

2. Schnitterkennung

Videoanalyse

Stephan Kopf

Inhalt

- Definition: Schnitt
- Klassifikation eines Schnittes
- Vorgehensweise bei der automatischen Schnitterkennung
 - Pixelbasierte Verfahren
 - Histogramme
 - Aggregierte Bilddaten
 - Kantenbasierte Verfahren (Canny, ECR)
- Bewertung der Qualität
- Ergebnisse zur Erkennung harter und weicher Schnitte

Definition: Schnitt

Schnitte (*cut*)

- liefern Informationen über den Produktionsprozess eines Filmes.
- trennen kontinuierliche Aufnahmen, die als *Kameraeinstellungen* (*shot*) bezeichnet werden.
- sind die kleinste Einheit eines Filmes, bei der die zeitliche Dimension noch enthalten ist.
- sind Voraussetzung für die im Folgenden vorgestellten Verfahren zur Analyse von Videos.

Szenen (*scene*)

- Inhaltlich ähnliche und zeitlich aufeinander folgende Kameraeinstellungen

Dialoge

- Szenen, bei denen das Bild wiederholt zwischen zwei oder mehreren Personen wechselt

Klassifikation eines Schnittes (I)

Harter Schnitt

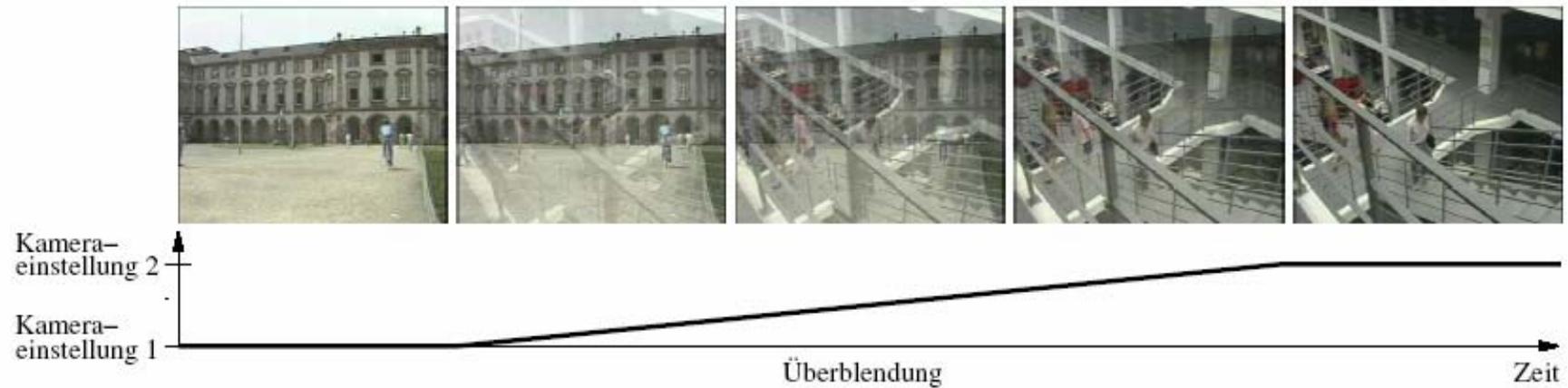
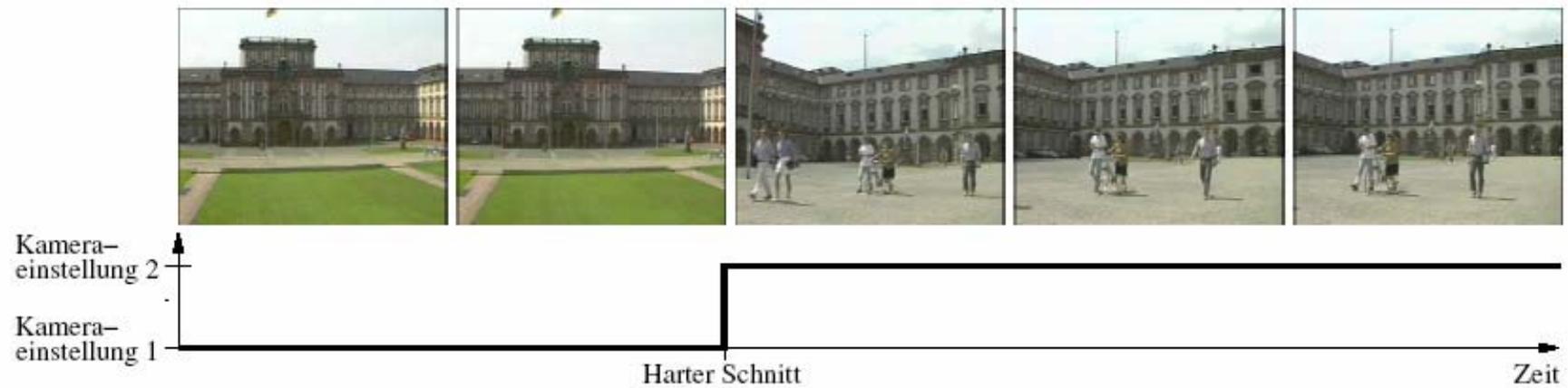
- kein Übergang zwischen zwei Kameraeinstellungen

