

it
informatik

Rainer Malaka
Andreas Butz
Heinrich Hußmann

Medieninformatik

Eine Einführung

Bilder

3.1	Bilder sind überall	86
3.2	Es werde Licht!	87
3.3	Farben im Kopf	89
3.4	Farbaddition, Farbsubtraktion und Farbräume	91
3.5	Weitere Farbmodelle	94
3.6	Digitalisierung und Kodierung	96
3.7	Kompression von Bildern	98
3.8	Bearbeitung von Bildern	107
	Zusammenfassung	113
	Übungen	114

Einleitung

» Digitale Bilder sind neben Texten die am meisten verwendeten digitalen Medien. Als digitale Fotografien, Illustrationen und in zahlreichen zusammengesetzten Medien werden sie von Profis und Laien umfangreich erstellt, bearbeitet und verwendet. In diesem Kapitel sollen die Grundlagen digitaler Bilder vorgestellt werden. Da Bilder meist von Menschen betrachtet werden, ist es wichtig, zunächst einige Phänomene bei der Wahrnehmung von Bildern zu beleuchten. Bei der digitalen Verarbeitung ist die erste Stufe die Digitalisierung von Bildern. Dabei werden die Bilder zunächst in einem zweidimensionalen Raster von Pixeln erfasst. Jedes Pixel speichert entweder Schwarzweißinformationen, Grauwerte oder Farbwerte. Verschiedene Farbmodelle bilden die Farbwerte auf drei bis vier diskrete Werte ab, die unterschiedlich viele Werte je nach Farbtiefe annehmen können. Die wichtigsten Farbmodelle und die ihnen zugrunde liegenden psychophysischen Grundlagen sowie ihre Anwendungsbereiche werden vorgestellt. Damit sind die Grundlagen zur Erfassung und Repräsentation von digitalen Bildern gelegt. Zur Speicherung digitaler Bilder sind effiziente Verfahren notwendig, um das Datenvolumen deutlich zu reduzieren. Verlustfreie und verlustbehaftete Verfahren nutzen unterschiedliche Eigenschaften der Bilder und der menschlichen Wahrnehmung aus, so dass der Speicher- und Übertragungsaufwand drastisch reduziert werden kann. Zur Abrundung des Kapitels werden noch « einige grundlegende Verfahren zur digitalen Bildbearbeitung vorgestellt.

Lernziele

In diesem Kapitel werden zunächst einige Grundlagen der **visuellen Wahrnehmung** vermittelt. Sie erklären zum Beispiel, wie Menschen Farben wahrnehmen. Diese Zusammenhänge erklären, warum man bei Farbmodellen in der Regel drei Grundfarben wie z.B. Rot, Grün und Blau, benötigt. Für digitale Bilder, die aus Pixeln aufgebaut sind, wird dann gezeigt, wie diese digitalisiert und codiert werden können. Damit das Datenvolumen reduziert werden kann, können **Kompressionsverfahren** eingesetzt werden, die verlustfrei oder verlustbehaftet sind. Sie werden am Beispiel von JPEG nachvollziehen können, wie verschiedene Kompressionsmethoden in einem Verfahren zusammenwirken können. Zum Abschluss lernen Sie einige einfache Verfahren zur **Bildbearbeitung**, wie z.B. Filter, kennen.

3.1 Bilder sind überall

Viele Menschen halten das Sehen für den wichtigsten menschlichen Sinn. Zweifellos sind visuelle Eindrücke für Menschen ausgesprochen wichtig, auch wenn wir genau genommen im Vergleich zu vielen Tieren gar nicht so gut sehen können und unsere Sehleistung oft mit dem Alter nachlässt. Unsere Umwelt ist voll von Bildern auf Plakaten, in Zeitschriften und im Web. In den letzten Jahren hat sich bei der Produktion und Verarbeitung von Bildern ein dramatischer Wandel hin zur digitalen Technik vollzogen. Klassische analoge Fotografie auf Negativ- oder Diapositivfilm spielt mit-

terweile im professionellen und privaten Bereich nur noch eine Außenseiterrolle. Digitale Kameras aller Leistungsstufen sind erschwinglich und dank billiger Speichermedien werden so viele Bilder erstellt wie nie zuvor.

Ob digitale Spiegelreflexkamera oder Fotohandy, die Prinzipien bei der Digitalisierung und Speicherung der Bilder sind immer die gleichen. Dabei müssen physikalische Signale gemessen und in einem digitalen Raster abgebildet werden. Die Digitalisierung eines Bildes muss wie bei anderen Medientypen auch eine geeignete Abtastung und Quantisierung leisten. Wenn Bilder als Dateien (z.B. über das Web oder per MMS) verschickt werden sollen, müssen sie kompakt gespeichert und komprimiert werden. Aber viele Fotos werden nicht so, wie sie gemacht werden, weiterverarbeitet, sondern mit vielfältigen Werkzeugen bearbeitet und verändert.

Doch bevor wir genauer auf diese Punkte eingehen, wollen wir uns Bilder erst einmal genauer ansehen und klären, was ein Bild eigentlich ist und was wir sehen, wenn wir ein Bild sehen.

Obwohl wir täglich Hunderte Fotos sehen und unserem Sehvermögen so viel zutrauen, spiegeln uns unsere Augen und unser Gehirn oft etwas vor und vieles ist im wahrsten Sinne des Wortes nicht so, wie es scheint. Der Begriff „fotografieren“ heißt auf Deutsch eigentlich „mit Licht schreiben“. Bei Fotos oder Bildern geht es also eigentlich um Licht, das physikalisch als Welle oder Teilchen gesehen werden kann. Nicht jedes Licht ist gleich. Es kann mehr oder weniger hell sein und mit unterschiedlichen Farben leuchten. Außerdem kann Licht polarisiert sein. Licht kann man mischen und aus bestimmten Grundfarben können andere Farben entstehen. Um zu verstehen, warum das nur scheinbar der Fall ist, müssen wir die Physik des Lichtes und die Mechanismen der Wahrnehmung genauer verstehen.

3.2 Es werde Licht!

Licht ist ein Phänomen, das einige Rätsel aufgibt, die hier nicht alle gelöst werden können. In diesem Kapitel sollen Photonen und die Relativitätstheorie vernachlässigt werden und wir betrachten Licht nur als eine **elektromagnetische Welle**. Eine solche Welle kann von ganz unterschiedlichen Quellen ausgestrahlt werden und bewegt sich geradlinig durch den Raum. ►Abbildung 3.1 zeigt zwei solche Wellen, die sich in x-Richtung ausbreiten. Die Wellen sind durch folgende Eigenschaften charakterisiert: die Ausbreitungsrichtung, die Frequenz oder Wellenlänge, die Amplitude und die Polarisation.

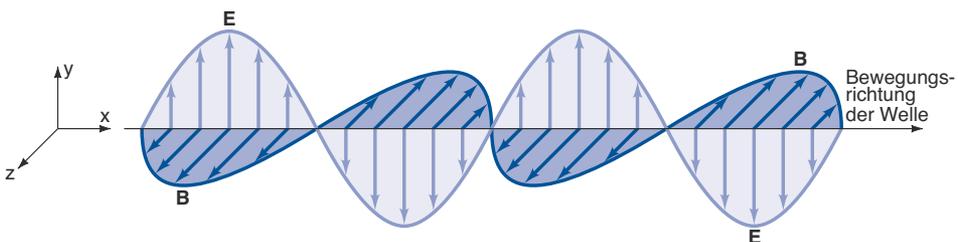


Abbildung 3.1: Zwei Wellen E und B, die sich in x-Richtung ausbreiten

Ausbreitungsrichtung: Elektromagnetische Wellen breiten sich von einer Quelle linear und mit konstanter Geschwindigkeit aus. Je nach Wellenlänge und Beschaffenheit eines Hindernisses kann eine solche Welle von einem solchen Hindernis abgelenkt bzw. reflektiert werden, absorbiert werden oder das Hindernis durchdringen. Im sichtbaren Bereich des Lichts sind solche Hindernisse z.B. Spiegel, schwarze Wände oder Glas. Bei den meisten Quellen, wie zum Beispiel einer Glühbirne, werden diese Wellen gestreut ausgesandt und viele Wellen strahlen in alle möglichen Richtungen. Es gibt aber auch gerichtete Quellen wie zum Beispiel beim Laser, wo viele Wellen gebündelt alle in die gleiche Richtung strahlen. Für unsere Wahrnehmung ist entscheidend, dass eine ausreichende Menge Lichtwellen im Auge ankommen. Da Lichtwellen Energie übertragen, dürfen aber auch nicht zu viele Wellen auf das Auge treffen, da sie sonst Schaden anrichten können.

F

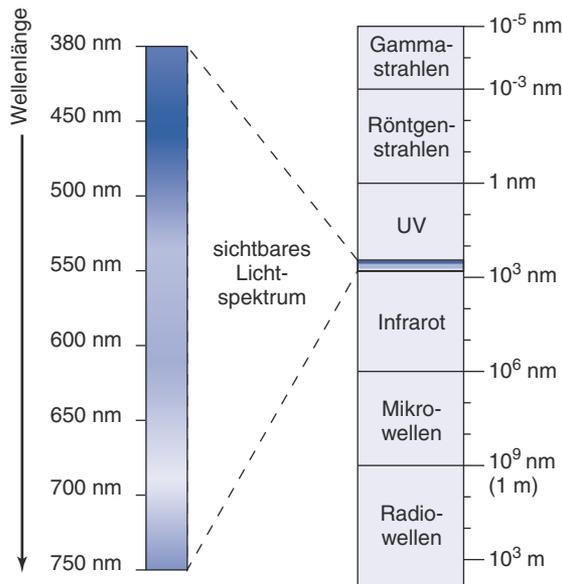


Abbildung 3.2: Elektromagnetische Wellen und sichtbares Licht

Frequenz: Die Frequenz entscheidet darüber, ob und wie die elektromagnetischen Wellen von Menschen wahrgenommen werden können. Die Frequenz wird in Schwingungen pro Sekunde angegeben. Die physikalische Einheit dafür ist Hertz. Im **sichtbaren Bereich** elektromagnetischer Wellen, also in dem Bereich, in dem wir Licht wahrnehmen, haben diese Frequenzen von 385 bis 789 THz (1 Tera Hertz = 1 Billion Hertz). Wenn man berücksichtigt, dass sich das Licht mit Lichtgeschwindigkeit ausbreitet, also mit ca. 300 Millionen Metern pro Sekunde, dann kann man leicht ausrechnen, dass eine Schwingung des Lichts ungefähr 380 bis 780 nm (1 Nanometer = 1 Milliardstel Meter) lang ist. Da die Lichtgeschwindigkeit konstant ist, kann man entweder die Frequenz angeben oder die **Wellenlänge**, also die Länge einer einzelnen Schwingung. Sichtbares Licht ist nur ein kleiner Bereich der elektromagnetischen Strahlung (►Abbildung 3.2). Strahlungen mit höheren Frequenzen reichen von UV-Licht über

Röntgenstrahlung bis hin zu radioaktiver Strahlung. Bei niedrigeren Frequenzen – also höheren Wellenlängen – erreicht man nach Infrarotstrahlung schließlich Radiowellen, die bis zu 1km Wellenlänge erreichen können. Die Wahrnehmung des Lichts wird später noch genauer erklärt, aber hier sei schon erwähnt, dass nicht alle Lebewesen genau den gleichen Ausschnitt von Wellenlängen als sichtbares Licht wahrnehmen. Manche Tiere sehen noch UV-Licht, andere auch noch Bereiche im Infrarotbereich. Innerhalb des sichtbaren Bereiches entscheidet die Wellenlänge über die Qualität des Lichts als **Farbwahrnehmung**, die auch etwas später diskutiert wird.

