Ein integrierter Ansatz zur Lokalisierung von Personen in Bildfolgen

Stefan Feyrer, Andreas Zell
Wilhelm-Schickard-Institut für Informatik, Abt. Rechnerarchitektur
Universität Tübingen
Köstlinstraße 6, D-72074 Tübingen
{feyrer, zell}@informatik.uni-tuebingen.de

Zusammenfassung

Gegenstand dieses Beitrags ist die Beschreibung eines Systems, das in der Lage ist, Personen visuell zu detektieren und mittels Kameranachführung schritthaltend zu verfolgen. Als Basiskomponenten kommen Farb-, Bewegungs- und Konturdetektionsverfahren zum Einsatz. Die hohe Robustheit des Systems wird zum einen durch die gegenseitige Beeinflussung dieser Basisverfahren auf Detektionsebene, zum anderen durch die Kombination der Detektionsergebnisse beim Selektionsprozeß erreicht. Verschiedene Kombinationsmöglichkeiten der Einzelkomponenten werden beschrieben und evaluiert.

1 Einführung

Mobile Serviceroboter, die für Aufgaben eingesetzt werden sollen, bei denen es wichtig ist, daß die Interaktion mit Menschen in möglichst natürlicher Art und Weise geschieht, müssen in der Lage sein, menschliche Benutzer, also Personen, als solche zu erkennen. Eine solche passive Detektionskomponente muß allerdings durch eine aktive Blicksteuerung, die den Bewegungen des Benutzers folgen kann, ergänzt werden. Erst dadurch wird dem Prinzip des aktiven Sehens [1] Rechnung getragen und eine Ausgangsbasis geschaffen, die weitere Erkennungsschritte und Wahrnehmungsstufen wie beispielsweise eine Personenidentifikation sinnvoll einsetzbar macht.

Die meisten in der Literatur beschriebenen Ansätze zur visuellen Detektion von Personen basieren auf einer von mehreren möglichen Verfahrensklassen. Bewegungsorientierte Verfahren [10] [11] haben den Nachteil, daß bei bewegter Kamera eine Kompensation der Eigenbewegung durchgeführt werden muß, die meist zu einer Verminderung der Detektionsgenauigkeit führt. Das Problem farbbasierter Ansätze [13] ist die hohe Empfindlichkeit gegenüber Beleuchtungseinflüssen und Änderungen der Farbtemperatur. Struktur- oder featureorientierte Verfahren [14] [16] sind häufig sehr rechenintensiv, wenn zuvor keine Eingrenzung auf potentielle Bildbereiche vorgenommen werden kann. Innerhalb der einzelnen Verfahrensklassen wurden sowohl konventionelle als auch neuronal arbeitende Ansätze [2] [9] [12] beschrieben. Einige Verfahren setzen eine stationäre Kamera und gleichbleibende Beleuchtungsverhältnisse voraus [8] [15].

Um eine robuste Erkennung auch in natürlichen Umgebungen zu gewährleisten, genügt es nicht, die einzelnen Verfahren unabhängig voneinander zu realisieren und die Einzelergebnisse in einen sich anschließenden separaten Selektionsprozeß einfließen zu lassen. Das System muß vielmehr auf allen Ebenen eine gegenseitige Beeinflussung der einzelnen Komponenten zulassen. Des weiteren muß es möglich sein, ein geometrisches Personenmodell zu integrieren.

Neben diesen Gesichtspunkten wurde bei der Realisierung des Systems besonderer Wert auf die Echtzeitfähigkeit im Hinblick auf den Einsatz auf einer mobilen und somit ressourcenbeschränkten Roboterplattform gelegt.

2 Merkmalsdetektion

Im folgenden werden die drei Basiskomponenten vorgestellt, auf die das Gesamtsystem aufgebaut ist. Bereits hier wird an einigen Stellen die enge Verzahnung der Einzelverfahren untereinander deutlich.

