Projektive Tiefenrepräsentationen aus partiell kalibrierten Stereokamerasystemen

Diplomarbeit
am Lehrstuhl für Kognitive Systeme
des Instituts für Informatik und Praktische Mathematik
der Christian–Albrechts–Universität zu Kiel

vorgelegt von
Derek Buck

Betreuer: Dr. Kostas Daniilidis
Prof. Dr. Gerald Sommer

vorgelegt am
24. 2. 1997

Zusammenfassung

Eine der Hauptaufgaben des Maschinensehens ist die Analyse der 3D-Struktur einer Szene mittels eines Stereokamerasystems. Die klassische Vorgehensweise ist die Kalibrierung der Kameras mit einem Kalibrierungsobjekt und die anschließende Bestimmung der „Euklidischen Struktur" des Raumes in einem kartesischen Koordinatensystem. Dieser Kalibrierungsschritt ist jedoch fehlerbehaftet und oft nicht praktikabel. Deshalb haben in letzter Zeit unkalibrierte Verfahren zur Strukturerschließung viel Beachtung gefunden. Diese liefern üblicherweise die „projektive Struktur" des Raumes, also die Szene bis auf eine projektive Transformation.

Der Ansatz des qualitativen oder aktiven Sehens liefert weitergehende Betrachtungen durch die Möglichkeit der Fixierung und damit der Realisierung vereinfachter Systemgeometrien. Hier gewinnt man die Szene bis auf eine Relieftransformation, die die Tiefenordnung erhält. Im Gegensatz zu schwach kalibrierten Verfahren, die mehrere Punktekorrespondenzen benötigen, kann hier direkt aus einem einzigen Bildpunktepaar rekonstruiert werden.

Wir stellen hier solche bekannten Verfahren vor und erweitern sie auf (partiell) unbekannte interne Kameraparameter auf dem Weg zu einer total unkalibrierten Vorgehensweise. Wir präsentieren neue Varianten und Eigenschaften der Rekonstruktionen. Wir liefern analytische und experimentelle Ergebnisse ihrer Leistungen, auch im Vergleich mit einem klassischen Verfahren mit Kalibrierung.

Danksagung

Ich bedanke mich bei Stephanie Willemsen für die Mitarbeit am Kalibrierungsprogramm. Ich bedanke mich bei Dr. K. Daniilidis und Prof. Dr. G. Sommer für die Betreuung meiner Arbeit.

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Problematik

Eine der Hauptaufgaben beim Maschinensehen ist die Gewinnung von Informationen über die dreidimensionale Struktur der abgebildeten Szene. Dies ist das Gebiet des sogenannten *3D-Maschinensehens*. Ihr klassisches Vorgehen ist nach [Mar82] die Detektion und Lokalisation von Änderungen von Bild zu Bild, die sogenannte *primäre Repräsentation*, die eine Interpretation bezüglich der 3D-Struktur erlaubt, in der *$2\frac{1}{2}$D-Repräsentation*. Die $2\frac{1}{2}$D-Repräsentation kann auch Tiefenkarte genannnt werden und enthält die beobachtbare Information, da man in der Regel ja nicht in die Objekte hineinsehen kann. Dieses Wissen kann man jedoch zu einer *3D-Repräsentation* unter Hinzunahme von Modellierungen der Szene nutzen. Wir sind hier lediglich an einer $2\frac{1}{2}$D-Repräsentation interessiert, da sie für viele Aufgaben wie die Navigation und Manipulation ausreicht. Ist also von Struktur die Rede, so meinen wir Information über die Lage und Gestalt von Oberflächen. Man beachte, daß alle diese Schritte von der Wahl geeigneter Modelle abhängen.

Die Verwendung von a-priori-Wissen und geeigneten Modellen ist kritisch für den Prozeß des 3D-Maschinensehens. Einerseits versucht man, die Allgemeinheit der Lösung zu gewährleisten, um sie in vielen, sich ändernden Situationen einzusetzen. Zudem ist die Gewinnung von Information, die als Eingabe für Lösungsverfahren dient, oft ein komplexer, fehlerhafter Vorgang. Andererseits hat sich in Analogie zur Künstlichen Intelligenz gezeigt, daß viele Aufgabenstellungen zu schwierig, ihre Lösungen zu zeitaufwendig und nicht robust genug sind, um in Realweltsituationen zu bestehen. Informationen und Annahmen über die Szene sind hier notwendige Bestandteile praktikabler Algorithmen. Als in diesem Zusammenhang hinderlich hat sich vor allem erwiesen, den Beobachter als reines Berechnungssystem zu beschreiben. Es ist jedoch sinnvoll zu berücksichtigen, daß ein System mit der Umgebung über seine physikalische Erscheinung in Beziehung steht und daß seine Komponenten über Sensoren Informationen liefern. So versteht man unter einem *aktiven Beobachter* nach [Baj88] ein System, daß in der Lage ist, die Parameter seines sensorischen Systems zu steuern. Dieser Paradigmenwechsel zu *aktivem* oder *qualitativem Sehen* hat in jüngerer Zeit viel Beachtung gefunden (siehe z. B. [Som95] und [Alo93]) und elegante sowie robuste Lösungen ermöglicht. Die Kombination von vorhandenem Wissen und zu gewinnender Information ist aber häufig ein für jede konkrete Aufgabe neu zu lösendes Problem.

Unabhängig vom Beobachtermodell bleibt die Frage, worüber 3D-Information aus Bildern zu gewinnen ist. Die verschiedenen Konzepte werden unter dem Stichwort *Strukturerschließung aus X* (Structure-from-X) geführt. Hier beschäftigen wir uns mit den beiden gängigsten, *Strukturerschließung aus Stereosehen (Structure-from-Stereo, SfS)* und *Strukturerschließung aus Bewegungssehen (Structure-from-Motion, SfM).*

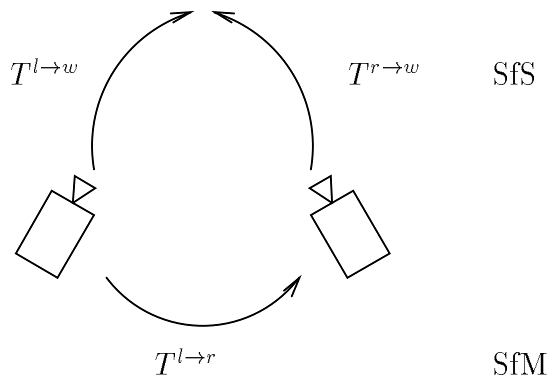

Abbildung 1.1: Prinzipielle Situation beim Problem der Strukturerschließung aus Stereosehen (SfS) und der Strukturerschließung aus Bewegungssehen (SfM). Der statische Aufbau ermöglicht eine Kalibrierung beim SfS, d. h. die einmalige Bestimmung der Transformationen der Abbildung von Welt- auf Kamerakoordinatensystem für alle beteiligten Kameras. Bei einer sich bewegenden Kamera ist diese Vorgehensweise nicht praktikabel. Hier versucht man, die Transformationen der relativen Bewegung zwischen den Kamerakoordinatensystemen und darüber die Struktur der Szene zu ermitteln (SfM).

Strukturerschließung aus Stereosehen ist das ältere Problem, da es auf frühe Anwendungen in der Photogrammetrie zurückgeht. Hier versucht man, die Struktur der Szene aus mehreren Aufnahmen zu gewinnen, wie es auch viele Tiere zu tun fähig sind, deren linkes und rechtes Gesichtsfeld überlappen. Binokulare Systeme sind dabei in fast allen Fällen ausreichend und auch am weitesten verbreitet. Man nutzt Wissen über die zeitlich konstante Systemgeometrie, um aus der örtlichen Verschiebung im Bild, der sogenannten Disparität, auf die Tiefe der abgebildeten Szenepunkte zu schließen. Erst diese örtliche Verschiebung ermöglicht eine Tiefeninterpretation. Probleme ergeben sich so aus der exakten Bestimmung dieser Verschiebung und der Systemgeometrie. Die Systemgeometrie kann hier durch die Bestimmung der Transformationen von einem Welt- auf die Kamerakoordinatensysteme ermittelt werden. Dies ist die sogenannte *Kalibrierung* (Abbildung 1.1).

Strukturerschließung aus Bewegungssehen ist eine jüngere Problematik, die bei mobilen Kamerasystemen im Vordergrund steht. Es existieren viele psychophysikalische Experimente, die auf SfM-Moduln bei Lebewesen schließen lassen; so ist auch das menschliche visuelle System in der Lage, nur über einzelne sich bewegende Punkte im Bild die zugehörige 3D-Struktur einzuordnen. Problemstellungen ergeben sich hier zumeist für monokulare, aber auch binokulare Systeme, sowie für mehrere einzeln gemachte Bilder und ganze „kontinuierliche“ Bildfolgen. Durch die Bewegung des Beobachters ist eine Kalibrierung hier nicht möglich. Man versucht daher, die relative Bewegung über die örtliche Verschiebung im Bild und damit die Struktur der Szene zu gewinnen (Abbildung 1.1). Darüber hinaus sind Betrachtungen in Szenen mit mehreren sich bewegenden Objekten nur lokal anzustellen, und es ist die Frage nach Gebieten unterschiedlicher Bewegung zu beantworten. Die Nicht-Trivialität des Problems hat dementsprechend eine Fülle von Lösungsmöglichkeiten hervorgebracht, die mit verschiedenen Annahmen und verschiedenem Aufwand betrieben werden.

Weitere Gebiete sind Strukturerschließung aus Reflektion (Structure-from-Shading) und Strukturerschließung aus Textur (Structure-from-Texture), wo mit Reflektanzeigenschaften bzw. der Lage von Textureinheiten gearbeitet wird, um nur noch zwei zu nennen.

1.2 Motivation und Beitrag dieser Arbeit

Ein wichtiger Aspekt des SfS und des SfM besteht darin, zu untersuchen, inwiefern die direkt aus dem Bild gewonnenen Informationen bereits etwas über die Struktur der Szene aussagen. Die klassische Vorgehensweise beim SfS ist, wie gesagt, die Kameras zu kalibrieren, das heißt, die Parameter zu bestimmen, die die Abbildung von einem Weltkoordinatensystem auf das Kamerakoordinatensystem charakterisieren. Dazu wird meist ein Kalibrierungsobjekt mit bekannten Abmessungen benutzt, so daß anschließend metrische Bestimmungen durchgeführt werden können. Dieser Kalibrierungsvorgang ist aber zeitaufwendig und in vielen Aufgabenstellungen nicht durchführbar. Zudem macht man sich eventuell unnötig von einem fehlerbehafteten Schritt abhängig. So taucht die Frage nach unkalibrierter Szenenanalyse auf. Damit verbunden ist die Beobachtung, daß es sich bei der Kamera um ein projektives Instrument handelt und daß das Aussehen der Szene geeignet im Rahmen der projektiven Geometrie diskutiert werden kann. In der Tat hat sich gezeigt, daß allein aus wenigen korrespondierenden Paaren von Bildpunkten die sogenannte *projektive Struktur* oder die sogenannte *affine Struktur* der Szene rekonstruiert werden kann ([Fau92]). Demnach ist zu unterscheiden, welche Struktur der Szene rekonstruiert werden soll, und welche Eigenschaften der Szene daran abzulesen sind. Es sind viele (Teil-)Aufgaben des 3D-Maschinensehens bereits mit wenigen Informationen zu lösen, die mit entsprechend weniger Aufwand, besonders im Hinblick auf aktive Systeme, zu gewinnen sind.

Ein Vorteil aktiver Systeme ist die Möglichkeit, sinnvolle geometrische Einschränkungen herbeizuführen, die die Bildinformation nicht beeinträchtigen, aber spezielle Verfahren zulassen, die im allgemeinen nicht angewendet werden können. Konkret beschäftigen wir uns hier mit (Varianten von) *Fixationsgeometrien*, bei denen sich die optischen Achsen der Kameras in einem Szenepunkt schneiden. In solchen Geometrien ist es sogar möglich, aus *einem einzelnen korrespondierenden Bildpunktepaar* eine projektive Rekonstruktion zu erhalten, die Informationen über die Tiefenordnung aus Sicht des Beobachters liefert. Solche besonderen Rekonstruktionen, die die Szene bis auf eine *Relieftransformation* wiedergeben, und ihre Bewerkstelligung werden in [GPMF95] und [GPFM96] im Überblick diskutiert, worauf wir hier aufbauen. Neben der grundsätzlichen Bedeutung der Tiefenordnung für Aufgaben wie die Navigation gilt es zu betonen, daß man eine projektive Rekonstruktion fast ohne Kalibrierungsinformation, lediglich unter Kenntnis interner Kameraparameter, erhält. Dies ist ein Vorteil gegenüber allgemeineren projektiven Rekonstruktionen ([Fau92]), die zwar kein Kalibrierungsobjekt, aber zumindest mehrere Punktkorrespondenzen für eine *schwache Kalibrierung* benötigen. In der Tat werden wir zeigen, daß in diesen Geometrien für eine Rekonstruktion, die die Tiefenordnung erhält, nur das Bildzentrum, für eine allgemeine projektive Rekonstruktion nur seine vertikale Komponente notwendig ist.

Diese Arbeit gliedert sich wie folgt: Das folgende Kapitel stellt die Prinzipien des SfS vor. Insbesondere behandelt es die sogenannte *Stratifikation* des Raumes nach [Fau95], die die einzelnen Strukturbeschreibungen miteinander in Beziehung bringt, und die Nutzbarkeit von *orientierter projektive Geometrie* nach [Sto91]. Dies ermöglicht uns die eingehendere Untersuchung der Eigenschaften der hier vorgestellten Rekonstruktionen. Im dritten Kapitel geht es um die Beschreibung einfacher fixierter Stereosysteme und ihrer Beziehung zum SfM-Problem. In diesen Systemen kann man einfach eine *skalierte relative Nähe (SRN)* oder eine *affine Nähe (AN)* gewinnen, wobei erstere eine 1-Punkt-Rekonstruktion bis auf eine Relieftransformation, letztere bis auf eine allgemeine projektive Transformation ermöglicht. Dies setzt allerdings eine vereinfachte Darstellung der Disparität voraus. Es wird mit exakten Fehlerformeln und einem experimentellen Vergleich analysiert, welche Darstellungen sich eignen und welche nicht. Das vierte Kapitel beschäftigt sich dann mit der unmittelbaren Gewinnung von 3D-Strukturinformation aus korrespondierenden Paaren von Bildpunkten. Es werden Rekonstruktionsmöglichkeiten aus SRN oder AN und ihre Eigenschaften vorgestellt und ergänzt. Dann werden bekannte Berechnungsmodelle und neue Varianten zur Gewinnung von SRN oder AN eingeführt. Insbesondere berücksichtigen wir den Fall unbekannter interner Kameraparameter für eine *total unkalibrierte* Rekonstruktion. Abschließend diskutieren wir anhand von Experimenten die Leistung der 1-Punkt-Rekonstruktionsverfahren im Stereofall bezüglich der Gewinnung der Euklidischen und der projektiven Struktur und ihrer Invarianten. Sie werden untereinander und mit einem klassischen SfS-Verfahren mit Kalibrierung verglichen. Dabei wird besonders der Einfluß von Fehlern im Bildzentrum, also in der verbleibenden a-priori-Kalibrierungsinformation, sowie die Möglichkeit einer robusten Gewinnung von relevanter Information wie Komplanarität und Tiefenordnung beleuchtet.

Kapitel 2

Strukturerschließung aus Stereosehen

Eine der Hauptaufgaben beim Maschinensehen ist die Analyse von Videobildern zur Gewinnung von Informationen über die 3D-Struktur der abgebildeten Szene, Aufgabengebiet des sogenannten *3D-Maschinensehens*. Es erscheint konsequent, sich, wie das menschliche visuelle System (MVS) oder vergleichbare Systeme in der Tierwelt, zusätzliche Informationen durch Verwendung mehrerer Kameras (Augen) zu verschaffen. So beschreibt *Strukturerschließung aus Stereosehen* (Structure-from-Stereo, SfS) die Aufgabe, die Struktur der Szene durch Aufnahmen aus verschiedenen Perspektiven zu rekonstruieren. Die allgemeine Lösung dieses Problems macht allerdings, wie viele Probleme des Maschinensehens, Schwierigkeiten in Bezug auf die Komplexität der Aufgabenstellung und in Bezug auf die Robustheit und den Berechnungsaufwand der Lösungen. Insbesondere ist die Einbeziehung von Informationen, die man im Prozeßverlauf gewinnt oder aus a-priori-Wissen bezieht, kritisch: Einerseits ermöglicht eine große Menge an Information besonders einfache und schnelle Lösungen, andererseits sind diese Informationen oft fehlerbehaftet und ihre Gewinnung schwierig und zeitaufwendig. Die Aktivität des Beobachters kann hier viele Probleme vereinfachen, wie wir im folgenden Kapitel sehen werden.

Grundsätzlich ist darauf zu achten, daß die verwendeten und gewonnenen Informationen mit der Aufgabenstellung im Einklang stehen. Da für viele Anwendungen eine vollständige Rekonstruktion der Szene nicht notwendig ist, kann man so auf Arbeitsschritte verzichten, die eventuell mit rechnerischem Aufwand oder Empfindlichkeit in den Eingabedaten versehen sind. Die in dieser Arbeit vorgestellten Lösungen werden dies berücksichtigen.

Die Richtigkeit dieses Ansatzes wird von psychophysikalischen Untersuchungen unterstützt, die nahelegen, daß die Leistungen des MVS in der Schätzung der exakten metrischen 3D-Struktur erstaunlich schlecht sind. Hingegen können relative Lagen oft sehr präzise geschätzt werden ([CSS91], [Joh91], zitiert in [GPMF95]).

Informationen und Informationsquellen können aber nicht beliebig eingeschränkt werden. So ist es nicht möglich, die Entfernung eines Objektes aus einer einzelnen Perspektive zu bestimmen, da z. B. ein doppelt so großes Objekt in doppelter Entfernung dasselbe Bild hat (Skalierungsmehrdeutigkeit). Erst zwei Projektionen desselben Szenepunktes und ihre örtliche Verschiebung ermöglichen eine Tiefeninterpretation. Es kann also einem erhöhtem Aufwand des Stereosehens gegenüber dem monokularen Sehen, der Soft- und

Hardwarelösungen betrifft, nicht immer aus dem Weg gegangen werden. Stereosehen ist unabdingbar für eine klassische Aufgabe der Photogrammetrie, die *Tiefenmessung*.

Dieses Kapitel ist so gegliedert: Der erste Abschnitt beschreibt das klassische Vorgehen beim SfS in einfacher, exemplarischer Form. Die drei darauf folgenden Abschnitte legen die mathematischen und konzeptionellen Grundlagen der Strukturgewinnung aus geometrischer Sicht. Dabei werden die hier verwendeten Modelle der Kamera, des Stereosystems und der realen Welt und wesentliche Hilfsmittel zu ihrer Modellierung vorgestellt. Der vorletzte Abschnitt stellt das klassische Rekonstruktionsverfahren im Rahmen der beschriebenen Modelle dar, bevor abschließend der wichtige Begriff der Disparität eingeführt wird.

2.1 SfS als 3-Stufen-Prozeß

Hier soll jetzt der allgemeine Ablauf des SfS-Prozesses dargestellt und exemplarisch vorgeführt werden. Er gliedert sich wie folgt in drei Phasen ([DA89]):

1. Präprozesse
2. Zuordnung
3. Bestimmung der 3D-Struktur

Manchmal spricht man auch unter Vernachlässigung des ersten Schrittes von einem 2-Stufen-Prozeß und vom dritten Schritt allgemeiner als Interpretation bezüglich der 3D-Struktur. Letzteres bezieht sich auf die angesprochene selektive Gewinnung von Eigenschaften der Szene und der Art ihrer Struktur, was in folgenden Abschnitten an Kontur gewinnen wird.

Ad 1. Präprozesse bereiten die Bilder für den Zuordnungsschritt vor. Hier werden interessierende Bildbereiche mit geeigneten Filtern behandelt. Zum Beispiel werden Eckpunkte oder Kantensegmente in jedem Bild extrahiert, die dann leichter ihrem Urbild in der Szene, also auch leichter ihrer Entsprechung in den Bildern, die aus anderer Perspektive aufgenommen wurden, zugeordnet werden können. Oder es wird mittels Glättungsfiltern Rauschen unterdrückt, und man gelangt so zu ähnlicheren Verläufen der einzelnen Bilder. Wie konkret vorgegangen wird, hängt in erster Linie vom Zuordnungsschritt ab.

Ad 2. Bei der *Zuordnung* geht es darum, die Korrespondenz zwischen einzelnen Bildpunkten oder Bildbereichen zu bestimmen, das heißt, die Punkte oder Bereiche zu ermitteln, die Bilder desselben Szeneausschnitts sind. Dies ist für die zumindest qualitative Einschätzung der Tiefe notwendig. Man unterscheidet dabei zwei Arten von Techniken:

(a) <u>Merkmals-basierte Techniken</u>
Diese Verfahren versuchen, Korrespondenz über die Zuordnung von Punkten auf signifikanten Bildbereichen wie Ecken oder Kanten zu etablieren. Hierzu ist eine Merkmalsextraktion im ersten Schritt erforderlich. Vorteile liegen in der Möglichkeit einer subpixelgenauen Lokalisierung und in der Robustheit gegenüber Beleuchtungsänderungen. Außerdem kann die Zuordnung oft über einfache Vergleiche der Merkmalsattribute in kurzer Zeit erfolgen. Allerdings müssen natürlich solche Merkmale vorhanden sein, und die betrachteten Punkte können nicht frei gewählt werden. Dies führt zu einem dünn besetzten Feld von Punkten.

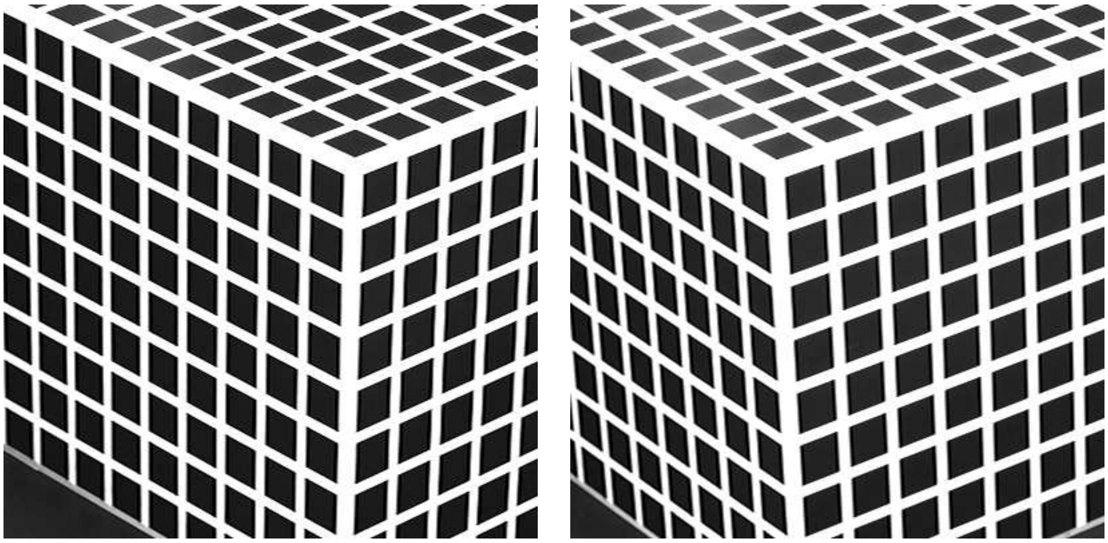

Abbildung 2.1: Zwei Bilder eines Kalibrierungswürfels als Beispiel für ein Stereobildpaar. Die örtliche Verschiebung zwischen den Bildern liefert Informationen über die 3D-Struktur. Diese Verschiebung wird im Zuordnungsschritt ermittelt.

(b) Gebiets-basierte Techniken
Hier werden Punkte nur **lokal** zugeordnet. Z. B. dient die Korrelation zwischen den Intensitätswerten in der Umgebung der betrachteten Bildpunkte als Maß für die Übereinstimmung des Urbildes, also der Korrespondenz. Da Intensitätswerte direkt benutzt werden, gibt es hier erhöhte Empfindlichkeit gegen Rauschen und Beleuchtungsänderungen. Zudem unterliegt die Berechnung der Korrelation hoher Komplexität und einer Unsicherheit in der Lokalisierung, und es muß eine Suche im Bild durchgeführt werden. Diese Verfahren können jedoch auf texturierten Gebieten angewendet werden, was die Berechnung eines dicht besetzten Feldes von Zuordnungen ermöglicht.

Eine andere wichtige Verfahrensgruppe sind die phasenbasierten: Hier wird eine globale örtliche Verschiebung in den Bildern als Phasenverschiebung im Fourierspektrum detektiert, beispielsweise als Antwort von Gaborfiltern ([VT95]). Diese Verfahren sind robust, kommen ohne eine Suche im Bild aus und können ohne jegliche Merkmale, auch ohne Textur, funktionieren. Allerdings darf die Bildpunktverschiebung nicht die Größe des Anwendungsbereiches des Filters überschreiten.

Die zugeordneten Punkte können einer globalen Konsistenzprüfung unterzogen werden, die mehrdeutige Zuordnungen ausschließen, indem sie Zuordnungen in der weiteren Umgebung und ihren Verlauf berücksichtigen. Dies ist vor allem bei Merkmals-basierten Techniken sinnvoll, da hier Merkmalseigenschaften einfließen können. Auch die Tiefenbestimmung und Szenenrekonstruktion des dritten Schrittes liefert Hinweise für die Korrektheit der gefundenen Korrespondenzen.

Die Systemgeometrie bedingt, daß die korrespondierenden Bildpunkte nicht beliebig liegen können. Zu jedem Bildpunkt existiert eine Gerade im anderen Bild, auf der die Zuordnung liegen muß. Diese Gerade heißt Epipolargerade und ergibt sich als Projektion der Gerade durch Bildpunkt und optisches Zentrum. Dadurch wird die Suche nach der

Korrespondenz auf einen eindimensionalen Raum beschränkt. Wir gehen darauf in §2.3 genauer ein.

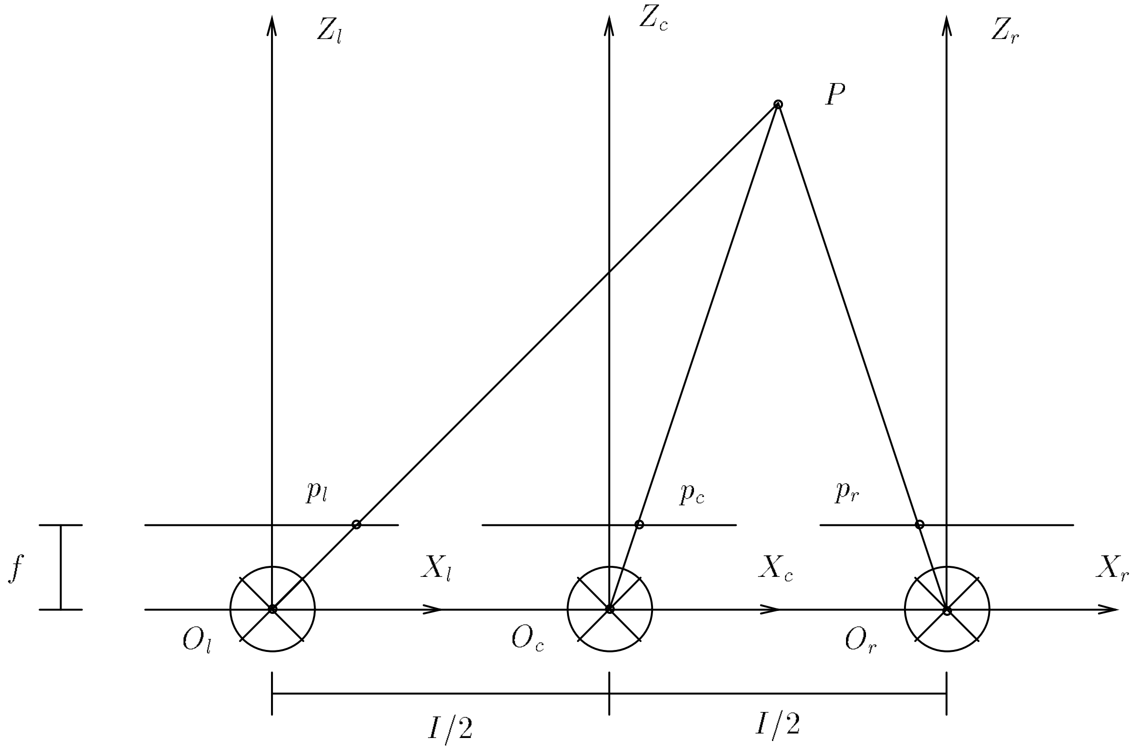

Abbildung 2.2: Binokulare parallele Stereogeometrie mit zwei Kameras der Brennweite f und Stereobasis der Breite I, von oben gesehen. Eingezeichnet sind die Achsen des linken, rechten und zyklopischen Koordinatensystems. Die Bildpunkte p_l, p_r, p_c sind die Schnittpunkte der Bildebenen mit den Sehstrahlen zum Szenepunkt P. Diese Sehstrahlen sind Geraden, die jeweils durch P und die Zentralprojektionspunkte O_l, O_r und O_c gehen.

Ad 3. Die *Bestimmung der 3D-Struktur* schließt den SfS-Prozeß ab. Hier werden die gefundenen korrespondierenden Bildpunkte zur Rekonstruktion des Szenepunktes genutzt. Unter Kenntnis der Kamera- und Systemgeometrie läßt sich so die Tiefe und mit dieser Information alle Koordinaten des 3D-Punktes bestimmen. Der Fall, wo anstelle von Bildpunkten ganze Bildbereiche wie Geraden auftauchen, sei hier zunächst vernachlässigt. Abbildung 2.2 zeigt das einfache Beispiel eines binokularen parallelen Stereosystems mit Kameras gleicher Brennweite. Hieran sei einmal die Vorgehensweise exemplarisch erläutert, wobei einige wichtige Begriffe vorgestellt werden.

Der Abbildungsvorgang wird durch eine Zentralprojektion modelliert. Der Bildpunkt $\boldsymbol{x} = (x, y, f)^T$, die Abbildung des Szenepunktes $\boldsymbol{X} = (X, Y, Z)^T$ durch eine Kamera mit der Brennweite f, ergibt sich aus dem Strahlensatz zu

$$(x, y) = (f\frac{X}{Z}, f\frac{Y}{Z}). \tag{2.1}$$

Die *horizontale Disparität* d_x, i. e. die Verschiebung des Bildpunktes in x-Richtung, ist dann

$$d_x \quad =_{\text{def}} \quad x_r - x_l = f\frac{X_r}{Z_r} - f\frac{X_l}{Z_l} = f\frac{X_c - I/2}{Z_c} - f\frac{X_c + I/2}{Z_c}$$

$$= \quad -f\frac{I}{Z_c}, \tag{2.2}$$

also erhält man die Tiefe aus der Beziehung

$$Z_c = -f\frac{I}{d_x}. \tag{2.3}$$

Hierbei ist I die Stereobasisbreite, und X_l, Y_l, Z_l, X_r, Y_r, Z_r, X_c, Y_c, Z_c sind die Weltkoordinaten bezüglich des linken, rechten und *zyklopischen Koordinatensystems.* Letzteres ist ein gleich ausgerichtetes Referenzkoordinatensystem, das einer virtuellen Kamera in der Mitte der Stereobasis entspricht. Seine Bildpunkte ergeben sich aus dem durch die Brennweite skalierten arithmetischen Mittel der Kamerabildpunkte:

$$\begin{aligned} 2x_c &= f\frac{X_c}{Z_c} + f\frac{X_c}{Z_c} = f\frac{X_r + I/2}{Z_r} + f\frac{X_c l - I/2}{Z_l} \\ &= f(x_r + x_l), \text{ also} \\ x_c &= f\frac{x_r + x_l}{2}. \end{aligned} \tag{2.4}$$

Da hier sowohl $Z_c = Z_l = Z_r$ als auch $Y_c = Y_l = Y_r$ ist, erhält man

$$y_c = fy_l = fy_r = f\frac{y_r + y_l}{2}. \tag{2.5}$$

Mit der Projektionsgleichung (2.1) ergeben sich dann die zyklopischen Koordinaten des 3D-Punktes aus der bekannten Tiefe (2.3) durch

$$(X_c, Y_c, Z_c) = (\frac{x_c Z_c}{f}, \frac{y_c Z_c}{f}, Z_c).$$

Mit (2.4, 2.5) bekommt man daraus die Beziehung zwischen den Bildpunkten und dem Szenepunkt

$$(X_c, Y_c, Z_c) = (\frac{x_r + x_l}{2} Z_c, \frac{y_r + y_l}{2} Z_c, Z_c),$$

was die Szenenrekonstruktion abschließt.

Die Einfachheit der Lösung beruht auf der einfachen geometrischen Darstellung eines parallelen Stereosystems. Im allgemeinen entsprechen parallele Stereogeometrien aber nicht der Aufnahmesituation, da so abgebildete Objekte meist nicht vollständig in beiden Bildern zu sehen sind. In den folgenden Abschnitten wird daher ein Rahmen für die Berechnungen in einer allgemeinen binokularen Stereogeometrie gegeben. Darüber hinaus ist zur Berechnung der Tiefe nach (2.3) exaktes a-priori-Wissen über die Brennweite und die Stereobasisbreite erforderlich. Auf diese Information kann unter Umständen verzichtet werden. Aber selbst an diesem einfachen Beispiel kann man typische Prinzipien der Rekonstruktion ablesen:

- Wie (2.2) zeigt, trägt die horizontale Disparität die wesentliche Tiefeninformation.
- Wie (2.3) zeigt, ist die Tiefenbestimmung abhängig von der Stereogeometrie.
- Die Einführung eines Referenzkoordinatensystems erleichtert die (rechnerische) Beschreibung des Stereosystems.

- Eine Aufnahme der Szene aus zwei Perspektiven reicht in fast allen Fällen für eine zumindest qualitative Bestimmung der Tiefe aus. Deswegen ist das binokulare Stereosystem das am meisten verwendete. Man beachte aber in (2.3) den Fall $d_x = 0$.

Gegenstand dieser Arbeit ist im wesentlichen der dritte und letzte Schritt. Im folgenden wird also untersucht, welche Interpretationen der 3D-Struktur korrespondierende Bildpunkte zulassen. Es wird dabei die Existenz einer genügenden Zahl von Bildpunktpaaren vorausgesetzt. Auswirkungen dieses Schrittes auf den Zuordnungsschritt sowie Art und Anzahl der Korrespondenzen werden aber diskutiert.

2.2 Das Projektionsverhalten einer Kamera

Trotz der technischen Komplexität moderner Kameras können sie zu Berechnungszwecken sehr effektiv durch das einfache Lochkameramodell mit einem Zentralprojektionspunkt, auch *optisches Zentrum* genannt, beschrieben werden (Abbildung 2.3). Dennoch erweist sich die einfache Beschreibung (2.1) über den Strahlensatz als zu simplifiziert für reale Kameras. In der Tat legt das Zentralprojektionsmodell die Beschreibung der Kamera als *projektives* Instrument nahe, das die Welt als dreidimensionalen projektiven Raum $\mathcal{P}^3$ auf die Bildebene als zweidimensionalen projektiven Raum $\mathcal{P}^2$ linear abbildet. Demnach läßt sich diese Funktion als 3×4-*Projektionsmatrix* P vom Rang 3 darstellen, und zwischen einem Szenepunkt M und seinem Bild m besteht die Beziehung

$$\boldsymbol{m} = P\boldsymbol{M}, \tag{2.6}$$

wobei $\boldsymbol{m}$ und $\boldsymbol{M}$ zugehörige Koordinatenvektoren der Länge 3 bzw. 4 sind. Man beachte, daß bei der algebraischen Beschreibumg projektiver Größen durch Koordinaten skalare Vielfache, die nicht Null sind, dieselbe Größe beschreiben. Somit ist P durch $12 - 1 = 11$ Parameter bestimmt. Die Kraft des projektiven Ansatzes liegt vor allem in der Möglichkeit, Lineare Algebra für Berechnungen und Beschreibungen einzusetzen.

Bei Kenntnis der Projektionsmatrix P einer Kamera lassen sich die Koordinaten des optischen Zentrums C ermitteln. Da die Zentralprojektion nur für das optische Zentrum undefiniert ist, sind seine Koordinaten durch den eindimensionalen Kern der Matrix P bestimmt, also durch

$$P\boldsymbol{C} = \mathbf{0}. \tag{2.7}$$

Man kann C auch in expliziter Form angeben. Dazu zerlegt man P in eine 3×3-Matrix $\bar{P}$ mit $\operatorname{rang}(\bar{P}) = 3$ und einen Vektor $\boldsymbol{p}$ der Länge 3:

$$P = (\bar{P}|\boldsymbol{p}) \tag{2.8}$$

Dann gilt für $\boldsymbol{C} = (\bar{\boldsymbol{C}}^T, 1)^T$:

$$\bar{\boldsymbol{C}} = -\bar{P}^{-1}\boldsymbol{p} \tag{2.9}$$

Des weiteren läßt sich der *Sehstrahl* zu einem Bildpunkt bestimmen. Dies ist die Gerade durch Bildpunkt und optisches Zentrum. Ein Punkt auf der Gerade, nämlich das optische Zentrum, ist bereits ermittelt. Als zweiten Punkt wählt man den unendlich fernen Punkt D der Gerade mit Koordinaten $(\bar{\boldsymbol{D}}^T, 0)^T$. Aus (2.6) und (2.8) folgt

$$\bar{\boldsymbol{D}} = \bar{P}^{-1}\boldsymbol{m}. \tag{2.10}$$

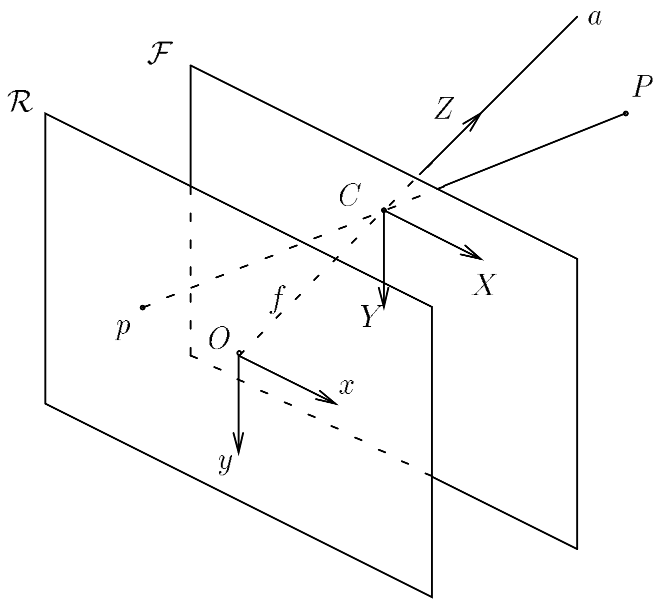

Abbildung 2.3: Kamerageometrie. Der Szenepunkt P wird durch Zentralprojektion mit Brennweite f am optischen Zentrum C auf den Bildpunkt p in der Bildebene $\mathcal{R}$ abgebildet. Punkte, die in der zu $\mathcal{R}$ parallelen Ebene durch C liegen, in der fokalen Ebene $\mathcal{F}$, werden auf die unendlich ferne Ebene abgebildet, lediglich C hat kein Bild unter der Zentralprojektion. Zur Euklidischen Beschreibung ist ein kartesisches Weltkoordinatensystem XYZ an C fixiert, wobei die Z-Achse mit der optischen Achse a der Kamera zusammenfällt. Damit ist ein Bildkoordinatensystem (x, y) entsprechend $(X, Y, -f)$ assoziiert, dessen Ursprung O im Schnittpunkt der optischen Achse mit der Bildebene liegt und Bildzentrum genannt wird.

Somit erhält man den Sehstrahl in parametrischer Form:

$$\boldsymbol{M} = \bar{P}^{-1}(-\boldsymbol{p} + \lambda \boldsymbol{m}), \quad -\infty \leq \lambda \leq \infty,$$

wobei λ ein projektiver Parameter ist.

Die Projektionsmatrix P läßt sich auch in geometrischer Weise beschreiben: Schreibt man P als drei Zeilenvektoren der Länge 4 untereinander

$$P = \begin{pmatrix} \boldsymbol{Q}_1{}^T \\ \boldsymbol{Q}_2{}^T \\ \boldsymbol{Q}_3{}^T \end{pmatrix},$$

so definiert für $i \in \{1, 2, 3\}$ $\boldsymbol{Q}_i{}^T \boldsymbol{M} = 0$ eine Ebene im $\mathcal{P}^3$. Wegen (2.7) gehen alle diese Ebenen durch das optische Zentrum. Nach Konvention liegen auf $\boldsymbol{Q}_3{}^T \boldsymbol{M} = 0$ die Punkte, die bei der Zentralprojektion auf die unendlich ferne Gerade in der Bildebene abgebildet werden. Es handelt sich hierbei also um die *fokale Ebene*, die parallel zur Bildebene durch das optische Zentrum geht (siehe Abbildung 2.3).

Bislang wurde die Kamera als abstraktes projektives Instrument beschrieben. Man kann ihrem Verhalten jedoch auch „von unten“ durch heuristische Überlegungen näherkommen. Dazu kehren wir noch einmal zur einfachen Zentralprojektionsgleichung und zu einem *Euklidischen* Modell zurück.

Zur rechnerischen Beschreibung sei ein kartesisches Koordinatensystem XYZ am Zentralprojektionspunkt fixiert, wobei die Z-Achse mit der optischen Achse der Kamera zu-

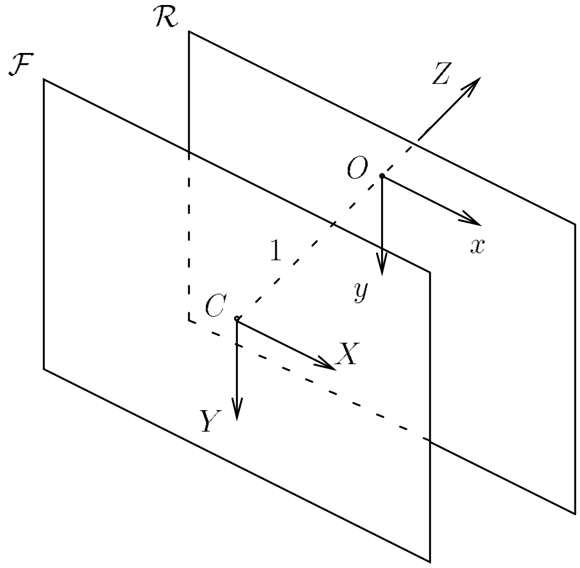

Abbildung 2.4: Geometrie einer normalisierten Kamera. Die Bildebene liegt in der Ebene $Z = 1$. Dies entspricht einer Brennweite 1. Das assoziierte Bildkoordinatensystem ist (x, y) entsprechend $(X, Y, 1)$. Man spricht dann von normalisierten Bildkoordinaten. Eine Kamera läßt sich bei Kenntnis der internen Parameter normalisieren und wird so zu einem Euklidischen Instrument.

sammenfalle. Die Zentralprojektionsgleichung für eine normierte Bildebene (Brennweite 1, Ebene $Z = 1$) ergibt sich aus dem Strahlensatz zu

$$(x, y) = (\frac{X}{Z}, \frac{Y}{Z}), \tag{2.11}$$

wobei X, Y, Z die Koordinaten des abgebildeten Szenepunktes im Kamerakoordinatensystem und x, y die Koordinaten des Bildpunktes, der dem Punkt $(x, y, 1)^T$ im Kamerakoordinatensystem entspricht, sind (Abbildung 2.4). Die Koordinaten von x und y heißen auch *normalisierte Bildkoordinaten*, die modellierte Kamera *normalisierte Kamera*. Den Punkt $\boldsymbol{x}_s = (x, y, 1)^T$ auf der normierten Bildebene nennt man ebenfalls *Sehstrahl*. Durch die Normierung der Bildebene ist die Gerade eindeutig bestimmt. Er steht zum betrachteten Szenepunkt $\boldsymbol{X} = (X, Y, Z)^T$ in der Beziehung

$$Z\boldsymbol{x}_s = \boldsymbol{X}. \tag{2.12}$$

Um reale Kameras besser modellieren zu können, muß jedoch die Koordinatentransformation komplexer beschrieben werden. Geht man nicht von Sehstrahlen, sondern von den Koordinaten der Bildpunkte in Pixeln aus, so muß die Geometrie des bildgebenden Sensors miteingehen. Sie äußert sich mathematisch in einer Transformation des Bildkoordinatensystems, nämlich einer Skalierung der x-Koordinate um s_x, einer Skalierung der y-Koordinate um s_y und einer Verschiebung des Punktes $(0, 0)$ nach (x_0, y_0). (x_0, y_0) ist das *Bildzentrum* in Pixelkoordinaten, also der Punkt, in dem sich optische Achse und Bildebene schneiden. In Koordinatenschreibweise ergibt sich so für Bildpunkte (x_p, y_p) in Pixeln

$$x_p = s_x x + x_0, \tag{2.13}$$
$$y_p = s_y y + y_0. \tag{2.14}$$

Nach Anwendung der Transformation

$$x = \frac{x_p - x_0}{s_x}, \tag{2.15}$$

$$y = \frac{y_p - y_0}{s_y} \tag{2.16}$$

kann man dann mit einer normalisierten Kamera arbeiten. Diese Transformation läßt sich aber auch auf die 3D-Punkte im Kamerakoordinatensystem anwenden. Definiert man

$$S =_{\text{def}} \begin{pmatrix} s_x & 0 & x_0 \\ 0 & s_y & y_0 \\ 0 & 0 & 1 \end{pmatrix},$$

so beschreibt S diese Transformation durch

$$\boldsymbol{X}_p = S\boldsymbol{X},$$

wobei sich dann (2.13) und (2.14) aus Anwendung der Projektionsgleichung (2.11) auf $\boldsymbol{X}_p$ ergeben. s_x, s_y, x_0, y_0 heißen *interne Kameraparameter*. Gilt $x_0 = y_0 = 0$ und $s_x = s_y = f$, so erhält man die Projektionsgleichung (2.1). Diese Beschreibung von S geht davon aus, daß das Bildkoordinatensystem stets orthogonale Achsen besitzt. Berücksichtigt man davon abweichende Koordinatensysteme, so äußert sich dies in einer *Scherung* s_{xy}, wobei dann $S_{12} = s_{xy} \neq 0$ ist. Die Transformation des Bildkoordinatensystems ist somit nicht mehr Euklidisch, sondern affin. Reale Kameras erfüllen jedoch meist $s_{xy} = 0$.