Amplitude: Die Amplitude einer Welle beschreibt die Stärke einer Welle. Hellere Licht entspricht also einer höheren Amplitude der Welle.

Polarisation: Betrachtet man in Abbildung 3.1 die beiden Wellen, kann man erkennen, dass beide Wellen zwar die gleiche Frequenz und Amplitude haben, also Qualität (Farbe) und Amplitude (Helligkeit) sind gleich, aber die Wellen sind unterschiedlich gedreht. Man sagt, beide sind unterschiedlich polarisiert. Im Alltag spielt die Polarisation für das menschliche Sehen kaum eine Rolle. Es gibt aber zwei Fälle, wo sie auch bei (digitalen) Medien relevant ist. Natürliches Sonnenlicht kann mit dem Eintritt in die Atmosphäre durch Reflexion und Streuung polarisiert werden. Manche Tiere können dies wahrnehmen und die Polarisation quasi als Kompass nutzen. Spezielle **Polarisationsfilter** für Kameras und Brillen lassen nur Licht einer bestimmten Polarisation durch. Dies kann man bei der Aufnahme von Bildern nutzen und so Kontraste, Spiegelungen und Farben von Bildern mit natürlichem Licht beeinflussen. Eine ganz andere Anwendung findet man bei **3D-Brillen**. Dabei werden zwei Bilder mit unterschiedlicher Polarisation generiert und überlagert und eine Brille filtert für das linke und rechte Auge jeweils das passende Bild heraus. Damit bekommt jedes Auge ein unterschiedliches Bild, wodurch ein **stereoskopischer Effekt** erzielt wird.

3.3 Farben im Kopf

Die Physik des Lichts erklärt einiges, aber viele Eigenheiten von Bildern entstehen erst durch die Art und Weise, wie unsere Augen und unser Gehirn aus elektromagnetischen Wellen **Sinneseindrücke** gewinnen. Bei der Wahrnehmung von Licht verwandeln in unseren Augen Nervenzellen die Lichtwellen in neuronale Aktivitäten, die das Gehirn verarbeitet. Wie bereits in Kapitel 1 dargestellt, befinden sich diese Nervenzellen auf der **Netzhaut**. Beim Menschen gibt es vier verschiedene Typen: drei sogenannte Zapfentypen erkennen farbiges Licht und die sogenannten Stäbchen unterscheiden nur zwischen hell und dunkel, reagieren also nur auf die Stärke und nicht die Frequenz des Lichts. ►Abbildung 3.3 zeigt die Frequenzbereiche, auf die **Stäbchen** und **Zapfen** am besten reagieren. Wir sehen, dass das Spektrum des sichtbaren Lichts abgedeckt wird und dass die „Rot“-Zapfen und die „Grün“-Zapfen recht ähnliche Spektren haben.

F

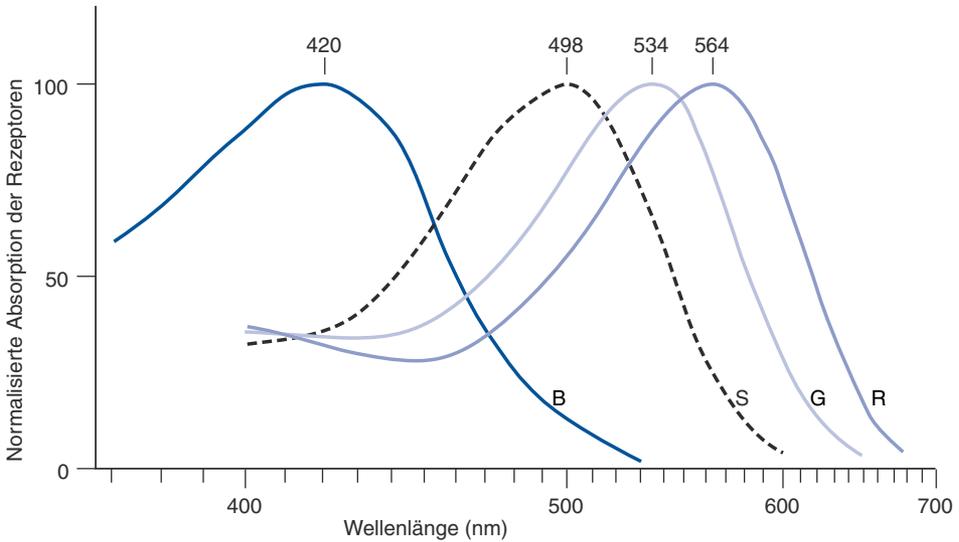


Abbildung 3.3: Antwortspektren von Stäbchen und Zapfen. Die Kurven zeigen die Antworten von blau-, grün-, rotsensitiven Zapfen (R, G, B) und von Stäbchen. Die Kurven sind so normalisiert, dass die Maximalwerte jeweils 100 betragen.

Zunächst könnte man vermuten, dass wir damit alle möglichen Lichteindrücke, die durch verschiedene Kombinationen von Wellenlängen des sichtbaren Lichts entstehen, erkennen und auch unterscheiden können. Doch das können wir gar nicht! Unsere Wahrnehmung reduziert die Welt der Wellenlängen dramatisch auf einen sehr viel kleineren Farbbereich, der eigentlich nur in unserem Gehirn existiert. Farben sind tatsächlich eine optische Täuschung!

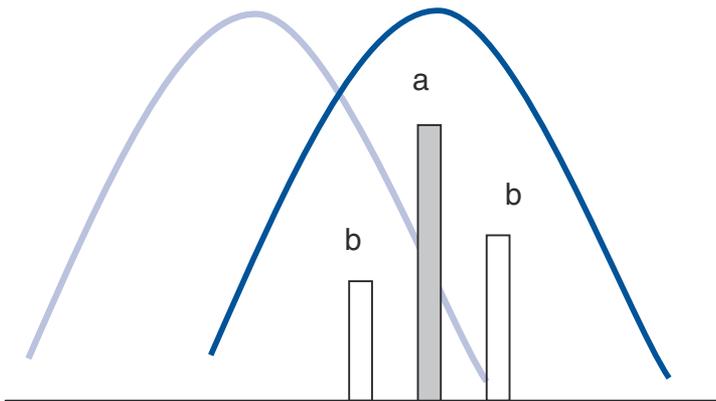


Abbildung 3.4: Zwei Signale (a) und (b) führen zu gleicher Farbwahrnehmung.

Das klingt zunächst eigenartig und scheint paradox zu sein. Zur Erklärung muss man sich die verschiedenen Typen von Stäbchen und Zapfen genauer ansehen und sich verdeutlichen, was bei der Überlagerung von Wellen passiert: Wenn Wellen mit verschiedener Frequenz überlagert werden, dann entsteht ein komplexeres Signal, in

dem die Ursprungsfrequenzen noch erhalten sind. Bei Tönen ist das ganz einfach nachvollziehbar: Wenn man auf einem Klavier zwei verschiedene Töne gleichzeitig spielt, dann kann man meist deutlich beide Töne hören und wenn man das resultierende Signal mit einer Fourier-Transformation umwandelt, dann kann man auch im Frequenzspektrum immer noch alle beteiligten Töne erkennen. Frequenzen mischen sich also nicht einfach. Was passiert aber im Auge? ►Abbildung 3.4 zeigt schematisch ein Beispiel, bei dem jeweils ein Signal (a) und (b) einmal aus Licht mit einer Wellenlänge besteht und einmal aus zwei Frequenzen zusammengesetzt ist. Offensichtlich sind die beiden Signale (a) und (b) deutlich verschieden (wenn wir auf einem Klavier entsprechende Töne spielen würden, dann würden die sehr unterschiedlich klingen). Wie reagieren aber die Rezeptoren darauf? Wenn wir uns eine mögliche Antwort des „Rot“-Zapfens in Abbildung 3.4 vorstellen, dann ist sehr gut vorstellbar, dass dieser auf (a) und (b) jeweils gleich stark reagiert. Entsprechendes gilt für den „Grün“-Zapfen. Der nimmt den rechten Balken von (b) zwar gar nicht wahr, aber dafür den linken etwas mehr und da er im sensitiveren Bereich als (a) ist, reagiert er insgesamt auf (b) genauso stark wie auf (a). Das heißt, die Antwort auf (a) und (b) ist gleich, obwohl die Signale völlig verschieden sind. In beiden Fällen wird, da „rot“ stark und „grün“ etwas reagiert, ein „Orange“-Ton wahrgenommen. Offensichtlich können wir nun denselben Farbton durch noch viel mehr verschiedene Kombinationen von Wellen erzeugen. Da das Auge nur drei Typen von Zapfen, also Farberkennern, hat, kommt es nur darauf an, dass die Erregung bei den drei Typen von Zäpfchen in der Summe jeweils übereinstimmt.

Die einfachste Variante, einen beliebigen Farbeindruck zu erzeugen, ist somit, für jeden Zapfentypen jeweils nur ein farbiges Lichtsignal zu nehmen. Somit kann man durch Regelung von drei unterschiedlichen Farbsignalen jeden Farbeindruck erzeugen. Wir brauchen also nur einen „Rot“- , einen „Grün“- und einen „Blau“-Wert, um im Auge den Eindruck beliebiger komplexer Überlagerungen von Lichtwellen zu erzeugen. Genau das geschieht, wenn wir **RGB-Monitore** oder Digitalkameras mit **RGB-Sensoren** verwenden. Und es klappt nur deshalb, weil wir eben gerade solche „Farbsensoren“ im Auge besitzen! Tatsächliche Bilder und die tatsächlichen Wellenlängen des Lichts in der Natur haben alle möglichen Wellenlängen. Nur weil wir mit unserem Auge die Wellenlängen des Lichts in drei Gruppen einteilen, können wir Farben aus diesen drei **Grundfarben** mischen und mit diesen **alle anderen Farben** erzeugen.

