Anwendung eines DCS-Netzwerkes zur Verifikation von charakteristischen Bildpunkten*

L. Witta[1], R. Herpers[2], J. Bruske[3] und G. Sommer[3]

[1]Lehrstuhl für Mensch-Maschine-Kommunikation, TU-München,
80290 München, E-mail: wta@mmk.e-technik.tu-muenchen.de
[2]GSF-Institut für Medizinische Informatik und Systemforschung, Neuherberg,
85764 Oberschleißheim, E-mail: herpers@gsf.de
[3]Institut für Informatik, Christian-Albrechts-Universität, 24105 Kiel

In diesem Beitrag wird eine Anwendung aus der Gesichtsbildverarbeitung vorgestellt, die auf der Basis von Dynamischen Zell-Strukturen (DCS) lokale Bildstrukturen an bestimmten markanten Punkten, wie z.B. Augenecken, klassifiziert. Die markanten Punkte werden im Voraus durch eine modellgetriebene sequentielle Suchstrategie detektiert. Um den Detektionsfehler durch die sequentielle Suchstrategie zu minimieren, werden die berechneten Bildpositionen durch das DCS verifiziert. Die Merkmalsextraktion für die modellgetriebene Suche, als auch für die datengetriebene Verifikation basieren auf dem Konzept der steuerbaren Filter. Durch die Anwendung des DCS-Netzwerkes sinkt in unseren Experimenten die durchschnittliche Wahrscheinlichkeit einer falsch positiven Verifikation auf weniger als 1%. Die entwickelten Verfahren werden am Beispiel der markanten Punkte in der Augenregion demonstriert.

Schlüsselwörter: Dynamische Zell-Strukturen, RBF-Netzwerke, Computer Vision

1 Einleitung

Charakteristische Bildpunkte in Gesichtsbildern, wie Augen– und Mundecken, stellen bedeutende Merkmale für eine Vielzahl von Anwendungen im Bereich der automatischen Gesichtserkennung dar [1]. Das grundlegende Problem bei einer automatischen Detektion ist, daß solche charakteristischen Punkte eher durch ihre semantische Bedeutung (Ecke des Mundes oder Auges) als durch eine einheitliche lokale Struktur in den Bilddaten definiert sind.

In vorausgegangenen Arbeiten haben wir eine sequentielle Suchstrategie entwickelt, welche automatisch charakteristische Punkte in Gesichtsbildausschnitten detektiert [2, 3]. Die Detektion dieser Punkte basiert dabei auf einer sequentiellen Auswertung der Kanten- und Linieninformation. Durch Integration von Modellwissen der untersuchten Bildregion ist eine robuste und sehr genaue Lokalisierung der gesuchten Punkte möglich. Das Ziel unserer Arbeiten ist die Ableitung von statistisch signifikanten Geometriemaßen aus Kindergesichtern, die eine Unterscheidung zwischen normalen Gesichtern und bestimmten Klassen von Dysmorphiesyndromen zulassen [4].

*Diese Arbeit wurde von der DFG unter den Aktenzeichen: So 320\1-1 und Ei 322\1-1 gefördert.

Durch Artefakte, Schatten oder andere unvorhergesehene Kantenelemente kann es allerdings vorkommen, daß das sequentielle Suchschema einen bestimmten markanten Punkt an einer falschen Stelle detektiert. Gerade im Zusammenhang mit der medizinischen Anwendung ist dies jedoch höchst unerwünscht. Es ist in diesem Fall besser, einen markanten Punkt nicht zu finden, als ihn ungenau oder an falscher Stelle zu lokalisieren.

Zur Reduktion von Fehldetektionen schließt sich daher an die Detektion eine Verifikation der gefundenen Punkte an. Dazu wird an den detektierten Punkten aus der lokalen Bildinformation ein Merkmalvektor generiert und durch ein DCS-Netzwerk klassifiziert. Ordnet das DCS-Netzwerk den Merkmalvektor des detektierten Bildpunktes der Klasse des gesuchten charakteristischen Punktes zu, so gilt der Punkt als gefunden. Nach dem Satz von Bayes läßt sich dadurch die Wahrscheinlichkeit einer Falschdetektion erheblich verringern.