2.1 Farbbasierte Detektion

Die Detektion von Hautfarbe basiert auf dem helligkeitsinvarianten HSI-Farbraum. Lediglich die HS-Komponente der einzelnen Bildpunkte wird für die Klassifikation herangezogen, der Intensitätsanteil wird nicht berücksichtigt. Leichte Schwankungen der Farbtemperatur und Beleuchtungseinflüsse werden durch ein globales Farbkorrekturverfahren [6] weitgehend eliminiert. Die Bereiche im zweidimensionalen HS-Raum, die als Hautfarbe klassifiziert werden sollen, sind über eine Lookup-Tabelle definiert. Diese Vorgehensweise findet ihre Begründung in empirischen Untersuchungen an Farbbildsequenzen von 13 Personen, die mittels manueller Segmentierung ergeben haben, daß Hautfarbbereiche nur ungenügend durch eine Normalverteilung im Farb-Unterraum, wie sie in anderen Systemen verwendet wurde [7] [13], beschrieben werden kann. In Abb. 1 ist dieses Ergebnis graphisch dargestellt. Sowohl der Hue-Bereich etwa bis zu einem Wert von 210 als auch der Einschnitt der Sättigungswerte in der Umgebung des Hue-Wertes 0 wären bei Verwendung einer Normalverteilung nicht adäquat repräsentiert. Um die Geschwindigkeit der Detektion zu erhöhen, wird ein zu klassifizierender Bildpunkt nicht einer rechnerisch aufwendigen Transformation in den HSI-Raum unterzogen. Es wird statt dessen eine in den RGB-Raum rücktransformierte Lookup-Tabelle verwendet.

Zur Clusterung von hautfarbenen Bereichen kommt ein Algorithmus zum Einsatz, der in Anlehnung an das Region-Growing-Verfahren von einem Startpunktgitter ausgeht, das im Laufe eines Iterationsprozesses sukzessive verfeinert wird. Die betrachteten Gitterpunkte sind folgendermaßen definiert:

$$P_{i,j} = \left[w \cdot \frac{i}{2^j}, h \cdot \frac{i}{2^j}\right] \quad \text{mit} \quad i < j \qquad \text{für} \quad j = 1, 2, \ldots \tag{1}$$

w bezeichnet dabei die Breite, h die Höhe des Bildes. Von jedem Gitterpunkt aus wird im Fall, daß es sich um einen hautfarbenen Punkt handelt, ein Flood-Fill-Algorithmus gestartet, der gleichzeitig einen 3x3-Open-Operator realisiert. Das bedeutet, daß jeweils le-

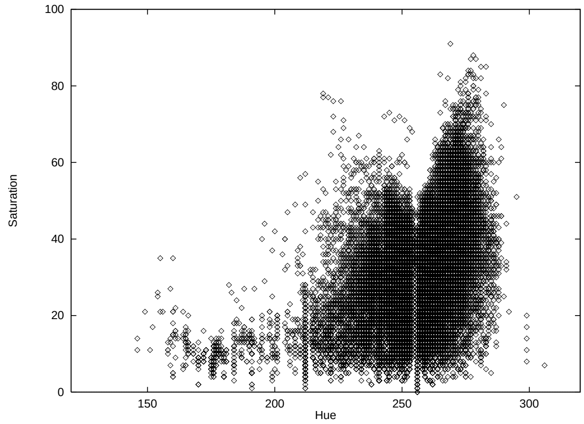

Abbildung 1. Farbverteilung der manuell segmentierten Hautfarbbereiche in den verwendeten Testsequenzen. Der Wertebereich von Hue und Sättigung erstreckt sich jeweils von 0 bis 255, wobei die Winkeleigenschaften des Hue-Wertes eine zyklische Interpretation erfordern.