3.4 Farbaddition, Farbsubtraktion und Farbräume

Wenn keine unserer Stäbchen und Zapfen im Auge durch Licht erregt werden, ist alles dunkel. „Wir sehen Schwarz“. Wenn mehr Licht das Auge erreicht, und alle Typen von Stäbchen und Zapfen gleichermaßen erregt werden, dann sehen wir „Weiß“. Wir haben eben erfahren, dass wir dazu auch gar nicht eine Mischung des gesamten Frequenzspektrums des sichtbaren Lichts benötigen, sondern dass eine Überlagerung von „Rot“, „Grün“ und „Blau“ reicht, um „Weiß“ zu erzeugen. Wir können also aus Rot, Grün und Blau die Farbe Weiß „mischen“.

F

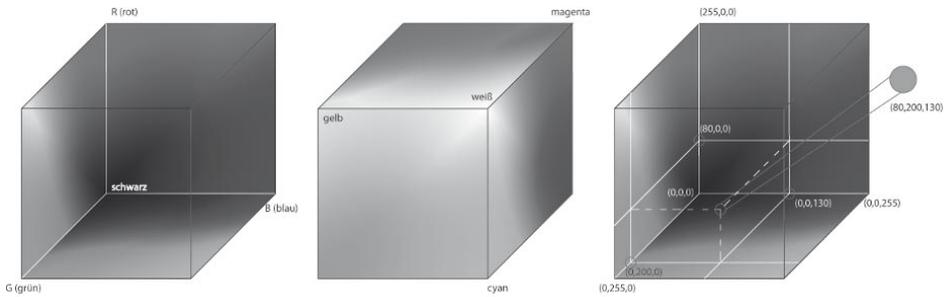


Abbildung 3.5: RGB-Farbwürfel

Wohlgemerkt: Unser Gehirn mischt die Farbe Weiß! Ein Lebewesen, das andere Farbsensoren hat, würde womöglich einen Dreiklang aus den drei Grundfarben sehen.

Diese Form der **Farbmischung** wird **additiv** genannt, da durch Hinzufügen von Licht aus Schwarz (also keinem Licht) einzelne Farben und schließlich weißes Licht „gemischt“ werden. Wenn nicht alle drei Grundfarben gleichermaßen gemischt werden, dann entstehen andere Farben (bzw. Farbeindrücke): Aus Rot und Grün wird Gelb, Rot und Blau ergeben Magenta und Grün und Blau werden zu Cyan. ►Abbildung 3.5 verdeutlicht in einem **Farbwürfel** die Möglichkeiten, Farben additiv zu mischen. Dabei sind alle aus R, G und B mischbaren Farben innerhalb des Würfels eingeschlossen. Den Raum in diesem Würfel mit allen Farben nennt man auch einen **Farbraum**. Die Ecken des Farbraumes sind die Grundfarben R, G, B, Weiß, Schwarz und an den weiteren Ecken Cyan, Gelb und Magenta. Jede Farbe innerhalb des Farbraumes kann über drei Koordinaten angegeben werden. Wenn wir beispielsweise für jede Grundfarbe 256 Abstufungen (Werte von 0 bis 255) haben, dann ist Schwarz mit dem Wert $(0,0,0)$ definiert, Rot mit $(255,0,0)$, Gelb mit $(255,255,0)$ etc.

Bei der additiven Mischung von Farben, wird aus verschiedenen farbigen Lichtquellen eine Lichtquelle mit einer helleren gemischten Farbquelle erzeugt. Dies geschieht z.B. bei Monitoren oder Projektoren. Wenn wir Farben auf Papier bringen, dann wird mit jeder neuen Farbe nicht Licht hinzugefügt, sondern abgezogen. Wenn wir einen grünen Farbklecks auf ein weißes Papier bringen, dann wird durch die grüne Farbe nicht mehr wie bei dem weißen Papier das gesamte (weiße) Licht reflektiert, sondern das blaue und rote Licht absorbiert und nur noch das grüne Licht reflektiert. Der Farbklecks erscheint somit grün. Auf Papier **subtrahiert** die Farbe Grün also Rot und Blau. Wenn wir uns den Farbwürfel aus Abbildung 3.5. ansehen, dann können wir sofort sehen, welche Farbe auf Papier welche anderen Farben wegfiltert: Jeweils die gegenüberliegenden Farben werden absorbiert. Bei Cyan wird Rot absorbiert, bei Gelb wird Blau herausgefiltert etc.

Wenn wir nun auf Papier Farben mischen, dann summiert sich die Absorption beider Farben: Wenn wir also Cyan und Gelb mischen, dann werden sowohl Blau als auch Rot absorbiert und im Resultat sehen wir Grün. Mit jeder hinzugemischten Farbe wird also von dem, was wir sehen, entsprechend Licht absorbiert bzw. subtrahiert. Deshalb spricht man hier auch von **subtraktiver** Farbmischung.

Mathematisch lassen sich das additive Farbmodell, bei dem mit den Farben Rot, Grün und Blau gemischt wird, und das subtraktive Farbmodell mit den Farben Cyan, Magenta und Gelb leicht ineinander umrechnen. Werte im ersten Modell nennt man **RGB-Werte**, die im zweiten Modell **CMY** (Y steht für Yellow, also Gelb). Ein RGB-Farbwert f_{RGB} muss einfach nur von der gegenüberliegenden Ecke des Farbwürfels her betrachtet werden:

$$f_{RGB} = \begin{pmatrix} f_R \\ f_G \\ f_B \end{pmatrix} = \begin{pmatrix} f_{\max} - f_C \\ f_{\max} - f_M \\ f_{\max} - f_Y \end{pmatrix} = \begin{pmatrix} f_{\max} \\ f_{\max} \\ f_{\max} \end{pmatrix} - \begin{pmatrix} f_C \\ f_M \\ f_Y \end{pmatrix} = \text{weiß}_{RGB} - \begin{pmatrix} f_C \\ f_M \\ f_Y \end{pmatrix} = \text{weiß}_{RGB} - f_{CMY}$$

Entsprechend gilt:

$$f_{CMY} = \begin{pmatrix} f_C \\ f_M \\ f_Y \end{pmatrix} = \text{weiß}_{RGB} - \begin{pmatrix} f_R \\ f_G \\ f_B \end{pmatrix} = \text{weiß}_{RGB} - f_{RGB}$$

Wenn wir für die Farbwerte also je Farbkanal mit 8 Bit 256 verschiedene Werte (bei insgesamt 24 Bit in RGB) haben, dann lässt sich ein Türkiston in RGB mit den Werten $(80,200,130)_{RGB}$ wie in Abbildung 3.5 dargestellt in CMY umrechnen:

$$f_{CMY} = \text{weiß}_{RGB} - \begin{pmatrix} f_R \\ f_G \\ f_B \end{pmatrix} = \begin{pmatrix} 255 \\ 255 \\ 255 \end{pmatrix} - \begin{pmatrix} 80 \\ 200 \\ 130 \end{pmatrix} = \begin{pmatrix} 175 \\ 55 \\ 125 \end{pmatrix}$$

Damit ist eigentlich schon alles Wesentliche gesagt. Trotzdem gibt es noch eine ganze Reihe von wichtigen Aspekten zu Farben und Farbmodellen, die ganze Bücher füllen können und über die man in zahlreichen Fotozeitschriften eine Unzahl von Artikeln finden kann.

Exkurs

Varianten von RGB- und CMY-Farbmodellen

RGB und CMY sind Farbmodelle, die für Geräte geeignet sind, die zum einen additiv und zum anderen subtraktiv Farben mischen. Die erste Gruppe umfasst also alle Geräte, die leuchten (Monitore, Beamer etc.) und die zweite Gruppe solche, die Licht reflektieren (also Prints, Drucke etc.). Eine Variante bei CMY ist das **CMYK-Modell**, wobei „K“ für „Black“ steht. Hierbei handelt es sich um eine für Drucker optimierte Variante, bei der ein zusätzlicher Farbkanal für schwarze Farbe helfen soll, Tinte zu sparen und die Schärfe der Darstellung schwarzer Schrift zu verbessern. Da bei Druckern sonst Schwarz nur durch Überlagerung von viel Farbe aus Cyan, Magenta und Gelb gemischt werden muss, ist dies zum einen ökonomischer und zum anderen im Resultat auch besser, da sonst meist kein richtiges Schwarz, sondern eher ein dunkles Braun entsteht.

Da aber jedes Gerät andere Grundfarben hat und außerdem je nach Umgebung die Grundfarben auch noch anders wirken können, sind weitere Farbprofile und Anpassungen nötig. Zum Beispiel können Drucker, je nachdem, wie die Tinten für C, M und Y genau beschaffen sind, sehr unterschiedliche Ergebnisse erreichen.

Zudem können die Farben auf unterschiedlichem Papier anders aussehen. Aus diesem Grund reicht es nicht, bei der Druckausgabe absolute Werte für die Komponenten C, M, Y and den Drucker anzugeben. Vielmehr muss der Drucker diese noch einmal umrechnen und an die tatsächlichen zur Verfügung stehenden Grundfarben anpassen. Damit die Anpassung richtig gelingt, muss der Drucker „kalibriert“ werden. Das heißt, es muss nachgemessen werden, welche Farbwirkung der Drucker auf einem gegebenen Papier mit seinen Grundfarben erzielt. Das Ergebnis einer solchen Kalibrierung ist ein „Farbprofil“, das dann im Druckvorgang die Umrechnung der theoretischen Farbwerte auf die tatsächlichen Mischungsverhältnisse der Druckerfarben steuert.

Entsprechend können auch unterschiedliche RGB-Geräte unterschiedliche Farben darstellen und müssen kalibriert werden, wenn standardisierte Darstellungen erzeugt werden sollen. Je nachdem, wo beim RGB-Modell die reinen Grundfarben tatsächlich physikalisch definiert sind, können die Modelle einen unterschiedlich großen RGB-Würfel aufspannen (Abbildung 3.5).

3.5 Weitere Farbmodelle

Bei den weiteren Farbmodellen lassen sich im Wesentlichen zwei Gruppen unterscheiden: zunächst alle Varianten der RGB- und CMY-Farb Räume und dann Farbmodelle, die die Farbinformation in einem Luminanz- und zwei weiteren Farbkanälen codieren. Über Sinn und Notwendigkeit der ersten Gruppe von Farbmodellen informiert der Kasten zu den Varianten der RGB- und CMY-Farbmodelle. Die zweite Gruppe arbeitet nicht mit drei Grundfarben, sondern trennt die **Farbinformation** von der **Helligkeitsinformation**. Etwas Ähnliches passiert bei der menschlichen Wahrnehmung, wo die Informationen der Farbwerte von den Zapfen direkt auf der nächsten Schicht von Neuronen entsprechend umcodiert werden. Das ist für die Wahrnehmung bei unterschiedlichen Helligkeiten vorteilhaft, weil die Empfindlichkeit auf diese Weise effizient entsprechend der vorherrschenden Helligkeit reguliert werden kann, ohne die Farbwahrnehmung zu beeinträchtigen. Insofern orientieren sich diese Modelle an der menschlichen Wahrnehmung. Sie können auch dazu genutzt werden, um ein Schwarzweißbild (nämlich den Helligkeitskanal) durch zwei zusätzliche Kanäle zu einem Farbbild anzureichern. Während bei RGB und CMY durch die Änderung eines Farbwertes auch die Helligkeit und Sättigung der Farbe verändert wird, werden bei wahrnehmungsorientierten Modellen Helligkeit, Sättigung, Farbton zum Teil als einzelne Kanäle codiert. Dies kann in vielen Anwendungen nützlich sein, wenn z.B. der Farbton gleich sein soll, aber Helligkeit und Sättigung variiert werden sollen. Die Umrechnung in diese Farbmodelle ist zum Teil recht aufwändig und nicht linear.