Weicher Schnitt

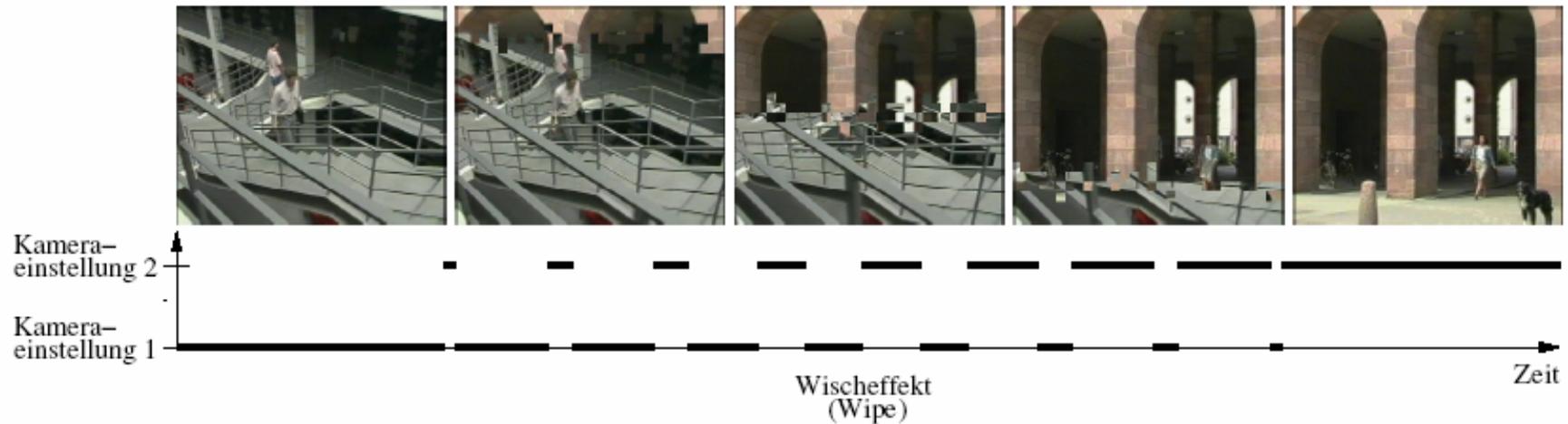
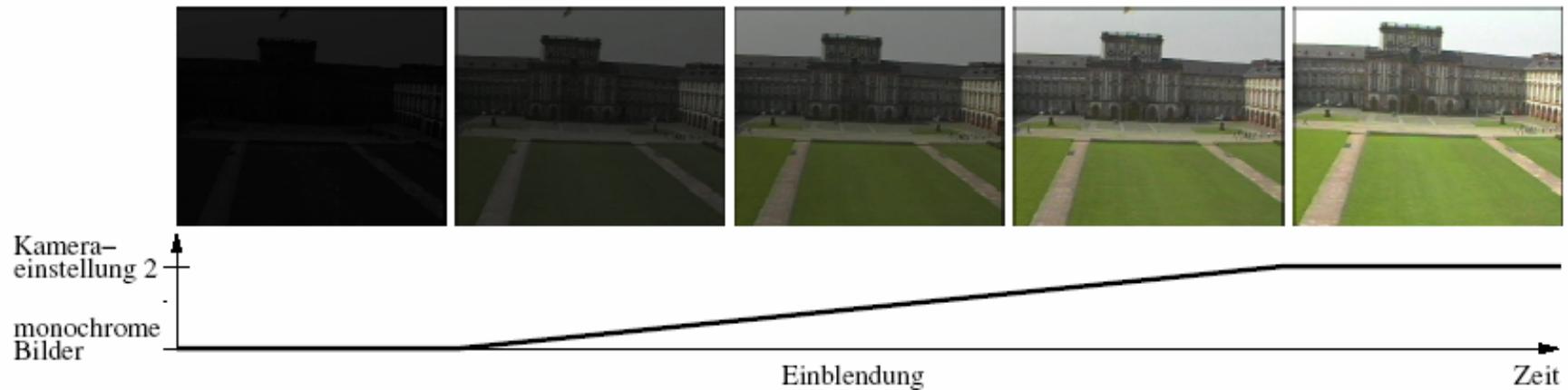
- künstlicher Übergang zwischen den beiden Kameraeinstellungen
- *Überblendung (dissolve)*: Der Wechsel von einer Kameraeinstellung zur nächsten erfolgt kontinuierlich.
- *Ein- und Ausblendungen (fade)* sind spezielle Überblendungen, bei denen eine der beiden Kameraeinstellungen aus monochromen – häufig schwarzen – Bildern besteht.
- Bei *Wischeffekten (wipe)* werden Pixel ausgewählter Bildregionen sofort verändert.

	Dauer eines Schnittes	
	Schnitt zwischen zwei Bildern	Schnitt über mehrere Bilder
Änderung der Werte (einzelner) Pixel zwischen benachbarten Bildern	Harter Schnitt	Wischeffekt
Kontinuierliche Änderung aller Pixelwerte über einen längeren Zeitraum		Überblendung Ein- oder Ausblendung

Klassifikation eines Schnittes (II)



Klassifikation eines Schnittes (III)



Automatische Schnitterkennung (I)

Idee

- Betrachte die Ähnlichkeit zweier benachbarter Bilder im Video
- Bilder einer Kameraeinstellung haben im Allgemeinen eine wesentlich stärkere Ähnlichkeit als Bilder unterschiedlicher Kameraeinstellungen.

Vorgehensweise

- Berechne Unterschied zwischen jeweils benachbarten Bildern
- Glätte Werte zur Reduktion von Klassifikationsfehlern durch Rauschen
- Hohe Werte deuten auf einen harten Schnitt hin

aber:

- Erkennung weicher Schnitte nur schwer möglich, da nur geringe Unterschiede zwischen benachbarten Bildern auftreten

Automatische Schnitterkennung (II)

Erkennung harter Schnitte

- Analysiere Ähnlichkeit zweier Bilder I_i und I_j mit $1 \leq i < j \leq N$ innerhalb einer Videosequenz $(I_1 \dots I_N)$

Ablauf

- Definiere Abbildung τ , die ein Bild I_i in einen Merkmalsraum transformiert:

$$\tau : \mathbb{N}^m \rightarrow F,$$

\mathbb{N}^m Raum, der durch alle Bilder $(I_n \in \mathbb{N}^m)$ aufgespannt wird.

F spezifiziert den Merkmalsraum mit $\tau(I_n) \in F$.

Automatische Schnitterkennung (III)

Erkennung harter Schnitte

2. Definiere robustes Distanzmaßes D zur Beurteilung der Ähnlichkeit zweier Bilder anhand der Merkmalswerte:

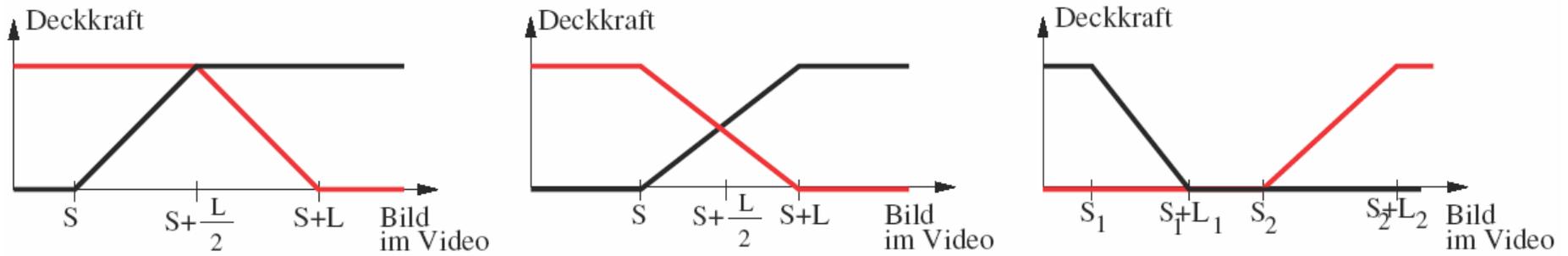
$$D : F \times F \rightarrow \mathbb{R}^+.$$

