

Eberhard Karls Universität Tübingen

Mathematisch-Naturwissenschaftliche Fakultät

Wilhelm-Schickard-Institut für Informatik

Dokumentation Praktikum Bildverarbeitung

Automatisches ausschneiden von Bildern aus Fotoalben

Mariella Dreißig, Tom Eckardt, Stefan Bergmann

4. Januar 2019

Betreuer

Andreas Karge, Prof. Andreas Schilling

Mariella Dreißig, Tom Eckardt, Stefan Bergmann:

Automatisches ausschneiden von Bildern aus Fotoalben

Dokumentation Praktikum Bildverarbeitung

Eberhard Karls Universität Tübingen

Zusammenfassung

Inhaltsverzeichnis

1	Einleitung	5
1.1	Problemstellung	5
1.2	Benutzte Technologien	5
1.3	Programmaufbau	5
2	Hintergrundentfernung	7
3	Rahmenentfernung	8
4	Gesichtserkennung	10
4.1	Cascade Classifier	11
5	Auswertung	13
5.1	Metriken zur Auswertung	13
5.1.1	Bildgröße	13
5.1.2	Bildmerkmale	13
5.1.3	Struktureller Ähnlichkeit (SSIM)	14
5.2	Ausgeschnittene Bilder	14
5.3	Gesichtserkennung	15
6	Fazit	20
7	Benutzung und Konfiguration	21
	Literaturverzeichnis	22

1 Einleitung

1.1 Problemstellung

1.2 Benutzte Technologien

Für die Entwicklung dieser Software wurde als Programmiersprache Python 3.6 verwendet zusammen mit der Bibliothek numpy in der Version 1.14.2. Numpy ermöglicht die schnellere Berechnung von Matrizen und wird für die Bibliothek OpenCV in der Version 3.4.0.12 benötigt. OpenCV bietet eine Reihe von Algorithmen für die Bildbearbeitung, die in diesem Programm verwendet wurden. Um eine Auswertung anzufertigen wurde aus der Bibliothek SciKit-image 0.13.1 einige Methoden verwendet.

1.3 Programmaufbau

Das Programm teilt sich in fünf Hauptkomponenten auf. Die erste ist in der *main.py* zu finden, die das Programm startet und die anderen Komponenten ausführt. Die zweite Komponente kümmert sich um das Entfernen des allgemeinen Hintergrunds und ist in der *backgroundremover.py* zu finden. In dieser Komponente werden die eigentlichen Fotos aus dem Fotoalbum grob ausgeschnitten. Danach wird in der nächsten Komponente für jedes Foto der Rahmen entfernt falls ein Rahmen vorhanden ist. Das Entfernen des Rahmens geschieht in der *rectextract.py*. Die letzte Komponente ermöglicht es Gesichter in den ausgeschnittenen Bildern zu erkennen, dies geschieht in der *facedetection.py*. Möchte man die ausgeschnittenen Bilder mit Beispielbildern vergleichen, so kann man die letzte Komponente in der Datei *compare.py* verwenden. Mit deren Hilfe Metriken aufgestellt werden zum Vergleich der Bilder. In dem Flussdiagramm 1.1 ist der genaue Ablauf des Programmes dargestellt.

