

Internationaler Studiengang Medieninformatik (IMI) – Master

**Masterarbeit**

Computergestützte Identifizierung von Pflanzen anhand ihrer Blattmerkmale

Wintersemester 2016/2017

Student: Christoph Franke

549642

Betreuer: Prof. Dr. Klaus Jung

Nico Hezel

**Inhaltsverzeichnis**

[1. Einleitung 1](#_Toc475891484)

[2. Grundlagen 1](#_Toc475891485)

[2.1. Morphologie von Pflanzen 1](#_Toc475891486)

[2.1.1. Bestimmungsmerkmale höherer Pflanzen 1](#_Toc475891487)

[2.1.2. Morphologie des Blattes 2](#_Toc475891488)

[2.1.2.1. Funktion und Aufbau des Blattes 2](#_Toc475891489)

[2.1.2.2. Blattformen 3](#_Toc475891490)

[2.1.2.3. Blattnervatur 4](#_Toc475891491)

[2.2. Digitale Bildverarbeitung 5](#_Toc475891492)

[2.2.1. Diskrete Bildrepräsentation 5](#_Toc475891493)

[2.2.2. Binarisierung 6](#_Toc475891494)

[2.2.3. Filter 8](#_Toc475891495)

[2.2.3.1. Allgemeine Definition 8](#_Toc475891496)

[2.2.3.2. Lineare Filter 9](#_Toc475891497)

[2.2.3.3. Kanten, Konturen, POI 11](#_Toc475891498)

[2.2.3.4. Morphologische Filter 11](#_Toc475891499)

[2.3. Maschinelles Lernen 13](#_Toc475891500)

[2.3.1. maschinelle Lernverfahren 13](#_Toc475891501)

[2.3.2. Klassifizierung 13](#_Toc475891502)

[2.3.3. Support-Vektor-Maschinen 16](#_Toc475891503)

[2.3.4. Künstliche Neuronale Netze 16](#_Toc475891504)

[2.3.4.1. Grundstruktur eines neuronalen Netzes 16](#_Toc475891505)

[2.3.4.2. Multilayer Perceptron und Backpropagation 18](#_Toc475891506)

[2.3.5. Features zur Beschreibung von Bildregionen 20](#_Toc475891507)

[2.3.5.1. Inarianz 20](#_Toc475891508)

[2.3.5.2. Geometrische Merkmale und Momente 21](#_Toc475891509)

[2.3.5.3. SIFT / SURF 21](#_Toc475891510)

[2.3.6. k-Means Clustering 21](#_Toc475891511)

[3. Konzeption 22](#_Toc475891512)

[3.1. Zielstellung 22](#_Toc475891513)

[3.2. Testumgebung 22](#_Toc475891514)

[3.3. Vorhandene Lösungen (oder Einordnung in Kontext oder so) 23](#_Toc475891515)

[3.3.1. Überblick 23](#_Toc475891516)

[3.3.2. Form 24](#_Toc475891517)

[3.3.3. Blatttextur 24](#_Toc475891518)

[3.3.4. Fourier-Transformation 24](#_Toc475891519)

[3.4. Lösungsansatz 24](#_Toc475891520)

[3.4.1. Datensatz 24](#_Toc475891521)

[3.4.2. Segmentierung der Blätter 26](#_Toc475891522)

[3.5. Wahl und Extraktion der Features 27](#_Toc475891523)

[3.5.1. Blattform 27](#_Toc475891524)

[3.5.1.1. Momente 28](#_Toc475891525)

[3.5.1.2. Shape Context 28](#_Toc475891526)

[3.5.1.3. Blattadern 28](#_Toc475891527)

[3.5.2. Blattadern 28](#_Toc475891528)

[3.5.2.1. Local binary pattern 28](#_Toc475891529)

[3.5.3. Blattadern 28](#_Toc475891530)

[3.5.4. Lernverfahren 29](#_Toc475891531)

[3.5.5. Wahl der Modelle 30](#_Toc475891532)

[3.5.6. Metriken zur Messung der Erkennungsgenauigkeit 30](#_Toc475891533)

[3.6. Systementwurf 30](#_Toc475891534)

[3.6.1. Systemkomponenten 30](#_Toc475891535)

[3.6.2. Datenbank? 30](#_Toc475891536)

[4. Implementierung 31](#_Toc475891537)

[4.1. Verwendete Bibliotheken 31](#_Toc475891538)

[4.2. allgemein 32](#_Toc475891539)

[4.3. Segmentierung 32](#_Toc475891540)

[4.4. Inner Distance Shape Context 34](#_Toc475891541)

[4.5. Multilevel Inner Distance Shape Context 36](#_Toc475891542)

[4.6. Codebooks 37](#_Toc475891543)

[4.6.1. Sparse dictionary learning 37](#_Toc475891544)

[4.7. Multilayer Perceptron 38](#_Toc475891545)

[4.8. Support Vector Machine 38](#_Toc475891546)

[4.9. Datenhaltung 38](#_Toc475891547)

[4.10. Evaluation 38](#_Toc475891548)

[5. Ergebnis 38](#_Toc475891549)

[5.1. Genauigkeit (welche Metrik?) 38](#_Toc475891550)

[5.2. Aufwandsanalyse? 39](#_Toc475891551)

[6. Zusammenfassung 39](#_Toc475891552)

[6.1. Ausblick 39](#_Toc475891553)

[I. Abbildungsverzeichnis 39](#_Toc475891554)

[II. Tabellenverzeichnis 41](#_Toc475891555)

[III. Codelistings 41](#_Toc475891556)

[a. Installationsanleitung 45](#_Toc475891557)

# Einleitung

Erkennung von Pflanzen herausforderndes Problem der computer vision, unregelmäßige Formen, hochvariable Texturen(?)

# Grundlagen

## Morphologie von Pflanzen

### Bestimmungsmerkmale höherer Pflanzen

Die als Kormophyten bezeichneten höheren Pflanzen weisen einen Kormus genannten Pflanzenkörper auf, der sich aus den Grundorganen Laubblatt, Wurzel und Sprossachse zusammensetzt. Die Gestalt der Organe steht in enger Wechselbeziehung mit ihrer Funktion. (Welle, 2014, S. 16)

Abhängig von den Lebensumständen des Organismus, variiert die Beschaffenheit der Organe zwischen den Arten. Anhand des Erscheinungsbildes ihrer Organe, können Pflanzen bestimmt und zugeordnet werden. Mit dieser Thematik beschäftigt sich die Morphologie, die Lehre von der äußeren Gestalt der Pflanzen. (Welle, 2014, S. 13)

Die Wurzel hat die Befestigung der Pflanze im Boden und die Aufnahme von Wasser und Nährsalzen zur Aufgabe. Die Gestalt des Wurzelsystems wird durch Standortfaktoren wie die Bodenbeschaffenheit, bestimmt (Welle, 2014, S. 18). Wurzeln können neben den ursprünglichen Aufgaben weitere Funktionen übernehmen, wie die Speicherung von Nährstoffen als Rüben oder Wurzelknollen. Die Anpassung an die zusätzliche Aufgabenstellung hat eine Veränderung der Grundform, das heißt eine Metamorphose zur Folge (Welle, 2014, S. 42).

Hauptaufgabe der Sprossachse ist die Laubblätter zu tragen und sie mit der Wurzel zu verbinden, um Wasser und Nährsalze zwischen diesen beiden Organen zu transportieren (Welle, 2014, S. 22). Wichtige Unterscheidungsmerkmale von Sprossachsen sind ihre Stellung, ihre Beschaffenheit und die Form ihres Achsenquerschnitts (Rothmaler, 1967, S. X). Metamorphosen der Sprossachse können, unter anderen, Sprossdornen, Knollen und Zwiebeln sein (Welle, 2014, S. 48, 50).

Bei Samenpflanzen können zusätzlich Früchte und Samen als Bestimmungsmerkmal dienen. Diese bilden sich nach der Befruchtung aus dem Fruchtknoten mit der eingeschlossenen Samenanlage. (Welle, 2014, S. 140)

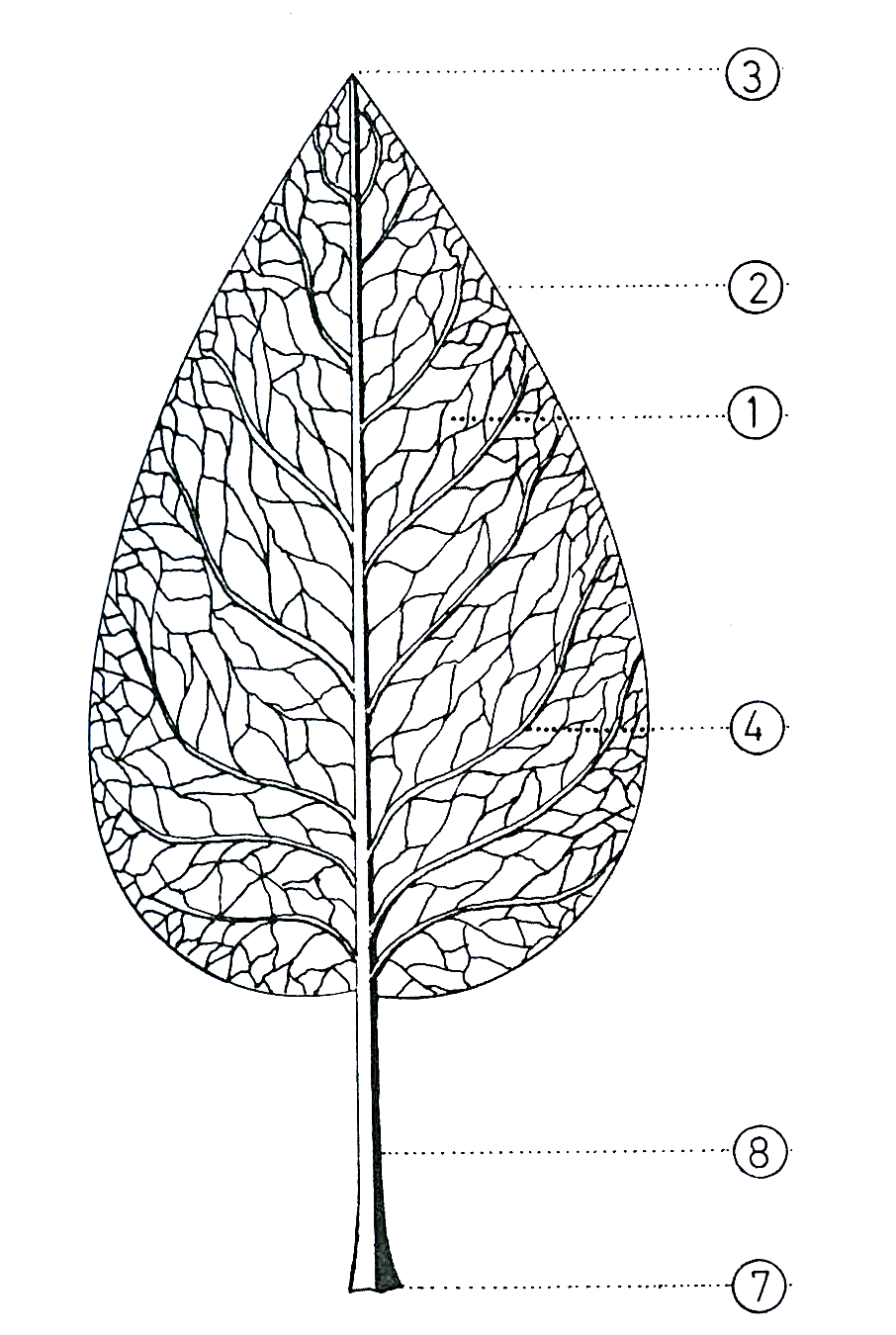
### Morphologie des Blattes

#### Funktion und Aufbau des Blattes

Laubblätter haben in ihrer Grundform die Funktionen des Gasaustauschs, der Transpiration und der Photosynthese inne. Die Photosynthese dient dem Aufbau von energiereichen Kohlenhydraten aus dem energieärmeren Kohlendioxid und Wasser unter dem Einfluss des Sonnenlichtes und der Anwesenheit von Chlorophyll. Die Konzentration von Letzterem bestimmt die Intensität der Grünfärbung der Blätter. (Welle, 2014, S. 26)

In Anpassung an bestimmte Gegebenheiten können Blätter als Metamorphosen die Gestalt von Blattranken als Halte- und Kletterorgane, von Blattdornen zum Schutz vor Tierfraß und von Fangblättern bei fleischfressenden Pflanzen annehmen. (Welle, 2014, S. 54)

Die sich aus metaphorisierten Laubblättern zusammensetzenden Blüten sind ein weiteres wichtiges Unterscheidungsmerkmal von sich geschlechtlich fortpflanzenden Samenpflanzen. Die Blütenhülle wird aus den umschließenden schützenden Kelchblättern und den gefärbten Blumen- oder Kronblättern gebildet. Die Staubblätter stellen das männliche und die Fruchtblätter das weibliche Geschlechtsorgan der Pflanze dar. (Welle, 2014, S. 118)



6

5

4

3

2

1

Abbildung 1 - Bau des Laubblattes: Blattspreite (1), Blattrand (2), Blattspitze (3), Blattnerv (4), Blattstiel (5), Blattgrund (6) (nach Welle, 2014, S. 27)

In Abbildung 1 ist der grundsätzliche Aufbau eines Blattes dargestellt. Die Fläche des Laubblattes wird als Blattspreite bezeichnet. Das Ende des Blattes bezeichnet man als Blattspitze. Abgegrenzt wird das Blatt vom Blattrand. Die Blattfläche wird von Blattnerven durchzogen. Über den Blattstiel ist das Blatt mit der Sprossachse verbunden. Der Blattgrund bezeichnet den direkt an die Sprossachse angrenzenden Teil des Blattes. (Welle, 2014, S. 26)

Im Folgenden wird betrachtet, anhand welcher Merkmale man Laubblätter morphologisch unterscheiden kann.

#### Blattformen

Die Form der Blattspreite ist sehr variantenreich und laut Welle (2014, S. 32) ein wichtiges Merkmal zur Bestimmung der Pflanzen. Ein Auszug aus den möglichen Formen ist in den Abbildungen 2, 3 und 4 zu sehen.

