

# 10. Stereo und Motion

## Stereo Vision

[EVC\\_Skriptum\\_CV, p.51](#)

- Zusammengesetzt aus griechisch "stereos" (räumlich, fest) und lateinisch "videre" (sehen).
- Bezeichnet räumliches Sehen bzw. binokulares Sehen.
- Ziel: Erstellung eines Tiefenbildes aus zwei Bildern einer Szene.
- Bilder sind 2D-Projektionen einer 3D-Szene, wobei eine Dimension (die Tiefe) verloren geht.
- Der Mensch kann Tiefeninformationen aus seiner Umgebung durch binokulares Sehen gewinnen.

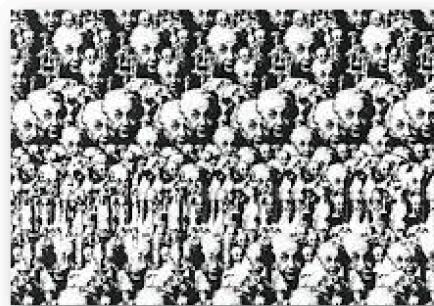
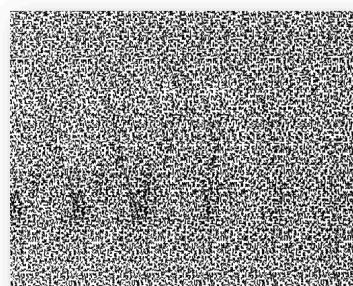
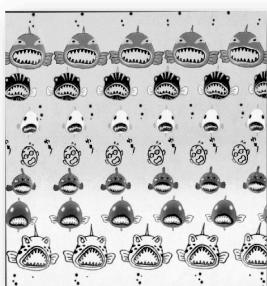
Prinzip der Stereo Vision:

- Nutzt zwei Kameras mit bekanntem Abstand zueinander.
- Anwendung geometrischer Prinzipien (Triangulation und Epipolargeometrie) zur Berechnung korrespondierender Bildpunkte.
- Rekonstruktion der Tiefenwerte.



For some people, Random Dot Stereograms are complicated to view:  
Autostereograms [Tyler77]:

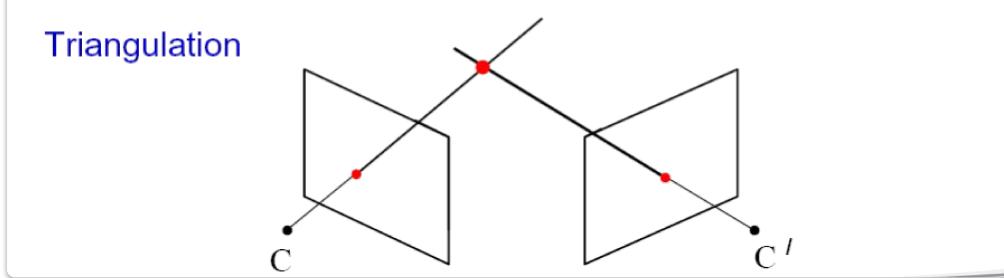
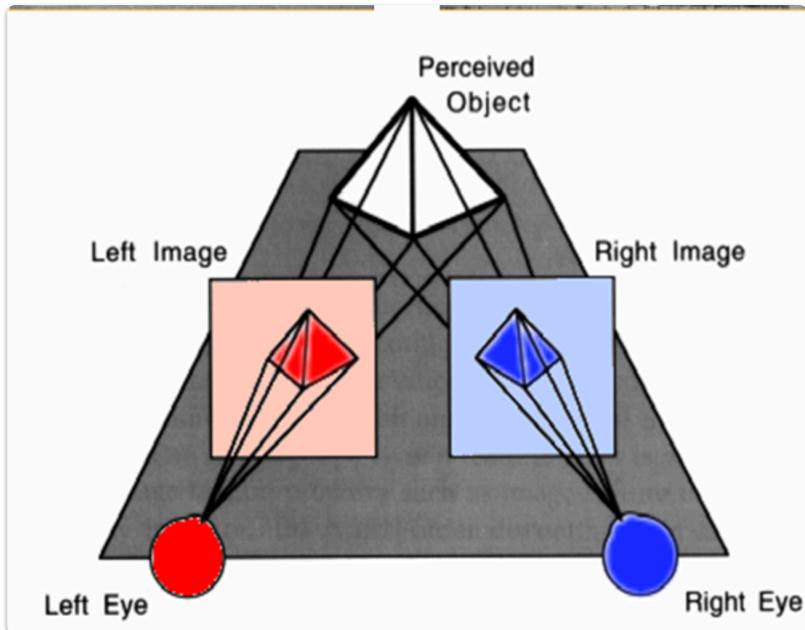
- Also called single image Stereogram
- Form of Art
- Is formed by repetition of patterns in specific intervals



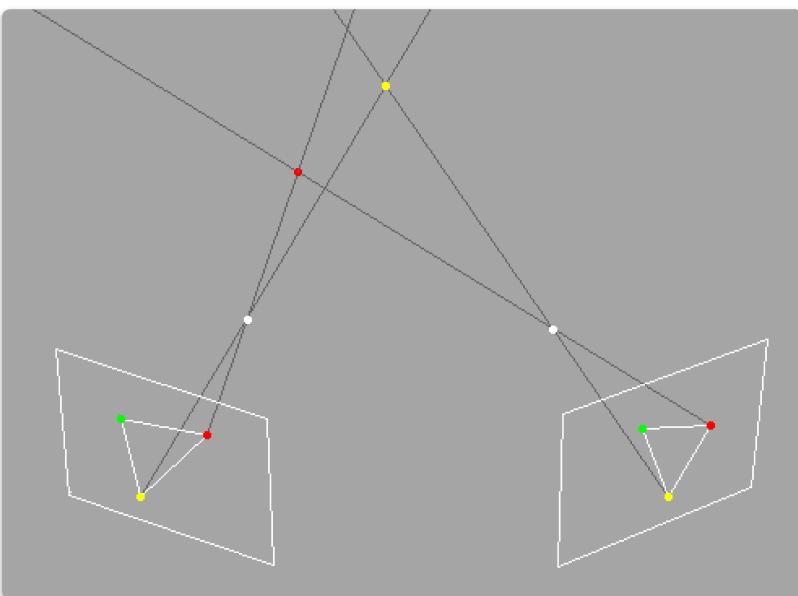
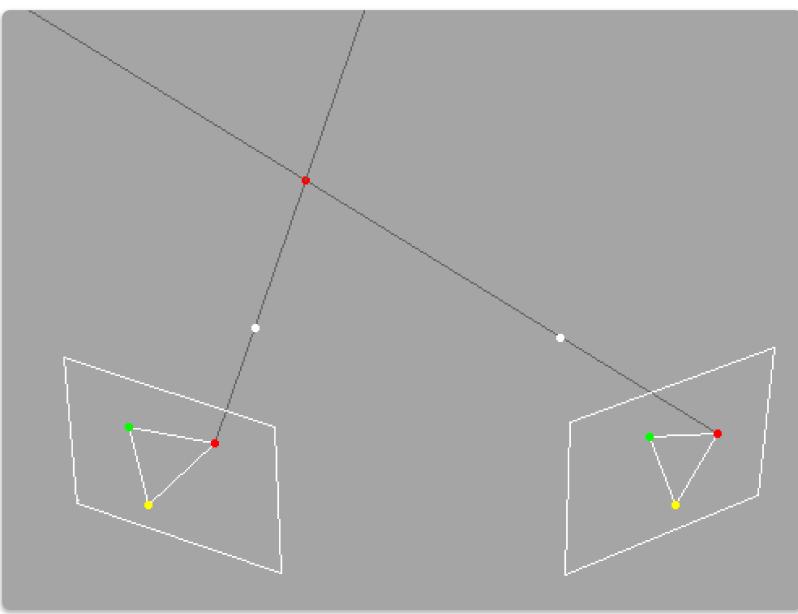
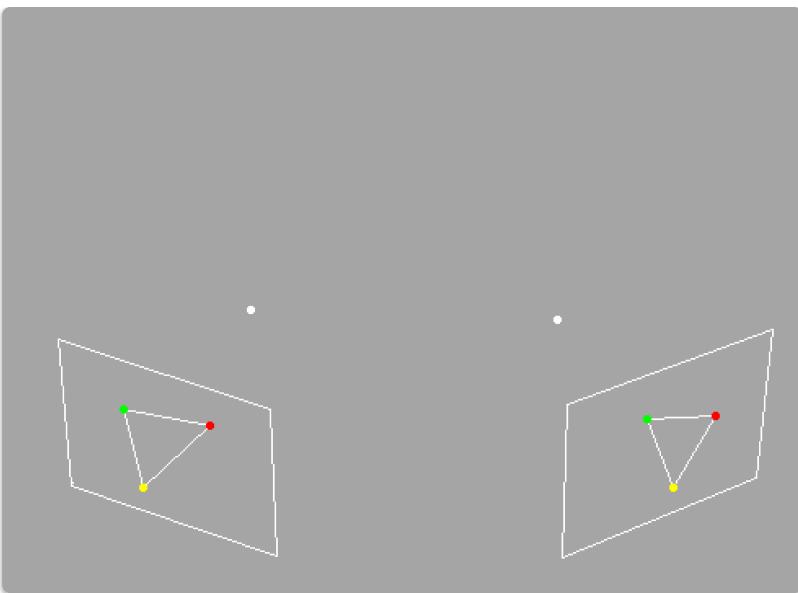
Tyler C.W. and Chang J.J. (1977) [Visual echoes: The perception of repetition in quasi-random patterns](#). *Vision Res.* 17, 109-116.

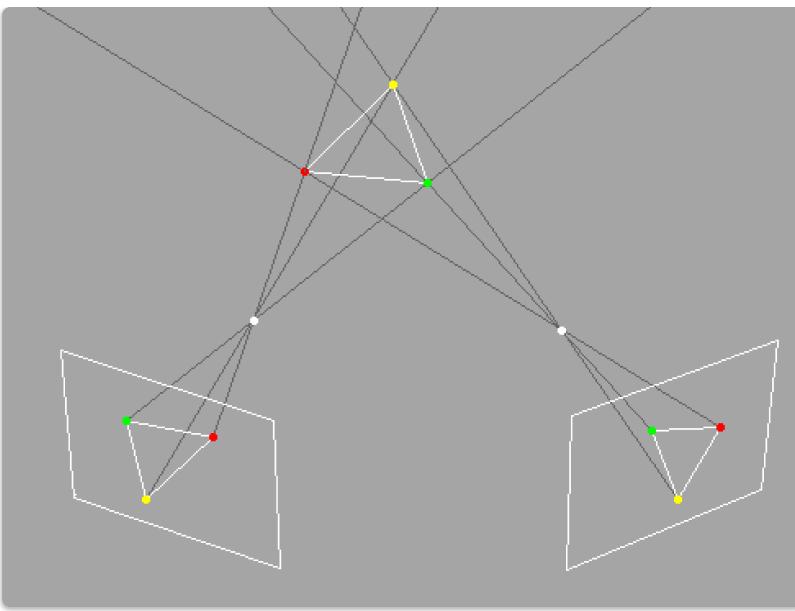
## Disparität:

- Leichter Versatz der beiden Kameras erzeugt unterschiedliche Bilder.
- Der Versatz in den Kamerabildern ist geringer für entfernte Objekte und größer für nahe Objekte.
- Dieser Versatz wird als Disparität bezeichnet.
- Durch Zuweisung der unterschiedlichen Disparitäten zu jedem Bildpunkt wird eine Disparitätsmatrix erstellt.
- Aus der Disparitätsmatrix lässt sich für jeden Bildpunkt die Tiefe ableiten.



**Triangulation: Fundament von Stereo Vision**

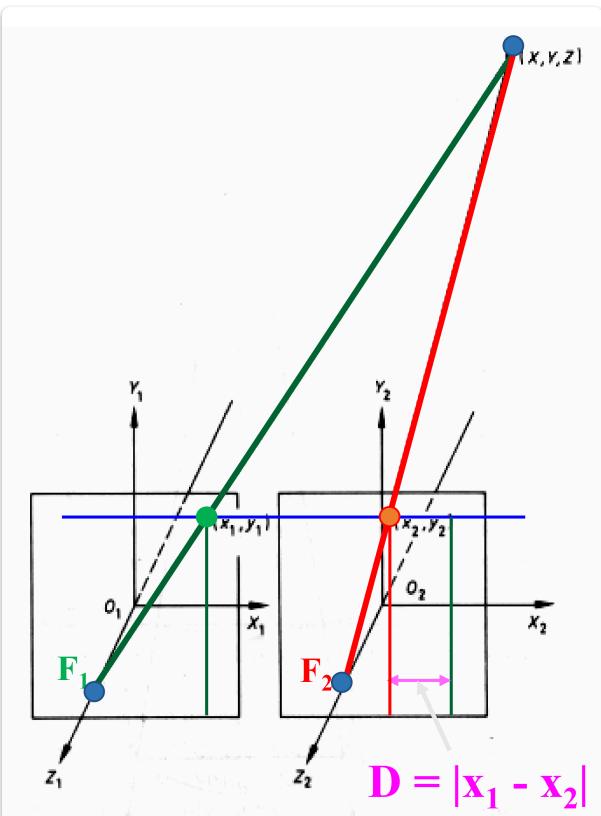




## Stereo Geometry

Grundlagen:

- Betrachtet werden *Perspektivische Projektionen* von Objektpunkten.
- **Spezifischer Fall (vereinfacht):** Beide Bildebene sind parallel zueinander ausgerichtet.
  - Dieser spezifische Fall kann zur Erklärung des generellen Falls herangezogen werden.
- Durch die Nutzung der geometrischen Position der Bildebene und der Information über die Tiefenprojektion des Objekts kann die Tiefe des Objekts berechnet werden.