1 Einleitung

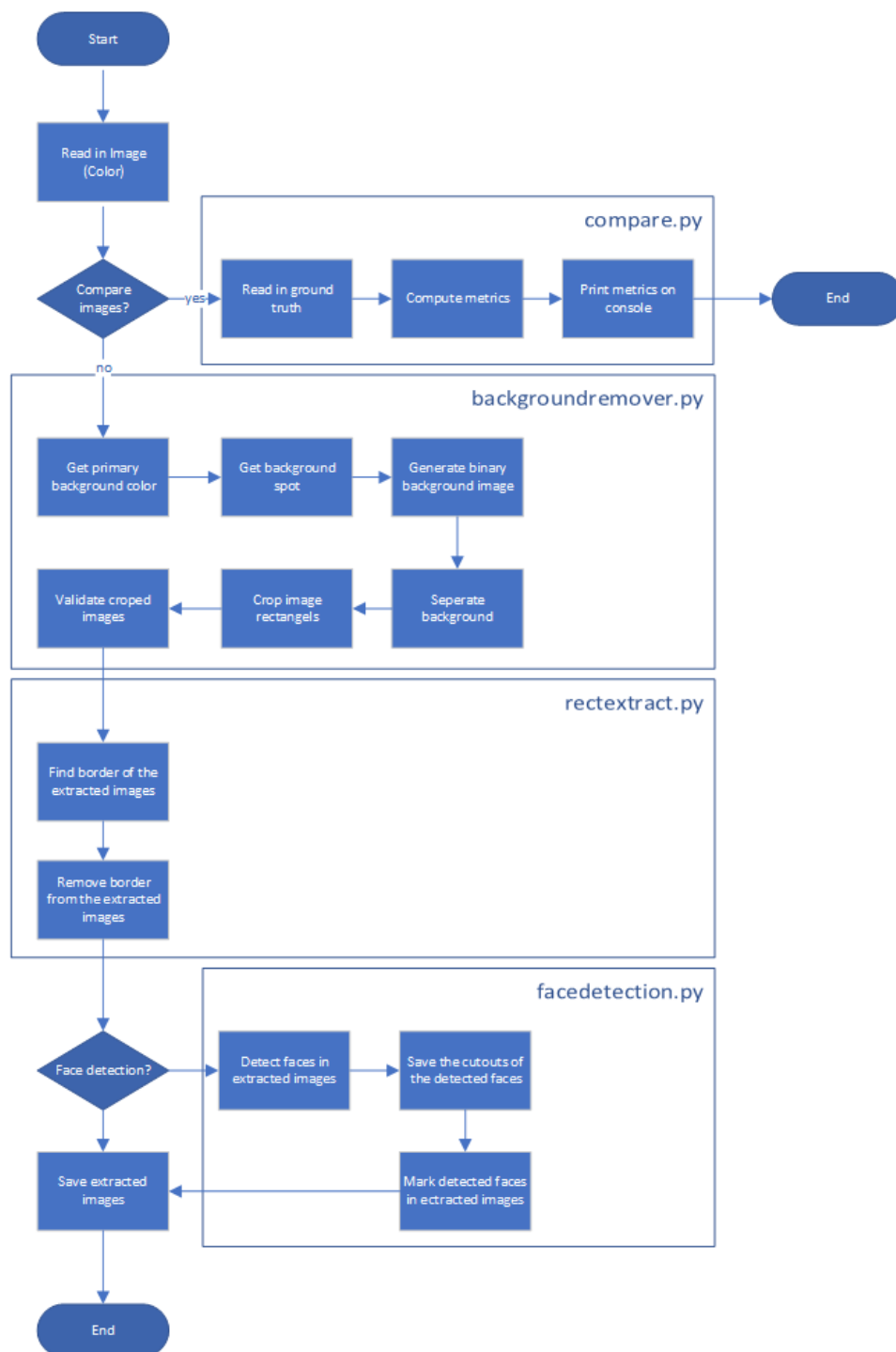


Abbildung 1.1: Flussdiagramm, das den Ablauf des Programmes darstellt. Verzweigungen können mit Hilfe von Parametern bei der Ausführung des Programmes gesteuert werden.

2 Hintergrundentfernung

3 Rahmenentfernung

$$G = \frac{1}{159} \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix}, C_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, C_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix},$$

Abbildung 3.1: Die Figur zeigt den Gausfilterkernel sowie die Faltungsmasken, die beim Canny-Algorithmus verwendet werden

Die Extraktion eines Bildes aus einem monotonen Rahmen geschieht in mehreren Schritten. Zu Beginn wird das gesamte Bild mit dem Canny-Kantendetektionsalgorithmus in ein binär Bild umgewandelt, dass sämtliche Pixel weiß darstellt bei denen eine hoher unterschied in der Intensität festzustellen ist. Dazu wird in dem Bild rauschen durch einen Gaußfilter entfernt dessen Kernel in Abbildung 3.1 zu erkennen ist. Um dann die Gradienten des Bildes mit zwei Faltungsmasken zu bestimmen, die auch in Abbildung 3.1 dargestellt sind. Die so gewonnen Gradienten werden durch eine Non-Maximum suppression überprüft, ob sie teil einer Kante sind. Daraus resultieren nur dünne Linien ([3]). Bei anderen Kantendetektoren wird dieser Schritt nicht ausgeführt, was im späteren dazu führen kann, dass es schwerer wird den Rahmen des Bildes zu erkennen.

