

PRIP-TR-43

13. September 1996

Rektifizierung oder direkte Tiefenberechnung?
Untersuchungen zur digitalen Objektvermessung
mit dem Stereoverfahren¹

Problemstellung der Diplomarbeit

Norbert Brändle

Abstract

Das Stereoverfahren ist als eine Methode zur Vermessung von dreidimensionalen Strukturen aus zweidimensionalen Bildern schon lange bekannt. Da die automatische, vom Rechner durchgeführte Stereoauswertung immer noch zahlreiche Probleme bereitet, insbesondere wenn es um die Bewältigung allgemeiner Szenen im dreidimensionalen Raum geht, sind Stereoalgorithmen weiterhin Gegenstand intensiver Forschungen. In den einschlägigen Publikationen wird sehr oft eine einfache Kameraanordnung, der sogenannte Normalfall, angenommen. Da die geometrischen Bedingungen des Normalfalls in realen Aufnahmesystemen nicht zutreffen, werden durch perspektivische Rektifizierung Bilder erzeugt, die diese Bedingungen erfüllen. Eine Alternative ist der Verzicht auf die Vorteile der einfachen Geometrie und die Stereoauswertung auf den Originalbildern unter Berücksichtigung der epipolaren Geometrie, was einen erhöhten Rechenaufwand bedeutet und daher oft nicht in Erwägung gezogen wird. Dieser Bericht behandelt das geometrische Kameramodell, die epipolare Geometrie einer Stereoanordnung, das Korrespondenzproblem, den Normalfall der Stereoauswertung und die perspektivische Rektifizierung von Stereobildern. Abschließend werden die Ziele der Diplomarbeit vorgestellt.

¹Diese Arbeit wird vom FWF unter der Projektnummer P9119-SPR gefördert

Inhaltsverzeichnis

1	Einführung	2
2	Geometrische Modellierung einer Kamera	2
2.1	Zentralprojektion im Raum	2
2.2	Linsenverzeichnung	4
2.3	Digitales Bild	4
2.4	Kamerakalibrierung	5
3	Grundlagen zur Stereoauswertung	6
3.1	Das Stereoproblem	6
3.2	Epipolare Geometrie	7
3.2.1	Epipolare Linien	7
3.2.2	Berechnung der epipolaren Linien	8
3.2.3	Der Normalfall der Stereoauswertung	9
3.2.4	Geometrische perspektivische Rektifizierung	10
3.3	Das Korrespondenzproblem	11
3.3.1	Flächenbasierte Stereoverfahren	11
3.3.2	Merkmalsbasierte Stereoverfahren	11
3.4	Das Rekonstruktionsproblem	12
3.4.1	Normalfall	12
3.4.2	Allgemeine Kameraanordnung	14
4	Problemstellung der Diplomarbeit	14
5	Zusammenfassung	15

1 Einführung

Im Rahmen des Projektes „Bildhafte Erfassung von archäologischen Fundgegenständen“ soll eine rechnerunterstützte Vermessung der dreidimensionalen Struktur von archäologischen Scherben erfolgen, um den Archäologen von Routinearbeit zu befreien: Bis zu 40.000 Scherben können an einem Ausgrabungsort gefunden werden und ein Großteil davon muß bisher manuell vermessen werden. Die Scherben werden fotografiert, vermessen, gezeichnet und katalogisiert. Durch die Automatisierung dieser Prozedur ergibt sich zusätzlich eine *objektive* Vermessung der Fundgegenstände, die individuelle Unterschiede bei Zeichnungen verschiedener Archäologen vermeidet. Überblicke über das Archäologieprojekt finden sich z.B. in [SMD93], [SMD94], [SM95] und [DKKF93]. Verschiedene Verfahren zur dreidimensionalen Rekonstruktion finden sich z.B. in [HS93], [Kra92],[Shi87] und [Hor86].

Das Rekonstruktionsverfahren, das in dieser Diplomarbeit untersucht werden soll, ist das Stereoverfahren, das schon seit Jahren Gegenstand intensiver wissenschaftlicher Forschung ist. Das Grundprinzip der Stereoskopie ist simpel und findet sich auch im menschlichen visuellen System ([VV90], [Mar82]). In der wissenschaftlichen Literatur wird sehr oft eine einfache Kamerageometrie, der sogenannte Normalfall angenommen. Im Rahmen der Diplomarbeit soll keine Einschränkung der Kamerageometrie auf diese einfache Kamerageometrie gemacht werden. In diesem Report wird in Kapitel 2 die geometrische Modellierung einer Kamera beschrieben. Kapitel 3 befaßt sich mit der geometrischen Beschreibung des Stereoproblems für zwei Kameras. In Kapitel 4 werden die Ziele meiner Diplomarbeit vorgestellt.

2 Geometrische Modellierung einer Kamera

In diesem Kapitel wird das Entstehen eines Bildes mathematisch modelliert.

2.1 Zentralprojektion im Raum

Das aufzunehmende Objekt liegt in einem dreidimensionalen kartesischen *Objekt-* oder *Weltkoordinatensystem* mit den Achsenbezeichnungen x , y , z . Die Kamera hat ihr eigenes dreidimensionales *Kamerakoordinatensystem*, das zum Weltkoordinatensystem durch eine Translation und eine Drehung in Beziehung steht. Ein Punkt (x, y, z) im Weltkoordinatensystem wird durch den Punkt (x', y', z') im Kamerakoordinatensystem ausgedrückt, wobei

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R(\omega, \phi, \kappa) \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix}. \quad (1)$$

Der Punkt $(x_0, y_0, z_0)^T$ ist das Projektionszentrum, $R(\omega, \phi, \kappa)$ ist eine Rotationsmatrix, die durch die drei Winkel ω (Rotation um x -Achse), ϕ (Rotation um y -Achse) und κ

