Masterarbeit am Institut für Informatik der Freien Universität Berlin
Arbeitsgruppe Künstliche Intelligenz

Erweiterung von Fast Retina Keypoints um Farbinformation

Anselm Brachmann

brachman@mi.fu-berlin.de

Eingereicht bei: Prof. Dr. Raúl Rojas

Betreut durch: Dr. Manfred Hild

Berlin, 11. Oktober 2013

Zusammenfassung

Die Arbeit beschäftigt sich mit dem *Fast Retina Keypoints*-Deskriptor (kurz FREAK), welcher ermöglicht, lokale Merkmale von Bildern so zu beschreiben, dass diese in anderen Bildern wiedergefunden werden können. Dabei verwirft dieser jegliche Farbinformation und nimmt die Beschreibung auf einer in Grauwerte konvertierten Version eines Eingabebildes vor. Beruhend auf der Annahme, dass Farbe zur Beschreibung von Bildern wichtige Informationen trägt, wird in dieser Arbeit geprüft, ob eine Hinzunahme der Farbe die Leistung des Deskriptors verbessern kann. In verschiedenen Experimenten werden dazu zwei Ansätze entworfen und anschließend evaluiert. Anhand einer konkreten Anwendung, in welcher Objekte in einem Kamerabild identifiziert werden sollen, werden die gewonnenen Erkenntnisse anschließend auf Relevanz überprüft. Abschließend kann so eine Empfehlung ausgesprochen werden, wie eine Erweiterung des FREAK-Deskriptors um Farbinformation möglich ist und unter welchen Umständen diese gewinnbringend sein kann.

Inhaltsverzeichnis

1 Einleitung **1**

2 Bildbeschreibung anhand lokaler Merkmale **5**

2.1 Einordnung des Verfahrens der Beschreibung lokaler Merkmale 5

2.2 Anforderungen an das Verfahren 7

2.3 Aufbau des Gesamtsystems 8

2.4 Detektion von Schlüsselpunkten 12

2.4.1 Integralbilder 12

2.4.2 Der Fast-Hessian Detektor 13

2.4.3 Weitere Detektoren 22

2.5 Deskriptoren 24

2.5.1 Reellwertige Deskriptoren 24

2.5.2 Die binären Deskriptoren BRIEF und BRISK 28

2.5.3 Der Fast Retina Keypoints Deskriptor 32

2.6 Vergleichsstrategien und Verifikation 38

2.7 Evaluation von Deskriptoren 40

3 Farbe und Farbräume **45**

3.1 Farbe und deren Wahrnehmung 45

3.2 Darstellung von Farbe durch Farbräume 46

3.3 Übertragungen der SIFT nach Farbe 53

4 Experimente **57**

4.1 Der Datensatz zur Evaluation 57

4.2 Farbe integriert durch Konkatenation von Deskriptorvektoren 60

4.2.1 Getestete Deskriptoren 60

4.2.2 Durchführung der Experimente 60

4.2.3 Ergebnisse 62

4.3 Farbe als zusätzliches Attribut 69

4.3.1 Extraktion und Vergleich von Farbe 69

4.3.2 Durchführung der Experimente 72

4.3.3 Ergebnisse 73

5 Praktische Anwendung **79**

5.1 Auswahl der Objekte 79

5.2 Durchführung des Experimentes 80

5.3 Ergebnisse 82

6 Diskussion der Ergebnisse **87**

6.1 Vergleich verschiedener Farbdarstellungen 87

6.2 Die Eignung von Farbabständen zum Vergleich 91

6.3 Die Bedeutung von Beleuchtung und Aufnahmequalität 92

7 Schlussbetrachtung **95**

Literatur **101**

A Anhang **103**

Abkürzungen

AGAST	Adaptive and Generic Accelerated Segment Test
BRIEF	Binary Robust Independent Elementary Features
BRISK	Binary Robust Invariant Scalable Keypoints
CIE	Commission internationale de l'éclairage
FAST	Features from Accelerated Segment Test
FREAK	Fast Retina Keypoints
NNDR	Nearest Neighbor Distance Ratio
ORB	Oriented FAST and Rotated BRIEF
SIFT	Scale-invariant Feature Transform
SURF	Speeded-Up Robust Features

1 Einleitung

Der Mensch kann sich ohne Schwierigkeiten in einer dreidimensionalen Welt orientieren. Er nimmt ihn umgebende Objekte wahr und erfreut sich am Anblick monumentaler Landschaften in der Ferne und am Farbenreichtum der Blüten in seiner direkten Umgebung. Dazu befähigt wird er durch seinen Sehsinn, welcher wahrgenommene Lichtreize derart verarbeitet, dass semantisch gehaltvolle Informationen daraus gewonnen werden. Die physiologische Funktionsweise des Auges ist größtenteils bekannt, genauso wie die Tatsache, dass elektrische Potentiale zur Informationsweiterleitung und -verarbeitung im Gehirn genutzt werden. Dennoch bleibt die Fähigkeit des Menschen unübertroffen bei dem Versuch, diese maschinell nachzubilden.

Auf dem Gebiet des maschinellen Sehens wird an Algorithmen geforscht, aus den Rohdaten aufgezeichneter Bilder semantisch wertvolle Informationen zu gewinnen. Ein universelles Verfahren, welches mit dem Menschen vergleichbare Leistungen erreicht, existiert jedoch nicht. Vielmehr wird das Problem visueller Wahrnehmung in Teilprobleme zerlegt, welche stark begrenzte Anwendungsfälle definieren: Eines ist beispielsweise das Auffinden von Gesichtern, ein anderes das Identifizieren von Objekten, welche in einem Bild zu sehen sein könnten. Eine Möglichkeit bekannte Objekte eines Bildes zu identifizieren ist es, diese durch *lokale Merkmale* zu beschreiben, womit sich die vorliegende Arbeit beschäftigt.

Für die Beschreibung lokaler Merkmale eines Bildes existieren verschiedene Verfahren. Die bekanntesten sind die von Lowe (2004) vorgestellte *Scale-invariant Feature Transform* (SIFT) und das wesentlich schnellere *Speeded-Up Robust Features* (SURF) Verfahren von Bay et al. (2008). Es handelt sich dabei um sogenannte *reellwertige Deskriptoren*, welche als etablierte Ansätze weite Verbreitung finden.

Binäre Deskriptoren ermöglichen ebenfalls die Beschreibung lokaler Merkmale eines Bildes und übertreffen die reellwertigen Verfahren sowohl bezüglich der Leistung als auch der Geschwindigkeit. Die Entwicklung binärer Deskriptoren wurde durch die Veröffentlichung des *Binary Robust Independent Elementary Features* (BRIEF) Deskriptors von Calonder et al. (2010) eingeleitet und fand ihren vorläufigen Höhepunkt mit *Fast Retina Keypoints* (FREAK) von Alahi et al. (2012). Der FREAK-Deskriptor steht im Mittelpunkt dieser Arbeit.

Fragestellung und Zielsetzung

Der FREAK-Deskriptor ermöglicht, lokale Merkmale von Bildern derart zu beschreiben, dass diese in anderen Bildern wiedergefunden werden können. Dabei verwirft der Algorithmus jegliche Farbinformation und nimmt die Beschreibung auf einer in Grauwerte konvertierten Version eines Eingabebildes vor. Beruhend auf der Annahme, dass Farbe zur Beschreibung von Bildern wichtige Informationen beitragen kann, wird in dieser Arbeit geprüft, ob sich die Leistung des FREAK-Deskriptors durch eine Hinzunahme von Farbinformation verbessern lässt. Zum Zeitpunkt der Erstellung der Arbeit ist hierfür kein Verfahren bekannt.
Die Untersuchung soll dabei zunächst theoretisch erfolgen, wobei zwei verschiedene Ansätze der Farberweiterung vorgeschlagen und evaluiert werden. Die theoretisch gewonnenen Erkenntnisse werfen jedoch die Frage auf, inwiefern diese in einer konkreten Anwendung von Relevanz sind. Ein weiteres Ziel der vorliegenden Arbeit ist demzufolge, in einer konkreten Anwendung zu prüfen, wie die theoretisch gewonnenen Erkenntnisse in der Praxis zu bewerten sind.

Aufbau der Arbeit

Die Arbeit gliedert sich in einen Grundlagenteil, einen experimentellen Teil und eine abschließende Diskussion, der Aufbau ist in Abbildung 1.1 dargestellt. Der Grundlagenteil setzt dabei zwei Schwerpunkte. Zunächst erfolgt die Darstellung des Verfahrens, Bilder anhand lokaler Merkmale zu beschreiben. Beginnend mit der Einordnung des Verfahrens auf dem Gebiet des Maschinellen Sehens wird im Anschluss ein Überblick über dessen Aufbau gegeben, da sich das Verfahren in verschiedene Schritte gliedern lässt. Die einzelnen Schritte werden jeweils in eigenen Abschnitten erläutert, sodass am Ende der gesamte Prozess dargestellt wurde. Der Teilschritt der *Beschreibung*, in welchem sich auch der FREAK-Deskriptor verortet, findet dabei besondere Beachtung. Da der FREAK-Deskriptor zum Zeitpunkt der Erstellung dieser Arbeit ein sehr neues Verfahren ist, werden auch dessen Vorgänger BRIEF und BRISK in deren Grundideen beschrieben. Das ermöglicht eine Abgrenzung dieser neuartigen binären Verfahren von den etablierten reellwertigen Methoden SIFT und SURF, welche ebenfalls in Grundzügen dargestellt werden sollen.

Das zweite Grundlagenkapitel beschäftigt sich mit Farbe, erklärt die Grundlagen der Farbwahrnehmung und legt verschiedene Möglichkeiten der Darstellung von Farbe als Farbräume dar. Auch wird hier eine Möglichkeit aufgezeigt, wie verschiedene Farben miteinander verglichen werden können. Abschließend wird ein kurzer Einblick gegeben, wie die SIFT in der Vergangenheit um Farbe erweitert wurde, was als Grundlage für einen Teil der in dieser Arbeit durchgeführten Experimente dient.

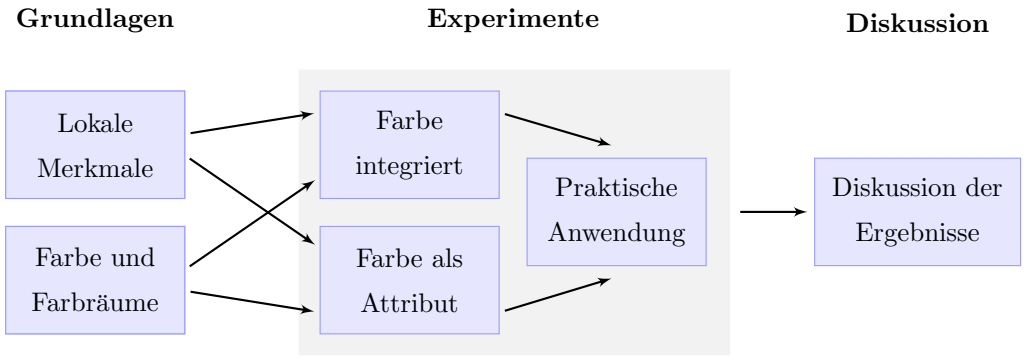


Abbildung 1.1: Dargestellt ist der schematische Aufbau der Arbeit. Nach den einführenden Grundlagenkapiteln zu lokalen Merkmalen und Farbe werden zwei verschiedene Ansätze aufgezeigt, den FREAK-Deskriptor um Farbe zu erweitern. Im Anschluss werden die gewonnen Erkenntnisse in einem Praxisversuch zur Anwendung gebracht, um die Relevanz der gewonnenen Erkenntnisse zu untersuchen. Die Diskussion der Ergebnisse schließt den inhaltlichen Teil der Arbeit ab.

Im experimentellen Teil werden die Grundlagen zusammengeführt. Zunächst wird dabei theoretisch geprüft, wie der FREAK-Deskriptor um Farbinformation erweitert werden kann. Dabei werden zwei verschiedene Ansätze verfolgt. Der eine Ansatz integriert Farbe direkt in den Deskriptor, der zweite behandelt Farbe als ein externes Attribut. Diese ersten Experimente werden anhand eines in der Forschung verbreiteten Datensatzes evaluiert.
Den ersten Experimenten ist ein praktischer Teil angeschlossen. Hier wird eine konkrete Anwendung demonstriert, in welcher Bilder anhand lokaler Merkmale beschrieben werden. Ziel ist es dabei, bekannte Objekte in einem Kamerabild zu identifizieren. Darüber hinaus dient diese konkrete Anwendung einer Überprüfung, inwiefern die in den ersten Experimenten gewonnenen Erkenntnisse in der Praxis bestätigt werden können.

Die Diskussion schließt den inhaltlichen Teil der Arbeit ab. Darin werden verschiedene in den Experimenten beobachtete Phänomene aufgegriffen und besprochen. In der Schlussbetrachtung kann darauf aufbauend sowohl bewertet werden, ob eine Erweiterung des FREAK-Deskriptors um Farbinformation gewinnbringend sein kann, als auch eine Empfehlung ausgesprochen werden, wie eine solche Erweiterung möglich ist.

2 Bildbeschreibung anhand lokaler Merkmale

In diesem Kapitel wird die Grundidee des Verfahrens erläutert, Bilder anhand *lokaler Merkmale* zu beschreiben. Zunächst wird dabei eine generelle Einordnung dieses Verfahrens auf dem Gebiet des maschinellen Sehens vorgenommen. Im Anschluss daran wird ein Überblick über das entstehende Gesamtsystems gegeben, welches sich wiederum in verschiedene Teilschritte gliedern lässt. Um im weiteren Verlauf der Arbeit auf diesen Grundlagen aufbauen zu können, werden die einzelnen Teilschritte in einem Detailgrad erläutert, der zum Verständnis der Arbeit nötig ist. Der Schwerpunkt der Arbeit liegt auf dem von Alahi et al. (2012) publizierten FREAK-Deskriptor, welcher zum Zeitpunkt der Erstellung dieser Arbeit die neueste Entwicklung auf dem Gebiet binärer Deskriptoren ist. Um dessen Eigenschaften besser verstehen zu können, werden zunächst binäre Deskriptoren von reellwertigen abgegrenzt. Im Anschluss daran werden auch die direkten Vorgänger des FREAK-Deskriptors beleuchtet.
Das Kapitel abschließen wird die Erläuterung der Methodik, welche einen Leistungsvergleich verschiedener Deskriptoren ermöglicht, da dies in den folgenden Experimenten von großer Bedeutung ist.

2.1 Einordnung des Verfahrens der Beschreibung lokaler Merkmale

Die *Beschreibung lokaler Bildmerkmale* ist ein Verfahren auf dem Gebiet des *Maschinellen Sehens*. Während bei der *Computergrafik* das Ziel ist, Szenen und Objekte auf Bildschirmen oder anderen Ausgabegeräten darzustellen, beschäftigt man sich auf dem Gebiet des Maschinellen Sehens mit dem "gegenteiligen" (Szeliski, 2011) Problem: Hier ist das Ziel, aus den rohen Daten, aufgezeichnet durch eine Kamera in Form von Pixelintensitäten, Informationen zu gewinnen: Im großen Maßstab können Satellitenbilder Auskunft geben über die Größe von Vegetationszonen, in einem kleineren Maßstab kann anhand von Kameraaufnahmen festgestellt werden, ob sich Personen oder bestimmte Objekte im Bild befinden.
Auf dem Gebiet der *Objekterkennung* unterscheidet man den *generischen* von dem *spezifischen Fall* (Grauman und Leibe, 2011). Im generischen Fall ist das Ziel, Objektkategorien im Bild zu finden, um dann Aussagen treffen zu können wie "Es befindet sich eine Person im Bild", im spezifischen Fall hingegen ist man daran interessiert, Instanzen

eines Objektes zu identifizieren. So könnte beispielsweise gewünscht sein, eine im Bild auftauchende Personen namentlich zu benennen.
Weder generische, noch spezifische Objekterkennung ist jedoch eine einfache Aufgabe. Verschafft man sich einen Überblick über die gegenwärtig populären Verfahren wird außerdem deutlich, dass es nicht *den* Standardalgorithmus gibt, welcher Kategorien erkennt und Objekte identifiziert. Welche Algorithmen genau sich anbieten, hängt stark von dem zu lösenden Problem ab. Zum Auffinden von Fußgängern in einem Bild hat sich beispielsweise das von Dalal und Triggs (2005) vorgestellte Verfahren der orientierten Gradienten etabliert, während ein weit verbreitetes Verfahren für das Finden von Gesichtern von Viola und Jones (2001) stammt. Betrachtet man die Details der Verfahren genauer, so werden deutliche Unterschiede sichtbar: Das erstgenannte Verfahren nutzt Gradienten und somit eine Art Silhouette von Objekten für deren Beschreibung, letzteres macht sich zunutze, dass die Oberfläche eines jeden Gesichtes charakteristische Merkmale wie etwa dunkle Augenhöhlen oder den Schatten von Nase und Mundpartie aufweist. Abhängig von der Problemstellung werden demnach problemspezifische Eigenschaften in großem Maße genutzt.
Auch das in dieser Arbeit verwendete Verfahren der Beschreibung lokaler Bildmerkmale ist in hohem Maße auf das zu lösende Problem angepasst. In der vorgestellten Ordnung gehört es in den Bereich der spezifischen Objekterkennung und ermöglicht, bekannte Objekte, wie etwa das Titelbild eines Buches, in einem anderen Bild zu finden. Für ein sehr ähnliches Problem, wie beispielsweise dem Identifizieren von bekannten Gesichtern, sind lokale Bildmerkmale hingegen ungeeignet und es bieten sich wiederum andere Verfahren an. Eine etablierte Methode ist hier die Anwendung von *Eigengesichtern* (Turk und Pentland, 1991), welche Gesichter global durch eine Linearkombination beschreiben und dadurch einen Abgleich ermöglichen. Dieser Umstand begründet sich dadurch, dass für die Beschreibung der Eingabedaten auch hier verschiedene Merkmale genutzt werden: Eigengesichter beschreiben die Form eines Gesichtes global, während lokale Merkmale, wie der Name schon ausdrückt, kleinere Bereiche eines Bildes betrachten.
Für eine genauere Einführung in die genannten Verfahren oder einen weitreichenderen Überblick zu Standardverfahren auf dem Gebiet des Maschinellen Sehens sei auf die Fachliteratur verwiesen, einen Überblick bietet beispielsweise Szeliski (2011). Zusammenfassend kann festgestellt werden, dass Objekterkennung viele Facetten hat und für

eine geplante Anwendung eine genaue Bestimmung des zu erreichenden Zieles nötig ist. Der nächste Abschnitt gibt einen Überblick über die einzelnen Teilschritte des Gesamtsystems zum Beschreiben von Bildern anhand lokaler Merkmale, worauf sich diese Arbeit konzentriert.

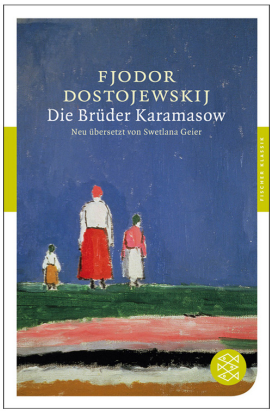
2.2 Anforderungen an das Verfahren

Ein Anwendungsszenario wurde bereits kurz genannt: Ein bekanntes Objekt oder Bild, beispielsweise das Titelbild eines Buches, gegeben als ein Referenzbild, soll innerhalb einer weiteren Aufnahme wiedergefunden werden. Die zweite Aufnahme könnte die eines Tisches sein, aufgenommen aus der Vogelperspektive oder aus einem Winkel, welcher ein menschlicher Betrachter einnehmen würde. Auf dem Tisch könnten verschiedene Gegenstände zu finden sein, darunter eventuell auch das gesuchte Buch. Einerseits könnte man herausfinden wollen, ob sich das Buch auf dem Tisch befindet, und wenn das der Fall ist, eine Angabe darüber, wo genau auf dem Tisch das Buch zu finden ist. Auch könnte dann von Interesse sein, in welcher Lage das Buch relativ zum Betrachter ausgerichtet ist.
Schnell werden verschiedene Schwierigkeiten deutlich. Während das Referenzbild eine unverzerrte Aufnahme des Titelbildes in einer bestimmten Auflösung zeigt, kann sich die Aufnahme der Tischoberfläche davon drastisch unterscheiden. Angenommen, das Buch befindet sich tatsächlich mit dem Titelbild nach oben sichtbar auf dem Tisch, so wird es im allgemeinen Fall nicht in der gleichen Auflösung wie im Referenzbild abgebildet sein. Eine Anforderung an das Verfahren ist somit *Skalierunginvarianz*: Bilder sollen auch gefunden werden, wenn diese in einer anderen Auflösung vorliegen, sei es eine höhere oder eine niedrigere im Vergleich zum Referenzbild.
Es sollte ebenfalls nicht vorausgesetzt werden müssen, dass das Buch immer zum Betrachter ausgerichtet liegt und auch gefunden werden, wenn es mit der Kopfseite nach unten ausgerichtet ist, oder gar in einem beliebigen Winkel. Diese geforderte Eigenschaft wird als *Rotationsinvarianz* bezeichnet. Eine außerdem geforderte Eigenschaft ist die Invarianz gegenüber *perspektivischer Verzerrung*, welche dann zustande kommt, wenn die Aufnahme der Szene räumlich nicht entsprechend des Referenzbildes ausgerichtet ist und sich der Betrachtungswinkel somit zwischen den Bildern unterscheidet. Die angesprochenen Transformationen sind in Abbildung 2.1 beispielhaft verdeutlicht.

Weitere Transformationen, gegenüber denen ein Verfahren zur Beschreibung lokaler Merkmale möglichst invariant sein sollte, können benannt werden: So sollte ein Bild auch wiedergefunden werden, wenn es *teilweise verdeckt* ist, was dann auftritt, wenn beispielsweise ein Stift oder ein anderer Gegenstand auf dem Buch abgelegt ist oder auf andere Weise die freie Sicht auf das komplette Titelbild verhindert. Fast immer auftretende Transformationen sind eine *Änderung der Beleuchtung*, sei es deren Richtung oder auch eine Änderung der Helligkeit und der Beleuchtungsfarbe. In der Formulierung der zu lösenden Aufgabe, ein Bild innerhalb eines anderen wieder zu finden, ist als weitere Transformation die *Translation* bereits implizit enthalten. Diese bezeichnet den Umstand, dass das Bild oder Objekt innerhalb des Eingabebildes an beliebiger Stelle zu finden sein kann.
Der nächste Abschnitt beschreibt überblicksartig den Aufbau des Systems, welches die angesprochenen Probleme in verschiedenen, aufeinander aufbauenden Schritten löst.

2.3 Aufbau des Gesamtsystems

Das Verfahren der Beschreibung lokaler Merkmale lässt sich in verschiedene Teilsysteme gliedern, welche zusammengesetzt das Gesamtsystem ergeben. Diese Teilsysteme sind durch alternative Verfahren austauschbar, da diese untereinander durch klar definierte Schnittstellen verbunden werden. Zunächst soll ein Überblick über das gesamte System gegeben werden, um dann in den sich anschließenden Abschnitten näher auf die einzelnen Teilschritte einzugehen. Begleitend zum Text wird auf die Abbildung 2.2 verwiesen, welche den beschriebenen Aufbau schematisch darstellt.
Betrachtet man die Anforderungen an das System, welche im vorangegangenen Abschnitt aufgestellt wurden, so wird schnell deutlich, dass der Versuch, Bilder auf Pixelebene direkt abzugleichen, zum Scheitern verurteilt ist. Die geforderten Invarianzen sind zu komplex, als dass diese durch Vergleiche auf Pixelebene sichergestellt werden können. Die Grundidee besteht vielmehr darin, *extreme Merkmale* eines Bildes zu identifizieren und deren nähere Umgebung möglichst eindeutig zu beschreiben.
Bevor die formale Darstellung dieses Vorgangs erfolgt, soll eine Intuition dafür gegeben werden, was unter einem extremen Merkmal zu verstehen ist. Greifen wir das bereits gezeigte Beispiel des Buches in Abbildung 2.1 auf. Auf dem Titelbild ist die dominie-



(a)

(b)

(c)

(d)

Abbildung 2.1: Zu sehen sind verschiedene Transformationen des Referenzbildes (a), welche auftreten, sobald der Betrachtungswinkel verändert wird. In Abbildung (b) ist das Titelbild im Bezug auf das Referenzbild skaliert, in (c) außerdem rotiert. In Bild (d) ist das Bild zusätzlich perspektivisch verzerrt. Auch eine leichte Veränderung der Beleuchtung ist zu beobachten. Gegenüber diesen Transformationen sollte ein Verfahren zur Beschreibung lokaler Merkmale möglichst invariant sein.

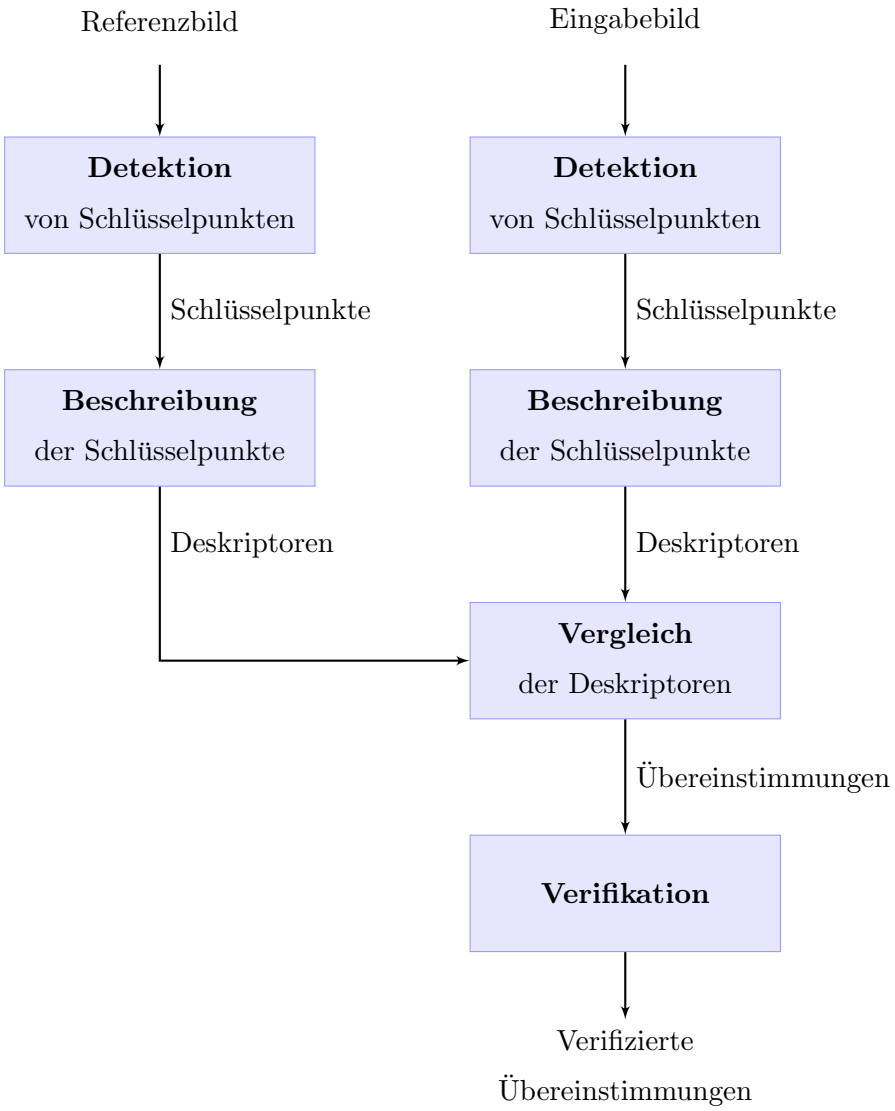


Abbildung 2.2: Schematische Darstellung des Ablaufes beim Abgleich von lokalen Merkmalen eines Eingabebildes mit einem Referenzbild. Zunächst wird in beiden Bildern nach Schlüsselpunkten gesucht (Abs. 2.4), deren unmittelbare Umgebungen in einem sich daran anschließenden Schritt durch Deskriptoren beschrieben werden (Abs. 2.5). Der Vergleich Abgleich von Deskriptoren verschiedener Bilder ermöglich das Auffinden von übereinstimmenden Merkmalen. Eine abschließende Verifikation (Abs. 2.6) der gefundenen Übereinstimmungen kann einen Großteil falscher Paarungen ausschließen.

rende Farbe das Blau des Himmels, welcher sich hinter den Figuren auf der linken Seite erstreckt. Extreme Merkmale oder *Extremstellen*, nachfolgend auch als *Schlüsselpunkte* bezeichnet, sind Teile des Bildes, welche sich stark von ihrer Umgebung abheben. Während der Himmel im Bild trotz leichter Änderungen der Schattierung eine homogene Fläche bildet, sticht das weiße Oberteil der mittleren Person deutlich hervor. Auch die Schrift, welche den Titel des Buches angibt, hebt sich stark vom Hintergrund ab. Betrachtet man in Abbildungen 2.1 die Bilder (b)-(c), so ist in jedem einzelnen, trotz der auftretenden Transformationen, das weiße Oberteil als markantes Merkmal weiterhin sichtbar. Der erste Schritt ist somit, genau solche Punkte im Bild zu detektieren, was in Abschnitt 2.4 genauer erläutert wird.

