



Stereo Vision

Anwendungen 2 / Sommersemester 2010

**Fakultät Technik und Informatik
Department Informatik**

Gregory Föll

Übersicht

- **Rückblick**
- **3D-Bildaufnahme**
- **Kamerakalibrierung**
- **Triangulation durch Active Vision**
- **Stereo Vision**
- **Korrespondenzproblem / Matching-Strategien**
- **Zusammenfassung und Ausblick**

- Zielsetzung: Entwicklung einer Hinderniserkennung für das **FAUST-onyx**-Fahrzeug
- Zur Hinderniserkennung und Entfernungsmessung, werden momentan eine monokulare Kamera und ein Linienlaser verwendet. Hindernisse werden nicht klassifiziert, es wird nur nach einer hellen Linie im Bild gesucht.
- Geplant ist, die Hinderniserkennung und Entfernungsmessung mit Hilfe einer Stereo-Kamera zu realisieren. Dadurch ergeben sich einige Vorteile:
 - Das zu erkennende Objekt kann gedreht oder teilweise verdeckt sein
 - Funktioniert auch bei gleichen Objekten verschiedener Größe
 - Benötigte Rechenkapazität ist relativ gering (gegenüber laser range finders)
 - Geringe Hardwarekosten
 - Annehmbare Komplexität der Algorithmen (gegenüber monokularem Ansatz)



Übersicht

- Rückblick
- **3D-Bildaufnahme**
- Kamerakalibrierung
- Triangulation durch Active Vision
- Stereo Vision
- Korrespondenzproblem / Matching-Strategien
- Zusammenfassung und Ausblick

Verfahren zur Gewinnung von Tiefeninformationen aus Bildern (1)

◆ **Tiefe aus Laufzeit:**

Ein Signal wird ausgesendet, es pflanzt sich mit einer bekannten Geschwindigkeit in Richtung Objekt fort, wird dort reflektiert und dann von einer Kamera empfangen. Die Signallaufzeit ist dann proportional zu dem zurückgelegten Weg.

◆ **Tiefe aus Phase (Interferometrie):**

- Spezialfall der Laufzeitmessung (periodischen Modulation)
- Misst Distanzen in Vielfachen der Wellenlänge unter Berücksichtigung von Amplitude und Phasenlage
- Phasenmessung durch Überlagerung kohärenter Lichtstrahlung
- Extrem genau durch die sehr kleine Wellenlänge des Lichtes

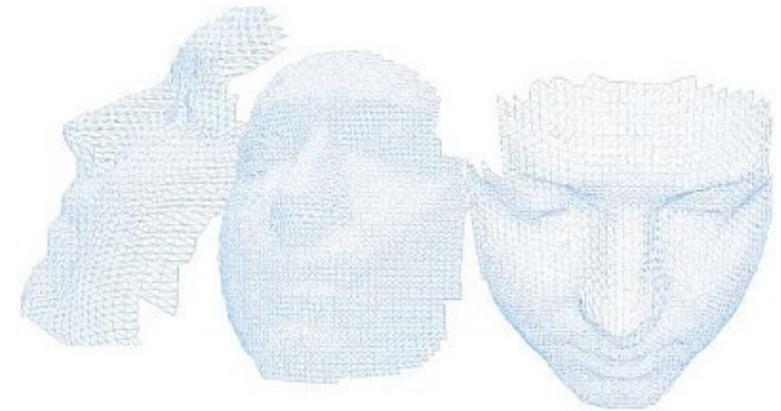
◆ **„Form aus Schattierung“:**

- Bestimmung der Form einer Oberfläche aus der lokalen Orientierung der Oberflächenelemente (Orientierung durch senkrecht zur Oberfläche stehende Vektoren)
- Keine Rekonstruktion der absoluten Tiefe möglich

Verfahren zur Gewinnung von Tiefeninformationen aus Bildern (2)

◆ Tiefe aus Triangulation:

- Das am häufigsten eingesetzte Verfahren
- Stereo Vision (Stereoskopie / passive Triangulation):
 - Gleichzeitige Aufnahme eines Objekts aus zwei unterschiedlichen Blickwinkeln
 - Berechnung der räumlichen Ausdehnung des Objekts mittels der Unterschiede in den beiden Aufnahmen
- Active Vision (aktive Triangulation):
 - Eine Kamera wird durch eine Lichtquelle ersetzt



Übersicht

- Rückblick
- 3D-Bildaufnahme
- **Kamerakalibrierung**
- Triangulation durch Active Vision
- Stereo Vision
- Korrespondenzproblem / Matching-Strategien
- Zusammenfassung und Ausblick

Vor der Aufnahme muss die Kamera/s kalibriert werden

➤ **Extrinsische Kalibrierung**

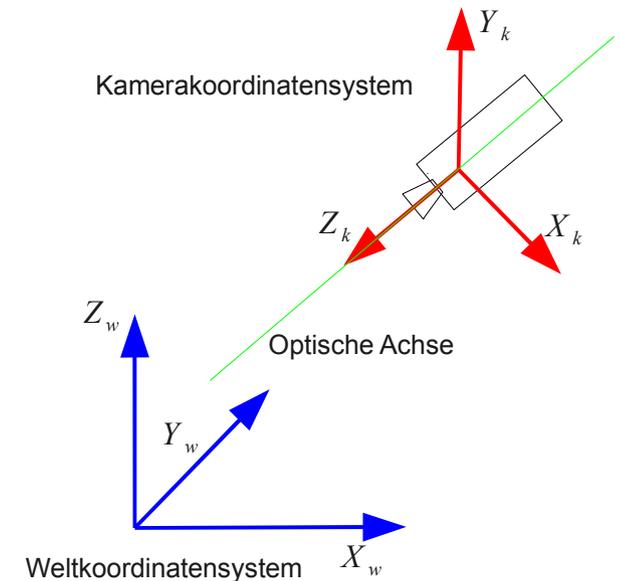
- Definiert den Zusammenhang zwischen dem, auf die betrachtete Szene bezogenen, 3D-Weltkoordinatensystem und dem, mit der bewegten Kamera verknüpften, 3D-Kamerakoordinatensystem
- Beschreibt also die Position und Richtung der Kamera bezüglich eines gegebenen Weltkoordinatensystems