(Rotation um z -Achse) gegeben ist:

$$R(\omega, \phi, \kappa) = \begin{pmatrix} \cos \phi \cos \kappa & -\cos \phi \sin \kappa & \sin \phi \\ \cos \omega \sin \kappa + \sin \omega \sin \phi \cos \kappa & \cos \omega \cos \kappa - \sin \omega \sin \phi \sin \kappa & -\sin \omega \cos \phi \\ \sin \omega \sin \kappa - \cos \omega \sin \phi \cos \kappa & \sin \omega \cos \kappa + \cos \omega \sin \phi \sin \kappa & \cos \omega \cos \phi \end{pmatrix}. \quad (2)$$

Um Beziehungen, die die Elemente der Rotationsmatrix R enthalten, ohne lange trigonometrische Ausdrücke darstellen zu können, wird folgende Konvention eingeführt:

$$R(\omega, \phi, \kappa) = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}.$$

Durch Gleichung (1) haben wir eine Darstellung eines 3D Weltpunktes P in Kamerakoordinaten Q . Angenommen, die Distanz von der Bildebene zum Projektionszentrum beträgt f , so wird der Bezug zwischen dem Kamerakoordinatensystem (x', y', z') zum *Bildkoordinatensystem* (u, v) durch eine *perspektivische Projektion* ausgedrückt:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{f}{z'} \begin{pmatrix} x' \\ y' \end{pmatrix}. \quad (3)$$

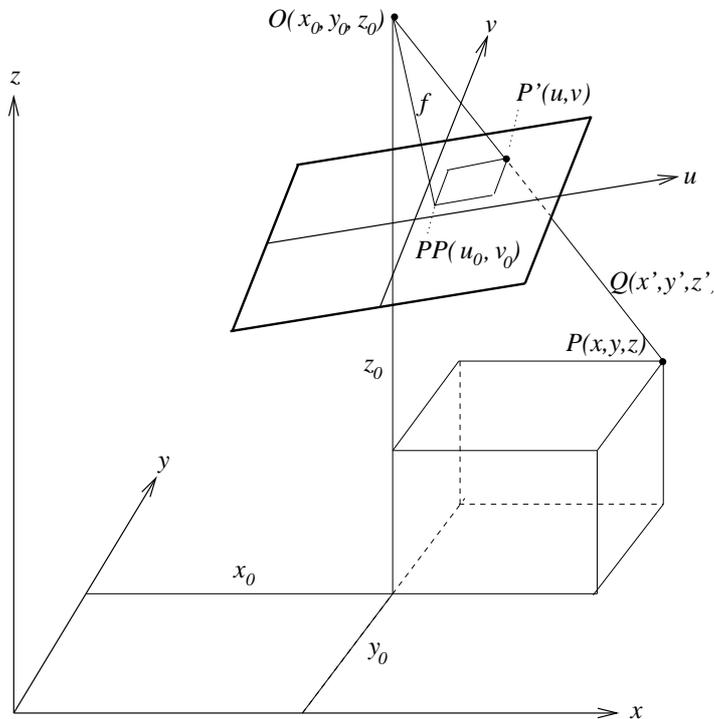


Abbildung 1: Beziehung zwischen Bild- und Objektkoordinaten

In der photogrammetrischen Literatur wird f *Kammerkonstante* genannt und ist mit der Brennweite der Linse verwandt. Da aber die Linse je nach Fokussierung eine unterschiedliche Distanz zur Bildebene hat, muß der effektive Wert durch eine Kamerakalibrierungsprozedur bestimmt werden.

Messungen werden in der Bildebene durchgeführt. Der *Bildhauptpunkt* ist der Punkt, an dem die *optische Achse* die Bildebene schneidet und ist wegen mechanischer Fehler der Optik nicht notwendigerweise $(0, 0)$. Liegt der Bildhauptpunkt an der Stelle (u_0, v_0) , so werden die Koordinaten eines projizierten Punktes folgendermaßen berechnet:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + \frac{f}{z'} \begin{pmatrix} x' \\ y' \end{pmatrix}. \quad (4)$$

Abbildung 1 zeigt die Zusammenhänge zwischen einem Punkt P im Weltkoordinatensystem und dem Punkt P' im Bildkoordinatensystem graphisch.

2.2 Linsenverzeichnung

Die bisherigen Ausführungen haben ein lineares Modell beschrieben, das die Kamera nur ungenügend modelliert. Durch Ungenauigkeiten bei realen optischen Linsen treten ortsvariante Abweichungen vom Modell der perspektivischen Projektion auf. Bei der *radialen Linsenverzeichnung* wird der Punkt besonders am Bildrand verzerrt abgebildet. Ein mathematisches Modell ist zum Beispiel in [Tsa86] beschrieben:

$$\begin{aligned} u_d &= u_u - D_u \\ v_d &= v_u - D_v \end{aligned} \quad (5)$$

wobei (u_d, v_d) die verzeichneten Bildkoordinaten auf der Bildebene sind und (u_u, v_u) die idealisierten unverzeichneten Bildkoordinaten sind, die gemäß Gleichung (4) berechnet werden. Der Fehler D_u und D_v zwischen den verzeichneten und unverzeichneten Bildkoordinaten ist eine vom Abstand zum Bildhauptpunkt abhängige Polynomfunktion mit den Koeffizienten κ_1 und κ_2 :

$$\begin{aligned} D_u &= u_d(\kappa_1 r^2 + \kappa_2 r^4) \\ D_v &= v_d(\kappa_1 r^2 + \kappa_2 r^4) \\ r &= \sqrt{u_d^2 + v_d^2}. \end{aligned}$$

Die Koeffizienten κ_1 und κ_2 müssen durch eine Kamerakalibrierungsprozedur bestimmt werden.

