Methoden der stochastischen Mustererkennung zur Detektion und Verfolgung von Personen in Videosequenzen

Gerhard Rigoll[*], Harald Breit[†]

In diesem Beitrag geben wir einen Überblick über verschiedene stochastische Methoden zur Mustererkennung, die wir an unserem Lehrstuhl zur Detektion und Verfolgung von Personen in Videosequenzen entwickelt haben. Die hier skizzierten Methoden sind pseudo-zweidimensionale Hidden-Markov-Modelle (P2DHMM), Kalman-Filterung, der Condensation-Algorithmus und Bewegungsdetektion. P2DHMMs sind besonders geeignet für Videosequenzen mit einem sich bewegenden Hintergrund, während typische Anwendungsgebiete für Bewegungsdetektionsverfahren Videosequenzen mit einem stillstehenden Hintergrund sind, wie sie typischerweise von Überwachungskameras erzeugt werden.

1 Einführung

Die Verfolgung von Objekten in beliebigen komplexen Umgebungen ist eines der Hauptprobleme und eine der größten Herausforderungen auf dem Gebiet der visuellen Überwachung. Während der letzten Jahre ist eine Vielfalt verschiedener Verfolgungsalgorithmen entwickelt worden. Eine gute Übersicht wird in [5] gegeben. Wenn man sich diese Vielfalt verschiedener Ansätze anschaut, stellt sich heraus, daß diese sich auch hinsichtlich ihrer Leistungsfähigkeit und geeigneter Anwendungsszenarien deutlich unterscheiden. Die einfachsten Algorithmen sind natürlich solche, die zusätzliche Sensoren wie z. B. Lampen oder spezielle Kleidung verwenden. In diesem Fall reduziert sich

[*] Technische Universität München, Lehrstuhl für Mensch-Maschine-Kommunikation, Arcisstr. 16, 80290 München, e-mail: rigoll@ei.tum.de

[†] Gerhard-Mercator-Universität Duisburg, Fachgebiet Technische Informatik, Bismarckstr. 90, 47057 Duisburg, e-mail: breit@fb9-ti.uni-duisburg.de

Quelle: Tagungsband VdS-Fachtagung "Videodetektion", 2002, Köln

das Verfolgungsproblem im wesentlichen auf das Problem der Verfolgung der Sensorsignale in jedem Einzelbild der Bildsequenz. Es stellt sich heraus, daß die Mehrheit der zur Zeit bekannten Verfolgungsalgorithmen hauptsächlich die Bewegungsinformation in der Bildsequenz auswertet (siehe [17]), entweder aus Differenzbildern hergeleitet oder aus einer Berechnung des optischen Flusses ([8]). Es gibt mehrere Beispiele für die erfolgreiche Verwendung dieser Ansätze (siehe z. B. [2, 7, 9, 20]). Sobald allerdings andere Bewegung neben derjenigen des sich bewegenden Objektes im Bild enthalten ist, kann dieser Ansatz zu ernsthaften Schwierigkeiten führen. Dies kann sehr schnell in realistischen Szenarien der Fall sein, wie z. B. bei der Überwachung des Verkehrs, von Korridoren, Einkaufszentren, Tankstellen oder Parkplätzen. An dieser Stelle muß unterschieden werden zwischen Szenarien mit statischem Hintergrund (siehe z. B. [16]) und solchen mit nichtstatischem Hintergrund, wie sie z. B. durch Kameraoperationen wie Vergrößerungvariation oder Schwenken verursacht werden. Das Vorhandensein eines statischen Hintergrundes ermöglicht den Einsatz einfacher Verfahren wie z. B. Hintergrundsubtraktion und Differenzbilder zur Detektion bewegter Objekte innerhalb des Bildes. In diesem Fall besteht sogar die grundsätzliche Aussicht, mehrere Objekte innerhalb des Bildes durch Lokalisierung der verschiedenen Bewegungsflüsse zu verfolgen, indem man die vorstehend genannten einfachen Verfahren anwendet. Weiterhin können neue Verfolgungsereignisse detektiert werden, die durch zusätzliche sich bewegende Objekte verursacht werden, welche im beobachteten Szenario erscheinen. Die Dinge werden aber komplizierter, falls typische Kameraoperationen wie Vergrößerungsvariation und Schwenken ausgeführt werden, die zu über das gesamte Bild verteilten Bewegungsinformationen führen. In diesem Fall wird es praktisch unmöglich, ein Objekt durch Auswertung der Bewegungsinformation zu verfolgen, weil die Bewegungsinformation fast überall im Bild vorhanden ist.

Ein weiterer wichtiger Punkt bei Überwachungsalgorithmen ist die Frage nach dem Modell, das zur Beschreibung der zu verfolgenden Objekte verwendet wird. Die Form des Objekts ist üblicherweise einer der wichtigsten Ansatzpunkte für diesen Zweck. Es ist aber auch möglich, Algorithmen zu entwickeln, die kein explizites Formmodell verwenden. Das vorangehend erwähnte Beispiel zur ausschließlichen Auswertung der Bewegungsinformation stellt eine Möglichkeit zur Objektverfolgung ohne Formmodell dar. In einem solchen Fall ersetzt die Verwendung von Farbinformationen (beispielsweise Hautfarben) oft die Forminformationen (siehe z. B. [18]). Die Farben werden dann meistens dazu benutzt, einen Segmentierungs- oder einen Blockzuordnungsprozeß durchzuführen, wobei jeder Block derart klassizifiert wird, daß er entweder zum zu verfolgenden Objekt oder zum Hintergrund gehört. Das Hauptproblem eines solchen Ansatzes ist die Tatsache, daß der blockweise arbeitende Zuordnungsprozeß auf Merkmalen basiert, die global im zu untersuchenden Block vorhanden sind. Auf diese Weise wird z. B. untersucht, ob der aktuelle Block einen ausreichend hohen Anteil an Hautfarben oder typische Frequenzen, die auf die Form eines Körpers hinweisen,