Das Ergebnis des Canny-Kantendetektionsalgorithmus hat dann die Eigenschaft, dass stellen des Bildes mit einer monotonen Fläche nur schwarz sind und dort keine starken Gradienten durch weiße Pixel markiert wurden. Mit Hilfe dieser Eigenschaft kann man einen monotonen Rahmen um das gegebene Bild erkennen. Hierzu betrachtet man jede Seite des Bildes einzeln und betrachtet für jede Seite einen bestimmten Ausschnitt. Das erste Fenster, das betrachtet wird, hat die Höhe beziehungsweise Breite der gewählten Seite des Bildes und geht eine gegebene Anzahl von Pixeln in das Bild. Nun kann man innerhalb dieses Fensters die weißen Pixel zählen. Nachdem dies geschehen ist vergrößert man den Ausschnitt indem man weitere reihen an Pixeln des Bildes hinzunimmt und man wieder die weißen Pixel zählt. Im Schaubild 3.2 sind die Ergebnisse eines solchen Vorgehens für ein Bild mit monotonem Rahmen dargestellt.

3 Rahmenentfernung

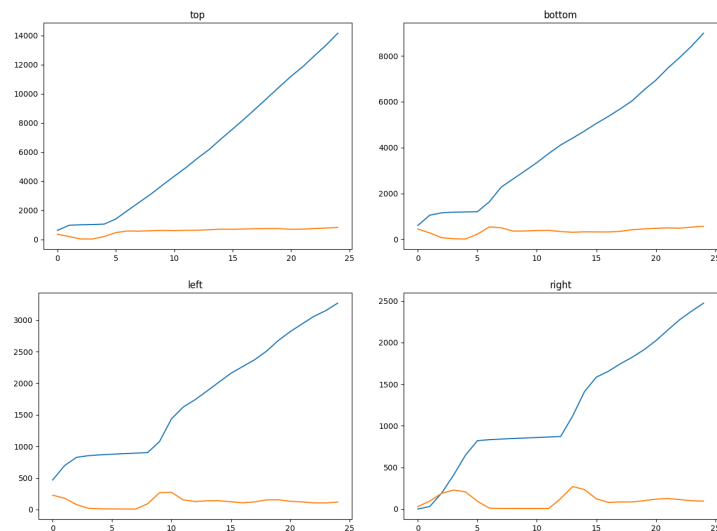


Abbildung 3.2: Die Diagramme zeigen in der oberen Linie die Zunahme der weißen Pixel in den immer größer werdenden Fenstern. Außerdem zeigen sie die jeweiligen Gradienten zur vorhergehenden Anzahl an weißen Pixeln in der darunter liegenden Linie.

Man kann deutlich erkennen, dass die Steigung deutlich abnimmt sobald der Rahmen erreicht ist. Direkt nach dem Rahmen nimmt die Steigung wieder zu. Der Rahmen muss sich dort befinden wo der Gradient am geringsten ist und endet an dem Punkt, wo der Gradient einen Ausschlag nach oben macht. Dies kann man an den gezeichneten Gradienten in dem Schaubild 3.2 erkennen. Um danach den Rahmen zu entfernen schneidet man das Bild an dieser Stelle aus.

4 Gesichtserkennung

Die Aufgabenstellung beinhaltet, in den extrahierten Fotos Gesichter zu erkennen. Hier muss die Gesichtserkennung (in der Fachliteratur "Face Detection") von der Wiedererkennung von bereits identifizierten Gesichtern (in der Fachliteratur "Face Recognition") unterschieden werden. In der vorliegenden Arbeit geht es ausschließlich darum, die Position von Gesichtern in den eingescannten Fotos zu erkennen und nicht, bereits erkannte Gesichter wiederzuerkennen oder zuzuordnen ¹.

In der klassischen Bildverarbeitung wurden lange einfache 2D-Verfahren verwendet, um Gesichter in Bildern zu erkennen ([1], [8]). Beim Template Matching werden beispielsweise Vorlagen (von Teilen) von Gesichtern erzeugt um diese dann im Anwendungsfall mit den Zielbildern zu vergleichen. Dieses Verfahren ist nicht sehr robust, besonders bei sich verändernden Lichtverhältnissen, aber gleichzeitig extrem rechenaufwändig.

Eine einfachere Methode der Gesichtserkennung in der klassischen Bildverarbeitung ist die Nutzung von geometrischen Merkmalen in den Gesichtern. Hierbei werden beispielsweise Grauwerte in einem bestimmten Bereich des Bildes aufsummiert oder Gradientenoperatoren zur horizontalen und vertikalen Kantenerkennung angewandt. So kann die Ausrichtung, die relative Lage und ähnliche Merkmale der Nase, der Augen und des Mundes identifiziert und zusammen als Gesicht klassifiziert werden.

