

Aufgabe 1

Computer Vision

Hochschule Darmstadt, Sommersemester 2020

Roman Kessler

30. April 2020

Aufgabe 1.1

Im folgenden wird der Vorgang skizziert, wie ich aus einem Bild der Fingerkuppe mit mittelmäßiger Auflösung und Beleuchtungsbedingungen den Fingerabdruck sichtbar gemacht habe. Untersuchungsobjekt ist mein linker Zeigefinger.

Zunächst war die Aufnahme des Bildes eine Herausforderung. Optimale Lichtbedingungen - wären wahrscheinlich Tageslicht gewesen. Es wurde stattdessen versucht, mit einer Schreibtischlampe die Fingerkuppe halbwegs gleichmäßig zu beleuchten. Es mussten mehrere Winkel fotografiert - und dadurch die Reflexion des Lichtes ausgenutzt werden - bis die Tiefen und Erhebungen des Fingers auf einem der Bilder gut genug erkennbar waren.

Für erste Übungen mit *OpenCV* und Bilderarbeitung in *Python 3* und zur besseren Reproduzierbarkeit der Ergebnisse wurde die Aufgabe in Python durchgeführt. Im ersten Schritt wurde das Schwarz-Weiß Bild auf den zu interessierenden Bereich zurecht geschnitten ("cropped"), und das Histogramm dargestellt (siehe Abbildung 1).

An dem Histogramm erkennt man schon eine bimodale Verteilung der Grauwerte in den Bereichen 100 und 200 (Abbildung 1, rechts). Um den Grauwertbereich besser auszunutzen, entscheide ich mich, das Histogramm zu strecken. Dadurch wird die Verteilungsfunktion zu einer (etwa) linearen Funktion gemacht. Die Dichte der Grauwerte wird von 0 zu 255 gestreckt (Abbildung 2).

Durch die Streckung des Histogramms werden die Muster der Fingerkuppe für das menschliche Auge schon besser ersichtlich. Als nächstes wurde eine adaptive Histogramm-Angleichung versucht. Dazu wird das Bild in kleinere Kacheln (8x8 Pixel) geteilt, und eine Histogrammangleichung für jede Kachel separat durchgeführt. Wir sehen noch einmal eine leichte Verbesserung in der Erkennung der Strukturen (Abbildung 3, links) und eine noch gleichmäßigere und stufenlosere Verteilung der Grauwerte (Abbildung 2, rechts, auch wenn diese für das menschliche Auge vermutlich kaum zu erkennen ist).

Es wurde noch versucht, das Bild zu binarisieren. Hierzu wurde - mit Trial-and-Error - ein Schwellwert zu finden, nach welchem die Täler und Hügel der

