

16. Multiscale Representations

Multiskalenanalyse

Nachbarschaftsoperationen sind nur der Ausgangspunkt für die nachfolgende Bildanalyse, da sie lediglich lokale Merkmale in der Größenordnung von höchstens einigen Pixeln extrahieren. Bilder enthalten aber auch großskalige Information. Zu deren Extraktion werden entsprechend größere Filtermasken benötigt. Allerdings erhöht sich der Rechenaufwand bei der Verwendung großer Filtermasken beträchtlich (eine Verdoppelung der Filtergröße führt zu einem Anstieg der Zahl der Rechenoperationen auf das Vierfache bei zweidimensionalen Bildern). Weiters können Grauwertänderungen durch Kontrastunterschiede zwischen den Objekten und dem Hintergrund nur schwer bestimmt werden. Dieses Problem tritt aufgrund einer Skalenfehlانpassung auf: Die Grauwerte ändern sich über größere Distanzen, welche die Operatoren, die sie detektieren sollen, nicht erfassen. Daher hängt die Detektion bestimmter Merkmale in einem Bild von der richtigen Skala ab. Diese wiederum hängt von den charakteristischen Größen im zu detektierenden Objekt ab. Wenn zum Beispiel Gesichtsdetektion das Ziel der Analyse ist, können Bilder wie im Beispiel rechts nur dann analysiert werden, wenn der Detektionsalgorithmus auf die unterschiedlichen Größen der Gesichter im Bild reagieren kann. Für die optimale Verarbeitung muss ein Bild also in unterschiedlichen Skalen vorliegen, was eine Darstellung in mehreren Auflösungsstufen voraussetzt - die Basis der Multiskalenanalyse. Die Multiskalenanalyse ist ein mathematisches Konzept, das die Signalanalyse auf unterschiedlichen Auflösungsstufen beschreibt. Während die Auswertung von feinen Details im Bild die maximale verfügbare Auflösung benötigt, lassen sich grobe Strukturen mit geringerem Aufwand bei reduzierter Auflösung analysieren.



Abtastung

Zuvor wollen wir nochmals kurz dem Vorgang der Diskretisierung von Signalen genauer auf den Grund gehen. Bekanntlich kann eine kontinuierliche Funktion in eine diskrete Funktion umgewandelt werden, was als Abtastung (*Sampling*) bezeichnet wird. Es ist die Entnahme von Abtastwerten der kontinuierlichen Funktion an bestimmten Punkten in der Zeit oder im Raum, üblicherweise in regelmäßigen Abständen. Um diesen Vorgang in einfacher Weise auch formal beschreiben zu können, wird die Impulsfunktion (Deltafunktion oder Dirac-Impuls) $\delta(x)$ verwendet. Diese Funktion, die einen kontinuierlichen, "idealen" Impuls modelliert, ist überall null mit Ausnahme des Ursprungs, wo ihr Wert zwar ungleich null, aber undefiniert ist, und ihr Integral ist eins:

$$\int_{-\infty}^{\infty} \delta(x) dx = 1$$

Man kann sich $\delta(x)$ als einzelnen Impuls an der Position null vorstellen, der unendlich schmal ist, aber endliche Energie aufweist. Der Abtastvorgang lässt sich mit der Impulsfunktion modellieren, indem eine kontinuierliche Funktion $g(x)$ mit der Impulsfunktion $\delta(x)$ punktweise multipliziert wird, dadurch erhält man einen einzelnen, diskreten Abtastwert der Funktion $g(x)$ an der Stelle $x = 0$.

Durch Verschieben der Impulsfunktion um eine Distanz x_0 kann $g(x)$ an jeder beliebigen Stelle $x = x_0$ abgetastet werden. Die Abtastung einer kontinuierlichen Funktion $g(x)$ an einer Folge von N Positionen $x_i = 1, 2, \dots, N$ kann als Summe der N Einzelabtastungen dargestellt werden:

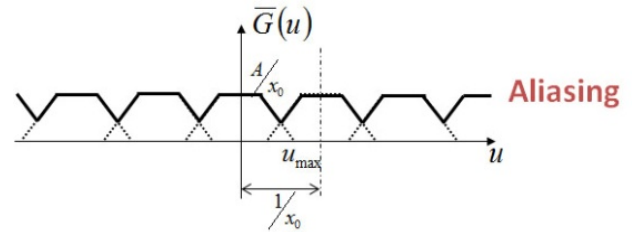
$$\begin{aligned} \bar{g}(x) &= g(x) \cdot [\delta(x-1) + \delta(x-2) + \dots + \delta(x-N)] \\ &= g(x) \cdot \sum_{i=1}^N \delta(x-i) \end{aligned}$$