Benutzt man neben dem Kamerakoordinatensystem $X_kY_kZ_k$ noch ein Bezugs-Weltkoordinatensystem $X_wY_wZ_w$, z. B. ein zyklopisches Koordinatensystem (siehe §2.1) oder das Koordinatensystem der anderen Kamera, so ergibt sich eine zusätzliche Euklidische Koordinatentransformation, die in der üblichen Weise durch eine Rotation R und eine Translation $\boldsymbol{t}$ beschrieben wird:

$$R^{k\to w} = \begin{pmatrix} \boldsymbol{R}_1^{k\to w\,T} \\ \boldsymbol{R}_2^{k\to w\,T} \\ \boldsymbol{R}_3^{k\to w\,T} \end{pmatrix} \qquad \boldsymbol{t}^{k\to w} = \begin{pmatrix} t_x^{k\to w} \\ t_y^{k\to w} \\ t_z^{k\to w} \end{pmatrix}$$

$$U^{k\to w} =_{\text{def}} (R^{k\to w}|\boldsymbol{t}^{k\to w})$$

Hier ist $R^{k\to w}$ eine 3×3-Rotationsmatrix und $\boldsymbol{t}^{k\to w}$ ein Translationsvektor der Länge 3. Somit ergibt sich eine 3×4-Transformationsmatrix $U^{k\to w}$. Die Indizierung $k \to w$ besagt, daß das Kamerakoordinatensystem in das Weltkoordinatensystem überführt wird, wobei der Vektor $\boldsymbol{t}^{k\to w}$ dann im Koordinatensystem von k beschrieben ist. Also:

$$\boldsymbol{X}_k = U^{k\to w}\boldsymbol{X}_w,$$

wobei $\boldsymbol{X}_k = (X_k, Y_k, Z_k)^T$ und $\boldsymbol{X}_w = (X_w, Y_w, Z_w, 1)^T$. $U^{k\to w}$ enthält die sogenannten *externen Kameraparameter*.

Fügt man interne und externe Kameraparameter zusammen, so ergibt sich insgesamt eine Koordinatentransformation $T^{k\to w}$ der Form

$$T^{k\to w} =_{\text{def}} SU^{k\to w} = (P^{k\to w}|\boldsymbol{p}^{k\to w}) = \left(\begin{array}{c|c} \boldsymbol{P}_1^{k\to w\,T} & p_x^{k\to w} \\ \boldsymbol{P}_2^{k\to w\,T} & p_y^{k\to w} \\ \boldsymbol{P}_3^{k\to w\,T} & p_z^{k\to w} \end{array}\right), \tag{2.17}$$

wobei $P^{k\to w} =_{\text{def}} SR^{k\to w}$, $\boldsymbol{p}^{k\to w} =_{\text{def}} S\boldsymbol{t}^{k\to w}$, also $T^{k\to w}$ eine 3×4-Matrix ist. Den Vektor $\boldsymbol{X}_{p_k} = (X_{p_k}, Y_{p_k}, Z_{p_k})^T$, der im Kamerakoordinatensystem beschrieben wird und interne Kameraparameter berücksichtigt, erhält man so aus

$$\boldsymbol{X}_{p_k} = T^{k\to w}\boldsymbol{X}_w, \tag{2.18}$$

wobei $\boldsymbol{X}_w = (X_w, Y_w, Z_w, 1)^T$. Auf diese transformierten Koordinaten kann jetzt wiederum die Zentralprojektionsgleichung (2.11) angewendet werden, um die Bildkoordinaten in Pixel zu erhalten.

$$(x_{p_k}, y_{p_k}) = (\frac{X_{p_k}}{Z_{p_k}}, \frac{Y_{p_k}}{Z_{p_k}}) \tag{2.19}$$

Die Analogie zum projektiven Modell wird hier deutlich. Die Matrix $T^{k\to w}$ beschreibt eine nicht-injektive, projektiv lineare Funktion, entsprechend der Projektionsmatrix P. Ihre Argumente sind „Euklidische Weltpunkte“ im Raum $\mathcal{P}^3$, die in eindeutiger Form durch homogene Koordinaten beschrieben werden können, deren letzte Komponente 1 ist. Ihre Resultate sind „Euklidische Bildpunkte“ im Raum $\mathcal{P}^2$, die in eindeutiger Form durch inhomogene Koordinaten gemäß (2.19) darstellbar sind, oder äquivalent durch Normierung der letzten Komponente auf 1. $T^{k\to w}$ ist durch 5 interne (x_0, y_0, s_x, s_y, s_{xy}) und 6 externe (3 Rotations-, 3 Translations-) Parameter, also insgesamt 11, bestimmt, wie es sich auch für P aus allgemeinen Überlegungen ergibt. Man beachte aber, daß $T^{k\to w}$ eine Anwendung auf Euklidische Strukturen voraussetzt. Mit dem Zusammenhang der verschiedenen 3D-Strukturen setzt sich der übernächste Abschnitt auseinander.

Den Vorgang, die Transformationsmatrix $T^{k\to w}$ bzw. P und eventuell getrennt externe und interne Parameter der Kamera zu bestimmen, nennt man *Kalibrierung*. P ist als nicht-injektive Funktion durch 6 Punkte, von denen je 5 nicht komplanar sind, eindeutig bestimmt, gezeigt in [Fau93]. Somit genügt die Kenntnis von 6 Szene- und Bildpunkten geeigneter Form, aus Robustheitsgründen werden jedoch meist weitaus mehr benutzt. Das klassische Kalibrierungsverfahren verwendet einen Kalibrierungswürfel (Abbildung 2.1) oder vergleichbare gemusterte Objekte mit bekannten Ausmaßen. In einem lokalen Euklidischen Koordinatensystem sind somit die Koordinaten von ausgezeichneten Eckpunkten bekannt, deren Bilder in Pixelkoordinaten dann im Bild detektiert und zugeordnet werden müssen. Ein ausführliche Darstellung des Kalibrierungsvorgangs und seiner Grundlagen findet sich in [Fau93].

2.3 Die Fundamentalmatrix

Ein wichtiges algebraisches Mittel, um eine allgemeine binokulare Stereogeometrie (Abbildung 2.5) zu beschreiben, ist die sogenannte *Fundamentalmatrix*. Die Informationen, die sie enthält und die mit ihr gewonnen werden können, wurden in letzter Zeit häufig untersucht und ihre Bedeutung für SfS erkannt (siehe [LF96b]). Es ergeben sich ähnliche Betrachtungen für ein Stereosystem, wie sie die Untersuchung der Projektionsmatrix für eine Kamera liefert.

Die Fundamentalmatrix ermöglicht die algebraische Manifestierung der *epipolaren Bedingung* (Abbildung 2.5): Man betrachte zwei korrespondierende Bildpunkte m und m' im linken und rechten Bild, die Bilder desselben Szenepunktes M sind. Da M auf dem Sehstrahl von m liegt, muß m' auf der Projektion dieses Strahls auf die rechte Bildebene

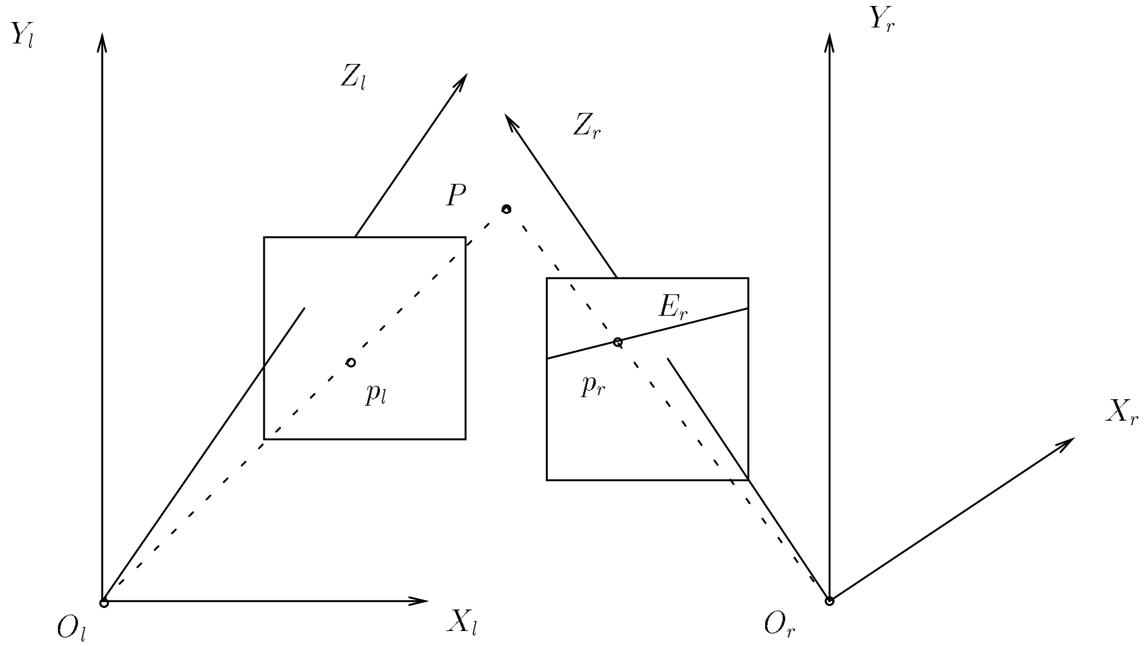

Abbildung 2.5: Binokulare nicht-parallele Stereogeometrie. Der Punkt P wird durch Zentralprojektion am Punkt O_l auf den Punkt p_l in der linken Bildebene abgebildet. p_l ist der Schnitt der linken Bildebene mit dem Sehstrahl, i. e. die Gerade durch O_l und P. Der Bildpunkt p_r in der rechten Bildebene muß deswegen auf der Projektion des linken Sehstrahls auf die rechte Bildebene liegen. Diese ergibt sich als Schnitt der rechten Bildebene mit der Ebene durch O_l, O_r, P und heißt Epipolargerade E_r. Ihre Kenntnis beschränkt die Suche im Zuordnungsschritt auf einen eindimensionalen Raum.

liegen. Sie ergibt sich als Schnitt der Ebene durch beide optische Zentren und M (bzw. m) mit der rechten Bildebene. Diese Schnittgerade heißt *Epipolargerade*. Sie beschränkt das Suchgebiet im Zuordnungsschritt auf einen eindimensionalen Raum. *Epipole* heißen die Projektionen der optischen Zentren bezüglich der jeweils anderen Kamera. Aus der geometrischen Interpretation der Epipolargeraden schließt man, daß alle diese Geraden durch den Epipol der Kamera gehen.

Die grundlegende Beobachtung ist nun, daß die Abbildung eines Bildpunktes auf seine korrespondierende Epipolargerade projektiv linear ist: Man kann nach dem Dualitätsprinzip den Raum der Bildpunkte mit dem Raum der Sehstrahlen identifizieren. Die Beziehung zu den Epipolargeraden im anderen Bild ist dann projektiv linear, da es sich um eine Zentralprojektion handelt. Die Fundamentalmatrix F ist nun die 3×3-Matrix, die diese Abbildung von $\mathcal{P}^2$ nach $\mathcal{P}^2$ beschreibt durch

$$\boldsymbol{l}'_{\boldsymbol{m}} = F\boldsymbol{m},$$

wobei $\boldsymbol{m}$ einen Bildpunkt und $\boldsymbol{l}'_{\boldsymbol{m}}$ die zugehörige Epipolargerade darstellen. Solche Abbildungen, die Punkte auf Geraden abbilden, heißen *Korrelationen*. Da die Epipolargerade zu einem Epipol $\boldsymbol{e}$ nicht definiert ist, gilt

$$F\boldsymbol{e} = \boldsymbol{0}$$

und also $\det(F) = 0$. Der zu m korrespondierende Bildpunkt m' mit Koordinatenvektor $\boldsymbol{m}'$ liegt auf $\boldsymbol{l}_{\boldsymbol{m}'}$, woraus folgt:

$$\boldsymbol{m}'^T F\boldsymbol{m} = 0 \tag{2.20}$$

Dies ist die epipolare Bedingung in algebraischer Form. Vertauschung der Rollen der Kameras liefert analog

$$\boldsymbol{m}^T F^T \boldsymbol{m}' = 0.$$

Die Fundamentalmatrix läßt sich auch mit Hilfe der (zerlegten) Projektionsmatrizen beider Kameras ausdrücken. Wir kennen bereits den Sehstrahl zu m, festgelegt durch das optische Zentrum C (2.9) und den unendlich fernen Punkt D (2.10) des Strahls. Die Projektionen auf die andere Bildebene sind der Epipol

$$\boldsymbol{e}' = P'\boldsymbol{C} \stackrel{(2.9)}{=} (\bar{P}'|\boldsymbol{p}') \begin{pmatrix} -\bar{P}^{-1}\boldsymbol{p} \\ 1 \end{pmatrix} = \boldsymbol{p}' - \bar{P}'\bar{P}^{-1}\boldsymbol{p}$$

und

$$\boldsymbol{d}' = P'\boldsymbol{D} \stackrel{(2.10)}{=} (\bar{P}'|\boldsymbol{p}') \begin{pmatrix} \bar{P}^{-1}\boldsymbol{m} \\ 0 \end{pmatrix} = \bar{P}'\bar{P}^{-1}\boldsymbol{m}.$$

Die Epipolargerade ergibt sich dann durch

$$\boldsymbol{l}'_{\boldsymbol{m}} = \boldsymbol{e}' \times \boldsymbol{d}' = (\boldsymbol{p}' - \bar{P}'\bar{P}^{-1}\boldsymbol{p}) \times (\bar{P}'\bar{P}^{-1}\boldsymbol{m}) = [\boldsymbol{p}' - \bar{P}'\bar{P}^{-1}\boldsymbol{p}]_\times \bar{P}'\bar{P}^{-1}\boldsymbol{m},$$

wobei $[]_\times$ die aus dem Kreuzprodukt hervorgehende antisymmetrische Matrix bezeichnet. Da dies für alle Bildpunkte gilt, existiert die Zerlegung

$$F = [\boldsymbol{p}' - \bar{P}'\bar{P}^{-1}\boldsymbol{p}]_\times \bar{P}'\bar{P}^{-1}.$$

Die Beschreibung der Projektionsmatrizen in der Euklidischen Form (2.17) liefert die analoge Beschreibung der Fundamentalmatrix mit den Parametern der Euklidischen Koordinatentransformation:

$$\begin{aligned}
F &= [\boldsymbol{p}^{r\to w} - P^{r\to w}(P^{l\to w})^{-1}\boldsymbol{p}^{l\to w}]_\times P^{r\to w}(P^{l\to w})^{-1} \\
&= [S_r\boldsymbol{t}^{r\to w} - S_r R^{r\to w}(S_l R^{l\to w})^{-1} S_l \boldsymbol{t}^{l\to w}]_\times S_r R^{r\to w}(S_l R^{l\to w})^{-1} \\
&= [S_r(\boldsymbol{t}^{r\to w} - R^{r\to l}\boldsymbol{t}^{l\to w})]_\times S_r R^{r\to l} S_l^{-1} \\
&= [S_r \boldsymbol{t}^{r\to l}]_\times S_r R^{r\to l} S_l^{-1} \qquad (2.21)
\end{aligned}$$

Mit $Q =_{\text{def}} S_r R^{r\to l} S_l^{-1}$ und $\boldsymbol{s} =_{\text{def}} S_r \boldsymbol{t}^{r\to l}$ heißt $F = [\boldsymbol{s}]_\times Q$ auch die *Q-$\boldsymbol{s}$-Repräsentation* der Fundamentalmatrix. An der Rechnung liest man die Darstellungen für die Epipole ab:

$$\boldsymbol{e}_r = S_r \boldsymbol{t}^{r\to l} \qquad \boldsymbol{e}_l = S_l \boldsymbol{t}^{l\to r} \qquad (2.22)$$

Aus historischer Sicht liegt jedoch eine andere Vorgehensweise vor, die von normalisierten Bildpunktkoordinaten ausgeht. Seien Sehstrahlen $\boldsymbol{x}_l$ und $\boldsymbol{x}_r$ zu einem Szenepunkt gegeben. Dann ergibt sich über (2.12) die Beziehung zwischen Sehstrahlen, Tiefen und relativer Bewegung zu

$$Z_r \boldsymbol{x}_r = R^{r\to l} Z_l \boldsymbol{x}_l + \boldsymbol{t}^{r\to l}.$$

Bildet man nun das Vektorprodukt mit $\boldsymbol{t}^{r\to l}$ von links

$$\boldsymbol{t}^{r\to l} \times Z_r \boldsymbol{x}_r = \boldsymbol{t}^{r\to l} \times R^{r\to l} Z_l \boldsymbol{x}_l,$$

so eliminiert das Skalarprodukt mit $\boldsymbol{x}_r$ die Tiefen:

$$\boldsymbol{x}_r{}^T(\boldsymbol{t}^{r\to l} \times R^{r\to l}\boldsymbol{x}_l) = 0$$

Man erhält die *essentielle Matrix*

$$E = [\boldsymbol{t}^{r\to l}]_\times R^{r\to l}$$

und die Longuet-Higgins-Gleichung

$$\boldsymbol{x}_r{}^T E \boldsymbol{x}_l = 0.$$

Analog erhält man die Beschreibung

$$E = [\boldsymbol{t}^{l\to r}]_\times R^{l\to r}. \tag{2.23}$$

Diese Gleichung war Ausgangspunkt für viele Betrachtungen zu SfM, da sie ohne Tiefeninformation Bedingungen an die Bewegung stellt, was dann auch zur Theorie über die Fundamentalmatrizen führte. Wiedereinführung von Pixelkoordinaten liefert eine letzte Darstellung der Fundamentalmatrix:

$$F = S_r^{-T} E S_l^{-1} = S_r^{-T} [\boldsymbol{t}^{r\to l}]_\times R^{r\to l} S_l^{-1} \tag{2.24}$$

Die Darstellungen (2.21) und (2.24) stehen miteinander über $A\boldsymbol{a} \times A\boldsymbol{b} = \det(A) A^{-T}(\boldsymbol{a} \times \boldsymbol{b})$ für reguläre 3×3-Matrizen A und entsprechende Vektoren $\boldsymbol{a}$ und $\boldsymbol{b}$ in Beziehung.

Stereosysteme, für die nur die Fundamentalmatrix bekannt ist, heißen *schwach kalibriert*. Da F eine projektive Größe ist und aufgrund der Nebenbedingung $\det(F) = 0$, ist F durch $9 - 1 - 1 = 7$ Parameter bestimmt. Da man im allgemeinen nur Punktkorrespondenzen aus dem Zuordnungsschritt erhält, bestimmt man F aus mindestens 7, besser mehr korrespondierenden Bildpunktepaaren über die Epipolarbedingung (2.20). Die essentielle Matrix ist aufgrund der Skalierungsmehrdeutigkeit der Translation durch $6 - 1 = 5$ Parameter der Bewegung bestimmt. Man erhält somit für die essentielle Matrix außer ihrer Skalierungsmehrdeutigkeit Nebenbedingungen, nämlich die verschwindete Determinante und die Gleichheit der zwei nicht verschwindenden Singulärwerte (siehe [Fau93]). Die Fundamentalmatrix bildet eine gute Grundlage für Strukturanalyse in unkalibrierten Stereosystemen, besonders falls die internen Parameter unbekannt sind.

2.4 Stratifikation des Raumes

Die bisherigen geometrischen Beschreibungen eines Stereosystems machen eine genauere Betrachtung des Problemrahmens notwendig. Aus mathematischer Sicht ist mit einer exakten Problemdefinition immer ein mathematischer Raum verbunden, in dem Beziehungen zwischen den Objekten studiert und Berechnungen durchgeführt werden. Gegenstand der SfS-Problematik ist die reale Welt und die Bildebene, die man sich in der Regel als Euklidische Räume vorstellt, in denen es Sinn macht, von Winkeln und Entfernungen zu sprechen und diese zu messen. Die Einführung einer Kamera als projektives Instrument, durch eine Zentralprojektion beschrieben, führt aber bereits zu Schwierigkeiten, da im projektiven Raum Winkel und Entfernungen nicht sinnvoll definiert werden können.

Die klassischen geometrischen Beschreibungen des Raumes als projektiv, affin und Euklidisch sind oft unabhängig voneinander untersucht worden. Gleichwohl ist bekannt, daß der projektive Raum den allgemeinsten Fall darstellt, daß man durch affine Ableitung einen affinen Raum erhält, und daß ein Euklidischer Raum einen Spezialfall eines affinen

Raumes darstellt (siehe z. B. [Fis92]). Die Arbeit [Fau95] macht sich dies zunutze. Faugeras schlägt vor, sich die Welt als drei Strukturen vorzustellen, die als Hüllen übereinander liegen. Dabei ist die projektive Hülle die allgemeinste, die affine Hülle in ihr enthalten, die Euklidische Hülle wiederum in der affinen. Dies ist die sogenannte *Stratifikation* des Raumes. Mit dieser Idee verbunden ist das alte geometrische Konzept, in den Strukturen nach Transformationen zu suchen, die die interessierenden Aussagen invariant lassen. So ist mit jeder Hülle eine Transformationsgruppe verbunden, so daß Strukturen immer „bis auf eine Transformation“ beschrieben sind. Deshalb behalte man immer im Auge, daß eine gewonnene Koordinatendarstellung einer Szene von der Wahl des Koordinatensystems bzw. einer geeigneten Basis, abhängig ist. Tabelle 2.1 zeigt die mit jeder Struktur verbundenen Invarianten und Transformationen. Dabei ergibt sich auch für die Transformationen eine entsprechende Hüllenbeziehung. Die Darstellungen können als Folgerungen spezieller Invarianzen begriffen werden: Eine affine Transformation wird als projektive Transformation aufgefaßt, die die unendlich ferne Hyperebene invariant läßt. Eine Euklidische Transformation oder Ähnlichkeit wird als affine Transformation angesehen, die die absolute Quadrik $\sum_{i=1}^{n} x_i = x_n = 0$ invariant läßt. Mit Hilfe dieser invarianten Strukturen können dann die üblichen invarianten Eigenschaften wie Teilverhältnis oder Winkel berechnet werden.

Noch ein Wort zur Koordinatendarstellung: Punkte, die nicht auf der unendlich fernen Hyperebene liegen, können durch eine einfache Koordinatenabbildung

$$(X_1, \ldots, X_n, X_{n+1}) \mapsto (\frac{X_1}{X_{n+1}}, \ldots, \frac{X_n}{X_{n+1}})$$

in affinen statt in projektiven Koordinaten dargestellt werden. Umgekehrt können Punkte des affinen Raumes in den projektiven Raum mittels der Koordinatenabbildung

$$(X_1, \ldots, X_n) \mapsto (X_1, \ldots, X_n, 1)$$

eingebettet werden. Diese Tatsache wurde bereits bei der projektiven und Euklidischen Darstellung der Projektionsmatrix ausgenutzt. In der Tat stellen projektiv lineare Funktionen im gewöhnlichen Euklidischen Raum $\boldsymbol{R}^n$ nichts anderes dar als gebrochen rationale Funktionen in jeder Komponente. Diese haben zwar nicht die Invarianten Euklidischer Transformationen, aber man kann auf sie den gesamten Apparat der klassischen Analysis, insbesondere der Differentialgeometrie, loslassen.

Im Sinne des SfS-Problems können jetzt die einzelnen Strukturen aus unterschiedlichen Informationen gewonnen werden. Eine projektive Rekonstruktion ist allein mit Hilfe des Fundamentalmatrix möglich. Eine affin invariante Repräsentation der Szene kann mit Hilfe der unendlich fernen Ebene ermittelt werden. Dabei nutzt man aus, daß diese wie fast jede andere Ebene eine Kollineation zwischen den Bildebenen induziert. Die Euklidische Struktur der Szene, genauer die Struktur bis auf eine Ähnlichkeit, kann durch Kenntnis der Bilder der absoluten Quadrik bestimmt werden. Diese zusätzliche Information ermöglicht die Berechnung von Winkeln zwischen Sehstrahlen und Teilverhältnissen als Euklidischen Invarianten. Darüber hinaus liefert die Quadrik-Gleichung eine affine Transformation der Bildebene, die im Sinne der internen Parameter interpretiert werden kann, ohne eine heuristische Einführung.

Hier zeigt sich die Stärke, die Transformationen über spezielle Invarianten im projektiven Raum auszuzeichnen. In [Fau95] wird nämlich gezeigt, daß rein aus Bilddaten alle

Struktur	Transformation	Invarianten	Matrix-Darstellung	Parameter
projektiv	projektiv linear	Kollinearität Komplanarität Doppelverhältnis	M: 3×3-Matrix $\boldsymbol{u}$, $\boldsymbol{v}$: 3D-Vektor μ: Skalar $H = \begin{pmatrix} M & \boldsymbol{u} \\ \boldsymbol{v}^T & \mu \end{pmatrix}$	15
affin	affin	Parallelität Schwerpunkt Teilverhältnis	N: reg. 3×3-Matrix $\boldsymbol{u}$: 3D-Vektor ν: Skalar $\neq 0$ $A = \begin{pmatrix} N & \boldsymbol{u} \\ \mathbf{0}^T & \nu \end{pmatrix}$	12
Euklidisch	Ähnlichkeit	Winkel relative Entfernung	R: Rotationsmatrix $\boldsymbol{t}$: 3D-Vektor ν: Skalar $\neq 0$ $S = \begin{pmatrix} R & \boldsymbol{t} \\ \mathbf{0}^T & \nu \end{pmatrix}$	7
Euklidisch	Kongruenz	Entfernung	R: Rotationsmatrix $\boldsymbol{t}$: 3D-Vektor $D = \begin{pmatrix} R & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{pmatrix}$	6

Tabelle 2.1: Geometrische Räume und ihre Transformationsgruppen. Nach dem Stratifikationsprinzip sind die Strukturen von unten nach oben ineinander enthalten. Ebenso stehen die Transformationen in Teilgruppenbeziehung. Daher gelten Invarianzen in „höheren Gruppen" selbstverständlich auch in „niedrigeren".

(nach [LV94])

hier erwähnten Strukturen mit ihren Invarianten gewonnen werden können. Bei relativ schwachen Anforderungen wie in der projektiven Struktur genügt dazu relativ wenig Information, die schnell gewonnen werden kann. Je mehr Information in Form von Invarianten die Struktur aber enthält, desto schwieriger ist ihre Gewinnung. So ist zur Ermittlung der affinen und Euklidischen Struktur a-priori-Wissen und Aktivität des Beobachters zwar nicht notwendig, erleichtert aber viele Prozesse.

Im Licht dieser Überlegung stellt sich das SfS-Problem neu dar. In erster Linie ist es wichtig, zwischen projektiver, affiner und Euklidischer Struktur zu unterscheiden und dabei die Wahl der Basis bzw. des Koordinatensystems zu berücksichtigen. Dann stellt sich die Frage, welche Struktur die benötigte Information enthält. Da ohne a-priori-Wissen die projektive Struktur am einfachsten zu bestimmen ist und die affine immer noch leichter als die gesamte Euklidische Struktur, wurde in letzter Zeit ausgiebig untersucht, welche Informationen in zuverlässiger Form vorliegen und welche konkreten Aufgaben des Maschinensehens damit zu lösen sind. Beispielsweise ist aber Tiefeninformation aus rein projektiver Struktur nicht erhältlich. Später werden wir sehen, wie man a-priori-Wissen und Euklidische sowie projektive Strukturen kombinieren kann, um schnell an nützliche Informationen heranzukommen.

2.4.1 Orientierte Projektive Geometrie

Ein relativ neu diskutierter Ansatz, der die Allgemeinheit projektiver Geometrie bewahrt, aber weiterführende geometrische Betrachtungen ermöglicht, ist der der orientierten projektiven Geometrie ([Sto91], Anwendungen bezüglich 3D-Maschinensehens in [LF96a], [Car96]). Dabei stellt man sich einen *orientierten projektiven Raum* $\mathcal{T}^n$ als zwei Kopien des gewöhnlichen projektiven Raumes $\mathcal{P}^n$ vor, weshalb man auch von einem *zweiseitigen Raum* spricht. Jeder Punkt p hat dabei seine *Antipode* $\neg p$ auf der anderen Seite. Algebraisch wird dies durch die Unterscheidung der homogenen Koordinatenvektoren $\boldsymbol{X}$ und $-\boldsymbol{X}$ erreicht. Dementsprechend gilt für die Koordinatendarstellung $\boldsymbol{x}$ und $\boldsymbol{y}$ orientierter projektiver Größen

$$\boldsymbol{x} \equiv \boldsymbol{y} \quad \Leftrightarrow_{\text{def}} \quad \exists \lambda > 0 : \boldsymbol{x} = \lambda \boldsymbol{y},$$

wobei $\equiv$ die „projektive Gleichheit" bezeichnet. Es ergibt sich ein Raum, der topologisch äquivalent zu einer n-Sphäre S^{n+1} im $\boldsymbol{R}^{n+1}$ ist, in der diametral liegende Punkte *nicht* identifiziert werden. Dies ermöglicht die Definition von zwei Orientierungen für jede geometrische Figur (Abbildung 2.6) und von zwei Seiten einer Hyperbene, die topologisch getrennt sind und je nach Orientierung der Ebene „links" oder „rechts" liegen. Abbildung 2.7 zeigt zwei Modellvorstellungen des zweidimensionalen Raumes. Zur Erfassung des dreidimensionalen Raumes stelle man sich etwa vor, von einem Punkt aus eine Pistole abzufeuern. Die Kugel würde, physikalische Gegebenheiten vernachlässigt, in einer geraden Bahn diese Welt durch die unendlich ferne Ebene verlassen, die gesamte entgegengesetzt orientierte Welt auf derselben Route durchqueren, durch die unendlich ferne Ebene wieder in diese Welt eintreten und den Schützen von hinten treffen.

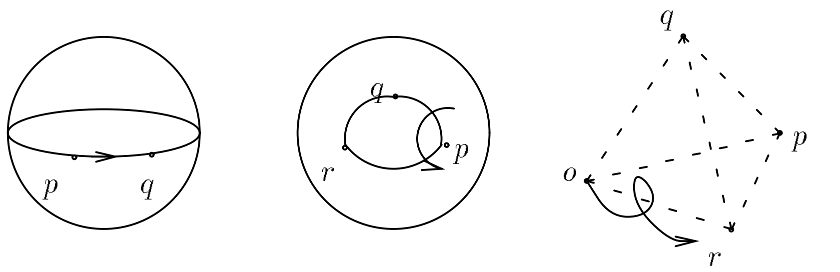

Abbildung 2.6: Orientierung von echten 2-, 3- und 4-Ecken. Formal wird Orientierung als zwei Äquivalenzklassen von orientierten n-Ecken eingeführt. Links: Die Orientierung von $(p; q)$ ergibt sich durch die kürzeste Verbindung von p nach q. Mitte: Die Orientierung von $(p; q; r)$ ergibt sich als Richtung der Drehung der Strecke $\bar{pq}$ nach $\bar{pr}$ um den kürzesten Winkel. Rechts: Die Orientierung von $(o; p; q; r)$ ergibt sich durch die Drehung der Strecke $\bar{op}$ nach $\bar{oq}$ um den kürzesten Winkel, wobei sich gleichzeitig von o nach r bewegt wird.
(aus [Sto91])

Die Modellierungen geben bereits den entscheidenden Hinweis im Hinblick auf die Fortsetzung der Stratifikation. Man stelle sich den gewöhnlichen projektiven Raum als positive Seite der unendlich fernen Ebene eingebettet vor. Algebraisch läßt sich dies durch eine Beschreibung der Punkte mittels Koordinatenvektoren mit positiver letzter Koordinate erreichen, wie es durch $\boldsymbol{X} = (X, Y, Z, 1)^T$ bei Euklidischen Punkten ohnehin getan wird. Jeder dieser Punkte hat dann seine Antipode $-\boldsymbol{X} = (-X, -Y, -Z, -1)^T$ (nicht $(X, Y, Z, -1)^T$!) im negativen Halbraum, der topologisch durch die unendlich ferne Ebene vom gewöhnlichen positiven Raum getrennt ist. Analog ist bei projektiven Transformationen P im orientierten Fall zwischen P und $-P$ zu unterscheiden. Alle projektiven Transformationen stellen in ihrer Matrixform bereits orientierte projektive Transformationen dar, funktionieren aber auf einem erweiterten Definitions- und Wertebereich und lassen sich nicht kanonisch einbetten, da die orientierten projektiven Transformationen, eingeschränkt auf den gewöhnlichen projektiven Raum, nicht identisch mit den ursprünglichen projektiven Transformationen sind. Man kann aber affine und Euklidische Transformationen als orientierte projektive Transformationen auffassen, die die unendlich ferne Ebene invariant lassen und jede ihrer Seiten auf sich abbilden.

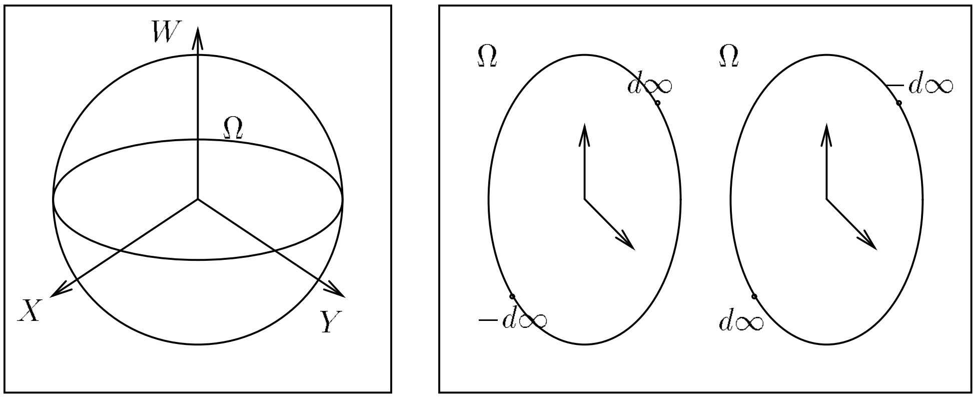

Abbildung 2.7: Modelle für den zweidimensionalen orientierten projektiven Raum. Links: Kugelmodell mit kartesischem Koordinatensystem im Zentrum. Die unendlich ferne Gerade Ω ergibt sich als Schnitt der Kugel mit der Ebene $W = 0$. Rechts: Ebenenmodell mit zwei Seiten bezüglich der unendlich fernen Gerade Ω. $d\infty$ und $-d\infty$ sind diametral auf Ω liegende Punkte, die im orientierten Fall nicht identifiziert werden.

(nach [Sto91])

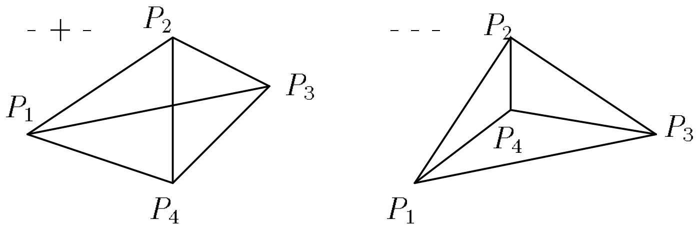

Abbildung 2.8: Signaturen zweier orientierter projektiver Basen (P_1, P_2, P_3, P_4) in der orientierten projektiven Ebene.

Vorteile der orientierten Modellierung liegen in der Möglichkeit der Definition von Orientierung, von zwei Seiten einer Hyperebene und, darauf aufbauend, von Segmenten und Konvexität, die eingehendere geometrische Beschreibungen von Objekten im projektiven Rahmen erlauben. Dazu betrachten wir im $\mathcal{T}^n$ eine *orientierte projektive Basis*, gegeben durch die Koordinatenvektoren $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n+1}, \boldsymbol{x})$, von denen jeweils $n+1$ linear unabhängig sind. Die Lage des *Einheitspunktes* $\boldsymbol{x}$ bezüglich $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n+1})$ ist gegeben durch die *Signatur* der Basis

$$(\sigma_i)_{i=1,\ldots,n+1} =_{\mathrm{def}} \operatorname{sign}(\det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}, \boldsymbol{x}, \boldsymbol{v}_{i+1}, \ldots, \boldsymbol{v}_{n+1})),$$

wobei jedes σ_i die Lage von $\boldsymbol{x}$ bezüglich der durch alle Punkte außer $\boldsymbol{v}_i$ definierten orientierten Hyperebene angibt. In Abbildung 2.8 sind Beispiele für Signaturen angegeben. Für uns wird später wichtig sein, wie man aus gegebenen Matrixdarstellungen projektiv linearer Funktionen durch Auffassung des Raumes als orientiert projektiv zusätzliche Eigenschaften der Funktion gewinnen kann. Daher ist folgende Bemerkung von Bedeutung:

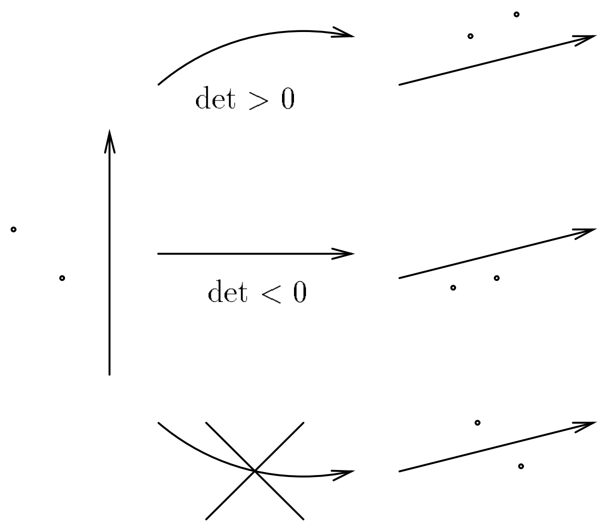

Abbildung 2.9: Mögliche und unmögliche Konfigurationen nach der Anwendung einer orientierten projektiven Transformation in der orientierten projektiven Ebene.

Bemerkung 2.1 Sei eine orientierte projektive Transformation im $\mathcal{T}^n$ gegeben durch die reguläre $(n+1)\times(n+1)$-Matrix P. Dann gilt: **Zwei Punkte liegen genau dann vor der Transformation mit P in demselben Halbraum bezüglich einer Hyperebene, wenn sie es danach tun.** Dabei behalten sie ihre Lage bezüglich der transformierten orientierten Hyperebene bei, falls $\det(P) > 0$, und wechseln sie, falls $\det(P) < 0$.

Beweis: Sei π eine Hyperebene im $\mathcal{T}^n$ und $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ ein Erzeugendensystem von π. Für alle Koordinatenvektoren $\boldsymbol{w}$ gilt dann:

$$\det(P\boldsymbol{v}_1, \ldots, P\boldsymbol{v}_n, P\boldsymbol{w}) = \det(P(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{w})) = \det(P) \cdot \det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{w})$$

Seien $\boldsymbol{x}$ und $\boldsymbol{y}$ Koordinatenvektoren zu zwei Punkten. Dann gilt:

$$\begin{aligned} & \operatorname{sign}(\det(P\boldsymbol{v}_1, \ldots, P\boldsymbol{v}_n, P\boldsymbol{x})) = \operatorname{sign}(\det(P\boldsymbol{v}_1, \ldots, P\boldsymbol{v}_n, P\boldsymbol{y})) \\ \Leftrightarrow \quad & \operatorname{sign}(\det(P) \cdot \det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{x})) = \operatorname{sign}(\det(P) \cdot \det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{y})) \\ \Leftrightarrow \quad & \operatorname{sign}(\det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{x})) = \operatorname{sign}(\det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{y})) \end{aligned}$$

Je nach Vorzeichen von $\det(P)$ bleibt $\operatorname{sign}(\det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{x}))$ und damit $\operatorname{sign}(\det(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n, \boldsymbol{y}))$ erhalten, und die Punkte bleiben auf ihrer Seite der transformierten orientierten Hyperebene, oder nicht. □

Insbesondere besagt diese Bemerkung, daß nur zwischen Basen gleicher oder inverser Signatur orientierte projektive Transformationen existieren. Dies ist plausibel, da zwischen diesen Signaturen ohne Kenntnis der Orientierung des zugrundeliegenden Raumes nicht unterschieden werden kann. Abbildung 2.9 zeigt mögliche und unmögliche Konfigurationen in der orientierten projektiven Ebene.

Ein Nachteil des orientierten Ansatzes ist darin zu sehen, daß man bei projektiven Rekonstruktionen auch aus dieser Welt hinausgelangen kann, was einer anschaulichen Interpretation im Weg steht. Es ergeben sich weitere kleinere Unterschiede, wie ein etwas anderes Doppelverhältnis. Die Darstellung in [Sto91] ist ausführlich und berücksichtigt besonders Berechnungsmodelle.

2.5 Euklidische 3D-Rekonstruktion bei Kenntnis der Kamera-Transformationsmatrizen

Wir wollen jetzt das klassische Verfahren zur Euklidischen Rekonstruktion der Szene für ein binokulares Stereosystem vorstellen. Dabei wird vorausgesetzt, daß linke und rechte Kamera kalibriert wurden. Man erhält hier dann sogar die Szene bis auf eine Kongruenz, da durch die Kalibrierung Referenztiefen vorliegen und die Skalierungsmehrdeutigkeit entfällt. Wie bereits erwähnt, ergeben sich bei allgemeiner Stereogeometrie nicht so einfache geschlossene Formen zur Tiefenberechnung wie in §2.1. Eine Möglichkeit, die sich bei bekannten Transformationsmatrizen der Kameras anbietet, ist die Bestimmung des Szenepunktes als Schnittpunkt zweier korrespondierender Sehstrahlen. Diesen Vorgang nennt man *Triangulation*. Diese Sehstrahlen bekommt man bei Kenntnis der internen Kameraparameter aus den Koordinaten der Bildpunkte in Pixel. Dann benötigt man noch die Beziehung zu einem, nicht notwendigerweise dritten, Koordinatensystem, in dem man den Schnittpunkt der Sehstrahlen berechnen kann.

Seien $X_lY_lZ_l$, $X_rY_rZ_r$ das linke und rechte Kamerakoordinatensystem und $X_wY_wZ_w$ ein Bezugs-Weltkoordinatensystem, so daß die Transformationen $T^{l\to w}$ und $T^{r\to w}$ bekannt sind. Sei $\boldsymbol{X}_w = (X_w, Y_w, Z_w, 1)^T$ ein (Euklidischer) Szenepunkt, ansonsten mögen die Bezeichnungen aus §2.2 gelten. Setzt man (2.18) in die Projektionsgleichung (2.19) ein, so ergibt sich für jede Kamera

$$x_{p_k} = \frac{\boldsymbol{P}_1^{k\to w^T}\boldsymbol{X}_w + p_x^{k\to w}}{\boldsymbol{P}_3^{k\to w^T}\boldsymbol{X}_w + p_z^{k\to w}}, \tag{2.25}$$

$$y_{p_k} = \frac{\boldsymbol{P}_2^{k\to w^T}\boldsymbol{X}_w + p_y^{k\to w}}{\boldsymbol{P}_3^{k\to w^T}\boldsymbol{X}_w + p_z^{k\to w}}. \tag{2.26}$$

Definiert man

$$\begin{array}{ll} \boldsymbol{A}_x^k =_{\mathrm{def}} x_{p_k}\boldsymbol{P}_3^{k\to w} - \boldsymbol{P}_1^{k\to w} & b_x^k =_{\mathrm{def}} p_x^{k\to w} - x_{p_k}p_z^{k\to w} \\ \boldsymbol{A}_y^k =_{\mathrm{def}} x_{p_k}\boldsymbol{P}_3^{k\to w} - \boldsymbol{P}_2^{k\to w} & b_y^k =_{\mathrm{def}} p_y^{k\to w} - y_{p_k}p_z^{k\to w}, \end{array}$$

so sieht man, daß (2.25) und (2.26) äquivalent zu den Ebenengleichungen

$$\boldsymbol{A}_x^{k^T}\boldsymbol{X}_w = b_x^k \tag{2.27}$$

$$\boldsymbol{A}_y^{k^T}\boldsymbol{X}_w = b_y^k \tag{2.28}$$

sind, die durch $T^{k\to w}$ und x_{p_k} bzw. y_{p_k} festgelegt werden. Per Definition gehen beide Ebenen durch den Szenepunkt. Beide Ebenen gehen ebenfalls durch den Zentralprojektionspunkt C_w, für den gilt $T^{k\to w}\begin{pmatrix}\boldsymbol{C}_w\\1\end{pmatrix} = \boldsymbol{0}$. Denn für $\boldsymbol{C}_w$ ist

$$
\begin{aligned}
\boldsymbol{A}_x^{k\,T}\boldsymbol{C}_w &= x_{p_k}\boldsymbol{P}_3^{k\to w\,T}\boldsymbol{C}_w - \boldsymbol{P}_1^{k\to w\,T}\boldsymbol{C}_w = -x_{p_k}p_z^{k\to w} + p_x^{k\to w}\\
&= b_x^k\\
\boldsymbol{A}_y^{k\,T}\boldsymbol{C}_w &= y_{p_k}\boldsymbol{P}_3^{k\to w\,T}\boldsymbol{C}_w - \boldsymbol{P}_2^{k\to w\,T}\boldsymbol{C}_w = -y_{p_k}p_z^{k\to w} + p_y^{k\to w}\\
&= b_y^k.
\end{aligned}
$$

Die Ebenen (2.27) und (2.28) sind aber nicht gleich, da $\det(P^{k\to w}) = \det(S)\det(R^{k\to w}) \neq 0$ ist, also $\boldsymbol{A}_x^k$ und $\boldsymbol{A}_y^k$ nicht linear abhängig sind. Somit bestimmt die Schnittgerade von (2.27) und (2.28) den Sehstrahl bezüglich der betreffenden Kamera. Rechnerisch wurde nachgewiesen, was anschaulich klar ist, nämlich daß durch die Projektionsgleichungen (2.25) und (2.26) der Sehstrahl als Gerade im Weltkoordinatensystem bestimmt wird.

