

# Ein Mehrgitterverfahren zur Bewegungssegmentierung in Bildfolgen

Horst Haußecker<sup>1</sup> und Bernd Jähne<sup>2</sup>

<sup>1</sup> Institut für Umweltphysik der Universität Heidelberg

Im Neuenheimer Feld 366, 69120 Heidelberg

<sup>2</sup> Scripps Institution of Oceanography, PORD

La Jolla, CA 92093-0230 USA

Zusammenfassung

doi: 10.5281/zenodo.14549

Es wird ein Verfahren vorgestellt, mit dem die Bewegung einzelner Bildbereiche in ausgedehnten Bildfolgen zuverlässig segmentiert werden kann. Dabei werden zuerst die ruhenden Strukturen in Bildsequenzen durch eine zeitliche Gaußpyramide ermittelt. Zum anschließenden Vergleich dieser Bilder mit den Originalbildern wird ein lokales Ähnlichkeitsmaß berechnet und dieses mit Hilfe von Fuzzy-Logic bewertet. Das Resultat ist ein fehlertolerantes Bewertungsverfahren, welches selbst mit bloßem Auge schwer zu erkennende Objekte detektiert. Der Bildinhalt wird somit, unabhängig vom Inhalt des Einzelbildes, allein aufgrund der Eigenschaft "Bewegung in der Sequenz" segmentiert.

## 1 Einleitung

Die exakte Bestimmung der Geschwindigkeiten von Objekten in Bildfolgen ist ein grundlegendes Problem der Digitalen Bildfolgenanalyse. Eine quantitative Bewegungsanalyse liefert jedoch selbst bei ausgedehnten Bildfolgen im allgemeinen nicht die wirkliche Bewegung der Objekte im 3D-Raum und ist bei Analyse von nur zwei aufeinanderfolgenden Bildern meist nicht möglich, [Verri, 1989]. In vielen Fällen ist es allerdings nicht notwendig, Geschwindigkeiten von Objekten in Bildern exakt zu bestimmen. Oft reicht eine mehr qualitative Analyse aus, d. h. eine *Bewegungssegmentierung*, die Bildbereiche bestimmt, in denen Bewegung stattfindet. Beispiele für solche Anwendungen sind die Überwachung von Verkehrsanlagen und Bahnsteigen [Haass, 1984] oder die Kontrolle von Fußgängerwegen [Shio, 1991]. Der hier beschriebene Anwendungsfall ist die Beobachtung der Ablösung von Sedimenten in fließenden Gewässern. Die Wasserbewegung im Untergrund kann zur Verlagerung von Bodenpartikeln und damit zu Bodenformationen führen. Begünstigt wird diese Partikelverlagerung an Schichtgrenzen zweier Bodenarten oder der Grenze von Boden zu festen oder flexiblen Bauteilen [Köhler, 1992]. Die automatische Detektion von Partikelbewegungen in diesen Bereichen der Bodenschichten in Abhängigkeit von hydrodynamischen und mechanischen Belastungen liefert wichtige Informationen über die Stabilität von Uferbefestigungen. Typische Bilder solcher Grenzschichten zwischen einer Sandoberfläche und einem aufliegenden Geotextil sind in Abb. 6 a) und b) zu sehen.

Zur Detektion von sich bewegenden Partikeln ist es notwendig, sie von den ruhenden Strukturen im Bild zu unterscheiden. Da es sich bei beiden zum Teil um gleichartige Objekte handelt (hier Sandkörner), kann man nicht durch eine Vorsegmentierung des Bildinhaltes Teilbereiche dem ruhenden Hintergrund zuordnen, wie dies z. B. bei Straßenszenen der Fall ist. Gleichzeitig ändert sich der ruhende Hintergrund im Laufe der Zeit durch Materialverlagerungen, so daß man nicht von einem zeitlich invarianten Hintergrundbild ausgehen kann, sondern zu jedem Bild die zeitlich lokalen, ruhenden Strukturen bestimmen muß.