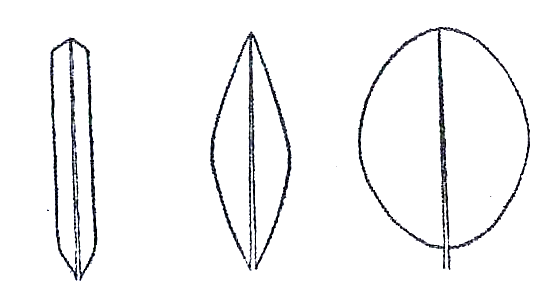
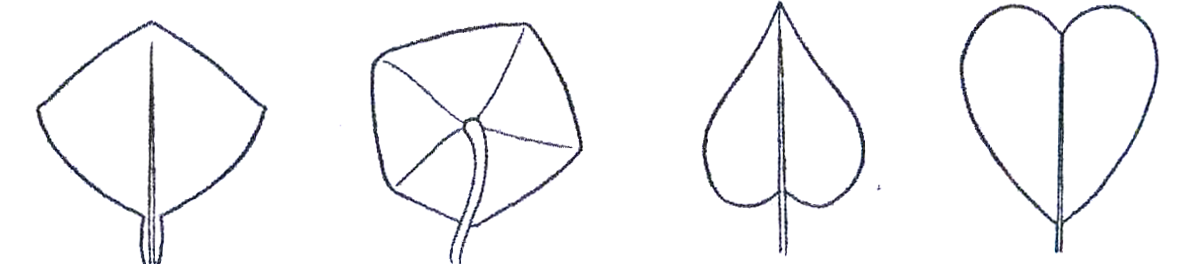
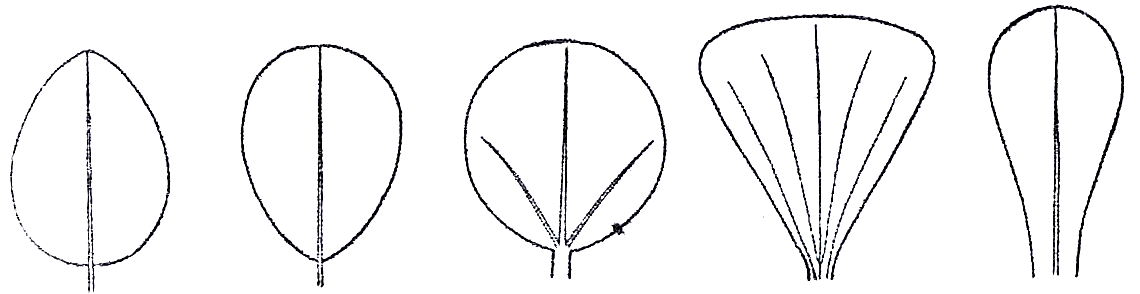
 

Abbildung 2 - Beispiele möglicher Formen einfacher ungeteilter Blätter, von links nach rechts: linealisch, lanzettlich, elliptisch, eiförmig, verkehrt eiförmig, kreisrund, keilig, spatelig, rautenförmig, schildförmig, herzförmig, verkehrt herzförmig (Rothmaler, 1967, S. XV, XVI)

|  |  |
| --- | --- |
| Abbildung 3 - Beispiele geteilter und gespaltener Laubblätter, von links nach rechts: handförmig gespalten, fiederförmig gespalten, handförmig geteilt, fiederteilig (Rothmaler, 1967, S. XIV) | Abbildung 4 - Beispiele zusammengesetzter Blätter, von links nach rechts: 3-zählig, 5-zählig, fußförmig, gefiedert, unpaarig gefiedert (Rothmaler, 1967, S. XIV) |

Grundsätzlich kann die Form unterschieden werden in einfache ungeteilte Blätter, bei denen der Rand nicht oder nur leicht eingeschnitten oder gelappt ist (Abbildung 2), und geteilte oder gespaltene Laubblätter (Abbildung 3), deren Blattspreite durch mehr oder weniger tiefe Einschnitte geteilt sind (Welle, 2014, S. 32, 34). Besteht die Blattspreite aus mehreren selbstständigen Blattteilen, spricht man von zusammengesetzten Blättern (Rothmaler, 1967, S. XIV). Die Anordnung der Blättchen genannten Teile ist eine weitere Möglichkeit der Differenzierung (Abbildung 4).

Zur Variation innerhalb der Blattformen tragen auch die Form des Spreitengrundes und der Spreitenspitze bei, welche in Abbildung 5 zu sehen sind.

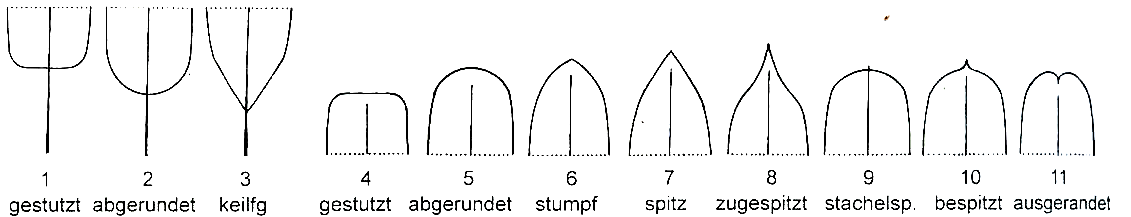


Abbildung 5 - Formen des Spreitengrundes und der Spreitenspitze, Spreitengrund (1 bis 3) von links nach rechts: gestutzt, abgerundet, keilig, | Spreitenspitze (4 bis 11) von links nach rechts: gestutzt, abgerundet, stumpf, spitz, zugespitzt, stachelspitz, bespitzt, ausgerandet (Rothmaler, 2011, S. 886)

Der Blattrand kann als weiteres Unterscheidungsmerkmal dienen. Die Gliederung nach Rothmaler (1967) ist in Abbildung 6 dargestellt.

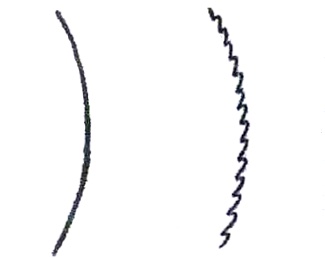


Abbildung 6 - Formen von Blatträndern, von links nach rechts: ganzrandig, gesägt, doppelt gesägt schrotsägeförmig, gezähnt, gekerbt, gebuchtet, geschweift (Rothmaler, 1967, S. XVI, XVII)

Nicht alle Blätter weisen, abhängig von der Art der Anheftung an die Sprossachse, einen Stiel auf. Bei gestielten Blättern ist der Blattstiel deutlich ausgebildet, während stengelumfassende, durchwachsene oder sitzende Blätter keinen Stiel ausbilden. (Rothmaler, 1967, S. XII)

Weiter kann man Blätter anhand ihrer Anordnung an der Sprossachse in wechselständige, gegenständige, quirlständige und bodenständige Blätter differenzieren (Rothmaler, 1967, S. XIII). Ferner dienen der Unterscheidung die Beschaffenheit des Blattes, die sich in ledrig, krautig, fleischig und häutig unterteilen lässt (Rothmaler, 1967, S. XVII) und der Spreitenquerschnitt, der die Art der Rollung oder Faltung des Blattes beschreibt (Rothmaler, 2011, S. 887).

#### Blattnervatur

Die Blattnerven, oder synonym Blattadern, sind das Blatt durchziehende Leitbündel. Ihre Aufgabe ist der Transport von Stoffen und die Wahrung der Festigkeit der Blattspreite (Welle, 2014, S. 86). Die Anordnung der Nerven kann zwischen den Arten variieren.

Nach Rothmaler (1967, S. XIII) unterscheidet man nach dem Verlauf der Blattnerven Blätter in:

* streifennervig: gleich starke Nerven verlaufen vom Blattgrund ohne Verzweigung nebeneinander zur Blattspitze
* netznervig: von einer oder mehreren Hauptnerven gehen kleinere Seitennerven ab, die in ein engmaschiges Adernetz übergehen. Weiter untergliedert man netznervige Blätter in:
  + fiedernervig: die Seitennerven gehen zu beiden Seiten von einem Hauptnerv ab
  + fingernervig: die größeren Nerven gehen strahlenförmig vom Blattstiel aus

Beispiele von Blättern, unterschieden in der Art des Verlaufs ihrer Blattnerven, sind in Abbildung 7 aufgeführt.

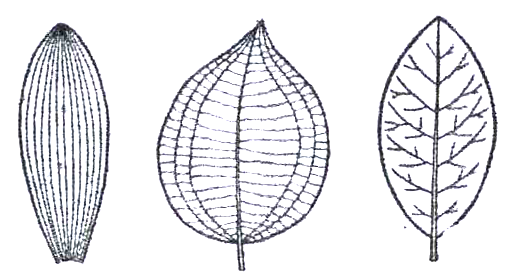
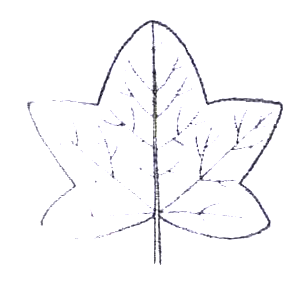
 

Abbildung 7 - mögliche Verläufe der Blattnerven in Laubblättern, von links nach rechts: streifennervig, netzervig, fiedernervig, fingernervig

## Digitale Bildverarbeitung

### Diskrete Bildrepräsentation

Da Computer keine kontinuierlichen analogen Bilder verarbeiten können, werden diese im einfachsten Fall in ein zweidimensionales Gitter überführt. Die meist mit der Matrixnotation angegebene Position innerhalb des Gitters deckt eine rechteckige Region des Bildes ab, die von einem Pixel repräsentiert wird. Der Wert des Pixels gibt die mittlere Intensität der Bestrahlung in dieser Bildregion wieder. (Jähne, 2012, S. 110) Ist das verständlich?

Die Auflösung des diskreten Bildes bestimmt die Anzahl der Pixel und somit den Detailgrad der Darstellung. Laut Jähne (2012, S. 111) „sollte die Pixelgröße kleiner sein als die kleinsten Objekte, die untersucht werden sollen“. Dagegen steht, dass der Aufwand der Bildverarbeitung mit steigender Pixelzahl zunimmt. (Jähne, 2012, S. 111).

Ein weiterer Aspekt der Auflösung ist die Tiefe. Diese bezieht sich auf die Anzahl der Bits, die einem Pixel zur Verfügung stehen, um den Intensitätswert zu speichern. Bei 8-Bit Tiefe lassen sich beispielsweise 28 = 256 verschiedene Intensitäts- oder Farbwerte abbilden. (Burdick, 1997, S. 17)

Mit einem Wert je Pixel können lediglich Graustufen dargestellt werden. Für die Repräsentation von Farben werden jedoch mehr Dimensionen benötigt. Bei den Farbräumen handelt es sich um dreidimensionale orthogonale Räume, deren Achsen senkrecht zueinander stehen. Im Falle des RGB-Farbraums bilden die Achsen die Intensitäten der Farben Rot, Grün und Blau. Der RGB-Farbraum ist additiv, das heißt alle Farben werden durch Addition der einzelnen Farbwerte abgeleitet, beginnend bei Schwarz, wenn kein Farbwert gesetzt ist. In der Bildverarbeitung werden häufig RGB-Bilder verwendet, da Computermonitore ihre Darstellung nativ, also direkt unterstützen. (Burdick, 1997, S. 22)

|  |  |
| --- | --- |
| Abbildung 8 - RGB-Farbraum (Burdick, 1997, S. 23) | C:\Users\chris\AppData\Local\Microsoft\Windows\INetCache\Content.Word\1280px-HSV_color_solid_cylinder_alpha_lowgamma.png  Abbildung 9 - HSV-Farbraum als Zylinder (Wikipedia, 2017) |

Das RGB-System entspricht allerdings nicht der menschlichen Farbwahrnehmung. Eine natürlichere Beschreibung von Farben ergibt sich durch ihre Unterteilung in Farbton (Hue), Sättigung (Saturation) und Intensität (Intensity) im sogenannten HSI-Raum, der ein Polarkoordinatensystem verwendet. (Jähne, 2012, S. 54)

Es gibt diverse Methoden, um Bilder vom RGB-Raum in den HSI-Raum und zurück zu transformieren. Eine spezielle Implementierung ist das HSV-Farbmodell, dessen Farbraum in Abbildung 9 zylindrisch dargestellt ist. Das Modell generiert den Farbton zwischen 0° und 360°, beginnend mit Rot. Die Sättigung liegt im Wertebereich von 0 bis 1, wobei 0 die Abwesenheit von Farbe codiert, also Grau, in Abbildung 9 entlang der zentralen Achse zu sehen. Der Value ist eine Variation der Intensität und liegt zwischen 0 (schwarz) und 1 (weiß). (Burdick, 1997, S. 29)

### Binarisierung

Binärbilder sind Bilder mit nur einem Bit je Pixel. Jeder Pixel kann folglich nur einen von zwei Werten annehmen, schwarz oder weiß. In der Bildverarbeitung werden die beiden Werte auch häufig als Vorder- beziehungsweise Hintergrund bezeichnet. (Burger & Burge, 2015, S. 223)

Die Trennung von Vorder- und Hintergrund bezeichnet man als Binarisierung, oder auch Segmentierung, und kann zur Freistellung von Bildobjekten dienen (Jähne, 2012, S. 541). Anwendung finden die resultierenden Binärbilder bei morphologischen Filtern (Abschnitt 2.2.3.4), zur Extraktion von Konturen im Bild (Abschnitt??) und zur Bestimmung momentbasierter geometrischer Eigenschaften von segmentierten Objekten (Abschnitt **Fehler! Verweisquelle konnte nicht gefunden werden.**).

Um ein Bild, meist ein Graustufenbild, zu binarisieren, wendet man Schwellwertoperationen an. Die Pixel werden, anhand ihrer Intensität *α*, in einer Punktoperation in zwei Klassen unterteilt. Die Klasse dient der Zuweisung von einem von zwei festen Werten, α0 und α1. Das heißt, dass Pixel mit einer Intensität unterhalb eines gewählten Schwellwertes *q* den Wert von *α0* erhalten, darüber den Wert *α1* (Burger & Burge, 2015, S. 61):

(1)

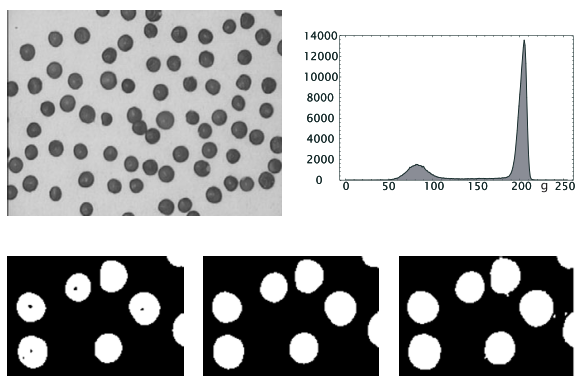


Abbildung 10 - Graustufenbild (links oben) mit seinem Histogramm (rechts oben), Binarisierung mit globalen Schwellwerten von 110, 147 und 185 (unten, von links nach rechts) (Jähne, 2012, S. 543)

Die Ermittlung eines geeigneten Schwellwertes ist die Aufgabe von automatischen Schwellwertoperationen. Viele Verfahren zur Schwellwertberechnung werten dazu die Histogramme des Ausgangsbildes aus. Histogramme sind Darstellungen von Häufigkeitsverteilungen bestimmter Merkmale. Im Falle von Bildern zeigen Histogramme die Häufigkeit der auftretenden Intensitätswerte an (Burger & Burge, 2015, S. 42).