3. Ein harter Schnitt existiert zwischen den Bildern i und j , falls die Distanz zweier Bilder einen Schwellwert T überschreitet:

$$D_{i,j} = D(\tau(I_i), \tau(I_j)) > T_{i,j}.$$

Automatische Schnitterkennung (IV)

Erkennung weicher Schnitte



Additive Überblendung

Kreuz-Überblendung

Aus-/Einblendung

S: Startzeitpunkt der Überblendung, L: Dauer der Überblendung

Annahmen:

- Veränderung der Transparenz kann durch eine lineare Funktion approximiert werden
- Bildinhalt beider Kameraeinstellungen verändert sich nicht

Automatische Schnitterkennung (V)

Erkennung weicher Schnitte

- Bildinhalt während einer Überblendung:

$$I_k = \alpha_k \cdot I_S + \beta_k \cdot I_{S+L} \quad \text{mit} \quad S \leq k \leq S + L.$$

- bei additiver Überblendung/Ausblendungen:

$$\alpha_k = \begin{cases} 1 & \text{für } S \leq k \leq S + L/2 \\ 1 - \frac{k-S-L/2}{L/2} & \text{für } S + L/2 < k \leq S + L \end{cases}$$
$$\beta_k = \begin{cases} \frac{k-S}{L/2} & \text{für } S \leq k \leq S + L/2 \\ 1 & \text{für } S + L/2 < k \leq S + L, \end{cases}$$

- Einblendungen werden durch ein Vertauschen der Parameter α und β modelliert

Automatische Schnitterkennung (VI)

Erkennung weicher Schnitte

- bei Kreuz-Überblendungen/Ausblendungen:

$$\alpha_k = 1 - \frac{k - S}{L} \quad \text{und} \quad \beta_k = 1 - \alpha_k \quad \text{mit} \quad S \leq k \leq S + L.$$

- bei Ausblendungen kann statt schwarzer Bilder auch $\beta_k = 0$ gesetzt werden

Regeln für die Erkennung weicher Schnitte:

- gleichmäßige Änderung der Distanzmaße zwischen jeweils zwei benachbarten Bildern:

$$D_{i,i+1} \approx D_{i+1,i+2} \quad \forall \quad S \leq i < S + \frac{L}{2} - 1 \quad \text{und} \quad S + \frac{L}{2} \leq i < S + L - 1.$$

- Differenz mit steigendem Abstand zwischen den Bildern zu:

$$D_{i,i+j} < D_{i,i+k} \quad \forall \quad S \leq i < i + j < i + k \leq S + L.$$

Automatische Schnitterkennung (VII)

Berechnung eines Distanzmaßes zweier Bilder

- Pixelbasierte Verfahren
- Histogramme
- Aggregierte Bilddaten
- Kantenbasierte Verfahren

Pixelbasierte Verfahren (I)

Summe der absoluten Pixeldifferenzen

$$D_{SAD} = \frac{1}{N_x \cdot N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} |I_i(x, y) - I_j(x, y)|.$$

Harter Schnitt, falls: $D_{SAD} > T$

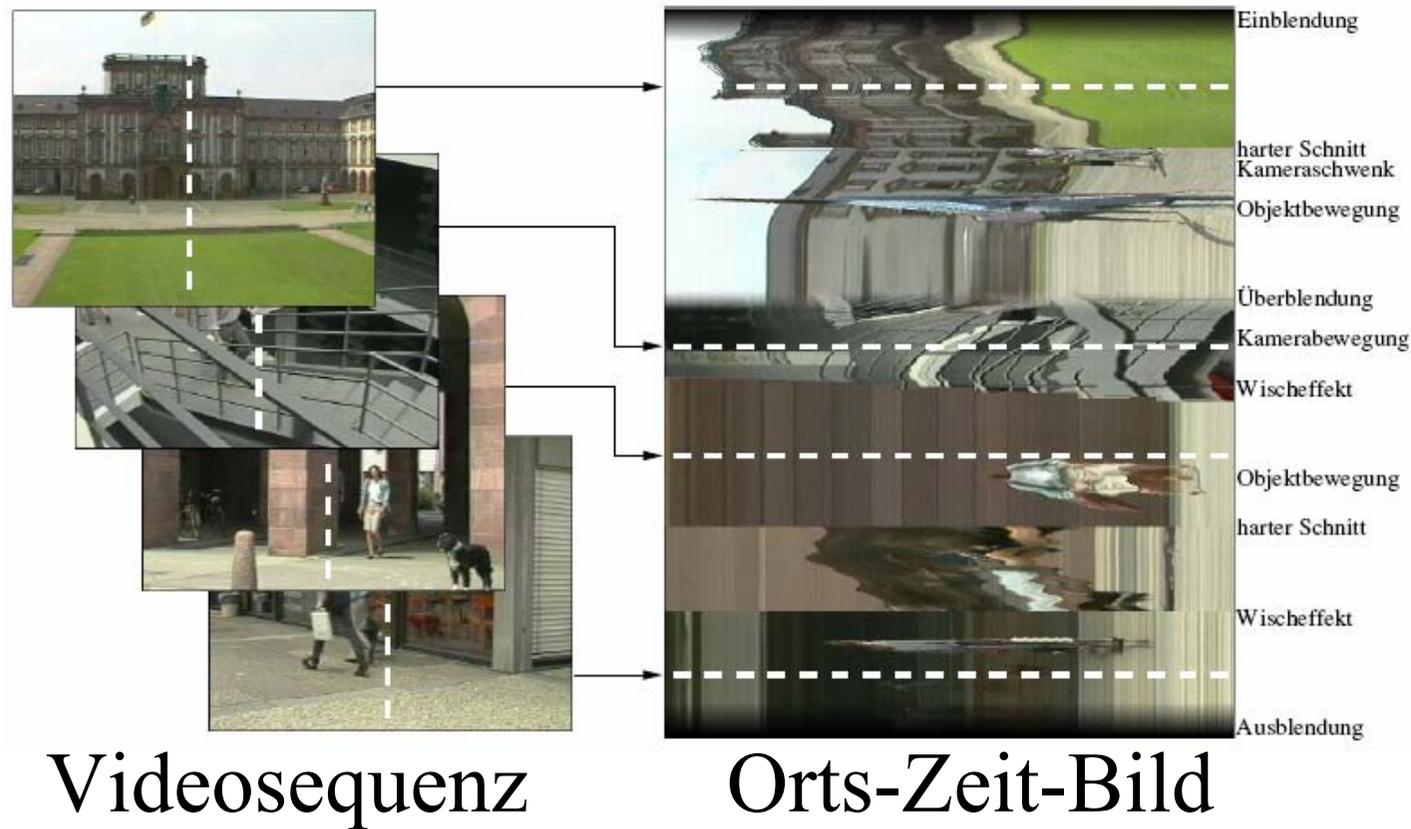
Pixelbasierte Verfahren (II)

Vorteile / Nachteile

- Allgemein sehr robuste und zuverlässige Ergebnisse, jedoch hohe Fehlerraten bei starken Objekt- und Kamerabewegungen
- Geringe Komplexität und dadurch schnell zu berechnen

Pixelbasierte Verfahren (III)

Orts-Zeit-Bilder



Histogramme (I)

Idee

- Ein *Histogramm* speichert für jeden Grau- bzw. Farbwert die absolute oder relative Anzahl der Pixel dieser Helligkeit bzw. Farbe im Bild.