Doch auch diese Methode ist rechen- und zeitintensiv. Deswegen wird häufig eher auf komplexere Methoden zur Gesichtserkennung gesetzt, wie beispielsweise das Elastic Bunch Graph Matching, das eine robuste Waveletanalyse nutzt ([9]) oder die auf der Hauptkomponentenanalyse basierende Eigenfaces-Methode ([6]). Weiterentwicklungen dieser beiden Methoden werden heute noch eingesetzt, obwohl die Ansätze schon vergleichsweise alt sind.

Im Folgenden wird ein populärer Ansatz vorgestellt, der in den vorliegenden Arbeit verwendet wurde.

¹ Aufgrund der geringen Datenmenge und der schlechten Qualität der alten Fotos ist es nicht möglich, Gesichter wiederzuerkennen.

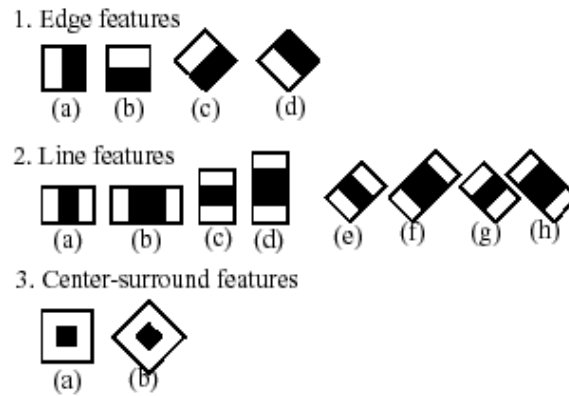


Abbildung 4.1: Alle vom Cascade Classifier Algorithmus genutzten Haar-ähnlichen Features zur Objekterkennung. Zur Gesichtserkennung werden nur fünf Features genutzt: 1(a), 1(b), 2(a), 2(c) und eine Kombination aus zwei 1(a) Features, die zusammen ein schachbrettartiges Muster ergeben.

4.1 Cascade Classifier

Die OpenCV Programmibliothek beinhaltet vorimplementierte Funktionen zur Gesichtserkennung. Sie basiert auf einem 2001 entwickelten Ansatz zur Objekterkennung in Bildern ([7], [2]). Es ist ein trainierter Classifier, der einfache geometrische Merkmale im Bild erkennt und auf deren Basis die gewünschten Objekte erkennen kann ([4]).

Die Methode nutzt sogenannte Haar-ähnliche Features, mit denen grundlegende Strukturen (Helligkeitsunterschiede) im Bild erkannt werden. Im Bild 4.1 werden alle Features aufgelistet, die der Classifier nutzen kann. Für die Gesichtserkennung werden in der Praxis aber nicht so viele genutzt, sondern nur einige wenige. Mit diesen können allgemeine Eigenschaften von menschlichen Gesichtern repräsentiert werden. So ist zum Beispiel die Nase heller als die Bereiche um die Augen, das entspricht dem Feature 2(a) in Bild 4.1.

Nachdem die zu nutzenden Features ausgewählt wurden, wird ein Integralbild erzeugt, das in der Lage ist, die Features in konstanter Zeit auszuwerten. Dadurch wird eine relativ geringe Laufzeit erzielt.

In der Trainingsphase wird der Classifier mit Positiv- und Negativbeispielen von dem zu erkennenden Objekt trainiert, in diesem Fall mit Gesichtern². Dieser Vorgang wird mit verschiedenen Arten des AdaBoostings beschleunigt und verbessert. Dadurch werden die besten Features für die Klassifizierung identifiziert.

²In der OpenCV Bibliothek ist sowohl ein Trainer als auch ein Detector enthalten. Der Detector ist vortrainiert kann direkt zur Gesichtserkennung verwendet werden. Wenn man genug Daten zur Verfügung hat, kann man aber auch einen selbst trainieren.

