

2. Schnitterkennung

Video - Inhaltsanalyse

Stephan Kopf

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

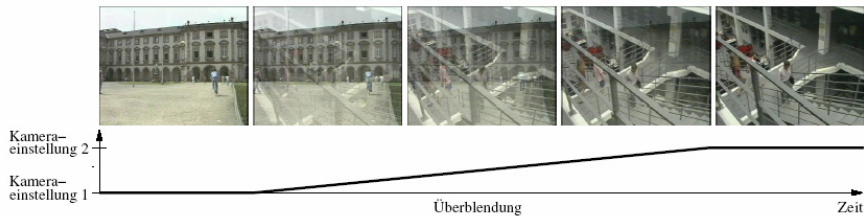
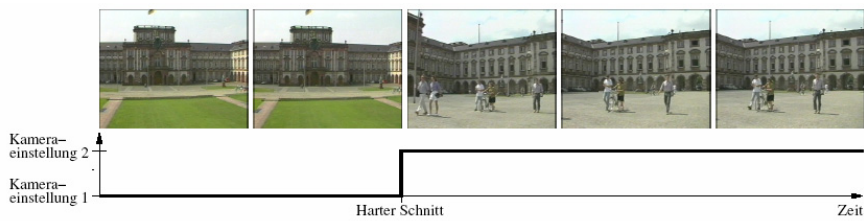
- künstlicher Übergang zwischen den beiden Kameraeinstellungen
- Überblendung (dissolve)*: Der Wechsel von einer Kameraeinstellung zur nächsten erfolgt kontinuierlich.
- Ein- und Ausblendungen (fade)* sind spezielle Überblendungen, bei denen eine der beiden Kameraeinstellungen aus monochromen – häufig 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 (einzeln) Pixel zwischen benachbarten Bildern	Harter Schnitt	Wischeffekt
Kontinuierliche Änderung aller Pixelwerte über einen längeren Zeitraum		Überblendung Ein- oder Ausblendung

Video-Inhaltsanalyse – Stephan Kopf – HS2006 – Kapitel 2: Schnitterkennung

3

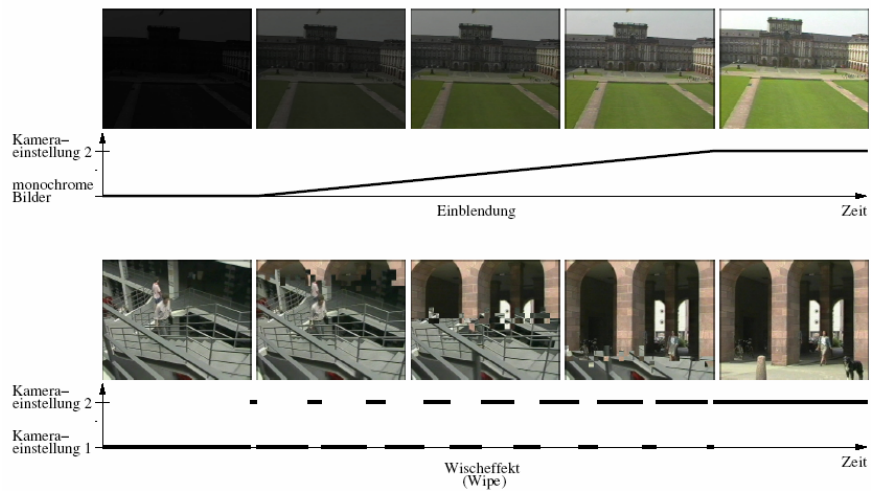
Klassifikation eines Schnittes (II)



Video-Inhaltsanalyse – Stephan Kopf – HS2006 – Kapitel 2: Schnitterkennung

4

Klassifikation eines Schnittes (III)



Video-Inhaltsanalyse – Stephan Kopf – HS2006 – Kapitel 2: Schnitterkennung

5

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

Video-Inhaltsanalyse – Stephan Kopf – HS2006 – Kapitel 2: Schnitterkennung

6

Automatische Schnitterkennung (II)

