

PRIP-TR-1

14. Dezember 1993

DIGITALES SEHEN MIT BILDPYRAMIDEN

W. G. Kropatsch

Abstract

Digital Computers are nowadays capable to acquire and to process visual information. The research area of 'Computer Vision' has as its goal to teach the machine abilities comparable to the human eye. At first we shall consider characteristical performance achieved by natural vision in order to derive constraints for digital vision. Then the analysis of digital images is classified into different classes of processing steps ordered according to the different degree of abstraction of the involved descriptions. An image pyramid is presented as a hierarchical data structure which is able to contain both images of various resolutions as well as image descriptions of various degrees of abstraction. Several realizations of pyramidal systems do exist and have shown that simple tasks of image analysis can be performed in real time. In dual pyramids a curve pyramid complements a pyramid of images. Some examples of this concept indicate the potential that the model of the pyramid still offers for computer vision.

Zusammenfassung

Digitale Computer sind heute in der Lage, visuelle Information zu erfassen und zu verarbeiten. Ein Ziel des Forschungsbereichs 'Computer Vision' ist es, der Maschine dem menschliche Auge ähnliche Fähigkeiten beizubringen. Als Vorgabe für Digitales Sehen werden zunächst einige Leistungsmerkmale des natürlichen Sehens betrachtet. Danach wird die Analyse von digitalen Bildern in Klassen von Verarbeitungsschritten eingeteilt, die nach dem Abstraktionsgrad der beteiligten Beschreibungen geordnet ist. Die Bildpyramide wird vorgestellt als eine hierarchische Systemarchitektur, die sowohl verschiedene Auflösungen eines Bildes als auch verschiedene Abstraktionsgrade von Bildbeschreibungen aufnehmen kann. Einige Realisierungen von Pyramidensystemen existieren und haben gezeigt, daß für einfache Bildanalyseaufgaben Echtzeitforderungen erfüllt werden können. Beispiele mit dem Konzept dualer Pyramiden, wo eine Bildpyramide durch eine Kurvenpyramide ergänzt wird, deuten die Möglichkeiten an, die das Modell der Pyramide in Zukunft für das digitale Sehen hat.

1 Vorbild Natur

Das Auge ist eines der wichtigsten Sinnesorgane, mit dem Mensch und Tier in der Lage sind, ihrer Umgebung Informationen zu entnehmen. Als Sehen bezeichnet man jenen Prozeß, der die durch das Auge empfangenen bildlichen Sinnesdaten in Empfindungen umsetzt. Sie entstehen durch die Einwirkung von Licht auf die etwa 130 Millionen Rezeptoren des menschlichen Auges. Dieses Licht wird von den Gegenständen der Umwelt reflektiert und dabei von ihnen in spezifischer Weise verändert. Z.B. bestimmt die Absorption gewisser Wellenlängen die Farbe des Gegenstandes; und die räumliche Lage der Oberfläche drückt sich in bezug auf Lichtquelle und Auge durch die Intensität des reflektierten Lichtes aus. Durch solche Veränderungen des Lichts wird eine Beziehung zwischen den Gegenständen der Umwelt und den Sinnesempfindungen hergestellt. Dieser Vorgang versetzt Mensch und Tier in die Lage, die Umwelt wahrzunehmen und in ihr dynamisch zu agieren.

Die Effizienz dieses Sehsystems wird durch einige experimentelle Meßwerte deutlich: So benötigt ein Mensch für die Wahrnehmung eines einfachen Szenenausschnitts, z.B. des Porträts einer bekannten Person in einem dunklen Raum, etwa 30 Millisekunden. Eine komplexe Szene dagegen, z.B. eine belebte Straßenkreuzung, braucht etwa 1 Sekunde, um wahrgenommen zu werden. Bei einer Schaltzeit von etwa 1-2 Millisekunden zwischen den etwa 12 Milliarden Neuronen des menschlichen Gehirns ergeben sich daraus für einen einfachen Szenenausschnitt etwa 15-30 Verarbeitungsschritte. Für eine komplexe Szene würde man viel mehr Verarbeitungsschritte erwarten, in der Zeit von einer Sekunde sind aber nur einige 100 möglich. Diese biologische Seheffizienz ergibt sich aus der Struktur der Neuronenverbindungen ("neurales Netzwerk", [?]), die eine starke Verflechtung aufweist und dadurch viele gleichzeitig ablaufende und zugleich verschiedenartige Prozesse ermöglicht.