4 Gesichtserkennung

Danach wird der Classifier in mehreren Stufen auf die Bilder angewandt ("cascade"), in denen das Objekt identifiziert werden soll. Jeder Classifier funktioniert dabei wie ein Filter, der die negativen Regionen aus dem Bild herausfiltert. Die verbleibenden Bereiche werden Stück für Stück durch immer komplexere Classifier gegeben, bis sie entweder aussortiert werden oder das Objekt (ein Gesicht) identifiziert wurde. Somit hat der Cascade Classifier die Struktur eines Entscheidungsbaumes und Rechenzeit wird gespart. Die Methode ist so robust, dass das Objekt in verschiedenen Größen in dem Bild erkennen kann, unabhängig von der Größe der Bilder, mit der der Classifier trainiert wurde.

Wenn die OpenCV Methode `detectMultiScale` ein Gesicht im Input-Bild erkennt, gibt es die Position in x- und y-Koordinaten und dessen Abmessungen (Höhe, Breite) zurück. In dem vorliegenden Programm werden zwei verschiedene Classifier auf die Bilder angewandt: einmal frontale Gesichter und einmal Gesichter im Profil. Die Rückgabewerte der Funktionen werden dafür genutzt, die Gesichter auszuschneiden und Dopplungen zu erkennen. Da es passieren kann, dass beide Classifier die gleichen Gesichter detektieren, wird durch eine Hilfsfunktion festgestellt, ob dieses Gesicht bereits in der Liste der erkannten Gesichter vorhanden ist. Dafür wird überprüft, ob der Mittelpunkt des kleineren Ausschnitts in dem größeren liegt. Wenn das der Fall ist, kann angenommen werden, dass es sich um das gleiche Gesicht handelt und eines davon wird verworfen. So wird sichergestellt, dass keine Gesichter verworfen werden, die lediglich nah beieinander liegen, so wie beispielsweise in einem Gruppenbild.

5 Auswertung

5.1 Metriken zur Auswertung

5.1.1 Bildgröße

Da es in dem Projekt um das Ausschneiden von Bildern geht ist die Bildgröße eine nahe liegende Metrik um die Genauigkeit unserer Software zu messen. Die Größe wird berechnet indem die Höhe mit der Breite multipliziert wird. Von der so ermittelten Größe des ausgeschnittenen Bildes wird die Größe des Ground Truth Bildes abgezogen. Aus dieser Differenz wird der Prozentuale unterschied zum Ground Truth Bild berechnet. Sollte das Ergebnis negativ sein ist das ausgeschnittene Bild kleiner als der Ground Truth.

5.1.2 Bildmerkmale

Eine weitere Metrik ist das erkennen gleicher Bildmerkmale. Dazu wird der patent freie ORB (Orientated FAST and Rotated BRIEF) detektor von OpenCV verwendet der Rotations invariant ist. ORB ist eine fusion zwischen dem FAST Schlüsselpunkt Detektor and BRIEF Beschreibung der Schlüsselpunkte mit vielen Modifikationen um die Geschwindigkeit des Algorithmus zu verbessern ([5]). Die so erkannten Merkmale werden dann der reihe nach anhand der Distanz zum Bildzentrum verglichen. Es wird für jedes erkannte Bildmerkmal die Distanz zum Zentrum des dazu gehörigen Bildes berechnet und das dann mit den Ergebnissen des Ground Truth verglichen. Solange zwischen den beiden Merkmalen keine zu großer Unterschied in der Distanz ist werden sie als gleich angesehen. Durch dieses Verfahren kann erkannt werden das sich der Ausschnitt des ausgeschnittenen Bildes an der selben stelle befindet, wie der Ground Truth und nicht nur die gleiche Größe ausgeschnitten wurde.

5.1.3 Struktureller Ähnlichkeit (SSIM)

Um ein Bild auf Gleichheit zu untersuchen verwenden wir den Strukturellen Ähnlichkeitsindex (Structural Similarity index, SSIM). Dieser berechnet die Ähnlichkeit zwischen verschiedenen Bildausschnitten mit folgender Gleichung:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

mit μ als Durchschnitt und σ als Varianz bzw. Kovarianzmatrix des Bildausschnittes. c sind da um die Division zu stabilisieren ([10]). Das Verfahren lässt sich nur auf gleich große Bilder anwenden. Aus diesem Grund verkleinern wir die Bilder mit einer linearen Interpolation auf eine Größe von 64 x 64 Pixeln. Auf die so verkleinerten Bilder wird dann der SSIM angewandt. Um so näher das Ergebnis an 1 ist um so ähnlicher sind die beiden Bilder.

