

Fusion textur- und konturbasierter Methoden zur Segmentierung

Uwe Handmann*

Zusammenfassung

In der vorliegenden Arbeit wird ein Verfahren vorgestellt, welches textur- und konturbasierte Verfahren zur Segmentierung fusioniert. Als Kopplungsmatrix wird eine selbstorganisierte Karte nach Kohonen verwendet. Eine verbesserte Objekt- zu Hintergrundtrennung im Vergleich zu Einzelalgorithmen wird demonstriert.

1 Einführung

In der Bildverarbeitung werden mit Hilfe verschiedener Algorithmen Grauwertbilder abhängig von der Aufgabenstellung bearbeitet. Diese Algorithmen sind meist spezialisiert, sie haben einen direkten Problembezug.

In wissensbasierten Bildanalyse-Systemen ist es deshalb notwendig, datenbasierte und modellbasierte Methoden der Bildverarbeitung auf unterschiedlichen Verarbeitungsstufen zur Steigerung der Verallgemeinerbarkeit und Effizienz zu koppeln. Diese Kopplung muß aufgabenabhängig sequenziell, parallel oder rückgekoppelt durchgeführt werden.

Die vorliegende Arbeit beschreibt einen Ansatz verschiedene datenbasierte Algorithmen zu koppeln, um einerseits ein verallgemeinertes Bildanalyseverfahren und andererseits eine Steigerung der Robustheit und Leistungsfähigkeit gegenüber einzelnen Algorithmen zu erhalten.

Im Gegensatz zu dem von DRAPER ET AL. in [2] vorgestellten Verfahren *VISIONS*, wird beim vorliegenden Ansatz die Kopplungsstruktur als zentrales Element eines Bildanalyse-Systems betrachtet. Es wird die Kopplung der Verarbeitungsmodule gelernt. In [2] ist bei festgelegter Kopplung die Kontrolle über den Einfluß einzelner Verfahren auf das System in den Mittelpunkt gestellt. DRAPER ET AL. beschreiben nach einer allgemeinen Einführung in die Bildverarbeitung Kontrollmechanismen für Bildverarbeitungsmodulare. Neben Ansätzen basierend auf Bayes-Netzwerken und Entscheidungsbäumen präferieren sie ein Kontrollsystem der Bildverarbeitungsmodulare mit Hilfe von Markov-Entscheidungsprozessen.

Einen Überblick über andere Fusionsansätze gibt VARSHNEY in *The Special Issue On Data Fusion* [9].

In der vorliegenden Arbeit wird ein Kopplungsmodell vorgestellt und eine Integration mit Hilfe einer *Kohonenkarte* besprochen. Ein Überblick über die verwendete Vorverarbeitung wird gegeben und der Fusionsprozess beschrieben. Experimente zur Objekt- zu Hintergrundtrennung werden vorgestellt und abschließend diskutiert.

2 Modell

Der vorliegende Lösungsansatz beschreibt die Fusion pixelorientierter Daten. Es wird eine Objekt- zu Hintergrundtrennung durchgeführt. In einem Vorverarbeitungsschritt werden Kontur- und Texturinformationen aus dem Grauwertbild berechnet und in den Verarbeitungsprozess eingekoppelt. Es werden die Merkmale der betrachteten Vorverarbeitungsverfahren extrahiert und über einen *Feature In - Decision Out* (DASARATHY [1]) Kopplungsprozess fusioniert. Hierbei sind die Merkmale in der Art gekoppelt, daß eine Bildsegmentierung durchgeführt werden kann.

Eine Prinzip-Darstellung ist in Abbildung 1 gegeben. Vorverarbeitende Bildverarbeitungsmodulare erhalten über eine Kamera K ein Grauwertbild der Umwelt. Die Ergebnisse der Vorverarbeitung werden über

*Institut für Neuroinformatik, Ruhr Universität Bochum, 44780 Bochum

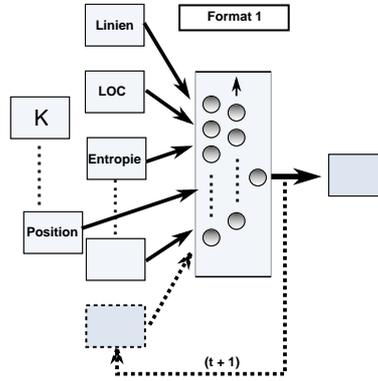


Abbildung 1: Kopplungsmodell

eine Kopplungsmatrix zur Segmentierung (*Format 1*) fusioniert. Die Kopplungsstruktur wird mit Hilfe einer *Kohonenkarte* (KOHONEN [5]) trainiert. Eine Rückkopplung des Segmentierungsergebnisses über der Zeit wird zum gegenwärtigen Zeitpunkt nicht betrachtet.