2 Neuronale Verifikation

In diesem Kapitel wird das zur Verifikation verwendete DCS-Netzwerk vorgestellt. Außerdem werden die Grundlagen der Merkmalserzeugung, die Klassifikation und schließlich die Verknüpfung der Verifikationsstufe mit der Detektionsstufe über den Satz von Bayes beschrieben.

2.1 Grundlagen der DCS-Netzwerke

Die in [5] eingeführten *Dynamic Cell Structures* (DCS) gehören zu der Klasse der RBF-basierten Approximationsverfahren, von denen sie sich dadurch unterscheiden, daß sie zusätzlich die Topologie der Eingabe-Mannigfaltigkeit in Form von *Perfekt Topologie-erhaltenden Merkmalskarten* (PTM) erlernen [6]. Diese nutzen sie für verbesserte Approximationseigenschaften und Lernverfahren. DCS-Netzwerke weisen folgende Architektur auf:

- eine verdeckte Schicht von RBF Knoten (evtl. dynamisch wachsend oder schrumpfend),
- eine sich dynamisch an die Topologie anpassende laterale Verbindungsstruktur zwischen den Knoten und
- eine Schicht meist linearer Ausgabeeinheiten.

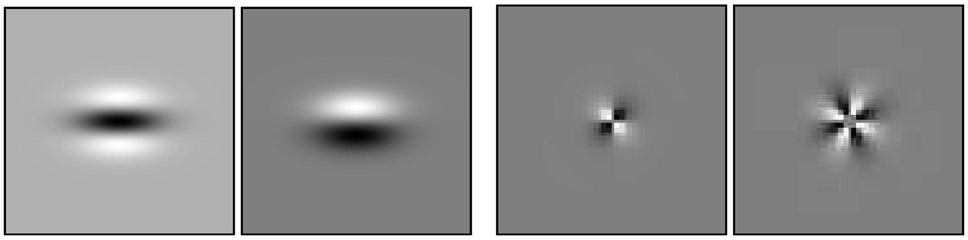
Die in diesem Artikel verwendeten *Growing Dynamic Cell Structures* (GDCS), die ähnlich zu den von Fritzke [7] durch fehlerbasiertes Einfügen neuer RBF Knoten wachsen, vermeiden damit die Unzulänglichkeiten der klassischen Kohonen-Karten: Die starre Topologie des Netzes (2- oder 3-dimensional), die vorgegebene Anzahl von Knoten und die rein von der Eingabewahrscheinlichkeitsdichte abhängige Verteilung der Knoten. Ihre Eleganz, Einfachheit und problemabhängige Struktur gaben den Ausschlag für die Verwendung des DCS-Netzwerkes als Verifikator, zumal sie sich in einer Reihe von Klassifikations-Benchmark-Tests herkömmlichen Klassifikatoren überlegen erwiesen [5].

2.2 Verwendete Merkmale

Der Merkmalvektor eines zu klassifizierenden Punktes wird durch Projektion der lokalen Bildumgebung des Punktes auf einen Satz von 70 orthogonalen Basisfunktionen gewonnen. Diese Basisfunktionen stammen aus einem bereits für die Detektion der markanten Punkte entwickelten Filterschema auf der Basis steuerbarer Filter [8]. Da die Projektionskoeffizienten bereits für die Detektion benötigt werden, müssen sie nicht erneut berechnet werden, was die Effizienz dieses Ansatzes erhöht.

In den Projektionskoeffizienten ist die wesentliche Kanten- und Linieninformationen der lokalen Bildstruktur, die auch für eine Verifikation notwendig ist, enthalten. Die Verwendung der Projektionskoeffizienten anstelle der direkten Bildinformation (originale Grauwertverteilung in der 27 x 27 Umgebung) ermöglicht darüber hinaus eine erhebliche Dimensionsreduzierung des Merkmalraumes, ohne auf die wesentliche Information der lokalen Bildstruktur verzichten zu müssen.