Ein einfaches lineares Modell, das auch einen Helligkeitskanal besitzt, ist das **YCbCr-Modell**, das einige sehr nützliche Eigenschaften hat. Im Prinzip handelt es sich um eine einfache lineare Abbildung, bei der das RGB-Signal mit einer Transformationsmatrix und einer Verschiebung in einen neuen Vektor umgewandelt wird:

$$f_{YCbCr} = \begin{pmatrix} f_Y \\ f_{Cb} \\ f_{Cr} \end{pmatrix} = \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix} + \begin{pmatrix} 0,299 & 0,587 & 0,114 \\ -0,168736 & -0,331264 & 0,5 \\ 0,5 & -0,418688 & -0,081312 \end{pmatrix} \begin{pmatrix} f_R \\ f_G \\ f_B \end{pmatrix}$$

wobei in dieser Formulierung alle Werte mit einem Byte also Werten zwischen 0 und 255 codiert werden. Der Y-Kanal, der die Helligkeit codiert, besteht zu einem hohen Anteil aus dem Grünwert und zu geringeren Anteilen aus den Rot- und Blauwerten. Die beiden Farbkanäle Cb und Cr codieren jeweils vor allem den Blau- und den Rotanteil.



Abbildung 3.6: Ein Farbbild (oben) zerlegt in die Farbkanäle R, G, B (zweite Zeile und in die Kanäle Y, Cr, Cb (untere Zeile)

In ►Abbildung 3.6 wird ein Bild gezeigt und einmal in die Kanäle R, G und B zerlegt und einmal in die Kanäle Y, Cr, Cb. Dabei werden einige Effekte offensichtlich: Während die RGB-Kanäle alle als Schwarzweißbilder noch für Menschen gut erkennbare Information enthalten, ist dies im YCrCb-Fall nur für den Y-Kanal der Fall. Die beiden Farbkanäle sind wenig kontrastreich und die Bilder wirken sehr verfremdet. Das Y-Bild ist dem G-Bild am ähnlichsten, aber durch die Mischung aus allen Farbkanälen ist ein noch besseres Schwarzweißbild entstanden. Diesen Effekt hat man sich bei der Kodierung von Bildinformationen in verschiedenen Fernschnormen zu Nutze gemacht. Bei **PAL** z.B. wird in einer dem YCrCb-Modell ähnlichen Kodierung ein Farbbild übertragen,

bei dem der Y-Kanal quasi das **Schwarzweißbild** überträgt. Einfache Fernseher, die nur schwarzweiß wiedergeben können, brauchen also nur dieses Bild anzuzeigen. Farbfernsehgeräte nutzen dann die beiden weiteren Farbkanäle, um ein buntes Bild zu erzeugen.

Aber noch etwas wird deutlich: Die beiden Farbkanäle Cr und Cb sind viel „flächiger“, d.h., benachbarte Bildpunkte haben viel öfter den gleichen Wert und zudem haben die meisten Bildpunkte mittlere Werte, d.h. sehr kleine und sehr große Werte sind relativ selten. Damit bietet die Kodierung in YCrCb-Werte gute Möglichkeiten, Bilder kompakter zu speichern. Dies wird beim JPEG-Verfahren genutzt, das später vorgestellt werden soll.

Der Effekt, dass im Helligkeitskanal wesentlich mehr Information steckt als in den beiden Farbkanälen, lässt sich an fast allen natürlichen Bildern zeigen. In Abbildung 3.6 wurde ein normales Urlaubsbild gewählt. Beliebige andere Bilder hätten ähnlich deutliche Ergebnisse gezeigt. Das heißt, dass mehr Information in den Helligkeitswerten als in den Farbwerten steckt, ist ein Phänomen, das fast überall in der Natur auftritt: Farben sind gleichmäßig verteilt, Helligkeiten eher nicht. Dies liegt daran, dass die meisten Gegenstände eher einheitlich gefärbt sind, das heißt, die Farbstoffe, die bestimmte Frequenzbereiche des sichtbaren Lichts entweder reflektieren oder absorbieren, sind in den meisten natürlichen Gegenständen über größere Bereiche recht einheitlich. Im Gegensatz dazu können Licht und Schatten starke Kontraste und Helligkeitsunterschiede auf dreidimensionalen Gegenständen erzeugen. Das Phänomen ist also einer eher einheitlichen Färbung bei dreidimensionaler Struktur geschuldet. Da dies in der Natur so ist, hat sich auch unser Sehsystem darauf spezialisiert und es verarbeitet Helligkeitsunterschiede und Kontraste besser als Farbvariationen. Besonders auffällig wird das nachts, wenn wir quasi ganz ohne Farben sehen. Dieser Zusammenhang wird aber oft falsch verstanden: Es wird behauptet, dass die Farbkanäle weniger Information hätten, weil wir diese schlechter wahrnehmen. Es ist aber gerade umgekehrt: Die Farbkanäle haben weniger Information und deshalb passt sich unser Sehsystem daran an. Dass in Abbildung 3.6 die Cr und Cb Kanäle so flach und informationsleer erscheinen, ist also nicht ein Resultat aus unserem Farbsehen, sondern ein Resultat aus der Natur, an das sich unser Sehsystem angepasst hat.

3.6 Digitalisierung und Kodierung

Digitale Bilder setzen sich aus Bildpunkten zusammen, den **Pixeln**, die in einem Raster angeordnet sind. Jeder Bildpunkt besteht aus Helligkeits- und Farbinformationen. Im einfachsten Fall besteht ein Pixel aus einem Bit, das sagt, ob der Punkt schwarz oder weiß ist. Für Farbbilder braucht man offensichtlich mehr Speicherkapazität als ein Bit pro Pixel, die dann auch über die Zahl der möglichen Farben entscheidet. Diese wird auch **Farbtiefe** genannt.

Typische Werte für Farbtiefen sind:

- 2 Farben (1 Bit) – entspricht **Schwarz/Weiß**,
- 16 Farben (4 Bit),

- 256 Graustufen (8 Bit),
- 256 Farben (8 Bit),
- 16,7 Millionen Farben (24 Bit) – „**True Color**“.

Bei einer True-Color-Farbtiefe werden die 24 Bit in drei Byte aufgeteilt, die jeweils einen Farbkanal (z.B. in RGB oder CMY) codieren. In vielen Formaten gibt es noch einen vierten Farbkanal mit weiteren 8 Bit, den **Alphakanal**. Er gibt in jedem Pixel an, wie transparent dieses Pixel sein soll. Der Alphakanal ist besonders wichtig, wenn verschiedene Bilder überlagert werden sollen. Er wird häufig in Bildbearbeitungsprogrammen genutzt.

Bilder mit noch höherer Farbtiefe codieren jedes Pixel mit 16 Bit, also insgesamt jedes Pixel mit 48 Bit. Manche Formate geben sich auch damit nicht zufrieden und definieren einen Farbraum, der je Farbkanal 32 Bit besitzt (sog. **High Dynamic Range**).

Typische digitale Fotografien kommen allerdings mit 24 Bit aus und in vielen Fällen reichen sogar deutlich weniger Farben. Besonders in Grafiken werden oft nur eine kleine Zahl von Farben verwendet. In diesem Fall reicht es, eine **Farbpalette** anzulegen, in der die verwendeten Farben hinterlegt sind und jede Farbe einen Index bekommt. Bei **GIF (Graphics Interface Format)** werden solche Farbtabellen mit maximal 256 Farben verwendet. Zusätzlich kann GIF eine Farbe als „**Transparenzfarbe**“ nutzen. Pixel mit dieser Farbe sind dann transparent. Einen echten Alphakanal ersetzt das zwar nicht, aber es erlaubt das Überlagern von Bildern. Im Gegensatz zu GIF können Bilder im moderneren **PNG-Format** (Portable Network Graphics) einen echten Alphakanal besitzen und wahlweise Farben in Farbpaletten oder in Farbkänen mit einer Farbtiefe von bis zu 24 Bit speichern.

Neben der Farbtiefe ist die **Auflösung** des Bildes entscheidend. Die meisten digitalen Kameras liefern heute Bilder mit einigen Megapixeln. Je nach Auflösung und Farbtiefe kann ein digitales Bild somit ein erhebliches Datenvolumen erzeugen. Bei 10 Megapixeln und 48 Bit pro Pixel hat das Bild 480 Millionen Bit oder 60 MByte. Eine Kamera mit 2 Megapixeln und 24 Bit Farbtiefe kommt immerhin auf 6 MByte pro Bild.

Bei solchen Datenmengen wird offensichtlich, dass Bilder für viele Zwecke sinnvoll komprimiert werden müssen. Prinzipiell kann man alle Kompressionsverfahren anwenden, die in Kapitel 2 besprochen wurden. Viele Kameras speichern von vorneherein in einem verlustbehafteten komprimierten Format. Und sogar bei hochwertigen Kameras, die eine verlustfreie Speicherung des Originalbildes anbieten (das sogenannte **RAW-Format**), werden zumindest verlustfreie Kompressionsverfahren angewendet, die das Datenvolumen beschränken.

Exkurs

Dithering

Wenn 256 Farben nicht reichen, kann man einerseits die Farbtiefe erhöhen und True Color verwenden. Man kann aber auch mit dem sogenannten **Dithering** Farbverläufe simulieren, indem Punkte gestreut werden. ►Abbildung 3.7 zeigt verschiedene Dithering-Verfahren, bei denen Grauwertbilder mit nur einem Bit Farbtiefe simuliert werden. Durch die Verteilung von weißen und schwarzen Pixeln entsteht der Eindruck von Grautönen. Entsprechend können auch mit einer begrenzten Zahl von Farbwerten neue Farben und Farbverläufe simuliert werden.

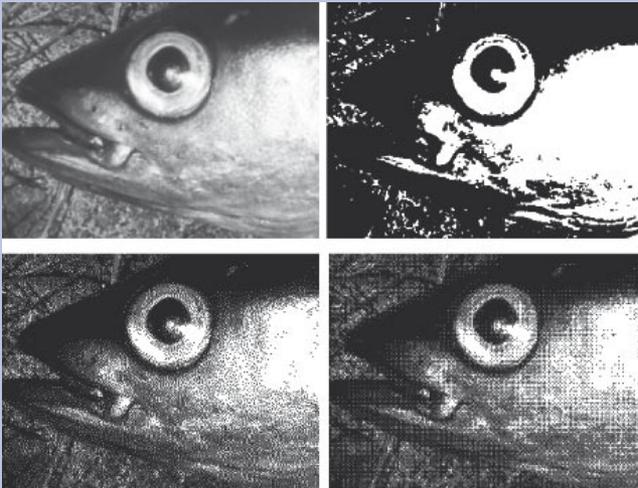


Abbildung 3.7: Dithering. Links oben: Originalbild mit 256 Grautönen. Rechts oben: Schwarzweißbild ohne Dithering. Link unten: Dithering mit Floyd-Steinberg-Verfahren. Rechts unten: Dithering mit festem Muster. Alle Bilder außer dem Originalbild haben 1 Bit Farbtiefe (Schwarz/Weiß).