Größe eines Histogramms

- 8-Bit Graustufenbilder: 256 Elemente
- Farbbilder mit 24-Bit Farbtiefe: > 16 Millionen Elemente

Lösung

- einzelne Histogramme für jeden Farbkanal
- Verringerung der Farbtiefe (entfernen der least significant bits für jeden Farbkanal)

Histogramme (II)

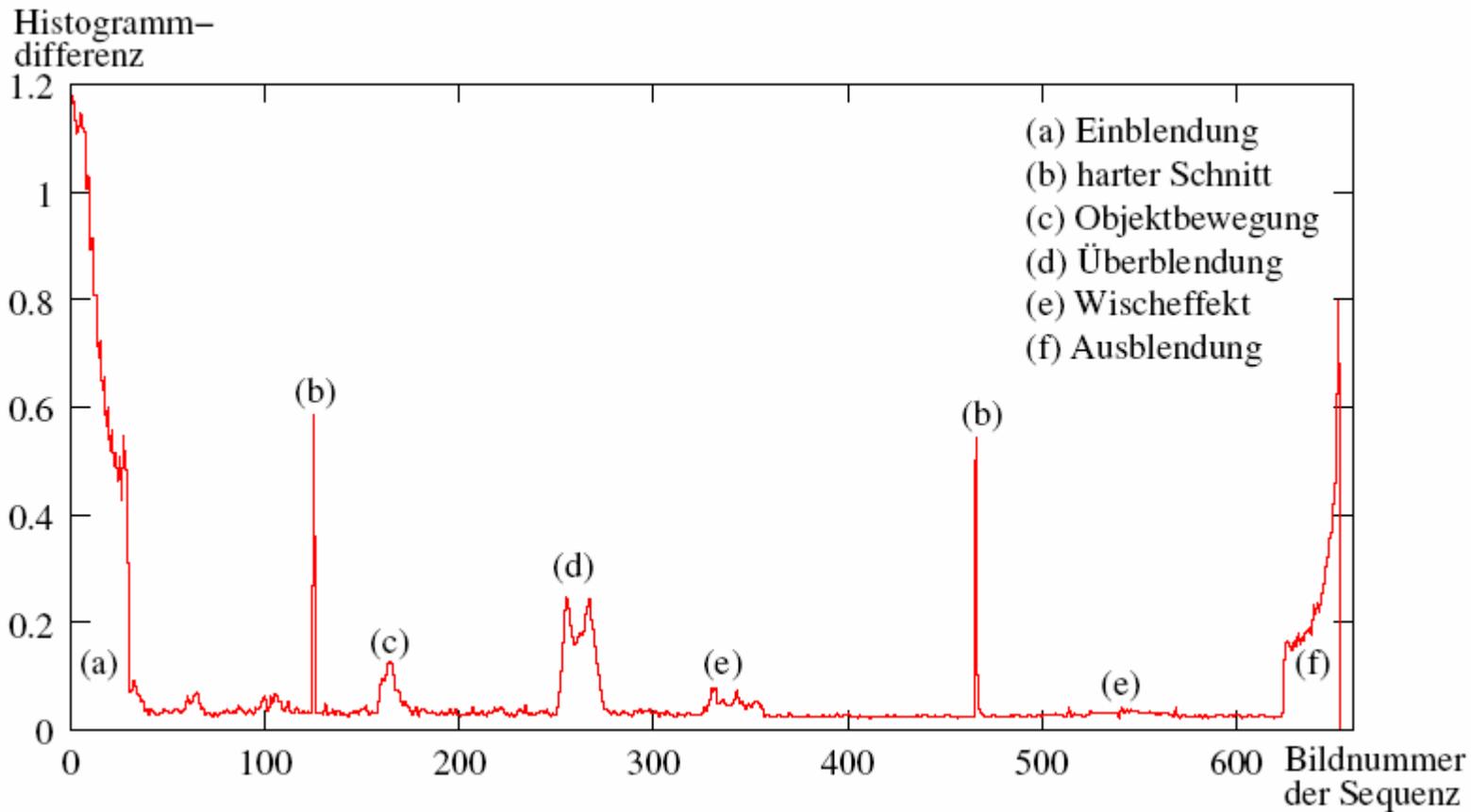
Histogrammdifferenz

- *Minkowski*-Metrik:

$$L_p(H_1, H_2) = \left(\sum_{m=1}^M |H_1(m) - H_2(m)|^p \right)^{\frac{1}{p}}$$

- $P=1$: L_1 -Norm: Summe der absoluten Histogrammdifferenzen
- $P=2$: L_2 -Norm (euklidische Norm): Summe der quadrierten Histogrammdifferenzen

Histogramme (III)



L1-Norm einer Videosequenz mit unterschiedlichen Schnitten

Histogramme (IV)

Vorteile / Nachteile

- Robuste und zuverlässige Ergebnisse sind möglich. Ähnlich den pixelbasierten Distanzmaßen treten Fehlklassifikationen insbesondere bei Helligkeitsänderungen und schnellen Bewegungen großer Objekte auf.
- Geringe Komplexität und dadurch schnell zu berechnen

Aggregierte Bilddaten (I)

