannten *Kameramatrix* C^2 . Da wir davon ausgehen, daß die Kamera des linken Bildes im Ursprung des Weltkoordinatensystems liegt, ergibt sich C_1 als:

$$C_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (2.9)$$

C_2 setzt sich aus einer *Rotation* R und einer *Translation* $\mathbf{t} = (t_x, t_y, t_z)^T$ zusammen, die die Transformation des Weltpunktes für die rechte Kamera beschreibt:

$$C_2 = (\mathbf{R} \mid \mathbf{t}) \quad (2.10)$$

Die *Kamerapose* beschreibt also die Orientierung der beiden Kameras zueinander. Diese beiden Parameter nennt man die *extrinsischen Kameraparameter* und es leitet sich die folgenden Eigenschaften her:

$$\mathbf{q}^{cT} \underbrace{[\mathbf{t}]_x \mathbf{R}}_E \mathbf{p}^c = 0 \quad (2.11)$$

mit $[\mathbf{t}]_x = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$

2.5 Essential Matrix

Wie aus den Gleichungen (2.4) und (2.5) hervorgeht, ergibt sich die Projektion der Kamerakoordinaten in homogene Bildkoordinaten durch $\tilde{\mathbf{p}}^i = \mathbf{P}_{\text{per}} \tilde{\mathbf{p}}^c$. Daher folgt mit (2.11):

$$\begin{aligned} (\mathbf{P}_{\text{per}} \tilde{\mathbf{q}}^c)^T \mathbf{E} (\mathbf{P}_{\text{per}} \tilde{\mathbf{p}}^c) &= 0 \\ \tilde{\mathbf{q}}^{iT} \mathbf{E} \tilde{\mathbf{p}}^j &= 0 \end{aligned} \quad (2.12)$$

²In der Literatur wird häufig der Bezeichner \mathbf{P} verwendet, um die Kameramatrizen zu kennzeichnen. Dies wurde aber aus Konsistenzgründen in diesem Dokument vermieden, da es sonst zu Verwechslungen mit der Projektionsmatrix aus 2.2 kommen könnte.

Gleichung (2.12) ist die sogenannte *Epipolarbedingung*, die für *normalisierte Bildkoordinaten* zutrifft. *Normalisierte Bildkoordinaten* ergeben sich durch eine Brennweite $F = 1$. Die Matrix \mathbf{E} wird *Essentialmatrix* genannt. Die *Essentialmatrix* besitzt die folgenden Eigenschaften:

- Die Multiplikation von \mathbf{E} mit einer Konstanten verändert die *Epipolarbedingung* nicht. Daher wird \mathbf{E} entweder durch die Frobeniusnorm $\|\mathbf{E}\|_F = \sqrt{\sum_{i,j} |e_{ij}|^2} = 1$ normiert, oder die Länge des Translationsvektors $\|\mathbf{t}\| = \sqrt{\sum_{i=1}^3 |x_i|^2}$ auf 1 gesetzt.
- Der Rang von \mathbf{E} ist 2, denn sie setzt sich aus dem Produkt einer Rotationsmatrix (Rang 3) und $[\mathbf{t}]_x$ (Rang 2) zusammen.
- \mathbf{E} hat fünf Freiheitsgrade. Drei aus Rotation und drei aus Translation, da die Skalierung der Szene jedoch nicht ermittelt werden kann¹, reduziert sich die Anzahl um 1.
- \mathbf{E} besitzt 3 Singulärwerte, wobei $\sigma_1 = \sigma_2$ und $\sigma_3 = 0$
- \mathbf{E} hängt nur von den extrinsischen Kameraparametern ab.

2.6 Fundamental Matrix

Ersetzt man in Gleichung (2.12) die Bildkoordinaten durch ihre entsprechenden Pixelkoordinaten nach (2.8) sieht man:

$$\tilde{\mathbf{p}}^i = \mathbf{K}^{-1} \tilde{\mathbf{p}}^p \text{ und } \tilde{\mathbf{q}}^i = \mathbf{K}^{-1} \tilde{\mathbf{q}}^p$$

$$\tilde{\mathbf{q}}^{pT} \underbrace{\left((\mathbf{K}^{-1})^T \mathbf{E} \mathbf{K}^{-1} \right)}_F \tilde{\mathbf{p}}^p = 0 \quad (2.13)$$

F ist die sogenannte *Fundamentalmatrix*. Sie hat die nachstehenden Merkmale:

¹Da es sich bei der \mathbf{E} -Matrix um eine projektive Größe handelt, lässt sich nicht sagen, wie groß die betrachtete Szene in der Realität ist. Die \mathbf{E} -Matrix kann aus der Abbildung eines Puppenhauses wie auch aus der eines Wolkenkratzers resultieren.



Bild 2.5: Glanzlichter in einer realen Lapraskopieaufnahme

- F ist nur bis auf die Skalierung eindeutig, es gilt:

$$\tilde{\mathbf{q}}^{\text{pT}} \rho \mathbf{F} \tilde{\mathbf{p}}^{\text{p}} = 0$$

Daher wird durch $\|\mathbf{F}\|_{\text{F}} = \sqrt{\sum_{i,j} |f_{ij}|^2} = 1$ normiert.

- F hat sieben Freiheitsgrade, da sich deren Anzahl wie schon bei der Essentialmatrix durch die Skalierung reduziert.
- F besitzt Rang 2
- F hängt von den extrinsischen und den intrinsischen Kameraparametern ab.
- Alle epipolaren Linien schneiden sich im Epipol.

2.7 Glanzlichtentstehung

Wie eingangs erwähnt kommt es bei endoskopischen Aufnahmen verstärkt zu dem Effekt, dass Glanzlichter entstehen. Ein reales Beispiel zeigt Bild 2.5. Diese Glanzlichter resultieren aus der Tatsache, dass das abgestrahlte Licht des Endoskops auf menschliches Gewebe trifft, welches aus viel Wasser besteht und somit eine sehr glatte Oberfläche aufweist.

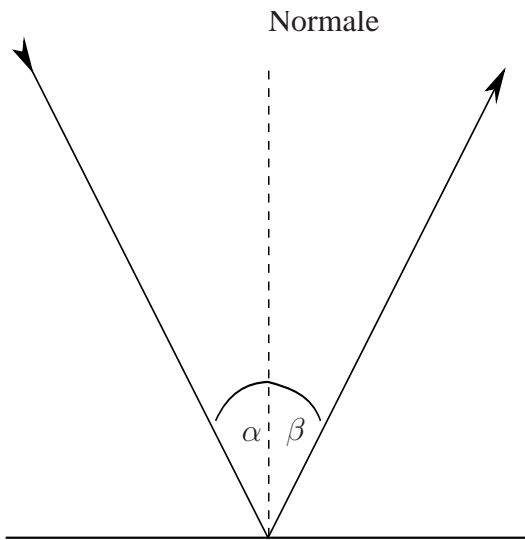


Bild 2.6: Reflexionseigenschaften

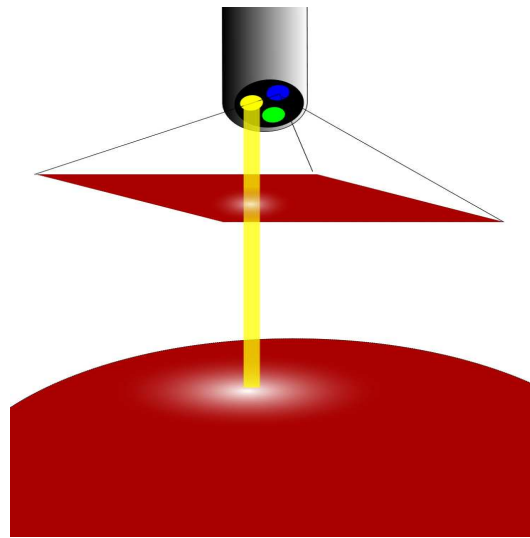


Bild 2.7: Bezug zur Endoskopie

Da das Gewebe oft mit Blut überzogen ist und die Wunde zudem während des Eingriffs gespült wird, bildet sich eine spiegelnde Oberfläche, die das Licht an einigen Stellen reflektiert.

Zur Verdeutlichung der Situation möchte ich etwas näher auf die physikalischen Gesetzmäßigkeiten und deren Bezug zur Aufnahmesituation eingehen. Wenn ein einfallender Lichtstrahl auf eine reflektierende Oberfläche trifft, so entspricht der Einfallswinkel des Lichts dem Ausfallwinkel des reflektierten Lichts gemessen am Lot, also der Oberflächennormale. Diese Situation lässt sich Bild 2.6 entnehmen. Der Lichtstrahl trifft von links oben auf die spiegelnde Oberfläche und wird in die Gegenrichtung reflektiert, die Winkelgrößen des Einfallswinkels α und des Ausfallwinkels β sind exakt gleich.

Im Zusammenhang mit der Endoskopie bedeutet das Folgendes: Wenn das vom Sensorkopf abgestrahlte Licht orthogonal zur Gewebeoberfläche auftritt, wird es direkt in Richtung zur Lichtquelle zurückgeworfen. Man spricht hierbei von einer *Retroreflexion*. Aufgrund der großenbedingten Nähe des Lichtleiters und der Kamera am Sensorkopf entstehen dadurch die Glanzlichter bei der Aufnahme. Daher ist davon auszugehen, dass wo immer ein Glanzlicht im aufgenommenen Bild auftritt, die Blickrichtung des Endoskops senkrecht zur Gewebeoberfläche war und somit Blickrichtung und Oberflächennormale

identisch sind (vgl.: Bild 2.7). Zusammenfassend ergibt sich folgende Relation, die bei ausgesendetem Licht vom Endoskopiekopf und Glanzlicht im Bild zutrifft:

$$\text{Einfallswinkel} = \text{Ausfallswinkel} \quad \Rightarrow \quad \text{Oberflächennormale} = \text{Blickrichtung} \quad (2.14)$$

Diese Formel kann anhand des schematischen Aufbaus des Endoskops (Bild 1.2) verifiziert werden.

Demnach lässt sich die Oberflächennormale aus den Glanzlichtern rekonstruieren, wenn man die Epipolargeometrie errechnet. Die Verfahren zur Schätzung der Epipolargeometrie werden allgemein unter dem Begriff *Struktur aus Bewegung* zusammengefasst und ich möchte in den nun folgenden Abschnitten näher auf sie eingehen.

Kapitel 3

Stand der Technik

Das folgende Kapitel soll einen Überblick über bereits vorhandene Techniken zur Bearbeitung der Teilschritte des Verfahrens geben. Da sich die behandelte Thematik in dieser Form nicht in der Literatur finden läßt, werden nur Konzepte vorgestellt, die als verwandt angesehen werden können. Das Verfahren ist in seiner Vorgehensweise allerdings einzigartig und als Innovation zu betrachten.

3.1 Punktverfolger

Für die Punktverfolgung gibt es verschiedene Tracker, die eingesetzt werden können. Die unter dem Oberbegriff *Optical Flow* zusammengefassten Tracker werden in den folgenden Dokumenten beschrieben:

Berthold K.P. Horn und Brian G. Schunck stellen in [HS81] einen Ansatz vor, der die Pixelbewegungen durch Helligkeitsunterschiede detektiert und stellen die Bedingung auf, dass die Bewegungen relativ gleichmäßig verlaufen sollten. Sie bewerten ihren Algorithmus allerdings selbst als anfällig gegenüber Rauschen.

Eine weitere Möglichkeit stellen Bruce D. Lucas and Takeo Kanade in [KL81] vor. Dieses gradientenbasierte Verfahren detektiert robust Features innerhalb bestimmter Pixelnachbarschaften. Erweitert wurde es in [Bou] durch eine Auflösungshierarchie. Dieses Verfahren fand in der Diplomarbeit Anwendung, da sich damit auch große Positionsänderungen

tracken lassen können.

Ergänzend zum *Optical Flow* gibt es einfache Methoden wie den in [Can86] beschriebenen Kantendetektor, wobei man nach der Kantendetektion über eine Blockmatchingmethode eine Bildkorrelation erstellt und er somit zum Featuretracken verwendet werden kann. Dies erweist sich aber bei großen Bewegungen als nicht robust.

Rotations- und skalierungsinvariante Featuretracker bieten den Vorteil, auch bei massiven Unterschieden in den Aufnahmemodalitäten gute Ergebnisse zu liefern. Zu den bekanntesten gehört der ebenfalls verwendete *SIFT* aus [Low04] von Dawid Lowe und dessen Erweiterung *SURF*, der die Laufzeit des Algorithmus weiter reduziert und in [BTVG06] näher erläutert wird.

