

Effiziente Filterverfahren auf Mehrgitter-Datenstrukturen

H. HAUSSECKER, B. JÄHNE

Interdisziplinäres Zentrum für wissenschaftliches Rechnen der Universität Heidelberg

H.-J. KÖHLER

Bundesanstalt für Wasserbau, Karlsruhe

Symposium Bildverarbeitung '95, Forschen, Entwickeln, Anwenden

Technische Akademie Esslingen, 29.11 - 01.12.1995, R.-J. Ahlers (Hrsg.)

doi: 10.5281/zenodo.14709

ZUSAMMENFASSUNG

Klassische Nachbarschafts-Operationen verknüpfen die Eigenschaften lokaler Umgebungen eines einzelnen Pixels um aus der umgebenden Grauwertstruktur Bildeigenschaften zu extrahieren. Trotz ihrer enormen Bedeutung für die Bildverarbeitung sind sie jedoch nur auf kleinen Umgebungen sinnvoll einsetzbar, da der Rechenaufwand mit wachsender Umgebungsgröße explosiv ansteigt. Eine nähere Betrachtung der verwendeten Faltungen zeigt aber, daß sich viele Nachbarschaftsoperationen auch auf große Masken bei minimalem Rechenaufwand erweitern lassen. Dies geschieht durch die Verwendung spezieller Mehrgitterstrukturen, zu deren Aufbau die Eigenschaften der Transferfunktionen geschickt ausgenutzt werden. Man erreicht hier (ähnlich wie bei der FFT) eine Reduktion des Rechenaufwandes von der Größenordnung N^2 auf ldN . Der hier beschriebene Anwendungsfall ist die Beobachtung der Bewegung von Sedimenten in fließenden Gewässern. Durch die langsamen Strukturveränderungen im Boden werden Filterungen über lange Bildsequenzen hinweg nötig, die nur durch die beschriebene zeitliche Mehrgitterstruktur realisierbar sind.

1 EINLEITUNG

Eine der grundlegendsten Bildverarbeitungsoperationen ist die Glättung oder allgemeiner, die Faltung mit einer symmetrischen Faltungsmaske. Sie verknüpft lokale Umgebungen einzelner Pixel durch eine gewichtete Mittelung benachbarter Grauwerte. Als low-level Bildverarbeitungsoperator wird sie zur Rauschunterdrückung eingesetzt. Die wahre Bedeutung dieser Operation darf jedoch nicht unterschätzt werden. Sie wirkt in jede beliebige Dimension komplexer Datenstrukturen und verknüpft sowohl elementare Grauwerte als auch – in Eigenschaftsbildern – die Ergebnisse vorangegangener, beliebiger Bildverarbeitungsoperationen und ist ein elementarer Bestandteil fast jeder höheren Bildverarbeitungsroutine. Jede Einsparung an Rechenzeit durch eine effiziente Berechnung wirkt sich somit auf mehrere Stufen der gesamten Bildauswertung aus.

Der elementaren Bedeutung der Faltung in der Bildverarbeitung wurde in der Vergangenheit dadurch Rechnung getragen, daß spezielle Hardware entwickelt wurde, die selbst Faltungen mit mehrdimensionalen, nicht separablen Faltungsmasken der Größe $32 \times 32 \times 32$ durchführen konnte. Diese *Hardwareconvolver* zeichneten sich durch eine hohe Rechengeschwindigkeit aus, selbst bei Faltungen dieser Größe mit über 32.000 Multiplikationen und Additionen, die mit keiner der zu dieser Zeit verfügbaren Standardhardware möglich gewesen wäre. Der entscheidende Nachteil dieser Hardware-Realisierung lag in ihrer Unflexibilität. Trotz ihrer enormen Leistungsfähigkeit bei großen Faltungsmasken war die Performance bei kleinen und gutmütigen Faltungsmasken schlecht. Sie konnte weder die Tatsache ausnutzen, daß bei kleinen Faltungsmasken der größte Teil der Filterkoeffizienten Null ist, noch konnte sie separable Masken als solche erkennen.

Dieser Beitrag widmet sich, im Gegenzug zu starrer Hardware, der Frage, wie man die mathematischen Eigenschaften von Faltungen ausnutzen kann und somit eine effiziente und schnelle Implementierung selbst großer Faltungen auf jeder beliebigen Standardhardware erreichen kann.

Zur Veranschaulichung wird die grundlegende Eigenschaft der Glättung anhand der Faltung mit einer Binomialmaske wiederholt. Sie vereint in idealer Weise optimale Glättungsergebnisse bei minimaler Maskengröße. Eine erste Einsparung an Rechenschritten erhält man durch Ausnutzen der Separabilität des Faltungskernels. Danach wird erläutert, wie große Masken durch ein sukzessives Falten mit kleinen Masken aufgebaut werden können. Dieser Schritt wird auf unterschiedliche Auflösungsstufen erweitert. Dies eröffnet eine neue Möglichkeit mit minimalem Rechenaufwand eine beliebig große Glättungsmaske in Form einer Gaußpyramide aufzubauen. Abschließend wird anhand eines konkreten Anwendungsbeispiels gezeigt, wie dieses Ergebnis dazu verwendet werden kann, auf verschiedenen zeitlichen Auflösungsstufen innerhalb ausgehnter Bildsequenzen selbst langsame Bewegungen sicher zu detektieren.

2 MATHEMATISCHE EIGENSCHAFTEN VON BINOMIALMASKEN

Ein optimaler Glättungsfilter sollte die folgenden Eigenschaften aufweisen:

- *Isotropie.* Die Transfer Funktion des Filter kernels sollte nur vom Betrag der Wellenzahl abhängen und nicht von der Richtung, bzw. Orientierung der Grauwertstruktur.
- *Optimale Balance zwischen Orts- und Wellenzahlauflösung.* Diese Forderung bezieht sich auf eine fundamentale Beziehung, bekannt als *Unschärferelation* in Physik und computer-vision und als *Zeit-Bandbreite Produkt* in der Signalverarbeitungsliteratur.
- *Effiziente Berechnung.* Diese Forderung ist Thema dieses Beitrags.

Die beiden ersten Kriterien werden von Gaußfiltern erfüllt. Im folgenden soll gezeigt werden, wie die dritte Bedingung über die Standardimplementierung hinaus optimal erfüllt werden kann. Ohne auf die analytische Form der Gaußfilter einzugehen, wird nun gleich die für die Anwendung relevante Realisierung in Form von Binomialfiltern vorgestellt.