enthält. Dies führt dazu, daß der Abgleichprozeß fehlerhaft sein kann, besonders falls irreführende Form- oder Farbmerkmale im Bild vorhanden sind [13]. Alternativ ist es auch möglich, explizite Formmodelle zu verwenden (siehe z. B. [1, 3]), aber die Konstruktion solcher Modelle ist eine ziemlich mühsame Aufgabe, und wegen der großen Flexibilität der menschlichen Körperbewegungen ist es sehr schwierig, ein Formmodell zu erstellen, welches mit all diesen Varianten zurechtkommt.

2 Pseudo-zweidimensionale Hidden-Markov-Modelle

Ein möglicher Ansatz zum Verfolgen von Objekten besteht in dem Versuch, die vorangehend erwähnten Probleme durch die Verwendung eines lernenden Ansatzes zu lösen. Das statistische Modell wird von einem sogenannten pseudo-zweidimensionalen Hidden-Markov-Modell (P2DHMM, siehe [11]) gebildet. Zusätzlich wird dieses P2D-HMM mit einem Kalman-Filter zur Bewegungsprädiktion kombiniert. Wie später noch detaillierter gezeigt werden wird, hat ein solcher Ansatz folgende Vorteile:

- Ähnlich wie das explizite Formmodell ist das statistische Formmodell in der Lage, A-priori-Wissen über die Form des menschlischen Körpers (z. B. die grobe Zerlegung in Kopf, Rumpf, Beine) auszunutzen. Deshalb behält es einige der Vorteile dieses Ansatzes bei, während es dessen Flexibilität und Robustheit erweitern kann.

- Gleichzeitig kann der Vorteil des modellfreien Ansatzes, die problemrelevanten Merkmale automatisch zu lernen, ausgenutzt werden. Somit kombiniert es die Vorteile des modellbasierten und des modellfreien Ansatzes.

- Die Tatsache, daß das System nicht auf Bewegungsinformationen beruht, hat einige Vorteile, z. B. die Fähigkeit, Personen unabhängig davon zu verfolgen, ob sie sich bewegen oder nicht.

- Eine weitere Konsequenz ist der Vorteil, daß die Verfolgung eines Objektes auch beim Vorhandensein von anderen sich im Hintergrund bewegenden Objekten möglich ist.

- Der Ansatz funktioniert sogar bei Kameraoperationen wie Vergrößerungsvariationen oder Schwenks, die Bewegungsinformationen über die gesamte Bildsequenz erzeugen.

- Durch die sogenannte „HMM multi stream“ -Technik ist es möglich, verschiedene Merkmale, wie z. B. Farb- oder Formmerkmale, zu kombinieren und diesen eine unterschiedliche Gewichtung zu geben.

- Durch die vorhergehend genannte Fähigkeit ist es möglich, das System entweder für einen personenunabhängigen Modus (z. B. durch die Betonung allgemeinerer Formmerkmale) oder für einen personenspezifischen Modus (z. B. durch die Betonung der Farbe der Kleidung) auszulegen. Im letztgenannten Fall könnte man eine bestimmte Person in der Gegenwart anderer sich bewegender Menschen verfolgen, z. B. in einer Fußgängerzone oder in einem Einkaufszentrum.

- Aufgrund spezifischer Fähigkeiten von P2DHMMs könnte das System lokale Informationen anstelle globaler Informationen verwerten. Dadurch könnte das System beispielsweise eine Person mit einem roten Hemd und einer blauen Hose in Gegenwart einer anderen Person verfolgen, die ein blaues Hemd und eine rote Hose trägt. Dies wäre bei Systemen, die globale Farbinformationen wie z. B. Farbhistogramme auswerten, nicht möglich.

3 Kurze Darstellung eines P2DHMM-basierten Verfolgungsalgorithmus

Das P2DHMM erzeugt einen Meßvektor, der als Eingabe für das Kalman-Filter verwendet wird. Die Komponenten dieses Vektors sind der Schwerpunkt der detektierten Person innerhalb des Bildes und die Breite und Höhe des einschließenden Rechtecks (bounding box). Zu diesem Zweck werden die folgenden Schritte ausgeführt: Zuerst wird das Bild mit einer DCT-basierten Merkmalsextraktionsmethode bearbeitet, die von einem Gesichtserkennungssystem [4] übernommen wurde. Das Bild wird mit einem Abtastfenster von oben nach unten und von links nach rechts abgetastet. Eine dreieckförmige Maske extrahiert aus jedem Abtastfenster die ersten 10 DCT-Koeffizienten, die in einem Vektor angeordnet werden. Eine Überlappung zwischen benachbarten Abtastfenstern verbessert die Fähigkeiten des HMM zur Modellierung der Nachbarbeziehungen zwischen den Fenstern. Das Ergebnis dieser Merkmalsextraktion ist ein zweidimensionales Feld von Vektoren mit der Dimension 10. Dieses Feld wird dem P2DHMM präsentiert, wie in Abb. 1 dargestellt.