Der zweite Schritt besteht darin, die detektierten Schlüsselpunkte zu *beschreiben*. Das geschieht, indem die unmittelbare Nachbarschaft dieser Punkte in Form eines *Deskriptors* abgebildet wird. Auch hier sind verschiedene Verfahren einsetzbar, an welche die Anforderung gestellt wird, dass sie eine Menge von Schlüsselpunkte als Eingabe akzeptieren und für diese einen Deskriptor als Ausgabe berechnen. Ausführlicher dargestellt werden Deskriptoren im Abschnitt 2.5 dieser Arbeit.

Wie in Abbildung 2.2 erkennbar, werden die bisher genannten Schritte sowohl auf das Referenzbild, als auch auf das Eingabebild angewendet. Anschließend ist es möglich, die Deskriptoren der einzelnen Bilder miteinander zu vergleichen und so in ihrer Umgebung ähnliche Schlüsselpunkte zu finden. Gleiche Schlüsselpunkte sollten dabei trotz eventuell auftretender Transformationen ähnliche Deskriptoren besitzen. Wie genau Deskriptoren miteinander verglichen werden hängt von deren Form ab. Im Falle der *Scale-invariant Feature Transform* (kurz SIFT), ist der Deskriptor ein 128-stelliger Vektor von Gleitkommazahlen, verwendet man hingegen das *Speeded-Up Robust Features* (SURF) Verfahren, so ist dieser Vektor nur 64 Stellen lang. SIFT und SURF werden als Repräsentanten reellwertiger Deskriptoren in Abschnitt 2.5.1 vorgestellt. Der Schwerpunkt dieser Arbeit wird auf einem *binären Deskriptor*, den sogenannten *Fast Retina Keypoints* (kurz FREAK, Abschnitt 2.5.3) liegen, dessen Deskriptor aus einem 512-stelligen, binären Vektor besteht.

Als zusätzlichen Schritt kann an den Vergleich der Deskriptoren die Verifikation der gefundenen Übereinstimmungen angeschlossen werden. Diese filtert die gefundene Menge an Paarungen, um so die Wahrscheinlichkeit des Auftretens gültiger Paarungen zu erhö-

hen. In Abschnitt 2.6 werden dazu verschiedene Möglichkeiten vorgestellt. Im folgenden Abschnitt soll als erster Teilschritt die Detektion von Schlüsselpunkten am Beispiel den Hesse-Detektors erläutert werden.

2.4 Detektion von Schlüsselpunkten

Für die Detektion von Schlüsselpunkten existieren verschiedene Ansätze. In dieser Arbeit im Detail besprochen wird der *Fast-Hessian Detektor* von Bay et al. (2008), da dieser auch in den späteren Experimenten zur Anwendung kommen wird und damit ein elementarer Bestandteil im Gefüge des dargestellten Gesamtsystems ist. Zunächst wird jedoch mit Integralbildern ein grundlegendes Konzept eingeführt, welches an verschiedenen Stellen der Arbeit wiederkehren wird, sowohl bei den Detektoren, als auch bei den Deskriptoren.

2.4.1 Integralbilder

Für häufig vorzunehmende Berechnungen ist es von Vorteil, effiziente Algorithmen einzusetzen. Eine dieser Berechnungen ist das Summieren von Pixelintensitäten beliebiger, durch ein Rechteck begrenzter Bereiche eines Bildes. Es bietet sich die Verwendung von *Integralbildern* an, ein von Viola und Jones (2001) in der Bildverarbeitung populär gemachtes Verfahren, welches wiederum auf *Summed Area Tables* von Crow (1984) zurückgeht. Mit Hilfe von Integralbildern ist diese Berechnung in konstanter Zeit möglich, unabhängig von der Größe des betrachteten Bereiches oder der des Bildes, nachdem das Integralbild einmalig erstellt wurde. Als Intensität eines Pixels wird dessen Helligkeitswert verstanden, es wird also vorausgesetzt, dass es sich bei der Eingabe um ein Bild mit genau einem Kanal handelt, beispielsweise ein Grauwertbild oder aber ein einzelner Farbkanal eines Bildes.
Sei $I(x, y)$ die Intensität des Pixels an Position (x, y) eines Bildes I. Das zugehörige Integralbild I_Σ lässt sich wie folgt berechnen:

$$I_\Sigma(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(x, y)$$

An jeder Koordinate des Integralbildes ist somit die Summe aller Pixel zwischen dieser und dem Bildursprung des Originalbildes abgelegt. Soll nun ein Bereich begrenzt durch die Koordinaten A, B, C und D berechnet werden, so ist dies anhand folgender Formel möglich:

$$\text{Bereich}_{\text{ABCD}} = I_\Sigma(A) + I_\Sigma(D) - (I_\Sigma(C) + I_\Sigma(D))$$

Die Abbildung 2.3 veranschaulicht das Verfahren. Integralbilder werden bei der Anwendung von SURF an verschiedenen Stellen genutzt, sowohl im Detektor, welcher im folgenden Abschnitt erläutert wird, als auch im Deskriptor. Binäre Deskriptoren setzen zur Verbesserung der Laufzeit ebenfalls Integralbilder ein, worauf an den entsprechenden Stellen hingewiesen werden wird.

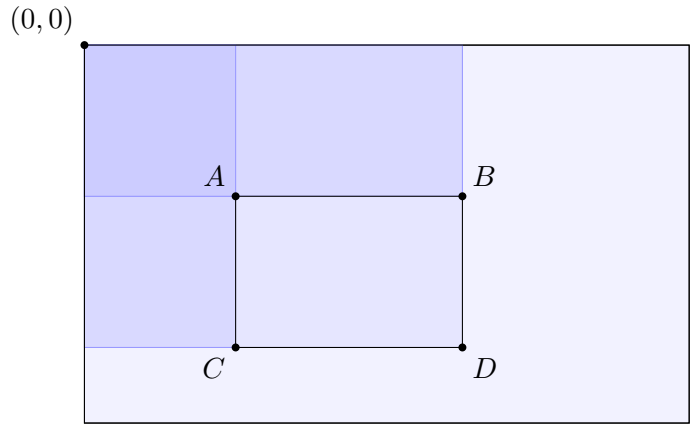


Abbildung 2.3: Verdeutlicht ist das Konzept eines Integralbildes. Soll die Summe der Pixelintensitäten des zugehörigen Originalbildes des durch die Koordinaten A, B, C und D begrenzten Bereiches berechnet werden, so ergibt sich diese aus der Formel $I_\Sigma(A)+I_\Sigma(D)-(I_\Sigma(C)+I_\Sigma(D))$.

2.4.2 Der Fast-Hessian Detektor

Der Fast-Hessian Detektor löst das Problem, Schlüsselpunkte im Bild zu identifizieren, welche bis zu einem gewissen Grad invariant gegenüber den in Abschnitt 2.2 geforderten Transformationen sind. Er geht auf den *Hesse-Detektor* von Beaudet (1978) zurück, ist jedoch weniger laufzeitintensiv als dieser.

Digitale Bilder können als zweidimensionale, diskrete Funktionen verstanden werden, deren Definitionsbereich die Koordinaten eines Bildes sind, der Wertebereich hingegen die Menge aller möglichen Intensitäten eines einzelnen Bildpunktes. Für Grauwertbilder werden die gegebenen Koordinaten auf diese Intensitäten abgebildet, bei Farbbildern bildet die Funktion Koordinaten entsprechend auf die Werte der einzelnen Farbkanäle ab. Schlüsselpunkte sind Extremstellen dieser Funktion, welche mit Mitteln der Analysis berechnet werden können. Dazu muss zunächst die *Hesse-Matrix* eingeführt werden, welche dem Detektor seinen Namen gibt.

Die Hesse-Matrix Die Hesse-Matrix ermöglicht es, eine Funktion $f(x, y)$ zweier unabhängiger Variablen x und y, deren zweite partielle Ableitungen existieren, auf Extremwerte zu untersuchen. Sie ist dabei im Punkt (x, y) wie folgt definiert:

$$\mathrm{H}_{f(x,y)} = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x,y) & \frac{\partial^2 f}{\partial x \partial y}(x,y) \\ \frac{\partial^2 f}{\partial y \partial x}(x,y) & \frac{\partial^2 f}{\partial y^2}(x,y) \end{pmatrix} = \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{pmatrix}$$

Nach dem Satz von Schwarz, welcher besagt, dass die Reihenfolge der Bildung partieller Ableitung keinen Einfluss auf das Ergebnis hat wenn die Funktion total differenzierbar ist, gilt $f_{xy} = f_{yx}$. Die Hesse-Matrix ist demzufolge nicht nur quadratisch, sondern auch symmetrisch.
Für die Bestimmung der Extremwerte einer Funktion werden die Eigenwerte der Hesse-Matrix betrachtet: Sind beide Eigenwerte positiv, so handelt es sich um ein lokales Minimum, da sich die Funktion von oben an den Punkt anschmiegt. Sind beide hingegen negativ, so handelt es sich um ein lokales Maximum. Ist das Vorzeichen der Eigenwerte verschieden, so liegt an der Stelle kein Extrempunkt, sondern ein Sattelpunkt vor. Die Hesse-Matrix ist quadratisch, daher ist das Produkt ihrer Eigenwerte gleich der Determinante:

$$\det \mathrm{H}_{f(x,y)} = \lambda_1 \cdot \lambda_2 = f_{xx} f_{yy} - (f_{xy})^2$$

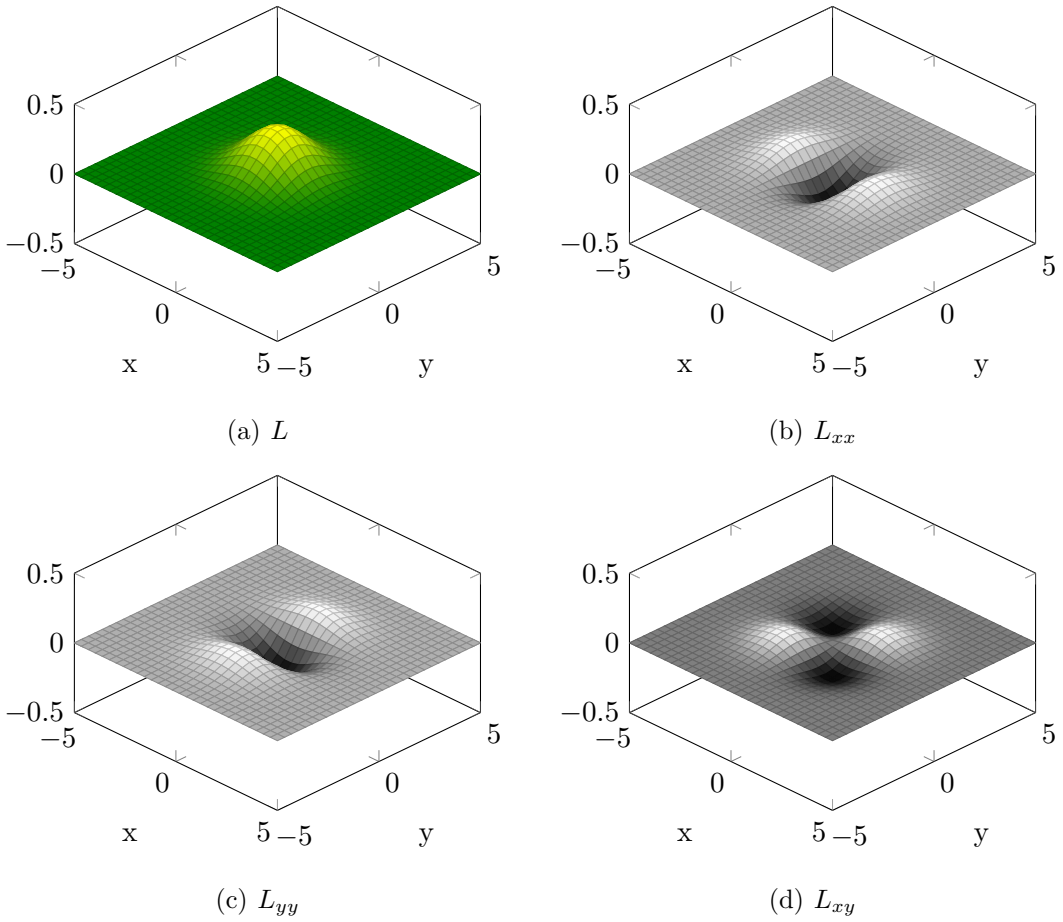
Es ist somit ausreichend, die Determinante zu betrachten, um zu untersuchen, ob ein Extremwert in dem Punkt vorliegt. Die Eigenwerte müssen nicht explizit berechnet werden. Ist die Determinante negativ, so sind die Vorzeichen der Eigenwerte verschieden,

es liegt kein Extrempunkt vor. Ist das Vorzeichen hingegen positiv, so haben die Eigenwerte das gleiche Vorzeichen und es handelt sich entweder um ein Minimum oder ein Maximum und somit um einen Extrempunkt. Gilt $\det \mathrm{H}_{f(x,y)} = 0$, so ist eine Entscheidung über das Vorliegen einer Extremstelle in dem Punkt anhand der Hesse-Matrix allein nicht möglich.

Der beschriebene Ansatz zur Ermittlung von Extrempunkten kann auch auf Bilder übertragen werden. Es wurde bereits beschrieben, wie Bilder als diskrete Funktionen verstanden werden können. Anhand der Faltung eines Bilder unter Verwendung spezieller Filter ist es möglich, ein Analogon für die Ableitungen der Bilder zu berechnen, um so die Hesse-Matrizen anzunähern. Man bedient sich dazu der Ableitungen der Gaußfunktion, deren Graph sowie der ihrer zweiten Ableitungen in Abbildung 2.4 dargestellt ist. Die genauen Funktionsvorschriften finden sich im Anhang A.1.

Betrachtet man die Ableitungen L_{xx} und L_{yy} so fällt auf, dass diese einander gleichen, einzig der Richtung nach zueinander rechtwinklig ausgerichtet sind. Bei deren Nutzung als Filter für die Faltung (auch *Konvolution*) eines Bildes schlagen diese bei Kanten in x- bzw. in y-Richtung an und erzeugen so hohe Werte. L_{xy} hingegen findet Kanten, welche im Bild ungefähr diagonal liegen.
Die Konvolution eines Bildes ist eine teure Operation. Mit dem *Konvolutionssatz* lässt sich zwar die Laufzeit verringern, in der Praxis lohnt sich dessen Anwendung jedoch erst für große Bilder, da das Eingabebild zunächst mithilfe der Fouriertransformation vom Bild- in den Frequenzraum gewandelt werden muss und entsprechend zurück. Ausführungen zum Thema der Faltung von Bildern, der Fouriertransformation oder dem Konvolutionssatz würden den Rahmen der Arbeit sprengen, daher sei an dieser Stelle für eine genauere Darstellung auf die Fachliteratur verwiesen. Eine gute Einführung in das Thema bietet beispielsweise Solomon (2011).
Der Hesse-Detektor bietet dennoch ein gutes Laufzeitverhalten. Anstatt die exakten Ableitungen der Gaußfunktion zur Filterung zu verwenden, werden diese diskretisiert und angenähert, so dass diese in der Folge sehr schnell berechnet werden können. Eine Darstellung der diskretisierten Ableitungen findet sich in Abbildung 2.5.

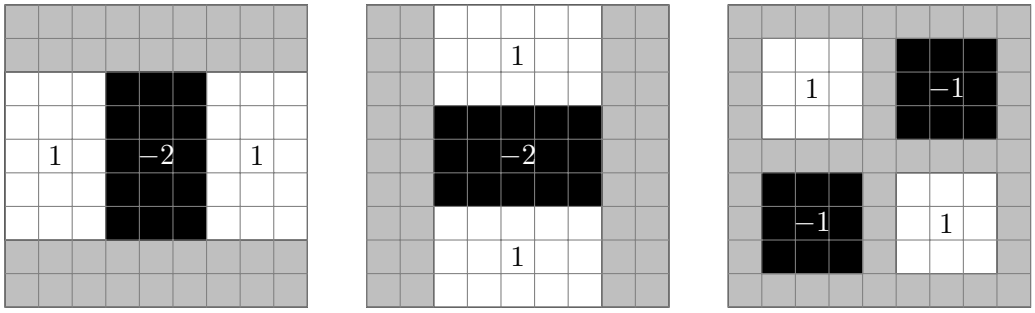
Der Vorteil der Diskretisierung der Ableitungen besteht darin, dass die einzelnen *Lappen* eines jeden Filters eine rechteckige Grundfläche besitzen. Es ist somit möglich, diese mit Hilfe der in Abschnitt 2.4.1 vorgestellten Integralbilder in konstanter Zeit zu berechnen.



(a) L (b) L_{xx}

(c) L_{yy} (d) L_{xy}

Abbildung 2.4: Dargestellt ist oben links in grün die Gaußfunktion L, hier für $\sigma = 1.2$, was der kleinsten Filtergröße des Hesse-Detektors entspricht. Rechts daneben deren zweite Ableitung L_{xx}. Unten links die Ableitung L_{yy}, sowie unten rechts die Ableitung L_{xy}.



(a) D_{xx} (b) D_{yy} (c) D_{xy}

Abbildung 2.5: Schematische Darstellung der Filter, welche sich aus den diskretisierten Ableitungen der Gaußfunktion ergeben. Die Zahlen geben den Gewichtungsfaktor des darunterliegenden Lappens an, grau gefärbte Flächen werden nicht betrachtet und gehen mit dem Faktor 0 ein. In (a) findet sich D_{xx}, in (b) D_{yy} und in (c) der Filter D_{xy}.

Der damit einhergehende Verlust der Genauigkeit gegenüber der Filterung mit den exakten Ableitungen wird in Bay et al. (2008) genauer untersucht und gegenüber dem Gewinn an Geschwindigkeit der Berechnung als unproblematisch bewertet.

Ein Vorteil der Nutzung der Gaußfunktion als Filter besteht darin, dass das Resultat gegenüber Rauschen bis zu einem gewissen Grad robust ist, da diese einer *Tiefpassfilterung* des Bildes entspricht. Da auch die Größe des Filters als Parameter σ in die Berechnung einfließt, ergibt sich die approximierte Hesse-Matrix wie folgt:

$$\mathrm{H}_{\mathrm{app}}(x, y, \sigma) = \begin{pmatrix} D_{xx}(x, y, \sigma) & D_{xy}(x, y, \sigma) \\ D_{yx}(x, y, \sigma) & D_{yy}(x, y, \sigma) \end{pmatrix}$$

Bay et al. (2008) schlagen vor, die Berechnung der Determinante anzupassen, um dem Größenverhältnis der Lappen der diskretisierten Ableitungen zu den exakten Werten der Gaußfunktion gerecht zu werden. Benutzt wird dafür folgende Näherung:

$$\det \mathrm{H}_{\mathrm{app}} = D_{xx} D_{yy} - (0.9 D_{xy})^2$$

Da die Gaußfunktion eine zirkulär symmetrische Funktion ist, werden Extrempunkte unabhängig von der Ausrichtung des Bildes gefunden, was eine erste Voraussetzung für die Rotationsinvarianz des Verfahrens ist. Die Skalierungsinvarianz hingegen wird auf eine andere Art und Weise sichergestellt. Genutzt wird hierfür der sogenannte *Scale-space*, welcher im Folgenden näher erläutert wird.

Eine Möglichkeit, Schlüsselpunkte unabhängig von der Auflösung eines Bildes zu finden ist, eine *Bildpyramide* aufzubauen. Eine Bildpyramide hält das gleiche Bild in verschiedenen Auflösungen vor und erlaubt so, auf den einzelnen Ebenen nach Schlüsselpunkten zu suchen und so Schlüsselpunkte verschiedener Größen zu finden. Das Skalieren eines Bildes ist jedoch eine laufzeitintensive Operation, weswegen diese im Fast-Hessian Detektor vermieden wird. Stattdessen erfolgt eine Veränderung der Größe der Filter, welche dann auf das unveränderte Bild angewendet werden, im Resultat entspricht das der Filterung des Bildes in verschiedenen Auflösungen. Für eine gegebene Seitenlänge ist die Struktur der Filter leicht zu berechnen. Das zugrunde liegende Prinzip ist dargestellt in Abbildung 2.6.

Die einzelnen Ebenen einer Bildpyramide werden im Kontext der Detektion von Schlüsselpunkten als *Oktaven* bezeichnet. Auf der ersten Oktave wird mit einem 9×9 Filter begonnen, auf der zweiten mit einem 15×15, auf der dritten mit einem 27×27 Pixel großem Filter, es findet somit jedes Mal eine Verdoppelung der Vergrößerung zur letzten Oktave statt, begonnen mit 6 Pixeln. Diese Größenänderung ergibt sich aus der Anforderung, die Filterstruktur mit einem zentralen Pixel bei dessen Vergrößerung zu erhalten, wie dargestellt in Abbildung 2.6(b). Dennoch ist das Filtern auf Oktavebene sehr statisch, was zum Nachteil wird, wenn sich die Filtergrößen im Verhältnis zur Auflösung des Eingabebildes nicht mit der des Referenzbildes decken. Lowe (2004) schlägt daher als Verbesserung vor, Schlüsselpunkte im *Subpixelbereich* zu finden, welche orts- und skalierungsstabil sind. Die Idee besteht darin, jede Oktave in weitere Ebenen, genannt *Intervalle*, zu zerlegen. Die Faltung des Bildes erfolgt dabei mit weiteren Filtern, deren Größe sich auf Intervallebene jeweils sich um eine Konstante unterscheidet. Eine Aufschlüsselung der Filtergrößen auf den unterschiedlichen Oktaven findet sich in Tabel-

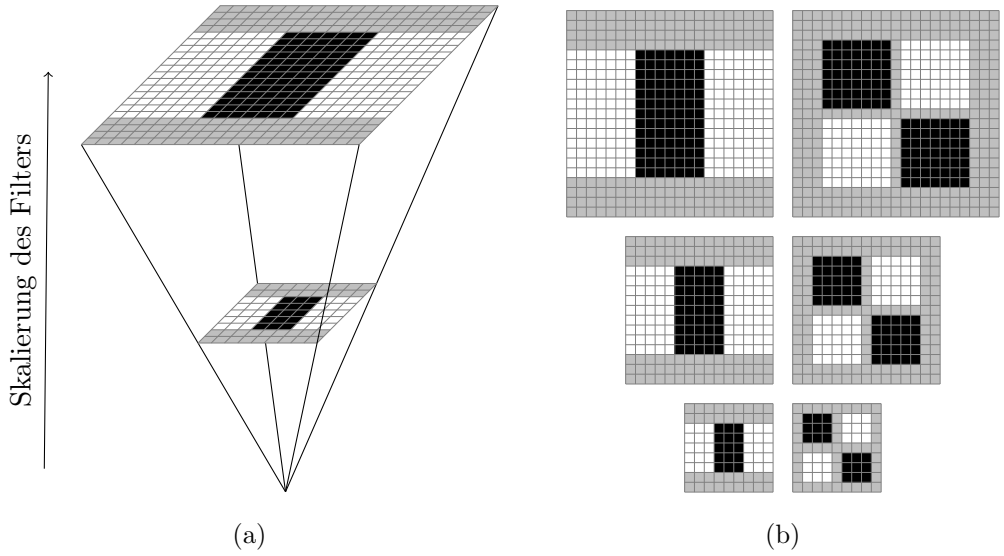
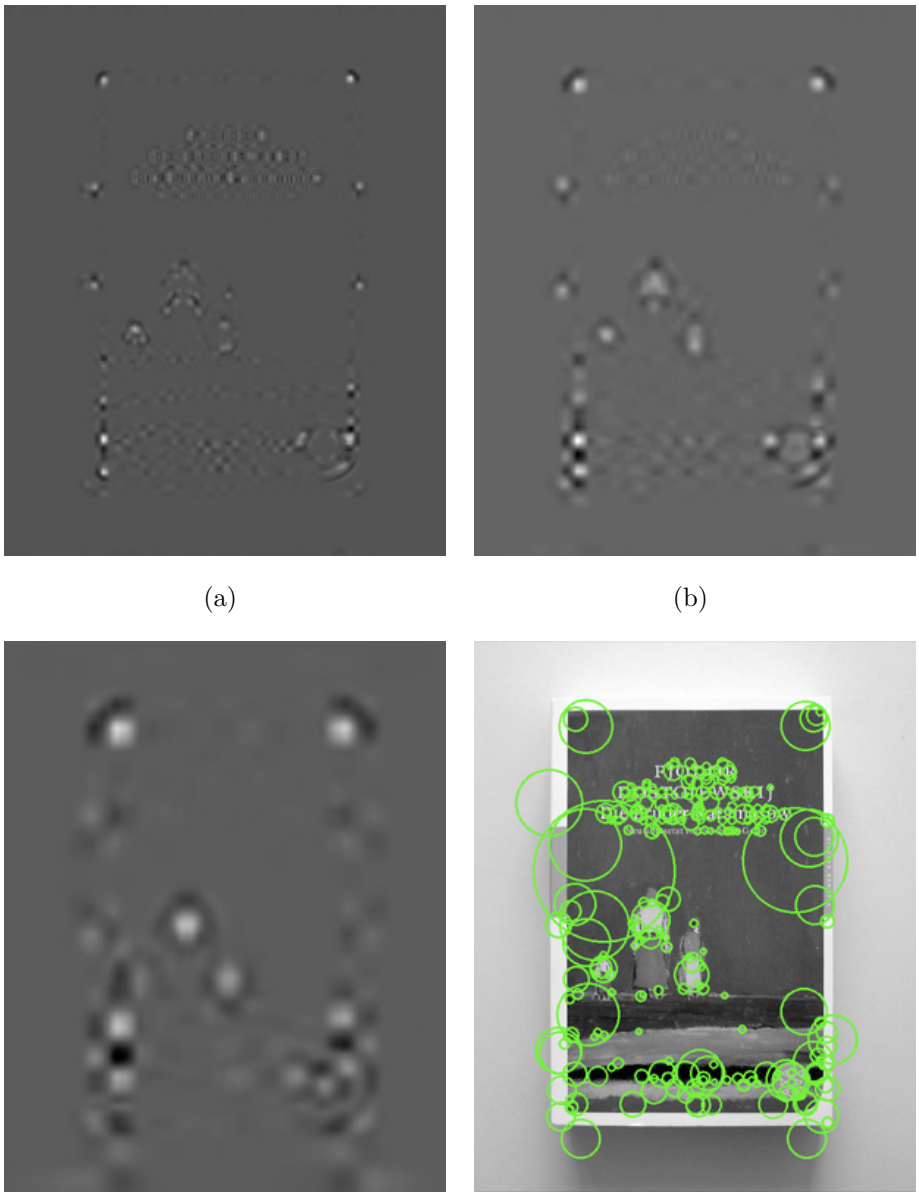


Abbildung 2.6: (a) Dargestellt ist das Prinzip, eine Bildpyramide durch Skalierung der Filter zu ersetzen. Die Auflösung des Eingabebildes bleibt erhalten, die Filter werden jedoch in verschiedenen Auflösungen darauf angewendet, da deren Struktur schnell berechnet werden kann. (b) Beispiele für Filter in verschiedenen Auflösungen. Die Struktur des Filtern muss dabei trotz der Skalierung erhalten bleiben. Abbildung nach (Evans, 2009)

le 2.1. Der Vorteil dieser Intervalle zwischen den Oktaven soll im Folgenden verdeutlicht werden.

Ein Extremum $\mathbf{x} = (x, y, \sigma)^\top$ im Scale-Space definiert sich durch dessen Position, gegeben als x und y Koordinaten, sowie dessen Größe σ. Die Zerlegung einer Oktave in Intervalle ermöglicht es, sowohl die Position als auch die genaue Größe des Schlüsselpunktes zu interpolieren und damit den angesprochenen Nachteil statischer Filtergrößen zu kompensieren. Mit Hilfe der vorgestellten Filter (siehe Abbildung 2.5) lässt sich die Determinante der Hessematrix in einem Punkt bestimmen. Das geschieht, indem das Bild mit den Filtern gefaltet, und aus der Ausgabe einzelner Filter die Determinante berechnet wird. Nachfolgend wird das sich daraus ergebende Bild, welches den Wert der Determinanten in einem Punkt zeigt, als *Filtermap* bezeichnet. Beispiele für Filtermaps verschiedener Größen finden sich in Abbildung 2.7.