2.1 AUFBAU VON BINOMIALMASKEN

Die Faltungsmasken enthalten die Werte der *diskreten Binomialverteilung*. Die Filterkoeffizienten für Binomialmasken B^n des Grades n mit $n + 1$ Filterkoeffizienten entsprechen den Werten des *Pascal'schen Dreiecks*:

n	f		σ^2
0	1	1	0
1	1/2	1 1	1/4
2	1/4	1 2 1	1/2
3	1/8	1 3 3 1	3/4
4	1/16	1 4 6 4 1	1
5	1/32	1 5 10 10 5 1	5/4
6	1/64	1 6 15 20 15 6 1	3/2
7	1/128	1 7 21 35 35 21 7 1	7/4
8	1/256	1 8 28 56 70 56 28 8 1	2

Dabei bezeichnet f den Skalierungsfaktor 2^{-n} , der zur Normierung auf 1 nötig ist und σ^2 die Varianz, d. h. σ entspricht der effektiven Reichweite der Glättung in Pixeln. Für die Varianz der Faltungsmasken gilt:

$$\sigma^2 = \frac{n}{4}. \quad (1)$$

Von praktischer Bedeutung sind hauptsächlich die Faltungsmasken mit ungerader Koeffizientenzahl, da sie nicht zu einer Verschiebung des Faltungsergebnisses führen.

2.2 SEPARABILITÄT

Zwei- und höher-dimensionale Binomialfilter lassen sich aus eindimensionalen Faltungen entlang der entsprechenden Koordinatenachsen aufbauen. Dies entspricht dem Aufbau der mehrdimensionalen Faltungsmasken durch eine Faltung der eindimensionalen Masken miteinander:

$$B^n = B_x^n * B_y^n \quad \text{und} \quad B^n = B_x^n * B_y^n * B_z^n \quad \text{bzw.} \quad B^n = B_x^n * B_y^n * B_t^n, \quad (2)$$

oder in Operator-Schreibweise:

$$\mathcal{B}^n = \mathcal{B}_x^n \mathcal{B}_y^n, \quad \mathcal{B}^n = \mathcal{B}_x^n \mathcal{B}_y^n \mathcal{B}_z^n \quad \text{und} \quad \mathcal{B}^n = \mathcal{B}_x^n \mathcal{B}_y^n \mathcal{B}_t^n. \quad (3)$$

Die kleinste 2D-Maske dieser Art ist ein 3×3 Binomialfilter ($n = 2$):

$$B^2 = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} * \frac{1}{4} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \quad (4)$$

Diese Faltungsmasken erfüllen die Bedingung der *isotropie* sehr gut. Bereits bei $n = 5$ ist die Anisotropie vernachlässigbar. Durch diese *Separabilität* wird der Rechenaufwand erheblich reduziert. Während bei einer Faltung mit einer zweidimensionalen Maske der Größe 5×5 schon 25 Multiplikationen und 24 Additionen pro Bildpunkt benötigt werden, reduzieren sich die Rechenschritte bei einer separablen Faltung auf $2 * (5$ Multiplikationen und 4 Additionen). Mit steigender Dimensionalität können die Einsparungen bis zu einige Größenordnungen betragen. Für die direkte Berechnung der Glättung mit einer Binomialmaske B^n der Dimension k werden allgemein $(n + 1)^k$ Multiplikationen und $(n + 1)^k - 1$ Additionen benötigt, im Gegensatz zu nur $k * (n + 1)$ Multiplikationen und $k * n$ Additionen bei Ausnutzung der Separabilität.

2.3 KASKADIERUNG VON ELEMENTARMASKEN

Eine weitere Eigenschaft macht die Binomialfilter zu idealen Faltungsmasken im Sinne einer effizienten Implementierung. Alle Masken B^n beliebiger Größe lassen sich aus der kleinsten Maske B^1 aufbauen, welche die Grauwerte zweier benachbarter Bildpunkte mittelt. Dies geschieht durch sukzessives Falten der Form:

$$B^m = \frac{1}{2^m} \underbrace{\left[\begin{array}{cc} 1 & 1 \end{array} \right] * \left[\begin{array}{cc} 1 & 1 \end{array} \right] * \cdots * \left[\begin{array}{cc} 1 & 1 \end{array} \right]}_{m \text{ mal}}. \quad (5)$$

Dadurch reduziert sich der Rechenaufwand deutlich weiter. Nach der Separabilität bleibt jeweils eine Multiplikation pro nötiger Addition, wobei die Multiplikation sehr rechenaufwendig ist. Die Faltungsmaske B^1 enthält jedoch nur 1 als Filterkoeffizient. Dadurch reduziert sich die Faltung auf eine einzige Addition mit anschließender Normierung. Da der Normierungsfaktor jedoch eine 2er-Potenz darstellt, kann die Division durch eine shift-Operation ersetzt werden, die kaum Rechenzeit benötigt. Abhängig von der Speichertiefe der Datenstruktur kann diese Operation nach mehreren Faltungen mit B^1 durchgeführt werden kurz bevor ein Überlauf stattfindet.

Durch diese *Kaskadierung* der Faltung, unter Ausnutzung der Separabilität, verringert sich der Rechenaufwand für eine Glättung mit einer Binomialmaske B^n der Dimension k auf nur $k * n$ Additionen und einige shift-Operationen pro Bildpunkt.

Wichtig für das Verständnis der folgenden Algorithmen ist die Tatsache, daß sich die negativen Eigenschaften der einzelnen kleinen Faltungsmasken nicht fortpflanzen. Dies bedeutet, daß die Gesamtfaltung als Ganze das Endergebnis bestimmt und nicht die Zwischenschritte. Obwohl eine einzelne Faltung mit B^1 das Ergebnis um einen Pixel verschiebt, wirkt die anschließende Faltung auf dieses Ergebnis unabhängig von der Position des Bildinhaltes und schiebt das Faltungsergebnis wieder auf den ursprünglichen Platz zurück. Dazu muß das Ergebnis der Faltung abwechselnd an die Position des rechten und linken Pixels der Faltungsmaske gespeichert werden. Auch die Tatsache, daß die Maske B^1 eine Rechteckmaske darstellt mit Überschwingern bei hohen Wellenzahlen, hinterläßt keine Spuren nach der zweiten Faltung, nach der der Frequenzgang bereits den glatten Verlauf einer Cosinus-Funktion aufweist, die zu hohen Wellenzahlen auf Null abfällt. Dies ist eine Folge der Linearität und Verschiebungsinvarianz der Faltung: Es macht keinen Unterschied, ob man zunächst große Masken durch sukzessives Falten mit kleinen Masken aufbaut, oder ob das Bild direkt mit einer Serie kleiner Faltungsmasken bearbeitet wird.