3.2 Autokalibrierung

Standardverfahren zur Autokalibration finden sich in [HZ03]. Diese befassen sich mit der Bestimmung der absoluten Quadrik und über die Kruppa-Gleichungen (siehe: [FLM92]). Beide Ansätze erfordern allerdings eine Triangulation von Punkten, die wir vermeiden wollten.

Eine effektive und intuitiv verständliche Methode stellen Gerhard Roth and Anthony Whitehead in [RW04] vor. Sie verwenden die Singulärwerte der Essentialmatrix als Kriterium zur Minimierung für die Schätzung der intrinsischen Kameraparameter. Diese Technik baut auf dem von Michael de la Maza and Deniz Yuret entwickelten Minimierungsalgorithmus *Dynamic Hill Climbing* auf, den [dIMY94] beschreibt. Die positive Bewertung der Autoren führte zu einer Übernahme in dieser Arbeit. Eine weitere Technik zur Autokalibration mittels Singulärwertzerlegung führen Manolis Lourakis und Rachid Deriche in [LD99] an. Hierbei ergibt sich eine Vereinfachung der Kruppa-Gleichungen. Die Autoren sprechen von einer akzeptablen Genauigkeit ihrer Ergebnisse.

3.3 Glanzlichtdetektion

Für die Glanzlichtdetektion gibt es die Möglichkeit, anhand des *dichromatischen Reflexionsmodells* Glanzlichtkanten in Bildern zu detektieren. Dieses fand in [Kli01] und [Vog06] seine Anwendung.

Nach [Sch04] zu urteilen, ist die zuverlässige Detektion von Glanzlichtern aufgrund von Fehlklassifikationen des *dichromatischen Reflexionsmodells* allerdings besser durch eine HSV-Farbraumanalyse zu erreichen, dementsprechend entschieden wir uns für diese.

3.4 Normalenbestimmung

Ein verwandtes Projekt stellt sich in [BBS04] vor. Die Autoren beschreiben darin, wie sie durch einen homographiebasierten Ansatz versuchen, die Ausrichtung eines Flugzeugs im Raum zu bestimmen. Sie benutzen dazu Felder in der aufgenommenen Landschaft und deren Deformation während des Flugs. Zur Dekomposition kommt ein von Bill Triggs entwickeltes Verfahren zum Einsatz, näheres dazu findet sich in [Tri98]. Die Analogie zu den Feldern ist allerdings im endoskopischen Umfeld nicht gegeben.

Kapitel 4

Eigener Ansatz

Im folgenden Kapitel wird näher auf die Details des Verfahrens eingegangen. Diese entsprechen zum großen Teil den in [HZ03] verwendeten Methoden zur Schätzung der Epipolargeometrie. Da vorwiegend unkalibrierte Bilddatensätze verfügbar waren, war es nötig, eine Autokalibrierung vorzunehmen. Hierzu wurde ein alternativer Optimierungsalgorithmus aus dem Feld der genetischen Algorithmen verwendet, um die intrinsischen Kameraparameter aus den unkalibrierten Bildfolgen zu schätzen. Abbildung 4.1 verdeutlicht den Ablauf des Verfahrens.

4.1 Punktverfolger

Wie bereits erwähnt, erfordert die Berechnung der Fundamentalmatrix mindestens sieben ideale Punktepaare. Zum Finden dieser Korrespondenzen zwischen jeweils zwei aufeinander folgenden Bildern wurden zwei unterschiedliche Punktverfolger eingesetzt, um die Robustheit der Autokalibrierung zu testen. Die Punktverfolger werden hier nur kurz vorgestellt, weitere Informationen können aber den angegebenen Referenzen entnommen werden.

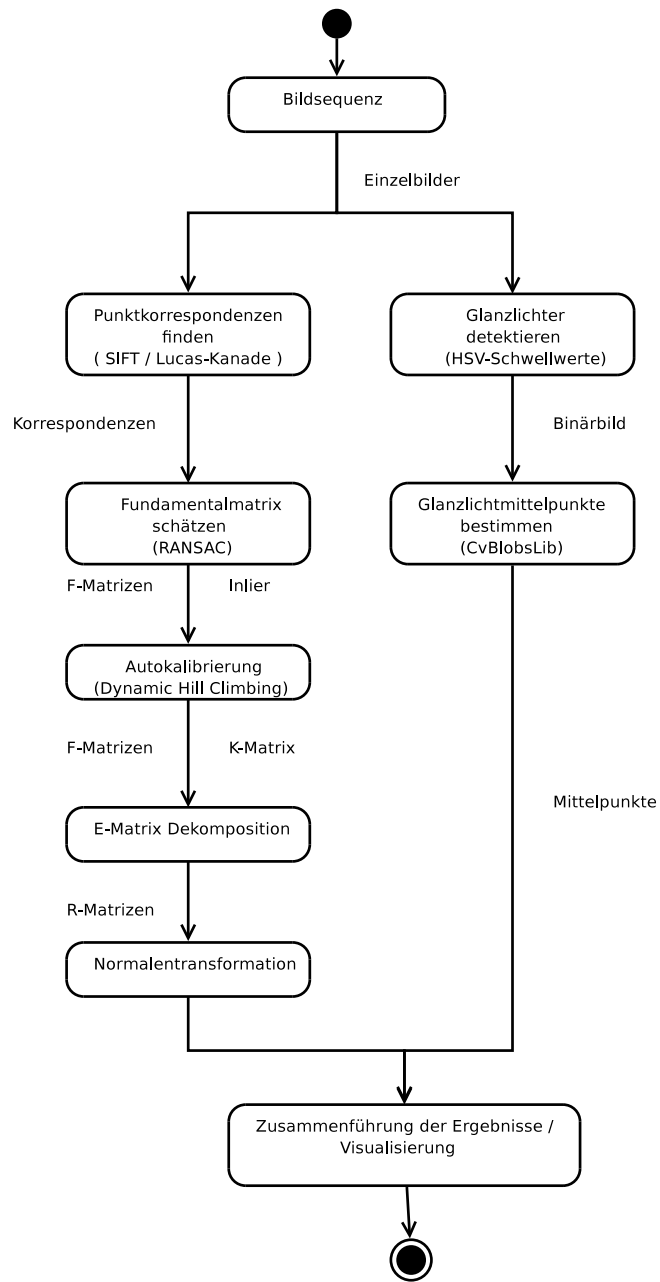


Bild 4.1: Schematischer Ablauf des Verfahrens

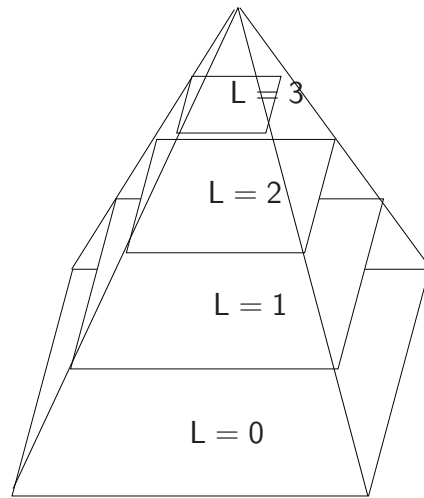


Bild 4.2: Auflösungshierarchie mit Pyramiden

4.1.1 Lucas-Kanade mit Pyramiden

Zum Verfolgen der Punkte wurde die OpenCV-Implementation des Lucas-Kanade Featuretrackers benutzt. Das von Bruce D. Lucas und Takeo Kanade entwickelte Verfahren versucht, die Korrespondenzen innerhalb der Bilder anhand einer gradientenbasierten Suche in Pixelnachbarschaften zu finden. Bei der Suche wird jeweils ein Fenster von $r \times r$ Pixeln im ersten Bild aufgespannt und im zweiten Bild gesucht. Bei großen Bewegungen ist intuitiv klar, dass man die Fenstergröße vergleichsweise groß wählen müsste, um diese zu detektieren. Hierbei ergeben sich aber potentiell drei Probleme, nämlich:

- kleine Bewegungen werden verschluckt
- entgegengesetzte Pixelbewegungen werden geglättet
- das Integrationsfenster vergrößert sich, dementsprechend der Aufwand der Berechnung

Um dem entgegenzuwirken, wird das Bild in mehreren Auflösungsstufen abgetastet, um zum einen die Größe des Integrationsfensters klein zu halten, zum anderen hat eine Auflösungshierarchie den Vorteil, auch große Pixelbewegungen innerhalb der Bilder subpixelgenau zu erkennen. Die Suche startet auf der Ebene mit der geringsten Auflösungsstufe

und wird iterativ fortgesetzt bis zur untersten Auflösungsebene, dem Bild selbst, wobei das Ergebnis der nächsthöheren Ebene immer als Initiallösung verwendet wird.

Die gradientenbasierte Suche hat allerdings den Nachteil, dass in Bereichen, die wenig Textur aufweisen, auch keine Korrespondenzen gefunden werden können.¹

Näheres zum Algorithmus und zu seiner Implementation lässt sich in [KL81] und [Bou] erfahren.

4.1.2 SIFT

Einen anderen Weg zum Finden der Korrespondenzen geht der von David Lowe entworfene Algorithmus *SIFT*. *SIFT* (engl.: **S**cale **I**nvariant **F**eature **T**ransform) hat sich aufgrund seiner Robustheit in weiten Teilen der Bildverarbeitung als Punktracker durchgesetzt.

SIFT ist rotations- und skalierungsinvariant, d.h. selbst wenn in den Aufnahmen zweier Bilder eine Rotation oder eine Skalierung vorkommt, sollten die Korrespondenzen zuverlässig gefunden werden. *SIFT* speichert die Position markanter Punkte in einem Bild unabhängig von den anderen Bildern der Sequenz und weist ihnen einen 128-wertigen Merkmalsvektor zu. Diese Merkmalsvektoren werden dann mit denen in einem anderen Bild verglichen, indem deren Differenz gebildet wird. Diejenigen Vektoren, die jeweils die geringste Differenz untereinander aufweisen, werden als korrespondierende Punkte in zwei Bildern betrachtet.

Ebenso, wie beim Lucas-Kanade-Tracker findet leider auch *SIFT* kaum Korrespondenzen in wenig texturierten Bildgebenden. Dies hängt mitunter auch damit zusammen, dass die zugrundeliegende Technik zum Bestimmen der Features gradientenbasiert ist.

Die genaue Vorgehensweise von *SIFT* wird in [Low04] beschrieben.

4.2 Fundamentalmatrixschätzung

Nachdem die korrespondierenden Punkte in den Bildern gefunden sind, ist es möglich, die Fundamentalmatrix zu berechnen. Sie bildet die Grundlage zur Schätzung der intrinsi-

¹Auf diesen Mangel des Algorithmus wird hingewiesen, da sich daraus bei der Berechnung der Fundamentalmatrix Probleme ergeben können, auf diese wird in Kapitel 5 eingegangen.

schen Kameraparameter und der Essentialmatrix, die in den darauffolgenden Abschnitten besprochen wird.

4.2.1 Berechnung der Fundamentalmatrix

Anhand von korrespondierenden Punkten in zwei Bildern derselben Szene lässt sich die Fundamentalmatrix ermitteln. Da sie nur sieben Freiheitsgrade besitzt, reichen theoretisch² sieben Korrespondenzpaare, um eine exakte Berechnung vorzunehmen. Für jedes der sieben Punktepaare kann man eine Gleichung der Art

$$\tilde{\mathbf{q}}^{\text{pT}} \mathbf{F} \tilde{\mathbf{p}}^{\text{p}} = 0$$

aufstellen. Eine Umformung in Vektornotation ergibt:

$$\begin{pmatrix} \tilde{p}_x^{\text{p}} & \tilde{q}_x^{\text{p}} & \tilde{p}_y^{\text{p}} & \tilde{q}_y^{\text{p}} & \tilde{q}_x^{\text{p}} & \tilde{p}_x^{\text{p}} & \tilde{q}_y^{\text{p}} & \tilde{p}_y^{\text{p}} & \tilde{q}_y^{\text{p}} & \tilde{q}_x^{\text{p}} & \tilde{p}_x^{\text{p}} & \tilde{p}_y^{\text{p}} & 1 \end{pmatrix} * \mathbf{f} = 0$$

$$\text{mit } \mathbf{f} = (F_{11} \ F_{12} \ F_{13} \ F_{21} \ F_{22} \ F_{23} \ F_{31} \ F_{32} \ F_{33})^{\text{T}}$$

Trägt man nun die sieben gewonnenen Zeilenvektoren in eine Matrix \mathbf{A} ein, erhält man:

$$\mathbf{A} * \mathbf{f} = 0 \tag{4.1}$$

Eine Lösung für \mathbf{f} ergibt sich durch die *Singulärwertzerlegung*³ SVD (engl.: **S**ingular **V**alue **D**ecomposition). Wenn \mathbf{A} sich mittels *Singulärwertzerlegung* zu einem Produkt $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^{\text{T}}$ aufteilen lässt, so ist \mathbf{f} der letzte Spaltenvektor von \mathbf{V} .

