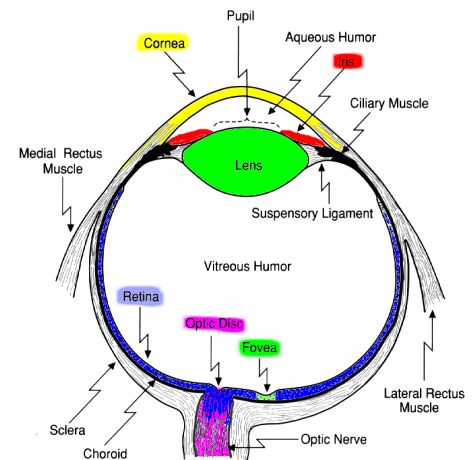


3. Bildaufnahme

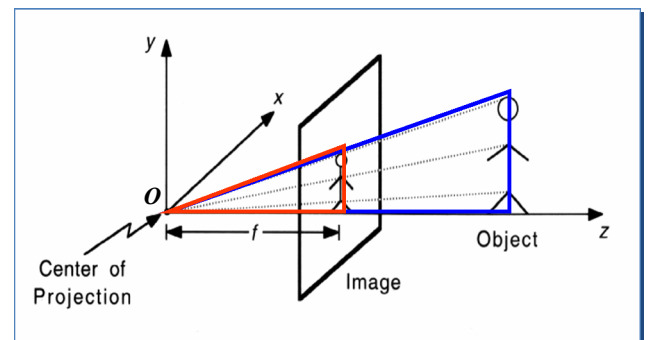
Das menschliche Auge

Das Auge des Menschen ist ein Organ, welches auf Lichtreize reagiert und für den Menschen das Sehen ermöglicht. Erreicht sichtbares Licht das Auge, kann dieses Informationen über die Umgebung wahrnehmen. Verschiedene physiologische Bausteine sind am Sehen beteiligt und werden unter dem Begriff "visuelles System" zusammengefasst. Der Vorgang des Sehens funktioniert folgendermaßen: Die Linse des Auges dient zur Fokussierung des Abbildes der Umgebung, welches auf eine lichtempfindliche Membran (Retina/Netzhaut) an der Rückseite des Auges seitenverkehrt und auf dem Kopf stehend projiziert wird. Die Netzhaut besteht aus Fotorezeptoren, welche durch Lichtreize angeregt werden, neuronale Impulse generieren und diese zu verschiedenen Hirnbereichen für die Verarbeitung weiterleiten. Chemische und elektrische Vorgänge sind an der Generierung des neuronalen Impulses beteiligt. Die Retina besteht aus mehreren Schichten von Neuronen, die durch Synapsen miteinander verbunden sind. Bei den lichtempfindlichen Fotorezeptorzellen können zwei Typen unterschieden werden: Stäbchen und Zäpfchen. Stäbchen ermöglichen das Schwarz-Weiß-Sehen, während Zäpfchen für die Wahrnehmung von Farbe verantwortlich sind. Die Fovea centralis liegt im Zentrum des Gelben Fleckes und stellt den Bereich des schärfsten Sehens dar. 50% der Nervenfasern des Sehnervs leiten Impulse der Fovea an das Gehirn weiter, die anderen 50% der Fasern Impulse der restlichen Rezeptoren in der Netzhaut.



Perspektivische Projektion

Das Modell einer Lochkamera beschreibt die mathematische Beziehung zwischen einem 3D Punkt und dessen Projektion auf eine Bildebene einer idealen Lochkamera. Die Kameraablenke wird ohne Verwendung einer Linse zur Fokussierung von Licht verwendet und durch einen Punkt beschrieben. Dies ist das einfachste Bildaufnahmegerät, das die Geometrie der perspektivischen Projektion genau wiedergibt. Das Modell beinhaltet keine geometrischen Verzerrungen oder Ungenauigkeiten von unfokussierten Objekten, welche durch Linsen und endlich dimensionierte Blenden hervorgerufen wird. Daher wird das Modell nur für eine vereinfachende Annäherung der echten Abbildung einer 3D Szene auf ein 2D Bild herangezogen.