2.3 Digitales Bild

Bis jetzt wurde die Abbildung eines Punktes im Weltkoordinatensystem auf einen Punkt im Bildkoordinatensystem beschrieben. Die Bildebene entspricht der Ebene der

Halbleitersensoren in einer CCD-Kamera (siehe z. B. [Tsa86]). Die diskreten digitalen Bilder werden durch einen Framegrabber erzeugt, weshalb noch folgende Transformation zu berücksichtigen ist:

$$\begin{aligned} u_f &= s_u \frac{N_{fu}}{N_{cu}} \frac{1}{d_u} u_d \\ v_f &= \frac{1}{d_v} v_d, \end{aligned} \quad (6)$$

wobei

- (u_f, v_f) : Zeilen- und Spaltenindex eines Pixels im Framegrabber
- (u_d, v_d) : Verzeichnete Bildkoordinaten gemäß Glg. (5)
- s_u : Unsicherheitsfaktor
- N_{fu} : Anzahl der Pixel einer Zeile des Framegrabbers
- N_{cu} : Anzahl Sensorelemente in u -Richtung
- d_u : Abstand zwischen benachbarten Sensorelementen in u -Richtung
- d_v : Abstand zwischen benachbarten Sensorelementen in v -Richtung

Der Unsicherheitsfaktor s_u resultiert aus Timingfehlern (jitter) zwischen Kamera und Framegrabber, da in gegenwärtiger Hardware kein Synchronisationsignal für den Zeilenanfang definiert ist [Len90]. Bei Verwendung einer digitalen Kamera fällt dieser Unsicherheitsfaktor weg (bzw. $s_u = 1$), da in diesem Fall direkt das auf dem CCD-Sensor entstehende Bild vom Rechner verwendet wird.

2.4 Kamerakalibrierung

Eine Kamerakalibrierung wird durchgeführt, um die Parameter des geometrischen Modells einer Kamera zu bestimmen. Die Parameter werden dabei in zwei Klassen geteilt. Die *äußere Orientierung* ist durch die sechs Parameter gegeben, die die Position einer Kamera im Weltkoordinatensystem bestimmen. Dazu zählen die Position des Linsenzentrums und die Orientierung der Bildebene:

- drei Translationsparameter x_0, y_0, z_0
- drei Rotationswinkel ω, ϕ, κ

Die *innere Orientierung* einer Kamera ist durch die restlichen sechs Parameter gegeben, die die Geometrie von 3D-Strahlen auf die Bildebene bestimmen:

- Kammerkonstante f
- Bildhauptpunkt (u_0, v_0)
- Koeffizienten der Linsenverzeichnung κ_1 und κ_2
- Unsicherheitsfaktor s_u

Die gesamte Spezifikation der Orientierung einer Kamera ergibt sich durch die innere und äußere Orientierung. Weitere Informationen über das Kalibrieren von Kameras finden sich z. B. in [Fau93], [Hor86] und [HS93].

3 Grundlagen zur Stereoauswertung

Dieses Kapitel befaßt sich mit der geometrischen Beschreibung der Stereoauswertung für zwei Kameras.

3.1 Das Stereoproblem

Abb. 2 zeigt das grundlegende Problem, das beim Stereoverfahren gelöst werden muß. Ein physikalischer Punkt P im Weltkoordinatensystem (x, y, z) wird auf den Punkt P' im Bildkoordinatensystem (u', v') der linken Kamera und auf den Punkt P'' im Bildkoordinatensystem (u'', v'') der rechten Kamera abgebildet.

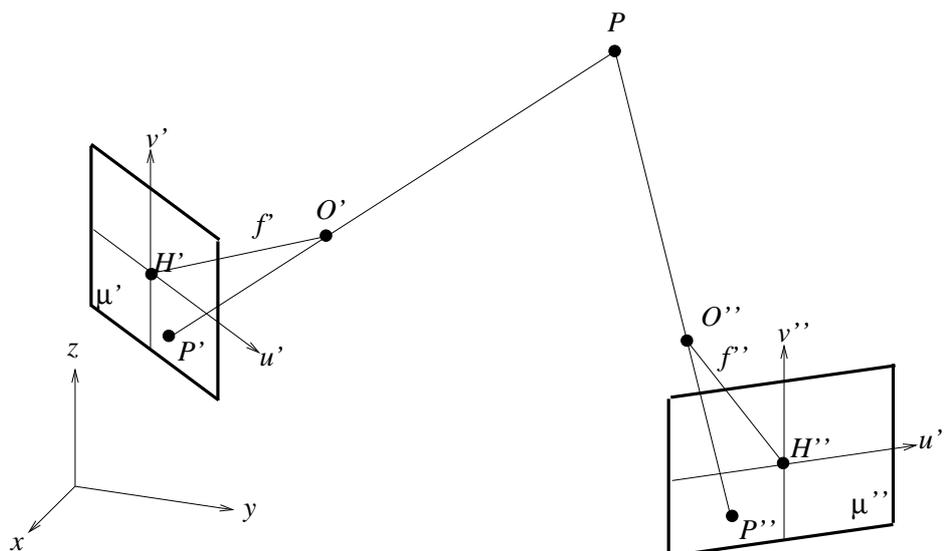


Abbildung 2: Das Stereoproblem

Auf Grund der Bildinformation in den Ebenen μ' und μ'' sind zwei Probleme zu lösen:

1. Für einen Punkt P' in der Bildebene μ' muss entschieden werden, zu welchem Punkt P'' in der Bildebene μ'' der Punkt P' korrespondiert. Die *Korrespondenz* zweier Punkte bedeutet, daß sie Bilder desselben physikalischen Punktes P sind. Dieses Problem wird als *Korrespondenzproblem* bezeichnet.
2. Für zwei korrespondierende Punkte P' und P'' müssen die 3D-Koordinaten des Punktes P im Weltkoordinatensystem berechnet werden. Dieses Problem wird als *Rekonstruktionsproblem* bezeichnet.

