

Gaborfilterbasierte visuelle Personenlokalisierung mit dreidimensionalen Feldern dynamischer Neuronen

Ulf-Dietrich Braumann, Anja Brakensiek
Hans-Joachim Böhme, Horst-Michael Groß

Fachgebiet Neuroinformatik
Technische Universität Ilmenau
PF 10 05 56
D-98684 Ilmenau
Tel.: +49 3677 69 1305, Fax.: +49 3677 69 1665
E-Mail: ulf@informatik.tu-ilmenau.de

Zusammenfassung

Für die *visuell basierte Interaktion* zwischen unpräparierten Personen und autonom agierenden Systemen innerhalb ebenso unpräparierter Umgebungen bildet die *Lokalisation* eines Nutzers durch das System die notwendige Voraussetzung. Die vorliegende Arbeit behandelt die Lokalisationsproblematik eines Nutzers für das *autonome mobile Robotersystem MILVA*¹. Da dieses System unter Real-World-Bedingungen eingesetzt wird, werden insbesondere zur Lokalisation Verfahren benötigt, die eine adäquate Robustheit besitzen. Vorgestellt wird hierfür ein gaborfilterbasierter Multi-Resolution-Ansatz zur konturbasierten Nutzerlokalisierung anhand eines einfachen Konturmodells für die Kopf-Schulter-Partie von Personen. Die Selektion eines Bildausschnitts (einschließlich Tiefeninformation), in dem mit höchster Wahrscheinlichkeit ein Nutzer zu finden ist, erfolgt mit einem dreidimensionalen Feld dynamischer Neuronen (dynamisches WTA nach Amari [1]).

Schlüsselwörter: visuell basierte Mensch-Maschine-Interaktion, visuelle Personenlokalisierung, Multi-Resolution-Analyse, Filter vom Gabor-Typ, dynamische neuronale Felder

1 Einleitung

Das mobile Robotersystem MILVA des FG Neuroinformatik der TU Ilmenau dient nicht allein als Experimentalplattform für Arbeiten zur autonomen Navigation in Real-World-Umgebungen (siehe dazu [12] und [15]), sondern ebenso für Arbeiten zur gestenbasierten Interaktion zwischen Nutzer und Roboter (siehe [2]). Sowohl für Navigation als auch Interaktion steht dabei zunehmend im Vordergrund, ohne spezielle Vorkehrungen innerhalb der Einsatzumgebungen des Roboters auskommen zu können und damit mehr und mehr die Komplexität realer Umgebungen (z. B. hinsichtlich des Arrangements von Gegenständen, der Beleuchtungsbedingungen usw.) zuzulassen. *Übergeordnete Zielstellung* ist, daß MILVA innerhalb einer *neuronalen Architektur* durch *aktive Lernprozesse* ein *intelligentes Verhalten* entwickelt, was sich nach unserer Ansicht nur erfolgreich in ständiger Interaktion mit der jeweiligen Umwelt vollziehen kann. MILVA (siehe Abbildung 1) ist technisch dafür ausgelegt, sich in Innenräumen (etwa Labors, Supermärkten usw.) ebenso wie im Außenbereich (alle befestigten Wege ohne Treppen) bewegen zu können. Die Herausforderung und der Reiz, in derlei Umgebungen insbesondere unter der genannten Zielstellung zu arbeiten ist also groß, zumal komplexitätsarme Umgebungen praktisch kaum vorzufinden sind, selbst scheinbar schlichte Räumlichkeiten weisen bei genauerem Betrachten auf kleinem Raum oft viele verschiedene Bedingungen auf.



Abbildung 1: Die mobile Roboterplattform MILVA

Eine naheliegende Aufgabenstellung für MILVA ist die Unterstützung von Personen (Nutzern) bei Transportaufgaben. Mögliche Einsatzszenarien dafür sind öffentliche Einrichtungen, in denen ein entsprechender Transportbedarf besteht. Wir haben das Zielszenario *Bahnhof* gewählt, da darin nach unserer Ansicht eine gesten- und damit visionbasierte Schnittstelle deutliche Vorteile gegenüber einer akustischen Schnittstelle aufweist. Zu diesem Zweck ist MILVA mit einem aktiven Zwei-Kamera-System ausgerüstet (gut erkennbar in Abbildung 1), mit Hilfe dessen die Umgebung zugleich im Detail als auch im Überblick erfaßt werden kann.

¹Multisensorielles intelligentes lernfähiges Vehikel in neuronaler Architektur

Eine zielgerichtete und damit intelligente Detailanalyse (aufmerksamkeitsbasierte Wahrnehmung, z. B. für die Analyse einer Geste) kann MILVA aber nur dann entwickeln, wenn sie für die visuelle Verarbeitung grundlegende *Auffälligkeitsmechanismen* besitzt, die nicht erst entwickelt zu werden brauchen und anhand derer das Bild einer Szene *auffälligkeitsbasiert* untersucht und damit vorselektiert werden kann. Für die Interaktion mit Personen sind Auffälligkeitsmechanismen zu deren *Lokalisation* nötig. Abbildung 2 zeigt im Schema alle Komponenten, anhand deren Fusion ein visuelles *Multi-Cue-Auffälligkeitssystem* gebildet wird: Neben der hier vorgestellten *konturbasierten Lokalisation* – einem speziellen Figure-Ground-Problem – stehen drei weitere Auffälligkeitskomponenten zur Verfügung (Personenlokalisierung anhand von hautfarbenen Bildbereichen, gesichtstypischen Bildausschnitten – Eigenfaces und anhand von Bewegung i. S. v. Winkbewegungen) für deren Einzelheiten auf [4] verwiesen wird. Außer bei der Hautfarbe handelt es sich bei allen anderen um grauwertbasiert zu extrahierende Merkmale.

