

Bildkompression am Beispiel JPEG

Medientechnologie IL

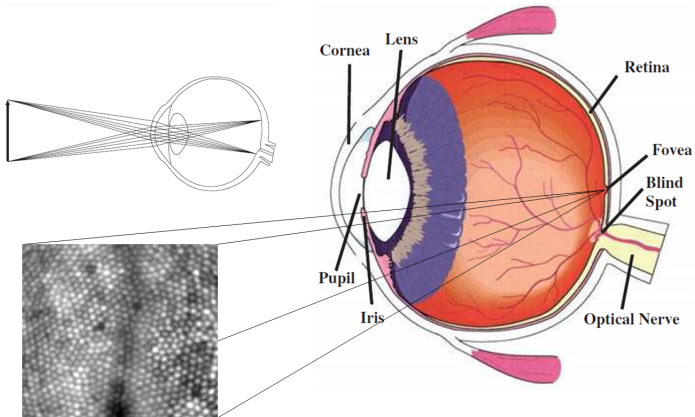
Andreas Unterweger

Vertiefung Medieninformatik und Bildverarbeitung
Studiengang ITS
FH Salzburg

Sommersemester 2023

Menschliche Bildwahrnehmung I

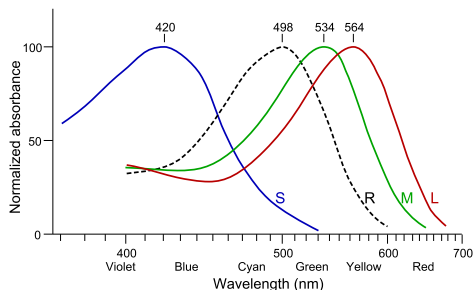
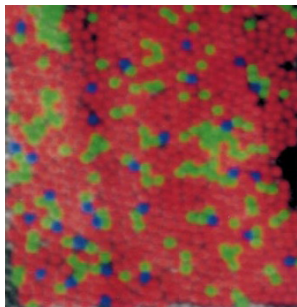
- Rezeptoren auf Netzhaut in Augen registrieren Photonen
- Gehirn interpretiert Lichtintensität als Bildinformation



Quellen: Roorda, A. und Williams, D.: The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520-522, 1999; Nadenau, M.: Integration of Human Color Vision Models into High Quality Image Compression. Doktorarbeit, Ecole Polytechnique Federale de Lausanne, 2000.

Menschliche Bildwahrnehmung II

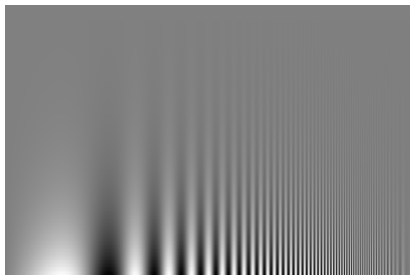
- Stäbchen (Rezeptoren): Nachtsicht (nur in Graustufen)
- Zapfen (Rezeptoren): Drei Arten → Farbwahrnehmung
- Helligkeitsempfinden folgt Stevenschem Potenzgesetz
- Kontrastempfindlichkeit höher als Helligkeits- und Farbempfindlichkeit



Quellen: Roorda, A. und Williams, D.: The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520-522, 1999; <http://commons.wikimedia.org/wiki/File:Cone-response.svg>

Menschliche Bildwahrnehmung III

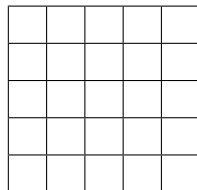
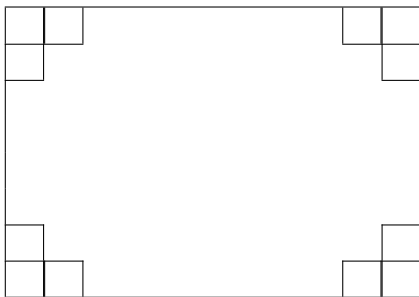
- Ortsfrequenz: Kehrwert der Länge der Periode eines (sinusförmigen) Hell-/Dunkelwechsels (Beispiel: Blauer Himmel vs. Gebüsch)
- Kontrast: Helligkeitsunterschied zwischen hellster und dunkelster Stelle (bzw. hellstem und dunkelstem Bereich)
- Kontrastempfindlichkeit abhängig von Ortsfrequenz



Quelle: Nadenau, M.: Integration of Human Color Vision Models into High Quality Image Compression. Doktorarbeit, Ecole Polytechnique Federale de Lausanne, 2000.