Ein technisches System, z.B. einen 'sehenden' Roboter, mißt man unwillkürlich an den Fähigkeiten des natürlichen Systems. Soll es ähnliche Leistungen wie dieses erbringen, so muß es aus sehr vielen Prozessoren bestehen, obwohl moderne Computer eine wesentlich kürzere Zykluszeit pro Arbeitsschritt aufweisen. Außerdem müssen die Prozessoren miteinander verbunden sein und gleichzeitig ("parallel") arbeiten.

Im Abschnitt 2 wird die zu lösende Aufgabe näher analysiert und in Verarbeitungsstufen gegliedert, die mit bekannten Methoden erreicht werden. Danach wird in Abschnitt 3 eine Verarbeitungsstruktur vorgestellt, die einige der Teilaufgaben in sehr kurzer Zeit bewältigen kann. Außerdem weist diese Rechnerarchitektur große Ähnlichkeit mit ihrem natürlichen/biologischen Vorbild auf [?].

2 Analyse von digitalen Bildern

Wie beim menschlichen Auge entsteht ein digitales Bild durch Abtasten der (projizierten) Bildebene an diskreten Punkten. Die meisten technischen Systeme ordnen diese Abtaststellen regelmäßig in Form eines Quadrat- oder Rechteckrasters an. Jedes der Bildelemente, meist Pixel (engl. 'picture element') genannt, erhält einen oder mehrere Meßwerte (Grauwerte), die die Intensität des auftretenden Lichtes in einer oder mehreren Wellenlängen angeben. Durch Quantisierung auf 64 oder 256 Stufen benötigt ein Grauwert 6 bzw. 8 bit bei der Speicherung. Bei der üblichen Bildgröße von 512×512 Pixel stellt ein digitales Grautonbild eine Datenmenge von 1/4 Megabyte (1 Byte = 8 Bit) dar. Zur Darstellung von Farbe werden 3 Spektralwerte (meist Rot, Grün und Blau) benötigt. Eine weitere Vervielfachung des Datenumfanges wird bei Zeitserien von Bildern erreicht. Um den Eindruck einer kontinuierlichen Bewegung zu erreichen, werden beispielsweise 25-40 Bilder pro Sekunde benötigt! Die dabei anfallenden Datenmengen machen nochmals deutlich, daß die elektronische Datenverarbeitung zur Lösung der Aufgaben der digitalen Bildanalyse ganz neue Wege gehen muß.

Eine abstrakte Aufgabe der Sinnesorgane ist es, eine Beziehung zwischen der realen Welt und einem internen Modell dieser Welt herzustellen. Im technischen System hat dieses interne Modell die Funktion, der Maschine Reaktionen und Entscheidungen aufgrund des optischen Eindruckes zu erlauben. Ein industrielles Sichtsystem soll zum Beispiel in der Lage sein, Fehler auf einer Leiterplatte zu erkennen und auszuscheiden. Dazu ist eine Reihe von Schritten notwendig, die in Abb. 1 dargestellt sind.

Zur richtigen Interpretation eines digitalen Bildes ist es sehr oft wichtig zu verstehen, welche physikalischen Vorgänge zur Entstehung des Bildes führen. Dazu gehört auch die Aufteilung der Welt in unterscheidbare Gegenstände, deren Oberflächen bestimmte optische Eigenschaften haben, die von einem optischen Sensor (z.B. einer Digitalkamera) abgebildet werden.