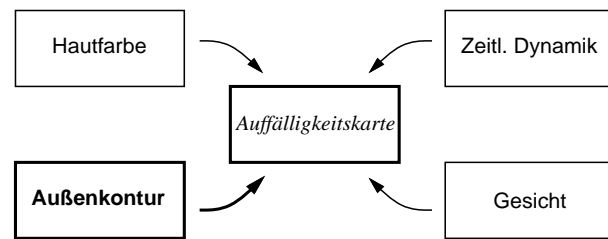


Abbildung 2: *Komponenten des visuellen Multi-Cue-Auffälligkeitssystems von MILVA*

Die Idee, die hinter der konturbasierten Lokalisation steht, soll Abbildung 3 verdeutlichen. Ihr liegt eine wichtige einschränkende Annahme zugrunde: es wird davon ausgegangen, daß ein Nutzer, der mit MILVA in Interaktion treten möchte, diese Absicht auch dadurch ausdrückt, daß er sich frontal zum Roboter gerichtet aufstellt². Weiterhin hat sich gezeigt, daß unter einigen Umgebungsbedingungen durch Beleuchtungseinflüsse beispielsweise Schlagschatten entstehen, die die Anwendung der Eigenfaces aber auch des Hautfarbmodells zur Personenlokalisierung anhand des Gesichts erheblich erschweren. Das Beispielbild in Abbildung 3 (links) enthält zwar prinzipiell dieselbe Flurszene wie in den Abbildungen 8 bis 11, allerdings ist dieses Bild hier offenbar stärker abgeblendet aufgenommen worden, da die Leuchten die Kamera ansonsten zu noch größerer Übersteuerung in einigen Bildbereichen geführt hätten. Außerdem wird in der momentanen Aufstellung der Person deren Gesicht kaum frontal beleuchtet, praktisch ein Resultat des Blendschutzes der Leuchten, was zu relativ schlechten Kontrasten im Bereich des Gesichts der Person führt. Viel makroskopischer dagegen die Außenkontur im Bereich der Kopf-Schulter-Partie! Sie hebt sich zwar nicht durchgängig gut vom Hintergrund ab (siehe Kopfbereich), ist aber mindestens im Schulterbereich aber noch akzeptabel deutlich ausgeprägt sichtbar.

Die Lokalisationsproblematik umfaßt neben einer Schätzung einer Richtung (2D-Koordinaten) noch einen weiteren Aspekt: die Schätzung einer Tiefe (Entfernung) für die richtige Ansteuerung der Kamera zur Detailanalyse. Die Suche nach typischen Konturen schließt die Bestimmung der Größe indirekt ein, so daß damit (neben der Nutzung von Eigenfaces) ein Zugang zur Entfernungsschätzung eines Nutzers gegeben ist.

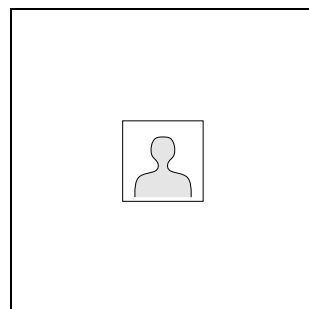


Abbildung 3: *Nutzerlokalisierung anhand des Umrisses der frontalen Kopf-Schulter-Partie: links ein Beispiel eines aufgrund der konkreten Umgebungsbedingungen eines Ganges wenig gesättigten und kontrastarmen Bildes eines Nutzers; rechts die schematische Darstellung des Umrisses der Kopf-Schulter-Partie in korrespondierender Lage und Größe.*

Welche Möglichkeit besteht, um diese Kontur zu erfassen und damit zur Lokalisation von Personen nutzbar werden zu lassen?

2 Gaborfilter zur visuellen Merkmalsextraktion

Gaborfilter [5], eine Variante der Quadraturfilter (vgl. [7]), werden aus der Überlagerung einer Sinus-/Cosinusfunktion und gaußförmiger Einhüllender gebildet (siehe Abbildung 4). Trennt man die Funktion des Gaborfilters in die

²Diese Annahme erscheint deshalb als gerechtfertigt, da nach Beobachtungen auch Personen untereinander sich intuitiv frontal zu anderen Personen (Zielperson) ausrichten (und ggf. stehenbleiben oder auf die Zielperson zugehen), um leichter registriert zu werden.

reelle und imaginäre Komponente, erhält man den geraden und den ungeraden Teilfilter und damit ein Filterpaar:

$$\operatorname{Re}(\phi(\vec{x})) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2}{2\sigma^2}x^2\right) \cos \vec{k}\vec{x} \quad (1)$$

$$\operatorname{Im}(\phi(\vec{x})) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2}{2\sigma^2}x^2\right) \sin \vec{k}\vec{x} \quad (2)$$

Hierin sind k die Wellenzahl, \vec{k} der Wellenzahlvektor (Richtung des Gaborfilters), σ die Halbwertsbreite der gaußschen Einhüllenden, \vec{x} der Ortsvektor (Koordinaten auf Filter) und j die imaginäre Einheit.