Wird jetzt die Notation auf die binokulare Situation mit einer linken und einer rechten Kamera übertragen, ergibt sich der Szenepunkt $\boldsymbol{X}_w$ als Schnittpunkt des linken und rechten Sehstrahls rechnerisch durch Lösung des überbestimmten Gleichungssystems

$$
\begin{pmatrix}\boldsymbol{A}_x^{l\,T}\\ \boldsymbol{A}_y^{l\,T}\\ \boldsymbol{A}_x^{r\,T}\\ \boldsymbol{A}_y^{r\,T}\end{pmatrix}\boldsymbol{X}_w = \begin{pmatrix}b_x^l\\ b_y^l\\ b_x^r\\ b_y^r\end{pmatrix}.
$$

Hier endet das klassische Vorgehen zur Szenenrekonstruktion. Es behandelt den allgemeinen Fall der Kamerageometrie, setzt aber die Kenntnis aller Kameraparameter in Form von Transformationsmatrizen voraus. Die folgenden Abhandlungen berücksichtigen, daß man sich oft in folgenden Situationen befindet:

- Die Kamerageometrie unterliegt Einschränkungen.
- Die Beziehung zwischen den Kamerakoordinatensystemen ist unbekannt.
- Man möchte auf möglichst viel Kalibrierungsinformation verzichten, da sie fehlerbehaftet ist.
- Zur Lösung der gestellten Aufgabe genügen qualitative Informationen über die Szene, insbesondere über die Tiefe von Szenepunkten. Oder die exakte Rekonstruktion der Szene ist nicht praktikabel, z. B. zu zeitaufwendig.
- Der Beobachter ist aktiv und kann die Gewinnung von Information erleichtern.

2.6 Disparität

Bevor wir dazu kommen, Lösungen für die angesprochenen Probleme vorzustellen, sei noch ein wichtiger Begriff eingeführt, der schon in §2.1 Verwendung gefunden hat, jetzt aber an Bedeutung gewinnen wird. Anstelle von korrespondierenden Bildpunkten läßt sich auch ein Bildpunkt und seine Verschiebung bis zur Position seines korrespondierenden Bildpunktes im anderen Bild betrachten. Dieser zweidimensionale Verschiebungsvektor wird *Disparität* $\boldsymbol{d}$ genannt. O. B. d. A. beschreiben wir hier die Verschiebung des linken Bildpunktes, also

$$\boldsymbol{d} =_{\mathrm{def}} (x_r - x_l, y_r - y_l).$$

Entsprechend der üblichen Assoziation heißt d_x *horizontale Disparität* und d_y *vertikale Disparität.*

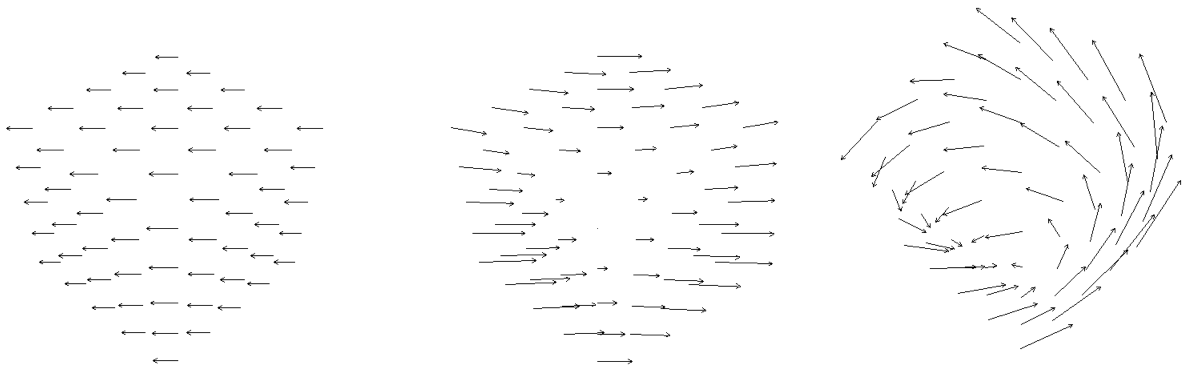

Abbildung 2.10: Disparitätsfelder. Links: Disparitätsfeld einer parallelen Stereogeometrie, entsprechend einer horizontalen Translationsbewegung von links nach rechts. Mitte: Disparitätsfeld einer einfachen Stereogeometrie. Links und rechts wird derselbe Szenepunkt ins Bildzentrum abgebildet, so daß dort die Disparität verschwindet. Um diese Fixierung zu realisieren, müssen nicht alle Freiheitsgrade der Bewegung des Systems genutzt werden. Rechts: Disparitätsfeld einer allgemeinen Stereogeometrie mit Translations- und Rotationsanteilen.

Offensichtlich sind die Informationen von $(\boldsymbol{x}_l, \boldsymbol{x}_r)$ und $(\boldsymbol{x}_l, \boldsymbol{d})$ gleichwertig. Letztere Darstellung legt jedoch die Sichtweise einer stattgefundenen Bewegung nahe, auf die wir noch zurückkommen werden. Abbildung 2.10 zeigt drei Disparitätsfelder, wie sie aus drei verschiedenen Bewegungen bzw. Stereogeometrien enstehen können.

Von besonderem Interesse wird sein, welche Informationen man aus einem einzigen korrespondierenden Paar von Bildpunkten bzw. einem einzigen Disparitätsvektor erhalten kann, wenn man auf möglichst viel Wissen, insbesondere auf Kalibrierungsinformation und die Beziehung zwischen den Kamerakoordinatensystemen, verzichtet, aber bestimmte Systemgeometrien voraussetzt.

Kapitel 3

Binokulare fixierte Stereosysteme

Die Abhandlungen des letzten Kapitels haben einen sehr allgemeinen Rahmen für die Rekonstruktion in binokularen Stereosystemen gegeben. In diesem Kapitel geht es um binokulare Systeme, deren Geometrie einer Einschränkung unterliegt, nämlich daß beide Kameras einen Szenepunkt *fixieren*. Das heißt, die optischen Achsen der Kameras schneiden sich in einem Szenepunkt, dem *Fixierungspunkt*. Systeme, die in der Lage sind zu fixieren, haben ebenfalls in der letzten Zeit viel Beachtung gefunden. Probleme und Beschreibungen der Struktur- und Bewegungsanalyse konnten unter dieser Voraussetzung vereinfacht werden. Solche Systeme werden unter dem Paradigma des *aktiven Sehens* geführt. Dabei kann ein *aktiver Beobachter* die geometrischen Parameter seines sensorischen Systems steuern. Das versetzt ihn in die Lage, sinnvolle geometrische Einschränkungen herbeizuführen, die in unserem Fall die Systembeschreibung und Gewinnung von 3D-Information erleichtern, ohne auf Bildinformation zu verzichten. Vor allem aber verschwindet die Disparität im Bildzentrum und ist in der Umgebung entsprechend gering, so daß phasenbasierte Zuordnungsverfahren eingesetzt werden können.

Abbildung 3.1 zeigt eine solche Geometrie. Das System sei so beschaffen, daß in der Ruhelage beide X-Achsen und die Stereobasis dieselbe Gerade beschreiben. Die optischen Achsen werden mit den Z-Achsen identifiziert, und der Schnittpunkt der Rotationsachsen der Kameras falle mit dem Zentralprojektionspunkt zusammen. Die Fixierung kann dann so realisiert werden: Die linke Kamera rotiere den Winkel ϕ_l um die Y_l-Achse, die rechte Kamera den Winkel ϕ_r um die Y_r-Achse. Diese Winkel heißen *Vergenzwinkel*. Dazu rotieren beide Kameras denselben Winkel um die X-Achsen, so daß die Z-Achsen in derselben Ebene liegen. Dieser Winkel heißt *Neigungswinkel*. Das System fixiere aber ohne *Torsion*, d. h. ohne Rotation der Kameras um ihre Z-Achsen, so daß die X-Achsen in derselben Ebene liegen. Da dies keine Auswirkungen auf die Bildinformation hat, kann auf diesen zusätzlichen Freiheitsgrad bei Stereosystemen in den meisten Fällen verzichtet werden, was ihre Beschreibung erleichtert.

Es liegt nahe, daß hier einfachere Formen und Verfahren zur Szenenrekonstruktion gewonnen werden können, da a-priori-Wissen vorliegt. Dies wird im folgenden ausführlich untersucht. Wir gehen dabei von der Disparität als Träger der Strukturinformation aus: **Jeder einzelne Disparitätsvektor trägt Strukturinformation, die im Fixierungsfall eine besonders einfache Darstellung erfährt.** Die Darstellung, die diese Information geeignet widerspiegelt, wird im ersten Abschnitt dieses Kapitels gesucht. Der zweite Abschnitt führt ein Referenzkoordinatensystem ein, das die Berechnungen ver-

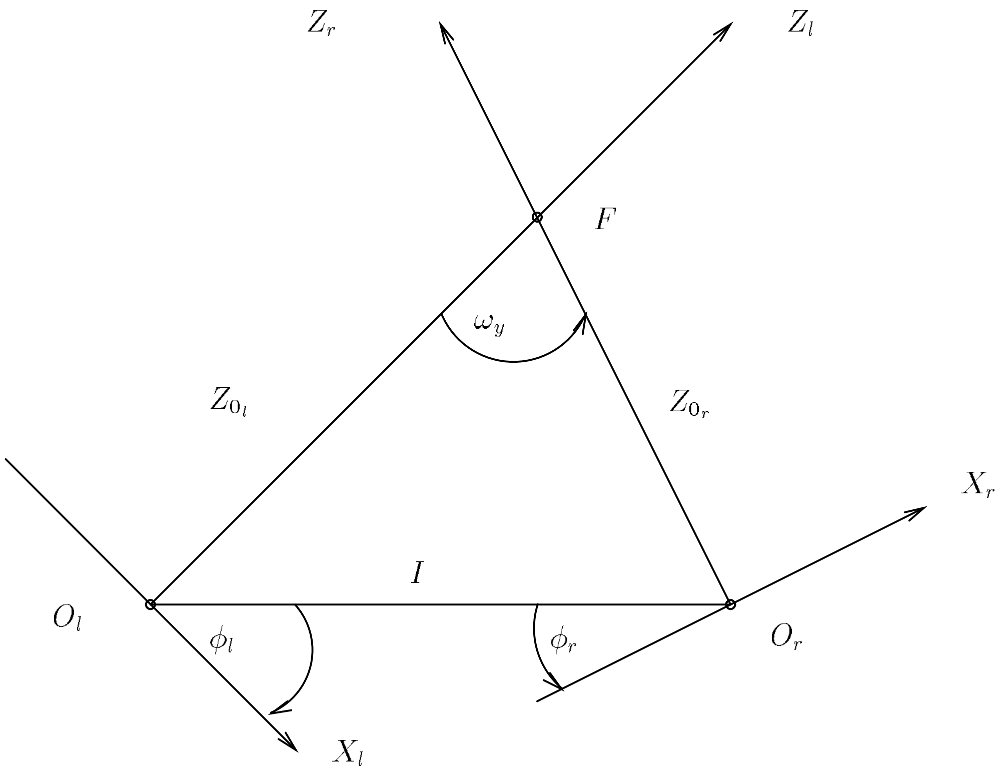

Abbildung 3.1: Binokulares fixiertes Stereosystem. Die optischen Achsen der Kameras schneiden sich im Fixierungspunkt F. Die Tiefe des Fixierungspunktes ist Z_{0_l} im linken und Z_{0_r} im rechten Koordinatensystem. I bezeichnet die Stereobasisbreite, ϕ_l und ϕ_r die Vergenzwinkel der linken bzw. rechten Kamera. Die Transformation zwischen linkem und rechtem Koordinatensystem besteht aus einer Translation entlang der Stereobasis und einer Rotation um die Y-Achse mit dem Winkel $\omega_y = \phi_r - \phi_l$.

einfacht. Der dritte Abschnitt diskutiert mit Hilfe von experimentellen Vergleichen die unterschiedlichen Disparitätsdarstellungen. Gegenstand des vierten Abschnittes ist eine eingehendere Betrachtung der Fixierungsbedingung. Der fünfte und letzte Abschnitt zieht die Verbindung zur SfM-Problematik.

3.1 Disparität vs. Verschiebungsrate

Zur Darstellung der Beziehung der Punktkorrespondenzen zur Tiefe und zu den Kameraparametern wird im folgenden also die Disparität genutzt. Da in ihr über die Bildpunktkoordinaten die beiden Kamerakoordinatensysteme vorkommen, sei zunächst ihre Beziehung beschrieben, wie sie sich im Fixierungsfall darstellt.

Das rechte Koordinatensystem geht aus dem linken durch eine Rotation $R^{l\to r}$ um die Y_l-Achse und eine Translation $\boldsymbol{t}^{l\to r}$ in die X_l und Z_l-Richtung hervor. Seien $\boldsymbol{X}_l = (X_l, Y_l, Z_l)^T$ und $\boldsymbol{X}_r = (X_r, Y_r, Z_r)^T$ die Koordinaten desselben Szenepunktes im linken und rechten Koordinatensystem. Dann gilt

$$\boldsymbol{X}_l = R^{l\to r}\boldsymbol{X}_r + \boldsymbol{t}^{l\to r}. \tag{3.1}$$

Da aber die Stereobasisbreite im Verhältnis zur Tiefe des fixierten Punktes klein ist (*An-*

nahme einer kleinen Stereobasis), sind die Vergenzwinkel klein, und die Rotationsmatrix kann durch die antisymmetrische Matrix der Winkelgeschwindigkeiten approximiert werden:

$$R^{l\to r} = \begin{pmatrix} \cos(\mu_y) & 0 & \sin(\mu_y) \\ 0 & 1 & 0 \\ -\sin(\mu_y) & 0 & \cos(\mu_y) \end{pmatrix} \qquad R^{l\to r} \approx \begin{pmatrix} 1 & 0 & \omega_y \\ 0 & 1 & 0 \\ -\omega_y & 0 & 1 \end{pmatrix}$$

Hier sind μ_y der diskrete Winkel und ω_y die Winkelgeschwindigkeit bei Rotation um die Y-Achse. Seien v_x^l und v_z^l die X- und Z-Komponenten des Verschiebungsvektors im linken Koordinatensystem. Dann erhält man aus (3.1) in guter Näherung

$$\boldsymbol{X}_l = \begin{pmatrix} 1 & 0 & \omega_y \\ 0 & 1 & 0 \\ -\omega_y & 0 & 1 \end{pmatrix} \boldsymbol{X}_r + \begin{pmatrix} v_x^l \\ 0 \\ v_z^l \end{pmatrix}. \tag{3.2}$$

Die Disparität steht dann mit den Bildpunktkoordinaten, der Tiefe des Szenepunktes und den Parametern der Kamerakoordinatentransformation in Beziehung. Arbeitet man mit normalisierten Bildkoordinaten, so ergibt sich:

$$\begin{aligned} d_x &= x_r - x_l \overset{(2.11)}{=} \frac{X_r}{Z_r} - \frac{X_l}{Z_l} \overset{(3.2)}{=} \frac{X_r Z_l - (X_r + \omega_y Z_r + v_x^l) Z_r}{Z_l Z_r} \\ &\overset{(2.12)}{=} \frac{x_r Z_l - x_r Z_r - \omega_y Z_r - v_x^l}{Z_l} \end{aligned} \tag{3.3}$$

$$\begin{aligned} d_y &= y_r - y_l \overset{(2.11)}{=} \frac{Y_r}{Z_r} - \frac{Y_l}{Z_l} \overset{(3.2)}{=} \frac{Y_r Z_l - Y_r Z_r}{Z_l Z_r} \\ &\overset{(2.12)}{=} \frac{y_r Z_l - y_r Z_r}{Z_l} \end{aligned} \tag{3.4}$$

In diesen Gleichungen sind noch die Tiefen bezüglich beider Koordinatensysteme enthalten. Ersetzung von Z_l gemäß (3.2) liefert

$$\begin{aligned} d_x &= \frac{x_r(-\omega_y X_r + Z_r + v_z^l) - x_r Z_r - \omega_y Z_r - v_x^l}{-\omega_y X_r + Z_r + v_z^l} \\ &= \frac{-v_x^l + v_z^l x_r - \omega_y Z_r (x_r^2 + 1)}{Z_r(-\omega_y x_r + 1) + v_z^l}, \end{aligned} \tag{3.5}$$

$$\begin{aligned} d_y &= \frac{y_r(-\omega_y X_r + Z_r + v_z^l) - y_r Z_r}{-\omega_y X_r + Z_r + v_z^l} \\ &= \frac{v_z^l y_r - \omega_y Z_r x_r y_r}{Z_r(-\omega_y x_r + 1) + v_z^l}. \end{aligned} \tag{3.6}$$

Diese exakten Gleichungen für die Disparität sind leider relativ komplex. Vor allem die Komplexität des Nenners, in dem die Tiefe in Abhängigkeit der x-Koordinate des Bildpunktes skaliert und eine Konstante addiert wird, erschwert die Gewinnung der Parameter aus der Disparität. Wir nehmen daher einen anderen Standpunkt ein und erhalten so eine einfachere Beschreibung.

Die Approximation der diskreten Rotationsmatrix durch die Matrix der Winkelgeschwindigkeiten, legt folgende Interpretation der binokularen Geometrie nahe: Das rechte Bild ist nach einer Bewegung der linken Kamera in die Position der rechten Kamera entstanden. Die Annahme einer kleinen Stereobasis besagt, daß die Bewegung gering ist und durch Geschwindigkeiten approximiert werden kann, was im Fall der Rotation bereits geschehen ist. Konsequenterweise kann man dann dazu übergehen, die bekannten Bewegungsgleichungen für Eigenbewegung (Kamerabewegung) auf das Stereoproblem anzuwenden (siehe z. B. [Hor86]).

Ein Beobachter bewege sich mit Translationsgeschwindigkeit $\boldsymbol{v} = (v_x, v_y, v_z)^T$ und Rotationsgeschwindigkeit $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^T$. Dann ist die Geschwindigkeit im Punkt $\boldsymbol{X} = (X, Y, Z)^T$

$$\dot{\boldsymbol{X}} = -\boldsymbol{v} - \boldsymbol{\omega} \times \boldsymbol{X}. \tag{3.7}$$

Die als *Verschiebungsrate* bezeichnete Geschwindigkeit eines Bildpunktes in normalisierten Bildkoordinaten ist dann in Komponentenschreibweise

$$\begin{aligned} \dot{x} &= (\frac{\dot{X}}{Z}) = \frac{\dot{X}Z - X\dot{Z}}{Z^2} \\ &= \frac{-v_x + v_z x}{Z} + \omega_x xy - \omega_y(x^2+1) + \omega_z y, && (3.8) \\ \dot{y} &= (\frac{\dot{Y}}{Z}) = \frac{\dot{Y}Z - Y\dot{Z}}{Z^2} \\ &= \frac{-v_y + v_z y}{Z} + \omega_x(y^2+1) - \omega_y xy - \omega_z x. && (3.9) \end{aligned}$$

Wird die Fixierung wie beschrieben realisiert, so verschwinden die v_y-, ω_x- und ω_z-Terme, und man erhält

$$\begin{aligned} \dot{x} &= \frac{-v_x + v_z x}{Z} - \omega_y(x^2+1), && (3.10) \\ \dot{y} &= \frac{v_z y}{Z} - \omega_y xy. && (3.11) \end{aligned}$$

Hierbei sind also v_x, v_z und ω_y die Parameter der Stereogeometrie, x und y die normalisierten Bildpunktkoordinaten, also der Sehstrahl, und Z eine Tiefenfunktion in Abhängigkeit von x und y. Diese Formeln bieten eine wesentlich geeignetere Grundlage zur Analyse der Disparitäten, da sie den Einfluß der Parameter einfacher und direkter wiedergeben. Zusätzlich hat man die *Fixierungsbedingung*, die besagt, daß im Bildzentrum $(0,0)$ die Verschiebung 0 ist, also für die x-Komponente (3.10)

$$\dot{x}|_{(x,y)=(0,0)} = -\frac{v_x}{Z_0} - \omega_y = 0,$$

wobei Z_0 die Tiefe des Fixierungspunktes ist, dessen Bildpunkt im Bildzentrum liegt. Hierüber erhält man eine Beziehung zwischen Rotation und Translation:

$$\omega_y = -\frac{v_x}{Z_0}$$

Es bleibt die Frage, wie gut die Approximation der Disparität durch die Verschiebungsrate ist. Die exakten Disparitätsformeln ergeben sich im Vergleich aus (3.5) und (3.6) zu

$$\begin{aligned} d_x &= \frac{\frac{-v_x^l + v_z^l x_r}{Z_r} - \omega_y(x_r^2+1)}{1 - \omega_y x_r + \frac{v_z^l}{Z_r}}, && (3.12) \\ d_y &= \frac{\frac{v_z^l y_r}{Z_r} - \omega_y x_r y_r}{1 - \omega_y x_r + \frac{v_z^l}{Z_r}}. && (3.13) \end{aligned}$$

Wir definieren den Fehler zwischen Disparität und Verschiebungsrate als

$$\begin{aligned}\Delta\dot{x} &=_{\text{def}} \dot{x} - d_x,\\ \Delta\dot{y} &=_{\text{def}} \dot{y} - d_y.\end{aligned}$$

Beschreibt man die Verschiebungsraten wie die Disparitäten mit v_x^l, v_z^l, x_r, y_r, Z_r durch

$$\dot{x}_r =_{\text{def}} \frac{-v_x^l + v_z^l x_r}{Z_r} - \omega_y(x_r^2 + 1), \tag{3.14}$$

$$\dot{y}_r =_{\text{def}} \frac{v_z^l y_r}{Z_r} - \omega_y x_r y_r, \tag{3.15}$$

so gilt angesichts von (3.12, 3.13)

$$\Delta\dot{x}_r = \left(\frac{v_z^l}{Z_r} - \omega_y x_r\right) d_x, \tag{3.16}$$

$$\Delta\dot{y}_r = \left(\frac{v_z^l}{Z_r} - \omega_y x_r\right) d_y. \tag{3.17}$$

Mit Hilfe von

$$d_y = 0 \quad \Leftrightarrow \quad \frac{v_z^l}{Z_r} = \omega_y x_r \ \vee \ y_r = 0,$$

was man direkt aus (3.13) erhält, bekommt man

$$\begin{aligned}\Delta\dot{x}_r = 0 \quad &\Leftrightarrow \quad \frac{v_z^l}{Z_r} = \omega_y x_r \ \vee \ d_x = 0 \quad \Leftrightarrow \quad X_r = \frac{v_z^l}{\omega_y} \ \vee \ d_x = 0,\\ \Delta\dot{y}_r = 0 \quad &\Leftrightarrow \quad \frac{v_z^l}{Z_r} = \omega_y x_r \ \vee \ d_y = 0 \quad \Leftrightarrow \quad X_r = \frac{v_z^l}{\omega_y} \quad \Leftrightarrow \quad d_y = 0.\end{aligned}$$

Man sieht: Die Verschiebungsrate gibt verschwindende Disparität korrekt wieder. Die Szenepunkte, für die die Disparität verschwindet, liegen auf dem *Vieth-Müller-Zylinder*. Dieser hat senkrecht zu der XZ-Ebene verlaufende Mantellinien und einen Basiskreis, der durch die optischen Zentren und den Fixierungspunkt geht (siehe §3.2). Falls die Disparität jedoch nicht Null ist, die Szenepunkte also entsprechend abweichende Tiefen haben, so ist der relative Fehler $\frac{v_z^l}{Z_r} - \omega_y x_r$ nicht unproblematisch für eine quantitative Beschreibung.

Hieran wird ein grundsätzliches Problem deutlich. Da bei den Verschiebungsraten differentiell bezüglich der Zeit vorgegangen wird, werden zu ihrer Berechnung für ein Bild in einer Bildfolge, Bilder, die zeitlich vorangehen und folgen, herangezogen, um so über den zeitlichen Verlauf zu mitteln. Dieses Vorgehen ist beim Stereoproblem mit zwei Kameras nicht möglich. Es werden daher im folgenden zwei andere Möglichkeiten beschrieben, eine einfache, aber möglichst exakte Darstellung der Disparität zu bekommen.

Wir benutzen zunächst eine Kombination von linken und rechten Bildpunktkoordinaten zur Beschreibung der Verschiebungsrate:

$$\dot{x}_{lr} =_{\text{def}} \frac{-v_x^l + v_z^l x_r}{Z_l} - \omega_y(x_r^2 + 1) \tag{3.18}$$

$$\dot{y}_{lr} =_{\text{def}} \frac{v_z^l y_r}{Z_l} - \omega_y x_r y_l \tag{3.19}$$

Um hier den Fehler bezüglich der exakten Disparität zu ermitteln, schreiben wir diese noch einmal um durch teilweise Ersetzung von Z_l:

$$\begin{aligned}
d_x &\overset{(3.3)}{=} \frac{x_r Z_l - x_r Z_r - \omega_y Z_r - v_x^l}{Z_l} \\
&\overset{(3.2)}{=} \frac{x_r(-\omega_y x_r Z_r + Z_r + v_z^l) - x_r Z_r - \omega_y Z_r - v_x^l}{Z_l} \\
&= \frac{-v_x^l + v_z^l x_r}{Z_l} - \frac{Z_r}{Z_l}\omega_y(x_r^2 + 1) \\
d_y &\overset{(3.4)}{=} \frac{y_r Z_l - y_r Z_r}{Z_l} \\
&\overset{(3.2)}{=} \frac{y_r(-\omega_y x_r Z_r + Z_r + v_z^l) - y_r Z_r}{Z_l} \\
&= \frac{v_z^l y_r}{Z_l} - \frac{Z_r}{Z_l}\omega_y x_r y_r
\end{aligned}$$

Man bekommt daraus sofort:

$$\Delta \dot{x}_{lr} = -\omega_y(x_r^2 + 1)(1 - \frac{Z_r}{Z_l}) \tag{3.20}$$

$$\Delta \dot{y}_{lr} = -\omega_y x_r(y_l - \frac{Z_r}{Z_l} y_r) \overset{(2.12)}{=} 0 \tag{3.21}$$

Also:

$$\begin{aligned}
&\Delta \dot{x}_{lr} = 0 \quad \Leftrightarrow \quad Z_l = Z_r \quad \Leftrightarrow \quad X_r = \frac{v_z^l}{\omega_y} \\
&\Delta \dot{y}_{lr} = 0
\end{aligned}$$

Die vertikale Disparität wird also bei beschriebener Fixierung, bis auf die Approximation der Rotation, exakt durch (3.19) dargestellt. Betrachtet man nur den Translationsanteil, so ist auch $\Delta \dot{x}_{lr} = 0$. Im allgemeinen ist die horizontale Disparität gleich der Darstellung (3.18), falls Szenepunkte auf der Ebene betrachtet werden, die senkrecht zu der XY-Ebene liegt und durch den Schnittpunkt der X-Achsen geht. Dadurch hat sie allerdings einen gravierenden Nachteil: Die Fixierungsbedingung wird nicht erfüllt, in dem Sinne, daß für den Fixierungspunkt Verschiebungsrate und Disparität nicht übereinstimmen. Somit ist für Punkte in der Nähe des Fixierungspunktes und für den Fixierungspunkt selbst, die normalerweise von besonderem Interesse sind, die Darstellung fehlerbehaftet. Dies äußert sich auch in $\Delta \dot{x}$, wo der Fehler nicht wie in (3.16) durch die exakte Disparität skaliert wird. Deswegen weicht hier die horizontale Verschiebungsrate im allgemeinen zu sehr von der Disparität ab. Der folgende Ansatz stellt die Fixierungsbedingung auf andere Weise sicher.

3.2 Einführung eines zyklopischen Auges

Wir haben gesehen, daß die Beschreibung der Disparität durch Verschiebungsraten darunter leidet, daß bei der Bewegungsinterpretation zeitlich nicht gemittelt werden kann. Um dies doch zu ermöglichen, wird ein Bezugskoordinatensystem, ein sogenanntes *zyklopisches Koordinatensystem*, eingeführt, das sich zwischen linkem und rechtem Koordinatensystem befindet und sozusagen zeitlich in der Mitte der Bewegung liegt. Man erwartet in diesem Koordinatensystem eine symmetrische Darstellung, die die bisher aufgetretenen Fehler sehr klein hält.

Ein zyklopisches Koordinatensystem wurde bereits in §2 beschrieben. Der Zentralprojektionspunkt lag dort in der Mitte der Stereobasis. Hier werden wir mit einem Koordinatensystem arbeiten, dessen Zentralprojektionspunkt der Schnittpunkt des Kreises, der durch die Zentralprojektionspunkte der linken und rechten Kamera und den Fixierungspunkt geht, mit der Winkelhalbierenden des Winkels ist, der von den optischen Achsen der Kameras eingeschlossen wird. Abbildung 3.2 beschreibt diese Geometrie. Vorteile der Lage des Koordinatensystems liegen in der Symmetrie bezüglich der Kameras und der einfachen geometrischen Beschreibung. Allerdings variiert die Lage des zyklopischen Auges leicht mit der Entfernung des Fixierungspunktes. Für große Distanzen sind die Unterschiede zwischen den Modellen aber zu vernachlässigen.

Zunächst wird das System aus Abbildung 3.2 geometrisch beschrieben, bevor die Verschiebungsraten, wie sie sich hier darstellen, analysiert werden. Diese Beschreibungen gehen auf [GPMF95] zurück.

Die folgenden Bezeichnungen ergeben sich aus Abbildung 3.2. Der Index c kennzeichne dabei das zyklopische Auge. Zwei Winkel sind für die Geometrie wichtig: Definiere den *Vergenzwinkel* μ und den *Versionswinkel* γ durch

$$\begin{aligned} \mu &=_{\text{def}} \frac{1}{2}(\phi_r - \phi_l), \\ \gamma &=_{\text{def}} \frac{1}{2}(\phi_r + \phi_l). \end{aligned}$$

Man beachte, daß alle Winkel als Rotationswinkel um die Y-Achsen zu verstehen sind und ein entsprechendes Vorzeichen erhalten. Durch Betrachtung der Winkel bei F sieht man, daß die Z_l- und die Z_r-Achse den Winkel $\phi_r - \phi_l = 2\mu$ einschließen. Ebenso ergibt sich $\angle MO_cS = \phi_l + \mu = \gamma$. Da die Dreiecke O_cSO_l und FSO_r ähnlich sind, ebenso die Dreiecke O_rSO_c und O_lSF, ergibt sich $\angle SO_lO_c = \mu$ und $\angle O_cO_rS = \mu$. Daraus folgt $|\overline{O_lO_c}| = |\overline{O_rO_c}| = h$. Der Winkel, den die X_l-Achse und die Gerade durch O_l und O_c einschließen, ist $\mu + \phi_l = \gamma$; der Winkel, den die X_r-Achse und die Gerade durch O_r und O_c einschließen, ist $\phi_r - \mu = \gamma$. Damit ist die Korrektheit der Abbildung gezeigt. Auch der abgebildete Kreis durch O_l, O_r und F, und eben auch durch O_c, der sogenannte *Vieth-Müller-Kreis*, hat eine besondere Eigenschaft: Punkte, die auf ihm liegen, haben Disparität 0. Dies kann man sich klarmachen, indem man von der Ähnlichkeit der Dreiecke, die durch die Sehstrahlen des betrachteten Punktes und des Fixierungspunktes definiert werden, auf die Gleichheit der Winkel zwischen den Z-Achsen und den Sehstrahlen schließt. Da wir von Bildebenen ausgehen, die parallel zu den XY-Ebenen liegen, gilt das Argument auch für Punkte auf den Geraden, die parallel zu den Bildebenen durch den Vieth-Müller-Kreis gehen. Man erhält so sämtliche Szenepunkte mit Disparität Null und den bereits erwähnten *Vieth-Müller-Zylinder*.

Ein Ziel, das durch die Einführung eines solchen Koordinatensystems erreicht werden sollte, war die Erfüllung der Fixierungsbedingung. Beschreibt man jetzt die Bewegung von links nach rechts „gemittelt“ im zyklopischen Koordinatensystem, so erhält man aus

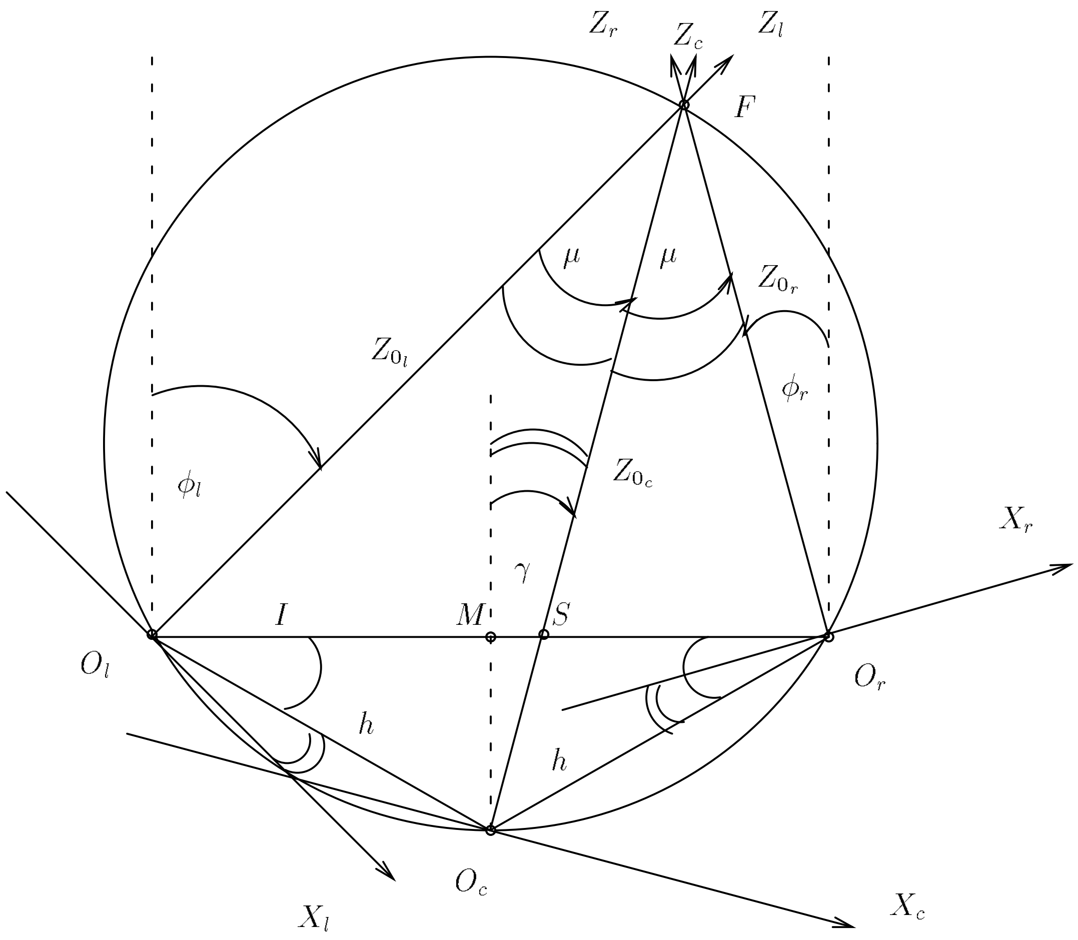

Abbildung 3.2: Fixiertes Stereosystem mit linkem, rechtem und zyklopischem Koordinatensystem. Fixierung des Punktes F wird realisiert durch Rotation der Kameras um die Vergenzwinkel ϕ_l und ϕ_r. Dann ist die Stereogeometrie bestimmt durch die Vergenz μ, die Version γ und die Stereobasisbreite I, i. e. die Länge der Strecke $\overline{O_l O_r}$. Die Tiefe des Fixierungspunktes ist in den verschiedenen Koordinatensystemen bezeichnet durch Z_{0_l}, Z_{0_r} und Z_{0_c}. Die Z-Achse des zyklopischen Koordinatensystems ist gegeben durch die Winkelhalbierende des Winkels am Fixierungspunkt. Der Schnitt dieser Gerade mit dem Kreis durch den Fixierungspunkt und den linken und rechten Zentralprojektionspunkt definiert den Zentralprojektionspunkt der zyklopischen Kamera. Er hat zu den anderen Zentralprojektionspunkten denselben Abstand h.

(nach [GPMF95])

der Betrachtung der Winkel bei F

$$\begin{aligned}
\boldsymbol{v}^c &= \begin{pmatrix} v_x^c \\ 0 \\ v_z^c \end{pmatrix} = I \begin{pmatrix} \cos(\gamma) \\ 0 \\ \sin(\gamma) \end{pmatrix}, && (3.22) \\
\omega_y &= \sin(2\mu). && (3.23)
\end{aligned}$$

Die Fixierungsbedingung lautet dann

$$\omega_y = -\frac{v_x^c}{Z_{0_c}}. \qquad (3.24)$$

Wir wollen versuchen, diese Gleichung rein geometrisch herzuleiten: Nach dem Satz des Thales schneidet die Senkrechte zur Geraden Z_c durch F den Vieth-Müller-Kreis auf dem Durchmesser durch O_c. Dann gilt

$$\cos(\gamma) = \frac{Z_{0_c}}{2r_u},$$

wobei r_u der Radius des Umkreises ist. Man erhält ihn aus den Größen des Dreiecks über

$$r_u = \frac{I}{-2\sin(2\mu)}.$$

Insgesamt ergibt sich

$$\begin{aligned}
\cos(\gamma) &= \frac{-2Z_{0_c}\sin(2\mu)}{2I} = -\frac{Z_{0_c}\sin(2\mu)}{I}, \text{ also} \\
\sin(2\mu) &= -\frac{I\cos(\gamma)}{Z_{0_c}}, && (3.25)
\end{aligned}$$

womit (3.24) gezeigt ist. Man erhält damit die Beziehung zwischen Translation und Rotation im Fixierungsfall:

$$\begin{aligned}
v_x^c &= -\frac{\omega_y}{Z_{0_c}} && (3.26) \\
v_z^c &= -\tan(\gamma)\frac{\omega_y}{Z_{0_c}} && (3.27)
\end{aligned}$$

Was ergibt sich jetzt für Verschiebungsraten in beliebigen Bildpunkten? $\boldsymbol{x}_c = (x_c, y_c, 1)^T$ bezeichne einen zyklopischen Sehstrahl. Dann sind die Verschiebungsraten im zyklopischen Koordinatensystem

$$\begin{aligned}
\dot{x}_c &=_{\text{def}} \frac{-v_x^c + v_z^c x_c}{Z_c} - \omega_y(x_c^2 + 1), && (3.28) \\
\dot{y}_c &=_{\text{def}} \frac{v_z^c y_c}{Z_c} - \omega_y x_c y_c. && (3.29)
\end{aligned}$$

Um die exakte Disparität in diesen Termen auszudrücken, müssen Koordinatentransformationen durchgeführt werden, die die Beziehung zu den zyklopischen Koordinaten herstellen. Seien mit $\boldsymbol{u}^l$ bzw. $\boldsymbol{u}^r$ die Translationen von O_l bzw. O_r nach O_c im linken bzw. rechten Koordinatensystem bezeichnet. Dann gilt mit den üblichen Approximationen:

$$\begin{aligned}
\boldsymbol{X}_l &= \begin{pmatrix} 1 & 0 & \frac{\omega_y}{2} \\ 0 & 1 & 0 \\ -\frac{\omega_y}{2} & 0 & 1 \end{pmatrix} \boldsymbol{X}_c + \begin{pmatrix} u_x^l \\ 0 \\ u_z^l \end{pmatrix} && (3.30) \\
\boldsymbol{X}_r &= \begin{pmatrix} 1 & 0 & -\frac{\omega_y}{2} \\ 0 & 1 & 0 \\ \frac{\omega_y}{2} & 0 & 1 \end{pmatrix} \boldsymbol{X}_c + \begin{pmatrix} u_x^r \\ 0 \\ u_z^r \end{pmatrix} && (3.31)
\end{aligned}$$

Aus der Geometrie ergibt sich

$$\boldsymbol{u}^l = -\boldsymbol{u}^r. \tag{3.32}$$

Wir definieren jetzt

$$\boldsymbol{v}^c =_{\text{def}} \begin{pmatrix} 2h\cos(\gamma) \\ 0 \\ 2h\sin(\gamma) \end{pmatrix},$$

so daß gilt:

$$\boldsymbol{u}^l = \begin{pmatrix} h\cos(\gamma) \\ 0 \\ h\sin(\gamma) \end{pmatrix} = \frac{1}{2}\boldsymbol{v}^c \tag{3.33}$$

Man beachte, daß u_z^l, u_z^r und v_z^c bei symmetrischer Fixierung verschwinden. Die Disparitäten lassen sich damit herleiten zu

$$\begin{aligned}
d_x &= x_r - x_l = \frac{X_r Z_l - X_l Z_r}{Z_l Z_r} \\
&\overset{(3.30,3.31)}{=} \frac{(X_c - \frac{\omega_y}{2} Z_c + u_x^r)(-\frac{\omega_y}{2} X_c + Z_c + u_z^l) - (X_c + \frac{\omega_y}{2} Z_c + u_x^l)(\frac{\omega_y}{2} X_c + Z_c + u_z^r)}{(-\frac{\omega_y}{2} X_c + Z_c + u_z^l)(\frac{\omega_y}{2} X_c + Z_c + u_z^r)} \\
&\overset{(3.32)}{=} \frac{-2u_x^l Z_c + 2u_z^l X_c - \omega_y (X_c^2 + Z_c^2)}{Z_c^2 - \omega_y^2 \frac{X_c^2}{4} + \omega_y u_z^l X_c - {u_z^l}^2} \\
&\overset{(3.33)}{=} \frac{-v_x^c Z_c + v_z^c X_c - \omega_y (X_c^2 + Z_c^2)}{Z_c^2 - \frac{1}{4}(\omega_y X_c - v_z^c)^2}, && (3.34) \\
d_y &= y_r - y_l = \frac{Y_r Z_l - Y_l Z_r}{Z_l Z_r} \\
&\overset{(3.30,3.31)}{=} \frac{Y_c((-\frac{\omega_y}{2} X_c + Z_c + u_z^l) - (\frac{\omega_y}{2} X_c + Z_c + u_z^r))}{(-\frac{\omega_y}{2} X_c + Z_c + u_z^l)(\frac{\omega_y}{2} X_c + Z_c + u_z^r)} \\
&\overset{(3.32)}{=} \frac{2u_z^l Y_c - \omega_y X_c Y_c}{Z_c^2 - \omega_y^2 \frac{X_c^2}{4} + \omega_y u_z^l X_c - {u_z^l}^2} \\
&\overset{(3.33)}{=} \frac{v_z^c Y_c - \omega_y X_c Y_c}{Z_c^2 - \frac{1}{4}(\omega_y X_c - v_z^c)^2}. && (3.35)
\end{aligned}$$

Wie man leicht mittels der Zentralprojektionsgleichung sieht, sind dies genau die Verschiebungsraten mit einem zusätzlichen Summanden $-\frac{1}{4}(\omega_y X_c - v_z^c)^2$ im Nenner. Man erhält

$$\Delta \dot{x}_c = \frac{\frac{1}{4}(\omega_y X_c - v_z^c)^2}{Z_c^2} d_x, \tag{3.36}$$

$$\Delta\dot{y}_c = \frac{\frac{1}{4}(\omega_y X_c - v_z^c)^2}{Z_c^2} d_y, \tag{3.37}$$

also

$$\begin{aligned}
\Delta\dot{x}_c = 0 \;\Leftrightarrow\; \omega_y X_c = v_z^c \vee d_x = 0 \;&\overset{(3.27)}{\Leftrightarrow}\; X_c = -\frac{\tan(\gamma)}{Z_{0_c}} \vee d_x = 0, \\
\Delta\dot{y}_c = 0 \;\Leftrightarrow\; \omega_y X_c = v_z^c \vee d_y = 0 \;&\overset{(3.27)}{\Leftrightarrow}\; X_c = -\frac{\tan(\gamma)}{Z_{0_c}} \vee d_y = 0.
\end{aligned}$$

Da $\frac{1}{4}(\omega_y X_c - v_z^c)^2$ klein bezüglich Z_c^2 ist, ist der Fehler im allgemeinen sehr klein. Er verschwindet für Disparität 0, also wiederum für Punkte auf dem Vieth-Müller-Zylinder und für Punkte auf einer Ebene parallel zur $Y_c Z_c$-Ebene. Diese geht aber nahe durch das Bildzentrum, da $\tan(\gamma)$ klein bezüglich Z_{0_c} ist. Diese Ebene ist sogar die $Y_c Z_c$-Ebene bei symmetrischer Fixierung.

Wie alle bisher betrachteten Fehler wird auch dieser kleiner mit kleiner werdendem ω_y, also mit kleiner werdender Stereobasisbreite bezüglich der Fixierungsdistanz. Da im Fixierungsfall über (3.26, 3.27) eine Beziehung zwischen Translation und Rotation besteht, kann die Genauigkeit der Darstellung bezüglich ω_y mathematisch auch noch anders ausgedrückt werden: Man betrachte die Disparitäten als Funktionen von ω_y bei festem γ und Z_{0_c} und führe einer Taylor-Entwicklung bezüglich 0 durch. Mit der Bezeichnung $\lambda_{0_c} =_{\text{def}} 1/Z_{0_c}$ gilt dann:

$$\begin{aligned}
d_x &\overset{(3.34,3.26,3.27)}{=} \frac{\omega_y \lambda_{0_c} Z_c - \omega_y \tan(\gamma)\lambda_{0_c} X_c - \omega_y (X_c^2 + Z_c^2)}{Z_c^2 - \frac{1}{4}\omega_y^2 X_c^2 - \frac{1}{2}\omega_y^2 \tan(\gamma)\lambda_{0_c} X_c - \frac{1}{4}\omega_y^2 \tan^2(\gamma)\lambda_{0_c}^2} \\
&\overset{\text{T.-E.}}{=} \omega_y \frac{(\lambda_{0_c} Z_c - \tan(\gamma)\lambda_{0_c} X_c - (X_c^2 + Z_c^2))Z_c^2}{Z_c^4} + \frac{\omega_y^2}{2} 0 + o(\omega_y^3) \\
&= \dot{x}_c + o(\omega_y^3) \\
d_y &\overset{(3.35,3.27)}{=} \frac{-\omega_y \tan(\gamma)\lambda_{0_c} Y_c - \omega_y X_c Y_c}{Z_c^2 - \frac{1}{4}\omega_y^2 X_c^2 - \frac{1}{2}\omega_y^2 \tan(\gamma)\lambda_{0_c} X_c - \frac{1}{4}\omega_y^2 \tan^2(\gamma)\lambda_{0_c}^2} \\
&\overset{\text{T.-E.}}{=} \omega_y \frac{(-\tan(\gamma)\lambda_{0_c} Y_c - X_c Y_c) Z_c^2}{Z_c^4} + \frac{\omega_y^2}{2} 0 + o(\omega_y^3) \\
&= \dot{y}_c + o(\omega_y^3)
\end{aligned}$$

Da nur noch Terme dritter Ordnung vorkommen, ist auch diese Approximation sehr gut [GPMF95].

