

4. Segmentierung von Objekten

Video - Inhaltsanalyse

Stephan Kopf

Inhalt

- Vorgehensweise
- Berechnung der Kamerabewegungen zwischen beliebigen Bildern
- Transformation eines Bildes
- Erzeugung von Hintergrundbildern / Panoramabildern
- Segmentierung von Objekten
- Ergebnisse

Vorgehensweise (I)

Ziel

- Objekte des Bildvordergrundes ausschneiden (segmentieren)

Annahmen

- Die Bewegung des Objektes unterscheidet sich von der Bewegung des Bildhintergrundes.
- Es besteht ein deutlicher Helligkeits- oder Farbunterschied zwischen Objekt und Hintergrund.
- Mindestens die Hälfte der Pixel im Bild gehören zum Bildhintergrund.
- Ein Objekt bewegt sich kontinuierlich über den Zeitablauf.

Vorgehensweise (II)

Segmentierungsalgorithmus

1. Der Bildhintergrund aller Bilder einer Kameraeinstellung wird deckungsgleich ausgerichtet.
 - Herleitung der Modellparameter zwischen beliebigen Bildern der Kameraeinstellung
 - Transformation eines Bildes
2. Erzeugung eines Hintergrundbildes
3. Segmentierung eines Objektes durch Vergleich der Bilder der Kameraeinstellung mit dem Hintergrundbild

Kamerabewegung zwischen beliebigen Bildern (I)

Annahme

- Für zwei aufeinander folgende Bilder i und $i+1$ einer Kameraeinstellung sind die Parameter des Kameramodells $T_{i,i+1}$ bekannt.

Gesucht

- Kameramodell zwischen beliebigen Bildern: $T_{i,j}$
- Wird ein Bild i mit $T_{i,j}$ transformiert, so ist Bild i passend zu dem Referenzbild j ausgerichtet.
- Ein Hintergrundbild wird erzeugt, indem jedes Bild i der Kameraeinstellung mit $T_{i,j}$ transformiert wird.

Kamerabewegung zwischen beliebigen Bildern (II)

Berechnung des Kameramodells $T_{i,j}$ (für $i < j$)

- Die Transformationen $T_{i,i+1}$ sind bekannt. Beliebige Koordinaten (x, y) im Bild i werden ausgewählt und mit $T_{i,i+1}$ auf die neue Position (x', y') transformiert.
- (x', y') wird mit $T_{i+1,i+2}$ transformiert und ergibt die Position (x'', y'') im Bild $i + 2$.
- Der Bewegungsvektor von (x, y) nach (x'', y'') entspricht der Verschiebung eines Hintergrundpixels über zwei Bilder.

Kamerabewegung zwischen beliebigen Bildern (III)

Berechnung des Kameramodells $T_{i,j}$ (für $i < j$)

(Fortsetzung)

- Vier unterschiedliche Punkte werden mit $T_{i,i+1}$ und $T_{i+1,i+2}$ transformiert und ergeben vier Bewegungsvektoren.
- Vier dieser Vektoren ermöglichen eine eindeutige Berechnung der acht Parameter des Kameramodells $T_{i,i+2}$.
- Mit dem gleichen Verfahren können für beliebige Bilder i und j ($i < j$) alle Transformationen $T_{i,j}$ berechnet werden.

Kamerabewegung zwischen beliebigen Bildern (IV)

