

Videoverarbeitung

Erläutern Sie die Grundzüge des in der Vorlesung vorgestellten Stereo Matching Algorithmus.

Ziel: Bestimmung der Verschiebungsvektoren zwischen dem rechten und dem linken Bild. Wichtig ist das Block Matching Verfahren. Schwierig: Einheitliche Flächen(texturlose Gebiete) und Verdeckungen

Prinzipien:

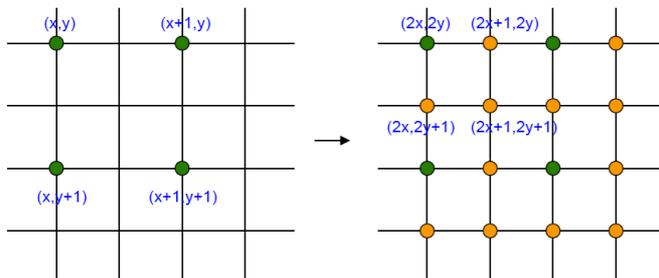
- Farbsegmentierung des Ausgangsbildes (ein Objekt hat meist mehrere Farbsegmente)
- Vergleich von Disparity-Werten(Unterschiede) eines Standardstereoverfahrens zwecks Clusterfindung um Ebenen zu bestimmen
- Zuordnung der Layer zu Farbsegmentierungen mittels Kostenfunktion -> Optimierung

Erläutern Sie den Unterschied zwischen Block-basierten und Netzbasierten (mesh based) Verfahren zur Bewegungserkennung.

Beim blockbasierten Verfahren haben alle Pixel eines Blocks den gleichen Verschiebungsvektor. Beim netzbasierten Verfahren hat jeder Netzknoten einen eigenen Verschiebungsvektor. Die Vektoren für die Pixel in einem Netz werden mittels den Netzknotenvektoren interpoliert. Dadurch erreicht man stetige Übergänge an den Rändern der Netzelemente. Artefakte an den Blockgrenzen werden eliminiert (bei blockbasierten Verfahren kann es zu solchen Artefakten kommen).

Wie kann man bei der Bestimmung der Bewegungsvektoren Subpixelgenauigkeit erreichen?

Stichwort: Bilineare Interpolation.



Bewirkt eine höhere Genauigkeit bei der Bestimmung der Bewegungsvektoren. Im Suchframe wird hier das Raster verfeinert. Hier werden Grauwerte zwischen den eigentlichen Pixelwerten interpoliert.

Was versteht man unter „Optical Flow“

$$\frac{\partial \psi}{\partial x} d_x + \frac{\partial \psi}{\partial y} d_y + \frac{\partial \psi}{\partial t} d_t = 0 \quad \text{or} \quad \frac{\partial \psi}{\partial x} v_x + \frac{\partial \psi}{\partial y} v_y + \frac{\partial \psi}{\partial t} = 0 \quad \text{or} \quad \nabla \psi^T \mathbf{v} + \frac{\partial \psi}{\partial t} = 0$$

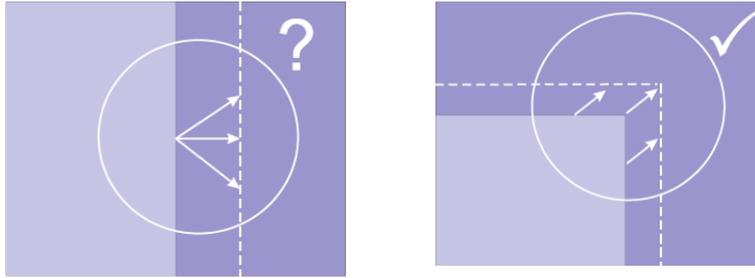
ψ ...Helligkeit bzw. Grauwerte der Bildpunkte. Annahme: Ein Bildpunkt hat in verschiedenen Frames den gleichen Grauwert hat.

Der Optical Flow ist eine wahrgenommene Bewegung, die in Wirklichkeit keine Bewegung ist, sondern nur als solche wahrgenommen wird.