Die Grundlage für die verwendeten Basisfunktionen bilden die in Abbildung 1a,b dargestellten orientierungsselektiven Kanten- und Linienfilter. Diese Filter werden in Orientierung und Skala gesteuert. Der Begriff der Steuerung bezieht sich dabei auf die lineare Rekonstruktion der Filter aus einem festen und kleinen Satz von Basisfunktionen (Beispiele Abb.1c,d) mit geeignet gewählten Superpositionskoeffizienten.



(a) (b) (c) (d)

Abbildung 1: Linien- und Kantendetektionsfilter (a,b). Es handelt sich dabei um die erste (b) und zweite (a) Ableitung einer Gaußfunktion mit einem Seitenverhältnis von 2:1. Der Faltungskern hat eine Größe von 27x27 Pixeln. Beispiele für Basisfunktionen (c,d) zur Steuerung der Filter von (a,b) in Orientierung und Skala.

2.3 Klassifikation

In unserer Anwendung unterscheiden wir 8 Klassen charakteristischer Bildpunkte: Innerer und äußerer Augenwinkel (jeweils für das rechte und linke Auge getrennt) und 4 Schnittpunkte des Irisrandes mit den Lidern. Obwohl die meisten Pixel in einer Augenregion keine charakteristischen Punkte sind, wird eine Restklasse nicht explizit berücksichtigt. Entsprechend werden nur Merkmalvektoren von vorselektierten charakteristischen Bildpunkten für das Training des DCS-Netzwerks verwendet. Damit wird verhindert, daß das DCS-Netzwerk nicht die meisten Neurone zur Repräsentation einer Klasse verwendet, an der wir nicht interessiert sind. Stattdessen wird die Restklasse implizit durch einen relativen, klassenspezifischen Schwellwert definiert. Liegt die Netzaktivierung einer Bildposition bei der Klassifikation unterhalb dieses Schwellwertes, wird sie der Restklasse zugeordnet. Als Schwellwert dient die durchschnittliche Aktivität der jeweils zu einer Klasse gehörenden Bildpunkte des untersuchten Bildausschnitts.

2.4 Der Bayes'sche Verifikationsansatz

Die Detektion und die Verifikation eines charakteristischen Punktes der Klasse C_i können als hintereinander ausgeführte, unabhängige Zufallsexperimente betrachtet werden. Im Falle der sequentiellen Suche besteht das Zufallsexperiment X aus dem Finden der Punktkoordinaten des markanten Punktes, der gerade aufgesucht werden soll. Bei der Verifikation wird die Klasse des gerade betrachteten Bildpunktes bestimmt und der Zufallsvariablen Y zugewiesen. Die Wahrscheinlichkeit, daß eine Bildposition, an der die sequentielle Suche terminiert, der Klasse C_i angehört, ist somit die a priori Wahrscheinlichkeit $P(X = C_i)$.

Wenn nun Y die Klasse bezeichnet, die der betrachteten Bildposition durch die Verifikation zugeordnet wird, dann ist $P(Y = C_i|X \neq C_i)$ die bedingte Wahrscheinlichkeit für eine falsch positive Klassifikation und $P(Y \neq C_i|X = C_i)$ die bedingte Wahrscheinlichkeit für eine falsch negative Klassifikation. Nach dem Satz von Bayes läßt sich nun die a-posteriori-Fehlerwahrscheinlichkeit, daß eine detektierte Bildposition nicht der gesuchten Klasse angehört, obwohl die Verifikation positiv war, berechnen als:

$$P(X \neq C_i|Y = C_i) = \frac{P(Y = C_i|X \neq C_i)\ P(X \neq C_i)}{P(Y = C_i)} \tag{1}$$

$$= \frac{P(Y = C_i|X \neq C_i)\ P(X \neq C_i)}{P(Y = C_i|X \neq C_i)\ P(X \neq C_i) + P(Y = C_i|X = C_i)\ P(X = C_i)}$$

Eine Verifikation ist nur dann sinnvoll, wenn die a posteriori Fehlerwahrscheinlichkeit $P(X \neq C_i|Y = C_i)$ kleiner ist, als die a priori Fehlerwahrscheinlichkeit $P(X \neq C_i)$.