3.2 Epipolare Geometrie

Dieser Abschnitt beschreibt, wie der zweidimensionale Suchbereich für korrespondierende Punkte auf eine Dimension beschränkt werden kann. Weiters erfolgt die Beschreibung der Kameraanordnung (Normalfall), für die der Suchbereich sehr einfach ist. Schließlich wird gezeigt, wie man aus allgemeinen Stereobildern Stereopaare erzeugen kann, für die der Normalfall gilt.

3.2.1 Epipolare Linien

Eine wichtige geometrische Einschränkung zur Lösung des Korrespondenzproblems ist durch die *epipolaren Linien* gegeben, die in Abb. 3 zu sehen sind.

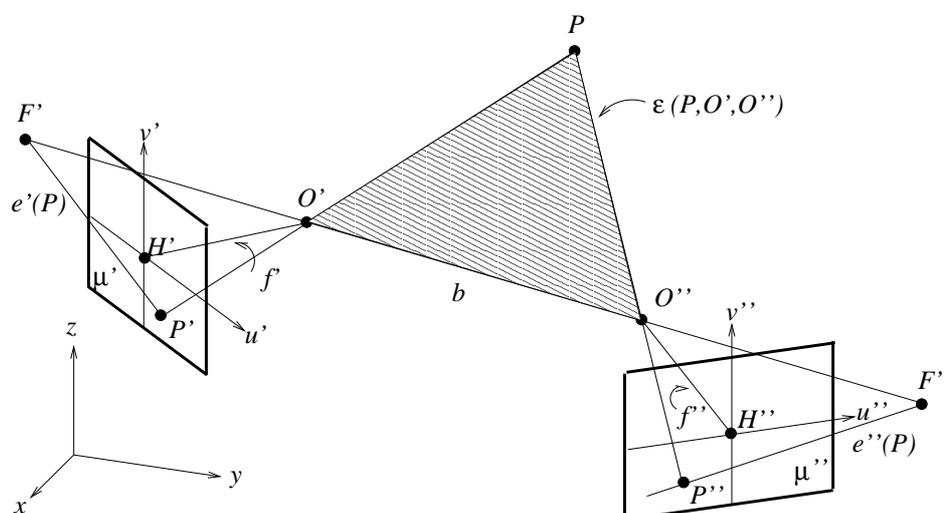


Abbildung 3: Epipolare Geometrie

Die Projektionszentren O' und O'' definieren den Abstand der Länge b zwischen den Kameras, der als *Basis* bezeichnet wird. Die Bildhauptpunkte H' und H'' sind die Ursprünge der Bildkoordinaten (u', v') und (u'', v'') . Die fünf Punkte P , O' , O'' , P' und P'' liegen alle in einer Ebene, der dem Punkt P zugeordneten *epipolaren Ebene* $\epsilon(P, O', O'')$. Die Schnittlinien von $\epsilon(P, O', O'')$ mit μ' und μ'' bilden die zwei dem Punkt P zugeordneten *epipolaren Linien* $e'(P)$ und $e''(P)$. Alle epipolaren Ebenen gehen durch die Basis $b = (O'O'')$. Die epipolaren Linien schneiden alle die *Epipole* F' und F'' , die durch den Schnitt der Basis b mit den Bildebenen μ' bzw. μ'' definiert sind. Epipolare Linien sind daher im allgemeinen Fall in einem Bild nicht parallel.

Der Hauptvorteil dieser geometrischen Beziehung resultiert daraus, daß die epipolare Ebene $\epsilon(P, O', O'')$ durch P , O' und O'' definiert ist. Wenn nur ein Bildpunkt P' gegeben ist, dann ist die epipolare Linie $e''(P)$ fixiert und P'' muß auf dieser Linie liegen. Die Suche nach korrespondierenden Punkten ist daher nur in einer Dimension notwendig.

3.2.2 Berechnung der epipolaren Linien

Die epipolaren Linien können folgendermaßen berechnet werden: Gegeben sei die relative Orientierung zwischen zwei Kameras durch eine Rotationsmatrix

$$R \equiv \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad (7)$$

und einen Translationsvektor

$$T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \quad (8)$$

Ein Punkt, der die Koordinaten (u', v') im Koordinatensystem der linken Bildebene mit Abstand f' zum Linsenzentrum hat, hat die 3D Koordinaten (u', v', f') (vgl. Glg. 3). Da das Linsenzentrum der Ursprung des Kamerakoordinatensystems ist, besteht ein Strahl, der durch den Punkt (u', v', f') und den Ursprung geht aus allen Vielfachen von (u', v', f') . Weiters hat jeder 3D-Punkt auf diesem Strahl die perspektivischen Projektionen (u', v') . Die Linie L ,

$$L = \left\{ \left(\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \middle| \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \lambda \begin{pmatrix} u' \\ v' \\ f' \end{pmatrix} \right) \right\} \quad (9)$$

ist die *inverse perspektivische Projektion* des Punktes (u', v') .

Im rechten Kamerakoordinatensystem werden die Koordinaten auf dieser Linie durch

$$\begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} = RL + T, \quad (10)$$

oder

$$\begin{aligned} x'' &= (r_{11}u' + r_{12}v' + r_{13}f')\lambda + t_x \\ y'' &= (r_{21}u' + r_{22}v' + r_{23}f')\lambda + t_y \\ z'' &= (r_{31}u' + r_{32}v' + r_{33}f')\lambda + t_z \end{aligned} \quad (11)$$

ausgedrückt und sie werden durch die perspektivische Projektion auf

$$\frac{u''}{f''} = \frac{x''}{z''} \quad \text{und} \quad \frac{v''}{f''} = \frac{y''}{z''} \quad (12)$$

abgebildet. Wenn die Abkürzungen

$$x'' = a\lambda + m, \quad y'' = b\lambda + n, \quad z'' = c\lambda + l,$$

verwendet werden, dann gilt:

$$\frac{u''}{f''} = \frac{a}{c} + \frac{cm - al}{c} \frac{1}{c\lambda + l} \quad \text{und} \quad \frac{v''}{f''} = \frac{b}{c} + \frac{cn - bl}{c} \frac{1}{c\lambda + l}. \quad (13)$$

Das ist eine Gerade, die den Punkt $(m/l, n/l)$, für $\lambda = 0$, mit dem Punkt $(a/c, b/c)$, für $\lambda \rightarrow \infty$, verbindet. Der erste dieser Punkte ist der Epipol F'' des rechten Bildes und der zweite Punkt ist der Fluchtpunkt der Linie L . Entlang dieser Gerade wird ein korrespondierender Punkt (u'', v'') für den Punkt (u', v') gesucht.