Bereits hier deutet sich jedoch an, daß für eine effiziente Implementierung der Glättung sehr einfache Masken zur *Vorglättung* verwendet werden können, die als einzelne Faltung Fehler verursachen. Durch eine *Nachglättung* werden die Fehler der Vorglättung anschließend bereinigt. Dieses Prinzip wird im folgenden auf mehrere Auflösungsstufen erweitert, um eine Mehrgitterstruktur bei möglichst geringem Rechenaufwand aufzubauen.

3 KASKADIERTE BINOMIALFILTER AUF EINER MEHRGITTER-STRUKTUR

Trotz der optimalen Eigenschaften der Binomialfilter bezüglich ihres Frequenzganges und der sehr effizienten Implementierungsmöglichkeit durch ihre Separabilität und die Kaskadierung, macht sich eine unangenehme Nebenerscheinung bemerkbar, sobald über eine größere Umgebung gemittelt werden soll. Die Standardabweichung des Filters und damit die effektive Reichweite der Mittelung wächst nur proportional zur Quadratwurzel von n (1). Das bedeutet, daß man zum Mitteln über eine doppelt so große Umgebung, eine vier mal größere Maske und damit

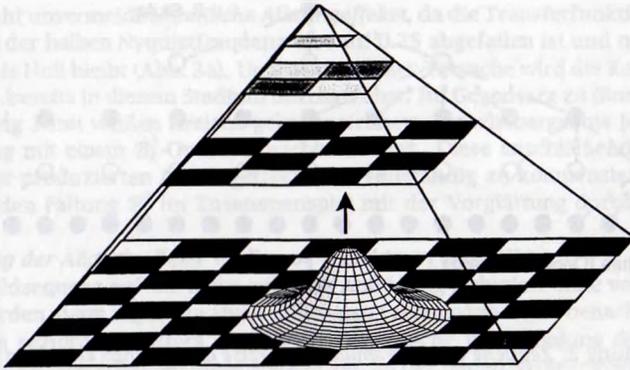


Abbildung 1: Schematischer Aufbau einer Gaußpyramide für 2D-Bilder. Die Quadrate entsprechen einzelnen Pixeln.

den vierfachen Rechenaufwand benötigt! Dies liegt daran, daß sich bei einer sukzessiven Anwendung zweier Binomialmasken die Varianzen addieren und nicht die effektiven Reichweiten und führt schnell zu einem unüberwindlichen Rechenaufwand, wenn man in mehreren Dimensionen oder über lange Bildsequenzen mitteln will.

3.1 ALLGEMEINE GAUSSPYRAMIDEN

Den Schlüssel zum Verringern des Rechenaufwandes liefert eine konsequente Ausnutzung des Abtasttheorems (Jähne, 1993a). Wird von einem vorgegebenen Abstand Δx_i bzw. Δt_i diskreter Grauwertstichproben in Raum und Zeit ausgegangen, so liefert das Abtasttheorem eine Zwangsbedingung an die maximale Frequenz (Nyquistfrequenz) einer periodischen Struktur, die von diesem diskreten Gitter unverfälscht wiedergegeben wird. Umgekehrt erhält man für eine vorgegebene Struktur einen maximalen Abstand der Abtastpunkte. Verringert sich die größte Frequenz der im Bild enthaltenen Strukturen, so vergrößert sich gleichzeitig dieser maximale Abstand. Genau dieser Fall tritt bei einer Glättung eines Bildes ein. Werden hohe Frequenzen durch Glättungsmasken so stark unterdrückt, daß nur noch Frequenzen unterhalb der halben Nyquistfrequenz vorhanden sind, dann wird der dazugehörige maximale Abstand der Abtastpunkte verdoppelt. Ohne Verlust an Information kann danach jeder zweite Bildpunkt bzw. jedes zweite Bild (bei zeitlicher Glättung) verworfen werden. Wird dieser Vorgang sukzessive fortgeführt, so erhält man eine Bildstruktur, in der von einer Ebene zur nächsten die Anzahl der Gitterpunkte und die maximale Auflösung halbiert wird: eine sogenannte *Gaußpyramide* (Abb. 1).

Durch eine Reduktion der Bildstruktur auf die halbe Auflösung kann die Zusammensetzung großer Binomialmasken nun viel effizienter gestaltet werden. Auf dem gröberen Gitter hat eine Glättungsmaske eine vier mal große Varianz als auf dem feineren Gitter. Ein Mitteln der auflösungsreduzierten Bildstruktur mit der gleichen Binomialmaske entspricht damit einer Maske mit doppelter Reichweite.

3.2 AUFBAU DURCH GEZIELTES UNTERABTASTEN

Um eine Gaußpyramide aufzubauen, muß zunächst eine geeignete Glättungsmaske gefunden werden, die alle Frequenzen oberhalb der halben Nyquistfrequenz entfernt. Die von (Burt, 1984)

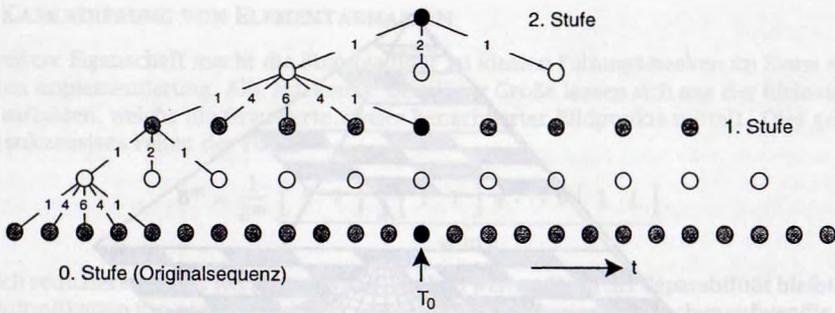


Abbildung 2: Zeitliche Gaußpyramide. Die Kreise entsprechen einzelnen Bildern.

zur Berechnung von Pyramiden in Ortsauflösung vorgeschlagene B^4 -Glättungsmaske ist zu klein, um diese Voraussetzung zu erfüllen (Abb. 3a). Zur Vermeidung von Aliasingeffekten wird mindestens eine B^{16} -Binomialmaske mit 17 Filterkoeffizienten benötigt. In (Jähne, 1993b) wird ein Verfahren zur effektiven Iteration großer 2D-Filtermasken durch Verwendung kleinerer Masken beschrieben, das den Rechenaufwand noch weiter reduziert. Dabei werden kleine Faltungsmasken auf größere Gitterabstände gespreizt. Dadurch entstehende Aliasingeffekte werden verhindert, wenn durch zusätzliche Glättungen mit kleineren Faltungsmasken bei diesen Frequenzen bereits Nullstellen der Transferfunktionen vorliegen. In zwei Dimensionen stehen dazu zwei Parameter der Glättungsmasken zur Verfügung. Sie können skaliert und gleichzeitig unter verschiedenen Winkeln angewendet werden. Um nicht die Effizienz der Separabilität zu verlieren, soll nun ein entsprechendes Verfahren für eindimensionale Faltungen gefunden werden. Damit wird die Form der Glättungsmasken unabhängig von der Dimension der Bildstruktur und kann entlang der Achsen aller Dimensionen angewendet werden, um eine mehrdimensionale Glättung durchzuführen. Obwohl für die eindimensionale Glättung keine Möglichkeit besteht, die Richtung der Faltung zu ändern, kann eine sinnvolle Zerlegung der Binomialmasken durchgeführt werden. Dies geschieht durch *gezieltes Unterabtasten* unter Ausnutzung der Nullstellen der Transferfunktionen. In Anlehnung an den im Anschluß gezeigten Anwendungsfall wird im folgenden der Aufbau einer Pyramide in Richtung der Zeitachse einer Bildsequenz aufgezeigt. Abbildung 2 zeigt schematisch den Aufbau der Gaußpyramide.