➤ **Intrinsische Kalibrierung**

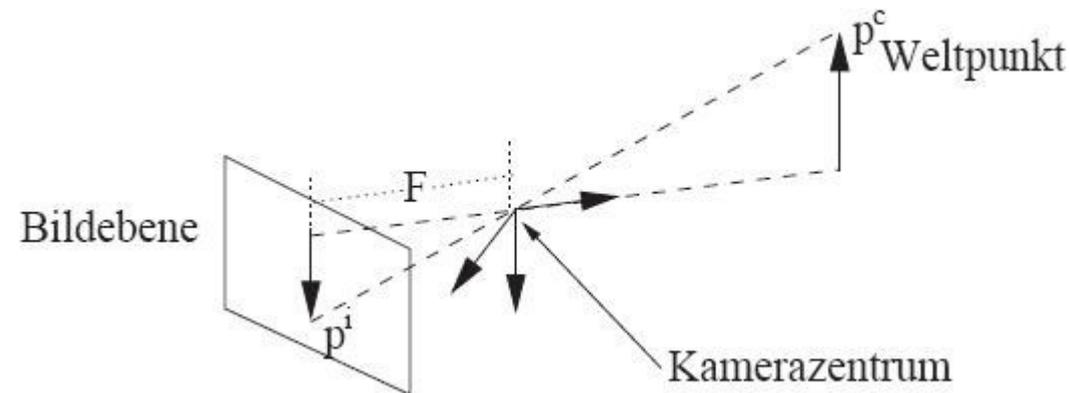
- Definiert die Abbildung zwischen dem 3D-Kamerakoordinatensystem (Maßeinheit: Meter) und dem 2D-Bildkoordinatensystem (Maßeinheit: Pixel)
- Definition durch die Parameter der internen Kamerageometrie
- Gleicht radiale Linsenverzerrung aus

Extrinsische Kalibrierung

- ◆ Position der Kamera entspricht dem optischen Zentrum und der Richtung der optischen Achse, die äquivalent zur Z-Achse des Kamerakoordinatensystems ist
- ◆ Ein Objekt besitzt Weltkoordinaten, die zuerst durch eine *Translation* und eine *Rotation* in Kamerakoordinaten umgewandelt werden müssen
- ◆ Zuerst wird der Ursprung des Weltkoordinatensystems über einen Vektor \mathbf{T} in den Ursprung des Kamerakoordinatensystems geschoben = Translation
- ◆ Dann wird das verschobene Weltkoordinatensystem so gedreht, dass es mit dem Kamerakoordinatensystem zusammenfällt = Rotation
- ◆ Es gibt somit 6 extrinsische Kameraparameter:
 - Translation in x-Richtung (T_x)
 - Translation in y-Richtung (T_y)
 - Translation in z-Richtung (T_z)
 - Rotation um die x-Achse mit Winkel alpha (R_x)
 - Rotation um die y-Achse mit Winkel beta (R_y)
 - Rotation um die z-Achse mit Winkel gamma (R_z)



Intrinsische Kalibrierung



Vereinfacht lässt sich die Abbildungsgeometrie anhand eines Lochkameramodells erläutern, welches eine „ideale Abbildung“ erzeugt

Dabei ergibt sich durch Anwendung der Strahlensätze die folgende Beziehung zwischen den 3D-Welt- und den 2D-Bildkoordinaten:

$$X_B = -F \frac{X_K}{Z_K} \quad Y_B = -F \frac{Y_K}{Z_K}$$

F bezeichnet die Brennweite (fokale Länge)

Linsenverzerrung

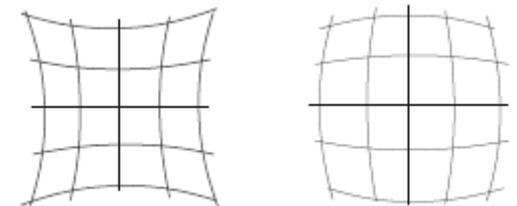
Leider erzeugt ein reales Objektiv im Gegensatz zu einer Lochkamera keine idealen Abbildungen, es weist Verzerrungen auf

◆ Tangentiale Verzerrung

→ Kann nach Tsai vernachlässigt werden

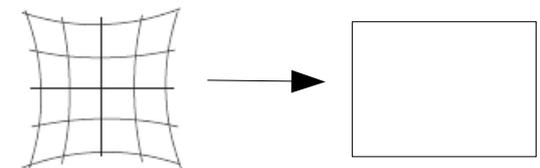
◆ Radiale Verzerrung

- Zur Bildmitte hin → kissenförmig
- Von der Bildmitte weg → tonnenförmig
- Entsteht aufgrund der Geometrie der CCD-Elemente bzw. durch die Optik selbst
- Wächst mit dem Abstand vom optischen Zentrum



Rektifizierung:

Die sogenannte Rektifizierung ist ein Teilschritt der intrinsischen Kalibrierung, bei dem Verzerrungen, unter Berücksichtigung der Kamerakonstanten, ausgeglichen werden