Der erste Schritt der genauen Bestimmung der Schlüsselpunkte ist es, die einzelnen Werte der Filtermaps mit einem Schwellwert zu vergleichen und nur jene Punkte beizu-



(a) (b)

(c) (d)

Abbildung 2.7: (a)-(c) Dargestellt sind Filtermaps eines Bildes nach aufsteigender Filtergröße. Schwarz sind lokale Minima, weiß sind lokale Maxima des Bildes. Mit zunehmender Filtergröße verschwinden immer mehr Details, wodurch Extrempunkte unterschiedlicher Größe gefunden werden. In (d) findet sich das Eingabebild, wobei Schlüsselpunkte entsprechend ihrer Größe und Position durch grüne Kreise eingezeichnet sind.

	Filtergrößen			
Oktave 1	9	15	21	27
Oktave 2	15	27	39	51
Oktave 3	27	51	75	99
Oktave 4	51	99	147	195

Tabelle 2.1: Gelistet sind die Seitenlängen der Filter einzelner Intervalle. Jede der vier Oktaven ist in weitere vier Intervalle unterteilt, die sich bezüglich der Filtergröße um einen konstanten Wert unterscheiden. Die Intervalle ermöglichen eine Interpolation der genauen Position eines Schlüsselpunktes im Subpixelbereich.

behalten, welche den Schwellwert überschreiten. Auf diese Weise kann die Sensibilität des Detektors gesteuert werden: Hohe Schwellwerte haben zur Folge, dass nur wenige, dafür aber sehr stabile Extrempunkte gefunden werden, niedrige Werte lassen sehr viel mehr Punkte zu.

Um sicherzustellen, dass jedes Extremum nur durch einen einzigen Punkt repräsentiert wird, wendet man zunächst eine *non-maximal supression* an, es werden also alle Punkte gelöscht, die verglichen mit ihren unmittelbaren Nachbarn nicht maximal sind. Dies geschieht, indem jeder Punkt mit seinen 8 direkten Nachbarn, sowie den 9 Pixeln des Intervalls darüber, sowie denen des Intervalls darunter verglichen wird. Auf Oktavebene dient die Filtermap des Größten, sowie die des kleinsten Filters nur zu Vergleichszwecken, auf denen selbst nicht nach Extrempunkten gesucht wird.

Mit Hilfe einer *Taylor-Entwicklung* ist es möglich, Funktionen in einem Punkt als Polynom anzunähern. Voraussetzung dafür ist es, deren Ableitungen in diesem Punkt zu kennen. Im Kontext des Fast-Hessian Detektors kann durch die Taylor-Entwicklung der Scale-Space Funktion $H(x, y, \sigma)$, welche die Stärke der Determinante in diesem Punkt angibt, die Position des Extremums im Subpixelbereich gefunden werden. Die Ableitungen in einem gefundenen Extrempunkt $\mathbf{x}$ sind durch die darüber und die darunterliegende Filtermap einer Oktave bestimmbar. Die Idee geht auf Brown und Lowe (2002) zurück, welche für die Interpolation den folgenden Ausdruck angeben:

$$H(\mathbf{x}) = H + \frac{\partial H^\top}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^\top \frac{\partial^2 H}{\partial \mathbf{x}^2}\mathbf{x}$$

Setzt man die Ableitung dieses Ausdrucks gleich Null und löst dann nach dem Argument auf, ermöglicht dies die Bestimmung des interpolierten Extremums $\hat{\mathbf{x}}$:

$$\hat{\mathbf{x}} = -\frac{\partial^2 H^{-1}}{\partial \mathbf{x}^2}\frac{\partial H}{\partial \mathbf{x}}$$

Der Nachteil statischer Filtergrößen wird so ausgeglichen. Die Position eines Extrempunktes kann auf diese Weise genauer bestimmt werden, als durch Oktav- und Intervallauflösungen eigentlich beschränkt ist. Das trägt auch zur Robustheit und Wiederholbarkeit des Auffindens der Schlüsselpunkte bei, da so der Auflösung des Eingabebildes und damit dessen Verhältnis zu den einzelnen Filtergrößen einen geringerer Einfluss zuteil wird.

2.4.3 Weitere Detektoren

Innerhalb der maschinellen Sehens ist der Fast-Hessian Detektor dem Gebiet der *blob detection*, einem Teilgebiet der *Feature detection*, zuzuordnen. Dieses Teilgebiet beschäftigt sich damit, markante Bereiche, sogenannte *blobs* (von engl. "Klecks") im Bild zu finden, die sich von ihrer Umgebung abheben. Der Fast-Hessian Detektor löst dies anhand der Determinante der Hesse-Matrix und ist durch den Einsatz von Integralbildern und vereinfachten Filtern deutlich effizienter als sein Vorgänger, der bereits genannte Hesse-Detektor.
Keinesfalls soll jedoch der Eindruck entstehen, der im Detail vorgestellte Detektor sei die einzige Möglichkeit, markante Bereiche eines Bildes zu detektieren. Ein weiteres Verfahren ist beispielsweise der *Harris Detektor* (Harris und Stephens, 1988), welcher eher Strukturen wie Ecken im Bild findet und Bereiche mit starker Varianz in Bezug auf die Textur hingegen nicht detektiert (Grauman und Leibe, 2011).
Ein Detektor, welcher ohne das Filtern des Bildes auskommt, ist der *Adaptive and Generic Accelerated Segment Test* (AGAST) von Mair et al. (2010). Der Detektor baut auf dem *Features from Accelerated Segment Test* (FAST) von Rosten und Drummond (2006) auf, welcher Ecken innerhalb eines Bildes nach der Grundidee des "20-Fragen Tests" findet: Um einen Pixel herum werden alle Pixel auf einem Kreis eines festen Radius (bezeichnet als *Bresenham's Circle*) betrachtet und mit dem Pixel in der Mitte verglichen.

Sind dabei mindestens S auf dem Kreis zusammenhängende Pixel heller oder dunkler als der Pixel in der Mitte, so bewertet der Detektor die untersuchte Stelle als Ecke. Dabei wird jedoch nicht jeder Pixel einzeln untersucht, vielmehr wird der Test anhand gezielter Betrachtungen einzelner, weniger Pixel vollzogen. Soll beispielsweise auf einem Kreis mit dem Umfang von 16 Pixeln S zusammenhängende Pixel gefunden werden, so reicht es aus, nur eine Teilmenge der auf dem Kreis liegenden Pixel zu betrachten. Je größer S dabei ist, desto weniger Punkte müssen tatsächlich verglichen werden. Dieser Entscheidungsbaum wird im Falle des FAST Detektors anhand einer Trainingsmenge gelernt. Dies hat jedoch den Nachteil, dass bereits eine leichte Rotation der Kamera ein erneutes Lernen der Verteilungen erfordert. Der AGAST Detektor kommt ohne den Trainingsschritt aus und arbeitet anhand zweier Entscheidungsbäume, welche sich zur Laufzeit an die getestete Umgebung anpassen.
Leutenegger et al. (2011) nutzen den AGAST Detektor wie beschrieben für die Detektion von Schlüsselpunkten im Scale-Space, um so auch verschiedene markante Punkte verschiedener Größen im Bild zu finden. Der resultierende Detektor wird im allgemeinen als *Multi-scale* AGAST referenziert.
Für Details zu den genannten Detektoren sei auf die entsprechenden Originalveröffentlichungen verwiesen. Der FREAK-Deskriptor selbst stellt keinen eigenen Detektor vor, die Autoren nutzen den Multi-scale AGAST Detektor. Sie weisen auch darauf hin, dass es für den Vergleich von Deskriptoren keine Rolle spielt, welcher Detektor zur Detektion der Schlüsselpunkte genutzt wird.

Der folgende Abschnitt wird sich mit Deskriptoren beschäftigen. Wie bereits angemerkt ermöglicht eine klare Definition der Schnittstellen, verschiedene Detektoren und Deskriptoren innerhalb des Gesamtsystems auszutauschen. Der Detektor übergibt an den Deskriptor eine Menge von Schlüsselpunkten, welche durch ein Tripel definiert sind: Ihre Position in Koordinaten (x,y), sowie die Position σ im Scale-Space, welche die Größe des Schlüsselpunktes angibt.

2.5 Deskriptoren

Die Aufgabe des Deskriptors ist es, die unmittelbare Umgebung eines gegebenen Schlüsselpunktes in Form eines Vektors zu beschreiben. Die Beschreibung sollte dabei invariant gegenüber den in Abschnitt 2.2 beschrieben Transformationen Skalierung, Rotation, Verzerrung sowie Beleuchtungsänderungen sein. Geringfügige Änderungen im Bild sollen also auch nur geringe Änderungen des Deskriptors zur Folge haben. Eine Ähnlichkeit der Umgebung zweier Schlüsselpunkte sollte sich damit direkt an der Ähnlichkeit ihrer Deskriptoren erkennen lassen. Da die Deskriptoren als Vektor vorliegen, bietet es sich an, die Ähnlichkeit mathematisch als Abstand der Vektoren zueinander messbar zu machen.
Es existieren verschiedene Ansätze zur Lösung dieses Problems, grundsätzlich lassen sich dabei *binäre Deskriptoren* als neue Entwicklungen von Verfahren abgrenzen, welche den Umkreis eines Schlüsselpunktes in Form *reellwertiger* Vektoren beschreiben. Etablierte reellwertige Verfahren sind die Deskriptoren SIFT und SURF, welche in Abschnitt 2.5.1 vorgestellt werden. Im Anschluss daran werden im Abschnitt 2.5.2 mit BRIEF und BRISK binäre Deskriptoren eingeführt. Der Schwerpunkt dieser Arbeit liegt auf dem FREAK-Deskriptor, welcher in Abschnitt 2.5.3 im Detail erläutert wird. In der Reihe der binären Deskriptoren ist dieser die neuste Entwicklung, welcher andere binäre Deskriptoren im Vergleich übertrifft. Die Methodik des Vergleichs von Deskriptoren bezüglich ihrer Robustheit gegenüber verschiedenen Transformationen wird in Abschnitt 2.7 genauer dargestellt.

2.5.1 Reellwertige Deskriptoren

Die *Scale-invariant Feature Transform* (SIFT) wurde von Lowe (2004) vorgeschlagen und beschreibt Schlüsselpunktumgebungen[1] als einen 128-stelligen Vektor von Gleitkommazahlen. Eine schnellere Alternative bietet das von Bay et al. (2008) vorgestellte *Speeded-Up Robust Features*-Verfahren (SURF), welches Schlüsselpunkte ebenfalls durch einen Vektor von Gleitkommazahlen beschreibt, welcher jedoch nur 64 Stellen lang ist. Beide Verfahren sind sehr ähnlich, als wesentlicher Unterschied ist jedoch festzuhalten,

[1] Wenn nachfolgend von der "Beschreibung eines Schlüsselpunktes" die Rede ist, bezieht sich das stets auf dessen Umgebung.

dass SURF häufigen Gebrauch von Integralbildern macht, während die SIFT auf Ebene der Pixel arbeitet. Aus diesem Unterschied begründet sich der Geschwindigkeitsvorteil von SURF gegenüber der SIFT.
Die Skalierungsinvarianz beider Verfahren wird erreicht, indem alle Berechnungen, sei es die Bestimmung der dominanten Orientierung oder die Beschreibung des Schlüsselpunktes selbst, immer in Abhängigkeit der vom Detektor übergebenen Position σ eines Schlüsselpunktes im Scale-Space ausgeführt wird, welche sich mit jeder Skalierung des Bildes entsprechend ändert.
Die Rotationsinvarianz hingegen wird auf andere Weise sichergestellt. Die Idee hierbei ist es, die dominante Orientierung des Schlüsselpunktes zu bestimmen und die Berechnung des Deskriptors entsprechend auszurichten. Wird ein Eingabebild rotiert, so ändert sich zwangsläufig auch die festgestellte Orientierung des Schlüsselpunktes und das Verfahren insgesamt ist invariant gegenüber Rotationen. Im Falle des FREAK-Deskriptors wird die Feststellung der Orientierung im Detail dargestellt (siehe Abschnitt 2.5.3), für SIFT und SURF sei auf die angegebenen Veröffentlichungen verwiesen (Lowe, 2004; Bay et al., 2008).
Die Berechnung des SIFT Deskriptors ist bildlich dargestellt in Abbildung 2.8. Die Stärke der Gradienten in der Umgebung des Schlüsselpunktes, zusammengefasst in einem 16×16 Gitter, werden zunächst anhand einer Gaußfunktion gewichtet, welche auf dem Schlüsselpunkt selbst zentriert ist. Auf diese Weise haben Gradienten näher am Zentrum einen höheren Einfluss als jene, die davon weiter entfernt sind. Das Gitter wird anschließend zu 4×4 Histogrammen gebündelt, welches jedes einzeln 8 Richtungen zusammenfasst. Daraus ergibt sich der finale SIFT Deskriptor von $4 \times 4 \times 8 = 128$ Stellen, welcher die Umgebung eines Schlüsselpunktes in Abhängigkeit von dessen Ausrichtung und Größe beschreibt.

Der SURF Deskriptor kodiert ebenfalls die Gradienten um den Schlüsselpunkt herum als Vektor, allerdings ist dieser mit 64-Stellen nur halb so lang wie der SIFT Deskriptor. Für die Berechnung des SURF Deskriptors wird ein 4×4 Zellen großes Gitter über den Schlüsselpunkt gelegt und jede dieser Zellen in weitere 5×5 Quadrate unterteilt. Für jedes dieser Quadrate wird dessen Orientierung berechnet, welche dann, ebenfalls mit einer Gaußfunktion gewichtet, als $\sum dx$, $\sum |dx|$, $\sum dy$ und $\sum |dy|$ aufsummiert werden,

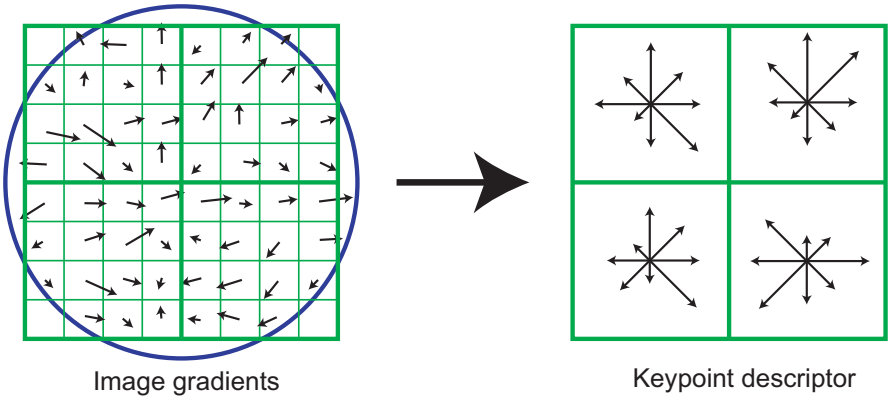


Abbildung 2.8: Schematische Darstellung der Berechnung des SIFT Deskriptors in verkleinerter Darstellung. Die durch einen blauen Kreis angedeutete Gaußfunktion gewichtet Gradienten in der Umgebung eines Schlüsselpunktes, welche dann in Histogramme zusammengefasst werden, aus denen sich der finale Deskriptor ergibt. Gegenüber dem Original ist die Darstellung verkleinert und zeigt einen Deskriptor mit nur $2 \times 2 \times 8 = 32$ Stellen. Abbildung entnommen aus (Lowe, 2004).

wobei dx für den Gradienten in x-Richtung und dy entsprechend in y-Richtung steht. Die einzelnen Summen des großen 4×4 Gitters werden zum Schluss konkateniert und ergeben so den finalen Vektor mit $4 \times 4 \times 4 = 64$ Stellen. Als letzter Schritt wird der Vektor normiert, was ihn zusätzlich invariant gegenüber einer Kontraständerung des Eingabebildes macht. Eine schematische Abbildung des Verfahrens findet sich in 2.9.

Laut der Autoren ist SURF in der Praxis nicht nur effizienter, sondern auch wesentlich robuster ist als die SIFT. Sie schreiben letzteres dem Fakt zu, dass der SURF-Deskriptor durch die Summenbildung innerhalb der einzelnen Zellen gegenüber Rauschen weniger anfällig ist. Beim SIFT Deskriptor jedoch beeinflusst Rauschen die Richtung der Gradienten, und kann diese damit stören.

Wie Rotations- und Skalierungsinvarianz erreicht wird, wurde beschrieben. Es existieren auch Versuche, perspektivische Änderungen in Form einer Homographie zu detektieren und bei der Berechnung des Deskriptors zu beachten. Für einen Überblick über diese Verfahren sei beispielsweise auf Tuytelaars und Mikolajczyk (2007) verwiesen. SIFT und SURF hingegen behandeln perspektivische Transformationen nicht gesondert, verringern jedoch deren Einfluss, indem sie unter Verwendung der Gaußfunktion näher am Schlüs-

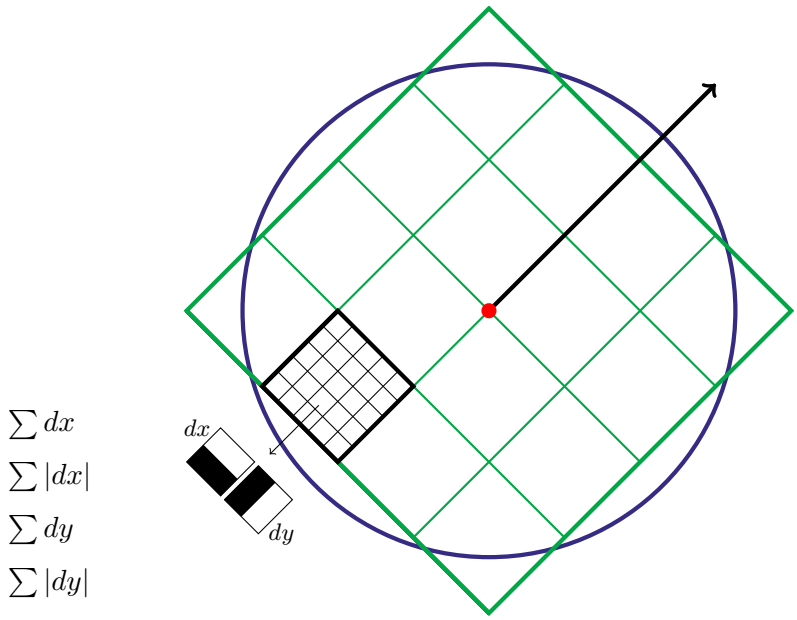


Abbildung 2.9: Schematische Darstellung des SURF-Deskriptors. Auf jedem Schlüsselpunkt wird, abhängig von dessen Größe und Richtung,hier dargestellt durch den schwarzen Pfeil, ein 4×4 zelliges Quadrat gelegt. In den einzelnen Zellen wird anhand von Haar-Wavelets die Richtung des Gradienten an 25 Punkten bestimmt, mit einer Gaußfunktion gewichtet, und innerhalb einer Zelle aufsummiert. Die Summen konkateniert ergeben den Deskriptor, bestehend aus $4\times 4\times 4 = 64$ Gleikommazahlen. Abbildung nach (Evans, 2009).

selpunkt liegende Gradienten höher gewichten. Der weitreichende praktische Einsatz beider Verfahren zeigt, dass dies eine akzeptable Vorgehensweise ist.

Die reellwertigen Deskriptoren verbindet, dass sie Schlüsselpunkte anhand eines Vektors reeller Zahlen $\mathbf{x} \in \mathbb{R}^n$ beschreiben.[2] Als Maß für die Ähnlichkeit zweier Deskriptoren $\mathbf{x}$ und $\mathbf{y}$ wird daher deren euklidischer Abstand zueinander verwendet:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Der SURF Deskriptor ist ein Punkt einem 64-dimensionalen Raum. Die interne Darstellung der Gleitkommazahlen eines Computers erfolgt jedoch binär. Somit könnte der SURF Deskriptor gleichzeitig als binärer Vektor in einem sehr viel höher dimensionalen Raum interpretiert werden. Im nächsten Abschnitt werden binäre Deskriptoren eingeführt, welche in diesen Raum direkt eingebettet sind.

2.5.2 Die binären Deskriptoren BRIEF und BRISK

Die Repräsentation von Deskriptoren als hochdimensionaler Vektor von Gleitkommzahlen hat Nachteile. Einerseits ist deren Berechnung selbst mitunter nicht sehr effizient, andererseits ist auch der Vergleich verschiedener Deskriptoren miteinander, also die Berechnung des euklidischen Abstandes, eine teure Operation.
Eine weitere Möglichkeit der Darstellung von Deskriptoren ist es, *Binärstrings* anstelle der reellwertigen Vektoren zu verwenden. Das erfordert jedoch auch den Entwurf neuer Algorithmen für deren Berechnung und Vergleich. In diesem Abschnitt soll die Grundidee erläutert werden, lokale Merkmale anhand binärer Deskriptoren zu beschreiben.

Der erste, im Jahr 2010 veröffentlichte binäre Deskriptor trägt den Namen *Binary Robust Independent Elementary Features* (BRIEF) von Calonder et al. (2010). Der Ansatz unterscheidet sind grundsätzlich von dem der reellwertigen Deskriptoren und leitete eine Entwicklung weiterer, darauf aufbauender Verfahren ein.

[2] Durch die Darstellung von Gleitkommazahlen im Computer sind diese zwangsläufig natürlich nie echt reell, sondern immer rational.

Die zu Grunde liegende Neuerung bestand darin, die Umgebung von Schlüsselpunkten anhand von paarweisen Intensitätsvergleichen einzelner Pixel eines geglätteten Bildes zu kodieren. Das Ergebnis eines solchen Vergleiches, welcher intuitiv verständlich ist als die Aussage, ein bestimmter Punkt sei dunkler als der damit verglichene, wird dann in Form eines einzigen Bits abgelegt, je nachdem ob der Test positiv oder negativ ausfällt. Da eine hohe Anzahl einzelner Pixelvergleiche stattfindet, werden die berechneten Bits anschließend konkateniert und ergeben einen Binärstring, welcher dann als binärer Deskriptorvektor genutzt wird. Sei τ der Test auf einem bestimmten Bereich $\mathbf{p}$ der Größe $S \times S$, so ist dieser definiert als:

$$\tau(\mathbf{p}; \mathbf{x}, \mathbf{y}) = \begin{cases} 1, & \text{falls } \mathbf{p}(\mathbf{x}) < \mathbf{p}(\mathbf{y}) \\ 0, & \text{sonst} \end{cases}$$

wobei $\mathbf{p}(\mathbf{x})$ für die Intensität eines geglätteten Eingabebildes an Position $\mathbf{x} = (u, v)^\top$ steht. Für eine Menge von n_d Positionen $\mathbf{x}$ und $\mathbf{y}$ setzt sich der Deskriptor zusammen als der Bitstring

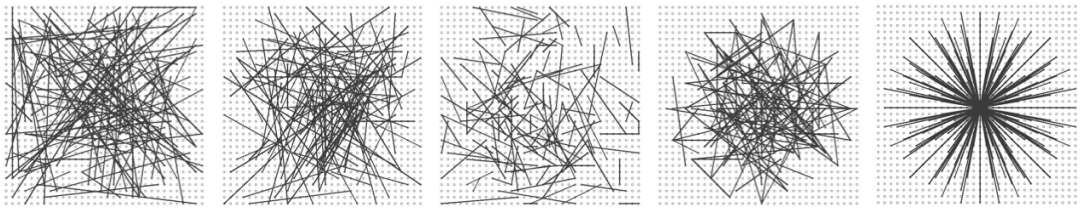
$$f_{n_d}(\mathbf{p}) = \sum_{1 \le i \le n_d} 2^{i-1} \cdot \tau(\mathbf{p}; \mathbf{x}_i, \mathbf{y}_i)$$

Für den Vergleich von Deskriptoren wird als Abstandsmaß der euklidische Abstand durch den *Hamming-Abstand* abgelöst. Ursprünglich wurde dieser von Hamming (1950) eingeführt um den Unterschied zweier Signale $\mathbf{x}$ und $\mathbf{y}$ zu errechnen, indem die Zahl voneinander verschiedener Zeichen festgestellt wird:

$$d(\mathbf{x}, \mathbf{y}) = \sum_{x_i \neq y_i} 1, \quad i = 1, ..., n$$

Der Hamming-Abstand ist nur definiert für zwei Zeichenketten der gleichen Länge n. Für binäre Zeichenketten, also Zeichenketten über dem Alphabet $\{0, 1\}$, lässt sich der Hamming-Abstand sehr effizient in Form eines *exklusiven Oders* (“xor”) beider Signale mit angeschlossenem Zählen der Einsen implementieren.

Bezüglich der Anordnung der zu vergleichenden Punkte $(\mathbf{x}_i, \mathbf{y}_i)$ führten die Autoren des BRIEF.Deskriptors verschiedene Experimente durch um festzustellen, welche Anordnung den diskriminativsten Deskriptor ergibt. Sie variierten dabei den Abstand der Vergleichspunkte zueinander, deren Verteilung über den Bereich allgemein, und testeten auch Muster mit regelmäßiger Anordnung der Vergleichspunkte (siehe Abbildung 2.10). Die Autoren stellen fest, dass Muster (b) am besten abschneidet, ebenfalls sehr gute Ergebnisse werden mit Muster (a), (c) und (d) erzielt. Muster (e) hingegen schneidet deutlich schlechter ab als alle Muster im Test.



(a) (b) (c) (d) (e)

Abbildung 2.10: Gezeigt sind beispielhaft verschiedenen Muster, welche von den Autoren von BRIEF zur Anordnung der zu vergleichenenden Punkte getestet wurden. Bei den Mustern (a)–(d) folgt die Anordnung der Punkte verschiedenen zufälligen Verteilungen, (e) ist ein Muster mit regelmäßiger Anordnung der Vergleichspunkte. Die Abbildungen sind entnommen aus (Calonder et al., 2010)

BRIEF selbst sieht nicht vor, invariant gegenüber Rotationen und der Skalierung des Eingabebildes zu sein. Durch eine manuelle Drehung der Eingabebilder selbst testen die Autoren, ebenfalls unter Verwendung des Fast-Hessian Detektors, die Wiedererkennungsrate von Schlüsselpunkten gegenüber SURF und zeigen, dass BRIEF nicht nur wesentlich effizienter, sondern auch diskriminativer ist als SURF. Verglichen wurde dabei mit der OpenCV[3] Implementierung von SURF, den Geschwindigkeitsvorteil geben die Autoren mit dem bis zu 41-fachen an. Auf BRIEF aufbauend veröffentlichen Rublee et al. (2011) einen Deskriptor inklusive eines Detektors, genannt *Oriented* FAST *and Rotated* BRIEF (ORB), welcher BRIEF um Rotationsinvarianz erweitert.

[3]OpenCV ist eine freie Bibliothek mit verschiedenen Algorithmen der Bildverarbeitung für freie und kommerzielle Nutzung. Siehe `http://opencv.org/`.

Binary Robust Invariant Scalable Keypoints (BRISK), ein weiterer binärer Deskriptor, veröffentlicht von Leutenegger et al. (2011), löst das Problem der Skalierungsinvarianz. Die Autoren schlagen einen Detektor vor, welcher auf dem bereits genannten AGAST beruht und interpolieren die genaue Position eines Schlüsselpunktes im Scale-Space, ähnlich dem vorgestellten Fast-Hessian Detektor.