Dabei werden folgende Masken verwendet:

$$\begin{aligned}
 B_t^4 &= \frac{1}{16} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix} \\
 B_t^2 &= \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \\
 B_{2t}^2 &= \frac{1}{4} \begin{bmatrix} 1 & 0 & 2 & 0 & 1 \end{bmatrix}.
 \end{aligned}$$

Definiert man analog zu (Jähne, 1993a) den *zeitlichen Reduktionsoperator* R_t , der nur jeden zweiten Bildpunkt bzw. jedes zweite Bild einer Bildfolge herausgreift, so entspricht die gesamte Glättungsmaske B_t^{casc} zum Aufbau einer Pyramidenstufe der *Filterkaskade*

$$B_t^{casc} = B_{2t}^2 B_t^4 = B_t^2 R_t B_t^4. \quad (6)$$

Als Ausgangspunkt der *Filterkaskade* dient eine Glättung der Originalsequenz mit dem Operator B_t^4 . Eine anschließende Reduktion dieser geglätteten Bildfolge auf die halbe zeitliche Auflösung

mit \mathcal{R}_1 verursacht unvermeidbar zeitliche Aliasingeffekte, da die Transferfunktion dieser kleinen Filtermaske bei der halben Nyquistfrequenz erst auf 0.25 abgefallen ist und noch weit darüber hinaus größer als Null bleibt (Abb. 3a). Ungeachtet dieser Tatsache wird die Reduktion der Bildsequenz mit \mathcal{R}_1 bereits in diesem Stadium durchgeführt. Im Gegensatz zu (Burt, 1984) wird dieses, in Abbildung 2 mit weißen Kreisen gekennzeichnete Zwischenergebnis jedoch durch eine weitere Glättung mit einem \mathcal{B}_1^2 -Operator nachbearbeitet. Diese anschließende Filterung dient dazu, die vorher produzierten Aliasingeffekte fast vollständig zu kompensieren. Die Wirkung der nachfolgenden Faltung \mathcal{B}_1^2 im Zusammenspiel mit der Vorglättung durch \mathcal{B}_1^4 ist in Abb. 3 illustriert.

Lokalisierung der Aliasingeffekte im Frequenzspektrum (Abb. 3b)

Wird eine Bildsequenz auf die halbe zeitliche Auflösung reduziert ohne vorher ausreichend geglättet zu werden, dann ragen die abgeschnittenen Frequenzen in die benachbarte Elementarzelle des neuen reziproken Gitters. Durch die periodische Wiederholung des Frequenzspektrums im Fourierraum erscheinen diese Frequenzen an der neuen Nyquistfrequenz bei $\tilde{\omega} = 1/2$ gespiegelt. Dort addieren sich im Spektrum die verfälschten Frequenzen zu den bereits vorhandenen korrekten Frequenzen hinzu.

Korrektur der Aliasingeffekte durch Nachglättung (Abb. 3c)

Eine Faltung der Bildsequenz mit \mathcal{B}_1^2 nach der Reduktion entspricht einer Faltung mit \mathcal{B}_{2t}^2 vor der Reduktion. Da der Operator auf einem Gitter mit doppeltem Gitterabstand arbeitet, sieht man die periodische Wiederholung der Transferfunktion bereits in der Elementarzelle des ursprünglichen Gitters. Der \mathcal{B}_1^2 -Operator alleine wäre daher kein guter Glättungsoperator, da er hohe Frequenzen identisch erhält. In einer Filterkaskade ergänzt er jedoch ideal den \mathcal{B}_1^4 -Operator. Die Transferfunktion von \mathcal{B}_1^2 hat bei der kritischen Grenzfrequenz $\tilde{\omega} = 1/2$ eine Nullstelle und ist in einer Umgebung dieser Frequenz, in welcher der Hauptteil der verfälschten Frequenzen liegt, kleiner als 10%. Der Anstieg oberhalb von $\tilde{\omega} = 1/2$ spielt kaum noch eine Rolle, da die Transferfunktion des \mathcal{B}_1^4 -Operators an diesen Stellen schon fast identisch Null ist. Die Nachglättung der Bildsequenz unterdrückt somit genau die kritischen Frequenzen, bei denen Aliasingeffekte zu erwarten sind. Sie unterdrückt natürlich auch die korrekten Frequenzen, die in diesem Bereich liegen. Dies würde aber auch bei einer Glättung mit einer großen Binomialmaske geschehen, da die Transferfunktion jeder geeigneten Filtermaske bei $\tilde{\omega} = 1/2$ identisch Null sein muß und sich durch die Unschärferelation zwischen Orts- und Fourierraum diesem Wert nur langsam nähern kann. Abbildung 3d zeigt die resultierende Transferfunktion der Filterkaskade im Vergleich zu einem \mathcal{B}_1^4 -Operator. Es ist deutlich zu erkennen, daß die Bereiche oberhalb von $\tilde{\omega} = 1/2$ stark reduziert wurden. Das Maximum des Überschingers ist kleiner als 2% des Maximalwertes bei $\tilde{\omega} = 0$. Dadurch wird Aliasing fast vollständig unterdrückt.

3.3 EFFIZIENZ DER FILTERKASKADE

Eine weitere Eigenschaft der Filterkaskade ist bemerkenswert: Obwohl die resultierende Filtermaske bei direkter Anwendung auf nur 9 Bilder der untersten Pyramidenstufe zugreift, hat sie die gleiche Effektivität wie eine \mathcal{B}_1^{12} -Maske mit 13 Filterkoeffizienten (Abb. 3e). Dies liegt daran, daß die $\mathcal{B}_1^{\text{asc}}$ -Maske weiter auseinanderliegende Bilder stärker wichtet und damit eine größere Varianz hat. In Abschnitt 3.6 wird die effektive Varianz dieser Maske berechnet.