diglich Punkte zum aktuellen Bereich hinzugefügt werden, die im Sinne des Erosionsoperators innere Punkte darstellen, d.h. die in ihrer Umgebung ausschließlich Hautfarbpixel aufweisen. Derartige Punkte werden hier als primäre Punkte bezeichnet. Sekundäre Hautfarbpunkte, die am Rand des erodierten Bereiches liegen und durch die im Open-Operator enthaltene Dilation zur Region hinzugefügt würden, werden ebenfalls als zur Region gehörend gekennzeichnet. Der Füllalgorithmus jedoch beschränkt sich auf die Verwertung primärer Punkte. Daß durch diese Vorgehensweise keine topologischen Fehlsegmentierungen entstehen, ist durch die Tatsache abgesichert, daß erodierte Bereiche topologisch äquivalent zu ihrer dilatierten Version sind.

Vorteil dieses Verfahrens ist die stark reduzierte Rechenzeit. Sowohl die Klassifikation als auch die morphologische Open-Operation müssen mit Ausnahme von Randbereichen und Einzelpunkten lediglich für die tatsächlichen Hautfarbbereiche durchgeführt werden.

Die auf diese Weise segmentierten Bereiche werden anschließend einem strukturbasierten Auswahlverfahren unterzogen, auf das in Abschnitt 3.2 genauer eingegangen wird.

2.2 Bewegungserkennung

Bewegungsorientierte Detektionsalgorithmen haben den Nachteil, daß sie sich nur mit großem zusätzlichen rechnerischen Aufwand auf Systemen realisieren lassen, die mit bewegter Kamera arbeiten. Hinzu kommt, daß die dabei erzielbare Detektionsgenauigkeit deutlich unterhalb der Genauigkeit liegt, die sich mit stationärer Kamera erreichen läßt. Aus diesem Grund wurde in der hier vorgestellten Version des Systems auf eine Kompensation der Eigenbewegung, wie sie beispielsweise in [3] beschrieben ist, verzichtet. Anstelle dessen kommt ein einfaches Differenzbildverfahren zum Einsatz, das mit zwei bei temporär stationärer Kamera in direkter Folge aufgenommenen Bildern arbeitet. Mittels eines Schwellwertfilters werden diejenigen Bereiche im Bild extrahiert, in denen während dieser Zeitspanne signifikante Grauwertveränderungen stattgefunden haben. Voraussetzung für diese Vorgehensweise ist, daß die Kameranachführung mit sehr hoher Geschwindigkeit erfolgt. Die verwendete Pan-Tilt-Unit besitzt eine maximale Geschwindigkeit von 300°/s, woraus sich selbst unter Berücksichtigung von Beschleunigungs- und Verzögerungsphasen bei der momentan möglichen Verarbeitungsgeschwindigkeit eine maximale Auslenkung pro Verarbeitungszyklus ergibt, die deutlich unterhalb des durch das verwendete Objektiv vorgegebenen Sichtbarkeitsbereichs liegt. Somit stellt zum gegenwärtigen Zeitpunkt die Pan-Tilt-Geschwindigkeit noch keinen limitierenden Faktor dar. Würde die Verarbeitungsgeschwindigkeit allerdings signifikant erhöht, müßte das Verfahren durch eine Eigenbewegungskompensation ersetzt werden.

2.3 Konturdetektion

Charakteristisches Merkmal von Personen ist der Konturverlauf des Kopf-/Schulterbereichs. Eine relativ hohe Invarianz gegenüber Veränderungen der Körperposition im Vergleich zu den Extremitäten prädestiniert diesen Konturbereich für die Erkennung von Personen.