Diese Summe von verschobenen Einzelpulsen wird als "Pulsfolge" bezeichnet. Wenn die Pulsfolge in beiden Richtungen ins Unendliche erweitert wird, erhalten wir die Kammfunktion oder Shah-Funktion:

$$S(x) = \sum_{i=-\infty}^{\infty} \delta(x-i)$$

Die Abtastung einer kontinuierlichen Funktion hat auch Auswirkungen auf das Frequenzspektrum des resultierenden (diskreten) Signals. Der Einsatz der Kammfunktion als formales Modell des Abtastvorgangs macht es einfach, diese spektralen Auswirkungen zu interpretieren. Die Kammfunktion besitzt, ähnlich der Gauß-Funktion, die seltene Eigenschaft, dass ihre Fouriertransformierte wiederum eine Kammfunktion ist, also den gleichen Funktionstyp hat. Das Produkt zweier Funktionen in einem Raum (entweder im Orts- oder im Spektralraum) entspricht einer linearen

Faltung im jeweils anderen Raum, das Fourierspektrum der Abtastfunktion ist wiederum eine Kammfunktion und besteht daher aus einer regelmäßigen Folge von Impulsen. Die Faltung einer beliebigen Funktion mit einem Impuls $\delta(x)$ ergibt aber wiederum die ursprüngliche Funktion, also: $f(x) * \delta(x) = f(x)$. Das hat zur Folge, dass im Fourierspektrum des abgetasteten Signals $\bar{G}(u)$ das Spektrum $G(u)$ des ursprünglichen, kontinuierlichen Signals unendlich oft, nämlich an jedem Puls im Spektrum der Abtastfunktion, repliziert wird. Das daraus resultierende Fourierspektrum ist daher periodisch mit der Periodenlänge $2\pi/x_0$, also im Abstand der Abtastfrequenz $1/x_0$. Solange sich die durch die Abtastung replizierten Spektralkomponenten in $\bar{G}(u)$ nicht überlappen, kann das ursprüngliche Spektrum $G(u)$ – und damit auch das ursprüngliche, kontinuierliche Signal $g(x)$ – ohne Verluste aus einer beliebigen Replika von $G(u)$ aus dem periodischen Spektrum $\bar{G}(u)$ rekonstruiert werden. Zur Diskretisierung eines kontinuierlichen Signals $g(x)$ mit Frequenzanteilen im Bereich $0 \leq 1/x_0 \leq u_{\max}$ benötigen wir eine Abtastfrequenz u_s , die mindestens doppelt so hoch wie die maximale Signalfrequenz $u_{\max}$ ist. Wird diese Bedingung nicht eingehalten, dann überlappen sich die replizierten Spektralteile im Spektrum des abgetasteten Signals und das Spektrum wird verfälscht mit der Folge, dass das ursprüngliche Signal nicht mehr fehlerfrei aus dem Spektrum rekonstruiert werden kann. Dieser Effekt wird als **Aliasing** bezeichnet.

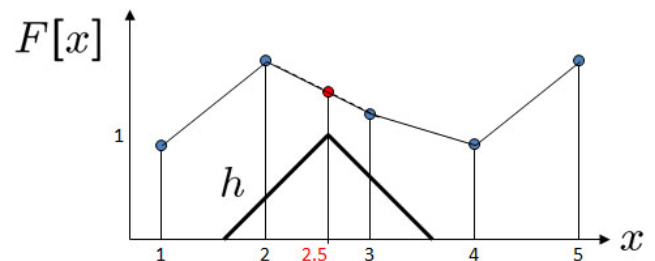


Bildskalierung

Zunächst gehen wir von der Frage aus, wie Bilder größenmäßig angepasst werden können, wenn sie mit ihrer Originalgröße zum Beispiel nicht auf den Bildschirm passen. Wie kann man zum Beispiel ein Bild erzeugen, das nur ein Viertel so groß ist wie das Original (auch *Subsampling* genannt)? Löschen wir einfach jede zweite Zeile und jede zweite Spalte, haben wir ein Abtastproblem, da wir dadurch Aliasing-Effekte erzeugen. Daher müssen die Bilder zuerst mit einem Gaußfilter gefiltert werden, um anschließend verkleinert zu werden, wobei der Filter nach dem Abtasttheorem doppelt so groß wie die gewünschte Verkleinerung sein muss. Dieser Theorie folgen im Weiteren auch alle Betrachtungen der Multiskalenanalyse.

