|  |
| --- |
| IU |
| Einführung in Computer Vision |
| DLBAIICV01 |

# Übergeordnete Lernziele

Computer Vision beschreibt den Prozess und die Techniken, die bei der Verwendung von Computern zur Gewinnung von Informationen aus Bildern oder Videos zum Einsatz kommen. Diese Informationen werden zur einfachen Objekterkennung, zur Empfehlung bestimmter Handlungen oder sogar zur Ausführung komplexer Aktionen, wie z. B. dem autonomen Fahren von Fahrzeugen, verwendet. Diese Fähigkeiten erweitern den Anwendungsbereich von Computern. Autonome Fahrzeuge, Qualitätskontrolle in der Produktion und Gesichtserkennung sind nur einige wenige Beispiele für die Möglichkeiten von Computer Vision. Jeden Tag entdecken mehr und mehr Unternehmen neue Anwendungsbereiche für Computer Vision, wodurch dieses Forschungsfeld immer attraktiver wird. Dieses Studienskript stellt Ihnen die grundlegenden Prinzipien und Technologien vor, die in moderner Computer Vision verwendet werden.

**Einführung in Computer Vision** behandelt die wichtigsten Fakten und Schlüsselkonzepte rund um die Bilderfassung sowohl beim menschlichen Auge als auch bei technischen Systemen. Dabei lernen Sie, die Bedeutung der Filterung in der Bildverarbeitung und deren praktische Anwendungen zu beschreiben. Wir werden die Funktion von Low-Level-Merkmalen, wie Kanten oder markanten Punkten, bei der Bildverarbeitung betrachten. Sie lernen auch, wie Deep-Learning-Methoden erfolgreich bei High-Level-Sehaufgaben eingesetzt werden. Schließlich werden in diesem Studienskript die Besonderheiten der Videobearbeitung erläutert und es wird gezeigt, wie gängige Probleme bei der Interpretation von Videomaterial gelöst werden können.

# Lektion 1 – Grundlagen des Sehens

**Lernziele**

Nach dem Abschluss dieser Lektion werden Sie...

... verstehen, wie das menschliche visuelle System funktioniert.

... beschreiben können, wie eine Kamera eine Szene einfängt und sie in ein Bild umwandelt.

... in der Lage sein, eine Szene in einem digitalen Bild durch Zahlen darstellen.

# 1. Grundlagen des Sehens

## Aus der Praxis

Sie beschäftigen sich mit Computer Vision und möchten Ihren Computer trainieren, die Nummern von Spielkarten zu erkennen, die auf Ihrem Schreibtisch liegen. Für einen Menschen ist dies eine scheinbar sehr einfache Aufgabe, aber für einen Computer beinhaltet sie mehrere Aspekte.

Zuerst gestalten Sie die Situation – Spielkarten auf Ihrem Schreibtisch – mit den Zahlen nach oben. Sie erkennen sofort, was Sie vor sich haben, aber einem Computer muss man es zunächst beibringen. Computer sind superschnell und verarbeiten Hunderte von Millionen von Nullen und Einsen pro Sekunde. Sie wissen wahrscheinlich schon, dass Bilder durch Pixel dargestellt werden und dass jedes Pixel einen numerischen Wert hat. Zum Beispiel hat eine 12-MP-Kamera insgesamt 12 Millionen Pixel.

Die meisten Smartphones können Farbbilder aufnehmen. Machen Sie mit Ihrem Smartphone ein Foto von den Karten auf Ihrem Schreibtisch. Danach laden Sie das Bild auf Ihren Computer hoch, installieren Python und die Bibliothek OpenCV-python und lesen das Bild in eine Variable ein.

Sehen Sie sich die Variable genau an, um die Darstellung des Bildes besser zu verstehen. Das Ergebnis ist ein dreidimensionales Array. Unabhängig von den Farben auf der Karte repräsentiert jede Dimension eine der Farben rot, blau und grün. Der Ort eines jeden Pixels wird durch je eine x- und eine y-Koordinate angegeben. Zu jedem Pixel gehören drei numerische Werte, die seine genaue Farbe definieren.