Beschreiben Sie das „Apertur-Problem“

Der Bewegungsvektor ist nicht eindeutig bestimmt.

Dieses Problem tritt auf, wenn nur kleine Bildausschnitte betrachtet werden. Verlaufen gerade Kanten durch einen solchen Ausschnitt kann ein Bildpunkt nicht eindeutig wieder zugeordnet werden (Siehe Abbildung). Das führt dazu, dass die Bewegung eines Objektes nicht oder nur in eine Richtung eindeutig bestimmt werden kann.



Diskutieren Sie die Rolle der Beleuchtung bei der Aufnahme von Hochgeschwindigkeitsvideos.

Da sehr viele Bilder pro Sekunde gemacht werden, können die einzelnen Bilder nur sehr kurz belichtet werden. Daher muss für ausreichend Beleuchtung gesorgt sein. Die Intensität des Lichtes, die gleichmäßige Ausleuchtung und der richtige Kontrast sind entscheidend für die Qualität und Verwertbarkeit der Aufnahmen. Mehr Licht ermöglicht kürzere Belichtungszeiten und damit eine schärfere Darstellung der Bewegung. Je mehr Licht zur Verfügung steht, desto kleiner kann die Blende gewählt werden, was wiederum eine größere Schärfentiefe zur Folge hat. Der mögliche Schärfbereich (zeitlich und räumlich) ist also direkt von der Beleuchtungsintensität abhängig. Der Kontrast hängt von der Lichtführung ab, aber auch von der Wahl des Hintergrundes, vor dem sich eine Bewegung abspielt. Das Licht spielt also eine entscheidende Rolle, besonders wenn es darum geht, verwendbare Messdaten aus einer Hochgeschwindigkeitsaufnahme zu gewinnen.

Erläutern Sie Anwendungsbeispiele, für welche die dreidimensionale Rekonstruktion der Videoszene Voraussetzung ist.

- Automatic Video Editing: Das Einfügen, Löschen und Modifizieren von Videoobjekten (bzw. Einfügen von künstlichen Videoobjekten) erfordert Kenntnis der 3D Szenengeometrie (Verdeckungen, perspektivische Verzerrung, etc.).
- Interactive Stereo Video: Stereodarstellung auf geeigneten Displays erzeugt 3D Eindruck beim Betrachter („Immersive TV/Video Conferencing“).
- Free Viewpoint Video: Die Szene wird von mehreren Videokameras aufgenommen, dadurch können beliebige Ansichten der Szene generiert/interpoliert werden. Der Benutzer kann den Blickpunkt frei wählen, unabhängig von der tatsächlichen Aufnahmeposition der verwendeten Videokameras.

Erläutern Sie die charakteristischen Eigenschaften von Block-basierten Verfahren zur Bewegungsbestimmung.

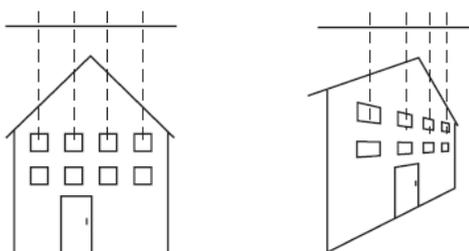
Generell bei Bewegungsbestimmung: die Bewegung hat mehr Freiheitsgrade als bei Stereoverschiebung. Man geht davon aus, dass alle Pixel in einem Block den gleichen Motion-Vector haben. Eine exakte Bestimmung des Motionvektors ist nicht immer möglich, der Fehler (=über Block aufsummierte Grauwertabweichung) des Motionvektors sollte möglichst gering sein.

EBMA = Exhaustive (gründlich) Block Matching Algorithm. Es wird eine Suchregion definiert, in dem gesucht wird. In dieser Suchregion werden alle Möglichkeiten untersucht. Es wird das beste Ergebnis genommen. Das Ergebnis kann verbessert werden, wenn man die Pixelpositionen nochmals unterteilt, ein genaueres Suchen ist dadurch möglich, siehe Bilinear Interpolation.