Es bleibt jedoch ein Problem dieser Darstellung. Existiert keine reale Kamera, die die Rolle des zyklopischen Auges übernehmen kann und für die dann, falls das Korrespondenzproblem gelöst ist, Bildpunkte im zyklopischen Koordinatensystem vorliegen, so müssen die Koordinaten des Bildpunktes (x_c, y_c) geschätzt werden. Der Ansatz, der sich aus der Lage des zyklopischen Auges ergibt, ist die Bildung des arithmetischen Mittels aus den linken und rechten Bildpunktkoordinaten, also

$$(\tilde{x}_c, \tilde{y}_c) =_{\text{def}} \frac{1}{2}(x_r + x_l, y_r + y_l).$$

In §2.1 wurde gezeigt, daß für eine parallele binokulare Geometrie dies exakt die zyklopischen Bildpunktkoordinaten sind. Auch für das Bildzentrum gilt $(\tilde{0}_c, \tilde{0}_c) = (0, 0) = (0_c, 0_c)$, und somit bleibt die Fixierungsbedingung auch für diese angenäherten Koordinaten erhalten. Allgemein ergibt sich:

$$
\begin{aligned}
\tilde{x}_c &= \frac{1}{2}\frac{X_r Z_l + X_l Z_r}{Z_l Z_r} \\
&\overset{(3.30,3.31)}{=} \frac{1}{2}\frac{(X_c - \frac{\omega_y}{2}Z_c + u_x^r)(-\frac{\omega_y}{2}X_c + Z_c + u_z^l) + (X_c + \frac{\omega_y}{2}Z_c + u_x^l)(\frac{\omega_y}{2}X_c + Z_c + u_z^r)}{(-\frac{\omega_y}{2}X_c + Z_c + u_z^l)(\frac{\omega_y}{2}X_c + Z_c + u_z^r)} \\
&\overset{(3.32)}{=} \frac{X_c Z_c + [-\frac{1}{4}\omega_y^2 X_c Z_c - \frac{1}{2}\omega_y u_z^l Z_c - \frac{1}{2}\omega_y u_x^l X_c - u_x^l u_z^l]}{Z_c^2 + [-(\frac{1}{2}\omega_y X_c - u_z^l)^2]}, \\
\tilde{y}_c &= \frac{1}{2}\frac{Y_r Z_l + Y_l Z_r}{Z_l Z_r} \\
&\overset{(3.30,3.31)}{=} \frac{1}{2}\frac{Y_c((-\frac{\omega_y}{2}X_c + Z_c + u_z^l) + (\frac{\omega_y}{2}X_c + Z_c + u_z^r))}{(-\frac{\omega_y}{2}X_c + Z_c + u_z^l)(\frac{\omega_y}{2}X_c + Z_c + u_z^r)} \\
&\overset{(3.32)}{=} \frac{Y_c Z_c}{Z_c^2 + [-(\frac{1}{2}\omega_y X_c - u_z^l)^2]}.
\end{aligned}
$$

Die Terme in eckigen Klammern enthalten den Fehler bezüglich der exakten Koordinaten. Auch hier ist der Fehler skaliert durch ω_y, und eine Taylor-Entwicklung wie bei der Disparitätsfunktion liefert

$$
\begin{aligned}
\tilde{x}_c &\overset{(3.33,3.22)}{=} \frac{X_c Z_c + \frac{1}{4}\omega_y^2 X_c Z_c + \frac{1}{4}\omega_y^2 \tan(\gamma)\lambda_{0_c} Z_c + \frac{1}{4}\omega_y^2 \lambda_{0_c} X_c - \frac{1}{4}\omega_y^2 \tan(\gamma)\lambda_{0_c}^2}{Z_c^2 - \frac{1}{4}\omega_y^2 X_c^2 - \frac{1}{2}\omega_y^2 \tan(\gamma)\lambda_{0_c} X_c - \frac{1}{4}\omega_y^2 \tan^2(\gamma)\lambda_{0_c}^2} \\
&\overset{\text{T.-E.}}{=} \frac{X_c Z_c}{Z_c^2} + \omega_y 0 + o(\omega_y^2) \\
&= x_c + o(\omega_y^2), \\
\tilde{y}_c &\overset{(3.33,3.22)}{=} \frac{Y_c Z_c}{Z_c^2 - \frac{1}{4}\omega_y^2 X_c^2 - \frac{1}{2}\omega_y^2 \tan(\gamma)\lambda_{0_c} X_c - \frac{1}{4}\omega_y^2 \tan^2(\gamma)\lambda_{0_c}^2} \\
&\overset{\text{T.-E.}}{=} \frac{Y_c Z_c}{Z_c^2} + \omega_y 0 + o(\omega_y^2) \\
&= y_c + o(\omega_y^2).
\end{aligned}
$$

Auch hier tauchen immerhin nur noch Terme quadratischer Ordnung auf. Somit ist diese Approximation geeignet. Dies werden auch die folgenden Experimente bestätigen.

3.3 Experimenteller Vergleich der Disparitätsdarstellungen

In Experimenten wurden den Formeln für die Verschiebungsraten die korrekten Werte für ω_y, $\boldsymbol{v}^l$, $\boldsymbol{v}^c$, $\boldsymbol{x}_r$, $\boldsymbol{x}_l$, Z_r, Z_l, Z_c zugrundegelegt und mit den richtigen Disparitätswerten verglichen (Tabelle 3.1, Abbildung 3.3). Wir bezeichnen Darstellung (3.28, 3.29) als „zyklopisch" (c), (3.14, 3.15) als „rechts" (r) und (3.18, 3.19) als „links-rechts" (l-r). Trotz der zusätzlichen Approximation der Bildpunktkoordinaten hat sich die zyklopische Darstellung im allgemeinen als sehr exakt und den anderen Darstellungen rechts und links-rechts weit überlegen erwiesen. Ihre Symmetrie bewirkt minimale Fehler, die auch mit zunehmender Entfernung vom Bildzentrum nicht wesentlich wachsen. Darüber hinaus sieht man, da (3.19) die vertikale Disparität exakt bis auf die Approximation der kleinen Rotationswinkel, vor allem $\cos(2\mu) \approx 1$, bestimmt, daß diese Approximation durchaus große relative Fehler hervorrufen kann, falls diese Winkel nicht sehr klein sind. Durch die symmetrische Position des zyklopischen Auges werden aber auch diese Fehler kompensiert, unabhängig von der Lage des Fixierungspunktes. Im allgemeinen werden die Darstellungen bei kleineren Vergenzwinkeln, also auch kleineren Rotationswinkeln bei der Koordinatentransformation, genauer.

mittlerer relativer Fehler [%]		Verschiebungsrate					
		rechts (3.14, 3.15)		links-rechts (3.18, 3.19)		zyklopisch (3.28, 3.29)	
		x	y	x	y	x	y
$\mu= \quad -5°$	$\gamma= \quad 0°$	3.4	42.7	31.6	37.5	0.7	0.7
	$\gamma= \; -2.5°$	3.4	75.1	32.1	71.4	0.7	0.8
	$\gamma= \quad 2.5°$	3.5	75.5	31.9	71.1	0.7	0.8
$\mu=-2.5°$	$\gamma= \quad 0°$	0.9	38.7	26.6	37.5	0.2	0.2
	$\gamma=-1.25°$	0.9	72.4	26.8	71.5	0.2	0.2
	$\gamma= \; 1.25°$	0.9	72.4	26.7	71.4	0.2	0.2

Tabelle 3.1: Mittlerer relativer Fehler der verschiedenen Darstellungen der Verschiebungsrate bezüglich der richtigen Disparität für 5^3 Punkte $\boldsymbol{X}_c$ in der Umgebung des Fixierungspunktes mit $\boldsymbol{X}_c = (id, jd, Z_{0_c} + kd)^T, i, j, k \in \{-2, -1, 0, 1, 2\}, d = 10$. Grundlage sind exakte Werte für die Parameter der Formeln. Für die binokulare Geometrie wurde eine Stereobasis der Breite 10 gewählt.

Abbildung 3.3 zeigt, daß im Bildzentrum die zyklopische Darstellung geringe Fehler aufweist. Dies liegt daran, daß bei den Experimenten keine Rücksicht auf die Sichtbarkeit der Punkte genommen wurde. Deshalb wurden mehrere Szenepunkte auf denselben Bildpunkt projiziert. Man erkennt, daß nur für den Punkt, an dem sich die optischen Achsen schneiden, die Fixierungsbedingung erfüllt und die Darstellung exakt ist. Sonst kann es passieren, daß das Bildzentrum nicht-verschwindende Disparität aufweist. Werden also andere Punkte auf dem Sehstrahl betrachtet, schneiden sich die optischen Achsen also nicht in einem abgebildeten Punkt, so verschwindet der Fehler nicht, ist aber immer noch gering, da die Fehlerbetrachtungen auf der Systemgeometrie beruhen und immer noch gelten. Dies ist Thema des folgenden Abschnittes. Weiter sieht man, daß der absolute Fehler mit zunehmender Tiefe abnimmt, da auch die Disparität abnimmt, sich an den Relationen aber nichts ändert. Da bei symmetrischer Fixierung die vertikale Disparität sehr gering ist, ist die rechts-Darstellung hier noch viel besser als die links-rechts-Darstellung.

An dieser Stelle seien einmal zur Verdeutlichung der experimentellen Ergebnisse zwei exakte Fehlerformeln, ohne Approximation der Rotationsmatrix, genannt. Es gilt für die exakte Disparität unter Berücksichtigung der Kosinus-Terme:

$$d_y \quad = \quad y_r - y_l = \frac{Y_r}{Z_r} - \frac{Y_l}{Z_l} = \frac{Y}{Z_l Z_r}(Z_l - Z_r)$$

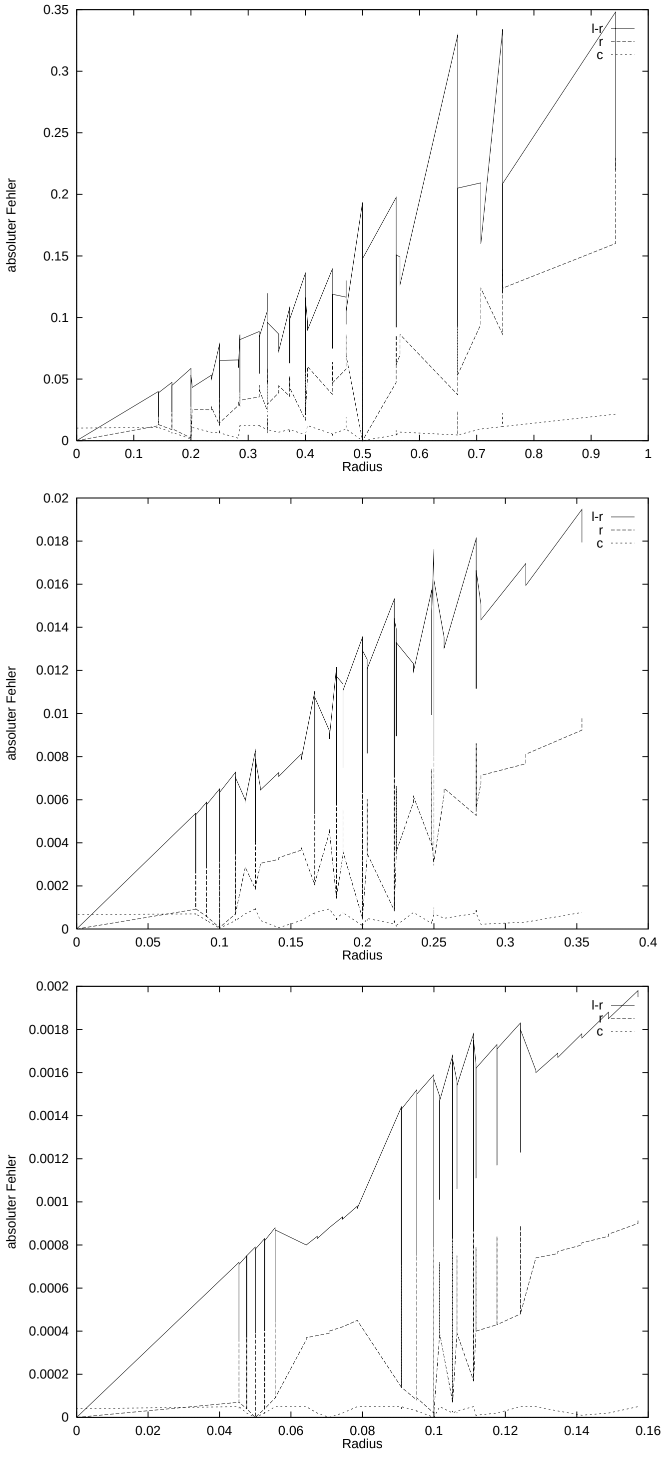

Abbildung 3.3: Mittlerer absoluter Fehler der verschiedenen Darstellungen der Verschiebungsrate bezüglich der richtigen Disparität für 5^3 Punkte $\boldsymbol{X}_c$ in der Umgebung des Fixierungspunktes, $\boldsymbol{X}_c = (id, jd, Z_{0_c} + kd)^T, i, j, k \in \{-2, -1, 0, 1, 2\}, d = 10$. Fehler und Radius wurden bezüglich der Euklidischen Vektornorm bestimmt: Radius $r =_{\text{def}} \|\boldsymbol{x}_c\|_2$, absoluter Fehler $\Delta\dot{\boldsymbol{x}} =_{\text{def}} \|(\Delta\dot{x}, \Delta\dot{y})\|_2 = \|(\dot{x} - d_x, \dot{y} - d_y)\|_2$. Stereobasisbreite ist 25. Z_0 ist von oben nach unten 50, 100 und 200. Das Stereosystem fixiert dabei symmetrisch.

$$\begin{aligned}
&= \frac{Y}{Z_l Z_r}(-\sin(2\mu)X_r + \cos(2\mu)Z_r + v_z^l - Z_r) \\
&= \frac{Y}{Z_l Z_r}(-\sin(2\mu)X_r + v_z^l) + \frac{Y}{Z_l Z_r}(\cos(2\mu) - 1)Z_r \\
&= \frac{v_z^l}{Z_r} - \omega_y x_r y_l + (\cos(2\mu) - 1)y_l \\
&= \dot{y_{lr}} + (\cos(2\mu) - 1)y_l
\end{aligned}$$

Für den absoluten Fehler gilt somit

$$\Delta\dot{y}_{lr} = \dot{y}_{lr} - d_y = (1 - \cos(2\mu))y_l.$$

Man sieht, daß für $\cos(2\mu) = 1$ der Fehler $\Delta\dot{y}_{lr}$ verschwindet. Durch die Multiplikation mit y_l ist dieser Fehler aber nicht vernachlässigbar und durchaus in der Größenordnung von d_y, was zu großen relativen Fehlern führen kann. Zum Vergleich wollen wir jetzt den Fehler bezüglich der zyklopischen Darstellung berechnen. Dazu benötigen wir ein paar Formeln:

Lemma 3.1 Mit den Bezeichnungen aus diesem Kapitel gilt:

1. $\tilde{y}_c = \frac{Y}{Z_l Z_r}\frac{Z_l+Z_r}{2}$
2. $Z_c = \frac{Z_l+Z_r}{2\cos(\mu)}$

Beweis:
Ad 1.

$$\tilde{y}_c = \frac{y_l + y_r}{2} = \frac{1}{2}(\frac{Y_l}{Z_l} + \frac{Y_r}{Z_r}) = \frac{1}{2}\frac{YZ_r + YZ_l}{Z_l Z_r} = \frac{Y}{Z_l Z_r}\frac{Z_l + Z_r}{2}$$

Ad 2.

$$Z_l + Z_r = (-\sin(\mu)X_c + \cos(\mu)Z_c + u_z^l) + (\sin(\mu)X_c + \cos(\mu)Z_c + u_z^r) \overset{(3.32)}{=} 2\cos(\mu)Z_c$$

□

Um die Formeln einfacher zu halten, schreiben wir $\dot{y}_c$ mit x_c und $\tilde{y}_c$. Für den Fehler kommt dann noch die Approximation $\tilde{x}_c \approx x_c$ hinzu, was aber an der Größenordnung nichts ändert. Man erhält:

$$\begin{aligned}
\dot{y}_c &\approx \frac{v_z^c \tilde{y}_c}{Z_c} - \omega_y x_c \tilde{y}_c \overset{\text{Lemma 3.1.1}}{=} \frac{Y}{Z_l Z_r}(\frac{v_z^c}{Z_c}\frac{Z_l + Z_r}{2} - \omega_y x_c \frac{Z_l + Z_r}{2}) \\
&\overset{\text{Lemma 3.1.2}}{=} \frac{Y}{Z_l Z_r}(\cos(\mu)v_z^c - \sin(2\mu)\cos(\mu)X_c) \\
d_y &= y_r - y_l = \frac{Y_r}{Z_r} - \frac{Y_l}{Z_l} = \frac{Y}{Z_l Z_r}(Z_l - Z_r) \\
&= \frac{Y}{Z_l Z_r}((-\sin(\mu)X_c + \cos(\mu)Z_c + u_z^l) - (\sin(\mu)X_c + \cos(\mu)Z_c + u_z^r)) \\
&\overset{(3.32,3.33)}{=} \frac{Y}{Z_l Z_r}(v_z^c - 2\sin(\mu)X_c)
\end{aligned}$$

$$\begin{aligned}
\Delta \dot{y}_c &= \dot{y}_c - d_y = \frac{Y}{Z_l Z_r}((2\sin(\mu) - \sin(2\mu)\cos(\mu))X_c + (\cos(\mu) - 1)v_z^c) \\
&= \frac{Y}{Z_l Z_r}(2\sin^3(\mu)X_c + (\cos(\mu) - 1)v_z^c) \\
&= (2\sin^3(\mu)X_c + (\cos(\mu) - 1)v_z^c)\frac{1}{Z_r}y_l
\end{aligned}$$

Die zusätzliche Skalierung des Fehlerterms $\Delta\dot{y}_c$ mit $\frac{1}{Z_r}$ gegenüber $\Delta\dot{y}_{lr}$ ist die entscheidende Größe, da das relativ große X_c durch die dritte Potenz des Sinus und v_z^c durch $\cos(\mu) - 1$ skaliert wird.

Wenn im folgenden die allgemeine Form der Verschiebungsrate (3.10, 3.11) verwendet wird, so kann diese also in der zyklopischen Darstellung (3.28, 3.29) am besten bestimmt werden. Der Fehler in der Approximation der Disparität ist dabei in der Relation so gering, daß mit diesen Formeln in nahezu allen denkbaren Anwendungen gearbeitet werden kann, in denen nicht in Y-Richtung translatiert und nicht um X- und Z-Achse rotiert wird. Dabei ist ihre Einfachheit der entscheidende Vorteil gegenüber den Disparitätsformeln. Berechnungen in den folgenden Kapiteln werden dies zeigen. Zudem rekonstruierten Garding et al. [GPMF95] eine frontoparallele Ebene über die Tiefe aus der Approximation der horizontalen Disparität. Sie berichten von einem relativen Fehler in der Tiefe, der im Bildzentrum verschwindet und selbst für $|x_c| = 0.4$ nicht 0.5% überschreitet.

3.4 Zur Fixierungsbedingung

Die hier vorgestellte einfache Beschreibung der Disparität durch die Verschiebungsrate ist, wie gesagt, nur unter der Voraussetzung erfolgt, daß unter den Bedingungen $v_y = 0$, $\omega_x = 0$, $\omega_z = 0$ fixiert wird. Jetzt sollen diese Voraussetzungen genauer untersucht werden. Wir unterscheiden grundsätzlich drei Bedingungen:

Fixierungsbedingung (FIX):
Die Disparität im Bildzentrum verschwindet.

Bedingung der schneidenden optischen Achsen (OPT):
Die optischen Achsen beider Kameras schneiden sich.

Bedingung der einfachen Stereogeometrie (SIS):
$X-$ und Z-Achsen beider Kameras liegen in derselben Ebene. In kontinuierlichen Bewegungsparametern bedeutet das $v_y = 0$, $\omega_x = 0$, $\omega_z = 0$.

Mit den Bedingungen **(FIX)** und **(OPT)** ist dabei jeweils ein 3D-Punkt verbunden, der Fixierungspunkt bzw. der Schnittpunkt der optischen Achsen.

Zunächst ist zu bemerken, daß **(FIX)** und **(OPT)** keineswegs **(SIS)** implizieren. Umgekehrt aber impliziert **(SIS)**, daß die optischen Achsen, in unserem Koordinatensystem die Z-Achsen, sich schneiden, wenn man einschließt, daß sich parallele Achsen im Unendlichen schneiden. Als zweites ist festzustellen, daß ganz allgemein gilt: **(FIX)** impliziert **(OPT)**. Doch obwohl **(OPT)** die allgemeinere Aussage ist, sind beide doch unterschiedlich in ihrer Realisierung: **(FIX)** läßt sich anhand der Bilddaten überprüfen und in aktiver

Form schrittweise erzwingen, **(OPT)** ist in der Regel nur an sensorischen Daten abzulesen, die man aus Messungen am Stereosystem erhält. Insofern ist es eventuell sinnvoll, die Bedingung an das Stereosystem bewußt einzuschränken. Man merke:

$$\begin{aligned}(\textbf{SIS}) &\Rightarrow (\textbf{OPT})\\ (\textbf{FIX}) &\Rightarrow (\textbf{OPT})\end{aligned}$$

Auch die Überprüfbarkeit von **(SIS)** anhand von Bilddaten ist problematisch. Eine nowendige Bedingung für **(SIS)** ist offensichtlich

$$\dot{y}|_{y=0} = 0.$$

Gilt obige Bedingung, so folgt daraus für alle Bildpunkte $(x, 0)$:

$$-\frac{v_y}{Z} + \omega_x - \omega_z x = 0$$

Ist $v_y = 0$, so folgt für das Bildzentrum $\omega_x = 0$, und daraus wiederum $\omega_z = 0$ und somit **(SIS)**. Ist $v_y \neq 0$, so gäbe die Tiefenfunktion im fraglichen Bereich eine Gerade

$$Z(x, 0) = \frac{v_y}{\omega_x - \omega_z x}$$

wieder, deren Vorhandensein zumindest unwahrscheinlich ist. Haben also mehrere Bildpunkte auf der horizontalen Achse des Bildkoordinatensystems verschwindende Disparität, so ist dies ein starkes Indiz dafür, daß eine einfache Stereogeometrie vorliegt.

Was bedeutet dies jetzt für die präsentierten Modelle und Formeln? Die Berechnungen für die Disparitäten wurden nur unter Voraussetzung von **(SIS)** durchgeführt. Auch das geometrische Modell berücksichtigt nicht, ob der Schnittpunkt der optischen Achsen in den Bildern sichtbar ist. Und die Beziehung $\omega_y = -\frac{v_x}{Z_0}$ wurde im zyklopischen Fall rein geometrisch verifiziert. Allerdings führt eine nicht verschwindende Disparität zu einem nicht verschwindenen Fehler in der Approximation durch die Verschiebungsrate (siehe Abbildung 3.3), der im Rahmen der Fehler für die anderen Bildpunkte liegt. Daraus folgt: **Alle Berechnungen und Modelle sind immer noch gültig, wenn man nur $v_y = 0$, $\omega_x = 0$, $\omega_z = 0$ und damit nur schneidende optische Achsen voraussetzt. Im Bildzentrum kann es aber im Gegensatz zu fixierten Systemen zu leichten Fehlern in der Disparitätsapproximation durch die Verschiebungsrate kommen.** Man beachte jeweils die andere Interpretation des Schnittpunktes und der Größe Z_0.

Aufgrund der leichteren Überprüfbarkeit von **(FIX)** aus reiner Bildinformation ist dies die verbreitetere Formulierung, die auch hier weitgehend verwendet wird. In dem Moment, wo man aber zur Durchsetzung von **(SIS)** auf sensorische Information oder mechanisch eingeschränkte Systeme zurückgreifen möchte, ist **(OPT)** ebenso realisierbar.

3.5 Beziehung zum SfM-Problem

Das Stereoproblem wird zum Bewegungsproblem, falls die Parameter der relativen Bewegung zwischen den Kameras unbekannt sind. Das heißt, daß, wie bei der Disparität zu sehen, die unbekannten Bewegungsparameter die Strukturparameter, in dem Fall die Tiefe, beeinflussen, was die Strukturgewinnung erschwert. Hiermit ist eine Beziehung zum Problem der *Strukturerkennung aus Bewegungssehen* (Structure-from-Motion, SfM) geschaffen. Beim SfM wird versucht, die 3D-Struktur über Bewegung in der Szene und eine ganze Bildfolge zu ermitteln. Das Problem entspricht prinzipiell dem Stereoproblem mit mehreren Kameras. Auch die 3-stufige Einteilung des Strukturerschließungsprozesses bleibt gültig. Man gewinnt beim SfM jedoch die unbekannte Struktur mit der unbekannten Bewegung aus Punktekorrespondenzen über mehrere Bilder hinweg. Dieses Vorgehen ist z. B. bei einem sich bewegenden Beobachter angebracht, wo wegen des sich ständig ändernden Koordinatensystems eine (externe) Kalibrierung grundsätzlich nicht praktikabel ist. Es wird meist von einer lokal konstanten Bewegung ausgegangen, die auch über sensorische Informationen nicht vollständig ermittelbar ist. Die gewonnene Struktur wird dann in einem Koordinatensystem, das einer Aufnahme zugeordnet ist, beschrieben, und es muß kein zyklopisches Koordinatensystem eingeführt werden.

Die Annahme, die zumeist gemacht wird, um die Fülle von 3D-Interpretationen einzuschränken, ist die einer *starren Bewegung*, die sich in Translation und Rotation zerlegen läßt. Man beachte dabei, daß diese Annahme nur lokal gemacht werden kann, falls es sich bewegende Objekte in der Szene gibt (Segmentierungsproblem). SfM-Verfahren lassen sich in zwei Gruppen einteilen: Zum einen gibt es sogenannte diskrete Verfahren, die einfach mit Punktekorrespondenzen in mehreren Bildern arbeiten (siehe [Fau93]). Zum anderen gibt es die sogenannten kontinuierlichen Verfahren: Sie gehen von einer geringen Bewegung aus, die durch Geschwindigkeiten approximiert werden kann, und man gelangt so zu den Darstellungen (3.7) und (3.8, 3.9) (siehe [Hor86]). Die Verschiebungsrate ist hier Eingabe von SfM-Verfahren und wird meist durch den *optischen Fluß* (u, v) bestimmt. Man macht dabei die Annahme, daß Grauwerte in der Bildfolge erhalten bleiben, so daß das totale Zeitdifferential der Grauwertfunktion $g(x, y)$ verschwindet:

$$\frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial y}\frac{dy}{dt} + \frac{\partial g}{\partial t} = 0$$

Der optische Fluß wird dann über $(u, v) = (\frac{dx}{dt}, \frac{dy}{dt})$ definiert. Die Unterbestimmtheit dieser Gleichung kann z. B. durch Annahmen über den Verlauf der Flußvektoren (Glattheit) oder weitere Anforderungen an die Grauwertfunktion ($\nabla g(x, y) = \mathbf{0}$) überkommen werden. Auf jeden Fall wird hier mit lokalen Zuordnungsverfahren gearbeitet, was aufgrund der geringen Bewegung im Bild konsequent ist.

Sowohl der diskrete als auch der kontinuierliche Ansatz benötigt zur Strukturgewinnung wenigstens 5 korrespondierende Bildpunktpaare, falls die internen Kameraparameter bekannt sind. Dies ergibt sich aus der Bestimmung der 3 Rotationskomponenten und der 2 Komponenten der Translationsrichtung. Besonders an den Translationsanteilen der Verschiebungsraten (3.8, 3.9) erkennt man, daß ohne Referenztiefe eine Skalierungsmehrdeutigkeit zwischen Tiefe und Translationskomponenten besteht, so daß nur die Richtung der Translation, also v_x/v_z und v_y/v_z bis auf einen konstanten Faktor, zu ermitteln ist. Somit bestimmen die klassischen SfM-Algorithmen, die versuchen, die Struktur über die Bestimmung der Bewegung zu ermitteln, die Szene bis auf eine Ähnlichkeit, also die Euklidische Struktur.

Die Verschiebungsrate kann also sowohl allgemein in SfM-Modellen als auch in den hier beschriebenen SfS-Modellen als Eingabe zur Strukturinterpretation genutzt werden, was im nächsten Kapitel auch geschehen wird. **Man beachte, daß sich die Verschiebungsrate im Stereofall als geeignete Beschreibung der Disparität herausgestellt hat, und zwar unabhängig von der Größe der Bewegung im Bild, aber abhängig von der einfachen Stereogeometrie. Umgekehrt ist im Bewegungsfall die Eignung der Verschiebungsrate abhängig von einer kleinen Bildpunktbewegung, sie ist aber auf allgemeine starre Bewegungen anwendbar.** Für eine Einführung in die Bewegungsanalyse sei auch auf [Jäh91] verwiesen.

3.5.1 Kontinuierliche Analoga

Abschließend einige Bemerkungen zu den hier vorgestellten Konzepten: Die Modelle für Kamera und Welt bleiben uneingeschränkt erhalten für Betrachtungen zu SfM. Man sucht jedoch beim SfM nach kontinuierlichen Analoga zu den meist diskreten Konzepten des SfS, und umgekehrt, um die Vorgehensweise nicht von der Beschreibung der Bewegung abhängig zu machen. So sind Disparität und Verschiebungsrate ein entsprechendes Begriffspaar. Die Fixierungsbedingung wurde als kontinuierliche Bedingung an die Verschiebungsrate eingeführt, entsprach aber im Fall der einfachen Stereogeometrie der entsprechenden Bedingung an die Disparitäten.

Wir wollen nun das Konzept der Fundamentalmatrix näher untersuchen. Es stellt sich die Frage, ob auch die Fundamentalmatrix und die epipolare Bedingung ein kontinuierliches Analogon besitzen. Dies ist in der Tat der Fall. Dazu betrachten wir die Verschiebungsraten in ihrer Vektorform. Wir definieren $\hat{\boldsymbol{z}} =_{\text{def}} (0,0,1)^T$ und erhalten aus (3.7) für einen Szenepunkt $\boldsymbol{X}$ mit zugehörigem Sehstrahl $\boldsymbol{x}$ unter Ausnutzung von Eigenschaften des Skalarprodukts und des Vektorprodukts, vor allem $\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a}^T\boldsymbol{c})\boldsymbol{b} - (\boldsymbol{a}^T\boldsymbol{b})\boldsymbol{c}$:

$$
\begin{aligned}
\dot{\boldsymbol{x}} &= [\dot{\overline{\frac{1}{\hat{\boldsymbol{z}}^T\boldsymbol{X}}\boldsymbol{X}}}] = [\dot{\overline{\frac{1}{\hat{\boldsymbol{z}}^T\boldsymbol{X}}}}]\boldsymbol{X} + \frac{1}{\hat{\boldsymbol{z}}^T\boldsymbol{X}}\dot{\boldsymbol{X}} = -\frac{\hat{\boldsymbol{z}}^T\dot{\boldsymbol{X}}}{(\hat{\boldsymbol{z}}^T\boldsymbol{X})^2}\boldsymbol{X} + \frac{1}{\hat{\boldsymbol{z}}^T\boldsymbol{X}}\dot{\boldsymbol{X}} \\
&= \frac{1}{\hat{\boldsymbol{z}}^T\boldsymbol{X}}(\hat{\boldsymbol{z}}^T\boldsymbol{v})\boldsymbol{x} + (\hat{\boldsymbol{z}}^T(\boldsymbol{\omega} \times \boldsymbol{x}))\boldsymbol{x} - \frac{1}{\hat{\boldsymbol{z}}^T\boldsymbol{X}}\boldsymbol{v} - \boldsymbol{\omega} \times \boldsymbol{x} \\
&= -\frac{1}{\hat{\boldsymbol{z}}^T\boldsymbol{X}}\hat{\boldsymbol{z}} \times (\boldsymbol{v} \times \boldsymbol{x}) - \hat{\boldsymbol{z}} \times (\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{\omega}))
\end{aligned}
$$

Das Skalarprodukt mit $(\boldsymbol{v} \times \boldsymbol{x})$ liefert

$$
\begin{aligned}
\dot{\boldsymbol{x}}^T(\boldsymbol{v} \times \boldsymbol{x}) &= -\hat{\boldsymbol{z}} \times (\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{\omega}))^T(\boldsymbol{v} \times \boldsymbol{x}) \\
&= -[(\hat{\boldsymbol{z}}^T(\boldsymbol{x} \times \boldsymbol{\omega}))\boldsymbol{x} - (\hat{\boldsymbol{z}}^T\boldsymbol{x})(\boldsymbol{x} \times \boldsymbol{\omega})]^T(\boldsymbol{v} \times \boldsymbol{x}) \\
&= (\boldsymbol{x} \times \boldsymbol{\omega})^T(\boldsymbol{v} \times \boldsymbol{x}).
\end{aligned}
$$

Daraus folgt die tiefenfreie Beziehung

$$
(\boldsymbol{v} \times \boldsymbol{x})^T(\dot{\boldsymbol{x}} - \boldsymbol{x} \times \boldsymbol{\omega}) = 0. \tag{3.38}
$$

Wegen

$$
\begin{aligned}
(\boldsymbol{v} \times \boldsymbol{x})^T(\boldsymbol{x} \times \boldsymbol{\omega}) &= (\boldsymbol{v}^T\boldsymbol{x})(\boldsymbol{x}^T\boldsymbol{\omega}) - (\boldsymbol{x}^T\boldsymbol{x})(\boldsymbol{v}^T\boldsymbol{\omega}) \\
&= \boldsymbol{x}^T(\frac{\boldsymbol{\omega}\boldsymbol{v}^T + \boldsymbol{v}\boldsymbol{\omega}^T}{2} - \boldsymbol{\omega}^T\boldsymbol{v}I_3)\boldsymbol{x}
\end{aligned}
$$

erhält man auch die Form

$$(\boldsymbol{x} \times \dot{\boldsymbol{x}})^T \boldsymbol{v} - \boldsymbol{x}^T \left(\frac{\boldsymbol{\omega}\boldsymbol{v}^T + \boldsymbol{v}\boldsymbol{\omega}^T}{2} - \boldsymbol{\omega}^T \boldsymbol{v} I_3\right)\boldsymbol{x} = 0, \tag{3.39}$$

wobei I_n die $n \times n$-Einheitsmatrix denotiert. Die Gleichungen (3.38) und (3.39) sind Ausdruck der epipolaren Bedingung im kontinuierlichen Fall. Sie beschränken die Bewegungsparameter in Abhängigkeit von der Bildpunktgeschwindigkeit, aber unabhängig von Tiefeninformation. Zudem bietet (3.39) den Zugang zur Bewegungsanalyse über die Matrix $K =_{\text{def}} \frac{\boldsymbol{\omega}\boldsymbol{v}^T + \boldsymbol{v}\boldsymbol{\omega}^T}{2} - \boldsymbol{\omega}^T \boldsymbol{v} I_3$, die symmetrisch ist und für die gilt: $\boldsymbol{v}^T K \boldsymbol{v} = \boldsymbol{0}$ und $\boldsymbol{\omega}^T K \boldsymbol{\omega} = \boldsymbol{0}$. Die 6 Parameter der Matrix K und die 3 Parameter des Vektors $\boldsymbol{v}$ bilden die *essentiellen Parameter im kontinuierlichen Fall*. Sie hängen wiederum lediglich von den 5 Bewegungsparametern ab. Dementsprechend gibt es neben der Skalierungsmehrdeutigkeit der Translation 3 Nebenbedingungen, die sich aus der konkreten Gestalt der essentiellen Parameter in Abhängigkeit der Bewegungsparameter ergeben. Dazu und zu weiteren Einzelheiten siehe [Dan92].

Führt man auch hier unbekannte interne Parameter ein, so erhält man bei Voraussetzung ihrer zeitlichen Konstanz für die Geschwindigkeit der Bildpunkte in Pixeln

$$\dot{\boldsymbol{x}}_p = (\dot{S\boldsymbol{x}}) = S\dot{\boldsymbol{x}}.$$

Setzt man die Pixelkoordinaten in (3.39) ein, so ergibt sich, wiederum unter Benutzung von $A\boldsymbol{a} \times A\boldsymbol{b} = \det(A)A^{-T}(\boldsymbol{a} \times \boldsymbol{b})$:

$$\begin{aligned}
(\boldsymbol{x} \times \dot{\boldsymbol{x}})^T \boldsymbol{v} - \boldsymbol{x}^T K \boldsymbol{x} &= ((S^{-1}\boldsymbol{x}_p) \times (S^{-1}\dot{\boldsymbol{x}}_p))^T (S^{-1}\boldsymbol{e}) - (S^{-1}\boldsymbol{x}_p)^T K (S^{-1}\boldsymbol{x}_p) \\
&= \det(S^{-1})((S^T(\boldsymbol{x}_p \times \boldsymbol{x}_p))^T (S^{-1}\boldsymbol{e}) - {\boldsymbol{x}_p}^T S^{-T} K S^{-1} \boldsymbol{x}_p \\
&= \frac{1}{\det(S)}(\boldsymbol{x}_p \times \dot{\boldsymbol{x}}_p)^T \boldsymbol{e} - {\boldsymbol{x}_p}^T S^{-T} K S^{-1} \boldsymbol{x}_p
\end{aligned}$$

Somit erhält man die Darstellung

$$(\boldsymbol{x}_p \times \dot{\boldsymbol{x}}_p)^T \boldsymbol{e} - {\boldsymbol{x}_p}^T M \boldsymbol{x}_p = 0,$$

wobei $\boldsymbol{e} =_{\text{def}} S\boldsymbol{v}$ den Expansionspunkt in Pixelkoordinaten bezeichnet und $M =_{\text{def}} \det(S) S^{-T} K S^{-1}$ die unkalibrierte Enstprechung der Matrix K ist. In $\boldsymbol{e}$ und der symmetrischen Matrix M sind die 9 *Fundamentalparameter* enthalten, die von den 5 Bewegungsparametern und den 5 internen Parametern abhängen. Es gibt hier außer der Skalierungsmehrdeutigkeit noch die Nebenbedingung $\boldsymbol{e}^T M \boldsymbol{e} = \boldsymbol{0}$. Daher können die Fundamentalparameter aus 7 Bildpunkten mit ihren Geschwindigkeiten, jeweils in Pixeln beschrieben, bestimmt werden, was der diskreten Beschreibung als Fundamentalmatrix entspricht.

Geht man von der hier beschriebenen einfachen Stereogeometrie **(SIS)** aus, so werden diese Darstellungen ebenfalls einfacher. Durch $v_y = 0$, $\omega_x = 0$, $\omega_z = 0$ gilt im kalibrierten Fall: $\boldsymbol{v} = (v_x, 0, v_z)^T$ und

$$K \overset{\boldsymbol{v}^T\boldsymbol{\omega}=\boldsymbol{0}}{=} \frac{1}{2}(\boldsymbol{\omega}\boldsymbol{v}^T + \boldsymbol{v}\boldsymbol{\omega}^T) = \frac{1}{2}\begin{pmatrix} 0 & v_x\omega_y & 0 \\ v_x\omega_y & 0 & v_z\omega_y \\ 0 & v_z\omega_y & 0 \end{pmatrix}.$$

Die 4 essentiellen Parameter $(u_1, u_2, u_3, u_4) =_{\mathrm{def}} (v_x\omega_y, v_z\omega_y, v_x, v_z)$ müssen die Nebenbedingung $u_1u_4 = u_2u_3$ erfüllen. Mit der Skalierungsmehrdeutigkeit kommt man so auf die 2 bestimmenden Bewegungsparameter v_x/v_z und ω_y. Im unkalibrierten Fall ergibt sich allgemein aus der Struktur der Matrizen S^{-T}, S^{-1} und K: $M_{11} = 0$. Dies erhält man als zusätzliche Nebenbedingung zum Fall der uneingeschränkten Bewegung. Somit kommt man auf die Abhängigkeit von 6 Fundamentalparametern, die sich aus den 5 internen und 2 Bewegungsparametern zusammensetzen. Es ergibt sich im allgemein üblichen Fall von $s_{xy} = 0$ sogar noch mehr. MAPLE berechnet explizit:

$$\begin{aligned} \boldsymbol{e} &= \begin{pmatrix} s_xv_x + x_0v_z \\ y_0v_z \\ v_z \end{pmatrix} \\ M &= \frac{1}{2s_xs_y} \begin{pmatrix} 0 & v_x\omega_y & -y_0v_x\omega_y \\ v_x\omega_y & 0 & \omega_y(-x_0v_x + s_xv_z) \\ -y_0v_x\omega_y & \omega_y(-x_0v_x + s_xv_z) & -2y_0\omega_y(-x_0v_x + s_xv_z) \end{pmatrix} \end{aligned}$$

Daraus gewinnt man direkt $y_0 = \frac{e_y}{e_z} = -\frac{M_{13}}{M_{12}} = -\frac{M_{33}}{2M_{23}}$. Mit diesen 2 Nebenbedingungen und $M_{11} = M_{22} = 0$ kommt man auf 4 Fundamentalparameter. Diese bestehen aus 2 Bewegungsparametern und 4 internen Parametern.

Wir schließen mit der Analogie der einfachen Stereogeometrie zum diskreten Fall. Die essentielle Matrix bekommt hier die Form der Matrix K. Aus (2.23) erhält man

$$\begin{aligned} E &= \begin{pmatrix} 0 & -t_z^l & 0 \\ t_z^l & 0 & -t_x^l \\ 0 & t_x^l & 0 \end{pmatrix} \begin{pmatrix} \cos(\mu_y) & 0 & \sin(\mu_y) \\ 0 & 1 & 0 \\ -\sin(\mu_y) & 0 & \cos(\mu_y) \end{pmatrix} \\ &= \begin{pmatrix} 0 & -t_z^l & 0 \\ t_z^l\cos(\mu_y) + t_x^l\sin(\mu_y) & 0 & t_z^l\sin(\mu_y) - t_x^l\cos(\mu_y) \\ 0 & t_x^l & 0 \end{pmatrix} = \begin{pmatrix} 0 & -t_z^l & 0 \\ -t_z^r & 0 & t_x^r \\ 0 & t_x^l & 0 \end{pmatrix}, \end{aligned}$$

wobei $(t_x^l, 0, t_z^l)$ bzw. $(t_x^r, 0, t_z^r)$ die Translationen von links nach rechts bzw. von rechts nach links im jeweiligen Koordinatensystem und μ_y den Rotationswinkel um die Y-Achse bezeichnen. Diese Matrix ist, wie K und $\boldsymbol{v}$ oben, nur durch die 2 Parameter der relativen Bewegung bestimmt, wobei die nicht verschwindenden Koeffizienten der Matrix der Nebenbedingung $(E_{21}^2 + E_{23}^2) - (E_{12}^2 + E_{32}^2) = 0$ unterliegen. Die Fundamentalmatrix hat im Falle gleicher interner Parameter die Gestalt $F = S^{-T}ES^{-1}$. Neben der Bedingung $\det(F) = 0$ ergibt sich hier wiederum aus der Form der Matrizen $F_{11} = 0$. Somit hängt in diesem Fall die Fundamentalmatrix nur von 6 Parametern ab, die sich aus den 2 Parametern der Bewegung und den 5 internen Parametern zusammensetzen. Eine Vereinfachung liefert wiederum $s_{xy} = 0$. Man gewinnt mit Hilfe von MAPLE:

$$\begin{aligned} F &= \frac{1}{s_x}s_y \begin{pmatrix} 0 & -t_z^l & y_0t_z^l \\ -t_z^r & 0 & s_xt_x^r + x_0t_z^r \\ y_0t_z^r & s_xt_x^l + x_0t_z^l & -y_0(s_xt_x^l + x_0t_z^l + s_xt_x^r + x_0t_z^r) \end{pmatrix} \\ &= \begin{pmatrix} 0 & -e_{l_z} & e_{l_y} \\ -e_{r_z} & 0 & e_{r_x} \\ e_{r_y} & e_{l_x} & -y_0(e_{l_x} + e_{r_x}) \end{pmatrix} \end{aligned}$$

Dies folgt mit der Komponentendarstellung der Epipole $\boldsymbol{e} = S\boldsymbol{t} = (s_x t_x + x_0 t_z, y_0 t_z, t_z)^T$. Man erhält hier $-y_0 = \frac{F_{13}}{F_{12}} = \frac{F_{31}}{F_{21}} = \frac{F_{33}}{F_{23}+F_{32}}$ und somit 2 Nebenbedingungen zusätzlich zu $F_{11} = 0$ und $F_{22} = 0$. Aus diesen Bedingungen folgt allerdings bereits $\det(F) = 0$, wie sich leicht zeigen läßt. Die Fundamentalmatrix ist also in diesem Fall durch 4 Parameter bestimmt, die sich aus den beiden Bewegungsparametern μ_y, t_x/t_z und 3 internen Parametern zusammensetzen, da s_y globaler Skalierungsfaktor und damit unbestimmbar ist. Man sieht weiter, daß hier das Problem der Bestimmung der Fundamentalmatrix im wesentlichen ein Problem der Epipolbestimmung ist und von dieser Unsicherheit abhängt. Wir werden im nächsten Kapitel allerdings mit nur einem einzigen Punkt eine projektive Rekonstruktion ermöglichen.

Kapitel 4

3D-Strukturinformation aus Verschiebungsraten

Wir wollen jetzt untersuchen, welche Informationen über die 3D-Struktur direkt aus der Verschiebungsrate zu gewinnen sind. Wir verzichten zunächst bewußt auf die Kombination mehrerer Vektoren zur Kalibrierung des Systems. Vorläufig seien aber die internen Parameter beider Kameras bekannt und identisch.

Im Zuordnungsschritt sei für einen Bildpunkt $\boldsymbol{x} = (x, y)$ die Verschiebungsrate $\dot{\boldsymbol{x}} = (\dot{x}, \dot{y})$ bestimmt worden. Betrachten wir zunächst den Fall der einfachen Stereogeometrie **(SIS)**. Die Verschiebungsrate steht dann mit den Systemparametern und der Tiefe Z des abgebildeten Szenepunktes wie folgt in Beziehung:

$$\dot{x} = -\frac{v_x}{Z} + \frac{v_z}{Z}x - \omega_y(x^2 + 1) \qquad (4.1)$$

$$\dot{y} = \frac{v_z}{Z}y - \omega_y xy \qquad (4.2)$$

Zunächst betrachten wir die klassische Aufgabe der Tiefenrekonstruktion. Sind v_x, v_z und ω_y bekannt, so läßt sich die Tiefe fast immer sowohl aus der horizontalen als auch der vertikalen Verschiebungsrate bestimmen.