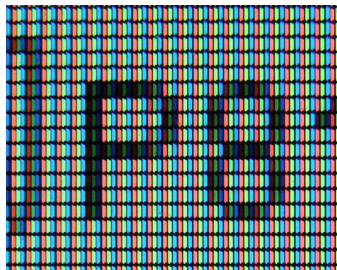
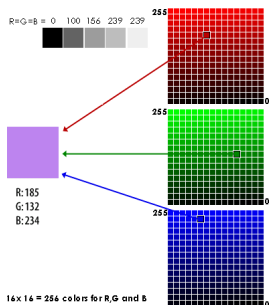
Digitale Bilder: Pixel

- Digitalisierung der Lichtintensität liefert digitales Bild
- Abtastung im Ortsbereich (zwei Dimensionen)
- Kleinste Einheit: Bildelement (englisch **P**icture **E**lement, *Pixel*)
- Pixelgröße beschränkt Ortsfrequenz (Abtasttheorem!)
- Quantisierung der Intensität → Wertebereich pro Pixel



Digitale Bilder: Farbpixel

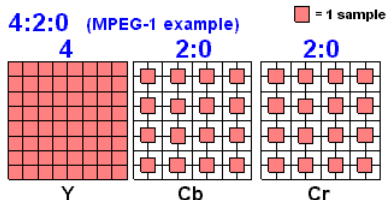
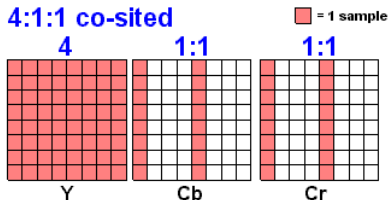
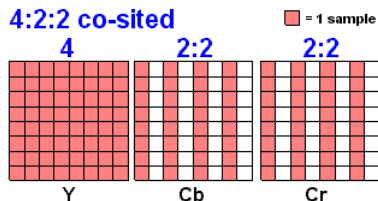
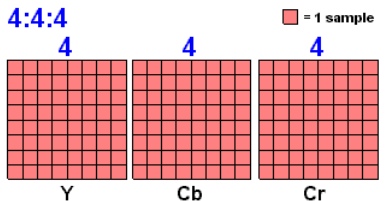
- Lichtintensität alleine würde nur Graustufenbilder ermöglichen
- Wellenlängenfilter für Rot-/Grün- und Blaubereich bei Aufnahme
- Lichtintensität wird für jeden Bereich separat quantisiert
- Drei Intensitätswerte (R, G und B) pro (Farb-)Pixel
- Farbwahrnehmung durch additive Mischung (RGB-Farbraum)



Quellen: Unbekannt; http://en.wikipedia.org/wiki/File:Closeup_of_pixels.JPG

- Mensch ist empfindlicher auf Helligkeit als auf exakte Farbe
- In RGB-Farbraum nicht ausnutzbar, da nicht getrennt
- Andere Farbräume als RGB möglich; Beispiel YCbCr („YUV“)
 - Luminanzkanal (Y): Entspricht gewichtetem Mittel von R, G und B
 - Zwei Chrominanz-(Farbdifferenz-)Kanäle zu blau und rot (Cb und Cr)
 - Kein dritter Differenzkanal notwendig (aus Rest berechenbar)
 - Prinzipiell verlustlos von und in RGB-Farbraum umrechenbar
- Vorteile für weitere Verarbeitung:
 - Luminanzkanal kann feiner quantisiert werden
 - Chrominanzkanäle können gröber quantisiert werden
 - Chrominanzkanäle können unterabgetastet werden

Chrominanzunterabtastung (Chroma Subsampling) I



Quelle: <http://www.pcmag.com/encyclopedia/term/57460/chroma-subsampling>

Chrominanzunterabtastung (Chroma Subsampling) II



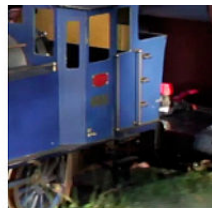
4:1:1



4:2:0



4:2:2



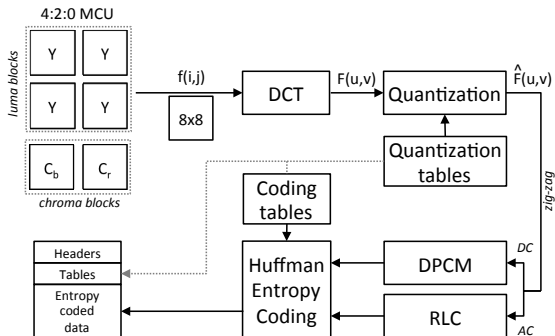
4:4:4



Quelle: <http://en.wikipedia.org/wiki/File:Colorcomp.jpg>

Übersicht JPEG

- Joint Photographic Experts Group (JPEG)
- Stillbildkompressionsstandard (ITU-T T.81)
- Verarbeitet Bilder üblicherweise im YCbCr-Farbraum (andere möglich)
- Unterteilt Bilder in $8 \cdot 8$ große Blöcke vor Transformation



Quelle: Temmermans, F., Bruylants, T., Schelkens, P., und Ebrahimi, T.: An introduction to JPEG Standards for Digitization and Archiving Applications. Jornadas Imatge i Recerca, vol. 14, pp. 19-33, 2016.