3.2.3 Der Normalfall der Stereoauswertung

Einen Sonderfall bei der Berechnung der epipolaren Linien ergibt eine Kameraanordnung, bei der die beiden Bildebenen μ' und μ'' in einer gemeinsamen Ebene liegen, die parallel zur Linie $(O'O'')$ ist (Abb. 4). In der gemeinsamen Ebene ist die rechte Bildebene μ'' nur horizontal zur linken Bildebene μ' verschoben. Gilt außerdem $f' = f''$, so wird diese Kameraanordnung als *Normalfall* bezeichnet.

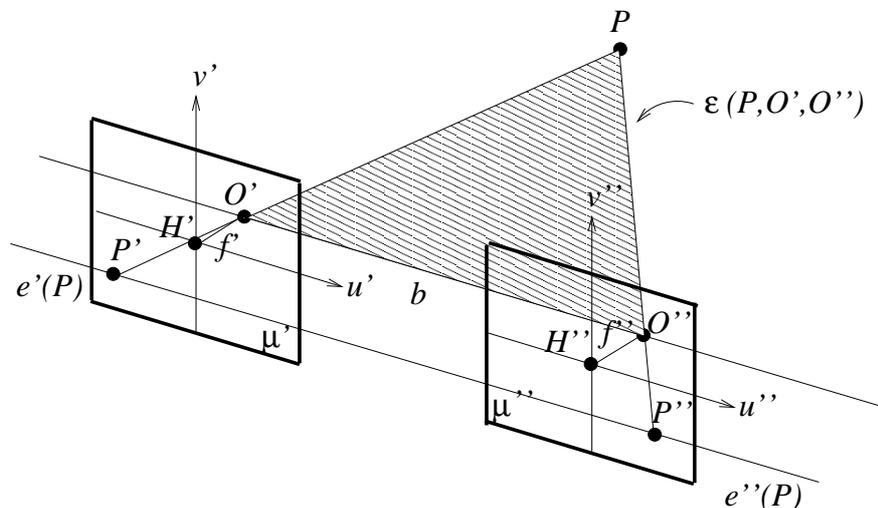


Abbildung 4: Normalfall

Beim Normalfall verlaufen die Schnittlinien der epipolaren Ebene $\epsilon(P, O', O'')$ mit den Bildebenen μ' und μ'' für jeden Punkt P des Weltkoordinatensystems in horizontaler Richtung. Die Suche nach korrespondierenden Punkten beschränkt sich auf dieselbe Zeile der beiden digitalen Bilder.

In der Praxis ergeben sich folgende Probleme bei einer Kameraanordnung im Normalfall:

- Die durch Kamerakalibrierung errechneten Parameter der perspektivischen Projektion sind durch Meßfehler gestört, sodaß im allgemeinen für korrespondierende Punkte $v' \neq v''$ gilt.
- Die beiden Bildebenen haben oft unterschiedliche Orientierung
- Werden zwei Kameras verwendet, so gilt für die Kammerkonstanten fast immer $f' \neq f''$

3.2.4 Geometrische perspektivische Rektifizierung

Aus zwei Bildern einer allgemeinen Kameraanordnung können durch geometrische perspektivische Rektifizierung zwei neue Bilder erzeugt werden, für die der Normalfall gilt. Das Grundprinzip besteht darin, die beiden Bilder in den Ebenen μ' und μ'' durch Bilder in einer gemeinsamen Ebene v zu ersetzen, wobei die Geometrie der Strahlenbündel, die durch die Punkte in μ' und μ'' und die Projektionszentren O' und O'' aufgespannt werden, erhalten bleibt.

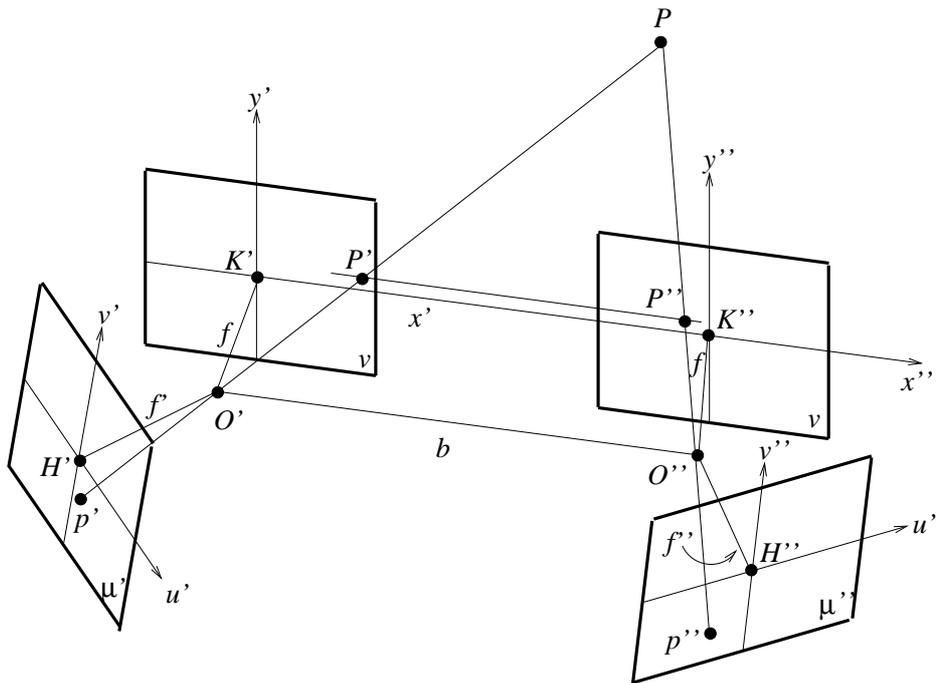


Abbildung 5: Rektifizierung

Abb. 5 zeigt die Ebenen μ' und μ'' , die neue Ebene v und die Bilder p', p'', P' und P'' von P . Die Rektifizierung muß garantieren, daß p' auf P' und p'' auf P'' abgebildet wird:

1. Es wird eine Ebene v gewählt, die parallel zu zur Basis b ist. Dies fixiert die Kamerkonstante f der beiden neuen Bilder. Da es unendlich viele solcher Ebenen gibt, muß ein zusätzliches Kriterium eingeführt werden: Die Ebene v soll auch parallel zur Linie sein, die durch den Schnitt der beiden Ebenen μ' und μ'' definiert ist. Neue Bildkoordinatensysteme (x', y') und (x'', y'') mit den Ursprüngen K' und K'' werden gewählt, wobei x' und x'' kollinear und parallel zu b sind. K' und K'' sind die Bildhauptpunkte der neuen Bilder und $K'O'$ und $K''O''$ sind parallel und orthogonal zu v und $K'K''$ muß die Länge der Basis b haben.
2. Die beiden neuen Bilder werden jeweils folgendermaßen berechnet:

- (a) Die Koordinaten der neuen Punkte auf der neuen Ebene gemessen in Pixel sind (x_i, y_i) . Diese Punkte werden durch die bekannte Orientierung der Kamera auf das Originalbild projiziert.
- (b) Dadurch hat jedes Pixel (x_i, y_i) im neuen Bild eine Position (u_i, v_i) im originalen Bild und leitet daraus seine Intensität ab. Da u_i und v_i im allgemeinen keine ganzzahligen Werte sind, wird die Intensität durch Interpolation mit den ganzzahligen Nachbarwerten von u_i und v_i ermittelt. Ein Überblick ueber Interpolationsmethoden findet sich z. B. in [PTVF92].

Wird die Linsenverzeichnung nicht berücksichtigt, so reicht es, vier Punkte (die z.B. ein Quadrat formen) im Originalbild zurückzuprojizieren und daraus eine projektive Transformation zu berechnen [HS93].

3.3 Das Korrespondenzproblem

Die Suche nach korrespondierenden Punkten entlang der epipolaren Linien erfolgt auf Grund von semantischen Informationen in den beiden Stereobildern. In der Literatur (z.B. [Hor86],[Na193],[DA89],[Kos93]) wird oft die Unterscheidung zwischen *flächenbasierten* und *merkmalsbasierten* Stereoverfahren gemacht.

3.3.1 Flächenbasierte Stereoverfahren

Bei der Korrespondenzanalyse in Stereobildern wird häufig die Annahme getroffen, daß korrespondierende Pixel einen ähnlichen Intensitätswert besitzen. Diese Annahme ist jedoch nicht ausreichend für eine eindeutige Zuordnung, da in der Regel eine große Anzahl identischer Intensitätswerte in einem Bild auftreten. Es werden deshalb mehrere benachbarte Pixel in einem Fenster zu einem Block zusammengefaßt und die Zuordnung zwischen den Pixeln erfolgt anhand eines Ähnlichkeitsmaßes zwischen den Intensitätswerten der Blöcke.

Flächenbasierte Verfahren haben den Nachteil, daß sie anfällig für photometrische Variationen und perspektivische Verzerrungen sind.

Abb. 6 illustriert die geometrischen Verzerrungen, die entstehen, wenn eine Oberfläche relativ zur Basis geneigt ist. In diesem Fall wird diese Oberfläche in einem Bild kürzer als im anderen Bild abgebildet. Ein Beispiel für einen Stereoalgorithmus im Rahmen des Archäologieprojektes findet sich in [MB95].

3.3.2 Merkmalsbasierte Stereoverfahren

Merkmalsbasierte Stereoverfahren ordnen Merkmale im linken Bild Merkmalen im rechten Bild zu. Als Merkmale werden besonders markante Teile des Bildes, wie z.B. Kanten, Eckpunkte oder Geradensegmente verwendet.

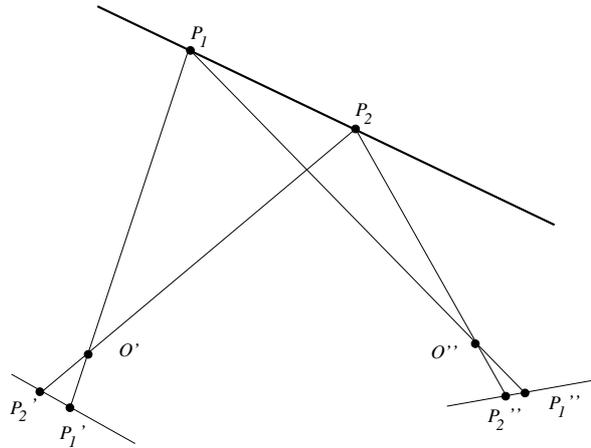


Abbildung 6: Verzerrungen bei einer geneigten Ebene

Merkmalsbasierte Stereoverfahren haben den Vorteil, weniger anfällig für photometrische Variationen zu sein und sind zudem schneller als flächenbasierte Verfahren, weil weniger Korrespondenzpunkte ausgewertet werden müssen. Andererseits kann dichte Tiefeninformation nur durch Interpolation berechnet werden.

3.4 Das Rekonstruktionsproblem

In diesem Abschnitt wird die Rekonstruktion der Weltkoordinaten aus zwei korrespondierenden Bildpunkten beschrieben.