Bei Verfahren zur automatischen Schwellwertbestimmung wird zwischen globalen und adaptiven Methoden unterschieden. In Abbildung 10 ist ein 8-Bit Graustufenbild (links oben) mit zugehörigem Histogramm (rechts oben) und die aus globalen Schwellwertoperationen resultierenden Binärbilder (unten) zu sehen. Globale Schwellwertmethoden bestimmen für ein Graustufenbild einen einzigen optimalen Schwellwert, anhand dessen die Pixel dem Vorder- oder Hintergrund zugeordnet werden (Burger & Burge, 2015, S. 268). Globale Schwellwerte können bei ungleichmäßiger Beleuchtung im Bild unerwünschte Ergebnisse erzielen. Bei adaptiven Methoden wird daher für jeden einzelnen Bildpunkt ein angepasster Schwellwert bestimmt (Burger & Burge, 2015, S. 291).

Otsu ist ein Beispiel für eine globale histogrammbasierte Schwellwertbestimmung. Das Verfahren versucht die Varianz, also die Streuung der Grauwerte, innerhalb der beiden zu bildenden Klassen zu minimieren. Die Varianz kann direkt aus dem Histogramm des Bildes abgeleitet werden. Gleichzeitig wird sichergestellt, dass die Trennung zwischen den Klassen möglichst groß ist. Dazu müssen ihre Mittelwerte möglichst weit auseinander liegen. (Burger & Burge, 2015, S. 275)

Laut Burger & Burge (2015, S. 278) arbeitet die Otsu-Methode *„sehr effizient“* und liefert in der Praxis *„recht brauchbare Ergebnisse“*.

### Filter

#### Allgemeine Definition

Filter, auch Nachbarschaftsoperationen genannt, betrachten, im Gegensatz zu Punktoperationen, nicht nur den einzelnen Pixel, sondern auch eine kleine Region um ihn herum, die Nachbarschaft. Sie kombinieren die benachbarten Pixel in einer bestimmten Weise und liefern so ein geändertes Bild. Wichtige Aufgaben, die Filter übernehmen können, sind beispielsweise die Unterdrückung von Rauschen, die Korrektur von Störungen oder die Detektion und Unterscheidung lokaler Strukturen im Bild, wie Kanten oder Linien. (Jähne, 2012, S. 296, 297)

Die Größe der Filterregion, also die Größe der Nachbarschaft, ist ein wichtiger Parameter des Filters. Sie bestimmt, wie viele umgebende Pixel zur Berechnung des neuen Wertes für den betrachteten Pixel herangezogen werden. Mit unterschiedlichen Größen erzielt man unter Umständen auch unterschiedlich starke Effekte. (Burger & Burge, 2015, S. 94)

Die Anwendung des Filters erfolgt für jeden Bildpunkt, beispielsweise durch zeilenweises Ablaufen des Bildes, wie in Abbildung 11 zu sehen ist. Die grauen Bereiche in der Abbildung stellen bereits betrachtete und geänderte Pixel dar. Wichtig ist, dass für die Betrachtung der Umgebung nur die ursprünglichen Werte des Bildes *I* herangezogen werden. Die Änderungen werden deshalb separat in einem Ergebnisbild *I‘* gespeichert werden. (Jähne, 2012, S. 303)

|  |  |
| --- | --- |
| Abbildung 11 - Anwendung eines Filters durch zeilenweises Verschieben der Filtermaske (Jähne, 2012, S. 303) | Abbildung 12 - Koordinatensystem der Filtermatrix *H* mit „Hot Spot“ genanntem Ursprung in der Mitte (Burger & Burge, 2015, S. 96) |

#### Lineare Filter

Lineare Filter verknüpfen die Werte der Pixel in der Umgebung in linearer Form durch eine gewichtete Summation. Die Filterfunktion wird durch eine Matrix abgebildet, in der die Gewichtungen der Pixelwerte der Umgebung gespeichert sind. Die Dimension der Matrix gibt die Größe der betrachteten Umgebung an. (Burger & Burge, 2015, S. 95)

Die Filtermatrix besitzt ein eigenes Koordinatensystem, dessen Ursprung üblicherweise in ihrer Mitte liegt (Abbildung 12). Bei Anwendung der Filtermatrix *H*, wird sie mit ihrem Ursprung über dem aktuell betrachteten Pixel *I(u, v)* positioniert. Darauf folgt eine Multiplikation aller gewichteten Elemente *H(i, j)* innerhalb der Matrixregion *R* mit den darunter liegenden Bildpunkten *I(u+i, v+j)* und eine anschließende Summation aller Produkte. Das Ergebnis wird im Ergebnisbild *I‘(u, v)* gespeichert, an derselben Position wie im Ausgangsbild *I(u, v)*. (Burger & Burge, 2015, S. 96)

Auf alle Bildpunkte angewendet, lässt sich die Operation wie folgt formulieren:

|  |  |
| --- | --- |
|  | (2) |

Die durchgeführte Operation entspricht einer Faltung *I‘ = I \* H* mit dem Faltungsoperator (\*). In der Mathematik beschreibt die Faltung einen Operator, der zwei Funktionen gleicher Dimension miteinander verknüpft. (Burger & Burge, 2015, S. 105)

Wichtige mathematische Eigenschaften von linearen Faltungen sind die Kommutativität *I \* H = H \* I,* also die Vertauschbarkeit der Argumente, die Linearität und die Assoziativität *(I \* H1) \* H2 = I \* (H1 \* H2)*. Auf Filter bezogen bedeutet die Assoziativität, dass man das gleiche Ergebnisbild erhält, auch wenn die Reihenfolge der Anwendung verschiedener Filter verändert wird, oder sie zu durch paarweise Faltung zu neuen Filtern zusammengefasst werden. (Burger & Burge, 2015, S. 107)

Aus der Assoziativität folgt auch die Separierbarkeit eines einzelnen Filters in kleinere Filter, die nacheinander ausgeführt werden können. Voraussetzung ist, dass sich der zweidimensionale Filter als Produkt von zwei eindimensionalen Filtern darstellen lässt. Die Zerlegung ist dahingehend relevant, dass sich der Berechnungsaufwand mit aus einer großen Filtermatrix separierten einzelnen Filtern reduzieren lässt. (Burger & Burge, 2015, S. 108)

Ein Beispiel für einen solchen separierbaren Filter ist der Gaußfilter *Gσ(x, y)*, der in zwei einzelne Funktionen zerlegt werden kann (Burger & Burge, 2015, S. 109):

|  |  |
| --- | --- |
|  | (3) |

Die Gaußfunktion ist glockenförmig, dargestellt in Abbildung 13 links. Die Standardabweichung σ bestimmt den Radius und somit auch die Größe der Filtermatrix. Der Ursprung der Matrix erhält das maximale Gewicht, die Größe der übrigen Koeffizienten nimmt zu den Rändern hin kontinuierlich ab (Abbildung 13 rechts).

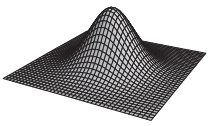
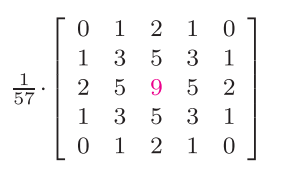
 

Abbildung 13 - Beispiel für einen Gaußfilter, Darstellung der Filterfunktion *H* als 3D-Plot (links) und als Filtermatrix (rechts) (Burger & Burge, 2015, S. 103)

Der Gaußfilter ist ein sogenannter Glättungsfilter, oder auch Tiefpassfilter. In der Bildverarbeitung werden durch Glättung die Intensitätsunterschiede benachbarter Bildelemente verringert. Die Unterdrückung hoher Signale führt zu einem Verwischen von Kanten, der Filter ist somit nicht kantenerhaltend. In der menschlichen Wahrnehmung wirken Bilder dadurch unschärfer. (Burger & Burge, 2015, S. 104)

zu Canny überleiten? Gaborfilter?

#### Kanten, Konturen, POI

#### Morphologische Filter

Mit morphologischen Filtern können gezielt Strukturen von Bildern verändert werden (Burger & Burge, 2015, S. 191). Mit ihnen lassen sich die Formen der durch Binarisierung vom Hintergrund getrennten Objekte ändern, indem diesen Bildpunkte hinzugefügt oder aus ihnen gelöscht werden (Jähne, 2012, S. 556).

Das Verhalten eines morphologischen Filters wird durch eine Matrix bestimmt, ähnlich der Filtermatrix bei linearen Filtern. Die Matrix wird hier als Strukturelement *H* bezeichnet (Abbildung 14, Mitte). (Burger & Burge, 2015, S. 194)

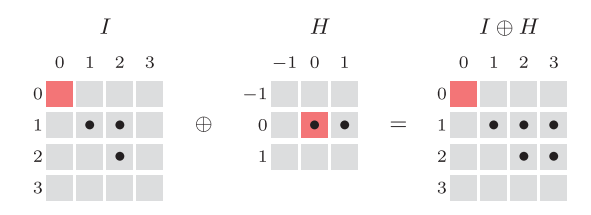


Abbildung 14 - Beispiel einer Dilation auf Binärbild *I* mit Strukturelement *H* (Mitte) und dem Ergebnisbild *I* ⊕ *H* (rechts), Elemente mit dem Wert 1 sind mit • markiert, die Ursprünge der jeweiligen Koordinatensysteme sind rot markiert. (Burger & Burge, 2015, S. 196)

Die grundlegenden Operationen der morphologischen Filter sind Dilation („Wachsen“) und Erosion („Schrumpfen“) (Burger & Burge, 2015, S. 193).

In der Mengennotation ausgedrückt, entspricht die Dilation der Summe aller Kombinationen von Koordinatenpaaren aus den Mengen der Punkte des Binärbildes *QI*und der Punkte des Strukturelementes *QH* (Burger & Burge, 2015, S. 196):

|  |  |
| --- | --- |
|  | (4) |

Laut Burger & Burge (2015, S. 196) kann man *„die Operation auch so interpretieren, dass das Strukturelement H an jedem Vordergrundpunkt des Bilds I repliziert wird“*. Eine Dilation auf einem Binärbild *I* mit dem Strukturelement *H* ist beispielhaft in Abbildung 14 illustriert.

Die Erosion ist in Mengennotation definiert als (Burger & Burge, 2015, S. 196):

|  |  |
| --- | --- |
|  | (5) |

Nach Burger & Burge (2015, S. 196) lässt sich die Erosion folgendermaßen interpretieren: *„Eine Position p ist im Ergebnis I* ⊖ *H dann (und nur dann) enthalten, wenn das Strukturelement H – mit seinem Ursprung positioniert an der Position p – vollständig im ursprünglichen Bild eingebettet ist, d. h., wo sich für jedes Element in H auch ein entsprechendes Element in I findet“*.

Morphologische Filter teilen viele, aber nicht alle Eigenschaften mit den linearen Filtern. So sind morphologische Operationen, im Allgemeinen nicht kommutativ. Im Speziellen ist die Erosion, im Gegensatz zur Dilation, nicht kommutativ: *I* ⊖ *H ≠ H* ⊖ *I*. Es gelten aber bei aufeinanderfolgender Ausführung von Erosion und Dilation auf das gleiche Bild folgende Regeln: *(I* ⊖ *H1)* ⊖ *H2* *= I* ⊖ *(H1* ⊕ *H2)* und *(I* ⊕ *H1)* ⊕ *H2* *= I* ⊕ *(H1* ⊕ *H2)*. Wie auch lineare Filter, lassen sich morphologische Filter zu neuen Filtern zusammenfassen oder auch in einzelne Filter separieren. (Jähne, 2012, S. 559)

Die elementaren Filter Dilation und Erosion lassen sich zu verschiedenen Operationen kombinieren. Die Operation des Öffnens (Opening) führt eine Erosion zur Entfernung kleiner Strukturen durch, gefolgt von einer Dilation zum Ausgleich der durch die Erosion bewirkten Verkleinerung der Objekte: *I* ∘ *H = (I* ⊖ *H)* ⊕ *H* (Abbildung 15, Mitte). (Jähne, 2012, S. 561)

Bei der Operation des Schließens (Closing) folgt auf eine Dilation zur Schließung kleiner Löcher, eine Erosion, die die Vergrößerung der Objekte, als Folge der Dilation, durch Schrumpfung wieder ausgleicht (Abbildung 15, rechts). (Jähne, 2012, S. 562)

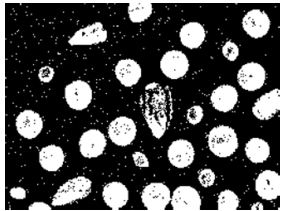
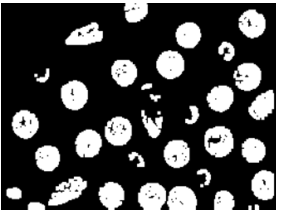
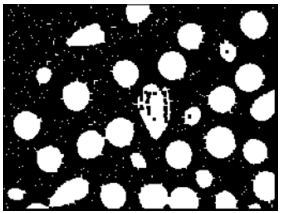
  

Abbildung 15 - Originalbild (links) nach Ausführung von Opening (Mitte) beziehungsweise Closing (rechts) mit Strukturelementen der Dimension 3 x 3 (nach Jähne, 2012, S. 561, 562)

## Maschinelles Lernen

### maschinelle Lernverfahren

Die Entwicklung maschineller Lernverfahren ist ein wichtiger Teilbereich der künstlichen Intelligenz (Ertel, 2016, S. 191). Es handelt sich beim maschinellen Lernen nicht um simples Auswendiglernen, sondern um die Generalisierung gelernter Beispiele, um das Gelernte auf theoretisch unendlich viele neue ungesehene Beispiele anwenden zu können. (Ertel, 2016, S. 192)

Beim maschinellen Lernen unterscheidet man zwischen überwachtem und unüberwachtem Lernen. Bei unüberwachten Lernverfahren sind zu den Trainingsdaten keine Zieldaten verfügbar. Ziel von unüberwachtem Lernen kann unter anderem sein, in den Daten Gruppen von ähnlichen Strukturen mittels Clustering (Kapitel 2.3.6) zu finden oder hochdimensionale Daten auf wenige Dimensionen zu projizieren. Dagegen ist bei überwachtem Lernen zu den zu trainierenden Beispielen als Eingabe die zugehörige Ausgabe bereits bekannt. Zum überwachten Lernen zählen Klassifizierungsprobleme (Kapitel 2.3.2). (Bishop, 2006, S. 3)

### Klassifizierung

Bei der Klassifizierung wird aus bereits klassifizierten Trainingsdaten, eine Funktion generiert, die die Klasse eines neuen Objektes anhand seiner Merkmale berechnen kann (Ertel, 2016, S. 193). Die Aufgabe sowohl des Trainings, als auch der Abbildung eines Merkmalsvektors auf einen Klassenwert, übernimmt ein Klassifizierer, schematisch in Abbildung 16 dargestellt. Die Merkmalsvektoren werden im Folgenden auch Features genannt.