Bemerkung 4.1 Für ein binokulares Stereosystem, das wie beschrieben fixiert, gilt:

1. (a) Die Tiefe läßt sich nicht aus $\dot{x}$ bestimmen genau dann, wenn $v_z \neq 0$ und $x = \frac{v_x}{v_z}$.
   (b) Die Tiefe läßt sich nicht aus $\dot{y}$ bestimmen genau dann, wenn $v_z = 0$ oder $y = 0$.
   (c) Die Tiefe läßt sich nicht aus $\dot{\boldsymbol{x}}$ bestimmen genau dann, wenn $v_z \neq 0$ und $(x, y) = (\frac{v_x}{v_z}, 0)$.
2. (a) Falls $x \neq \frac{v_x}{v_z}$, so gilt
   $$Z = \frac{-v_x + v_z x}{\dot{x} + \omega_y(x^2 + 1)}.$$
   (b) Falls $v_z \neq 0$ und $y \neq 0$, so gilt
   $$Z = \frac{v_z y}{\dot{y} + \omega_y xy}.$$

Beweis:
Ad 1(a), 2(a). Ist $\dot{x} + \omega_y(x^2+1) \neq 0$, so gilt $Z = \frac{-v_x+v_z x}{\dot{x}+\omega_y(x^2+1)}$. Ansonsten gilt

$$\dot{x} + \omega_y(x^2+1) = 0 \quad \Leftrightarrow \quad \frac{-v_x + v_z x}{Z} = 0 \quad \overset{Z\neq 0}{\Leftrightarrow} \quad -v_x + v_z x = 0$$
$$\overset{v_x\neq 0}{\Leftrightarrow} \quad x = \frac{v_x}{v_z} \;\wedge\; v_z \neq 0.$$

Die Fälle $Z = 0$ und $v_x = 0$ können aus geometrischen Gründen ausgeschlossen werden.
Ad 1(b), 2(b). Ist $\dot{y} + \omega_y xy \neq 0$, so gilt $Z = \frac{v_z y}{\dot{y}+\omega_y xy}$. Ansonsten gilt

$$\dot{y} + \omega_y xy = 0 \quad \Leftrightarrow \quad v_z y = 0 \quad \Leftrightarrow \quad v_z = 0 \;\vee\; y = 0.$$

Ad 1(c). Folgt aus der Konjunktion der Bedingungen aus 1(a) und 1(b). □

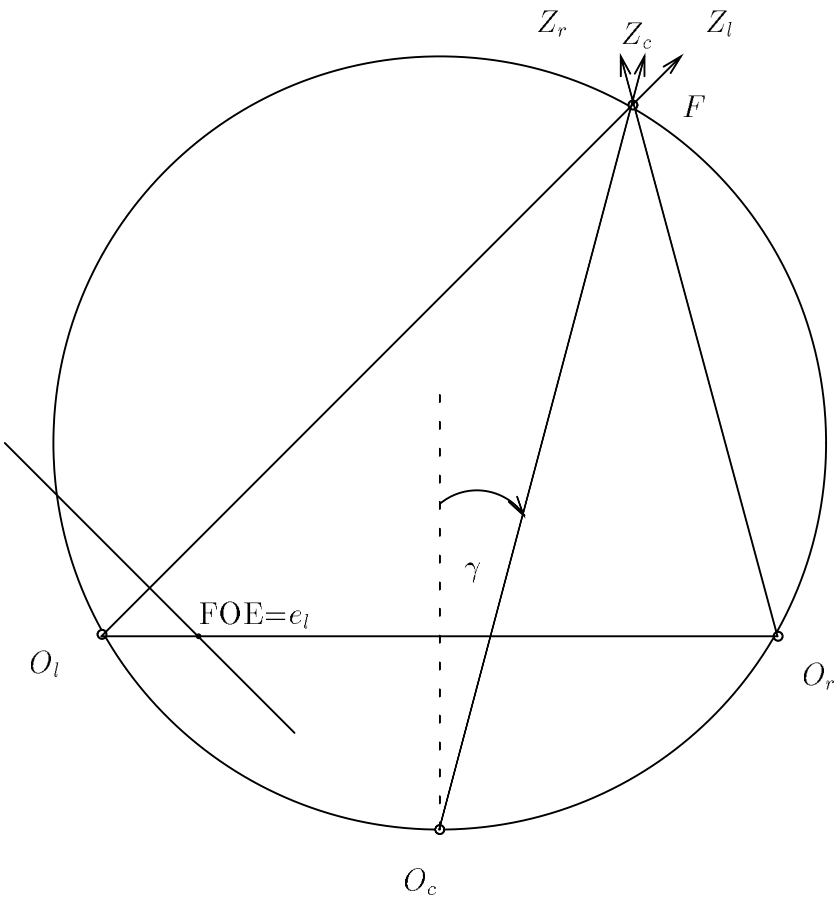

Abbildung 4.1: Für Punkte auf dem Sehstrahl $(\frac{v_x}{v_z}, 0, 1)^T$ kann aus den Verschiebungsraten die Tiefe nicht bestimmt werden. Der zugehörige Punkt auf der normierten Bildebene heißt Expansionspunkt FOE. Für ein binokulares Stereosystem ist er gleich dem Epipol e, i. e. die Projektion des optischen Zentrums der anderen Kamera.

Abbildung 4.1 zeigt die Situation, in der die Tiefe eines Szenepunktes nicht bestimmt werden kann. Man sieht, daß dieser Punkt in der Regel außerhalb des Sichtfeldes liegt. Sein Bildpunkt ist der sogenannte *Expansionspunkt (Focus Of Expansion, FOE)*, der im

Falle allgemeiner Bewegung die Form $(\frac{v_x}{v_z}, \frac{v_y}{v_z})$ hat und die Information über die Translationsrichtung trägt. Hat man es mit einem binokularen Stereosystem zu tun und beschreibt die Translation im linken Koordinatensystem, was auch unsere zyklopische Darstellung tut, so liegt der Zentralprojektionspunkt der rechten Kamera in Translationsrichtung, und der Expansionspunkt ist gleich dem Epipol. Es sei für den Bewegungsfall bemerkt, daß sich im FOE auch im allgemeinen, nicht fixierten Fall die Tiefe nicht ermitteln läßt. Man mache sich dies klar, indem man Bemerkung 4.1 für die allgemeinen Verschiebungsraten formuliert und den Beweis analog führt.

In der Regel erhält man also sogar eine Überbestimmtheit für die Tiefe aus den Verschiebungsraten. **Mit Kenntnis der Tiefe kann man dann über die Sehstrahlgleichung $\boldsymbol{X} = Z\boldsymbol{x}$ einfach alle Koordinaten des Szenepunktes bestimmen.** Da in den Gleichungen für die Disparität die Tiefe immer reziprok auftaucht, führt man den Begriff der *Nähe* λ ein, der äquivalent gebraucht werden kann.

$$\lambda =_{\mathrm{def}} \frac{1}{Z}$$

Die Nähe des Fixierungspunktes sei

$$\lambda_0 =_{\mathrm{def}} \frac{1}{Z_0}.$$

Zur Berechnung der Tiefe benötigt man die Systemparameter v_x, v_z, ω_y, die z. B. Ergebnis einer Kalibrierung sein können. Wir wollen, wie bereits erwähnt, einen Ansatz verfolgen, bei dem diese Werte nicht vorliegen. Der Rest des Kapitels beschäftigt sich damit, welche Informationen man dann bekommen kann, und wie sie zu berechnen sind. Grundlage ist dabei die Präsentation in [GPMF95].

Die ersten beiden Abschnitte dieses Kapitels beschreiben, welche Informationen man aus der Verschiebungsrate gewinnen kann. Der dritte, vierte und fünfte Abschnitt beantworten die wie-Frage in verschiedenen Ansätzen und erweitern diese insbesondere auf den Fall unbekannter interner Kameraparameter. Das Kapitel schließt mit einem experimentellen Vergleich der verschiedenen Verfahren und damit verbundener Diskussion.

4.1 Skalierte relative Nähe und Relieftransformation

Da in der Verschiebungsrate die Nähe immer durch die Parameter der relativen Translationsbewegung skaliert auftritt, ist ohne ihre Kenntnis direkte Tiefenbestimmung nicht möglich. Wie wir später sehen werden, sind aber einfache Abbildungen der Tiefe bzw. der Nähe ermittelbar, deren Eigenschaften zunächst zu untersuchen sind.

Wir definieren die *skalierte relative Nähe (scaled relative nearness, SRN)* durch

$$\sigma =_{\mathrm{def}} A\lambda + B,$$

wobei A und B Konstanten sind mit $A \neq 0$. [1] Dies ist eine Funktion der Menge der Bildpunkte, was man durch $\sigma(x, y) =_{\mathrm{def}} A\lambda(x, y) + B$ ausdrücken kann. Welche Informationen liegen nun in Form der SRN vor?

[1] Garding et al. [GPMF95] definieren $\sigma =_{\mathrm{def}} A(B - \lambda)$, um die aus der Verschiebungsrate gewonnenen Daten direkt den Konstanten A und B zuzuordnen.

Bemerkung 4.2

1. **Die skalierten relativen Nähen erhalten das Teilverhältnis der Nähen,** das heißt, für drei Szenepunkte mit den Nähen λ_1, λ_2, λ_3 und den skalierten relativen Nähen σ_1, σ_2, σ_3 gilt
$$\frac{\sigma_1 - \sigma_2}{\sigma_1 - \sigma_3} = \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_3}.$$

2. **Die skalierten relativen Nähen erhalten die Tiefenordnung, falls $A < 0$, und kehren sie um, falls $A > 0$.**

Beweis: Seien drei Szenepunkte mit den Nähen λ_1, λ_2, λ_3 und den skalierten relativen Nähen σ_1, σ_2, σ_3 gegeben.

Ad 1.
$$\frac{\sigma_1 - \sigma_2}{\sigma_1 - \sigma_3} = \frac{(A\lambda_1 + B) - (A\lambda_2 + B)}{(A\lambda_1 + B) - (A\lambda_3 + B)} = \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_3}$$

Ad 2.

1. Fall: $A < 0$. $\sigma_1 \leq \sigma_2 \Leftrightarrow A\lambda_1 + B \leq A\lambda_2 + B \Leftrightarrow \lambda_2 \leq \lambda_1 \Leftrightarrow Z_1 \leq Z_2$

2. Fall: $A > 0$. $\sigma_1 \leq \sigma_2 \Leftrightarrow A\lambda_1 + B \leq A\lambda_2 + B \Leftrightarrow \lambda_1 \leq \lambda_2 \Leftrightarrow Z_2 \leq Z_1$

□

Die SRN enthält also bereits wichtige Informationen über die Szene, die in Anwendungen benötigt werden, wie die Tiefenordnung. Der zweite wichtige Aspekt der SRN ist, daß es möglich ist, direkt daraus die Szene zu rekonstruieren. Wie gezeigt, läßt sich der gesuchte Szenepunkt aus dem Sehstrahl und der Tiefe rekonstruieren (siehe 2.12). Der Sehstrahl ist hier ebenfalls als bekannt vorausgesetzt. Die SRN liefert aber nicht direkt Euklidische Tiefe, sondern eine „projektive Transformation der Tiefe“:

$$\sigma = \frac{A + BZ}{Z} \qquad \frac{1}{\sigma} = \frac{Z}{A + BZ}$$

Dieser Sprachgebrauch rührt daher, daß, wie in §2.4 dargestellt, projektive Transformationen, angewendet auf Euklidische Punkte, gebrochen rationale Funktionen in den einzelnen Komponenten sind. Eine Rekonstruktion der Szene mit σ anstelle von Z liefert dann auch eine projektive Rekonstruktion statt einer Euklidischen. Dazu betrachten wir zunächst real beobachtbare Euklidische Punkte $\boldsymbol{X} = (X, Y, Z, 1)^T$ mit $Z > 0$ im projektiven Raum $\mathcal{P}^3$ und die projektive Gleichheit

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \equiv \begin{pmatrix} x \\ y \\ 1 \\ \lambda \end{pmatrix}.$$

Die projektiven Rekonstruktionen werden induziert durch die projektive Transformation von real beobachtbaren Euklidischen Szenepunkten, wie sie die SRN bewirkt. Wir stellen insgesamt drei vor.

Die σ-W-Rekonstruktion:
Die 3D-Rekonstruktion

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} =_{\text{def}} \begin{pmatrix} x \\ y \\ 1 \\ \sigma \end{pmatrix}$$

entspricht einer projektiven Transformation der Originalszene

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} = \mathcal{R}_1 \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & B & A \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} I_3 & \mathbf{0} \\ \boldsymbol{B}^T & A \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix},$$

die zwischen Euklidischen Szenepunkten als

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \frac{1}{A + BZ} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

geschrieben werden kann.

Diese Rekonstruktion wird als „σ-W“ bezeichnet, da sich die Transformation auf die W-Koordinate bezieht. Die Bruchschreibweise $\frac{1}{A+BZ}$ ist lediglich Ausdruck der projektiven Abbildung und bedeutet nicht, daß $A + BZ \neq 0$ für alle betrachteten Punkte ist.

Die σ-Z-Rekonstruktion:
Die 3D-Rekonstruktion

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} =_{\text{def}} \begin{pmatrix} x \\ y \\ \sigma \\ 1 \end{pmatrix}$$

entspricht einer projektiven Transformation der Originalszene

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} = \mathcal{R}_2 \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & B & A \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} I_2 & 0 \\ 0 & M_{AB} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix},$$

die zwischen Euklidischen Szenepunkten als

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} X \\ Y \\ A + BZ \end{pmatrix}$$

geschrieben werden kann.

Diese Rekonstruktion wird als „σ-Z“ bezeichnet, da sich die Transformation auf die Z-Koordinate bezieht. Für alle betrachteten Punkte ist die Euklidische Schreibweise stets definiert. Diese beiden Rekonstruktionen haben jedoch mit der realen Szene nur sehr wenig gemeinsam. So werden bei der σ-W-Rekonstruktion eventuell unendlich ferne Punkte und bei der σ-Z-Rekonstruktion Euklidische Punkte mit negativer Z-Koordinate rekonstruiert. Man möchte vielmehr die Parameter A und B selbst festsetzen durch $A' \neq 0$ und B' und damit eine geschätzte Nähe λ' erhalten, die eine physikalisch sinnvolle Rekonstruktion ermöglicht:

$$\lambda' =_{\text{def}} \frac{\sigma - B'}{A'} = \frac{A}{A'}\lambda + \frac{B - B'}{A'} = a\lambda + b = \frac{a + bZ}{Z}$$

mit $a =_{\text{def}} \frac{A}{A'} \neq 0$ und $b =_{\text{def}} \frac{B-B'}{A'}$. In Abhängigkeit der gewonnenen σ-Werte können dann A' und B' so gewählt werden, daß für alle Punkte $\lambda' > 0$, also $Z' > 0$ gilt. Dies ist theoretisch immer möglich, da diese Bedingung für $A = A'$ und $B = B'$, also $\lambda = \lambda'$, erfüllt ist. Wir werden bei den Experimenten sehen, wie man dies in der Praxis verwirklicht.

Die σ-λ'-Rekonstruktion:
Die 3D-Rekonstruktion

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} =_{\text{def}} \begin{pmatrix} x \\ y \\ 1 \\ \lambda' \end{pmatrix}$$

entspricht einer projektiven Transformation der Originalszene

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} = \mathcal{R}_3 \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & b & a \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} I_3 & \mathbf{0} \\ \boldsymbol{b}^T & a \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix},$$

die zwischen Euklidischen Szenepunkten als

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \frac{1}{a + bZ} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

geschrieben werden kann.

Alle diese Rekonstruktionen liefern also die projektive Struktur der Szene. Insbesondere interessiert man sich für die projektiv linearen, invertierbaren Funktionen $\mathcal{R}_1$, $\mathcal{R}_2$ und $\mathcal{R}_3$, die die rekonstruierte mit der realen Szene verbinden.

Bemerkung 4.3

1. **$\mathcal{R}_1$, $\mathcal{R}_2$ und $\mathcal{R}_3$ sind projektiv lineare, invertierbare Funktionen und erhalten daher z. B. Inzidenz, Doppelverhältnis, Kollinearität, Komplanarität.**

2. **$\mathcal{R}_2$ und $\mathcal{R}_3$ stellen invertierbare C^∞-Funktionen von $\{(X, Y, Z) \in \boldsymbol{R}^3 | Z > 0\}$ nach $\boldsymbol{R}^3$ dar.**

Beweis:
Ad 1. klar nach den Definitionen.
Ad 2. Die C^∞-Eigenschaft ergibt sich aus den Euklidischen Darstellungen der Funktionen. □

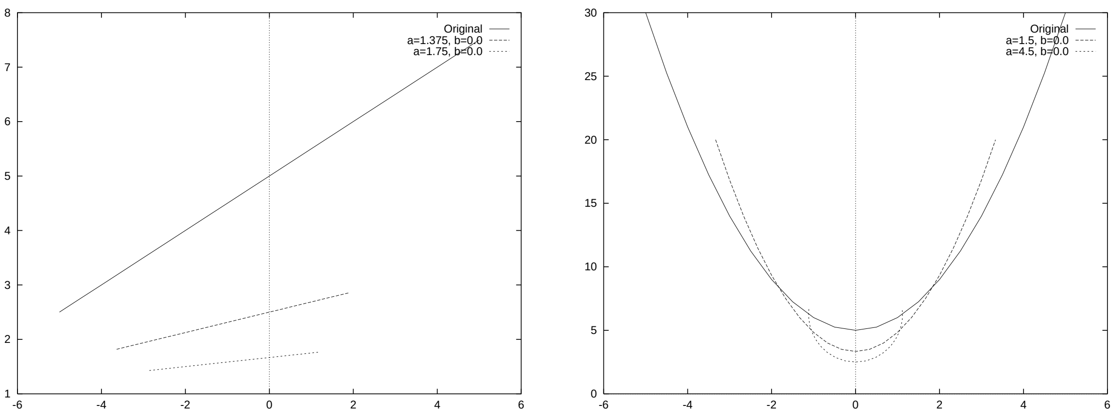

Abbildung 4.2: Relieftransformationen. Die Diagramme zeigen horizontale Schnitte einer Gruppe von Flächen, die sich über Relieftransformation in Relation befinden. Man beachte, daß Ebenen auf Ebenen abgebildet werden und die Tiefenordnung erhalten bleibt.

Zu 2. ist noch zu sagen, daß projektive Räume über Körpern auch differenzierbare Mannigfaltigkeiten sind, und die Menge der projektiv linearen, invertierbaren Abbildungen in der Menge der Diffeomorphismen enthalten ist. 2. ist ein typisches Beispiel dafür, sich von der Allgemeinheit und einfachen Beschreibung der projektiven Geometrie zu lösen und unter Berücksichtigung von Punkten, die im Endlichen (Euklidischen bezüglich des hier verwendeten Koordinatensystems) und vor der Kamera ($Z > 0$) liegen, zu weiteren Erkenntnissen zu gelangen. Dies wird in der folgenden, wesentlichen Bemerkung dieses Abschnittes fortgesetzt.

Bemerkung 4.4 Als Funktionen des Euklidischen Raumes $\boldsymbol{R}^3$ aufgefaßt, gilt:

$$\left.\begin{array}{c}\mathcal{R}_2\\ \mathcal{R}_3\end{array}\right\} \textbf{ erhält die Tiefenordnung genau dann, wenn } \left\{\begin{array}{l} A < 0 \\ \operatorname{sign}(A') = \operatorname{sign}(A)\end{array}\right. .$$

Ansonsten kehren die Abbildungen die Tiefenordnung um.

Beweis: Im Fall $\mathcal{R}_2$ ist dies in Bemerkung 4.2.2 für die SRN bereits gezeigt worden. Man betrachte jetzt $\mathcal{R}_3$. Seien Z_1, Z_2 reale und Z_1', Z_2' rekonstruierte Tiefen. Dann gilt

$$Z_1' \leq Z_2' \quad \Leftrightarrow \quad \lambda_2' \leq \lambda_1' \quad \Leftrightarrow \quad \frac{\sigma_2 - B'}{A'} \leq \frac{\sigma_1 - B'}{A'},$$

also, falls $A' < 0$,

$$Z_1' \leq Z_2' \quad \Leftrightarrow \quad \sigma_1 \leq \sigma_2$$

und, falls $A' > 0$,

$$Z_1' \leq Z_2' \quad \Leftrightarrow \quad \sigma_2 \leq \sigma_1.$$

Mit Bemerkung 4.2.2 folgt dann die Behauptung. □

Dies ist die entscheidende Eigenschaft dieser projektiven Transformationen, die sie gegenüber allgemeinen projektiven Transformationen auszeichnen. Man nennt eine solche Abbildung *Relieftransformation*. Es ist klar, daß die Information über die Tiefenordnung viele Aufgaben im 3D-Maschinensehen erleichtert, z. B. die Navigation. Abbildung 4.2 zeigt Flächen, die über eine Relieftransformation in Beziehung stehen.

Die Rekonstruktion aus Sehstrahlen läßt weitere Rückschlüsse auf die Gestalt der Szene zu. Wir betrachten das Fixgebilde der Transformationen, das sich allgemein für beliebige Parameter ergibt.

Bemerkung 4.5

1. **$\mathcal{R}_1$ und $\mathcal{R}_3$ lassen alle Punkte der Ebene $(A-1)W+BZ=0$ und den Koordinatenursprung fix, $\mathcal{R}_2$ die Punkte mit $Z=W=0$.**
2. **$\mathcal{R}_1$ und $\mathcal{R}_3$ haben außerdem genau die Sehstrahlen als Fixgeraden. $\mathcal{R}_2$ hat im allgemeinen keine Fixgeraden.**

Beweis:

Ad 1. Seien (X,Y,Z,W) die Koordinaten eines Fixpunktes. Betrachten wir zunächst $\mathcal{R}_1$, was auch den Fall $\mathcal{R}_3$ erledigt. Dann gibt es ein $\mu \in \boldsymbol{R} \setminus \{0\}$ mit

$$X=\mu X \quad Y=\mu Y \quad Z=\mu Z \quad W=\mu(AW+BZ).$$

1. Fall: $X \neq 0 \vee Y \neq 0 \vee Z \neq 0$

   Dann ist $\mu=1$ und $(A-1)W+BZ=0$.

2. Fall: $X=0 \wedge Y=0 \wedge Z=0$

   Dann ist $\mu=\frac{1}{A}$.

Einsetzen beider Fälle in die Transformationsgleichung liefert die Behauptung.

Jetzt zu $\mathcal{R}_2$. Es gibt dann ein $\mu \in \boldsymbol{R} \setminus \{0\}$ mit

$$X=\mu X \quad Y=\mu Y \quad Z=\mu(AW+BZ) \quad W=\mu Z.$$

1. Fall: $X \neq 0 \vee Y \neq 0$

   Dann ist $\mu=1$ und $W=Z$. Daraus folgt: $(A+B-1)Z=0$. Ohne Einschränkung von A und B erhält man also $Z=W=0$.

2. Fall: $X=0 \wedge Y=0$

   Dann gilt $Z \neq 0$ und $W \neq 0$ und $\mu=\frac{W}{Z}$. Daraus folgt $Z^2-BWZ-AW^2=0$ und somit die Bedingung $A>0 \vee B \neq 0$ von der Existenz reeller Lösungen der quadratischen Gleichung.

Einsetzen des 1. Falles liefert die Behauptung.

Ad 2. Seien $(l_{41},l_{42},l_{43},l_{23},l_{31},l_{12})$ die Plücker-Koordinaten einer Fixgerade. Betrachten wir zunächst $\mathcal{R}_1$. Es gibt ein $\mu \in \boldsymbol{R} \setminus \{0\}$ mit

$$\begin{array}{c} l_{41}=\mu(Al_{41}+Bl_{31}) \quad l_{42}=\mu(Al_{42}+Bl_{23}) \quad l_{43}=\mu Al_{43} \\ l_{23}=\mu l_{23} \quad l_{31}=\mu l_{31} \quad l_{12}=\mu l_{12} \end{array} .$$

Aus den Gleichungen für l_{41} und l_{42} folgt, daß diese Gerade nicht im Unendlichen liegt. Seien $(X_1,Y_1,Z_1,1)$, und $(X_2,Y_2,Z_2,1)$ die Koordinaten ihrer erzeugenden Punkte.

1. Fall: $l_{23} \neq 0 \vee l_{31} \neq 0 \vee l_{12} \neq 0$

   Dann ist $\mu=1$ und $l_{43}=0$, also $Z=Z_1=Z_2$. Aus den Gleichungen für l_{41} und l_{42} folgt damit:

$$\begin{aligned} (A-1+BZ)(X_2-X_1) &= 0 \\ (A-1+BZ)(Y_2-Y_1) &= 0 \end{aligned}$$

   Wegen $Z_1=Z_2$ ist $X_1 \neq X_2$ oder $Y_1 \neq Y_2$, also $(A-1)+BZ=0$. Damit liegt die Gerade in der Fixpunktebene.

2. Fall: $l_{23} = 0 \;\wedge\; l_{31} = 0 \;\wedge\; l_{12} = 0$

   Dann ist $\mu = \frac{1}{A}$ und $Y_1Z_2 = Z_1Y_2$, $Z_1X_2 = X_1Z_2$, $X_1Y_2 = Y_1X_2$, wodurch genau die Geraden durch den Koordinatenursprung festgelegt sind.

Die Probe durch Einsetzen der gewonnenen Werte zeigt wiederum die Behauptung.

Für $\mathcal{R}_2$ gibt es ein $\mu \in \boldsymbol{R} \setminus \{0\}$ mit

$$\begin{array}{c} l_{41} = \mu l_{31} \quad l_{42} = \mu(-l_{23}) \quad l_{43} = \mu(-Al_{43}) \\ l_{23} = \mu(-Al_{42} + Bl_{23}) \quad l_{31} = \mu(Al_{41} + Bl_{31}) \quad l_{12} = \mu l_{12} \quad . \end{array}$$

Wäre die Gerade im Unendlichen, so bekäme man aus den Gleichungen $l_{23} = 0$ und $l_{31} = 0$, wodurch keine Gerade definiert wäre. Seien daher wieder $(X_1, Y_1, Z_1, 1)$ und $(X_2, Y_2, Z_2, 1)$ die Koordinaten ihrer erzeugenden Punkte.

1. Fall: $l_{43} = 0$

   Also ist $Z = Z_1 = Z_2$. Aus den Gleichungen für l_{41}, l_{42}, l_{31} und l_{23} folgt:

$$\begin{array}{rcl} (\mu Z - 1)(X_2 - X_1) & = & 0 \\ (\mu Z - 1)(Y_2 - Y_1) & = & 0 \\ (Z - \mu(A + BZ))(X_2 - X_1) & = & 0 \\ (Z - \mu(A + BZ))(Y_2 - Y_1) & = & 0 \end{array}$$

   Wegen $Z_1 = Z_2$ ist $X_1 \neq X_2$ oder $Y_1 \neq Y_2$. Somit ist $\mu = \frac{1}{Z} = \frac{Z}{A+BZ}$. Die Existenz von reellen Lösungen der quadratischen Gleichung $Z^2 - BZ - A = 0$ stellt wie im Fixpunktfall Bedingungen an A und B.

2. Fall: $l_{43} \neq 0$

   Dann ist $\mu = -\frac{1}{A}$. Damit folgt, falls man nicht besondere Anforderungen an A und B stellt, daß alle übrigen Koordinaten verschwinden, wodurch keine Gerade mehr definiert ist.

□

Daß die Sehstrahlen unter den Transformationen invariant bleiben, die anstelle von Nähe mit skalierter relativer Nähe rekonstruieren, ist keine Überraschung. Nicht so selbstverständlich ist hingegen die Existenz einer Ebene aus Fixpunkten. Man weiß somit, daß sich rekonstruierte Gerade, beispielsweise über zwei Punkte definiert, und Originalgerade in dieser Ebene schneiden, die durch die beiden Parameter der Relieftransformation bestimmt ist. Diese Tatsache sollte man bei einer Ermittlung der Reliefparameter als Kriterium für ihre Exaktheit hinzuziehen.

Die Auffassung des Raumes als orientiert projektiv ermöglicht die Gewinnung zusätzlicher Eigenschaften.

Bemerkung 4.6 $\mathcal{R}_1$, $\mathcal{R}_2$ und $\mathcal{R}_3$ sind orientierte projektive Transformationen. Dabei gilt:

$$\left.\begin{array}{l} \mathcal{R}_1 \\ \mathcal{R}_2 \\ \mathcal{R}_3 \end{array}\right\} \textbf{ ist Orientierungs-erhaltend genau dann, wenn } \left\{\begin{array}{l} A > 0 \\ A < 0 \\ \mathrm{sign}(A') = \mathrm{sign}(A) \end{array}\right. .$$

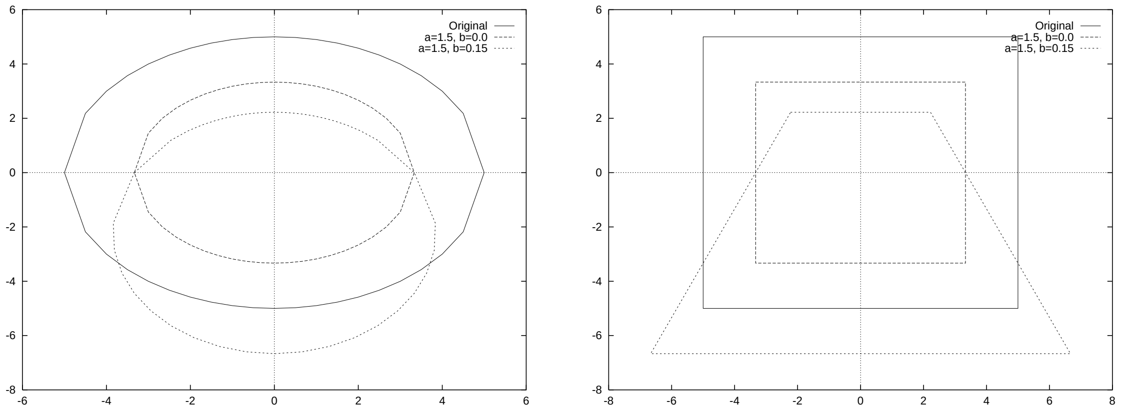

Abbildung 4.3: Relieftransformationen konvexer Flächen. Die Diagramme zeigen horizontale Schnitte einer Gruppe von Flächen, die sich über Relieftransformation in Relation befinden. Man beachte, daß die Konvexität der Flächen und die Lage von Punkten bezüglich orientierter Geraden in der Ebene erhalten bleibt.

Ansonsten sind die Abbildungen Orientierungs-umkehrend. **Insbesondere erhalten sie die Konvexität und die Halbraumeigenschaft aus Bemerkung 2.1.**

Beweis: Es gilt: $\det(\mathcal{R}_1) = A$, $\det(\mathcal{R}_2) = -A$, $\det(\mathcal{R}_3) = a = \frac{A}{A'}$. $\square$

Aus der Erhaltung der Tiefenordnung folgt bei der Rekonstruktion aus Sehstrahlen die Erhaltung der Orientierung. Somit ist Bemerkung 4.6 konsistent mit Bemerkung 4.4. Abbildung 4.3 zeigt konvexe Flächen, die über eine Relieftransformation in Beziehung stehen, wobei ihre Konvexität nicht verletzt wird. Abschließend sei bemerkt, daß man durch die σ-Z- und σ-λ'-Rekonstruktion bei den betrachteten Punkten immer in dieser Welt bleibt, die σ-W-Rekonstruktion aber auch Punkte auf der anderen Seite der unendlich fernen Ebene liefern kann.

4.2 Affine Nähe

Betrachtet man genauer, welche Informationen zu einer projektiven Rekonstruktion aus einem Sehstrahl notwendig sind, so stellt man fest, daß die Gewinnung von skalierter relativer Nähe eine zu strenge Anforderung ist. Wir definieren die *affine Nähe (affine nearness, AN)* durch

$$\alpha =_{\text{def}} A\lambda + B + Cy + Dx,$$

wobei A, B, C und D Konstanten sind mit $A \neq 0$. [2] Darüber erhält man ebenfalls eine projektive Transformation der Tiefe:

$$\alpha = \frac{A + BZ + CY + DX}{Z} \qquad \frac{1}{\alpha} = \frac{Z}{A + BZ + CY + DX}$$

Daraus ergeben sich ganz analog drei projektive Rekonstruktionen der Szene:

[2] Garding et al. bezeichnen in [GPFM96] die skalierte relative Nähe als affine Nähe.

Die α-W-Rekonstruktion:
Die 3D-Rekonstruktion

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} =_{\text{def}} \begin{pmatrix} x \\ y \\ 1 \\ \alpha \end{pmatrix}$$

entspricht einer projektiven Transformation der Originalszene

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} = \mathcal{A}_1 \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ D & C & B & A \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} I_3 & \mathbf{0} \\ \boldsymbol{B}^T & A \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix}.$$

Die α-Z-Rekonstruktion:
Die 3D-Rekonstruktion

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} =_{\text{def}} \begin{pmatrix} x \\ y \\ \alpha \\ 1 \end{pmatrix}$$

entspricht einer projektiven Transformation der Originalszene

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} = \mathcal{A}_2 \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ D & C & B & A \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} I_2 & 0 \\ 0 & M_{AB} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix}.$$

Die α-λ'-Rekonstruktion:
Die 3D-Rekonstruktion

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} =_{\text{def}} \begin{pmatrix} x \\ y \\ 1 \\ \lambda' \end{pmatrix}$$

entspricht einer projektiven Transformation der Originalszene

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ W' \end{pmatrix} = \mathcal{A}_3 \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ d & c & b & a \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} = \begin{pmatrix} I_3 & \mathbf{0} \\ \boldsymbol{b}^T & a \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix}.$$

Dabei ist bei der α-λ'-Rekonstruktion

$$\begin{aligned} \lambda' &=_{\text{def}} \frac{\alpha - B' - C'x - D'y}{A'} = \frac{A}{A'}\lambda + \frac{B-B'}{A'} + \frac{C-C'}{A'}y + \frac{D-D'}{A'}x \\ &= a\lambda + b + cy + dx = \frac{a + bZ + cY + dX}{Z} \end{aligned}$$

mit $a =_{\text{def}} \frac{A}{A'} \neq 0$, $b =_{\text{def}} \frac{B-B'}{A'}$, $c =_{\text{def}} \frac{C-C'}{A'}$, $d =_{\text{def}} \frac{D-D'}{A'}$. Die Bedingung $\lambda' > 0$ ist hier allerdings nicht prinzipiell zu realisieren. Dies ist nur möglich, falls alle zu betrachtenden Sehstrahlen vorher bekannt sind und die Parameter entsprechend gewählt werden können.

Die sich hieraus ergebenden projektiven Abbildungen haben die allgemeinen Eigenschaften, die in §4.1 beschrieben sind, sind aber im allgemeinen keine Relieftransformationen, erhalten also nicht die Tiefenordnung.

Bemerkung 4.7 Die Bemerkungen 4.3 und 4.6 gelten für $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ anstelle von $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$. Jedoch erhalten $\mathcal{A}_1$, $\mathcal{A}_2$ und $\mathcal{A}_3$ im allgemeinen nicht die Tiefenordnung. $\mathcal{A}_1$ und $\mathcal{A}_3$ lassen Punkte der Ebene $(A-1)W + BZ + CY + DX = 0$ und die Sehstrahlen invariant. $\mathcal{A}_2$ hat im allgemeinen weder Fixpunkte noch Fixgeraden.

Beweis: Die Bemerkung 4.3 ist so allgemein formuliert, daß sie auch für $\mathcal{A}_1$, $\mathcal{A}_2$ und $\mathcal{A}_3$ gilt.

Bemerkung 4.6 gilt, da sich durch die zusätzlichen nicht verschwindenden Koeffizienten der Matrizen nichts an den Determinanten ändert, wovon man sich leicht durch Bildung von Blockmatrizen überzeugt.

Der Beweis von Bemerkung 4.5 läßt sich einfach um die zusätzlichen Koeffizienten der Transformation erweitern. Man beachte, daß die Fixpunktebene dadurch eine etwas andere Gestalt erhält, daß die Sehstrahlen als Fixgeraden nicht von der letzten Zeile der Matrix abhängen und daß jetzt auch Punkte mit $Z = W = 0$ keine Fixpunkte von $\mathcal{A}_2$ sind.

Wichtigster Unterschied: Diese Abbildungen erhalten jedoch im allgemeinen nicht die Tiefenordnung, falls $C \neq 0$ oder $D \neq 0$ (bzw. $c \neq 0$ oder $d \neq 0$) ist. Hängt dann die Tiefe nicht geeignet von den Bildpunktkoordinaten ab, hat man es also nicht mit speziellen Flächen zu tun, so ergibt sich keine Monotonie von α bezüglich λ. Seien z. B. $B = 0$, $C = 0$, $D > 0$ und α_1, α_2, α_3 affine Nähen zu den drei Bildpunkten $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$. Dann gilt:

$$\alpha_1 < \alpha_2 < \alpha_3 \quad \Leftrightarrow \quad A + Dx_1 < A + Dx_2 < A + Dx_3 \quad \Leftrightarrow \quad x_1 < x_2 < x_3$$

Man kann beliebig viele (gewölbte) Oberflächen angeben, für die $x_1 < x_2 < x_3$, aber $\lambda_1 < \lambda_3 < \lambda_2$ gilt. □

Man sieht in dem Beweis, daß projektive Transformationen, deren nicht verschwindenden Koeffizienten in der dritten oder vierten Zeile die X- oder Y-Koordinate betreffen, im allgemeinen die Tiefenordnung nicht erhalten. Insbesondere gilt das auch für allgemeine affine und Euklidische Transformationen. Somit sind Relieftransformationen besonders ausgezeichnet. Man erhält aber auch in dem Fall, daß nur affine und nicht skalierte relative Nähe vorhanden ist, Informationen über die Szene, insbesondere ihre projektive Struktur.

Die Koeffizienten der affinen Nähe bestimmen nach Konvention die Ebene im Unendlichen $AW + BZ + CY + DX = 0$. Sind diese Koeffizienten bekannt, sie können sich zum Beispiel durch eine Kalibrierung ergeben, so kennt man die Ebene im Unendlichen, und man erhält die affine oder die Euklidische Struktur der Szene. Dies läßt sich auch an den Transformationsgleichungen ablesen, da Einsetzen der Werte für die Koeffizienten den zum Beispiel in einem Euklidischen Koordinatensystem beschriebenen Originalpunkt liefert.

Wir haben jetzt eine Reihe von Eigenschaften der Szene aus der SRN und AN abgeleitet. Im nächsten Abschnitt werden endlich SRN-Berechnungsmodelle vorgestellt, später auch Verfahren, die AN liefern. Wir werden sehen, welches Vorzeichen und welche Gestalt die Parameter der Nähen haben. Von besonderem Interesse ist dabei der Parameter A bzw. a, der auch im allgemeineren Fall der Rekonstruktion aus affiner Nähe allein die Orientierungserhaltung festlegt.

4.3 SRN-Berechnungsmodelle

Wie bereits erwähnt, kann SRN direkt aus der Verschiebungsrate gewonnen werden. Auf der Gerade $x = 0$ wird die horizontale Disparität zu

$$\dot{x}|_{x=0} = -v_x\lambda - \omega_y + \omega_z y. \tag{4.3}$$

Bereits so kann man auf einfache Weise eine SRN gewinnen. Wir halten fest: **Falls $\omega_z = 0$, so liefert die horizontale Verschiebungsrate für Bildpunkte auf der horizontalen Achse des Bildkoordinatensystems eine SRN mit $A = -v_x$ und $B = -\omega_y$.** Im Fixierungsfall erhält diese SRN eine anschauliche Interpretation:

$$\sigma = A\lambda + B = -v_x\lambda - \omega_y = v_x(\lambda_0 - \lambda)$$

Ist diese SRN gleich 0, so liegt der betrachete Punkt auf der Ebene $Z = Z_0$. Hat die SRN negatives Vorzeichen, so liegt der Punkt im Halbraum der Kameras, bei positivem Vorzeichen im komplementären Halbraum. Um in anderen Bildbereichen die SRN gewinnen zu können, bedarf es jedoch anderer Berechnungsmodelle.

Auf der Gerade $y = 0$ wird die vertikale Disparität zu

$$\dot{y}|_{y=0} = -v_y\lambda + \omega_x - \omega_z x.$$

Falls $\omega_z = 0$, so liefert die vertikale Verschiebungsrate für Bildpunkte auf der vertikalen Achse des Bildkoordinatensystems eine SRN mit $A = -v_y$ und $B = \omega_x$. Die entsprechende Fixierungsbedingung lautet

$$\dot{y}|_{(x,y)=(0,0)} = -v_y\lambda_0 + \omega_x = 0 \quad \Leftrightarrow \quad \omega_x = \frac{v_y}{Z_0}.$$

Also ergibt sich im Fixierungsfall eine SRN

$$\sigma = A\lambda + B = -v_y\lambda + \omega_x = v_y(\lambda_0 - \lambda),$$

die je nach Vorzeichen von v_y eine Interpretation der Lage des betracheten Szenepunktes ermöglicht. Allerdings ist im allgemeinen bei einer Stereogeometrie $v_y = 0$, wobei dann $\dot{y}$ für $y = 0$ keine und für $y \neq 0$ nur wenig Tiefenabhängigkeit besitzt. Daher wird hier mit der horizontalen Verschiebungsrate gearbeitet. **Man muß darauf achten, daß bei der Anwendung auf den Bewegungsfall der Parameter A, hier und später vor allem v_x, nicht Null sein darf, was im Stereofall mit $v_x \neq 0$ automatisch gewährleistet ist.** Spezialfälle der Bewegung, etwa mit $v_x = 0$ und $v_y \neq 0$, die ein Vorgehen über die vertikale Verschiebungsrate geeignet erscheinen lassen, werden hier nicht behandelt, lassen sich aber ganz analog ableiten.

Die Parameter $A = -v_x$ und $B = -\omega_y$ werden in ähnlicher Form immer wieder auftauchen. Allein durch Kenntnis des Vorzeichens von v_x, das sich im Stereofall allein aus der Beschreibung ergibt, aber auch im Bewegungsfall eine geringe Anforderung darstellt, kann man auf die Erhaltung oder Umkehrung der Orientierung der Transformationen schließen. Im Falle der Beschreibung der Bewegung von links nach rechts ist $A < 0$.

4.3.1 Das PAD-Modell

In [Wei90] und [LSS94] werden Informationen aus der Verschiebungsrate nicht mit kartesischen, sondern mit polaren Bildkoordinaten gewonnen. Die Autoren brachten die Vorteile, die die Darstellung in diesen Koordinaten bringt, in Zusammenhang mit dem hier vorliegenden Problem. Dies führt zum *PAD-Modell* (polar angle diparity, Polarwinkeldisparität).

Wir führen *ebene Polarkoordinaten* in der üblichen Weise ein durch

$$\begin{aligned} x &= \rho\cos(\theta), && (4.4)\\ y &= \rho\sin(\theta). && (4.5)\end{aligned}$$

Umgekehrt gilt dann

$$\begin{aligned} \theta &= \tan^{-1}(\frac{y}{x}),\\ \rho &= \sqrt{x^2+y^2}.\end{aligned}$$

Die *polare Verschiebungsrate* $(\dot\rho, \dot\theta)$ steht dann mit der kartesischen Verschiebungsrate wie folgt in Beziehung [Dan95]:

$$\begin{aligned} \dot\rho &= \frac{1}{2\sqrt{x^2+y^2}}(2\dot x x + 2\dot y y) = \frac{1}{\rho}(\dot x\cos(\theta) + \dot y\sin(\theta))\\ &= \dot x\cos(\theta) + \dot y\sin(\theta)\\ \dot\theta &= \cos^2(\theta)\frac{\dot y x - y\dot x}{x^2} = \frac{1}{\rho^2}(\dot y\rho\cos(\theta) - \dot x\rho\sin(\theta))\\ &= \frac{1}{\rho}(\dot y\cos(\theta) - \dot x\sin(\theta)) && (4.6)\end{aligned}$$

Einsetzen von (3.8, 3.9) liefert die Beziehung zu den Bewegungsparametern.

$$\begin{aligned} \dot\rho &= (\frac{-v_x+v_z x}{Z} + \omega_x xy - \omega_y(x^2+1) + \omega_z y)\cos(\theta)\\ &\quad +(\frac{-v_y+v_z y}{Z} + \omega_x(y^2+1) - \omega_y xy - \omega_z x)\sin(\theta)\\ &\overset{(4.4,4.5)}{=} -\frac{v_x}{Z}\cos(\theta) - \frac{v_y}{Z}\sin(\theta) + \frac{v_z}{Z}\rho + \omega_x\sin(\theta)(\rho^2+1) - \omega_y\cos(\theta)(\rho^2+1)\\ \dot\theta &= \frac{1}{\rho}[(\frac{-v_y+v_z y}{Z} + \omega_x(y^2+1) - \omega_y xy - \omega_z x)\cos(\theta)\\ &\quad -(\frac{-v_x+v_z x}{Z} + \omega_x xy - \omega_y(x^2+1) + \omega_z y)\sin(\theta)]\\ &\overset{(4.4,4.5)}{=} \frac{v_x}{Z}\frac{\sin(\theta)}{\rho} - \frac{v_y}{Z}\frac{\cos(\theta)}{\rho} + \omega_x\frac{\cos(\theta)}{\rho} + \omega_y\frac{\sin(\theta)}{\rho} - \omega_z && (4.7)\end{aligned}$$

Man sieht, daß die ρ-Verschiebungsrate invariant gegenüber Torsion, also Rotation um die Z-Achse, die θ-Verschiebungsrate invariant gegenüber Translation längs der optischen Achse, also der Z-Achse, ist. Letzteres macht man sich für unser Problem zunutze. Mit $v_y = 0$, $\omega_x = 0$, $\omega_z = 0$ gilt nämlich

$$\begin{aligned} \dot\rho &= (-v_x\cos(\theta) + v_z\rho)\lambda - \omega_y\cos(\theta)(\rho^2+1),\\ \dot\theta &= v_x\frac{\sin(\theta)}{\rho}\lambda + \omega_y\frac{\sin(\theta)}{\rho}.\end{aligned}$$

Man erhält also eine SRN σ für einen Bildpunkt (ρ, θ) direkt aus der θ-Verschiebungsrate mittels

$$\sigma = -\frac{\rho}{\sin(\theta)}\dot{\theta} = -v_x\lambda - \omega_y. \tag{4.8}$$

Für kartesische Bildkoordinaten (x, y) benötigt man x- und y-Disparität zur SRN-Berechnung gemäß

$$\begin{aligned}\sigma &= -\frac{\rho}{\sin\theta}\dot{\theta} \overset{(4.6)}{=} -\frac{\rho}{\sin\theta}[\frac{1}{\rho}(\dot{y}\cos\theta - \dot{x}\sin\theta)] = \dot{x} - \frac{\cos\theta}{\sin\theta}\dot{y} \\ &\overset{(4.4,4.5)}{=} \dot{x} - \frac{x}{y}\dot{y},\end{aligned}$$

wie man vielleicht auch schon an (4.1, 4.2) erkannt hat. Durch Betrachtung anderer Verschiebungsraten bekommt man also SRN auch in anderen Bildbereichen, leider nicht für $\sin\theta = 0$ bzw. $y = 0$. Ergebnis ist also: **Falls $v_y = 0$, $\omega_x = 0$, $\omega_z = 0$, so liefert die θ-Verschiebungsrate für Bildpunkte (ρ, θ) mit $\sin\theta \neq 0$ in ebenen Polarkoordinaten eine SRN mit $A = -v_x$ und $B = -\omega_y$ mittels $-\frac{\rho}{\sin(\theta)}\dot{\theta}$. Für Bildpunkte (x, y) mit $y \neq 0$ in kartesischen Bildkoordinaten liefert die x- und y-Disparität mittels $\dot{x} - \frac{x}{y}\dot{y}$ dieselbe skalierte relative Nähe.** Es ist keine Überraschung, daß sich hier dieselbe SRN wie bei 4.3 ergibt. Die Terme, die in der horizontalen Verschiebungsrate noch von den Bildpunktkoordinaten abhängen, werden hier mit Hilfe der vertikalen Disparität eliminiert, was bei Bildpunkten mit $x = 0$ ebenfalls getan wurde.