5.2 Ausgeschnittene Bilder

In der Tabelle 5.1 sind die Ergebnisse für einige der Testbilder dargestellte den direkten Vergleich kann man in Abbildung ?? sehen dort sind die Bilder nebeneinander aufgezeigt.

Bild	Größe	Merkmale	SSIM
01	1.92	3.64	0.82
02	1.60	2.39	0.77
03	-5.52	10.12	0.36
04	3.15	16.54	0.68
05	1.48	2.65	0.81
06	-4.42	23.58	0.55
07	2.96	4.40	0.80
08	8.82	8.07	0.39
09	-1.24	2.06	0.70
10	-0.58	5.57	0.76
11	-4.35	28.46	0.49
12	2.96	1.74	0.80
13	1.26	3.24	0.88
14	3.01	1.94	0.72
15	0.76	19.51	0.60

Tabelle 5.1: Ergebnisse der Vergleichsmetriken der ausgeschnittenen Bilder.

5.3 Gesichtserkennung

Die von OpenCV mitgelieferten Methoden zur Gesichtserkennung auf Bildern funktioniert grundsätzlich gut. Sie ist einfach anwendbar, hat eine vertretbare Laufzeit (der Algorithmus läuft nur einige Sekunden auf einer ganzen eingescannten Albumseite) und erkennt zumindest einen Großteil der Gesichter in den Fotos.

Um die Performanz der Classifier einzeln und in Kombination auszuwerten, wurden Wahrheitsmatrizen aufgestellt. Ausgewertet werden jeweils die korrekt erkannten Gesichter (True Positives (TP): positive - positive), die nicht erkannten Gesichter (False Negatives (FN): positive - negative) und die fälschlicherweise erkannten Gesichter (False Positives (FP): negative - positive). Korrekt nicht erkannte Gesichter (True Negatives (TN): negative - negative) gibt es in unserem Anwendungsfall nicht.

Tabellen 5.2 und 5.3 zeigen die Metriken in dem ersten und dem zweiten Datensatz wenn nur der frontal faces Classifier angewendet wird. Die daraus errechnete Genauigkeit ¹ beträgt im ersten Datensatz 0.75 und im zweiten 0.59.

Wenn der zweite Classifier zur Erkennung von Profilen mit dazu genommen wird, erhöht sich

		Reference	
Prediction		Positive	Negative
	Positive	270	1
	Negative	88	-

Tabelle 5.2: Wahrheitsmatrix im ersten Datensatz, wenn nur der frontal-Classifier zum Einsatz kommt. Die Genauigkeit der Gesichtserkennung beträgt 0.75.

		Reference	
Prediction		Positive	Negative
	Positive	282	7
	Negative	172	-

Tabelle 5.3: Wahrheitsmatrix im zweiten Datensatz, wenn nur der frontal-Classifier zum Einsatz kommt. Die Genauigkeit der Gesichtserkennung beträgt 0.59.

die Anzahl der erkannten Gesichtern ein wenig. Das ist in Tabellen 5.4 und 5.5 zu sehen. Die Gesichter, die bereits von dem frontal-Classifier erkannt wurden, werden automatisch heraus gefiltert. Danach bleibt ein kleiner Mehrgewinn durch die Anwendung des profile-Classifiers übrig, die Performanz verbessert sich auf 0.76 und 0.65 jeweils.

¹Formel für die Berechnung der Genauigkeit ("Accuracy") ist: $\frac{TP+TN}{TP+FN+TN+FP}$, ($TN = 0$ in unserem Fall).

5 Auswertung

		Reference	
		Positive	Negative
Prediction	Positive	274	2
	Negative	84	-

Tabelle 5.4: Wahrheitsmatrix im ersten Datensatz, wenn frontal- und profile-Classifer kombiniert werden. Die Genauigkeit der Gesichtserkennung beträgt 0.76.

		Reference	
		Positive	Negative
Prediction	Positive	301	8
	Negative	153	-

Tabelle 5.5: Wahrheitsmatrix im zweiten Datensatz, wenn frontal- und profile-Classifer kombiniert werden. Die Genauigkeit der Gesichtserkennung beträgt 0.65.