Das vorliegende Verfahren gliedert sich somit in zwei Schritte: Im ersten Schritt wird aus einer zeitlichen Umgebung des betrachteten Bildes der ruhende Hintergrund extrahiert. Im zweiten Schritt wird dieses Hintergrundbild mit dem Originalbild verglichen und es werden Bildbereiche bestimmt, in denen ein signifikanter Unterschied zwischen beiden Bildern besteht. Dadurch werden Gebiete detektiert, in denen mit großer Wahrscheinlichkeit Bewegung stattfindet.

## 2 Extraktion ruhender Strukturen mittels einer zeitlichen Gaußpyramide

Zur Entfernung von sich bewegenden Strukturen aus der Bildsequenz wird eine zeitliche Glättung der Form

$$H(\mathbf{x}, t_0) = \int_{t_0 - \Delta t}^{t_0 + \Delta t} w(t_0 - t)g(\mathbf{x}, t) dt, \quad (1)$$

durchgeführt. Als Faltungsmaske wird eine *Gaußmaske*

$$w(t) = g_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} \quad (2)$$

verwendet. Sie bietet eine minimale Unschärfe der Transferfunktion im Fourierraum bei minimaler Ausdehnung der Faltungsmaske im Zeitraum. Zur diskreten Realisierung definieren wir, analog zu [Jähne, 1991], die *zeitlichen Binomialmasken*

$$B_t^2 = \frac{1}{4} [1 \ 2 \ 1] \quad B_t^4 = \frac{1}{16} [1 \ 4 \ 6 \ 4 \ 1] \quad (3)$$

und den *Reduktionsoperator*  $R_t$ , der nur jedes zweite Bild aus einer Bildfolge herausgreift.

Eine effektive und sukzessive Berechnung der Glättung mit einer Gaußmaske wird durch die in Abb. 1 skizzierte *zeitliche Gaußpyramide* mit den in (3) definierten Binomialmasken angenähert. Dazu glättet man die Originalsequenz mit der  $B_t^4$ -Maske. Eine Reduktion der Zeitauflösung des Ergebnisses dieser Faltung durch  $R_t$  führt zu Aliasingeffekten, die aber dadurch kompensiert werden können, daß nach der Reduktion zusätzlich mit einer  $B_t^2$ -Maske geglättet wird. Insgesamt wird die nächste Ebene der Pyramide durch den zusammengesetzten Operator ( $B_t^2 R_t B_t^4$ ) erzeugt. Die Wirkung dieser Glättung ist in Abb. 2 anhand

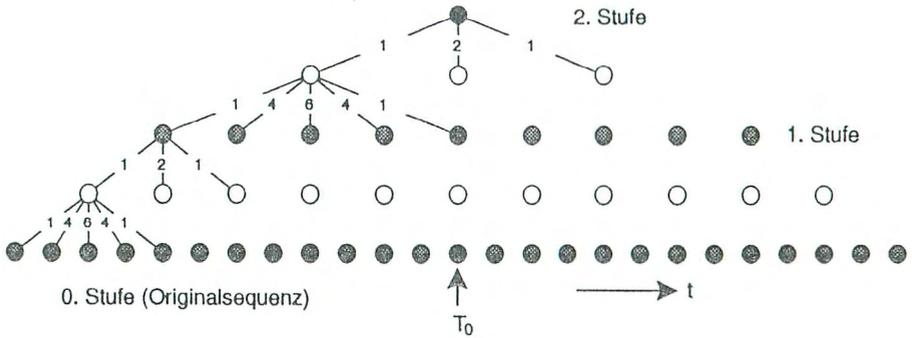


Abb. 1: Zeitliche Gaußpyramide. Die Kreise entsprechen einzelnen Bildern.

der resultierenden Transferfunktion im Vergleich zu einer  $B_1^4$ -Maske dargestellt. Dieses in [Jähne, 1992] beschriebene *gezielte Unterabtasten* ermöglicht eine sehr recheneffektive und aliasingfreie Iteration von Binomialmasken.