Eine Faltung eines Signals mit Gaborfiltern führt zum idealen Kompromiß zwischen zeitlicher bzw. örtlicher und frequenzmäßiger Auflösung (Orientierung und Betrag/Phase). Eine Vielzahl von Untersuchungen stellte heraus, daß es eine überaus große Ähnlichkeit zwischen Profilen rezeptiver Felder von Neuronen des primären visuellen Cortex von Primaten einerseits und zweidimensionalen Gaborfunktionen andererseits gibt (s. z. B. [9]). Ferner konnte festgestellt werden, daß häufig jeweils zwei Zellen die gleiche lokale Lage ihres rezeptiven Feldes besitzen, sich aber in ihrem Antwortverhalten ein Phasenversatz von $\frac{\pi}{2}$ ergibt (siehe [14]). Neuronen mit derartigen rezeptiven Feldern, die sich gegenseitig nur hinsichtlich Orientierung und Wellenzahl (Frequenz) unterscheiden, erlauben eine umfassende Frequenz- und Orientierungszerlegung und tragen damit außerdem zu einer Datenkompression bei.

Für die hier benötigten Orientierungsmerkmale wird keine Phaseninformation benötigt, es wird der Betrag der Ergebnisse von Gleichung 1 und 2 gebildet. Dafür reicht es aus, Filter für Orientierungen von 0 bis π einzusetzen.

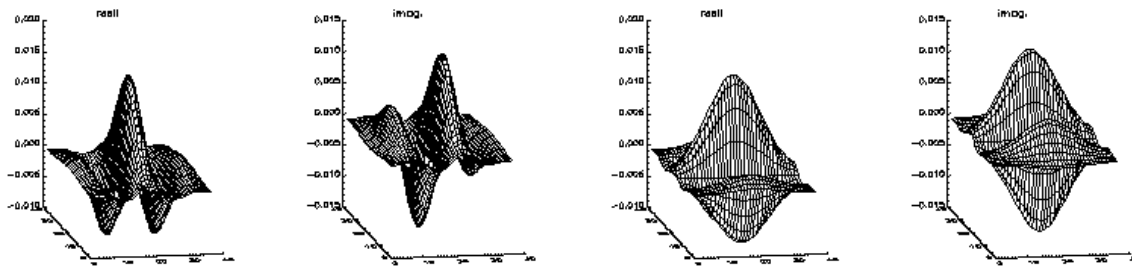
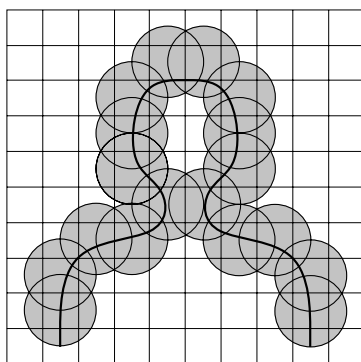


Abbildung 4: Gaborfilterpaare exemplarisch in zwei der genutzten vier ausgewählten Orientierungen, links das 0°-Paar und rechts das 90°-Paar, je mit Real- und Imaginärteil.

3 Arrangements von Gaborfiltern



Orientierung ↑	0°	0	0	0	0	1	0	0	0	0
	45°	0	0	1	0	0	0	0	1	1
	90°	0	0	1	0	0	0	0	0	0
	135°	1	1	0	0	1	0	1	1	0
		linearer Filterindex →								
		0								9

Abbildung 5: Schematische Darstellung von Gaborfilterarrangement und Abstandsmaß. Ergebnis der vier Filterungen an jedem Aufsatzpunkt ist die dominante Orientierung an diesem Punkt.

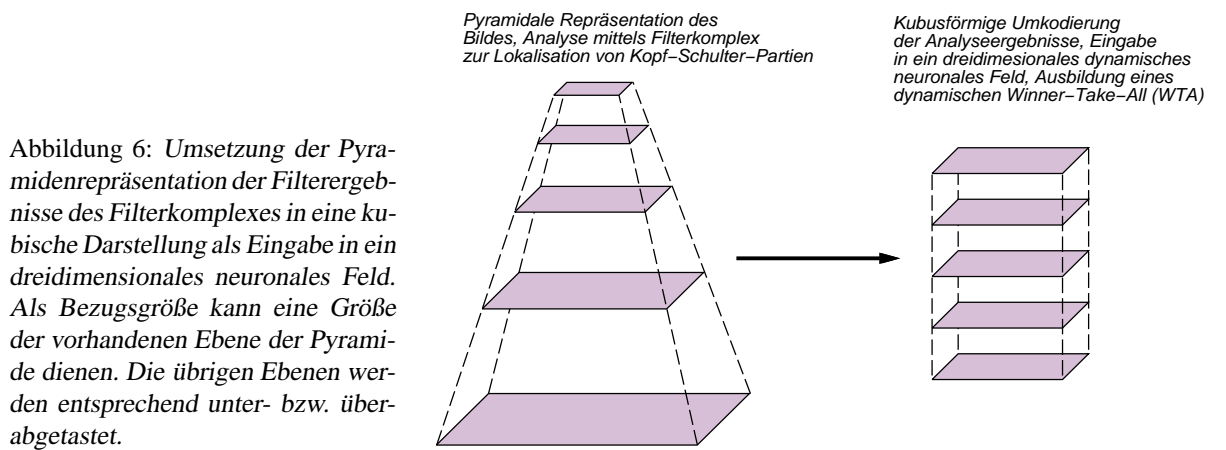
Für das hier vorgestellte Verfahren werden Gaborfilter zur Approximation der Kontur in bestimmter räumlich verteilter Anordnung (Arrangement) mit definierten Ausrichtungen (4 in 45°-Schritten) angewendet, um damit die von Kopf und Schultern einer frontal zur Kamera ausgerichteten Person gebildete typische Kontur zu detektieren und damit die Person zu lokalisieren. Arrangementbildung (Wahl der Gaborfilterparameter, Orientierungsanzahl und der räumlichen Verteilung) einschließlich der Bestimmung eines Abstandsmaßes definieren das gaborfilterbasiertes Modell³ einer Kontur vollständig. Sehr wesentlich ist dabei, daß für das Modell nicht die Filter-Responses