3 Extraktion der Texturinformation und Kanten

Für die Vorverarbeitung werden Grauwertbilder einer Autobahnszene herangezogen. Es werden Verfahren ausgewählt, die sowohl Informationen über die Kontur sowie über die Grauwertstatistik geben. Die datenbasierten Module der *Entropieberechnung*, die *Local Orientation Coding (LOC)*, sowie das *Linienbild* werden verwendet (siehe Abbildung 2).

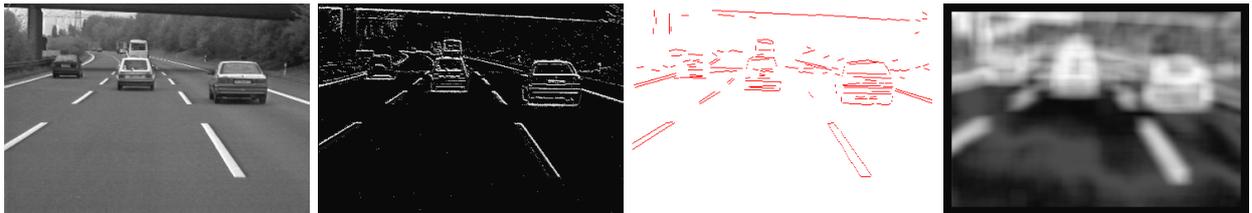


Abbildung 2: Verfahren: Grauwert, LOC, Linien, Entropie

LOC (GOERICK [3]) ist ein mehrstufiges Verfahren, welches eine Kodierung lokaler Grauwertdifferenzen durchführt. Diese Kodierung ist abhängig vom Verlauf der Kontur. Es wird die 4er oder 8er-Nachbarschaft eines Pixels betrachtet. Bild 3 beschreibt die Additonsterme der 3×3 Kodematrix $r(j)$, welche nach Betrachtung der Kontur einer 4er-Nachbarschaft in der Summe den *LOC*-Kode $R(x_i)$ des Zentralpixels x_i ergeben. In Gleichung (1) wird dieser Zusammenhang verdeutlicht. Für $q(j) = 0$ ist das betrachtete Pixel Konturpixel, für $q(j) = 1$ ist das betrachtete Pixel kein Konturpixel:

$$R(x_i) = \sum_j r(j) \cdot q(j), \quad j \in \{1, 2, \dots, 9\}, \quad q(j) \in \{0, 1\}. \quad (1)$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 2 & R(x_i) & 4 \\ 0 & 8 & 0 \end{bmatrix}$$

Abbildung 3: Kode-Matrix $r(j)$ für 4er Nachbarschaft

Wie in Abbildung 4 an Beispielen dargestellt, wird dem zentralen Pixel ein Kode R zugeordnet, welcher eindeutig den Verlauf der Kontur beschreibt. Mögliche Kodes sind $\{0, 1, \dots, 15\}$.

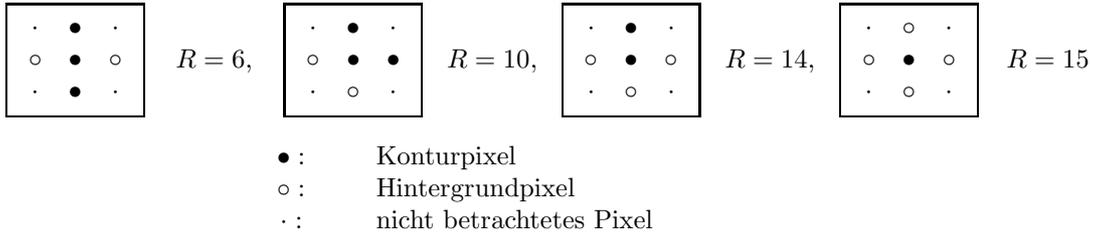


Abbildung 4: *LOC* (Beispiel: 4er-Nachbarschaft)

Das Liniensbild wird mit Hilfe eines Sobelfilterkerns, Ausdünnung und Konkatenation (*Contour Matching Hardware* der Firma *ELTEC, Mainz*) aus dem Grauwertbild berechnet. Es werden die (x, y) -Koordinaten von Anfang und Endpunkt der Linienelemente l_i erzeugt, wobei l die Linielänge und i den Index der einzelnen Linien beschreibt. Der Betrag $|l_i|$ der Linienelemente stellt ein Maß für deren Signifikanz dar. Beträge $|l_i| < 10$ werden vernachlässigt, um eine Rauschunterdrückung zu erreichen.