**Merkmal 1**

**Merkmal n**

**Klasse**

Klassifizierer

**…**

**klassifizierte Trainingsdaten**

**(Trainingsphase)**

Abbildung 16 - funktionale Struktur eines Klassifikators (nach Ertel, 2016, S. 194)

Bei der Klassifizierung können Fehlentscheidungen auftreten, wenn ein Objekt der falschen Klasse zugeordnet wird. Die Summen der gewichteten Fehler bezeichnet man als Kosten. Ziel beim Training des Klassifizierers ist es, die aus den Fehlentscheidungen entstehenden Kosten zu minimieren. (Ertel, 2016, S. 20)

Zur Minimierung der Kosten lassen sich Entscheidungsregeln definieren und variieren (Duda, Hart, & Stork, 2001, S. 4). Sie lassen sich veranschaulichen, indem man die verschiedenen Eingangsmerkmale in Relation zueinander setzt. Es spannt sich ein mehrdimensionaler Featureraum auf, der anhand der Entscheidungsregel so partitioniert werden kann, dass ausreichend viele Punkte einer partitionierten Region nur einer einzelnen Klasse zugeordnet werden können.

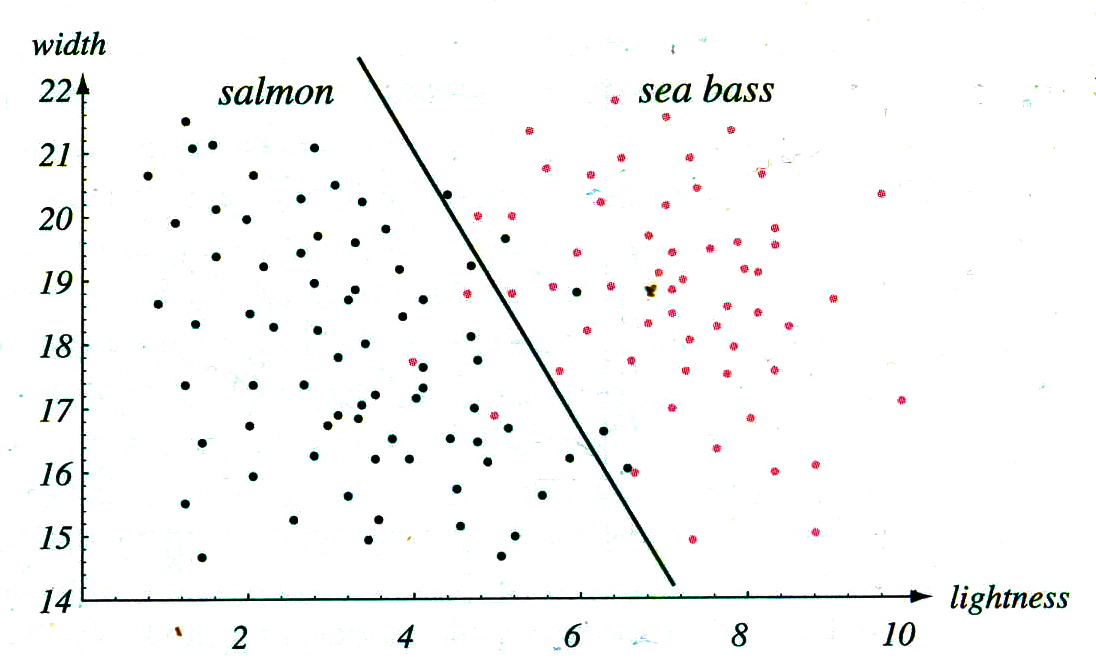
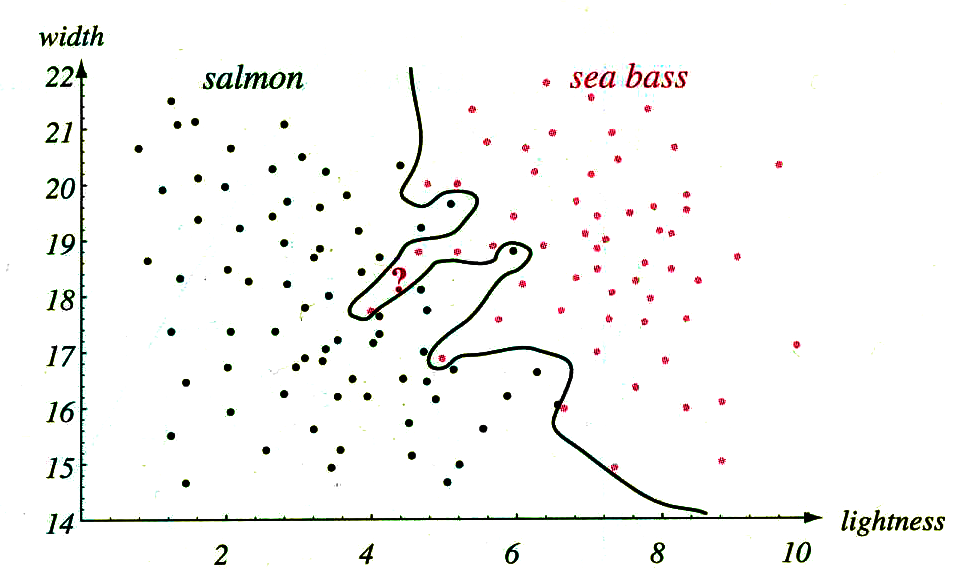
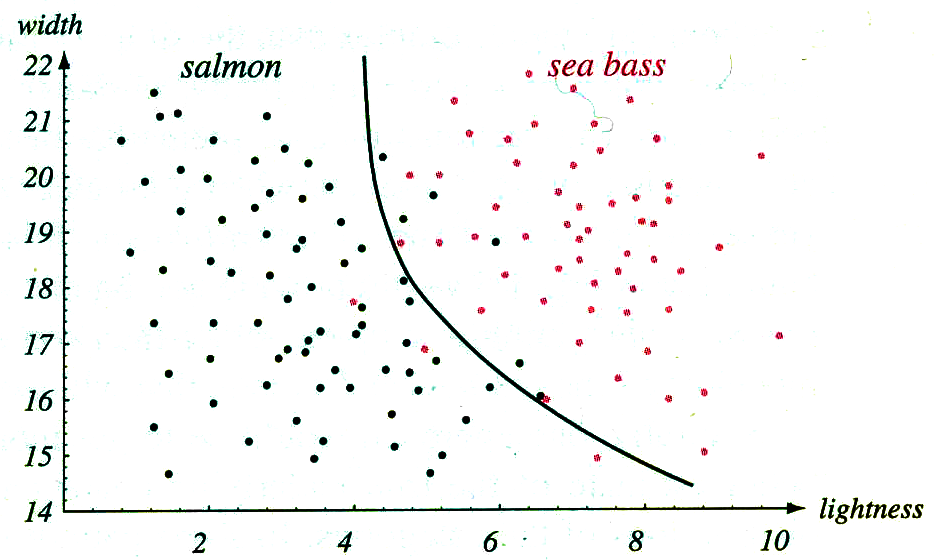
  

Abbildung 17 - verschieden komplexe Entscheidungsgrenzen in einem zweidimensionalen Featureraum zur Unterscheidung zwischen zwei Klassen - Lachs („salmon“, schwarze Punkte) und Wolfsbarsch („sea bass“, rote Punkte) - anhand zweier gegenübergestellter Merkmale - Breite („width“) und Helligkeit („lightness“) (nach Duda, Hart, & Stork, 2001, S. 5, 6)

Ein zweidimensionaler Featureraum ist in Abbildung 17 illustriert. Hier wird beispielhaft zwischen zwei Klassen anhand von zwei Merkmalen unterschieden. Als Entscheidungsregel dient eine Entscheidungsgrenze, die den Featureraum in die zwei zu unterscheiden Klassen teilt. Dargestellt ist auch die Problematik, eine geeignete Komplexität des Modells zu wählen. Die lineare Grenze in Abbildung 17 (links) lässt noch relativ viele Fehler zu, zu beiden Seiten finden sich auch Punkte der jeweils anderen Klasse. Während in der Mitte die Entscheidungsgrenze so gestaltet ist, dass die Grenze die Klassen perfekt trennt, erscheint diese Trennung zu stark an die Trainingsdaten angepasst. Es könnte sich das Problem ergeben, dass die Klassifizierung zwar auf den Trainingsdaten zu 100 Prozent korrekt ist, neu hinzukommende ungesehene Objekte aber falsch klassifiziert werden. Man spricht auch von „Overfitting“ oder Überanpassung. Das widerspricht der gewünschten Generalisierung des Erlernten. Die rechte Entscheidungsgrenze in Abbildung 17 bildet seinen Kompromiss aus den beiden vorangegangenen Grenzen.

Die Wahl eines geeigneten Modells ist nur ein Teil im Entwurf eines Systems, das anhand von Merkmalen Muster erkennen und Entscheidungen zur Klassifizierung treffen soll. Nach Duda, Hart, & Stork (2001, S. 14) lässt sich der Prozess des Entwerfens, in fünf Schritte unterteilen, dargestellt in Abbildung 18.

Start

Daten sammeln

Wahl der Features



Wahl des Modells

Vorwissen

Training des Klassifizierers

Evaluation des Klassifizierers

Ende

Abbildung 18 - Schema des Entwerfens eines Systems zur Mustererkennung und Klassifizierung (nach Duda u. a., 2001, S. 14)

Laut Duda, Hart, & Stork (2001, S. 14) ist die Wahl der Features ein kritischer Entwurfsschritt und hängt von der Charakteristik der zu untersuchenden Problematik ab. Vorwissen über Merkmale, mit denen sich die Klassen möglicherweise gut differenzieren lassen, kann bei der Auswahl helfen und fließt auch in die Wahl des Modells mit ein. Die Evaluation hat schließlich die Aufgabe, die Performanz des Systems zu überprüfen und Möglichkeiten der Verbesserungen des Systems aufzuzeigen, was wiederum Anpassungen der vorherigen Schritte erfordert. (Duda u. a., 2001, S. 14, 15)

Es gibt eine Vielzahl an Modellen, die Entscheidungen im Sinne einer Klassifizierung treffen können. Anhand ihrer Komplexität kann man die zugrundeliegenden in Lazy-Learning (faules Lernen) und Eager-Learning (eifriges Lernen) unterteilen . Beim Lazy-Learning liegt der größte Aufwand nicht in der Lernphase, sondern in der Anwendung auf neue Beispiele. Eager-Learning-Verfahren besitzen deutlich aufwändigere Lernphasen, aber eine effiziente Anwendung. (Ertel, 2016, S. 214)

Beispiele für Lazy-Learning sind die k-Nearest-Neighbour-Methode, lokal gewichtete Regression oder fallbasiertes Lernen. Zu den Modellen mit Eager-Learning zählen Entscheidungsbäume, Bayes-Netze, neuronale Netze und Support-Vektor-Maschinen. (Ertel, 2016, S. 258)

Die beiden letztgenannten werden in den folgenden Kapiteln 2.3.3 beziehungsweise 2.3.4 näher betrachtet

### Support-Vektor-Maschinen

Support-Vektor-Maschinen basieren auf der Idee, dass sich Daten zweier Kategorien, nach einer geeigneten nicht-linearen Abbildung in einen Raum mit ausreichend großer Dimension, immer durch eine Hyperebene trennen lassen. Die Dimension ist dabei typischerweise deutlich höher als die des originalen Featureraumes. (Duda u. a., 2001, S. 259)

Da es meist unendlich viele solcher Ebenen gibt, wird die Hyperebene gesucht, „*die zu beiden Klassen einen möglichst großen minimalen Abstand hat*“ (Ertel, 2016, S. 298). Das bedeutet, dass die optimal trennende Hyperebene so gelegt wird, dass ihr parallel laufender Rand, oder auch Margin, möglichst groß ist, ohne dass ein Punkt innerhalb des Randes liegt (siehe Abbildung 19). Je größer der Rand ist, desto besser ist die Generalisierung des Klassifizierers (Duda u. a., 2001, S. 262). Die Punkte auf dem Rand nennt man Support-Vektoren. Sie parametrisieren die Hyperebene und haben natürlich alle den gleichen Abstand zu ihr (Ertel, 2016, S. 298).

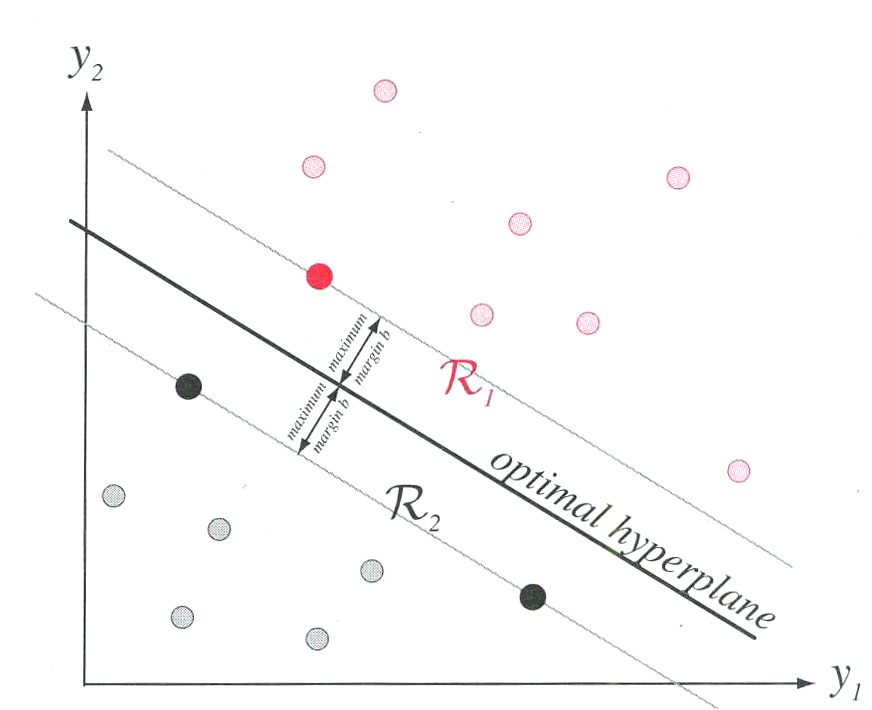


Abbildung 19 - die zwei Klassen R1 und R2 trennende optimale Hyperebene („hyperplane“) mit ihren Support-Vektoren (hervorgehobene Punkte) und dem Rand (zwischen Support-Vektoren und Hyperebene)

### Künstliche Neuronale Netze

#### Grundstruktur eines neuronalen Netzes

Künstliche neuronale Netze haben ihr Vorbild meist in der Natur. Künstliche Neuronen (Abbildung 21) bilden biologische Neuronen (Abbildung 20) ab. Sie sind allerdings nur äußerst simple Abstraktionen, realisiert als Elemente von Soft- oder Hardware. (Keller, Liu, & Fogel, 2016, S. 8)

|  |  |
| --- | --- |
| Synapse  Axon  Zellkörper  Dendriten  Abbildung 20 - schematische Darstellung von zwei biologischen Neuronen (nach Keller, Liu, & Fogel, 2016, S. 8) | Abbildung 21 - schematische Darstellung von zwei künstlichen Neuronen mit Eingabe- und Ausgabesignalen (Keller u. a., 2016, S. 8) |