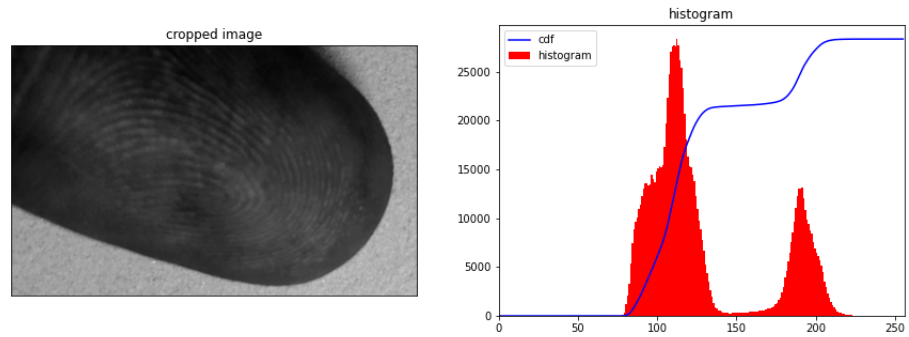


Abbildung 1: Ursprungsbild

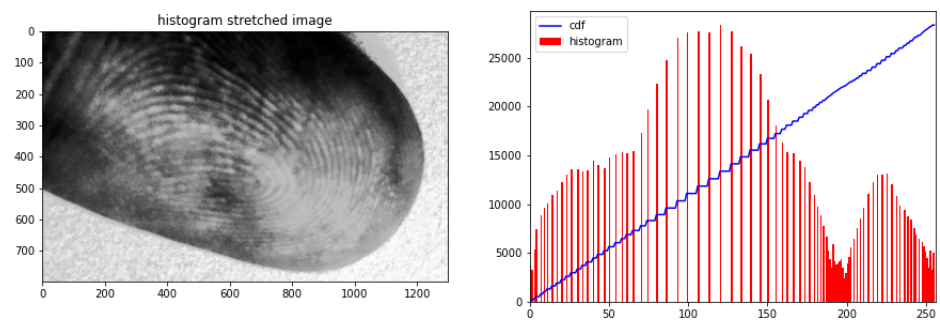


Abbildung 2: Histogramm-gestrecktes Bild

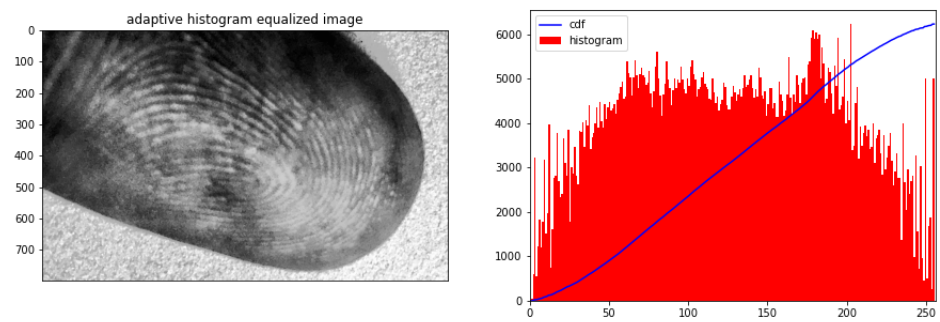


Abbildung 3: Adaptive Histogrammangleichung

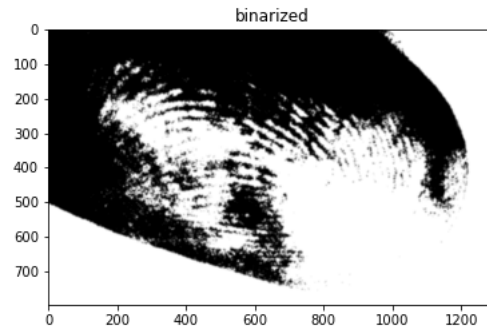


Abbildung 4: Binarisierung

Fingerkuppen gut unterscheidbar sind. In Abbildung 4 wurde der Schwellwert 120 probiert.

Man sieht klar, dass das Bild global noch sehr ungleichmäßige Grauwertverteilungen aufweist. Um das auszugleichen, könnte man weiterführend z.B. eine Grauwertangleichung durchführen, die wie ein Gradient über das Bild läuft, von links oben (dunkle Bereiche) nach rechts unten (helle Bereiche), um das Bild homogener zu machen.

Anschließend könnte die Binarisierung vielleicht bessere Ergebnisse liefern.

Fazit

Die Wahl der Lichtquelle und der Winkel Licht-Objekt-Kamera bedarf einer guten Vorbereitung und einigen Versuchen. Das hier präsentierte Ergebnis ist nicht sehr gut. Es ist besonders schwer, durch die Wölbung des Fingers, eine homogene Ausleuchtung der gesamten Oberfläche zu erreichen. Die Bildvorverarbeitung wäre um einiges einfacher, wenn ein gutes Bild als Ausgangsbasis vorliegt.

Aufgabe 1.2

Um die Rillen des Innengewindes einer Schraubenmutter zu überprüfen, würde ich mich für die Fourier-Transformation entscheiden.

Bei einer sauberen Fräsung, sollte ich in der Gewindenfrequenz, und in den Oberwellen dazu große Amplituden erhalten. Weiterhin sollte ich weniger starke Amplituden in den Frequenzen dazwischen erhalten. Je besser das Signal-zu-Rausch Verhältnis ist (also hohe Amplitude in den "guten" Frequenzen, niedrige in den umliegenden Frequenzen dazwischen), desto sauberer der Schliff. Ich sollte dabei jedoch auf die Richtung achten (orthogonal zur Bohrung). Um vielleicht einen Phasenverschiebung bei der Gewindefräsung auszuschließen, sollte

ich ebenfalls die Phase der Fourier Transformation betrachten und mit meiner zu erwartenden Phase abgleichen.