Die Konturdetektion bedient sich eines Verfahrens, das unter dem Begriff *aktive Strahlen* bekannt ist [4] [5] und in sehr enger Verwandtschaft zu den sogenannten *aktiven Konturen* [1] steht. Der Vorteil aktiver Strahlen gegenüber den aktiven Konturen liegt in der Reduktion einer zwei- auf eine eindimensionale Energieminimierung und folglich in einer Verminderung der benötigten Rechenzeit. Grundsätzliche Vorgehensweise bei aktiven Strahlen ist die Untersuchung von Grauwertverläufen auf Geraden, die sämtlich durch einen innerhalb der Objektkonturen befindlichen Referenzpunkt verlaufen. Mit Hilfe einer Energieminimierung wird versucht, eine Punktkette zu ermitteln, die den tatsächlichen Objektkonturverlauf beschreibt. Die Energiefunktion setzt sich zusammen aus einer äußeren Energie, die in diesem Fall durch den negativen Gradienten des Grauwertverlaufs bestimmt wird, und einer inneren Energie, die bestimmte Verlaufseigenschaften der Konturpunktkette charakterisiert. Die verwendete Funktion für die innere Energie wurde gegenüber [4] um einen Term erweitert und hat folgendes Aussehen:

$$E = \int_{-\frac{\pi}{4}}^{\frac{5\pi}{4}} \left(\alpha \left| \frac{\partial}{\partial \varphi} \lambda(\varphi) \right|^2 + \beta \left| \frac{\partial^2}{\partial \varphi^2} \lambda(\varphi) \right|^2 + \gamma D_m^2(\lambda, \varphi) \right) d\varphi \qquad (2)$$

Der erste Term beschreibt die Spannung, der zweite Term die Glattheit der Kontur. In den dritten, neuen Term fließen die Ergebnisse der Bewegungsdetektion ein (siehe Abschnitt 3.1). Durch das eingeschränkte Integrationsintervall wird erreicht, daß der Kopf-Rumpf-Übergang bei der Kontursuche nicht berücksichtigt wird, da dort meist kein eindeutiger Konturverlauf festzustellen ist.

Durch einen Vergleich mit einer empirisch ermittelten Durchschnittskontur wird ein Ähnlichkeitsmaß bestimmt, das wie folgt definiert ist:

$$Q_{Contour} = \int_{-\frac{\pi}{4}}^{\frac{5\pi}{4}} a(\tau\lambda_{min}(\varphi) - \lambda_{ave}(\varphi))^2 + b\left(\frac{\partial}{\partial\varphi}\lambda_{min}(\varphi) - \frac{\partial}{\partial\varphi}\lambda_{ave}(\varphi)\right)^2 d\varphi \qquad (3)$$

Der durch λ_{min} beschriebene momentane Konturverlauf erfährt dabei mittels eines Skalierungsfaktors τ eine Größennormierung. Die Werte für λ_{ave} sind jeweils der relativen Referenzpunktposition zur detektierten Kontur angepaßt. Zum gegenwärtigen Zeitpunkt wird vorausgesetzt, daß sich Personen aufrecht im Bild befinden. Eine strukturelle Gesichtsanalyse mit Bestimmung der Richtung der Augenlinie könnte diese Rotationsabhängigkeit beseitigen.

3 Kombination der Einzelverfahren

Die beschriebenen Einzelkomponenten weisen an vielen Stellen Punkte auf, die eine gegenseitige Beeinflussung ermöglichen bzw. erfordern. Neben dieser direkten Kombination werden bei der anschließenden Selektion Ergebnisse aller drei Komponenten zusammengeführt.

3.1 Kombination auf Detektionsebene

Die Konturdetektion hängt sehr stark von den Ergebnissen von Farb- und Bewegungserkennung ab. Der für das Verfahren der aktiven Strahlen notwendige Referenzpunkt beispielsweise wird von der Farbsegmentierung geliefert. Dazu wird der Mittelpunkt des umgebenden achsenparallelen Rechtecks der Farbregion herangezogen. Dies stellt in fast allen Fällen sicher, daß sich dieser Ausgangspunkt innerhalb der Kopfkontur befindet. Des weiteren wird die Größe des umgebenden Rechtecks verwendet, um die Größennormierung in (3) zu bewerkstelligen.

Auch die Bewegungsdetektion wird von der Farbsegmentierung beeinflußt. Lediglich in erweiterten Umgebungen von Farbregionen wird die Differenzbildanalyse durchgeführt. Dies bewirkt zwar keine qualitative Verbesserung des Verfahrens, wohl aber eine Effizienzsteigerung.

