Konturbasierte Personenlokalisation mittels dreidimensionaler neuronaler Felder und steuerbarer Filter

Ulf-Dietrich Braumann, Andrea Corradini*, Hans-Joachim Böhme, Horst-Michael Groß
Fachgebiet Neuroinformatik, Technische Universität Ilmenau, PF 10 05 65, 98684 Ilmenau
http://cortex.informatik.tu-ilmenau.de/

*gefördert durch ein TMR-Stipendium der Europäischen Union: Marie Curie Research Training Grant # ERB FMBI CT 97 2613

Präsentiert auf dem 20. DAGM-Symposium Mustererkennung 1998, 29. September - 1. Oktober 1998, Stuttgart

Zusammenfassung

Es wird ein Verfahren zur **Lokalisation von Personen** innerhalb unpräparierter visueller Szenen anhand typischer äußerer Konturen der **Kopf-Schulter-Partie** in frontaler Ausrichtung vorgestellt. Diese Konturen werden durch ein **räumlich verteiltes Arrangement steuerbarer Filter approximiert**, das auf einer Anzahl von pyramidal abgestuften Auflösungen eines Bildes angewendet wird. Durch den Bezug auf eine typische Kontur steht ein visuelles Merkmal zur Verfügung, das sich im Auftreten als hinreichend invariant erweist und ohne besondere Details auskommt.
Unter den Bedingungen realer Umwelten führt der Einsatz dieses Filterarrangements allerdings häufig zu einem unscharfen und teilweise artefaktbehafteten Ergebnis, so daß ein Kriterium zur Selektion als Bedingung für eine gute Lokalisationsleistung benötigt wird. Daher werden zur Bildung einer optimalen Lokalisationsentscheidung **dreidimensionale neuronale Felder** verwendet, innerhalb derer eine **dynamische Musterbildung** (AMARI 1977) mit dem Ziel der **Selektion einer einzelnen Person** erfolgt. Mittels einfacher homogener feldinterner Wechselwirkungsmechanismen können robuste Lokalisationsentscheidungen (Richtung und Entfernung) gefunden werden. Die Aktivitäten der Feldneuronen können als interne Zustände angesehen werden, die eine permanente Lokalisation (und damit eine visuelle Verfolgung der Person) wesentlich vereinfachen.
Diese Arbeit ordnet sich thematisch in ein **Gesamtkonzept zur visuell basierten Mensch-Maschine-Interaktion** ein und schafft dafür notwendige Voraussetzungen. Die Zielplattform ist der autonome mobile Roboter **MILVA**, der als autonomer Transportassistent eingesetzt und dabei anhand eines bestimmten Gestenalphabets durch Nutzer manövriert werden soll.

Motivation und Hintergründe

Der hier vorgestellte Ansatz zur Personenlokalisation beruht auf einer Beschreibung der äußeren Kontur von Kopf und Schultern. Unabhängig davon haben OREN et al. 1997 einen interessanten verwandten Ansatz entworfen, der sich unter Verwendung sog. **Haar Wavelet Templates** auf den kompletten Umriß von Personen (Fußgänger) bezieht. Der gemeinsame Aspekt besteht darin, daß räumlich verteilte orientierte lokale Filteroperatoren benutzt werden. Interessante Bezüge bestehen auch zur Arbeit von MELLER 1997, der u. a. mittels verschiedener **gitterartig angeordneter Gaborfilter** bestimmte Teilmerkmale auf Personen extrahiert, deren lokale Relationen zur Personenlokalisation verwendet werden.
Für einen konturbasierten Ansatz zur Bildung eines Auffälligkeitsmaßes sei auf **physiologische Fakten** verwiesen. Im visuellen Kortex von Säugern existieren in den verschiedenen Arealen Zellen, die **orientierte rezeptive Felder** (RF) aufweisen und mittels derer eine Dekomposition des visuellen Inputs erfolgen kann. Zur Modellierung solcher RF erweisen sich neben Gaborfunktionen (JONES 1987) auch Gaußfunktionen (einschl. Ableitungen niederer Ordnung, s. KOENDERINK 1990 oder YOUNG 1991) als geeignet. Den von uns verwendeten steuerbaren Filtern liegen die **zweiten Richtungsableitungen von Gaußfunktionen** zugrunde (FREEMAN 1991). Alle diese Modelle von RF sind zur Extraktion entsprechend orientierter Konturkomponenten geeignet.
Psychophysische Aspekte unseres Ansatzes sind bspw. **gute Fortsetzung** und **Symmetrie**, beides Bestandteile der Gestaltsgesetze (KOFFKA 1935).
Zur Modellierung einer bestimmten Kontur durch stückweise Approximation mittels eines **Arrangements orientierter Filter** wird u. U. eine **Vielzahl unterschiedlicher Orientierungen** benötigt. Da das bei diskreter Realisierung sehr rechenaufwendig wäre, verwenden wir **steuerbare Filter**.