Aufgabe 1.3

Es sollte sowohl mit weißen Gesichtern auf schwarzem Hintergrund, als auch mit schwarzen Gesichtern auf weißem Hintergrund trainiert werden. Dazu könnten z.B. schwarze Gesichter auf weißem Hintergrund einfach invertiert werden (und umgekehrt) um die Anzahl der Trainingsdaten zu erhöhen. Andere Farbkombinationen sollten auch möglich sein, aber das Training könnte vielleicht weniger effizient erfolgen, wenn der zusätzliche Faktor Farbe mit ins Spiel kommt. Ebenfalls sollten auch Bilder im Trainingsdatensatz sein, welche mehr als **ein** Gesicht gleichzeitig zeigen, damit der resultierende Algorithmus sich nicht gezwungen sieht, sich für eines der Gesichter zu entscheiden.

An der Stelle kommt es natürlich ganz drauf an, worauf der Algorithmus trainiert worden ist. Wenn er nur eine binäre Entscheidung treffen soll, ob ein Gesicht vorhanden ist, oder nicht, dann reichen wahrscheinlich Bilder mit je einem Gesicht im Trainingsdatensatz. Wenn auch die Orte von potentiell mehreren Gesichtern bestimmt werden sollen, dann benötigt man auch viele Trainingsdaten mit mehreren Gesichtern.

Die Gesichter sollten auch nicht alle frontal sein, da im hier vorliegenden Testbild die Gesichter eher im Profil gezeigt sind. Ebenfalls sind die Gesichter in ihrer Form abstrahiert. All das muss beim Training berücksichtigt werden.

Aufgabe 1.4 - Gruppenaufgabe

Zunächst spielen für unsere Einordnung des Bildes markante Merkmale des Stiers, wie spitze Hörner, Schweif, längliche Gesichtsform - nach unten hin spitz zulaufend -, Körperform, und Anzahl der Beine eine Rolle. Weiterhin die gesamte Haltung des Körpers, die auf Aggression oder Angriff schließen lässt.

Den Menschen erkennen wir auch sofort an typischen Merkmalen (Kopf, Körper, Arme, Schnitt der Kleidung). Da er eine gerade, gestreckte, selbstbewusste und exponierte Körperhaltung hat, leiten wir ab, dass die Situation gewollt ist, und die Szenerie keine Fluchtsituation darstellt. Die Merkmale des Toreros sind nicht so stark ausgeprägt. Man kann den Torero nur als einen solchen erkennen, da der Stier einen Kontext gibt. Beides zusammen lässt erahnen, dass es sich um einen Stierkampf handelt.

Ein CNN könnte hier versuchen, anhand der Körper-Merkmale der beiden Figuren die einzelnen Figuren zu erkennen. Bei dem Mensch könnte es jedoch schwieriger werden, diesen als solchen zu erkennen, außer es wurde entsprechend trainiert. Für ein CNN könnte der Torero zu abstrakt gezeichnet sein, um als Mensch klassifiziert zu werden. Arme und Beine und der Kopf sind beispielsweise nur angedeutet. Es ist im Gegensatz zum Stier kaum ein eindeutiges Merkmal eines Menschen zu sehen. Der Stier würde von einem CNN wahrscheinlich gut

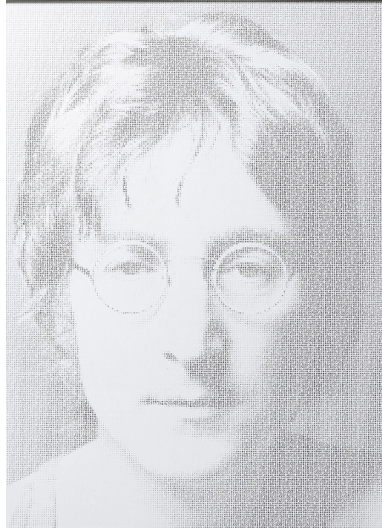


Abbildung 5: John Lennon

erkannt werden - vorausgesetzt es wurde mit entsprechenden Bildern trainiert.

Das Gesamtbild einordnen zu können, käme auf die Art des Netzwerkes und die Daten an, mit denen es trainiert wurde, und, ob es auf Szenen wie z.B. (nennen wir es:) Sportarten oder auf einzelne Objekte trainiert wurde.

Systematisch, um einzelne Objekte zu erkennen, würde man zunächst alle Kanten im Bild segmentieren. Weiterhin Anhand der Helligkeitswerte der Objekte (hier: Stier dunkel, Hintergrund hell) die verschiedenen Teilbereiche des Bildes (z.B. Pixel) zu zusammengehörigen Objekten clustern (=segmentieren).

Anschließend würde man die einzelnen Teilbilder mit bestimmten Templates abgleichen, die man von Objekten hat, und erkennen, um welches Objekt es sich handeln könnte.