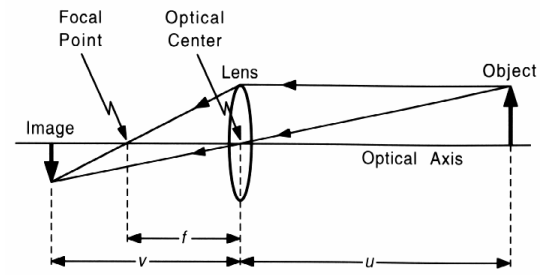
Zur Vereinfachung der Herleitung der Gleichung zur perspektivischen Projektion werden folgende Annahmen getroffen (siehe Abbildung oben):

- (1) Das Zentrum der Projektion stimmt mit dem Ursprung der Welt überein.
- (2) Die Kameraachse (optische Achse) ist an der z-Achse der Welt ausgerichtet.
- (3) Die Bildebene ist im Vordergrund des Zentrums der Projektion, um eine Bildumkehrung zu vermeiden.

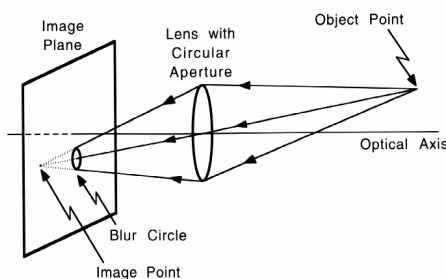
Das Modell besteht aus einer Ebene (Bildebene) und einem 3D Punkt O (Zentrum der Projektion). Die fokale Länge f , die Entfernung zwischen der Bildebene und dem Zentrum der Projektion O stellt die Brennweite dar (z.B. die Entfernung von der Linse zum CCD Array). Die optische Achse ist jene Gerade die durch O führt die normal auf die Bildebene steht. Folgende ähnliche Dreiecksbeziehungen bestehen dabei: $\frac{y}{Y} = \frac{f}{Z}$, y stellt die Projektion des Objektes in der Bildebene dar, Y die Objektgröße und Z die Entfernung des Objektes zu O . Folgende ähnliche Zusammenhänge können daraus abgeleitet werden: $x = \frac{f}{Z}X$, und $y = \frac{f}{Z}Y$, welche die Gleichungen der perspektivischen Projektion darstellen. Diese Projektion ist nicht linear (die Entfernung zu einem Objekt ist umgekehrt proportional zu dessen Bildgröße), ist eine $n:1$ Abbildung (mehrere 3D Punkte können auf denselben Punkt abgebildet werden), 3D Linien werden auf 2D Linien projiziert, Entfernungen und Winkel bleiben nicht erhalten und parallele Linien werden nicht auf parallele Linien projiziert (sofern diese nicht parallel zur Bildebene liegen). Im Allgemeinen geht die 3D Informationen in einer 2D Projektion für immer verloren.

Linsen

Der größte Nachteil einer Lochkamera, die zu geringe Menge an eindringendem Licht, wird durch die Einführung der Linse verbessert. Eine Linse ist ein optisches Bauteil mit annähernd perfekter axialer Symmetrie, welches Licht bricht und somit die von einem Punkt ausgehenden Lichtstrahlen sammelt. Einfache Linsen bestehen aus einem optischen Element, welches parallele Strahlen, die in Richtung der optischen Achse einfallen, im Brennpunkt, der die Entfernung f zur Linse hat, fokussieren. Liegt ein Objekt in einer endlichen Entfernung u , werden dessen Lichtstrahlen zu einem Bild hinter dem Brennpunkt in der Entfernung v fokussiert. Die zur Linsenachse senkrecht liegende Ebene, welche mit einer Entfernung v von der Linse entfernt liegt, wird Bildebene genannt. Die Beziehung zwischen einer Entfernung u vom Objekt zur Linse und der Entfernung v von der Linse zum Bild wird durch die vereinfachte Linsengleichung beschrieben: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$. Dies besagt, dass Objekte die sehr weit weg sind ($u = \infty$, $v = f$) im Brennpunkt scharf abgebildet sind, alle Objekte die näher sind, dahinter. Die fokale Länge f bestimmt auch den Öffnungswinkel der Kamera. Ist f nahe an der Linse (= klein) ist der Öffnungswinkel groß (Weitwinkel), ist f groß, so ist der Öffnungswinkel klein (Zoom).