Eine weitere Optimierungsmöglichkeit ist der Hierarchical Block Matching Algorithm (HBMA). Hierbei wird ein und dasselbe Frame in verschiedenen Auflösungen in einer Bildpyramide gespeichert (kleinste Auflösung befindet sich oben). Die Suche in Frames mit geringer Auflösung erfolgt schnell, pro Ebene kann das Suchergebnis verfeinert werden.

Erklären Sie die Begriffe *chirping effect* und *converging effect* (oder „*Keystone effect*“) an Hand einer Skizze. In welchem Zusammenhang wurden die Begriffe in der Vorlesung besprochen?

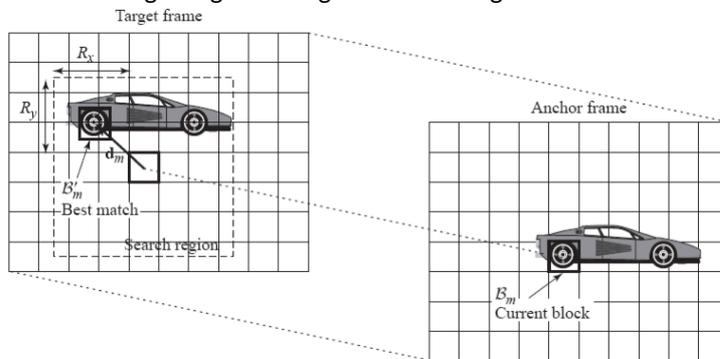
Chirping Effekt: Die beobachtete räumliche Frequenz ändert sich



Keystone Effekt: Parallel Linien treffen sich im Unendlichen

Beschreiben Sie ein Block-basiertes Verfahren zur Verfolgung von Bewegung in aufeinanderfolgenden Frames mit eigenen Worten und mittels Pseudo-Code

Siehe vorherige Frage und folgende Abbildung



Minimierungsfunktion:

$$E_{DFD}(\mathbf{d}_m) = \sum_{\mathbf{x} \in B_m} |\psi_2(\mathbf{x} + \mathbf{d}_m) - \psi_1(\mathbf{x})|^p \rightarrow \min$$

- DFD (Displaced Frame Difference: Unterschied zwischen 2 betrachteten Frames)
- B_m (betrachteter Block B [mit Index m]),
- \mathbf{d}_m (Bewegungsvektor, zugehörig zum Blocks B_m)
-

In diesem Modell wird angenommen, dass alle Pixel eines Blocks ein- und dieselbe Translation, beschrieben durch \mathbf{d}_m , ausführen).

ψ_1 und ψ_2 sind die Grauwertfunktionen der Frames vor und nach der Bewegung

EDFD ist die Fehlerfunktion (Error), welche die Grauwertabweichungen zwischen dem ursprünglichen Block im ersten Frame und jenem Block im 2. Frame, der vom ursprünglichen Frame die Positionsabweichung \mathbf{d}_m besitzt, angibt.

Die Suchaufgabe ist, für jeden Block B_m jenen Bewegungsvektor \mathbf{d}_m zu finden, welcher die Fehlerfunktion E (d.h. die Abweichung im Grauwertbild-Blockinhalt) minimiert. Setzt man den Parameter p auf 1, so wird die Fehlerfunktion durch einfaches Aufsummieren der Absolutwerte der Grauwertdifferenzen berechnet. Dies wird auf der Folie als MAD (Maximum Absolute Difference) Verfahren bezeichnet. Aufsummiert wird über alle Pixel \mathbf{x} , die zum betrachteten Block B_m gehören.

```

For i=1 bis Breite //alle Blöcke
  For j=1 bis Höhe //alle Blöcke
    Setze MADmin
    For k=1 bis Suchregion //Suchregion durchsuche
      For l=1 bis Suchregion //Suchregion durchsuchen
        MAD berechnen
        If MAD < MADmin
          MADmin = MAD
        End
      End
    End
  End
  VorausgesagterFrame(i+Blocksize,j+Blocksize) = TargetFrame //Put the best matching block in the predicted image
  lblk und jblk setzen //Blockindes setzen
  Mvx und mvy setzen //Speichern des geschätzten Bewegungsvektor
End
End

```

In welchen Bildregionen erwarten Sie gute bzw. schlechte Ergebnisse der Block-basierten Bewegungsbestimmung (Begründung)?