Bei der Entropieberechnung wird, basierend auf der Informationstheorie nach SHANNON [8], durch Schätzung der lokale Informationsgehalt einer Region berechnet. Hierbei wird über einen Bildausschnitt der Größe 20×20 Pixel, zentriert im Punkt x_i , die Entropie $G(x_i)$ nach Gleichung (2) als Maß der zu erwartenden Information über die im Bildausschnitt dargestellten k Pixel bestimmt. Mit Hilfe der relativen Häufigkeit $p(g_k)$ der Grauwerte g_k kann ein Maß für $G(x_i)$ hergeleitet werden,

$$G(x_i) = - \sum_k p(g_k) \log p(g_k) \quad (2)$$

wobei der Wert $G(x_i)$ dem Zentralpixel x_i des betrachteten Bildausschnitts zugeordnet wird. Eine nähere Einführung gibt KALINKE ET AL. [4].

4 Fusion

4.1 Basisrepräsentation und Merkmalextraktion

Bei der Kopplung von Modulen zur Bildanalyse ist eine geeignete Definition der Repräsentation ein wesentliches Problem, um eine hohe Flexibilität hinsichtlich der Erweiterbarkeit um zusätzliche Module einer gegebenen Integrationsstruktur zu erhalten. Die zukünftige Einkopplung von Methoden in eine gegebene Struktur muß einfach realisierbar, sowie die Einbindung zusätzlicher Sensorik flexibel möglich sein.

Als Basisrepräsentation der Vorverarbeitungsstufen ist daher die Pixeldarstellung in der Bildebene gewählt. Dabei bleibt der geometrische Bezug zum Bild erhalten. Jedem Pixel wird ein Merkmalcode aus der Vorverarbeitung zugeordnet und an die Kopplungsmatrix übergeben.

Zur Fusion pixelorientierter Repräsentationen ist eine Reduzierung der Dimension des Merkmalraumes anzustreben, damit die Dimension des Merkmalsraums für die folgende Kopplungsstruktur handhabbar ist. Es muß jedoch darauf geachtet werden, daß die Vorteile der Basisrepräsentationen, wie Nachbarschaft und geometrische Konsistenz beibehalten werden.

Daher wird ein Merkmalvektor \mathbf{v} erzeugt, welcher die Ergebnisse der Module geeignet kodiert beinhaltet (Gleichung (3))

$$\mathbf{v} = (v_i)^T, \text{ mit } i \in \{1, 2, 3\}. \quad (3)$$

Die im Modul *LOC* berechneten Richtungscodes $R(x_i)$ der Kontur können auf vier, für Objekte wesentliche Richtungen zu v_1 zusammengefaßt werden. Horizontale Merkmale $R(x_i) \in \{1, 8, 9\}$, vertikale Merk-

male $R(x_i) \in \{3, 4, 6\}$, Merkmale mit positiver $R(x_i) \in \{3, 12\}$ und negativer Steigung $R(x_i) \in \{5, 10\}$ des *LOC*-Kodes werden gemäß Definition (4) kodiert,

$$v_1(x_i) = \begin{cases} 1 & R(x_i) \in \{1, 8, 9\} \\ 2 & R(x_i) \in \{3, 4, 6\} \\ 3 & \text{für } R(x_i) \in \{3, 12\} \\ 4 & R(x_i) \in \{5, 10\} \\ 0 & \text{sonst} \end{cases} . \quad (4)$$

Die Entropiewerte werden gröber quantisiert und nach Definition (5) kodiert. Man erhält eine Reduktion auf sechs mögliche Entropiewerte für v_2 ,

$$v_2(x_i) = \begin{cases} 0 & G(x_i) = 0 \\ 5 & G(x_i) = 1 \\ 6 & \text{für } G(x_i) = 2 \\ 7 & G(x_i) = 3 \\ 8 & G(x_i) = 4 \\ 9 & G(x_i) = 5 \end{cases} . \quad (5)$$

Die Linienelemente l_i werden in Pixel in der Bildebene zerlegt ($L(x_i) = 1$ falls l_i das Pixel x_i berührt) und nach Definition (6) mit v_3 kodiert,

$$v_3(x_i) = \begin{cases} 0 & \text{für } L(x_i) = 0 \\ 10 & L(x_i) = 1 \end{cases} . \quad (6)$$

Man erhält so eine nicht überlappende Kodierung der Module im Wertebereich. Abschließend wird eine Histogrammberechnung im Wertebereich $\{1..10\}$ über ein 9×9 -Fenster für jedes Zentralpixel durchgeführt. Durch Erweiterung um die normierten Pixelkoordinaten wird ein zwölfdimensionaler Merkmalvektor erzeugt. Jeder Merkmalvektor hat dadurch einen direkten Bezug zur Lage des zugehörigen Pixels x_i in der Bildebene.