Abbildung 3f zeigt die resultierende Transferfunktion für eine 3- bzw. 4-stufige Pyramide bezüglich des Gitters der Originalsequenz. Die Überschwinger wurden durch die mehrfache Anwendung der Filterkaskade noch weiter reduziert und die Transferfunktion nähert sich immer mehr der einer idealen Gaußmaske an. Dies ist eine direkte Folge der Linearität der Faltung. Nach dem Zentralen Grenzwertsatz der Statistik hat eine lineare Überlagerung vieler Einzelprozesse immer die Normalverteilung als Grenzwert. Die Form der Einzelprozesse ist dabei von un-

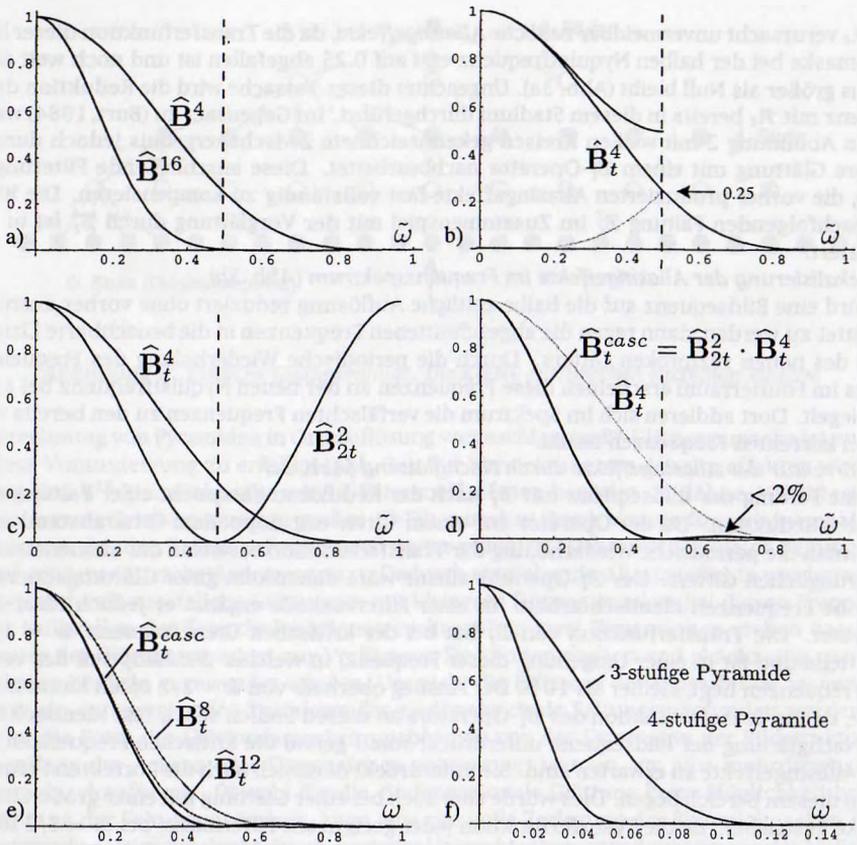


Abbildung 3: Transferfunktion der verwendeten Filterkaskade im Vergleich zu konventionellen Binomialmasken: a) Vergleich $B_t^4 \leftrightarrow B_t^8$. b) Lokalisierung der Aliasingeffekte im Frequenzspektrum. c) Transferfunktionen der beiden Filter, die zum Aufbau der Gaußpyramide verwendet werden. d) Resultierende Transferfunktion der Filterkaskade im Vergleich zu B_t^4 . e) Vergleich der Filterkaskade mit B_t^8 und B_t^{12} . f) Resultierende Transferfunktionen für 3- und 4-stufige Pyramiden.

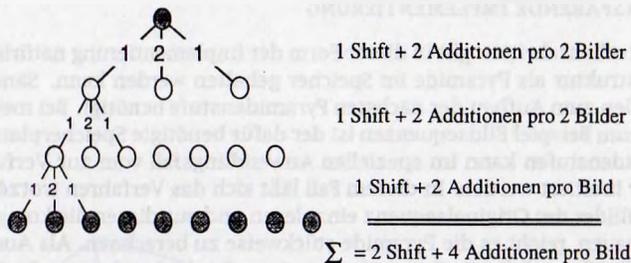


Abbildung 4: Detaillierte Berechnung der Gaußpyramide.

tergeordneter Bedeutung¹. Der Überschwinger an der Transferfunktion von B_i^{asc} zeigt jedoch, daß die Filterkaskade selbst keine lineare Operation ist. Die Nichtlinearität wird durch den Reduktionsoperator R_i verursacht, der eine nichtlineare Operation darstellt. Dies bedeutet, daß die Reihenfolge von Glättungen und Reduktion wichtig ist und nicht vertauscht werden darf.

Gaußpyramide mit B_i^{asc}			Binomialmaske gleicher Varianz	
Stufenzahl	Shift	Additionen	Multiplikationen	Additionen
1	2	4	13	12
2	3	6	61	60
3	3.5	7	253	252
n	$4(1 - 0.5^n)$	$8(1 - 0.5^n)$	$4^{n+1} - 3$	$4^{n+1} - 4$
⋮	⋮	⋮	⋮	⋮
∞	4	8	∞	∞

Tabelle 1: Anzahl der Operationen pro Bild für Gaußpyramiden verschiedener Stufen im Vergleich zu einer direkten Faltung mit Binomialmasken gleicher Varianz.

Die gesamte Effizienz der Filterkaskade wird deutlich, wenn man die B_i^4 -Maske weiter in 2 B_i^2 -Masken zerlegt und die Reduktion so früh wie möglich durchführt (Abb. 4). Die Berechnung einer nachfolgenden Pyramidenstufe benötigt nur 2 Shift Operationen und 4 Additionen pro Bild im Gegensatz zu 9 Multiplikationen und 8 Additionen pro Bild bei einer direkten Berechnung der Filterkoeffizienten! Für jede nachfolgende Pyramidenstufe reduziert sich durch die halbierte Auflösung die Anzahl der zusätzlich benötigten Operationen jeweils um die Hälfte. In Tabelle 1 ist für verschiedene Stufen die Anzahl der Filteroperationen pro Bild der untersten Pyramidenstufe im Vergleich zu einer direkten Berechnung der Filterkoeffizienten zusammengestellt. Während die Anzahl der Operationen und somit die Rechenzeit bei der direkten Berechnung mit der Stufenzahl stark ansteigt, konvergiert sie bei der Filterkaskade schnell und führt dazu, daß eine beliebig starke Glättung einer Bildsequenz mit maximal 4 Shift Operationen und 8 Additionen pro Bild der untersten Pyramidenstufe durchgeführt werden kann!