Ein solches P2DHMM kann als zweidimensionales stochastisches Modell für ein Objekt innerhalb eines Bildes aufgefaßt werden, das das Auftreten einer Merkmalsvektorsequenz modelliert, welche von dem Objekt abgeleitet werden kann, falls dieses wie oben beschrieben vorverarbeitet wird (siehe [11]). Die Parameter des P2DHMMs bestehen aus den Übergangs- und den Ausgabewahrscheinlichkeiten der einzelnen HMM-Zustände und können zur Modellierung verschiedener Objekte trainiert werden. Das Training auf Personenformen kann folgendermaßen erreicht werden: Bilder, die Personen enthalten und geeignet vorverarbeitet wurden, werden dem P2DHMM präsentiert, um die Struktur eines menschlichen Körpers zu modellieren, indem man Parameterschätzmethoden wie z. B. den Vorwärts-Rückwärts-Algorithmus auf das P2DHMM

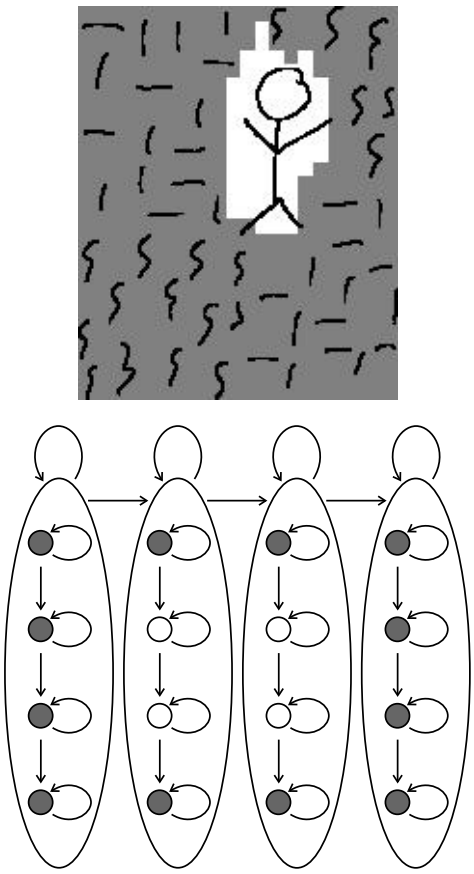

Abbildung 1: P2DHMM als stochastisches Modell für ein zweidimensionales Objekt.

anwendet (siehe z. B. [14]). Weil das P2DHMM als elastisches Modell angesehen werden kann (siehe auch [19] wegen anderer Ansätze für deformierbare Modelle), besitzt es die Fähigkeit, den menschlichen Körper in verschiedenen Positionen zu modellieren. Dies wird in Abb. 1 durch eine Handskizze oberhalb des P2DHMM illustriert, die eine Person innerhalb eines komplexen Hintergrundes zeigt. Das Szenario der Modellierung von Handskizzen wurde intensiv in [12] untersucht.

Bei der Modellierung einer solchen Skizze durch ein P2DHMM wird jeder vertikale Streifen des oben gezeigten Bildes einem der Metazustände des P2DHMM zugeordnet, die durch die verikalen Ellipsen in Abb. 1 symbolisiert werden. Zusätzlich werden die Blöcke in jedem Streifen den Zuständen innerhalb der Metazustände in vertikaler Richtung zugeordnet. Somit kann das P2DHMM das Bild nichtlinear zweidimensional verzerren.

In Abb. 1 wird angenommen, daß alle hell markierten Zustände Körperteilen zugeordnet werden, während die dunkel markierten Zustände dem Bildhintergrund zugeordnet werden. In [15] wird beschrieben, wie der Viterbi-Algorithmus (siehe [14]) dazu benutzt werden kann, zu einem gegebenen Bild die Zuordnung von Bildblöcken zu Zuständen zu berechnen und somit eine automatische Segmentierung des Bildes in Körper und Hintergrund zu erhalten. Diese integrierte Segmentierungsprozedur wird dadurch illustriert, daß die den weißen Zuständen des P2DHMM zugeordneten Berei-

che in Abb. 1 weiß dargestellt wurden.

Die eigentliche Verfolgungsprozedur wird dadurch realisiert, daß jedes Einzelbild der Videosequenz dem P2DHMM präsentiert wird. Dann wird mit dem Viterbi-Algorithmus die Zuordung der Bildblöcke zu den Zuständen des P2DHMM berechnet. Von den Bildblöcken, die den Personenzuständen zugeordnet wurden, wird der Schwerpunkt, dargestellt als x_s und y_s, berechnet sowie ein einschließendes Rechteck mit der Breite w und der Höhe h. Diese Daten dienen als Eingabemessung für das Kalman-Filter.

Um die sich bewegende Person zu beschreiben und um das Ergebnis der Verfolgungsprozedur darzustellen, wird der Zustandsvektor $\mathbf{x}$ eingeführt, der aus den folgenden Komponenten besteht:

$$\mathbf{x} = \begin{bmatrix} x_s\text{: x-Koordinate des Schwerpunktes} \\ y_s\text{: y-Koordinate des Schwerpunktes} \\ v_x\text{: horizontale Geschwindigkeit} \\ v_y\text{: vertikale Geschwindigkeit} \\ w\text{: Breite des einschließenden Rechtecks} \\ h\text{: Höhe des einschließenden Rechtecks} \end{bmatrix} \tag{1}$$