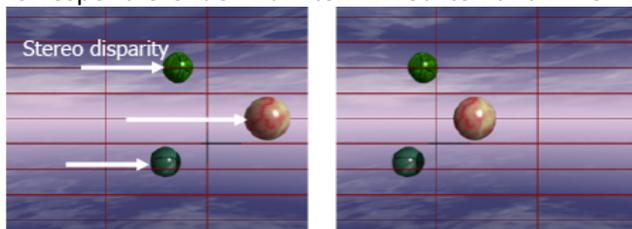
Bildregionen, welche durch gerade Linien definierte Bildinhalte beinhalten, sind mit blockbasierte Verfahren gut zu bestimmen d.h. Gruppen von Bewegungsvektoren sind scharf durch eine Kante unterteilt, dadurch kommt es nicht zu störenden Artefakten an den Blockgrenzen, da diese Teil des Bildes sind und nicht durch die blockbasierte Bewegungsbestimmung erschaffen werden.

Erläutern Sie den Begriff ‚stereo disparity‘.

Beim menschlichen Sehen entsteht der Stereoeindruck dadurch, dass die von den 2 Augen erfassten Bilder nicht völlig identisch sind, sondern gewisse geometrische Abweichungen (stereo disparities) aufweisen.

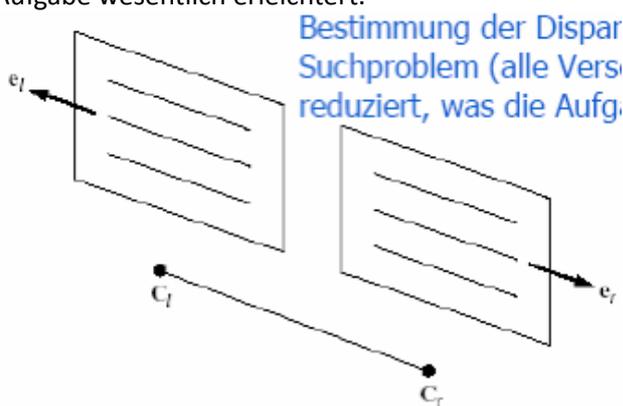
Das Ausmaß dieser Abweichungen liefert Information über die Tiefe eines Szenenpunktes, d.h. seine Entfernung vom Auge (bzw. der Kamera).

Disparity (auch Verschiebungsvektor oder Parallaxe genannt): Gibt die Verschiebung zwischen korrespondierenden Punkten im rechten und linken Bild an.