Tiefenschärfebereich (DOF)



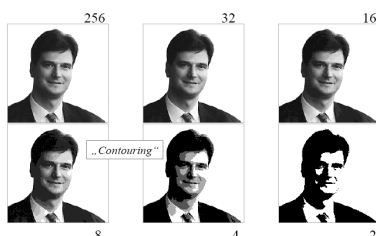
Theoretisch sind auf Grund der vereinfachten Linsengleichung immer nur Objekte mit einer bestimmten Entfernung zu Kamera scharf abgebildet, alle anderen unscharf. In der Praxis werden jedoch Objekte in einem gewissen Tiefenschärfebereich (engl.: Depth Of Field (DOF)) als scharf abgebildet bezeichnet. Der DOF ist somit die Entfernung zwischen dem nächsten und weitesten entfernten Objekt in einer Szene, welches vertretbar scharf in einem Bild dargestellt werden kann. Der Tiefenschärfebereich eines Objekts kann durch die Sensorauflösung und Blendenöffnung beeinflusst werden. Die Blende eines optischen Systems

ist eine ringförmige Öffnung vor der Linse, welche den Kegelwinkel des Strahlenbündels bestimmt, welches auf die Bildebene trifft. Die Blende bestimmt, wie viele einfallende Strahlen durchgelassen werden und daher wie viel Licht die Bildebene erreicht (je kleiner die Blende, desto weniger Strahlen treffen die Bildebene und desto dunkler das Bild für eine gegebene Belichtungszeit). Der Tiefenschärfebereich ist durch die Größe der Linse und damit auch der maximalen Blende bestimmt. Je größer die Blende, desto mehr Strahlen eines Objektpunktes gelangen auf die Bildebene. Ist diese jetzt nicht im richtigen Abstand v zur Linse (zu weit vorne oder hinten) fallen die Strahlen nicht mehr in einem Punkt zusammen, sondern bilden einen Kreis (Blur Circle) auf der Bildebene. Benachbarte Objektpunkte bilden überlagerte Kreise auf der Bildebene – das Bild wird unscharf. Der Bereich der scharf abgebildet wird ist somit abhängig von der Größe der Unschärfekreise, die wiederum von der Größe der Blende abhängen (d.h. eine größere Blende führt somit zu einem kleineren DOF). Weiters spielt auch die Sensorauflösung eine Rolle, je höher die Auflösung, desto besser können auch die Unschärfekreise abgebildet werden (DOF wird kleiner).

Radiometrie

Radiometrie bezeichnet die Menge an Methoden zur Messung elektromagnetischer Strahlung (inklusive des sichtbaren Lichtes). Die radiometrische Beziehung zwischen der Welt und deren Projektionen wird durch die Menge des gesamten reflektierten Lichtes eines Oberflächenpunktes beschrieben, welche auch als Strahldichte (engl. radiance) bezeichnet wird. Die Menge an Licht, die von diesem Punkt auf das Bild projiziert wird, wird Bestrahlungsstärke (engl. irradiance) genannt. Diese radiometrischen Größen charakterisieren die Strahlung aller Frequenzen, betrachtet man jeden Frequenzbereich getrennt, nennt man das Spektrum. Das Strahlungsaufkommen einer Oberfläche wird spektrale Bestrahlungsstärke (engl. spectral irradiance) genannt. Die Bestrahlungsstärke wird in der Einheit Watt pro Quadratmeter (W/m^2) angegeben und wird bei Bildern als Helligkeit bezeichnet.

Sampling



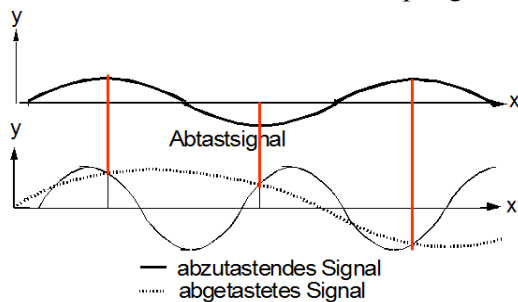
Die Bildauflösung ist ein Sammelbegriff für den Detailreichtum eines Bildes und wird von der *radiometrischen Bildauflösung*, *Pixelauflösung*, *räumlichen* und *zeitlichen Auflösung* bestimmt. Die radiometrische Auflösung bestimmt wie fein ein System Helligkeitsunterschiede repräsentiert oder unterscheidet und wird durch die Anzahl von Stufen oder Bits beschrieben (typischerweise 8 bits oder $2^8 = 256$ Helligkeitsstufen). Der Mensch kann ca. 120 verschiedene Grauwerte erkennen,