4.3.2 Das Deformationsmodell

Ein Ansatz, der zur Informationsgewinnung nicht die Verschiebungsraten, sondern ihre ersten örtlichen Ableitungen betrachtet, ist das *Deformationsmodell* [KvD76]. Auch diese Daten geben die Beziehung zwischen Bildpunktkoordinaten, Tiefe und Systemparametern wieder. Im Bildzentrum bekommen sie jedoch eine besonders einfache Gestalt. So kann mit Hilfe der sogenannten Deformation eine skalierte Nähe für das Bildzentrum, und nur dort, ermittelt werden. Nimmt man aber an, daß die Punkte in der Umgebung des abgebildeten Szenepunktes, i. e. im Falle des beschriebenen Stereosystems der Fixierungspunkt, auf einer Ebene liegen, so erhält man auch für ihre Bildpunkte eine SRN, und zwar dann direkt über ihre Koordinaten.

Die ersten örtlichen Ableitungen der Verschiebungsraten enthalten Informationen zur lokalen Bewegung von Flächenelementen in der Bildebene. Diese äußern sich in der *Divergenz* div, der *Rotation* curl und den *Scherungen* shear_1, shear_2.

$$\begin{aligned}\operatorname{div}\dot{\boldsymbol{x}} &=_{\mathrm{def}} \frac{\partial\dot{x}}{\partial x} + \frac{\partial\dot{y}}{\partial y} \\ \operatorname{curl}\dot{\boldsymbol{x}} &=_{\mathrm{def}} -\frac{\partial\dot{x}}{\partial y} + \frac{\partial\dot{y}}{\partial x} \\ \mathrm{shear}_1\,\dot{\boldsymbol{x}} &=_{\mathrm{def}} \frac{\partial\dot{x}}{\partial x} - \frac{\partial\dot{y}}{\partial y} \\ \mathrm{shear}_2\,\dot{\boldsymbol{x}} &=_{\mathrm{def}} \frac{\partial\dot{x}}{\partial y} + \frac{\partial\dot{y}}{\partial x}\end{aligned}$$

Die beiden Scherungskomponenten werden auch in einem zweidimensionalen Vektor, der *Deformation* **def**, zusammengefaßt:

$$\mathbf{def}\ \dot{\boldsymbol{x}} =_{\text{def}} (\text{shear}_1\ \dot{\boldsymbol{x}}, \text{shear}_2\ \dot{\boldsymbol{x}}) = (\frac{\partial \dot{\mathrm{x}}}{\partial \mathrm{x}} - \frac{\partial \dot{\mathrm{y}}}{\partial \mathrm{y}}, \frac{\partial \dot{\mathrm{x}}}{\partial \mathrm{y}} + \frac{\partial \dot{\mathrm{y}}}{\partial \mathrm{x}})$$

Der Vorteil dieser Beschreibungen liegt in ihrer Interpretation. Die Divergenz spezifiziert die lokale isotrope Expansion oder Kontraktion, die Rotation die lokale Rotation. Die erste Scherung beschreibt eine Expansion in x- und Kontraktion in y-Richtung, die zweite Scherung eine Expansion in y- und Kontraktion in x-Richtung, wobei der Flächeninhalt jeweils gleich bleibt. Entsprechend bestimmt der Deformationsvektor die Expansion längs einer Achse und die Kontraktion senkrecht zu dieser Richtung bei Erhaltung des Flächeninhalts. Es wird sich zeigen, daß die Scherungen im Bildzentrum nur von der lokalen Orientierung der 3D-Oberfläche und der um die Nähe skalierten Translation in x-Richtung abhängen.

Einsetzen der Verschiebungsraten (3.8, 3.9) liefert für die Scherungen

$$\begin{aligned}
\text{shear}_1\ \dot{\boldsymbol{x}} &= (\frac{v_z Z - (-v_x + v_z x)\frac{\partial Z}{\partial x}}{Z^2} + \omega_x y - 2\omega_y x) \\
&\quad -(\frac{v_z Z - (-v_y + v_z y)\frac{\partial Z}{\partial y}}{Z^2} + 2\omega_x y - \omega_y x) \\
&= \frac{p(v_x - v_z x) - q(v_y - v_z y)}{Z(1 - px - qy)} - \omega_x y - \omega_y x, \\
\text{shear}_2\ \dot{\boldsymbol{x}} &= (\frac{-(-v_x + v_z x)\frac{\partial Z}{\partial y}}{Z^2} + \omega_x x + \omega_z) \\
&\quad +(\frac{-(-v_y + v_z y)\frac{\partial Z}{\partial x}}{Z^2} - \omega_y y - \omega_z) \\
&= \frac{q(v_x - v_z x) + p(v_y - v_z y)}{Z(1 - px - qy)} + \omega_x x - \omega_y y.
\end{aligned}$$

Hierbei ist $(p, q, -1) = (\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y}, -1)$ eine Flächennormale im betrachteten Punkt, und es gilt aufgrund der Kettenregel die Beziehung:

$$\frac{\partial Z}{\partial x} = \frac{\partial Z}{\partial X}\frac{\partial X}{\partial x} + \frac{\partial Z}{\partial Y}\frac{\partial Y}{\partial x} = \frac{\partial Z}{\partial X}(Z + x\frac{\partial Z}{\partial x}) + \frac{\partial Z}{\partial Y} y \frac{\partial Z}{\partial x},$$

also

$$\frac{\partial Z}{\partial x}(1 - px - qy) = pZ, \tag{4.9}$$

und analog

$$\frac{\partial Z}{\partial y}(1 - px - qy) = qZ. \tag{4.10}$$

Man liest als erste wichtige Eigenschaft der Scherungen ab, daß sie ganz allgemein torsionsunabhängig sind. Im Bildzentrum bekommen sie die erwähnte einfache Form:

$$\begin{aligned}
\text{shear}_1\ \dot{\boldsymbol{x}}(0,0) &= \frac{p_0 v_x - q_0 v_y}{Z_0} \\
\text{shear}_2\ \dot{\boldsymbol{x}}(0,0) &= \frac{q_0 v_x + p_0 v_y}{Z_0}
\end{aligned}$$

Dabei ist $(p_0, q_0, -1)$ eine Flächennormale im Fixierungspunkt. Dies ist bereits eine skalierte Nähe. Man benutzt aber noch $v_y = 0$, um über die Scherungen respektive die Deformation eine skalierte Nähe zu erhalten, die den direkten Bezug zur Flächennormalen liefert:

$$\begin{aligned} \text{shear}_1\, \dot{\boldsymbol{x}}(0,0) &= \frac{v_x}{Z_0} p_0 \\ \text{shear}_2\, \dot{\boldsymbol{x}}(0,0) &= \frac{v_x}{Z_0} q_0 \\ \mathbf{def}\, \dot{\boldsymbol{x}}(0,0) &= \frac{v_x}{Z_0}(p_0, q_0) \end{aligned}$$

Es ist klar, daß die Flächennormale im Fixierungspunkt etwas über die Lage der Punkte aussagt, falls man lokal von einer Ebene als 3D-Fläche ausgeht. Ohne Kenntnis des unbekannten Skalierungsfaktors $\frac{v_x}{Z_0}$ kann die Flächennormale aber nicht bestimmt werden. Dennoch enthalten auch die skalierten Werte von p_0 und q_0, wie sie sich aus der Deformation ergeben, Informationen über die Lage der Punkte. Die Deformation steht sogar mit der skalierten relativen Nähe $\sigma = v_x(\lambda_0 - \lambda)$, die die vorigen Modelle im Fixierungsfall bestimmt haben, über

$$\nabla\sigma(0,0) = \mathbf{def}\, \dot{\boldsymbol{x}}(0,0)$$

in Beziehung wegen

$$\begin{aligned} \nabla\sigma(0,0) &= \left(\frac{\partial\sigma}{\partial x}(0,0), \frac{\partial\sigma}{\partial y}(0,0)\right) = v_x\left(-\frac{\partial\lambda}{\partial x}, -\frac{\partial\lambda}{\partial y}\right) \\ &\overset{(4.9,4.10)}{=} v_x\left(\frac{p_0 Z_0}{Z_0^2}, \frac{q_0 Z_0}{Z_0^2}\right) = \frac{v_x}{Z_0}(p_0, q_0) = \mathbf{def}\, \dot{\boldsymbol{x}}(0,0). \end{aligned}$$

Um über die skalierte Nähe $\nabla\sigma$ im Bildzentrum Aussagen für benachbarte Bildpunkte treffen zu könnenn, nimmt man an, daß in einer kleinen Umgebung um den Fixierungspunkt sich die abgebildete Oberfläche durch eine Ebene

$$Z = p_0 X + q_0 Y + Z_0 \tag{4.11}$$

beschreiben läßt. Für Bildpunkte gilt dann

$$p_0 x + q_0 y = 1 - \frac{Z_0}{Z}.$$

Multipliziert man diese Gleichung mit $\frac{v_x}{Z_0}$, so ergibt sich

$$\left(\frac{v_x}{Z_0} p_0\right) x + \left(\frac{v_x}{Z_0} q_0\right) y = v_x(\lambda_0 - \lambda).$$

Also erhält man die skalierte relative Nähe σ aus der Deformation durch

$$\begin{aligned} \sigma &= \text{shear}_1\, \dot{\boldsymbol{x}}(0,0)\text{x} + \text{shear}_2\, \dot{\boldsymbol{x}}(0,0)\text{y} \\ &= \mathbf{def}\, \dot{\boldsymbol{x}}(0,0)^{\mathrm{T}} \boldsymbol{x}. \end{aligned}$$

Man weiß darüber hinaus, daß sich für die Bewegung einer Ebene allgemein Verschiebungsraten ergeben, die quadratisch in den Bildkoordinaten sind, und somit lineare erste örtliche Ableitungen (siehe §4.3.3). Dies kann zur Berechnung der Deformation aus den Verschiebungsraten genutzt werden. Eine andere Möglichkeit besteht in der Anwendung eines Ableitungsfilters auf das Feld der Verschiebungsraten. Dies setzt allerdings die Bestimmung eines dichten Feldes voraus. Insgesamt ergibt sich so für das Deformationsmodell: **Falls $v_y = 0$ und falls die betrachteten Szenepunkte auf einer Ebene mit dem Fixierungspunkt liegen, so liefert die Deformation eine skalierte relative Nähe mit den Konstanten $A = -v_x$ und $B = v_x\lambda_0$ für die entsprechenden Bildpunkte $\boldsymbol{x}$ um das Bildzentrum mittels def $\dot{\boldsymbol{x}}(0,0)^{\mathrm{T}}\boldsymbol{x}$.**

Für eine kleine Umgebung um das Bildzentrum ist diese Annahme in den meisten Fällen näherungsweise korrekt, allerdings doch eine Einschränkung gegenüber dem Polarwinkeldisparitätsmodell. Ein weiterer kleiner Nachteil besteht darin, daß die Disparitäten nicht direkt verwendet werden können. Man macht sich somit aber auch automatisch nicht abhängig von einzelnen, eventuell fehlerbehafteten Disparitätswerten. Ist die Ebenenbedingung erfüllt, so kann hier die skalierte relative Nähe in allen in Frage kommenden Bildbereichen, also auch für die x-Achse, bestimmt werden.

4.3.3 Das RDC-Modell

Das *RDC-Modell* (regional disparity correction, regionale Disparitätskorrektur) [3] [GPFM96] stellt eine Kombination des PAD-Modells und des Deformationsmodells dar. Wie beim PAD-Modell wird die vertikale Verschiebungsrate genutzt, um die Terme in der horizontalen Verschiebungsrate, die von den Bildpunktkoordinaten abhängen, zu eliminieren. Wie das Deformationsmodell nutzt RDC dazu die vereinfachte Darstellung für Punkte auf einer Ebene. Die SRN, die sich dabei ergibt, ist dieselbe wie beim PAD-Modell.

Die Verschiebungsraten einer sich bewegenden Ebene π ergeben sich aus (4.11) durch Einsetzen von

$$\lambda = \lambda_0(1 - p_0 x - q_0 y) \tag{4.12}$$

in (4.1, 4.2):

$$\begin{aligned}
\dot{x}_\pi &= -\lambda_0(1 - p_0 x - q_0 y)v_x + \lambda_0(1 - p_0 x - q_0 y)v_z x + \omega_x xy - \omega_y(x^2+1) + \omega_z y \\
\dot{y}_\pi &= -\lambda_0(1 - p_0 x - q_0 y)v_y + \lambda_0(1 - p_0 x - q_0 y)v_z y + \omega_x(y^2+1) - \omega_y xy - \omega_z x
\end{aligned}$$

Als Funktion der Bildpunktkoordinaten lauten sie

$$\begin{aligned}
\dot{x}_\pi &= (-\lambda_0 v_x - \omega_y) + (\lambda_0 p_0 v_x + \lambda_0 v_z)x + (\lambda_0 q_0 v_x + \omega_z)y \\
&\quad + (-\lambda_0 q_0 v_z + \omega_x)xy + (-\lambda_0 p_0 v_z - \omega_y)x^2 \\
&= A_x + B_x x + C_x y + D_x xy + E_x x^2 \\
\dot{y}_\pi &= (-\lambda_0 v_y + \omega_x) + (\lambda_0 p_0 v_y - \omega_z)x + (\lambda_0 q_0 v_y + \lambda_0 v_z)y \\
&\quad + (-\lambda_0 p_0 v_z - \omega_y)xy + (-\lambda_0 q_0 v_z + \omega_x)y^2 \\
&= A_y + B_y x + C_y y + D_y xy + E_y y^2,
\end{aligned}$$

wobei $D_x = E_y$ und $E_x = D_y$. Sind nun die Koeffizienten der vertikalen Verschiebungsrate bekannt, so definiert man

$$g_\pi =_{\mathrm{def}} -C_y x + B_y y - D_y x^2 - E_y xy$$

[3] Dieses Modell weicht von dem RDC-Modell in [GPMF95] ab.

und gewinnt

$$\begin{aligned}\dot{x}_\pi + g_\pi &= A_x + (B_x - C_y)x + (C_x + B_y)y \\ &= (-\lambda_0 v_x - \omega_y) + \lambda_0(p_0 v_x - q_0 v_y)x + \lambda_0(q_0 v_x + p_0 v_y)y \\ &= (-\lambda_0 v_x - \omega_y) + v_x\lambda_0(p_0 x + q_0 y) + v_y\lambda_0(p_0 v_y - q_0 v_x).\end{aligned}$$

Eine SRN erhält man unter der Bedingung $v_y = 0$ und unter Ausnutzung von $\lambda_0(p_0 x + q_0 y) = \lambda_0 - \lambda$, was aus (4.12) folgt:

$$\sigma = \dot{x}_\pi + g_\pi = -v_x\lambda - \omega_y$$

Die Koeffizienten sind Ergebnis einer Ebenenanpassung und lassen sich durch Lösung eines linearen Gleichungssystems in den Bildpunkten und den Verschiebungsraten ermitteln. Wir halten zunächst fest: **Falls $v_y = 0$ und falls die betrachteten Szenepunkte auf einer Ebene liegen, so liefert eine Ebenenanpassung $g_\pi(\boldsymbol{x})$ und für die entsprechenden Bildpunkte $\boldsymbol{x}$ eine SRN mit $A = -v_x$ und $B = -\omega_y$ mittels $\dot{x}_\pi + g_\pi(\boldsymbol{x})$.**

Man hofft jedoch auch im Fall, daß es sich nicht um eine Ebene handelt, zumindest eine gute Approximation der SRN erhalten kann. Man gewinnt analog

$$\hat{\sigma} = \dot{x}_\pi + g_\pi \approx -v_x\lambda - \omega_y.$$

g_π ergibt sich durch eine Ebenenanpassung nur an die vertikale Verschiebungsrate. Im Stereofall hängt $\dot{y}$ kaum von der Tiefe ab, da sie durch das kleine v_z skaliert wird. Deshalb ist hier der Approximationsfehler gering. Vernachlässigt man diesen Fehler, nimmt man also an, daß g_π exakt vorliegt, so läßt sich der Fehler ϵ in der SRN bestimmen:

$$\begin{aligned}\hat{\sigma} &= \dot{x} + g_\pi = (\dot{x} - \dot{x}_\pi) + (\dot{x}_\pi + g_\pi) = (\lambda - \lambda_\pi)(-v_x + v_z x) + \sigma \\ &= \sigma + v_z(\lambda - \lambda_\pi)x = \sigma + \epsilon\end{aligned}$$

Im Stereofall verschwindet v_z sogar bei symmetrischer Fixierung. Deshalb ist hier das RDC-Modell besonders geeignet.

Bei diesem Modell müssen ebenfalls Verschiebungsraten in mehreren Punkten herangezogen werden. Die Bestimmung der Koeffizienten der vertikalen Verschiebungsrate stellt eine Kalibrierung dar, da dadurch die Parameter der relativen Bewegung gewonnen werden. Will man sich von der Ebenenbedingung lösen, so legt das Fehlerverhalten die Anwendung im SfS nahe. Dabei rufe man sich aber in Erinnerung, daß, für $v_y = 0$, ω_x und ω_z nahe bei Null sein müssen, um eine gute Approximation der Disparität durch die Verschiebungsrate zu gewährleisten. Bezüglich das PAD-Modells kann man sagen, daß eine fehlerbehaftete Bestimmung der SRN eine Anwendung im gesamten Bildbereich ermöglicht.

4.4 Varianten des PAD-Modells

Wir kommen jetzt zu drei Varianten des in §4.3.1 vorgestellten Modells. Zum ersten handelt es sich um eine diskrete Version, die ohne zyklopisches Koordinatensystem auskommt. Zum zweiten wird die Möglichkeit der Approximation durch die horizontale Verschiebungsrate untersucht. Zum dritten stellen wir eine Verallgemeinerung des Modells vor.

4.4.1 Diskrete Versionen

Es stellt sich bei einer binokularen Geometrie die Frage, ob die Einführung eines zyklopischen Koordinatensystems und die Approximation der Disparität durch die Verschiebungsrate nicht umgangen werden können. Wir gehen dabei von einer einfachen Stereogeometrie wie in §3 aus und betrachten das diskrete Analogon $d_x - \frac{x_r}{y_r} d_y$.

$$\begin{aligned} \frac{x_r}{y_r} d_y &= \frac{X_r}{Y_r}(y_r - y_l) \overset{Y_r = Y_l}{=} X_r\left(\frac{1}{Z_r} - \frac{1}{Z_l}\right) = x_r - \frac{X_r}{Z_l} \\ d_x - \frac{x_r}{y_r} d_y &= (x_r - x_l) - \left(x_r - \frac{X_r}{Z_l}\right) = \frac{X_r - X_l}{Z_l} \end{aligned}$$

Die Beziehung vom rechten zum linken Koordinatensystem liefert $X_r = \cos(2\mu)X_l - \sin(2\mu)Z_l + v_x^r$, und man erhält so

$$\begin{aligned} \alpha_l &= d_x - \frac{x_r}{y_r} d_y = \frac{1}{Z_l}(X_r - X_l) = \frac{1}{Z_l}[v_x^r - \sin(2\mu)Z_l + (\cos(2\mu) - 1)X_l] \\ &= v_x^r \lambda_l - \sin(2\mu) + (\cos(2\mu) - 1)x_l, \end{aligned}$$

also eine affine Nähe. **Falls $v_y = 0$, $\omega_x = 0$, $\omega_z = 0$, so liefert die Disparität für korrespondierende Bildpunkte $(\boldsymbol{x}_l, \boldsymbol{x}_r)$ mit $y_r \neq 0$ eine AN bezüglich des linken Koordinatensystems mit $A = v_x^r$, $B = -\sin(2\mu)$ und $D = \cos(2\mu) - 1$ mittels $d_x - \frac{x_r}{y_r} d_y$.**

Das analoge Vorgehen für den linken Bildpunkt liefert die affine Nähe

$$\begin{aligned} \alpha_r &= d_x - \frac{x_l}{y_l} d_y = (x_r - x_l) - \frac{X_l}{Y_l}(y_r - y_l) = (x_r - x_l) - \left(\frac{X_l}{Z_r} - x_l\right) = \frac{X_r - X_l}{Z_r} \\ &= \frac{1}{Z_r}[(1 - \cos(2\mu))X_r - \sin(2\mu)Z_r - v_x^l] \\ &= -v_x^l \lambda_r - \sin(2\mu) + (1 - \cos(2\mu))x_r. \end{aligned}$$

Falls $v_y = 0$, $\omega_x = 0$, $\omega_z = 0$, so liefert die Disparität für korrespondierende Bidpunkte $(\boldsymbol{x}_l, \boldsymbol{x}_r)$ mit $y_l \neq 0$ eine AN bezüglich des rechten Koordinatensystems mit $A = v_x^l$, $B = -\sin(2\mu)$ und $D = 1 - \cos(2\mu)$ mittels $d_x - \frac{x_l}{y_l} d_y$.

Diese Informationen haben den Vorteil, daß sie völlig fehlerfrei berechnet werden können, da sie keine Approximation der zyklopischen Bildpunktkoordinaten enthalten. Darüber hinaus besteht eine gewisse Wahlfreiheit bezüglich des linken und rechten Koordinatensystems, was besonders die Singularität $y = 0$ betrifft. Allerdings gewinnt man hier nur AN und somit keine Information über die Tiefenordnung. Man erhält eine SRN, falls $\cos(2\mu) = 1$ zumindest näherungsweise gilt, was bei geringer Rotation der Fall ist.

4.4.2 Die horizontale Verschiebungsrate als Approximation

Bereits zu Beginn der Betrachtungen über die SRN-Gewinnung wurde gesehen, daß die horizontale Verschiebungsrate im Falle $\omega_z = 0$ auf der Geraden $x = 0$ eine SRN liefert. Es gilt sogar, ohne jegliche Einschränkung der Bewegungsparameter: **Die horizontale Verschiebungsrate liefert für Bildpunkte auf der horizontalen Achse des Bildkoordinatensystems eine AN mit $A = -v_x$, $B = -\omega_y$ und $C = \omega_z$.**

Taugt $\dot{x}$ auch sonst zur SRN- oder AN-Gewinnung? Unter den Bedingungen $v_z = 0$, $\omega_x = 0$ ist

$$\dot{x} = -v_x\lambda - \omega_y + \omega_z y - \omega_y x^2.$$

Man gewinnt also eine AN und im Falle $\omega_z = 0$ eine SRN bis auf einen Fehlerterm ϵ, der aber quadratisch in x, also relativ klein ist:

$$\begin{aligned}
\hat{\alpha} &= \dot{x} = -v_x\lambda - \omega_y + \omega_z y - \omega_y x^2 = \alpha + \epsilon \\
\hat{\sigma} &= \dot{x} = -v_x\lambda - \omega_y - \omega_y x^2 = \sigma + \epsilon
\end{aligned}$$

mit $\alpha = -v_x\lambda - \omega_y + \omega_z y$, $\sigma = -v_x\lambda - \omega_y$, $\epsilon = -\omega_y x^2$. Also: **Falls $v_z = 0$, $\omega_x = 0$, liefert die horizontale Verschiebungsrate eine AN mit $A = -v_x$, $B = -\omega_y$ und $C = \omega_z$, und im Falle von $\omega_z = 0$ eine entsprechende SRN mit additivem Fehler $-\omega_y x^2$.**

Ähnlich wie bei der Verwendung des Ebenenmodells erkauft man sich hier die uneingeschränkte Information im gesamten Bild durch einen inhärenten Fehler. Die Einschränkung $v_z = 0$ schmerzt dabei weniger im Stereofall (symmetrische Fixierung) als im Bewegungsfall, da die meisten Bewegungen in der Praxis eine „Vorwärts"- oder „Rückwärts"-Komponente besitzen. Auf die Vorzüge, die die reine Betrachtung der horizontalen Verschiebungsrate trotzdem hat, werden wir im anschließenden Abschnitt eingehen.

4.4.3 Eine Erweiterung des PAD-Modells

Zum Schluß wollen wir genauer untersuchen, welche Anforderungen an das PAD-Verfahren zu stellen sind. Die Voraussetzung einer einfachen Stereogeometrie mit $v_y = 0$, $\omega_x = 0$ und $\omega_z = 0$ scheint nicht genau den Punkt zu treffen. Wir betrachten noch einmal die θ-Verschiebungsrate (4.7):

$$\rho^2\dot{\theta} = \frac{v_x}{Z}y - \frac{v_y}{Z}x + \omega_x x + \omega_y y - \omega_z\rho^2$$

Um den quadratischen Term zu eliminieren, muß $\omega_z = 0$ gefordert werden. Daneben verlangen wir aber nur das Senkrechtstehen von Translationsrichtung und Rotationsachse $\boldsymbol{v}^T\boldsymbol{\omega} = \mathbf{0}$. Mit $v_y = -v_x\frac{\omega_x}{\omega_y}$ bekommt man dann:

$$\begin{aligned}
\rho^2\dot{\theta} &= \frac{v_x}{Z}y + \frac{v_x}{Z}\frac{\omega_x}{\omega_y}x + \omega_x x + \omega_y y \\
&= (\omega_x x + \omega_y y)(\frac{v_x}{\omega_y}\lambda + 1)
\end{aligned}$$

Nicht exakte Kenntnis von ω_x und ω_y, sondern nur Kenntnis der Richtung der Rotationsachse $\frac{1}{\|\boldsymbol{\omega}\|}\boldsymbol{\omega} = (\frac{\omega_x}{\sqrt{\omega_x^2+\omega_y^2}}, \frac{\omega_y}{\sqrt{\omega_x^2+\omega_y^2}}, 0)^T$ ist notwendig, um daraus eine SRN zu erhalten.

$$\begin{aligned}
\sigma &= \rho^2\dot{\theta}/(\frac{\omega_x}{\|\boldsymbol{\omega}\|}\rho\cos(\theta) + \frac{\omega_y}{\|\boldsymbol{\omega}\|}\rho\sin(\theta)) \stackrel{(4.6)}{=} (x\dot{y} - y\dot{x})/(\frac{\omega_x}{\|\boldsymbol{\omega}\|}x + \frac{\omega_y}{\|\boldsymbol{\omega}\|}y) \\
&= \|\boldsymbol{\omega}\|(\frac{v_x}{\omega_y}\lambda + 1) = \|\boldsymbol{\omega}\|\frac{v_x}{\omega_y}\lambda + \|\boldsymbol{\omega}\|
\end{aligned}$$

Wir bekommen also: **Falls $\omega_z = 0$, $\boldsymbol{v}^T\boldsymbol{\omega} = \mathbf{0}$ und die Richtung der Rotationsachse bekannt ist, so liefert die θ-Verschiebungsrate für Bildpunkte (ρ, θ) bzw. (x, y)**

mit $\omega_x x+\omega_y y \neq 0$ **eine SRN mit** $A = \|\boldsymbol{\omega}\| \frac{v_x}{\omega_y}$ **und** $B = \|\boldsymbol{\omega}\|$ **mittels** $\rho^2\dot{\theta}/(\frac{\omega_x}{\|\boldsymbol{\omega}\|}\rho\cos(\theta)+\frac{\omega_y}{\|\boldsymbol{\omega}\|}\rho\sin(\theta)) = (x\dot{y} - y\dot{x})/(\frac{\omega_x}{\|\boldsymbol{\omega}\|}x + \frac{\omega_y}{\|\boldsymbol{\omega}\|}y)$.

Dies ist eine echte Erweiterung des bisherigen PAD-Verfahrens, da aus $v_y = 0$, $\omega_x = 0$ und $\omega_z = 0$ insbesondere $\boldsymbol{v}^T\boldsymbol{\omega} = \mathbf{0}$ folgt und die Rotationsachse $(0,1,0)^T$ bekannt ist. Rotations- und Translationsachse stehen senkrecht aufeinander, falls für $\omega_z = 0$ die Fixierungsbedingungen $\omega_y = -\frac{v_x}{Z_0}$ und $\omega_x = \frac{v_y}{Z_0}$ erfüllt sind, was auch bei schneidenden optischen Achsen gilt. Allerdings ist in der Praxis oft die Ermittlung der Richtung der Rotationsachse nicht leichter zu realisieren als die zusätzliche Bedingung $\omega_x = 0$, die im Stereofall ohnehin angebracht ist. Man behalte diese Erweiterung aber für spezielle Anwendungen und als Grundlage des PAD-Modells im Auge.

4.5 Verfahren bei unbekannten internen Kameraparametern

Alle bisherigen Berechnungsmodelle gingen von einem Sehstrahl $\boldsymbol{x}$ aus, also den normalisierten Bildkoordinaten, um die Szene zu rekonstruieren. Dazu wurde die Kenntnis aller internen Kameraparameter und ihre Gleichheit in allen Perspektiven vorausgesetzt. Diese müssen jedoch Ergebnis einer sorgfältigen Kalibrierung sein, und die Forderung ihrer Gleichheit ist für manche Verfahren vielleicht eine unnötige Einschränkung. So kann man mit dem gleichen Argument, mit dem man auf die explizite Ermittlung der externen, also der relativen Bewegungsparameter verzichten möchte, versuchen, möglichst viel Wissen über die internen Parameter entbehrlich zu machen. Ziel ist ein *total unkalibriertes* Vorgehen, bei dem man keine oder wenigstens beinahe keine Kalibrierungsinformation zur Gewinnung von Strukturinformation heranzieht. Vorteile liegen in einer leichteren, schnelleren, fehlerfreieren Vorgehensweise, da der Kalibrierungsschritt entfällt, und in der Anwendbarkeit in Situationen, wo er aus verschiedenen Gründen nicht durchführbar ist. Ein Nachteil liegt in der Rauschanfälligkeit bei den Verfahren, die wirklich nur *ein einziges* korrespondierendes Punktepaar nutzen. Zudem ist natürlich nicht jede Information über die Szene ohne Kalibrierung zu gewinnen, die mit Kalibrierung zu gewinnen ist. Und der Einsatzbereich (fast) total unkalibrierter Verfahren ist wie bei kalibrierten nicht beliebig groß, was sich bereits gezeigt hat. Welche Anforderungen werden aber an die Verfahren gestellt, wenn man auch auf interne Kameraparameter verzichten will? Diese Frage zu beantworten, die bislang in der Literatur wenig untersucht wurde, hat sich dieser Abschnitt zum Ziel gesetzt.

Gemäß §2.2 stehen im Fall von senkrecht aufeinander stehenden Bildkoordinatenachsen die normalisierten Bildpunkte $\boldsymbol{x} = (x, y)$ mit den direkt beobachtbaren Bildpunkten in Pixelkoordinaten $\boldsymbol{x}_p = (x_p, y_p)$ über die Skalierungsfaktoren s_x, s_y und das Bildzentrum in Pixeln (x_0, y_0) wie folgt in Beziehung (siehe (2.13–2.16)):

$$\begin{aligned} x_p &= s_x x + x_0 & \qquad y_p &= s_y y + y_0 \\ x &= \frac{x_p - x_0}{s_x} & \qquad y &= \frac{y_p - y_0}{s_y} \end{aligned}$$

Die Darstellung als 3D-Punkt-Transformation durch die Matrix S (§2.2) führt zu folgender wichtiger Bemerkung.

Bemerkung 4.8 Die Eigenschaften der Transformationen $\mathcal{R}_i$ und $\mathcal{A}_i$, $i \in \{1,2,3\}$, bleiben bei der Multiplikation von links mit der Transformation, die durch die internen Parameter induziert wird, erhalten. Daher haben die Rekonstruktionen mit $\boldsymbol{x}_p$ anstelle von $\boldsymbol{x}$ dieselben Invarianzen. Lediglich das Fixgebilde ändert sich.

Beweis: (exemplarisch) Wir betrachten zum Beispiel die σ-Z-Rekonstruktion mit $\boldsymbol{X}' = (X', Y', Z', 1)^T =_{\text{def}} (x_p, y_p, \sigma, 1)^T$. Dann steht der rekonstruierte Punkt $\boldsymbol{X}'$ mit dem Originalpunkt $\boldsymbol{X} = (X, Y, Z, 1)^T \equiv (x, y, 1, \lambda)^T$ wie folgt in Beziehung:

$$
\begin{aligned}
\boldsymbol{X}' &= \begin{pmatrix} s_x & 0 & 0 & x_0 \\ 0 & s_y & 0 & y_0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ \sigma \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & 0 & x_0 \\ 0 & s_y & 0 & y_0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathcal{R}_2 \boldsymbol{X} \\
&= \begin{pmatrix} s_x & 0 & 0 & x_0 \\ 0 & s_y & 0 & y_0 \\ 0 & 0 & B & A \\ 0 & 0 & 1 & 0 \end{pmatrix} \boldsymbol{X} = S_2 \mathcal{R}_2 \boldsymbol{X}, \quad S_2 =_{\text{def}} \begin{pmatrix} s_x & 0 & 0 & x_0 \\ 0 & s_y & 0 & y_0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}
$$

Man sieht, daß sich wieder eine projektive Transformation ergeben hat. Mehr noch hat die Multiplikation mit S_2 nur die ersten beiden Zeilen der Relieftransformation $\mathcal{R}_2$ betroffen. Daher ergibt sich wieder eine Relieftransformation. Eine solche Gestalt, bei der in den ersten beiden Zeilen die internen Parameter auftauchen, die Parameter der affinen bzw. skalierten relativen Nähe aber nicht betroffen sind, erhält man auch für die anderen Transformationen. Der Grund dafür ist, daß die Transformation durch die internen Parameter nicht projektiv, sondern nur affin in der Bildebene ist und die Tiefe der 3D-Punkte nicht beeinflußt. Zudem bleiben die Vorzeichen der Determinanten und somit die Behandlung der Orientierung erhalten, weil s_x, s_y, s_{xy} positiv sind. Da man aber nicht mehr aus Sehstrahlen rekonstruiert, bleiben diese bei den Transformationen nicht mehr fix. □

Die Bemerkung ist deshalb wichtig, weil sie besagt, daß für eine Rekonstruktion über eine AN oder SRN nur die Pixelkoordinaten der Bildpunkte benötigt werden. Die Notwendigkeit der Kenntnis der internen Parameter hängt also allein von den AN- bzw. SRN-Berechnungsverfahren ab. Dies soll im folgenden für die einzelnen Modelle untersucht werden.

Da diese Modelle mit den Verschiebungsraten arbeiten, stellt sich die Frage, inwiefern die bei der Zuordnung bestimmte Verschiebung in Pixeln benutzt werden kann. Es gilt

$$
\begin{aligned}
(\dot{x}_p, \dot{y}_p) &= (s_x \dot{x}, s_y \dot{y}) \qquad (4.13) \\
(d_{x_p}, d_{y_p}) &= (x_{p_r} - x_{p_l}, y_{p_r} - y_{p_l}) = (s_x d_x, s_y d_y)
\end{aligned}
$$

nur, falls die internen Parameter in den betrachteten Aufnahmen gleich sind, was wir kurz mit Int = const. notieren. Der Stereofall ist oft nicht so restriktiv, liefert dann aber kompliziertere Formeln. Der wichtigste Fall ist der gleicher Skalierungsfaktoren, notiert mit Skal = const., in dem gilt:

$$
\begin{aligned}
(d_{x_p}, d_{y_p}) &= (s_x d_x + (x_{0_r} - x_{0_l}), s_y d_y + (y_{0_r} - y_{0_l})) \\
&= (s_x d_x + \Delta x_0, s_y d_y + \Delta y_0), \qquad (4.14) \\
&\quad (\Delta x_0, \Delta y_0) =_{\text{def}} (x_{0_r} - x_{0_l}, y_{0_r} - y_{0_l})
\end{aligned}
$$

In diesem Licht muß die Approximation der zyklopischen Bildpunktkoordinaten $(x_c, y_c) \approx \frac{1}{2}(x_l + x_r, y_l + y_r)$ aus §3.2 neu gesehen werden. Zum Beispiel ist

$$x_{p_c} \approx \frac{1}{2}((s_{x_l} x_l + s_{x_r} x_r) + (x_{0_l} + x_{0_r})).$$

Auch hier ist lediglich Skal = const. erforderlich, um zu einer Darstellung

$$(x_{p_c}, y_{p_c}) \approx \frac{1}{2}(x_{p_l} + x_{p_r}, y_{p_l} + y_{p_r}) = (\tilde{s}_{x_c} x_c + \tilde{x}_{0_c}, \tilde{s}_{y_c} y_c + \tilde{y}_{0_c})$$

mit $\tilde{s}_{x_c} =_{\text{def}} s_{x_l} = s_{x_r}$, $\tilde{s}_{y_c} =_{\text{def}} s_{y_l} = s_{y_r}$, $\tilde{x}_{0_c} =_{\text{def}} \frac{1}{2}(x_{0_l} + x_{0_r})$, $\tilde{y}_{0_c} =_{\text{def}} \frac{1}{2}(y_{0_l} + y_{0_r})$ zu kommen, die realen Bildebenen entspricht.

Wir fassen zusammen: Im Bewegungsfall sollte man von Int = const. ausgehen und mit (4.13) rechnen, was wir bei den in Frage kommenden Verfahren auch tun werden, um die Formeln einfach zu halten. Im Stereofall genügt es oft, Skal = const. zu fordern und mit (4.14) zu arbeiten. Soweit ich es überblicke, können alle hier besprochenen Verfahren bis auf das PAD-Modell eine zusätzliche additive Konstante verkraften, weil sie nur die Darstellung und nicht die qualitative Ausgabe beeinflußt. Um die Flut an Fällen einzudämmen, werden wir darauf bei den einzelnen Berechnungsmodellen aber nicht immer eingehen.

4.5.1 Das Deformationsmodell

Das Deformationsmodell nutzt die einfachere Darstellung der Deformation im Bildzentrum aus. Daher muß es bekannt sein. Folgende Definition von Bildpunktkoordinaten, die nur Skalierungen enthalten, ist sinnvoll:

$$\bar{\boldsymbol{x}} = (\bar{x}, \bar{y}) =_{\text{def}} (x_p - x_0, y_p - y_0) = (s_x x, s_y y)$$

Da in den Scherungen horizontale und vertikale Verschiebungsrate kombiniert werden, ist zusätzlich noch zu fordern $f = s_x = s_y$. Damit ergibt sich:

$$\begin{aligned} \textbf{def}\ \dot{\boldsymbol{x}}_{\mathrm{p}} &= (\mathrm{shear}_1\ \dot{\boldsymbol{x}}_{\mathrm{p}}, \mathrm{shear}_2\ \dot{\boldsymbol{x}}_{\mathrm{p}}) = (\frac{\partial \dot{\mathrm{x}}_{\mathrm{p}}}{\partial \mathrm{x}} - \frac{\partial \dot{\mathrm{y}}_{\mathrm{p}}}{\partial \mathrm{y}}, \frac{\partial \dot{\mathrm{x}}_{\mathrm{p}}}{\partial \mathrm{y}} + \frac{\partial \dot{\mathrm{y}}_{\mathrm{p}}}{\partial \mathrm{x}}) \\ &= (f(\frac{\partial \dot{x}}{\partial x} - \frac{\partial \dot{y}}{\partial y}), f(\frac{\partial \dot{x}}{\partial y} + \frac{\partial \dot{y}}{\partial x})) = f\textbf{def}\ \dot{\boldsymbol{x}} \end{aligned}$$

Man erhält also in analoger Form eine SRN σ_p, die noch die Brennweite f enthält:

$$\sigma_p = \textbf{def}\ \dot{\boldsymbol{x}}_{\mathrm{p}}(\mathrm{x}_0, \mathrm{y}_0)^{\mathrm{T}}(\bar{\mathrm{x}}, \bar{\mathrm{y}}) = [\mathrm{f}\textbf{def}\ \dot{\boldsymbol{x}}(0,0)]^{\mathrm{T}}[\mathrm{f}(\mathrm{x}, \mathrm{y})] = \mathrm{f}^2[\textbf{def}\ \dot{\boldsymbol{x}}(0,0)^{\mathrm{T}}\boldsymbol{x}] = \mathrm{f}^2\sigma$$

Also: **Ist das Bildzentrum (x_0, y_0) bekannt und $f = s_x = s_y$, so liefert die Deformation eine SRN mit $A = -f^2 v_x$ und $B = f^2 v_x \lambda_0$ mittels def $\dot{\boldsymbol{x}}_{\mathrm{p}}(\mathrm{x}_0, \mathrm{y}_0)^{\mathrm{T}}(\bar{\mathrm{x}}, \bar{\mathrm{y}})$.**

4.5.2 Das RDC-Modell

Garding et al. [GPFM96] stellten das RDC-Modell bereits für eine unbekannte Brennweite f vor. Man fordert hier also wieder ein bekanntes Bildzentrum und $f = s_x = s_y$. Für

$\bar{\boldsymbol{x}} = (\bar{x}, \bar{y})$ haben dann die Verschiebungsraten folgende Gestalt:

$$\begin{aligned}
\dot{x}_{p\pi} &= f(-\lambda_0 v_x - \omega_y) + (\lambda_0 p_0 v_x + \lambda_0 v_z)\bar{x} + (\lambda_0 q_0 v_x + \omega_z)\bar{y} \\
&\quad + \frac{1}{f}(-\lambda_0 q_0 v_z + \omega_x)\bar{x}\bar{y} + \frac{1}{f}(-\lambda_0 p_0 v_z - \omega_y)\bar{x}^2 \\
&= \bar{A}_x + \bar{B}_x\bar{x} + \bar{C}_x\bar{y} + \bar{D}_x\bar{x}\bar{y} + \bar{E}_x\bar{x}^2 \\
\dot{y}_{p\pi} &= f(-\lambda_0 v_y + \omega_x) + (\lambda_0 p_0 v_y - \omega_z)\bar{x} + (\lambda_0 q_0 v_y + \lambda_0 v_z)\bar{y} \\
&\quad + \frac{1}{f}(-\lambda_0 p_0 v_z - \omega_y)\bar{x}\bar{y} + \frac{1}{f}(-\lambda_0 q_0 v_z + \omega_x)\bar{y}^2 \\
&= \bar{A}_y + \bar{B}_y\bar{x} + \bar{C}_y\bar{y} + \bar{D}_y\bar{x}\bar{y} + \bar{E}_y\bar{y}^2
\end{aligned}$$

Es gilt wiederum $\bar{D}_x = \bar{E}_y$ und $\bar{E}_x = \bar{D}_y$, vor allem aber $\bar{B}_x = B_x$, $\bar{C}_x = C_x$, $\bar{B}_y = B_y$, $\bar{C}_y = C_y$, was im Falle eines unbekannten Bildzentrums nicht erfüllt würde. Definiert man dann analog

$$\bar{g}_\pi =_{\text{def}} -\bar{C}_y\bar{x} + \bar{B}_y\bar{y} - \bar{D}_y\bar{x}^2 - \bar{E}_y\bar{x}\bar{y},$$

so gewinnt man unter der Bedingung $v_y = 0$

$$\begin{aligned}
\dot{x}_{p\pi} + \bar{g}_\pi &= \bar{A}_x + (\bar{B}_x - \bar{C}_y)\bar{x} + (\bar{C}_x + \bar{B}_y)\bar{y} = \bar{A}_x + (B_x - C_y)\bar{x} + (C_x + B_y)\bar{y} \\
&= f(-\lambda_0 v_x - \omega_y) + v_x\lambda_0(p_0\bar{x} + q_0\bar{y}).
\end{aligned}$$

Eine SRN σ_p bekommt man dann unter Ausnutzung der Ebenenbedingung (4.12) über $\lambda_0(p_0\bar{x} + q_0\bar{y}) = f(\lambda_0 - \lambda)$:

$$\sigma_p = \dot{x}_{p\pi} + \bar{g}_\pi = f(-v_x\lambda - \omega_y) = f\sigma$$

Insgesamt kann man schließen: **Ist das Bildzentrum** (x_0, y_0) **bekannt und** $f = s_x = s_y$**, so liefert eine Ebenenanpassung** $\bar{g}_\pi(\bar{\boldsymbol{x}})$ **eine SRN mit** $A = -fv_x$ **und** $B = -f\omega_y$ **mittels** $\dot{x}_{p\pi} + \bar{g}_\pi(\bar{\boldsymbol{x}})$**.**

4.5.3 Das PAD-Modell

Wir beschränken uns auf die kartesische Berechnung der SRN mittels $\dot{x} - \frac{x}{y}\dot{y}$. Wegen der Singularität $y = 0$ muß y_0, aber in der Tat nicht mehr, bekannt sein, um wenigstens eine AN zu bestimmen. Sind beide Komponenten des Bildzentrums bekannt, so erhält man sogar eine SRN:

$$\sigma_p = \dot{x}_p - \frac{\bar{x}}{\bar{y}}\dot{y}_p = s_x\dot{x} - \frac{\bar{x}}{\bar{y}}s_y\dot{y} = s_x(\dot{x} - \frac{x}{y}\dot{y}) = s_x\sigma$$

Ist das Bildzentrum bekannt, so liefert die Verschiebungsrate eine SRN mit $A = -s_xv_x$ **und** $B = -s_x\omega_y$ **mittels** $\dot{x}_p - \frac{\bar{x}}{\bar{y}}\dot{y}_p$**.** Wird hier $\dot{y}_p$ durch $s_y\dot{y} + \Delta y_0$ ersetzt, ergibt sich jedoch weder eine SRN noch eine AN, weshalb hier Int = const. gefordert werden muß. [4]

Dieses Vorgehen legt nahe, die Berechnung ohne Kenntnis von x_0 durchzuführen. Dies liefert:

$$\begin{aligned}
\alpha_p &= \dot{x}_p - \frac{x_p}{\bar{y}}\dot{y}_p = (\dot{x}_p - \frac{\bar{x}}{\bar{y}}\dot{y}_p) - \frac{x_0}{\bar{y}}\dot{y}_p = \sigma_p - \frac{x_0}{y}\dot{y} \\
&= \sigma_p - x_0v_z\lambda + x_0\omega_yx = (-s_xv_x - x_0v_z)\lambda - s_x\omega_y + x_0\omega_yx
\end{aligned}$$

[4] Genau genommen, darf $x_{0_l} \neq x_{0_r}$ sein, da $\dot{x}_p$ nicht von $\frac{\bar{x}}{\bar{y}}$ beeinflußt wird.