Schließlich bedient sich die Konturanalyse der Ergebnisse der Bewegungsdetektion, indem die Konturverläufe im bewegungssegmentierten Bild mit in die Funktion der in (2) definierten inneren Konturenergie über folgende Definition einfließen:

$$D_m(\lambda, \varphi) = \begin{cases} \min\{|\lambda(\varphi) - \lambda_m| \mid \lambda_m \in M(\varphi)\} & \text{für} \quad M(\varphi) \neq \emptyset \\ 0 & \text{für} \quad M(\varphi) = \emptyset \end{cases} \tag{4}$$

$M(\varphi)$ ist dabei definiert als die Menge derjenigen Parameterwerte λ, bei denen auf dem betrachteten Strahl mit Zentrum (x_0, y_0) ein Übergang zwischen einem als bewegt und einem als statisch detektierten Bereich stattfindet:

$$M(\varphi) = \{\lambda \mid \nabla f_m(x_0 + \lambda\cos\varphi, y_0 + \lambda\sin\varphi) \neq 0\} \tag{5}$$

Der dritte Term ist somit ein Maß für die Entfernung des Punktes zum nächstgelegenen Bewegungskonturpunkt.

3.2 Selektion von Kopfregionen

Ziel des Selektionsprozesses ist es, Gesichtsbereiche von anderen detektierten Regionen zu trennen. Dazu werden zunächst Merkmale der durch die Farbsegmentierung entstehenden Bereiche bestimmt. Das Höhen-Breiten-Verhältnis des umgebenden achsenparallelen

Rechtecks wird ebenso herangezogen wie dessen Füllgrad. Ferner wird der Durchschnitt der relativen Häufigkeiten der HS-Farbkombinationen der zum Bereich gehörenden Bildpunkte berechnet. Dieser Wert gibt Aufschluß darüber, wie stark die Farbe des Bereichs von der durch die manuelle Segmentierung ermittelten „mittleren Hautfarbe“ abweicht. Des weiteren fließen der Betrag der Positionsänderung und der Größenänderungsfaktor im Verhältnis zu den im jeweiligen Vorgängerbild selektierten Bereichen ein. Dies verhindert ein Springen der selektierten Regionen und ermöglicht ein gleichmäßiges Tracking. Beim Vergleich mit den Ergebnissen des Vorgängerbildes werden sowohl mögliche Zeitdifferenzen zwischen aufeinanderfolgenden Bildern als auch die Tatsache berücksichtigt, daß weiter von der Kamera entfernt stattfindende Bewegungen zu einer geringeren Positionsveränderung führen.

Die einzelnen Kriterien werden in einem Merkmalsvektor x zusammengefaßt und fließen über folgende Gleichungen in die farbbasierte Bewertung ein:

$$Q_{Color}(x) = \sum_i \alpha_i q_i(x_i) \qquad \text{mit} \qquad q_i(x_i) = e^{-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)} \tag{6}$$

μ_i und σ_i bezeichnen den Mittelwert bzw. die Standardabweichung des jeweiligen Merkmals x_i. Die Einzelbewertungen werden über konstante Faktoren α_i gewichtet. Eine Region wird von der Selektion ausgeschlossen, wenn ihre absolute Größe einen von der Bildgröße abhängigen Minimalwert unterschreitet. Tritt der Fall ein, daß die detektierte Region sich an einem der Bildränder befindet, geht das Höhen-Breiten-Verhältnis nicht in die Bewertung mit ein, da in diesem Fall keine Aussage über die tatsächliche Objektform gemacht werden kann. In Tab. 1 sind die durch Auswertung der Testsequenzen ermittelten Bewertungskennwerte angegeben..