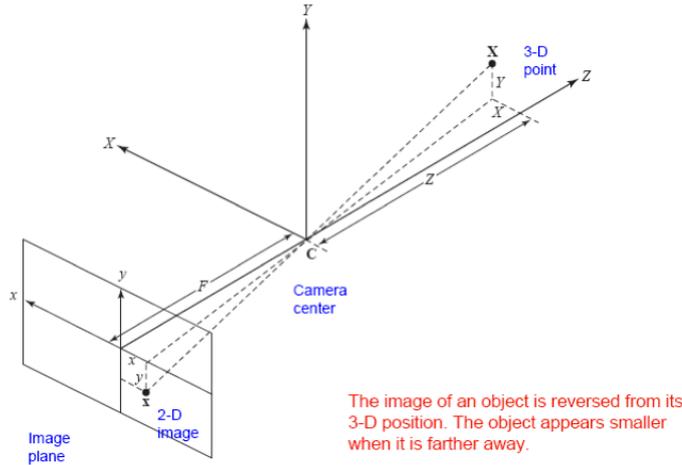


Was versteht man unter ‚Epipolargeometrie‘ und wozu wird sie verwendet

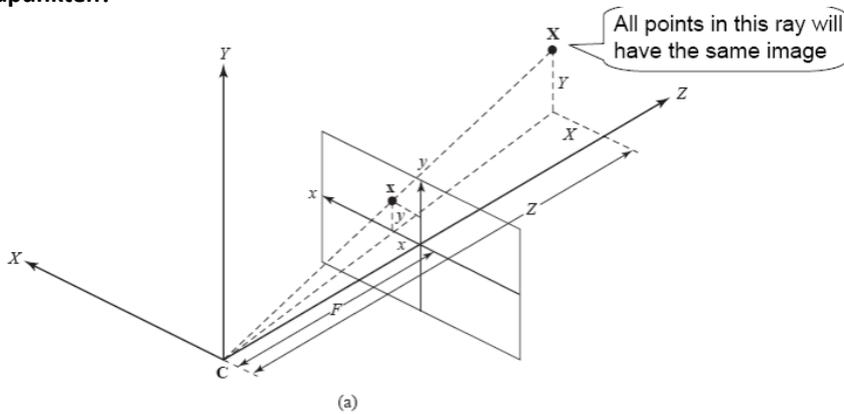
Sind die Kameraposition und weitere Kameraparameter bekannt (kalibrierte Kameras), so kann das Stereobildpaar in Epipolargeometrie transformiert werden. In Epipolargeometrie liegen korrespondierende Punkte auf horizontalen Geraden. Die Bestimmung der Disparity wird dadurch von einem 2D auf ein 1D Suchproblem (alle Verschiebungsvektoren sind horizontal) reduziert, was die Aufgabe wesentlich erleichtert.



Beschreiben Sie das Modell einer Abbildung mittels Lochkamera (pinhole camera) an Hand einer Skizze



Welcher Zusammenhang besteht dabei (Pinhole Camera) zwischen den 3D Raumpunkten und 2D Bildpunkten?



$$\frac{x}{F} = \frac{X}{Z}, \frac{y}{F} = \frac{Y}{Z} \Rightarrow x = F \frac{X}{Z}, y = F \frac{Y}{Z}$$

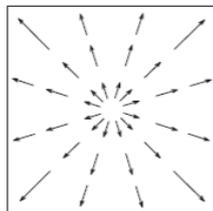
x, y are inversely related to Z

Welches Problem der Mehrdeutigkeit tritt bei der Abbildung auf?

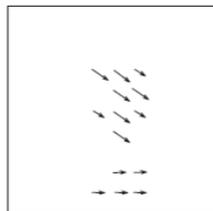
Alle Punkte auf Strahl haben selbe Abbildung in Bild.

Erläutern Sie den Unterschied zwischen folgenden Bewegungsmodellen mittels geeigneter Skizzen: global / block-based / pixel-based / region-based

Global:
Entire motion field is represented by a few global parameters



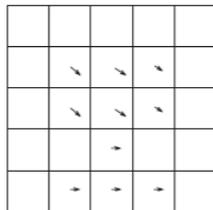
(a)



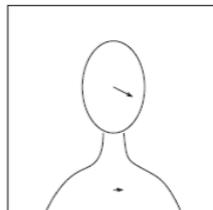
(b)

Pixel-based:
One MV at each pixel, with some smoothness constraint between adjacent MVs.

Block-based:
Entire frame is divided into blocks, and motion in each block is characterized by a few parameters.