Das in BRISK genutzte Sampling-Muster für die Vergleiche ist dargestellt in Abbildung 2.11. Im Gegensatz zu BRIEF wird ein deterministisches Muster gewählt, einzelne Punkte dienen dabei gleichzeitig mehreren paarweisen Vergleichen, was den Vorteil hat, dass die Zahl der Punkte insgesamt reduziert wird. Die Werte an einzelnen Punkten können somit mehrfach genutzt werden, während die Berechnung der Intensität nach der Glättung in diesem Punkt nur ein einziges Mal nötig ist. Auch wird nicht das gesamte Bild anhand eines Gaußfilters mit konstantem Parameter σ geglättet, vielmehr unterscheidet sich dieser Parameter in Abhängigkeit des Abstandes eines Abtastpunktes zur Mitte. Im Kontext der Detektoren wurde bereits angesprochen, dass die Glättung anhand einer Gaußfunktion eine teure Operation ist. Binäre Deskriptoren lösen dieses Problem, indem sie in der praktischen Anwendung die Glättung durch einen einfachen Mittelswertsfilter ersetzen. Dieser betrachtet die durchschnittliche Intensität eines quadratischen Bereiches um einen Abtastpunkt herum. Mithilfe der in Abschnitt 2.4.1 vorgestellten Integralbilder ist die Berechnung in konstanter Zeit möglich und selbst die Verwendung verschieden großer Bereiche wirkt sich nicht nachteilig auf die Laufzeit aus. Insgesamt nutzt BRISK 60 Abtastpunkte, wobei sich die mögliche Anzahl der Paarungen aus dem Vergleich jeden Punktes mit jedem anderen als $(60 \cdot 59)/2 = 1770$ ergibt. Für den 512-Bit Vektor von Vergleichen, welche den Deskriptor ergeben, werden die 512 Paarungen mit der kürzesten Entfernung der Punkte zueinander genutzt. Die Autoren weisen darauf hin, dass bezüglich der Wahl der Vergleiche Forschungsbedarf bestände um festzustellen, welche Strategie sich am besten eignet. Diesbezüglich bietet der FREAK-Deskriptor, welcher in Abschnitt 2.5.3 im Detail vorgestellt werden soll, einen guten Ansatz.

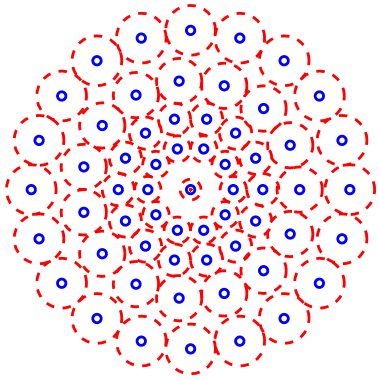


Abbildung 2.11: Das BRISK Muster verwendet 60 Abtastpunkte für die Intensitätsvergleiche, hier dargestellt als blaue Punkte. Die Größe des Einzugbereiches dieser Punkte variiert und wird zum Zentrum hin kleiner. Gesteuert wird dies durch eine Glättung mit einem Paramter σ, hier dargestellt durch rote Kreise. Abbildung entnommen aus (Leutenegger et al., 2011).

2.5.3 Der Fast Retina Keypoints Deskriptor

Der *Fast Retina Keypoints*-Deskriptor (FREAK) von Alahi et al. (2012) baut auf den Ideen der bereits beschriebenen binären Deskriptoren BRIEF und BRISK auf. Es handelt sich um einen reinen Deskriptor ohne einen eigenen Detektor, dieser ist jedoch sowohl invariant gegenüber Skalierungen des Eingabebildes als auch gegenüber Rotationen desselben. Da das Thema dieser Arbeit den Hauptfokus auf den FREAK-Deskriptor richtet, soll dieser im folgenden Abschnitt im Detail dargestellt werden.

Mit dem Wort *Retina* findet sich im Namen des Deskriptors ein Verweis auf das Auge. Bei der Retina (von lat. *rete* “das Netz”, daher auch “Netzhaut” genannt) handelt es sich um die lichtempfindliche Struktur im Inneren des Auges von Wirbeltieren. Auf ihr sitzen die Rezeptoren, die darauf treffende Lichtsignale wahrnehmen. Rezeptoren können einerseits sogenannte Stäbchen sein, welche sehr lichtempfindlich sind, und damit ermöglichen, auch in der Dämmerung zu sehen, andererseits sogenannte Zapfen, welche für das Farbsehen zuständig sind. Im Punkt des schärfsten Sehens, der sogenannten *Fovela centralis*, ist die Dichte der Zapfen am höchsten. Mit zunehmendem Abstand nimmt die Zapfendichte jedoch auf der *Parafovealen*- und schließlich der *Perifovealen* Schicht exponentiell ab. Gleichzeitig werden die Signale zusammenfassend verschaltet, wobei der Grad der Verschaltung von innen nach außen zunimmt. Es werden somit

mit zunehmendem Abstand zur Fovea centralis mehr Lichtreize zusammengefasst, bevor diese durch die sogenannten Ganglienzellen an das Gehirn weitergeleitet werden. Diese Verschaltung wird von binären Deskriptoren durch die Intensitätsvergleiche reproduziert. Abbildung 2.12 stellt die Funktionsweise binärer Deskriptoren als Analogie zum beschriebenen Aufbau des menschlichen Auges dar.

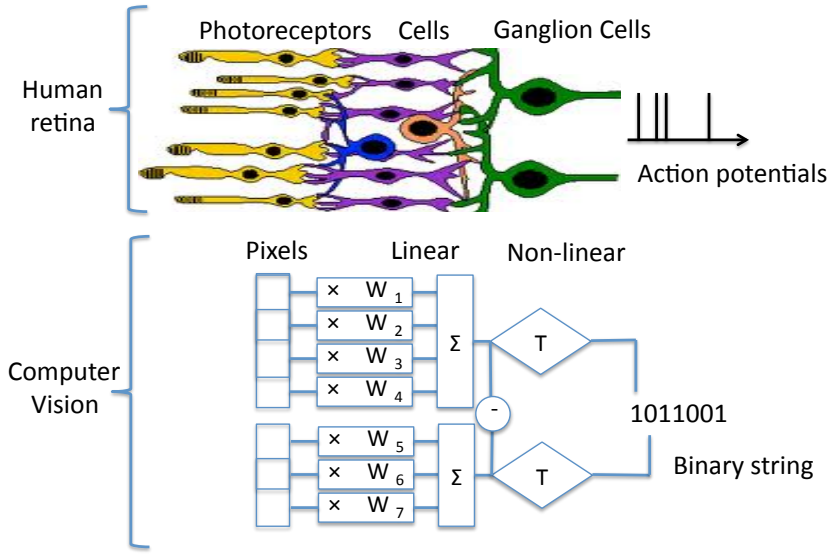


Abbildung 2.12: Die Abbildung zeigt binäre Deskriptoren als Analogie zur Funktionsweise der Retina. Die Signale der Fotorezeptoren werden parallel miteinander verschaltet, um dann als Aktionspotential an das Gehirn weitergeleitet zu werden. Das entspricht der Berechnung eines Binärstrings anhand von Intensitätsvergleichen bei binären Deskriptoren, schematisch dargestellt im unteren Bereich des Bildes. Die einzelnen Bits ergeben sich aus dem Vergleich summierter Bereiche eines Eingabebildes, hier dargestellt als die Differenz der Summen mit nachgeschaltetem Vorzeichentest T. Abbildung entnommen aus (Alahi et al., 2012).

Der FREAK-Deskriptor geht noch einen Schritt weiter und reproduziert die Funktion des Auges nicht nur in Form der Intensitätsvergleiche, welche allen binären Deskriptoren gemein ist, sondern lässt sich darüber hinaus auch von der Verteilung der Rezeptoren auf der Netzhaut inspirieren. Abbildung 2.13 zeigt das Abtastmuster des FREAK-Deskriptors. Aufgebaut ist dieses aus insgesamt 43 Abtastpunkten, welche in 7 Ringen zu je 6 Rezeptoren um einen zentralen Punkt angeordnet sind. Von innen nach außen

nimmt die Größe der Abtastpunkte zu, was hier wieder als Parameter σ der verwendeten Glättung zu verstehen ist und damit den Bereich der in den Punkt einfließenden Informationen definiert.

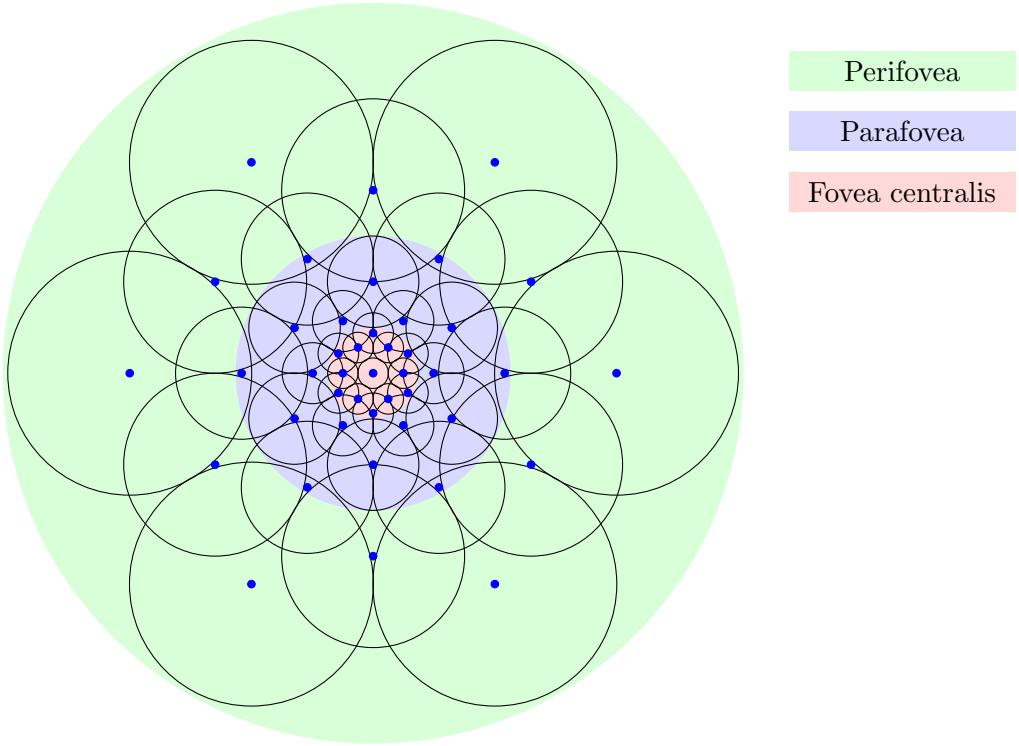


Abbildung 2.13: Das FREAK-Muster, bestehend aus jeweils 6 Punkten auf 7 Ringen um das Zentrum herum. Im Unterschied zum Muster des BRISK-Deskriptors nimmt die Stärke der Glättung mit größerem Abstand zur Mitte stark zu, außerdem überlappen die 43 Einzugsbereiche der Abtastpunkte einander zum Teil. Die Gliederung der Ringe in Perifovea, Parafovea und Fovea centralis ist dem menschlichen Auge nachempfunden.

Verglichen mit dem vorgestellten Muster des BRISK-Deskriptors in Abbildung 2.11 fällt zwar eine leichte Ähnlichkeit auf, zwei grundsätzliche Unterschiede sind jedoch deutlich zu erkennen: Zum Einen ist der Größenunterschied der Abtastpunkte sehr viel stärker ausgeprägt und reproduziert damit den exponentiellen Abfall der Dichte an Ganglienzellen im menschlichen Auge. Zum Anderen überlappen die Einzugsbereiche der Punkte mitunter stark. Dass diese Unterschiede der Leistung des Deskriptors zuträglich sind, wurde von den Autoren experimentell bestätigt.

Analog zum BRIEF-Deskriptor setzt sich der Binärstring F des FREAK-Deskriptors wie folgt zusammen:

$$F = \sum_{0 \leq a < N} 2^a T(P_a)$$

wobei P_a ein Paar rezeptiver Felder ist und N die Länge des Deskriptors in Bits. $T(P_a)$ steht für den Intensitätsvergleich

$$T(P_a) = \begin{cases} 1, & \text{falls } I(P_a^{r_1}) - I(P_a^{r_2}) > 0 \\ 0, & \text{sonst.} \end{cases}$$

Der Ausdruck $I(P_a^{r_1})$ bezeichnet die Intensität des ersten rezeptiven Feldes des Paares P_a. Mit 43 rezeptiven Feldern existieren $43 \cdot 42/2 = 903$ mögliche Paarungen. Zur Bildung eines 512 Bit langen Deskriptors muss somit eine Teilmenge an Paarungen gefunden werden, welche die möglichst diskriminativsten Paare enthält. Die Autoren wählten folgende Strategie:

1. Stelle eine Matrix über eine große Menge von Schlüsselpunkten mehrerer Testbilder auf, wobei in jede Reihe der Deskriptor eines Schlüsselpunktes eingetragen wird, bestehend aus allen 903 möglichen Paarungen. In den Spalten finden sich somit die Ergebnisse der Tests $T(P_a)$. Als Richtwert geben die Autoren eine Zahl von 50000 Schlüsselpunkten an.

2. Berechne für jede Spalte den Mittelwert. Ein Mittelwert nahe 0.5 steht dabei für eine hohe Varianz, d. h. 0 und 1 treten in etwa mit gleicher Häufigkeit auf. Ein solcher Test ist innerhalb der Beispiele sehr diskriminativ.

3. Ordne die Spalten absteigend, entsprechend der höchsten Varianz als Kriterium.

4. Die ersten 512 Spalten enthalten nach der Sortierung die diskriminativsten Tests und werden zur Bildung des 512 Bit langen Deskriptors gewählt.

Die Autoren beobachten, dass die durch den vorgeschlagenen Algorithmus ausgewählten Paare einem "von grob nach fein"-Prinzip folgen: Während für die ersten Bits vor allem

Paaren auf den äußeren Schichten des Deskriptors miteinander verglichen werden, werden für weitere Bits immer mehr rezeptive Felder der inneren Schichten einbezogen. Das ähnelt ebenfalls der Funktionsweise des Auges, welches für die ungefähre Lokalisierung eines Objektes die Perifovealen Rezeptoren nutzt, welche weit von der Fovea centralis entfernt liegen. Für die Fokussierung eines Objektes wird hingegen die Parafoveale Schicht einbezogen und schließlich für das scharfe Sehen die Fovea centralis (Alahi et al. (2012)). Dieses Prinzip kann auch beim Abgleich eines Deskriptors mit einer sehr großen Datenbank genutzt werden. Laut der Autoren ist es eine funktionierende Strategie, beim Vergleich von Deskriptoren schon anhand der ersten 16 Bits eine Vorauswahl zu treffen. Ein vollständiger Vergleich aller 512 Bits muss dann nur auf den so ausgewählten Deskriptoren vollzogen werden. Das spart Rechenzeit, erzielt aber dennoch gute Ergebnisse.

Der FREAK-Deskriptor ist rotationsinvariant, schlägt also ein Verfahren vor, welches die Richtung eines Schlüsselpunktes detektiert, um das Abtastmuster entsprechend auszurichten. Auch dies wird anhand von Intensitätsvergleichen der Paarungen erreicht. Zunächst wird dafür eine Teilmenge G aller Paarungen gewählt, welche 45 Paare enthält, wie dargestellt in Abbildung 2.14. Anhand folgender Formel wird die Richtung O des Schlüsselpunktes bestimmt:

$$O = \frac{1}{M} \sum_{P_o \in G} (I(P_o^{r_1}) - I(P_o^{r_2})) \frac{P_o^{r_1} - P_o^{r_2}}{\|P_o^{r_1} - P_o^{r_2}\|}$$

$P_o^{r_1}$ steht dabei für die euklidischen Koordinaten des Punktes $P_o^{r_1}$, wobei das Muster auf dem Koordinatenursprung zentriert ist. M ist die Zahl der Paare in G, im Fall des Vorschlages der Autoren also 45. Der Ausdruck summiert die in G vorkommenden Richtungen auf, wobei diese jedoch nach der Stärke ihres Intensitätsabfalls gewichtet werden. O zeigt somit nach der Berechnung die dominierende Richtung des Schlüsselpunktes an und das Muster kann entsprechend dieser ausgerichtet werden. Die genauen Koordinaten aller Abtastpunkte für verschiedene Ausrichtungen und Skalierungen des Musters werden in Form einer Lookup-Tabelle abgelegt, um bei deren mehrfachen Anwendung eine konstante Zugriffszeit zu erreichen.

Wie bereits angesprochen wurde, ist BRISK ebenfalls invariant gegenüber Rotationen.

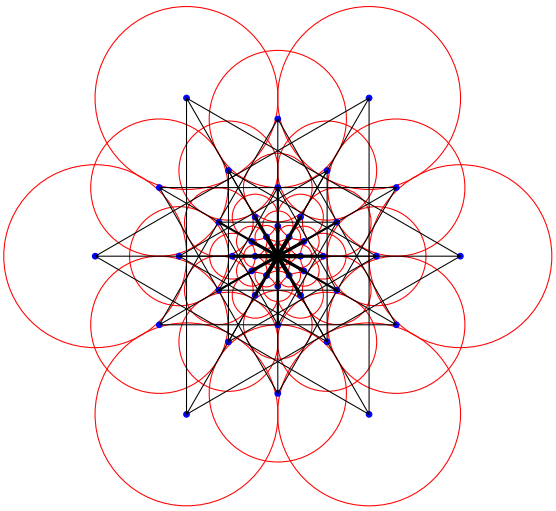


Abbildung 2.14: Das FREAK Muster mit eingezeichneten Paarungen zur Feststellung der Rotation, verbunden durch schwarze Linien. Der Ringstruktur folgend werden direkt gegenüberliegende Bereiche verglichen, sowie Bereiche, welche auf einem Ring liegend ein Dreieck bilden.

Erreicht wird das auf die selbe Art, wie für den FREAK-Deskriptor beschrieben. Die Struktur des FREAK-Deskriptors bringt jedoch an dieser Stelle einen großen Vorteil mit sich: Während das BRISK-Muster auf dem äußeren Ring 20 Abtastpunkte besitzt (siehe Abbildung 2.11), sind es beim FREAK-Muster nur 6 Punkte. Die Zahl der verschiedenen Winkel, welche in der Lookup-Tabelle vorgehalten werden müssen, kann somit drastisch reduziert werden, da sich Rotationen damit für das FREAK-Muster nicht so stark bemerkbar machen wie das bei BRISK der Fall ist. Während die Lookup-Tabelle für BRISK 40 MB an Speicher benötigt, ist bei FREAK mit 7 MB nur ein Bruchteil dessen nötig.
Perspektivische Verzerrungen behandelt der FREAK-Deskriptor, genau wie die reelwertigen Deskriptoren SIFT und SURF, nicht gesondert in Form einer Feststellung der Homographie. Während SIFT und SURF dieses Problem lösen, indem sie anhand einer Gewichtung der zentralen Gradienten Störungen dieser Art auffangen, ist diese Gewichtung der Struktur des FREAK-Deskriptors bereits inhärent: Durch die immer kleiner werdenden Einzugsbereiche der Abtastpunkte zur Mitte hin werden auch hier Störungen durch perspektivische Verzerrungen aufgefangen.
Den Angaben der Autoren nach erzielt FREAK bessere Ergebnisse bezüglich der Beschreibung von Schlüsselpunkten als SIFT und SURF, und übertrifft im Vergleich auch die vorgestellten binären Deskriptoren BRISK und BRIEF.

Der FREAK-Deskriptor ermöglicht eine Entscheidung über die Gleichheit zweier Schlüsselpunktumgebungen. Auf dem Gebiet der *Entscheidungsfindung* lässt er sich den sogenannten *Fast and Frugal*-Verfahren (Gigerenzer und Todd, 1999) zuordnen. Grundsätzlich unterscheidet man dabei *unbegrenzte* und *begrenzte Rationalität*: Bei unbegrenzter Rationalität spielen die Kosten einer Suche keine Rolle, üblicherweise dienen Wahrscheinlichkeitsverteilungen zur Modellierung einer Situation. Menschen hingegen treffen Entscheidungen auf der Grundlage begrenzter Rationalität. Einerseits sind sie beschränkt durch die Fähigkeiten ihres Gehirns, andererseits durch die Tatsache, dass nicht immer alle relevanten Informationen bekannt sind. Man unterscheidet hier *Satisficing*- und die genannten Fast and Frugal-Heuristiken. Erste bezeichnen die Strategie, die erstbeste Lösung zu akzeptieren und die Suche danach abzubrechen. Ein hungriger Mensch beispielsweise mag sich seine Leibspeise erträumen, es reicht aber auch schon ein einfaches Mahl, um seinen Hunger zu stillen – die Lösung ist ausreichend, wenn auch nicht optimal. Fast and Frugal Heuristiken treffen Entscheidungen auf Grundlage begrenzter Informationen und Zeit, einfache Regeln dienen dabei zur Entscheidungsfindung. Im Falle des FREAK-Deskriptors sind dies die Intensitätsvergleiche zur Feststellung der Gleichheit zweier Schlüsselpunktumgebungen. Zwar wird die Lösung mit jedem weiteren Vergleich etwas besser, man kann aber wie beschrieben schon nach 16 Vergleichen abbrechen und auf Grundlage der gesehenen Vergleiche eine Entscheidung treffen.

Im nächsten Abschnitt sollen mit den Vergleichsstrategien von Deskriptoren und der Verifikation der gefundenen Übereinstimmungen die letzten Schritte des Gesamtsystems dargestellt werden.

2.6 Vergleichsstrategien und Verifikation

Bezüglich des Vergleiches von Deskriptoren wurde bereits kurz angesprochen, dass bei den reellwertigen Deskriptoren der euklidische Abstand zweier Deskriptorvektoren zueinander als Maß dient, bei den binären Deskriptoren der wesentlich schneller zu berechnende Hamming-Abstand. Ein geringer Abstand steht dabei für eine große Ähnlichkeit der Schlüsselpunktumgebungen, ein hoher hingegen für große Unterschiede. Um anhand des Abstandes eine Entscheidung über "Übereinstimmung" (engl. "match") und "kei-

ne Übereinstimmung" zu treffen, existieren verschiedene Strategien, welche in diesem Abschnitt vorgestellt werden sollen.

Grundsätzlich lassen sich drei verschiedene Strategien unterscheiden (Mikolajczyk und Schmid, 2005). Eine sehr einfache Strategie des Vergleiches ist das *Schwellwertverfahren*. Der theoretisch maximale Abstand zweier FREAK-Deskriptoren ist 512, was gleichbedeutend wäre mit der Aussage, dass alle Bits verschieden sind. Man könnte nun als Schwellwert einen maximalen Abstand von 16 setzen und so festlegen, dass man alle Deskriptoren mit einer Distanz von 16 und darunter als Übereinstimmung wertet, alle darüber liegenden als verschieden. Genauso gut könnte dieser Abstand aber auch bei 32 oder gar bei 47 liegen. Ein Problem wird dabei deutlich: Das Setzen dieses Schwellwertes ist nicht allgemein gültig möglich. Während ein zu niedriger Schwellwert sehr viele Schlüsselpunkte als falsch zurückweist, kann ein zu hoher Schwellwert Schlüsselpunkte fälschlicherweise als Übereinstimmung werten. Eine bessere Strategie ist daher das sogenannte *Nearest Neighbor Matching*. Dabei wird nur der nächste Nachbar eines Deskriptors als Übereinstimmung gewertet, und das auch nur, wenn dessen Abstand unterhalb eines bestimmten Schwellwertes liegt. Bei Anwendnung dieser Strategie hat jeder Deskriptor maximal einen Nachbar. Mikolajczyk und Schmid (2005) schlagen eine weitere Strategie vor, und führen dazu das Maß der *Nearest Neighbor Distance Ratio* (NNDR) ein. Hierbei ist die Idee, Nearest Neighbour Matching einzusetzen, jedoch das Verhältnis des Abstandes zum ersten mit dem zum zweiten Nachbarn mit einem Schwellwert zu vergleichen. Seien D_A, D_B und D_C Deskriptoren der Regionen A, B und C. Die NNDR berechnet sich dabei wie folgt:

$$\text{NNDR} = \frac{d(D_A, D_B)}{d(D_A, D_C)}$$

wobei $d(D_A, D_B)$ für den Abstand der Deskriptoren D_A zu D_B steht, sei es der euklidische- oder auch der Hamming-Abstand. Eine Veranschaulichung der verschiedenen Strategien anhand eines Beispiels findet sich in Abbildung 2.15.

Bei der Beschreibung des Aufbau des Gesamtsystems zum Abgleich lokaler Merkmale (siehe Abschnitt 2.3) wurde als optionaler letzter Schritt die *Verifikation* genannt. Die Grundidee dabei ist es, anhand der gefundenen Übereinstimmungen die geometrische Transformation des gefundenen Objektes zu schätzen, und diese als Grundlage für eine

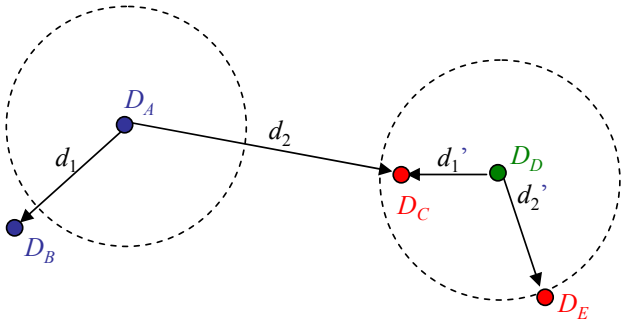


Abbildung 2.15: Veranschaulichung verschiedener Vergleichsstrategien. D_A korrespondiert zu D_B (blau), D_D weder zu D_C noch zu D_E (rot). Bei einem festen Schwellwert, dargestellt durch die gestrichelten Kreise, wird D_B von D_A fälschlicherweise zurückgewiesen, D_D jedoch betrachtet D_C als Übereinstimmung, was ebenfalls nicht korrekt ist. Bei der Nearest-Neighbor Strategie wertet D_A den Deskriptor D_B als Übereinstimmung, D_D den Deskriptor D_C jedoch ebenfalls, auch wenn D_E jetzt korrekt zurückgewiesen wird. Beim Einsatz von NNDR-Matching weist D_D alle Punkte korrekt zurück, während für D_A der Deskriptor D_B weiter eine Übereinstimmung ist. Abbildung entnommen aus (Szeliski, 2011).

Bewertung der Plausibilität der gefundenen Übereinstimmung zu nutzen. Zum Einsatz kommt dabei der sogenannte RANSAC-Algorithmus (Fischler und Bolles, 1981; Szeliski, 2011), welcher durch eine Auswahl von Datenpunkten imstande ist, die vorliegende Homographie zu schätzen und so Ausreißer zu eliminieren. Die detaillierte Darstellung dieses Schrittes würde den Rahmen der Arbeit sprengen und wäre in der Gliederung eher als Exkurs zu betrachtet, daher sei an dieser Stelle auf die zuvor genannte Fachliteratur verwiesen.

2.7 Evaluation von Deskriptoren

Das Grundlagenkapitel zu den Deskriptoren abschließen soll eine Darstellung der Methodik zum Vergleich der Leistung verschiedener Deskriptoren untereinander. In den Experimenten wird dies für die Bewertung der Farbdeskriptoren von großer Bedeutung sein.

Deskriptoren mit einer gewählten Strategie zum Vergleich, sind Klassifikatoren[4], dem-

[4]Wenn im Folgenden von der "Bewertung eines Deskriptors" die Rede ist, so bezieht sich das auf den

zufolge können diese auch anhand der Methoden der Signalentdeckungstheorie (Swets, 1969) evaluiert werden. Grundsätzlich wird für eine Eingabe zweier Schlüsselpunkte eine Ausgabe der Form "die Schlüsselpunkte stimmen überein" (positiv) oder "keine Übereinstimmung" (negativ) generiert und damit eine binäre Entscheidung getroffen. Diese Entscheidung wiederum kann korrekt oder falsch sein, was die Definition der folgenden vier Kennzahlen erlaubt:

Richtig Positive (TP von engl. *true positves*) geben die Zahl der vom Klassifikator korrekt gefundenen Übereinstimmungen an.

Falsch Positive (FP von engl. *false positves*) geben die Zahl der Schlüsselpunkpaare an, welche der Klassifikator fälschlicherweise als Übereinstimmung ausgibt.

Richtig Negative (TN engl. *true negatives*) geben die Zahl der korrekt als "nicht übereinstimmend" bewerteten Beispiele an.

Falsch Negative (FN engl. *false negatives*) geben die Zahl der fälschlicherweise als "nicht übereinstimmend" bewerteten Beispiele an.

Bei einem idealen Klassifikator, welcher keine Fehler macht und immer die richtige Entscheidung trifft, ist die Zahl der falsch Positiven, sowie die Zahl der falsch Negativen stets gleich null. Beobachten lässt sich, dass sich die Zahl der echt positiven Beispiele ergibt als die Summe $\mathrm{TP} + \mathrm{FN}$, die Zahl aller echt negativen Beispiele aus der Summe $\mathrm{TN} + \mathrm{FP}$. Abbildung 2.16 stellt den Zusammenhang der Kennzahlen graphisch dar. In Abhängigkeit eines beobachtbaren Merkmals wird ein Schwellwert gesetzt, anhand dessen je nachdem ob ein Beispiel darüber unter darunter liegt, die Entscheidung "positiv" oder "negativ" gefällt wird. Das betrachtete Merkmal ist im Falle der Klassifikatoren der Abstand zweier Deskriptoren zueinander, welcher zwischen 0 und 512 im Falle des FREAK-Deskriptors liegt. Die Beispiele ergeben sich aus dem Vergleich jedes Deskriptors des Referenzbildes mit jedem des Eingabebildes. Bei einem guten Deskriptor ist der Abstand der Deskriptorvektoren gleicher Schlüsselpunkte gering, der ungleicher Schlüsselpunkte hingegen hoch, die Kurven sind somit klar separiert und lassen sich gut trennen.