Zweidimensionale DCT I

- Menschliche Bildwahrnehmung ist frequenzabhängig → DCT
- Bild hat zwei Dimensionen, DCT ist eindimensional → Erweitern
- Ausgangspunkt: Symmetrische DCT-II (normalisiert):

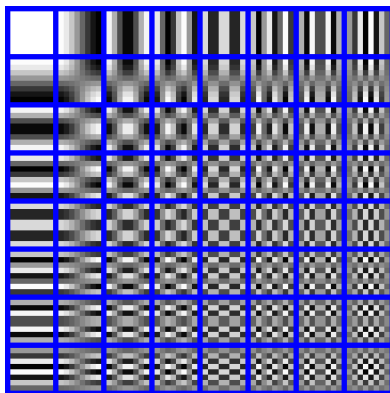
$$Y_k = \sum_{n=0}^{N-1} \eta_k X_n \cos\left(\frac{\pi k (n + \frac{1}{2})}{N}\right), \eta_i = \begin{cases} \sqrt{\frac{1}{N}} & i = 0 \\ \sqrt{\frac{2}{N}} & \text{sonst} \end{cases}$$

- Ohne Details: DCT(transformierte Zeilen) = 2-D-DCT
- Eingangsdaten: $X_{m,n}$, Ausgangsdaten: $Y_{k,l}$ (mit je $M \cdot N$ Elementen)
- m : Spalte, n : Zeile → k : horizontale, l : vertikale Ortsfrequenz

$$Y_{k,l} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \eta_k \eta_l X_{m,n} \cos\left(\frac{\pi k (m + \frac{1}{2})}{N}\right) \cos\left(\frac{\pi l (n + \frac{1}{2})}{N}\right)$$

Zweidimensionale DCT II

- JPEG: $M = N = 8$ (quadratische Blöcke)
 - Eingangswerte sind rein positiv ($\forall i, j \in \mathbb{N}^+ : X_{i,j} \in \mathbb{N}^+$), z.B. 0–255
 - cos-Funktionsterme erfordern positive und negative Werte
- Eingangswertebereich durch Subtraktion verschieben, z.B. -128–127



Quelle: <http://en.wikipedia.org/wiki/File:Dctjpeg.png>

- Lineare Quantisierung (unabhängig pro DCT-Koeffizient)
 - Menschliche Bildwahrnehmung empfindlicher bei niedrigen Frequenzen
 - Niedrige Frequenzen feiner quantisieren
 - Menschliche Bildwahrnehmung unempfindlicher bei hohen Frequenzen
 - Hohe Frequenzen gröber quantisieren

$$Y'_{k,l} = \left\lfloor \frac{Y_{k,l}}{Q_{k,l}} + \frac{1}{2} \right\rfloor$$

→ 8 · 8-Matrix mit Quantisierungswerten notwendig

- JPEG-Standard schlägt Standardmatrizen vor (Beispiel für Y-Kanal):

$$Q = \begin{pmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{pmatrix}$$

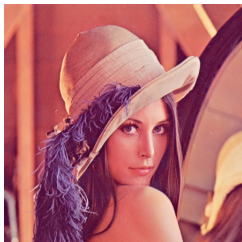
- In der Praxis: Skalierung von Q (JPEG-Referenzsoftware)
- Qualitätsstufen q zwischen 1 und 99%

$$Q'_{k,l} = \left\lfloor \frac{\lambda Q_{k,l}}{100} + \frac{1}{2} \right\rfloor, \lambda = \begin{cases} \left\lfloor \frac{5000}{q} \right\rfloor & 1 \leq q \leq 50 \\ 200 - 2q & 51 \leq q \leq 99 \end{cases}$$

- Wertelimitierung auf Bereich zwischen 0 und 255
- Höhere Qualität → Feinere Quantisierung
- Niedrigere Qualität → Gröbere Quantisierung
- Separate Quantisierungsmatrix für Cb- und Cr-Kanäle (siehe Anhang K von ITU-T T.81, RFC 2435 u.a.)