Tsai'sches Kalibrierungsverfahren

Es gibt verschiedene Verfahren für die Kalibrierung von Kameras. Oft wird, wegen seiner guten Eigenschaften, das Verfahren von Roger Tsai eingesetzt, es wird wie folgt durchgeführt:

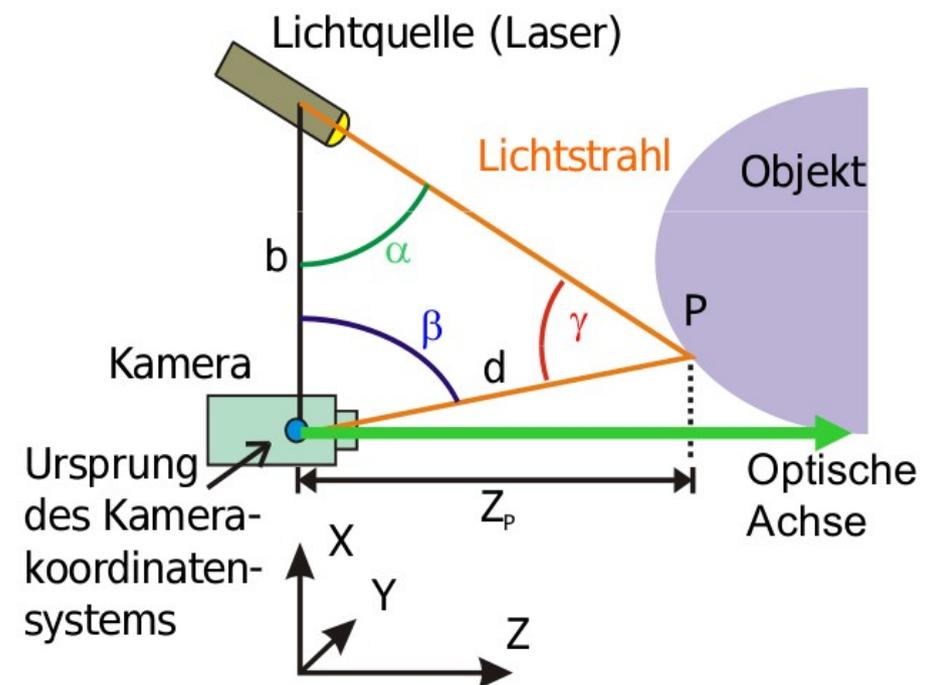
1. Weltkoordinatensystem in das Kamerakoordinatensystem transformieren
2. Kamerakoordinaten in unverzerrte Sensorkoordinaten transformieren
3. Unverzerrte Sensorkoordinaten in verzerrte transformieren
4. Verzerrte Sensorkoordinaten in digitale Bildkoordinaten transformieren

Übersicht

- Rückblick
- 3D-Bildaufnahme
- Kamerakalibrierung
- **Triangulation durch Active Vision**
- Stereo Vision
- Korrespondenzproblem / Matching-Strategien
- Zusammenfassung und Ausblick

Lichtpunkttechnik (1)

- ◆ Eine Lichtquelle emittiert fokussiertes Licht auf die Oberfläche eines Objekts an Ort P
- ◆ Das Objekt wird durch die Kamera aufgenommen
- ◆ Die Lichtquelle liegt auf der X-Achse des Kamerakoordinatensystems
- ◆ Abstand b und Winkel α sind bekannt (werden mechanisch gemessen)
- ◆ X- und Y-Achse sind parallel zu x- und y-Achse in Bildkoordinaten



Lichtpunkttechnik (2)

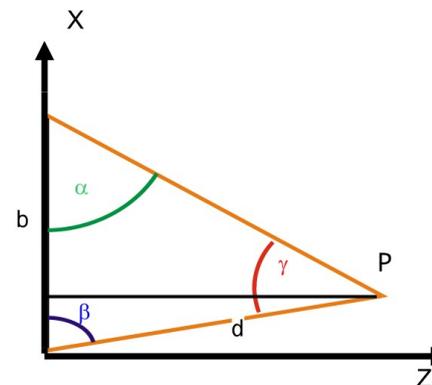
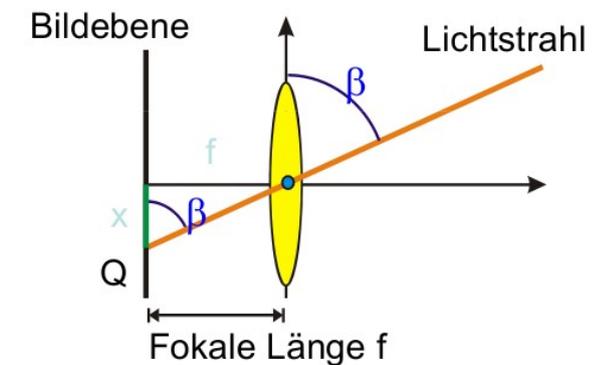
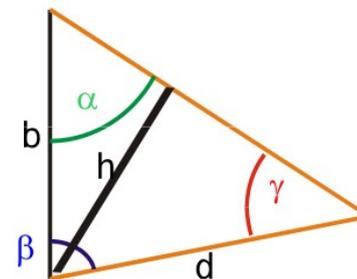
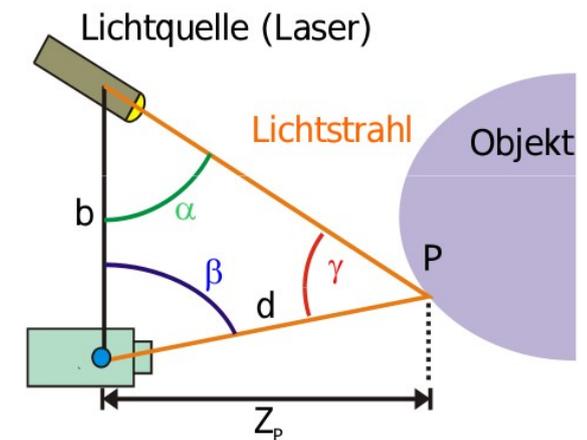
Wenn P in der X-Z-Ebene liegt, dann kann Winkel β , aus der Projektion Q im Bild, wie folgt berechnet werden:

$$\tan \beta = f/x \Rightarrow \beta = \arctan(f/x)$$

Und d kann jetzt einfach mit Hilfe der Sinus- und Kosinusgesetze in Dreiecken berechnet werden