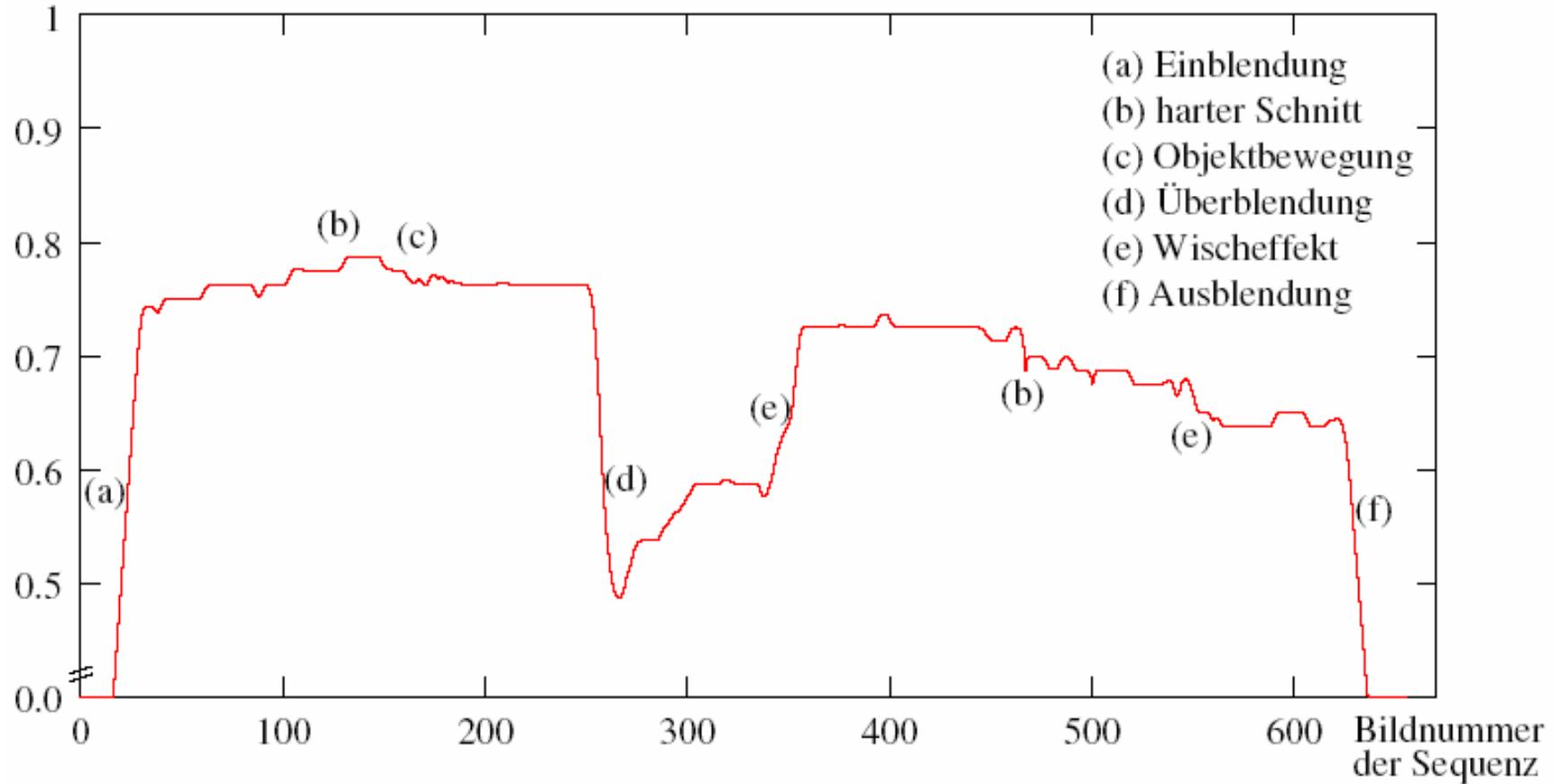
**Standardabweichung I der Helligkeitswerte
aller Pixel eines Bildes:**

$$\sigma_I = \sqrt{\frac{1}{N_x \cdot N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} (I(x, y) - \bar{I})^2}$$

- Die Standardabweichung sinkt an den Rändern von Ein- bzw. Ausblendungen und in der Mitte einer Überblendung, da die Pixel in diesen Bereichen durchschnittliche Helligkeits- bzw. Farbwerte annehmen.
→ Erkennung von Ein-, Aus- und Überblendungen

Aggregierte Bilddaten (II)

Standardabweichung
der Helligkeit



Aggregierte Bilddaten (III)

Probleme

- schnelle Kamera- oder Objektbewegungen:
 - Kameraeinstellungen sind häufig unscharf
 - Standardabweichung sinkt
 - viele falsch erkannte Schnitte

Ursachen für Unschärfe

- Durch die Aufnahme
- durch die Kompression des digitalen Videos

Aggregierte Bilddaten (IV)

Kantenbasierter Kontrast (*edge contrast*)

schwache Kanten $w_I = \sum_{x,y} \begin{cases} I(x,y) & \text{falls } \theta_w \leq I(x,y) < \theta_s, \\ 0 & \text{sonst} \end{cases}$

starke Kanten $s_I = \sum_{x,y} \begin{cases} I(x,y) & \text{falls } I(x,y) \geq \theta_s, \\ 0 & \text{sonst.} \end{cases}$

$$EC = 1 + \frac{s_I - w_I - 1}{s_I + w_I + 1}, \quad EC \in [0, 2].$$

Aggregierte Bilddaten (V)

Kantenbasierter Kontrast

$$EC = 1 + \frac{s_I - w_I - 1}{s_I + w_I + 1}, \quad EC \in [0, 2].$$

Anteil starker und schwacher Kanten	EC
$s_I = 0$	0
$s_I < w_I$	$0 < EC < 1$
$s_I \approx w_I > 0$	1
$s_I > w_I$	$1 < EC < 2$
$s_I \gg w_I$	2

Aggregierte Bilddaten (VI)

Vorteile/Nachteile

- Ein wesentlicher Vorteil der Schnitterkennung mit aggregierten Bilddaten ist der geringe Einfluss der Kamera- bzw. Objektbewegung und die geringe Komplexität der Berechnung.
- Nur bei schnellen Bewegungen treten höhere Fehlerraten auf, da das Bild hierbei häufig an Schärfe verliert.