¹Dies ist auch der Grund dafür, daß Binomialmasken durch Faltung einzelner Rechteckmasken miteinander aufgebaut werden können.

3.4 SPEICHERSPARENDE IMPLEMENTIERUNG

Die Anzahl an Rechenschritten gilt in dieser Form der Implementierung natürlich nur, wenn die gesamte Datenstruktur als Pyramide im Speicher gehalten werden kann. Sämtliche Zwischenergebnisse werden zum Aufbau der nächsten Pyramidenstufe benötigt. Bei mehrdimensionalen Bilddaten, wie zum Beispiel Bildsequenzen ist der dafür benötigte Speicherplatz enorm. Die Anzahl der Pyramidenstufen kann im speziellen Anwendungsfall vom zur Verfügung stehenden Arbeitsspeicher limitiert werden. In diesem Fall läßt sich das Verfahren trotzdem anwenden.

Anstatt alle Bilder der Originalsequenz einzulesen und aus diesen die komplette Mehrgitterstruktur aufzubauen, reicht es die Pyramide stückweise zu berechnen. Als Ausgangspunkt dienen alle Originalbilder, auf die die Pyramide, von einem Punkt der obersten Stufe ausgehend, auf die Basis zugreift (25 Bilder für die 2-stufige Pyramide in Abbildung 2). Das Ergebnis wird dem mittleren Bild der obersten Stufe zugeordnet. Das Grundprinzip der Mehrgitterstruktur ist nun, daß diese Berechnung nicht für alle benachbarten Bilder durchgeführt werden muß. Es reicht eine neue Pyramide für das nächste Bild aufzubauen, welches der Auflösung der obersten Stufe entspricht, z. B. nur jedes vierte Bild bei einer 2-stufigen Pyramide. Die Ergebnisse für die dazwischenliegenden Bildern können durch Interpolation aus diesen gewonnen werden. Die stückweise Berechnung der Pyramide hat zusätzlich den Vorteil, daß ein großer Teil der Zwischenergebnisse, die beim Aufbau der ersten Teilpyramide berechnet wurden, für die benachbarte Teilpyramide verwendet werden kann. Benachbarte Teilpyramiden überlappen sich stark und unterscheiden sich nur im Randbereich. Nur die Originalbilder dieses Randbereiches müssen neu eingelesen und deren darüberliegende Zwischenergebnisse neu berechnet werden. Da auf der gegenüberliegenden Seite der Pyramide genau die gleiche Anzahl an Bildern und Zwischenergebnissen nicht mehr benötigt werden, kann deren Speicherplatz dazu verwendet werden. Eine genaue Berechnung zeigt, daß in dieser Form exakt der gleiche Rechenaufwand benötigt wird wie für den weiter oben beschriebenen Aufbau einer kompletten Pyramide. Der Speicherbedarf ist jedoch deutlich reduziert, da insgesamt nur noch eine Teilpyramide mit Zwischenergebnissen aufbewahrt werden muß.

3.5 2D-PYRAMIDEN

Die Filterkaskade, die in diesem Abschnitt zur effizienten Berechnung eindimensionaler zeitlicher Glättungen entwickelt wurde, kann auch für zwei- und mehrdimensionale räumliche Glättungen verwendet werden. Zum Aufbau einer 2D-Gaußpyramide kann die 1D-Filterkaskade abwechselnd auf Spalten und Zeilen des Bildes angewendet werden. Da bei der ersten Glättung bereits eine Reduktion des Bildes in dieser Richtung stattfindet, muß die zweite Glättung nur noch für die Hälfte der Spalten bzw. Zeilen durchgeführt werden. Bei quadratischen Bildern erhöht sich der Rechenaufwand pro Pixel daher nur auf das 1.5-fache der Operationen in Tabelle 1. Die Transferfunktionen der 2D-Filterkaskade für ein- und zweistufige Pyramiden sind in Abbildung 5a und 5b dargestellt. Die Abbildungen 5c und 5d zeigen einen vergrößerten Ausschnitt in z-Richtung von 0 - 0.05. Man erkennt die Überschwinger bei hohen Wellenzahlen, die bereits bei der eindimensionalen Transferfunktion zu sehen waren. Durch diese lokalen Maxima wird die Transferfunktion anisotrop. Die anisotropen Teile sind jedoch kleiner als 2 % und somit vernachlässigbar klein. Der Hauptteil oberhalb von 5 % ist in guter Näherung isotrop.

3.6 EFFEKTIVE REICHWEITE DER GLÄTTUNG

Zum Abschluß soll nun noch gezeigt werden, welcher Ordnung einer herkömmlichen Binomialmaske eine Pyramidenstufe entspricht. Sei B_i^n der Glättungsoperator, der durch eine bestimmte

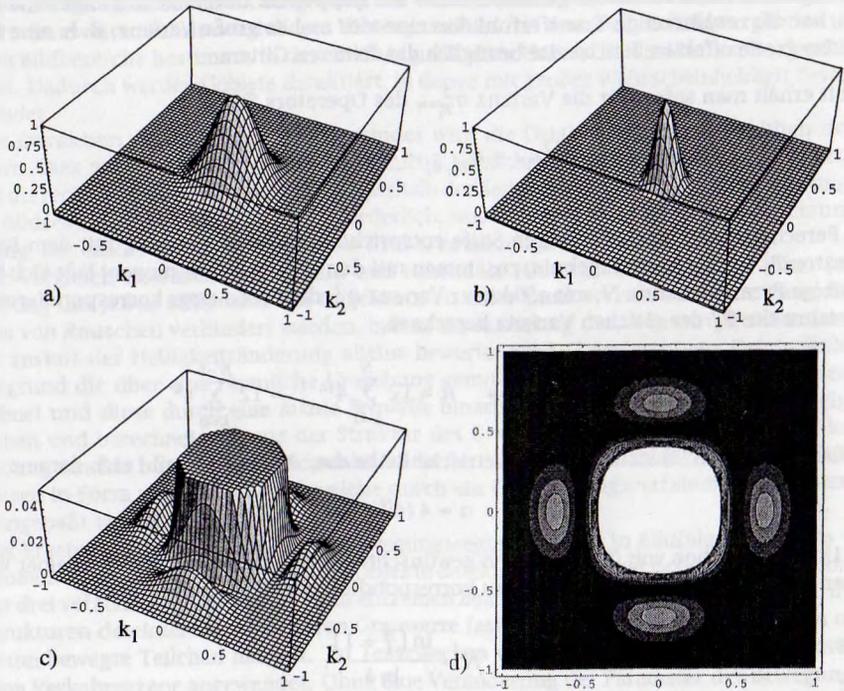


Abbildung 5: Transferfunktionen der Filterkaskade B_i^{acc} zum Aufbau zweidimensionaler Pyramiden: a) Transferfunktion der ersten Pyramidenstufe. b) Effektive Transferfunktion einer 2-stufigen Pyramide. c) Transferfunktion der ersten Pyramidenstufe mit kleinerem Ausschnitt der z-Achse von 0 - 0.05 um die Überschwinger bei hohen Wellenzahlen hervorzuheben. d) Contourplot von b).