Wichtige Elemente in der Abbildung:

- $(X, Y, Z)$ : Ein Punkt im 3D-Raum.
- $F_l$ : Projektionszentrum der linken Kamera.
- $F_r$ : Projektionszentrum der rechten Kamera.
- Bildebenen (parallel zueinander).
- $x'_l$ : Projektion des 3D-Punktes auf der linken Bildebene.
- $x'_r$ : Projektion des 3D-Punktes auf der rechten Bildebene.
- $D = |x'_l - x'_r|$ : Die *Disparität* – der horizontale Abstand zwischen den korrespondierenden Bildpunkten.

Zusammenfassend: Stereo Vision nutzt zwei leicht versetzte Kameras, um durch die Analyse der Disparität zwischen den beiden resultierenden Bildern Tiefeninformationen zu gewinnen. Die geometrische Beziehung zwischen den Kameras ermöglicht die Berechnung der 3D-Positionen der Punkte in der Szene.

## Stereoskopie

---

[EVC\\_Skriptum\\_CV](#), p.51

Stereoskopie (griechisch "skopeo" = betrachten):

- Wiedergabe von Bildern mit einem räumlichen Eindruck von Tiefe.
- Die Tiefe ist physikalisch nicht vorhanden.
- Umgangssprachlich fälschlich als "3D" bezeichnet, obwohl es sich um zweidimensionale Abbildungen handelt, die einen räumlichen Eindruck vermitteln.

Prinzip des räumlichen Sehens:

- Bereits im 3. Jh. v. Chr. von dem griechischen Mathematiker Euklid beschrieben.
- Viele Wissenschaftler (u.a. Leonardo da Vinci) beschäftigten sich mit diesem Phänomen.

Geschichte der Stereoskopie:

- 19. Jh.: Charles Wheatstone entdeckte die Stereoskopie.
  - Hielt 1838 einen bahnbrechenden Vortrag über "einige merkwürdige und bisher nicht beobachtete Erscheinungen beim beidäugigen Sehen".
  - Berechnete und zeichnete Bildpaare.
  - Konstruierte das Stereoskop, ein Apparat, um diese Bildpaare räumlich betrachten zu können



Weitere Entwicklung:

- 1849: David Brewster stellte die erste Stereokamera vor.
  - Ermöglichte erstmals, ein bewegtes Motiv aufzunehmen (allerdings noch nicht für Fotografie im heutigen Sinne).
- 1851: Durchbruch auf der Weltausstellung in London.

- Hardware for Viewing (3D-TV sets):
  - Anaglyph
  - Polarized
  - Field-sequential (Active shutter)
  - Lenticular display

Anaglyph

Polarizing

Active shutter

Lenticular

Zusammenfassend: Die Stereoskopie erzeugt einen räumlichen Eindruck aus zwei leicht unterschiedlichen, zweidimensionalen Bildern. Historisch gesehen reicht die Beobachtung dieses Phänomens bis in die Antike zurück, die eigentliche Erfindung des Stereoskops und der Stereokamera erfolgte jedoch im 19. Jahrhundert.

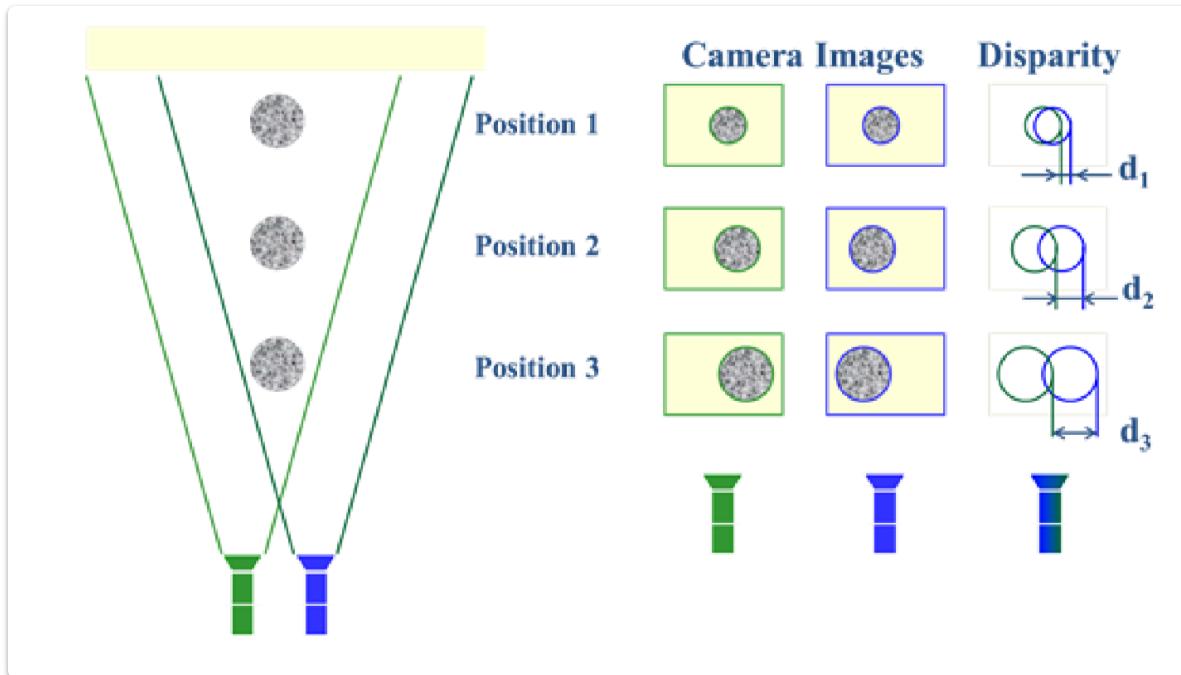
## Disparität

[EVC\\_Skriptum\\_CV, p.52](#)

Definition:

- Ein einzelner Punkt wird in beiden Bildern des Stereosystems auf unterschiedliche Bildkoordinaten abgebildet.

- Die **horizontale Differenz** zwischen diesen Bildkoordinaten nennt man **Disparität**.



Erläuterung anhand paralleler Kameras (siehe Abbildung 44):

- Eine Kugel wird an drei verschiedenen Positionen (Abständen zum Kamerapaar) aufgenommen.
- In der Überlagerung der beiden Kamerabilder kann die Disparität beobachtet werden.
- Position 1 (weit entfernt):**
  - Unterschied der Position der Kugel im überlagerten Bild ( $d_1$ ) ist klein.
- Position 2 und 3 (näher):**
  - Die Bilder der Kugel werden größer, da sie näher sind.
  - Der Unterschied der Punkte in den überlagerten Bildern ( $d_2, d_3$ ) wird größer.
  - Die Disparität wird größer ( $d_3 > d_2 > d_1$ ).

Zusammenhang zwischen Disparität und Tiefe:

- Je näher das Objekt zur Kamera ist, desto größer ist die Disparität.
- Im Unendlichen ist die Disparität 0.
- Die Disparität ist somit *umgekehrt proportional* zur Tiefe.

## Normalfall (Achsparalleles Stereosystem)

[EVC\\_Skriptum\\_CV, p.52](#)

Definition:

- Zeichnet sich durch zwei Kameras aus, die horizontal verschoben sind.
- Deren Koordinatensysteme sind nicht gegeneinander verdreht.
- Der Abstand zwischen den beiden optischen Zentren wird **Basislinie** ( $B$ ) genannt.

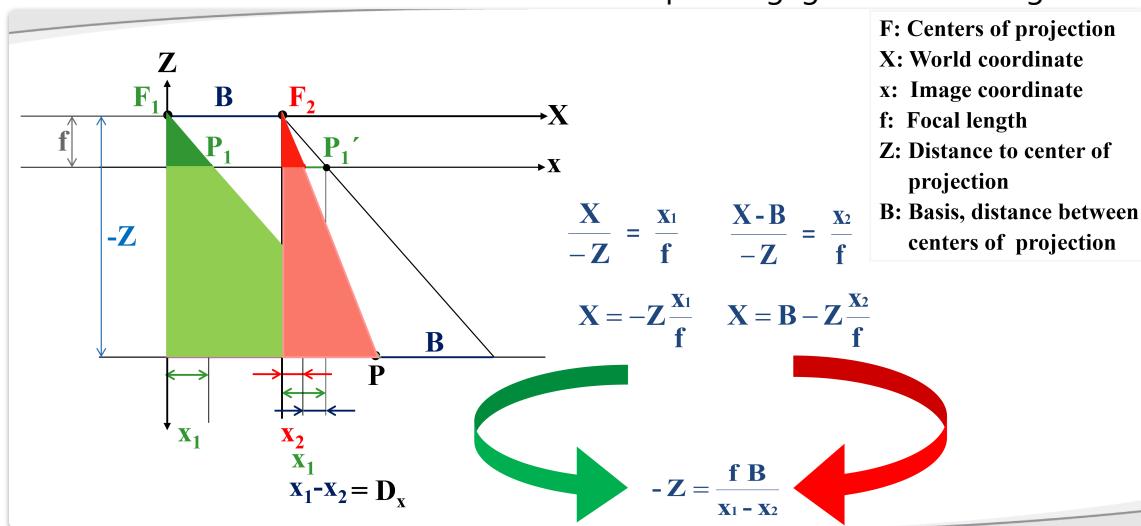
- Die Brennweite ( $f$ ) legt den Abstand der beiden Brennpunkte zu ihren Bildebenen fest und ist für beide Kameras als identisch vorausgesetzt.
- Ein 3D-Punkt  $X$  wird somit über die beiden optischen Zentren in den Abbildungen  $x$  und  $x'$  projiziert.
- Beim achsparallelen Stereosystem sind die Bildzeilen identisch, was für die unterschiedliche Perspektive der Kameras hinsichtlich des 3D-Punktes  $X$  nur zu einer horizontalen Disparität  $D$  in der Abbildung führt.
- Die Disparität wird im Allgemeinen in Bildkoordinaten berechnet, sodass die Einheit Pixel ist:  $D = |x_l - x_r|$ .

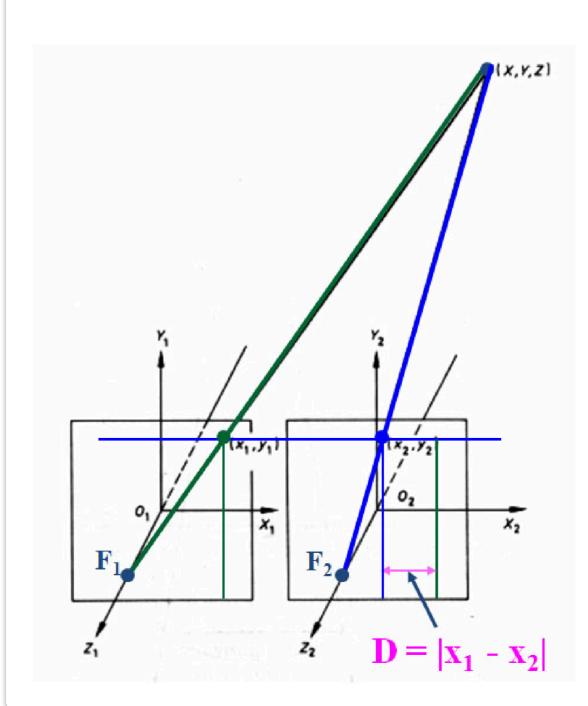
Tiefenberechnung:

- Der Abstand  $Z$  eines Punktes  $X$  von der Kamera lässt sich aus den bekannten konstanten Kameraparametern  $f$ ,  $B$  sowie der Disparität  $D$  berechnen:

$$Z = \frac{B \cdot f}{D}$$

- Damit stellt die Disparität ein Maß für die Raumtiefe des 3D-Punktes  $X$  dar und verhält sich *umgekehrt proportional* zu ihr.
- Für Punkte im Unendlichen muss daher die Disparität gegen Null konvergieren.