Kantenbasierte Verfahren: Canny (I)

Idee der kantenbasierten Verfahren

- Veränderung der Kanten zweier Bildern analysieren

Berechnung einer Kantenbildes: *Canny*-Kantendetektor

- Kanten entstehen durch Helligkeitsänderungen zwischen benachbarten Pixeln
1. Um Bildrauschen zu reduzieren, wird das Bild mit einer Annäherung der Gaußschen Normalverteilung **geglättet**:

Maske:
$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

Kantenbasierte Verfahren : Canny (II)

2. Die **Gradienten** der einzelnen Pixel werden mit dem Sobeloperator ermittelt:

$$\text{Maske: } \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

Grad_X

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

Grad_Y

3. **Kantenstärke**: Summe der beiden Gradienten

Kantenbasierte Verfahren : Canny (III)

4. Berechne **Kantenanstiege** für jedes Pixel:

$$\text{Kantenanstieg} = \begin{cases} \arctan\left(\frac{\text{Grad}_Y}{\text{Grad}_X}\right) & \text{falls } \text{Grad}_X \neq 0 \\ 0^\circ & \text{falls } \text{Grad}_X = 0, \text{Grad}_Y = 0 \\ 90^\circ & \text{falls } \text{Grad}_X = 0, \text{Grad}_Y \neq 0 \end{cases}$$

4 Kantenrichtungen sind zulässig:

0° , 45° , 90° und 135°

Kantenbasierte Verfahren : Canny (IV)

5. Kante soll nur 1 Pixel breit sein:

→ Suche lokaler Maxima (*non maxima suppression*):
setze Kantenstärke auf Null (Pixel soll keine Kante sein), falls gilt:

- ein Nachbarpixel hat eine höhere Kantenstärke und
- der Kantenanstieg führt vom aktuellen Pixel nicht zu diesem Pixel

Kantenbasierte Verfahren : Canny (V)

6. Festlegung der Kantenpixel (**Hysteresis**)

- 2 Schwellwerte werden definiert: T_{low} und T_{high}
- falls Kantenstärke $> T_{\text{high}}$:
→ setze Pixel als Kantenpixel
- falls $T_{\text{low}} < \text{Kantenstärke} < T_{\text{high}}$
und
Pixel grenzt an ein Kantenpixel:
→ setze Pixel als Kantenpixel
- Ergebnis: Kantenbild

Kantenbasierte Verfahren : Canny (VI)

Beispiel: *Canny*-Kantendetektor



Originalbild



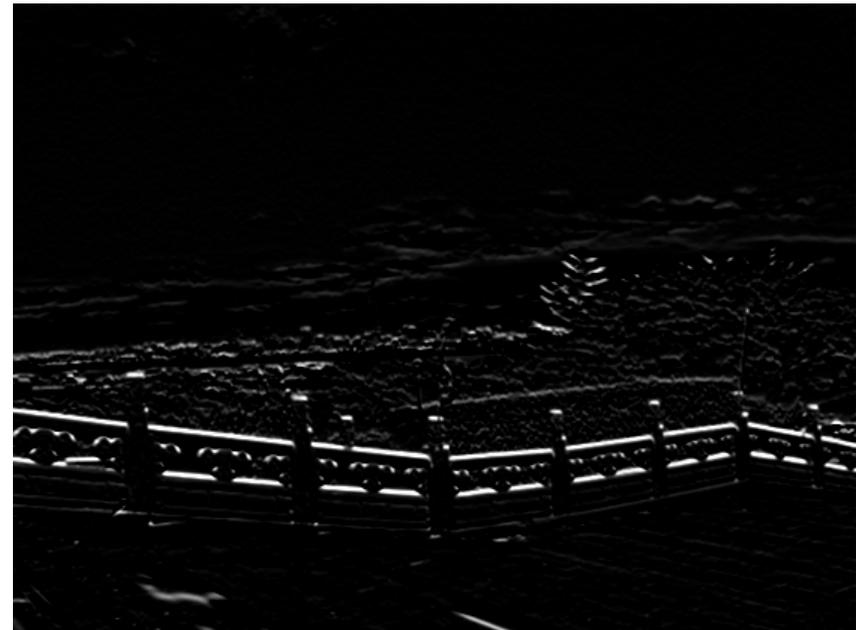
geglättetes Bild

1. Glättung

Kantenbasierte Verfahren : Canny (VII)



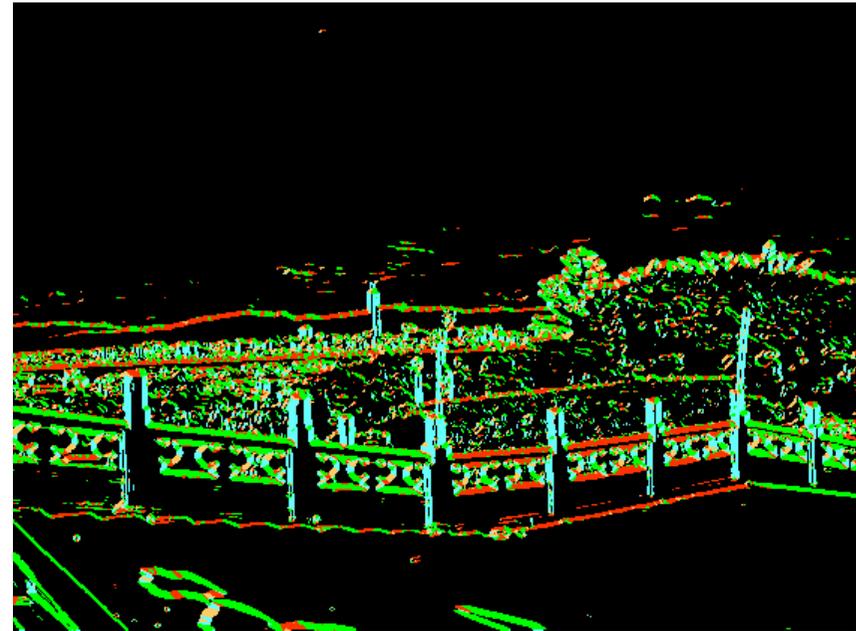
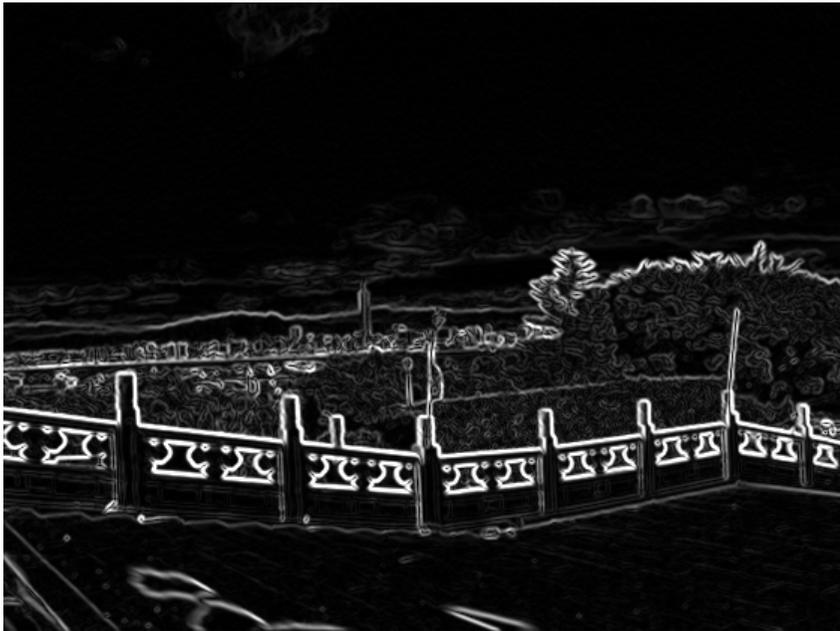
Gradient in x-Richtung