---

entsprechenden Klassifikator, nicht auf den Deskriptorvektor eines Schlüsselpunktes.

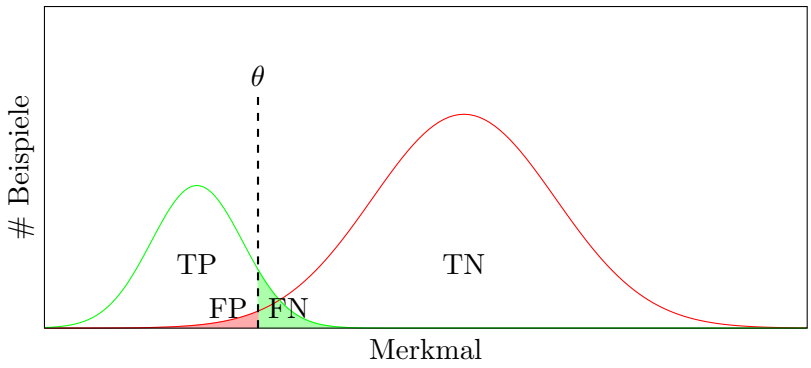


Abbildung 2.16: Dargestellt ist die Klassifikation anhand eines Schwellwertes θ. Die grüne Kurve ist als Histogramm aller echt positiven Beispiele zu verstehen, die rote als eines aller echt negativen, beide sind über ein zur Klassifikation genutztes Merkmal abgetragen. Der Klassifikator trifft die Entscheidung in Abhängigkeit des Schwellwertes. Ist dieser zu gering wächst die Zahl der falsch negativ klassifizierten Beispiele (grün gefüllte Fläche), ist dieser zu hoch die der falsch positiven (rot gefüllte Fläche). Ein Merkmal ist umso besser, je klarer es die Kurven separiert.

Aus den vier vorgestellten Kennzahlen lassen sich nun weitere Messgrößen definieren wie etwa die *Trefferquote* (engl. *recall*), sowie die *Genauigkeit* (engl. *precision*). Im weiteren Verlauf der Arbeit werden die englischen Begriffe eingedeutscht verwendet. Definiert sind Precision und Recall wie folgt:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Im Kontext von Schlüsselpunktdeskriptoren ist es üblich, statt der Precision den Wert *1-Precision* zu verwenden (Mikolajczyk und Schmid, 2005), welcher als *Ungenauigkeit* verstanden werden kann. Der Wert ergibt sich als:

$$1 - \text{Precision} = 1 - \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{FP}}{\text{TP} + \text{FP}}$$

Mit der Veröffentlichung “A Performance Evaluation of Local Descriptors” (Mikolajczyk und Schmid, 2005) haben dessen Autoren ein Verfahren für den Vergleich von Deskrip-

toren vorgeschlagen, welches sich seither als Standard etabliert hat. Voraussetzung sind als Datengrundlage ein Referenzbild sowie weitere Eingabebilder, welche verschiedenen Transformationen unterzogen sind. Für jedes dieser Bilder muss die perspektivische Transformation in Bezug auf das Eingabebild bekannt sein, um daraus ableiten zu können, welche Schlüsselpunktpaare im Referenz- und im Eingabebild korrespondieren. Iteriert man nun über alle möglichen Schwellwerte θ als Merkmal, im Falle des FREAK-Deskriptors also über die Werte von 0 bis 512, so lässt sich die Zahl der falsch positiv und negativ, sowie die der richtig positiv und negativ zugeordneten Schlüsselpunkte in Eingabe und Referenzbild feststellen, was wiederum die Berechnung des Recalls sowie der 1-Precision erlaubt. Diese Werte werden dann graphisch gegenübergestellt, wie beispielhaft zu sehen in Abbildung 2.17.

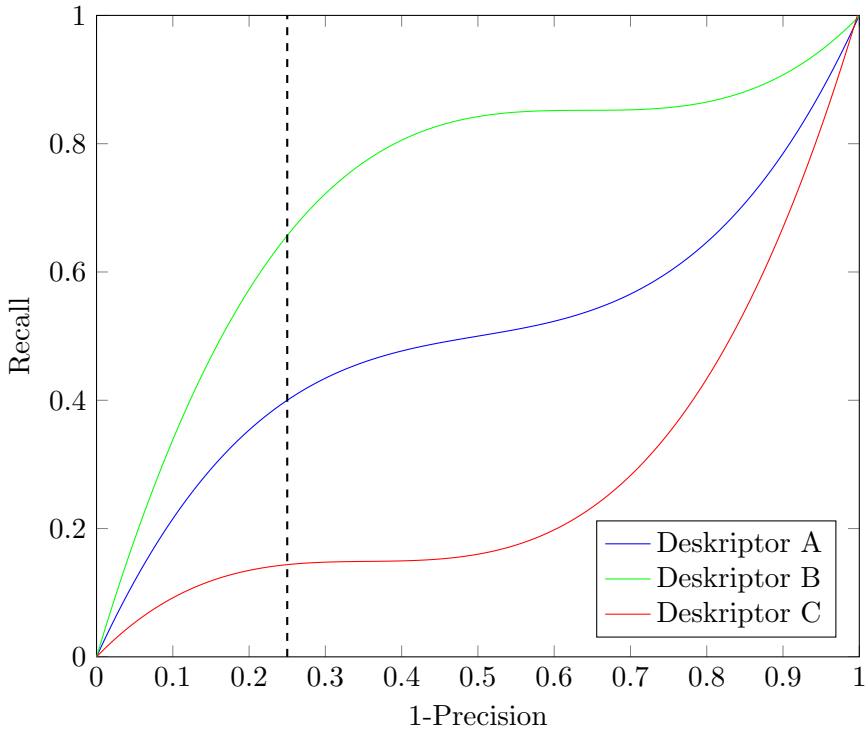


Abbildung 2.17: Beispielhafter graphischer Vergleich dreier Deskriptoren. Betrachtet man die Werte für 1–Precision = 0.25, so erreicht A einen Recall von knapp 0.4, während B mit etwa 0.65 darüber liegt, d. h. weniger falsch negative Beispiele zählt. Der im Vergleich schlechteste Klassifikator ist C, dessen Kurve sich durchgängig unter denen der anderen hält. Einen hohen Recall kann dieser Deskriptor somit nur mit einhergehender hoher Ungenauigkeit erreichen.

Eine Gegenüberstellung dieser Kurven zweier verschiedener Deskriptoren erlaubt eine Beurteilung, welcher Deskriptor die Schlüsselpunkte besser beschreibt. Das wiederum bedeutet, dass diese im Referenzbild besser wiedergefunden werden. Ein sehr guter Deskriptor erreicht früh, d. h. bei einem geringen Schwellwert θ einen hohen Recall nahe eins, wobei die Ungenauigkeit, also die Zahl der falsch Positiven, gering bleibt. Anhand der Kurven in Abbildung 2.16 wird das plausibel: Die Zahl der falsch negativen Beispiele bleibt genau dann klein, wenn die Varianz der grünen Kurve gering ist und diese weit links liegt. Ein schlechter Klassifikator hingegen produziert Merkmale, bei denen die Zahl der falsch negativen Beispiele mit wachsendem Schwellwert θ nur sehr langsam gegen null konvergiert. Dadurch bleibt der Recall niedrig, die Kurve liegt also unter der eines besseren Deskriptors. Je mehr falsch positive Beispiele hinzukommen, desto höher wird die Ungenauigkeit. Beispielhafte Kurven, welche das beschriebene Verhalten veranschaulichen, finden sich in Abbildung 2.17.

3 Farbe und Farbräume

In diesem Grundlagenkapitel soll in das Thema Farbe eingeführt werden. Zunächst erfolgt eine Bestimmung, was unter dem Phänomen Farbe zu verstehen ist. Im Anschluss daran wird die Darstellung von Farbe in Form von Farbräumen erläutert, wobei verschiedene Farbräume vorgestellt werden. Das Kapitel wird durch einen Einblick abgeschlossen, wie die SIFT in der Vergangenheit um Farbe erweitert wurde.

3.1 Farbe und deren Wahrnehmung

Das elektromagnetisches Spektrum lässt sich anhand der Wellenlänge der Strahlung in verschiedene Bereiche unterteilen. So findet sich beispielsweise im Bereich von 10pm bis 1nm die Röntgenstrahlung, welche in der Medizin eine wichtige Bedeutung in der Diagnostik spielt, oder etwa von 10m bis 10km die Radiowellen, welche für die Übertragung von Hörfunk und Fernsehen genutzt werden. Im Bereich von 400nm bis 750nm, also zwischen den genannten Beispielen liegend, findet sich das *Licht*. Licht bezeichnet den für den Menschen sichtbaren Bereich des elektromagnetischen Spektrums. In Abhängigkeit der Wellenlänge lassen sich innerhalb dieses Bereiches verschiedene *Farben* unterscheiden: Bei 400nm findet sich die Farbe violett, mit zunehmender Wellenlänge findet sich blau, grün und schließlich rot bei 700nm, wie dargestellt in Abbildung 3.1.

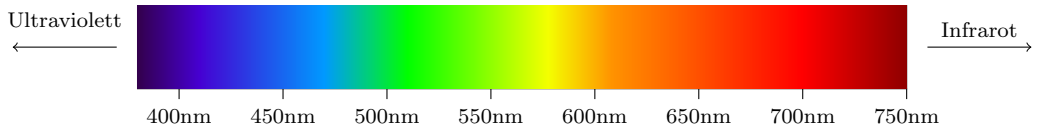


Abbildung 3.1: Dargestellt ist das elektromagnetische Spektrum im Bereich von 400nm bis 700nm, welches vom Menschen als Licht in Form von Farbe wahrnehmbar ist.

Farbe hingegen ist ein Sinneseindruck, welcher Aufgrund der wahrgenommenen Wellenlängen vom menschlichen Gehirn erzeugt wird und ist nicht "im Licht enthalten". Diese Tatsache wird durch ein interessantes Experiment deutlich: Zerlegt man weißes Licht anhand eines Prismas, so fehlt in dem sich auffächernden Spektrum das Magentarot. Dennoch können wir Magentarot wahrnehmen und unterscheiden es als eigene Farbe etwa von Blau oder Rot. Zu erklären ist dies durch eine Betrachtung, wie die Farbwahrnehmung des Menschen funktioniert, was dieser Abschnitt darlegen soll.

Bei der Besprechung des FREAK-Deskriptors und dessen biologischer Motivation wurden bereits Stäbchen und Zapfen als Sinneszellen im menschlichen Auge unterschieden. Von den für das Farbsehen zuständigen Zapfen besitzt der Mensch drei Arten, was ihn zu einem *Trichromaten* macht. Die drei Zapfenarten sind für verschiedene Bereiche von Wellenlängen zuständig: Eine Art ist empfindlich für Violett und Blau, eine für Grün und eine weitere für Orangerot. Daraus ergeben sich die drei Grundfarben Blau, Grün und Rot. Werden nun zwei verschiedene Zapfen gleichzeitig angesprochen, so lässt das den Eindruck der Sekundärfarben entstehen: Eine Mischung von Blau und Grün ergibt Cyan, Grün und Rot ergibt Gelb und Blau und Rot ergibt Magenta. Betrachtet man erneut Abbildung 3.1, so lassen sich die ersten beiden Sekundärfarben tatsächlich zwischen den zugehörigen Primärfarben liegend entdecken. Für Magenta hingegen ist dies nicht möglich. Dies verdeutlicht, dass Farbe keine dem Licht inhärente Eigenschaft, sondern ein vom Gehirn erzeugter Sinneseindruck ist. Magenta wird genau dann gesehen, wenn die Zapfen für Rot und Blau gleichzeitig angesprochen werden.
Ausgehend von den vom Menschen wahrgenommenen Farben lassen sich verschiedene *Farbräume* definieren, welche eine Ordnung der Farben vorschlagen. Im Folgenden sollen die für diese Arbeit relevanten Farbräume vorgestellt werden.

3.2 Darstellung von Farbe durch Farbräume

Anhand von Farbräumen (auch als *Farbmodelle* bezeichnet) ist es möglich, Farben zu ordnen. Da der Mensch drei verschiedene Zapfentypen zur Wahrnehmung von Farbe nutzt, lassen sich Farben in Form eines dreispaltigen Vektor beschreiben (Plataniotis und Venetsanopoulos, 2000). Dazu jedoch gibt es verschiedene Möglichkeiten, in der Literatur werden folgende Familien von Farbmodellen unterschieden:

Psychophysikalische Modelle entsprechen der Art und Weise, wie Menschen Farbe psychologisch wahrnehmen. Ein Beispiel ist der *HSV-Farbraum*, welcher Farbe in Form des Farbtons, der Sättigung und der Helligkeit beschreibt.

Physiologisch inspirierte Modelle sind motiviert durch physiologische Aspekte der Farbwahrnehmung. So wie der Mensch drei Zapfen zur Unterscheidung der Wellenlänge des Lichtes besitzt, so basiert auch der *RGB-Farbraum* auf drei Grundfarben.

Gegenfarbmodelle stellen Farbe anhand eines Verhältnisses von Primärfarben dar. Beispielsweise lassen sich Farben so als Rot–Grün und Blau–Gelb Kontrast kodieren. Sie beruhen auf Experimenten, wie ein menschlicher Betrachter Farbe unterscheidet und enthalten damit sowohl physiologische, als auch psychophysikalische Aspekte.

Die Modelle berücksichtigen auf verschiedene Art und Weise, wie der Mensch Farbe wahrnimmt, sei es psychologisch oder physiologisch. Dies erscheint sinnvoll, da "Farbe" bereits als menschliche Empfindung definiert wurde. Darüber hinaus gibt es außerdem Farbräume, welche auf rein physikalischen Messungen der Spektrums basieren. Diese sind jedoch aus dem oben genannten Grund in dieser Arbeit nicht weiter von Interesse. Im Folgenden sollen verschiedene Farbmodelle eingeführt werden, welche im späteren Verlauf der Arbeit Anwendung finden werden. Begonnen wird dabei mit dem RGB-Modell.

RGB Der RGB-Farbraum beschreibt Farbe als Zusammensetzung der drei Grundfarben Rot, Grün und Blau. Motiviert ist dies durch die physiologische Erkenntnis, dass die unterschiedlichen Zapfentypen im menschlichen Auge für diese Wellenbereiche empfindlich sind, und gleichzeitig gereizt werden können. Eine Farbe ist somit als dreispaltiger Vektor $\mathbf{v} = (R, G, B)$ darstellbar, wobei der Wertebereich der einzelnen Komponenten sowohl nach unten als auch nach oben begrenzt sind. Die Elemente des Vektors geben den Grad der Aktivierung der jeweiligen Grundfarbe an. Man spricht in diesem Zusammenhang auch von drei *Farbkanälen*. Das führt auf das Prinzip der *additiven Farbmischung*: Aktiviert man die genannten Grundfarben einzeln, so werden genau diese vom Menschen wahrgenommen. Darüber hinaus ist es möglich, Farben zu mischen, indem die Grundfarben in einem bestimmten Verhältnis zueinander aktiviert werden. Beispielsweise ist Magenta in Form des Vektors $(1, 0, 1)$ darstellbar, also maximal aktives Blau und Rot, jedoch kein Grünanteil. Werden alle Grundfarben maximal aktiviert, so sieht der Mensch Weiß, ist keine Farbe aktiviert, so wird Schwarz gesehen. Darstellbar ist der RGB-Farbraum in Form eines Würfels, dessen Ecken die Primärfarben, die Sekundärfarben, sowie Weiß und Schwarz repräsentieren, eine Darstellung findet sich in Abbildung 3.2. Sind alle Farben immer zu einem gleichen Teil aktiviert, so erhält man verschiedene Grautöne, der Grad der Aktivierung steuert dabei dessen

Helligkeit. Im RGB-Würfel liegen diese somit auf der Geraden $(0, 0, 0)$ bis $(1, 1, 1)$, die Diagonale, welche Schwarz mit Weiß verbindet.

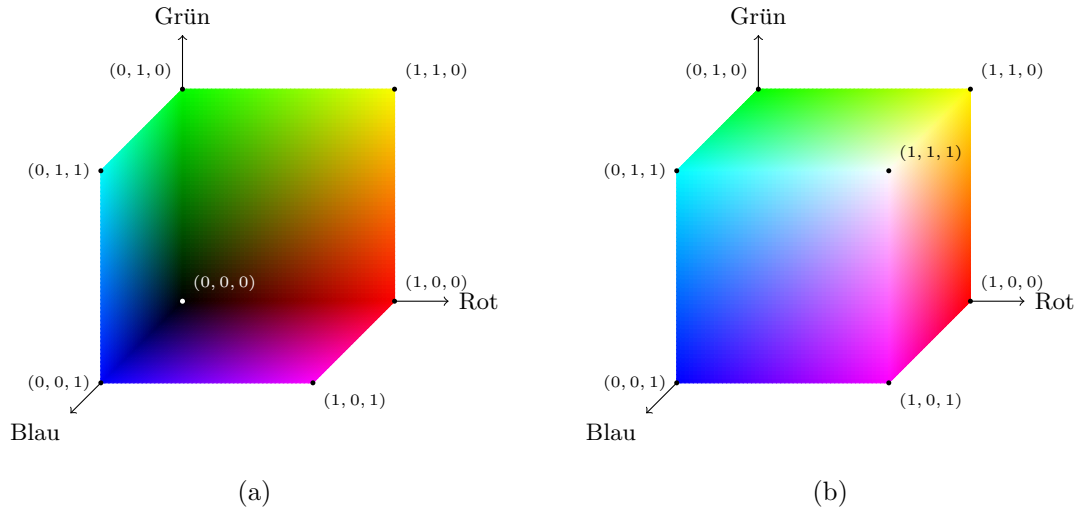


Abbildung 3.2: Der RGB-Farbwürfel in der Innenansicht (a) sowie in der Außenansicht (b). Im Nullpunkt sind alle Grundfarben inaktiv, was gleichbedeutend ist mit Schwarz. Anhand der drei Achsen lässt sich die Aktivierung der jeweiligen Farbe verbildlichen.

In Computersystemen ist RGB der meist eingesetzte Farbraum, da Ausgabemedien wie beispielsweise Bildschirme oder Projektoren Farbe ebenfalls additiv reproduzieren. Dazu werden LEDs sehr dicht nebeneinander platziert und je nach gewünschter Farbe aktiviert. Ein 24-Bit Farbwert sieht für jede Grundfarbe 8-Bit vor, was den Werten von 0 (keine Aktivierung) bis 255 (maximale Aktivierung) entspricht. Verschiedene digitale Kameras funktionieren auf ähnliche Weise, allerdings wird dabei nicht für jeden Pixel des Bildes die Farbinformation optisch ermittelt. Genutzt wird beispielsweise das sogenannte *Bayer-Muster* (engl. *Bayer-Pattern*), dargestellt in Abbildung 3.3.

Da die Grün-empfindlichen Zapfen des menschliches Auges einen höheren Anteil beim Kontrastsehen haben als Rot und Blau, ist der Grün-Sensor im Bayer-Muster häufiger vertreten. Erreicht wird damit eine für die menschliche Empfindung höhere Bildqualität. Das Problem der Interpolation fehlender Farbe, also das Ermitteln des Grün- oder Blauanteils für einen Punkt, für den nur Rot optisch ermittelt wurde, ist als *Demosaicing*

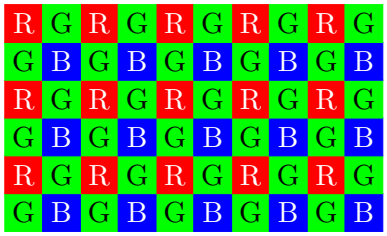


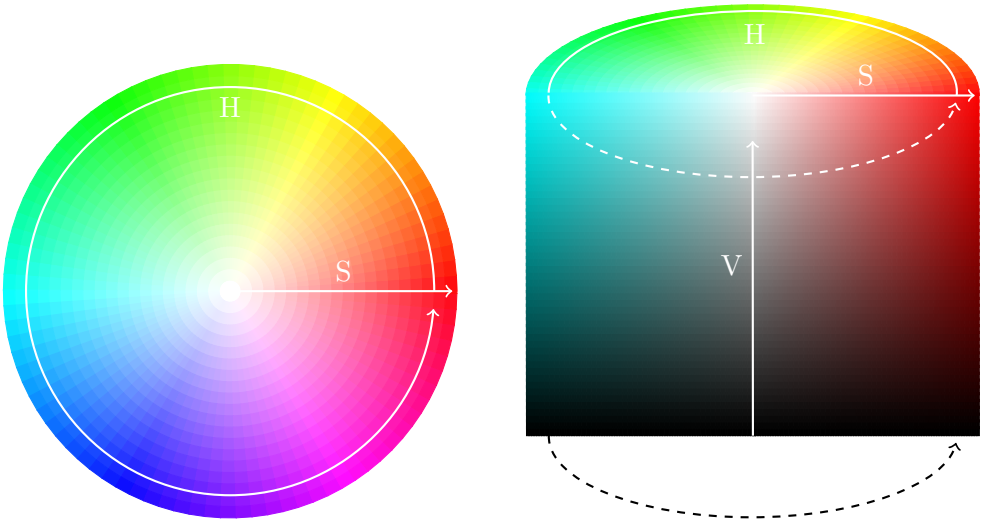
Abbildung 3.3: Darstellung des Bayer-Musters, wie es in Digitalkameras Anwendung findet. Die einzelnen Quadrate sind Sensoren, welche die verschiedenen Grundfarben Rot, Grün und Blau aufzeichnen. Mithilfe von speziellen Algorithmen werden dann alle Farbinformationen für die Punkte interpoliert, welche nicht in der jeweiligen Farbe vorliegen. Grün ist aufgrund seiner Bedeutung für das Kontrastsehen stärker vertreten als Blau und Rot.

bekannt. Für weitere Information sei auf die Fachliteratur verwiesen, einen Einstieg bietet beispielsweise Kimmel (1999). An dieser Stelle soll ein einfaches Verständnis darüber ausreichen, wie digitale Farbbilder entstehen.

HSV HSV steht für die englischen Begriffe *Hue* (Farbton), *Saturation* (Sättigung) und *Value* (Helligkeit). Die Kodierung von Farbe anhand dieser Eigenschaften kommt eher der Ordnung nahe, wie Menschen Farbe benennen, weswegen es den Psychophysikalischen Modellen zuzuordnen ist. Eine Farbbezeichnung wie "ein sattes, dunkles Grün" lässt sich mithilfe des RGB-Farbraums schlecht abgrenzen, da die Eigenschaften "satt" und "dunkel" durch den Anteil der von Grün verschiedenen Farben Rot und Blau eingeregelt werden müssen. Dennoch können Menschen sich unter dieser Farbe etwas vorstellen. Im HSV-Raum ist diese Farbe sehr viel leichter zu beschreiben: Der Farbton ist Grün, die Sättigung ist hoch, das Grün hat also einen geringen Weißanteil, und die Helligkeit ist eher gering.

Der Farbton wird in Form eines Winkels zwischen 0° und 360° beschrieben, Sättigung und Helligkeit hingegen als Wert zwischen 0 und 1. Das führt zu der Darstellung des HSV-Farbraums als Zylinder, wie er in Abbildung 3.4 zu finden ist: Die Deckfläche enthält alle Farben entsprechend des Spektrums und führt Blau über Magenta wieder auf Rot zurück. Der Abstand eines Punktes zur Mitte, parallel zur Deckfläche, kodiert die Sättigung der Farbe. Der Abstand eines Punktes zur Grundfläche und somit die Höhe des Zylinders kodiert die Helligkeit.

Alternativ kann der HSV-Farbraum auch als Kegel dargestellt werden, wobei die Kegel-



(a) (b)

Abbildung 3.4: (a) zeigt die Deckfläche des HSV-Zylinders, welcher den Farbton in Form eines Winkels sowie die Sättigung kodiert. Die Helligkeit einer Farbe wird erst in einem senkrechten Schnitt entlang der Geraden $H = 0°$ durch den Zylinder sichtbar, wie er in (b) abgebildet ist.

spitze Schwarz repräsentiert. Das deutet auf verschiedene Probleme des HSV-Farbraumes hin: Zum Einen macht es wenig Sinn, einen Farbton oder eine Sättigung für eine Farbe mit der Helligkeit $V = 0$ anzugeben, die gesamte Grundfläche des Zylinders ist an dieser Stelle schwarz. Zum Anderen besteht das Problem, dass eine Farbe, je geringer ihr Sättigungswert wird, weniger unterscheidbar ist von anderen Farben. An dieser Stelle spricht man spricht man auch von der *Instabilität des Farbtons* in der Umgebung der Grauwertachse (van de Weijer und Schmid, 2006). Die genauen Formeln zur Umrechnung von RGB nach HSV sind im Anhang A.2 gegeben.

Gegenfarben Die Beobachtung, dass Farben als Gegensatzpaare wahrgenommen werden, geht auf Hering (1878) zurück. Er bemerkte, dass Menschen sich kein "gelbes Blau" und kein "rotes Grün" vorstellen können. Das ließ ihn zu der Annahme gelangen, diese Farben seien durch chemische Prozesse so gekoppelt, dass sie sich bei Überlagerung gegenseitig aufheben. Er definierte die vier Grundfarben als die Gegensatzpaare Rot–Grün und Blau–Gelb, welche heute auch als *Urfarben* bezeichnet werden. Schwarz die Gegenfarbe von Weiß. Auch die Entstehung von *Nachbildern* sprach für seine Theorie:

Betrachtet man lange ein Farbbild und anschließend eine weiße Fläche, so scheint das zuvor gesehene Bild in seinen jeweiligen Gegenfarben für einen weiteren Moment sichtbar. Für Hering war dieses Phänomen durch ein chemisches Gleichgewicht erklärbar, welches sich erst allmählich wieder einstellt. Tatsächlich konnte die Theorie von Hering bestätigt werden Klinke et al. (2005). Abbildung 3.5 stellt schematisch dar, wie die Farbinformationen im visuellen System des Menschen verschaltet werden. Das Schema entspricht den von Hering beobachteten Gegenfarben.

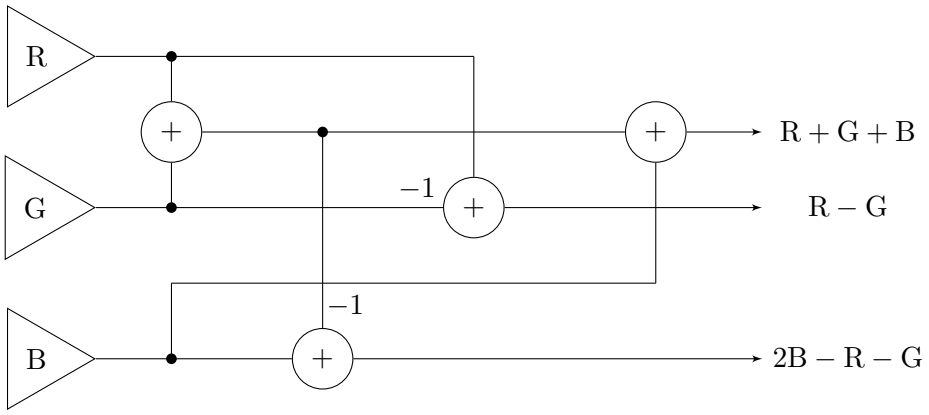


Abbildung 3.5: Die Verschaltung der Farbinformationen im menschlichen Sehsystem entspricht der Gegenfarbtheorie Herings, welche Farben als Rot–Grün, Blau–Gelb und Schwarz–Weiß Kontrast kodiert.