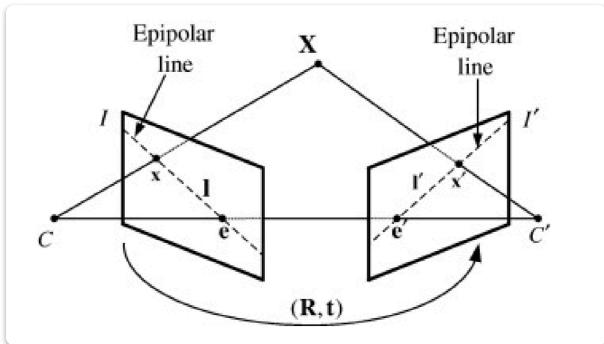


## Epipolargeometrie

[EVC\\_Skriptum\\_CV, p.52](#)

Kameraanordnungen mit zwei Kameras im Stereosystem:

- Können bezüglich ihrer räumlichen Anordnung in zwei grundlegende Klassen eingeteilt werden:
  - Das bereits vorgestellte *achsparallele Stereosystem (Normalfall)*.
  - Die *konvergente Anordnung* (Ausrichtung der optischen Achsen auf einen Konvergenzpunkt).
- Bei der allgemeineren Stereogeometrie, auch *Epipolargeometrie* genannt, sind die beiden Kameras nicht nur zueinander verschoben, sondern auch zueinander gedreht.



Wichtige Begriffe:

- **Epipole (e und e')**: Die Schnittpunkte der Verbindungsgeraden der beiden Kamerazentren (Basislinienebene) mit den jeweiligen Bildebenen. Die Epipole können auch als Projektion des optischen Zentrums der einen Kamera in der Bildebene der anderen Kamera aufgefasst werden.

- **Epipolarebene:** Die Ebene, die durch den 3D-Punkt  $X$  und die beiden Brennpunkte  $C$  und  $C'$  aufgespannt wird.
- **Epipolarlinien ( $l$  und  $l'$ ):** Die Schnittlinien der Epipolarebene mit den beiden Bildebenden. Betrachtet man die beiden Sehstrahlen des 3D-Punktes in den beiden Kameras als Gummiband, so bewegt man diesen Punkt innerhalb der Epipolarebene, wobei diese jedoch immer auf den Epipolarlinien liegen.

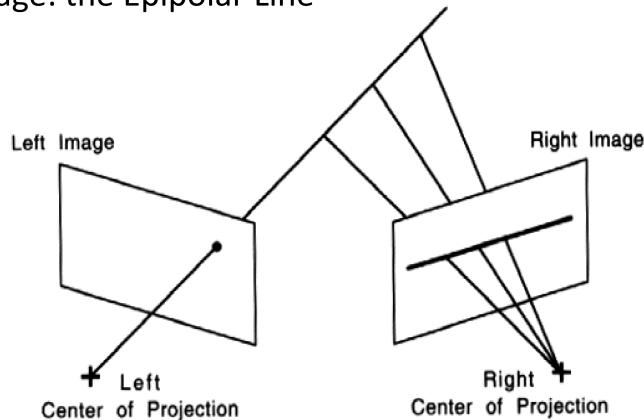
Bedeutung der Epipolarlinien:

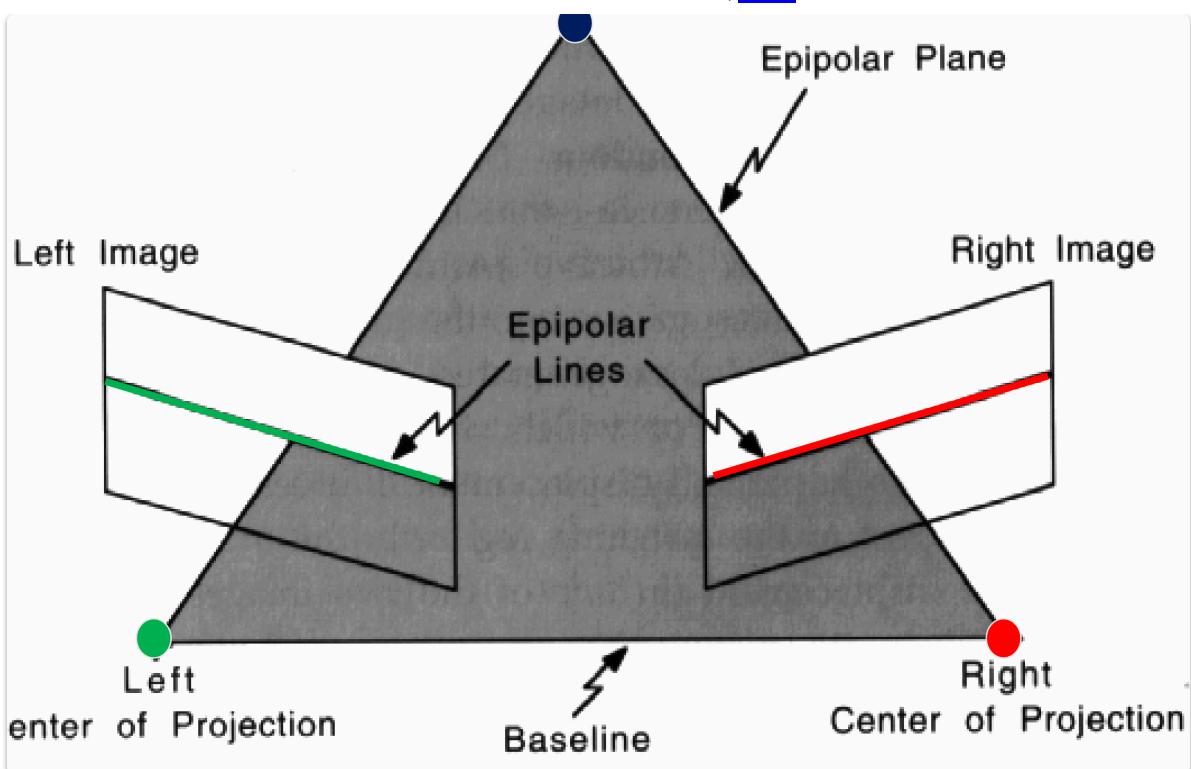
- Lässt man den 3D-Punkt  $X$  entlang seines Sehstrahles (z.B. in Richtung Kamera 1) laufen, so ergibt sich immer die gleiche Abbildung  $x$  in Kamera 1, während in Kamera 2 die Abbildung  $x'$  entlang der Epipolarlinie  $l'$  wandert.



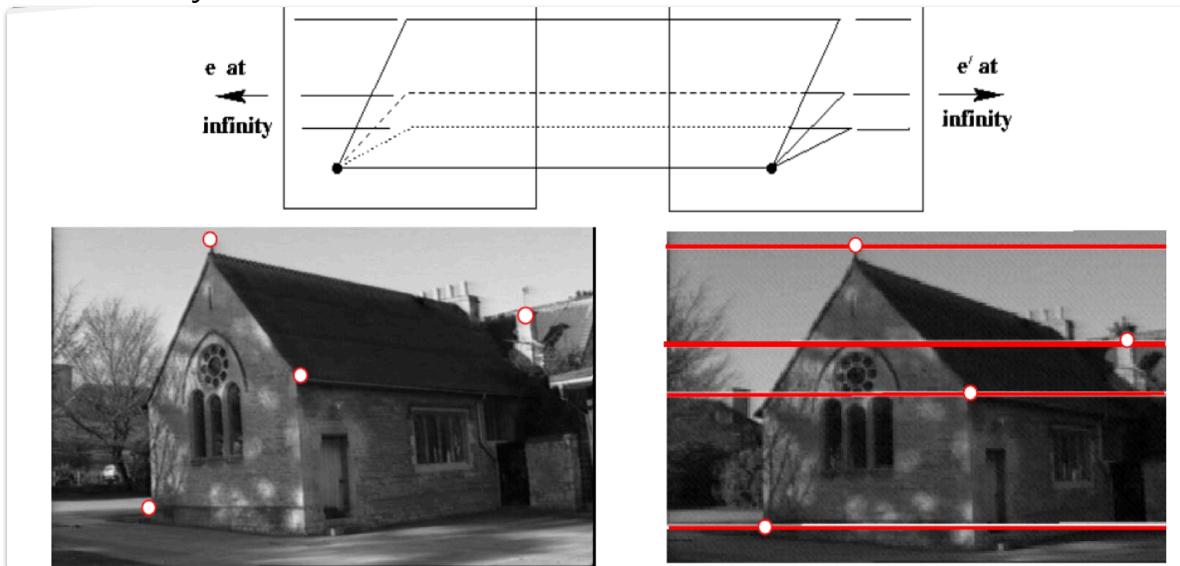
- Der Sehstrahl von jedem 3D-Punkt in einer Kamera liefert somit als Projektion in der anderen Kamera die entsprechende Epipolarlinie.
- Folglich muss auch für jeden Bildpunkt in einer Kamera der korrespondierende Punkt auf einer der Epipolarlinien in der anderen Kamera liegen.

- **Epipolar Constraint:** Each point of the left image can lie only on a specific line in the right image: the Epipolar Line





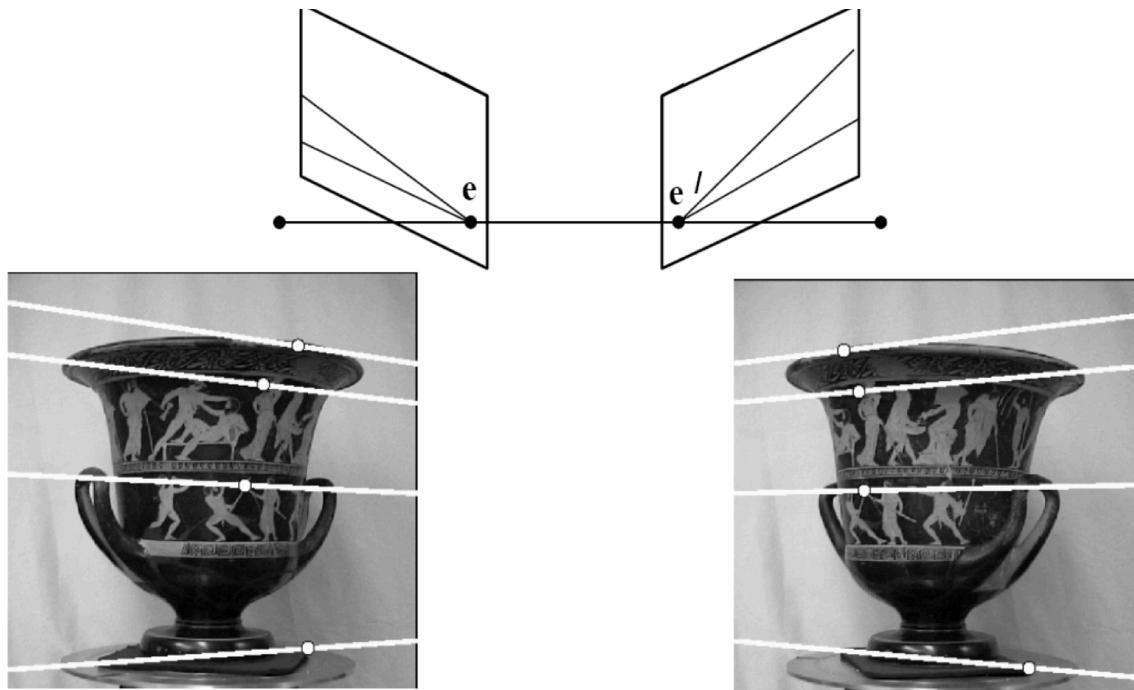
Das heißt für jeden dieser Punkte muss ich nur eine Zeile fahren:



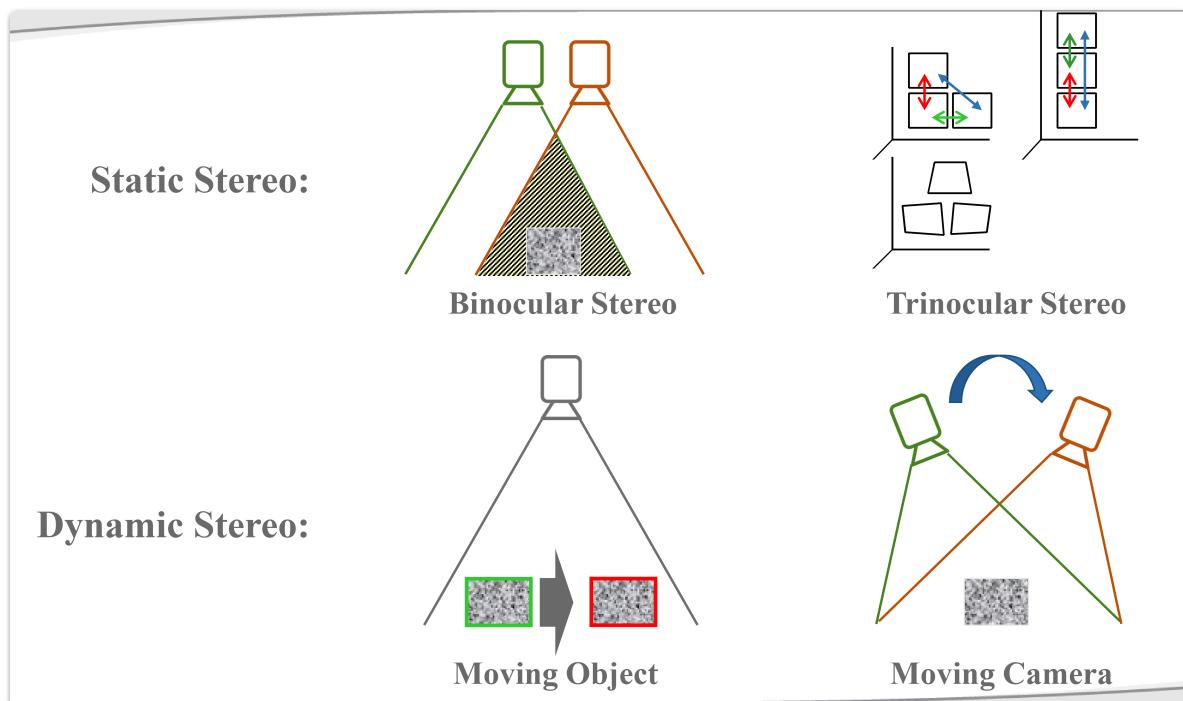
Wichtige Eigenschaften:

- Die Epipolargeometrie hängt **nur** von der relativen Pose (Position und Orientierung) und den internen Parametern der beiden Kameras ab.
  - Dazu gehören die Position der Kamerazentren und der Bildebenen.
- Sie hängt **nicht** von der Szenenstruktur ab (den 3D-Punkten außerhalb der Kameras).