Die Kenntnis des Bildgebungsprozesses erlaubt es in der Folge, Störungen zu beseitigen, die ihre Ursache im Aufnahmeverfahren haben. Die optische Qualität der Bilder kann durch geeignete Bildoperationen wesentlich verbessert werden (z.B. durch Filterung, Kontrastverstärkung, etc. siehe [?]). Das Ergebnis ist wieder ein digitales Bild, das eine Reproduktion auf einem Bildanzeigergerät erlaubt.

Durch Segmentation wird ein Bild in Regionen unterteilt, die jene Pixel zusammenfassen, die einander bezüglich einer Eigenschaft (hell, dunkel) ähnlich sind (z.B. alle Betonflächen (rot) auf dem Satellitenbild von Graz, Abb. ??). Die Geometrie zusammenhängender Regionen kann durch ihre Kontur beschrieben werden. Weitere geometrische Bildeigenschaften lassen sich aus Eigenschaften der Gegenstände in der Wirklichkeit und aus der Abbildungsgeometrie (Kameraparameter wie Brennweite, usw.) ableiten und werden in der Folge dazu

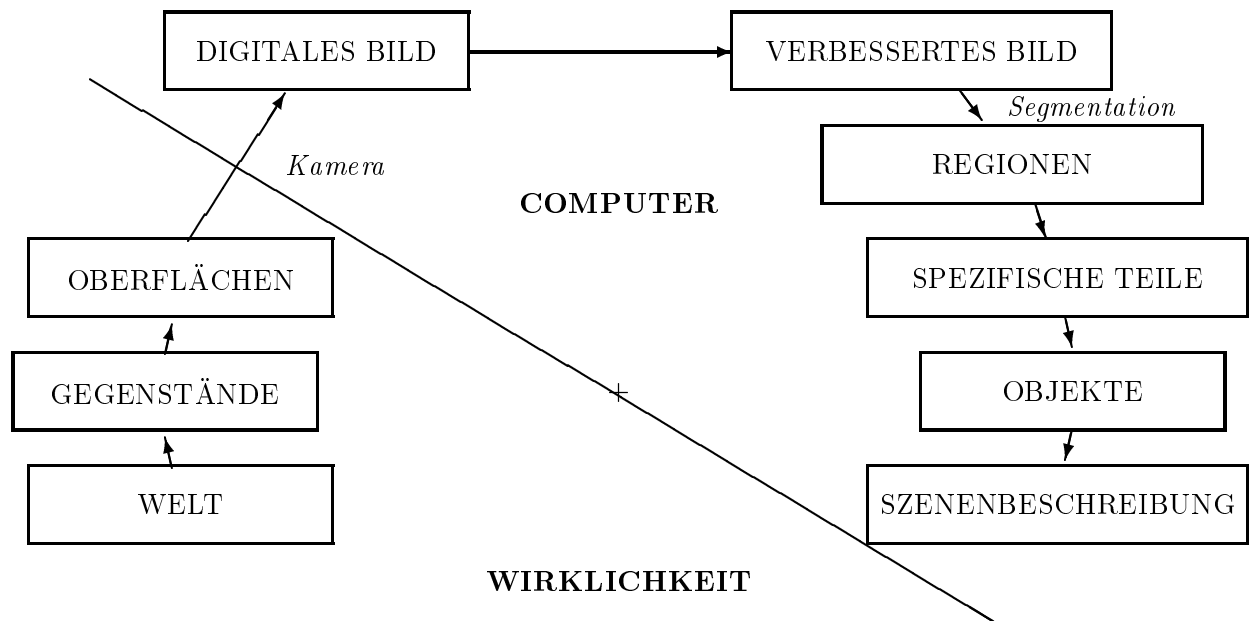


Abbildung 1: Verarbeitungsschritte bei der digitalen Bildanalyse

verwendet, um Regionen als spezifische Teile (z.B. sehr lange, schmale Betonstreifen) zu identifizieren. So haben die Autobahn und der Flughafen im Satellitenbild Graz die Eigenschaft, die längsten zusammenhängenden (länglichen) Betonflächen zu sein (Abb. ??). Die Kenntnis, aus welchen Teilen und Oberflächen sich ein Gegenstand zusammensetzt und unter welchen Randbedingungen die Einzelteile das Objekt bilden (z.B. relative Lage zueinander), erlaubt in ihrer Übertragung auf das Bild die Zusammenfassung der identifizierten Teile zu Objekten. Objekte stellen die Beziehung zum internen Modell der Welt her. Die Beziehung ermöglicht einen Vergleich zwischen dem Bild der Realität und dem Modell, der zum Beispiel dazu verwendet werden kann, das Modell der Wirklichkeit besser anzupassen (z.B. Korrektur einer digitalen Landkarte). Ein Sicherheitsüberwachungssystem wird Alarm schlagen, wenn es eine verdächtige Bewegung im Blickfeld erkennt.

Diese Gliederung der Analyse digitaler Bilder soll die Komplexität der Aufgabenstellung illustrieren und kann gleichzeitig dazu dienen, die verschiedenen Arbeitsschritte zu ordnen. Sie muß jeder speziellen Problemstellung angepaßt werden, es können Einzelschritte wegfallen und auch neue dazukommen.

3 Bildpyramide

æ