Gegenfarbmodelle eignen sich auf Grund ihres Bezuges auf die wahrgenommen Farbunterschiede, um ein Abstandsmaß zwischen Farben zu definieren. Dabei soll messbar gemacht werden, wie stark sich zwei Farbreize voneinander unterscheiden: Blau und Hellblau sollen also beispielsweise einen geringeren Abstand haben als Blau zu hellgrün. Genutzt wird hierfür der L*a*b* Farbraum, welcher wiederum den XYZ-Farbraum der *Commission internationale de l'éclairage*[5] (CIE) als Ausgangspunkt nutzt.
Im Jahr 1932 führte die CIE Experimente zu wahrgenommenen Farbunterschieden durch. Dabei wurde beobachtet, dass im Bereich Blau-Grün gemischter Farben teilweise Rot von diesen "subtrahiert" werden müsste, damit ein Beobachter diese als gleich zu Cyan

[5]Die CIE ist eine in Wien ansässige Gesellschaft zur Förderung der Zusammenarbeit von Wissenschaft auf dem Gebiet der Beleuchtung. Siehe `http://www.cie.co.at/`.

als Spektralfarbe empfindet. Das kann erreicht werden, indem der Spektralfarbe wiederum rot hinzugefügt wird. Zu erklären ist diese Phänomen durch die Physiologie des Auges: Die Wahrnehmung der Farbe Cyan kann durch "reines", also spektrales Licht der Wellenlänge 480nm erreicht werden, oder aber durch Reizung mit blauem und grünen Licht gleichzeitig. In beiden Fällen werden die Zapfen für Blau und Grün angesprochen. Der Grund für die Wahrnehmung eines Rotanteils bei gemischten Farben liegt darin, dass die Reizung mit grünem Licht die roten Zapfen ebenfalls leicht anspricht, da sich hier die Bereiche der Empfindlichkeit überlappen. Bei spektralem Licht ist das nicht der Fall. Farben, welche als gleich empfunden werden, jedoch spektral unterschiedlich zusammengesetzt sind, bezeichnet man als *metamere Farben*.
Um diesem Umstand gerecht zu werden, definierte die CIE den XYZ-Farbraum. Dieser stellt Farben in Form einer X-, einer Y- und einer Z-Komponente dar und umfasst alle wahrnehmbaren Farben, und somit auch die Spektralfarben, ohne dass diese durch negative Komponenten zusammengesetzt werden müssen. Die Umwandlung ist anhand einer Lineartransformation möglich, welche im Anhang A.2 zu finden ist.
Der euklidische Abstand zweier XYZ-kodierter Farben ist jedoch nicht ohne Weiteres auf einen vom Menschen wahrgenommenen Abstand zu übertragen. Diese Eigenschaft besitzt der L*a*b* Farbraum, welcher wiederum auf dem XYZ-Farbraum basiert. Die Vorschriften zur Umrechnung von RGB nach L*a*b* sind ebenfalls im Anhang zu finden. Die Komponente L* zeigt die Luminanz einer Farbe an, die a-Achse enthält von -150 bis 100 ungefähr die Gegenfarben Grün und Rot, während sich auf der b-Achse von -100 bis 150 ungefähr die Farben Blau und Gelb gegenüber liegen. L*a*b* verzerrt den XYZ-Farbraum in einer Art und Weise, dass der euklidische Abstand zweier L*a*b* kodierter Farben tatsächlich dem Grad des vom Menschen wahrgenommenen Unterschiedes entspricht. Verwendet wird dazu das sogenannte ΔE^* Maß, welches für zwei gegebene Farben (L_1^*, a_1^*, b_1^*) und (L_2^*, a_2^*, b_2^*) wie folgt definiert ist:

$$\Delta E^* = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2}$$

Die Auswahl und der Detailgrad der hier beleuchteten Farbräume erfolgte mit dem Ziel, deren Grundideen darzustellen, um auf diese im späteren Verlauf der Arbeit aufbauen zu können. Eine vergleichende Darstellung darüber, wie einzelne Kanäle der eingeführten

Farbräume ein bestimmtes Bild kodieren, findet sich in Abbildung 3.6. Es existieren darüber hinaus weitere Farbräume, welche hier nicht vorgestellt wurden, da dies den Umfang der Arbeit sprengen würde und zum Verständnis selbst nicht nötig ist.

3.3 Übertragungen der SIFT nach Farbe

Von verschiedenen Autoren gibt es Ansätze, die in Kapitel 2.5.1 vorgestellte SIFT um Farbinformation zu erweitern.[6] Da es sich dabei um ein aktuelles Forschungsthema handelt, erhebt die die folgende Liste keinen Anspruch auf Vollständigkeit. Vielmehr soll sie zur Orientierung dienen und eine Grundlage schaffen für die Erweiterung des FREAK-Deskriptors. Die folgenden Ausführungen orientieren sich dabei an einer Veröffentlichung von van de Sande et al. (2010), welche verschiedene Ansätze der Erweiterung der SIFT um Farbinformation miteinander vergleicht.

HSV-SIFT Bosch et al. (2008) nutzen den HSV-Farbraum und berechnen den Deskriptorvektor aller drei Kanäle des Farbmodells. Diese werden im Anschluss konkateniert, was einen 3×128-dimensionalen Deskriptorvektor ergibt. Die angesprochenen Probleme, wie etwa die Instabilität des Farbtons im Bereich der Grauwertachse im HSV-Farbraum, werden von den Autoren nicht gesondert behandelt.

Hue-SIFT Die Autoren van de Weijer und Schmid (2006) hingegen schlagen mit Hue-SIFT ein Verfahren vor, in welchem die Instabilität des Farbtons Beachtung findet. Anhand der Information zur Sättigung wird eine Aussage über die Stabilität des Farbtons getroffen, welche anschließend als Gewicht bei der Berechnung der Histogramme einfließt.

Opponent-SIFT Für die Berechnung des Deskriptorvektors schlagen van de Sande et al. (2010) vor, das Gegenfarbmodell zu verwenden, indem auch hier die Deskriptoren der einzelnen Kanäle konkateniert werden. In Bezug auf verschiedene photometrische

[6] Die Erweiterung des SURF-Deskriptors um Farbinformation scheint für die Forschung nicht von großem Interesse zu sein, zumindest wenn man die Zahl der Veröffentlichungen betrachtet. Das mag mit Ähnlichkeit der Deskriptoren SIFT und SURF zusammenhängen, inwieweit die Ergebnisse für SIFT jedoch auf SURF übertragbar sind, muss untersucht werden.

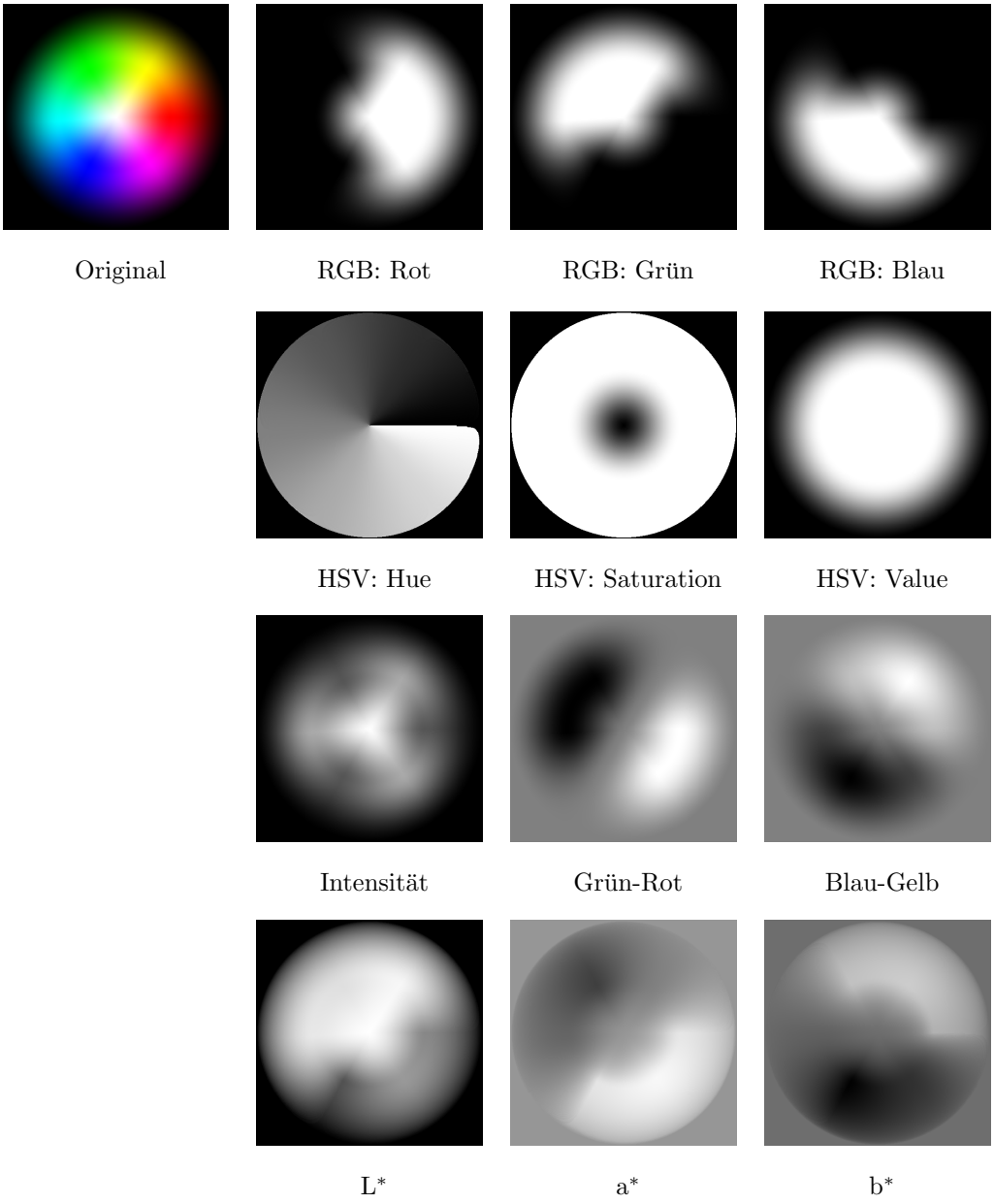


Abbildung 3.6: Links oben findet sich das Originalbild, rechts daneben sind die einzelnen Kanäle im RGB-Farbraum dargestellt. Darunter sind die Kanäle des HSV-Raums abgebildet, wiederum darunter die des Gegenfarbmodells. Ganz unten finden sich einzelnen Kanäle des Originalbildes im L*a*b* Farbraum. Deutlich wird, dass in diesem Blau als dunkelste Farbe gesehen wird, während das in den anderen Farbräumen nicht der Fall ist. In jedem Bild wurden die Werte in den darstellbaren Bereich von 0 bis 255 gewandelt.

Transformationen, beispielsweise eine Änderung der Helligkeit eines Bildes, spricht er die Empfehlung aus, Opponent-SIFT als den diesbezüglich am robustesten Deskriptor einzusetzen.

RGB-SIFT Bei der RGB-SIFT, ebenfalls vorgeschlagen von van de Sande et al. (2010), wird der SIFT-Deskriptor für den Rot-, den Grün- und den Blau-Kanal einzeln berechnet und dann konkateniert, was wiederum einen Deskriptor mit 3×128 Dimensionen hervorbringt.

Darüber hinaus gibt es weitere Ansätze, die SIFT um Farbinformation zu erweitern. Die hier gezeigte Auswahl führt für jede der vorgestellten Familien von Farmodellen jeweils einen Repräsentanten auf. HSV-SIFT deckt mit dem HSV-Farbraum ein Psychophysikalisches Modell ab, RGB-SIFT ein Physiologisch inspiriertes Modell. Mit Opponent-SIFT sind auch die Gegenfarbmodelle mit einem Repräsentanten vertreten. Allen gemein ist die Eigenschaft, dass eine Konkatenation der Deskriptorvektoren einzelner Kanäle als Ausgangsprinzip genutzt wird. Diese Art, den Deskriptorvektor zu berechnen, wird in dem ersten Experiment in Abschnitt 4.2 aufgegriffen.

4 Experimente

Auf Grundlage der vorgestellten Farbräume sollen in diesem Abschnitt der Arbeit verschiedene Experimente durchgeführt werden, wobei das Ziel ist, den FREAK-Deskriptor um Farbe zu erweitern. Es soll untersucht werden, welche der in Abschnitt 3.2 vorgestellten Repräsentationen von Farbe sich dabei am besten eignet. Im ersten Ansatz werden dabei die Deskriptorvektoren verschiedener Kanäle berechnet und konkateniert, die Farbe also direkt in den Deskriptorvektor integriert. Im zweiten Ansatz wird die Farbe als unabhängiges Attribut zu dem FREAK-Deskriptor des Intensitätsbildes hinzugefügt.

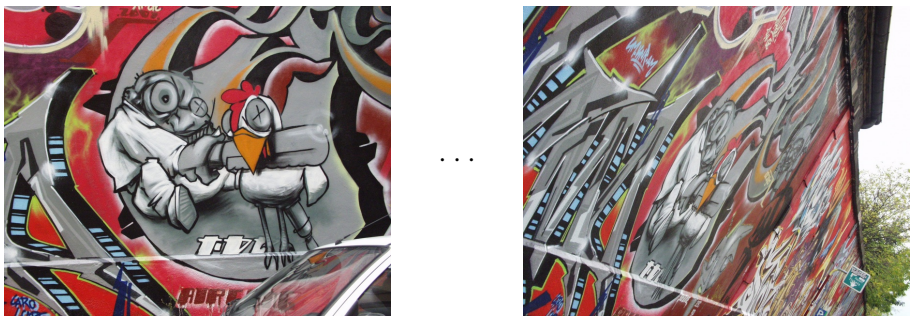
4.1 Der Datensatz zur Evaluation

Für die Evaluation der Experimente wird der Datensatz von Mikolajczyk und Schmid (2005) genutzt, welcher sich zu einem Standard für den Vergleich von Deskriptoren entwickelt hat. Enthalten sind darin acht verschiedene Bildreihen mit jeweils sechs verschiedenen Bildern, wie dargestellt in Abbildung 4.1. Das erste Bild eines Motivs ist dabei jeweils das Originalbild, während die anderen fünf verschiedenen Transformationen unterzogen sind, bezüglich derer die Robustheit eines Deskriptors getestet werden kann. Die Transformationen sind Rotation, Skalierung, perspektivische Verzerrung, Weichzeichnung, Beleuchtungsänderung und Artefakte, wie sie bei JPEG-komprimierten Bildern auftreten.
Hinsichtlich der Farbe unterscheiden sich die Bilder stark. Mit der Bildreihe `boat` ist darüber hinaus ein Bild enthalten, welches keine Farbinformation enthält und als einzigen Kanal Graustufen bietet. In anderen Bildern dominieren mitunter verschiedene Farben, so etwa in der Bildreihe `bikes` Blau, in `wall` Rot und Grau und in `bark` Grün und Grau.
Die perspektivische Änderung eines jeden Bildes im Bezug auf das Originalbild ist in Form einer Homographie gegeben. Für jeden Punkt auf dem Original kann somit berechnet werden, an welchen Koordinaten dieser nach der Transformation zu finden ist. Der Grad der Transformation schwankt und reicht von keiner Transformation, in welcher die angegebene Homographie die Identität ist, bis hin zu starker perspektivischer Veränderung, beispielsweise in der Bildreihe `graf` (siehe Abbildung 4.1(d)).

(a) `bark`: Skalierung und Rotation

(b) `bikes`: Weichzeichnung

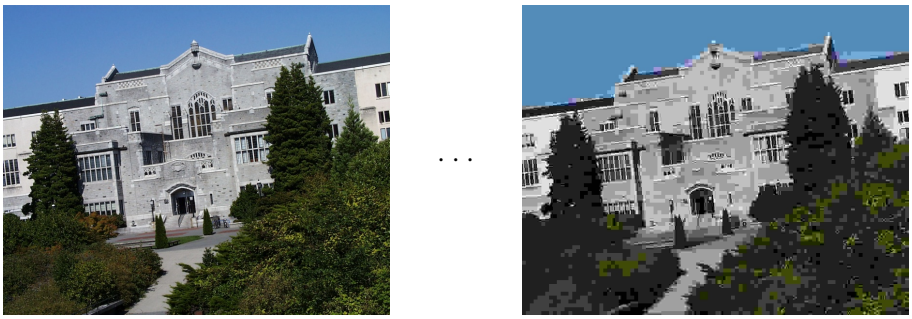
(c) `boat`: Skalierung und Rotation



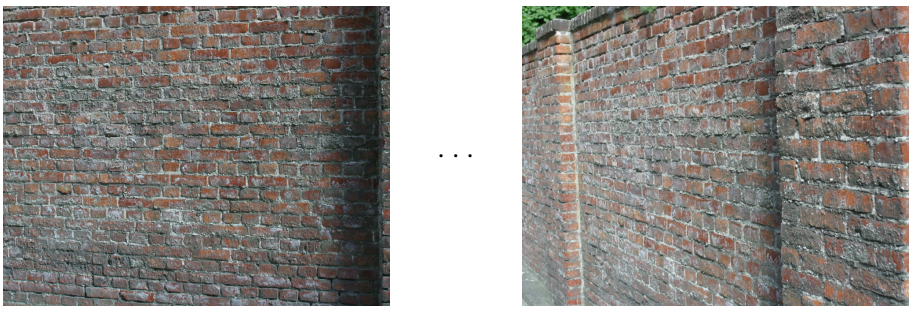
(d) `graf`: Perspektivische Verzerrung

(e) `leuven`: Beleuchtungsänderung

(f) `trees`: Weichzeichnung



(g) `ubc`: JPEG-Kompression



(h) `wall`: Perspektivische Verzerrung

Abbildung 4.1: Der verwendete Datensatz enthält acht verschiedene Bildreihen mit jeweils sechs unterschiedlich stark transformierten Bildern. Links zu sehen ist das Originalbild, rechts das am stärksten durch die genannten Transformationen veränderte Bild der jeweiligen Reihe.

4.2 Farbe integriert durch Konkatenation von Deskriptorvektoren

Im ersten Experiment werden Farbdeskriptoren durch Konkatenation von Deskriptorvektoren verschiedener Farbkanäle gebildet. Welches Farbmodell zur Darstellung der Farbe sich dabei am besten eignet, ist Gegenstand der Untersuchung.

4.2.1 Getestete Deskriptoren

In diesem Experiment werden die Deskriptoren, analog zu den in Abschnitt 3.3 vorgestellten Ansätzen die SIFT um Farbe zu erweitern, auf den einzelnen Farbkanälen berechnet und anschließend konkateniert. Es werden fünf verschiedene Ansätze verglichen, aufgeführt sind diese in Tabelle 4.1. Die Namen der dadurch entstehenden Deskriptoren geben Aufschluss über die Farbkanäle, welche zu dessen Bildung genutzt werden: rgbFREAK beispielsweise bezeichnet den Deskriptor, welcher sich aus drei 512-stelligen, binären FREAK-Deskriptoren, berechnet auf den Kanälen Rot, Grün und Blau eines Bildes zusammensetzt. Für den Deskriptor oppFREAK werden die zwei Farbkanäle des Gegenfarbenmodells genutzt, oppiFREAK berücksichtigt als einen weiteren Kanal die Intensität. Der hueFREAK-Deskriptor wird auf dem Farbwinkel berechnet, während hsvFREAK mit dem Farbwinkel, der Sättigung und der Helligkeit alle drei Kanäle des HSV-Farbraums zur Bildung des Deskriptors nutzt.
Ein Ziel ist hierbei explizit, den FREAK-Deskriptor so wenig wie möglich zu verändern. So wird neben Anordnung und Größe der Abtastpunkte auch die Reihenfolge der Vergleiche so beibehalten, wie von den Autoren Alahi et al. (2012) ermittelt und empfohlen. Die paarweisen Farbvergleiche bei der Bildung des Deskriptorvektors folgen somit ebenfalls dem vorgestellten "von grob nach fein"-Prinzip (vgl. Abschnitt 2.5.3).

4.2.2 Durchführung der Experimente

Die Deskriptoren werden nach dem in Abschnitt 2.7 vorgestellten Verfahren evaluiert, welches die Werte Recall und 1-Precision ermittelt und die Kurve der gegeneinander aufgetragenen Werte als Diskussionsgrundlage nutzt. Da untersucht werden soll, inwieweit eine Erweiterung um Farbinformation eine Verbesserung des FREAK-Deskriptors bedeuten kann, werden diese Kurven mit denen des ursprünglichen Deskriptors von Alahi et al. (2012) verglichen.

Name	Farbmodell	Verwendete Kanäle
FREAK	Graustufen	Graustufen
rgbFREAK	RGB	Rot, Grün, Blau
oppFREAK	Gegenfarben	Grün-Rot, Blau-Gelb
oppiFREAK	Gegenfarben	Grün-Rot, Blau-Gelb, Intensität
hueFREAK	HSV	Farbton
hsvFREAK	HSV	Farbton, Saturation, Value

Tabelle 4.1: Auflistung der getesteten Deskriptoren. Der Name des jeweiligen Deskriptors gibt Aufschluss über das verwendete Farbmodell und die Kanäle, welche zu dessen Bildung konkateniert werden.

Schlüsselpunkte werden durch den in Abschnitt 2.4 vorgestellten Fast-Hessian Detektor auf dem nach Grauwerten konvertierten Eingabebild ermittelt, Details zu der Konvertierung finden sich im Anhang A.2. Der FREAK-Deskriptor selbst definiert keinen eigenen Detektor, somit muss eine Entscheidung bezüglich der Wahl eines Detektors getroffen werden. Zwar hängt die Leistung eines Deskriptors auch von den der Art der Schlüsselpunkte ab und es macht einen Unterschied, ob beispielsweise Ecken (engl. “corners”) oder Bereiche mit starker Varianz der Textur (engl. “blobs”) gefunden werden (siehe Abschnitt 2.4.3). Im Vergleich von Deskriptoren ist dieser Unterschied jedoch nicht ausschlaggebend, wie Alahi et al. (2012) feststellen, die Wahl des Detektors stellen die Autoren somit frei. In dieser Arbeit fiel die Wahl auf den Fast-Hessian Detektor, da dieser anschaulich die Idee von *Extrempunkten als Schlüsselpunkte* darstellt und darüber hinaus anhand eines Schwellwertes einfach zu konfigurieren ist. Der Schwellwert wurde dabei so eingestellt, dass zwischen 900 und 1000 Schlüsselpunkte auf den Ausgangsbild detektiert werden.

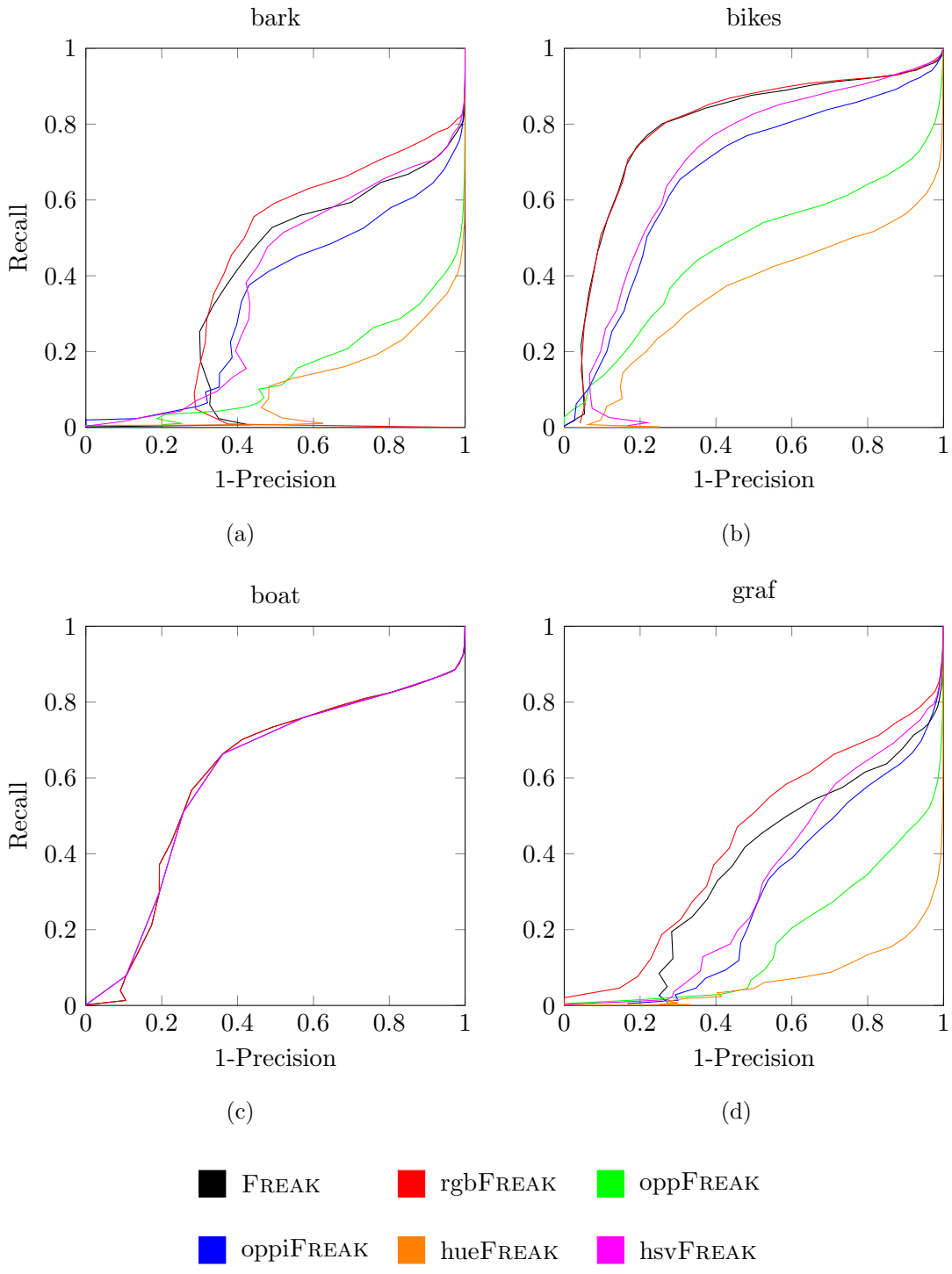
Analog zu Alahi et al. (2012) werden nicht alle Bilder einer Reihe des Datensatzes von Mikolajczyk und Schmid (2005) verwendet, sondern sich auf die ersten vier Bilder beschränkt. Begründet wird das durch den Umstand, dass beispielsweise die perspektivische Verzerrung in späteren Bildern einer Reihe zu stark wird, als dass noch sinnvolle Erkenntnisse aus den Kurven gewonnen werden können, da bei zu starker Transformation kaum noch Punkte wiedergefunden werden. Zwischen den einzelnen Bildern eines Motivs wird nicht unterschieden, da in der Folge die vierfache Zahl an zu betrachtenden

Kurven entstehen würde, ohne dass dies einen Erkenntnismehrwert generiert. Die Gesamtzahl der möglichen Übereinstimmungen einer Bildreihe entspricht damit der Summe aller möglichen Übereinstimmungen der transformierten Bilder zum Originalbild.
Die Länge der getesteten Deskriptoren reicht von 512 Bit, etwa mit hueFREAK, bis zu $3 \cdot 512 = 1536$ Bit, beispielsweise mit rgbFREAK. Somit schwankt auch der Wertebereich des Hamming-Abstands, was berücksichtigt werden muss. Zur Ermittlung der Kurve werden im Experiment für jeden Deskriptor 100 Punkte erhoben. Der Schwellwert θ beginnt bei 0 und wird für jeden Deskriptor somit in jedem Schritt um ein hundertstel seines maximalen Abstandes erhöht. Das entspricht der Vergleichsstrategie des Schwellwertverfahrens (siehe Abschnitt 2.6): In jedem Schritt wird dabei für jedes transformierte Bild jeder einzelne Deskriptorvektor mit jedem Deskriptorvektor des Ausgangsbildes verglichen und anhand des Schwellwertes geprüft, ob eine Übereinstimmung bezüglich des Abstandes vorliegt. Diese Entscheidung wird anschließend anhand der Homographie verifiziert: Liegt sowohl seitens der Deskriptoren als auch der bekannten Homographie eine Übereinstimmung vor, so wird die Zahl der echt positiven Übereinstimmungen erhöht, im Falle der fehlenden Übereinstimmung seitens der Homographie die der falsch positiven. Ist der Abstand zweier Deskriptoren nicht kleiner gleich des Schwellwertes, laut Homographie handelt es sich jedoch um die selben Schlüsselpunkte, so liegt ein falsch negatives Beispiel vor. Liegt jedoch laut Homographie auch keine Übereinstimmung vor, so erhöht das die Zahl der echt negativen Übereinstimmungen.