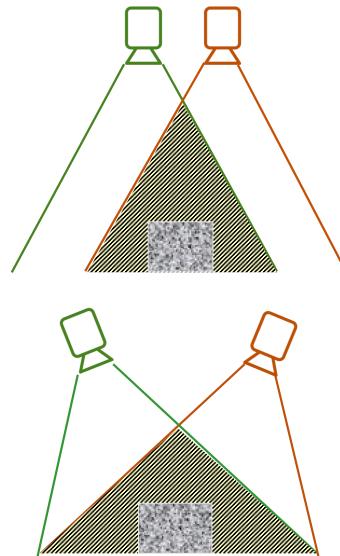
Im Normalfall sind diese Epipole nicht parallel:



## Camera Setup



- Baseline
  - Distance between cameras (focal points)
- Trade-off
  - Small baseline: Matching easier
  - Large baseline: Depth precision better

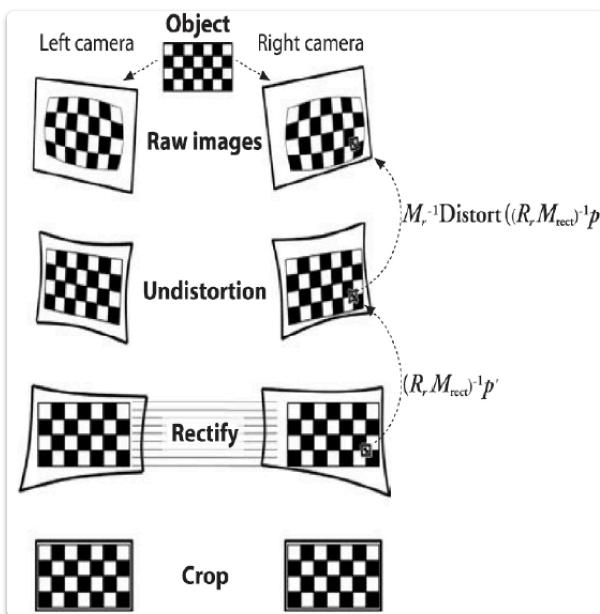


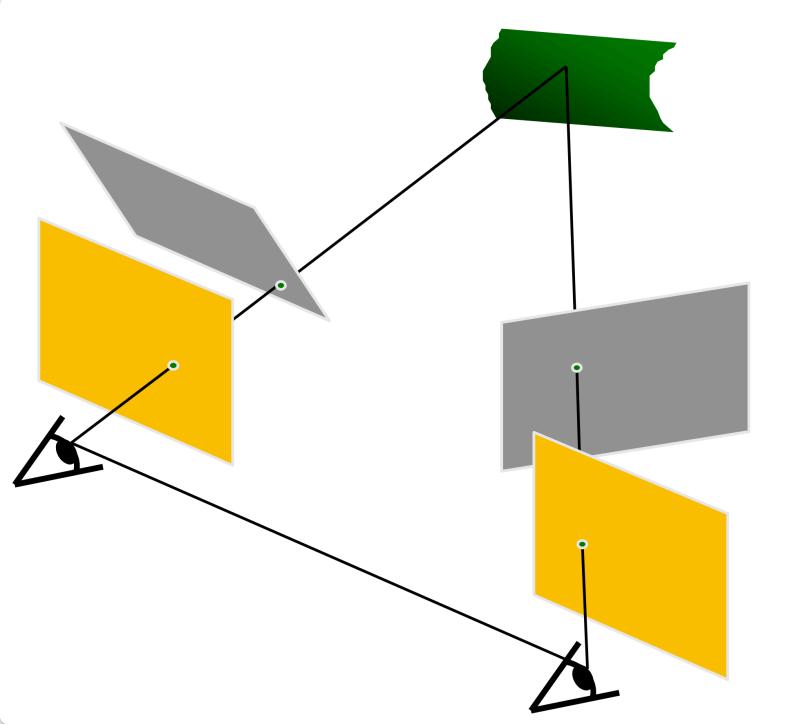
## Image Rectification

Ziel: Erreichen einer vereinfachten Stereo-Geometrie (ähnlich dem Normalfall) durch Bildrektifikation.

Image Re-projection:

- Die Bildebene werden auf eine gemeinsame Ebene re-projiziert.
- Diese gemeinsame Ebene ist parallel zur Linie zwischen den optischen Zentren (Basislinie).
- Wichtig: Es ist zu beachten, dass hauptsächlich der Brennpunkt der Kamera relevant ist.





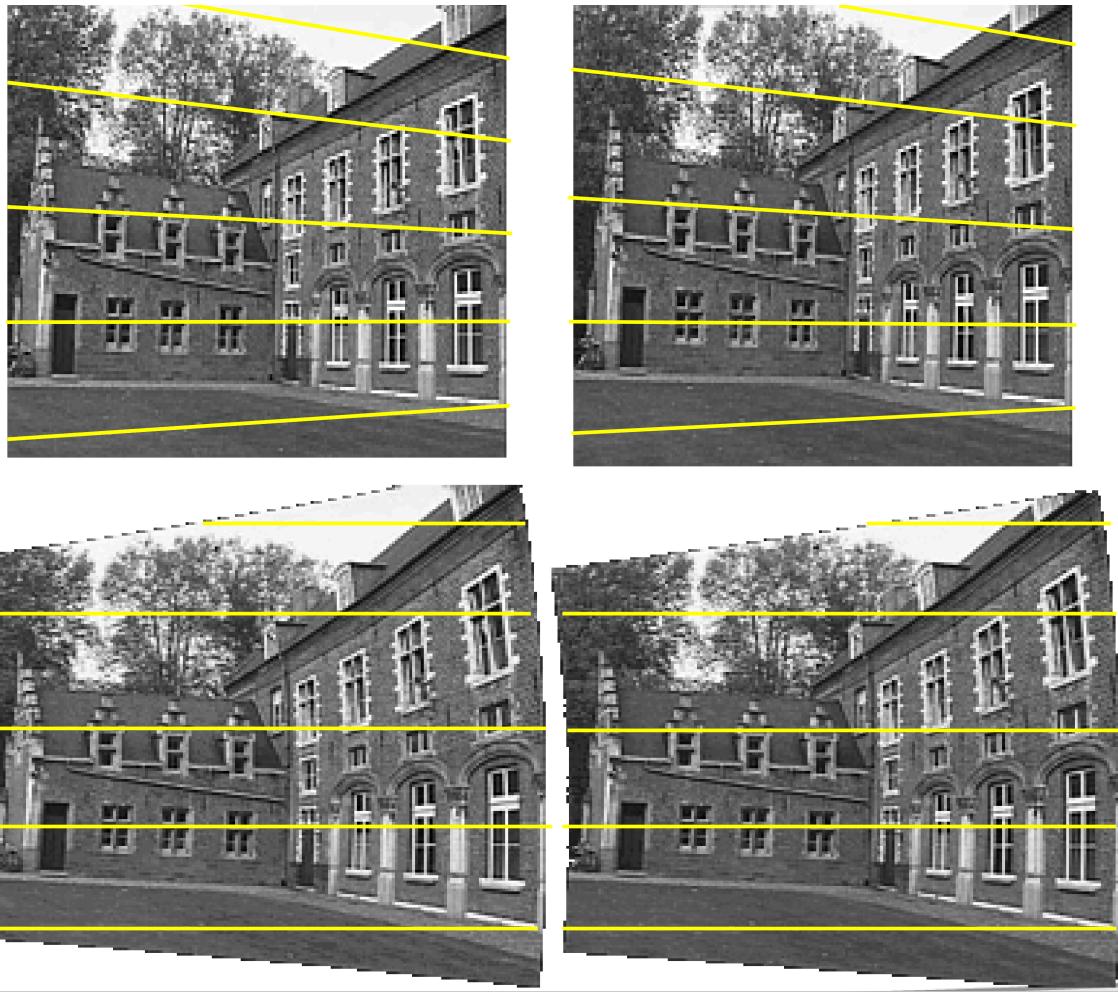
Prozess der Bildrektifikation (illustriert in der Abbildung):

1. **Raw Images:** Die ursprünglichen, möglicherweise verzerrten Bilder der linken und rechten Kamera.
2. **Undistortion:** Korrektur der Linsenverzerrung in den Rohbildern.
3. **Rectify:** Transformation der undistorrierten Bilder, sodass korrespondierende Punkte in den beiden Bildern auf der gleichen horizontalen Zeile liegen. Dies entspricht der Geometrie mit parallelen Bildebenen.
4. **Crop (optional):** Beschneidung der rektifizierten Bilder, um Bereiche ohne gültige Informationen zu entfernen.

Vorteil der Bildrektifikation:

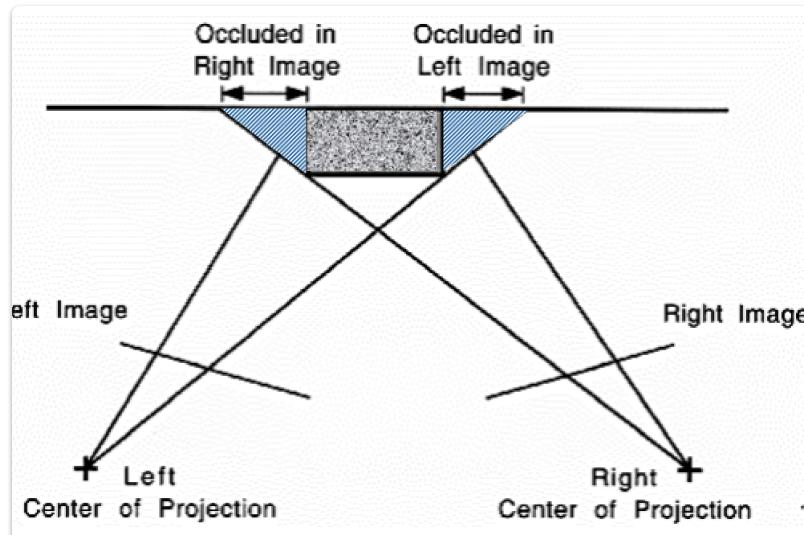
- Vereinfacht die Suche nach korrespondierenden Punkten erheblich, da diese nun auf horizontalen Epipolarlinien liegen (die zu horizontalen Zeilen im Bild werden).
- Ermöglicht die direkte Anwendung von Algorithmen, die für den Normalfall der Stereo-Geometrie entwickelt wurden.

**Beispiel:**



## Okklusionen

- **Ansichtsabhängig (View dependent):** Okklusionen treten auf, weil verschiedene Kameras unterschiedliche Perspektiven auf die Szene haben.
- **Verdeckte Punkte können nicht berechnet werden:** Bereiche, die in einem Bild verdeckt sind, liefern keine direkten Korrespondenzen im anderen Bild und somit keine Tiefeninformationen durch Stereo-Matching.



Erläuterung anhand der Abbildung:

- Ein Objekt (grauer Kasten) verdeckt einen bestimmten Bereich der Szene für die rechte Kamera (als "Occluded in Right Image" markiert).
- Gleichzeitig verdeckt das Objekt einen anderen Bereich der Szene für die linke Kamera (als "Occluded in Left Image" markiert).
- Diese verdeckten Bereiche können in den jeweiligen Bildern nicht mit korrespondierenden Punkten im anderen Bild abgeglichen werden.

Folge von Okklusionen:

- Führt zu fehlenden Tiefeninformationen in den Bereichen, die in mindestens einer der Kameras verdeckt sind.
- Die Größe und Position der okkludierten Bereiche hängen von der relativen Position und Orientierung der Kameras sowie der Geometrie der Szene ab.