Die Bewegung der Person wird durch ein einfaches dynamisches Modell beschrieben, welches davon ausgeht, daß sich die Person zwischen den Abtastpunkten k und $k+1$ mit konstanter Geschwindigkeit bewegt. In diesem Fall kann das dynamische Verhalten der Person folgendermaßen beschrieben werden:

$$\mathbf{x}_{k+1} = \mathbf{A} \cdot \mathbf{x}_k \tag{2}$$

mit

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Nur ein Teil der Größen des Zustandsvektors kann direkt gemessen werden, sodaß sich die folgende Gleichung für den Meßvektor ergibt:

$$\mathbf{y}_k = \mathbf{H} \cdot \mathbf{x}_k \tag{4}$$

Das Kalman-Filter rekonstruiert den Zustandsvektor $\mathbf{x}$ aus dem Meßvektor $\mathbf{y}$ gemäß

folgender Gleichungen (siehe z. B. [6, 16]):

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}(k) \cdot \left[\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_k^-\right] \quad (5)$$

$$\mathbf{K}(k) = \frac{\mathbf{P}^-(k)\,\mathbf{H}^T}{\mathbf{H}\,\mathbf{P}^-(k)\mathbf{H}^T + \mathbf{R}(k)} \quad (6)$$

$$\mathbf{P}^+(k) = [\mathbf{I} - \mathbf{K}(k) \cdot \mathbf{H}] \cdot \mathbf{P}^-(k) \quad (7)$$

$$\hat{\mathbf{x}}_{k+1}^- = \mathbf{A} \cdot \hat{\mathbf{x}}_k^+ \quad (8)$$

$$\mathbf{P}^-(k+1) = \mathbf{A} \cdot \mathbf{P}^+(k) \cdot \mathbf{A}^T + \mathbf{Q}(k) \quad (9)$$

Auf diese Weise wird die Verstärkungsmatrix $\mathbf{K}(k)$ für jeden Zeitschritt k gemäß den gegebenen Gleichungen des Kalman-Filters aktualisiert. Der Meßvektor $\mathbf{y}$ ist in diesem Fall

$$\mathbf{y} = [x_s, y_s, w, h]^T \quad (10)$$

woraus sich folgende Meßmatrix ergibt:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

Aus den im Meßvektor $\mathbf{y}$ enthaltenen Daten, die vom P2DHMM geliefert werden, schätzt das Kalman-Filter den Zustandsvektor $\mathbf{x}$ und prädiziert auf diese Weise die Informationen über das einschließende Rechteck, die in den letzten beiden Komponenten von $\mathbf{x}$ enthalten sind. Die dritte und vierte Komponente von $\mathbf{x}$ liefern die Geschwindigkeit der Person und dienen hauptsächlich der Unterstützung des mathematischen Modells der Bewegung der Person und der Stabilität des Systems.

Somit schätzt das Kalman-Filter den neuen Zustandsvektor $\hat{\mathbf{x}}_k^+$ durch Vereinigung der Daten des Systemmodells und der Messung mit einer variablen Verstärkung. Es arbeitet mit dem P2DHMM folgendermaßen zusammen: Das vom Kalman-Filter geschätzte einschließende Rechteck wird mit dem Faktor 1,3 und einem Bias von 20 Bildpunkten vergrößert. Die Suche des P2DHMMs wird auf diesen eingeschränkten Bereich beschränkt. Diese Limitierung des Suchbereichs beschleunigt die Segmentierungsprozedur erheblich und liefert eine gute Segmentierungsfläche, auch wenn die Person ihre Form während ihrer Bewegung ändert (z. B. indem sie ihre Arme schwingt).

4 Der Condensation-Algorithmus

Der Condensation-Algorithmus soll hier nur kurz beschrieben werden. Eine detailliertere Beschreibung ist in [10] zu finden. Der Name ist eine künstliche Konstruktion und

stammt von den englischen Wörtern „conditional density propagation“ , die einen wesentlichen Aspekt dieses Algorithmusses beschreiben, nämlich die Fortpflanzung bedingter Wahrscheinlichkeitsdichten. Der Algorithmus stelt eine Verallgemeinerung des Kalman-Filters für nicht-gaußsche Verteilungen dar und wird auch als Partikelfilter bezeichnet.

Dieser Algorithmus dient zur Beschreibung der zeitlichen Fortpflanzung von bedingten Dichten. Dies wird z. B. auch vom Kalman-Filter gemacht, aber der Condensation-Algorithmus hat den Vorteil, daß er aus mathematischer Sicht einfacher ist und daher eine recht einfache Kombination mehrerer Meßverfahren erlaubt, wie später gezeigt werden wird.

Die zeitliche Fortpflanzung der bedingten Wahrscheinlichkeitsdichten kann in die nachstehenden zeitlich aufeinaderfolgenden Schritte zerlegt werden:

1. Deterministische Drift
   Es wird angenommen, daß das zu verfolgende Objekt sich gemäß einem dynamischen Modell bewegt, sodaß der der neue Zustand in erster Näherung aus dem alten Zustand berechnet werden kann. Dies wird als Prädiktion bezeichnet.

2. Stochastische Diffusion
   Das dynamische Modell ist üblicherweise nur eine Näherung, sodaß eine Abweichung zwischen dem neuen Zustand, den das Modell vorhergesagt hat, und dem tatsächlichen Zustand bestehen wird. Da diese Differenz nicht deterministisch beschrieben werden kann, wird sie als stochastischer Prozeß beschrieben.

3. Reaktiver Effekt der Messung
   Um die im vorhergehenden Schritt erhaltene Fortpflanzung der Dichte mit der Realität in bessere Übereinstimmung zu bringen, wird eine Messung durchgeführt. Diese Messung ist üblicherweise nicht exakt und muß in das System eingebracht werden. Das Ergebnis ist nun eine bedingte Wahrscheinlichkeitsdichte des neuen Zustands.