³Das hier benutzte Modell wurde manuell anhand einiger weniger Testbilder entworfen. Es laufen Vorbereitungen, das Modell anhand von quantitativ beschriebenem Bildmaterial (s. u.) durch Bewertung der Lokalisierungsleistung adaptiv zu optimieren.

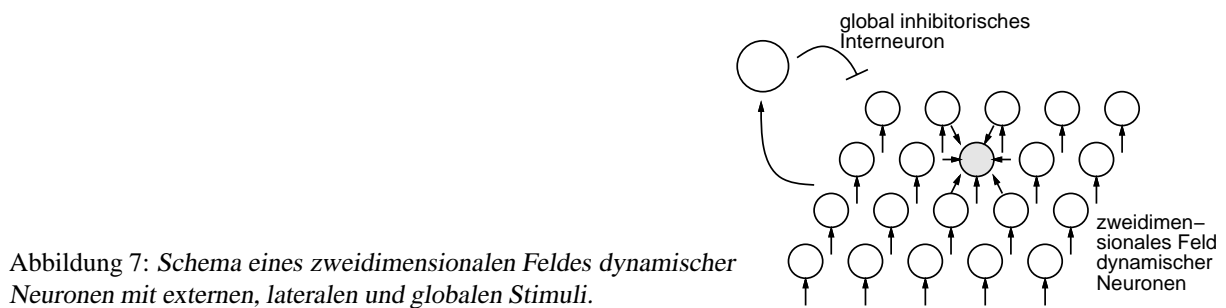
aller Orientierungen herangezogen werden, sondern daraus die *dominante Orientierung* bestimmt wird (siehe Abbildung 5). Als Abstandsmaß zwischen Modell und Daten wird eine einfache Hammingdistanz bestimmt, die allerdings bei richtungsmäßig benachbarten Orientierungen (45° -Differenz) mit 0.2 gewichtet wird. Insgesamt besteht das verwendete Arrangement aus 56 verteilten Aufsatzpunkten entlang des Konturverlaufes.

Die Filterung erfolgt in einer AuflösungsPyramide [8] mit fünf diskreten Abstufungen, so daß die Lokalisation auch hinsichtlich der Entfernung der Person erfolgen kann. Was gewünscht ist, ist die Lokalisation *einer* Person im Bild. Eine Eigenschaft einer idealen Konturkante im Bild ist, daß sie mit $\frac{1}{k}$ abfallende spektrale Anteile enthält, so daß eine ideale Kante zunächst auf allen Ebenen der Pyramide abgebildet wird. Jedes Gaborfilter bringt (s. o.) eine lokale Unschärfe mit und hat gleichzeitig ein frequenz- und richtungsmäßiges Bandpaßverhalten. Das führt für das gesamte Arrangement ebenfalls zu einer Art Unschärfe mit Bandpaßverhalten, so daß das Filterarrangement mindestens auf benachbarten Ebenen zu lokal koinzidenten Responses führen kann, wobei die Translationsinvarianz der Responses besonders dort hoch ist, wo sie insgesamt gut mit einer Kontur übereinstimmen. Solche Koinzidenzen sind ausgehend von Störungen im Bildhintergrund nicht in demselben Maße zu erwarten.

Um eine gute Lokalisationsschätzung ausführen zu können, bedarf es eines Selektionsverfahrens, das die genannten Eigenschaften des Filterarrangements ausnutzt. Für dieses dreidimensionale Lokalisationsproblem können dreidimensionale Felder dynamischer Neuronen verwendet werden. Zur Sicherung der Korrespondenzen aus der Pyramidenrepräsentation für das dynamische Feld ist eine Umsetzung erforderlich (siehe Abbildung 6). Die hier verwendete Größe des dynamischen Feldes bezieht sich auf die feinste Auflösungsebene, damit entsteht ein Feld von $96 \times 71 \times 5$ Neuronen.