Hierbei gilt es allerdings zu beachten, dass die Punktkorrespondenzen nicht alle in einer Ebene liegen, also linear unabhängig sind. Sonst ist das Gleichungssystem schlecht konditioniert und die Komponenten der \mathbf{F} -Matrix lassen sich nicht berechnen. Überprüfen lässt

²“Theoretisch” deswegen, weil von idealen Punktkorrespondenzen ausgegangen wird. In der Praxis erweist sich das Finden solcher Korrespondenzen als schwierig, da die Bilder verrauscht sind und zudem Linsenverzerrungen und Beleuchtungsdifferenzen zu berücksichtigen sind. Dies führt bei der Analyse zu sogenannten Ausreißern, also der falschen Zuordnung der Punkte.

³Die Singulärwertzerlegung ist eine effiziente Methode zur Lösung linearer Gleichungssysteme, näheres zur Singulärwertzerlegung findet man in [TB97]

sich dies durch eine Ranganalyse der Matrix A anhand der SVD. Die Anzahl der Singulärwerte in Σ ungleich Null kennzeichnet den Rang von A . Dieser muss exakt 7 betragen.

4.2.2 RANSAC

Leider kommt es bei der Korrespondenzanalyse von realen Bildern häufig zu einer Fehlzurordnung der Punkte zweier aufeinander folgender Bilder einer Sequenz. Diese sogenannten *Outlier* resultieren aus Rauschen in den Bilddaten, veränderten Beleuchtungsbedingungen oder zu starken Bewegungen des Sensors. Um dieser falschen Zuordnung entgegenzuwirken wurde eine robuste Methode zur Berechnung der Fundamentalmatrix verwendet: *RANSAC*.

Unter *RANSAC* (engl.: **R**andom **S**ampling **C**onsensus) versteht man die iterative Berechnung der Fundamentalmatrix mittels des in 4.2.1 vorgestellten Verfahrens. So werden in einer Iteration aus den ermittelten Punktkorrespondenzen jeweils zufällig sieben Punktepaare ausgewählt und die Fundamentalmatrix geschätzt. Danach wird geprüft, wieviele der Gesamtkorrespondenzen die Epipolarbedingung $\tilde{q}^p \mathbf{F} \tilde{p}^p = 0$ erfüllen. In der Praxis verwendet man hierbei einen Schwellwert, da das exakte Einhalten der Epipolarbedingung nur durch ideale Punkte erreicht werden kann. Diejenigen Punkte, die die Epipolarbedingung erfüllen, werden als *Inlier* klassifiziert. Der Algorithmus iteriert nun solange, bis entweder eine bestimmte Anzahl an *Inliern* erreicht ist oder eine vorher definierte Iterationszahl überschritten wird. Als Ergebnis erhält man die Fundamentalmatrix zurückgeliefert, die die meisten *Inlier* aufweist. Man spricht hier auch von der Fundamentalmatrix mit dem größten *Support*. Diese wird im weiteren Verlauf des Verfahrens verwendet.

Sei P die Wahrscheinlichkeit, dass alle Punktkorrespondenzen *Inlier* in mindestens einem der Iterationsschritte sind und ϵ die erwartete Zahl an *Outliern*, dann lässt sich die Anzahl der maximalen Iterationen m wie folgt berechnen:

$$m = \frac{\log(1 - P)}{\log(1 - \epsilon^7)} \quad (4.2)$$

RANSAC zeichnet sich durch seine Robustheit in Bezug auf die verwendeten Punktkorrespondenzen aus, selbst in dem Fall, dass über 50% der gefundenen Punkte Fehlkorrespondenzen sind.

4.3 Autokalibrierung

Eine weitere Voraussetzung, um die Essentialmatrix und damit die Kamerageometrie zu schätzen, ist die Kenntnis der intrinsischen Kameraparameter. Diese werden wie in 2.3 bereits erklärt in der Kalibrationsmatrix \mathbf{K} zusammengefasst. Im Rahmen der Diplomarbeit wurden durchweg unkalibrierte Bildfolgen verwendet, was die Zuhilfenahme eines Autokalibrierungsalgorithmus erforderlich machte.

Die Entscheidung fiel dabei auf ein von Anthony Whitehead und Gerhard Roth entwickeltes Verfahren, welches in [RW04], [RW02a] und [RW02b] erläutert wird und in [GMR02] seine Umsetzung fand. Die intuitive Verständlichkeit des Algorithmus, die Abweichung von den in [HZ03, pp. 458-501] vorgestellten Standardverfahren und die guten Ergebnisse der genannten Quellen motivierten zu einer Übernahme der Technik in dieser Arbeit.

Ein Autokalibrierungsalgorithmus schätzt die intrinsischen Kameraparameter aus den Bildern selbst, er erfordert also keine umständlichen Kalibrationsaufnahmen oder a-priori Wissen über deren Beschaffenheit.

4.3.1 Verfahren - zwei Bilder

Die Kalibrationsmatrix \mathbf{K} hat nach 2.3 den folgenden Aufbau:

$$\mathbf{K} = \begin{pmatrix} F_x & s & H_x \\ 0 & F_y & H_y \\ 0 & 0 & 1 \end{pmatrix} \quad (4.3)$$

Erfahrungsgemäß erweist es sich laut [RW04] für die Visualisierung als sinnvoll, nur einige Parameter der Kalibrationsmatrix zu schätzen. So wird im Allgemeinen davon ausgegangen, daß die Produktionstechnik bei der Herstellung von Kameras so präzise arbeitet, dass die Scherung s der CCD-Sensoren nicht mehr vorhanden ist. Auch die Koordinaten des Hauptpunktes sind üblicherweise in der Mitte des Bildkoordinatensystems angesiedelt. Somit reduziert sich die Suche der Kameraparameter auf die effektiven Brennweiten F_x

und F_y und die zu bestimmende Matrix auf:

$$\mathbf{K} = \begin{pmatrix} F_x & 0 & \frac{width}{2} \\ 0 & F_y & \frac{height}{2} \\ 0 & 0 & 1 \end{pmatrix} \quad (4.4)$$

Wie man [RW04] entnehmen kann, befinden sich die möglichen Werte für F_x und F_y im Intervall zwischen 1 und 5000 Pixeln.

Während die Fundamentalmatrix die Zusammenhänge zwischen den Pixelkoordinaten zweier Bilder beschreibt und somit abhängig von intrinsischen und extrinsischen Kameraparametern ist, beschreibt die Essentialmatrix die Zusammenhänge zwischen den Bildkoordinaten. Somit ist die Essentialmatrix nur von den extrinsischen Parametern abhängig. Da die Bilder im behandelten Fall aus Sequenzen stammen, kann man weiterhin davon ausgehen, dass \mathbf{K} bei allen Bildern identisch ist und sich die intrinsischen Parameter während der Aufnahme nicht ändern. Die Essentialmatrix kann man dann wie folgt aus der Fundamentalmatrix errechnen:

$$\mathbf{E} = \mathbf{K}^T \mathbf{F} \mathbf{K} \quad (4.5)$$

Eine wichtige Eigenschaft der Essentialmatrix geht aus 2.5 hervor. Ihre ersten beiden Singulärwerte σ_1 und σ_2 müssen gleich sein, während der letzte Singulärwert σ_3 Null ist.

$$\sigma_2 - \sigma_1 = 0 \quad (4.6)$$

Diese Eigenschaft kann man sich zunutze machen, indem man sie für die Autokalibrierung der Bilder verwendet. Aus (4.6) ergibt sich ein Minimierungskriterium für die zu bestimmende \mathbf{K} -Matrix:

$$1 - \frac{\sigma_2}{\sigma_1} \quad (4.7)$$

Um nun die \mathbf{K} -Matrix zu ermitteln, geht man wie folgt vor:

1. Setze $F_x = a * 4999 + 1$ und $F_y = b * 4999 + 1$ mit $a, b \in [0, 1]$
2. Berechne $\mathbf{E} = \mathbf{K}^T \mathbf{F} \mathbf{K}$

3. Bestimme die Singulärwertzerlegung von E
4. Werte $1 - \frac{\sigma_2}{\sigma_1}$ aus

Hierbei werden jetzt a und b durch einen Minimierungsalgorithmus solange variiert, bis das Minimum der Funktion (4.7) gefunden ist. Im Minimum liegt dann die bestapproximierte K -Matrix für die beiden Bilder.

4.3.2 Verfahren - Sequenz

Da nicht nur zwei Bilder betrachtet werden, sondern gleich eine ganze Reihe von M aufeinanderfolgenden Bildern, lässt sich das oben genannte Verfahren noch erweitern. Die Fundamentalmatrix je zweier Sequenzbilder hat durch die Berechnung mit RANSAC (vgl. 4.2.2) eine ihr zugeordnete Supportgröße, die Anzahl der Inlier, die die Epipolarbedingung erfüllen. Diese Supportgröße spiegelt das *Vertrauen* in die Korrektheit der Fundamentalmatrix wieder. Eine Fundamentalmatrix mit vielen Inliern wird mit großer Sicherheit die Bildgeometrie besser wiedergeben, als eine mit geringem Support. Dieses *Vertrauen* drücken wir mathematisch durch eine Gewichtung ω_i aus:

$$\omega_i = 1 - \frac{\text{Inliers}_i}{\text{MaxInliers}} \quad \text{mit} \quad i = 0 \dots M - 1 \quad (4.8)$$

Inliers_i bezeichnet hierbei die Supportgröße der Fundamentalmatrix zwischen den Bildern i und $i + 1$. MaxInliers gibt die maximale Supportgröße aller gefundenen Fundamentalmatrizen an.

Wie man sieht, bewegt sich der Gewichtungsfaktor ω im Bereich zwischen 0 und 1. Die Gewichtung verhält sich antiproportional zum *Vertrauensfaktor*, da das Ziel des Algorithmus die Minimierung einer Funktion ist. Eine niedrige Gewichtung lässt dementsprechend auf ein hohes *Vertrauen* in die Korrektheit der Fundamentalmatrixberechnung schließen und umgekehrt.

Für die Sequenz gilt es dann, die folgende Funktion zu minimieren:

$$\sum_{i=0}^{M-1} \omega_i \left(1 - \frac{\sigma_2}{\sigma_1}\right) \quad (4.9)$$

Dies geschieht analog zum Fall mit zwei Bildern.

4.3.3 Dynamic Hill Climbing

Zur Minimierung von (4.9) wurde eine von Michael de la Maza und Deniz Yuret entwickelte Technik verwendet, das sogenannte *Dynamic Hill Climbing*, kurz DHC. Dieser Minimierungsalgorithmus leitet sich aus dem Umfeld der evolutionären Algorithmen ab. Evolutionäre Algorithmen versuchen, durch Simulation natürlicher Vorgänge wie Selektion und Mutation Modelle der Wirklichkeit abzubilden. In diesem Zusammenhang heißt das, die errechnete K -Matrix wird als Individuum betrachtet, welches mutiert und dann mit den anderen Individuen verglichen wird. Sobald ein Individuum als *bestes Individuum* seiner Generation klassifiziert ist, startet die Suche neu, wobei hierbei als Ausgangspunkt für die weitere Mutation ebendieses Individuum gewählt wird. Diese Suche läuft solange, bis das *beste Individuum* aller Generationen gefunden ist.

Bei Minimierungsalgorithmen besteht allgemein die Gefahr, in einem lokalen Minimum *steckenzubleiben*. Der Algorithmus terminiert dann und liefert das lokale Minimum zurück, obwohl das globale Minimum der Funktion noch nicht gefunden ist. DHC hat hier einige Möglichkeiten, um ein solches Szenario zu verhindern. Das Verfahren wird im Folgenden sinngemäß nach [dlMY94] und [dlMY03] erläutert.