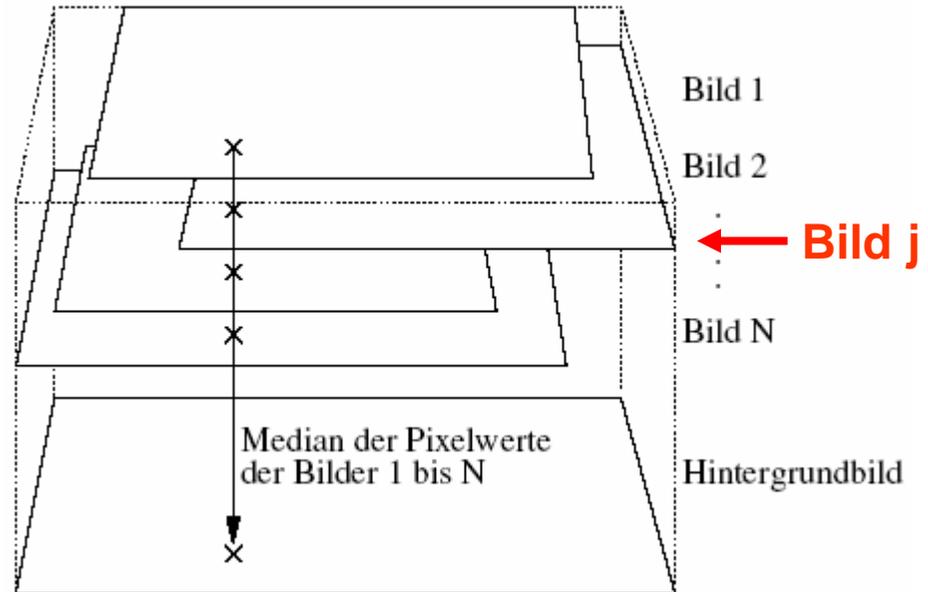
Berechnung des Kameramodells $T_{i,j}$ (für $i > j$)

- Um aus der Transformation $T_{i,j}$ die inverse Transformation $T_{j,i}$ abzuleiten, werden vier Bewegungsvektoren von Bild i nach j bestimmt.
- Die Richtungen der vier Bewegungsvektoren werden umgedreht, d. h. Startpunkte und Endpunkte werden vertauscht.
- Die vier Vektoren definieren durch Lösen des Gleichungssystems die Transformation $T_{j,i}$.

Transformation eines Bildes (I)

Idee

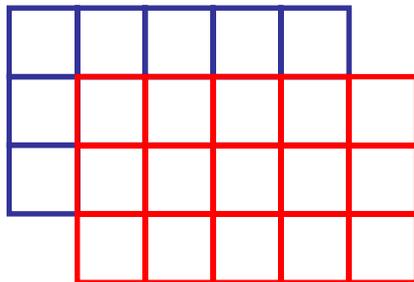
- Ein beliebiges Bild j der Kameraeinstellung wird als **Referenzbild** ausgewählt, an dem alle anderen Bilder ausgerichtet werden sollen.
- Der Bildhintergrund ist deckungsgleich, falls Bild i wird mit T_{ij} transformiert wird.



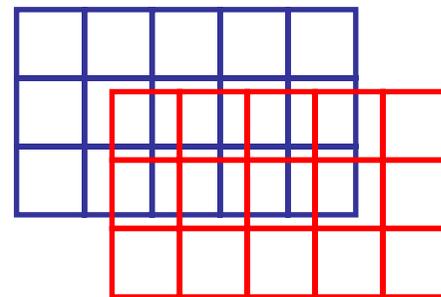
Transformation eines Bildes (II)

Vorgehensweise bei der Erstellung eines Hintergrundbildes

- Jedem Pixel des Hintergrundbildes wird ein Pixelwert aus den transformierten Bildern zugeordnet.
- Die inverse Transformation $T_{j,i}$ liefert – ausgehend von der Pixelposition (x', y') im Hintergrundbild – die Position (x, y) im ursprünglichen Bild.
- Jedem Pixel an der Position (x', y') im transformierten Bild I' wird der Helligkeitswert an der Position (x, y) aus dem ursprünglichen Bild zugewiesen.
- (x, y) entspricht jedoch nicht genau einer ganzzahligen Pixelposition, sondern wird im Allgemeinen zwischen vier Pixeln liegen.



ganzzahlige Verschiebung



Subpixel-Verschiebung

Transformation eines Bildes (III)