Literatur

AMARI, SHUN-ICHI: *Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields.* Biological Cybernetics 27, Seiten 77-87, 1977

FREEMAN, WILLIAM T. und ADELSON, EDWARD H.: *The Design and Use of Steerable Filters.* IEEE Transactions on Pattern Analysis and Machine Intelligence 13, Seiten 891-906, 1991

JONES, JUDSON P. und PALMER, LARRY A.: *An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex.* Journal of Neurophysiology 58, Seiten 1233-1258, 1987

KOENDERINK, JAN J. und DOORN, ANDREA J.: *Receptive Field Families.* Biological Cybernetics 63, Seiten 291-297, 1990

KOFFKA, KURT: *Principles of the Gestalt Psychology.* Harcourt, Brace & World, New York, 1935

MELLER, SEBASTIAN: *Segmentierungsfreie Personendetektion.* Diplomarbeit, Friedrich-Alexander-Universität Erlangen-Nürnberg, 1997

OREN, MICHAEL; PAPAGEORGIOU, CONSTANTINE; SINHA, PAWAN; OSUNA, EDGAR und POGGIO, TOMASO: *Pedestrian Detection Using Wavelet Templates.* Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97), Seiten 193-199, 1997

YOUNG, RICHARD A.: *Oh Say, Can You See? The Physiology of Vision.* TR GMR-7364, General Motors Research Laboratories, Warren (MI), 1991

Danksagung Die Autoren danken Markus Krabbes, Anja Brakensiek und Heiko Kempe für Anregungen und Diskussionen.

Steuerbare Filter

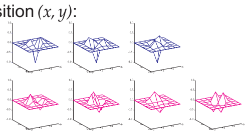
Beim Einsatz von steuerbaren Filtern wird ein Satz von M kopierten, jedoch verschieden orientierten Basisfiltern f^{α} eingesetzt, anhand derer eine Interpolation eines Filterergebnisses einer beliebigen Orientierung erfolgen kann, indem mit Hilfe von Interpolationsfunktionen eine **Linearkombination aus den Basisfiltern** gebildet wird: $f^{\theta}(\cdot) = \sum_{j=1}^{M} I^{(\alpha_j)}(\theta) f^{\alpha_j}(\cdot)$

Aus Gründen der Phasenlageunabhängigkeit wird damit ein **Quadraturpaar** aus zweiter Richtungsableitung der Gaußfunktion und einer Approximation der Hilberttransformierten verwendet, dessen orientierte Energie maximiert wird: $E^{(x,y)}(\theta) = \left(G^{(x,y)}(\theta)\right)^2 + \left(H^{(x,y)}(\theta)\right)^2$

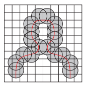
Dafür wird eine vereinfachte Bildungsvorschrift verwendet, wobei die Koeffizienten c_1 und b_1 aus Linearkombinationen von Produkten der G^{α} und H^{α} gebildet werden: $E^{(x,y)}(\theta) \approx a_0 + b_1 \sin(2\theta) + c_1 \cos(2\theta)$
Über die Nullstellen von $dE^{(x,y)}(\theta)/d\theta$ wird der Winkel θ_{MAX} **analytisch** gefunden, für den die **orientierte Energie E maximal** wird. Dieser Winkel steht für die **dominante Orientierung** an der Position (x,y):

$$\theta_{MAX}^{(x,y)} = \arg\max \left\{ E(\theta), \quad \forall\theta \quad \middle| \quad \theta \in \left\{ \pm\frac{1}{2}\arccos\left(\pm c_1 / \sqrt{b_1^2 + c_1^2}\right) \right\} \right\}$$

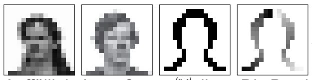
Zur Veranschaulichung sind die sieben verwendeten Basisfilter (drei für die Grundfunktion und vier für deren Hilbertransformierte, hier mit einer Wellenzahl von $k=0.22$) dargestellt:



Filterarrangements



Um mittels der steuerbaren Filter die zu detektierende Kontur von Kopf-Schulter-Partien bestimmen zu können (siehe Prinzipskizze rechts), wurde der **typische Verlauf** solcher Konturen anhand eines Datensatzes von *60* Portraitaufnahmen (siehe die zwei links dargestellte Beispiele) ermittelt. Es erfolgte eine Kantenfilterung mit einem Sobeloperator und eine anschließende Mittelung über diese Kantenbilder. Die Orientierung



entlang der als Binärbild vorliegenden **mittleren Kontur** wurde mit den steuerbaren Filtern bestimmt (siehe rechtes Bild, grauwertkodierte Orientierungswinkel). Damit liegt ein Template Λ für ein Filterarrangement vor, das als Grundlage für die Bildung eines konturbasierten Auffälligkeitsmaßes $m^{(x,y)}$ dient. Die Bestimmung dieses Maßes für einen Bildausschnitt (Matrix Θ) erfolgt unter Berücksichtigung der π-Periodizität des Ergebnisses der Orientierungsfilterung:

$$m^{(x,y)} = \frac{\sum_{\substack{i=0 \\ \lambda_{i,j}\neq 0}}^{I-1} \sum_{j=0}^{J-1} m^{*}\left(\lambda_{i,j} - \theta_{MAX}^{(x+i-\frac{I}{2},\, y+j-\frac{J}{2})}\right)}{\operatorname{card}\left(\operatorname{supp}(\Lambda)\right)}$$