Dynamic Hill Climbing besteht aus einer äußeren und einer inneren Schleife. Die äußere Schleife bestimmt die Startpunkte für die Suche der inneren Schleife nach dem lokalen Minimum. Durch die äußere Schleife wird sichergestellt, dass der Suchraum der Parameter uniform abgetastet wird, wodurch einer Konzentration der Suche auf einen bestimmten Bereich entgegengewirkt wird.

Wenn wir die Dimensionalität des Suchraums mit n bezeichnen, so verwaltet die innere Schleife $2 * n$ Vektoren, von denen n eine orthogonale Basis im Suchraum beschreiben. Die anderen n Vektoren sind deren Negation. Zusätzlich gehören zur inneren Schleife noch zwei Vektoren, die beschreiben, in welche Richtung des Suchraums sich der Algorithmus bewegt hat und die Koordinaten des aktuell betrachteten Punktes \boldsymbol{x} . Im Fall der Autokalibrierung besteht der Vektor \boldsymbol{x} aus den Komponenten a und b , also $\boldsymbol{x} = (a, b)^T$.

In jeder Iteration l wird nun der Vektor mit der größten Länge \boldsymbol{v}_l herausgesucht. Dann wird das Ergebnis der Minimierungsfunktion $f(\boldsymbol{x}_l + \boldsymbol{v}_l)$ überprüft. Stellt sich heraus, dass $f(\boldsymbol{x}_l + \boldsymbol{v}_l) < f(\boldsymbol{x}_l)$ ist, so gilt:

$$\boldsymbol{x}_{l+1} = \boldsymbol{x}_l + \boldsymbol{v}_l \quad (4.10)$$

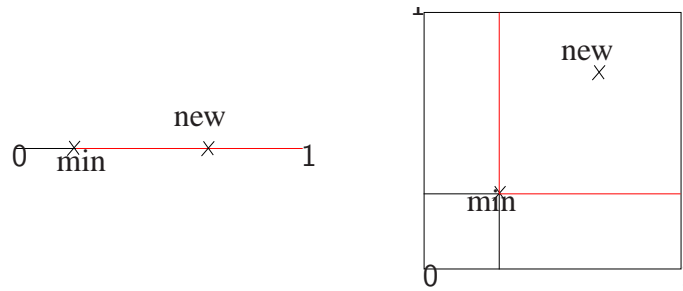


Bild 4.3: Finden des neuen Initialisierungsvektors für die innere Schleife

Danach wird zur nächsten Iteration fortgeschritten. Wählt man zweimal hintereinander v_1 als Versatzvektor, so wird seine Länge verdoppelt.

$$v_{1+1} = 2 * v_{1+1} \quad \text{wenn} \quad v_{1+1} == v_1 \quad (4.11)$$

Ist allerdings $f(x_1 + v_1) \geq f(x_1)$ so wird v_1 halbiert.

Die Iterationen laufen solange, bis die Größe aller Vektoren einen vorher gesetzten Schwellwert θ_{DHC} unterschreiten. Dann terminiert die innere Schleife und x_1 wird als lokales Optimum zurückgeliefert. Die beiden zusätzlichen Vektoren d_1^1 und d_1^2 speichern die letzten beiden unterschiedlichen Richtungen, in die sich der Algorithmus bewegt hat, und deren Negation:

$$d_1^1 = v_{1-2} + v_{1-1} \quad \text{mit} \quad v_{1-2} \neq v_{1-1} \quad (4.12)$$

$$d_1^2 = -d_1^1 \quad (4.13)$$

Wenn die innere Schleife terminiert, merkt sich die äußere Schleife die Position des lokalen Minimums x_1 und sucht einen neuen Startpunkt, der im Suchraum möglichst weit entfernt von den bereits gefundenen Minima liegt.

Wir finden einen weit entfernten Punkt, indem der Suchraum mittels der bereits gefundenen Minima in Intervalle eingeteilt wird und wählen dann einen zufälligen Punkt im größten Teilintervall. Diese Situation ist in Bild 4.3 exemplarisch für den 1D- bzw. 2D-Fall dargestellt. Der neue Startpunkt wird aus den rot gekennzeichneten Intervallen selektiert. Hierdurch lässt sich die uniforme Abtastung des Suchraums gewährleisten.

Als Ergebnis enthält die äußere Schleife am Ende des Algorithmus alle gefundenen lokalen Minima, aus denen sich anhand der Funktionswerte leicht das globale Minimum der Funktion finden lässt.

4.4 Essentialmatrix Dekomposition

Nach Ablauf der Minimierung erhält man also die benötigten Kameraparameter zur Schätzung der Essentialmatrix. Die Essentialmatrix wird durch die gefundenen F -Matrizen und die Kameramatrix K zwischen zwei aufeinanderfolgenden Bildern berechnet. Hiernach folgt die Dekomposition der E -Matrix in eine Rotationsmatrix R und einen Translationsvektor t , die die Geometrie zwischen den Ansichten der beiden Bilder repräsentieren (vgl. Bild 2.4 und Gleichung (2.11)). Der Translationsvektor t gibt bei der Dekomposition nur die Richtung der Translation wieder, nicht die absolute Länge. Die Skalierung der Szene ist unbestimmt und kann erst durch aufwendigere Techniken wie den trifokalen Tensor bestimmt werden.⁴

4.4.1 Dekomposition - Zwei Bilder

Betrachten wir die vorerst die Situation zweier aufeinander folgender Bilder der Sequenz zum Zeitpunkt t und $t + 1$ und deren Fundamentalmatrix F_t . Die Essentialmatrix ergibt sich dann, wie in (4.5) bereits erwähnt durch:

$$E_t = K^T F_t K \quad (4.14)$$

Die Essentialmatrix E_t ist zusammengesetzt aus einer Rotationsmatrix R_t und der Kreuzproduktmatrix $[t_t]_x$.

$$E_t = [t_t]_x R_t \quad (4.15)$$

Nun kommt ein Verfahren zum Einsatz, das von Hartley und Zisserman in [HZ03, pp. 257-259] zur Dekomposition der E_t -Matrix in R_t und $[t_t]_x$ vorgeschlagen wird. Zunächst ist es

⁴siehe dazu auch [HZ03, pp. 364-390]

nötig, zwei Hilfsmatrizen \mathbf{W} und \mathbf{Z} zu definieren:

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

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

Jetzt läßt sich die normierte Essentialmatrix mittels Singulärwertzerlegung aufspalten in:

$$\mathbf{SVD}(\mathbf{E}_t) = \mathbf{U}_t \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{V}_t^T \quad (4.16)$$

Dann gibt es folgende Möglichkeiten für \mathbf{R}_t und $[\mathbf{t}_t]_x$:

$$[\mathbf{t}_t]_x = \mathbf{U}_t \mathbf{Z} \mathbf{U}_t^T \quad (4.17)$$

$$\mathbf{R}_t = \begin{cases} \mathbf{U}_t \mathbf{W} \mathbf{V}_t^T = \mathbf{R}_{t1} \\ \mathbf{U}_t \mathbf{W}^T \mathbf{V}_t^T = \mathbf{R}_{t2} \end{cases} \quad (4.18)$$

Da Rotationsmatrizen eine Determinante von 1 besitzen müssen, wird diese Bedingung geprüft und im Fall einer Determinante von -1 invertiert.

$$\mathbf{R}_{t1} = -\mathbf{R}_{t1} \quad \text{wenn} \quad \det(\mathbf{R}_{t1}) == -1 \quad (4.19)$$

$$\mathbf{R}_{t2} = -\mathbf{R}_{t2} \quad \text{wenn} \quad \det(\mathbf{R}_{t2}) == -1 \quad (4.20)$$

Somit ergeben sich noch vier Möglichkeiten für die Kameramatrix \mathbf{C}_2 (vgl. Kapitel 2.4):

$$\mathbf{C}_2 = \begin{cases} (\mathbf{R}_{t1} \mid \mathbf{t}_t) \\ (\mathbf{R}_{t2} \mid \mathbf{t}_t) \\ (\mathbf{R}_{t1} \mid -\mathbf{t}_t) \\ (\mathbf{R}_{t2} \mid -\mathbf{t}_t) \end{cases} \quad (4.21)$$

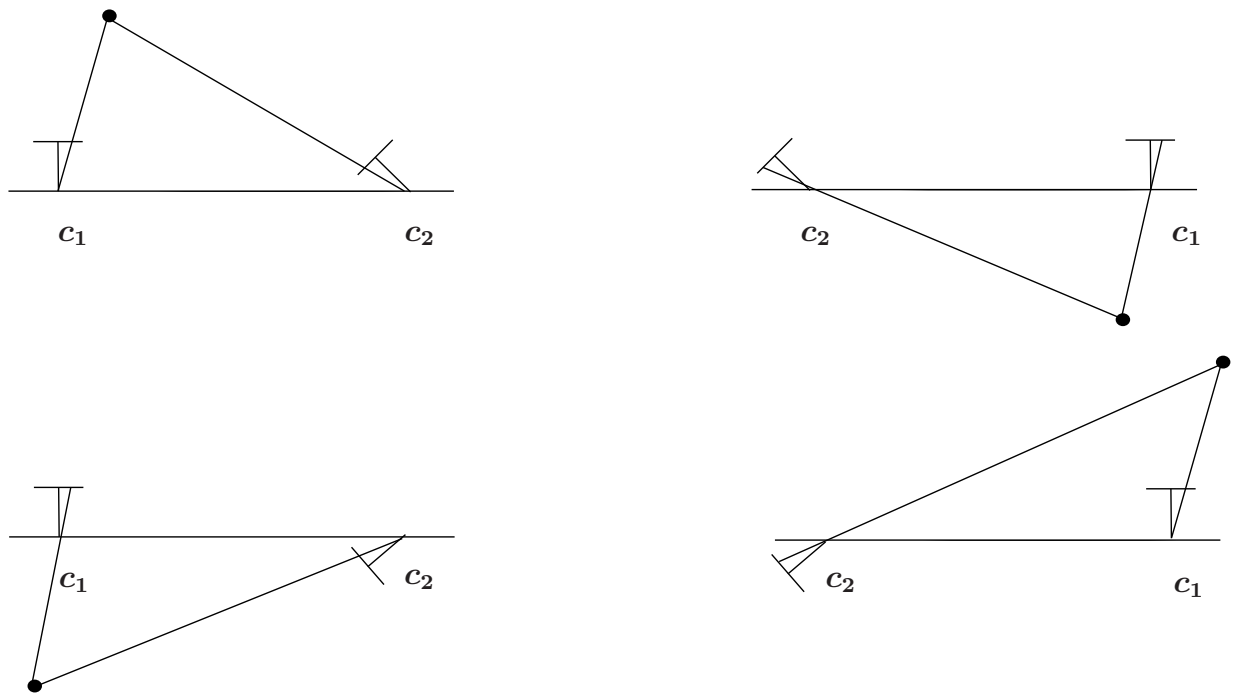


Bild 4.4: Die vier möglichen Faktorisierungen von \mathbf{E} . Rechts-Links: Invertierung von \mathbf{t}_t .
 Oben-Unten: \mathbf{R}_{t_1} bzw. \mathbf{R}_{t_2}

Diese vier Möglichkeiten sind in Bild 4.4 dargestellt. Man erkennt deutlich die Invertierung des Translationsvektors \mathbf{t}_t und die Anwendung der beiden Rotationsmatrizen. Außerdem wird ersichtlich, daß es nur eine Möglichkeit für die korrekte Lösung gibt: Nur in einer der vier Konfigurationen liegt der betrachtete Punkt **vor** beiden Kameras.

Um nun die korrekte Lösung zu finden reicht es aus, eines der gefundenen Korrespondenzpaare rückzuprojizieren und zu prüfen, ob der Punkt vor beiden Kameras liegt, also eine positive z-Koordinate aufweist⁵.

⁵Auf das Verfahren zur Rückprojektion wird hier nicht näher eingegangen. Für weitere Informationen zur Rückprojektion von Punkten sei auf [HZ03, pp 311-324] verwiesen.