## 1.1 Das menschliche visuelle System

Das menschliche visuelle System besteht aus den folgenden Komponenten: Hornhaut, Linse, Netzhaut, Sehnerv und Pupille, wie in der folgenden Abbildung dargestellt. Der visuelle Kortex befindet sich im Gehirn und verarbeitet die vom Auge erfassten Informationen. (Rill et al., 2014).

The Human Eye



Cornea

Quelle: Basierend auf Erin Silversmith (2019) CC BY-SA 3.0. [https://commons.wikimedia.org/wiki/File:Human\_eye\_cross\_section\_detached\_retina.svg](https://commons.wikimedia.org/wiki/File%3AHuman_eye_cross_section_detached_retina.svg)

Der erste Schritt im optischen System ist die Erzeugung eines Bildes auf der Netzhaut. Der Aufbau und die Funktionsweise von Hornhaut, Linse, Sehnerv und Pupille sind komplex und werden im Folgenden beschrieben:

### Hornhaut

Die harte weiße Haut an der Außenseite des Augapfels wird als Lederhaut (Sklera) bezeichnet. Die äußere durchsichtige Schicht an der Vorderseite des Auges ist die Hornhaut, die eine Fortsetzung der Sklera bildet. Diese transparente Hülle besteht hauptsächlich aus fein miteinander verwobenen Kollagenfasern. Der dünne Tränenfilm, der den äußeren Teil der Hornhaut bedeckt und für gleichmäßige Lichtbrechung sorgt, wird durch das gelegentliche Schließen des Augenlids aufrechterhalten. Die auftretenden Unterschiede im Tränenfilm führen zu Streuungen und geringfügigen Veränderungen der visuellen Anpassungsfähigkeit. Die Dicke der Hornhaut beträgt 0,5 - 0,6 mm.

### Pupille

Die Pupille hat im menschlichen visuellen System zwei Aufgaben. Erstens begrenzt sie die Lichtmenge, die die Netzhaut erreicht, zweitens ist sie für die Veränderung der Blendengröße im Abbildungssystems des menschlichen Auges verantwortlich. Eine größere Pupillenerweiterung führt aufgrund von Abbildungsfehlern und der Beugung im Auge nicht unbedingt zu einer besseren Bildqualität. Die Größe der Pupille ist nicht genau festgelegt, aber für die notwendige räumliche Wahrnehmung des Menschen variiert ihr Durchmesser üblicherweise zwischen 2 und 8 mm. Die Pupille, die als Blende dient, liegt zwischen Linse und Hornhaut, so dass eine Veränderung der Pupillengröße das Sehfeld nicht beeinflusst.

### Linse

Akkommodation

Akkommodation bezeichnet die automatische Anpassung des Auges, die es ermöglicht, in verschiedenen Entfernungen zu sehen, und wird in erster Linie durch Veränderungen der Konvexität der Augenlinse erreicht.