somit reichen 256 Graustufen aus, um ein Bild realitätsnah zu erzeugen. Werden weniger Graustufen verwendet, sinkt der Speicherbedarf, aber auch die Qualität, Contouringeffekte entstehen. Die Pixelauflösung steht für die Anzahl der Bildelemente (Pixel) in digitalen Bildern, d.h. die Anzahl der Pixel pro Zeile (Breite) mal Anzahl der Pixel pro Spalte (Höhe), z.B. 640 x 480, und wird üblicherweise in Megapixel angegeben.

Eine andere Möglichkeit, die Auflösung zu beschreiben, ist Pixel per Längeneinheit (engl. dots per inch (dpi) oder pixels per inch (ppi)). Diese ist aber nur für Flachbettscanner gültig, da dort der Abstand zwischen Objekt und Sensor bekannt ist, für Kameras im Allgemeinen nicht. Diese sogenannte räumliche (geometrische) Auflösung gibt an, wie nahe beieinander liegende Linien in einem Bild aufgelöst werden können und hängt nicht nur von der Pixelauflösung sondern auch von den Eigenschaften des Bildaufnahmesystems ab. Sie wird durch die minimale Entfernung zwischen zwei benachbarten Merkmalen oder die minimale Größe eines Merkmals definiert.

Die zeitliche Auflösung gibt an, wie viele Bilder pro Sekunde aufgenommen werden (z.B. 25 frames per second).

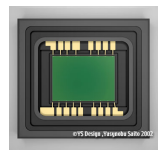
In digitalen Kameras wird anhand der zur Verfügung stehenden Auflösungen eine analog/digital Wandlung im Raum, in der Zeit und im Spektrum durchgeführt. Die Auflösungsgrenze wird durch das *Nyquist-Shannon Sampling Theorem* berechnet. Unter Sampling versteht man den Umwandlungsprozess von einem Signal (z.B. eine



kontinuierliche Funktion in Zeit oder Raum) zu einer numerischen Folge (eine diskrete Funktion in Zeit oder Raum). Das Theorem besagt, dass ein bandbegrenzt analoges Signal perfekt von einer unendlichen Folge von Samples rekonstruiert wird, wenn die Samplingrate $2 \cdot B$ Samples pro Sekunde (engl. samples per second) übersteigt. B steht für die höchstvorkommende Frequenz im Originalsignal. Das heißt um ein Signal $f(t)$ exakt wiederzugewinnen, muss man $f(t)$ mit einer Rate abtasten, die größer als das Doppelte seiner höchsten Frequenzkomponente ist ($f_{\text{abtast}} > 2 \cdot f_{\text{maxSignal}}$).