Will man ein Bild vergrößern (*Resampling/ Upsampling*), muss man sich auch wieder die Abtastfunktion ansehen und bei zusätzlich hinzugefügten Datenpunkten deren Werte aus den benachbarten interpolieren. Die diskreten Grauwerte der Matrix des vergrößerten Bildes werden aus benachbarten Pixeln des Ausgangsbildes interpoliert. Die neue Matrix ist geometrisch durch das gewählte Bezugssystem definiert.

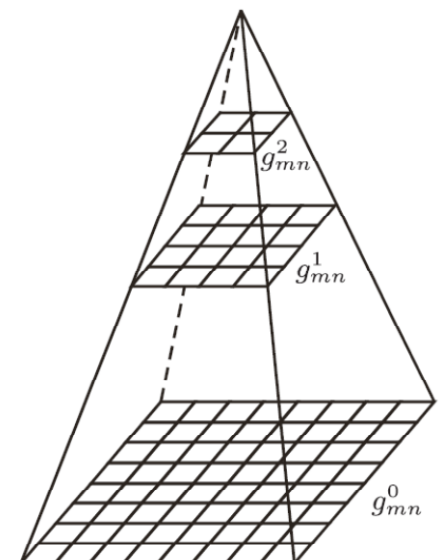
Die neuen Bildelemente setzen sich aus Teilstücken von Bildelementen der Matrix des Eingabebildes zusammen und können daher wiederum durch Filter definiert werden, die genau dieses tun. Es muss daher ein Filterkern definiert werden, nach dem die Grauwertzuweisung erfolgen soll. Dazu sind zwei Resamplingverfahren allgemein verbreitet, **bilineare** sowie **bikubische** Interpolation. Bilineare Interpolation errechnet den Wert einfach aus den Nachbarwerten wie in der Abbildung rechts gezeigt. Die Filterfunktion $h(x)$ wird über $F(x)$ gelegt und erzeugt den neuen Wert. Bei der bikubischen Interpolation geht man von der vergrößerten neuen Bildmatrix aus und rechnet mit Hilfe der Transformationsgleichungen von den Mitten der dortigen Bildelemente in das Eingabebild zurück.



Bei der bikubischen Interpolation geht man von der vergrößerten neuen Bildmatrix aus und rechnet mit Hilfe der Transformationsgleichungen von den Mitten der dortigen Bildelemente in das Eingabebild zurück.

Bildpyramiden

Für viele Aufgaben der Bildauswertung ist es wünschenswert, Strukturen auf unterschiedlichen Skalen effizient zu analysieren, um verschiedene Objekteigenschaften untersuchen zu können und eine gewisse Invarianz bezüglich der Größe zu erreichen. Wendet man jedoch Bildauswertungsverfahren naiv auf mehreren Skalen an, so ist dies mit beträchtlichem Rechenaufwand verbunden. Soll beispielsweise mit einem Korrelationsfilter in einem Bild der Größe $N \times N$ ein Objekt der Ausdehnung $L \times L$ detektiert werden, so entsteht für die Faltung im Ortsbereich ein Aufwand von ca. $O(N^2 \cdot L^2)$. Um den Aufwand für die Untersuchung großer Bildstrukturen zu reduzieren, wurde daher das Konzept der Bildpyramiden eingeführt: Wenn man Signale in unterschiedlichen Skalen verarbeiten will,



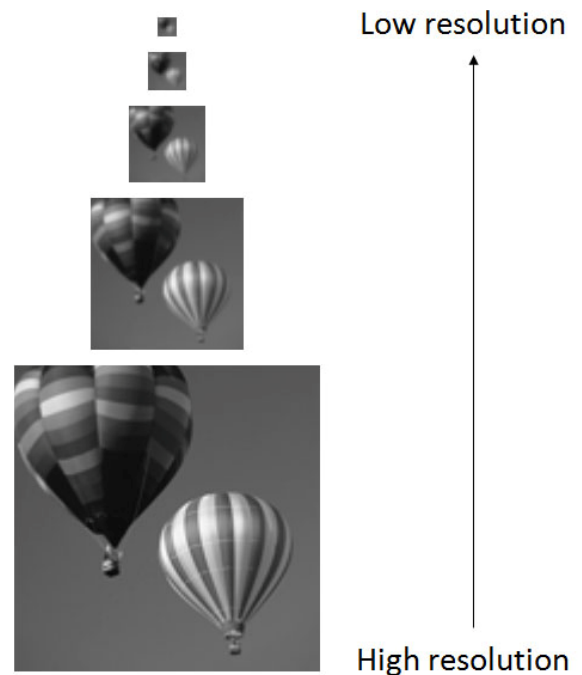
dann kann dies am effektivsten in einer Mehrgitterdarstellung erfolgen. Die zugrunde liegende Idee ist einfach. Während die Darstellung feiner Skalen die volle Auflösung erfordert, reicht für grobe Strukturen eine niedrigere Auflösung. Dies führt zu einer Folge von Bildern, die von Stufe zu Stufe immer kleiner werden. Das wird durch eine iterative Filterung und Unterabtastung des Bildes erreicht. Man erhält eine Pyramide von Bildern, wobei jedes Bild eine bestimmte Skala repräsentiert.