Darin ist $I=J=16$ die Größe und $\lambda_{i,j}$ ein Element des Templates Λ und $\theta^{(\cdot)}_{MAX}$ die an der korrespond. Bildkoordinate bestimmte dominante Orientierung. Zwischen beiden wird ein elementw. Maß $m^*(\gamma)$ bestimmt: $m^*(\gamma) = \frac{\cos(2|\gamma|)+1}{2}$

Die Kardinalität des Supports der Matrix Λ entspricht genau der Anzahl ihrer besetzten Elemente. Die Normierung sichert $m^{(x,y)}$ für die folgende dynamische Selektion den Wertebereich *[0,1]*.

3D Dynamische Felder

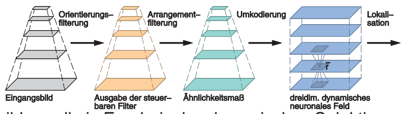
Um eine akzeptable Lokalisation von Personen zu erzielen, reicht jedoch die Bildung dieses Ähnlichkeitsmaßes allein nicht aus. Dafür bedarf es eines zusätzlichen **Selektionsmechanismus**. Dynamische neuronale Felder eignen sich gut für eine dynamische Selektion. Da hier insgesamt **fünf Auflösungsebenen** (Halboktavabstand) verwendet werden, um bei der Lokalisation auch die Entfernung einer Person abschätzen zu können, muß das dynamische Feld zur Selektion entsprechend auch **dreidimensional** sein. Ziel der feldinternen Selektion (dyn. Musterbildung) ist die **Herausbildung eines einzelnen Blobs aktiver Neuronen** in einem Gleichgewichtszustand. Ein dyn. Feld ist charakterisiert durch:

- ❍ einfache homogene feldinterne Wechselwirkungsmechanismen
- ❍ interne Zustandsspeicherung in den Neuronenaktivitäten, das vorherige Selektionserg. fließt ständig mit ein
- ❍ Beschreibung als rückgekoppeltes nichtlineares dynamisches System erster Ordnung:

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

Darin wird als x der externe Input bezeichnet, der hier konkret dem umkodierten (s.u.) Ähnlichkeitsmaß m entspricht; die Konstanten c_h, c_l und c_i sind Parameter des Systems. Die Funktion w ist die Mexikanerhut-ähnliche laterale Kopplung: $w(\vec{r}-\vec{r}') = 2\exp\left(\frac{-3|\vec{r}-\vec{r}'|^2}{2\sigma^2}\right) - \exp\left(\frac{-|\vec{r}-\vec{r}'|^2}{\sigma^2}\right)$ und y die sigmoidale Ausgabefunktion eines Feldneurons: $y(\vec{r},t) = \left(1+\exp(-z(\vec{r},t))\right)^{-1}$ Zusätzlich erfolgt eine globale Inhibition durch ein hemmendes lineares Interneuron: $h(t) = \int_{\mathbf{F}} y(\vec{r}'',t)\, d^3\vec{r}''$

Um die Korrespondenzen zwischen den Ebenen nicht zu verletzen, erfolgt eine Umkodierung von m in eine quaderförmige Struktur. Nebenstehend ist der gesamte Ablauf der Verarbeitung skizziert:



Ergebnisse

Anhand eines typischen Beispielbildes soll ein Ergebnis des dynamischen Selektionsprozesses zur Personenlokalisation dargestellt werden. Von links nach rechts sind in den sieben Spalten in jeweils fünf Auflösungsstufen folgende Teilergebnisse zu sehen:

- ❍ Grauwertbild
- ❍ domin. Orient. $(-22.5^{\circ}..+22.5^{\circ}]$
- ❍ domin. Orient. $(+22.5^{\circ}..+67.5^{\circ}]$
- ❍ dom. Orient. $(+67.5^{\circ}..+112.5^{\circ}]$
- ❍ dom. Orient. $(+112.5^{\circ}..+157.5^{\circ}]$
- ❍ Ähnlichkeitsmaß m
- ❍ Selektionsergebnis

Da der Blob aktiver Neuronen sich auf der Ebene feinster Auflösung konzentriert, ist für die im Beispiel vorgenommene Dimensionierung bereits die größtmögliche Lokalisationsdistanz erreicht. Im Mittel werden zur Relaxation des dynamischen Systems (Dim. $16\times16\times5$ Elemente) *11* Iterationsschritte (Eulerverfahren) benötigt. Der Rechenzeitbedarf dafür liegt bei ca. 0.2s auf einem Pentium 133MHz.