Merkmal	μ	σ
Positionsänderung	0,0623	0,0791
Größenänderung	1,0139	0,0809
Höhen-Breiten-Verhältnis	0,7474	0,1236
Füllgrad	0,6163	0,0850
Farbwertigkeit	0,2831	0,0519

Tabelle 1. Statistische Kennwerte der Strukturmerkmale zur Selektion von Kopfregionen. Die Änderungsraten beziehen sich auf ein im Abstand von 1/24 Sekunde aufgenommenes Vorgängerbild. Die Positionsänderung ist größenbereinigt.

Die Ergebnisse der Bewegungsdetektion fließen in die Gesamtbewertung über den prozentualen Anteil der bewegungsdetektierten Bildpunkte bezogen auf die um den Faktor 1,5 vergrößerten Farbregionen ein. Die gewichtete Summe der Einzelbewertungskomponenten ergibt schließlich den Wert, der für die Selektion der „Gewinnerregion“ herangezogen wird:

$$Q = c_1 Q_{Color} + c_2 Q_{Motion} + c_3 Q_{Contour} \tag{7}$$

4 Ergebnisse

Einsatzplattform des beschriebenen Systems ist ein mobiler Roboter des Typs RWI B21. Der Roboter besitzt zwei Doppelprozessor-PentiumPro-PCs, von denen allerdings lediglich einer für die Bildverarbeitung genutzt werden kann. Die Bildakquisition erfolgt über zwei CCD-Farbkameras, die auf einer Pan-Tilt-Unit montiert sind. Zum gegenwärtigen Zeitpunkt werden nur die Bilder einer der beiden Kameras verarbeitet.

Aufgrund der sehr starken Abhängigkeit von Szenenbeschaffenheit und -dynamik kann keine objektive Aussage über die Erkennungsleistung des Systems in Form von Erkennungsraten gemacht werden. Die Analyse mehrerer unterschiedlicher Bildsequenzen läßt jedoch den qualitativen Schluß zu, daß durch Hinzunahme der Bewegungskomponente zur Farbdetektion eine signifikante Erhöhung der Erkennungsleistung erzielt werden kann. Die zusätzliche Integration von Konturinformation führt zu einer verhältnismäßig geringen Verbesserung der Gesamtleistung. Tab. 2 zeigt exemplarisch

Erkennungsleistung	Konturdetektion aus	Konturdetektion ein
Bewegungsdetektion aus	92,5%	95,0%
Bewegungsdetektion ein	97,9%	98,9%

Tabelle 2. Erkennungsleistung bei verschiedenen Kombinationen von Einzelkomponenten bezogen auf eine ausgewählte Bildsequenz.

für eine ausgewählte Bildsequenz mit starker Szenendynamik die ermittelten Erkennungsraten. Bei einer Auflösung von 192x144 Bildpunkten kann momentan eine Verarbeitungsgeschwindigkeit von 8 Bildern pro Sekunde erreicht werden. Abb. 1 zeigt Einzelbilder einer Beispielsequenz mit den detektierten Kopfbereichen.

Abbildung 2. Ermittelte Kopfregionen und detektierte Konturverläufe. Die Einzelbilder wurden im Abstand von 5 Sekunden aufgenommen, um größere Bewegungsänderungen zu demonstrieren.

5 Zusammenfassung und Ausblick

Es wurde ein Ansatz zur visuellen Lokalisierung von Personen vorgestellt, der auf der Integration von Farb-, Form- und Bewegungsinformation beruht. Die Erkennungsleistung des Gesamtsystems beruht sehr stark auf der Verzahnung der Einzelkomponenten sowohl auf Detektionsebene als auch auf Evaluations- und Selektionsebene. Das System erreicht PC-basiert eine Verarbeitungsgeschwindigkeit von 8 Bildern pro Sekunde und ermöglicht so ein schritthaltendes Tracking durch Kameranachführung.

Geplante Erweiterungen sind die Verbesserung der Bildfolgenanalyse durch den Einsatz eines α-β-γ-Trackers, die Ergänzung des Personenmodells um die Schulter- und Rumpfpartie sowie die Berücksichtigung struktureller Gesichtsmerkmale.