Wollen wir die Größe eines Bildes reduzieren, können wir das Bild nicht einfach abtasten, indem wir jeden zweiten Bildpunkt in jeder zweiten Zeile nehmen (siehe oben). Zum Beispiel würde eine Struktur, die im Originalbild dreimal pro Wellenlänge abgetastet wird, im nächsten Bild nur anderthalbmals abgetastet werden. Es würde sich ein fehlerhaftes Muster ergeben. Also müssen wir sicherstellen, dass alle weniger als viermal pro Wellenlänge abgetasteten Strukturen durch einen passenden Glättungsfilter unterdrückt werden. Für die Erzeugung eines Skalenraums bedeutet das, dass die Größenreduktion mit einer angemessenen Glättung Hand in Hand gehen muss. Der Aufbau einer Bildpyramide für ein (o. B. d. A.) quadratisches Bild g_{mn} der Größe $N \times N$ beginnt mit der Berechnung der darüber liegenden Ebene, die zum Beispiel nur mehr ein Viertel der Originalgröße ausmacht. Zuerst erfolgt eine Tiefpassfilterung von g_{mn}^l auf halbe Bandbreite. Danach erfolgt eine Unterabtastung um den Faktor 2. Wegen der vorangegangenen Tiefpassfilterung stellt dies keine Verletzung der Abtastbedingung dar. Es resultiert das gefilterte und unterabgetastete Bild g_{mn}^l der Größe $N/2 \times N/2$. Iterativ kann nun dieser Vorgang wiederholt werden um die fehlenden Ebenen der Pyramide der Höhe K zu erstellen: $v = 2, \dots, K$; g^{v-1}_{mn}

In der Pyramide liegt das Bild in verschiedenen Auflösungen vor.

Dadurch können sowohl grobe als auch feine Bildstrukturen mittels kleiner lokaler Operatoren ausgewertet werden. Die Anwendung des Operators auf einer oberen Stufe der Pyramide ($v > 0$) entspricht näherungsweise der Anwendung eines virtuell entsprechend vergrößerten Operators auf das Originalbild ($v = 0$). Durch das Pyramidenkonzept gelangen grobe Strukturen sozusagen in Reichweite der lokalen Operatoren. Betrachtet man erneut das Beispiel der Detektionsaufgabe, so sinkt der Aufwand auf der Stufe $v > 0$ der Pyramide auf ca. $O(2^{-4v} \cdot N^2 \cdot L^2)$, denn die Abmessungen von Bild und Objekt verringern sich jeweils um den Faktor 2^{-v} auf $2^{-v}N \times 2^{-v}N$ bzw. $2^{-v}L \times 2^{-v}L$. Der Rechenaufwand verringert sich also mit wachsendem v erheblich. Somit ist auch die Detektion großer Objekte effizient möglich, allerdings um den Preis des Verlustes feiner Details aufgrund der Tiefpassfilterung zwischen den Pyramidenstufen.