3 Ergebnisse

668 Merkmalvektoren wurden aus insgesamt 98 Augenregionen von rechten und linken Augen abgeleitet, davon bilden 527 Merkmalvektoren die Trainingsmenge und 141 Merkmalvektoren die Testmenge.

Der Klassifikationsfehler des Netzwerkes (gemittelt über alle Klassen) wird sowohl bezüglich der Trainingsmenge als auch bezüglich der Testmenge in Abhängigkeit zur Anzahl der Neurone, die während der Trainingsphase in das Netz eingefügt wurden, in Tabelle 1 dargestellt.

Netzwerkgröße						
Neurone	23	32	59	87	115	127
Trainingsmenge						
Fehler(%)	9.7	8.2	5.5	3.4	1.5	0
Testmenge						
Fehler(%)	8.5	7.8	7.0	6.3	5.6	4.9

Tabelle 1: Klassifikationsfehler bezüglich der Größe des DCS-Netzwerkes.

Tabelle 2 (letzte Spalte im Vergleich zur ersten) zeigt, daß die Rate der Falschdetektion durch die zusätzliche Anwendung des Verifikationssystems erheblich reduziert werden konnte. Dies gilt sowohl für die klassenspezifischen Fehlerwahrscheinlichkeiten als auch für die gesamte Fehlerwahrscheinlichkeit, die von 6.5% auf unter 1% gesunken ist. Eine anschauliche Erklärung dafür kann aus Abb. 2 entnommen werden. 2a und b zeigen die Klassen- und Aktivitätskarte für eine Augenregion ohne die Anwendung des Aktivitätsschwellwertes. Nach Einschränkung auf eine bestimmte Klasse und Anwendung des klassenspezifischen Schwellwertes (Abb.2c) gibt es nur noch sehr wenige Bildpositionen, die der gesuchten Klasse (hier: Schnittpunkt Iris/Lid) zugeordnet werden. Darunter ist auch der markante Punkt selbst, einige Punkte in seiner direkten Nachbarschaft und einige Punkte an weiter entfernten Stellen, deren Struktur der des markanten Punktes ähnelt. Nur an letzteren Punkten kann überhaupt eine falsch-positive Klassifikation auftreten (Tab.2). Eine direkte Detektion der markanten Punkte mit Hilfe des DCS-Netzwerkes (ohne das sequentielle Suchschema) ist zwar ebenfalls möglich, kommt dagegen über eine durchschnittliche Erfolgsquote von ca. 50% nicht hinaus [9].

Markanter Punkt	$\hat{P}(X \neq C_i)$	$\hat{P}(Y = C_i \mid X \neq C_i)$	$\hat{P}(Y \neq C_i \mid X = C_i)$	$\hat{P}(X \neq C_i \mid Y = C_i)$
innere Augenecke	14.4	6.7	2.2	1.1
äußere Augenecke	8.3	11.1	1.0	1.0
Schnittp.Iris/Oberlid	3.2	14.3	0.9	0.5
Schnittp.Iris/Unterlid	4.1	0.0	0.9	0.0
Gesamt	6.5	8.3	1.2	0.2

Tabelle 2: Empirische Fehlerwahrscheinlichkeiten in %

(a) (b) (c)

Abbildung 2: (a) Klassenkarte einer Augenregion ohne Schwellwert. Jeder Bildposition ist durch das DCS-Netzwerk einer der 8 Klassen (dargestellt als Graustufen) zugeordnet worden. (b) Zugehörige Aktivitätskarte. Bildstrukturen, die markanten Punkten ähneln, erzielen eine hohe Aktivierung (dunkle Bereiche) des Netzwerkes. Helle Bereiche bedeuten geringe Aktivierung. (c) Augenregion mit überlagerter Klasssenkarte nach der Schwellwertoperation. Die weiß eingefärbten Bereiche gehören der Klasse "Schnittpunkt Iris/Lid" an.