4 Dynamische Neuronale Felder



Die Modellierung von Neuronen als *formale dynamische Neuronen* (mit Ratenkodierung des Feuerverhaltens) trägt dem Umstand Rechnung, daß Neuronen über Synapsen erregt werden, die ein Trägheitsverhalten besitzen, so daß einlaufende Spikes in kontinuierliche Potentialverläufe umgesetzt werden. Daß führt nach Aufsummierung am Soma zu einem trägheitsbehafteten Membranpotential der Zelle, weshalb sich ihre Feuerrate nicht sprunghaft ändern kann. Verbände aus interagierenden formalen dynamischen Neuronen lassen sich als Systeme von nichtlinearen gekoppelten Differentialgleichungen mathematisch beschreiben. In der Modellierung neuronaler Informationsverarbeitung spielen solche dynamischen Systeme eine wichtige Rolle (siehe z. B. [1], [10]) bei kompetitiven und

Selektionsprozessen, wobei die Parametrisierung des jeweiligen Systems wesentlich den Typ des auftretenden dynamischen Verhaltens beeinflusst. Ausgehend von der konkreten Problemstellung der Lösung einer Lokalisationsaufgabe ist ein dynamisches Selektionsverhalten gewünscht, bei dem sich *eine* lokale Erregung gegenüber anderen Erregungen des Feldes in einem kompetitiven Prozeß durchsetzt (WTA). Für neuere Anwendungen dynamischer Systeme für Selektionsaufgaben in zweidimensionalen Feldern sei auf [17], [11], [16] & [6] verwiesen.

Demgegenüber wird in der von uns vorgestellten Arbeit neben der Selektion im dreidimensionalen insbesondere eine nicht-isotrope konstante Mexikanerhut-Funktion zur lateralen Verschaltung der Neuronen gewählt, d. h. die Reichweite der Verschaltung ist abhängig von der Raumrichtung unterschiedlich. Während in zwei Raumrichtungen (x und y) gleichgroße Verschaltungsweiten bestehen, wird in z-Richtung eine wesentlich kürzere Verschaltungsweite gewählt. Damit wird dem Umstand Rechnung getragen, daß in dieser Richtung entsprechend der Anzahl der Pyramidenebenen 5 Schichten vorliegen, wobei je zwei benachbarte Schichten Stimuli abbilden, die wesentlich weiter räumlich auseinanderliegen als die zweier benachbarter Neuronen bei festem z.

$$\tau \frac{d}{dt} z(\vec{r}, t) = -z(\vec{r}, t) - c_h h(t) + c_l \int_R w(\vec{r} - \vec{r}') y(\vec{r}', t) d^2 \vec{r}' + c_i x(\vec{r}, t) \quad (3)$$

$$\text{mit } w(\vec{r} - \vec{r}') = 2 \exp\left(\frac{-3|\vec{r} - \vec{r}'|^2}{2\sigma^2}\right) - \exp\left(\frac{-|\vec{r} - \vec{r}'|^2}{\sigma^2}\right) \quad , \quad (4)$$

$$y(\vec{r}, t) = \frac{1}{1 + \exp(-z(\vec{r}, t))} \quad \text{und} \quad (5)$$

$$h(t) = \int_R y(\vec{r}'', t) d\vec{r}'' \quad (6)$$

Darin ist $\vec{r} = (x, y, z)^T$ die Koordinate eines Neurons, $z(\vec{r}, t)$ die Aktivierung eines Neurons \vec{r} zur Zeit t , $y(\vec{r}, t)$ die Aktivität dieses Neurons, $x(\vec{r}, t)$ sein externer Input (entspricht der Differenz aus Anzahl der zum Gaborfilterarrangement gehörenden Filter und dem zum \vec{r} -ten Neuron korrespondierenden Abstandsmaß), $h(t)$ die Aktivität des globalen inhibitorischen Interneurons, $w(\vec{r} - \vec{r}')$ die Funktion zur lateralen Aktivierung des Neurons \vec{r} aus der Nachbarschaft R . τ ist die Zeitkonstante des dynamischen Systems und σ die Halbwertsbreite der Gaußschen Glockenkurven, aus denen die Nachbarschaftsfunktion bestimmt wird. In den Untersuchungen wurden folgende Werte für die Konstanten gewählt: $c_h = 0.025$, $c_l = 0.1$, $c_i = 0.1$, $\sigma = 2$ (in z-Richtung halbiert), $\tau = 10$ bei $\Delta T = 1$ (ΔT : Abtastrate). Der Bereich R der Nachbarschaftsfunktion erstreckt sich über 5 Pixel bzw. 3 Pixel in z-Richtung (anisotrope Nachbarschaft).

5 Ergebnisse und Diskussion

Die Leistungsfähigkeit des Ansatzes kann an Real-World-Kamerabildern aus Innenraum⁴(Flur) als auch Außenbereich (Hof) dargestellt werden. Die Diskussion muß hier weitgehend qualitativ erfolgen, da quantitativ zu beschreibendes Bildmaterial für die Kopf-Schulter-Partie bisher nicht vorliegt und auch generell schwierig zu erstellen ist, da etwa die Eigenschaften des Bildhintergrundes ebenfalls quantitativ eingehen müßten.

Dazu wurden zugunsten eines einfacheren Verständnisses auch alle Zwischenergebnisse einschließlich der dominanten Orientierung⁵ dargestellt. Die Ergebnisse in den Abbildungen stellen Momentaufnahmen zu dem Zeitpunkt dar, an dem sich das maximal aktive Neuron aus dem Feld nicht mehr als 1% beträgt. Durchschnittlich werden für die Selektion 11 Iterationsschritte benötigt.