Die Zellkörper biologischer Neuronen können elektrische Spannung speichern. Die Spannung steigt, je mehr Impulse von anderen Neuronen ankommen. Ist ein bestimmter Schwellwert überschritten, feuert das Neuron und leitet einen Impuls über das Axon an andere verknüpfte Nervenzellen. (Ertel, 2016, S. 266)

Biologisch inspirierte künstliche neuronale Netze bilden diesen Vorgang mit einem Aktivierungspotential in einer diskreten Zeitskala ab. Das Aufladen des Potentials erfolgt durch gewichtete Summation aller eingehenden Verbindungen *x1,…,xn*. Für ein Neuron *i* wird dies wie folgt formuliert (Ertel, 2016, S. 269):

|  |  |
| --- | --- |
|  | (6) |

wobei *wij* die Gewichtung zwischen Neuron *i* und dem am Eingang verknüpften Neuron *j* ist. Die Aktivierungsfunktion *f* wird auf die gewichtete Summe angewandt und bestimmt welches Ergebnis *xi* als Ausgabe an die benachbarten Neuronen weitergeleitet wird. Ein einfaches Beispiel für eine Aktivierungsfunktion, das ist vor allem bei binären Neuronen sinnvoll ist, ist die Schwellwert- oder Stufenfunktion, die in Abbildung 22 links abgebildet ist (Ertel, 2016, S. 269):

|  |  |
| --- | --- |
|  | (7) |

Sigmoid-Funktionen sind die am häufigsten eingesetzten Aktivierungsfunktionen in neuronalen Netzen. Sie halten „eine exzellente Balance zwischen linearem und nicht-linearem Verhalten“ (nach Keller u. a., 2016, S. 29). Ein Beispiel für eine Sigmoid-Funktion, wie in Abbildung 22 in der Mitte zu sehen, ist die logistische Funktion:

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | (8) |
| wobei a die Steigung der Sigmoid-Funktion parametrisiert. | | |  |
| f(x)  x | x  f(x) | x  f(x) | | |

Abbildung 22 - Aktivierungsfunktionen von links nach rechts: Schwellwertfunktion, logistische Funktion, hyberbolischer Tangens (nach Keller u. a., 2016, S. 28, 29, 30)

#### Multilayer Perceptron und Backpropagation

Wie biologische Neuronen lassen sich auch künstliche Neuronen zu komplexen Netzwerken zusammenfügen, die einen hohen Grad der Verknüpfung vorweisen (Keller u. a., 2016, S. 9). Netze mit mindestens einer versteckten Schicht sind in der Lage, nicht-lineare Abbildungen zu erlernen (Ertel, 2016, S. 293).

Dazu zählen die sogenannten Multilayer-Perceptrons, schematisch in Abbildung 23 dargestellt. Ihre Merkmale sind nach Keller u. a. (2016, S. 35):

* Jedes Neuronenmodell im Netzwerk beinhaltet eine nicht-lineare, differenzierbare Funktion
* Das Netzwerk enthält eine oder mehr Schichten, die versteckt sind vor den Eingangs- und Ausgangsmodi
* Das Netzwerk hat einen hohen Grad an Konnektivität, deren Umfang durch die synaptischen Gewichte des Netzwerks bestimmt wird

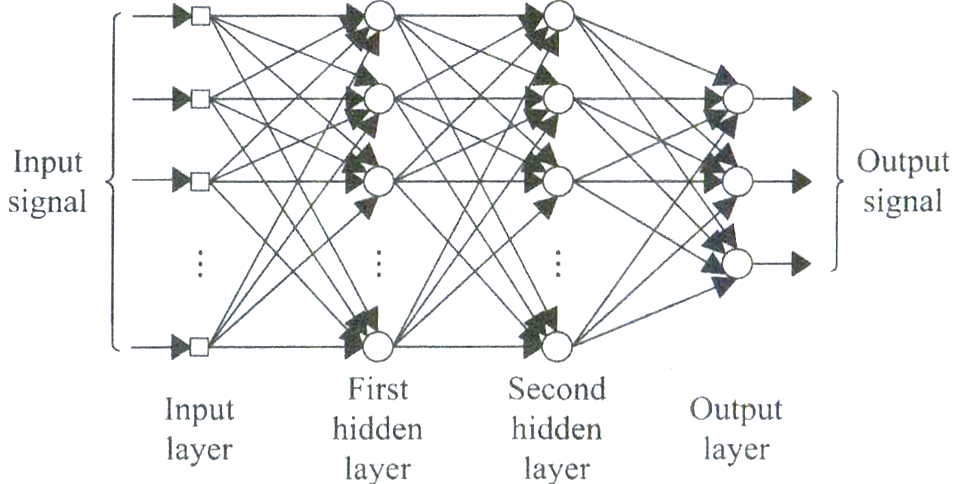


Abbildung 23 -Schema eines Multilayer-Perceptron mit zwei versteckten Schichten („hidden layer“), der Eingabeschicht („input layer“) und der Ausgabeschicht („output layer“). Das Eingangssignal („input signal“) wird vorwärts durch das Netz propagiert, mit dem Ausgangssignal („output signal“) als Ergebnis (Keller u. a., 2016, S. 36)

Üblicherweise erfolgt das Training von Multilayer Perceptrons mit der Backpropagation genannten Lernmethode. Lerneffekte werden während des Trainings durch Änderungen der Gewichtungen *w* zwischen den Neuronen erzielt, in Abhängigkeit von den auftretenden Fehlern. Die Fehlerrate wird anhand einer summierten quadratischen Fehlerfunktion ermittelt (Ertel, 2016, S. 292):

|  |  |
| --- | --- |
|  | (9) |

wobei *tp* die gewünschte Antwort und *xp* die tatsächliche Antwort des Netzes für ein Trainingsmuster *p* darstellt. Ziel ist es den Wert der Fehlerfunktion zu minimieren.

Die Richtung des stärksten Anstiegs der Fehlerfunktion wird mit einem Fehlergradienten angezeigt. Zur Bestimmung eines Minimums wird deshalb der Richtung des negativen Gradienten gefolgt (Ertel, 2016, S. 289). Die Anpassung der Gewichte anhand des negativen Gradienten der Fehlerfunktion *Ep*aus (9) wird folgendermaßen formuliert (Ertel, 2016, S. 292):

|  |  |
| --- | --- |
|  | (9) |

Der Parameter *η* ist die Lernrate zur Festlegung der Stärke der Gewichtsänderung und bestimmt die Geschwindigkeit der Konvergenz Richtung Minimum der Fehlerfunktion (Ertel, 2016, S. 289). ist das Fehlersignal des Neurons *j* und abhängig davon, ob dieses ein Neuron der Ausgabe- oder der versteckten Schicht ist. Für ein verstecktes Neuron *j* errechnet sich aus den Änderungen aller der nächsthöheren Schicht. Bei einem Ausgabeneuron ist proportional zum Fehler des Netzes.

Der generelle Ablauf der Backpropagation beginnt mit einer Initialisierung aller Gewichte mit zufälligen Werten. Darauf folgen die Schritte:

* das Trainingsmuster wird an der Eingabeschicht angelegt
* das Muster wird vorwärts durch das Netz propagiert, das heißt die Aktivierungen werden ab der ersten versteckten Schicht für alle Neuronen berechnet
* der summierte quadratische Fehler *Ep* wird berechnet
* der Fehler wird durch alle Schichten rückwärts propagiert und die Gewichte angepasst

Die Schritte werden für jedes Trainingsmuster wiederholt, bis die Gewichte konvergiert sind oder eine festgelegte diskrete Zeitschranke erreicht ist.

### Features zur Beschreibung von Bildregionen

#### Inarianz

Invarianz!

This approach may be impractical, however, if the number of training examples is limited, or if there are several invariants (because the number of combinations of transformations grows exponentially with the number of such transformations). We therefore seek alternative approaches for encouraging an adaptive model to exhibit the required invariances. These can broadly be divided into four categories:

1. The training set is augmented using replicas of the training patterns, trans- formed according to the desired invariances. For instance, in our digit recog- nition example, we could make multiple copies of each example in which the

digit is shifted to a different position in each image.

2. A regularization term is added to the error function that penalizes changes in the model output when the input is transformed. This leads to the technique of tangent propagation, discussed in Section 5.5.4.

3. Invariance is built into the pre-processing by extracting features that are invari- ant under the required transformations. Any subsequent regression or classi- fication system that uses such features as inputs will necessarily also respect these invariances.

4. The final option is to build the invariance properties into the structure of a neu- ral network (or into the definition of a kernel function in the case of techniques such as the relevance vector machine). One way to achieve this is through the use of local receptive fields and shared weights, as discussed in the context of convolutional neural networks in Section 5.5.6.

(Bishop 262)

#### Geometrische Merkmale und Momente

Regionen eines Binärbildes können mit geometrischen Merkmalen beschrieben werden, dazu zählen:

* der Umfang, errechnet aus den Kontursegmenten einer zusammenhängenden Region
* die Fläche als Summe aller enthaltenen Bildpunkte
* die Kompaktheit als Relation zwischen Fläche und ihrem Umfang

#### SIFT / SURF

### k-Means Clustering

Ziel des Clusterings ist es, in einem multidimensionalen Raum Gruppen oder „Cluster“ von Daten zu identifizieren (Bishop, 2006, S. 424). Cluster sind Anhäufungen von benachbarten Punkten, deren Abstand untereinander typischerweise kleiner ist als der Abstand zwischen Punkten aus unterschiedlichen Clustern (Ertel, 2016, S. 245).

Ein sehr simples Verfahren zur Clusterbildung ist k-Means, bei dem die Anzahl der zu bildenden *k* Cluster von vornherein bekannt sein muss. Die *k* Clusterzentren werden dabei zunächst zufällig oder per Hand initialisiert. Darauf folgen zwei Schritte (Ertel, 2016, S. 246):

* Zuordnung der Datenpunkte zu einem Clusterzentrum
* Neuberechnung der Clusterzentren

Diese beiden Schritte werden wiederholt, bis sich die Position der Clusterzentren nicht mehr verändert. Beispielhaft mit *k = 2* ist der Ablauf in Abbildung 24 skizziert, wobei die Konvergenz nach drei Iterationen erreicht ist.

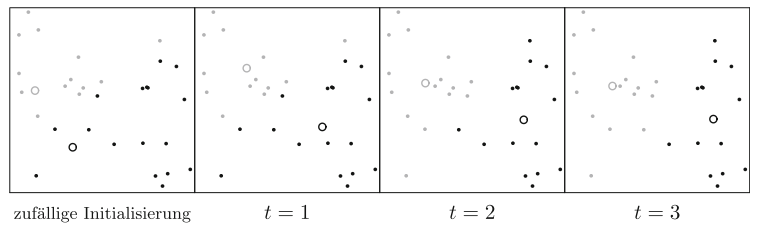


Abbildung 24 - k-Means mit zwei zu bildenden Clustern (hellgrau und schwarz) nach der Initialisierung (links) und nach *t* Iterationen (ab 2. von links). Die Zuordnung der Datenpunkte zu den Clusterzentren (Kreise) ist durch die Graustufe (hellgrau oder schwarz) codiert. (Ertel, 2016, S. 247)

Die Zuordnung der Punkte zum nächstgelegenen Zentrum erfolgt Abstandsfunktionen wie beispielsweise dem eukidlischen Abstand (Ertel, 2016, S. 245). Das jeweilige Clusterzentrum *μ* für die *x1,…,xl* dem Zentrum zugeordneten Punkte wird in jeder Iteration wie folgt neu berechnet :

|  |  |
| --- | --- |
|  | (10) |

# Konzeption

## Zielstellung

System entwerfen, Muster erkennen, die der klassifizierung dienen

abbilden, wie Mensch Blätter unterscheidet -> möglichst viel prior knowledge(Vorwissen) in features packen

keine Aussage über Größe möglich

Konzentration auf Form der Blattspreite, da sehr variabel

Blätter einzeln betrachten, Anordnung fällt somit weg. Aber: zusammengesetzte Blätter auch als solche betrachten (nicht Blättchen einzeln)

keine Metamorphosen

Fotos einzelstehender Blätter

Ausreichend viele Features wählen, Optionen auch bei Modellen, siehe Entwurf des Systems

## Testumgebung

aus Oberfläche heraus Segmentierung testen, Features extrahieren/kombinieren, Training starten

Unittests

## Vorhandene Lösungen (oder Einordnung in Kontext oder so)

### Überblick

Wie in Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.** bereits erwähnt, sind verschiedene Arbeitsschritte notwendig, um die Ausgangsdaten für die vorzubereiten.

Um die Blätter einheitlich beschreiben zu können, müssen die Deskriptoren unabhängig von Skalierung, Rotation und ? beschreiben. Chaki, Parekh und Bhattacharya (2015) schlagen vor das Blatt zuvor zu rotieren.

Datenbanken für Blätter

Einen interessanten Ansatz bietet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Autoren | preprocessing | Deskriptoren | Klassifikatoren | Datensatz | Schluss |
| (Chaki, Parekh, & Bhattacharya, 2015) | - Graustufen, Binarisierung, Kontur  - Blatt rotiert, so dass Hauptachse ausgerichtet zu horizontaler Achse  - Beschneidung auf BB-Größe, Skalierung auf einheitl. Auflösung (7 „slots“ mit max. 300 px) | -Textur: Gabor-Filter, Gray level co-occurrencematrix  -Form: Curvelet-Transform, Hu-Momente  - komibinierter Featurevektor aus beiden | -MLP  -neuro fuzzy classifier | -Flavia |  |
| (Du, Wang, & Zhang, o. J.) | Graustufen, Binarisierung, Kontur, Region of Interest | geometrische features, invariante Features (Hu-Momente) | moving median center hyperspheres (MMC), KNN | ? |  |
| (Han, Chee, Chan, Wilkin, & Remagnino, o. J.) | ? | unüberwachte Features ermittelt durch convolutional neural networks (CNN) | multi-layered deconvolutional network | Royal Botanic Gardens Kew |  |
| (Kalyoncu & Toygar, 2015) | Segmentierung, noise reduction, Kontur, Glättung and corner region detection | invariante Momente, Multi-scale distance matrix, margin distance (Distanz zwischen Punkten des Original-Blattrands und des geglätteten Rands) | Linear Discriminant Classifier | Flavia, Leafsnap |  |
| (Kumar u. a., 2012) | Farbbasierte Segmentierung, | Curvature-based Shape Features, | Nearest Neighbours | Leafsnap |  |
| (Wang, Sun, Zhang, Ying, & Ma, 2016) | ? | PCNN, Hu-Momente, Zernike-Momente | SVM | Intelligent Computing Lab dataset |  |
| (Reyes, Caicedo, & Camargo, o. J.) | ? | fine-tuned CNN | CNN | ImageCLEF |  |
| (Satti, Satya, & Sharma, 2013) | Binarisierung, Kontur, Glättung | Farbe  Form: Geometrie, Morphologische Features (Adern, Glättungsfaktor …), Zahnung (über Winkel von Randpunkten) | KNN, ANN (keine Angabe welches) | Flavia | ANN schneller und genauer als KNN |
| (Cao, Wang, & Brown, 2016) | Binarisierung, Kontur, Beschreibung durch gleichmäßig verteilte Punkte | R-Angles (Skalierungsinv. über 7 Stufen) | L1-Norm, Punkt zu Punkt Dynamic Programming | Flavia | hohe Genauigkeit (speziell mit DP) |
|  |  |  |  |  |  |

### Form

R-Angles [Similarity based leaf image retrieval using multiscale R-angle description]

FFT

### Blatttextur

-zB Entropy, Kontrast etc. (Shabanzade, Zahedi, & Amin Aghvami, 2011) Extraktion der Blattadern

### Fourier-Transformation

## Lösungsansatz

### Datensatz

Es sind Datenbanken verfügbar, die unter freier Lizenz zum Training für die automatische Bestimmung genutzt werden können. Unter anderem veröffentlichten Wu, Bao, Xu, Wang, Chang und Xiang (2007) den Flavia-Datensatz, den sie im Rahmen der Entwicklung eines eigenen Algorithmus zur Blatterkennung erstellten. Der Datensatz beinhaltet 1800 Bilder von Blättern 32 unterschiedlicher, hauptsächlich chinesischer Arten. Die Bilder liegen vorverarbeitet vor. Die Blätter sind bereits freigestellt und die Blattstiele wurden entfernt (siehe Abbildung 10a).