Gradient in y-Richtung

2. Gradientenbilder

Kantenbasierte Verfahren : Canny (VIII)

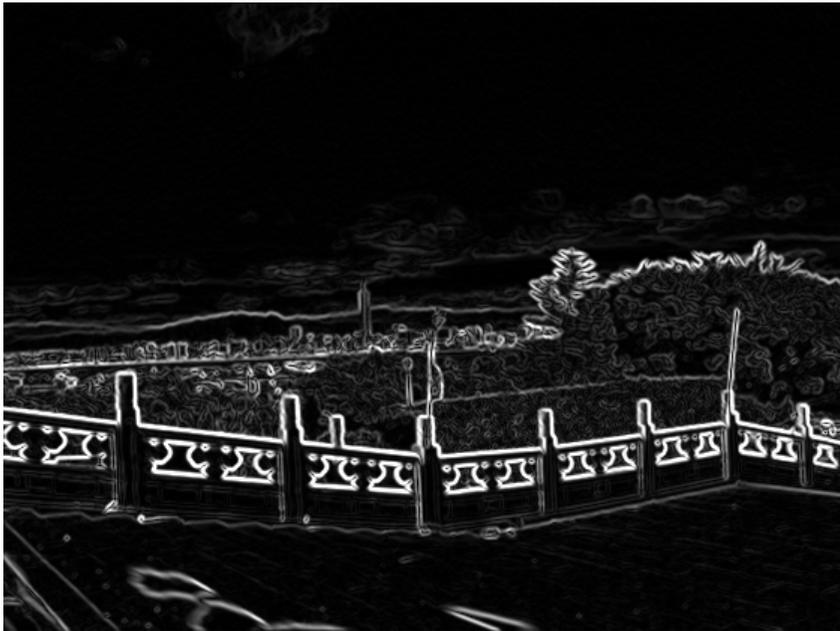


0 Grad, 45 Grad, 90 Grad, 135 Grad

3. Kantenstärke

4. Kantenanstieg

Kantenbasierte Verfahren : Canny (VII)



ohne Suche lokaler Maxima



Suche lokaler Maxima

5. Suche lokaler Maxima

Kantenbasierte Verfahren : Canny (VII)



Kantenbild ohne Schwellwerte



starke und schwache Kanten

6a. Kantenbild vor Hysterese

Kantenbasierte Verfahren : Canny (VII)



für das Kantenbild ausgewählte
schwache und starke Kanten



Kantenbild

6b. Kantenbild nach Hysterese

Kantenbasierte Verfahren : ECR (I)

Kantenänderungsrate (*edge change ratio, edge change fraction*)

- Kanten zweier Bilder i und j eines Videos werden mit Hilfe des *Canny*-Kantendetektors ermittelt.
- Kantenpixel, die im ersten aber nicht im zweiten Bild enthalten sind, werden als *ausgehende Kantenpixel* bezeichnet, die im zweiten Bild neu hinzukommenden Kantenpixel als *eingehende Kantenpixel*.

$$\rho_{out}(i) = \frac{E_{out}(i)}{S_i}$$

$$\rho_{in}(j) = \frac{E_{in}(j)}{S_j}$$

$$ECR_{i,j} = \max \{ \rho_{out}(i), \rho_{in}(j) \}.$$

- Anzahl der aus- und eingehenden Kantenpixel: $E_{out}(i)$, $E_{in}(i)$
- Anteil der ausgehenden und eingehenden Kantenpixel zur gesamten Anzahl der Kantenpixel S_i : $p_{out}(i)$, $p_{in}(i)$

Kantenbasierte Verfahren : ECR (II)

Problem

- Kamera- oder Objektbewegungen erhöhen die Anzahl der eingehenden und ausgehenden Kanten deutlich
→ Ausgleich der Kamerabewegung (Kapitel 3)
- Weiterhin geringe Verschiebung der Kanten zweier Bilder möglich
→ Kanten des Bildes werden verbreitert (Dilatation)

Dilatation / Erosion

- Operatoren, die für ein Graustufenbild I und ein Strukturelement B mit Radius r (z.B. Kreis oder Quadrat) folgendermaßen definiert sind:

$$\text{Dilatation: } (I \circ B)(x) = \sup\{I(x - r), r \in B\}$$

$$\text{Erosion: } (I \bullet B)(x) = \inf\{I(x - r), r \in B\}$$

- Für diskrete Pixelwerte kann Maximum/Minimum verwendet werden.

Kantenbasierte Verfahren : ECR (III)