Ein wichtiger Parameter eines Bildes ist die **Auflösung**. Ihre Bedeutung sei anhand eines Satellitenbildes und einer Luftaufnahme exemplarisch erläutert. Im Landsat-Thematic-Mapper (TM) Bild der Region Graz (Abb. ??) entspricht einem Pixel im Bild etwa ein

Quadrat mit Seitenlänge $30m$ auf der Erdoberfläche. Die digitale Luftaufnahme des Daedalus-Scanners der NASA (Abb. ??) erzielt hingegen eine Auflösung von rund $2m$. Betrachtet man denselben Ausschnitt der Erdoberfläche, z.B. $10km \times 10km$ mit beiden Systemen, so benötigt die Luftaufnahme 5000×5000 Pixel, ein Landsat TM Bild rund 300×300 Pixel. Der Unterschied beider Systeme liegt in der Datenmenge und in der Qualität der Daten. Die höhere Auflösung der Luftaufnahme gibt den Bildern einen größeren Detailreichtum, Straßen und Häuser sind deutlich erkennbar. Das TM-Bild hingegen gibt einen besseren Überblick, nur breite Straßen wie die Autobahn sind erkennbar.

Die größere Datenmenge der hochauflösenden Luftaufnahme erfordert entsprechend mehr Zeit für die Verarbeitung als das TM-Bild. Das TM-Bild mit niedriger Auflösung hat also den Vorteil, große Objekte mit wenigen Daten zu beschreiben und daher auch schneller zu erkennen, während die höhere Auflösung eine größere geometrische Genauigkeit in Form und Lage von Objekten liefert. Diese Vorteile werden systematisch in einer Bildpyramide vereinigt.

Eine Bildpyramide ist eine geordnete Menge von digitalen Bildern unterschiedlicher Auflösung, die alle denselben Ausschnitt der Realität abbilden. Die Ordnung wird durch die Auflösung definiert, man spricht von den Ebenen der Pyramide. Jedem Pixel der größeren Auflösung entspricht ein gleich großes Fenster (=quadratische Anordnung von benachbarten Pixeln) in der jeweilig feineren Auflösung. Die Größe dieses **Reduktionsfensters** charakterisiert die Struktur der Pyramide [?]. Die **Reduktionsfunktion** faßt die Pixel des Reduktionsfensters zusammen und ordnet das Ergebnis einer Zelle der darüberliegenden Pyramidenstufe zu. Das Bild der jeweils größeren Auflösung hat eine um einen Faktor geringere Anzahl Pixel. Dieser Faktor ist zwischen je zwei benachbarten Ebenen gleich und wird **Reduktionsfaktor** genannt. Er bestimmt gleichzeitig die Flächenreduktion der Auflösung. Dem Faktor 4 entsprechen Bilder mit den Auflösungen in Ebene 1: $30m \times 30m = 900m^2$, in Ebene 2: $60m \times 60m = 3600m^2$, in Ebene 3: $120m \times 120m = 14400m^2$, usw.