Lineare Interpolation

- g_x und g_y : ganzzahliger Anteil von x bzw. von y

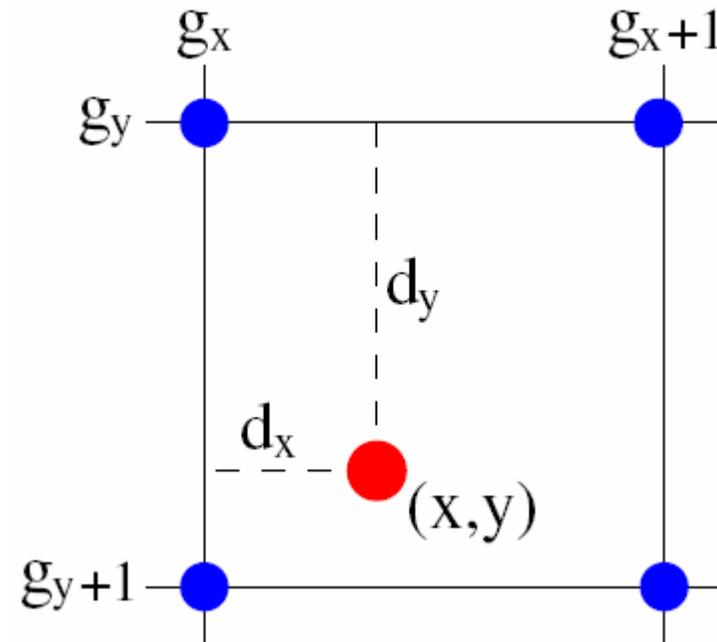
- Rest:

$$d_x := x - g_x$$

$$d_y := y - g_y$$

- Helligkeit I' im transformierten Bild

$$I'(x', y') = [(1-d_x) \cdot I(g_x, g_y) + d_x \cdot I(g_x + 1, g_y)] \cdot (1-d_y) + [(1-d_x) \cdot I(g_x, g_y + 1) + d_x \cdot I(g_x + 1, g_y + 1)] \cdot d_y$$



Transformation eines Bildes (IV)

Merkmale der linearen Interpolation

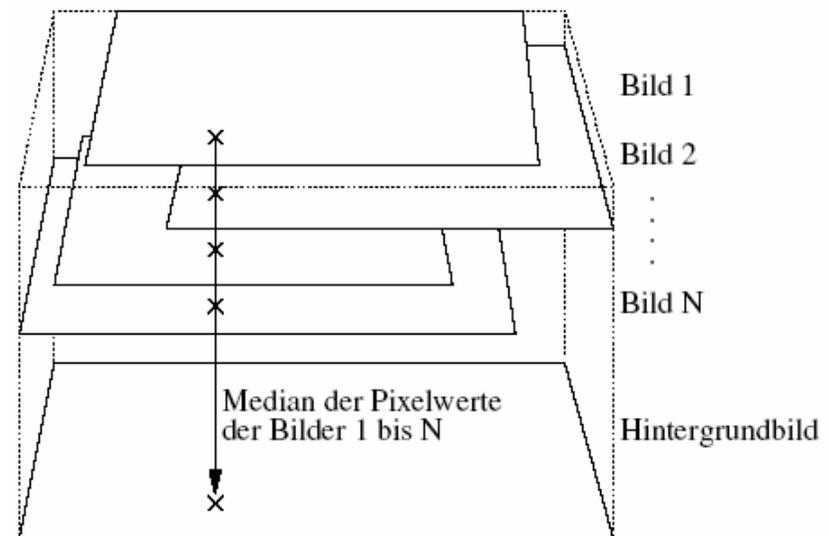
- Je weiter das Pixel (x, y) von einer ganzzahligen Pixelposition entfernt ist, desto größer ist der Einfluss der anderen Pixel auf die Helligkeit des Pixels.
- Durch die hohe Genauigkeit der Ausrichtung der Bilder mit der linearen Interpolation ist eine exakte Segmentierung von Objekten möglich.
- Die Unschärfe nimmt bis zu einer Verschiebung von 0,5 Pixel zu (Glättung des Bildes).
- Für Panoramabilder in hohen Auflösungen sind Unschärfen und fehlerhafte Farben nicht akzeptabel und die Interpolation somit nicht geeignet.