## Testähnliches Beispiel

- Eine Szene wird mit einem Stereo-Setup aufgenommen. Der Abstand der beiden Kameras mit einer fokalen Länge von **450 Pixeln** beträgt **8cm**. Für einen Bildpunkt wird eine Disparität von **10 Pixeln** festgestellt. Wie weit ist der zugehörige Szenenpunkt entfernt?

$$\begin{aligned} \bullet & f = 450 \\ \bullet & B = 8\text{cm} \quad Z = \frac{f * B}{x_1 - x_2} \quad Z = \frac{450 * 8\text{cm}}{10} = 360\text{cm} \\ \bullet & D = 10 \end{aligned}$$

- Welche Disparität hat ein Szenenpunkt, der doppelt so weit entfernt ist?

$$D = \frac{f * B}{Z} \quad D = \frac{450 * 8\text{cm}}{720\text{cm}} = \frac{45}{9} = 5$$

## Korrespondenzproblem

[EVC\\_Skriptum\\_CV, p.53](#)

Definition:

- Das Korrespondenzproblem stellt das zentrale Problem der Stereo Vision dar.
- Es bezeichnet die Aufgabe, für jeden Bildpunkt im linken Bild jenen Punkt im rechten Bild zu finden, der denselben Objektpunkt abbildet.
- Entsprechende Suchverfahren werden als *Korrespondenzanalyse* oder auch als *Stereo Matching* bezeichnet.

Vereinfachung durch Epipolarkorrektur (Bildrektifikation):

- Aufgabe der Epipolarkorrektur ist es, Stereobildpaare anhand der Epipolargeometrie so zu transformieren, dass zusammengehörige Bildpunkte auf derselben horizontalen Linie liegen.

- Statt in zwei Dimensionen muss ein korrespondierender Punkt hier nur mehr entlang einer einzigen Scanline gesucht werden.
- Unter dieser Voraussetzung wird das Korrespondenzproblem wesentlich vereinfacht und die Korrespondenzanalyse somit beschleunigt.
- Gängige Stereo Matching Verfahren gehen in der Regel davon aus, dass die Bildpaare in *rektifizierter* Form vorliegen.

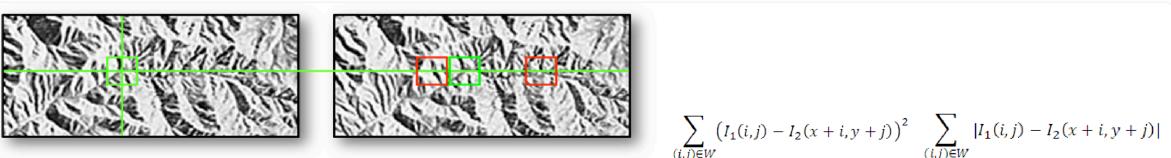
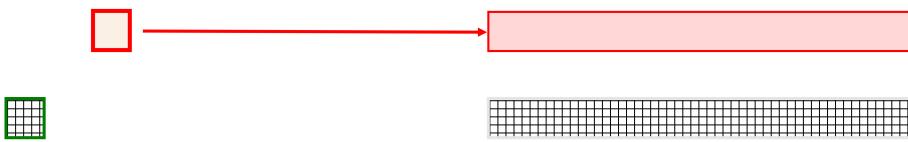
## Regionenbasiertes Matching (Area-Based Matching - ABM)

EVC\_Skriptum\_CV, p.53

Grundprinzip:

- Vergleicht kleine Ausschnitte (Beobachtungsfenster) zwischen den *rektifizierten* Bildern.
- Für jede Position eines Beobachtungsfensters im linken Bild wird das entsprechende Beobachtungsfenster im rechten Bild entlang der Epipolarlinie (jetzt eine horizontale Scanline) bewegt.
- Für jede Position wird geprüft, wie gut die Grauwerte mit den zu vergleichenden Grauwerten im linken Bild übereinstimmen.
- Dies geschieht durch eine Ähnlichkeitsmessung.

- The observation window in the left image is fixed
- For each position in the right image, the correlation function is calculated
- The window will be "slid" from left to right across the image
- Is performed for each position in the left image



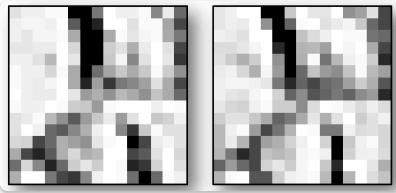
### Einfluss der Fenstergröße:

- **Zu kleine Fenster:** Beinhaltet zu wenig Information für eine korrekte Korrespondenzzuordnung, was zu erhöhten Fehlzuordnungen führen kann.
- **Zu große Fenster:** Erhöht die Rechenzeit.

### Gängige Ähnlichkeitsmaße:

- Summe der absoluten Differenzen (Sum of Absolute Differences - SAD)

- Summe der quadrierten Differenzen (Sum of Squared Differences - SSD)
- Normalisierte Kreuzkorrelation (Normalized Cross Correlation - NCC)



Formeln für Ähnlichkeitsmaße:

- **SSD (Sum of Squared Differences):**

$$SSD(\Delta m, \Delta n) = \sum_{i,j \in R} [I_1(i, j) - I_2(i - \Delta m, j - \Delta n)]^2$$

- $I_1(i, j)$ : Pixelintensität am Koordinaten  $(i, j)$  im ersten Fenster.
- $I_2(i - \Delta m, j - \Delta n)$ : Pixelintensität am verschobenen Koordinaten im zweiten Fenster.
- $R$ : Bereich des Fensters.
- $\Delta m, \Delta n$ : Verschiebung des zweiten Fensters relativ zum ersten.

- **CC (Cross-Correlation):**

$$CC(\Delta m, \Delta n) = \sum_{i,j \in R} [I_1(i, j) \cdot I_2(i - \Delta m, j - \Delta n)]$$

- Hier wird die Korrelation (Ähnlichkeit im Muster) direkt berechnet. Höhere Werte deuten auf größere Ähnlichkeit hin.

Anmerkung: Die gezeigte Formel für CC ist die unnormalisierte Kreuzkorrelation. Oft wird eine normalisierte Version (NCC) verwendet, um die Ergebnisse robuster gegenüber Helligkeits- und Kontrastunterschieden zu machen.

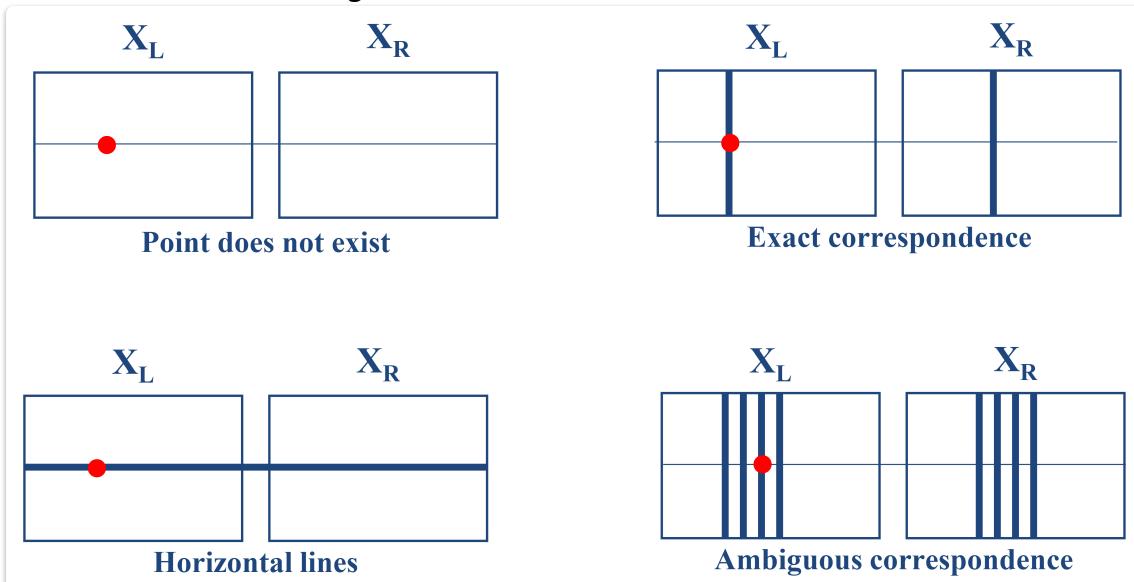
### Funktionsweise der Ähnlichkeitsmessung:

- Prinzipiell wird bei allen diesen Verfahren die Ähnlichkeit durch einen Vergleich von Pixeln innerhalb einer quadratischen Nachbarschaft zwischen dem linken und dem rechten Bild berechnet.
- Wenn das linke und rechte Bild exakt aufeinander passen, erhält man als Resultat ein Maximum (bei NCC) oder Minimum (bei SAD und SSD) in der Ähnlichkeitsfunktion.
- Mit Hilfe der Position des Maximums oder Minimums der Ähnlichkeitsfunktion wird die Position des korrespondierenden Punktes bestimmt (Disparität).

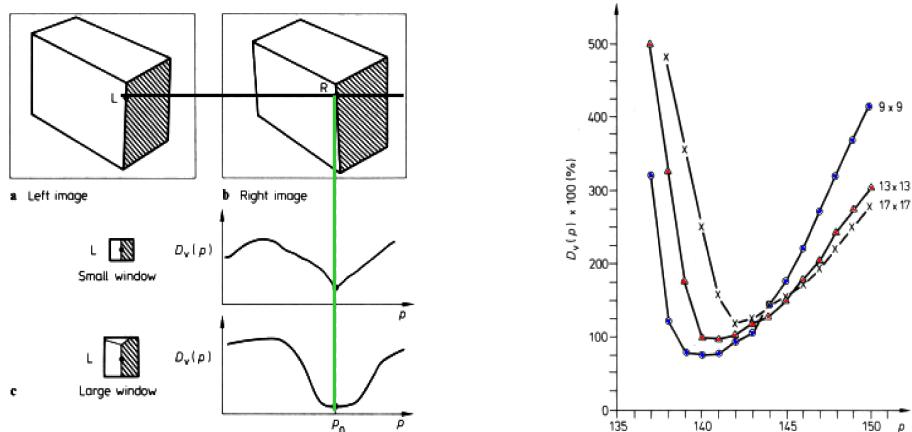
### Probleme:

- Es ist möglich, dass kein eindeutiges Minimum oder Maximum gefunden werden kann.
- Dies kann zum Beispiel durch **Verdeckungen** passieren, wenn ein Punkt von einer Kamera aus sichtbar ist und von der anderen nicht.

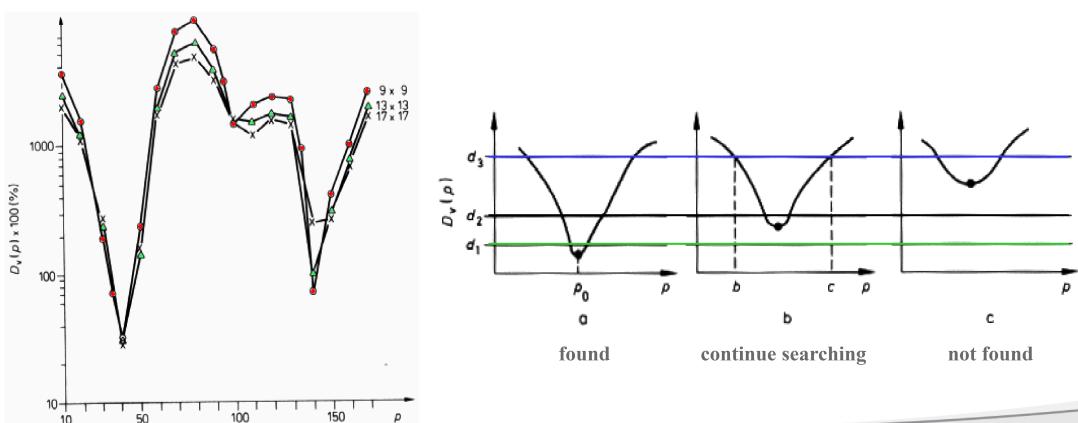
- Intensitäten in den beiden Bildern müssen nicht zwingend passen, der korrespondierende Punkt ist nicht notwendig.



- Correspondence is strongly dependent on window size used
  - Different algorithms like adaptive matching, pyramids ....