4.4.2 Dekomposition - Sequenz

Die Dekomposition für eine Sequenz von M Bildern läuft analog zum Fall mit zwei Bildern. Für jedes aufeinanderfolgende Bildpaar berechnet man die zugehörige Essentialmatrix und zerlegt sie in Rotation und Translation. Somit erhält man insgesamt $M - 1$ Rotationsmatrizen und Translationsvektoren. Die Translationsvektoren spielen in unserem Fall für die Normalenbestimmung der Bildebene keine Rolle, da sich durch eine Skalierung die Rotationsmatrizen nicht ändern und allein diese ausschlaggebend für die relative Orientierung der Bildebene sind.

4.5 Normalenbestimmung

Ist die korrekte Lösung für die Rotationsmatrix R_t gefunden, lassen sich die Normalen zur Bildebene bestimmen. Dies wird nun exemplarisch für die beiden ersten Bilder der Sequenz, I_0 und I_1 zum Zeitpunkt $t = 0$ einmal vorgeführt. Hierfür sei eine neue Notation für die Normalen eingeführt: n_t^m . Das Subskript gibt hierbei an, zu welchem Bild die Normale gehört, das Superskript macht deutlich, in welchem Bild die Normale angezeigt wird.

I_0 wird bei der Essentialmatrixzerlegung die Kameramatrix C_1 und I_1 die Kameramatrix C_2 zugordnet:

$$C_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$C_2 = (R_0 \mid t_0)$$

Nach [HZ03, pp 158-164] ergibt sich Normale n_0^0 zu C_1 definitionsgemäß als der Vektor, in dessen Richtung die Kamera schaut und somit bei einer angenommenen Verschiebung in den Ursprung des Weltkoordinatensystems als:

$$n_0^0 = (0, 0, 1)^T \tag{4.22}$$

Die Normale \mathbf{n}_1^0 zur Bildebene von C_2 errechnet sich dann als:

$$\mathbf{n}_1^0 = \mathbf{R}_0 (0, 0, 1)^T \quad (4.23)$$

Möchte man nun die Normale von I_0 im Bild im Bild I_1 darstellen, so muss man \mathbf{R}_0 transponieren. Die Normalentransformation wird also invertiert:

$$\mathbf{n}_0^1 = \mathbf{R}_0^T (0, 0, 1)^T \quad (4.24)$$

$$\mathbf{n}_1^1 = (0, 0, 1)^T \quad (4.25)$$

Für die weitere Darstellung der Normalen \mathbf{n}_t^m in beliebigen Bildern m einer Sequenz mit M Bildern lässt sich folgende Gleichung aufstellen:

$$\mathbf{n}_t^m = \left(\prod_{i=m-1}^t \mathbf{R}_i^T \right) * (0, 0, 1)^T \quad \text{mit} \quad m = 1 \dots M, t = 0 \dots M - 1 \quad (4.26)$$

Anzumerken bleibt, dass alle Transformationen immer relativ zum ersten Bild der Sequenz zu sehen sind. Das erste Bild dient als Grundlage aller Berechnungen und definiert die Richtung seiner Normalen zum Zeitpunkt $t = 0$ anhand der Kameramatrix ($\text{Id}_3 \mid (0, 0, 0)^T$) als $(0, 0, 1)^T$. Da keine Trackinginformationen der Kamera vorliegen, ist die absolute Ausrichtung im Raum also unklar.

4.6 Glanzlichtdetektion

Zum Detektieren der Bilder wurde ein Schwellwertverfahren im HSV-Farbraum angewendet. Diese einfache Methode hat sich den Ergebnissen von [Sch04] zufolge als robuste Technik erwiesen, um Glanzlichter in endoskopischen Aufnahmen zu detektieren. Zwar werden noch weitere Techniken evaluiert, diese erreichen jedoch nicht die Stabilität und Effizienz der HSV-Farbraumanalyse.

4.6.1 HSV-Schwellwertanalyse

Der HSV-Farbraum ist ein System zur Beschreibung von Farben. Im HSV-Farbraum werden Farben mittels eines *Farbwinkels*, der *Sättigung* und der *Dunkelstufe* definiert. Der

Farbwinkel (engl. hue) liegt im Bereich von 0 bis 360 Grad und bestimmt den Farbton. Die *Sättigung* (engl. saturation) gibt den Weißanteil der Farbe an. Gesättigte Farben erscheinen intensiver, während ungesättigte Farben beim Betrachter einen verblassten Wahrnehmungseindruck hervorrufen. Der Sättigungswert einer Farbe bewegt sich zwischen 0% (ungesättigt) und 100% (vollgesättigte, reine Farbe). Die *Dunkelstufe* (engl. value) steht für die Helligkeit der Farbe und wird ebenso wie die Sättigung durch einen Prozentwert zwischen 0 und 100 spezifiziert.

Wie eingangs erwähnt, heben sich Glanzlichter in endoskopischen Bildern durch ihre starke Helligkeit im Vergleich zum umliegenden Gewebe ab. Mittels des HSV-Farbmodells und einem Schwellwertverfahren lassen sie sich daher leicht detektieren. Hierzu interpretiert man die Glanzlichter als ungesättigte, helle Stellen im Bild.

Zur Detektion der Glanzlichter werden die Endoskopiebilder zuerst in den HSV-Farbraum transformiert. Nun untersucht man die einzelnen Kanäle des transformierten Bildes. Da die Farbtoninformation für das Glanzlicht infolge der Überbelichtung des Gewebes keine Rolle spielt, erfolgt die Untersuchung auf Glanzlichter nur auf dem Sättigungs- und dem Dunkelstufenkanal. Gesucht werden die Stellen im Bild, an denen zugleich eine geringe Sättigung und eine hohe Dunkelstufe vorhanden ist. Dies geschieht mit zwei Schwellwerten, die für beide Kanäle getrennt festgelegt werden. Hierbei ist besonderes Augenmerk auf die Auswahl geeigneter Werte für die Schwellen zu legen, da bei unsachgemäßer Anwendung des Schwellwertverfahrens entweder einerseits zu viele helle Bereiche als Glanzlicht klassifiziert werden, die keine sind, oder andererseits Glanzlichter verloren gehen können.

Seien $S(x, y)$ der Wert des Sättigungskanals und $V(x, y)$ der Wert des Dunkelstufenkanals an der Pixelposition x, y und θ_S respektive θ_V die gewählten konstanten Schwellwerte für die beiden Kanäle, so verdeutlichen die folgenden Gleichungen deren Anwendung:

$$f_S(S(x, y)) = \begin{cases} 1, & \text{wenn } S(x, y) \leq \theta_S \\ 0, & \text{andernfalls} \end{cases}, S(x, y) \in [0,100], \theta_S \in [0,100] \quad (4.27)$$

$$f_V(V(x, y)) = \begin{cases} 1, & \text{wenn } V(x, y) \geq \theta_V \\ 0, & \text{andernfalls} \end{cases}, V(x, y) \in [0,100], \theta_V \in [0,100] \quad (4.28)$$

Als Ergebnis erhält man zwei Binärbilder, deren weiße Bereiche jeweils die Stellen mit niedriger Sättigung beziehungsweise hoher Dunkelstufe kennzeichnen. Diese kann man nun durch ein logisches UND miteinander verknüpfen, wodurch wiederum ein Binärbild entsteht, in dem die Farbinformationen zusammengeführt werden:

$$f_{\text{HIGHLIGHT}}(x, y) = f_S(S(x, y)) \wedge f_V(V(x, y)) \quad (4.29)$$

Das resultierende Bild codiert also die detektierten Glanzlichter im ursprünglichen Bild durch weiße Pixel, der Rest bleibt schwarz.

Ein Phänomen, welches beim vorgeschlagenen Verfahren allerdings auftritt, ergibt sich aus der nicht uniformen Beleuchtung des Gewebes durch das Endoskop und die Wahl eines konstanten Schwellwertes: die zusammenhängende Glanzlichtfläche kann in kleine Teilflächen zerfallen. Dies stellt insofern ein Problem dar, als dass im nächsten Schritt des Verfahrens das Zentrum der Glanzlichtfläche ermittelt werden soll. Um diesem Zerfall entgegenzuwirken, setzt man morphologische Filter ein. Das Bild wird mehrfach dilatiert und erodiert, um die entstandenen Lücken zu schließen.

4.6.2 Bestimmen des Zentroids

Um später die Normale an der Stelle des Glanzlichts zu kennzeichnen ist es wichtig den Mittelpunkt, auch *Zentroid* genannt, der Glanzlichtfläche zu kennen. Die Glanzlichtflächen werden im Binärbild als zusammenhängende weiße Pixel gekennzeichnet. Diese zusammenhängenden Pixel bilden eine Punktmenge P mit L Punkten. Der i -te Punkt in P sei mit $p^i = (p_x^i, p_y^i)$ bezeichnet. Nun kann man nach dem folgenden Schema vorgehen, um den Zentroid $z = (z_x, z_y)$ des Glanzlichtes zu finden:

1. Bestimmen der x-Komponente des Zentroids:

Addiere alle x-Komponenten der Punkte und Teile durch die Gesamtanzahl der Punkte.

$$z_x = \frac{1}{L} \sum_{i=1}^L p_x^i \quad (4.30)$$

2. Bestimmen der y-Komponente des Zentroids:

Addiere alle y-Komponenten der Punkte und Teile durch die Gesamtanzahl der Punkte.

$$z_y = \frac{1}{L} \sum_{i=1}^L p_y^i \quad (4.31)$$

Kapitel 5

Experimente und Ergebnisse

5.1 Implementetation

Das vorgestellte Verfahren ist in C++ implementiert. Hierbei wurde in weiten Teilen auf die Funktionalität der OpenCV-Routinen zurückgegriffen. Zur einfachen Anpassung der Parameter des Programms wird eine Konfigurationsdatei im XML-Format eingelesen, die das Programm steuert. So ist es möglich, bestimmte Teile abzuschalten, falls nur eine reduzierter Funktionsumfang notwendig ist. Es wurde ein QT-Interface zur Justierung der Schwellwerte für die HSV-Farbraumanalyse entwickelt. Als Punkttracker kamen sowohl die in OpenCV vorhandene Implementetation des pyramidalen Lucas-Kanade Algorithmus zum Einsatz, als auch die Implementetation des SIFT-Featuretrackers von Rob Hess¹. Die Geometrieanalyse verläuft anhand der von OpenCV gelieferten Daten. Zur Identifikation der Glanzlichter wurde auf Routinen der OpenCV-nahen CVBlobsLib zurückgegriffen. Eine detaillierte Dokumentation zum Programm, den verwendeten Bibliotheken und Systemvoraussetzungen findet sich im Anhang.

¹<http://web.engr.oregonstate.edu/~hess/index.html>

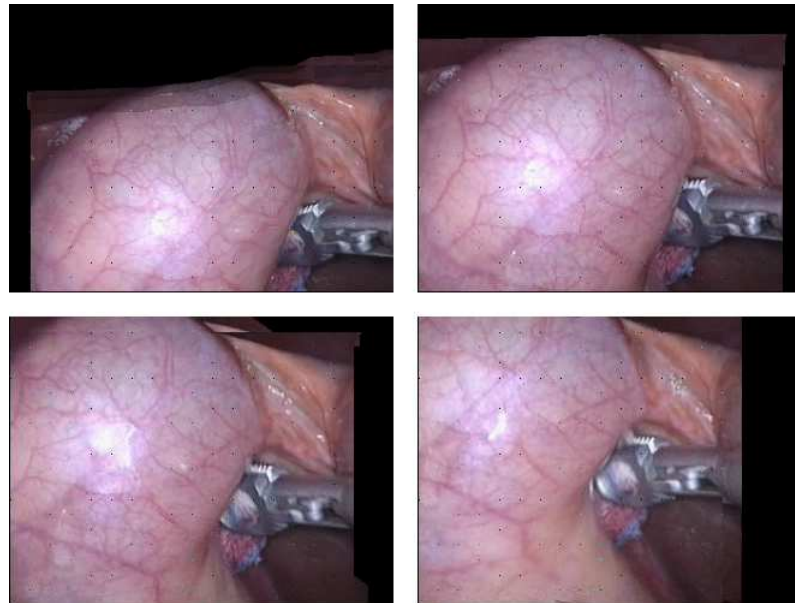


Bild 5.1: Die Frames 1, 30, 60 und 90 der Gallensequenz

5.2 Versuchsaufbau und Durchführung

5.2.1 Testbilder

Zur Evaluation des Systems kamen vier Testsequenzen zum Einsatz. Die erste Sequenz zeigt eine Reihe von Bildern während eines laparoskopischen Eingriffs an der Galle. Zu dieser Testsequenz lagen leider keine Kalibrationsdaten vor, sie konnte allerdings erfolgreich zum Testen des HSV-Schwelwertverfahrens genutzt werden. Die Sequenz besteht aus 99 Bildern der Größe 320x240 Pixel. Abbildung 5.1 zeigt einige ausgewählte Frames. Die zweite Sequenz wurde aus [RW04] übernommen, um die Ergebnisse der Autokalibrierung abzugleichen². Sie zeigt das Schloss von Leuven aus verschiedenen Blickwinkeln. Die 28 Bilder sind jeweils 768x576 Pixel groß. Hier war die Brennweite [RW04] zu entnehmen, wodurch ein direkter Vergleich möglich war. Ein paar Bilder sind in Abbildung 5.2 dargestellt.