Vermeidung von Unschärfe

- Statt des interpolierten Wertes wird der Helligkeitswert des nächstgelegenen Pixels verwendet.
- Damit ist der Nachteil verbunden, dass lediglich eine Verschiebung um ganzzahlige Werte möglich ist und die Genauigkeit der Transformation abnimmt.

Konstruktion von Hintergrundbildern (I)

- In einem *Hintergrund-* oder *Panoramabild* sollen Vordergrundobjekte nicht oder höchstens einmal enthalten sein.
- Alle N Bilder einer Kameraeinstellung werden transformiert:
- Als Auswahl für jedes Pixel im Hintergrundbild stehen bis zu N Pixel aus den transformierten Bildern zur Verfügung.
- Der **Median** liefert einen guten Wert für die Hintergrundpixel.



Konstruktion von Hintergrundbildern (II)

Berechnung des Medians

- Füge Pixel aller ausgerichteten Bilder an einer Position in Liste ein, sortiere Liste und wähle mittleres Element der Liste als Median aus
→ hoher Rechenaufwand pro Pixel: $O(N \log N)$

Verbesserung

- Erzeuge aus den Pixelwerten an einer Bildposition ein Histogramm.
- Durchlaufe Histogramm aufsteigend und summiere die Anzahl der Pixel.
- Der Median entspricht dem Indexeintrag, bei dem die Summe die Hälfte aller Pixel des Histogramms überschreitet.

Konstruktion von Hintergrundbildern (III)

Annahme

- Mindestens die Hälfte der Pixel beschreibt den Bildhintergrund
 - Der *Median* dieser N Helligkeitswerte ist eine gute Heuristik für ein Hintergrundpixel.

Problem

- Objekt bewegt sich so langsam durch das Bild, dass einzelne Objektpixel mehr als die Hälfte der Zeit an einer Pixelposition verweilen.
 - fehlerhafte Bereiche entstehen in den Hintergrundbildern.

Konstruktion von Hintergrundbildern (IV)

Bsp: Hintergrundbild mittels gewichteter Mittelwerte



Originalvideo



Gewichtung der letzten 10 Bilder



Konstruktion von Hintergrundbildern (V)

Bsp: Hintergrundbild durch Berechnung des Medians



Fehler durch langsame Objektbewegung

Konstruktion von Hintergrundbildern (VI)

Verbesserung der Segmentierung durch Differenzbilder

- Vergleich zweier benachbarter und durch die Transformation entsprechend ausgerichteter Bilder
→ Objektbewegungen führen zu Bildunterschieden in mindestens zwei Regionen
- Schätzung der Position und Größe des Objektes durch den Schwerpunkt (S_x, S_y) der Differenzpixel $D(x,y)$:

$$S_x = \frac{1}{\sum_{x,y} D(x,y)} \sum_{x,y} x \cdot D(x,y) \quad S_y = \frac{1}{\sum_{x,y} D(x,y)} \sum_{x,y} y \cdot D(x,y)$$

- Festlegung eines Rechtecks um den Schwerpunkt (S_x, S_y) abhängig von den Varianzen der Positionen der Differenzpixel.

Konstruktion von Hintergrundbildern (VII)



Segmentierung von Objekten (I)

- Der Medianfilter entfernt Objekte des Bildvordergrunds.
- Die Segmentierung eines Objektes erfolgt durch den Vergleich des transformierten Bildes mit dem Hintergrundbild.
- Die Position und Form der Objektes im ursprünglichen Bild der Kameraeinstellung wird durch die inverse Transformation ermittelt.
- Fehler der Segmentierung im Bereich der Objektgrenzen werden durch die morphologischen Operatoren *Opening* und *Closing* reduziert.
- Das größte Objekt wird ausgewählt.
- Kanten in den Randbereichen eines Objektes werden als tatsächliche Objektgrenzen verwendet.