Pyramidenstufe N ersetzt werden soll. Um den Zusammenhang zwischen n und N zu finden, wird die effektive Varianz der Filterkaskade $\mathcal{B}_t^{casc} = \mathcal{B}_t^2 \mathcal{R}_t \mathcal{B}_t^1$ mit der Varianz von \mathcal{B}_t^n verglichen. Die effektive Varianz von \mathcal{B}_t^{casc} ergibt sich aus den folgenden beiden Eigenschaften von Binomialmasken:

- Zwei Glättungen mit Binomialmasken entsprechen einer einzigen Glättung mit einer größeren Binomialmaske, deren Varianz durch die Summe der Varianzen der beiden Glättungsmasken gegeben ist.
- Wird ein Filter auf einem größeren Gitter mit doppeltem Gitterabstand angewendet, dann hat die resultierende Transferfunktion eine vier mal so große Varianz, d. h. eine doppelt so große effektive Reichweite bezüglich des feineren Gitters.

Damit erhält man sofort für die Varianz $\sigma_{\mathcal{B}_t^{casc}}^2$ des Operators \mathcal{B}_t^{casc} :

$$\sigma_{\mathcal{B}_t^{casc}}^2 = \sigma_{\mathcal{B}_t^2 \mathcal{B}_t^1}^2 = 4 \underbrace{\sigma_{\mathcal{B}_t^2}^2}_{0.5} + \underbrace{\sigma_{\mathcal{B}_t^1}^2}_1 = 3. \quad (7)$$

Die Berechnung der nächst höheren Stufe entspricht somit einer Glättung mit dem Binomialoperator \mathcal{B}_t^{12} . Mit diesem Ergebnis, zusammen mit den obigen Überlegungen, läßt sich für eine beliebige Pyramidenstufe N , mit effektiver Varianz σ_N^2 , die Größe einer korrespondierenden Binomialmaske \mathcal{B}_t^n der gleichen Varianz berechnen:

$$\sigma_{\mathcal{B}_t^n}^2 = \sigma_N^2 \iff n = 12 \sum_{k=1}^N 4^{(k-1)} = 12 \sum_{k=0}^{N-1} 4^k. \quad (8)$$

Der letzte Term in (8) stellt eine geometrische Reihe dar. Aufgelöst ergibt sich daraus:

$$n = 4(4^N - 1). \quad (9)$$

Die Umkehrfunktion von (9) liefert den gewünschten Zusammenhang zwischen einer vorgegebenen Binomialmaske \mathcal{B}_t^n und der dazu korrespondierenden Pyramidenstufe N :

$$N = \frac{\ln\left(\frac{n}{4} + 1\right)}{\ln 4}. \quad (10)$$

4 KURZBESCHREIBUNG EINES ANWENDUNGSBEISPIELS

Ein Anwendungsfall, für dessen Realisierung eine Filterung über ausgedehnte Bildsequenzen notwendig wird, ist die Beobachtung der Ablösung von Sedimenten in fließenden Gewässern. Die Wasserbewegung im Untergrund kann zur Verlagerung von Bodenpartikeln und damit zu Boden-deformationen führen. Begünstigt wird diese Partikelverlagerung an Schichtgrenzen zweier Bodenarten oder der Grenze von Boden zu festen oder flexiblen Bauteilen (Köhler, 1993). 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. Diese werden mit Hilfe eines starren Endoskopes mit integrierter Glasfaserbeleuchtung und angeflanschter Videokamera direkt aus tieferen Sedimentschichten gewonnen.

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 verwendete Verfahren (Haußecker, Jähne, 1993) gliedert sich 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.

Zur Extraktion des ruhenden Hintergrundes wird die Originalbildsequenz zeitlich tiefpaßgefiltert. Dies geschieht durch die in Abschnitt 3.1 beschriebene zeitliche Mehrgitterstruktur. Durch die recht langsamen Bewegungen innerhalb der Sedimentstruktur wird eine Filterung über lange Bildsequenzen (ca. 300 Bilder) erforderlich, was bei einer direkten Implementierung der Glättung, mit einer Faltungsmaske dieser Größe, kaum möglich wäre.

Der Vergleich zwischen Originalbild und ruhendem Hintergrund erfolgt auf einer lokalen Umgebung um jeden Bildpunkt. Im Gegensatz zu einer punktwisen Differenz kann die Detektion von Rauschen verhindert werden, indem die Änderung der Grauwertstruktur um einen Punkt anstatt der Helligkeitsänderung alleine bewertet wird. Dazu wird aus Originalbild und Hintergrund die über eine räumliche Umgebung gemittelte, zeitliche Varianz der Bildsequenz berechnet und diese durch eine *lokale Schwelle* binarisiert. Diese Schwelle ist abhängig vom Bildinhalt und berechnet sich aus der Struktur des Grauwertverlaufes innerhalb einer lokalen Umgebung um den betrachteten Bildpunkt. In die Berechnung der lokalen Schwelle fließt Expertenwissen in Form von *Fuzzy Logik*, welche durch ein Optimierungsverfahren der Problemstellung angepaßt wurde (Haußecker, 1993).

Das Ergebnis ist ein Verfahren zur Bewegungssegmentierung in Bildfolgen, welches selbst mit bloßem Auge schwer zu erkennende Objekte detektiert, sofern sie sich bewegen. Abbildung 6 zeigt drei verschiedene Bilder mit ihren einzelnen Bearbeitungsschritten. Man erkennt, daß dabei Strukturen detektiert werden, deren Grauwerte fast im Rauschen untergehen, wenn es sich dabei um bewegte Teilchen handelt. Zu Testzwecken wurde der Algorithmus noch zusätzlich auf eine Verkehrsszene angewendet. Ohne eine Veränderung der Parameter der Bewegungssegmentierung werden dort im rechten oberen Viertel deutlich zwei Fahrzeuge und im rechten unteren Viertel ein Fußgänger detektiert (Abb. 6c).