Wir bezeichnen die Struktur einer regelmäßigen Pyramide mit F/k , wobei F das Reduktionsfenster und k den Reduktionsfaktor darstellt. Ist $F/k = 1$, so liegen die Reduktionsfenster dicht nebeneinander. Ist $F/k > 1$, so überlappen einander benachbarte Reduktionsfenster. Bei einer überlappenden Pyramide, deren Grundebene endlich und begrenzt ist (z.B. $2 \times 2/2$, Abb. ??), verringert sich die Anzahl der Zellen etwas mehr als durch den Reduktionsfaktor k angegeben. Der Grund dafür sind die Zellen am Rand jeder Ebene, die nur in einem Reduktionsfenster berücksichtigt werden, während alle anderen Zellen in mehreren Reduktionsfenstern liegen. Die Flächenreduktion stimmt aber immer präzise mit dem Reduktionsfaktor überein.

Die klassische Bildpyramide wird auf einem hochauflösenden digitalen Bild folgendermaßen aufgebaut (vgl. Abb. 2). Je 2×2 Pixel des Eingabebildes werden gemittelt und in einem Pixel der nächsten Pyramidenstufe gespeichert. Da die Fläche des 2×2 Reduktionsfensters genau der Fläche des berechneten Pixels der nächsthöheren Pyramidenstufe entspricht, liegen die Reduktionsfenster dicht nebeneinander auf der Grundebene, aber ohne einander zu überlappen. Diese klassische $2 \times 2/4$ Pyramide [?] ist strukturell äquivalent einem vollständigen Quadtree [?], einer vor allem in geographischen Informationssystemen häufig verwendeten Datenstruktur.

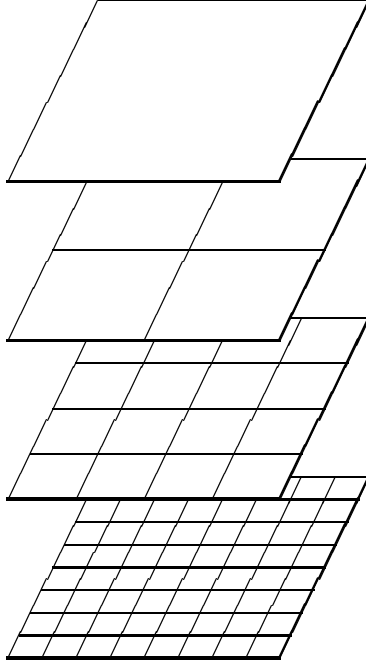


Abbildung 2: Klassische $2 \times 2/4$ Pyramide

4 Numerische Berechnungen in einer Pyramide

Zur Illustration der Effizienz dieser Datenstruktur diene ein kleines Beispiel. Wir wollen die Summe, den Mittelwert und die Varianz von 8 benachbarten Pixeln berechnen (siehe Abb. 3). Zur Bestimmung der Summe der Grauwerte wird als Reduktionsfunktion die Addition (+) verwendet. Die vier Additionen von Ebene 1 zu Ebene 2 können von vier Rechenwerken gleichzeitig in einem Schritt ausgeführt werden. Genauso geschehen die beiden Additionen von Ebene 2 zu Ebene 3 gleichzeitig. Gemeinsam mit der letzten Addition, die das Ergebnis $\boxed{32}$ liefert, sind also 3 Schritte notwendig, um die 8 ($= 2^3$) Grauwerte in der Basis zu addieren.