Exkurs: Morphologische Operatoren

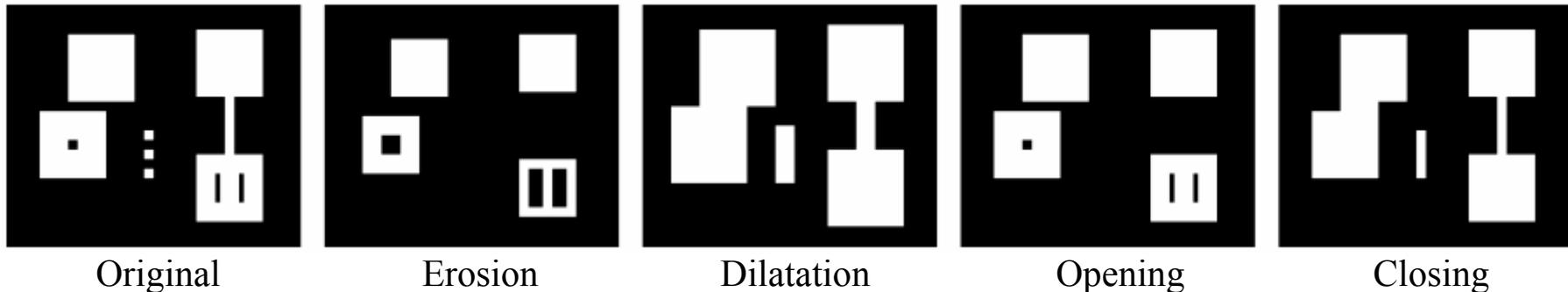
Opening / Closing

$$\textit{Dilatation} : D_B(I(x)) = \max \{I(x+r) \mid r \in B\},$$

$$\textit{Erosion} : E_B(I(x)) = \min \{I(x+r) \mid r \in B\}.$$

$$\textit{Opening} : O_B(x) = D_B [E_B (I(x))],$$

$$\textit{Closing} : C_B(x) = E_B [D_B (I(x))].$$



- Opening: Kleine und schmale Objektregionen werden entfernt.
- Closing: Löcher werden gefüllt und Lücken zwischen benachbarten Regionen geschlossen.

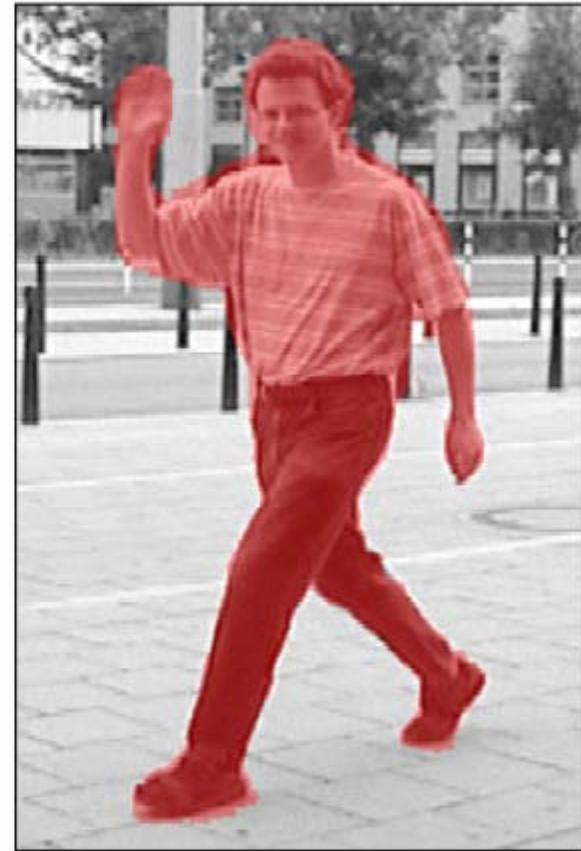
Segmentierung von Objekten (II)