3.7 Kompression von Bildern

Wirklich hohe Kompressionsraten erreicht man erst mit **verlustbehafteten** Verfahren. Während das oben erwähnte GIF verlustfrei arbeitet und auf einer reduzierten Zahl von Farben basiert, können verlustbehaftete Verfahren auch bei hoher Farbtiefe die Datenmenge deutlich reduzieren. Wie in Kapitel 2 beschrieben, hängt die Menge der Daten eines digitalisierten Signals von der

- Abtasthäufigkeit (Samplingrate, bzw. bei Bildern Anzahl der Pixel) und der
- Abtastgenauigkeit (Quantisierung, Farbtiefe)

ab. Verlustbehaftete Verfahren versuchen nun, die Abtastung etwas ungenauer zu machen, also die Zahl der Abtastungen oder die Genauigkeit der Abtastung herunterzusetzen. Allerdings soll dies nicht global geschehen, sondern nur an den Stellen, wo

es am wenigsten auffällt. Würde man global die Abtastung vergrößern, könnte man einfach ein Bild verkleinern und mit weniger Farben arbeiten. Stattdessen soll das Bild aber die Größe behalten und auch noch die ursprüngliche Farbigkeit besitzen.

Das **JPEG-Format** eignet sich besonders gut, um einige Ideen und Verfahren kennenzulernen, die aus großen Bilddateien kompakte Datenmengen machen, ohne dass der visuelle Eindruck besonders darunter leidet.

Exkurs

JPEG

JPEG steht für die Joint Photographic Expert Group, eine Gruppe von Experten, bei der zwei Standardisierungsorganisationen, die International Telecommunication Union (ITU) und die International Organization for Standardization (ISO), zusammenarbeiten. Seit Mitte der 80er Jahre entwickelten sie einen Standard für die Kompression von digitalen Bildern. Im Jahr 1988 wurde schließlich der JPEG-Standard beschlossen, der seitdem sehr weite Verbreitung gefunden hat.

Im Wesentlichen kann man die JPEG-Kompression in folgende Schritte einteilen:

- **Chroma-Subsampling**
- **Umcodierung in den Frequenzraum**
- **Quantisierung**
- **Kompression**

Diese Schritte sollen im Folgenden erläutert werden.

3.7.1 Chroma-Subsampling

Chroma-Subsampling ist ein **verlustbehafteter Kompressionsschritt**, bei dem die Abtastrate für einen Teil des Bildes verringert wird. Die zugrunde liegende Idee ist leicht verständlich, wenn man die Kodierung eines Farbbildes im YCrCb-Modell ansieht (Abbildung 3.6). Da die beiden Farbkanäle Cr, Cb viel „flächiger“ sind, also viel öfter benachbarte Pixel dieselben Werte haben, kann man mehrere Pixel durch eines repräsentieren, ohne dass der Verlust stark auffällt. ►Abbildung 3.8 zeigt dies für einen kleinen Ausschnitt des Bildes. Beim Chroma-Subsampling gibt es verschiedene Varianten. Man kann einen Block aus vier Pixeln oder nur zwei benachbarte Pixel (horizontal oder vertikal) zusammenfassen. Damit werden die Cr- und Cb-Kanäle auf die Hälfte bis ein Viertel komprimiert. Da der Y-Kanal mehr Information enthält, bleibt er gleich und wird nicht verändert. Das heißt, die Vergrößerung der Abtastrate wird beim Subsampling nur auf den Farbkanälen durchgeführt, da man im Resultat kaum einen Unterschied bemerkt. Bei einer Kompression von jeweils vier Pixeln auf eines in beiden Farbkanälen wird somit die Anzahl der Pixel in den Farbkanälen geviertelt und da die Pixel im Y-Kanal erhalten bleiben, wird das gesamte Datenvolumen insgesamt halbiert.

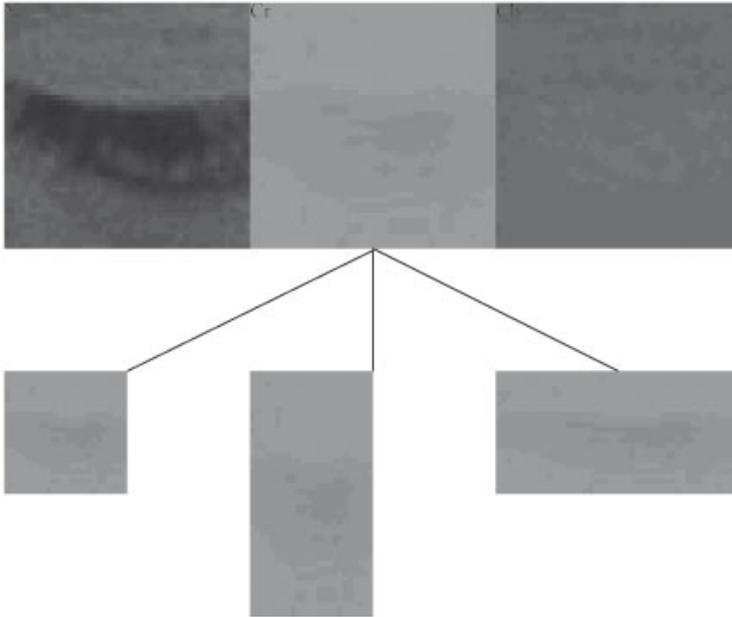


Abbildung 3.8: Vergrößerung aus ►Abbildung 3.6, Y-, Cr-, Cb-Kanal, jeweils 48x48 Pixel und Chroma-Subsampling für den Cr-Kanal auf 24x48, 48x24 und 24x24 Pixel (zweite Zeile)

Bei JPEG kann man bei der Kodierung wählen, wie stark komprimiert werden soll. Je nach eingestellter Kompressionsrate wird das Chroma-Subsampling bei JPEG entweder gar nicht oder mit einer Halbierung oder Viertelung der Farbkanäle durchgeführt. Das Subsampling ist eine verlustbehaftete Kompression, da nach dem Subsampling Information gelöscht wurde, auch wenn dies in den meisten Fällen kaum auffällt.

3.7.2 Umcodierung in den Frequenzraum

Die folgenden Schritte geschehen für alle drei Farbkanäle getrennt. Das heißt, es werden Y, Cr und Cb jeweils unabhängig voneinander untersucht und komprimiert. Wenn man nur einen Farbkanal betrachtet, hat man es also jeweils mit Grauwertbildern zu tun, bei denen jedes Pixel im Fall von JPEG mit einem Byte, also Werten von 0 (Schwarz) bis 255 (Weiß), codiert wird. Je nach Subsampling im vorigen Schritt haben die drei Bilder der einzelnen Kanäle unterschiedliche Größen.

Die Bilder bzw. genauer die drei Farbkanäle werden nun **umcodiert**, um anschließend die Quantisierung, also die Genauigkeit der Abtastung, zu vergrößern. Dabei sollen wie beim Chroma-Subsampling nicht alle Werte vergrößert werden, sondern vor allem solche, bei denen eine Vergrößerung am wenigsten auffällt. Aus diesem Grund werden die **in einem Bild vorkommenden Frequenzen** betrachtet. Die einführenden Kapitel betrachteten bereits Abtastung und Frequenzen und vor allem Audiosignale. Aber auch für Bilder ist es sinnvoll, von Frequenzen zu sprechen. Ein Bild mit senkrechten

Streifen wie in ►Abbildung 3.9 (ganz links) ist in horizontaler Richtung von wechselnden hellen und dunklen Pixeln bestimmt, während es keine Veränderung gibt, wenn man das Bild an einer bestimmten Stelle der x-Achse in vertikaler Richtung „liest“. Es hat somit horizontale Frequenzanteile, aber die vertikale Frequenz ist null. Entsprechend variieren die Beispiele in Abbildung 3.9 in ihren Anteilen horizontaler und vertikaler Frequenzanteile.

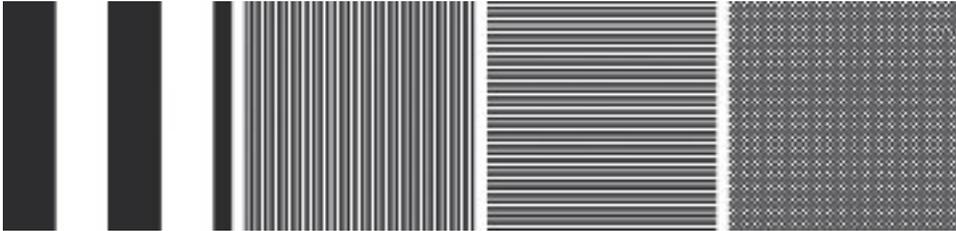


Abbildung 3.9: Bilder mit unterschiedlichen horizontalen und vertikalen Frequenzanteilen, von links nach rechts: (1) niedrige horizontale Frequenzen, (2) nur hohe horizontale Frequenzen, (3) nur hohe vertikale Frequenzen, (4) Mischung hoher horizontaler und vertikaler Frequenzen

Ähnlich, wie Audiosignale aus Überlagerungen von einfachen Schwingungen betrachtet werden können, lassen sich auch Bilder aus einfachen Bildern bzw. Mustern zusammensetzen. Insbesondere können Bilder aus einer Grundmenge von Bildern zusammengesetzt werden, die jeweils nur aus einer horizontalen und einer vertikalen Frequenz bestehen. ►Abbildung 3.10 zeigt 64 solcher Basismuster, aus denen man jedes beliebige 8 mal 8 Pixel große Grauwertbild erzeugen kann. Dazu muss man für jedes dieser Basismuster angeben, mit welchem Anteil es in dem zu erzeugenden Bild auftreten soll.

Ein Transformationsverfahren, das ein Bild in solche Basismuster zerlegt, ist die diskrete **Cosinustransformation (DCT)**. Mit ihr kann für ein Bild berechnet werden, welche Frequenzanteile vorkommen. Prinzipiell könnte man aber beliebig große Bilder mittels DCT transformieren. Bei JPEG werden allerdings die Bilder in 8 mal 8 Pixel große **Teilblöcke** zerlegt, die dann unabhängig voneinander betrachtet werden. Deshalb beschränken wir uns auch hier auf Grauwertbilder mit nur 8 mal 8 Pixeln.