Der sogenannte Leafsnap-Datensatz entstand in Zusammenarbeit der Columbia University, der University of Maryland und der Smithsonian Institution. Kumar, Belhumeur, Biswas, Jacobs, Kress, Lopez und Soares (2012) schufen die Grundlagen zur Entwicklung eines elektronischen Feldführers für Smartphones. Die zugrundeliegenden Trainingsdaten umfassen Fotografien von nach eigener Aussage allen 185 in den nordöstlichen USA beheimateten Baumarten. Die Bilder sind sowohl in ihrer ursprünglichen Form, als auch in segmentierter, binarisierter Form erhältlich (siehe Abbildung 10b).

Warum die beiden?

Farbe, Panaschierung … (Leafnet hat Farbschema an Seite)



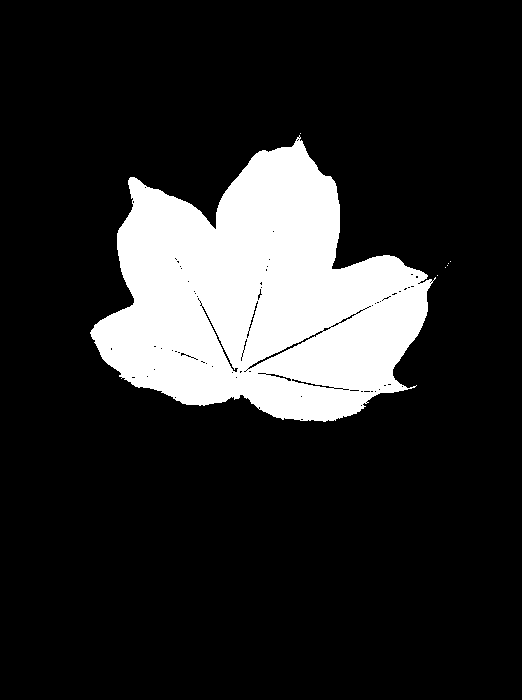


Abb. 10a: Beispielfoto aus der Flavia-Datenbank Abb. 10b: unbearbeitetes (links) und segmentier-

(Wu, Bao, Xu, Wang, Chang & Xiang, 2007) tes Blatt (rechts) aus dem Leafsnap-Datensatz

(Kumar, Belhumeur, Biswas, Jacobs, Kress, Lopez & Soares, 2012)

Den beiden Datensätzen ist gemein, dass nur Fotografien von den Blattoberseiten aufgenommen wurden falsch (ungeordnet bei Leafsnap). Es tauchen Bilder von Blattunterseiten zwischen denen von Blattoberseiten auf. Es soll in dieser Arbeit aber auch geprüft werden, ob sich die Blattunterseite besser eignet, um bestimmte Merkmale, im Besonderen die Blattadern, erkennen und trainieren zu können. geordnet, gleiche Menge an Ober- und Unterseiten, bessere Möglichkeit zu vergleichen

Daher wurde für die Bestimmung zusätzlich ein eigener Datensatz erstellt. Er enthält Daten 25 verschiedener Spezies mit jeweils 40 Fotografien je Ober- und Unterseite. Bei den Arten handelt es sich hauptsächlich um in Deutschland beheimatete Pflanzen. In Abbildung 11 ist ein Auszug aus den Bilddaten zu sehen. *(verweisen auf Anhang, Tabelle der Spezies)*



Abb. 11: Ober- (links) und Unterseite (rechts) eines Klarapfels (Malus Domestica)

Alle genannten Datensätze sollen beim Training des zu implementierenden Algorithmus Verwendung finden und die Ergebnisse der Erkennung verglichen werden.

AUFLÖSUNG! mehr Details

alles gestielte Blätter, netzadrige …

keine nadelförmigen Blätter von Nadelbäumen, wenig Textur-Informationen

Bild mit zusammengesetzter Blattform

### Segmentierung der Blätter

Die Blätter des selbsterstellten Datensatzes müssen zunächst vom Hintergrund freigestellt werden. Auch die Bilder des Leafsnap-Algorithmus von Kumar et. al. (2012) sollten vor der weiteren Verwendung erneut segmentiert werden. Neben den nicht segmentierten Bildern sind lediglich binarisierte Daten vorhanden (siehe Abbildung 10b). Diesen fehlen notwendige Informationen der Blatttextur, da sich die Autoren auf die Blattkontur konzentrieren und die Textur dafür nicht benötigen.

Kumar et. al. (2012) benutzen in ihrer Arbeit einen eigenen Algorithmus einer farbbasierten Segmentierung. Sie führen an, dass kanten- und regionsbasierte Verfahren gegenüber ihrem Verfahren langsamer arbeiten und Details der Blattränder verloren gehen können.

Wie in Abbildung 11 beispielhaft zu sehen, ist der Hintergrund im selbsterstellten Bildmaterial aber hell und gleichmäßig ausgeleuchtet, so dass sich die Blätter deutlich von ihm abheben. Dies bedeutet, dass sich auch in einem umgewandelten Graustufenbild zwei deutlich voneinander getrennte Maxima zwischen Vorder- und Hintergrund im Histogramm ergeben. Es wird angenommen, dass deshalb ein Schwellwert-Algorithmus basierend auf den Graustufen genügt, um das Blatt freizustellen ohne Details der Kontur zu verlieren. Zu diesem Zweck wird die Otsu-Methode verwendet, die mit bimodaler Graustufenverteilung gute automatisierte Schwellwerte liefert *(Quelle!)*.

da hintergrund heller -> invers (definition vorder/hintergrund)

Es muss allerdings getestet werden, ob diese Methode auch mit den Leafsnap-Bildern fehlerfrei arbeitet, da der Datensatz Bilder mit ungenügender Ausleuchtung und somit geringem Kontrast zwischen Vorder- und Hintergrund enthält *(Beispielbilder?)*.

### Wahl und Extraktion der Features

Wie aus den Grundlagen der Morph. von Blättern hervorgeht, sind Blätter sehr formenreich

Bezug auf Blattmerkmale, was nehmen, was nicht

zweidimensionale Abbildung

3D theoretisch möglich mit Bildserien aus verschiedenen Blickwinkeln, zu vernachlässigende Merkmale, Rollung fällt somit weg

invariant weil?

rotation

scale - verschiedene Winkel

Form und Textur möglichst getrennt betrachten

keine Farbe, weil Blätter zumeist grün, keine Blüttenblätter, die bunt sein können und bei denen es eine wichtige Rolle spielen würde

mit oder ohne Blattstiel? laut Grundlagen (Anheftung) auch Aussagekraft

### Blattform

Wie in Kapitel 2.1.2.2 angeführt, ist die Form der Blätter in der Botanik ein wichtiges Unterscheidungsmerkmal für die Bestimmung einer Pflanze. Daher soll sie auch bei der automatisierten Bestimmung eine besondere Beachtung finden.

Zur Beschreibung der Geometrie der Blätter eignen sich Momente. Es kann nicht davon ausgegangen werden kann, dass die als Ausgangsmaterial dienenden Blätter stets aus dem gleichen Winkel mit derselben Ausrichtung und Entfernung fotografiert werden. Daher müssen die beschreibenden Momente rotations-, skalierungs- und translationsinvariant gewählt werden. Dazu kommen die in Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.** erwähnten Hu-Momente in Betracht.

Daneben gibt es weitere invariante Momente, die zur Beschreibung in Frage kommen. Zulkifli (2009) vergleicht in seiner Arbeit Zernike-Momente, Legendre-Momente und Tchebichef-Momente als Eingangsdaten für ein neuronales Netz *(Spezialfall erwähnen: GENERAL REGRESSION NEURAL NETWORK*)*.* Er kommt zu dem Schluss, dass von den drei Verfahren die Tchebichef-Momente die geringste Fehlerrate bei der Erkennung aufweisen.

Aus diesem Grund werden sowohl Hu-Momente, als auch Tchebichef-Momente zum Einsatz kommen und deren Erkennungsgenauigkeit verglichen werden.

*… (Kontur?)*

#### Momente

Hu, Zernike

Zernike moments are not a texture feature, but rather a global measure of how the mass is distributed

#### Shape Context

#### Blattadern

### Blattadern

Texturbeschreibung

#### Local binary pattern

#### Blattadern

Als Schwerpunkt dieser Arbeit soll die Eignung der Struktur der Blattadern zur Erkennung der Pflanzenspezies überprüft werden.

Zur Klassifikation von Leguminosen nutzen Larese, Bayá, Craviotto, Arango, Gallo und Granitto (2014) in ihrer Abhandlung ausschliesslich die Blattadern. Als Besonderheit können sie durch ihre angebrachte Methode auch verschiedene Variationen innerhalb einer Art unterscheiden. Zur Segmentierung der Adern nutzen sie eine erweiterte Form der Hit or Miss Transform *(erklären!)* auf den Graustufenbildern. Sie wird für verschiedene Skalierungen durchgeführt, um eine skalierungsinvariante Beschreibung zu erhalten. Als Features dienen unter anderen die Anzahl der Kanten, Knoten und die Winkel der Kanten. Allerdings sind speziell die Winkel nicht rotationsinvariant. Aus diesem Grund müssen die Blätter bereits in der Vorverarbeitung entlang ihrer Längsachse gleich ausgerichtet werden.

Alternativ können die Blattadern auch durch einen Gabor-Filter segmentiert werden, wie es Katyal (2012) vorschlägt. *(mehr)* Seine Arbeit konzentriert sich auf die Freistellung der Adern, beinhaltet aber keine Methode zur Extraktion etwaiger Features.

Vorschlag für Features:

* Zahl der starken Kanten (Adern mit größter Breite, wie differenzieren?)
* Zahl der Areolen
* Volumen der Adern (Anzahl Pixel im Verhältnis Volumen Blatt?)
* Länge (Pixelzahl im Verhältnis Kontur?)
* wie Winkel rotationsunabhängig beschreiben? im Verhältnis zur Hauptachse? (ist das nicht immer gleich?)

(Larese u. a., 2014)

* Feature: 100 x 100 Patch für jede Skalierung (um Shape-Einfluss rauszunehmen)

52 traits (Tabelle) VMeO/VmO/VMO Nicht rotationsinvariant!

* sKlassifikation: Random Forests, Support Vector Machines with Gaussian kernel und Penalized Discriminant Analysis (linear nach Fisher)
* Recursive Feature Elimination (RFE)

Röntgenbilder stehen nicht zur Verfügung

* Nur Farbbild, Röntgenbilder besser, wird auch teilweise so angewendet (Areolenbestimmung)

### Lernverfahren

SVM, MLP

Deep Learning in letzter Zeit zwar der neueste heiße Scheiß, aber hier „Domänenwissen“ verwenden

pcnn - pulse-coupled neural networks (Wang u. a., 2016)

* 2D-neuronales Netz basierend auf visuellem Cortex von Katzen
* selbstständiges Finden von relevanten Features
* jedes Neuron repräsentiert Farbinfo eines Pixels
* Neuronen sind mit Nachbarn verbunden
* Puls setzt Stimulus zurück

### Wahl der Modelle

SVM dazu, damit vergleichbar

gleiche Dim. wie bei MLP der Eingangsdaten mögl.

probabilistisches neuronales Netz (MLP)

wenn zu langsam PNN

wenn nicht zufriedenstellend irgendwas anderes

support vector machine

### Metriken zur Messung der Erkennungsgenauigkeit

## Systementwurf

### Systemkomponenten

keine blosse Abfolge von Skripten, sonder strukturiert, damit austauschbare Komponenten (Interface, abstrakte Klassen)

### Datenbank?

Warum Datenbank?

* zentraler Zugriff
  + vereinfachter Zugriff über Netzwerk
  + Übersicht (besser als lauter Dateien irgendwo)
  + besser als HDF5, da immer ganz gelesen wird
* Option: pandas, LMDB

„Sacred“ to keep track? <http://sacred.readthedocs.io/>

HDF5:

* effiziente Serialisierung/Speicherung von numpy arrays
* cross languages
* common dataformat in science.
* einfacher physischer Austausch der Daten
* Nachteil: nur Arrays(?)

hdf5 vs pickle (Serialisierung): - pickle kann nur aus python gelesen werden, kein zukunftssicherer Standard(?), kleinere Dateien,

warum SQL

* strukturiert (besser für Auswertung der Ergebnisse, übersichtlicher)
* Datenmenge überschaubar (große Datenmengen -> besser NOSQL)
* Daten gleichförmig?

NOSQL

* Skalierbarkeit horizontal (auf Rechner verteilen)

# Implementierung

### Verwendete Bibliotheken

in Kapitel 3 rein!