Für eine mathematische Formulierung des Algorithmusses müssen zunächst einige Begriffe definiert werden: Wir benutzen die diskrete Zeit k, definiert durch $t_k = t_0 + k \cdot T_s$, wobei t_0 die Anfangszeit darstellt und T_s das Abtastintervall. Der Zustand des modellierten Objektes zur diskreten Zeit k wird als Zustandsvektor $\mathbf{x}_k = \mathbf{x}(t_k)$ geschrieben und seine Geschichte als $\mathbf{X}_k = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$. Auf analoge Weise wird eine Menge von Bildmerkmalen in einem Meß- oder Beobachtungsvektor $\mathbf{z}_k$ zusammengefaßt, dessen Geschichte als $\mathbf{Z}_k = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\}$ geschrieben wird.

Unter Verwendung dieser Symbole und dem Satz von Bayes kann das Verfolgungsproblem folgendermaßen formuliert werden:

$$p(\mathbf{x}_k|\mathbf{Z}_k) \propto p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Z}_{k-1}) \tag{12}$$

Der Condensation-Algorithmus verwendet eine Menge von Proben des Zustandsvektors, um dessen bedingte Wahrscheinlichkeitsdichtefunktion $p(\mathbf{x}_k|\mathbf{Z}_k)$ zu approximieren. Die Probenmenge besteht aus N Proben $\mathbf{s}_k^{(n)}, n = 1 \ldots N$, von denen jede mit der Wahrscheinlichkeit $\pi_k^{(n)}$ gewichtet wird. Die Wahrscheinlichkeiten werden aus einer Messung gewonnen und addieren sich zu 1 auf:

$$\pi_k^{(n)} = \frac{p(\mathbf{z}_k|\mathbf{x}_k = \mathbf{s}_k^{(n)})}{\sum_{j=1}^{N} p(\mathbf{z}_k|\mathbf{x}_k = \mathbf{s}_k^{(j)})} \tag{13}$$

Nun kann die bedingte Zustandsdichte $p(\mathbf{x}_k|\mathbf{Z}_k)$ durch die gewichtete Probenmenge $(\mathbf{s}_k^{(n)}, \pi_k^{(n)}, n = 1 \ldots N)$ repräsentiert werden, die aus der vorherigen Probenmenge folgendermaßen ermittelt werden kann:

Konstruiere aus der alten Probenmenge $\{\mathbf{s}_{k-1}^{(n)}, \pi_{k-1}^{(n)}, c_{k-1}(n), n = 1 \ldots N\}$ zum Zeitschritt $k-1$ eine neue Probenmenge $\{\mathbf{s}_k^{(n)}, \pi_k^{(n)}, c_k(n), n = 1 \ldots N\}$ zum Zeitschritt k wie folgt:

1. Wähle eine Probe $\mathbf{s}_k^{\prime(n)}$ wie folgt:

   a) Erzeuge eine gleichverteilte Zufallszahl $r \in [0, 1]$.

   b) Finde (z. B. durch binäre Unterteilung) das kleinste j, für das $c_{k-1}^{(j)} \geq r$.

   c) Setze $\mathbf{s}_k^{\prime(n)} = \mathbf{s}_{k-1}^{(j)}$.

2. Prädiziere durch Abtastung von

   $$p(\mathbf{x}_k|\mathbf{x}_{k-1} = \mathbf{s}_k^{\prime(n)}) \tag{14}$$

   um jedes $\mathbf{s}_k^{\prime(n)}$ zu wählen. Zum Beispiel in dem Fall, daß die Dynamik des Systems durch eine lineare stochastische Differenzengleichung beschrieben wird, kann die neue Probe gemäß der Gleichung

   $$\mathbf{s}_k^{(n)} = \mathbf{A}\mathbf{s}_k^{\prime(n)} + \mathbf{w}_k^{(n)}, \tag{15}$$

   erzeugt werden, wobei $\mathbf{A}$ die Systemmatrix darstellt, die die Objektdynamik beschreibt, und $\mathbf{w}_k^{(n)}$ ein stochastischer Prozeß ist, der das Prozeßrauschen beschreibt.

3. Messe und gewichte die neue Position unter Verwendung der gemessenen Merkmale $\mathbf{z}_k$:

   $$\pi_k^{(n)} = p(\mathbf{z}_k|\mathbf{x}_t = \mathbf{s}_k^{(n)}), \tag{16}$$

   dann normalisiere so, daß $\sum_n \pi_k^{(n)} = 1$ und speichere diese Werte zusammen mit der kumulativen Wahrscheinlichkeit als $(\mathbf{s}_k^{(n)}, \pi_k^{(n)}, c_k^{(n)})$,

wobei

$$c_k^{(0)} = 0, \tag{17}$$

$$c_k^{(n)} = c_k^{(n-1)} + \pi_k^{(n)} \quad (n = 1 \ldots N). \tag{18}$$

Nun besteht das Verfolgungsergebnis zum diskreten Zeitschritt k aus den aktualisierten Zustandsvektorproben $\mathbf{s}_k$ ($k = 1 \ldots N$) und deren aktualisierten Auftretenswahrscheinlichkeiten π_k ($k = 1 \ldots N$). Falls gewünscht, kann dieses Ergebnis weiter bearbeitet werden, indem man Momente der berechneten Position zur diskreten Zeit k gemäß

$$\mathrm{E}\left\{f(\mathbf{x}_k)\right\} = \sum_{n=1}^{N} \pi_k^{(n)} f\left(\mathbf{s}_k^{(n)}\right), \tag{19}$$

berechnet und beispielsweise eine mittlere Position (den Erwartungswert) erhält, wenn man $f(\mathbf{x}) = \mathbf{x}$ setzt.