Die Ausdehnung des aus der Selektion resultierenden Blobs ist in z-Richtung nicht auf eine Ebene beschränkt. Um eine genauere Entfernungsbestimmung vornehmen zu können, müßte der Schwerpunkt des Blobs gebildet werden, so daß die Entfernungsbestimmung somit zwischen den diskreten Abstufungen interpolieren kann.

Der vorgestellte Ergebnisse sind beispielhaft, der Ansatz hat sich als erfreulich robust erwiesen, und kann das Lokalisationsproblem lösen. Der WTA-Prozeß im dreidimensionalen Feld dynamischer Neuronen führt in einem sehr weiten Bereich verschiedener Bedingungen zu guten Selektionsergebnissen. Für den Einsatz in größeren Entfernungsbereichen ist eine feinere Grundaufklärung erforderlich, was den Bedarf an Rechenleistung allerdings erheblich steigert und den Einsatz von Spezialhardware rechtfertigt.

⁴Zur Orientierung in den Abbildungen 8-11: In jeder Zeile sind Ergebnisse für eine der fünf im Abstand $1/\sqrt{2}$ berechneten Auflösungsstufen dargestellt, von oben nach unten 96×71 , 68×50 , 48×36 , 34×25 und 23×18 Pixel, d. h. die hier genutzte feinste Auflösung ist bereits durch achtfaches Unterabtasten des Kamerabildes im Format 768×572 entstanden. In den sieben Spalten sind von links nach rechts dargestellt: Input, Ergebnisse der Orientierungsfilterung (0° , 45° , 90° und 135°), Ergebnis der Filterung mit dem Gaborfilterarrangement, Ergebnis der Selektion im dreidimensionalen Feld dynamischer Neuronen.

⁵Unterschreitet die Filterresponse der dominanten Orientierung eine Schwelle, wird keine Orientierung ausgegeben.



Abbildung 8: Lokalisationsergebnisse im Innenbereich, Teil 1: Die Lokalisation der Person erfolgt nicht scharf auf einer der Pyramidenebenen, der entstehende räumliche Blob ist am stärksten auf der mittelsten der fünf Ebenen ausgebildet. Neben dem dominanten Blob existieren weitere kleine Blobs auf der feinsten Auflösungsebene, die lokal durch hochfrequente ungerichtete visuelle Strukturen des Gangendes entstanden sind. Es ist zu schlußfolgern daß der WTA-Prozesses zu zeitig abgebrochen wurde. Weitere Erläuterungen werden im Text gegeben.

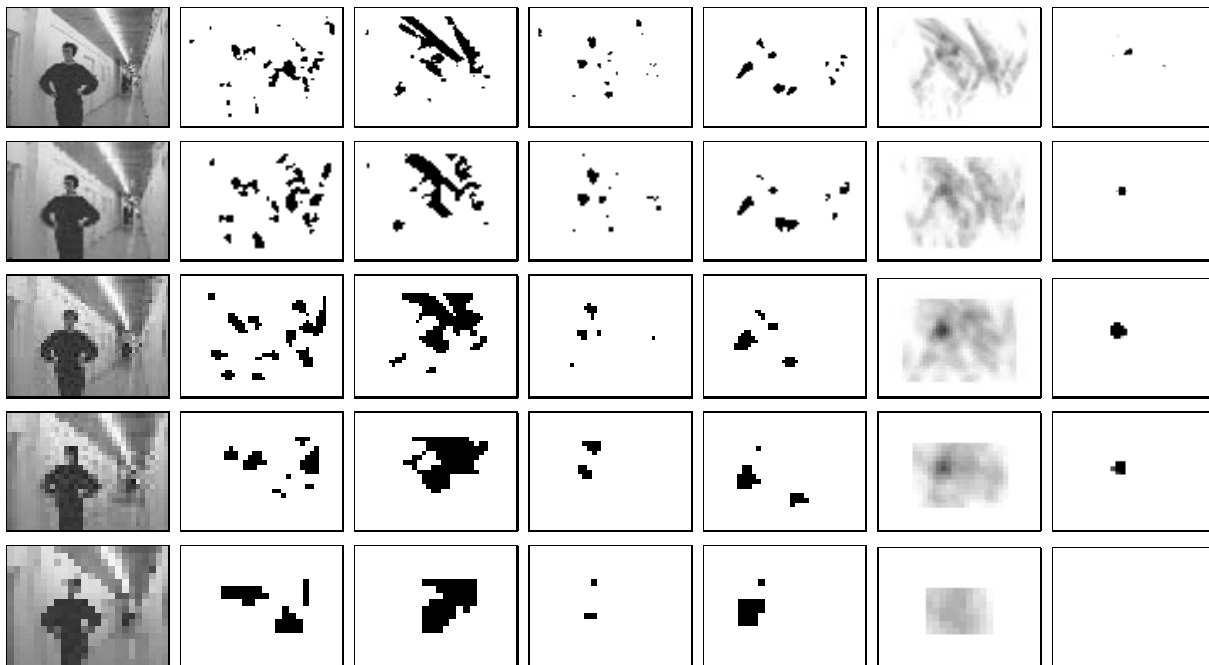


Abbildung 9: Lokalisationsergebnisse im Innenbereich, Teil 2: Im Unterschied zu Abbildung 8 ist die Entfernung der Person reduziert. Der Blob an aktiven Neuronen (siehe rechte Spalte) hat ist zwar weiterhin am stärksten auf der mittleren Ebene der Pyramide abgebildet, aber sein Schwerpunkt hat sich weiter hin zu gröberen Auflösungsebenen (in der Darstellung nach unten) verschoben. Die Entfernungsänderung fiel dabei offenbar geringer als ein $1/\sqrt{2}$ -Schritt aus (ca. Reduzierung auf das 0.7-fache), da die größte Blobausprägung sich nicht bis auf die nächstgrößere Ebene verlagert hat.

