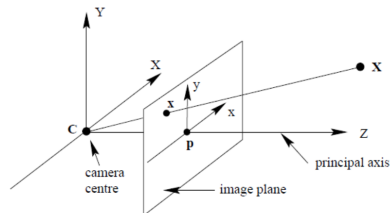
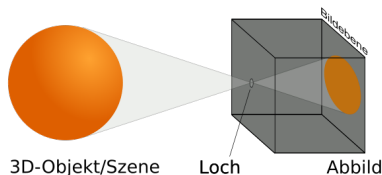


Computer Vision: 3D-Geometrie

Lochkamera Modell



C – Projektionszentrum, Optische Achse, Bildebene,

P – Hauptpunkt (optische Achse kreuzt die Bildebene), x – Bildpunkt, X – Weltpunkt

3D-Koordinatensystem: Ursprung im Kamerazentrum, Z -Achse – optische Achse

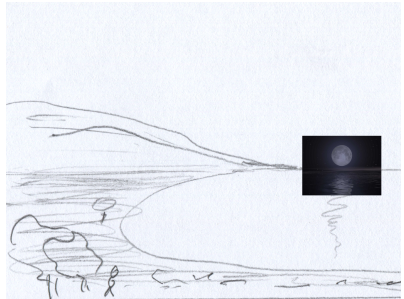
Weltpunkt $X = [X_1, X_2, Z]$

x ist ein Punkt im 3D, d.h. $x = [x_1, x_2, f]$ mit Brennweite f

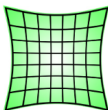
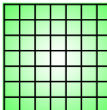
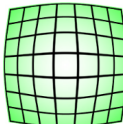
Projektive Abbildung:
$$\begin{aligned} x_1 &= X_1 \cdot f / Z \\ x_2 &= X_2 \cdot f / Z \end{aligned}$$

Nicht eineindeutig – einem x entsprechen alle X , die auf dem Strahl $\lambda(x - C) = \lambda x$ liegen.

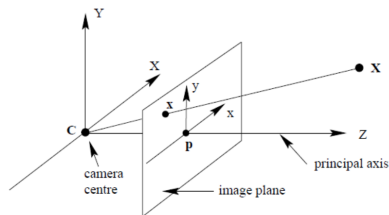
Zoom:



Radiale Verzerrung:

Kissenförmige Verzerrung	Verzerrungsfreie Abbildung (Idealfall)	Tonnenförmige Verzerrung
		

Homogene Koordinaten



Ein „Vektor“ $h \in \mathbb{R}^3$, d.h. $h = [h_1, h_2, h_3]$ beschreibt die Menge der Strahlen im $\mathbb{R}^3$ mit $X = \lambda \cdot [h_1/h_3, h_2/h_3, 1]$

Topologisch gesehen ist diese Menge zweidimensional – redundante Beschreibung.

Die Menge der Weltpunkte im $\mathbb{R}^3$ wird auf Teilmengen partitioniert.

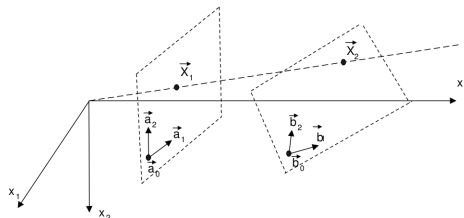
Zwei Punkte X und X' gehören einer Teilmenge (Äquivalenzklasse, Strahl), wenn $X_1/Z = X'_1/Z'$ und $X_2/Z = X'_2/Z'$

In homogenen Koordinaten gibt es keine projektive Abbildung mehr, denn X und x sind bereits in einer Äquivalenzklasse – ein Strahl.

Abbildung „Homogene Koordinaten $\rightarrow$ Bildkoordinaten“:

$$\begin{aligned} x_1 &= h_1 \cdot f/h_3 \\ x_2 &= h_2 \cdot f/h_3 \end{aligned}$$

Wie wird eine Ebene im $\mathbb{R}^3$ auf eine andere Ebene im $\mathbb{R}^3$ (z.B. Bildschirm) Abgebildet?



Sei $[\lambda_1, \lambda_2]$ Koordinaten eines Punktes im Koordinatensystem der abzubildenden Ebene.

Die Lage der abzubildenden Ebene im $\mathbb{R}^3$ wird durch: $b_0 \in \mathbb{R}^3$ (ein Punkt in der Ebene) und $b_1, b_2 \in \mathbb{R}^3$ (Basis, der die Ebene aufspannt) angegeben.