4.1 Berechnung der bedingten Wahrscheinlichkeiten

Die bedingten Wahrscheinlichkeiten in (13) müssen aus einer Messung innerhalb des aktuellen Bildes gewonnen werden. Die bedingten Wahrscheinlichkeiten stellen die entscheidenden Stelle dar, an der die verschiedenen Modi in einem wahrscheinlichkeitstheoretischen Rahmen innerhalb des Condensation-Algorithmusses zusammengefügt werden können. Daher sind sie für unsere Zwecke von besonderem Interesse. Im folgenden soll gezeigt werden, wie zwei Methoden zur Ermittlung dieser Meßdaten verwendet werden können, nämlich ein P2DHMM und ein Bewegungsdetektor.

Die Aufgabe besteht nun darin, einen Meßvektor, der von einem Meßmodus geliefert wird und beispielsweise den Schwerpunkt einer Person enthält, in einer solchen Weise auszuwerten, daß wir die bedingte Wahrscheinlichkeit dieses Meßvektors $\mathbf{z}_k$ unter der Bedingung einer angenommenen Probe für den Zustandsvektor $\mathbf{x}_k$, nämlich $p(\mathbf{z}_k|\mathbf{x}_k)$, berechnen können. Es ist möglich, den Zusammenhang zwischen $\mathbf{z}_k$ und $\mathbf{x}_k$ in Form einer Meßgleichung zu schreiben, wie sie auch beim Kalman-Filter verwendet wird:

$$\mathbf{z}_k = \mathbf{H} \cdot \mathbf{x}_k + \mathbf{v}_k. \tag{20}$$

Falls $\mathbf{v}_k$ weißes Rauschen beschreibt, ist es eine vernünftige Annahme, daß die Variable $\mathbf{z}_k$ einen stochastischen Prozeß darstellt, der durch eine Gaußsche Verteilung charakterisiert werden kann, wobei $\mathbf{H}\mathbf{x}_k$ als Mittelwert des Prozesses betrachtet werden kann. In diesem Fall kann die eben genannte Gaußsche Verteilung interpretiert werden als die Wahrscheinlichkeit des Meßvektors $\mathbf{z}_k$ unter der Annahme, daß die Probe $\mathbf{x}_k$ der wahre Zustandsvektor ist, woraus folgt:

$$p(\mathbf{z}_k|\mathbf{x}_k = \mathbf{s}_k^{(n)}) = \exp(-\frac{1}{2}(\mathbf{z}_k - \mathbf{H}\mathbf{x}_k)^T \mathbf{C}(\mathbf{z}_k - \mathbf{H}\mathbf{x}_k)) \tag{21}$$

In dieser Funktion bezeichnet $\mathbf{C}$ die Kovarianzmatrix, die geeignet gewählt werden muß. Die resultierenden Wahrscheinlichkeitswerte werden anschließend normalisiert, sodaß ihre Summe 1 ergibt.

Der Zustandsvektor $\mathbf{x}$ (und jeder Probenvektor $\mathbf{s}$) besteht aus den Komponenten

$$\mathbf{x} = [x_c, y_c, v_x, v_y, w, h]^T , \tag{22}$$

wobei x_c und y_c den Mittelpunkt des einschließenden Rechtecks mit der Breite w, der Höhe h und den Geschwindigkeitskomponenten v_x und v_y bezeichnen.

Die Funktionalität dieses Ansatzes kann man sich anhand der folgenden Überlegung plausibel machen: Falls der Meßvektor $\mathbf{z}_k$ fast mit $\mathbf{H}\mathbf{x}_k$ identisch ist, müssen Messung und Probe sehr nahe beieinanderliegen (d. h. $\mathbf{z}_k$ bestätigt $\mathbf{x}_k$ sehr gut) und genau dann wird (21) eine hohe Wahrscheinlichkeit für diese Probe liefern. Daher ist dies eine geeignete Gleichung für die wahrscheinlichkeitstheoretische Interpretation der Messung $\mathbf{z}_k$.

4.1.1 P2DHMM

Eine Methode für die Gewinnung von Meßdaten ist ein P2DHMM, wie in Abschnitt 2 beschrieben.

4.1.2 Bewegungsdetektor

Eine weitere Methode für die Gewinnung von Meßdaten ist die Verwendung eines Bewegungsdetektors. Hier soll gezeigt werden, wie eine weitere Messung in das Condensation-basierte Verfolgungssystem integriert werden kann und somit das Ergebnis verbessern kann. Der Bewegungsdetektor basiert auf einer Berechnung der Differenzen d zwischen Bildpunkten $\mathbf{i}(x, y)$ im aktuellen Bild und korrespondierenden Bildpunkten in einem Referenzbild gemäß

$$d_k(x, y) = \|\mathbf{i}_k(x, y) - \mathbf{i}_{\mathrm{ref}}(x, y)\| \tag{23}$$

und einer nachfolgenden Schwellwertbildung. Für die Bildpunkte, deren Differenz die Schwelle überschreitet, wird ein einschließendes Rechteck berechnet, dessen Parameter (Mittelpunkt, Breite, Höhe) zu einem Bewegungsmeßvektor mit folgenden Komponenten zusammmengefaßt werden:

$$\mathbf{z}_{\mathrm{m}} = [x_{\mathrm{cobb,m}}, y_{\mathrm{cobb,m}}, w_{\mathrm{bb,m}}, h_{\mathrm{bb,m}}]^T . \tag{24}$$

Die zugehörige Meßmatrix hat die Form

$$\mathbf{H}_{\mathrm{m}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{25}$$

Ein solcher Bewegungsdetektor ist nur wirklich sinnvoll für Szenarien mit einem konstanten Hintergrund. Falls die Kamera schwenken würde, enthielte das gesamte Bild Bewegungsinformationen, und folglich könnte keine bestimmte Bewegung detektiert werden.