Jeder 8x8-Block enthält 64 Werte. Der Wert $f(x,y)$ gibt für das Pixel an der Stelle (x,y) innerhalb des betrachteten Blockes den Grauwert an (was je nach Kanal ein Wert für Y, Cr oder Cb ist). Durch die DCT werden diese nun in 64 neue Werte, die **DCT-Koeffizienten** $F(u,v)$, umgerechnet, die angeben, mit welchem Anteil die Basisbilder der DCT in dem Block vorkommen.

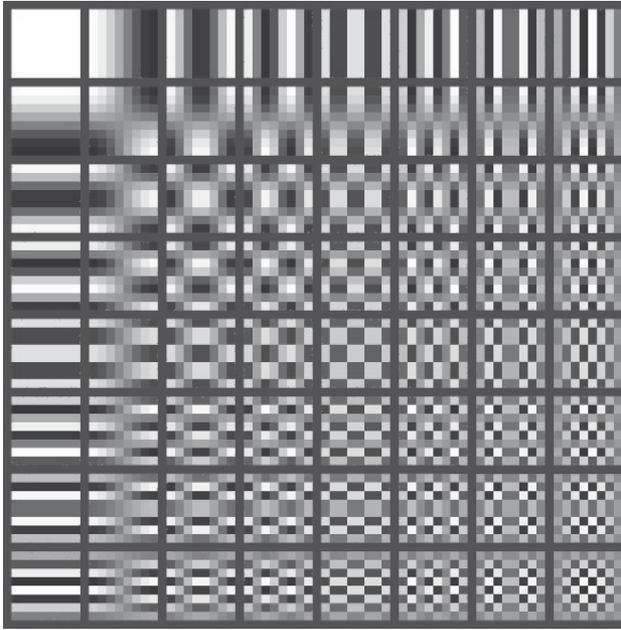


Abbildung 3.10: 64 Basismuster für die DCT, jedes Basismuster besteht aus 8x8 Werten, die unterschiedliche Frequenzanteile des Bildes repräsentieren. Ganz oben links ist die DC-Komponente. Die Muster weiter rechts entsprechen höheren horizontalen, die weiter unten höher werdenden vertikalen Frequenzen.

Exkurs

Berechnung der DCT

Zur Berechnung der DCT-Koeffizienten werden zunächst die Werte $f(x,y)$, die im Bereich von 0 bis 255 liegen, auf den Bereich -128 bis 127 verschoben (indem einfach bei jedem Pixel 128 subtrahiert wird). Mit den neuen Werten $f(x,y)$ werden nun die Koeffizienten, die dann auch im Bereich $[-128, 127]$ liegen, durch folgende Formeln berechnet:

$$F(u,v) = \frac{1}{4} c_u c_v \sum_{x=0}^7 \sum_{y=0}^7 f(x,y) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16}$$

$$c_k = \begin{cases} \frac{1}{\sqrt{2}}, & \text{falls } k = 0 \\ 1, & \text{sonst} \end{cases}$$

Dabei sind x, y die Koordinaten im 8x8-Teilbild und u, v die Indizes der 64 DCT-Koeffizienten.

Diese Transformation lässt sich leicht berechnen, da die Cosinus-Werte und die Werte für die c_k nur einmal berechnet werden müssen und nicht von $f(x,y)$ abhängen. Man kann also für alle 64 möglichen Werte von x, y und alle 64 Werte von u, v insgesamt 4096 Parameter $T(x,y,u,v)$ berechnen:

$$T(x, y, u, v) = \frac{1}{4} c_u c_v \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16}$$

Womit man die DCT einfach als

$$F(u, v) = \sum_{x=0}^7 \sum_{y=0}^7 T(x, y, u, v) f(x, y)$$

schreiben kann. Ein Teilbild muss also immer nur mit einer bestimmten Matrix bzw. einem Basismuster multipliziert werden, um den jeweiligen DCT-Koeffizienten zu berechnen. Der Wert $F(u,v)$ gibt an, wie stark dieses Muster aus horizontalen und vertikalen Frequenzen in dem Teilbild vorkommt. ►Abbildung 3.10 zeigt die bei der DCT verwendeten 64 Basismuster.

Um ein Bild aus den DCT-Koeffizienten zu rekonstruieren, muss es einfach nur wieder aus den Basismustern zusammengesetzt werden. Die DCT-Koeffizienten geben an, wie stark die einzelnen Basismuster gewichtet werden. Mit den Werten $T(x,y,u,v)$ lässt sich die **inverse DCT** ganz einfach formulieren als:

$$f(x, y) = \sum_{u=0}^7 \sum_{v=0}^7 T(x, y, u, v) F(u, v)$$

Daran sieht man auch, dass die DCT und die inverse DCT ganz symmetrisch ablaufen.

Der DCT-Koeffizient $F(0,0)$ hat eine besondere Rolle. Er berechnet den Mittelwert aller Pixel innerhalb des Teilbildes und wird **DC-Komponente** genannt. Die anderen Werte $F(u,v)$ nennt man **AC-Komponenten**.¹

Die DCT reduziert den Speicherbedarf noch nicht: Aus 64 Pixelwerten wurden 64 Koeffizienten. Im Prinzip ist die DCT-Transformation zusammen mit der inversen DCT eine verlustfreie Transformation des Bildes, solange keine Rundungsfehler bei den Berechnungen auftreten. Warum die DCT helfen kann, ein Bild zu komprimieren, kann man gut nachvollziehen, wenn man die typischen Werte der DCT-Koeffizienten genauer betrachtet.

1 DC steht Direct Current und heißt Gleichstrom. AC steht für Alternating Current und heißt auf Deutsch Wechselstrom.

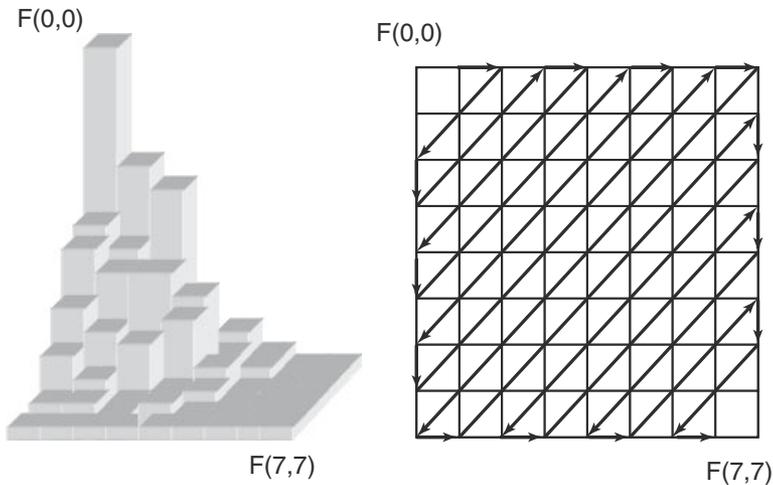


Abbildung 3.11: DCT-Koeffizienten. Links: typische Werteverteilung bei Bildern. Der DC-Anteil $F(0,0)$ ist hoch. Hohe Frequenzanteile (vorn und rechts) sind niedrig und oft null. Rechts: Zick-Zack-Scan zum Auslesen der AC-Koeffizienten.

Diese haben zwei wichtige Eigenschaften, die bei der Kompression genutzt werden können:

1. Bei vielen Bildern sind die hohen Frequenzanteile weniger stark vorhanden als die niedrigen. ►Abbildung 3.11 zeigt eine typische Verteilung der Werte der DCT-Koeffizienten. Bei einem einfarbigen Bild wären sogar außer dem DC-Koeffizienten alle anderen null.
2. Darüber hinaus bedeuten Änderungen bei den hohen Frequenzanteilen bei der Wahrnehmung eines Bildes keinen großen Unterschied.

Man kann also den zweiten Effekt nutzen, um den ersten Effekt noch zu verstärken, indem man solche Werte, die nahe bei null sind, gleich auf null setzt. Außerdem kann man höhere Frequenzanteile mit etwas weniger Bits codieren als die niedrigeren, um so Speicher zu sparen, womit man mehr Ungenauigkeit in Kauf nimmt. Anstatt die DCT-Koeffizienten mit 8 Bit (ebenso wie die Werte der Pixel je Farbkanal) zu codieren, können wir stattdessen für DCT-Koeffizienten höherer Frequenzen weniger Bits verwenden. Mit diesem Vorgehen wird die Auflösung digitalisierter Werte, also die Quantisierung, vergrößert und somit Speicherplatz gespart. Wie beim Subsampling wird die Vergrößerung auch in diesem Fall nicht auf alle Werte des Bildes gleichmäßig angewandt, sondern nur auf hohe Frequenzanteile. Damit dies möglich ist, musste das Bild mittels DCT zunächst transformiert werden.

Exkurs

Re-Quantisierung

Die neue **Quantisierung** der DCT-Koeffizienten wird mit einer **Quantisierungsmatrix** durchgeführt. Dabei gibt es für jeden der 8×8 DCT-Koeffizienten einen Wert $Q(u,v)$, durch den dieser geteilt wird:

$$F'(u,v) = \text{Round} \left(\frac{F(u,v)}{Q(u,v)} \right)$$

Durch die Rundung auf ganze Zahlen wird damit der neue Wertebereich der Koeffizienten $F'(u,v)$ geringer. Außerdem kommen kleinere Zahlen häufiger vor und somit kann man z.B. mit der Huffman-Kodierung solche Werte mit weniger Bit codieren. Da die Werte $F'(u,v)$ mit höherem u und v oft den Wert null haben, kann man dies zusätzlich für die Kompression ausnutzen. Die neue Quantisierung der DCT-Koeffizienten stellt den wesentlichen verlustbehafteten Schritt bei JPEG dar. Je nachdem, wie hoch die Werte $Q(u,v)$ sind, geht mehr oder weniger Information verloren. Dabei gibt es sowohl vordefinierte Quantisierungsmatrizen als auch solche, die ein Kompressionsprogramm selbst definiert haben kann. Soll keine Information verloren gehen, so können alle Werte auf $Q(u,v)=1$ gesetzt werden.

Bei der Rekonstruktion eines komprimierten Bildes muss die verwendete Quantisierungsmatrix bekannt sein, damit aus den Werten $F'(u,v)$ wieder DCT-Koeffizienten $\tilde{F}(u,v)$ berechnet werden können:

$$\tilde{F}(u,v) = F'(u,v) \cdot Q(u,v)$$

Aufgrund der verlustbehafteten Rundung bei der Konstruktion von $F'(u,v)$ werden diese Werte $\tilde{F}(u,v)$ im Allgemeinen nicht mehr genau den originalen Werten $F(u,v)$ entsprechen.