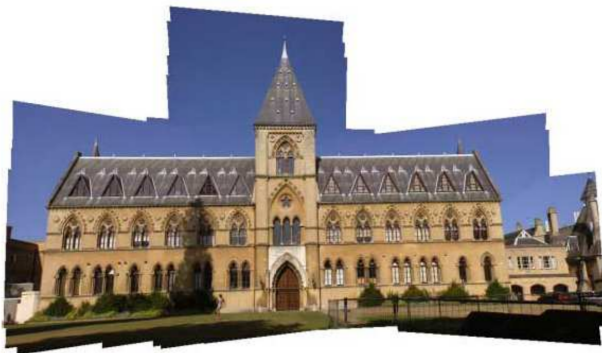
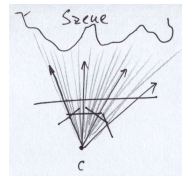
Die 3D-Punkte der Ebene sind somit $b_0 + \lambda_1 b_1 + \lambda_2 b_2$, d.h. $X = H \cdot [\lambda_1, \lambda_2, 1]$.

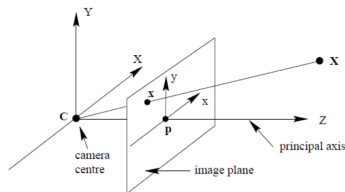
Das sind zugleich die homogenen Koordinaten.

Homographie ist eine lineare Transformation in homogenen Koordinaten.

Homographie – Panoramas

Viele Ebenen (Bildschirme) werden auf eine abgebildet.





Kameraparameter (keine Verzerrung, quadratische Pixel):

- intrinsische Parameter – Brennweite f , Hauptpunkt $[P_x, P_y]$
- extrinsische Parameter – Position des Kamerakoordinatensystem in der Welt (3 Winkel in Form der 3×3 Rotationsmatrix R und 3D-Verschiebung t)

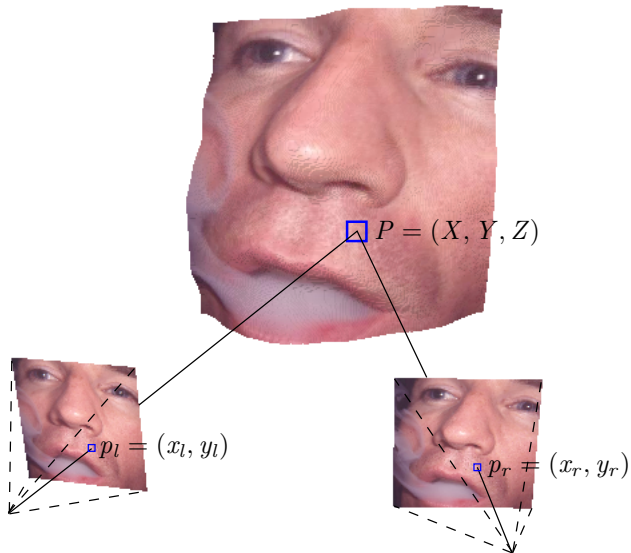
$$x = PX = \begin{bmatrix} f & 0 & 0 & P_x \\ 0 & f & 0 & P_y \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Kamerakalibrierung:

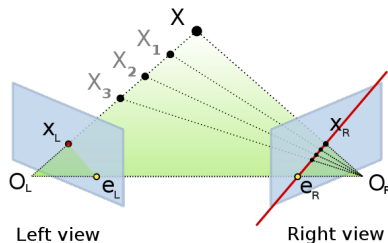
Gegeben sei die Lernstichprobe $((X, x) \dots)$, Gesucht werden die Kameraparameter.

Bestimmung der Lage eines Objektes:

Gegeben sind die Kameraparameter und Abbilder x , Gesucht werden die X .



Nicht jedes beliebige Paar von Bildpunkten entspricht einem Weltpunkt!!!



Sei $t = O_L - O_R$ der Verschiebungsvektor.

$x_l + t$ liegt in der (grünen) Ebene π (Koordinatensystem der linken Kamera wird um t verschoben)

$(x_l + t) \times t = (x_l + t)[t]_{\times}$ steht senkrecht zur π ($[t]_{\times}$ ist der Kreuzprodukt in der Matrixform)

Derselbe Vektor in der Koordinatensystem der rechten Kamera ist $a = (x_l + t)[t]_{\times} R$

x_r liegt in π , d.h. $\langle a, x_r \rangle = 0$

Schließlich: $(x_l + t)[t]_{\times} R x_r = x_l[t]_{\times} R x_r + t[t]_{\times} R x_r = x_l[t]_{\times} R x_r = \mathbf{x}_l \mathbf{E} \mathbf{x}_r = 0$
mit **Essential Matrix** $E = [t]_{\times} R$