4.2 Kombination mehrerer Modi

Ein interessanter Aspekt des Condensation-Algorithmusses ist die Möglichkeit, die Daten von mehreren Messungen zu integrieren. Eine solche Kombination kann es ermöglichen, die Nachteile einer einzelnen Methode zu überwinden und die Stärken mehrerer Methoden zu vereinigen.

Die Stelle, an der wir die Messungen im Condensation-Algorithmus vereinigt haben, ist die Berechnung der Wichtungen $\pi_k^{(n)}$ für die Probenvektoren $\mathbf{s}_k^{(n)}$. Wenn man beispielsweise aus (21) unter Verwendung von unterschiedlichen Meßvektoren und geeigneten Meßmatrizen zwei (normalisierte) Meßwahrscheinlichkeiten gewonnen hat, wird das resultierende Gewicht der Probe durch Multiplikation der einzelnen Meßwahrscheinlichkeiten gemäß der Gleichung

$$p(\mathbf{z}_1, \mathbf{z}_2|\mathbf{s}_k^{(n)}) = p(\mathbf{z}_1|\mathbf{s}_k^{(n)}) \cdot p(\mathbf{z}_2|\mathbf{s}_k^{(n)}) \tag{26}$$

und einer nachfolgenden Normalisierung berechnet. Durch diese Multiplikation wird die höchste Wahrscheinlichkeit erzielt, wenn beide Messungen eine hohe Wahrscheinlichkeit liefern, während die resultierende Wahrscheinlichkeit kleiner sein wird, wenn nur eine Messung eine hohe Wahrscheinlichkeit liefert, und schließlich sehr klein, wenn beide Messungen eine niedrige Wahrscheinlichkeit liefern. Nach der Multiplikation werden die erhaltenen Wahrscheinlichkeiten wieder normalisiert, damit ihre Summe 1 ergibt. Diese modifizierten Probengewichtungen haben einen günstigen Einfluß auf das Verfolgungsergebnis, welches nun das Ergebnis einer multimodalen Vereinigung von verschiedenen Informationskanälen ist, die den Verfolgungsprozeß unterstützen.

5 Beispielbilder aus Verfolgungsszenen

Einige interessante Ergebnisse unserer Verfolgungssysteme sind in Abb. 2 und Abb. 3 dargestellt. In Abb. 2 ist eine Szene gezeigt, die von einem System bearbeitet wurde, das auf einem P2DHMM und einem Kalman-Filter beruht. Die Sequenz zeigt einen Fußgänger, der eine Straße überquert. Während er geht, schwenkt die Kamera, und im Hintergrund fahren Fahrzeuge vorbei, sodaß sehr viel Bewegung im gesamten Bild enthalten ist und folglich ein Bewegungsdetektor nicht sehr hilfreich wäre, während dies kein Problem für das P2DHMM darstellt. Das weiße Rechteck zeigt das Ergebnis des Verfolgungssystems. Es ist zu sehen, daß das Verfolgungssystem den Kontakt zur Person während der gesamten Sequenz aufrechterhält, sogar wenn die Person zeitweilig von einem Mast verdeckt wird.

In Abb. 3 ist ein typisches Außenüberwachungsszenario mit einem stillstehenden Hintergrund dargestellt (Daten von der PETS 2001). In dieser Szene wird die Person, die verfolgt werden soll, teilweise von einem vorbeifahrenden Auto verdeckt. Für diese Sequenz verwendeten wir eine Kombination aus einem P2DHMM und einem Bewegungsdetektor, die von einem Condensation-Algorithmus zusammengefügt wurden. In der oberen Zeile ist ein Fall zu sehen, wo das System mit dem P2DHMM alleine den Kontakt zur Person nach einer Weile verliert (siehe das dritte Bild in dieser Zeile), während in der unteren Zeile zu sehen ist, daß nach Integration des Bewegungsdetektors der Kontakt zur Person erhalten bleibt. Im letzten Bild in der oberen Zeile ist das Ergebnis des Bewegungsdetektors mit den erkannten Bewegungsbereichen und dem zugehörigen einschließenden Rechteck detailliert und vergrößert dargestellt. In dieser Szene würde die Verwendung des Bewegungsdetektors alleine versagen, weil andere bewegte Objekte (man beachte das vorbeifahrende Auto im zweiten Bild) die Messung sehr stark stören

6 Zusammenfassung und Ausblick

In diesem Beitrag haben wir verschiedene stochastische Ansätze vorgestellt, die wir zur Verfolgung von Personen in Videosequenzen verwendet haben. Die Ansätze wurden beschrieben und einige positive Beispiele wurden gezeigt. Trotzdem gibt es viele Szenarien, bei denen die hier vorgestellten Systeme versagen. Deshalb forschen wir weiter an verbesserten Systemen, mit dem Ziel, Systeme zu entwickeln, die ebensogut wie ein Mensch oder sogar noch besser Objekte verfolgen können.

Abbildung 2: Beispiel für eine Personenverfolgung mit vorbeifahrenden Fahrzeugen, Verdeckung durch einen Mast und Kameraschwenks.