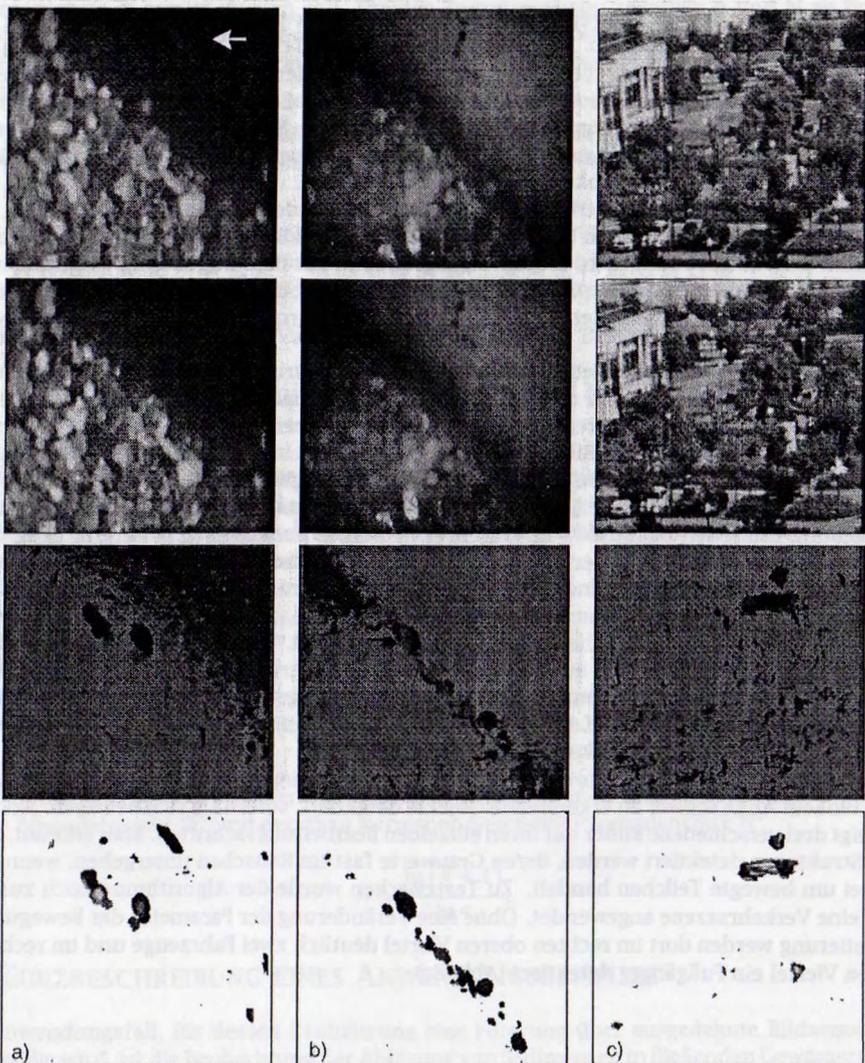


Abbildung 6: Beispiele zur Bewegungssegmentierung: Die drei Spalten zeigen verschiedene Beispielbilder. a) + b): Querschnitt einer Sandgrenzschicht. Der Bildausschnitt beträgt etwa 1 cm in horizontaler und vertikaler Richtung. c): Verkehrsszene. Von oben nach unten: Originalbild, Hintergrund, lokale Schwelle und Ergebnis der Segmentierung. Der Pfeil in Originalbild a) markiert ein Sandteilchen, welches sich kaum vom Hintergrund abhebt, aber im Ergebnis über seine gesamte Fläche detektiert wird.

5 ZUSAMMENFASSUNG UND DISKUSSION

In diesem Beitrag wird ein Verfahren vorgestellt um rechenaufwendige Glättungen effizient zu implementieren. Grundlage des Verfahrens bilden Binomialmasken auf verschiedenen räumlichen Auflösungsstufen.

Da die Glättung mit Binomialmasken gewöhnlich als elementare Grundlage der Bildverarbeitung gilt, wird im allgemeinen kein Versuch unternommen, diese Operation weiter zu untersuchen. Die mathematischen Eigenschaften der Binomialmasken erlauben es bereits die benötigten Rechenschritte deutlich zu reduzieren. Es besteht aber weiterhin das Problem, daß der Rechenaufwand quadratisch mit der Maskengröße anwächst und nicht linear, wie es wünschenswert wäre.

Eine genaue Untersuchung der Form der Transferfunktionen zeigt jedoch neue Möglichkeiten auf, selbst Glättungen über große räumliche Entfernungen effizient zu implementieren. Man erreicht eine Reduktion des Rechenaufwandes weit über das gewünschte Maß hinaus, wobei die Anzahl der Rechenoperationen pro Datenelement sehr schnell konvergiert und somit unabhängig von der Größe der Glättung ist.

Dieses Beispiel soll zeigen, wie eine kritische Überprüfung der Effizienz bisher verwendeter Algorithmen oft ungeahnte Möglichkeiten aufdecken kann ein Verfahren zu beschleunigen. Es soll dazu ermutigen selbst eingefahrene Bildverarbeitungsroutinen bis hin zu den elementarsten Operationen auf eventuelle Einsparungen an Rechenschritten zu untersuchen. Jede dadurch gewonnene Geschwindigkeitssteigerung spart kostenaufwendige Spezialhardware und läßt sich schnell realisieren.

Mit den Binomialmasken wurde ein elementarer Baustein des allgemeinen Filterbaukastens beschleunigt. Es bleibt zu hoffen, daß neben der Verbesserung bisher bekannter Verfahren der Bildverarbeitung und der Suche nach neuen Ansätzen, die Effizienz der Bildverarbeitungsoperatoren nicht vernachlässigt wird, auch wenn die rasante Steigerung der Performance von Standardhardware in den letzten Jahren dazu verleitet.

LITERATUR

- Burt, P. J.: *The pyramid as a structure for efficient computation*. Multiresolution image processing and analysis, A. Rosenfeld (Ed.), Springer Series in Information Sciences, Vol. 12, Springer, New York, 1984.
- Haußecker, H; Jähne, B.: *Ein Mehrgitterverfahren zur Bewegungssegmentierung in Bildfolgen*. Proc. 15. DAGM Symp. Mustererkennung 1993, S.J. Pöpll (Ed.), pp. 24-31, Springer, Berlin 1993.
- Haußecker, H: *Mehrgitter-Bewegungssegmentierung in Bildfolgen mit Anwendung zur Detektion von Sedimentverlagerungen*. Diplomarbeit, Institut für Umweltphysik, Universität Heidelberg, 1993.
- Jähne, B.: *Digital Image Processing - Concepts, Algorithms and Scientific Applications*. 2. Auflage, Springer, Berlin, 1993.
- Jähne, B.: *Spatio-Temporal Image Processing, Theory and Applications*. Lecture Notes in Computer Science, Vol. 751, Springer, Berlin, 1993.
- Köhler, H.-J.: *The influence of hydraulic head and hydraulic gradient on the filtration process*. Bundesanstalt für Wasserbau Karlsruhe 1993.