Essential Matrix E

entspricht der Relation zwischen „metrischen“ Koordinatensystemen
(z -Achse stimmt mit der optischen Achse überein, alle Abstände sind in mm gemessen)

Fundamentalmatrix F

entspricht der Relation zwischen „Kamerakoordinatensystemen“
(z -Achse ist zu der optischen Achse parallel, alle Abstände sind in Pixeln gemessen)

Zusammenhang:

$E = K_l F K_r$ bzw. $F = K_l^{-1} E K_r^{-1}$ mit den **Kameramatrizen** K_l und K_r

$$K_{l|r} = \begin{bmatrix} s_x & 0 & P_x \\ 0 & s_y & P_y \\ 0 & 0 & 1 \end{bmatrix}$$

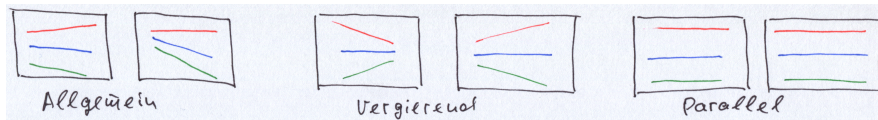
s_x und s_y sind Pixelgrößen, (P_x, P_y) ist der Hauptpunkt.

Für eine gegebene Konfiguration der Kameras **gibt es** eine Matrix F so, dass

$$x_l^T F x_r = \begin{bmatrix} x_{l1} & x_{l2} & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_{r1} \\ x_{r2} \\ 1 \end{bmatrix} = 0$$

für alle Paare (x_l, x_r) der korrespondierenden Punkte gilt.

Epipolarlinien: Betrachtet man zu einem bestimmten x_l alle gültigen x_r , so gilt $x_l F x_r = \langle x_l F, x_r \rangle = 0$ (Liniengleichung im rechten Bild).



Wichtiger Spezialfall: parallele Anordnung – die Bilder sind **rektifiziert**:

- die optischen Achsen der Kameras sind zu einander parallel,
- senkrecht zur Linie, die Projektionszentren verbindet,
- die Bildschirme sind nicht „verdreh“

Epipolargeometrie – Unterraum

Die Menge **aller Punktepaare** (x_l, x_r) ist $\mathbb{R}^4$,

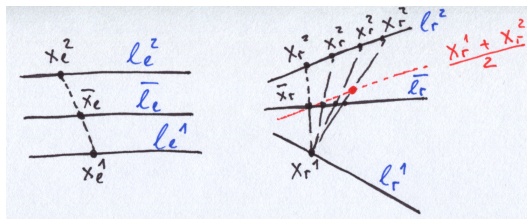
Die Menge **aller gültigen Korrespondenzpaare** (für die $x_l F x_r = 0$ gilt) ist dem 3D-Raum homeomorph, d.h. $\mathbb{R}^3$

$\Rightarrow$ die Bedingung $x_l F x_r = 0$ definiert eine 3D-Untermannigfaltigkeit im $\mathbb{R}^4$.

Frage: ist diese ein **affiner** Raum?

Ein Raum $\mathcal{R}$ heißt affin, wenn $x, y \in \mathcal{R} \Rightarrow \alpha x + (1 - \alpha)y \in \mathcal{R}$

Antwort: **Nein** im Allgemeinen

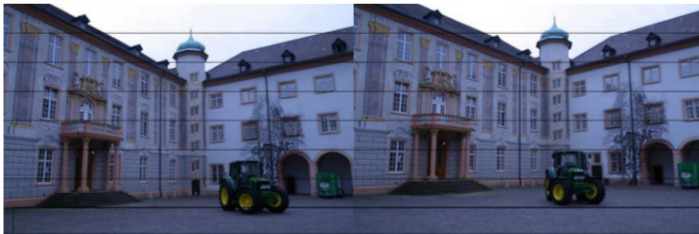


Man betrachtet zwei „Punkte“ im $\mathbb{R}^4$ (zwei Korrespondenzpaare (x_l^1, x_r^1) und (x_l^2, x_r^2)), die der 3D-Untermannigfaltigkeit gehören, d.h. $x_l^1 F x_r^1 = 0$ und $x_l^2 F x_r^2 = 0$.

Man untersucht z.B. den „Mittelpunkt“ $(\bar{x}_l, \bar{x}_r) = 1/2((x_l^1, x_r^1) + (x_l^2, x_r^2))$ und sieht, dass dieser nicht unbedingt die Bedingung der Epipolargeometrie erfüllt.

Rektifizierung

Auf beide Bilder werden Homographien angewendet so, dass die Bilder rektifiziert sind. Dafür ist das Wissen über die Szene nicht (unbedingt) benötigt.