4.2.3 Ergebnisse

In Abbildung 4.2 sind die den Deskriptoren entsprechenden Kurven 1-Precision gegen Recall für jede Bildreihe dargestellt. Wie im letzten Abschnitt beschrieben, wurden diese dabei nach den Bildreihen gruppiert, ohne jedoch einzelne Bilder der selben Reihe zu unterscheiden. Da die Ergebnisse in Abhängigkeit der getesteten Reihe sehr unterschiedlich ausfallen, sind die Kurven nach diesen aufgeschlüsselt. Getestet wurde jeweils jeder der in Tabelle 4.1 gelisteten Arten, den Farbdeskriptor zusammenzusetzen. Zu Vergleichszwecken ist auch die Kurve von FREAK als Urverfahren gezeigt, welches keinerlei Farbinformationen betrachtet, sondern auf Grauwertbildern arbeitet.

Zunächst lässt sich beobachten, dass die Kurven einen teilweise sehr unterschiedlichen


bark
bikes
boat
graf
Recall
1-Precision
0
0.2
0.4
0.6
0.8
1
(a)
(b)
(c)
(d)
FREAK
rgbFREAK
oppFREAK
oppiFREAK
hueFREAK
hsvFREAK

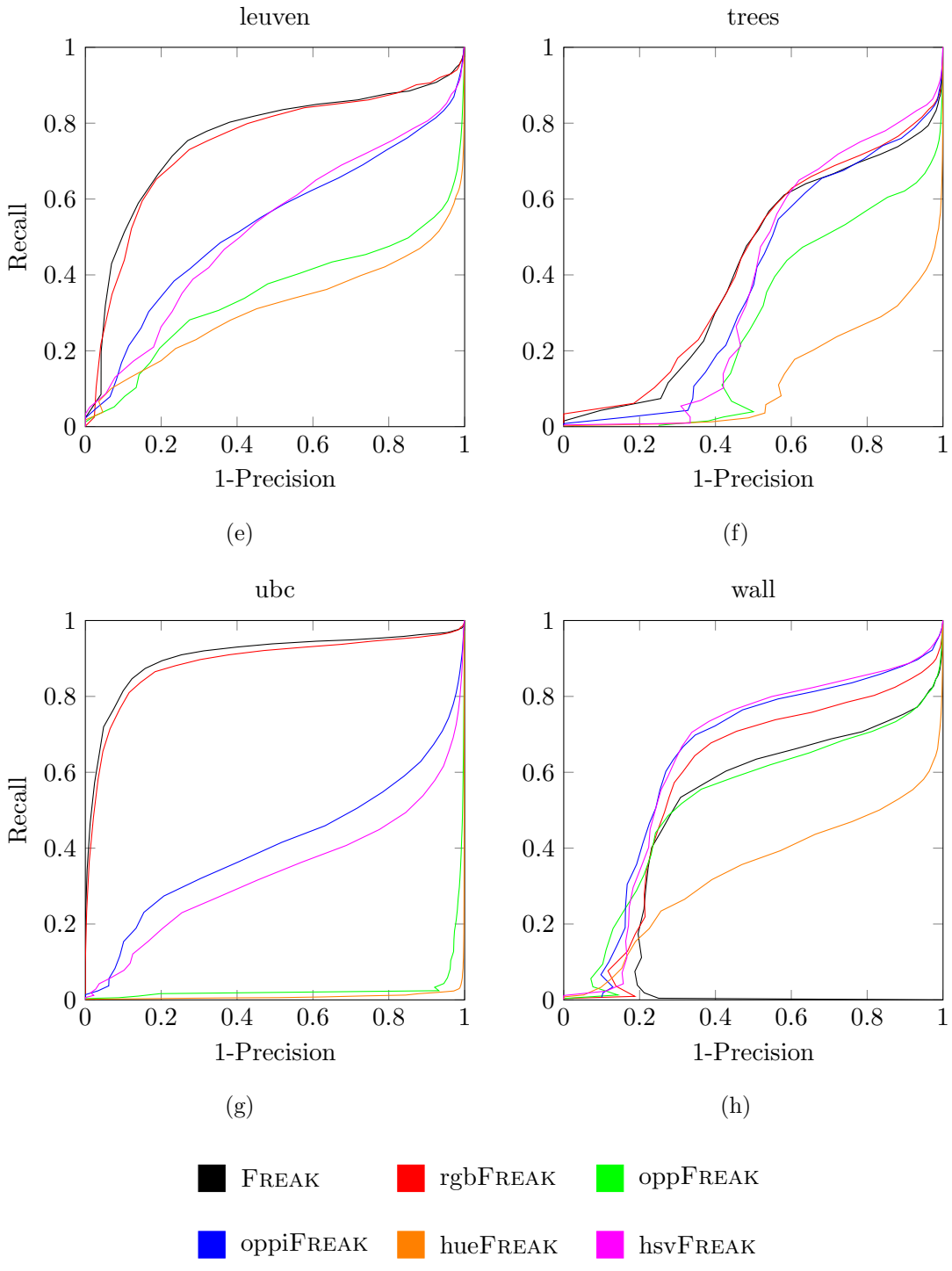
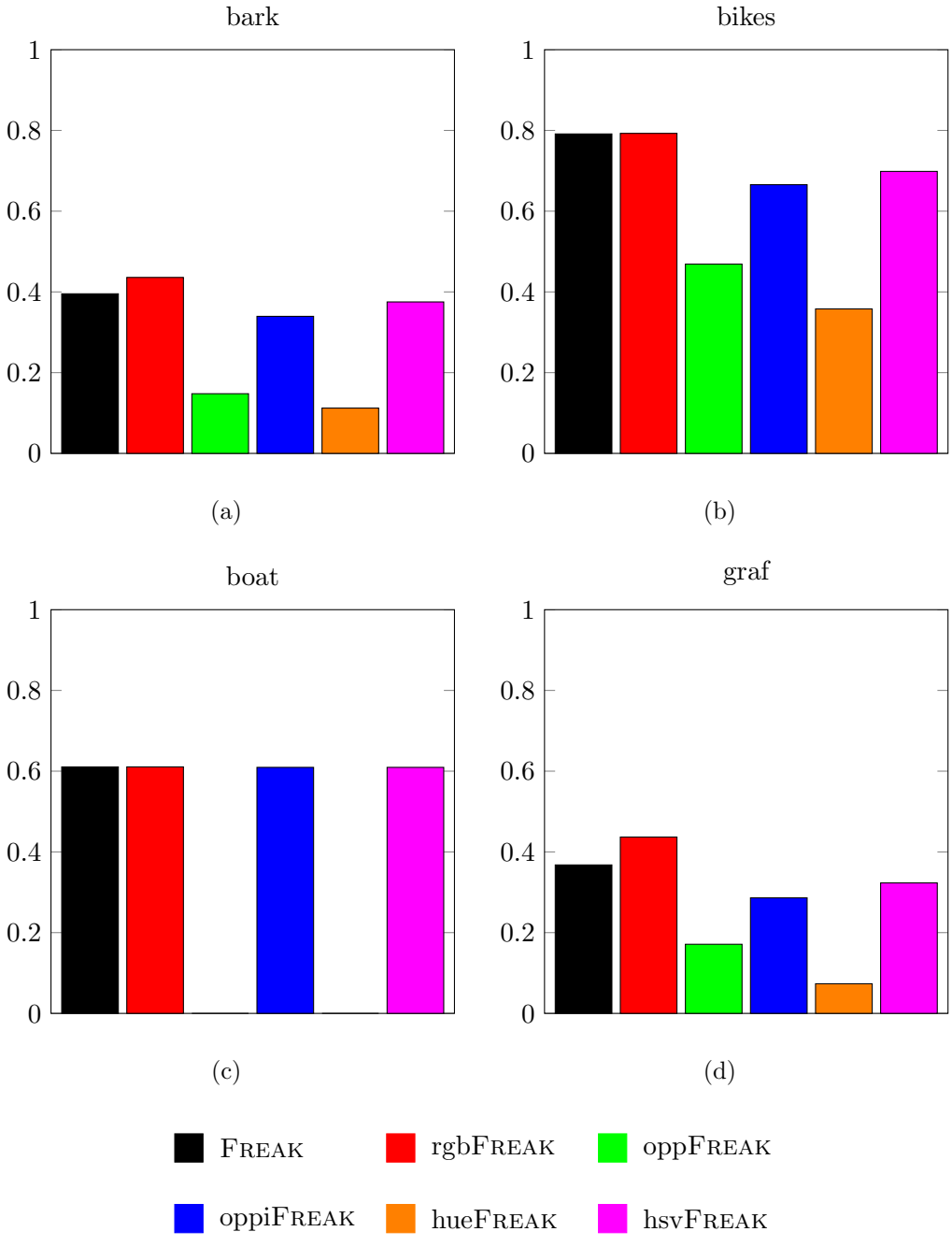


Abbildung 4.2: Gezeigt sind die experimentell ermittelten Kurven zum Vergleich der verschieden zusammengesetzten Deskriptoren.

Verlauf zeigen. Zum Verständnis sollen zunächst auftretende Besonderheiten geklärt werden. Beispielsweise fällt auf, dass die meisten Kurven scheinbar im Punkt $(0, 0)$ beginnen, um dann in $(1, 1)$ zu enden. Für einzelne Kurven trifft das jedoch nicht zu, diese beginnen scheinbar im Punkt $(1, 0)$ und weisen erst später eine geringere Ungenauigkeit auf, beispielsweise zu sehen bei FREAK auf dem Datensatz `bark` in Abbildung 4.2(a). Dieser Verlauf kommt zustande, wenn die erste positive Entscheidung des Deskriptors falsch ist, ohne dass bereits richtige Entscheidungen getroffen wurden. Erst wenn richtige Entscheidungen hinzukommen, wird der Wert 1-Precision, also die Ungenauigkeit, allmählich kleiner. Ist die erste Entscheidung des Deskriptors hingegen richtig, so beginnt die Kurve bei einem 1-Precision Wert von null und einem sehr kleinen Recall, beispielsweise zu sehen für rgbFREAK auf dem Datensatz `graf` in Abbildung 4.2(d).
Vereinzelt kommt es vor, dass die Kurve auf keiner der beiden Achsen beginnt, beispielsweise für FREAK in selbiger Abbildung. In diesem Fall liegen im ersten Schritt positiver Entscheidungen sowohl richtige als auch falsche Entscheidungen gleichzeitig vor. Exakt im Punkt $(0, 0)$ beginnen die Kurven nie. Der erste Punkt wird erst eingetragen, wenn mindestens eine positive Entscheidung vorliegt, welche richtig oder auch falsch sein kann. Der Quotient für die Berechnung des Wertes 1-Precision ist anderenfalls 0 und somit nicht definiert (vgl. dazu die Definitionen der Werte in 2.7). Wie angesprochen wird jeder der Schlüsselpunkte des Eingabebildes mit jedem des Referenzbildes verglichen. Sei n_e die Zahl der Schlüsselpunkte auf dem Eingabebild und n_r die Zahl derer auf dem Referenzbild. Die Zahl aller Vergleiche ist somit $n_e \cdot n_r$, wovon maximal das Minimum aus n_r und n_e korrespondierende Paarungen existieren können. Das Verhältnis von echt Positiven zu falsch Positiven ist somit bei maximalem Schwellwert θ unausgewogen zugunsten der falsch Positiven, woraus sich ein 1-Precision Wert sehr nahe an eins ergibt. In Abschnitt 2.7 wurde an einem fiktiven Beispiel erläutert, dass ein Deskriptor im Vergleich zu einem weiteren dann als besser zu bewerten ist, wenn dessen Kurve über der des anderen liegt. Um diese Tatsache numerisch fassen zu können, wird der Flächeninhalt unter den Kurven ermittelt und dient als Evaluationskriterium. Der Flächeninhalt wird exakt ermittelt für den Bereich vom ersten vorliegenden 1-Precision Wert bis zum letzten. Abbildung 4.3 stellt die Flächeninhalte in Form von Säulendiagrammen dar, auch hier aufgeschlüsselt nach der jeweiligen Bildreihe. Die exakten Werte finden sich im Anhang A.3 in der Tabelle 4.3.


bark
bikes
boat
graf
1
0.8
0.6
0.4
0.2
0
(a)
(b)
(c)
(d)
FREAK
rgbFREAK
oppFREAK
oppiFREAK
hueFREAK
hsvFREAK

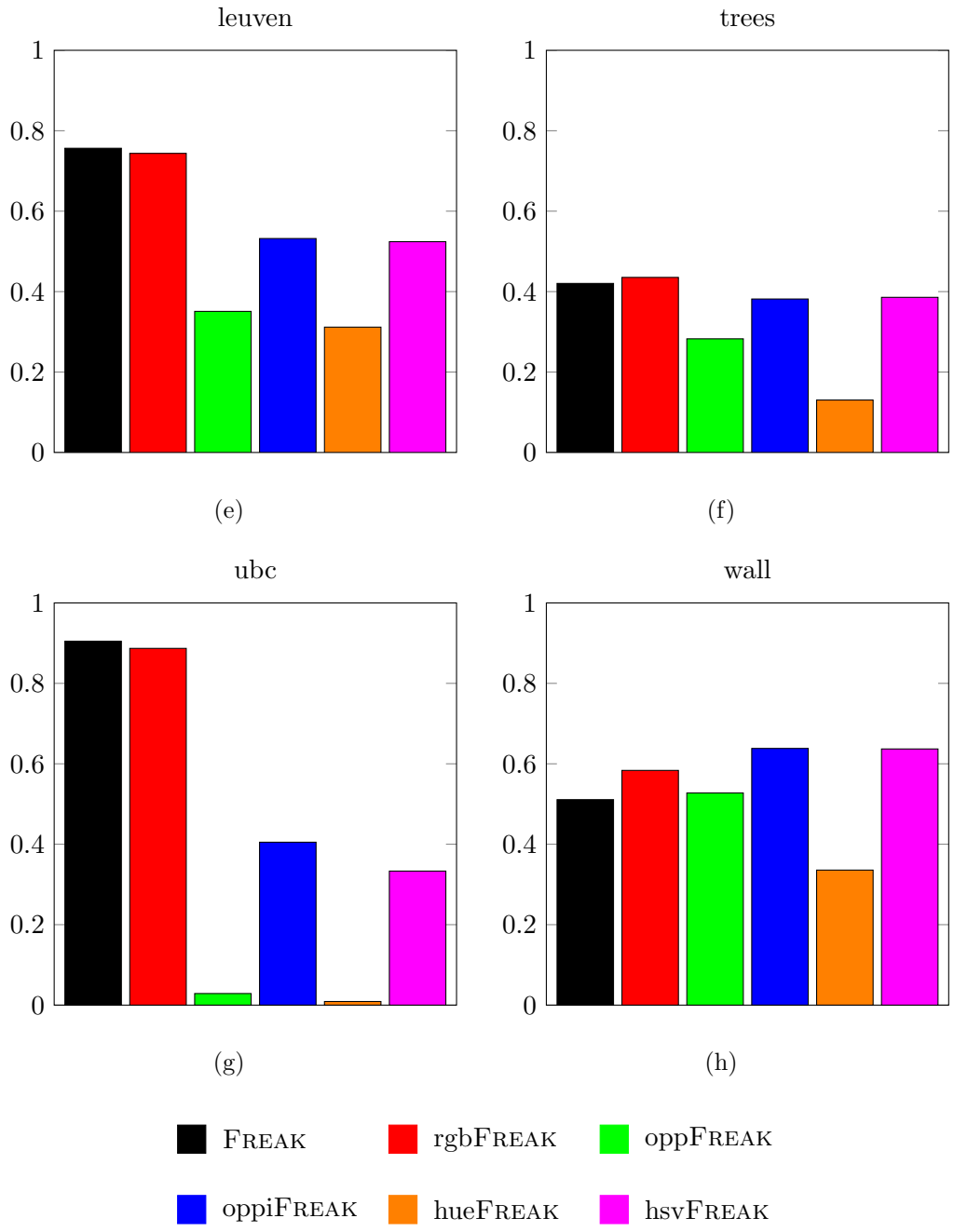


Abbildung 4.3: Darstellung der Flächeninhalte unter den Kurven 1-Precision gegen Recall aus Abbildung 4.2. Je höher die Säule, desto besser ist der bezeichnete Deskriptor.

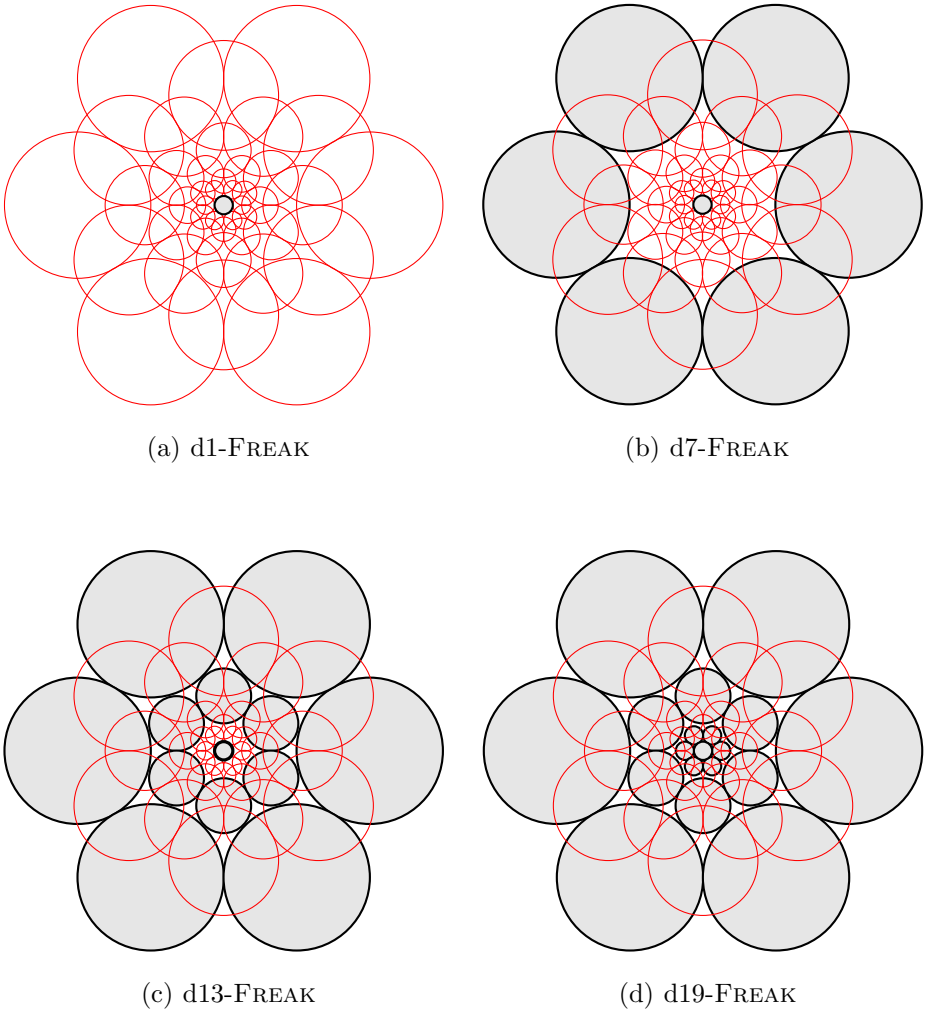
Es fällt auf, dass rgbFREAK meist einen höheren Flächeninhalt unter der Kurve erreicht als FREAK, welcher dahingehend wiederum besser abschneidet als hsvFREAK, oppiFREAK, oppFREAK und schließlich hueFREAK, siehe Datensätze `bark`, `bikes`, `graf` und `trees`. Auf den Datensätzen `leuven` und `ubc` ist rgbFREAK hingegen nicht besser als FREAK, welcher in diesen Fällen durch keinen der um Farbe erweiterten Deskriptoren übertroffen wird. Auf dem Datensatz `boat` sind die Deskriptoren FREAK und rgbFREAK gleichauf, nahezu gleich gefolgt von oppiFREAK und hsvFREAK. In der Darstellung der Kurven ist nicht zu erkennen, dass sich für hueFREAK und oppFREAK für `boat` keine Kurven erzeugen lassen, da für alle Schwellwerte θ alle Vergleiche positiv ausfallen und sie somit immer einen Recall von eins bei einer Ungenauigkeit sehr nahe an eins besitzen. Die Definition eines Farbwinkels oder einer Gegenfarbe ergibt auf Grauwertbildern keinen Sinn, was diesen Umstand erklärt. Auf dem Datensatz `wall` ist rgbFREAK erneut besser als FREAK, wird jedoch noch übertroffen von oppiFREAK und hsvFREAK.

4.3 Farbe als zusätzliches Attribut

In diesem Abschnitt wird eine alternative Möglichkeit untersucht, den FREAK-Deskriptor um Farbinformation zu erweitern. Anders als im vorangegangenem Experiment geschieht dies jedoch nicht integriert durch die Konkatenation der Deskriptoren der Kanäle verschiedener Farbräume. Vielmehr wird Farbe hier als zusätzliches Attribut an den FREAK-Deskriptor des Schlüsselpunktes auf dem entsprechenden Grauwertbild geknüpft.

4.3.1 Extraktion und Vergleich von Farbe

Im Grundlagenkapitel über Farben wurde der L*a*b* Farbraum als Möglichkeit eingeführt, vom Menschen empfundene Farbunterschiede anhand des euklidischen Abstandes zweier gegebener Farben messbar zu machen (siehe Abschnitt 3.2). In diesem Experiment wird dies aufgegriffen und es werden Möglichkeiten getestet, die Umgebung eines Schlüsselpunktes herkömmlich durch den FREAK-Deskriptor zu beschreiben, wie dies im Grundlagenkapitel 2.5.3 eingeführt wurde. Zusätzlich soll ein Farbvektor berechnet werden, welcher ebenfalls den Schlüsselpunkt beschreibt und so ermöglicht, die Ähnlichkeit von Schlüsselpunkten anhand des Abstandes ihrer Farbvektoren im L*a*b*-Raum zu bestimmen. Insgesamt werden vier verschiedene Möglichkeiten getestet, die Farbe in der Schlüsselpunktumgebung zu erheben, eine Darstellung findet sich in Abbildung 4.4. Das Muster d1 extrahiert die Farbe im Zentrum des Schlüsselpunktes, der Farbvektor ist somit eindimensional. Das dem FREAK-Deskriptor inhärente Prinzip der paarweisen Vergleiche von Abtastpunkten wird nicht beibehalten, vielmehr wird hier die reine Farbe eines Punktes im L*a*b*-Raum extrahiert, um diese später zusätzlich zu dem ursprünglichen FREAK-Deskriptor eines Schlüsselpunktes vergleichen zu können. Für die Konstruktion der Muster d7, d13 und d19 werden die aus Abbildung 2.13 bekannten Schichten der Perifovea, der Parafovea und der Fovea centralis in dieser Reihenfolge hinzugenommen, das Muster d19 extrahiert somit die meiste Farbe und erzeugt einen Farbvektor mit 19 Dimensionen. Dieser Entwurf folgt dem bei der Einführung des FREAK-Deskriptors beschriebenem "grob nach fein"-Prinzip (siehe Abschnitt 2.5.3), wonach die Aussagekraft der äußeren Schichten des FREAK-Musters am stärksten ist. Eine Auflistung der beschriebenen Farbdeskriptoren findet sich in Tabelle 4.2.
Der Abstand zweier Farben wird anhand des in Abschnitt 3.2 eingeführten ΔE^* Maßes



(a) d1-FREAK (b) d7-FREAK

(c) d13-FREAK (d) d19-FREAK

Abbildung 4.4: Dargestellt ist das FREAK-Muster, welches zur Berechnung des Deskriptorvektors auf einem Grauwertbild genutzt wird. Zusätzlich wird in diesem Experiment die Farbe im L*a*b*-Farbraum extrahiert, hier verdeutlicht durch die grau hinterlegten Bereiche. Der zentrale Abtastpunkt fließt in jeden der vier Farbdeskriptoren ein, bei den Mustern (b), (c) und (d) kommen zusätzliche abgetastete Bereiche von außen nach innen hinzu. Die Auswahl der Ringstruktur erfolgt nach der aus Abbildung 2.13 bekannten Gliederung des FREAK-Musters in Perifovea, Parafovea und der Fovea centralis.

Name	Struktur	Δ_{max}
d1	Zentrum	367.42
d7	Zentrum, Perifovea	972.11
d13	Zentrum, Perifovea, Parafovea	1324.76
d19	Zentrum, Perifovea, Parafovea, Fovea centralis	1601.56

Tabelle 4.2: Auflistung der getesteten Farbvektoren, welche den FREAK-Deskriptor in Form eines zusätzlichen Attributes erweitern. Mit der Anzahl der extrahierten Farben ändert sich neben der Dimension des Farbvektors auch der maximal mögliche Abstand Δ_{max} zweier Farbvektoren zueinander.

bestimmt, welches sich als der euklidische Abstand ihrer L*a*b*-Koordinaten berechnen lässt. Enthält ein Farbvektor mehr als eine Dimension, wird zunächst der paarweise Abstand aller korrespondierenden Abtastpunkte auf diese Art berechnet, was einen Abstandsvektor $\mathbf{d}$ ergibt. Der Abstand zweier Farbvektoren $\mathbf{A}$ und $\mathbf{B}$ ergibt sich als der Betrag dieses Abstandvektors:

$$\Delta(\mathbf{A}, \mathbf{B}) = \|\mathbf{d}\|_2 = \sqrt{d_1^2 + \ldots + d_n^2}$$

wobei

$$\mathbf{d} = \begin{bmatrix} \Delta E^*(a_1, b_1) \\ \Delta E^*(a_2, b_2) \\ \vdots \\ \Delta E^*(a_n, b_n) \end{bmatrix}$$

Es können somit nur Farbvektoren gleicher Länge verglichen werden, was in dem beschriebenen Anwendungskontext sinnvoll erscheint. Dabei fällt auf, dass der maximale Abstand zweier Farbvektoren mit deren Länge wächst. Der maximale Abstand der Farbvektoren, wie dieser jeweils in Tabelle 4.2 gelistet ist, berechnet sich für eine Dimension n als

$$\Delta_{max}(n) \approx \sqrt{n \cdot 367.42^2} = 367.42 \cdot \sqrt{n}$$

Die Konstante ergibt sich aus dem maximalen Abstand zweier Farben im L*a*b*-Farbraum:

$$\sqrt{(100-0)^2+(-150-100)^2+(-100-150)^2} \approx 367.42$$

Die Kenntnis des maximalen Farbabstandes ist für die Berechnung der Recall gegen 1-Precision Kurve nötig, da dieser den maximalen Wert des Schwellwertes θ definiert.