Mit dem Subsampling und der neuen Quantisierung sind die verlustbehafteten Schritte erklärt. Bei der Speicherung werden diese nun noch mit verlustfreien Verfahren kombiniert, wobei die DC-Komponenten $F(0,0)$ der Teilbilder anders als die anderen Komponenten $F(u,v)$ abgespeichert werden:

■ Kodierung der DC-Komponenten

- Da benachbarte Blöcke oft ähnliche Grundtöne haben, kann man die DC-Komponenten durch **prädiktive Kodierung** speichern. Ausgehend von einem Startwert werden für die folgenden Werte nur die Differenzen gespeichert.
- Es sind weniger Bits für die (kleineren) Differenzwerte nötig. Dies kann durch Huffman-Kodierung oder eine andere Entropiecodierung geschehen (z.B. arithmetische Kodierung).
- Die DC-Komponenten ergeben zusammen ein vergrößertes (Vorschau-)Bild.

- Kodierung der AC-Komponenten
- Die AC-Komponenten werden in einem **Zick-Zack-Scan** ausgelesen, zuerst die Werte mit niederen, dann die Werte mit hohen Frequenzen (Abbildung 3.11).
- Dabei kommen am Ende oft viele Nullen vor. Hier kann im ersten Schritt die Lauf­längencodierung genutzt werden.
- Im letzten Schritt können die Werte noch mit einer Entropiecodierung codiert werden (Huffman oder arithmetische Kodierung).
- In JPEG sind für die Huffmancodierung Standardverteilungen definiert. Es können aber auch eigene Verteilungen berechnet werden.

JPEG ist heute das wichtigste Verfahren der verlustbehafteten Kompression für digitale Bilder. Seit seiner Definition Ende der 80er Jahre wurden einerseits einige Schwächen deutlich, wie z.B. Block-Artefakte bei hoher Kompression. Andererseits wurden neue Verfahren zur Kodierung vorgeschlagen, die sehr vielversprechend sind. Mit **JPEG 2000** gibt es einen moderneren Standard, der z.B. anstelle der blockweisen DCT eine wesentlich flexiblere **Wavelet-Transformation** nutzt. Trotz seiner Überlegenheit gegenüber JPEG hat sich JPEG 2000 aber in der Praxis noch nicht durchgesetzt.

Exkurs

JPEG 2000

Die Joint Photographic Experts Group (JPEG) hat als Nachfolgestandard JPEG 2000 definiert, das eine Reihe von Vorteilen besitzt:

- bessere Leistung bei hoher Kompression (z.B. bei weniger als 0,25 Bit pro Pixel),
- verlustfreie und verlustbehaftete Kompression,
- progressiver Bildaufbau, das heißt, ein Bild kann mit den ersten Daten grob dargestellt und mit mehr Daten zunehmend verfeinert werden,
- Kodierung von Regions-of-Interest, das heißt, wichtige Bereiche in einem Bild können mit höherer Genauigkeit codiert werden,
- Robustheit bei Bitfehlern,
- Einbindung eines Alphakanals,
- Definition von Metadaten zur Beschreibung des Inhalts.

Wichtigste technische Änderung ist die Verwendung der Diskreten Wavelet-Transformation anstelle der DCT. Damit ist es möglich, das Bild in größere Teilbilder (tiles) zu zerlegen, als dies bei den 8x8-Blöcken möglich ist. Außerdem kann die DCT rekursiv angewandt werden, so dass immer ein Viertel eines Bildes eine vergrößerte Darstellung des Bildes darstellt, was den progressiven Bildaufbau ermöglicht.

3.8 Bearbeitung von Bildern

Mit der Kodierung, Kompression und Speicherung fängt für die meisten Menschen die Arbeit mit digitalen Bildern erst an. Bilder werden auf vielfältige Arten be- und verarbeitet:

- **Skalierung** für die richtige Anzeigegröße
- **Aufbereitung** für Druck, Grafik, Web etc.
- **Drehung, Entzerrung, Filterung** von Bildern
- **Schnitt, Montage, Manipulation**

Das Spektrum der Werkzeuge, Methoden und Verfahren für die Bildbearbeitung ist riesig und übersteigt bei weitem die Möglichkeiten, sie auch nur annähernd in diesem Buch zu behandeln.

Einige grundlegende Techniken sollen aber im Folgenden vorgestellt werden.

3.8.1 Punkt-Operationen

Bei **Punkt-Operationen** werden die Werte von einzelnen Pixeln verändert, ohne dabei die Nachbarpixel zu betrachten. Einfache Beispiele sind Methoden zur Änderung der Farbtiefe. Sollen z.B. aus Grauwertbildern mit 256 Werten pro Pixel **Schwarzweißbilder** mit nur einem Bit, also 2 Werten pro Pixel, gemacht werden, so kann man die neuen Werte für jedes Pixel einfach durch

$$f'(x, y) = \begin{cases} 1, & \text{falls } f(x, y) > w \\ 0, & \text{sonst} \end{cases}$$

berechnen. Dabei ist w ein beliebiger Grenzwert („Schwellwert“), der angibt, ab welchem Grauwert Pixel weiß sein sollen.

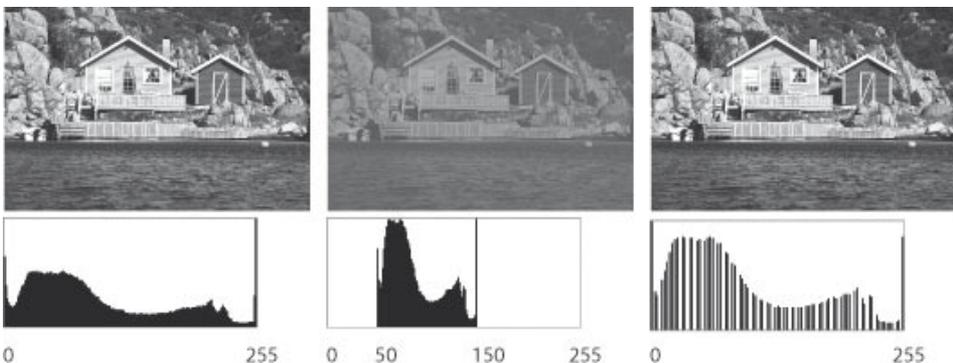


Abbildung 3.12: Grauwertbilder mit Histogrammen. Linkes Bild: Alle Grauwerte kommen vor. Mittleres Bild: Nur Grauwerte zwischen 50 und 150 kommen vor. Rechtes Bild: Spreizung der Grauwerte auf den ganzen Bereich, wobei das Ausgangsbild das mittlere Bild ist.

Bei vielen Operationen auf Bildern spielen **Histogramme** eine Rolle. Das Histogramm eines Bildes gibt an, welche Werte wie oft in den Pixeln eines Bildes vorkommen. Bei einem Grauwertbild mit 256 möglichen Werten pro Pixel wird das Histogramm durch einen Vektor $h(w)$, $w=0,\dots,255$ repräsentiert, bei dem $h(w)$ angibt, wie viele Pixel den Wert w haben. Ein Histogramm sagt einiges über ein Bild aus (►Abbildung 3.12, linkes Bild). Typischerweise soll ein Bild den gesamten Bereich eines Histogramms nutzen, sonst wirkt es flach und besitzt wenig **Kontrast** (►Abbildung 3.12, mittleres Bild). Nutzt ein Grauwertbild mit 8 Bit pro Pixel nur die Grauwerte 50 bis 150, dann könnte man die Werte der Pixel so spreizen, dass der minimale Wert bei 0 und der maximale Wert bei 255 liegen (►Abbildung 3.12, rechtes Bild). Dies leistet z.B. folgende einfache lineare Transformation:

$$f'(x, y) = \frac{f(x, y) - h_{\min}}{h_{\max} - h_{\min}} w_{\max}$$

Dabei sind h_{\min} und h_{\max} der minimale und der maximale Wert des Histogramms und w_{\max} der maximale mögliche Grauwert (im Beispiel $h_{\min}=50$, $h_{\max}=150$, $w_{\max}=255$). Das Bild wird dann wieder kontrastreicher aussehen, aber das Histogramm wird Lücken haben (Abbildung 3.12, rechtes Bild). Dies liegt daran, dass hier im Beispiel zwischen 50 und 150 nur 101 unterschiedliche Werte vorhanden sind, die dann zwar auf den gesamten Bereich zwischen 0 und 255 gespreizt werden, aber es bleiben trotzdem nur 101 unterschiedliche Werte.

Ähnliche Verfahren kann man nicht nur bei Grauwertbildern, sondern auch bei Farbbildern anwenden. Das kann entweder für alle Farbkanäle gemeinsam oder getrennt geschehen. Neben linearen Verzerrungen kann man verschiedene nichtlineare Kurven definieren, die die Pixelwerte des Ausgangsbildes auf die des Zielbildes abbilden. Viele Bildbearbeitungsprogramme bieten automatische Korrekturen an, die auf dieser Grundlage Farben und Kontraste eines Bildes optimieren.

3.8.2 Filter

Im Gegensatz zu Punkt-Operatoren werden bei **Filtern** die neuen Pixel des Bildes berechnet, indem die Informationen von mehreren Pixeln im Originalbild berücksichtigt werden. Meist werden nur Pixel in der Nachbarschaft betrachtet. Einer der einfachsten Filter ist ein **Glättungsfilter** oder **Weichzeichner**, bei dem jedes Pixel durch den Mittelwert aus seinem eigenen und der Werte seiner Nachbarn neu berechnet wird:

$$f'(x, y) = \frac{1}{9} \sum_{i=x-1}^{x+1} \sum_{j=y-1}^{y+1} f(i, j) \quad (1)$$

In diesem Fall wird eine Region von 9 Pixeln berücksichtigt. Dieser Filter ist ein **linearer Filter**. Das heißt, der neue Wert wird durch eine lineare Funktion berechnet. Man kann das mathematisch auch so beschreiben, dass die Umgebung des betrachteten Pixels an der Stelle (x, y) mit einer Matrix gefaltet wird:

$$f'(x, y) = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} * \begin{pmatrix} f(x-1, y-1) & f(x, y-1) & f(x+1, y-1) \\ f(x-1, y) & f(x, y) & f(x+1, y) \\ f(x-1, y+1) & f(x, y+1) & f(x+1, y+1) \end{pmatrix} \quad (2)$$

Bei einer Faltung von zwei Matrizen werden jeweils die Werte der beiden Matrizen, die an derselben Stelle stehen, paarweise miteinander multipliziert und dann diese Produkte aufsummiert. Die beiden Gleichungen (1) und (2) berechnen also genau den gleichen Wert. Allerdings hat die mathematische Darstellung in (2) den Vorteil, dass der Filter als Matrix dargestellt wird, die über das Bild gelegt wird.