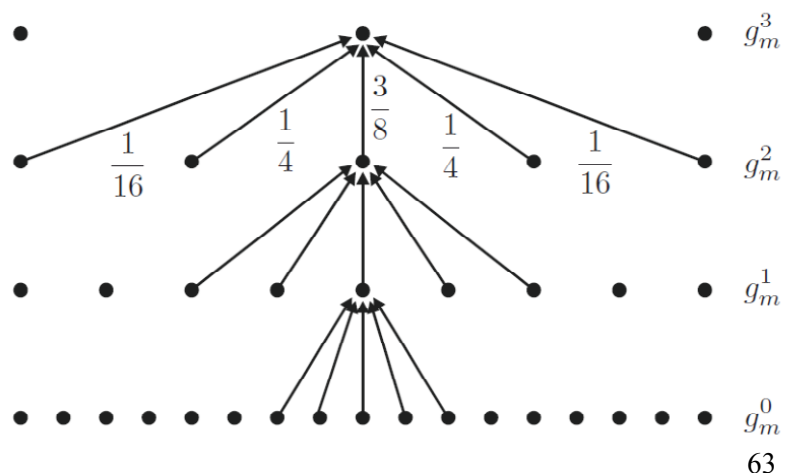
Trotz der Speicherung von Bildinformation auf mehreren Skalen bleibt der Speicheraufwand für eine Bildpyramide überschaubar. Das erste unterabgetastete Bild g_{mn}^1 hat die Größe $N/2 \times N/2$, das zweite g_{mn}^2 hat die Größe $N/4 \times N/4$, usw. Die Gesamtzahl der Bildpixel beträgt $4/3 \cdot N^2$, wie man mit Hilfe der geometrischen Reihe abschätzen kann. Ebenso effektiv ist die Berechnung der Pyramide. Derselbe Glättungsfilter wird auf jede Ebene der Pyramide angewandt. Damit erfordert die Berechnung der gesamten Pyramide lediglich $4/3$ der Rechenoperationen für ein zweidimensionales Bild. Wenn die Pyramide einmal berechnet ist, können wir Nachbarschaftsoperationen mit großen Skalen in den oberen Ebenen der Pyramide durchführen.



Gaußpyramide

Eine mit einem Tiefpass (Gaußfilter) konstruierte Bildpyramide heißt Gaußpyramide. Nebenstehende Abbildung veranschaulicht die Vorgehensweise der Tiefpassfilterung und Unterabtastung um den Faktor 2 an einem eindimensionalen Signal g_m . Das darüber liegende Pixel wird jeweils aus fünf darunter liegenden Pixeln durch eine Gewichtung des 1×5 Gaussfilters $[1/16 \ 1/4 \ 3/8 \ 1/4 \ 1/16]$ erzeugt. Die kombinierte Glättung und Größenreduktion bei der Berechnung der $(v+1)$ -ten Pyramidenstufe aus der v -ten Ebene kann mit einem einzigen Operator ausgedrückt werden:

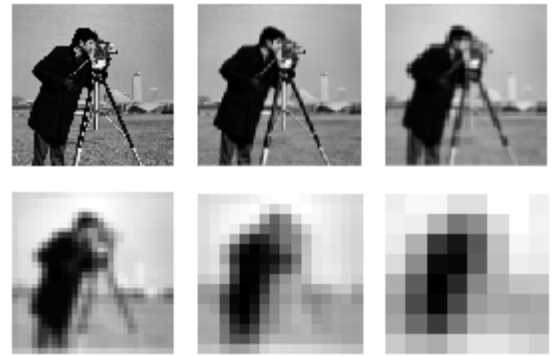
$$G^{(0)} = G, G^{(v+1)} = B_{\downarrow 2} G^{(v)}$$



Die Zahl hinter dem $\downarrow$ im Index bezeichnet die Abtastrate. Die nullte Ebene ist das Originalbild. Die durch mehrmalige Anwendung dieses Operators erhaltene Bildserie wird wie erwähnt als Gaußpyramide bezeichnet. Von einer Ebene zur nächsten nimmt die Auflösung auf die Hälfte ab. Wir können uns diese Bildserie ebenfalls in Form einer Pyramide angeordnet vorstellen.

Die Gaußpyramide stellt eine Serie von tiefpass-gefilterten Bildern dar, bei denen die Grenzfrequenz von Ebene zu Ebene auf die Hälfte (eine Oktave) abnimmt. Damit verbleiben nur zunehmend gröbere Details im Bild. Um den gesamten Bereich der Frequenzen zu umspannen, sind nur wenige Pyramidenebenen erforderlich. Aus einem $N \times N$ -Bild können wir eine Pyramide mit maximal $\log_2(N+1)$

Ebenen berechnen. Das kleinste Bild besteht aus nur einem Bildpunkt. In der Abbildung links wurde jede Ebene wieder auf die Originalauflösung skaliert, um Details an den oberen Ebenen zu zeigen.



