

# *Pfinder: Real-Time Tracking of the Human Body*

Christopher Wren, Ali Azarbayejani, Trevor Darrell, Alex Pentland

aus: IEEE Transactions on Pattern Analysis and Machine Intelligence (pp. 780-785)

12. April 2006

Seminar „Visuelle Überwachung“

Mareike Lissek

# Inhaltsverzeichnis

<b>1</b>	<b>Abstract</b>	<b>2</b>
<b>2</b>	<b>Einbindung in das Thema des Seminars: Überwachung</b>	<b>2</b>
<b>3</b>	<b>Zusammenfassung des Inhalts</b>	<b>2</b>
3.1	Modellierung der Person . . . . .	3
3.2	Modellierung der Szene . . . . .	3
3.3	Analysis Loop . . . . .	4
3.4	Initialisierung . . . . .	4
<b>4</b>	<b>Diskussion</b>	<b>5</b>
<b>5</b>	<b>Fazit</b>	<b>6</b>
	<b>Literaturverzeichnis</b>	<b>7</b>

# **1 Abstract**

Pfinder (Person Finder) ist ein Echtzeitsystem, welches eine Blob-Darstellung zur Personenverfolgung und Verhaltensinterpretation verwendet.[2]

## **2 Einbindung in das Thema des Seminars: Überwachung**

Das Tracking von Menschen ist die Bewegungsverfolgung und -interpretation. Als Anwendungsgebiete im Bereich der Überwachung sind die Überwachung von öffentlichen Plätzen sowie die Verfolgung von Menschen in z.B. Supermärkten zur Verhaltensanalyse denkbar. In dem Paper wird ein Verfahren vorgestellt, welches als Grundlage zur Erkennung von Menschen in Videobildern dienen kann.

Als praktischer Einsatz wird die Einbindung in das System ALIVE (Artificial Life IVE) angegeben. ALIVE ist eine Ganzkörperinteraktion mit animierten Agenten und benutzt Pfinder, um zu entscheiden wie der Agent reagiert.

## **3 Zusammenfassung des Inhalts**

1977 wurde von Pentland und Kauth die hier (und allgemein) verwendete Blob-Darstellung entwickelt [1].

Im Folgenden werde ich kurz einen Ablauf des Pfinder-Algorithmus darstellen:

### 3.1 Modellierung der Person

Hier wird bereits davon ausgegangen, dass schon eine Szene aufgebaut und ein Mensch detektiert wurde. Die Mittelwerte und Kovarianzmatrizen der 2D-Regionen werden mit  $\mu$  bzw.  $K$  bezeichnet. Die räumliche Blobstatistik wird als Gauß Modell wiedergegeben:

$$Pr(O) = \frac{e^{-\frac{1}{2}(O-\mu)^T K^{-1}(O-\mu)}}{(2\pi)^{\frac{m}{2}} |K|^{\frac{1}{2}}} \quad (1)$$

Außerdem gibt es eine „pixel-by-pixel support map“, die die Zugehörigkeit der einzelnen Pixel zu den Blobs aufzeigt. Für einen Blob  $k$  wäre das:

$$s_k = \begin{cases} 1 & (x,y) \in k \\ 0 & \text{sonst} \end{cases} \quad (2)$$

Jeder Blob besitzt eine Raum-  $(x,y)$  und eine Farbkomponente  $(Y,U,V)$ . Man könnte zur Blobbeschreibung auch noch Bewegungs- und Texturmessungen verwenden, allerdings scheint dies zu rechenintensiv zu sein.

Die „Blobstatistik“ wird rekursiv mit den Informationen aus der aktuellen Statistik und den Vorgaben (priors) erneuert.

### 3.2 Modellierung der Szene

Die Textur der Szene wird als Klasse „Null“ deklariert. Es wird also davon ausgegangen, dass es sich um eine statische Szene handelt wie zum Beispiel ein Büro.  $\mu_0$  ist der Mittelwert  $(Y,U,V)$  eines Punktes der Textur und  $K_0$  die dazugehörige Kovarianzmatrix. Eine der Hauptaufgaben von Pfänder ist festzustellen, welche

Pixel der Textur von einem Menschen verdeckt und welche sichtbar sind. In jedem neuen Frame wird die Statistik der sichtbaren Pixel folgendermaßen erneuert:

$$\mu_t = \alpha y + (1 - \alpha)\mu_{t-1} \quad (3)$$

Dabei ist  $y$  der Vektor  $(x,y,Y,U,V)$ .