Zur Extraktion des Hintergrundes eines weiteren Bildes muß nicht die gesamte Pyramide neu aufgebaut werden. Verschiebt man die Pyramide um ein Bild nach rechts, so werden die meisten der bereits berechneten Glättungen erneut benötigt. Wenn man alle Zwischenstufen der Pyramide aufbewahrt, so kann man diese Zwischenergebnisse verwenden, und es genügt die am rechten Rand neu hinzukommenden Faltungen zu berechnen.

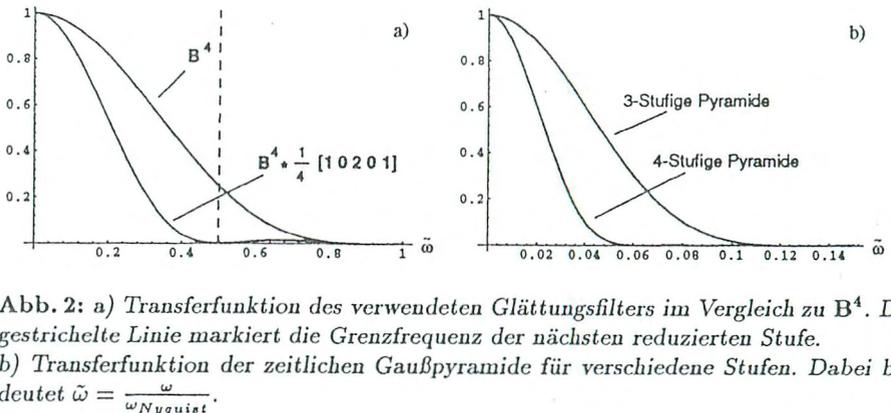


Abb. 2: a) Transferfunktion des verwendeten Glättungsfilters im Vergleich zu  $B^4$ . Die gestrichelte Linie markiert die Grenzfrequenz der nächsten reduzierten Stufe.

b) Transferfunktion der zeitlichen Gaußpyramide für verschiedene Stufen. Dabei bedeutet  $\tilde{\omega} = \frac{\omega}{\omega_{Nyquist}}$ .

### 3 Vergleich des Originalbildes mit dem Hintergrund

Nachdem mit dem in Abschnitt 2 vorgestellten Verfahren die ruhenden Strukturen in einem Bild detektiert wurden, muß nun das so erhaltene Hintergrundbild mit dem Originalbild verglichen werden, um die Bildbereiche zu segmentieren,

in denen Bewegung stattfindet. Dies geschieht durch einen *qualitativen* Vergleich zwischen den *lokalen Strukturen* in beiden Bildern. Er ähnelt der menschlichen Interpretation von Bildinhalten, durch die beim Vergleich zweier Bilder Strukturen als gleich bzw. ähnlich erkannt werden, auch ohne daß die Differenz im absoluten Grauwert an isolierten Punkten quantitativ bestimmt wird. Zur Realisierung wird zuerst ein *lokales Ähnlichkeitsmaß* für jeden Punkt berechnet. Dieses wird mit einer Schwelle verglichen, die darüber entscheidet, ob es sich um einen signifikanten Unterschied zwischen beiden Bildern handelt oder um Rauschen. Es zeigt sich dabei, daß nur eine Schwelle sinnvoll ist, die sich den lokalen Bildstrukturen anpaßt.

### 3.1 Lokales Ähnlichkeitsmaß

Ein lokales Ähnlichkeitsmaß  $S(x_0)$  am Punkt  $x_0$  wird folgendermaßen definiert:

$$S(x_0) = \int_G h(x_0 - x) (H(x) - O(x))^2 dx, \quad (4)$$

mit  $H(x)$  = Hintergrundbild und  $O(x)$  = Originalbild.  $G$  bezeichnet eine lokale Integrationsumgebung und  $h(x)$  eine normierte Fensterfunktion mit Schwerpunkt bei 0, welche die punktweise Differenz der beiden Bilder in der lokalen Umgebung verschieden stark wichtet. Dadurch werden einzelne Rauschpeaks unterdrückt und es wird berücksichtigt, daß es sich bei den bewegten Objekten um ausgedehnte Strukturen handelt.