Binärbilder trotzdem in 8Bit, definition: 0 Hintergrund >0 (1 oder 255) Vordergrund, opencv liefert beim thresholding 255

nicht Java sondern Python

OpenCV Konturpunkte, Binarisierung

skimage und opnecv, eigentlich alle benutzen numpy arrays als represäntation von Bildern, benutzt: 8bit farbrepräsentation

versuch schleifen zu vermeiden mit numpy, indizierung über mehrere Indizes

### allgemein

alle Features erwarten segmentiertes Bild, alle Segmentierer liefern segmentiertes Bild (Vereinheitlichung)

Unsupervised Feature genannt, um anzuzeigen, dass unsupervised lernverfahren folgen muss clustering oder codebook nötig

Segmentierte Bilder, Hintergrund weiß (abgeguckt von allen anderen)

einfache Binarisierung: da Hintergrund mit weiß codiert, gleich schwarz setzen (Binärbild Codierung!) und Rest auf 1 bzw. 255

Achtung: jpg macht Probleme! (irgendwo anders hier)

Feature Klasse übernimmt Vorbereitung des Bildes, delegiert an Unterklassen, wo eigentliche Berechnungen stattfinden

### Segmentierung

Ausgang immer maskiertes Bild

Die Segmentierung mittels Otsu-Thresholding wurde mit Hilfe der OpenCV-Bibliothek umgesetzt. Zunächst wird das Ausgangsbild im RGB-Farbraum eingelesen und in ein Graustufenbild umgewandelt. ??? Allerdings war das Ergebnis entgegen der Vermutung aus Kapitel unzureichend für eine anschließende Weiterverarbeitung. In einigen Bildern verfälschen lange Schatten die Kontur des binarisierten Blattes. Da bei der Aufnahme des Testdatensatzes teilweise ein Blitzlicht verwendet wurde, sind die vom Blatt geworfenen Schatten sehr dunkel. Da auch das Blatt dunkler ist als der Hintergrund, werden die Pixel der Schatten dem Vordergrund zugeordnet. (siehe Bilder)

Da die Fotos der Blätter vor einem neutralen farbarmen Hintergrund aufgenommen wurden, lassen sich die Blätter vor allem in der Sättigung von diesem unterscheiden, der eigentliche Farbwert ist vernachlässigbar. Somit ist der HSV-Farbraum, der die Helligkeit von der Farbsättigung und dem Farbwert trennt, für die Freistellung der farbigen Blätter besser geeignet als der RGB-Farbraum (siehe Abbildung). Ein Otsu-Thresholding des Sättigungsanteils trennt in einer Vielzahl der Bilder

Warum kmeans? Bilder raussuchen, Kmeans im Farbraum, mehr informationen

Es ist also notwendig

Zudem fransen die schattigen Ränder der Blätter aus.

Löcher schließen, Löcher sollen Merkmal der Textur, aber nicht der Form sein

Hintergrund eingefärbt vom Umgebungslicht. Einige Fotos in Dämmerung geschossen, farbiges Licht

Auch dies ergab

value dazu

Bestehendes Problem: farbige Schatten. Die Schatten sind hauptächlich grau mit einer geringen Sättigung. zwei Lichtquellen: Umgebungsbeleuchtung, Blitz

Da aber auf einigen Es verbleiben dennoch farbige Schatten,

Generell erfolgt vor allen Segmentierungen, bei denen Faltungen mit festen Kernelgrößen vorgenommen werden, zunächst eine Skalierung des Bildes auf eine vordefinierte Größe. So wird verhindert, dass die strukturierenden Elemente unterschiedlichen Einfluss haben (Erosion, Dilation).

kein Gauss, damit Kontur scharf bleibt (-> Segmentation verursacht deswegen Löcher)

DSC\_6982.JPG, DSC\_6813.JPG - Beispiel dafür, dass einfaches Otsu und k-means nicht funktionieren

H - viel noise

S + V - wenig noise, bisschen harte schatten

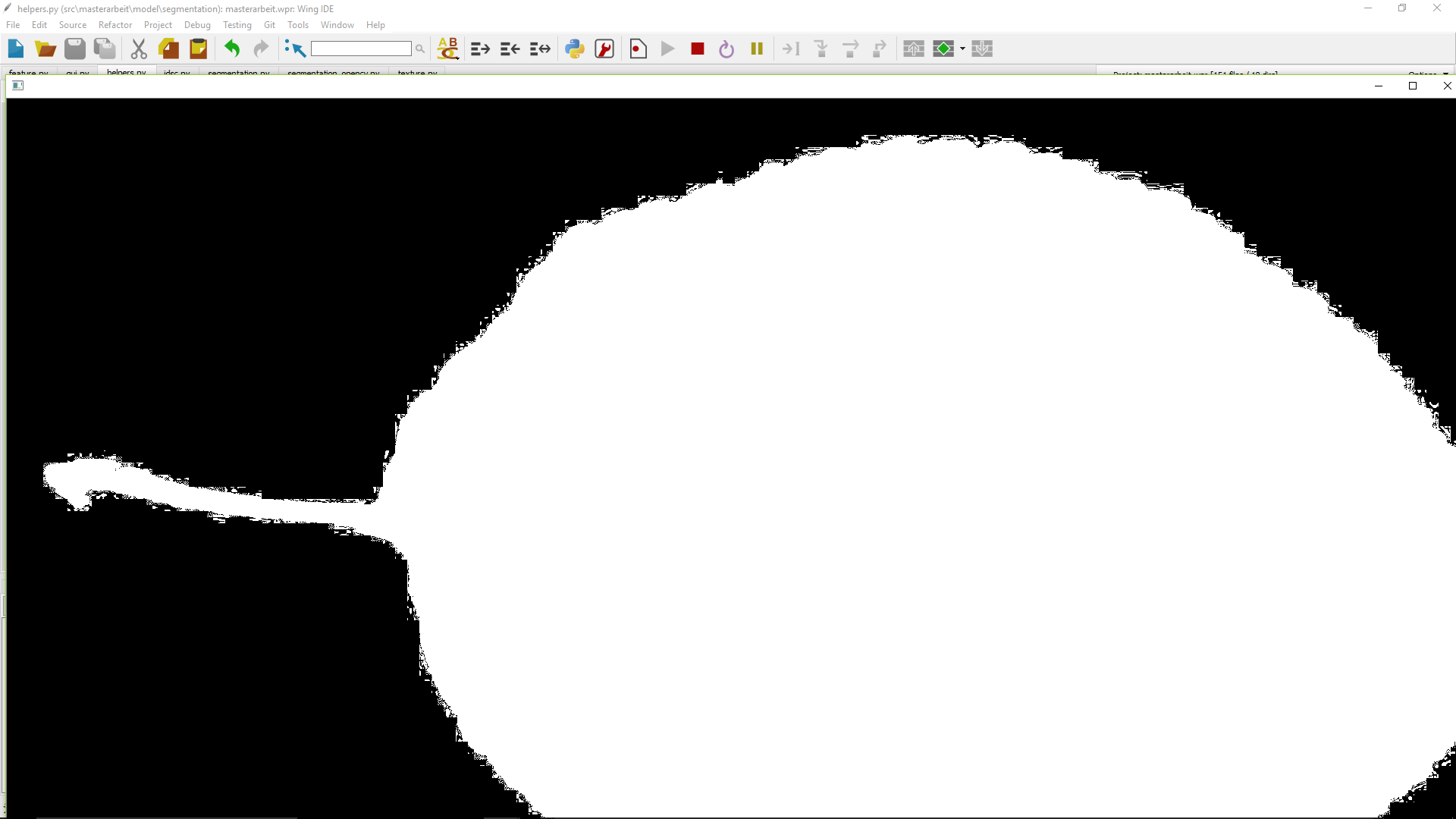
nur S - weniger schatten, noisy, da fragmente der schatten

Schwierigkeit bei entfernung stiel entfernt (vielleicht bezug auf Plant identification using leaf shapes—A pattern counting approach)

Stiel wichtig (Grundlage), Details von Blattgrund und Blattspitze kann verloren gehen

Entscheidung: Stiel bleibt dran

normalisieren, da nur erhältnis der werte zueinander entscheidend, nicht höhe der werte (wie oft taucht pattern x auf im verh. zu allen anderen)



Shape allgemein: immer längste kontur nehmen, falls mehrere Konturen, kleine Objekte bei Segmentierung zurückbleiben

### Inner Distance Shape Context

Für Python sind keine Bibliotheken verfügbar, die Methoden zur Extraktion des Inner Distance Shape Context bereitstellen. Daher wird der Algorithmus von ??? selbst implementiert.

Ausgangspunkt für die Berechnung jedweder Shape Contexts ist die aus einem Binärbild ermittelte Kontur. Dieser wird eine vordefinierte Anzahl an Punkten, die in regelmäßigen Abständen verteilt sind, entnommen. Keine besonderen Punkte

Die euklidischen Distanzen zwischen allen entnommenen Punkten werden berechnet und in einer mit Nullen initialisierten Distanzmatrix gehalten. Abweichend vom ursprünglichen Konzept des Shape Contexts, werden allerdings nur innere Distanzen berücksichtigt. Dies bedeutet, dass die direkte Linie zwischen zwei betrachteten Punkten komplett innerhalb der Form liegen muss, damit sie innerhalb der Form als verbunden gelten können und ihre Distanz eingetragen wird. Da der Hintergrund im Binärbild mit Nullen codiert ist, wird dazu programmatisch überprüft, ob Zwischenpunkte auf dieser Linie nur Werte größer Null annehmen. Ist dies der Fall, wird die Distanz eingetragen. Die verbleibenden Nullen geben an, dass keine Verbindung zwischen den jeweilig indizierten Punkten besteht.

Die Distanzmatrix bildet den Eingangsgraphen für die Ermittlung der kürzesten Wege zwischen den Punkten der Kontur, wobei die euklidische Distanz die Kantenlänge angibt. Über den Floyd-Warshall-Algorithmus wird eine Matrix erzeugt (scipy), die paarweise die kürzesten Wege enthält. Jeder kürzeste Weg ist die minimale Summe der Länge der Kanten im ungerichteten Graph, die abgelaufen werden müssen, um den Punkt zu erreichen. Gibt es keine Kante zwischen zwei Punkten, ist der kürzeste Weg, wie in der Distanzmatrix, gleich Null, beziehungsweise unendlich.

Pseudocode hierhin

Für jeden aus der Kontur entnommenen Punkte wird anschließend ein Histogramm erstellt. Das Histogramm beschreibt das Verhältnis zu allen anderen Punkten im Hinblick auf die relative Orientierung und die Länge des kürzesten Weges zu ihnen. Die relative Orientierung wird als Winkel zwischen der an der Kontur anliegenden Tangente am betrachteten Punkt und dem Zielpunkt definiert.

Für die Zuordnung der Punkte in das Histogramm werden Winkel und Wegelängen als Polarkoordinaten in jeweils acht Klassen unterteilt. Der Wertebereich der Winkel liegt zwischen 0 und 2π, der der Wegelängen zwischen 0 und dem Logarithmus der maximalen Distanz. Als maximale Distanz wird die Diagonale des Ausgangsbildes angenommen. entspricht der Vorlage

Jeder nicht erreichbare Punkt wird der Histogramm-Klasse, zugeordnet, die alle Punkte enthält, die im Hinblick auf kürzeste Wege den größten Abstand zum betrachteten Punkt haben. Somit ist das Histogramm jedes Punktes vollständig beschreibend, die Summe seiner Werte entspricht immer der Anzahl der entnommenen Punkte minus den betrachteten Punkt. Um den Wertebereich des Histogramms unabhängig von der Anzahl der Punkte zu halten, wird die L1-Norm, also die Summennorm, darauf angewendet.

Histogrammformel hier hin

Um die Konturen von Objekten anhand ihrer ermittelten Inner Distance Shape Contexts zu vergleichen und zu klassifizieren, verwenden bla et al, dynamische Programmierung. Da die Histogramme nicht geordnet, irgendwas Dies eignet sich nicht für die angestrebte Lösung, die Klassifizierung mit neuronalen Netzen oder Support Vector Machines durchzuführen.

Die erzeugten Histogramme können aber als sich wiederholende Muster gleicher Dimension angenommen werden. Diese Muster können als einzelne Fragmente für ein Bag-of-words-Modell dienen und zu visuellen Wörtern zusammengefügt werden, die wiederum die Form beschreiben. Diese Aufgabe übernehmen die in Kapitel ??? beschriebenen Codebooks.

Die einzelnen Histogramme mit einer Dimension von 8 x 8 werden konkateniert, so dass das Endergebnis der Feature-Extraktion eine Matrix der Dimension Anzahl der Punkte x 8 x 8 ist.

### Multilevel Inner Distance Shape Context

Angelehnt an die Idee des Independent Inner Distance Shape Context von ??? wurde der im vorigen Kapitel beschriebene Algorithmus erweitert, aus der Motivation heraus, die feinen Details der Konturen unabhängig von der Gesamtkontur beschreiben zu können. Anderer Name, da Umsetzung von dem vorgeschlagenen Algorithmus abweicht.

Wie auch beim Independent Inner Distance Shape Context, wird zunächst mehrere Stufen des binarisierten Bildes gebildet. Auf der niedrigsten, der feinsten Stufe, bleibt das Ausgangsbild erhalten, um alle Details der Kontur abzubilden. Auf den weiteren Stufen wird das Ausgangsbild mit einem Gauß-Filter *Gσ(x,y)* gefaltet, und das Bild somit geglättet (siehe 2.2.3.2). Zwischen den Stufen steigt die Standardabweichung *σ* stetig an, so dass auch der Grad der Glättung stetig ansteigt und die Form des Objektes so an Details verliert. Konkret wird *σ*, beginnend mit *σ = 8* auf der zweiten Stufe, von Stufe zu Stufe verachtfacht. Thresholding!

mean

Die maximalen Distanzen bestimmen hier zusätzlich die Größe der Nachbarschaft eines Punktes in Bezug auf die maximale Länge der kürzesten Wege. Auf dem feinsten Detailgrad, also der niedrigsten Stufe der Gauß-Pyramide, wird nur die unmittelbare Nachbarschaft betrachtet, die maximale Distanz muss dementsprechend klein sein. Auf dem gröbsten Detailgrad sollen alle Punkte der Kontur als benachbart angenommen werden, die maximale Distanz ist am größten.

Die geglätteten Konturen der Bilder der Gauß-Pyramide dienen als Eingangspunkte des Algorithmus des Inner Distance Shape Context (Kapitel ..). Der Algorithmus wurde dahingehend modifiziert, dass Punkte außerhalb der maximalen Distanz ignoriert werden, anstatt sie der letzten Histogramm-Klasse zuzuordnen. Das Histogramm ist somit nicht vollständig beschreibend, sondern beschreibt nur die Nachbarschaft der Punkte.

Abweichend vom Independent Inner Distance Shape Context von … werden die Stufen der Gauß-Pyramide unabhängig voneinander behandelt. Die Konturpunkte zwischen den Stufen sind nicht deckungsgleich. Stattdessen werden mehr Punkte betrachtet, je feiner der Detailgrad ist. Der Grund dafür ist, dass die Gesamtzahl an Konturpunkten erhöht werden muss, damit für eine detaillierte Beschreibung einer kleinen Nachbarschaft genug Punkte in diesem Ausschnitt der Kontur zur Verfügung stehen.

Die resultierenden Histogramme der verschiedenen Stufen sind als Basiselemente, aufgrund der verschieden definierten Nachbarschaften, nicht direkt miteinander kombinierbar. Daher wird für jede Stufe ein eigenes Codebook benötigt.