Der Mittelwert $\underline{\mu}$ berechnet sich als Summe der Grauwerte, geteilt durch die Anzahl n der beteiligten Pixel. Da es sich bei unserer Struktur um eine regelmäßige handelt, ist die Anzahl der an Ebene 2 beteiligten Pixel $n = 2$. An einem Pixel der Ebene 3 sind $n = 4$ Pixel der Grundebene beteiligt, und der Pixel auf Ebene 4 umfaßt alle $n = 8$ Pixel der ersten Ebene. Diese Zahlen lassen sich aber auch sehr einfach berechnen, indem alle Pixel der Ebene 1 auf 1 gesetzt und mit dem zuvor beschriebenen Verfahren die summiert werden. Der Mittelwert $\underline{\mu}$ ergibt sich also in jeder Zelle als Quotient der Grauwertsumme $\boxed{s}$ und der Pixelanzahl n .

Diese einfache Methode der Mittelwertbestimmung hat den Nachteil, in großen Pyramiden sehr große ganzzahlige Grauwertsummen speichern zu müssen. Dies kann vermieden werden, wenn der Mittelwert $\underline{\mu}_i^{(j+1)}$ der höheren Ebene aus den Mittelwerten $\underline{\mu}_{2i}^{(j)}$ und $\underline{\mu}_{2i+1}^{(j)}$ der Ebene darunter als arithmetisches Mittel berechnet wird:

$$\underline{\mu}_i^{(j+1)} = (\underline{\mu}_{2i}^{(j)} + \underline{\mu}_{2i+1}^{(j)})/2. \quad (1)$$

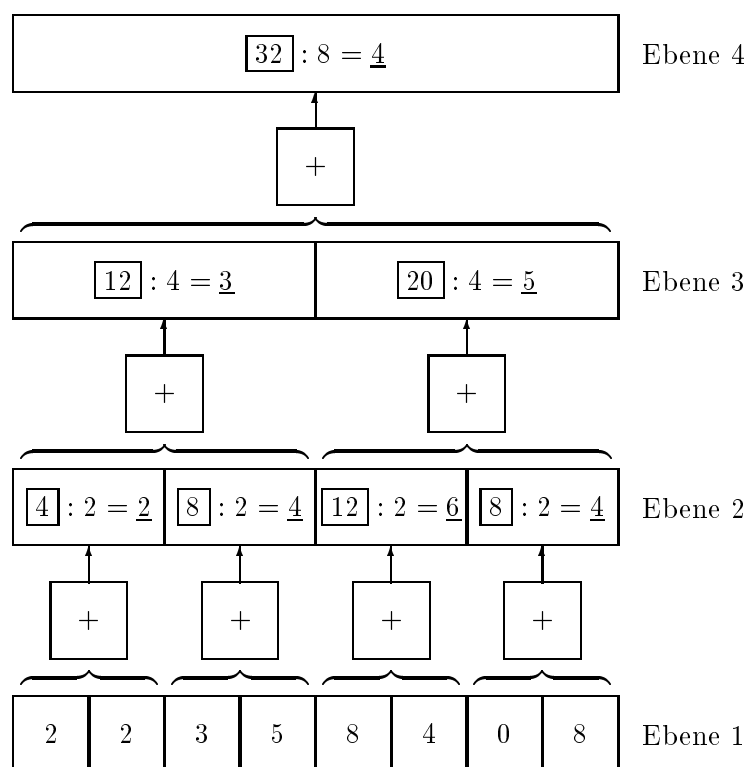


Abbildung 3: Berechnung von Summe $\boxed{s}$ und Mittelwert $\underline{\mu}$ in einer Pyramide

Bei dieser Methode bleibt der Wertebereich der Zelleninhalte beschränkt.

Bei Mittelwertsreduktion erscheint das reduzierte Bild als verkleinerte Version des Originalbildes, dessen Konturen von Ebene zu Ebene unschärfer werden (vgl. Abb. ??). Wie die Schärfe von Konturen erhalten werden kann, wird in Abschnitt ?? beschrieben.