Gesamtaufgabe wird oft in zwei Teilaufgaben zerlegt:

1. gegeben sei ein Bildpaar,
gesucht wird eine Lernstichprobe der Korrespondenzpaare $((x_l, x_r) \dots)$.
2. gegeben sei eine Lernstichprobe der Korrespondenzpaare,
gesucht wird die Fundamentalmatrix F

Es gibt Ansätze, die diese zwei Phasen nicht trennen, d.h. sowohl F als auch die „dazu passenden“ Korrespondenzpaare werden **gleichzeitig** ermittelt.

Zur zweiten Aufgabe:

- Die Matrix F ist bis auf Skalierung definiert: $x_l^T F x_r = 0 \Leftrightarrow x_l^T (aF) x_r = 0$
Ein möglicher Ausweg: $f_{33} = 1$ setzen – funktioniert nicht, wenn $f_{33} = 0$
(z.B. beim rektifizierten Fall :-))
Andere Variante: $\|F\| = 1 \dots$
- Es gilt $\text{rang}(F) = 2$

Die zu lösende Aufgabe hat **9** unbekannte Parameter (Koeffizienten von F)
aber nur **7 Freiheitsgrade**.

8-Punkte Algorithmus

Die Forderung $\text{rang}(F) = 2$ wird zunächst vernachlässigt.

$$x_l F x_r = \begin{bmatrix} x_{l1} & x_{l2} & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_{r1} \\ x_{r2} \\ 1 \end{bmatrix} = 0$$

kann komponentenweise wie folgt umgeschrieben werden:

$$x_{l1} x_{r1} \cdot f_{11} + x_{l1} x_{r2} \cdot f_{12} + x_{l1} \cdot f_{13} + x_{l2} x_{r1} \cdot f_{21} + x_{l2} x_{r2} \cdot f_{22} + x_{l2} \cdot f_{23} + x_{r1} \cdot f_{31} + x_{r2} \cdot f_{32} + f_{33} = 0$$

(eine bezüglich f_{ij} lineare Gleichung mit 9 unbekannten).

$\Rightarrow$ 8 Gleichungen (8 bekannte Korrespondenzpaare) zusammen mit $\|F\| = 1$ werden benötigt um F zu schätzen.

Problem: die Messungen x_l, x_r sind ungenau:

$\Rightarrow$ die resultierende Matrix hat den vollen Rang (d.h. entspricht nicht der Realität).

Einfache Lösung:

zerlege die resultierende F (Singular Value Decomposition) in $F = USV^T$

S ist eine Diagonalmatrix mit singulären Werten

setze den kleinsten singulären Wert auf Null $S_{,,} \rightarrow S'$ und

baue die endgültige $F = US'V^T$ wieder zusammen.

Sehr oft stehen viel mehr als 8 Korrespondenzpaare zur Verfügung
⇒ „Mitteln“ ⇒ Genauigkeit erhöhen.

$$\sum_i (x_l^i F x_r^i)^2 = \sum_i \langle \tilde{x}^i, F \rangle^2 \rightarrow \min_F$$

s.t. $\|F\| = 1$

Die Lösung – der Eigenvektor der Matrix $\sum_i \tilde{x}^i \otimes \tilde{x}^i$ zum kleinsten Eigenwert.
(Erinnere an PCA)

Die Forderung $\text{rang}(F) = 2$ wird explizit behandelt.

7 Korrespondenzpaare (minimal) werden benötigt + die Einschränkung $\det(F) = 0$
⇒ Polynom 3. Grades – analytisch lösbar.

Mehr als 7 Punkte – ein Gleichungssystem mit nicht linearen Einschränkungen (?).

Schwieriger zu lösen ↔ dafür aber in der Regel bessere Ergebnisse.

Suche nach den Korrespondenzpaaren

Typische Reihenfolge:

1. Besondere Punkte werden gesucht (für beide Bilder getrennt)
2. Die potentielle Korrespondenzpaare werden bewertet:
 - a) Quadratische Abstand zwischen SIFT Deskriptoren,
 - b) Block Matching Maße
 - c) ...
3. Korrespondenzpaare werden ermittelt – ↓

Etwas mehr dazu später bei der Erkennung.

Cross-Validierung:

Für jeden Punkt des linken Bildes wird der beste Punkt im rechten Bild gesucht. Analog für jedes Punkt des rechten Bildes

Haben sich zwei Punkte „gegenseitig gefunden“, so bilden sie ein Korrespondenzpaar, anderenfalls werden die Punkte verworfen.

⇒ viele „gute“ Punkte werden verloren,
es bleibt aber immer noch genügend viel für die Ermittlung der Epipolargeometrie.

Heiratsproblem, Matching von bipartiten Graphen etc.