### Codebooks

Die Klassifizierer benötigen als Eingangsdaten eindimensionale Feature-Vektoren, die statistische Eigenschaften eines Objektes in einer festen Reihenfolge beschreiben. Um ungeordnete Features zu beschreiben, können diese als Fragmente visueller Wörter aufgefasst werden. Über die Häufigkeit des Auftretens der visuellen Wörter, können Histogramme gebildet werden, welche dann wiederum als eindimensionale geordnete Feature-Vektoren nutzbar sind. Die Aufgabe der Bestimmung geeigneter visueller Codewörter übernehmen die im Folgenden beschriebenen Codebooks. (kommt in 3 rein)

Zunächst bei Extraktion der Features direkt angewendet, eigenes Feature für jeden Typ, getrennt gespeichert, on the fly codebook erstellung

später getrennt, da Extraktion lange dauert und Codebooks auf die gleichen Ausgangsdaten der Features zugreifen, für bessere Vergleichbarkeit. -> viel Speicherbedarf, on the fly noch nicht vollständig, da bisher keine möglichkeit, benutzten codebook typ mitzuspeichern (für predict wichtig), achtung: mischung vermeiden, on the fly ist deaktiviert

### Sparse dictionary learning

Speziell für die Umwandlung der Histogramme des Inner Distance Shape Context wurde ein Wörterbuch für das sogenannte „sparse dictionary learning“ umgesetzt. Die Aufgabe des Wörterbuchs ist es, die Menge der extrahierten Shape-Context-Vektoren zu reduzieren, indem diese in ihre Atome zerlegt werden. Mit Atomen werden die Linearkombinationen bezeichnet, aus denen sich die Vektoren komponieren lassen.

Löst das Problem der Formel

Vorteil: online dictionary (wird nicht benutzt), Dictionary kann erweitert werden

Angelehnt an die Arbeit von Zhao et al. sind die Kombinationen der Atome des Wörterbuchs sich wiederholende Muster und können somit als visuelle Wörter betrachtet werden. und in einem Histogramm gezählt. Abweichend nicht darauf geachtet, dass nur bestimmte Zahl auftauchen darf.

Atom als visuelles Wort

Die Repräsentation ist übererfüllt,

Dictionary. KMeans zusätzlich

erwartete Dimension

### Multilayer Perceptron

### Support Vector Machine

### Datenhaltung

### Evaluation

# Ergebnis

local binary pattern keine Aussagekraft, auch in Kombination nicht

weder

Annahme: Kombination besser, da jedes für sich nicht umfassend beschreibend

## Genauigkeit (welche Metrik?)

Vergleich verschiedene Blattformen

Vergleich steigende Anzahl an Klassen

Vergleich Vorder/Rückseite

Vergleich steigende Anzahl an Trainingsblättern (Erkennungsrate)

Kombination der Features

## Aufwandsanalyse?

# Zusammenfassung

## Ausblick

* Gleiche Spezies untersuchen – verschiedene Sorten (möglicherweise gleiche Blattform, aber Unterschiede in Blattadern), im Datenset nur Apfel (?)
* bugs/usability in Oberfläche (war für Zwecke hier egal9
* Smartphone App

Optimierung IDSC: Punkte auf Linie immer 10, kein check für unique Punkte(wenn 10 zuviel)

Performance generell (in C schreiben, python zu langsam, opencv interface?)

Textur integrieren in IDSC (SPT)

Trainingsfotos: Beleuchtung, eine diffuse, unfarbige Lichtquelle, DSC\_5821

bekannte Fehler: gelegentlicher hdf5 error beim Zugriff auf Datei

„H5F\_open - unable to open file“

# Abbildungsverzeichnis

[Abbildung 1 - Bau des Laubblattes: Blattspreite (1), Blattrand (2), Blattspitze (3), Blattnerv (4), Blattstiel (5), Blattgrund (6) (nach Welle, 2014, S. 27) 2](#_Toc475891558)

[Abbildung 2 - Beispiele möglicher Formen einfacher ungeteilter Blätter, von links nach rechts: linealisch, lanzettlich, elliptisch, eiförmig, verkehrt eiförmig, kreisrund, keilig, spatelig, rautenförmig, schildförmig, herzförmig, verkehrt herzförmig (Rothmaler, 1967, S. XV, XVI) 3](#_Toc475891559)

[Abbildung 3 - Beispiele geteilter und gespaltener Laubblätter, von links nach rechts: handförmig gespalten, fiederförmig gespalten, handförmig geteilt, fiederteilig (Rothmaler, 1967, S. XIV) 3](#_Toc475891560)

[Abbildung 4 - Beispiele zusammengesetzter Blätter, von links nach rechts: 3-zählig, 5-zählig, fußförmig, gefiedert, unpaarig gefiedert (Rothmaler, 1967, S. XIV) 3](#_Toc475891561)

[Abbildung 5 - Formen des Spreitengrundes und der Spreitenspitze, Spreitengrund (1 bis 3) von links nach rechts: gestutzt, abgerundet, keilig, | Spreitenspitze (4 bis 11) von links nach rechts: gestutzt, abgerundet, stumpf, spitz, zugespitzt, stachelspitz, bespitzt, ausgerandet (Rothmaler, 2011, S. 886) 4](#_Toc475891562)

[Abbildung 6 - Formen von Blatträndern, von links nach rechts: ganzrandig, gesägt, doppelt gesägt schrotsägeförmig, gezähnt, gekerbt, gebuchtet, geschweift 4](#_Toc475891563)

[Abbildung 7 - RGB-Farbraum (Burdick, 1997, S. 23) 6](#_Toc475891564)

[Abbildung 8 - HSV-Farbraum als Zylinder (Wikipedia, 2017) 6](#_Toc475891565)

[Abbildung 9 - Graustufenbild (links oben) mit seinem Histogramm (rechts oben), Binarisierung mit globalen Schwellwerten von 110, 147 und 185 (unten, von links nach rechts) (Jähne, 2012, S. 543) 7](#_Toc475891566)

[Abbildung 10 - Anwendung eines Filters durch zeilenweises Verschieben der Filtermaske (Jähne, 2012, S. 303) 9](#_Toc475891567)

[Abbildung 11 - Koordinatensystem der Filtermatrix *H* mit „Hot Spot“ genanntem Ursprung in der Mitte (Burger & Burge, 2015, S. 96) 9](#_Toc475891568)

[Abbildung 12 - Beispiel für einen Gaußfilter, Darstellung der Filterfunktion *H* als 3D-Plot (links) und als Filtermatrix (rechts) (Burger & Burge, 2015, S. 103) 10](#_Toc475891569)

[Abbildung 13 - Beispiel einer Dilation auf Binärbild *I* mit Strukturelement *H* (Mitte) und dem Ergebnisbild *I* ⊕ *H* (rechts), Elemente mit dem Wert 1 sind mit • markiert, die Ursprünge der jeweiligen Koordinatensysteme sind rot markiert. (Burger & Burge, 2015, S. 196) 11](#_Toc475891570)

[Abbildung 14 - Originalbild (links) nach Ausführung von Opening (Mitte) beziehungsweise Closing (rechts) mit Strukturelementen der Dimension 3 x 3 (nach Jähne, 2012, S. 561, 562) 12](#_Toc475891571)

[Abbildung 15 - funktionale Struktur eines Klassifikators (nach Ertel, 2016, S. 194) 13](#_Toc475891572)

[Abbildung 16 - verschieden komplexe Entscheidungsgrenzen in einem zweidimensionalen Featureraum zur Unterscheidung zwischen zwei Klassen - Lachs („salmon“, schwarze Punkte) und Wolfsbarsch („sea bass“, rote Punkte) - anhand zweier gegenübergestellter Merkmale - Breite („width“) und Helligkeit („lightness“) (nach Duda, Hart, & Stork, 2001, S. 5, 6) 14](#_Toc475891573)

[Abbildung 17 - Schema des Entwerfens eines Systems zur Mustererkennung und Klassifizierung (nach Duda u. a., 2001, S. 14) 15](#_Toc475891574)

[Abbildung 18 - die zwei Klassen R1 und R2 trennende optimale Hyperebene („hyperplane“) mit ihren Support-Vektoren (hervorgehobene Punkte) und dem Rand (zwischen Support-Vektoren und Hyperebene) 16](#_Toc475891575)

[Abbildung 19 - schematische Darstellung von zwei biologischen Neuronen (nach Keller, Liu, & Fogel, 2016, S. 8) 17](#_Toc475891576)

[Abbildung 20 - schematische Darstellung von zwei künstlichen Neuronen mit Eingabe- und Ausgabesignalen (Keller u. a., 2016, S. 8) 17](#_Toc475891577)

[Abbildung 21 - Aktivierungsfunktionen von links nach rechts: Schwellwertfunktion, logistische Funktion, hyberbolischer Tangens (nach Keller u. a., 2016, S. 28, 29, 30) 18](#_Toc475891578)

[Abbildung 22 -Schema eines Multilayer-Perceptron mit zwei versteckten Schichten („hidden layer“), der Eingabeschicht („input layer“) und der Ausgabeschicht („output layer“). Das Eingangssignal („input signal“) wird vorwärts durch das Netz propagiert, mit dem Ausgangssignal („output signal“) als Ergebnis (Keller u. a., 2016, S. 36) 19](#_Toc475891579)

[Abbildung 23 - k-Means mit zwei zu bildenden Clustern (hellgrau und schwarz) nach der Initialisierung (links) und nach *t* Iterationen (ab 2. von links). Die Zuordnung der Datenpunkte zu den Clusterzentren (Kreise) ist durch die Graustufe (hellgrau oder schwarz) codiert. (Ertel, 2016, S. 247) 21](#_Toc475891580)

# Tabellenverzeichnis

**Es konnten keine Einträge für ein Abbildungsverzeichnis gefunden werden.**

# Codelistings

**Es konnten keine Einträge für ein Abbildungsverzeichnis gefunden werden.**

1. Literaturverzeichnis

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer Science+Business Media.

Burdick, H. E. (1997). *Digital Imaging- Theory and Applications*. New York: McGraw-Hill.

Burger, W., & Burge, M. J. (2015). *Digitale Bildverarbeitung: eine algorithmische Einführung mit Java* (3. Aufl.). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-04604-9

Cao, J., Wang, B., & Brown, D. (2016). Similarity based leaf image retrieval using multiscale R-angle description. *Information Sciences*, *374*, 51–64. https://doi.org/10.1016/j.ins.2016.09.023

Chaki, J., Parekh, R., & Bhattacharya, S. (2015). Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognition Letters*, *58*, 61–68. https://doi.org/10.1016/j.patrec.2015.02.010

Du, J.-X., Wang, X.-F., & Zhang, G.-J. (o. J.). Leaf shape based plant species recognition. https://doi.org/10.1016/j.amc.2006.07.072

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2. Aufl.). New York: John Wiley & Sons.

Ertel, W. (2016). *Grundkurs Künstliche Intelligenz* (4. Aufl.). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-8348-9989-7

Han, S., Chee, L., Chan, S., Wilkin, P., & Remagnino, P. (o. J.). DEEP-PLANT: PLANT IDENTIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS.

Jähne, B. (2012). *Digitale Bildverarbeitung: und Bildgewinnung (German Edition)* (7. Aufl.). Heidelberg: Springer Verlag.

Kalyoncu, C., & Toygar, Ö. (2015). Geometric leaf classification. *Computer Vision and Image Understanding*, *133*, 102–109. https://doi.org/10.1016/j.cviu.2014.11.001

Katyal, V. (2012). *Leaf vein segmentation using Odd Gabor filters and morphological operations*. Amity University Indien.

Keller, J. M., Liu, D., & Fogel, D. B. (2016). *Fundamentals of Computational Intelligence*. New Jersey: John Wiley & Sons.

Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I., & Soares, J. V. B. (2012). Leafsnap: A Computer Vision System for Automatic Plant Species Identification. *Proceedings of the 12th European Conference on Computer Vision (ECCV)*.

Larese, M. G., Bayá, A. E., Craviotto, R. M., Arango, M. R., Gallo, C., & Granitto, P. M. (2014). Multiscale recognition of legume varieties based on leaf venation images. *Expert Systems with Applications*, *41*(10), 4638–4647. https://doi.org/10.1016/j.eswa.2014.01.029

Reyes, A. K., Caicedo, J. C., & Camargo, J. E. (o. J.). Fine-tuning Deep Convolutional Networks for Plant Recognition.

Rothmaler, W. (1967). *Exkursionsflora von Deutschland - Gefäßpflanzen* (6. Aufl.). Berlin: Volk und Wissen volkseigener Verlag.

Rothmaler, W. (2011). *Exkursionsflora von Deutschland - Gefäßpflanzen* (20. Aufl.). Heidelberg: Spektrum Akademischer Verlag.

Satti, V., Satya, A., & Sharma, S. (2013). AN AUTOMATIC LEAF RECOGNITION SYSTEM FOR PLANT IDENTIFICATION USING MACHINE VISION TECHNOLOGY. *International Journal of Engineering Science and Technology*, *5*(4), 975–5462.

Shabanzade, M., Zahedi, M., & Amin Aghvami, S. (2011). Combination of Local Descriptors and Global Features for Leaf Recognition. *Signal & Image Processing : An International Journal*, *2*(3), 23–31. https://doi.org/10.5121/sipij.2011.2303

Wang, Z., Sun, X., Zhang, Y., Ying, Z., & Ma, Y. (2016). Leaf recognition based on PCNN. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-015-1904-1

Welle, E. F. (2014). *Kleines Repetitorium der Botanik* (16. Aufl.). Hamburg: Verlag Dr. Felix Büchner - Handwerk und Technik.

Wikipedia, C. (2017). HSL and HSV. Abgerufen 22. Februar 2017, von https://en.wikipedia.org/w/index.php?title=HSL\_and\_HSV&oldid=765756890

Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. *ISSPIT 2007 - 2007 IEEE International Symposium on Signal Processing and Information Technology*, (December), 11–16. https://doi.org/10.1109/ISSPIT.2007.4458016

Zhao, C., Chan, S. S. F., Cham, W.-K., & Chu, L. M. (2015). Plant identification using leaf shapes—A pattern counting approach. *Pattern Recognition*, *48*(10), 3203–3215. https://doi.org/10.1016/j.patcog.2015.04.004

Zulkifli, Z. B. (2009). *PLANT LEAF IDENTIFICATION USING MOMENT INVARIANTS & GENERAL REGRESSION NEURAL NETWORK*. Faculty of Computer Science and Information Systems.

1. Anhang

## Installationsanleitung

-theano.orc (Windows: unter %USERPROFILE%):

[*global]*

*device =* gpu

- CUDA Toolkit installieren

u.U in theano.orc (wenn cl.exe nicht gefunden oder Value '2008' is not defined for option 'cl-version')

[*nvcc]*

*compiler\_bindir=E:\Programme\Microsoft Visual Studio 12.0\VC\bin*

utf-8 not supported: in theano.compat ändern (fett):

def decode\_iter(itr):

for x in itr:

yield x.decode(**'gbk'**)

header-Datei fehlt: in das angegebene Verzeichnis kopieren