3.4.1 Normalfall

Gegeben ist eine Kameraanordnung im Normalfall mit der Basis b , wobei sich das Projektionszentrum O' der linken Kamera im Weltkoordinatensystem an der Stelle

$$\begin{pmatrix} -b/2 \\ 0 \\ 0 \end{pmatrix}$$

befindet und das Projektionszentrum O'' der rechten Kamera im Weltkoordinatensystem an der Stelle

$$\begin{pmatrix} b/2 \\ 0 \\ 0 \end{pmatrix}$$

ist. Für beide Kameras gelte $R(\omega, \phi, \kappa) = I$ (Einheitsmatrix) (siehe Abb. 7).

Sei (x, y, z) ein 3D-Punkt im Weltkoordinatensystem und seien (u', v') und (u'', v'') seine perspektivischen Projektionen im linken bzw. rechten Bild. Dann gilt

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \frac{f}{z} \begin{pmatrix} x + b/2 \\ y \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} u'' \\ v'' \end{pmatrix} = \frac{f}{z} \begin{pmatrix} x - b/2 \\ y \end{pmatrix}. \quad (14)$$

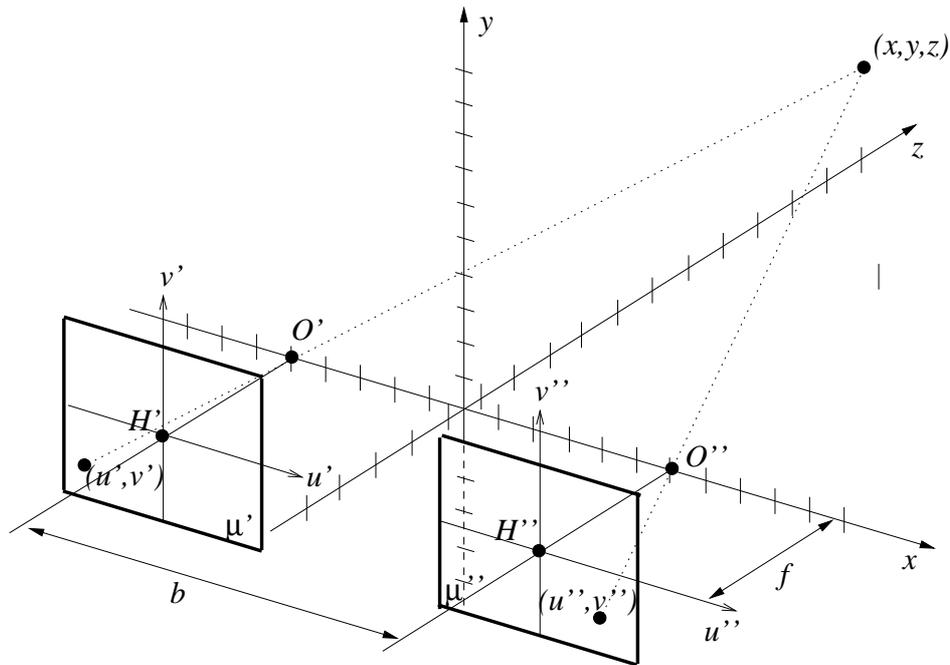


Abbildung 7: Kameraanordnung im Normalfall

Hat man ein Korrespondenzpaar (u', v') und (u'', v'') gegeben, so kann die Lösung für x, y, z aus der Differenz $u' - u''$, die auch *Disparität* genannt wird, aus Gleichung (14) berechnet werden.

$$u' - u'' = \frac{f}{z} \left[x + \frac{b}{2} - \left(x - \frac{b}{2} \right) \right] = \frac{f}{z} b$$

Es gilt daher

$$z = \frac{fb}{u' - u''} \quad (15)$$

Ist einmal z bestimmt, so können die (x, y) -Koordinaten durch die Gleichungen der perspektivischen Projektion bestimmt werden:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{z}{f} \begin{pmatrix} u' \\ v' \end{pmatrix} - \begin{pmatrix} b/2 \\ 0 \end{pmatrix} = \frac{z}{f} \begin{pmatrix} u'' \\ v'' \end{pmatrix} + \begin{pmatrix} b/2 \\ 0 \end{pmatrix} \quad (16)$$

3.4.2 Allgemeine Kameraanordnung

Gegeben sind zwei Kameras, die wie in Abschnitt 3.2.2 durch die Rotationsmatrix R und den Translationsvektor T relativ zueinander orientiert sind. f' und f'' sind die fokalen Längen der linken bzw. rechten Kamera, (u', v') und (u'', v'') sind zwei korrespondierende Punkte im linken und im rechten Bild. Dann gilt gemäß Gleichungen (11) und (12):

$$\begin{aligned} \left(r_{11} \frac{u'}{f'} + r_{12} \frac{v'}{f'} + r_{13} \right) z' + t_x &= \frac{u''}{f''} z'', \\ \left(r_{21} \frac{u'}{f'} + r_{22} \frac{v'}{f'} + r_{23} \right) z' + t_y &= \frac{v''}{f''} z'', \\ \left(r_{31} \frac{u'}{f'} + r_{32} \frac{v'}{f'} + r_{33} \right) z' + t_z &= z''. \end{aligned} \quad (17)$$

Zwei dieser Gleichungen können verwendet werden, um z' und z'' und anschließend die restlichen Kamerakoordinaten zu berechnen:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} u'/f' \\ v'/f' \\ 1 \end{pmatrix} z' \quad \text{und} \quad \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} = \begin{pmatrix} u''/f'' \\ v''/f'' \\ 1 \end{pmatrix} z'' \quad (18)$$

4 Problemstellung der Diplomarbeit

Ein Hauptmerkmal der zahlreichen publizierten Arbeiten über das Stereoverfahren ist die Annahme des Normalfalls und damit die Beschränkung des Suchraums für korrespondierende Punkte auf horizontale Bildzeilen. Es werden *Einschränkungen zur Kamerageometrie* gemacht, damit man vereinfachte Algorithmen findet, die wenig *Einschränkungen zur 3D-Szene* benötigen. Große Probleme bereiten zum Beispiel Verdeckungen und spiegelnde Oberflächen [BN95].