Anschließend wird P in Kamerakoordinaten umgerechnet:

$$X = d \cdot \cos \beta, \quad Y = 0, \quad Z = d \cdot \sin \beta$$

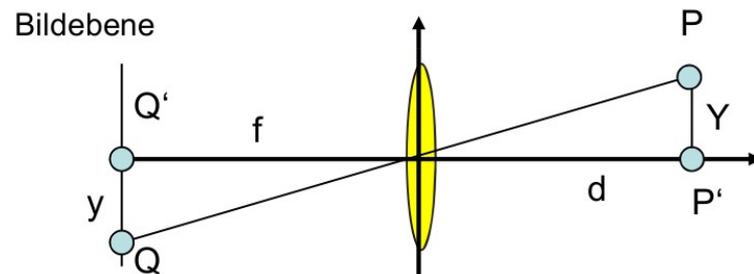
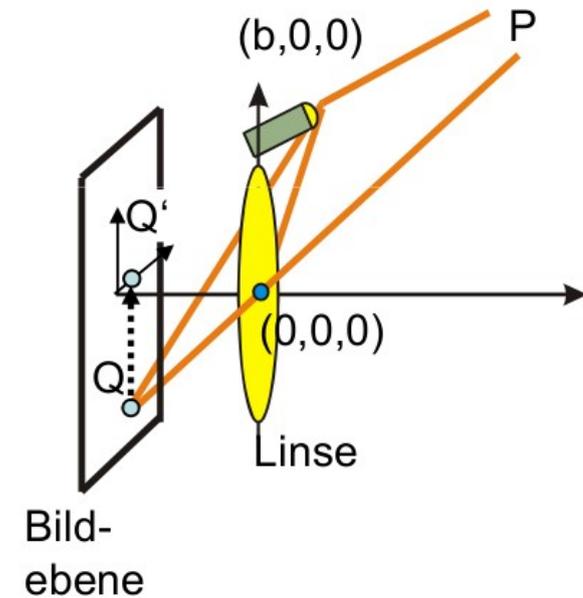


Lichtpunkttechnik (3)

Falls P nicht in der X-Z-Ebene liegt:

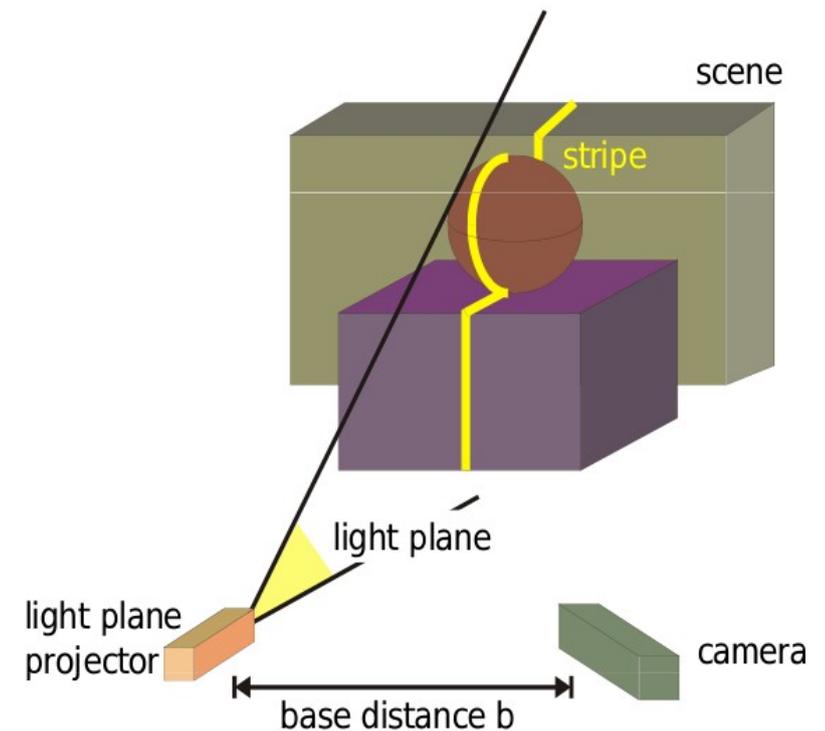
- ▶ Projektion von Q nach Q' in der X-Z-Ebene
- ▶ Berechne X und Z für P' von Q'
- ▶ Berechne Y nach den Strahlensätzen:

$$Y = \frac{d}{fy}$$



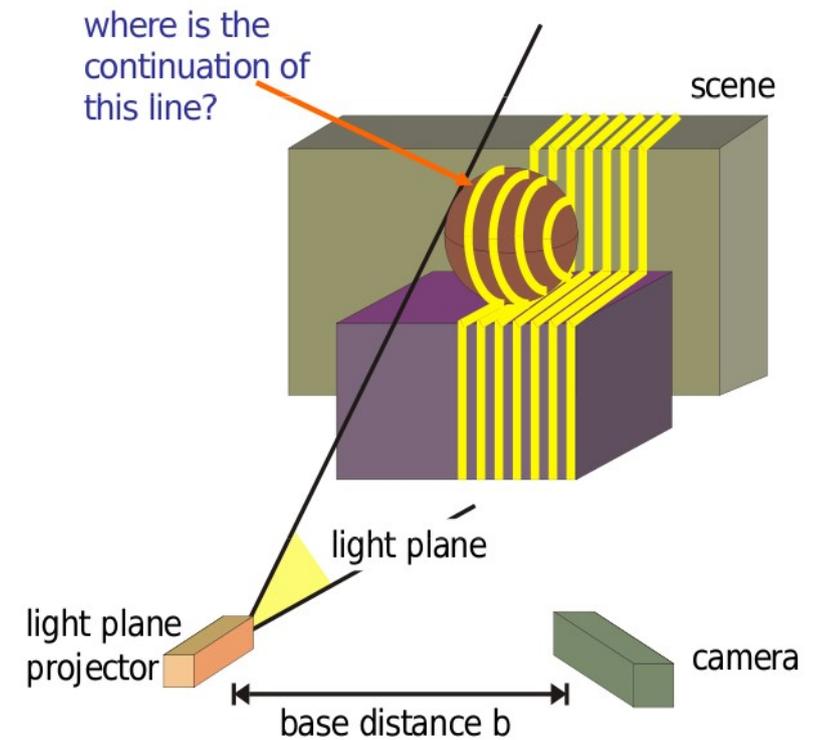
Streifenprojektion

- ◆ Anstatt eines Punktes wird ein Lichtstreifen projiziert
- ◆ Falls die Lichtebene orthogonal zur X-Achse des Kamerakoordinatensystems ist, wird nur ein Punkt auf jeder Linie mit konstantem y beleuchtet
- ◆ **Nachteil:** Man benötigt so viele Bilder wie Streifen



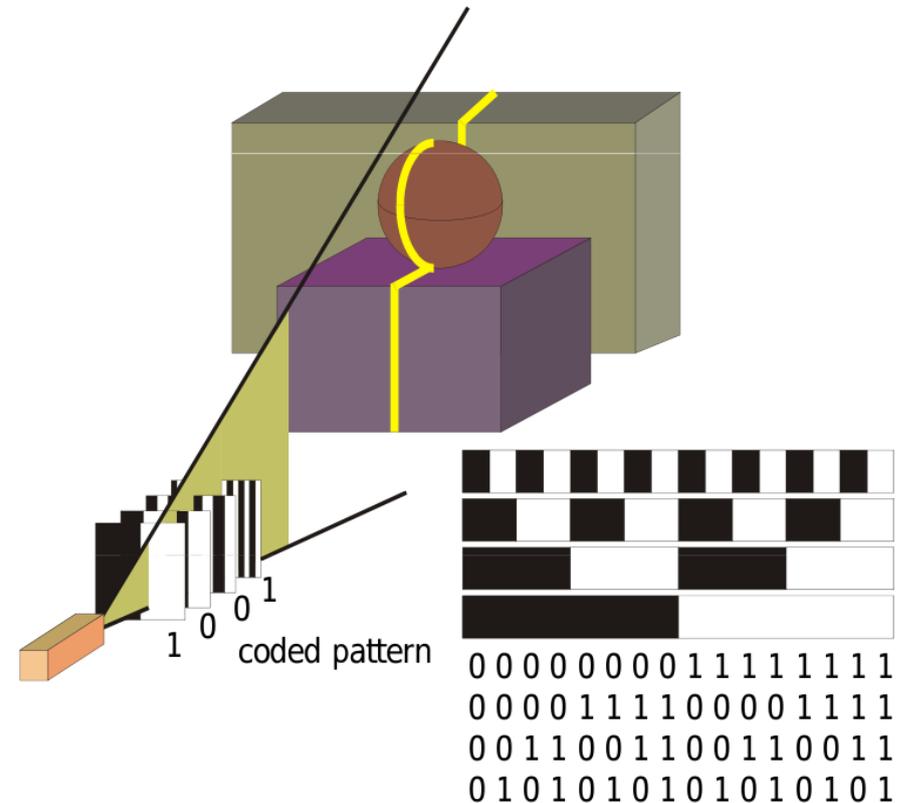
Projektion eines statischen Linienmusters

- Erweiterung der Streifenprojektion:
Projektion eines Streifen-musters
- **Problem:** Eindeutige Zuordnung der
Streifen



Projektion eines Streifen-Codes

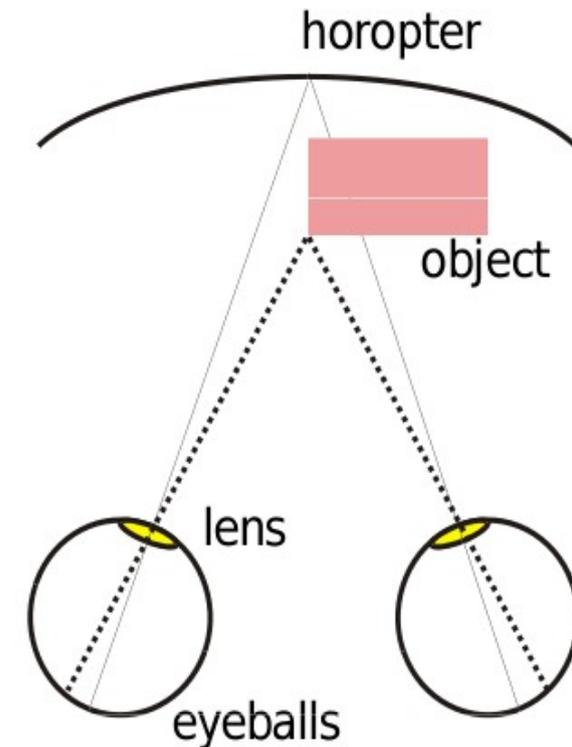
- Lösung des Identifikations-problems:
Zeitcodiertes S/W Muster
- Dadurch: Eindeutige Zuordnung von
Teilen des Musters zu Orten in der
3D-Szene
- Benötigt $\log_2 n$ Muster falls n
verschiedene Streifen unterschieden
werden sollen
- Binary Gray Code (BGC) wird oft zur
Encodierung genutzt. Der BGC ist ein
Code bei dem aufeinander folgende
Zahlen sich in genau einer Ziffer
unterscheiden