- Solution of ambiguity with threshold or additional constraints
- Different threshold values



**Vorteil:**

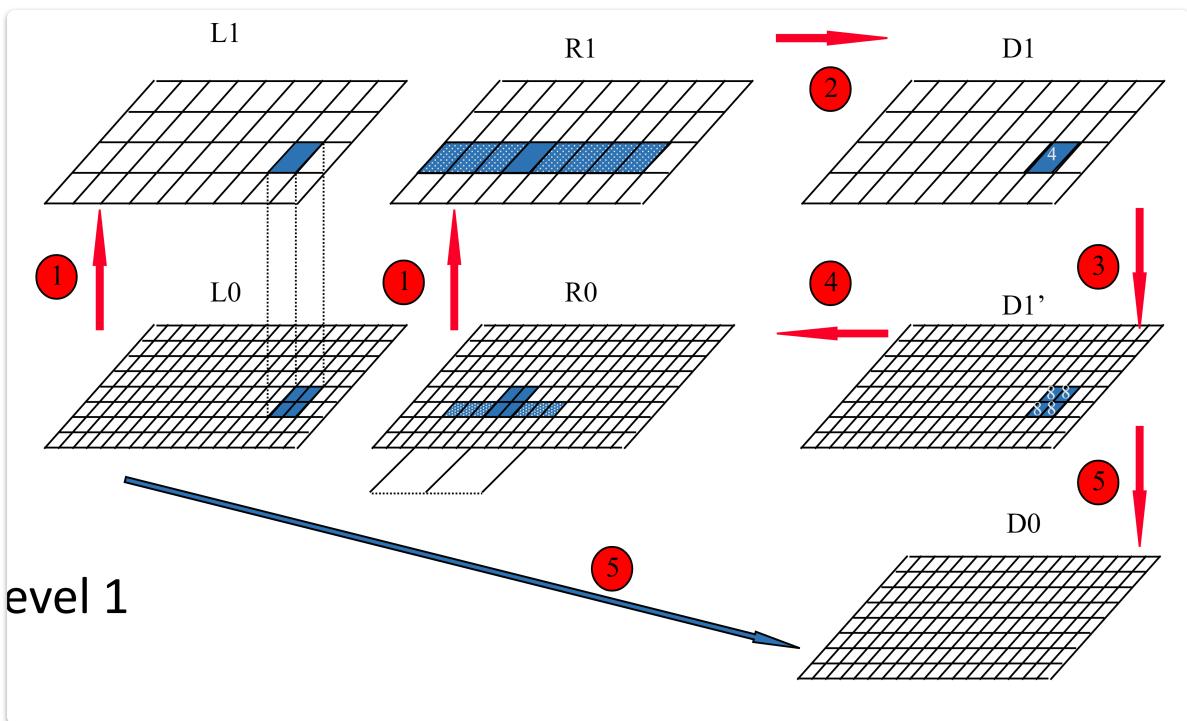
- Der Hauptvorteil des regionenbasierten Matching gegenüber den merkmalbasierten Verfahren ist ein dichteres Tiefenbild.
- Hier werden die Tiefenwerte für alle Pixel direkt ausgerechnet, nicht nur für einige ausgewählte Merkmalspunkte.

### Nachteil:

- Eine höhere Komplexität und ein entsprechend höherer Rechenaufwand.

### Hierarchical Matching

Ansatz: Verwendet Bildpyramiden (z.B. Gaussian Pyramids), um das Korrespondenzproblem effizienter zu lösen.

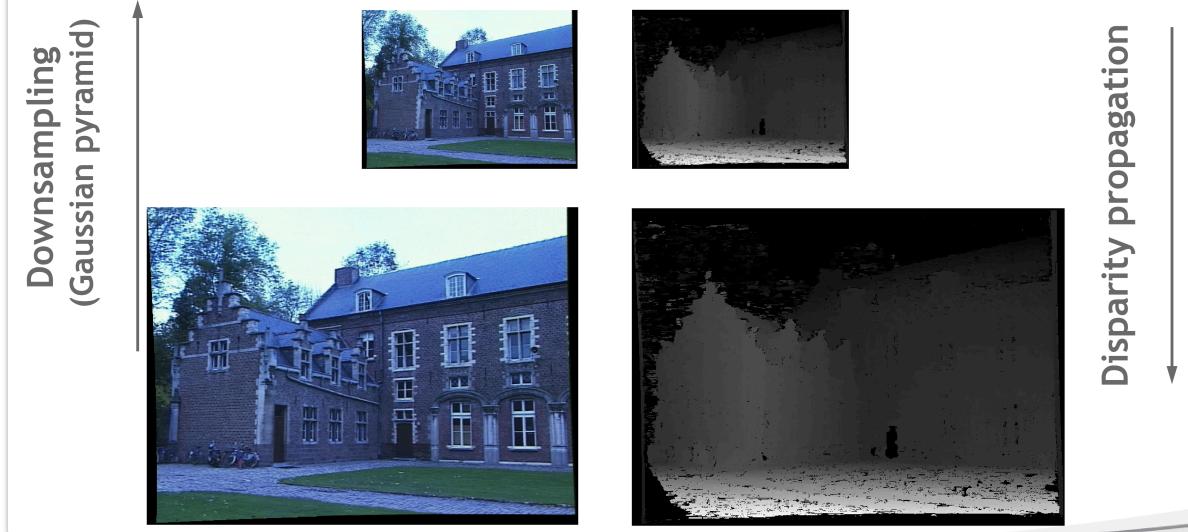


Schritte des hierarchischen Matchings (basierend auf der oberen Abbildung):

- Gaussian Pyramids:** Erstellung von Bildpyramiden für das linke und rechte Bild. Höhere Ebenen sind dabei niedrigere Auflösungsversionen der Originalbilder.
- Compute disparities of level 1:** Berechnung der Disparitäten auf der obersten (niedrigsten Auflösungs-) Ebene der Pyramide. Dies ist recheneffizienter und kann größere Disparitätsbereiche abdecken.
- Project values:** Die auf der höheren Ebene berechneten Disparitäten werden auf die nächstniedrigere Ebene projiziert und dienen dort als initiale Schätzwerte für die Disparitätssuche.
- Compute disparities of lower level:** Die Disparitäten werden auf der niedrigeren Ebene (mit höherer Auflösung) verfeinert, indem um die projizierten Schätzwerte herum gesucht wird. Dieser Prozess wird für alle Ebenen der Pyramide wiederholt, bis zur Originalauflösung.

tion

ity ranges

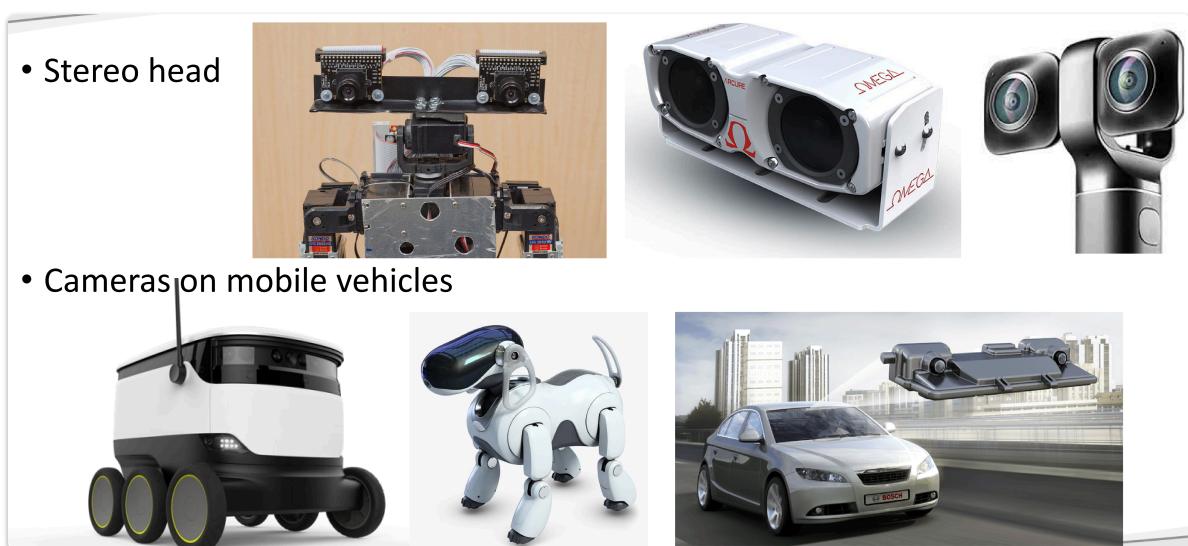


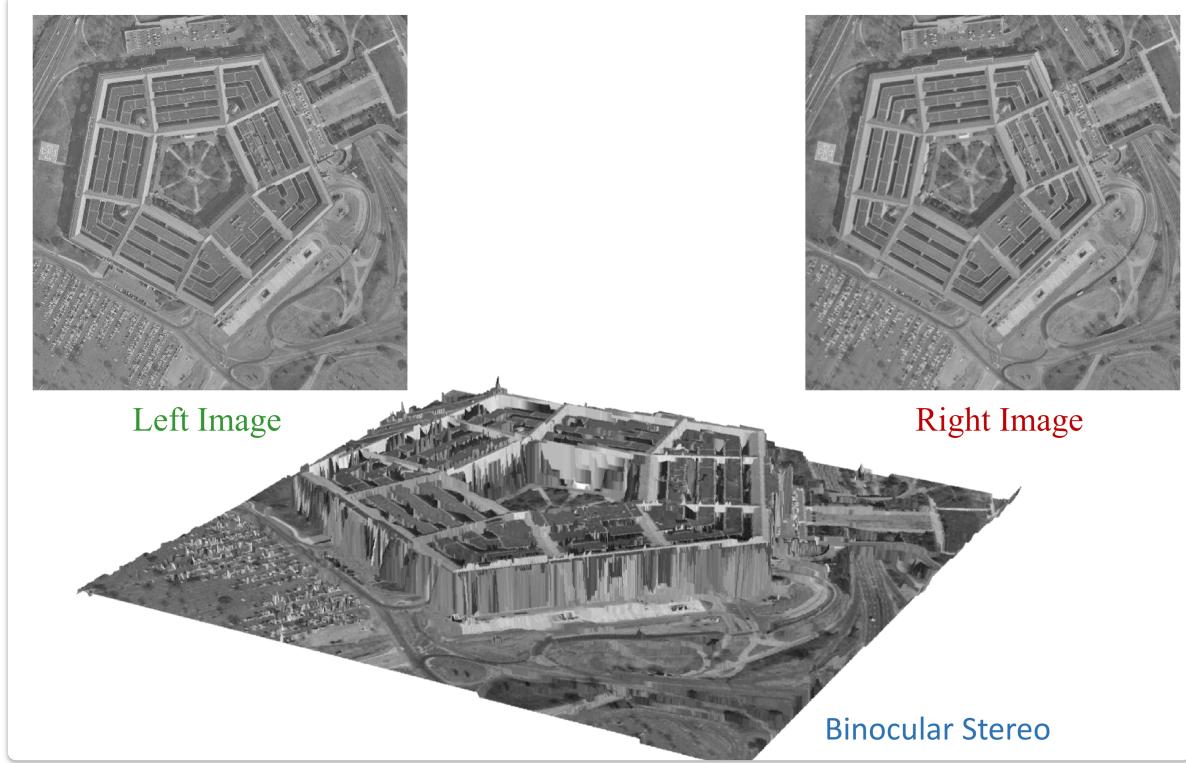
Vorteile des hierarchischen Stereo Matchings (basierend auf der unteren Abbildung):

- **Allows faster computation:** Die Suche nach Korrespondenzen beginnt auf einer niedrigen Auflösung, was die Anzahl der zu vergleichenden Pixel reduziert und somit die Berechnungszeit beschleunigt.
- **Deals with large disparity ranges:** Die grobe Suche auf niedriger Auflösung kann große Disparitätsunterschiede erfassen. Die anschließende Verfeinerung auf höheren Auflösungen ermöglicht eine präzisere Disparitätsschätzung.

Zusammenfassend: Hierarchisches Matching nutzt den Pyramidenansatz, um die Effizienz und den Suchbereich des Stereo Matchings zu verbessern. Durch die schrittweise Verfeinerung der Disparitäten von groben zu feinen Auflösungen können sowohl Rechenzeit reduziert als auch größere Disparitätsbereiche akkurat behandelt werden.

## Beispiele





## Merkmalbasiertes Matching

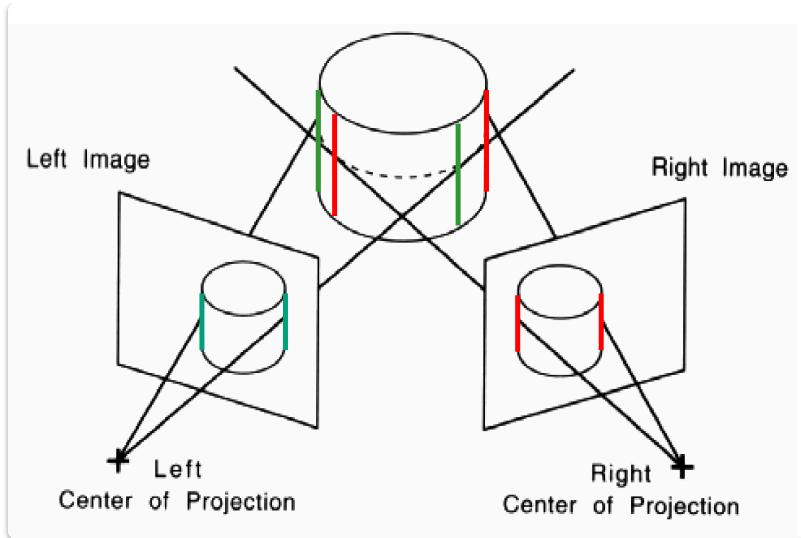
[EVC\\_Skriptum\\_CV, p.53](#)

Problem des ABM (Regionenbasiertes Matching):

- Homogene Bildbereiche enthalten sehr wenig Information.
- Werden aber in die Berechnung miteinbezogen und können zu Fehlern führen.