Im Rahmen des Archäologieprojektes sollen einzelne archäologische Scherben dreidimensional erfaßt werden. Dadurch gelten folgende Einschränkungen zur 3D-Szene:

- Die zu berechnende Tiefe z im Weltkoordinatensystem variiert in einem zum Kamerastand relativ kleinen Intervall. Dadurch kann der Suchraum entlang der epipolaren Linien stark eingeschränkt werden.
- Es gibt keine Verdeckungen durch andere Objekte.

Im Gegensatz zu diesen Einschränkungen zur Szene wird im Rahmen der Diplomarbeit keine Beschränkung der Kamerageometrie auf den Normalfall gemacht. Die Tiefenberechnung soll für folgende zwei Arten von Stereopaaren einer allgemeinen Kameraanordnung erfolgen, wobei kalibrierte Kameras vorausgesetzt werden:

1. Auf den Normalfall rektifizierte Stereobilder mit Korrespondenzsuche entlang von horizontalen Bildzeilen als epipolare Linien. Folgende Aufgaben sind geplant:
 - Entwicklung eines Rektifizierungsalgorithmus und Anwendung eines flächenbasierten Stereoalgorithmus (Abschn. 3.3.1) mit Berücksichtigung der Einschränkungen des Suchbereichs.
 - Untersuchung der Fehler.
 - Ausarbeitung der Vor- und Nachteile der Rektifizierung.

2. Originale Stereobilder mit Korrespondenzsuche entlang von von allgemeinen digitalen epipolaren Linien. Folgende Ziele sind geplant:
 - Entwicklung eines flächenbasierten Stereoalgorithmus mit Berücksichtigung der Einschränkungen des Suchbereichs.
 - Untersuchung des Performanceverlustes gegenüber der Rektifizierung.
 - Untersuchung, ob die Anzahl der gefundenen Korrespondenpaare erhöht werden kann.
 - Ausarbeitung der Vor- und Nachteile der direkten Tiefenberechnung.

5 Zusammenfassung

Neben der Modellierung der Bildentstehung wurden die Grundzüge der Stereoauswertung wie epipolare Geometrie, Normalfall, geometrische perspektivische Rektifizierung und das Korrespondenzproblem beschrieben. Die Aufgabe der Diplomarbeit besteht einerseits darin, die Rektifizierung auf den in der Literatur oft angenommenen Normalfall zu untersuchen, andererseits darin, eventuelle Vorteile gegenüber der Rektifizierung durch eine direkte Tiefenberechnung entlang von epipolaren Linien zu evaluieren.

Literatur

- [BN95] Dinkar N. Bhat and Shree. K. Nayar. Stereo in the Presence of Specular Reflection. In *Proceedings ICCV 1995*, pages 1086–1092, 1995.
- [DA89] Umesh R. Dhond and J.K. Aggarwal. Structure from Stereo – A Review. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1489–1510, 1989.
- [DKKF93] Petros Dintsis, Walter G. Kropatsch, Fritz Krinzinger, and Herwig Friesinger. Ansuchen zur Förderung des Projektes: Bildhafte Erfassung von archäologischen Fundgegenständen (Keramik), 1993.

- [Fau93] Olivier Faugeras. *Three Dimensional Computer Vision*. MIT Press, 1993.
- [Hor86] Berthold Klaus Paul Horn. *Robot Vision*. MIT Press, 1986.
- [HS93] Robert M. Haralick and Linda G. Shapiro. *Computer and Robot Vision*, volume II. Addison Wesley, 1993.
- [Kos93] Andreas Koschan. What is New in Computational Stereo Since 1989: A Survey on Current Stereo Papers. Technical Report 93-22, Technische Universität Berlin, 1993.
- [Kra92] Karl Kraus. *Photogrammetry Volume 1*. Duemmler Bonn, 1992.
- [Len90] R. Lenz. Grundlagen der Videometrie, angewandt auf eine ultra-hochauflösende CCD-Farbkamera. *Technisches Messen*, 57(10):366–380, 1990.
- [Mar82] David Marr. *Vision*. W. H. Freeman and Company, 1982.
- [MB95] Christian Menard and Norbert Brändle. Hierarchical Area-Based Stereo Algorithm for 3D Acquisition. In *International Workshop on Stereoscopic and Three Dimensional Imaging*, pages 195–201, 1995.
- [Nal93] V. S. Nalwa. *A Guided Tour to Computer Vision*. Addison Wesley, 1993.
- [PTVF92] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.
- [Shi87] Yoshiaki Shirai. *Three-Dimensional Computer Vision*. Springer Verlag, 1987.
- [SM95] Robert Sablatnig and Christian Menard. Computer based acquisition of archaeological finds: The first step towards automatic classification. In *3rd International Symposium on Computing in Archaeology, Rome (in press)*, 1995.
- [SMD93] Robert Sablatnig, Christian Menard, and P. Dintsis. A preliminary study on methods for a pictorial acquisition of archaeological finds. In P. M. Fischer, editor, *Archaeology and Natural Science (ANS)*, volume 1, pages 143–151. Paul Astroems Foerlag, Gothenburg, 1993.
- [SMD94] Robert Sablatnig, Christian Menard, and P. Dintsis. Bildhafte, dreidimensionale Erfassung von archäologischen Fundgegenständen als Grundlage für die automatisierte Klassifikation. In O. Stoll, editor, *Computer & Antike*, volume 3, pages 59–84. Scripta Mercaturae Verlag, 1994.
- [Tsa86] Roger Y. Tsai. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In *Proceedings IEEE CVPR*, pages 364–374, 1986.

- [VV90] Russell L. De Valois and Karen K. De Valois. *Spatial Vision*. Oxford University Press, 1990.