Übersicht

- Rückblick
- 3D-Bildaufnahme
- Kamerakalibrierung
- Triangulation durch Active Vision
- **Stereo Vision**
- Korrespondenzproblem / Matching-Strategien
- Zusammenfassung und Ausblick

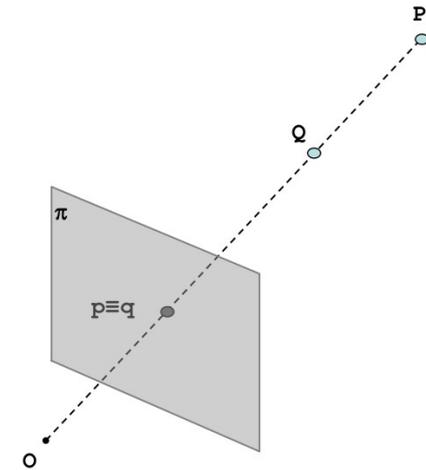
- Wird ein Objekt gleichzeitig von zwei Positionen aus betrachtet, die durch eine Basislinie voneinander getrennt sind (zwei Augen), so erscheint es unter verschiedenen Blickwinkeln
- Durch die Unterschiede in den beiden entstandenen Bildern, kann mittels Triangulation, die räumliche Ausdehnung des Objektes berechnet werden
- Menschen können die beiden Bilder meistens sehr gut automatisch integrieren
- Der Stereoeindruck entsteht dabei durch die Disparitätsunterschiede
- Fusion der Bilder erfolgt nur, wenn sich das Objekt nicht zu weit von der Horopterfläche entfernt befindet



©Virtual Clones Ltd.

Eine Kamera:

Zwei auf der selben Sichtlinie liegende, reale Punkte Q und P, werden auf den selben Bildpunkt abgebildet

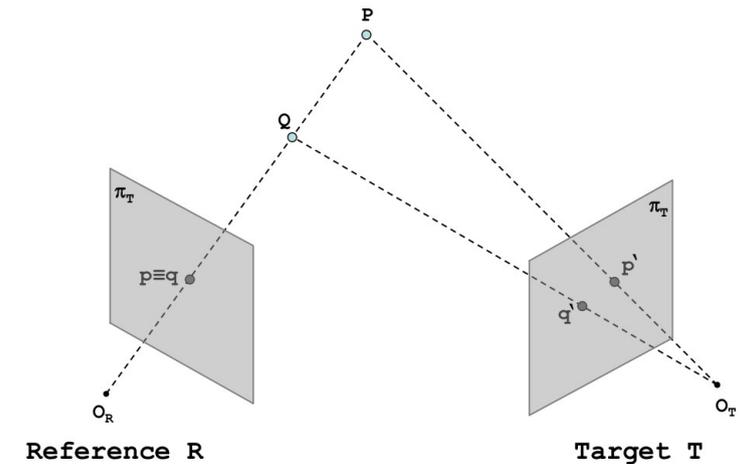


π : image plane

O : optical center

Zwei Kameras:

Fallst es gelingt, die korrespondierenden Punkte in den beiden Bildern zu finde, kann mittels Triangulation, die räumliche Ausdehnung des Objektes berechnet werden



Triangulation

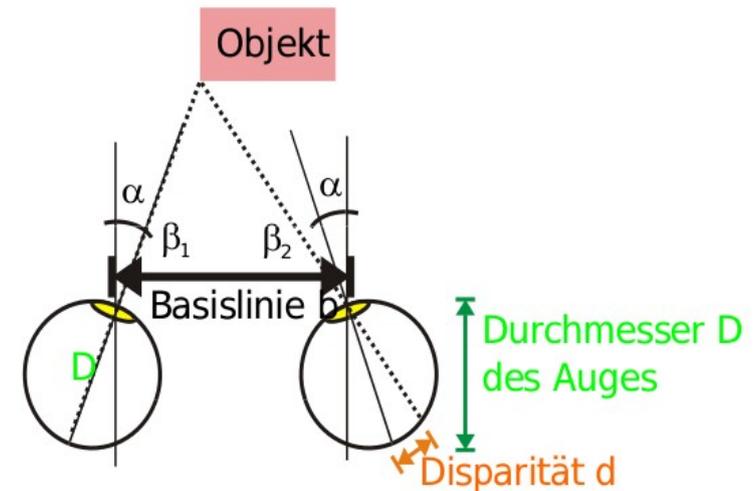
Bestimmen des Dreiecks durch zwei Winkel und eine Seite

Winkel:

Kombination zwischen Fokussierungsrichtung und Disparität

Seite:

Basislinie (Abstand zwischen Linsenzentren)



Als Menschen erkennen wir absolute Abstände ziemlich schlecht und können den Abstand zum Objekt nur schätzen (bis zu 450 m)

Korrespondenzproblem

Welche Pixel im Bild R stimmen mit welchen Pixeln im Bild T überein?