4 Schlußbemerkung

In diesem Beitrag wurde gezeigt, daß der Einsatz einer Verifikationsstufe erheblich zur Reduktion der Fehlerwahrscheinlichkeit der vorgeschalteten Detektionsstufe beitragen kann. Der zur Verifikation verwendete neurale Klassifikator auf der Basis von Dynamischen Zell-Strukturen bietet bei einer einfachen Architektur und geringem Trainingsaufwand sehr gute Klassifikationseigenschaften. Da die verwendeten Merkmale nur die lokale Umgebung der markanten Punkte beschreiben, und weder Modell- noch Kontextwissen in die Klassifikation einfließen, also rein datengetrieben ist, ist der Klassifikator für sich allein (wie alle anderen Verfahren, die diese Eigenschaften aufweisen) für eine Detektion der Punkte ungeeignet. Die Kombination jedoch von modellbasierter, sequentieller Suche mit einer datengetriebenen Verifikation übertrifft die Leistung der beiden Einzelteile.

Danksagung

Wir möchten uns bei Herrn Dr. M. Michaelis für die Bereitstellung des Steuerbarkeitsschemas und die vielen hilfreichen Diskussionen bedanken. Ebenso möchten wir uns bei Frau Prof. S. Stengel-Rutkowski vom Kinderzentrum München für die bereitwillige Unterstützung der Arbeiten an dem Projekt bedanken. Diese Arbeit wurde von der DFG unter den Aktenzeichen So 320\1-1 und Ei 322\1-1 gefördert.

Literatur

[1] A. Samal and P.A. Iyengar, *Automatic recognition and analysis of human faces and facial expressions: A survey*, Pattern Recognition, Vol. 25, 65-77, 1992.

[2] L. Witta, R. Herpers, M. Michaelis, G. Sommer und S. Stengel-Rutkowski, *Computergestützte Detektion charakteristischer Bildpunkte in dysmorphen Kindergesichtsbildern zur Unterstützung der Phänotypdiagnose*, in diesem Tagungsband Aachener Workshop "Bildverarbeitung für die Medizin", 1996.

[3] R. Herpers, M. Michaelis, L. Witta and G. Sommer, *Detection of keypoints in face images*, Tech. Report 24/95, GSF–Research Center, Germany, 1995.

[4] S. Stengel-Rutkowski und P. Schimanek, *Chromosomale und nicht-chromoso-male Dysmorphiesyndrome*, Enke Verlag, Stuttgart, 1985.

[5] J. Bruske and G. Sommer, *Dynamic cell structure learns perfectly topology preserving map*, Neural Computation, Vol. 7, no. 4, pp. 845-865, 1995.

[6] T. Martinetz, *Competitive Hebbian learning rule forms perfectly topology preserving maps*, Proc. ICANN 93, pp. 426-438, 1993.

[7] B. Fritzke, *Growing cell structures - a self organizing network for unsupervised and supervised training*, ICSI Berkeley, Tech.-Rep., tr-93-026, 1993.

[8] M. Michaelis, *Low level image processing using steerable filters*, PhD thesis, Christian-Albrechts-Universität, D-24105 Kiel, Germany, 1995.

[9] R. Herpers, L. Witta, J. Bruske and G. Sommer, *Evaluation of local image structures applying a DCS network*, in: Solving Engineering Problems with Neural networks, Proc. 2nd int. Conf. EANN'96, A.B. Bulsari et al. (Eds.), 305-312, 1996.