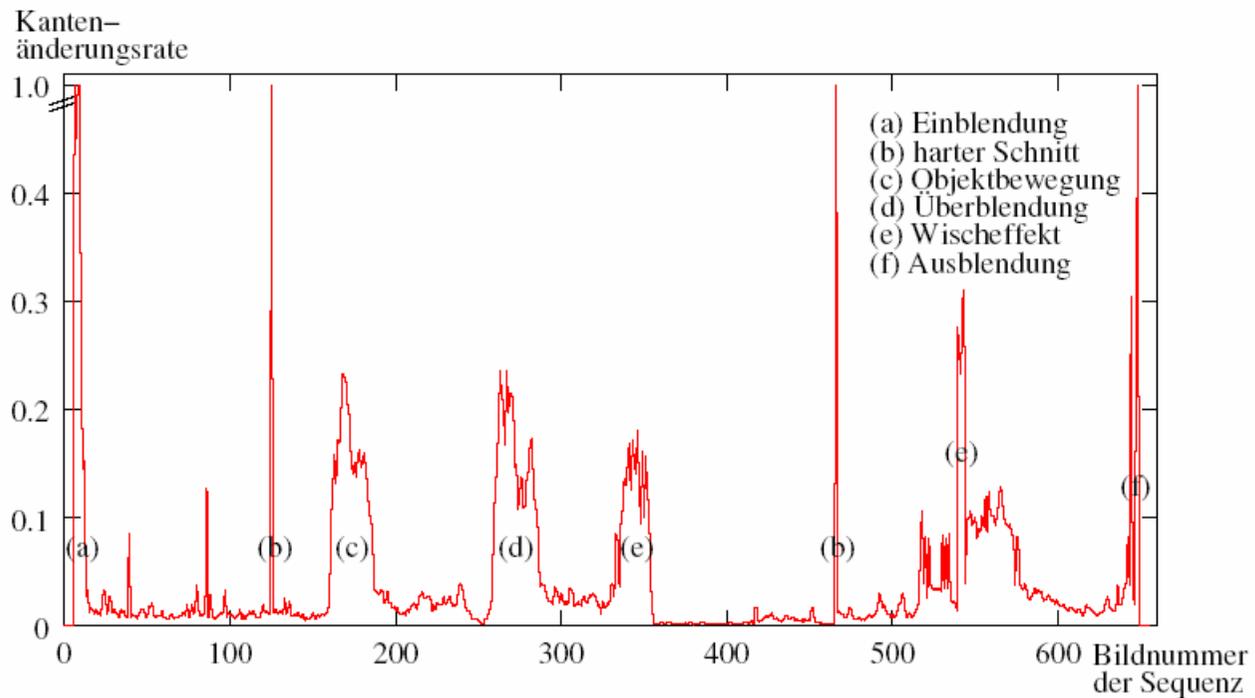
Erkennung von Schnitten

- einzelner hoher Wert
→ harter Schnitt
- mehrere zusammenhängende leicht erhöhte Werte
→ weicher Schnitt
- Zusätzlicher Indikator für weiche Schnitte:
 - Verhältnis der eingehenden zu den ausgehenden Kanten.
 - Kanten verschwinden bei einer Ausblendung oder in der ersten Hälfte einer Überblendung ($p_{out}(i) > p_{in}(i)$), bei einer Einblendung und in der zweiten Hälfte einer Überblendung ist das Verhältnis umgekehrt.

Kantenbasierte Verfahren : ECR (IV)

Qualität

- sehr hohe Anzahl an fehlerhaft erkannten Schnitten durch Objektbewegungen



Bewertung der Qualität (I)

- **Präzision** (*precision*)

$$P = \frac{C}{C+F} \in [0, 1]$$

- **Vollständigkeit eines Suchergebnisses** (*recall*)

$$V = \frac{C}{C+M} \in [0, 1].$$

- C: korrekt erkannten Schnitte
- F: fehlerhaft erkannte Schnitte
- M: nicht erkannte Schnitte
- Alle erkannten Schnitte sind echte Schnitte:
Präzision \rightarrow max.
- In der Menge der erkannten Schnitte sind alle Schnitte enthalten
Vollständigkeit \rightarrow max.

Bewertung der Qualität (II)

- Die einzelnen Maße ermöglichen keine Aussage über die Qualität eines Verfahrens.
- F1–Maß:

$$F1 = 2 \cdot \frac{P * V}{P + V} \in [0, 1] \quad \text{für } P, V \neq 0.$$

Erkennung harter Schnitte

Beispielhafte Ergebnisse

Verfahren	Präzision	Vollständigkeit	F1	Rechenzeit
Summe absoluter Differenzen	85,2 %	82,7 %	83,9 %	0,86
Kantenänderungsrate	76,1 %	86,5 %	81,0 %	7,78
Histogramm	60,4 %	79,2 %	68,5 %	0,67
Durchschnittlicher Farbwert	56,9 %	68,2 %	62,0 %	0,67
Kontrast	55,7 %	68,9 %	61,6 %	0,76
Kantenbasierter Kontrast	13,3 %	23,5 %	16,9 %	0,75

(Ergebnisse von 9 Testvideos)

Optimierungen für harte Schnitte (I)

- Ziel: Verringerung der Fehlerraten in Kameraeinstellungen mit starken Bewegungen
- Harter Schnitt, falls $D(i) > D_{avg}(i) + T$

dynamischer Schwellwert:

$$D_{avg}(i) = \frac{1}{N} \sum_{j=i-\frac{N}{2}, j \neq i}^{i+\frac{N}{2}} D(j)$$

Optimierungen für harte Schnitte (II)

Beispielhafte Ergebnisse bei einem dynamischen Schwellwert

Verfahren	Präzision	Vollständigkeit	F1
Summe absoluter Differenzen	94,4 %	94,2 %	94,3 %
Kantenänderungsrate	82,8 %	92,2 %	87,2 %
Histogramm	81,4 %	89,0 %	85,0 %
Durchschnittlicher Farbwert	74,1 %	76,6 %	75,3 %
Kontrast	72,7 %	74,6 %	73,6 %
Kantenbasierter Kontrast	21,1 %	20,4 %	20,7 %

Erkennung weicher Schnitte (I)

Erkennung weicher Schnitte

- Eine Unterscheidung zwischen Objekt- oder Kamerabewegung und einem weichen Schnitt ist aufgrund der Ähnlichkeit der Merkmalswerte häufig nicht möglich.
→ wesentlich höhere Fehlerraten

Vorgehensweise bei Kontrast / kantenbasiertem Kontrast

- Analyse eines Intervalls: Werte sinken durch Überlagerung zweier Bilder bis zur Mitte einer Überblendung und steigen anschließend wieder.
- Suche lokales Minimum in der Mitte einer Überblendung.
- Überprüfe, ob Differenz zwischen Minimum und Maximum innerhalb des Intervalls einen Schwellwert überschreitet.

Erkennung weicher Schnitte (II)

Vorgehensweise bei der ECR

$$M_{ECR}(i) = \sum_{j=i-\frac{N}{2}}^{i+\frac{N}{2}} ECR(j) - \max \left\{ ECR(j) : j = i - \frac{N}{2} \dots i + \frac{N}{2} \right\}.$$

- Summiere ECR-Werte innerhalb eines Intervalls und ziehe maximalen ECR-Wert ab
→ harte Schnitte werden ausgeschlossen

Erkennung weicher Schnitte (III)

Vorgehensweise bei Histogrammen

$$M_{HD}(i) = HD_N(i - \frac{N}{2}) - \max \left\{ HD_1(j) : j = i - \frac{N}{2} \dots i + \frac{N}{2} \right\}.$$

- $HD_1(j)$: Histogrammdifferenz zweier benachbarter Bilder j und $j + 1$
 - $HD_N(i)$: Histogrammdifferenz zwischen Bild i und $i + N$
- Die *modifizierte Histogrammdifferenz* M_{HD} hat nur innerhalb eines weichen Schnittes einen hohen Differenzwert.

Erkennung weicher Schnitte (IV)

Überblendungen

Verfahren	Präzision	Vollständigkeit	F1
Kantenänderungsrate	45,0 %	43,1 %	44,0 %
Histogramm	58,3 %	52,1 %	55,0 %
Kontrast	54,2 %	59,2 %	56,6 %
Kantenbasierter Kontrast	46,1 %	37,9 %	41,6 %

Ein- und Ausblendungen

Verfahren	Präzision	Vollständigkeit	F1
Kontrast	97,7 %	74,4 %	84,5 %
Kantenbasierter Kontrast	93,0 %	72,1 %	81,2 %

Fragen ?