Desweiteren wurden freundlicherweise von Peter Decker noch zwei Sequenzen aus seiner

²Herunterzuladen gab es sie unter: <http://www.cs.unc.edu/marc/>



Bild 5.2: Die Frames 0, 9, 18 und 27 der Schlossequenz

Diplomarbeit [Dec07] zum Testen der Autokalibrierung zur Verfügung gestellt. Die eine zeigt eine Tasse, aufgenommen mit unterschiedlichen Kamerastandpunkten. Es handelt sich um acht Bilder der Größe 1024x768 Pixel. Die andere ist eine synthetisch in Maya gerenderte Sequenz, die eine Parkhausszene darstellt. Diese Sequenz umfasst fünf Bilder mit 1024x768 Pixeln. Bei beiden waren Ground-Truth-Daten zur Kalibration vorhanden, somit eigneten sie sich sehr gut für die Evaluation der Autokalibration. Diese Bilder sieht man in Abbildung 5.3 und 5.4.



Bild 5.3: Die Tassensequenz

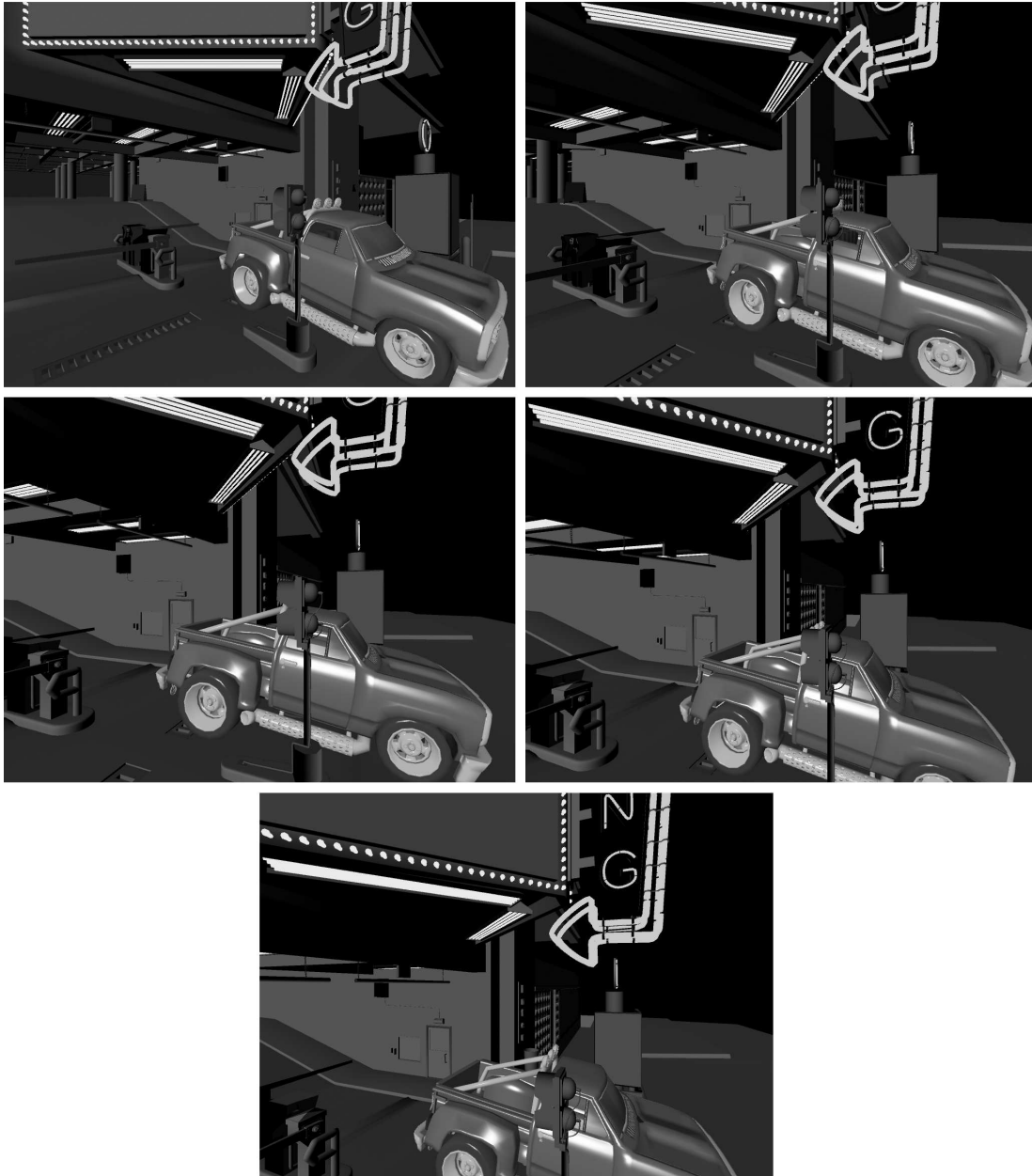


Bild 5.4: Die gerenderte Mayasequenz

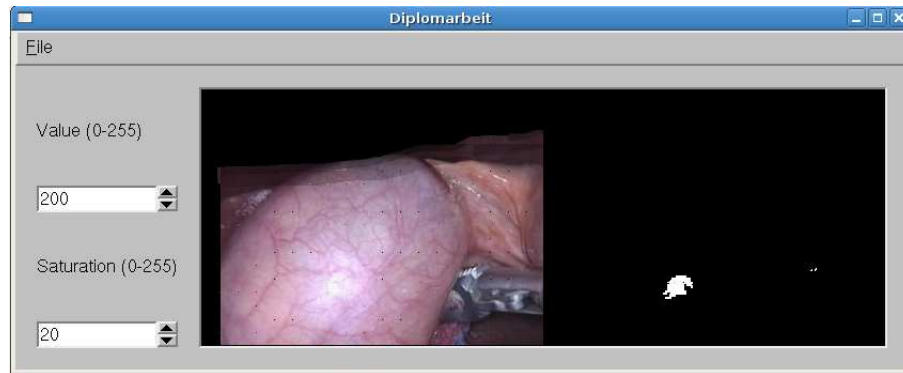


Bild 5.5: QT-Interface zum Justieren der HSV-Schwellwerte

5.2.2 Glanzlichtdetektion

Zum Setzen der Schwellwerte wurde ein QT-Interface implementiert, da diese situativ je nach Beleuchtungsbedingungen angepasst werden müssen und es keine allgemeingültige Vorgabe für adäquate Schwellwerte gibt. Hierzu sieht sich der Anwender als erstes dem in Bild 5.5 dargestellten GUI gegenüber.

Im linken Frame wird immer das erste Bild der Sequenz eingeblendet. Es besteht die Möglichkeit, die Schwellwerte auf dem Sättigungs- und Dunkelstufenkanal getrennt voneinander zu justieren, wobei das Ergebnis im rechten Frame betrachtet werden kann. Somit sind die Auswirkungen einer Schwellwertänderung direkt erkennbar. Sind die Parameter der Sequenz entsprechend eingestellt, so kann man das Interface verlassen und die Schwellwerte werden zur Glanzlichtanalyse für die Gesamtsequenz übernommen.

Getestet wurde das Verfahren vorwiegend mit der Gallensequenz. Wenn die Glanzlichter gefunden sind, können sie als zusammenhängende Einzelflächen, man spricht hierbei von *Blobs*, durch Routinen der *CvBlobsLib* identifiziert werden und ihr Mittelpunkt wird errechnet.

5.2.3 Punktverfolger

Anschließend wird direkt nach Punktkorrespondenzen gesucht. Hierbei gilt zu beachten, dass man zwei verschiedene Punkttracker verwenden kann, was sich über die Konfigu-

rationsdatei festlegen lässt. Im Zusammenhang mit der Gallensequenz wurde der Lucas-Kanade Tracker benutzt, der auf den OpenCV-Methoden *cvGoodFeaturesToTrack* und *cvCalcOpticalFlowPyrLK* aufbaut. Bei der Tassen- und der gerenderten Sequenz kam SIFT zum Einsatz. Bei der Schlosssequenz hingegen wurden beide Tracker getestet. Eine Begründung für die unterschiedliche Anwendung der Tracker findet sich in 5.3. Für den Lucas-Kanade-Tracker wurden folgende Parameter übergeben:

- Fenstergröße: 9x9 Pixel
- Abbruchkriterium: maximal 100 Iterationen oder Unterschreiten des Schwellwerts von 0.1

Für den SIFTTracker wurde als Schwellwert für die Distanz der Merkmalsvektoren ein Schwellwert von 0.49 festgelegt.

5.2.4 Fundamentalmatrixschätzung

Anhand der gewonnenen Punktkorrespondenzen wurde die Fundamentalmatrix mittels *cvFindFundamentalMat* geschätzt. Als Methode wurde *CV_FM_RANSAC_ONLY* gesetzt, was die Berechnung der F -Matrix mit RANSAC erlaubt. Eine Abweichung von 0.5 Pixeln von der Epipolarbedingung war für die Klassifikation als Inlier zulässig. Weiterhin forderten wir eine 99%-ige Wahrscheinlichkeit, dass zumindest eines der Samplesets nur Inlier enthält.

5.2.5 Autokalibrierung und Essentialmatrixdekomposition

Nach der Schätzung der F -Matrizen wurde die Brennweite f und damit die K -Matrix ermittelt. Es wurden 250 Iterationen des Dynamic-Hill-Climbing-Algorithmus durchgeführt, wie in [RW04] vorgeschlagen. Als initiale Schrittgröße wurde 1 gewählt und als minimale Vektorlänge 0.1.

Nach Ablauf der Autokalibration konnten durch die gefundene K -Matrix die entsprechenden E -Matrizen aus den F -Matrizen errechnet werden. Diese wurden nach dem in 4.4.1 vorgestellten Verfahren in ihre R -Matrizen zerlegt.



Bild 5.6: Auswirkungen der Schwellwerte auf die gefundenen Glanzlichter(v.l.o.n.r.u.): Originalbild, zu hohe Sättigung, angepasste Schwellwerte, zu niedrige Dunkelstufe

5.3 Ergebnisse

5.3.1 Glanzlichtdetektion

Die Glanzlichtdetektion funktioniert, dem visuellen Eindruck nach zu urteilen, recht gut. Die Auswirkungen unangepasster Schwellwerte sind in Bild 5.6 exemplarisch dargestellt. Wie man sieht, besteht die Gefahr, entweder eine zu kleine oder aber eine zu große Fläche zu detektieren. Bei zu kleinen Flächen zerfällt das Glanzlicht in Teilflächen. Dies ist insofern kritisch, als dass diese durch die sich anschließenden morphologischen Operationen verlorengehen könnten.

Zu groß gewählte Flächen ergeben schlichtweg eine Fehlklassifizierung von Gewebe als Glanzlicht. Somit ist bei der Wahl der Schwellwerte besonderes Augenmerk auf die Resultate des Justierungsinterface zu legen, was nicht zuletzt für diesen Zweck implementiert wurde. Als angepasste Schwellwerte für die Gallesequenz ergaben sich 20 für den



Bild 5.7: Detektierter Schwerpunkt (rot) und Bounding Box des Glanzlichts

Dunkelstufen- und 200 für den Sättigungskanal.

Bei der Einteilung in *Blobs* wurden nur Flächen mit mehr als 200 Pixeln betrachtet. Dieser Schritt wurde der Einfachheit halber für die Evaluation des Verfahrens mit der gegebenen Gallesequenz gegangen, um eine weitere Fehlklassifikation auszuschließen. Wie man in Bild 5.7 erkennt, wurde das primär in der Mitte angesiedelte, große Glanzlicht innerhalb der Sequenz verfolgt und dessen Schwerpunkt bestimmt. Durch die Bestimmung des Schwerpunktes ergibt sich eine zuverlässige Positionsschätzung in den weiteren Bildern der Sequenz.