Literatur

[1] A. Blake, A. Yuille: *Active Vision*, MIT Press, 1992

[2] A. Brakensiek, U.-D. Braumann, H.-J. Böhme, C. Rieck, H.-M. Groß: *Farb- und strukturbasierte neuronale Verfahren zur Lokalisation von Gesichtern in Real-World-Szenen*, in E. Paulus, F. M. Wahl (Hrsg.): Mustererkennung 1997, Springer, S. 113-120, 1997

[3] K. Daniilidis, M. Hansen, C. Krauss, G. Sommer: *Auf dem Weg zum künstlichen aktiven Sehen: Modellfreie Bewegungsverfolgung durch Kameranachführung*, in G. Sagerer, S. Posch, F. Kummert (Hrsg.): Mustererkennung 1995, Springer, S. 277-284, 1995

[4] J. Denzler: *Aktives Sehen zur Echtzeitobjektverfolgung*, Dissertation, Technische Fakultät der Universität Erlangen-Nürnberg, 1997

[5] J. Denzler, H. Niemann: *Echtzeitobjektverfolgung mit Aktiven Strahlen*, in B. Jähne, P. Geißler, H. Haußecker, F. Hering (Hrsg.): Mustererkennung 1996, Springer, S. 84-91, 1996

[6] S. Feyrer, O. Schimmel, A. Zell: *Adaptive farbbasierte Objektverfolgung mit dem mobilen Roboter ROBIN*, Workshop SOAVE 97, Selbstorganisation von adaptivem Verhalten, Fortschritt-Berichte VDI, Reihe 8, Nr. 663, S. 176-185, 1997

[7] W. Hafner, H. Kirchner, O. Munkelt: *Farbklassifikation im Projekt STABIL*, 1. Workshop Farbbildverarbeitung, Fachberichte Informatik, 15/95, S. 25-29, 1995

[8] K. Huggle, W. Eckstein: *Extraktion von Personen in Videobildern*, in G. Sagerer, S. Posch, F. Kummert (Hrsg.): Mustererkennung 1995, Springer, S. 134-144, 1995

[9] R. Kubisch, H. Ritter: *Erkennung menschlicher Kopfhaltungen mittels künstlicher neuronaler Netze*, in B. Jähne, P. Geißler, H. Haußecker, F. Hering (Hrsg.): Mustererkennung 1996, Springer, S. 109-117, 1996

[10] C. Ridder, O. Munkelt, H. Kirchner: *Adaptive Background Estimation and Foreground Detection Using Kalman-Filtering*, ICRAM 95, International Conference on Recent Advances in Mechatronics, Istanbul, 1995

[11] K. Rohr: *Towards Model-Based Recognition of Human Movements in Image Sequences*, CVGIP: Image Understanding, Bd. 59, Nr. 1, S. 94-115, 1994

[12] H. A. Rowley, S. Baluja, T. Kanade: *Neural Network-Based Face Detection*, IEEE PAMI, Bd. 20, Nr. 1, S. 23-38, 1998

[13] M. Sommerau, G. Mamier, A. Zell, M. Vogt, P. Levi: *Fast Face Localization and Tracking with Model-Based Time Synchronization of a Head System*, in G. Sagerer, S. Posch, F. Kummert (Hrsg.): Mustererkennung 1995, Springer, S. 516-523, 1995

[14] K.-K. Sung, T. Poggio: *Example-Based Learning for View-Based Human Face Detection*, IEEE PAMI, Bd. 20, Nr. 1, S. 39-51, 1998

[15] C. R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland: *Pfinder: Real-Time Tracking of the Human Body*, IEEE PAMI, Bd. 19, Nr. 7, S. 780-785, 1997

[16] K. C. Yow, R. Cipolla: *Feature-based human face detection*, Image and Vision Computing, Bd. 15, S. 713-735, 1997