Aufgabe 1.5

Wir betrachten die Abbildung 5. Wie können wir daraus ein zusammenhängendes Bild generieren? z.B. durch Weichzeichnen mit einer Faltung mit z.B. einem Gaußfilter oder Mittelwert-Filter. Es könnte aber auch ein "Maximum-Value" Filter verwendet werden, da die dunklen Bereiche anschließend durchgängig dunkel gemacht werden, und die hellen Bereiche durchgängig hell bleiben.

Ein CNN, welches das Ziel hat, das Objekt zu Klassifizieren (z.B. binär: Gesicht / kein Gesicht), würde implizit wahrscheinlich ähnlich vorgehen. Denn ein CNN hat üblicherweise mehrere Faltungsschichten (Convolutional Layers). Da Bilder üblicherweise verrauscht sind, würden einige der Convolution Filters auch während dem Training zu einer Verwaschung des Bildes führen, um dieses als Gesicht später kategorisieren zu können.

Aufgabe 1.6

Interpretationsmöglichkeiten

1. Viereck (Schlangenlinie ist uninteressant)
2. Schlangenlinie (Viereck ist uninteressant)
3. Zusammengesetztes Objekt aus beidem

Zu sehen ist ein zusammengesetztes Bild aus vermutlich 2 Objekten, einem Viereck und einer Schlangenlinie/Kringel. Die Interpretation, was es ist, hängt bei so einem Bild stark vom Kontext ab. Lese ich in einer Zeitschrift über Ergebnisse von Mikroskopieuntersuchungen, interessiert mich wahrscheinlich das schlängenförmige ("lebendige") Objekt. Das Viereck könnte dabei eine Referenzgröße darstellen. Interessiere ich mich dagegen für die Qualitätskontrolle von Produkten der Luftfahrtindustrie (z.B. das Viereck ist ein Teil ein Austauschteil eines Flugzeuges), dann ist das schlängenförmige Objekt vielleicht eher eine Halterung für das Bauteil, und soll nicht weiter untersucht werden.

Als ersten Eindruck - ohne Kontext - würden wir als Menschen aufgrund des Gestaltungsgesetzes "**Gesetz der Kontinuität**" wahrscheinlich die Schlangenlinie als vordergründiges / interessierendes Objekt wahrnehmen. Ein Computer Vision (CV) Algorithmus arbeitet hier i.d.R. anders, und würde eher das Viereck als Objekt von Interesse kategorisieren, da CV Algorithmen typischerweise besser mit geometrischen Objekten, klaren Kanten und Ecken umgehen können.

Jedoch kann das Bild auf viele verschiedene Weise segmentiert werden. Das Bild könnte auch in mehr als 2 Teile segmentiert werden. Aber schon bei 2 Einzelteilen, sind die Kombinationsmöglichkeiten theoretisch sehr vielfältig (siehe Abbildung 6). Wir hätten wahrscheinlich aus o.g. Gestaltungsgesetz kein großes Problem, die "korrekte Segmentierung" (Abbildung 6) durchzuführen. Ein CV Programm aber - je nach dem - schon eher.

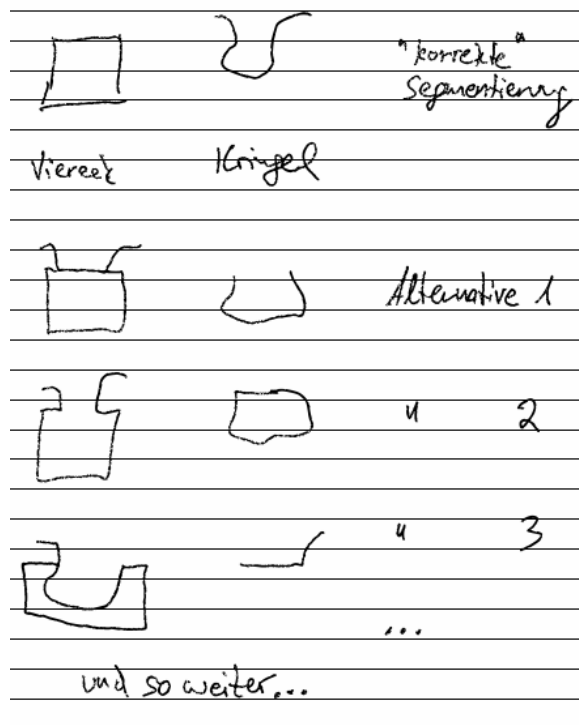


Abbildung 6: Kringel