Festlegung der Ground Truth

Die Referenz für die Auswertung der Gesichtserkennung festzulegen ist nicht trivial. Auf den alten Fotos sind aufgrund der schlechten Qualität der Aufnahmen viele Gesichter nicht tatsächlich zu erkennen. Vielmehr erschließt sich ein menschlicher Betrachter aus dem Kontext die Information, dass sich an gewissen Stellen ein Gesicht befinden muss. Das wurde bei der Festlegung der Ground Truth berücksichtigt. So wurden beispielsweise besonders kleine, verdeckte oder stark verschwommene Gesichter für die Berechnung der Metriken nicht berücksichtigt. Die Classifier haben außerdem Probleme, Gesichter zu erkennen, wenn starke Okklusionen auftreten, selbst wenn das Gesicht nur im Profil zu erkennen ist oder die Person eine Kopfbedeckung trägt. Diese wiederum wurden in die Gesamtsumme der Gesichter mit einberechnet. Im Folgenden werden exemplarisch einige Bilder gezeigt, die der Gesichtserkennungsalgorithmus nicht erkennen kann und die deswegen nicht in die Auswertung mit einbezogen wurden.

Bild 5.1 zeigt ein kleines Mädchen, jedoch ist die Aufnahme sehr verschwommen. Jeder Mensch, der das Bild betrachtet, weiß, dass das Mädchen ein Gesicht haben muss. Wenn jedoch die Kontextinformationen entfernt werden würde, wäre es selbst für einen Menschen schwer, ein Gesicht zu identifizieren.

Auf Bild 5.2 ist eine ähnliche Situation zu sehen: alle abgebildeten Menschen haben Gesichter, jedoch sind sie objektiv nicht wirklich zu sehen. Sie sind relativ klein, was normalerweise für den Classifier kein Problem darstellt, aber eine schlechte Beleuchtung und geringer Kontrast tragen dazu bei, dass die Gesichter nicht erkannt werden.

Um dennoch eine repräsentative Auswertung der Performanz der Gesichtserkennung zu gewährleisten, wurden viele nicht erkannte Gesichter mit in die Metriken einberechnet. Beispielsweise



Abbildung 5.1: Starke Bewegung im Bild verhindert, dass das Gesicht des Mädchens tatsächlich zu sehen ist, auch wenn menschliche Betrachter durchaus ihr Gesicht erkennen könnten.

können die Classifier keine Gesichter mit Kopfbedeckungen erkennen. Ansonsten sind aber die Gesichter gut zu identifizieren, was in Bild 5.3 zu sehen ist.

Bild 5.4 zeigt einen problematischen Grenzfall aus dem zweiten Datensatz. In diesem Bild wurden insgesamt drei Gesichter erkannt (zwei vom frontal-Classifier (grün) und eins vom profile-Classifier (rot)). Da alle Gesichter in etwa die gleiche Größe und Schärfe haben, sollte angenommen werden, dass auch die restlichen Personen erkannt werden. Es ist nicht ersichtlich, warum auf diesem Bild beide Classifier keine weiteren Gesichter erkennen. Von daher fließen alle Gesichter auf diesem Foto mit in die Bewertung ein und erhöhen die Anzahl der False Negatives enorm. Besonders im zweiten Datensatz kommen solche Fälle häufiger vor, was die schlechte Performanz des Algorithmus' im Vergleich zum ersten Datensatz erklärt.

5 Auswertung



Abbildung 5.2: Die Gesichtserkennung schlägt fehl, wenn die Gesichter zu schlecht beleuchtet sind und/oder zu wenig Kontrast aufweisen. Auch, wenn normalerweise auch Gesichter in dieser Größe von dem Algorithmus problemlos erkannt werden.



Abbildung 5.3: Kein Classifier erkennt Gesichter mit Hüten, auch wenn sie nur einen kleinen Teil vom Gesicht verdecken. Alle anderen Gesichter werden problemlos identifiziert.



Abbildung 5.4: In diesem Gruppenfoto werden nur wenige Gesichter erkannt, obwohl sie alle ähnlich aussehen. Das verschlechtert die Kennwerte für die Genauigkeit besonders im zweiten Datensatz stark.

6 Fazit

Das vorliegende Programm extrahiert einzelne Fotos aus eingescannten Albenseiten mithilfe von Methoden aus der klassischen Bildverarbeitung. Das passiert in zwei Schritten. Im ersten Schritt werden die Fotos mithilfe der Hintergrundfarbe und einem Binärbild grob ausgeschnitten. Diese grob ausgeschnittenen Bilder haben allerdings oft noch einen Rahmen, der dann im zweiten Schritt entfernt wird.

