Zusammenfassung von:

Ismail Haritaoglu, David Harwood, Larry S. Davis.

W^4S:

A Real-Time System for Detecting and Tracking People in $2\frac{1}{2}$D.

University of Maryland

Matthias Rolf

mrolf (at) techfak.uni-bielefeld.de

10. April 2006

In seinem Paper stellt Ismail Haritaoglu das Personenverfolgungssystem W^4S vor, das Personen auch nach Verdeckungen weitertracken soll und skizziert die verwendeten Verfahren.

Im Vordergrund dieser Zusammenfassung stehen die verschiedenen Teilprobleme, die beim Tracking zu lösen sind — und wie W^4S dies tut. Obwohl es sich um ein älteres System handelt (1998), lassen sich die grundlegenden Fragestellungen und Probleme des Trackings gut daran erkennen und diskutieren.

1 Einleitung

W^4S ist ein System zum Tracken von Personen und ihren Körperteilen. Langfristig soll das System so in der Lage sein, Aktionen zwischen Menschen oder von Menschen mit Objekten der Szene zu erkennen und zu beschreiben. Für W^4S wurde ein vorhandenes

Abbildung 1: W^4Serkennt Personen und ihre Extremitäten im Videostream und stellt das Ergebnis in $2\frac{1}{2}$D dar.

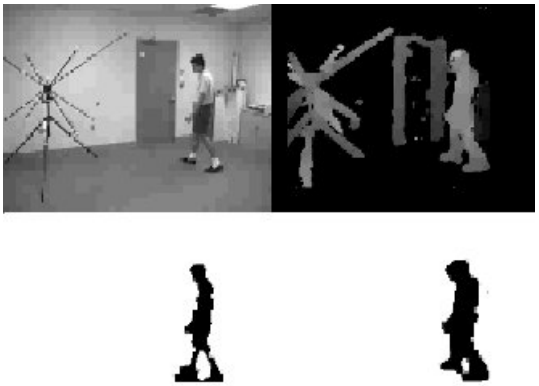

Abbildung 2: Das Ergebnis des Background Subtraction Verfahrens. Links im Grauwert-, rechts im Disparitätsbild.

System zur Personenverfolgung um Stereobildanalyse erweitert, um es robuster gegen einige typische Störeinflüsse zu machen. Aus den gewonnenen Daten kann W^4S eine Darstellung der Szene in $2\frac{1}{2}$D erzeugen, wie in Abbildung 1 zu sehen ist.

2 Detektion von Objekten

Die Aufgabe von W^4S ist das Verfolgen von Personen in einem Videostream. Eine getrackte Person muss also in jedem Frame wiedergefunden werden. Dazu bedient sich W^4S des so genannten *Background Subtraction*[1] Verfahrens. Dabei wird ein Modell des Hintergrundes aufgebaut, also der Bildteile, die sich zuletzt kaum geändert haben. Eine Bildregion zeigt potentiell eine Person, falls sie signifikant von diesem Modell abweicht.

Background Subtraction erfordert ein pixelweises Modell des Hintergrundes. Für jeden Pixel wird sein *Minimalwert* M, sein *Maximalwert* N und die größte absolute *Differenz* D zwischen zwei Frames gespeichert[2]. Das Modell muss regelmäßig aktualisiert werden, um z.B. Beleuchtungsänderungen gerecht werden zu können.

W^4S benutzt zunächst folgende Klassifikationregel: Ein Pixel x wird als Objekt-Pixel betrachtet, falls gilt:

$$|M(x) - I(x)| > D(x) \text{ oder } |N(x) - I(x)| > D(x) \tag{1}$$

Das Ergebnis dieses Prozesses sind zwei Binärbilder (siehe Abb. 2). Da die Bilder i.d.R. noch beträchtliches Rauschen enthalten, werden sie zunächst geglättet. Anschließend werden die Regionen aus dem Intensitäts- und Disparitätsbild kombiniert. Dazu wird im wesentlichen der Schnitt beider Bilder berechnet. Hier stellt sich die Kombination von Intensität und Stereo als vorteilhaft heraus: Das Disparitätsbild ist unempfindlich

[1] Dieser Ansatz wird in [3] näher beleuchtet

[2] Intensitäts- und Disparitätsbild werden dabei unabhägig von einander betrachtet.

Abbildung 3: Bestimmung der Position mit Hilfe der Silhouette.