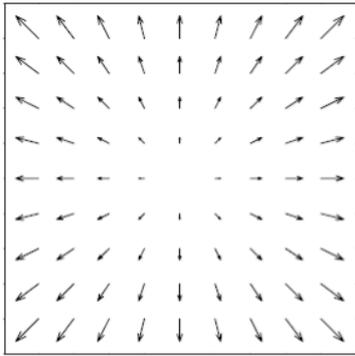
(c)



(d)

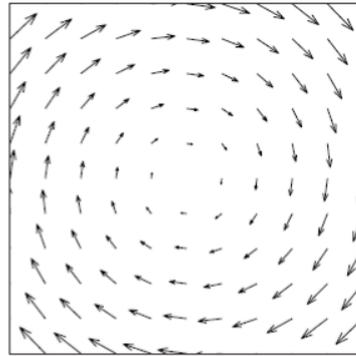
Region-based:
Entire frame is divided into regions, each region corresponding to an object or sub-object with consistent motion, represented by a few parameters.

Nennen Sie 3 verschiedene Arten der Kamerabewegung und skizzieren Sie das zugehörige Feld der Bewegungsvektoren.



(a)

Camera zoom

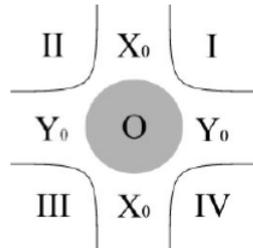


(b)

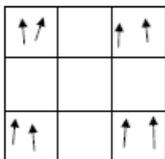
Camera rotation around Z-axis (roll)

Wie könnte man die Kamerabewegung aus einem beobachteten Feld von Bewegungsvektoren rekonstruieren?

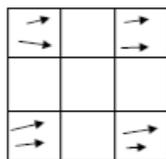
- Split image into 7 Regions



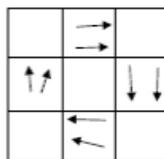
- Compute mean motion in each region
 - tilt, boom: little horizontal movement in Regions I, II, III, IV
 - pan, track: little vertical movement in Regions I, II, III, IV
 - roll: little vertical motion in X0 and little horizontal motion in Y0
 - zoom, dolly: vertical motion in X0 and horizontal motion in Y0



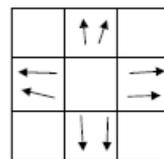
tilt, boom



pan, track

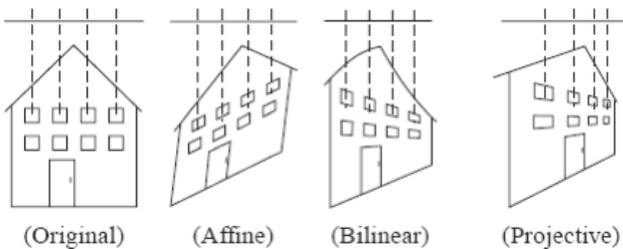


roll



zoom, dolly

Erläutern Sie die Begriffe projektive Abbildung, affine Abbildung und bilineare Abbildung.



(Original)

(Affine)

(Bilinear)

(Projective)

Affine und Bilineare Abbildungen sind Näherungen an die Projektive Abbildung!

Projektive Abbildung ist als Einzige vom Chirping betroffen, d.h. die räumlichen Frequenzen erhöhen sich mit zunehmender Distanz (Fensterabstand). Diese Abbildung ist eigentlich die Beste, jedoch die Komplizierteste, da hier Brüche vorkommen.

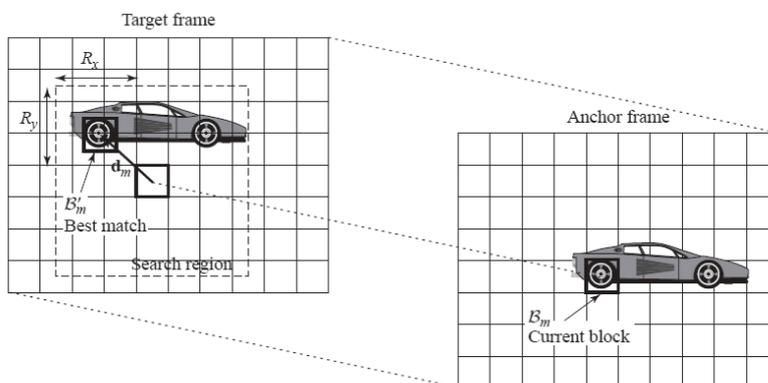
Converging (Keystone) tritt bei der **Bilinearen** Abbildung auf, da die Linien nicht mehr parallel sind und sich so im Endlichen treffen. Die Bilineare Abbildung wird auch (wie projektive) durch 8 Parameter bestimmt, allerdings sind diese rein polynomiell und es treten keine Brüche auf.