Ist y_0 bekannt, so liefert die Verschiebungsrate eine AN mit $A = -s_x v_x - x_0 v_z$, $B = -s_x\omega_y$ und $D = x_0\omega_y$ mittels $\dot{x}_p - \frac{x_p}{\bar{y}}\dot{y}_p$. Wegen $s_x, x_0 > 0$ ist lediglich im Bewegungsfall darauf zu achten, daß $A \neq 0$ ist.

4.5.4 Diskrete Versionen des PAD-Modells

Auch in der Eliminierung der internen Parameter ergeben sich Analogien zum PAD-Modell. Jedoch bewirkt die Unsymmetrie der Berechnung noch größere Freiheiten in linken und rechten Parametern. Wir verlangen nur $y_{0_l} = y_{0_r}$ und übertragen die Berechnungen:

$$
\begin{aligned}
\alpha_x &= d_{x_p} - \frac{\bar{x}_r}{\bar{y}_r} d_{y_p} = (s_{x_r}x_r - s_{x_l}x_l + x_{0_r} - x_{0_l}) - \frac{s_{x_r}x_r}{s_{y_r}y_r}(s_{y_r}y_r - s_{y_l}y_l) \\
&= s_{x_r}x_r\frac{s_{y_l}y_l}{s_{y_r}y_r} - s_{x_l}x_l + x_{0_r} - x_{0_l} = s_{x_r}\frac{s_{y_l}}{s_{y_r}}\frac{X_r}{Z_l} - s_{x_l}x_l + x_{0_r} - x_{0_l} \\
&= s_{x_r}\frac{s_{y_l}}{s_{y_r}}\frac{\cos(2\mu)X_l - \sin(2\mu)Z_l + v_x^r}{Z_l} - s_{x_l}x_l + x_{0_r} - x0_l \\
&= (s_{x_r}\frac{s_{y_l}}{s_{y_r}}v_x^r)\lambda_l + (-s_{x_r}\frac{s_{y_l}}{s_{y_r}}\sin(2\mu) + x_{0_r} - x_{0_l}) + (s_{x_r}\frac{s_{y_l}}{s_{y_r}}\cos(2\mu) - s_{x_l})x_l \\
\alpha_y &= d_{x_p} - \frac{x_{p_r}}{\bar{y}_r}d_{y_p} = (d_{x_p} - \frac{\bar{x}_r}{\bar{y}_r}d_{y_p}) - (\frac{x_{0_r}}{\bar{y}_r}d_{y_p} = \alpha_x - x_{0_r}(1 - \frac{s_{y_l}}{s_{y_r}}\frac{Z_r}{Z_l}) \\
&= \alpha_x - x_{0_r}(1 - \frac{s_{y_l}}{s_{y_r}}\frac{\sin(2\mu)X_l + \cos(2\mu)Z_l + v_z^r}{Z_l}) \\
&= (s_{x_r}\frac{s_{y_l}}{s_{y_r}}v_x^r + x_{0_r} + x_{0_r}\frac{s_{y_l}}{s_{y_r}}v_z^r)\lambda_l + (-s_{x_r}\frac{s_{y_l}}{s_{y_r}}\sin(2\mu) - x_{0_l} + x_{0_r}\frac{s_{y_l}}{s_{y_r}}\cos(2\mu)) \\
&\quad +(s_{x_r}\frac{s_{y_l}}{s_{y_r}}\cos(2\mu) - s_{x_l} + x_{0_r}\frac{s_{y_l}}{s_{y_r}}\sin(2\mu))x_l
\end{aligned}
$$

Ist das Bildzentrum bekannt, so liefert die Disparität eine AN mit $A = s_{x_r}\frac{s_{y_l}}{s_{y_r}}v_x^r$, $B = -s_{x_r}\frac{s_{y_l}}{s_{y_r}}\sin(2\mu) + x_{0_r} - x_{0_l}$ und $D = s_{x_r}\frac{s_{y_l}}{s_{y_r}}\cos(2\mu) - s_{x_l}$ mittels $d_{x_p} - \frac{\bar{x}_r}{\bar{y}_r}d_{y_p}$. Ist nur y_0 bekannt, so liefert die Disparität eine AN mit $A = s_{x_r}\frac{s_{y_l}}{s_{y_r}}v_x^r + x_{0_r}\frac{s_{y_l}}{s_{y_r}}v_z^r$, $B = -s_{x_r}\frac{s_{y_l}}{s_{y_r}}\sin(2\mu) - x_{0_l} + x_{0_r}\frac{s_{y_l}}{s_{y_r}}\cos(2\mu)$ und $D = s_{x_r}\frac{s_{y_l}}{s_{y_r}}\cos(2\mu) - s_{x_l} + x_{0_r}\frac{s_{y_l}}{s_{y_r}}\sin(2\mu)$ mittels $d_{x_p} - \frac{x_{p_r}}{\bar{y}_r}d_{y_p}$. Analoge affine Nähen ergeben sich durch Berechnung von $d_{x_p} - \frac{\bar{x}_l}{\bar{y}_l}d_{y_p}$ und $d_{x_p} - \frac{x_{p_l}}{\bar{y}_l}d_{y_p}$. Man beachte auch hier, daß die Parameter A in der Praxis immer ungleich Null sind wegen $v_x^r < 0$ und $v_x^l > 0$.

Da diese Verfahren nur im Stereofall angewendet werden, konnte hier exemplarisch ausführlich vorgeführt werden, wo und wie die Gleichheit linker und rechter interner Parameter eingeht. Die Parameter der berechneten Nähen sind wiederum vollkommen analog zum kontinuierlichen Fall in §4.5.3, wie man mit $\cos(2\mu) = 1$ und der Gleichheit von linken und rechten Parametern sieht.

4.5.5 Die horizontale Verschiebungsrate

Betrachtet man statt der normalisierten horizontalen Verschiebung die horizontale Verschiebung in Pixeln, so erhält man

$$\dot{x}_p = s_x\dot{x} = -s_x v_x\lambda - s_x\omega_y + s_x\omega_z y - s_x\omega_y x^2$$

$$\approx \quad -s_x v_x \lambda - s_x \omega_y + s_x \omega_z y.$$

Also: **Die horizontale Verschiebungsrate liefert eine AN mit $A = -s_x v_x$, $B = -s_x \omega_y$ und $C = s_x \omega_z$, und im Falle von $\omega_z = 0$ eine entsprechende SRN, mit additivem Fehler $-s_x \omega_y x^2$.** Der relative Fehler hat sich gegenüber der Verwendung von $\dot{x}$ natürlich nicht erhöht. Auffällig ist, daß zur, wenn auch fehlerhaften, Berechnung der SRN bzw. AN *keine Informationen über die internen Parameter* benötigt werden. Gleichzeitig rufe man sich aber in Erinnerung, daß die horizontale Verschiebungsrate nur in einer symmetrisch fixierten Systemgeometrie entsprechend $v_z = 0$ funktioniert.

4.6 Zusammenfassung und Überblick

Die Vielzahl der vorgestellten SRN- und AN-Verfahren und ihrer Modifikationen verlangen nach einer Zusammenstellung ihrer Anwendungsgebiete und Leistungen im Hinblick auf nötiges a-priori-Wissen, Anzahl der Eingabepunkte, etc. (Tabellen 4.1, 4.2). Zuvor aber noch eine Bemerkung zur Fixierung: Es wurde streng darauf geachtet, daß alle geometrischen Einschränkungen in Form von verschwindenden Bewegungsparametern formuliert wurden und sich nur darüber auswirken. Die Fixierung, als Bedingung an die Verschiebungsraten gestellt, erleichtert nur die Darstellungen, ihre Interpretation oder die Überprüfung und Realisierung der Geometrie und führt zur exakten Approximation der Disparität durch die zyklopische Verschiebungsrate im Bildzentrum. Da man also ohne Fixierungsbedingung auskommt, machen auch Verfahren Sinn, die ohne vollständige Kenntnis des Bildzentrums arbeiten.

Wir kommen jetzt zur Übersicht in Tabelle 4.1, aus der die hier verwendeten Bezeichnungen zu entnehmen sind. Die Spalte „Anwendung" nimmt Rücksicht auf die Praktikabilität des Verfahrens. So ist eine Kalibrierung im Bewegungsfall nicht durchführbar, andererseits schränken Bedingungen wie $v_z = 0$ nur im Stereofall nicht unangemessen ein. Der Übergang ist aber fließend, und so soll ein Eintrag in dieser Spalte nicht das letzte Wort über die Anwendbarkeit sprechen.

Die Mindestanzahl von Zuordnungen gibt die Zahl korrespondierender Paare von Bildpunkten an, die eine Rekonstruktion ermöglichen. Das *trans*-Verfahren stellt einen Sonderfall dar, da es für jede Transformationsmatrix Korrespondenzen zwischen 6 nicht komplanaren Punkten auf einem Kalibrierungsobjekt und ihren Bildern benötigt. Die klassischen *sfm*-Verfahren benötigen 5 Zuordnungen, um die 5 Parameter der relativen Bewegung zu bestimmen. Bei der Fundamentalmatrix kommen noch die unbekannten internen Parameter hinzu, was die Anzahl auf 7 erhöht. Hier zeigt sich die Stärke der SRN/AN-Verfahren, die direkt aus einem einzigen Punktepaar seine Lage in der projektiven Struktur ermitteln können. Das *RDC*-Modell liegt aufgrund der Ebenenanpassung im Bereich der klassischen Verfahren. Dadurch daß sie nicht auf einem einzelnen Vektor als Eingabe aufbauen, sind sie robuster, was sich jedoch erst weit über der Mindestanzahl bemerkbar macht. Das Deformationsmodell paßt nicht genau in diesen Rahmen: Es betrachtet zwar nur das Bildzentrum und seine Verschiebung, benötigt aber zur Berechnung des Gradienten den lokalen örtlichen Grauwertverlauf. Der Aufwand, der dabei betrieben wird, entspricht der einer Kalibrierung.

PADl/*PADr* haben eine gewisse theoretische Bedeutung dadurch, daß sie völlig fehlerfrei, da ohne Disparitätsapproximation, arbeiten. Der Fehler des *PAD*-Modells ist, wie

Name des Verfahrens	Kürzel	Anwendung	Mindestanzahl von Zuordnungen	Fehler
Euklidische Rekonstruktion mittels $T^{k \to w}$-Matrizen	*trans*	Stereo	6 (Kalibrierungs-objekt)	—
Euklidische Rekonstruktion mittels E-Matrix	*sfm*	Bewegung	5	—
Projektive Rekonstruktion mittels F-Matrix	*fund*	Stereo	7	—
Deformations-modell	*def*	Stereo Bewegung	lokales Disparitätsfeld	Ebenenmodell $[\tilde{\boldsymbol{x}}_c \approx \boldsymbol{x}_c]$
RDC-Modell	*RDC*	Stereo Bewegung	5	Ebenenmodell $[\tilde{\boldsymbol{x}}_c \approx \boldsymbol{x}_c]$
PAD-Modell	*PAD*	Stereo	1	$[\tilde{\boldsymbol{x}}_c \approx \boldsymbol{x}_c]$
diskrete PAD-Modelle	*PADl* *PADr*	Stereo	1	—
horizontale Disparität	*dx*	Stereo	1	$-\omega_y x^2$

Tabelle 4.1: Übersicht über die SRN/AN-Rekonstruktionsverfahren im Vergleich mit klassischen Verfahren. Einträge in eckigen Klammern betreffen nur den Stereofall.

in §3.2 und §3.3 gezeigt, sehr klein. Während das *def*-Modell nur lokal angewendet werden sollte, scheint der Fehlerterm des *RDC*-Modells gutartig und besser als bei *dx*. Die klassischen SfS- und SfM-Verfahren haben keinen inhärenten Fehler.

In Tabelle 4.2 sind die Modifikationen für unbekannte interne Parameter mitaufgenommen. Bei den Bedingungen, die das Verfahren an die Geometrie, die internen Parameter und an die a-priori-Kalibrierungsinformation stellt, wird die Allgemeinheit des *fund*-Ansatzes deutlich. Sind keine Einschränkungen und kein a-priori-Wissen vorhanden, so ist dieses Verfahren angezeigt. Das *trans*-Verfahren ermittelt hingegen alle Parameter durch eine vorgeschaltete Kalibrierung, das *sfm*-Verfahren verlangt die Eingabe von normalisierten Bildkoordinaten. Die letzten beiden Verfahren sind aber ebenso in allgemeinen Systemgeometrien anwendbar. Die SRN/AN-Verfahren decken hier einen weiten Bereich ab. Je nach Situation und Anforderung ist dem einen oder anderen der Vorzug zu geben. *def* und *RDC* schränken die Bewegung nur durch $v_y = 0$ ein, gehen aber von einer Ebene aus, was je nach Anwendung einen Fehler mit sich bringt. Beim *RDC*- wie beim *dx*-Modell ist darauf zu achten, daß die Approximation der Disparität durch die Verschiebungsrate im Stereofall für $\omega_x \neq 0$ oder $\omega_z \neq 0$ stark fehlerbehaftet sein kann und nicht mehr in den Rahmen fällt, der in §3 untersucht wurde. Die *PAD*-Modelle und *dx* verlangen schneidende Achsen. Dabei haben die *PAD*-Modelle den Nachteil, daß sie nicht im gesamten Bildbereich arbeiten können.

Die Bedingungen an die internen Parameter [5] werden interessant, wenn sie zumin-

[5] Streng genommen, muß nicht immer Int = const. gefordert werden, wenn die internen Parameter in allen Perspektiven bekannt sind.

Kürzel	geometrische Bedingungen	Bedingungen an interne Parameter	a-priori-Information	Rekon-struktion
trans	—	—	s_x, s_y, x_0, y_0, s_{xy} $\boldsymbol{v}$, $\boldsymbol{\omega}$	Euklidisch (lokal)
sfm	—	Int. = const.	s_x, s_y, x_0, y_0, s_{xy}	Euklidisch (Kamera)
fund	—	—	—	Projektiv (lokal)
def	Ebene $v_y = 0$ $[\omega_x \approx 0, \omega_z \approx 0]$	Int. = const.	s_x, s_y, x_0, y_0, s_{xy}	Projektiv (c, SRN)
def-f	– " –	Skal = const. $s_{xy} = 0$, $s_x = s_y$	x_0, y_0 [c]	– " –
RDC	Ebene $v_y = 0$ $[\omega_x \approx 0, \omega_z \approx 0]$	Int = const.	s_x, s_y, x_0, y_0, s_{xy}	Projektiv (c, SRN)
RDC-f	– " –	Skal = const. $s_{xy} = 0$, $s_x = s_y$	x_0, y_0 [c]	– " –
PAD	$v_y = 0$, $\omega_x = 0$, $\omega_z = 0$ $y \neq 0/\sin(\theta) \neq 0$	Int = const.	s_x, s_y, x_0, y_0, s_{xy}	Projektiv (c, SRN)
PAD-sx	$v_y = 0$, $\omega_x = 0$, $\omega_z = 0$ $y_p \neq y_0$	Skal = const. $s_{xy} = 0$, $y_{0_l} = y_{0_r}$	x_0, y_0 [c]	– " –
PAD-sx0	– " –	– " –	y_0 [c]	Projektiv (c, AN)
PADl *PADr*	$v_y = 0$, $\omega_x = 0$, $\omega_z = 0$ $y_l \neq 0/y_r \neq 0$	Int = const.	s_x, s_y, x_0, y_0, s_{xy} [l, r]	Projektiv (l/r, AN)
PADl-sx *PADr-sx*	$v_y = 0$, $\omega_x = 0$, $\omega_z = 0$ $y_{p_l} \neq y_0/y_{p_r} \neq y_0$	$s_{xy} = 0$, $y_{0_l} = y_{0_r}$	x_{0_l}/x_{0_r}, y_0	– " –
PADl-sx0 *PADr-sx0*	– " –	– " –	y_0	– " –
dx	$v_z = 0$, $\omega_x = 0$, $\omega_z = 0$	Int = const.	s_x, s_y, x_0, y_0, s_{xy}	Projektiv (c, SRN)
dx-AN	$v_z = 0$, $\omega_x = 0$ $[\omega_z \approx 0]$	– " –	– " –	Projektiv (c, AN)
dxp	$v_z = 0$, $\omega_x = 0$, $\omega_z = 0$	Skal = const. $s_{xy} = 0$	—	Projektiv (c, SRN)
dxp-AN	$v_z = 0$, $\omega_x = 0$ $[\omega_z \approx 0]$	– " –	– " –	Projektiv (c, AN)

Tabelle 4.2: Übersicht über die SRN/AN-Rekonstruktionsverfahren mit teilweise unbekannten internen Parametern im Vergleich mit klassischen Verfahren. Einträge in eckigen Klammern betreffen nur den Stereofall.

dest teilweise unbekannt sind. Die größtmögliche Veränderlichkeit der Internen erlaubt *PADl/PADr*, wo schließlich nur $y_{0_l} = y_{0_r}$ gefordert wird, gefolgt von *dxp* mit Skal = const., *def* und *RDC* verlangen darüber hinaus noch $s_x = s_y$, und *PAD* erlaubt nur $x_{0_l} \neq x_{0_r}$. Dabei wurden die Untersuchungen nur für $s_{xy} = 0$ durchgeführt, was durchaus realistisch ist.

Die SRN-Verfahren kommen nicht um die vollständige Kenntnis des (zyklopischen) Bildzentrums herum. Die *PAD*-Modelle liefern aber eine AN nur mit y_0. Die Gewinnung einer SRN über *dxp* kommt völlig ohne Kalibrierungsinformation aus, ist aber des Fehlers wegen mit Vorsicht zu genießen. Jedoch bleibt festzuhalten: *dxp* **ist das einzige total unkalibrierte Verfahren, daß bei schneidenden Achsen und symmetrischer Stereogeometrie eine (fehlerbehaftete) SRN liefert.**

Das *trans*- und das *sfm*-Verfahren rekonstruieren Euklidisch. Dabei liefert *trans* die Szene in dem von der Kalibrierung abhängigen Koordinatensystem, *sfm* in einem Kamerakoordinatensystem. Das *fund*-Verfahren liefert die projektive Struktur der Szene in einer Basis, die von den gewählten Zuordnungen abhängt. Die SRN- und AN-Verfahren rekonstruieren eine Szene, die sich um eine genau bestimmte Transformation von der im linken, rechten oder meist zyklopischen Koordinatensystem beschriebenen realen Szene unterscheidet. Dabei ermöglichen sie eine physikalisch sinnvolle Rekonstruktion mit $Z > 0$ und $W = 1$. Die SRN-Verfahren liefern zudem die Tiefenordnung.

Welchem Verfahren der Vorzug zu geben ist, hängt von der konkreten Aufgabe ab. Es sollte jedoch deutlich geworden sein, daß bereits die projektive Struktur und die SRN wichtige Informationen über die Szene enthalten. Bei Voraussetzung bekannter geometrischer Einschränkungen lassen sich diese äußerst schnell und einfach gewinnen, ohne alle internen Kameraparameter kennen oder bestimmen zu müssen. So lassen sich einfache Moduln entwickeln, die bei der Gewinnung von 3D-Information helfen.

4.7 Experimente und Diskussion

Zum Abschluß dieses Kapitels wollen wir einige der vorgestellten Verfahren bezüglich des SfS-Problems experimentell vergleichen. Die Konzentration liegt dabei auf den *PAD*- bzw. den *dx*-Verfahren, da sie ohne Kalibrierung arbeiten, höchstens unter Benutzung einiger interner Kameraparameter, und sich damit grundlegend von den anderen Verfahren unterscheiden. Es ist zu untersuchen, wie sicher sie Informationen trotz der Benutzung je nur eines einzigen korrespondierenden Bildpunktepaares bestimmen können, und wie sie gegeneinander und gegen das klassische *trans*-Verfahren bestehen. Dabei untersuchen wir die projektive und die Euklidische Rekonstruktion der Szene.

Die hier durchgeführten Experimente basieren auf künstlichen Daten. Dabei dienen als Szenepunkte $3 \cdot 4 \cdot 4 = 48$ regelmäßig auf einem Kalibrierungswürfel angeordnete Punkte (Abbildungen 4.4, 4.5). Die binokulare einfache Stereogeometrie ist bestimmt durch die Festlegung der Version γ, der Vergenz μ und der Breite der Stereobasis I. Diese Parameter und die Parameter der Kameraprojektion manifestieren sich in den Transformationsmatrizen, die die Beziehung zwischen lokalem Würfelkoordinatensystem und Kamerakoordinatensystem enthalten. Wir haben hier für links und rechts gleiche interne Kameraparameter gewählt: $s_x = 770.4$, $s_y = 747.6$, $x_0 = 253.9$, $y_0 = 293.9$, $s_{xy} = 0$. Diese Parameter, die Stereobasisbreite $I = 250$ und die Entfernung des Fixierungspunktes $Z_{0_c} = 1000$ bleiben in allen Versuchen gleich. Als Variable des Systems dient γ, woraus

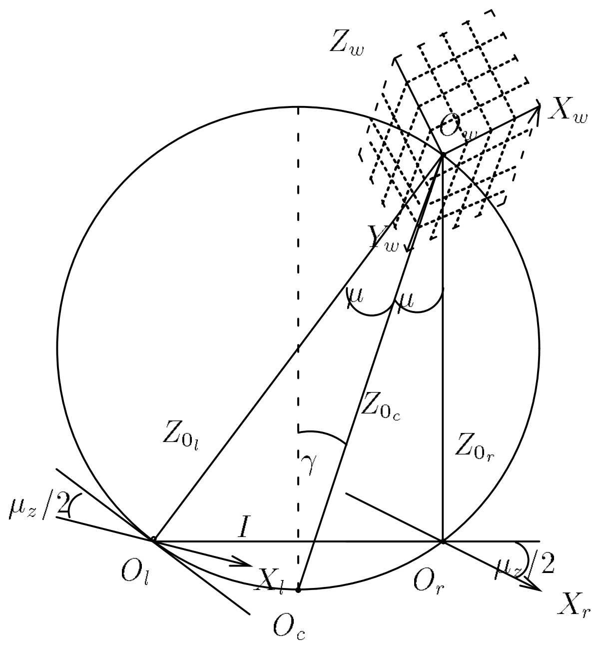

Abbildung 4.4: Aufbau der Experimente. Systemgeometrie nach §3. Die Parameter Z_{0_c} und I sind in den Experimenten fest, γ, μ und μ_z variabel. Der Ursprung des lokalen Weltkoordinatensystems des Würfels liegt im vordersten Eckpunkt O_w, der zugleich der Fixierungspunkt ist.

sich über die Fixierungsbedingung (3.25) μ ableiten läßt. Dann ergibt sich die Euklidische Transformation zwischen Kamera- und Weltkoordinatensystem zu:

$$R^{l\to w} = R_y(-\mu)R_z(\mu_z/2)R_{\text{lok}} \qquad \boldsymbol{t}^{l\to w} = (0, 0, \frac{Z_{0_c}}{\cos(\gamma)}\cos(\gamma+\mu))^T$$

$$R^{r\to w} = R_y(\mu)R_z(-\mu_z/2)R_{\text{lok}} \qquad \boldsymbol{t}^{r\to w} = (0, 0, \frac{Z_{0_c}}{\cos(\gamma)}\cos(\gamma-\mu))^T$$

Dabei bezeichnen R_x, R_y, R_z Rotationen um die X-, Y- bzw. Z-Achse. $R_{\text{lok}} = R_x(45°) \cdot R_y(-45°)$ ist eine immer gleiche Rotation des Würfelkoordinatensystems zur Sichtbarmachung verdeckter Punkte. μ_z bezeichnet die Torsion des Systems. Die Translationen in Z-Richtung ergeben sich aus der Entfernung des vorderen Würfeleckpunktes, das ist hier der Fixierungspunkt, von den Zentralprojektionspunkten. Z. B. ergibt sich aus dem Sinussatz

$$Z_{0_l} = \sin(-(90° + \gamma + \mu))\frac{I}{\sin(2\mu)} = \cos(\gamma+\mu)\frac{Z_{0_c}}{\cos(\gamma)}.$$

Man beachte, daß die relative Lage des Würfels bezüglich des zyklopischen Auges bei festem I und Z_{0_c} immer gleich ist.

Neben diesen geometrischen Parametern gibt es noch weitere Eingaben. Die *PAD*-Modelle benötigen einen Radius r in Pixeln um die Singularität $y = 0$ herum. Da besonders die Robustheit der Verfahren interessiert, wird zu den sich aus der Anwendung der Transformationsmatrizen ergebenden Bildpunkten in Pixeln ein Gaußsches Rauschen mit der Standardabweichung σ_{noise} (sigma) in Pixeln addiert. Ebenso werden die benötigten internen Parameter relativ gestört (ΔInt), um Kalibrierungsfehler zu simulieren. Der prinzipielle Ablauf der Experimente ist in Abbildung 4.6 dargestellt.

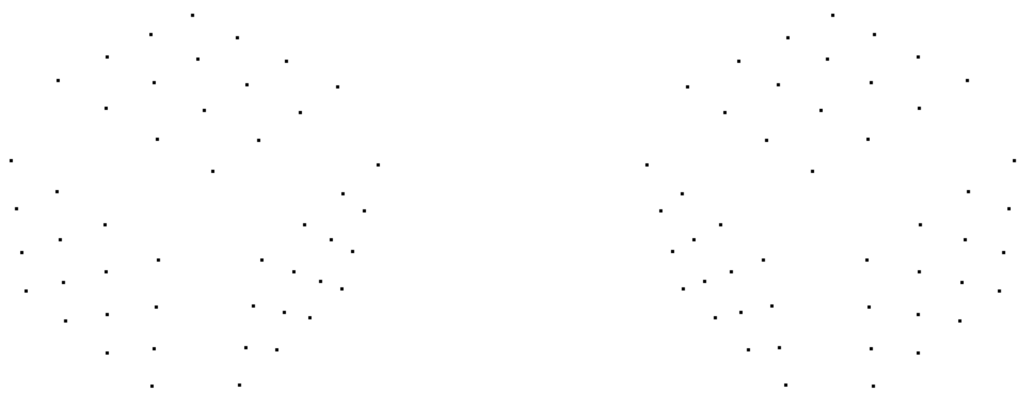

Abbildung 4.5: Linkes und rechtes 512×512-Bild der für die Experimente verwendeten Punkte auf einem Kalibrierungswürfel. Diese Abbildungen ergeben sich für verschwindende Torsion und Version.

4.7.1 Projektive Rekonstruktion

In diesem Abschnitt geht es um die Güte der projektiven Rekonstruktion der Szene. Zum ersten wird der relative Fehler für die σ/α-Z-Rekonstruktion bestimmt. Danach interessieren wir uns für zwei wesentliche invariante Eigenschaften, die Komplanarität und die Tiefenordnung, und wie sie wiedergegeben werden. Die Tiefenordnung wird durch die σ-Z-Rekonstruktion gewährleistet, für die Komplanarität wird die σ/α-λ'-Rekonstruktion genutzt, um eine in den Ausmaßen mit der Szene vergleichbare Rekonstruktion zu erhalten. Auf Details wird im einzelnen eingegangen.

Wir kommen zunächst zur σ/α-Z-Rekonstruktion. Der gewonnene σ- bzw. α-Wert dient hier zu einer Rekonstruktion $\boldsymbol{X}' =_{\text{def}} (x, y, \sigma, 1)^T$ bzw. $\boldsymbol{X}' =_{\text{def}} (x, y, \alpha, 1)^T$. Dabei wird zunächst der relative Fehler in der projektiven Rekonstruktion durch $\frac{\|\boldsymbol{X}-\boldsymbol{X}'\|_2}{\|\boldsymbol{X}\|_2}$ in den einzelnen Verfahren verglichen. Die Grundwahrheit $\boldsymbol{X}$ ergibt sich dabei aus den bekannten Sehstrahlen und der Grundwahrheit für σ bzw. α, die man aus den bekannten Parametern der Systemgeometrie bestimmt. Da sonst die Fehler unverhältnismäßig groß in den Bildpunkten werden, wurden die internen Parameter hier als exakt bekannt vorausgesetzt, so daß mit den Sehstrahlen rekonstruiert werden konnte. Abbildung 4.7 zeigt die Ergebnisse für das *PAD*-, *PADl*- und *dx*-Verfahren. In Abhängigkeit des Bildrauschens zeigt sich, daß der Fehler im *dx*-Verfahren durch den inhärenten Fehlerterm $-\omega_y x^2$ am größten ist, sich die Verfahren aber bei stärkerem Rauschen annähern. *PADl* ist bei sehr genauer Kenntnis der Bildpunkte besser, da hier die Sehstrahlen nicht approximiert werden, dann wirkt sich aus, daß *PADl* eine affine Nähe bestimmt, die von einem x-Term abhängt, was sich bei stärkerem Rauschen bemerkbar macht. Da *dx* unter der Bedingung $\gamma = 0°$ funktioniert, wirkt sich bei wachsender Version dies entsprechend aus. Die Empfindlichkeit ist aber immer noch geringer als die von *PAD* und *PADl* gegenüber nicht verschwindender Torsion. Diese Ergebnisse erhält man mehr oder weniger direkt durch die Fortpflanzung der Fehler in den Bildpunkten. Aufschlußreicher ist die Betrachtung der

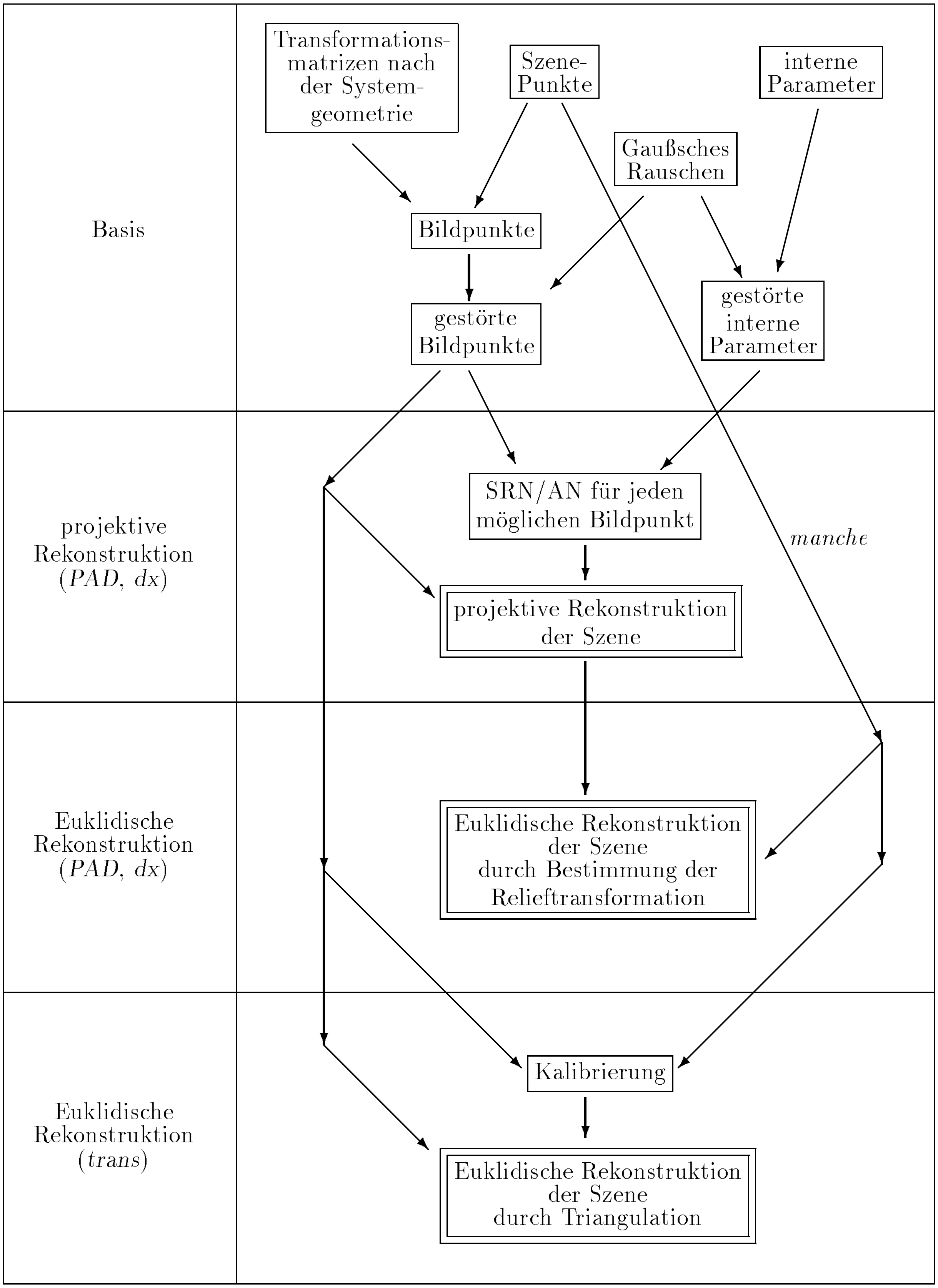

Abbildung 4.6: Prinzipieller Ablauf der Experimente für das *PAD*-, das *dx*- und das *trans*-Verfahren.

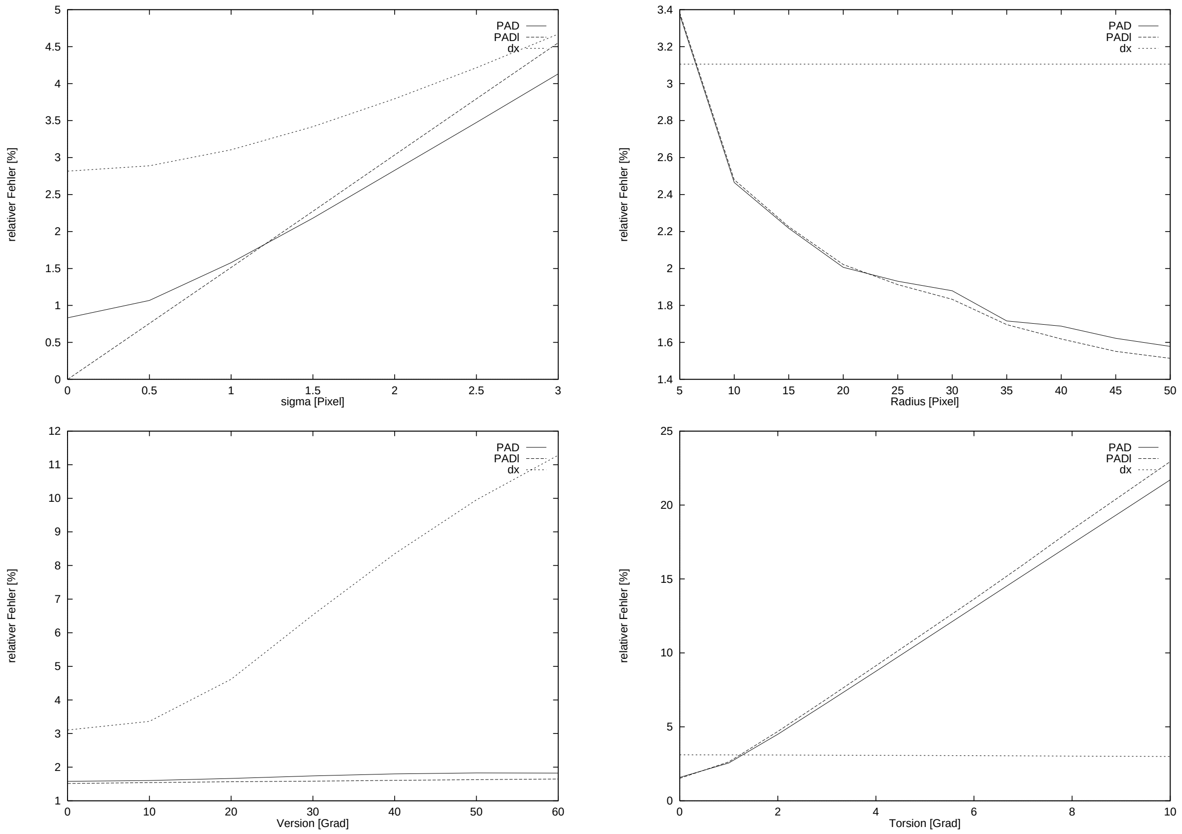

Abbildung 4.7: Relativer Fehler in der σ/α-Z-Rekonstruktion für *PAD*, *PADl* und *dx*. Falls nicht anders angegeben, ist $\sigma_{noise} = 1$, $\gamma = 0°$, $\mu_z = 0°$, $r = 50$. Es ist stets $\Delta\mathrm{Int} = 0$.

Leistung der *PAD*-Verfahren bezüglich des Radius. Bei einem sehr kleinen Radius sind sie sogar schlechter als das *dx*-Verfahren. Die Kurve wird erst etwa für $r = 25$ flacher, das Verfahren immer besser, da auch weniger Punkte rekonstruiert werden (siehe Tabelle 4.3). Dies weist auf das kritische, insbesondere numerische, Fehlerverhalten des Terms $\frac{x}{y}$ hin. Für Bildpunkte (x, y) mit y nahe bei 0 werden Fehler in x und der vertikalen Disparität entsprechend verstärkt.

Radius [Pixel]	rekonstruierte Punkte (von 48) [#]	Radius [Pixel]	rekonstruierte Punkte (von 48) [#]
5	46	30	38
10	46	35	38
15	44	40	36
20	42	45	36
25	42	50	33

Tabelle 4.3: Anzahl der von den *PAD*-Verfahren rekonstruierten Punkten in Abhängigkeit des Radius um $y_p = y_0$.

Der Fehler in der projektiven Rekonstruktion scheint zwar in vielen Fällen relativ gering zu sein, liefert aber kaum Hinweise darauf, wie sicher Informationen aus der projektiven Hülle zu bekommen sind. Deshalb werden wir darauf jetzt genauer eingehen. Die

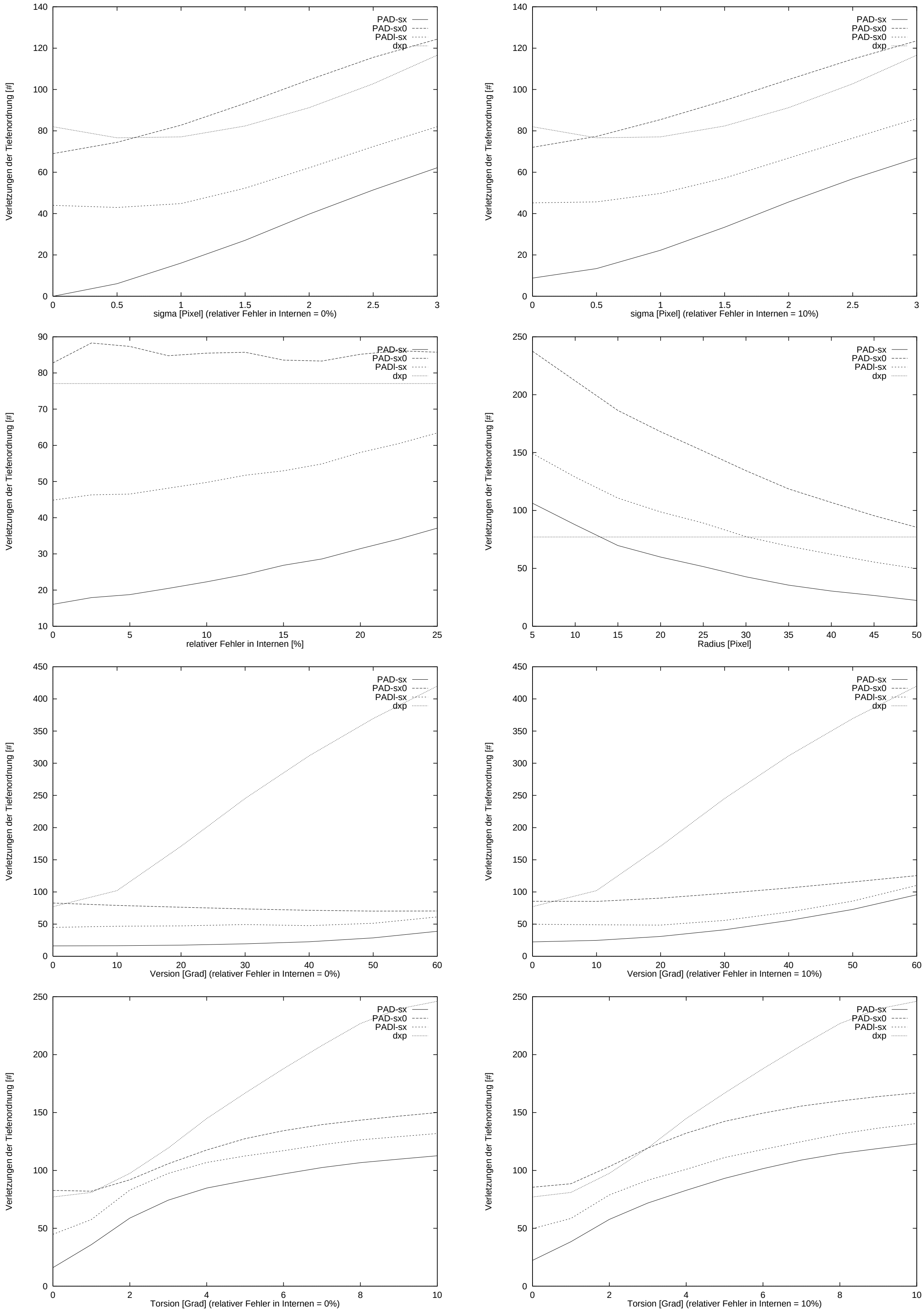

Abbildung 4.8: Anzahl der Verletzungen der Tiefenordnung in der σ/α-Z-Rekonstruktion für *PAD-sx*, *PAD-sx0*, *PADl-sx* und *dxp*. Falls nicht anders angegeben, ist $\sigma_{noise} = 1$, $\gamma = 0°$, $\mu_z = 0°$, $r = 50$, $\Delta\text{Int} = 10\%$.

signifikante Eigenschaft der Relieftransformation ist die Erhaltung der Tiefenordnung. Im folgenden werden wir die Anzahl der Verletzungen der Tiefenordnung bei der Rekonstruktion betrachten, also $\mathrm{card}(\{\{(\boldsymbol{X}_1, \boldsymbol{X}'_1), (\boldsymbol{X}_2, \boldsymbol{X}'_2)\} | (Z_1 < Z_2 \wedge Z'_1 > Z'_2) \vee (Z_1 > Z_2 \wedge Z'_1 < Z'_2)\})$. Es ergibt sich ein maximaler Wert von $\binom{48}{2} = 1128$. Da es hier prinzipiell nur um die Ermittlung skalierter relativer bzw. affiner Nähe geht, können hier die σ/α-Z-Rekonstruktionen aus Pixelkoordinaten und somit (partiell) unkalibrierte Verfahren eingesetzt werden. Dabei hat sich hier und später gezeigt, daß jeweils *PAD* und *PAD-sx*, *PADl* und *PADl-sx* sowie *dx* und *dxp* nahezu identische Ergebnisse liefern. Dies überrascht nicht, wenn man sich die Analogien der Berechnungen vor Augen hält. Also gilt: **Für die Rekonstruktion kann man im Falle von *PAD* und *PADl* auf die Kenntnis der Skalierungsfaktoren und im Falle von *dx* auf die Kenntnis aller interner Parameter verzichten.** Der Verzicht auf die Kenntnis von x_0 bei den *PAD*-Verfahren ist dagegen problematisch. Der Grund dafür ist vor allem die Größe des Ausdrucks $\frac{\bar{x}}{\bar{y}}$, der in unseren Experimenten wesentlich kleiner ist als $\frac{x_p}{\bar{y}}$, wodurch der Fehler in unakzeptabler Weise verstärkt wird. Um doch eine experimentelle Leistung der Verfahren vorzuweisen, ist *PAD-sx0* für die Tiefenordnung mit aufgenommen. Lediglich *PAD-sx* bestimmt eine SRN, *PAD-sx0*, *PADl-sx* und *dxp* nur eine AN. Angesichts von Abbildung 4.8 kann man allgemein feststellen, daß *PAD-sx* auch stets am besten abschneidet, gefolgt von *PADl-sx*, dann *PAD-sx0* und *dxp*. **Dabei sollte man aufgrund der Größenordnung der Fehler nur in den Fällen, in denen das Bildzentrum nicht einmal näherungsweise bekannt ist oder Bildpunkte mit y_p nahe bei y_0 von Interesse sind, andere Verfahren dem *PAD-sx*-Verfahren zur Bestimmung der Tiefenordnung vorziehen.**

Wir schließen die Eigenschaften der projektiven Hülle mit Betrachtungen zur Komplanarität ab. Es wurden $4 \cdot 4 = 16$ Punkte der oberen Ebene im Bild (siehe Abbildung 4.5) rekonstruiert, da die Gerade $y = 0$ dort nicht hindurchführt. Um zu einer Rekonstruktion zu gelangen, die die realen Ausmaße der abgebildeten Ebene widerspiegelt, wurde die σ-λ'-Rekonstruktion mit Sehstrahlen angewendet. Wir demonstrieren die geeignete Parameterwahl am Beispiel eines SRN-Verfahrens: Es ist nach Definition

$$\lambda' = \frac{\sigma - B'}{A'} = \frac{\sigma}{A'} - \frac{B'}{A'} = \frac{A}{A'}\lambda + \frac{B - B'}{A'}.$$

Für jeden möglichen Bildpunkt sei ein σ_i, $i \in \{1, \ldots, n\}$, bestimmt worden. Für eine physikalisch sinnvolle Rekonstruktion muß $\lambda' \geq 0$ sein. Damit λ' die Tiefenvariation geeignet widerspiegelt, muß der Term $\frac{\sigma - B'}{A'}$ aber auch entsprechend variieren. Führt man dann noch den Parameter $Z'_{max} > 0$ für die maximal zu rekonstruierende Tiefe ein, so ist eine sinnvolle Definition

$$\begin{aligned} A' &=_{\mathrm{def}} \; -(\frac{1}{n}\sum_{i=1}^{n} |\sigma_i|) Z'_{max} < 0, \\ B' &=_{\mathrm{def}} \; \max\{\sigma_i | i \in \{1, \ldots, n\}\} + \frac{1}{n}\sum_{i=1}^{n} |\sigma_i|. \end{aligned}$$

Es ergibt sich

$$\lambda' =_{\mathrm{def}} \frac{(\max\{\sigma_i | i \in \{1, \ldots, n\}\} - \sigma) + \frac{1}{n}\sum_{i=1}^{n} |\sigma_i|}{(\frac{1}{n}\sum_{i=1}^{n} |\sigma_i|) Z'_{max}}.$$

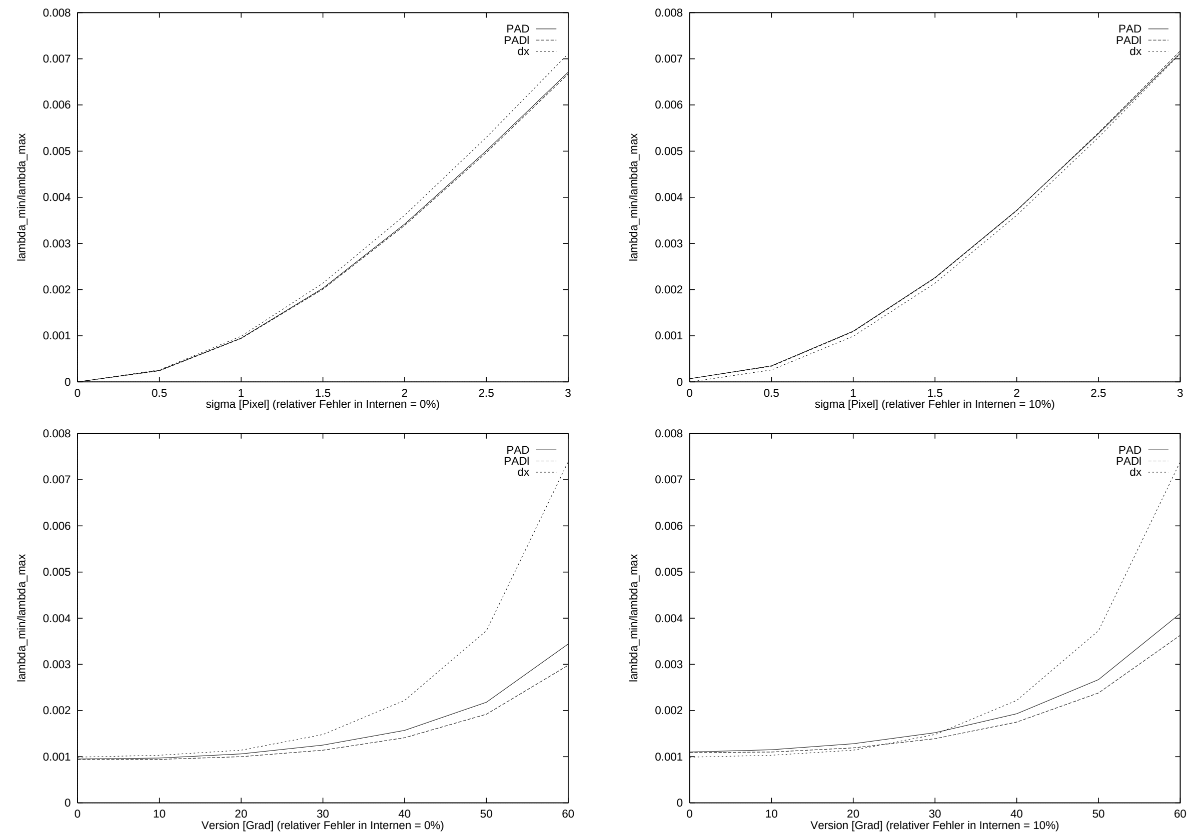

Abbildung 4.9: Das Komplanaritätsmaß $\lambda_{min}/\lambda_{max}$ in der σ-λ'-Rekonstruktion für *PAD*, *PADI* und *dx*. Rekonstruiert wurden alle 16 Punkte der oberen Ebene im Bild. Falls nicht anders angegeben, ist $\sigma_{noise} = 1$, $\gamma = 0°$, $\mu_z = 0°$. Der Radius spielt hier keine Rolle und ist Null.