2D-Suchraum?



Reference (R)



Target (T)

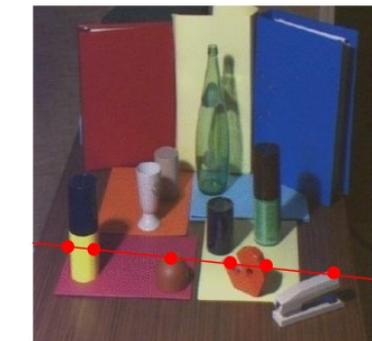
Nein!

Jeder Ort P im Bild R kann nur Orte im Bild T zeigen, die auf einer Linie liegen, die die Projektion von P nach T ist

Diese Linie heißt Epipolarlinie und beschränkt die Suche nach Korrespondenzpaaren auf einen 1D-Suchraum

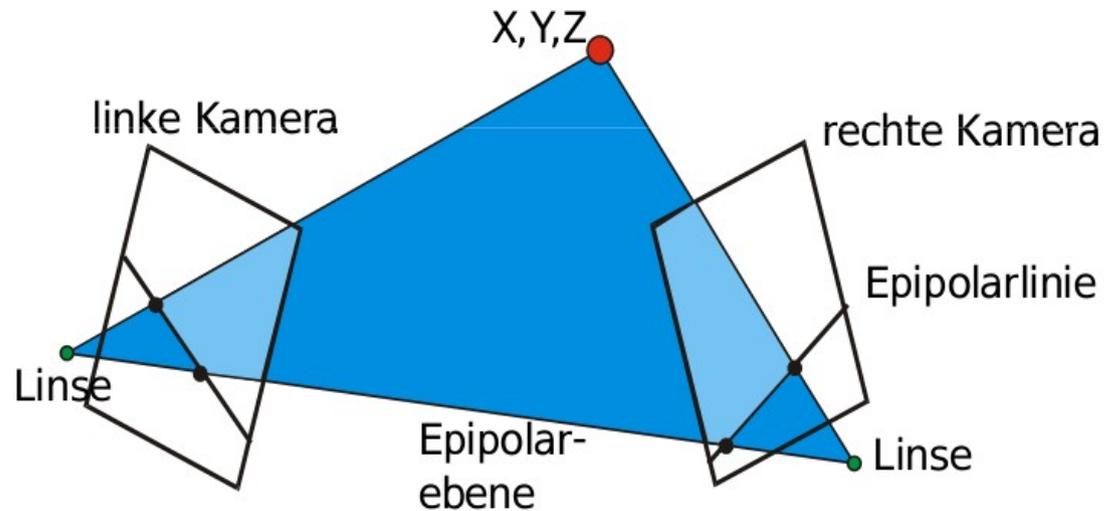


Reference (R)



Target (T)

Epipolarebene

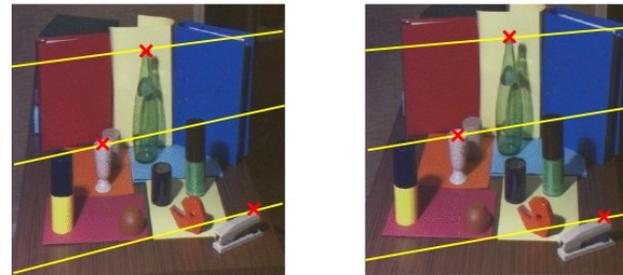
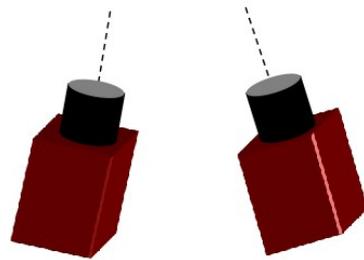


Die Ursprungsorte der beiden Kamerakoordinatensysteme und ein beliebiger Punkt im 3D-Raum, bestimmen eine Epipolarebene

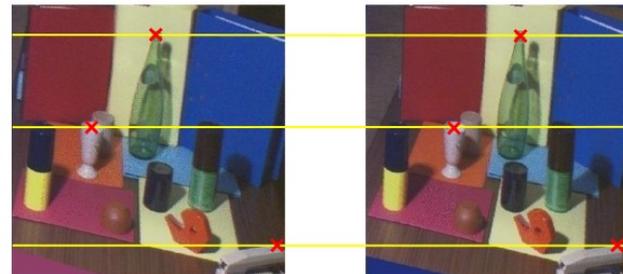
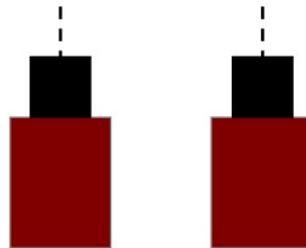
Jeder Punkt in dieser Ebene wird auf dieselben beiden Epipolarlinien in den beiden Bildern projiziert

Standardform / kanonische Stereogeometrie

Zur Vereinfachung wird das Stereobild so transformiert, als ob die optischen Achsen der beiden Kameras parallel zueinander sind und in derselben Ebene liegen:



Original stereo pair

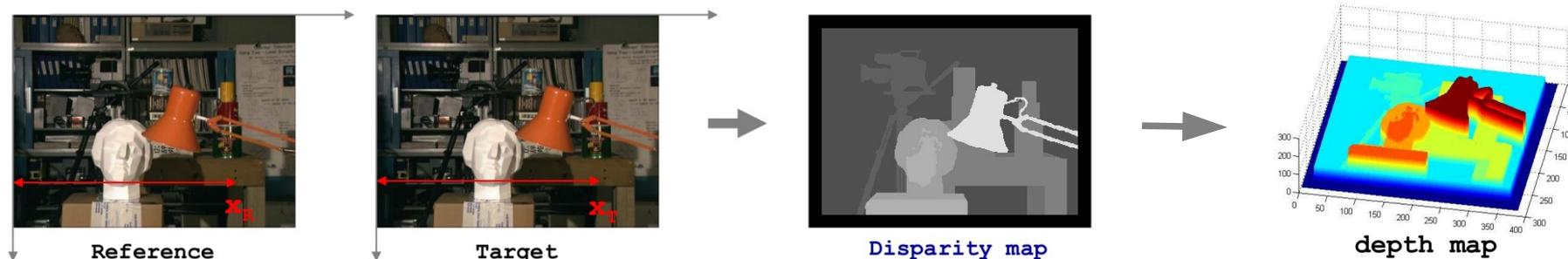


Stereo pair in standard form

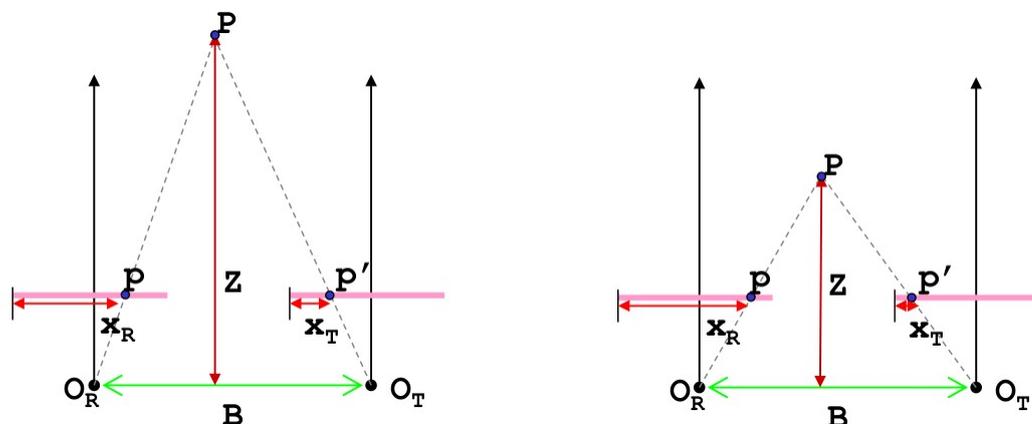
Danach liegen Punkte beider Bilder mit gleichem y-Wert auf der gleichen Epipolarlinie

Disparität und Tiefe

Die Disparität ist der Unterschied der x-Koordinaten zwei korrespondierender Punkte ($x_r - x_t$). Sie wird als Graustufenbild kodiert, wobei hellere Bereiche näher dran sind, daraus kann dann eine Tiefenkarte berechnet werden



Je näher ein Punkt an der Kamera liegt, desto größer ist die Disparität



Übersicht

- Rückblick
- 3D-Bildaufnahme
- Kamerakalibrierung
- Triangulation durch Active Vision
- Stereo Vision
- **Korrespondenzproblem / Matching-Strategien**
- Zusammenfassung und Ausblick

Das Lösen des Korrespondenzproblems ist nicht so trivial wie es scheint

Verzerrungen und Rauschen



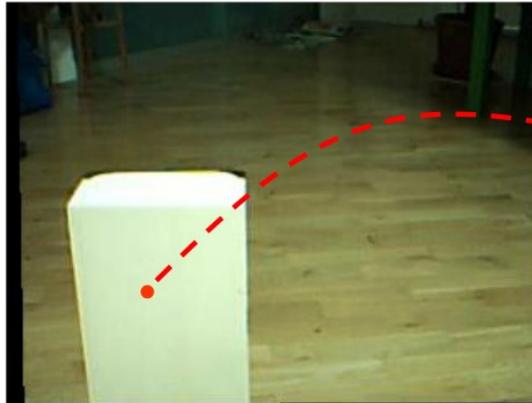
Spiegelnde Oberflächen



Verschiedene Ansichten



Homogene Oberflächen



Verfahren, die ich im Projekt 1 weiter untersuche:

◆ **Block Matching** nach Kurt Konolige

- Eines der schnellsten verfügbaren Verfahren
- Führt einen Ähnlichkeitsvergleich zwischen zwei Pixelblöcken durch
- Die erzeugte Disparitätskarte ist für manche Einsatzgebiete nicht genau genug

◆ **Graph-Cut base Matching** nach Vladimir Kolmogorov

- Langsam
- Liefert gute Ergebnisse

◆ **Semi-Global Block Matching** nach Heiko Hirschmüller

- Viel schneller als der Graph-Cut-Algorithmus bei annähernd gleich guten Ergebnissen
- Produziert eine sauberere Disparitätskarte als der Block-Match-Algorithmus

Übersicht

- **Rückblick**
- **3D-Bildaufnahme**
- **Kamerakalibrierung**
- **Triangulation durch Active Vision**
- **Stereo Vision**
- **Korrespondenzproblem / Matching-Strategien**
- **Zusammenfassung und Ausblick**

- ◆ Das große Problem beim Stereo Vision ist die Korrespondenz zwischen den Pixeln beider Bilder
- ◆ Es gibt unterschiedliche Verfahren, die das Problem mit unterschiedlicher Genauigkeit und Geschwindigkeit lösen
- ◆ Ziel für Projekt 1: Ein passendes Verfahren für die Lösung des Korrespondenzproblems finden und implementieren
- ◆ Ziel für Projekt 2: Ein passendes Verfahren für das Erkennen und Verfolgen von 3D-Objekten finden und implementieren

Fragen...