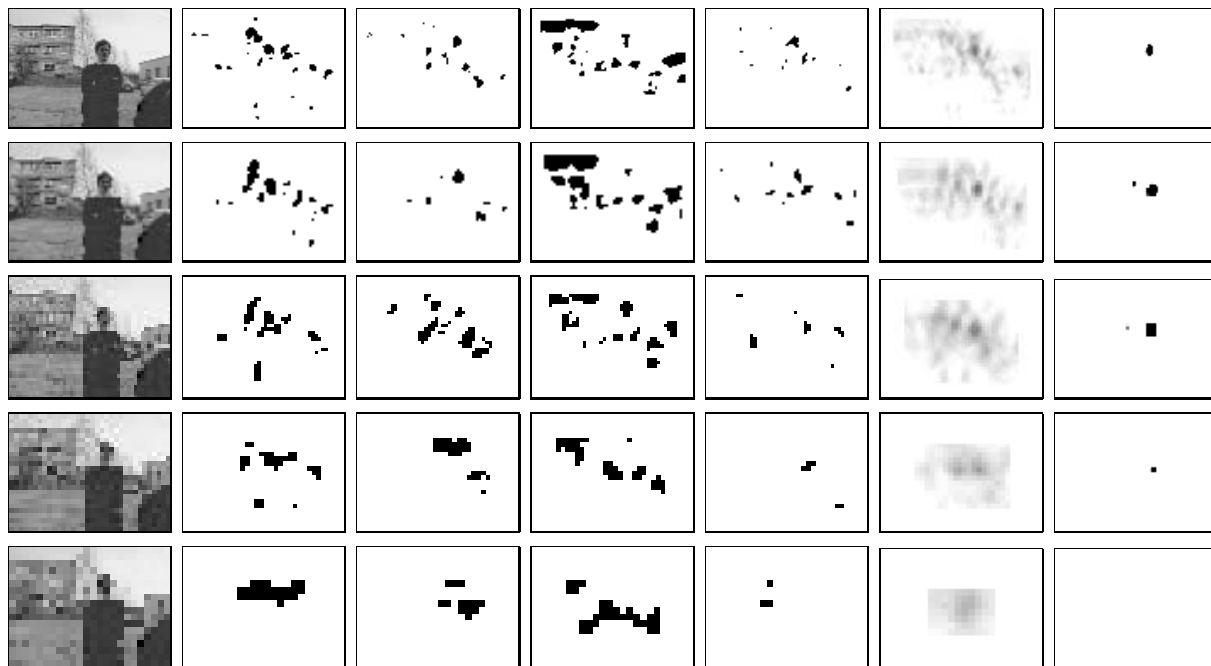


Abbildung 10: Lokalisationsergebnisse im Außenbereich, Teil 1: Diese Ergebnisdarstellung ist analog zu Abbildung 8.