Die endgültige Entscheidung darüber, ob ein Punkt dem Hintergrund oder dem Originalbild zuzuordnen ist, ergibt sich durch einen abschließenden Vergleich des Ähnlichkeitsmaßes  $S$  mit einer lokalen Schwelle. Sie bestimmt, ob der Unterschied zwischen beiden Bildern als signifikant betrachtet wird oder nicht. [Westberg, 1992] zeigt, daß ein Likelihood-Ratio-Test zum Bewerten der Differenz zweier Bilder unter bestimmten Voraussetzungen zu einem Schwellwertverfahren führt, bei dem ein Ähnlichkeitsmaß wie in (4) mit einer Schwelle verglichen wird. Im Gegensatz zu dieser *globalen Schwelle* paßt sich die hier verwendete *lokale Schwelle* den lokalen Strukturen in beiden Bildern an. Diese *strukturadaptive Bewertung* des Ähnlichkeitsmaßes entspricht dem menschlichen visuellen System, welches abhängig vom Bildinhalt manche Unterschiede sofort erkennt, bei anderen aber "genauer hinschauen" muß. Ein Ansatz, der dieser Intention folgt, wird im nächsten Abschnitt vorgestellt.

### 3.2 Lokale Schwelle durch Fuzzy-Logic

[Hsu, 1984] und [Nagel, 1982] verwenden einen Likelihood-Test zum Vergleich zweier aufeinanderfolgender Bilder auf Unterschiede bezüglich modellierter Grauwertverläufe. In [Bouthemy, 1989] wird das gleiche Testverfahren angewendet, um ein 3D-Volumen im  $xt$ -Raum auf Anwesenheit einer Kante und somit Bewegung zu testen. In der vorliegenden Arbeit werden in ähnlicher Art und Weise

statistische Eigenschaften der Grauwertverläufe korrespondierender lokaler Umgebungen im Bild und im Hintergrund verglichen. In erster Näherung kommen dafür der *mittlere Grauwert* und die *Varianz* der Grauwertstichproben in Betracht. Durch Verwenden höherer Momente kann die exakte Grauwertstatistik beliebig gut angenähert werden. Gute Ergebnisse wurden auch durch Vergleich der lokalen Orientierung zwischen Originalbild und Hintergrund erzielt. Im Gegensatz zu [Nagel, 1982] werden jedoch nicht direkt Grauwertverläufe modelliert und diese statistisch verglichen, sondern die Unterschiede der lokalen Grauwertstatistiken werden zur Bewertung des zuvor berechneten Ähnlichkeitsmaßes benutzt.

Die Berechnung der lokalen Schwelle erfolgt mit Hilfe der *Fuzzy-Logic*, die der menschlichen Entscheidungsfindung nachempfunden ist, [Zadeh, 1988]. Im vorliegenden Verfahren sind die Ausgangsgrößen die *Differenz des lokalen mittleren Grauwertes* (im folgenden kurz *GD*) und die *Differenz der lokalen Varianz* (im folgenden kurz *VD*) von Hintergrund und Originalbild. Als Ergebnis der Verknüpfung dieser Größen erhält man die lokale Schwelle. Da alle möglichen Werte von *GD* und *VD* von vornherein bekannt sind, läßt sie sich als *Look-Up-Tabelle* (LUT) abspeichern. Abb. 5a zeigt die lokale Schwelle in Abhängigkeit von *GD* und *VD*. In Abb. 3 ist das gesamte Verfahren der Bewegungssegmentierung im Überblick dargestellt.