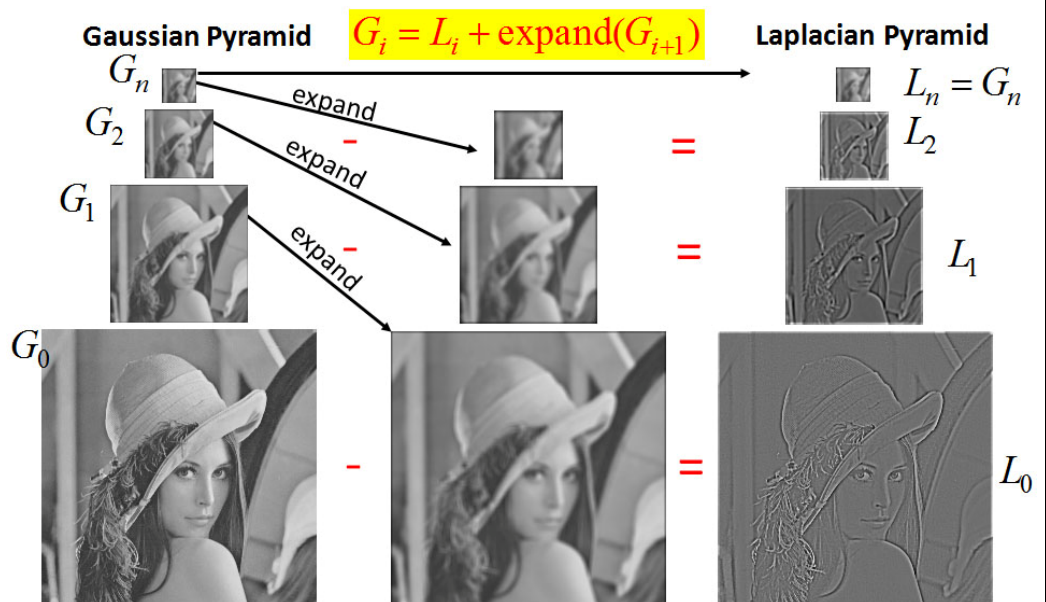
Laplacepyramide

Bisher wurden Bildpyramiden mittels einer Tiefpassfilterung konstruiert. In jedem Schritt werden Signalanteile mit hoher Ortsfrequenz herausgefiltert, sodass auf den oberen Stufen der Pyramide nur noch niedrige Ortsfrequenzen vorhanden sind.

Alternativ kann die Pyramide auch so konstruiert werden, dass ihre Stufen eine Bandpassfilterung des Originalbildes enthalten. Eine einfache Möglichkeit, um eine solche Bandpassfilterung zu realisieren, ist die Subtraktion zweier aufeinanderfolgender Bilder einer Gaußpyramide (Difference-of-Gaussians). Dafür muss das Bild in der größeren Ebene zuerst expandiert werden. Diese Operation wird vom Expansionsoperator $\uparrow 2$ durchgeführt. Wie beim reduzierenden Glättungsoperator wird der Grad der Expansion durch die Zahl nach dem $\uparrow$ -Zeichen im Index angegeben. Die Expansion ist schwieriger als die Größenreduktion, da die fehlende Information interpoliert werden muss. Zur Vergrößerung um den Faktor zwei in allen Richtungen muss jeder zweite Bildpunkt in jeder Zeile interpoliert werden und dann jede zweite Zeile. Die Erzeugung der v -ten Ebene der Laplacepyramide kann folgendermaßen beschrieben werden:

$$L^{(v)} = G^{(v)} - \uparrow 2 \cdot G^{(v+1)}, L^{(v)} = G^{(v)}$$

Die Laplacepyramide stellt die Approximationen der zweiten Ableitung des Originalbildes mit unterschiedlichen Glättungsparametern dar. Auf den ersten Stufen der Laplace-Pyramide werden feine Kantenstrukturen hervorgehoben, während auf den höheren Stufen gröbere Kantenstrukturen des Originalbildes sichtbar sind. Die Laplace-Pyramide ist eine Zerlegung des Bildsignals in Bandpassbereiche mit logarithmischer Frequenzstufung. Das Originalbild kann aus der Laplace-Pyramide sowie dem



obersten Bild der Gauß-Pyramide exakt rekonstruiert werden. Dies geschieht einfach durch eine Umkehrung des Konstruktionsschemas der Laplacepyramide.

Im Gegensatz zur Fouriertransformation führt die Laplacepyramide lediglich zu einer groben Frequenzzzerlegung ohne Richtungszerlegung. Alle Frequenzen innerhalb des Bereiches von ungefähr einer Oktave (Faktor 2) sind unabhängig von ihrer Richtung in einer Ebene der Pyramide enthalten.