Abbildung 3.13: Linkes Bild: Original, Mitte: einfacher Weichzeichner, rechts: einfacher Scharfzeichner (Boost-Filter)

Das Ergebnis dieses Weichzeichners zeigt ►Abbildung 3.13. Die Matrix, die einen solchen linearen Filter definiert, kann beliebig groß sein und beliebige Werte enthalten und somit lassen sich zahlreiche Filter definieren. Sollen Kontraste nicht weichgezeichnet, sondern verstärkt werden, dann kann man z.B. einen einfachen **Boost-Filter** definieren (►Abbildung 3.1, rechts):

$$f'(x, y) = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix} * \begin{pmatrix} f(x-1, y-1) & f(x, y-1) & f(x+1, y-1) \\ f(x-1, y) & f(x, y) & f(x+1, y) \\ f(x-1, y+1) & f(x, y+1) & f(x+1, y+1) \end{pmatrix}$$

Filter können auch eingesetzt werden, um Kanten zu detektieren. Die Kombination mehrerer Filter wird oft für **Scharfzeichner** in Programmen zur Bildbearbeitung eingesetzt. Neben den hier vorgestellten linearen Filtern können weitere nichtlineare Filter beliebige Funktionen auf den Werten von Pixeln in der Nachbarschaft eines Pixels berechnen. Solche Filter können z.B. zur Rauschunterdrückung oder aber für künstlerische Effekte zur Bildmanipulation verwendet werden.

3.8.3 Geometrische Operationen

Während Punkt-Operationen und Filter die einzelnen Pixel eines Bildes zwar neu berechnen, aber noch an ihrem Platz lassen, gibt es eine Reihe von Transformationen, bei denen Pixel an neuen Positionen des Bildes landen. Zu solchen Operationen gehören die **Verschiebung**, **Streckung** (Verkleinern oder Vergrößern), **Drehung** oder **Scherung** eines Bildes. Das größte Problem dabei ist, dass sowohl das Ausgangsbild als auch das Zielbild ein festes Pixelraster haben, aber diese Transformationen nicht immer für jedes Zielpixel genau ein Ausgangspixel zuordnen und umgekehrt auch nicht jedem Ausgangspixel genau ein Zielpixel (►Abbildung 3.14).

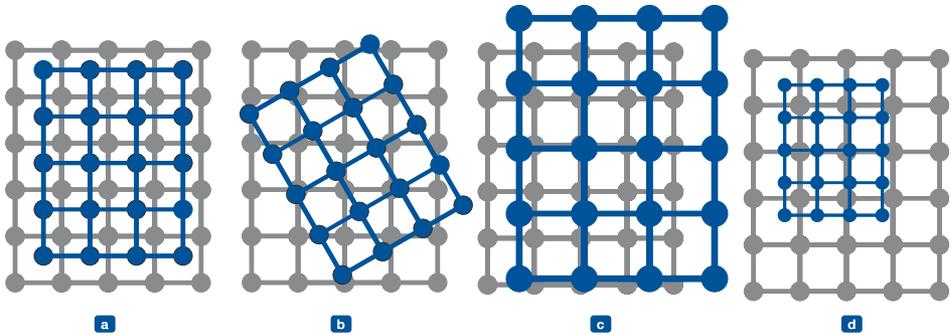


Abbildung 3.14: Probleme bei geometrischen Operationen. Blaues Raster: transformiertes Pixelraster des Originalbildes; graues Raster: Pixelraster des Zielbildes bei (a) Verschiebung, (b) Drehung, (c) Vergrößerung, (d) Verkleinerung

Wenn beim Verkleinern eines Bildes einem Zielpixel mehrere Ausgangspixel zugeordnet werden, kann man dies einfach dadurch ausgleichen, dass man dem Zielpixel den Mittelwert der Ausgangspixel zuordnet. In diesem Fall findet allerdings eine sogenannte **Unterabtastung** eines Bildes statt. In Kapitel 2 wurde bei der Digitalisierung erwähnt, dass bei solchen Unterabtastungen störende Frequenzen auftreten können. Bei Bildern äußern sich diese z.B. als **Moiré-Effekt**. ►Abbildung 3.15 zeigt einen solchen Effekt der Unterabtastung.



Abbildung 3.15: Unterabtastung beim Skalieren von Bildern. Linkes Bild: Originalbild, Mitte: Original mit Weichzeichner. Rechts oben: verkleinertes Originalbild, bei dem Moiré-Effekte sichtbar werden. Rechts unten: Verkleinerung des mittleren Bildes, bei dem der Weichzeichner den Moiré-Effekt verhindert hat

Eine Gegenmaßnahme ist die Anwendung eines **Weichzeichners** vor dem Verkleinern. Er wirkt als **Tiefpass** und verhindert solche Fehler.

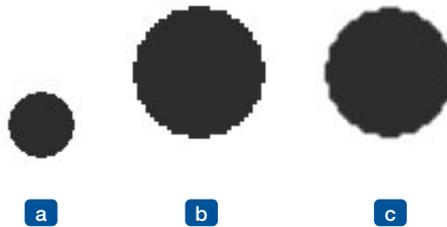


Abbildung 3.16: Alias-Effekt beim Vergrößern. (a) Originalbild, (b) Vergrößerung mit Zoom 200% ohne Antialiasing, (c) mit Antialiasing

Im umgekehrten Fall, wenn ein Pixel auf mehrere Zielpixel abgebildet wird, ist die einfachste Lösung nicht immer die beste. Wenn beim Vergrößern eines Bildes ein Pixel auf mehrere Pixel des neuen Bildes abgebildet wird, so kann man einfach den Wert des Ausgangspixels entsprechend oft in die Zielpixel kopieren. Dies führt allerdings oft zu unschönen Effekten. Das Resultat sieht „verpixelt“ aus. Pixelblöcke mit harten Kanten sind entlang der Konturen zu sehen. Gegen diesen Alias-Effekt kann man wiederum einen Weichzeichner als Anti-Aliasing nach der Skalierung anwenden, der die entstandenen Stufen etwas glättet (►Abbildung 3.16).

Es existieren zahlreiche Verfahren, mit denen bei verschiedenen geometrischen Operationen das Resultat optimiert werden kann. Oft gibt es keine optimale Variante der Optimierung, sondern je nach Anwendung und Bildmaterial kann zwischen verschiedenen Verfahren ausgewählt werden (z.B. bikubische oder bilineare Interpolation). In den meisten Programmen zur Bildbearbeitung werden solche Verfahren entweder automatisch oder von Hand ausgewählt, um die Resultate zu optimieren.

3.8.4 Weitere Operationen

Mit den einfachen geometrischen Operationen und den hier vorgestellten Filtern und Verfahren zur Bildverarbeitung geben sich moderne Programme zur Bildbearbeitung nicht zufrieden. Sie bieten eine Vielzahl weiterer Operationen an, die oft für viele Nutzer schon zur undurchschaubaren Vielfalt geworden sind. Aber auch damit sind noch lange nicht die Möglichkeiten erschöpft. Wissenschaftler entwickeln immer neue Verfahren, mit denen Bilder manipuliert werden können. Dazu gehören z.B. Verfahren zum Zusammenfügen von Bildern aus vielen Teilbildern, das Ersetzen von Texturen, die nachträgliche Veränderung der Beleuchtung oder die Löschung und Ersetzung von unerwünschten Inhalten auf Bildern.

3.8.5 Welches Format für welchen Zweck

In diesem Kapitel wurden einige Aspekte digitaler Bilder vorgestellt. Bei den unterschiedlichen Formaten und Auflösungen stellt sich natürlich die Frage, wie sollen Bilder am besten codiert und gespeichert werden, so dass möglichst viele Freiheiten bei der Bearbeitung erhalten bleiben.

- **Aufnahmeformat:** Bei der Aufnahme gilt eigentlich immer die Regel: so viel wie möglich speichern. Speicherkarten sind so erschwänglich, dass hier nicht gespart werden muss. Das heißt, man sollte
 - die maximale Auflösung verwenden, da man später vielleicht nur einen Ausschnitt verwenden will;
 - die maximale Farbtiefe verwenden, da dann nachträglich noch viel am Histogramm geändert werden kann und Über- bzw. Unterbelichtungen besser korrigiert werden können;
 - keine Kompression verwenden, also immer RAW-Bilder speichern, da zu frühe verlustbehaftete Kompression unter Umständen schon Details löscht.
- **Archivformat:** Wenn es irgendwie geht, sollte man die RAW-Bilder speichern und außerdem alle durch Bearbeitung entstandenen Varianten.
- **Komprimierte Formate für das Web und Anwendungen:** Wenn Bilder auf Computermonitoren betrachtet werden, reichen meist komprimierte Formate aus:
 - JPEG: eignet sich besonders für digitale Fotos. Im Web sollte so stark komprimiert werden, dass eine Seite auch bei geringer Bandbreite innerhalb weniger Sekunden geladen ist.
 - GIF: eignet sich besonders für Rastergrafiken, bei denen wenige Farben vorkommen. Es ist z.B. für Buttons und Icons ein sehr gutes Format.
 - PNG: ist nicht so verbreitet wie GIF, aber moderner und flexibler. Insbesondere, wenn ein echter Alphakanal benötigt wird, ist es sehr gut geeignet.

Auf die vielen weiteren Formate kann hier nicht erschöpfend eingegangen werden. Mit den Grundlagen dieses Kapitels wird man aber schnell auch bei anderen Verfahren eine Vorstellung gewinnen, wo deren Vor- und Nachteile liegen.

Zusammenfassung

Digitale Bilder und deren Erfassung, Kodierung und Verarbeitung können kaum ohne die menschliche Wahrnehmung verstanden werden. Vor allem die Wahrnehmung von Farben ist zentral für die Kodierung von Bildern in verschiedenen Farbmodellen. Da wir nur drei unterschiedliche Farbrezeptoren (die Zapfen) haben, reichen in den Farbmodellen auch drei unterschiedliche Farbwerte, um scheinbar alle möglichen Farben des Frequenzspektrums des sichtbaren Lichts darzustellen.

Bei der **Codierung** von Bildern können unterschiedliche Farbtiefen angeben, wie genau ein Pixel an einem Punkt des Bildes Werte erfassen kann. Die Zahl der Pixel bestimmt die Auflösung eines Bildes. Bei hoher Auflösung und hoher Farbtiefe entstehen große Datenmengen, die durch geeignete Komprimierungstechniken reduziert werden können, ohne dass Unterschiede zum Original auffällig werden. Dabei verändern verlustfreie Verfahren das Original nicht und verlustbehaftete Verfahren versuchen, dort viel einzusparen, wo beim Betrachten des komprimierten Bildes am wenigsten Veränderung bemerkt wird. Das verlustbehaftete JPEG-Verfahren nutzt als ersten verlustbehafteten Kompressionsschritt das Chroma-Subsampling, bei dem im YCrCb-Farbmodell die Chromakanäle auf weniger Pixel reduziert werden (auf 50% oder auf 25% der Pixel). Nach der diskreten Cosinustransformation ist die neue Quantisierung der DCT-Koeffizienten der zweite verlustbehaftete Schritt. Verlustfreie Verfahren runden die Kompression ab (Zick-Zack-Scan, Lauflängencodierung, Entropiecodierung).

Bildbearbeitung bietet eine Vielzahl von Operationen, die Bilder verändern. Punktoperationen, Filter und geometrische Operationen sind einige der wichtigsten und grundlegenden Klassen solcher Verfahren. Sie ermöglichen es unter anderem, Bilder zu drehen, Kontraste zu optimieren, zu schärfen und weichzuzeichnen.