Affine Abbildung wird nur durch 6 Parameter bestimmt und ist auch rein polynomiell.

Erläutern Sie das Prinzip eines blockbasierten Verfahrens zur Bestimmung der Bewegungsvektoren.

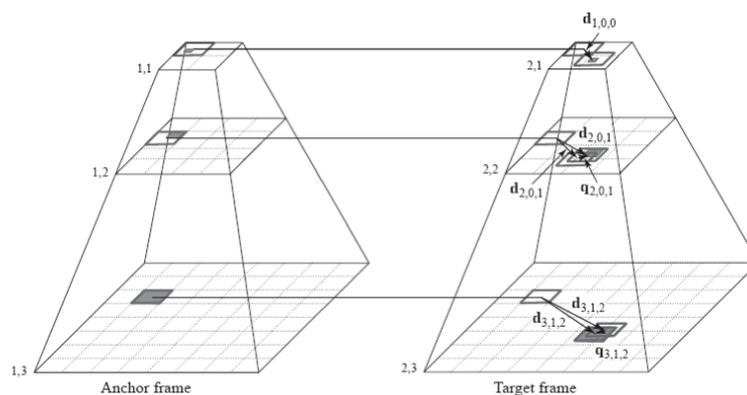
Was ist das Prinzip eines hierarchischen Verfahrens und welchen Vorteil hat es?

Ein Bildausschnitt wird im nächsten Bild gesucht und die Verschiebung ist der Bewegungsvektor. Bei der vollständigen Suche wird das gesamte Suchfenster nach dem Bildausschnitt durchsucht. Hierbei wird an jeder Position die Abweichung der Grauwerte zwischen dem Bildausschnitt und dem aktuellen Ausschnitt berechnet. Für eine Übereinstimmung wird nun der Bereich herangezogen, an dem diese Abweichung der Grauwerte am kleinsten ist.



Hierarchische Suche:

Auflösung der Bilder wird verkleinert, danach erfolgt Blockmatching in Bild mit verkleinerter Auflösung; das Ergebnis (ungefährer Bereich) wird als Wert für die nächsthöhere Auflösung verwendet. Dies erfolgt so lange, bis die vorhandene Auflösung erreicht ist. Vorteil: viel schneller als vollständige Suche



Beschreiben Sie anhand einer Skizze das Bayer Pattern

G1	R2	G3	R4	G5
B6	G7	B8	G9	B10
G11	R12	G13	R14	G15
B16	G17	B18	G19	B20
G21	R22	G23	R24	G25

Als Bayer-Sensor bezeichnet man einen [Fotosensor](#), der, ähnlich einem [Schachbrett](#), mit einem [Farbfilter](#) überzogen ist, das meist zu 50% aus Grün, und je 25% aus Rot und Blau besteht. Dabei wird berücksichtigt, dass das menschliche Auge auf Grün empfindlicher reagiert als auf andere Farben.

Erläutern Sie an Hand der Skizze bilineare Interpolation von Farbwerten

Siehe Skizze von voriger Frage.