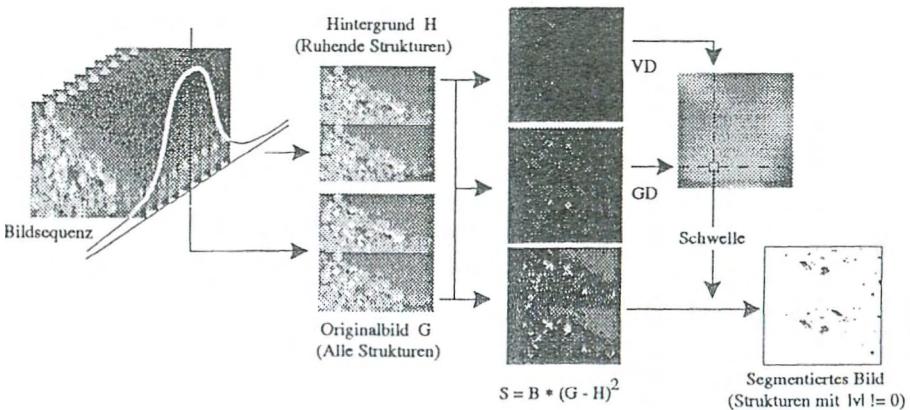


Abb. 3: Ablaufdiagramm des Verfahrens der Bewegungssegmentierung.

Insgesamt entspricht dies einer Verknüpfung der Struktureigenschaften der lokalen Umgebungen in beiden Bildern zu einem Binärbild. Darin bezeichnet 1 "Bewegung" und 0 "keine Bewegung". Für  $n$  verschiedene Merkmale  $\mu_i$  läßt sich dies durch die Abbildung

$$f(\mu_1, \mu_2, \dots, \mu_n) \rightarrow \{0, 1\} \quad (5)$$

des  $n$ -dimensionalen Merkmalraumes in den Binärraum der Bewegungssegmentierung beschreiben. Im unserem Fall ist  $n = 3$  und  $(\mu_1, \mu_2, \mu_3) = (GD, VD, S)$ . Anstelle der direkten Verknüpfung von  $GD, VD$  und  $S$  nach (5) wird aus den ersten beiden Eigenschaften eine lokale Schwelle durch Fuzzy-Logic bestimmt, mit der die dritte Eigenschaft binarisiert wird. Dies hat den Vorteil, daß eine analytisch nur schwer zu bestimmende Funktion durch wenige logische Verknüpfungen aufgebaut wird und sich somit leicht optimieren läßt.

## 4 Ergebnisse

In Abb. 6 werden drei verschiedene Bilder mit ihren einzelnen Bearbeitungsschritten gezeigt. Man sieht, daß dabei auch Strukturen detektiert werden, deren Grauwerte fast im Rauschen untergehen, wenn es sich dabei um bewegte Teilchen handelt. Dies wird noch deutlicher, wenn man Grauwertschnitte durch die Bilder betrachtet. In Abb. 4 ist je ein Schnitt durch das Originalbild und das Hintergrundbild an der in Abb. 6 markierten Stelle dargestellt.

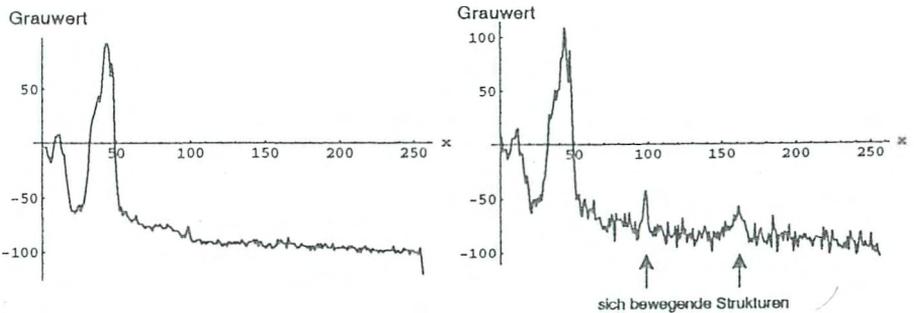


Abb. 4: Grauwertschnitte durch das Bild in Abb. 6 a). Rechts ist der Schnitt durch das Originalbild und links der durch das zeitlich geglättete Hintergrundbild dargestellt.