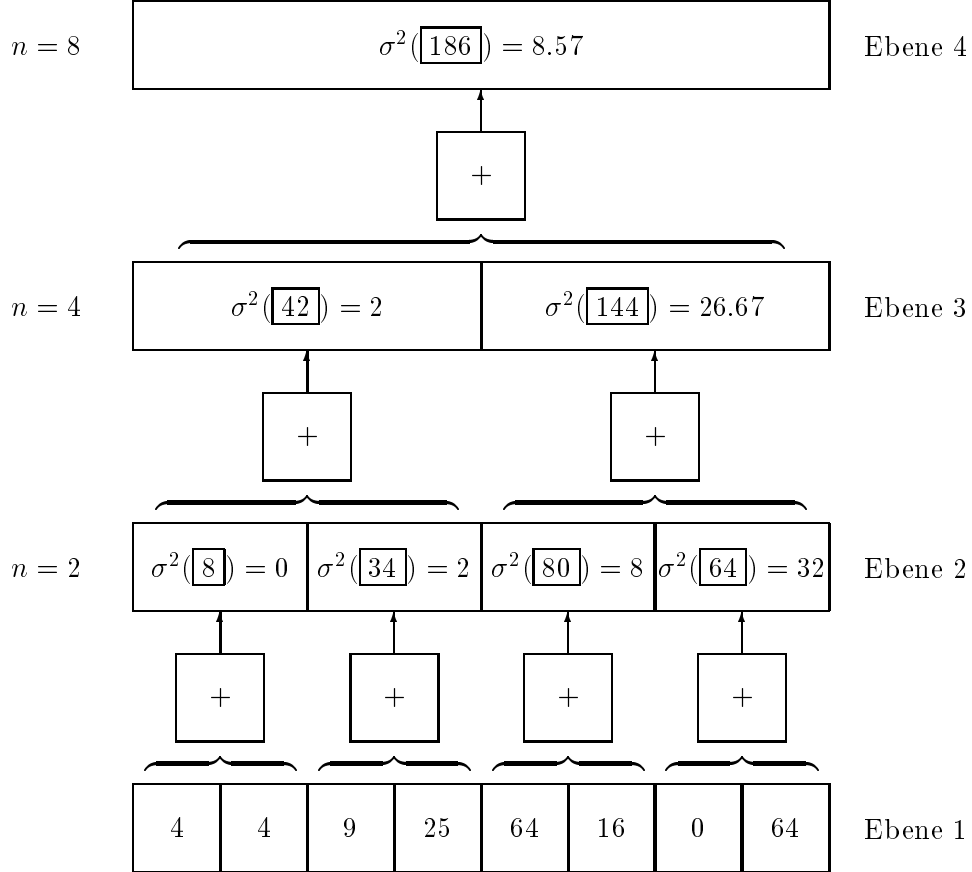


Abbildung 4: Berechnung von Quadratsumme $\boxed{q}$ und Varianz σ^2 in einer Pyramide

Die Varianz σ^2 der Grauwerte eines digitalen Bildes mißt die Abweichungen vom Mittelwert und wird oft als Maß für Textur verwendet. Sie kann nach folgender Formel berechnet werden [?]:

$$\sigma^2 = \frac{1}{n-1} \left(\sum_{j=1}^n x_j^2 - n\mu^2 \right) \quad (2)$$

Die Werte für n und μ wurden bereits vorher bestimmt. Die Summe der Quadrate kann wieder unter Verwendung des Summationsverfahrens berechnet werden, indem zuvor die Quadrate der Grauwerte in Ebene 1 gebildet werden (siehe Abb. 4). Die Werte für σ^2 können dann nach Gleichung 2 in allen Zellen gleichzeitig berechnet werden.

Die Komplexität der ausgeführten Rechenoperationen hängt größenordnungsmäßig vom Logarithmus (zur Basis 2) der Pixelanzahl in Ebene 1 ab ($\lg 8 = 3$). Die Anzahl der Ebenen ist

ausschlaggebend für die benötigten Rechenschritte. Diese Erkenntnis gilt nicht nur für dieses einfache eindimensionale Beispiel, sondern auch und ganz besonders für zweidimensionale Bilder (vgl. Abb. 2). Die klassische $2 \times 2/4$ Pyramide mit $2^n \times 2^n$ Pixeln in Ebene 1 hat nur n Ebenen, der Mittelwert eines 512×512 großen Bildes kann in 9 Schritten berechnet werden.