5.3.2 Punkttracking und Fundamentalmatrixschätzung

Für die Autokalibration und die Geometrieschätzung ist eine stabile Korrespondenzanalyse enorm wichtig. Beide Module bauen auf der von OpenCV errechneten Fundamentalmatrix auf, welche wiederum allein aus den Punktkorrespondenzen resultiert.

Sind die gegebenen Punktkorrespondenzen nicht einigermaßen gleichmäßig im Bild verteilt, so schlägt die Berechnung des RANSAC-Algorithmus fehl. Man spricht hierbei von *degenerierten Konfigurationen*. Bei *degenerierten Konfigurationen* approximiert RANSAC die F -Matrix beliebig, was von OpenCV nicht abgefangen wird. Diese können sich außerdem noch aus den Bildmodalitäten ergeben. Liegt beispielsweise eine pure Rotation, keine Translation oder gar keine Bewegung in den Bildern vor, nennt man dies auch

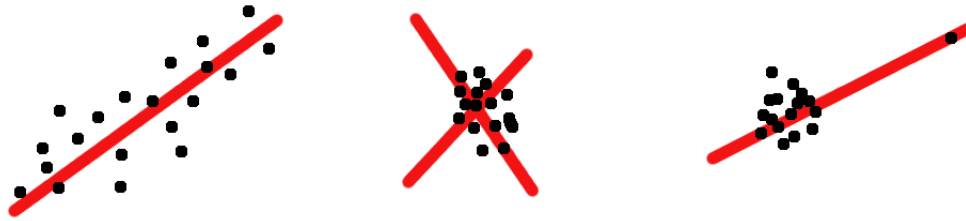


Bild 5.8: Approximationsprobleme bei RANSAC

eine *degenerierte Konfiguration*. Ebenso können linear abhängige Korrespondenzen die F -Matrixschätzung stark beeinflussen. Mehr zu degenerierten Konfigurationen lässt sich in [HZ03, pp. 295-297] erfahren.

Ein Beispiel sei gegeben: Man stelle sich vor, RANSAC solle eine Linie durch eine Punktemenge approximieren. Sind die Punkte gleichmäßig verteilt, so kann RANSAC die Linie korrekt schätzen. Sind die Punkte geclustert, so schafft RANSAC es nicht, die Linie zu errechnen und liefert ein willkürliches Ergebnis. Wenn zu der geclusterten Punktmenge noch ein Outlier hinzukommt, so wird er höchstwahrscheinlich maßgeblich den Verlauf der Linie beeinflussen. Diese Situation ist in Bild 5.8 dargestellt.

Deswegen wurden zwei Trackingverfahren genutzt, um die aufnahmespezifischen Modalitäten abzudecken. Der zuerst gewählte SIFT-Tracker erwies sich für die Gallesequenz als unbrauchbar, da die Punktkorrespondenzen nur geclustert in gewissen Teilen des Bildes gefunden wurden. Dies sind die Bereiche des Bildes, wo sich augenscheinlich die bestexturierten Flächen befinden. Daher wurde bei der Sequenz nur mit dem Lucas-Kanade-Tracker gearbeitet. Dieser liefert durchweg stabile Punktkorrespondenzen für die Gallesequenz. Bei der Schlosssequenz konnten beide Tracker zum Einsatz kommen, beide erkannten korrekte Punktkorrespondenzen in unterschiedlichen Bereichen der Bilder. Die Tassen- und die gerenderte Sequenz machten die Nutzung des SIFT-Trackers nötig, da die Bilder in stark unterschiedlichen Skalierungen vorliegen, mit denen der Lucas-Kanade-Tracker nicht robust genug zurechtkommt.

Die beiden problematischen Konfigurationen sind in den Bildern 5.9 und 5.10 gezeigt. Es wird offensichtlich, wie SIFT geclusterte Punktkorrespondenzen bevorzugt und der Lucas-Kanade-Tracker in Bezug auf die Skalierung versagt.

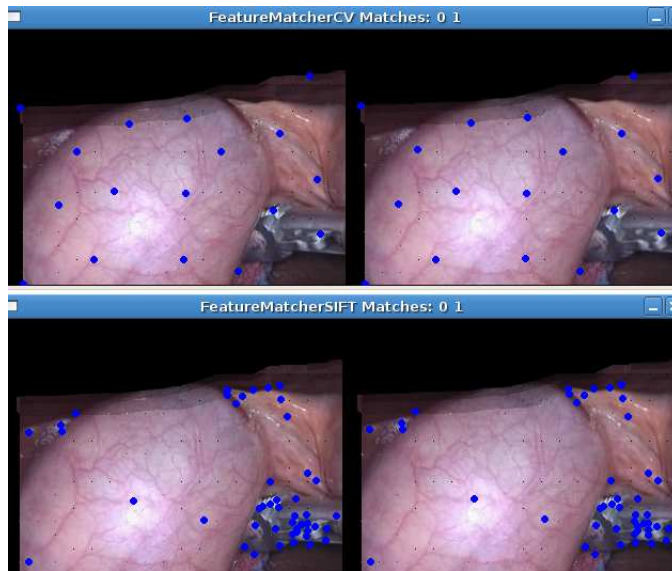


Bild 5.9: Gleichmäßige Verteilung beim LK-Tracker (oben) und Clusterbildung bei SIFT (unten)

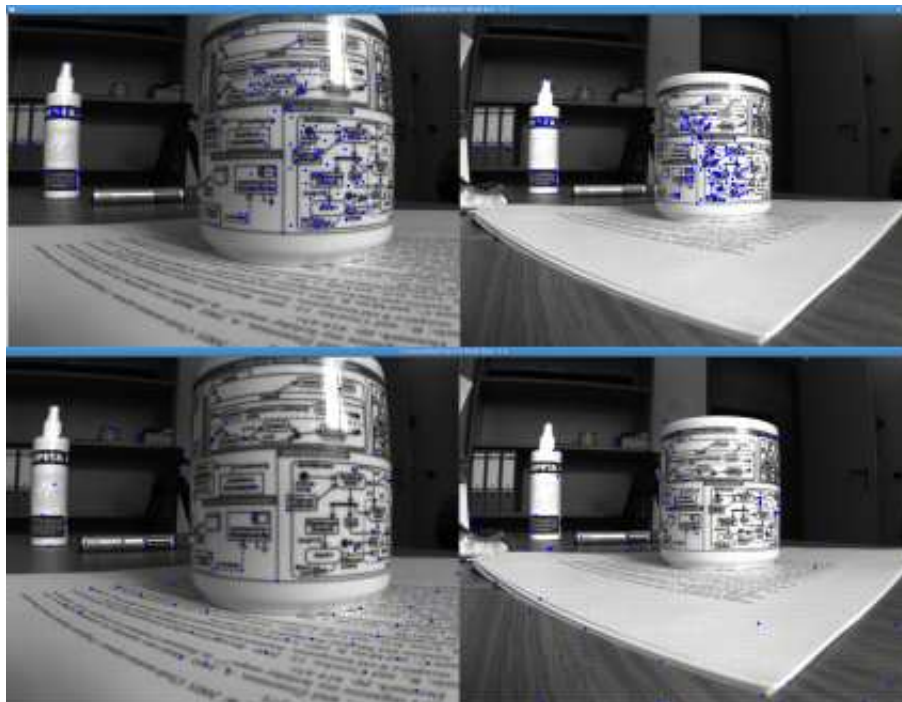


Bild 5.10: Gute skalierungsunabhängige Punktdetektion bei Sift (oben) und massive Fehlkorrespondenz beim LK-Tracker (unten)

Die Ergebnisse des RANSAC-Algorithmus sind in Tabelle 5.1 abgebildet. Die Tabelle zeigt jeweils die Anzahl der gefundenen Gesamtkorrespondenzen und die als Inlier klassifizierten Punkte, sowie deren prozentualen Anteil.

Sequenz	# SIFT	Inlier SIFT	Anteil	# Lucas-Kanade	Inlier Lucas-Kanade	Anteil
gerendert	329	175	53%	-	-	-
Tasse	1741	1381	79%	-	-	-
Schloß	20690	17586	84%	2949	2132	72%
Galle	-	-	-	1980	1381	70%

Tabelle 5.1: Ergebnisse der RANSAC-Inlieranalyse

5.3.3 Autokalibration

Unter der Einschränkung der Anwendbarkeit der Featuretracker lässt sich die Brennweite F bei gegebenen Punktkorrespondenzen mittels des Dynamic-Hill-Climbing approximieren. Die Ergebnisse werden in Tabelle 5.2 jeweils in Bezug auf den verwendeten Tracker dargestellt.

Wie zu sehen ist, bewegen sich die Abweichungen zwischen 3% und 22%. Auffällig

Sequenz	Ground Truth F	SIFT F	Abweichung %	Lucas-Kanade F	Abweichung %
gerendert	512 px	530 px	3.5%	-	-
Tasse	791 px	815 px	3%	-	-
Schloß	1100 px	1280 px	16%	1348px	22%
Galle	-	-	-	128 px	-

Tabelle 5.2: Ergebnisse der Autokalibration

ist die Abweichung zwischen SIFT- und Lucas-Kanade-Tracker bei der Schlosssequenz von 16% gegenüber 22%. Wie sich herausgestellt hat, kann man die Abweichung beim Lucas-Kanade-Tracker weiter auf 12% reduzieren, wenn man die Einhaltung der Epipolarbedingung strafft und nur einen Schwellwert von 0.1 zuläßt. Dies deutet darauf hin, daß der Lucas-Kanade-Algorithmus in OpenCV eine höhere Outlierquote produziert, während der SIFT-Algorithmus genauer zu arbeiten scheint. Diese Annahme wird durch Tabelle 5.1 bestätigt.

Hier gilt es, einen Mittelweg zwischen Genauigkeit und Berechenbarkeit der F -Matrix zu finden und die Parameter der Szene anzupassen.

5.3.4 Geometrieschätzung

Zur Geometrieschätzung selbst lassen sich keine direkt vergleichbaren Ergebnisse finden. Es ist davon auszugehen, dass der Fehler aus der Autokalibration weiter in die Geometrieschätzung miteinfließt. Dies ist allerdings als generelles Problem beim Umgang mit unkalibrierten Datensätzen, die eine Autokalibration erfordern, zu werten. Nach [RW04] kommt es bei anderen Verfahren ebenfalls zu Abweichungen im Bereich von 3% bis 15%, die allerdings den Autoren nach für die Visualisierung zu vernachlässigen sind.

Kapitel 6

Fazit

6.1 Ergebnisanalyse

Es wurde gezeigt, dass es prinzipiell möglich ist, die Oberflächennormalen an Stellen von Glanzlichtern durch die gegebenen physikalischen Umstände zu ermitteln. Während Glanzlichter in anderen Anwendungen eher als störend empfunden werden und man versucht, diese zu eliminieren oder zu ersetzen, wurde bei dieser Diplomarbeit der umgekehrte Weg verfolgt, um direkt zusätzliche Informationen über die Geometrie der Szene zu erhalten. Der Umgang mit unkalibrierten Datensätzen erschwert zwar die genaue Berechnung von Geometriedaten, allerdings halten sich die Abweichungen in einem für den Visualisierungskontext adäquaten Rahmen.

Die Glanzlichtdetektion durch eine HSV-Farbraumanalyse zählt nicht nur zu einem der schnellsten Verfahren dieser Art, sie liefert zudem auch die stabilsten Ergebnisse gegenüber anderen Verfahren.