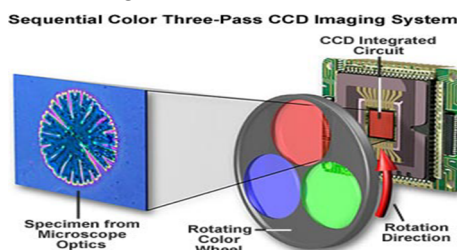
Bildsensoren

Ein Bildsensor ist ein Bauelement, das ein optisches Bild in ein elektronisches Signal umwandelt. Es sind zwei Arten von Pixelsensoren (*digitale Charge-Coupled Device (CCD)* oder *Complementary Metal-Oxide-Semiconductor (CMOS)*) in Verwendung. Ein CCD Sensor ist ein analoges Bauteil und besteht aus lichtempfindlichen Transistoren. Bei einer gegebenen Belichtungszeit (engl. exposure time) baut sich, während der Direktstrahlungsanteil des Lichtes auf den CCD Sensor auftrifft, dort eine elektrische Ladung auf und speichert diese. Zeile für Zeile wird diese ausgelesen und zu einer Spannung konvertiert, die anschließend in digitale Information umgewandelt wird. CCD Kameras arbeiten wie auch gewöhnliche analoge Filmkameras nach dem Prinzip des Flächenintegrals, und sind daher linear. CCD Sensoren (ausgenommen high-end frame transfer CCDs) sind empfindlicher für vertikale Abstriche von hellen Lichtquellen bei einer Überladung des Sensors. Ein CMOS Bildaufnahmechip hingegen besitzt aktive Pixelsensoren, das heißt, jedes Pixel ist individuell auslesbar, die Spannung, die der Helligkeit entspricht liegt immer an. CMOS Bausteine produzieren wenig Abwärme und verbrauchen nur beim Hin- und Herschalten von ON und OFF Zuständen der Transistoren Energie. Dieser Sensor hat durch die unabhängige Betrachtung der Nachbarn jedes Pixels einen nicht-linearen Output. CMOS haben im Vergleich zu CCD Sensoren einen geringeren Energieverbrauch, schnellere Lesegeschwindigkeit und sind in der Herstellung billiger.



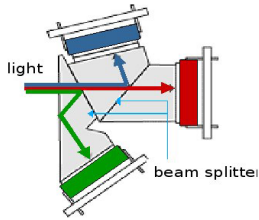
Farbe auf CCD Sensoren

Digitale Farbkameras können Farbe auf drei verschiedenen Wegen erfassen: Field Sequential Technik, 3 Chip und Color Filter Array (CFA). Alle diese verschiedenen Arten verwenden Farbfiler, da ein Fotosensor die Lichtintensität nur mit einer kleinen oder nicht ausreichenden Genauigkeit für Wellenlängen detektiert und daher keine Farbinformation herausfiltern kann. Die Farbfiler filtern das Licht aufgrund ihres Wellenlängenbereiches, sodass die einzelnen gefilterten Intensitäten Information über die Farbe des Lichtes beinhalten.



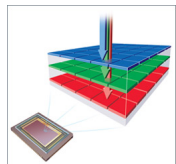
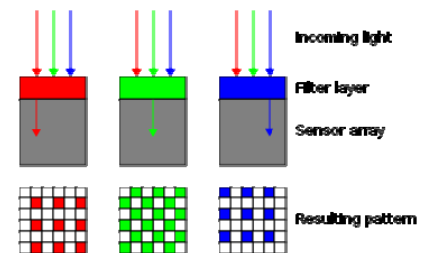
Die Field Sequential Technik wurde in den frühen 1900er Jahren durch Prokudin Gorskii erfunden, der die Farbinformation der Szene dadurch gewann, dass er mithilfe von Rot-, Grün- und Blau-Filtern drei Schwarz-Weiß Bildern aufnahm und auf Glasplatten entwickelte. Durch den erneuten Einsatz von Farbfiltren war er in der Lage diese Bilder mithilfe eines Lichtprojektionssystems (Lantern Projektor) in Farbe darzustellen.

Jede Bewegung in der Bildaufnahme resultiert aber in „Geisterbildern“, da die roten, grünen und blauen Bilder der Objekte zu verschiedenen, hintereinander folgenden Zeitpunkten gemacht werden. In den USA wurde 1940 ein ähnliches Prinzip (Field Sequential System) zum Einsatz im Bereich Farbfernsehen entwickelt. Die primären Farbinformationen wurden ebenfalls mit Hilfe von aufeinanderfolgenden Bildern aufgenommen. Hierfür wurde ein mechanisches System verwendet, welches aus Farbfilterplatten, die mit 1440 rpm (rounds per minute), das sind 24 Umdrehungen pro Sekunde (72Hz Abtastung), innerhalb der Kamera rotierten, und einem Empfänger bestand, der die roten, blauen und grünen Fernsehbilder aufnahm und in einer Sequenz ausstrahlte. Diese Art der Farbaufnahme wird vor allem bei Mikroskopen verwendet, da sich dort die Szene nicht bewegt.



Durch die Verwendung eines strahlenbrechenden Prisma kann eine korrekte Farbtrennung erzielt werden, da das Eingangssignal in einen roten, grünen und blauen Anteil aufgespalten wird. In 3-CCD Kameras ist jeder dieser drei CCDs so angeordnet, dass er genau einen Farbkanal aufnimmt. Das Problem dabei ist es, die CCD Chips so anzuordnen, dass ein eintretender Strahl genau auf denselben Pixel in allen 3 CCD Chips fällt, denn nur so kann die Farbtreue gewährleistet werden. Diese Technik wird ausschließlich für Videokameras und Camcorder verwendet.