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: $D_{SAD} > T$

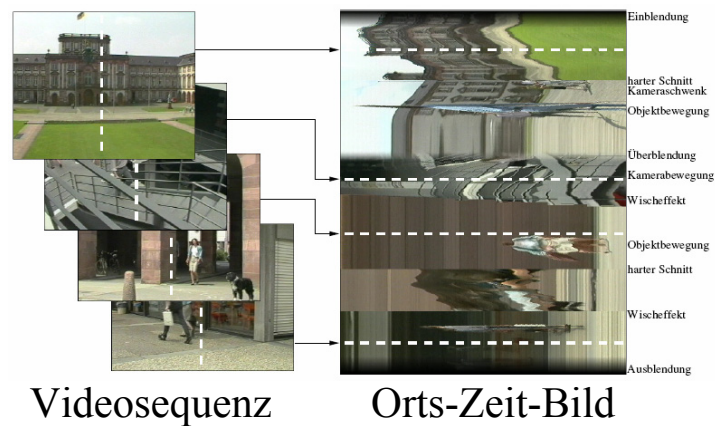
Pixelbasierte Verfahren (II)

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)

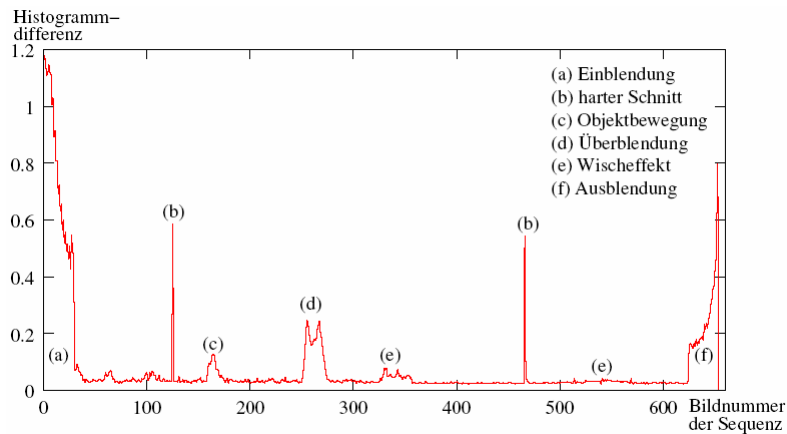
Histogrammdifferenz

- *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

Video-Inhaltsanalyse – Stephan Kopf – HS2006 – Kapitel 2: Schnitterkennung

13

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

Video-Inhaltsanalyse – Stephan Kopf – HS2006 – Kapitel 2: Schnitterkennung

14

Aggregierte Bilddaten (I)

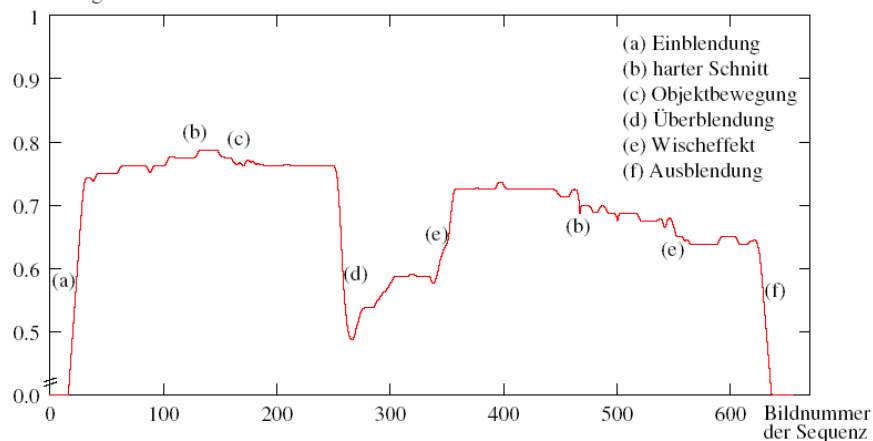
**Standardabweichung I der Helligkeitswerte
aller Pixel eines Bildes:**

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

- 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*)

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

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

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

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.