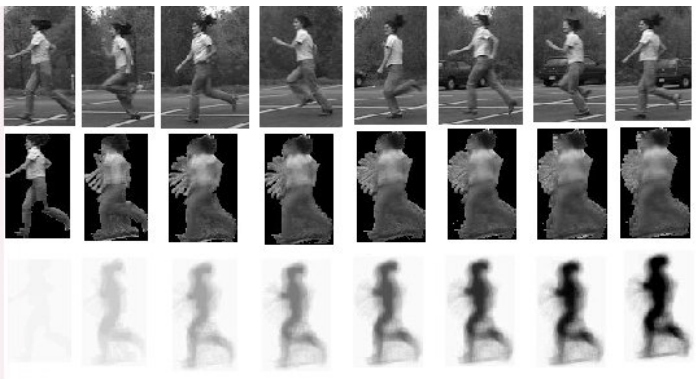

Abbildung 4: Oben: Videostream. Mitte: Texturmodell ψ. Unten: Frequenzbild w.

gegenüber Schatten, die sonst ein großes Problem für Background Subtraction sind. Dafür erzeugt das Intensitätsbild exaktere Konturen.

3 Tracking von Objekten

Der nächste Schritt ist die Zuordnung von bereits getrackten Objekten zu den gefundenen Regionen. W^4S verwaltet zur Positionsschätzung für jedes Objekt ein Bewegungsmodell zweiter Ordnung[3]. Beim Abgleich der Positionen kann es sein, dass mehrere Objekte einer Bildregion zugeordnet werden (oder anders herum).

Der einfachste Fall ist die eindeutige Zuordnung von einem Objekt zu einer Bildregion. Die wichtigste Aufgabe in diesem Fall ist das Aktualisieren des Bewegungsmodells. Eine erste Schätzung liefert der Median der Bildregion, der robuster gegen Störeinflüsse ist als der Mittelwert. Eine genauere Schätzung wird dann mit Hilfe der Objektsilhouette berechnet (siehe Abb. 3). Bestimmt wird die Bildkoordinate, an der die aktuelle Silhouette am besten mit der zuletzt bestimmten korreliert.

Es kommt vor, dass sich zwei oder mehr Personen im Bild verdecken. Nach der Überschneidung ist es wichtig festzustellen, wer wer ist. Um dieses Zuordnungsroblem zu lösen, benutzt W^4S unter anderem ein so genanntes *Temporal Texture Template* (Abb.

[3]Ein solches Modell besteht aus geschätzter Position, Geschwindigkeit und Beschleunigung

4). Gespeichert wird zum einen die durchschnittliche Textur ψ eines Objektes. In einem Frequenzbild w wird die Häufigkeit gespeichert, mit der ein Pixel zu dem Objekt zählte. Es dient als Gewichtung beim Update des Texturmodells:

$$\psi^t(x,y) = \frac{I(x,y) + w^{t-1}(x,y) * \psi^{t-1}(x,y)}{1 + w^{t-1}(x,y)} \tag{2}$$

Die Ausrichtung der Bildregion an das Texturmodell geschieht mit Hilfe des Medians.

4 Tracking von Körperteilen

Nach der Zuordnung der Objekte als Ganzes versucht W^4S die Positionen von Kopf, Oberkörper, Händen und Füßen zu bestimmen. Zur Bestimmung werden vor allem Merkmale der Silhouette sowie die relativen Abstände zwischen ihnen verarbeitet (siehe [4], *Ghost*).

5 Fazit

Bei W^4S handelt es sich, gemessen an den übrigen Seminarthemen um ein recht altes System, das zudem auf stationäre Kameras und Stereodaten angewiesen ist. Ein potentieller Schwachpunkt ist das recht simple Background-Modell, dass in [3] kritisiert wird. Nichtsdestotrotz erlangt das System durch die einbezogenen Stereodaten scheinbar eine recht hohe Robustheit. Erklärtes Langzeitziel ist das Erkennen von Interaktionen zwischen Personen im Bild. Auf der Homepage[4] ist bis heute leider nur zu sehen, dass Personen erkannt werden, die Objekte tragen.

Literatur

[1] Ismail Haritaoglu, David Harwood, Larry S. Davis. W^{4}S: A Real-Time System for Detecting and Tracking People in $2\frac{1}{2}$D. Technical report, University of Maryland.

[2] Ismail Haritaoglu, David Harwood, Larry S. Davis. W^4: Real-Time Surveillance of People and Their Activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), August 2000.

[3] Alan McIvor, Qi Zang, Reinhard Klette. The Background Subtraction Problem for Video Surveillance Systems. Technical report, University of Auckland, 2000.

[4] Ismail Haritaoglu, David Harwood, Larry S. Davis. *Ghost*: A Human Body Part Labeling System Using Silhouettes. Technical report, University of Maryland, 1998.

[4] http://www.umiacs.umd.edu/~hismail/