Die Linse befindet sich unmittelbar hinter der Iris. Zur Veranschaulichung dient die obige Abbildung. Der Brechwert der Linse wird in Dioptrien gemessen. Eine Dioptrie ist eine Maßeinheit für die Vergrößerungsstärke einer Linse. Dioptrien sind proportional zu eins geteilt durch die Brennweite. Die Brennweite ist der Abstand zwischen Linse und Sehgrube (Fovea centralis). Eine entspannte Linse hat einen Brechwert von 20 Dioptrien. Bei voller Akkommodation hat die Linse hingegen einen Brechwert von 30 Dioptrien. Akkommodation ist die Anpassung der Brechkraft, damit sowohl nahe als auch ferne Gegenstände vom Auge fokussiert und scharfgestellt werden können. Die Linse wird an ihrem Außenrand von Zonulafasern gehalten, die mit dem Ziliarkörper verbunden sind. Die Kontraktion der Ziliarmuskeln im Ziliarkörper führt zu einer Entspannung der Zonulafasern. Dadurch nimmt die Krümmung der Linse zu, so dass das Auge auf nähere Objekte fokussieren kann. Durch Entspannung der Ziliarmuskeln nimmt die Spannung der Zonulafasern zu, weshalb die Linse abgeflacht wird und das Auge auf weit entfernte Dinge fokussieren kann. Die Fähigkeit der Linse, ihre Form zu verändern, nimmt mit zunehmenden Alter ab, da die Linse dann zur Verhärtung neigt. Eine Kombination aus linsen- und nichtlinsenbedingten Faktoren, wie z. B. Linsenvergrößerung, Alterung der Ziliarmuskeln oder auch Veränderungen an den Ansatzpunkten der Zonulafasern, kann zu einem Zustand führen, der als „Presbyopie“ bezeichnet wird und den Verlust der Akkommodationsfähigkeit bedeutet.

Die Linse besteht aus mehreren Schichten langer Linsenfasern, die vom Rand ausgehen und zu den Linsenpolen hin gestreckt sind. Die in der Linse befindlichen kristallinen Schichten wachsen im Laufe des Lebens kontinuierlich weiter, wodurch die Linse altert. Die sich entwickelnden Nahtmuster tragen zur Entstehung einer kristallinen Linse bei, deren Brechungsindex einen gradientenförmigen Verlauf aufweist. Bei einfacheren Linsen werden die Gradienten des Brechungsindex mit Hilfe von sphärischen Fischaugenlinsen und durchschnittlichen optischen Linsen bestimmt. Der genaue Gradientenverlauf des menschlichen Auges ist unbekannt. Der Index hat jedoch in der Mitte mit 1,415 seinen Höchstwert und sinkt zunächst langsam und dann in der Nähe der Oberfläche schnell auf einen Wert von etwa 1,37. (Atchison & Thibos, 2016). Es wurde eine allmähliche Abnahme der Fähigkeit von Linsen beobachtet, unterschiedliche Objektabstände zu berücksichtigen. Beim Menschen nimmt die Fähigkeit zur **Akkommodation** ab und geht mit etwa 60 Lebensjahren verloren.

### Weitere Bauteile des Auges

Die undurchsichtige, faserige und schützende äußere Schicht des Auges ist als Sklera bekannt. Sie steht in ständiger und direkter Verbindung mit der Hornhaut auf der Vorderseite und der Hülle, die den Sehnerv umgibt, auf der Rückseite. Das lichtempfindliche Nervengewebe des Auges wird als Netzhaut bezeichnet. Sie wandelt die Bilder aus dem optischen System des Auges in elektrische Impulse um und leitet sie über den Sehnerv an das Gehirn weiter, wo sie in interpretiert, d. h. gesehen werden. Es handelt sich um eine dünne membranartige Innenschicht, die aus zwei Zellarten besteht: Zapfen und Stäbchen. Nur im Bereich des blinden Flecks sind keine Zapfen und Stäbchen vorhanden, weil dort der Sehnerv das Auge verlässt. Zapfen sind lichtempfindliche Rezeptorzellen der Netzhaut, die das Sehen bei Tageslicht und die Farbdifferenzierung ermöglichen. Stäbchen sind spezialisierte Rezeptorzellen der Netzhaut, die lichtempfindlich sind und für das Sehen bei Nacht zuständig sind, aber keine Farben erkennen. Außerdem sind die Stäbchen wichtig für das periphere Sehen und für die Erzeugung von Kontrast.

### Fragen zur Selbstkontrolle

1. Bitte wählen Sie die richtige(n) Aussage(n) aus.
* *Die Akkommodation des menschlichen visuellen Systems verändert sich im Laufe der Jahre.*
* Das menschliche Auge kann den Fokus zwischen nahen und fernen Objekten wechseln.
* Die Stäbchen sind notwendig, um Farbinformationen zu erhalten.