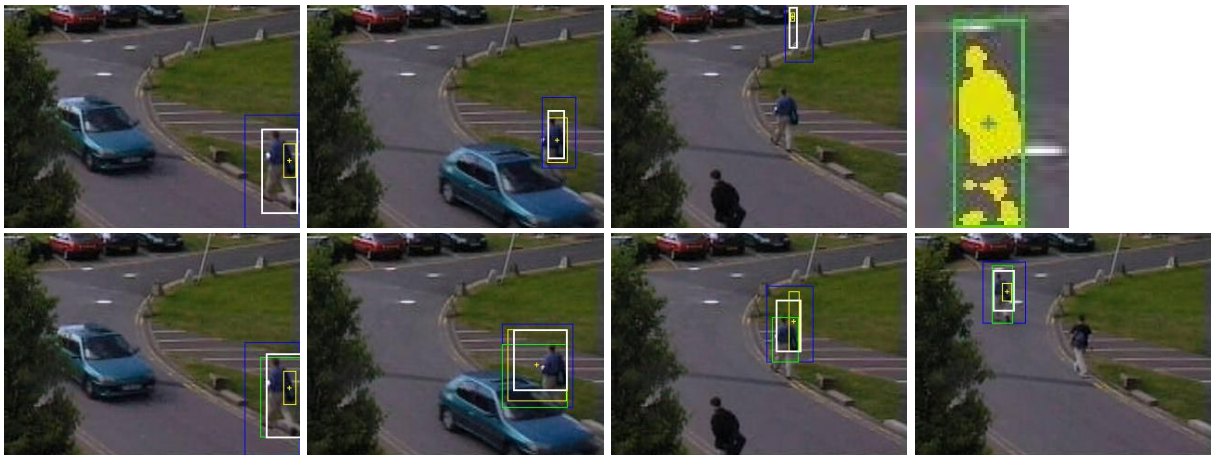

Abbildung 3: Verfolgungsergebnisse mit dem Condensation-Algorithmus. Obere Zeile: Nur P2DHMM. Untere Zeile: P2DHMM kombiniert mit einem Bewegungsdetektor (angedeutet durch ein zusätzliches Rechteck). Siehe Text.

Literatur

[1] A. M. Baumberg and D. C. Hogg. An Efficient Method for Contour Tracking Using Active Shape Models. Technical Report 94.11, School of Computer Studies, University of Leeds, Apr. 1994.

[2] Q. Cai and J. Aggarwal. Tracking Human Motion Using Multiple Cameras. In *Proceedings of ICPR*, volume 3, pages 68–72, Vienna, 1996.

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models — their training and application. *Computer Vision and Image Understanding: CVIU*, 61(1):38–59, Jan. 1995.

[4] S. Eickeler, S. Müller, and G. Rigoll. Recognition of JPEG Compressed Face Images

Based on Statistical Methods. *Image and Vision Computing Journal, Special Issue on Facial Image Analysis*, 18(4):279–287, Mar. 2000.

[5] D. Gavrila. The Visual Analysis of Human Movement: A Survey. *Computer Vision and Image Understanding*, 73(1):82–98, Jan. 1999.

[6] M. S. Grewal and A. P. Andrews. *Kalman Filtering Theory and Practice*. Prentice-Hall, Englewood Cliffs, 1993.

[7] B. Heisele and C. Wöhler. Motion-Based Recognition of Pedestrians. *International Conference on Pattern Recognition*, pages 1325–1330, 1998.

[8] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[9] A. Iketani, A. Nagai, Y. Kuno, and Y. Shirai. Detecting Persons on Changing Background. In *Proceedings of ICPR*, volume 1, pages 74–76, Brisbane, 1998.

[10] M. Isard and A. Blake. Condensation – Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision*, 29(1):5–28, Aug. 1998.

[11] S. Kuo and O. E. Agazzi. Keyword Spotting in Poorly Printed Documents Using Pseudo 2-D Hidden Markov Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 842–848, Aug. 1994.

[12] S. Müller, S. Eickeler, C. Neukirchen, and B. Winterstein. Segmentation and Classification of Hand-Drawn Pictograms in Cluttered Scenes - An Integrated Approach. In *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3489–3492, Phoenix, Mar. 1999.

[13] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian Detection Using Wavelet Templates. In *Proceedings of Computer Vision and Pattern Recognition*, pages 193–199, 1997.

[14] L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. of the IEEE*, 77(2):257–285, Feb. 1989.

[15] G. Rigoll, S. Eickeler, and S. Müller. Person Tracking in Real-World Scenarios Using Statistical Methods. In *IEEE Int. Conference on Automatic Face and Gesture Recognition*, pages 342–347, Grenoble, France, Mar. 2000.

[16] J. Segen and S. Pingali. Camera-Based System for Tracking People in Real Time. In *Proceedings of ICPR*, volume 3, pages 63–67, Vienna, 1996.

[17] M. Shah and R. Jain. *Motion-Based Recognition*. Computational Imaging and Vision. Kluwer Academic Publishers, 1997.

[18] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-Time Tracking of the Human Body. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 19, pages 780–785, July 1997.

[19] L.-Q. Xu and D. C. Hogg. Using neural network to learn spatial-temporal models for moving deformable objects training. In *International Workshop on Neural Networks for Identification, Control, Robotics and Signal/Image Processing*, pages 145–153, 1996.

[20] T. Yamane, Y. Shirai, and J. Miura. Person tracking by integrating optical flow and uniform brightness regions. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA-98)*, pages 3267–3272, Piscataway, May 16–20 1998. IEEE Computer Society.