Quantisierung III

$q = 99$



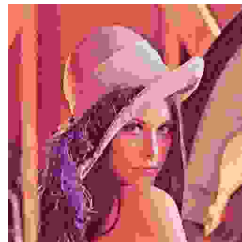
$q = 50$



$q = 20$

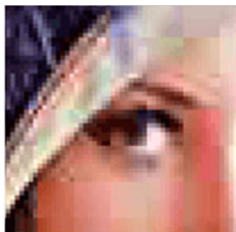


$q = 1$



Quantisierung IV

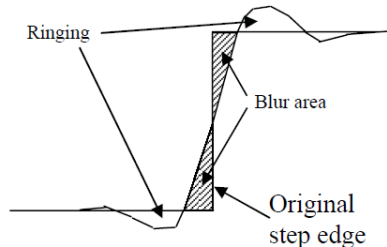
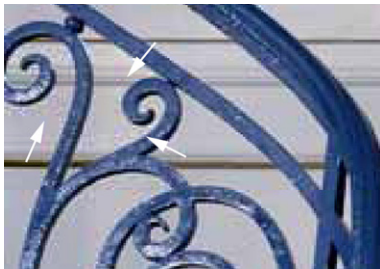
- Bei grober Quantisierung: Blockartefakte (englisch *Blocking*)
 - Blöcke werden unabhängig voneinander transformiert und quantisiert
- Diskrepanz der Rekonstruktionen an Blockrändern → Sichtbarwerden
- Bei grober Quantisierung: Unschärfeartefakte (englisch *Blurring*)
 - Verlust hoher Frequenzen erzeugt Tiefpasswirkung



Quelle: <http://fh-salzburg.ac.at>

Quantisierung V

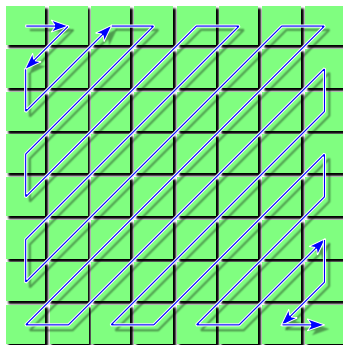
- Überschwingartefakte (englisch *Ringing*)
- Analogon zum Gibbsschen Phänomen bei Fourierreihen
- Tritt vor allem bei harten Kanten auf (auch bei feiner Quantisierung)



Quellen: Le Dinh, P. T. und Algorith, J. P.: Video compression artifacts and MPEG noise reduction. <http://www.embedded.com/design/real-time-and-performance/4013028/Video-compression-artifacts-and-MPEG-noise-reduction> (abgerufen am 6.6.2016), 2006; Punchihewa, G. A. D., Bailey, D. G. und Hogson, R. M.: Objective Evaluation of Edge Blur and Ringing Artefacts: Application to JPEG and JPEG 2000 Image Codecs. In Proceedings of Image and Vision Computing New Zealand 2005, pages 61-66, Dunedin, New Zealand, 2005.

Zickzack-Scan

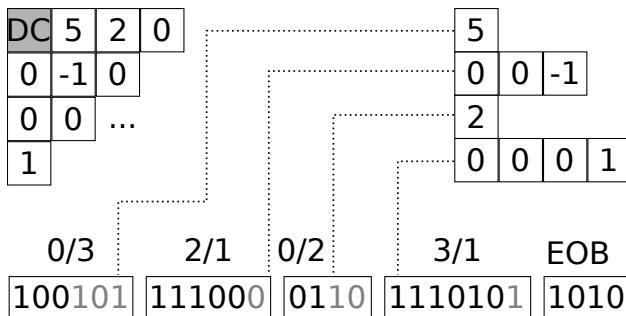
- Hochfrequente Koeffizienten sind nach Quantisierung großteils null
 - Lange Nullfolgen lassen sich leichter komprimieren
- AC-Koeffizienten von niedrig- nach hochfrequent scannen
(DC-Koeffizienten werden separat als Differenzen Huffman-kodiert)



Quelle: http://upload.wikimedia.org/wikipedia/commons/4/43/JPEG_ZigZag.svg

Laufängenkodierung (englisch *Run-Length Encoding*)

- Gruppierung von Nullen (Anzahl Nullen: *Run*) und je einem Wert
- Länge der Binärcodierung des Wertes (mit Vorzeichen!): *Length*
- Run-Length-Kombination (r, l) wird als Huffman-Codewort kodiert
- Wert w wird als w (wenn positiv) bzw. $2^l - 1 + w$ (wenn negativ) binär kodiert → Codewort 0 entspricht kleinstmöglichem Wert
- Symbol EOB (End of Block) ist Platzhalter für restliche Nullen



Fragen?