Die meistverwendeten Techniken zur Farbaufnahme mit CCDs sind jedoch Color Filter Array (CFA) oder Color Filter Mosaic (CFM). CFM verwendet ein Mosaik aus winzigen Farbfiltern, die über jedem Pixels eines Bildsensors platziert werden, um Farbinformation aufzunehmen. Die am häufigsten verwendeten CFAs, Bayer Filter genannt, bestehen aus 50% grün, 25% rot und 25% blau und werden daher auch als RGBG, GRGB oder RGGB bezeichnet. Der Filter wurde nach seinem Erfinder, Bryce E. Bayer von der Firma Eastman Kodak, benannt. Er verwendet doppelt so viele grüne Elemente wie rote oder blaue, um die Physiologie des menschlichen Auges nachzuahmen. Der Mensch ist für grünes Licht am empfindlichsten und der Wellenbereich für Grün liegt zwischen Blau und Rot. Der rohe (engl. raw) Output einer Bayer-Filter Kamera wird als Bayer Muster Bild (engl. Bayer pattern image) bezeichnet. Jedes Pixel nimmt nur eine der drei Farben auf, somit kann die tatsächlich Farbe des Pixels, das ja aus RGB Komponenten besteht, nicht bestimmt werden. Um ein voll farbiges Bild zu erhalten, werden Interpolationsalgorithmen verwendet, um die fehlenden Rot, Grün und Blauanteile je Farbkanal für jeden fehlenden Punkt aus dessen Nachbarn zu berechnen (engl. demosaicing). Der Rechenvorgang kann innerhalb der Kamera gemacht werden, z.B. durch Konvertierung des Raws in ein JPEG oder TIFF Format, oder außerhalb der Kamera, durch die direkte Verwendung der Rohdaten des Sensor in einem Bildanzeigeprogramm (wie zum Beispiel nomacs, www.nomacs.org).



Der Foveon X3 Sensor ist eine Kombination der 3CCD und der CFA Technik, verwendet ein Array von Fotoelementen, welches jeweils aus drei vertikal gestapelten Fotodioden besteht und in einem zweidimensionalen Raster angeordnet ist. Jeder dieser gestapelten Fotodioden reagiert auf verschiedene Wellenlängen des Lichtes, d.h. jede hat eine unterschiedliche spektrale Sensitivitätskurve. Das Signal der drei Fotodioden wird anschließend verarbeitet und resultiert in Daten, welche drei additive Primärfarben rot, grün und blau beinhalten und das Farbbild fehlerfrei darstellen kann.

Ideale vs. Reale Kamera

Reale Kameras weichen von idealen Kameramodellen ab, da sie Linsen mit optischen Phänomenen und nicht-ideale Bildsensoren mit Chipartefakten verwenden. Einige dieser Störungen sind die folgenden:

Die *optische Aberration* tritt auf, wenn Licht von einem Punkt auf einem Objekt nicht in einem einzigen Punkt nach der Transmission durch die Linse zusammenfällt.

Chromatische Aberration ist eine Art von Störung, bei der die Linse nicht im Stande ist, alle Farben im gleichen Konvergenzpunkt zu fokussieren. Diese Aberration tritt auf, da Linsen unterschiedliche Brechungsindizes für verschiedene Wellenlängen des Lichtes aufweisen. Chromatische Aberration manifestiert sich durch „Farbränder/-säume“ entlang der Abgrenzung zwischen dunklen und hellen Bildteilen.

Die *sphärische Aberration* tritt auf, da Kugeloberflächen nicht die ideale Form für Linsen sind, aber beim Schleifen von Glas die einfachste Form darstellen. Diese Aberration führt dazu, dass Strahlen, die weiter am Rand der Linse auftreffen in einem etwas anderen Punkt fokussiert werden, als Strahlen die nahe der Achse liegen.

Linsen Verzerrung ist eine Abweichung einer geradlinigen Projektion, gerade Linien in einer Szene bleiben nicht gerade im Bild, diese könne nach außen (tonnenförmig) oder nach innen (kissenförmig) verzerrt sein.