4.2 Kopplungsstruktur

Als Kopplungsstruktur werden selbstorganisierende Karten nach KOHONEN [5] [7] verwendet. Bei diesem Verfahren spielt die räumliche Anordnung der Neuronen innerhalb der Karte eine wesentliche Rolle. Die Aktivität der Neuronen wird über der räumlichen Verteilung bestimmt. Ziel ist, die Signalähnlichkeit angelegter Eingangssignale in eine Lagenachbarschaft abzubilden. Die Folge sind topographische Merkmalkarten der Eingangssignale.

5 Kohonenkarte

Der Kohonen-Algorithmus führt eine nichtlineare Projektion des zwölfdimensionalen Eingangsraums in ein zweidimensionales Gitter mit Neuronen durch. Dabei wird nach der kleinsten euklidischen Distanz $\|\mathbf{x} - \mathbf{m}_i\|$ zwischen dem Eingangsdatenvektor \mathbf{x} der Dimension n ($n = 12$) und dem Referenzvektor \mathbf{m}_i der Neuronen N_i bestimmt (Gleichungen (7))

$$\|\mathbf{x} - \mathbf{m}_c\| = \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\}. \quad (7)$$

Index c kennzeichnet den Referenzvektor \mathbf{m}_c des Neurons, welches minimale Distanz zum Merkmalsvektor \mathbf{x} hat (Gleichung (8)),

$$c = \arg \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\}. \quad (8)$$

Die Dimension der *Kohonenkarte* wurde auf 20×30 festgelegt, weil *Kohonenkarten* höherer Dimension Bereiche aufweisen, die keine Neuronenaktivität haben. Die Nachbarschaft der Neuronen N_i ist hexagonal, da jedes Neuron dadurch sechs direkte Nachbarn hat. Initial wurden die Neuronen zufällig belegt.

Die eingesetzte Lernregel (Gleichung (9)) zum Zeitpunkt $t+1$ des aktuellen Referenzvektors $m_i(t)$ hat Einfluß auf Nachbarneuronen unter Zuhilfenahme eines Gewichtungsterms $h_{ci}(t)$, auf die Nachbarschaft N_c

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(t)[\mathbf{x}(t) - \mathbf{m}_i(t)], \quad (9)$$

wobei h_{ci} eine monoton fallende Funktion $\alpha(t)$ (Lernrate) ist:

$$h_{ci} = \alpha(t), i \in N_c \quad (10)$$

$$h_{ci} = 0, i \notin N_c. \quad (11)$$

6 Experimente

Die Kopplungsmatrix (20×30) wurde in zwei Stufen gelernt. Zuerst wird eine grobe Ordnung der Neuronen bestimmt. Im zweiten Schritt wird die Feinstruktur der Karte gelernt. Es wird ein Trainingsdatensatz von 56604 Merkmalvektoren verwendet, wobei 55% der Vektoren dem Vordergrund zugeordnet sind.

Das Training der Neuronen wurde in zwei Stufen durchgeführt. In der ersten Stufe werden 20000 Vektoren und in der zweiten Stufe 15000 Vektoren trainiert. Ein initialer Radius des Einflußbereichs der h_{ic} von $c = 30$ ist vorgegeben, der im zweiten Trainingsschritt auf $c = 8$ verfeinert wird. Die Lernrate α wird initial auf $\alpha = 0.12$ gesetzt und im zweiten Schritt auf $\alpha = 0.06$ reduziert.

In Abbildung 5 ist eine Projektion der Neuronenaktivität auf eine Matrix der Größe (30×20) dargestellt.