Das Programm wurde auf zwei Fotoalben angewandt mit jeweils 32 und 49 Seiten. Insgesamt wurden 126 von 136 Bildern (92.65 %) im ersten und 192 von 211 Bildern (91.00 %) im zweiten Album erfolgreich extrahiert. Die Performanz des Programms kann anhand von verschiedenen Parametern ausgewertet werden.

Die Maße des Bildes (Bildgröße), die Bildmerkmale (ORB) und die strukturelle Ähnlichkeit (SSIM) verglichen mit der Ground Truth zeigen, wie erfolgreich der Code ist. Besonders der SSIM ist ein gutes Indiz für den Erfolg des Algorithmus, und in den meisten Bildern werden gute Werte erzielt. Insgesamt lässt sich sagen, dass in dem Datensatz die meisten Fotos erkannt und gut ausgeschnitten werden, bis auf einige Ausnahmen. In diesen Fällen sind oft schlechte Lichtverhältnisse oder ungewöhnliche Formen der Fotos der Grund dafür.

Optional kann das Programm menschliche Gesichter erkennen und extrahieren. Dafür wird ein einfacher machine learning Ansatz verwendet. Dieser erkennt erfolgreich die meisten Gesichter, trotz des Alters der Fotos und der mangelnden Qualität. Dennoch werden einige Gesichter nicht erkannt, meist, wenn sie zu klein und verschwommen sind oder die abgebildete Person eine Kopfbedeckung trägt.

In kommenden Arbeiten könnte das Programm verbessert werden, indem Sonderfälle wie ovale oder schlecht belichtete Fotos anders behandelt werden. Der aktuelle Stand kann gut für die Extraktion von Fotos aus eingescannten Albumseiten verwendet werden, ohne dass viel manuelle Nachbearbeitung nötig ist. Das Programm wurde vollständig in Python unter der Verwendung von OpenCV geschrieben.

7 Benutzung und Konfiguration

Literaturverzeichnis

- [1] Kees, Benjamin: *Gesichtserkennung*, 2012. <https://www.algoropticon.de/Dateien/Seminararbeit{ }Gesichtserkennung.pdf>.
- [2] Lienhart, Rainer, Alexander Kuranov, Vadim Pisarevsky und M R L Technical Report: *Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection*. Technischer Bericht, 2002. <http://www.multimedia-computing.de/mediawiki//images/5/52/MRL-TR-May02-revised-Dec02.pdf>.
- [3] OpenCV: *Canny Edge detector*. https://docs.opencv.org/3.4/da/d5c/tutorial_canny_detector.htmls, besucht: 2018-12-19.
- [4] OpenCV: *Haar Feature-based Cascade Classifier for Object Detection*. <https://docs.opencv.org/2.4/modules/objdetect/doc/cascade{ }classification.html{#}grouprectangles>, besucht: 2018-12-09.
- [5] OpenCV: *ORB: An efficient alternative to SIFT or SURF*. https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_orb/py_orb.html, besucht: 2018-12-19.
- [6] Turk, Matthew und Alex Pentland: *Eigenfaces for Recognition*. Journal of Cognitive Neuroscience, 3(1):71–86, 1991. <https://doi.org/10.1162/jocn.1991.3.1.71>.
- [7] Viola, Paul: *Rapid Object Detection using a Boosted Cascade of Simple Features*. In: *Conference on Computer Vision and Pattern Recognition*, 2001. <http://wearables.cc.gatech.edu/paper{ }of{ }week/viola01rapid.pdf>.
- [8] Wächter, Carsten und Stefan Römer: *Gesichtserkennung I - Biometrische Systeme*, 2001. <http://www.informatik.uni-ulm.de/ni/Lehre/WS01/HS-Biometrische-Systeme/ausarbeitungen/Gesichtserkennung{ }1{ }block.pdf>.

- [9] Wiskott, Laurenz, Norbert Kr und Christoph Von Der Malsburg: *Face Recognition by Elastic Bunch Graph Matching*. Intelligent Biometric Techniques in Fingerprint and Face Detection, (11):355–396, 1999. <http://www.face-rec.org/algorithms/ebgm/wisfelkrue99-facerecognition-jainbook.pdf>.
- [10] Z. Wang, A. C. Bovik, H. R. Sheikh und E. P. Simoncelli: *Image quality assessment: From error visibility to structural similarity*. IEEE Transactions on Image Processing, 13(4):600–612, 2004. <http://www.cns.nyu.edu/pub/lcv/wang03-preprint.pdf>.