Differenzbild



nach Anwendung
morphologischer Operatoren



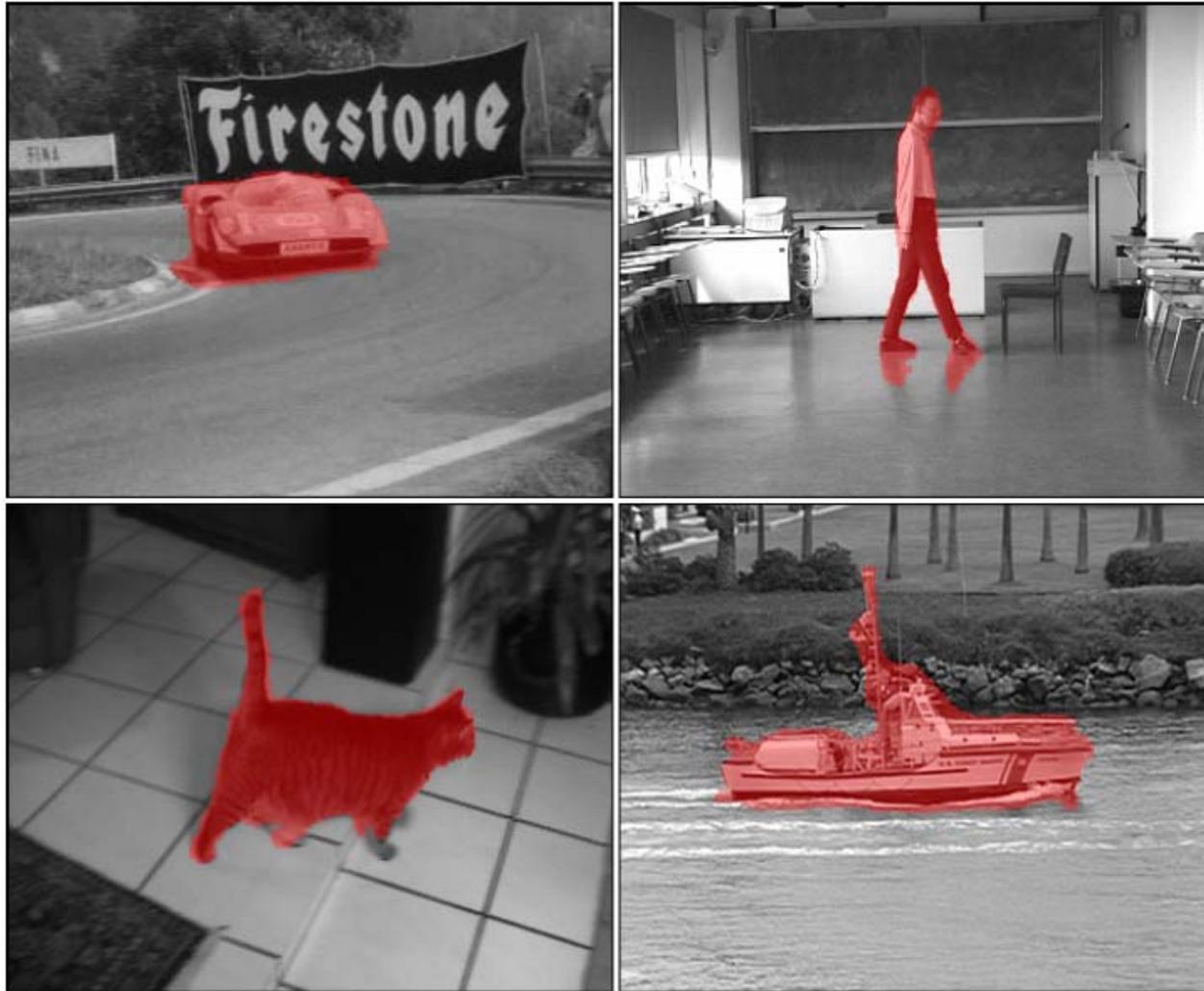
Optimierung der Ränder

Ergebnisse

Ursachen für Fehler

- sehr große Objekte
- geringe Objektbewegung
- Objekt nähert sich der Kamera (großes Objekt und Objekt bleibt an Bildposition)
- keine markanten Strukturen im Bildhintergrund
- Schatten eines Objektes
- Objekt und Hintergrund besitzen identische Helligkeitswerte

Beispiele segmentierter Objekte



Panoramabilder aus Videos



Fragen ?