Ähnlich wie Summe, Mittelwert und Varianz können statistische Momente und Modellapproximationen nach demselben Schema berechnet werden. Eine ausführliche Beschreibung solcher Verfahren zur Schätzung integraler Bildeigenschaften kann in [?] nachgelesen werden.

5 Gauß- und Laplacepyramide

Bei zwei aufeinanderfolgenden Umläufen des Satelliten über derselben Stelle der Erdoberfläche reicht die Aufnahmegenaugkeit nicht aus, um dieselbe Abbildungsgeometrie in beiden Aufnahmen zu erzielen. Die beiden Bilder werden leichte Unterschiede aufweisen, selbst wenn sich auf der Erdoberfläche dazwischen nichts verändert hat. Die klassische $2 \times 2/4$ Bildpyramide hat wie alle nichtüberlappenden Pyramiden die Eigenschaft, daß ihr Inhalt sehr von der relativen Lage ihres Mittelpunkts abhängt. Grauwertunterschiede werden nicht überall gleichmäßig geglättet, sodaß schon bei einer geringfügigen Verschiebung um 1 Pixel in den mittleren Ebenen der Pyramiden größere Unterschiede auftreten können.

Dieser Instabilität der nicht überlappenden Pyramidenstruktur kann durch Vergrößerung der Reduktionsfenster unter Beibehaltung des Reduktionsfaktors begegnet werden. Durch Hinzunahme eines Pixelringes zum 2×2 Reduktionsfensters der klassischen Pyramide entsteht die $4 \times 4/4$ Pyramide (z.B. [?]). Verwendet man in einer überlappenden Pyramide wie vorhin den Mittelwert als Reduktionsfunktion, so treten im Gesamtmittel die Pixel im Zentrum öfter auf als jene am Rand der Pyramide, da die Überlappungen benachbarter Reduktionsfenster im Zentrum stärker sind. Dieses Ungleichgewicht kann durch eine Gewichtung bei der Mittelung ausgeglichen werden. Die Gaußsche Glockenkurve hat für diesen Zweck eine wertvolle Eigenschaft [?]: Die wiederholte Mittelung eines Bildes mit Gaußgewichten kann auch durch einmalige Mittelung mit Gaußgewichten einer flacheren Gaußkurve erzielt werden. Somit sind auch Pixel auf höheren Ebene der **“Gaußpyramide”** Gauß-gewichtete Mittel ihres rezeptiven Feldes der Grundebene. Ein rezeptives Feld einer Pyramidenzelle umfaßt alle Pixel einer Ebene, die einen Einfluß auf den Wert dieser Zelle haben [?].

Alle in dieser Arbeit dargestellten Grauwertpyramiden (Abb. ??, ??, ??) haben eine $3 \times 3/2$ Struktur und verwenden eine Gauß-gewichtete Mittelung als Reduktionsfunktion. Diese Art der Pyramide unterscheidet sich von der klassischen dadurch, daß sie doppelt so viele Ebenen hat und bedingt durch den kleinen Reduktionsfaktor zwischen je zwei Ebenen eine 45° Rotation der Rasterachsen auftritt [?]. Dies wird besonders in ihrer Darstellung deutlich (siehe z.B. Abb. ??), weil auf Rasterschirmen nur achsparallele Raster angezeigt werden können.

Bei Betrachtung einer Gaußpyramide fällt auf, daß die großen Bildmerkmale von der untersten bis in höhere Ebenen auftreten. Diese Redundanz stellt sich als Nachteil bei Speicherung und Übertragung von Bildpyramiden heraus. Die Pyramide soll die Bildmerkmale ihrer Größe nach ordnen. Merkmale von höheren Ebenen sollen nicht auch in tieferen Ebenen auftreten. Dadurch soll ein Mehrfachauftreten desselben Merkmals verhindert und eine