Ansatz des merkmalbasierten Matching:

- Verwendet einzelne, ausgewählte Pixel (Merkmale), die sich gut zuordnen lassen.
- Merkmale werden aus jedem Bild individuell extrahiert, bevor sie verglichen werden.
- Lokale Merkmale können Ecken, Kanten oder andere *Interest Points* sein.
- Diese Interest Points werden durch lokale Operatoren wie z.B. Moravec oder SIFT extrahiert.



Vorteil des merkmalbasierten Matching:

- Der eigentliche Korrespondenzvergleich kann schneller durchgeführt werden, da bei der Merkmalsextraktion eine wesentliche Datenreduktion stattfindet.

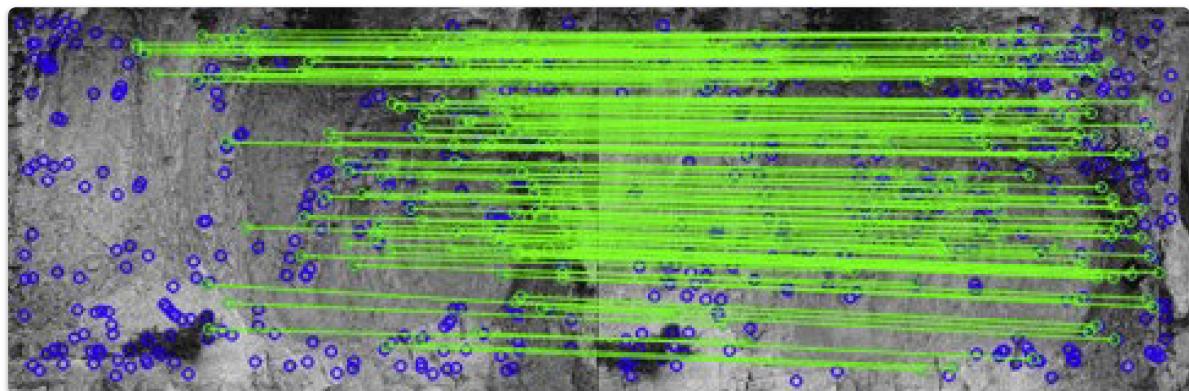
Nachteil des merkmalbasierten Matching:

- Der Hauptnachteil dieser Verfahren liegt aber darin, dass man nur zuverlässige Tiefeninformationen für diese ausgewählten Merkmale erhält (kein dichtes Tiefenbild wie bei ABM).
- Die Bereiche zwischen den Interest Points bleiben zunächst unberücksichtigt und müssen ggf. weiteren Verarbeitungen unterzogen werden (z.B. Interpolation).

### Matching Criteria (Merkmalbasiertes Matching)

Verwendete Merkmale:

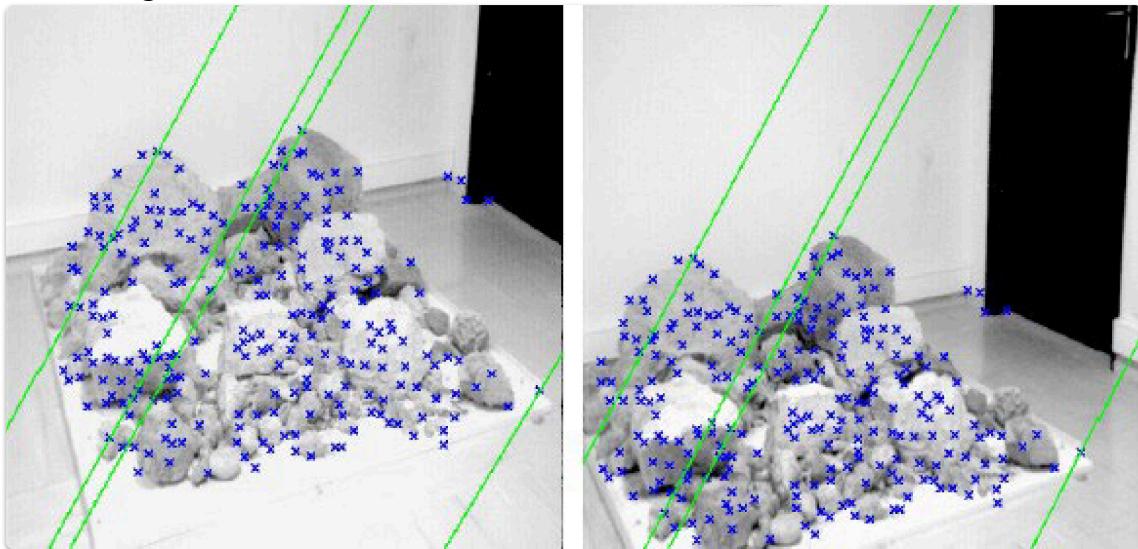
- "Corner"-artige Features (z.B. aus [Zhang, ...])
- Kanten (Edges) [viele Personen...]
- Gradienten [Seitz 89; Scharstein 94]
- Interest Points (z.B. SIFT)



### Feature-based Stereo

Prozess:

- **Match "corner" (interest) points:** Zuerst werden markante Punkte in beiden Bildern detektiert und einander zugeordnet (gematcht).
- **Interpolate complete solution:** Da nur für die Merkmale Tiefeninformationen vorhanden sind, muss die Tiefenkarte für die übrigen Bildbereiche interpoliert werden, um eine vollständige Tiefenkarte zu erhalten.



## Structure-from-Motion (SfM)

[EVC\\_Skriptum\\_CV, p.54](#)

Definition:

- Bezeichnet den Prozess der Gewinnung dreidimensionaler Informationen von Objekten oder einer ganzen Szene durch die Auswertung einer zeitlichen Folge von Bildern.

Zentrales Problem:

- Bestimmung der **Richtung** und des **Ausmaßes** der Kamerabewegung.
- Ansätze zur Lösung:
  - **Motion Field estimation:** Schätzung des Bewegungsfeldes der Punkte im Bild über die Zeit.
  - **Motion Field analysis:** Analyse des geschätzten Bewegungsfeldes zur Bestimmung der Kamerabewegung und der Szenenstruktur.
- Verschiebung des Blickpunkts führt zur scheinbaren Bewegung von Objekten in der Szene.

Unterschied zur Stereo Vision:

- Im Gegensatz zur Stereo Vision ist die Kamerageometrie (die relative Position und Orientierung der Kameras zueinander) zunächst **nicht bekannt**.

Vorgehensweise:

- Anhand der gefundenen Korrespondenzen werden Bewegungsfelder geschätzt.
- Diese Bewegungsfelder können dazu verwendet werden, die Kamerabewegungen zu bestimmen.
- Aus den Kamerabewegungen und den korrespondierenden Punkten kann dann die 3D-Struktur der Szene rekonstruiert werden.

Informationen, die aus der Bewegung gewonnen werden können:

- Information über die Bewegung des Betrachters (der Kamera).
- Tiefeninformationen der Umgebung (vgl. Stereo Vision).
- *Motion Parallax* = *Motion Disparity*: Nahe Objekte scheinen sich bei Bewegung des Betrachters schneller relativ zu entfernten Objekten zu bewegen.



## Bewegungsfeld

[EVC\\_Skriptum\\_CV, p.54](#)

Beobachtung:

- Fixiert ein Beobachter den Horizont, so scheinen sich Mond, Sterne und die gesamte obere Gesichtssphäre zu bewegen.
- Welt und Erdboden scheinen in einem kontinuierlichen Strom vorbeizuziehen.

Relativgeschwindigkeit und Entfernung (Helmholtz, 1950):

- Projiziert man die Umgebung auf eine Abbildungsebene vor dem Betrachter, so ist die relative Geschwindigkeit eines Objekts *umgekehrt proportional* zu seiner Entfernung vom Betrachter.
- Je weiter ein Objekt vom Betrachter entfernt ist, desto geringer ist dessen Bewegung.
- Sterne und der Mond bewegen sich nicht, die Straße direkt neben dem Beobachter bewegt sich sehr schnell.

Richtung des Bewegungsvektors:

- Die Richtung des Bewegungsvektors eines Ortspunktes ist abhängig von der Lage dieses Ortspunktes zum Betrachter.

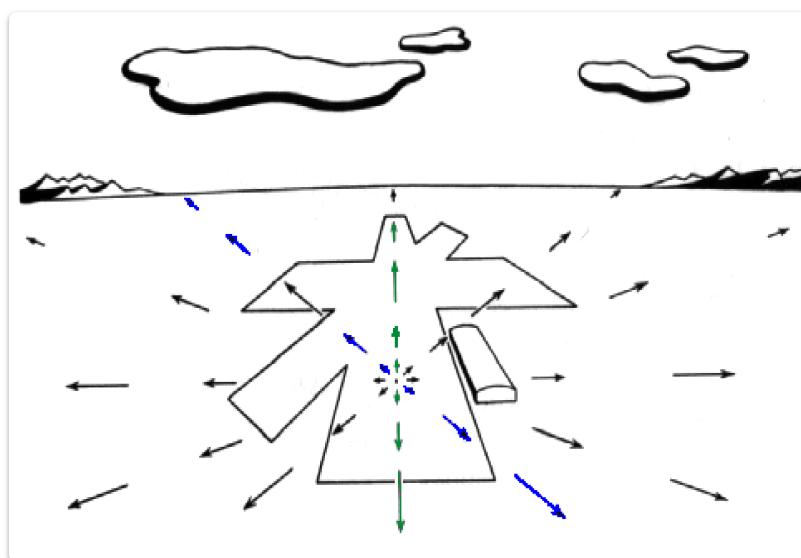
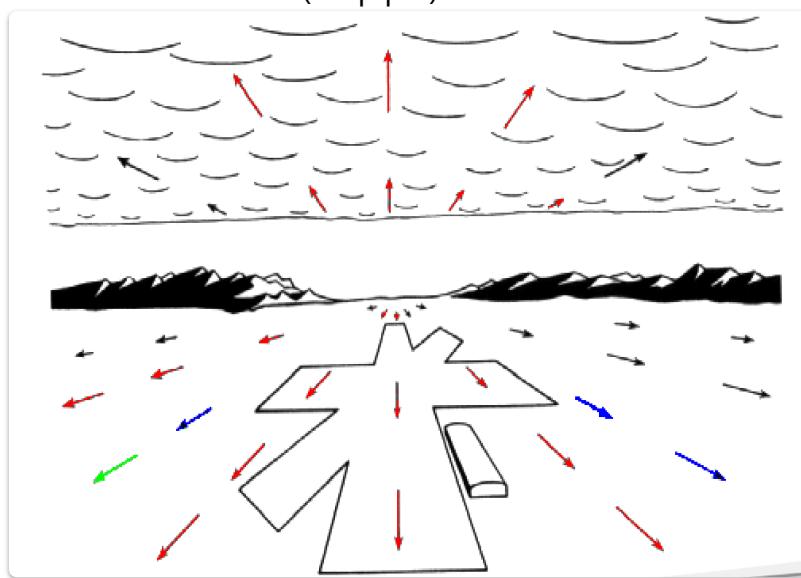
- In einer Umgebung, in der sich die Objekte zueinander nicht bewegen, kann die Eigenbewegung berechnet und eine relative Tiefenkarte der Umgebung erstellt werden.

Bestandteile des Bewegungsfeldes:

- Das Bewegungsfeld besteht aus den einzelnen Bewegungsvektoren aller Punkte im Bild.

Kamera ohne Rotation (Translation):

- Bewegt sich die Kamera ohne zu rotieren, bewirkt dies ein "nach außen" oder "nach innen" Zeigen aller Vektoren zu einem einzigen Punkt.
- Dieser Punkt wird *Focus of Expansion (FoE)* (bei Vorwärtsbewegung) oder *Focus of Contraction (FoC)* (bei Rückwärtsbewegung) genannt.
- Dieser Punkt befindet sich dort, wo sich die Verschiebungsvektor der Kamera mit der Bildebene schneidet (= Epipol).



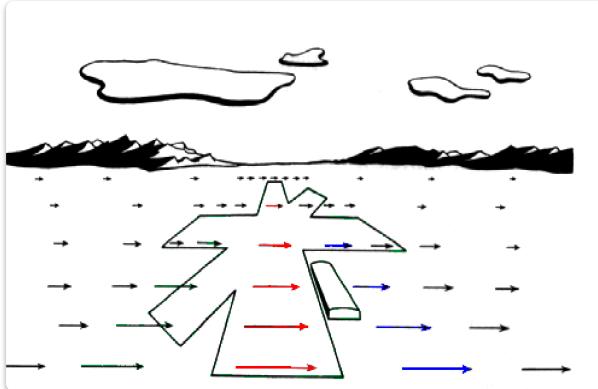
Größe der Bildbewegung eines Szenenpunktes (bei Kamerabewegung):

- Hängt *umgekehrt proportional* von der Entfernung eines Punktes zur Kamera ab.

- Hängt *direkt proportional* vom Sinus des Winkels zwischen der Richtung, in der dieser Szenenpunkt liegt, und der Richtung, in welcher die Kamera verschoben wird, ab.