Da der Zähler stark von σ abhängt, gibt λ' die Tiefenvariation gut wieder. Zum anderen gilt für alle $k \in \{1, \ldots, n\}$

$$\lambda'_k \geq 0 \quad \Leftrightarrow \quad \sigma_k \leq B' \quad \Leftrightarrow \quad \sigma_k \leq \max\{\sigma_i | i \in \{1, \ldots, n\}\} + \frac{1}{n} \sum_{i=1}^{n} |\sigma_i|,$$

$$Z'_k \leq Z'_{max} \quad \Leftrightarrow \quad \lambda'_k \geq \frac{1}{Z'_{max}} \quad \Leftrightarrow \quad \sigma_k \leq \frac{A'}{Z'_{max}} \quad \Leftrightarrow \quad \sigma_k \leq \max\{\sigma_i | i \in \{1, \ldots, n\}\},$$

wobei jeweils die letzten Ungleichungen trivialerweise erfüllt sind. Es wird also ein Punkt mit der Tiefe Z'_{max} rekonstruiert. Wegen $A' < 0$ ist $a = \frac{A}{A'} > 0$, und es gelten die entsprechenden Eigenschaften, die zu Beginn des Kapitels erläutert wurden. Rekonstruktionen mit diesen Parametern können der tatsächlichen Euklidischen Struktur bereits sehr nahe kommen, wenn man die maximale Tiefe der Szenepunkte kennt, wie Abbildung 4.10 zeigt. Durch die Kenntnis der Sehstrahlen verbleibt nur noch ein Fehler in der relativen Tiefe der Punkte, der durch die geeignete Wiedergabe der Tiefenvariation klein gehalten wird.

Als Maß für die Komplanarität der Ebene dient die Kovarianzmatrix, genauer das Verhältnis des kleinsten Eigenwertes λ_{min} (lambda_min) zum größten Eigenwert λ_{max} (lambda_max). $\lambda_{min}/\lambda_{max}$ ist zwar unabhängig von der Wahl von Z'_{max}, aber nicht von der Art der Rekonstruktion. Zum Vergleich sei angegeben, daß die unverrauschte Rekonstruktion aller drei Ebenen einen Wert von etwa 0.35 liefert. Man sieht an Abbildung 4.9, daß

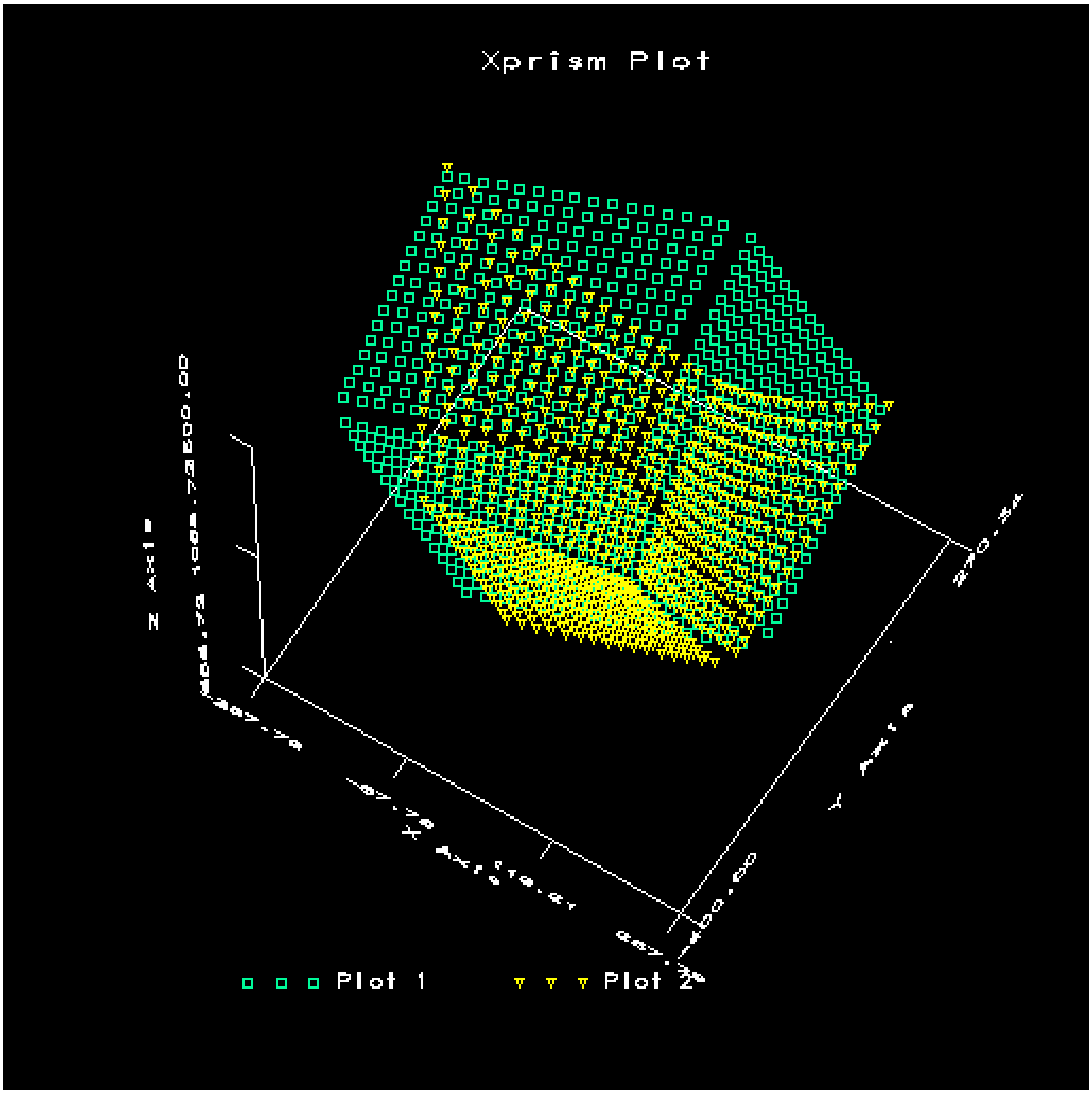

Abbildung 4.10: Original und projektive Rekonstruktion eines Kalibrierungswürfels. Mittels des dx-Verfahrens wurde eine σ-λ'-Rekonstruktion durchgeführt, wobei symmetrische Fixierung und exakte Kenntnis der Bildpunkte und internen Parameter angenommen wurde. Die Rekonstruktion (Plot 2) liegt hier dicht beim Original (Plot 1), da der Parameter Z'_{max} sehr nahe an der Grundwahrheit liegt.

die gelieferten Werte stets weit darunter, unter 0.01, liegen. Bei *PADl* wurde für die Rekonstruktion $D' = 0$ gesetzt. *PAD* und *PADl* liefern sehr ähnliche Ergebnisse und sind besser als dx, falls die internen Parameter exakt bekannt oder die Version groß, größer als 30°, ist. Im Falle von fehlerbehafteten internen Parametern ist dx aber trotz des Fehlerterms $-\omega_y x^2$ besser zur Komplanaritätsschätzung geeignet, da $|x|$ in diesen Experimenten auch sehr klein ist. Man kann schließen: **Die Vorteile von dx der Anwendung im gesamten Bildbereich und die Qualität der Leistung bei fehlerbehafteten internen Parametern und geringer Version erscheinen für einen schnellen Komplanaritätstest — nur mittels horizontaler Verschiebung! — geeignet. Lediglich bei großer Version sind, wenn möglich, *PAD* und *PADl* vorzuziehen.**

4.7.2 Euklidische Rekonstruktion

Ausgehend von einer projektiven Rekonstruktion kann man die Euklidische Struktur der Szene bestimmen. **Man beachte, daß bereits mit zwei oder drei Tiefen zu den Sehstrahlen die Transformationen und somit die Euklidische Struktur bestimmt werden kann.** Wird von einem Verfahren eine SRN bestimmt, so genügt $Z_1 \neq Z_2$ als Anforderung an die Tiefen. Wird eine AN bestimmt mit $C = 0$ bzw. $D = 0$, so dürfen (X_i, Z_i) bzw. (Y_i, Z_i) für $i \in \{1, 2, 3\}$ nicht kollinear sein, was sich leicht nachprüfen läßt.

Diese Information bezüglich des linken, rechten oder zyklopischen Koordinatensystems liegt aber in der Praxis nicht oder nur sehr ungenau vor. Realistischer ist, von einem lokalen Koordinatensystem auszugehen, wie es bei einem Kalibrierungswürfel gegeben ist. Dann gewinnt man die Euklidische Struktur, indem man die projektive Transformation zwischen der projektiven und der Euklidischen Basis bestimmt. Dabei läßt sich eine Euklidische Basis $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ in natürlicher Weise im Sinne der Stratifikation zu einer projektiven Basis erweitern:

$$\left(\begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \end{pmatrix}, \begin{pmatrix} v_{21} \\ v_{22} \\ v_{23} \end{pmatrix}, \begin{pmatrix} v_{31} \\ v_{32} \\ v_{33} \end{pmatrix} \right) \mapsto \left(\begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \\ 1 \end{pmatrix}, \begin{pmatrix} v_{21} \\ v_{22} \\ v_{23} \\ 1 \end{pmatrix}, \begin{pmatrix} v_{31} \\ v_{32} \\ v_{33} \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right)$$

Somit ist die Basis-Transformation eindeutig bestimmt. Wir geben nun zu mindestens 5 projektiv rekonstruierten Punkten ihre Szenepunkte auf dem Kalibrierungswürfel in lokalen Koordinaten an. Wendet man die daraus gewonnene Transformation $\mathcal{N}^{w \to k}$ auf alle projektiv rekonstruierten Punkte an, so erhält man eine Euklidische Repräsentation der Szene:

$$\boldsymbol{X}_w = \mathcal{N}^{w \to k} \boldsymbol{X}_k{}'$$

Dabei ist $\boldsymbol{X}_k{}'$ der bezüglich eines Kamerakoordinatensystems, aus den Pixelkoordinaten rekonstruierte Punkt und $\boldsymbol{X}_w = (X_w, Y_w, Z_w, 1)^T$ sein korrespondierender Punkt in einem Euklidischen Koordinatensystem. Die Transformation $\mathcal{N}^{w \to k}$ ergibt sich aus

$$\boldsymbol{X}_w = \tilde{T}^{w \to k} \boldsymbol{X}_k = \tilde{T}^{w \to k} \mathcal{N}^{-1} \boldsymbol{X}_k{}'.$$

$\mathcal{N}$ ist eine der durch skalierte relative oder affine Nähe induzierten Transformationen, d. h. $\mathcal{R}_i$ oder $\mathcal{A}_i$ für $i \in \{1, 2, 3\}$, $\tilde{T}^{w \to k}$ die „invertierbare Kamera-Transformationsmatrix“ von Welt- auf Kamerakoordinatensystem, $\tilde{T}^{w \to k} =_{\mathrm{def}} \begin{pmatrix} T^{k \to w} & \\ \mathbf{0}^T & 1 \end{pmatrix}$, also $\mathcal{N}^{w \to k} =_{\mathrm{def}} \tilde{T}^{w \to k} \mathcal{N}^{-1}$.

$\mathcal{N}^{w\to k}$ hat dann natürlich nur noch die Eigenschaften allgemeiner projektiver Transformationen. Die mit Hilfe von Referenzpunkten gewonnene Euklidische Struktur liefert dann aber exakte Angaben über Tiefe, Parallelität, Winkel etc. Im Vergleich zum *trans*-Verfahren, was *a-priori* kalibriert, handelt es sich bei diesem Vorgehen um eine *a-posteriori-Kalibrierung.*

Von besonderem Interesse ist neben dem internen Vergleich der 1-Punkt-Rekonstruktionsverfahren die Leistung bezüglich des klassischen *trans*-Verfahrens. Wir bestimmten dazu den relativen Fehler $\frac{\|\boldsymbol{X}-\hat{\boldsymbol{X}}\|_2}{\|\boldsymbol{X}\|_2}$ zwischen realem Punkt $\boldsymbol{X}$ und seiner Rekonstruktion $\hat{\boldsymbol{X}}$. Um es vorwegzunehmen: **In allen hier getesteten Situationen hat sich das *trans*-Verfahren in Leistung und Geschwindigkeit den *PAD*- und *dx*-Verfahren überlegen gezeigt.** Dabei kommt hier als zusätzliche Eingabe die Anzahl der zur Rekonstruktion genutzten Referenzpunkte hinzu, die beim *trans*-Verfahren zur Bestimmung von $T^{k\to w}$, bei den *PAD*- und *dx*-Verfahren zur Bestimmung von $\mathcal{N}^{w\to k}$ notwendig sind. Auch die Vermutung, daß das klassische *trans*-Verfahren bei wenigen Referenzpunkten versagt, da es 6 Punkte benötigt, gegenüber 5 Punkten bei Festlegung der Transformation von 3D auf 3D, konnte nicht bestätigt werden. Im Gegenteil verbietet der Radius der *PAD*-Verfahren eine zu kleine Auswahl von Referenzpunkten. Neben *trans* testen wir *PAD-sx*, *PADl-sx* und *dxp*. Hier hat sich, wie zuvor, gezeigt, daß *PAD* mit *PAD-sx*, *PADl* mit *PADl-sx* und *dx* mit *dxp* vergleichbar ist, und man zumindest auf die Skalierungsfaktoren verzichten kann. Andererseits sind die Leistungen von *PAD-sx0* und *PADl-sx0* aufgrund der Numerik von $\frac{x_p}{\bar{y}}$ zu schlecht.

Abbildung 4.11 zeigt den relativen Fehler in Abhängigkeit des Bildpunktrauschens bei unterschiedlicher Zahl von Referenzpunkten und unterschiedlichen Fehlern in den internen Parametern. Es wird von der einfachsten Systemgeometrie mit $\gamma = 0°$ und $\mu_z = 0°$ ausgegangen. Man sieht, daß die Schere zwischen *trans* und den anderen Verfahren unabhängig von der Anzahl der Referenzpunkte ist. Zudem ist zu erkennen, daß der Einfluß der Fehler in den internen Parametern, bis auf den Fall sehr geringen Bildpunktrauschens, klein ist. Dies liegt daran, daß bei $\gamma = 0°$, also bei symmetrischer Fixierung, die vertikale Disparität sehr gering ist, nämlich $\dot{y} = -\omega_y xy$. Damit wirken sich auch Fehler auf den Term mit der vertikalen Disparität kaum aus, der Term mit der horizontalen Disparität ist aber unabhängig von der internen Parametern (siehe *dxp*)! *PAD-sx* und *PADl-sx* sind wiederum nahezu identisch. Ab einem Bildpunktrauschen von etwa 1.5 is *dxp* sogar besser als *PAD-sx* und *PADl-sx*. Hierfür ist meiner Meinung nach auch an dieser Stelle das schlechte Fehlerverhalten von $\frac{\bar{x}}{\bar{y}}$ verantwortlich.

Die Unempfindlichkeit der *PAD*-Verfahren gegenüber Fehlern in den internen Parametern, falls $\gamma = 0°$, zeigt sich auch in Abbildung 4.12. Je nach Anzahl der Referenzpunkte sind *PAD-sx* und *PADl-sx* erst bei 20% bis 25% (!) Fehler schlechter als *dxp*. Dies ändert sich bei nicht verschwindender Version, wo der Einfluß der vertikalen Disparität und damit der Fehler über $\frac{\bar{x}}{\bar{y}}$ wächst. *PAD-sx* und *PADl-sx* haben einen bis zu vierfachen Fehler und sind bereits ab $\gamma = 20°$ schlechter als *dxp*, obwohl *dxp* dann eine stark fehlerbehaftete SRN liefert. Die schlechte Leistung der *PAD*-Verfahren äußert sich auch in der starken Empfindlichkeit gegenüber Torsion und dem hohen erforderlichen Radius. Trotz der Unabhängigkeit von *dxp* von jeder a-priori-Kalibrierungsinformation ist auch hier *trans* stets überlegen.

trans liefert trotz seiner Allgemeinheit auch in dieser geometrisch eingeschränkten Si-

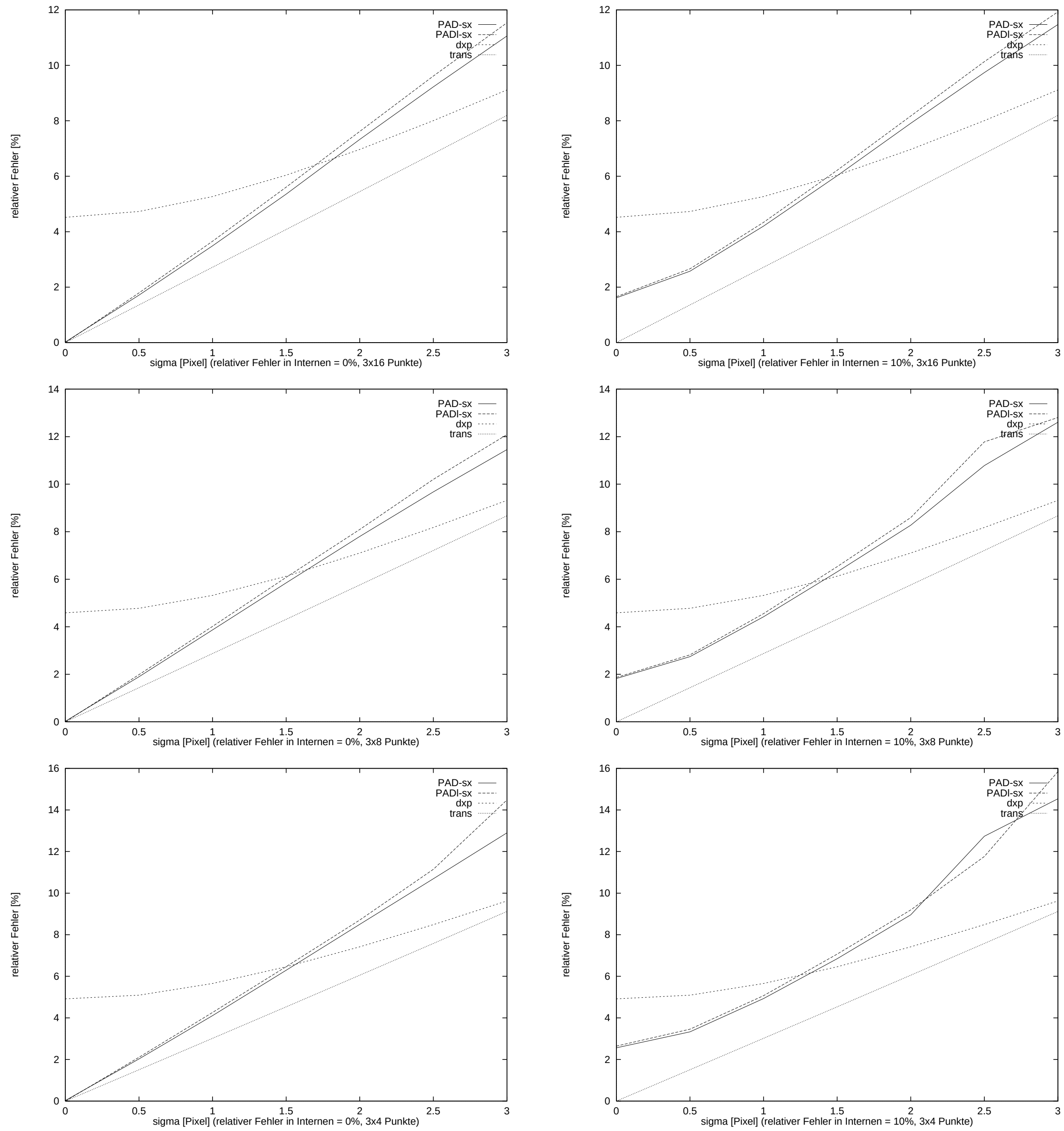

Abbildung 4.11: Relativer Fehler in der Euklidischen Rekonstruktion für *PAD-sx*, *PADl-sx*, *dxp* und *trans* in Abhängigkeit des Bildpunktrauschens. Es ist stets $\gamma = 0°$, $\mu_z = 0°$, $r = 50$.

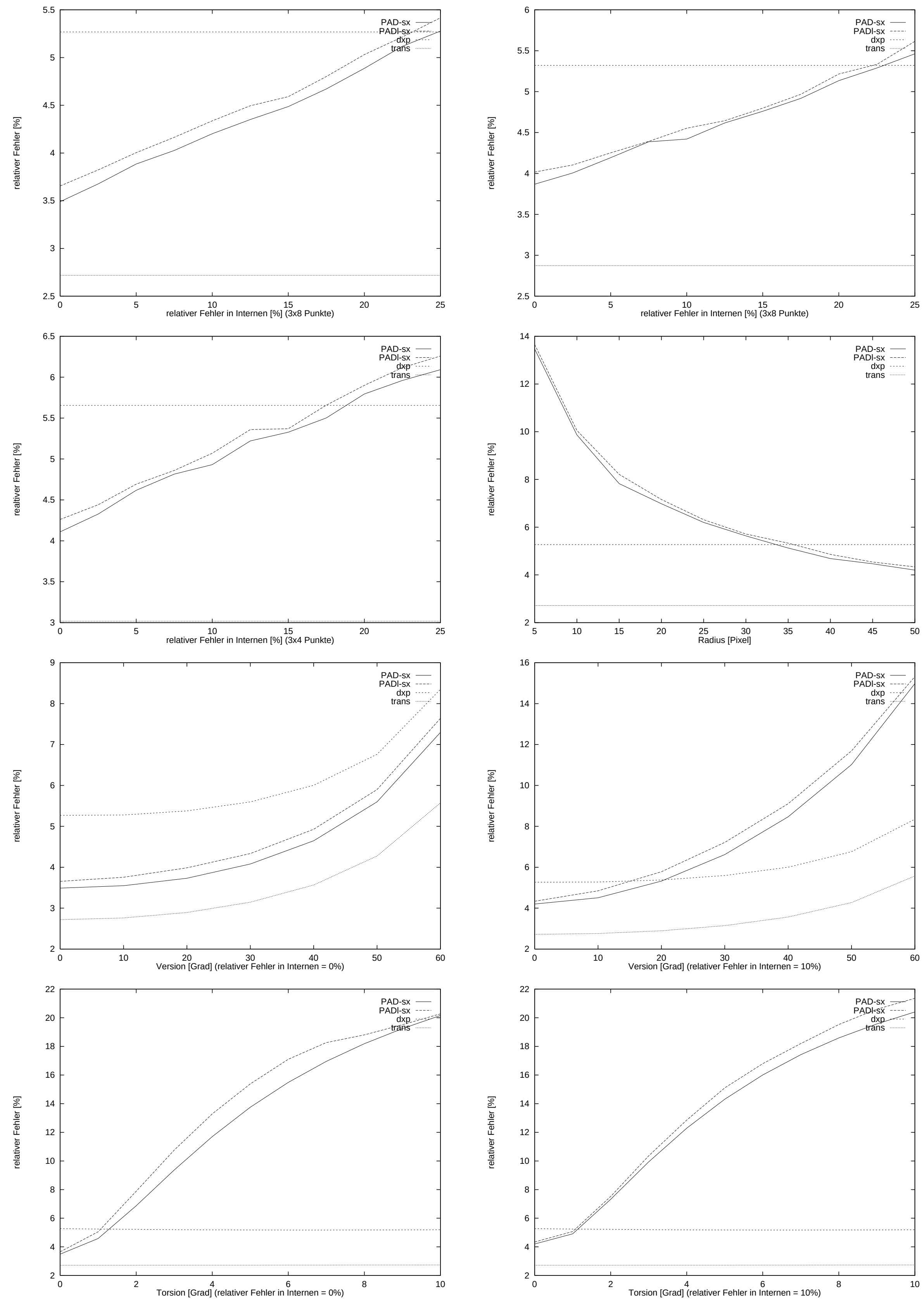

Abbildung 4.12: Relativer Fehler in der Euklidischen Rekonstruktion für *PAD-sx*, *PADl-sx*, *dxp* und *trans*. Falls nicht anders angegeben, ist $\sigma_{noise} = 1$, $\Delta\mathrm{Int} = 10\%$, $\gamma = 0°$, $\mu_z = 0°$, $r = 50$, und es werden sämtliche 48 Punkte als Referenzpunkte verwendet.

tuation immer die besten Ergebnisse. Warum ist dies so? Eine detailierte Fehleranalyse würde den Rahmen dieser Arbeit sprengen. Ich möchte jedoch starke Indizien vortragen, die in der Vorgehensweise liegen. Sowohl die SRN/AN-Verfahren als auch das *trans*-Verfahren lassen sich grob in zwei Schritte einteilen:

1. Schritt: Kalibrierung

trans: Bestimmung der Transformationsmatrizen aus den Referenzpunkten und ihren Bildpunkten

SRN/AN: Bestimmung der um die Euklidische Koordinatentransformation erweiterte Matrix der Relieftransformation $\mathcal{N}^{w \to k}$ aus den Referenzpunkten und ihren Bildpunkten *und* der affinen Nähe

2. Schritt: Rekonstruktion

trans: Bestimmung des 3D-Punktes aus $T^{l \to w}$, $T^{r \to w}$ und den korrespondierenden Bildpunkten durch Triangulation

SRN/AN: Bestimmung des 3D-Punktes aus $\mathcal{N}^{w \to k}$ und den korrespondierenden Bildpunkten *und* der affinen Nähe

Ich argumentiere nun, daß beide Schritte nahezu identisch sind bis auf die zusätzliche Nutzung einer „projektiven Tiefe“ der SRN/AN-Verfahren. Diesen Verfahren geht sozusagen ein fehlerbehafteter 0. Schritt voraus, der zwar eine 3D-Rekonstruktion ermöglicht, aber zur Bestimmung der Euklidischen Hülle bei der Verwendung von Referenzpunkten umgangen werden kann, wie es *trans* durch Benutzung von *zwei* Matrizen tut. Dadurch hat man im 2. Schritt einen Fehler, der einer Matrixanwendung wie bei den SRN/AN-Verfahren entspricht, ohne zuvor Tiefeninformation gewinnen zu müssen. K. Daniilidis wies zudem darauf hin, daß die Ermittlung von Tiefe aus Disparitäten über die Parameter der relativen Bewegung immer fehlerbehafteter ist als die Ermittlung über Transformationsmatrizen, bei deren Berechnung die exakte Position der Referenzpunkte eingegangen ist. Hier zeigt sich die Überlegenheit spezieller SfS-Verfahren gegenüber den allgemeineren SfM-Algorithmen, bei denen solche Referenzpunkte nicht zur Verfügung stehen.

Abschließend eine Bemerkung zum Berechnungsaufwand: Während die projektive Struktur unnachahmlich schnell aus den korrespondierenden Bildpunkten zu gewinnen ist, ist der Aufwand zur Bestimmung der 15 Parameter der Matrix $\mathcal{N}^{w \to k}$ relativ hoch. Die zweifache Berechnung der 11 Parameter einer Projektionsmatrix ist auch hier günstiger. Mit den hier genannten Einschränkungen sind aber die 1-Punkt-SRN/AN-Verfahren eine einfache und schnelle Informationsquelle für die Gewinnung projektiver Invarianten, die auch in Verbindung mit anderen SfM-Algorithmen genutzt werden können.

4.7.3 Zusammenfassung

Als wichtigste allgemeine Erkenntnis wollen wir festhalten: **Die Bestimmung einer SRN oder einer AN ermöglicht eine äußerst schnelle projektive Rekonstruktion der Szene, die Aussagen über die Komplanarität und im Falle der SRN**

über die Tiefenordnung zuläßt. Zur Euklidischen Rekonstruktion ist diese Information nicht unbedingt notwendig und das Verfahren mit den Kamera-Transformationsmatrizen exakter und schneller. Man sollte also in jedem Fall nur die unbedingt benötigten Informationen ermitteln. Im einzelnen hat sich ergeben:

- Bei den *PAD*-Verfahren kann auf die Kenntnis der Skalierungsfaktoren verzichtet werden. Die Verfahren, die nur mit y_0 arbeiten, haben ein zu schlechtes Fehlerverhalten gezeigt, sowohl bei der projektiven als auch bei der Euklidischen Rekonstruktion.

- Zur Gewinnung der Tiefenordnung sollte lediglich *PAD* verwendet werden.

- Zur projektiven Rekonstruktion sind die *PAD*-Verfahren, wo einsetzbar, den dx-Verfahren vorzuziehen. *PAD* und *PADl* sind hier äquivalent zu gebrauchen.

- Für einen Komplanaritätstest können zumeist die Vorteile der dx-Verfahren ohne große qualitative Einschränkungen genutzt werden. Man beachte die unterschiedlichen Testwerte für verschiedene Rekonstruktionsarten.

- Zur Euklidischen Rekonstruktion mit Referenzpunkten ist durchgehend das *trans*-Verfahren zu empfehlen.

- Die σ-λ'-Rekonstruktion liefert mittels des dx-Verfahrens schnell ein grobes Bild der Szene, falls die internen Kameraparameter bekannt sind.

Kapitel 5

Diskussion

Am Ausgangspunkt dieser Arbeit stand die Frage nach (partiell, total) unkalibrierter Strukturerschließung mittels eines Stereosystems. Um dieses Problem einordnen zu können, haben wir eine Einführung in die klassischen SfS-Konzepte gegeben. Dies enthielt die Beschreibung der prinzipiellen Vorgehensweise beim SfS-Prozeß, des Kameramodells, der Modellierung der Welt und der Bildebene sowie der Hilfsmittel für eine kalibrierte Szenenrekonstruktion, der perspektivischen Transformationsmatrix, und für eine unkalibrierte, der Fundamentalmatrix. Dabei gingen wir besonders auf die sogenannte Stratifikation des Raumes ein, in deren Rahmen die Strukturerschließungsproblematik erst richtig zu verstehen ist, und auf die zusätzlichen Eigenschaften, die durch Nutzung orientierter projektiver Geometrie zu gewinnen sind.

Im Zusammenhang mit aktiven Systemen haben wir dann binokulare fixierte Stereosysteme vorgestellt, die weniger Freiheitsgrade als allgemeine Systeme haben und einfacher zu beschreiben sind. Vor allem interessierte uns die Möglichkeit, die Disparität als Verschiebungsrate darzustellen. Dies sollte dazu dienen, die Beziehung zwischen der Strukturinformation, konkret der Tiefe der Szenepunkte, und den Parametern der relativen Bewegung, hier in kontinuierlicher Form, zu simplifizieren. Wir haben analytisch und experimentell gezeigt, daß bei Einführung eines zyklopischen Auges diese Darstellung fast fehlerfrei und für praktische Zwecke geeignet ist. Diese Darstellung konnte jedoch nicht durch geeignete Wahl von linken und rechten Bildpunktkoordinaten umgangen werden, da sich hier zum Teil große relative Fehler ergaben.

Es wurde ausgeführt, in welchen Formen fixierte Systeme auftreten, wie sie im einzelnen realisiert werden können und welche Vereinfachungen und Bedingungen an das System gefolgert werden können. Über das unkalibrierte Vorgehen und die Nutzung der Disparität und ihrer Darstellung zur Gewinnung von Strukturinformation findet man einen natürlichen Zugang zur SfM-Problematik. Wir haben gesehen, wie allgemeine SfM-Verfahren, die z. B. mit der essentiellen Matrix arbeiten, Eingang in das Stereoproblem finden. In diesem Zusammenhang wurden kontinuierliche Analoga, wie sie in der Bewegungsanalyse verwendet werden, und Darstellungen der essentiellen sowie der Fundamentalmatrix im Falle einer fixierten Geometrie hergeleitet. Auch hier kann die Strukturgewinnung in vereinfachter Form erfolgen.

Wir haben uns dann darauf konzentriert, die Verschiebungsrate zur Gewinnung von Strukturinformation zu nutzen. Wir haben gesehen, wie einfache Funktionen der Nähe (SRN, AN), wie sie in der Verschiebungsrate auftauchen, bereits „projektive Tiefeninformation“ enthalten. Neben einer projektiven Rekonstruktion (über AN) kann zum Teil eine Rekonstruktion bis auf eine Relieftransformation (über SRN) gewonnen werden. Damit stehen in diesem Fall nicht nur Informationen über Inzidenz, Kollinearität und Komplanarität zur Verfügung, sondern auch über die Tiefenordnung der Szenepunkte. Die Auffassung des Raumes als orientiert projektiv lieferte zusätzliche Eigenschaften von Objekten wie Konvexität. Als wesentliche neue Erkenntnis kam hinzu, daß Eigenschaften wie Komplanarität, Konvexität und Tiefenordnung auch dann invariant bleiben, wenn man von unbekannten internen Parametern, also nur den Pixelkoordinaten der Bildpunkte, ausgeht.

Wir haben dann bekannte SRN-Berechnungsmodelle vorgestellt und diese erweitert, vor allem um Verfahren bei unbekannten internen Parametern und AN-Modelle. Besondere Beachtung haben die Verfahren gefunden, die aus einem einzigen korrespondierenden Punktepaar rekonstruierten. Dabei konnten wir nur unter Kenntnis des Bildzentrums eine Reliefrekonstruktion, unter Kenntnis seiner vertikalen Komponente eine projektive Rekonstruktion erreichen, die man nur unter noch stärkeren Einschränkungen total unkalibriert gewinnt.

In Experimenten haben wir die hergeleiteten Eigenschaften wiedergefunden. Wir testeten dabei die vorgestellten 1-Punkt-Rekonstruktionsverfahren bezüglich ihrer Robustheit. In der projektiven Rekonstruktion interessierten wir uns für die Wiedergabe der Tiefenordnung und der Komplanarität. Diese Informationen konnten sehr schnell und einfach, zum ersten nur unter Kenntnis des Bildzentrums mit dem PAD-Verfahren, zum zweiten auch total unkalibriert mit der horizontalen Disparität, gewonnen werden. Für eine Euklidische Rekonstruktion führten wir eine a-posteriori-Kalibrierung durch und verglichen den relativen Fehler in der gewonnenen Koordinatendarstellung mit einem klassischen Verfahren. Hier war das klassische Verfahren in Genauigkeit und Geschwindigkeit stets überlegen.

Wir fassen den Beitrag dieser Arbeit zusammen:

- Erläuterung der Nutzbarkeit orientierter projektiver Geometrie im Rahmen der Stratifikation des Raumes

- analytische und experimentelle Ergebnisse zur Darstellung der Disparität als Verschiebungsrate im Falle einer fixierten binokularen Stereogeometrie

- Darstellung unterschiedlicher Fixierungsbedingungen bezüglich ihrer Realisierung, ihrer Beziehung untereinander und ihrer Auswirkung auf die Darstellung der Disparität

- Herleitung der essentiellen Matrix, der Fundamentalmatrix und ihrer kontinuierlichen Analoga im Falle einer fixierten Stereogeometrie

- ausführliche, ergänzte Zusammenstellung der Rekonstruktionsarten aus SRN und AN und ihrer Eigenschaften mit neuen Erkenntnissen zur Orientierungserhaltung, zum Fixgebilde und zur Rekonstruktion mit unbekannten internen Parametern

- Präsentation von Varianten des PAD-Modells: diskrete Versionen, eine Approximation, eine Erweiterung

- Erweiterung der vorgestellten SRN/AN-Verfahren um unbekannte interne Parameter

- Zusammenstellung der vorgestellten SRN/AN-Verfahren im Vergleich mit klassischen im Hinblick auf ihre Einsatzmöglichkeit, inhärente Fehler und gelieferte Rekonstruktion

- Experimente auf künstlichen Daten für die 1-Punkt-Verfahren bezüglich ihrer Robustheit gegenüber Bildpunktrauschen und Fehlern in der verbleibenden Kalibrierungsinformation: Gewinnung der projektiven Struktur, besonders Erhaltung der Komplanarität und der Tiefenordnung, und Gewinnung der Euklidischen Struktur im Vergleich mit einem klassischen Verfahren mit Kalibrierung

- Präsentation einer Möglichkeit zur Festsetzung der Parameter der λ'-Rekonstruktion und damit einer geeigneten Visualisierung der projektiven Rekonstruktion

Wir halten als wichtigste Ergebnisse fest:

- Die Nutzung orientierter projektiver Geometrie ermöglicht die Gewinnung zusätzlicher Invarianten der projektiven Transformationen der Szene, nämlich Konvexität, und den Ausschluß bestimmter Konfigurationen.

- Im Falle einer einfachen Stereogeometrie **(SIS)** läßt sich die Disparität in simplifizierter Form als Verschiebungsrate schreiben, falls man ein zyklopisches Koordinatensystem benutzt. In diesem Fall ist die Darstellung fast fehlerfrei, bei Verwendung linker und rechter Bildpunktkoordinaten im allgemeinen stark fehlerbehaftet.

- Rekonstruiert man aus SRN bzw. AN mit Bildpunkten in Pixelkoordinaten, so bleiben die Tiefenordnung, die Orientierungserhaltung und die allgemeinen projektiven Invarianzen bestehen.

- Im Falle einer einfachen Stereogeometrie **(SIS)** erhält man über den PAD-Ansatz eine Rekonstruktion bis auf eine Relieftransformation, falls das Bildzentrum bekannt ist, bis auf eine allgemeine projektive Transformation, falls nur seine vertikale Komponente bekannt ist, bis auf eine fehlerbehaftete Relieftransformation ohne Kenntnis der internen Parameter, falls symmetrisch fixiert wird.

- Im Falle einer einfachen Stereogeometrie **(SIS)** läßt sich schnell und einfach eine projektive Rekonstruktion mittels des PAD-Ansatzes gewinnen. Für die Gewinnung der Tiefenordnung benötigt man das Bildzentrum, für einen Komplanaritätstest kann man eventuell total unkalibriert vorgehen.

- Auch im Falle einer einfachen Stereogeometrie **(SIS)** ist das klassische Verfahren mit a-priori-Kalibrierung den 1-Punkt-Verfahren in der Gewinnung der Euklidischen Struktur immer überlegen.

Abbildungsverzeichnis

Tabellenverzeichnis

Literaturverzeichnis

[Alo93] Y. Aloimonos, editor. *Active Perception.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.

[Baj88] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76:996–1005, 1988.

[Car96] S. Carlsson. Combinatorial geometry for shape representation and indexing. In *ECCV-96 Workshop on Object Recognition*, 1996.

[CSS91] T.S. Collett, U. Schwartz, and E.C. Sobel. The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, 20:733–754, 1991.

[DA89] U.R. Dhond and J.K. Aggarwal. Structure from stereo — a review. *IEEE Trans. Systems, Man, and Cybernetics*, 19:1489–1510, 1989.

[Dan92] K. Daniilidis. Zur Fehlerempfindlichkeit in der Ermittlung von Objektbeschreibungen und relativen Bewegungen aus monokularen Bildfolgen. Dissertation, Fakultät für Informatik, Universität Karlsruhe (TH), 1992.

[Dan95] K. Daniilidis. Attentive visual motion processing: computations in the log-polar plane. *Computing Suppl.*, 11:1–20, 1995. Special Issue on Theoretical Foundations of Computer Vision.

[Fau92] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *Proc. Second European Conference on Computer Vision*, pages 563–578. Santa Margerita, Italy, May 23-26, G. Sandini (Ed.), Lecture Notes in Computer Science 588, Springer-Verlag, Berlin et al., 1992.

[Fau93] O. Faugeras. *Three-dimensional Computer Vision.* MIT-Press, Cambridge, MA, 1993.

[Fau95] O. Faugeras. Stratification of three-dimensional vision: projective, affine, and metric representations. *Journal Opt. Soc. Am. A*, 12:465–484, 1995.

[Fis86] G. Fischer. *Lineare Algebra.* Vieweg, 1986.

[Fis92] G. Fischer. *Analytische Geometrie.* Vieweg, 1992.

[GPFM96] J. Garding, J. Porrill, J.P. Frisby, and J.E.W. Mayhew. Uncalibrated relief reconstruction and model alignment from binocular disparities. In *Proc. Fourth European Conference on Computer Vision*, pages 427–438, Cambridge, UK, April 14-18, B. Buxton (Ed.), Springer LNCS 1064, 1996.

[GPMF95] J. Garding, J. Porrill, J.E.W. Mayhew, and J.P. Frisby. Steropsis, vertical disparity and relief transformations. *Vision Research*, 35:703–722, 1995.

[Hor86] B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[Jäh91] B. Jähne. *Digitale Bildverarbeitung*. Springer Verlag, Berlin et al., 1991.

[Joh91] E.B. Johnston. Systematic distortions of shape from stereopsis. *Visual Research*, 31:1351–1360, 1991.

[KvD76] J.J. Koenderink and A.J. van Doorn. Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, 21:29–35, 1976.

[LF96a] S. Laveau and O. Faugeras. Oriented projective geometry in computer vision. In *Proc. Fourth European Conference on Computer Vision*, pages 147–156, Cambridge, UK, April 14-18, B. Buxton (Ed.), Springer LNCS 1064, 1996.

[LF96b] Q.-T. Luong and O.D. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17:43–75, 1996.

[LSS94] L. Liu, S.B. Stevenson, and C.M. Schor. A polar coordinate system for describing binocular disparity. *Visual Research*, 34:1205–1222, 1994.

[LV94] Q. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. In *Proc. Third European Conference on Computer Vision*, pages 589–599. Stockholm, Sweden, May 2-6, J.O. Eklundh (Ed.), Springer LNCS 800, 1994.

[Mar82] D. Marr. *Vision*. W. H. Freeman and Co., San Francisco, 1982.

[Som95] G. Sommer. Verhaltensbasierter Entwurf technischer visueller Systeme. *Künstliche Intelligenz*, 5(3), 1995.

[Sto91] J. Stolfi. *Oriented projective geometry*. Academic Press, 1991.

[VT95] B. Völpel and W.M. Theimer. Localization uncertainty in area-based stereo algorithms. *IEEE Trans. Systems, Man, and Cybernetics*, 25:1628–1634, 1995.

[Wei90] D. Weinshall. Qualitative depth from stereo, with applications. *Computer Vision, Graphics, and Image Processing*, 49:222–241, 1990.

Erklärung

Hiermit versichere ich, daß ich die vorliegende Arbeit selbständig verfaßt und ausschließlich die angegebenen Hilfsmittel und Quellen verwendet habe.

Kiel, den