Man erkennt, daß sich die mit Pfeilen gekennzeichneten lokalen Grauwertmaxima nur knapp über den Rauschlevel erheben. Im Schnitt durch das Hintergrundbild sind diese trotz des erheblich geringeren Rauschens (das Rauschen wird bei der zeitlichen Tiefpaßfilterung unterdrückt) nicht zu erkennen. Dies deutet darauf hin, daß es sich bei den beiden Peaks um bewegte Strukturen handelt. Zusätzlich weist die lokale Schwelle an dieser Stelle ein deutliches Minimum auf (Abb. 5b). Dadurch wird die Detektion der bewegten Strukturen erleichtert. Im Ergebnisbild werden an diesen Stellen auch deutlich zwei Teilchen detektiert. Zu Testzwecken wurde der Algorithmus noch zusätzlich auf eine Verkehrsszene angewendet. Ohne eine Veränderung der Parameter der Fuzzy-Logic werden dort im rechten oberen Viertel deutlich zwei Fahrzeuge und im rechten unteren Viertel ein Fußgänger detektiert (Abb. 6c).

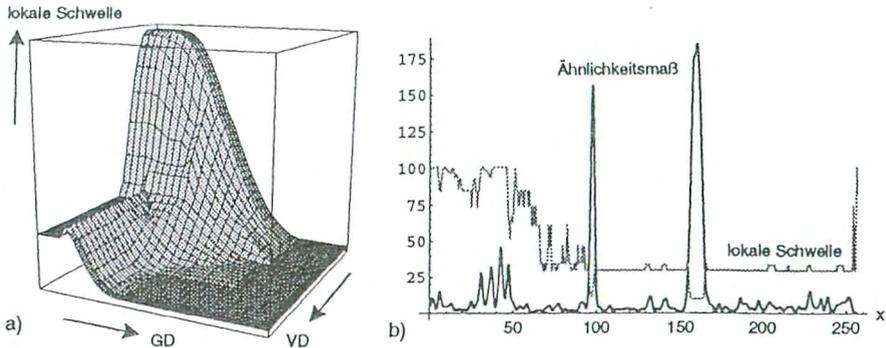


Abb. 5: a) LUT der lokalen Schwelle. b) Ähnlichkeitsmaß und lokale Schwelle im Grauwertschnitt der Abb. 4.

## Literaturverzeichnis

- Bouthemy, P.: *A Maximum Likelihood Framework for Determining Moving Edges*. IEEE Trans. Pattern Anal. Machine Intell., vol. 11, No. 5, pp. 499-511, May 1989
- Haass, U.L.: *Verfolgung dynamischer Änderungen in TV-Bildszenen zur Deutung von Bewegungen*. FhG-Berichte 2-84, 12, 1984
- Hsu, Y.Z., Nagel, H.-H., Rekers, G.: *New likelihood test methods for change detection in image sequences*. Comp. Vision, Graph. Image Proc. 26, 73, 1984
- Jähne, B.: *Digital Image Processing - Concepts, Algorithms and Scientific Applications*. Springer 1991
- Jähne, B.: *Spatio-Temporal Image Processing With Scientific Applications*. Habilitationsschrift, Technische Universität Hamburg-Harburg 1992
- Köhler, H.-J.: *Materialverlagerung an Grenzschichten unter hydrodynamischer Belastung*. Projektantrag, Bundesanstalt für Wasserbau Karlsruhe 1992
- Nagel, H.-H., Rekers, G.: *Moving object masks based on an improved likelihood test*. Proc. Int. Conf. Patt. Recogn., pp. 1140-1142, München 1982
- Shio, A., Sklansky, J.: *Segmentation of People in Motion*. IEEE Visual Motion, Proc., pp. 325-332, 1991
- Verri, A., Poggio, T.: *Motion Field and Optical Flow: Qualitative Properties*. IEEE Trans. Pattern Anal. Machine Intell., vol. 11, No. 5, pp. 545-558, May 1989
- Westberg L.: *Hierarchical Contour-Based Segmentation of Dynamic Scenes*. IEEE Trans. Pattern Anal. Machine Intell., vol. 14, No. 9, pp. 946-952, September 1992
- Zadeh, L.A.: *Fuzzy-Logic*. IEEE Computer April 1988