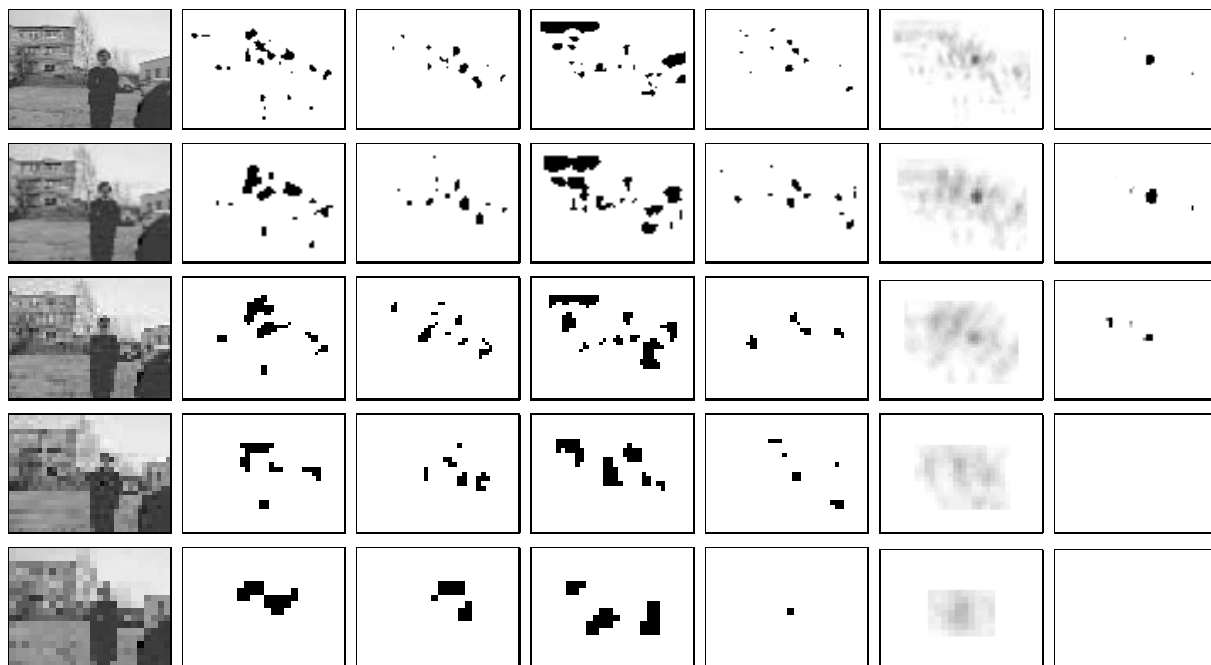


Abbildung 11: Lokalisationsergebnisse im Außenbereich, Teil 2: Die Entfernung der Person zur Kamera ist vergrößert worden, so daß sich der Blob im Vergleich zu Abbildung 10 in Richtung nächstfeinere Ebene (nach oben) verschoben hat.

Danksagung

Die Autoren danken Heiko Kempe vom Zentrum für Bild- und Signalverarbeitung e. V. Ilmenau für seine praktischen und theoretischen Anregungen sowie Carsten Rieck für seine implementationsseitige Unterstützung.

Literatur

- [1] Amari, Shun-ichi, 1977, Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. *Biological Cybernetics*, **27**, 77-87
- [2] Boehme, Hans-Joachim; Brakensiek, Anja; Braumann, Ulf-Dietrich; Krabbes, Markus und Gross, Horst-Michael, 1997, Neural Architecture for Gesture-Based Human-Machine-Interaction, *Proceedings of the Bielefeld Gesture Workshop*, Bielefeld, BR Deutschland, September, erscheint Anfang 1998
- [3] Brakensiek, Anja; Braumann, Ulf-Dietrich; Böhme, Hans-Joachim; Rieck, Carsten und Gross, Horst-Michael, 1997, Farb- und strukturbasierte neuronale Verfahren zur Lokalisierung von Gesichtern in Real-World-Szenen, *Mustererkennung 1997 – Tagungsband des 19. DAGM-Symposiums*, Braunschweig, BR Deutschland, September, 113-120
- [4] Brakensiek, Anja; Braumann, Ulf-Dietrich; Corradini, Andrea; Böhme, Hans-Joachim und Gross, Horst-Michael, 1998, Neuronale Verfahren zur Lokalisation und Bewertung von Handgesten in Real-World-Szenen, *dieser Band*
- [5] Gabor, Dennis, 1946, Theory of Communication. *Journal of the Institute of Electrical Engineers – Part 3: Radio and Communication Engineering*, **93**, 429-457
- [6] Hamker, Fred H. und Gross, Horst-Michael, 1997, Object Selection with Dynamic Neural Maps, *ICANN'97 – Proceedings of the 7th International Conference on Artificial Neural Networks*, Lausanne, Schweiz, Oktober, 919-924
- [7] Jähne, Bernd, 1993, *Digitale Bildverarbeitung*, 3. Auflage, Springer-Verlag, Berlin – Heidelberg – New York
- [8] Jolion, Jean-Michel und Rosenfeld, Azriel, 1994, *A Pyramid Framework for Early Vision: Multiresolutional Computer Vision*, Kluwer Academic Publishers, Boston
- [9] Jones, Judson P. und Palmer, Larry A., 1987, An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex. *Journal of Neurophysiology*, **58**, 1233-1258
- [10] Kaski, Samuel und Kohonen, Teuvo, 1994, Winner-Take-All Networks for Physiological Models of Competitive Learning. *Neural Networks*, **7**, 973-984
- [11] Kopecz, Klaus und Schöner, Gregor, 1995, Saccadic Motor Planning by Integrating Visual Information and Pre-Information on Neural Dynamic Fields. *Biological Cybernetics*, **73**, 49-60
- [12] Krabbes, Markus; Böhme, Hans-Joachim; Stephan, Volker und Gross, Horst-Michael, 1997, Handlungsorganisation intentionaler neuronaler Agenten, *SOAVE'97 – Tagungsband des Workshops „Selbstorganisation von Adaptivem Verhalten“*, Ilmenau, BR Deutschland, September, 35-44
- [13] Pentland, Alex P., 1996, Smart Rooms. *Scientific American*, **274**, 54-62
- [14] Pollen, Daniel A. und Ronner, Steven F., 1981, Phase Relationships Between Adjacent Simple Cells in the Visual Cortex. *Science*, **212**, 1409-1411
- [15] Seiler, Torsten; Stephan, Volker; Heinze, Andrea und Gross, Horst-Michael, 1997, Handlungsauswahl durch Antizipation sensomotorischer Konsequenzen, *SOAVE'97 – Tagungsband des Workshops „Selbstorganisation von Adaptivem Verhalten“*, Ilmenau, BR Deutschland, September, 109-118
- [16] Stephan, Volker und Gross, Horst-Michael, 1997, Formerhaltende sequentielle visuelle Aufmerksamkeit in columnar organisierten neuronalen Feldern, *Mustererkennung 1997 – Tagungsband des 19. DAGM-Symposiums*, Braunschweig, BR Deutschland, September, 411-418
- [17] Worth, Andrew J.; Lehar, Steve und Kennedy, David N., 1992, A Recurrent Cooperative/Competitive Field for Segmentation of Magnetic Resonance Brain Images. *IEEE Transactions on Knowledge and Data Engineering*, **4**, 156-161