Interpolation of green pixels

– The average value of the upper, lower, left and right neighbors yield the interpolated value of the regarded pixel: $G8 = (G3 + G7 + G9 + G13) / 4$

Interpolation of red/blue pixels

– At a pixel position that contains a green value: Average of neighboring pixels of the respective color: $B9 = (B8 + B10) / 2$ $R9 = (R4 + R14) / 2$

– At a pixel position that contains a red or blue value: Average of diagonally neighboring 4 pixels: $R8 = (R2 + R4 + R12 + R14) / 4$ $B12 = (B6 + B8 + B16 + B18) / 4$

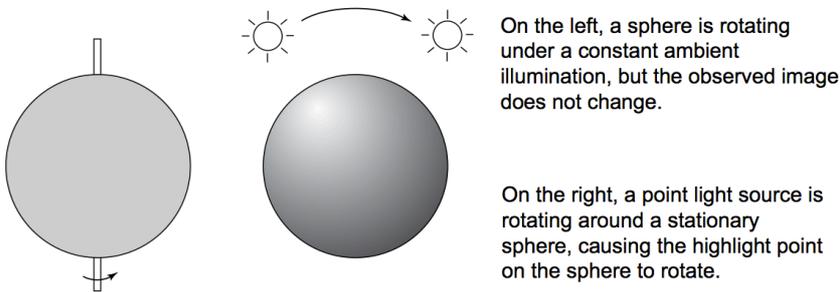
Beschreiben Sie den Deblocking Filtern. Insbesondere im Zusammenhang mit H.264

The filter operates on the edges of each 4×4 or 8×8 transform block in the luma and chroma planes of each picture. Each small block's edge is assigned a boundary strength based on whether it is also a macroblock boundary, the coding (intra/inter) of the blocks, whether references (in motion prediction and reference frame choice) differ, and whether it is a luma or chroma edge. Stronger levels of filtering are assigned by this scheme where there is likely to be more distortion. The filter can modify as many as three samples on either side of a given block edge (in the case where an edge is a luma edge that lies between different macroblocks and at least one of them is intra coded). In most cases it can modify one or two samples on either side of the edge (depending on the quantization step size, the tuning of the filter strength by the encoder, the result of an edge detection test, and other factors).

Mehr zu Optical Flow

Optical Flow is the pattern of *apparent motion* of objects, surfaces and edges in a visual scene caused by the relative motion between a camera and the scene. It depends on illumination and object surface texture.

- 2-D Motion: Projection of 3-D motion, depending on 3D object motion and projection operator
- Optical flow: “Perceived” 2-D motion based on changes in image pattern, also depends on illumination and object surface texture



sample exam questions:

How are the key problems solved for e.g. the HS method?

?

What is the brightness constancy? What is the Optical Flow Equation?

brightness constancy: the tendency for a visual object to be perceived as having the same brightness under widely different conditions of illumination. Führt zu Problemen mit texturlosen Regionen und Überdeckungen.

Explain the relationship of 2d motion and OF. Give example.

The goal of OF estimation is to compute approximation of the motion field (optical flow field) from a video sequence.

Compare the HS and LK method

Die Lucas-Kanade-Methode zur Berechnung des optischen Flusses geht auf die beiden Forscher Bruce D Lucas und Takeo Kanade zurück. Sie schlugen diese Methode erstmals 1981 vor. Die Methode ist ein beliebtes Verfahren, das noch heute weite Anwendung findet. Die Zusatzbedingung, die zur Berechnung des optischen Flusses benötigt wird, ist die Annahme der Gleichheit des Flusses in der lokalen Umgebung des zentralen Pixels, für den der Fluss bestimmt wird.

Vorteile: keine Block Artefakte, besser als Block Matching, Noise reduziert aber immer noch vorhanden. Leicht und schnell.

Nachteile: unbrauchbar in texturlosen Gebieten, Fehler an Grenzen

The Horn–Schunck method of estimating optical flow is a global method which introduces a global constraint of smoothness to solve the aperture problem.

Vorteile: smooth flow(?), Gute Resultate in texturfreien Gebieten

Nachteile: Oversmoothing an den Grenzen, langsam weil iterativ

Explain the challenge and a possible solution for it.

Challenges: Beide funktionieren nur gut bei kleinen Bewegungen. Lösung: Pyramide mit immer höher aufgelösten Bildern

Where and how can you evaluate an OF algorithm?

The Middlebury Dataset offers test sequences and their GT.