Berechnung von Kamerabewegung:

- Damit können somit die Richtung der Kamerabewegung (FoE bzw. FoC) und der Betrag der Bewegung berechnet werden.
- Dies führt dann über die Epipolargeometrie zur Stereorekonstruktion (obwohl hier nur eine einzelne bewegte Kamera verwendet wird, nicht zwei gleichzeitig).



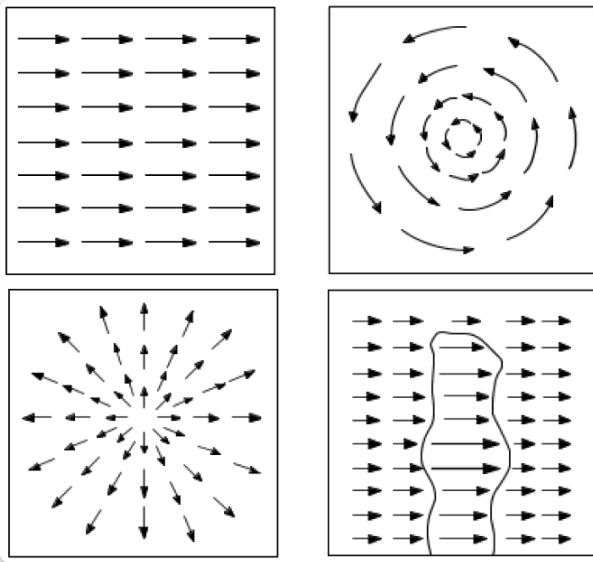
## Bewegungsfeld: Schätzung

Herausforderungen:

- **Sparsely occupied vector field:** Das resultierende Bewegungsfeld kann spärlich sein, d.h., nicht für jeden Pixel ist eine Bewegungsinformation vorhanden (insbesondere in homogenen Bereichen).
- **Same problem as stereo - just the moving direction of the camera is not known:** Das grundlegende Problem der Korrespondenzfindung ist ähnlich wie beim Stereo Matching, jedoch ist die relative Bewegung (die "Basislinie" und Ausrichtung) der Kamera zwischen den Zeitpunkten unbekannt.
- **Epipolar line not known in the beginning:** Da die Kamerabewegung unbekannt ist, sind auch die Epipolarlinien zunächst nicht bekannt, was die Suche nach korrespondierenden Punkten erschwert.

Ansätze zur Korrespondenzfindung zwischen Bildern in einer Sequenz:

- **High temporal sampling = low differences:** Bei einer hohen Bildrate (geringer Zeitabstand zwischen den Bildern) sind die Unterschiede zwischen aufeinanderfolgenden Bildern geringer, was die Korrespondenzsuche erleichtern kann.
- **Either unchanged intensities in both images or unchanged edges in both images:** Annahme, dass entweder die Intensitäten der korrespondierenden Punkte oder deren lokale Struktur (z.B. Kanten) sich zwischen den aufeinanderfolgenden Bildern nicht wesentlich ändern. Diese Annahme kann zur Einschränkung der Suche verwendet werden.



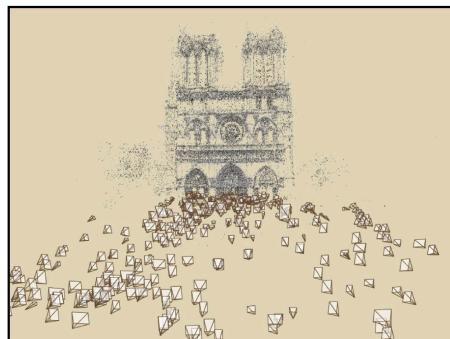
## Multi View Geometry

[EVC\\_Skriptum\\_CV, p.54](#)

### Definition:

- Eine weitere Möglichkeit zur 3D-Rekonstruktion, bei der mehr als zwei beliebige Bilder verwendet werden, um die Relation der Bildpunkte und der Punkte im 3D-Raum zu errechnen.

Structure from Motion (SfM)

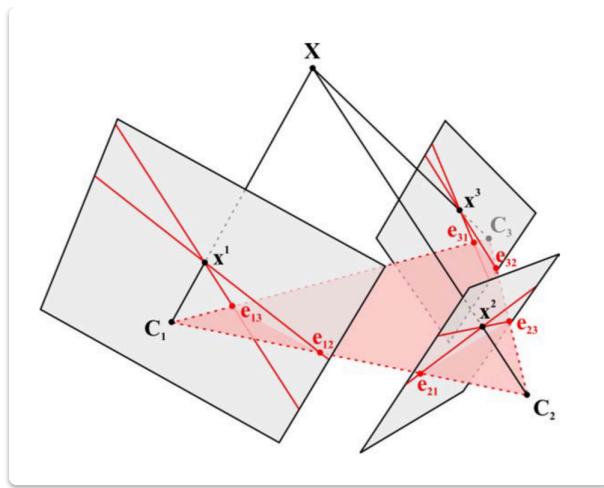


Dense multiview stereo



### Vorteil gegenüber Stereo mit nur zwei Bildern:

- Betrachtet man den Fall mit 3 Bildern: Wenn ein Objektpunkt in zwei Bildern identifiziert wurde, so kann seine geometrische Position im dritten Bild durch den Schnitt der entsprechenden Epipolar-Geraden vorhergesagt werden (siehe Abbildung 45).
- Im Unterschied zum Bildpaar existiert beim Bildtripel jedoch ein eindeutiges Ergebnis.
- Von den 2 oder 3 Basisbildern ausgehend, können bei der Mehrbildgeometrie dann stufenweise mehr Bilder der Szene aus mehreren von verschiedenen Orten aus aufgenommenen Bildern hinzugenommen werden.
- Durch Hinzunahme von mehr Bildern entsteht eine stabilere 3D-Punktwolke.
- Die Durch Bündelausgleich verbessert werden kann.



## Bündelausgleich (Bundle Adjustment):

- Stammt aus der Photogrammetrie.
- Erlaubt die gleichzeitige Bestimmung der internen und externen Kameraparameter sowie der 3D-Struktur der Szene aus mehreren verschiedenen Ansichten.
- Grundlegende Annahmen für den Bündelausgleich sind die Starrheit der Szene zwischen den einzelnen Ansichten und die Erfüllung der Kollinearitätsgleichung (Collinearity Condition).
- Kollinearitätsgleichung: Besagt, dass ein betrachteter 3D-Objektpunkt, sein zugehöriger Bildpunkt und das Projektionszentrum der Kamera auf einer gemeinsamen Gerade liegen müssen.

## Zielsetzung des Bündelausgleichs:

- Gleichzeitige Variation der 3D-Koordinaten der einzelnen Ansichten und der Transformationsparameter der einzelnen Ansichten.
- Ziel ist, eine möglichst gute Übereinstimmung zwischen erwarteten und gemessenen Bildpunkten zu erreichen.

## Zentralprojektion:

- Im Falle einer Zentralprojektion erzeugt jeder Szenenpunkt einen Strahl durch das Projektionszentrum.
- Für die Menge aller Punkte entsteht ein Strahlenbündel, welches im Projektionszentrum geschnürt wird.

## Structure from Motion

### Grundlagen:

- Gegeben viele korrespondierende Punkte über mehrere Bilder hinweg,  $\{(u_{ij}, v_{ij})\}$ .
- Ziel ist die simultane Berechnung der 3D-Positionen der Punkte  $x_i$  und der Kamera- (oder Bewegungs-) Parameter  $(K, R_j, t_j)$ .
  - $K$ : Intrinsische Kameraparameter (Kalibrierungsmatrix).

- $R_j$ : Rotationsmatrix der Kamera im  $j$ -ten Bild.
- $\mathbf{t}_j$ : Translationsvektor der Kamera im  $j$ -ten Bild.
- Die Bildkoordinaten  $(u_{ij}, v_{ij})$  sind eine Funktion der Kameraparameter und der 3D-Punktpositionen:

$$u_{ij} = f(K, R_j, \mathbf{t}_j, \mathbf{x}_i)$$

$$v_{ij} = g(K, R_j, \mathbf{t}_j, \mathbf{x}_i)$$

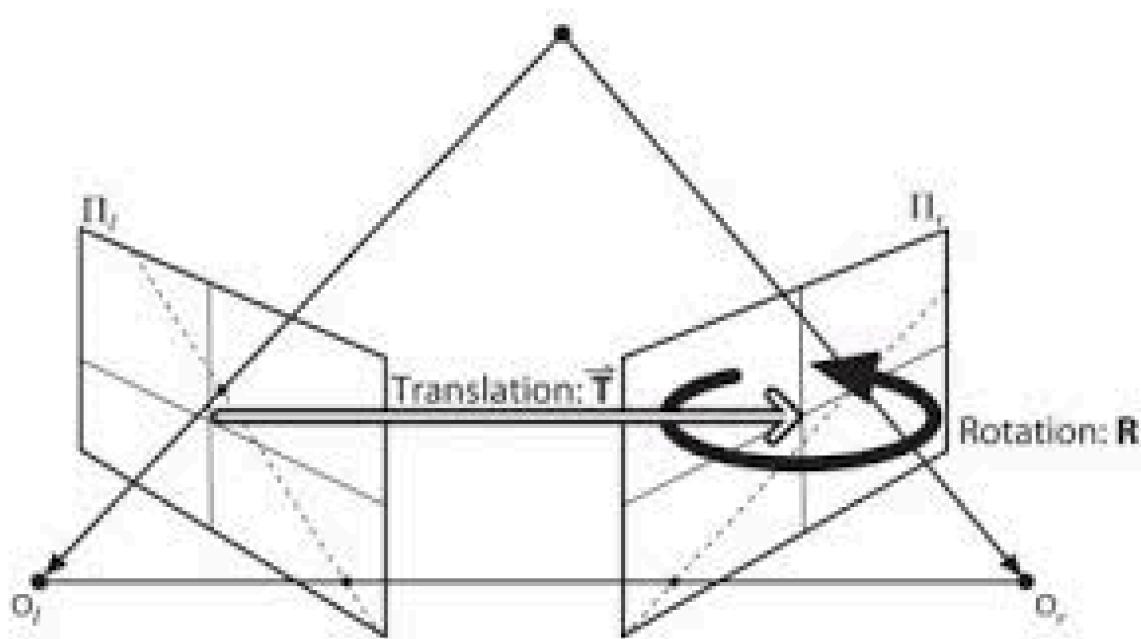
Hauptvarianten:

- **Kalibriert:** Die intrinsischen Kameraparameter  $K$  sind bekannt.
- **Unkalibriert:** Die intrinsischen Kameraparameter  $K$  sind unbekannt (manchmal assoziiert mit euklidischen und projektiven Rekonstruktionen).

## Automatische Berechnung von F (Fundamentale Matrix - typisch für unkalibrierte SfM)

Schritte:

1. **Interest points:** Detektion von markanten Punkten in den Bildern.
2. **Putative correspondences:** Erstellung initialer, potenzieller Korrespondenzen zwischen den Bildern (können fehlerhaft sein).
3. **RANSAC (RANdom SAmple Consensus):** Robustes Schätzverfahren zur Entfernung von Ausreißern (falschen Korrespondenzen) und zur Schätzung der Fundamentalen Matrix  $F$ .
4. **Non-linear re-estimation of F:** Nicht-lineare Optimierung zur Verfeinerung der Schätzung der Fundamentalen Matrix  $F$ .
5. **Guided matching:** Verwendung der geschätzten Epipolargeometrie (aus  $F$ ) zur Verbesserung der Korrespondenzsuche.
6. **Repeat (4.) and (5.) until stable:** Iterativer Prozess der Verfeinerung der Fundamentalen Matrix und der Korrespondenzen, bis eine stabile Lösung erreicht ist.



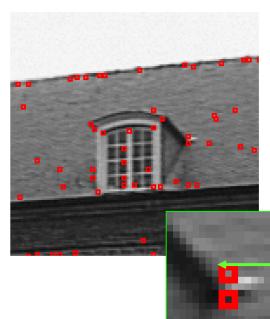
## Beispiel



Select strongest features (e.g. 1000/image)

Evaluate SSD and SAD for all features with similar coordinates

e.g.  $(x', y') \in [x - \frac{w}{10}, x + \frac{w}{10}] \times [y - \frac{h}{10}, y + \frac{h}{10}]$



Keep mutual best matches  
Still many wrong matches!



## RANSAC



Step 1. Extract features

Step 2. Compute a set of potential matches

Step 3. do

Step 3.1 select minimal sample (i.e. 7 matches)

Step 3.2 compute solution(s) for F

Step 3.3 determine inliers

(verify hypothesis)

}

(generate hypothesis)

until  $\Gamma(\#inliers, \#samples) < 95\%$

- Step 4. Compute F based on all inliers
- Step 5. Look for additional matches
- Step 6. Refine F based on all correct matches

$$\Gamma = 1 - \left(1 - \left(\frac{\#inliers}{\#matches}\right)^7\right)^{\#samples}$$

#inliers	90%	80%	70%	60%	50%
#samples	5	13	35	106	382