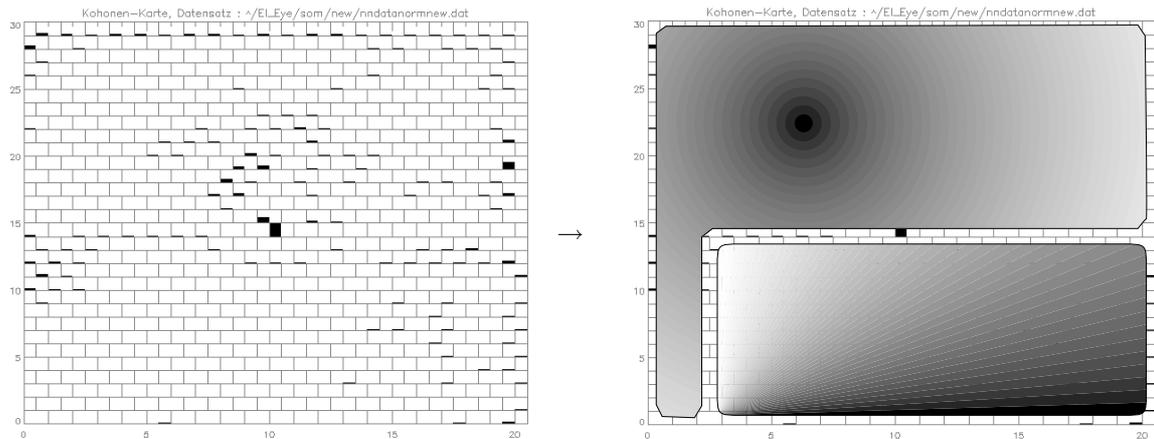


Abbildung 5: Kohonenkarte und Klassen

In Abbildung 5 (links) ist die Häufigkeit der maximalen Aktivität der einzelnen Neuronen N_i mit Hilfe von Balkendiagrammen eingezeichnet. Wie ersichtlich kann eine Zuordnung der Neuronen N_i in zwei Klassen durchgeführt werden. Eine Separierung ist möglich (Abbildung 5, rechts), wodurch die Segmentierungsaufgabe gelöst werden kann.

Es kann ein unbekannter Merkmalvektor eines Pixels, abhängig davon, in welcher Region der Karte er maximale Aktivität erzeugt, zugeordnet werden. Abbildung 6 (mitte) zeigt die so segmentierte Region in einem unbekanntem Bild durch Rückprojektion der Klassenzugehörigkeiten der Pixel in die Bildebene. Eine Verbesserung der Segmentierung der PKW im Vergleich zu Einzelalgorithmen (z.B. Segmentierung mit Hilfe der lokalen Entropie) wird erreicht (Bild 6, links).



Abbildung 6: Rückprojektion: Grauwertbild, Segmentierung Fusion, Segmentierung Entropie

7 Diskussion und Ausblick

Wie gezeigt wurde, ergibt sich eine Steigerung der Leistungsfähigkeit durch Kopplung mehrerer Algorithmen. Es fällt jedoch auf, daß eine Reduktion des Kontrastes im Grauwertbild, beispielsweise bei Schattenwurf, die Segmentierung verschlechtert kann, da Konturinformationen verloren gehen können.

Es scheint daher notwendig die Anzahl der Vorverarbeitungsmodule auszuweiten. Beispielsweise kann eine Analyse über der Zeit berücksichtigt werden. Zusätzlich kann ein verbessertes Kopplungsmodell eine Verbesserung bringen. Geeignet könnte eine höherdimensionale neuronale Struktur sein, die in ein hierarchisches Modell integriert wird. Hierbei ist die Kopplung lernbar, sowie ein *on-line*-Training möglich. Eine Konservierung einzelner Kopplungsstrukturelemente kann ebenfalls durchgeführt werden. Eine Rückkopplung über die Zeit ist ebenso möglich, wie die Integration von zusätzlichen Sensorinformationen.

Literatur

- [1] Belur V. Dasarathy; Decision Fusion; IEEE Computer Society Press; Los Alamitos 1994
- [2] B.A. Draper, A.R. Hanson, E.M. Riseman; Proceedings of the IEEE, November 1996 Vol.84 No.11 Knowledge-Directed Vision: Control, Learning, and Integration;
- [3] Christian Goerick; Local Orientation Coding and Adaptive Thresholding for Real Time Early Vision; IRINI94-04, Institut für Neuroinformatik, Bochum 1994
- [4] Thomas Kalinke, Werner von Seelen; Entropie als Maß des lokalen Informationsgehalts in Bildern zur Realisierung einer Aufmerksamkeitssteuerung; in Mustererkennung 1996 Hrsg. B. Jähne et al.; Springer-Verlag Berlin Heidelberg 1996
- [5] T. Kohonen; Self-Organizing Maps; Springer-Verlag Berlin 1995
- [6] T. Kohonen, et al.; The Self-Organizing Map Program Package; Helsinki University of Computer and Information Science 1995
- [7] H. Ritter, T. Martinetz, and K. Schulten; Neuronale Netze; Addison-Wesley Bonn 1994
- [8] C. E. Shannon; A Mathematical Theory of Communication; Bell Systems Technical Journal 1948
- [9] Pramod K. Varshney, Editor; Special Issue on Data Fusion; Proceedings of the IEEE, January 1997 Vol.85 No.1