### 3.3 Analysis Loop

Der erste Schritt ist das räumliche Modell der Blobs zu schätzen. Dies geschieht mit Hilfe von dynamischen Modellen und dem sogenannten Kalman-Filter.

Daraufhin werden für jedes Pixel die Wahrscheinlichkeiten bestimmt, ob es zu einem Blob  $k$  gehört oder zu der Szene. Dabei wird die Mahalanobis-Distanz verwendet.

Nun wird eine sogenannte „support map“  $s(x,y)$  erstellt, in der ein Pixel der Klasse zugeordnet wird, die die größte Wahrscheinlichkeit für das Pixel aufzeigt bzw. den kleinsten Distanzwert hat. Zum Schluss werden die Blob- und Texturmodelle geupdatet.

Um Fehlern in der Klassifikation vorzubeugen, kann man umgebungsabhängiges Vorwissen nutzen, wie zum Beispiel die Hautfarbe der Hände.

### 3.4 Initialisierung

Um die Szene zu modellieren, wird eine Videosequenz analysiert, die keine Person enthält. Nachdem dies geschehen ist achtet Pfänder auf große Abweichungen zu diesem Modell. Wenn eine genügend große Abweichung bemerkt wurde, wird ein Blob-Modell der Person aufgebaut. Um die Blobs zu initialisieren wird ver-

sucht Hände, Füße und den Kopf der Person zu erkennen, um dort dann neue Blobs aufzubauen.

## 4 Diskussion

Beim ersten Lesen macht das Paper einen guten Eindruck. Versucht man jedoch die angegebenen Formeln zu entziffern, d.h. jeder Variable eine Bedeutung zuzuweisen, stellt man fest, dass es nicht möglich ist, da nicht alles deklariert wird.

So richtig „neu“ ist das Verfahren nicht, eher eine Zusammenstellung von bereits bekannten Algorithmen, die sich alle in den Literaturangaben des Papers wiederfinden lassen.

Die folgende Liste zählt ein paar Dinge auf, die mir negativ aufgefallen sind:

- Im Abstract ist von „Verhaltensinterpretation“ die Rede, jedoch wird im gesamten Paper kein Bezug mehr darauf genommen.
- Viele Variablen werden gar nicht oder erst einige Seiten nach dem ersten Gebrauch erläutert.
- Es gibt Beschränkungen des Systems:
  - Es darf sich nur eine Person im Bild befinden.
  - Bei den Berechnung wird davon ausgegangen, dass es sich um eine relativ statische Szene handelt, d.h. große Veränderungen des Hintergrunds führen zu Problemen.

- Einige Dinge werden sehr ausführlich (z.B. das Gauß-Modell) dargestellt, andere wichtige Dinge eher in einem Nebensatz erwähnt (z.B. der Kalman-Filter)

Positiv anzumerken ist, dass das System sehr robust gegenüber Verdeckungen arbeitet.

## **5 Fazit**

Mein persönliches Fazit unter Berücksichtigung des Seminarthemas ist, dass ich nicht glaube, dass dieses System im Bereich der Überwachung Anwendung findet, da sich die Beschränkungen des Systems nicht mit der Überwachung von z.B. großen Plätzen vereinbaren lässt.

Das Paper an sich verspricht auch in der Einleitung viel mehr als es halten kann. Wenn man es genau durcharbeitet fallen einem viele Kleinigkeiten auf, die das genaue Verständnis erschweren.

## Literatur

- [1] R. J. Kauth, A. P. Pentland, and G. S. Thomas. Blob: An unsupervised clustering approach to spatial preprocessing of mss imagery. In *11th Int'l Symposium on Remote Sensing of the Environment*, April 1977.
- [2] Christopher Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland. Pfinder: Real-time tracking of the human body. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 780–785, July 1997.