Eine bedeutende Rolle nehmen die Featuretracker im ganzen Prozess ein. Sie beeinflussen die weiteren Schritte des Verfahrens maßgeblich durch ihre Robustheit. Fehlkorrespondenzen führen zu einer Fehlberechnung der Fundamentalmatrix, die wiederum die Grundlage für Autokalibration und Geometrieschätzung bildet. Im Rahmen der Diplomarbeit wurde durch die Anwendung zweier Trackingalgorithmen dafür Sorge getragen, dem verfügbaren Datenmaterial gerecht zu werden. Ebenso wurden Techniken wie RANSAC verwen-

det, um die Ergebnisse der Featuretracker im weiteren Kontext stabiler zu machen. Leider hat sich im Verlauf der Arbeit herausgestellt, dass die ursprüngliche Überlegung, mit OpenCV zu arbeiten, einige Hürden mit sich bringt, die vorher nicht absehbar waren. So war doch oft mangels ausreichender Dokumentation eine Quellcodestudie nötig, um die Arbeitsweise vorhandener Algorithmen zu verifizieren und zu verstehen. Der Support von OpenCV beschränkt sich auf eine Newsgroup¹ in der sich die Anwender gegenseitig bei Problemen helfen können, einen offiziellen Support gibt es nicht. Andererseits bietet OpenCV einen durchaus beachtlichen Funktionsumfang, wenn auch weiten Teilen nicht der Status *releasefertig* zugesprochen werden kann.

Dennoch lässt sich sagen, dass die Ergebnisse durchaus akzeptabel sind, zumal keinerlei Information über die verwendeten Bilddatensätze vorhanden sein muss und die Berechnungen allein auf gefundenen Punktkorrespondenzen basieren. Zusätzlich lassen sich alle Verfahren durch ihre intuitive Vorgehensweise einfach implementieren und validieren. In dieser Diplomarbeit wurde kein Wert auf die Laufzeitanforderungen gelegt, dennoch kann die Berechnung der Geometrie und der Glanzlichter nach der Korrespondenzfindung durch die verwendeten Verfahren relativ zeitnah erfolgen.

6.2 Zukunft

Für die Zukunft wäre es wünschenswert, die Schätzung der Korrespondenzanalyse und die Berechnung der Fundamentalmatrix robuster zu machen. Es wurde beispielsweise keine radiale Verzerrung der Linse modelliert und die Bilder wurden vor der Bearbeitung durch die Featuretracker nicht entzerrt.

Eine weitere Möglichkeit zur Evaluation der Geometrieschätzung wären Ground-Truth-Daten. Wir gehen im kalibrierten Fall von geringen Abweichungen zu den realen Gegebenheiten aus. Diese genaueren Ergebnisse könnten dann wichtige zusätzliche Informationen für die 3D-Rekonstruktion liefern, um dem Einzug des Systems im medizinischen Umfeld den Weg zu bereiten.

¹<http://tech.groups.yahoo.com/group/OpenCV/>

Anhang A

Mathematische Bezeichner und Symbole

Dieser Anhang gibt einen Überblick über die verwendeten Bezeichner und Symbole

Bezeichner	Bedeutung
α	Einfallswinkel
β	Ausfallswinkel
\mathbf{p}	Punkt in karthesischen Koordinaten
$\tilde{\mathbf{p}}$	Punkt in homogenen Koordinaten
$\tilde{\mathbf{p}}^w, \mathbf{p}^w$	Punkt im Weltkoordinatensystem (homogen, karthesisch)
$\tilde{\mathbf{p}}^c, \mathbf{p}^c$	Punkt im Kamerakoordinatensystem (homogen, karthesisch)
$\tilde{\mathbf{p}}^i, \mathbf{p}^i$	Punkt im Bildkoordinatensystem (homogen, karthesisch)
$\tilde{\mathbf{p}}^p, \mathbf{p}^p$	Punkt im Pixelkoordinatensystem (homogen, karthesisch)
F	Brennweite der Kamera
F_x	effektive horizontale Brennweite in Pixeln
F_y	effektive vertikale Brennweite in Pixeln
H	Hauptpunkt des Bildes
H_x	x-Koordinate des Hauptpunkts
H_y	y-Koordinate des Hauptpunkts

Bezeichner	Bedeutung
s	Scherung
d_x	Breite eines CCD-Sensorelements
d_y	Höhe eines CCD-Sensorelements
θ	Winkel zwischen den Achsen der CCD-Kamera
\mathbf{K}	Kalibrationsmatrix
\mathbf{c}_1	Kamerazentrum links
\mathbf{c}_2	Kamerazentrum rechts
l_1	epipolare Linie im linken Bild
l_2	epipolare Linie im rechten Bild
\mathbf{C}	Projektionsmatrix
\mathbf{C}_1	Kameramatrix der linken Kamera
\mathbf{C}_2	Kameramatrix der rechten Kamera
\mathbf{R}	Rotationsmatrix
\mathbf{t}	Translationsvektor
\mathbf{E}	Essentialmatrix
$[\mathbf{t}]_x$	Kreuzproduktmatrix des Translationsvektors
\mathbf{P}_{per}	perspektivische Projektionsmatrix
$\ \mathbf{E}\ _F$	Essentialmatrix in Frobeniusnorm
$\ \mathbf{t}\ $	Länge des Translationsvektors
\mathbf{F}	Fundamentalmatrix
ρ	Skalierungsfaktor der Szene
σ	Singulärwert
\mathbf{f}	Einträge der Fundamentalmatrix in Vektordarstellung
\mathbf{A}	Punktkorrespondenzmatrix
$\text{SVD}(\mathbf{E})$	Singulärwertzerlegung der Essentialmatrix
\mathbf{U}	linke Matrix der Singulärwertzerlegung
Σ	mittlere Matrix der Singulärwertzerlegung
\mathbf{V}	rechte Matrix der Singulärwertzerlegung

Bezeichner	Bedeutung
P	RANSAC Wahrscheinlichkeit
m	RANSAC Iterationen
ϵ	RANSAC erwartete Outlierquote
$width$	Bildbreite
$height$	Bildhöhe
a	erste Komponente des Zufallsvektors x
b	zweite Komponente des Zufallsvektors x
ω	Gewichtungsfaktor der Fundamentalmatrix in DHC
Inliers	Supportset einer Fundamentalmatrix
MaxInliers	maximales Supportset aller Fundamentalmatrizen
v	von DHC gewählter Vektor
d	Richtungsspeicher für DHC
n_t^b	Normale des t-ten Bildes im b-ten Bild
f_S	Schwellwertfunktion des Sättigungskanals
f_V	Schwellwertfunktion des Dunkelstufenkanals
$S(x, y)$	Intensität des Sättigungskanals
$V(x, y)$	Intensität des Dunkelstufenkanals
θ_S	Schwellwert des Sättigungskanals
θ_V	Schwellwert des Dunkelstufenkanals
$f_{HIGHLIGHT}$	Gesamtfunktion zur Glanzlichtdetektion
z	Zentroid eines Glanzlichts

Anhang B

Implementationsdetails

B.1 Benötigte Bibliotheken

Die folgenden Bibliotheken werden zum Zeitpunkt dieses Dokuments benötigt:

- OpenCV v. 1.0 (<http://sourceforge.net/projects/opencvlibrary/>)
- cvBlobsLib v. 6 (<http://opencvlibrary.sourceforge.net/cvBlobsLib>)
- GLUT (<http://www.opengl.org/resources/libraries/glut/>)
- QT3 (<http://trolltech.com/products/qt/qt3>)
- libxml2 (<http://xmlsoft.org/>)

B.2 Doxygen-Dokumentation

Im Folgenden findet sich die von Doxygen erstellte Klassendokumentation.

Verzeichnis der Bilder

1.1	Flexibles und starres Endoskop	3
1.2	Schematischer Querschnitt durch ein Endoskop	4
1.3	Aufnahmen aus dem OP bei einer Laparoskopie	5
2.1	Welt- und Kamerakoordinatensystem	10
2.2	Lochkameraprinzip	12
2.3	Pixelkoordinatensystem	13
2.4	Epipolargeometrie	15
2.5	Glanzlichter in einer realen Laparoskopieaufnahme	18
2.6	Reflexionseigenschaften	19
2.7	Bezug zur Endoskopie	19
4.1	Schematischer Ablauf des Verfahrens	26
4.2	Auflösungshierarchie mit Pyramiden	27
4.3	Finden des neuen Initialisierungsvektors für die innere Schleife	35
4.4	Die vier möglichen Faktorisierungen von E . Rechts-Links: Invertierung von t_t . Oben-Unten: R_{t1} bzw. R_{t2}	38
5.1	Die Frames 1, 30, 60 und 90 der Gallsequenz	46
5.2	Die Frames 0, 9, 18 und 27 der Schlossequenz	47

5.3	Die Tassensequenz	48
5.4	Die gerenderte Mayasequenz	49
5.5	QT-Interface zum Justieren der HSV-Schwellwerte	50
5.6	Auswirkungen der Schwellwerte auf die gefundenen Glanzlichter(v.l.o.n.r.u.): Originalbild, zu hohe Sättigung, angepasste Schwellwerte, zu niedrige Dunkelstufe	52
5.7	Detektierter Schwerpunkt (rot) und Bounding Box des Glanzlichts	53
5.8	Approximationsprobleme bei RANSAC	54
5.9	Gleichmäßige Verteilung beim LK-Tracker (oben) und Clusterbildung bei SIFT (unten)	55
5.10	Gute skalierungsunabhängige Punktdetektion bei Sift (oben) und massive Fehlkorrespondenz beim LK-Tracker (unten)	55

Literaturverzeichnis

- [BBS04] Eric Beets, Samia Boukir, and David Suter. Aircraft pose estimation from homography. Technical report, Department of Electrical and Computer Systems Engineering, Monash University, Australia, 2004.
- [Bou] Jean-Yves Bouguet. Pyramidal implementation of the lucas kanade feature tracker.
- [BTVG06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *ECCV*, 2006.
- [Can86] John Francis Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [Dec07] Peter Decker. Bildbasierte bewegungsschätzung aus kamerafahrten anhand prägnanter merkmale. Master’s thesis, Universität Koblenz-Landau, Campus Koblenz, Fachbereich 4 Informatik, Institut für Computervisualistik, 9 2007.
- [dlMY94] Michael de la Maza and Deniz Yuret. Dynamic hill climbing. Technical report, Massachusetts Institute of Technology, Computer Science and AI Lab, web.mit.edu/, 1994.
- [dlMY03] Michael de la Maza and Deniz Yuret. Dynamic hill climbing: Overcoming the limitations of optimization techniques. Technical report, Massachusetts Institute of Technology, Computer Science and AI Lab, web.mit.edu/, 2003.

- [FLM92] Olivier D. Faugeras, Quang-Tuan Luong, and Stephen J. Maybank. Camera self-calibration: Theory and experiments. In *European Conference on Computer Vision*, pages 321–334, 1992.
- [GMR02] Dmitry Gorodnichy, S. Malik, and Gerhard Roth. Affordable 3d face tracking using projective vision. In *Vision Interface 2002*, pages 383–390, 2002.
- [Hei03] Benno Heigl. *Plenoptic Scene Modelling From Uncalibrated Image Sequences*. PhD thesis, Technische Fakultät der Universität Erlangen-Nürnberg, 2003.
- [HS81] Berthold K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
- [HZ03] Richard I. Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [KL81] Takeo Kanade and Bruce Lucas. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, 4 1981.
- [Kli01] Caesar Klimowicz. Lichtfelderzeugung aus endoskopischen bildfolgen des bauchraums. Master's thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, <http://www.informatik.uni-erlangen.de>, 2001.
- [LD99] Manolis I.A. Lourakis and Rachid Deriche. Camera self-calibration using the singular value decomposition of the fundamental matrix: From point correspondences to 3d measurements, 1999.
- [Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [RW02a] Gerhard Roth and Anthony Whitehead. Evolutionary based autocalibration from the fundamental matrix, 4 2002.

- [RW02b] Gerhard Roth and Anthony Whitehead. Some improvements on two autocalibration algorithms based on the fundamental matrix. In *International Conference on Pattern Recognition (ICPR 2000)*. Volume 2, pages 312–315, 2002.
- [RW04] Gerhard Roth and Anthony Whitehead. Estimating intrinsic camera parameters from the fundamental matrix using an evolutionary approach. In *EURASIP Journal on Applied Signal Processing 2004:8*, pages 1113–1124, 2004.
- [Sch04] Philipp Schaer. Evaluation von glanzlichtdetektionsverfahren in der endoskopie, 2004.
- [TB97] Lloyd N. Trefethen and David Bau. Numerical linear algebra, 1997.
- [Tri98] Bill Triggs. Autocalibration from planar scenes, 1998.
- [Tsa87] Roger Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.
- [TV98] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, New York, 1998.
- [Vog06] Florian Vogt. *Augmented Light Field Visualization and Real-Time Image Enhancement for Computer Assisted Endoscopic Surgery*. PhD thesis, Technische Fakultät der Universität Erlangen-Nürnberg, 2006.
- [Zha00] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.