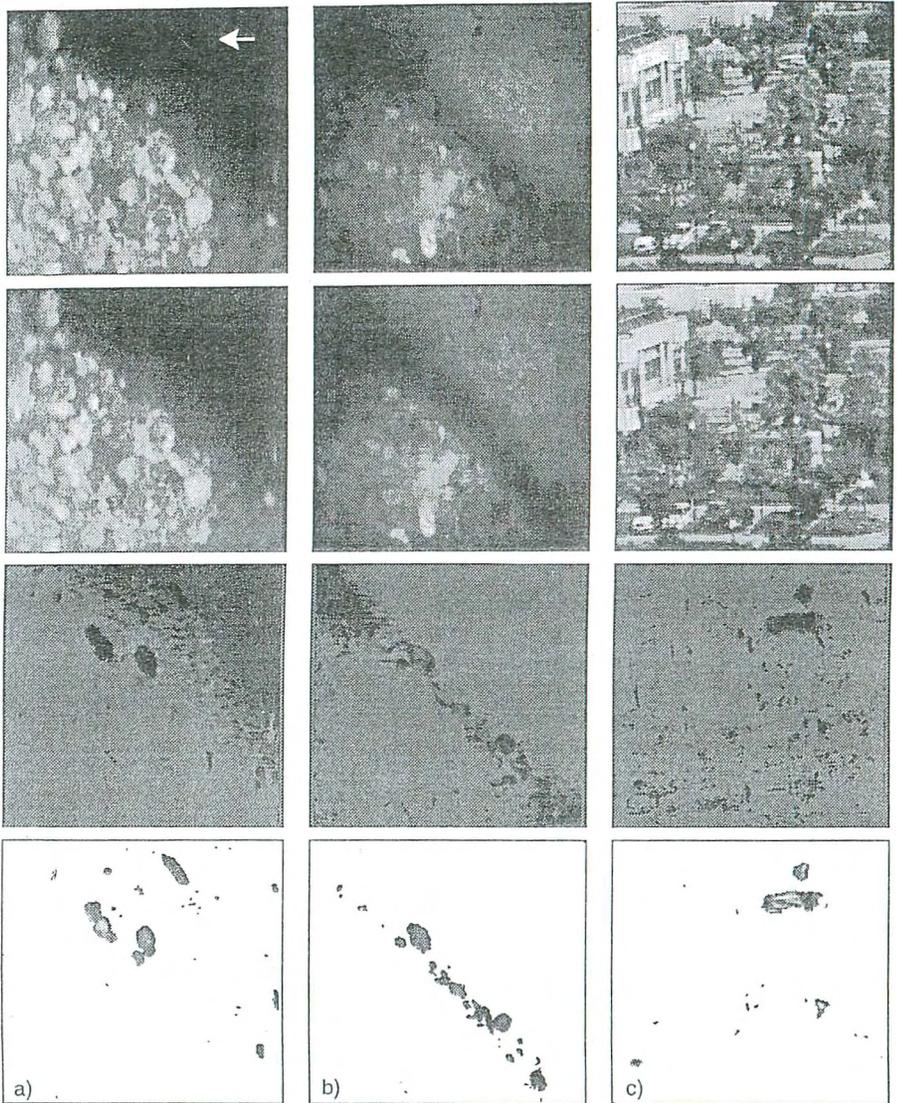


Abb. 6: Beispiele zur Bewegungssegmentierung: Die drei Spalten zeigen verschiedene Beispielbilder. a) + b): Querschnitt einer Sandgrenzschicht. Der Bildausschnitt beträgt etwa 1 cm in horizontaler und vertikaler Richtung. c): Verkehrsszene. Von oben nach unten: Originalbild, Hintergrund, lokale Schwelle und Ergebnis der Segmentierung. Der Pfeil in Originalbild a) markiert den horizontalen Grauwertschnitt, der in Abb. 4 dargestellt ist.