4.3.2 Durchführung der Experimente

Im vorangegangenen Experiment, in welchem Farbe direkt in den FREAK-Deskriptor integriert wurde, konnte der Hamming-Abstand für den Vergleich von Deskriptoren genutzt werden. Das ist in diesem Experiment nicht möglich, da sich hier das Format des FREAK-Deskriptors und das des Farbvektors voneinander unterscheiden. Auf der einen Seite existiert somit mit dem FREAK-Deskriptor ein Klassifikator, welcher mit Hilfe eines Schwellwertes eine Entscheidung anhand des Hamming-Abstandes der Deskriptorvektoren trifft. Auf der anderen Seite existiert ein Klassifikator, welcher anhand des Schwellwertes und des euklidischen Abstandes der Farbvektoren eine Entscheidung fällt. Die Entscheidungen beider Klassifikatoren müssen somit miteinander verknüpft werden, um daraus wiederum eine finale Entscheidung zu generieren.
Im Experiment wurde als Ansatz gewählt, die Klassifikatoren multiplikativ zu verknüpfen, also zu "verunden". Nur wenn beide Klassifikatoren eine positive Entscheidung treffen, ist auch das Endergebnis positiv, in jedem anderen negativ. Erreicht werden soll so, dass Schlüsselpunkte, welche durch den FREAK-Deskriptor nicht unterschieden werden können, da eine Betrachtung der Farbe nötig ist, auch durch den zweiten Klassifikator bewertet werden müssen.
In der Vorbereitung des Experimentes viel auf, dass sich die Kurven des zusammengesetzten Klassifikators mit dem maximal gesetzten Schwellwert des Farbvektors nicht von der Kurve des FREAK-Deskriptors alleine unterschieden. Als Grund dafür wurde festgestellt, dass der Bereich des maximalen Farbabstandes nur theoretischer Natur ist und in der Anwendung nicht auftritt. Der tatsächlich auftretende maximale Farbabstand liegt bei jeder Bildreihe weit unterhalb des theoretisch möglichen Abstandes. Diese Tatsache wirkt sich auf den finalen Klassifikator aus, da dieser in der Folge toleranter und damit

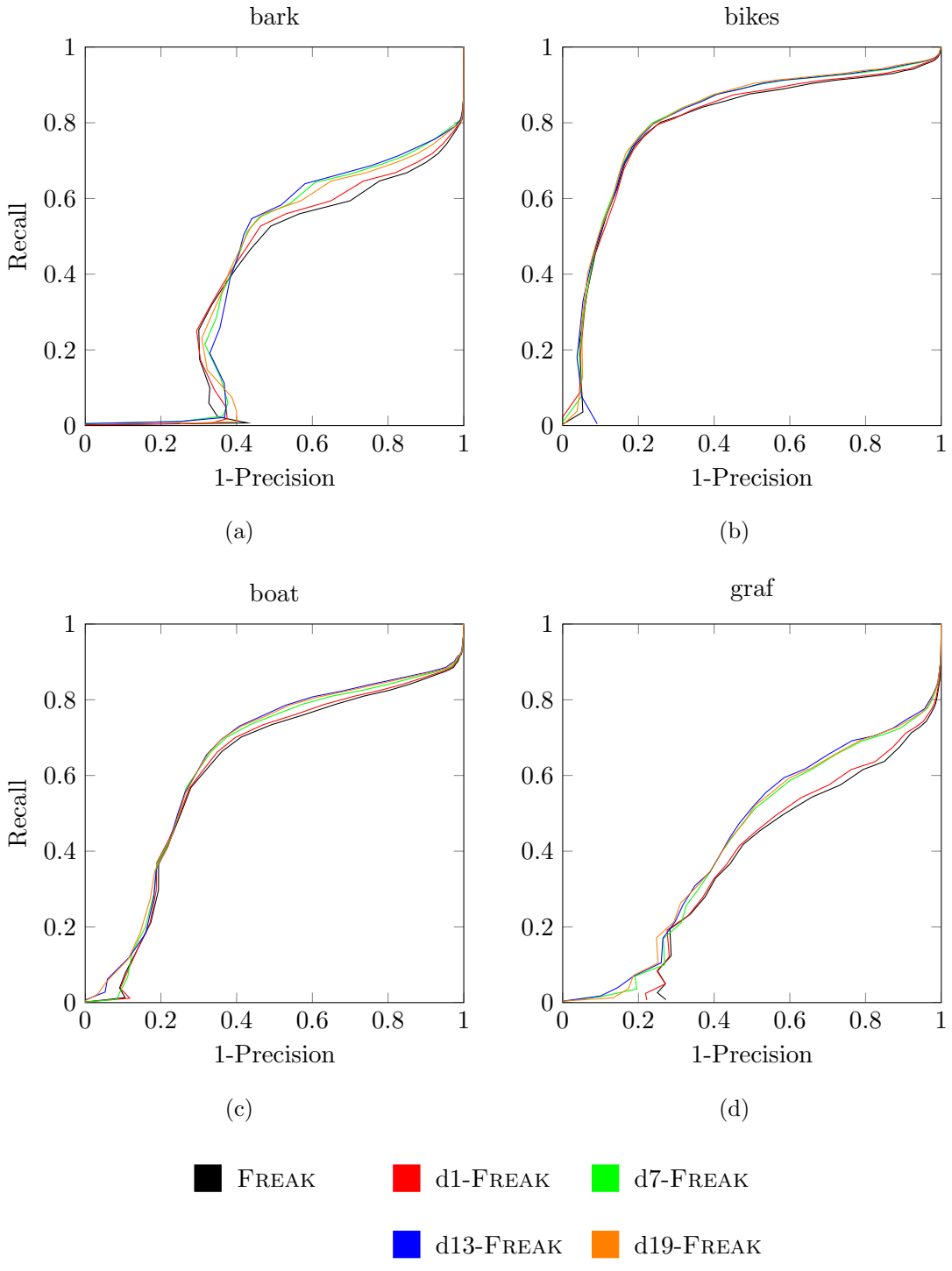
stets positiv entscheidet. Die Entscheidung bleibt damit einzig dem auf dem FREAK-Deskriptor basierenden Klassifikator überlassen, was das beschriebene Verhalten nicht unterscheidbarer Kurven erzeugt.
Um diesem Problem zu begegnen, wurde für jedes Motiv des Datensatzes der maximal auftretende Farbabstand in Abhängigkeit der Länge des Farbvektors bestimmt, und dieser als maximaler Wert des Schwellwertes gesetzt. Tabelle 4.3 listet die experimentell ermittelten, maximalen Farbabstände. Zu beobachten ist, dass diese mit der Länge der Farbvektoren zunehmen, was dem erwarteten Verhalten entspricht.

	d1	d7	d13	d19
`bark`	60	80	100	160
`bikes`	80	120	160	180
`boat`	80	120	140	180
`graf`	80	120	160	200
`leuven`	90	180	250	300
`trees`	60	120	140	180
`ubc`	90	160	200	250
`wall`	60	90	110	140

Tabelle 4.3: Gelistet sind die experimentell ermittelten maximalen Farbanstände in Abhängigkeit der Bildreihe und der Länge des Farbvektors.

4.3.3 Ergebnisse

Die Ergebnisse werden wie im vorangegangenem Experiment dargestellt. Die Kurven 1-Precision gegen Recall finden sich in Abbildung 4.5, ein direkter Vergleich der Kurven wird, wie bekannt, über den Flächeninhalt unter der Kurve vorgenommen. Die Säulendiagramme der Flächeninhalte finden sich in Abbildung 4.6. Erneut wird für den direkten Vergleich für jede Bildreihe auch die Kurve des unveränderten FREAK-Deskriptors gezeigt. Da sich die Kurven stark ähneln und meist schwer zu unterscheiden sind, listet die Tabelle A.2 im Anhang A.3 die Flächeninhalte unter den Kurven in gerundeter Form.
Zunächst fällt auf, dass die Ergebnisse weniger unterschiedlich ausfallen als im vorangegangenem Experiment der integrierten Farbe. Eine Beurteilung, inwiefern dabei


bark
bikes
boat
graf
Recall
1-Precision
0
0.2
0.4
0.6
0.8
1
(a)
(b)
(c)
(d)
Freak
d1-Freak
d7-Freak
d13-Freak
d19-Freak

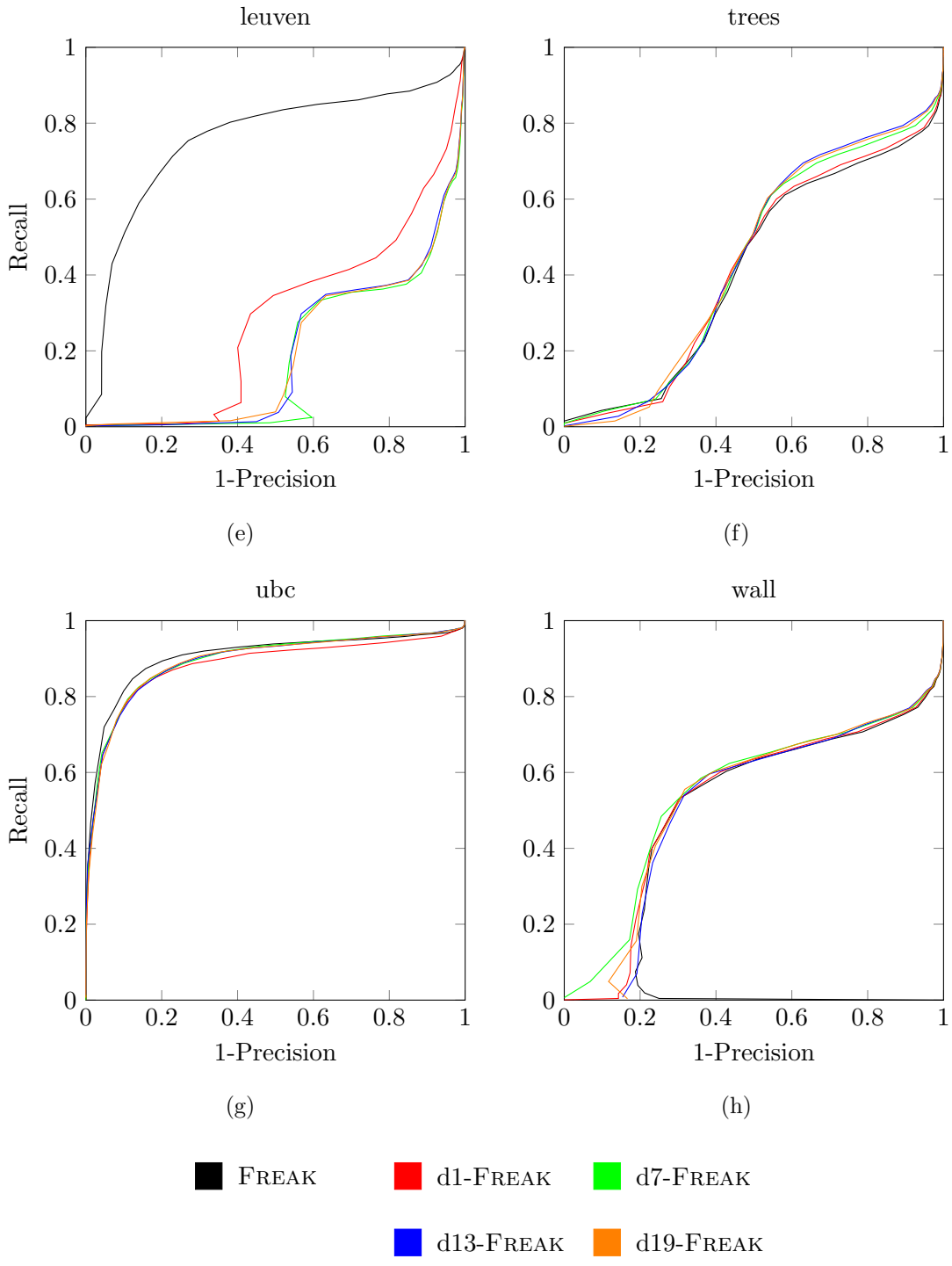
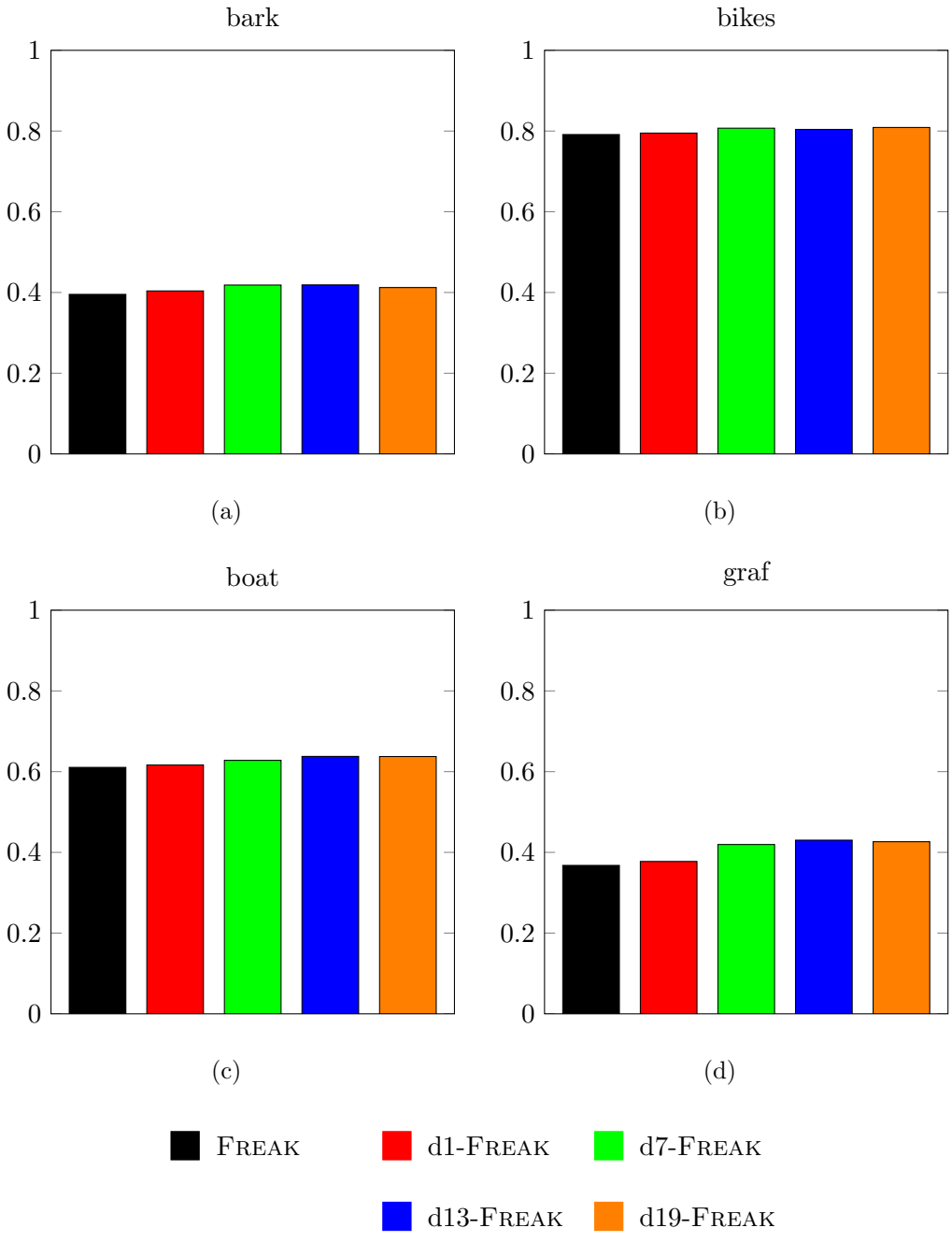


Abbildung 4.5


bark
bikes
boat
graf
1
0.8
0.6
0.4
0.2
0
(a)
(b)
(c)
(d)
Freak
d1-Freak
d7-Freak
d13-Freak
d19-Freak

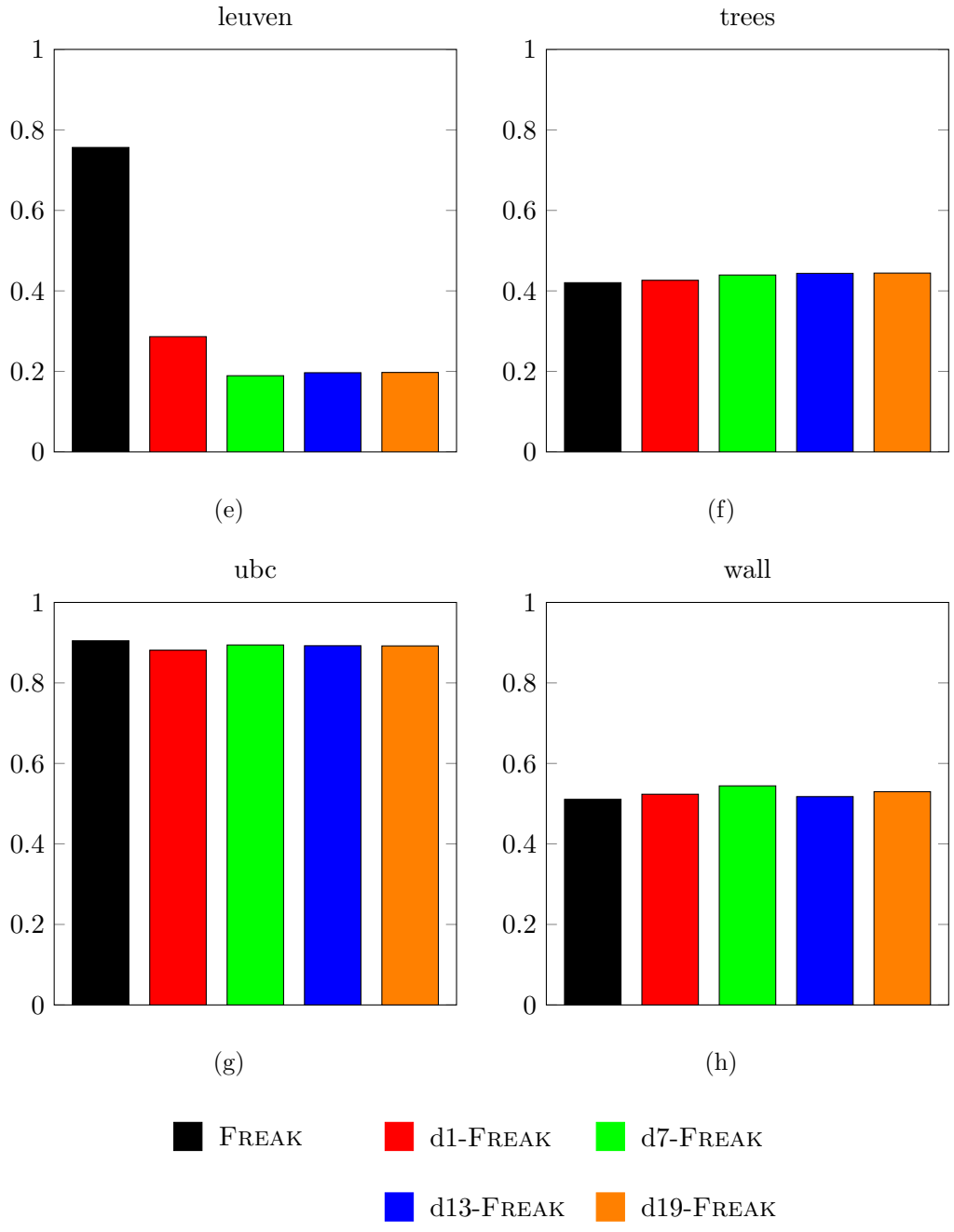


Abbildung 4.6: Darstellung der Flächeninhalte unter den Kurven aus Abbildung 4.5.

überhaupt von signifikanten Unterschieden gesprochen werden kann, lässt die kleine Zahl verfügbarer Werte nicht zu. Dennoch sollen die Flächeninhalte hier zumindest miteinander verglichen werden.

Jeder zusammengesetzte Klassifikator, welcher neben dem FREAK-Deskriptor auch Farbe berücksichtigt, weist auf sechs von acht Bildreihen ein besseres Ergebnis auf als der FREAK-Deskriptor, gemessen am Flächeninhalt. Allein auf den Bildreihen `leuven` und `ubc` bleibt der FREAK-Deskriptor in seiner ursprünglichen Form unübertroffen. Das deckt sich mit den Ergebnissen des vorangegangenem Experimentes, bei welchem für die Bildreihen `ubc` und `leuven` ebenfalls keine Verbesserung erzielt werden konnte. Während sich die Kurven in ihrem Verlauf oft ähneln (siehe etwa `bark`, `bikes` oder `graf`), zeigt die Kurve für das Motiv `leuven` sogar einen deutlich anderen Verlauf. Das geht mit einem erheblich kleineren Flächeninhalt der um Farbe erweiterten Deskriptoren unter den Kurven einher.

In den meisten Fällen schneidet d13-FREAK am besten ab, welcher auf den Motiven `bark`, `boat` und `graf` den höchsten Flächeninhalt erzielt. Der Deskriptor mit den meisten Farbinformationen, d19-FREAK, schneidet auf den Bildreihen `bikes` und `trees` am besten ab, während d7-FREAK nur auf dem Datensatz `wall` das beste Ergebnis erzielt. Wird nur die Farbe des Zentrums zur Bildung des Deskriptors berücksichtigt, wie geschehen für d1-FREAK, übertrifft dieser Deskriptor in keinem der getesteten Fälle alle anderen Deskriptoren.

5 Praktische Anwendung

Die Experimente des vorangegangenem Kapitels sind von theoretischer Natur und verwenden den zur Evaluation von Deskriptoren weit verbreiteten Datensatz von Mikolajczyk und Schmid (2005). Die Erkenntnisse aus diesen Experimenten nutzend soll abschließend auch die praktische Anwendung der Beschreibung von Bildern anhand lokaler Merkmale demonstriert werden. Darüber hinaus soll das praktische Experiment untersuchen, inwiefern die theoretisch gewonnenen Einsichten in einer konkreten Anwendung von Relevanz sind. Beispielsweise soll die Bedeutung der Beleuchtung und die der Farbartefakte beobachtet werden. Das Ziel der konkreten Anwendung soll sein, bekannte Objekte, welche als Referenzbilder vorliegen, in einem Kamerabild zu identifizieren.

5.1 Auswahl der Objekte

Die Objekte wurden für das Experiment bewusst so gewählt, dass sich diese in ihrem Motiv sehr ähneln, in der Farbe jedoch stark unterschieden, um einen eventuellen Vorteil eines Farbe berücksichtigenden Verfahrens gegenüber FREAK in unveränderter Form sichtbar zu machen. Die Wahl fiel auf Schokoladentafeln der selben Marke, welche jedoch einen untereinander verschiedenen Gehalt von Kakao aufweisen, was jeweils verdeutlicht ist durch die Farbe der Verpackung. Dargestellt sind diese in Abbildung 5.1. Im weiteren Verlauf der Arbeit werden die Tafeln nach den dominierenden Farben Blau, Gelb, Grün, Lila und Rot benannt. Alle Bilder haben eine Auflösung von 200×477 Pixeln und liegen im RGB-Format vor. Aufgenommen wurden die Bilder unter Verwendung eines Scanners, wodurch eine sehr gleichmäßige Beleuchtung erreicht werden konnte.
Die Struktur der Referenzbilder ist sehr ähnlich. In der oberen Hälfte findet sich ein Strichcode, darunter in weiß der Name der Sorte, welche wiederum gefolgt wird von einem Text in weißer Schrift, welcher sich jedoch in Inhalt und Länge unterscheidet. Mittig findet sich eine Darstellung zweier Stück Schokolade in einem Braunton, welcher den Gehalt an Kakao widerspiegelt. In der unteren Hälfte ist die Ernährungstabelle in der jeweilig dominierenden Farbe abgebildet, darüber findet sich erneut ein Text, welcher Aufschluss über die Zutaten gibt. Der Inhalt der Texte soll nicht von Bedeutung sein, einzig zu beachten ist die Tatsache, dass diese sich zum Teil in der Länge unterscheiden. Auffällig bleibt dieser Umstand auch, wenn die Bilder in geringer Auflösung vorliegen.

(a) Blau (b) Gelb (c) Grün (d) Lila (e) Rot

Abbildung 5.1: Zu sehen sind die Referenzbilder für das praktische Experiment. Zur Vereinfachung werden diese nach ihren dominierenden Farben benannt. Die Bilder sind in ihrer Struktur sehr ähnlich, bezüglich der Farben jedoch für den Menschen leicht zu unterscheiden.

5.2 Durchführung des Experimentes

Bei der Durchführung des Experimentes wird sich an Abbildung 2.2 des Grundlagenkapitels 2.3 orientiert. Zunächst werden die Bilder eingelesen und eine Grauwertkopie eines jeden Bildes erstellt. Auf diesen werden Schlüsselpunkte mithilfe des beschriebenen Fast-Hessian Detektors (siehe Abschnitt 2.4.2) gefunden. Im Anschluss daran werden diese einerseits durch den unveränderten FREAK-Deskriptor auf dem Grauwertbild beschrieben, andererseits durch den in Abschnitt 4.2 eingeführten rgbFREAK auf dem Farbbild. Dieser hatte sich im Experiment gegenüber den anderen Deskriptoren als am geeignetsten erwiesen. Die Farbe als externes Attribut zu beschreiben, wie gezeigt in Abschnitt 4.3, wurde verworfen, da hier wieder maximale Farbabstände hätten ermittelt werden müssen, um den Deskriptor überhaupt einsetzen zu können. Das macht diese Art, Farbe zu beschreiben für die Anwendung ungeeignet.

Der Detektor wurde so eingestellt, dass zwischen 240 und 300 Schlüsselpunkte auf jedem Referenzbild gefunden und beschrieben werden. Als Kamera wurde eine Webcam verwendet, welche Bilder der Auflösung 640×480 liefert. Für das Kamerabild wurden ebenfalls Schlüsselpunkte auf dessen Grauwertbild gefunden und dann sowohl durch FREAK auf selbigem Bild als auch durch rgbFREAK auf dem Farbbild beschrieben. Im Experi-

ment wurden für jeden Frame alle beschriebenen Schlüsselpunkte eines Deskriptors mit denen der Referenzbilder abgeglichen und die Zahl der gefundenen Übereinstimmungen festgestellt. Als Strategie des Vergleiches wurde Nearest neighbor matching mit einer experimentell ermittelten NNDR von 0.65 eingesetzt (siehe Abschnitt 2.6). Abbildung 5.2 zeigt eine Aufnahme der Anwendung.

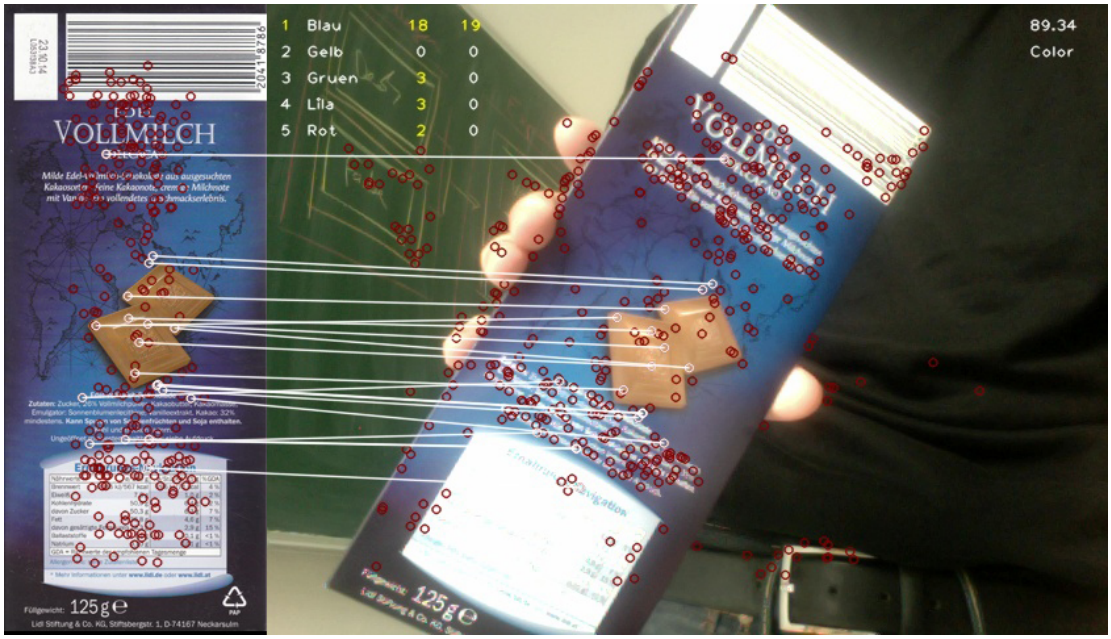


Abbildung 5.2: Darstellung der praktischen Anwendung. Links sichtbar ist das Referenzbild, rechts das Kamerabild. Schlüsselpunkte beider Bilder sind rot eingezeichnet, Übereinstimmungen weiß verbunden. Links oben im Kamerabild findet sich die Zahl der gefundenen Übereinstimmungen für FREAK in der ersten Spalte und rgbFREAK in der zweiten. In der Aufnahme findet FREAK fälschlicherweise auch Übereinstimmungen mit Referenzbildern anderer Farben, rgbFREAK hingegen korrekterweise nur Schlüsselpunkte der blauen Tafel.

Während des Experiments wurde die Zahl der gefundenen Übereinstimmungen von FREAK und rgbFREAK im Zeitverlauf festgehalten. In einem Zeitfenster von etwa 20 Sekunden wird jeweils eine Tafel in das Bild gehalten, beginnend mit Blau, dann Gelb, Grün, Lila und Rot. Mit diesem Versuchsaufbau gehen die Transformationen Skalierung, Rotation und perspektivische Verzerrung des Referenzbildes bezogen auf das Kamerabild einher. Darüber hinaus tritt auch eine Änderung der Beleuchtung auf. Dies wird einerseits durch die verschiedenen Winkel des präsentierten Objektes zur Kamera

bedingt, andererseits durch den glänzenden Druck, welcher manche Farben heller erscheinen lässt oder gänzlich weiß überblendet. Auch können Farbartefakte auftreten, was der geringen Farbauflösung der verwendeten Kamera geschuldet ist. So werden mitunter weiße Stellen durch andere Farben ersetzt, sehr gut zu beobachten beispielsweise in der Gewichtsangabe “125g” in Abbildung 5.2 unten.

5.3 Ergebnisse

Abbildung 5.3 stellt die Zahl der gefundenen Übereinstimmungen in Abhängigkeit der Zeit grafisch dar. Der farbige Balken am oberen Ende jeder Grafik zeigt an, dass sich das in der Überschrift benannte Objekt zu diesem Zeitpunkt tatsächlich im Bild befindet. Die schwarze Kurve zeigt die Ergebnisse für FREAK, die farbige jeweils für rgbFREAK. Eine optimale Kurve würde immer 0 Übereinstimmungen zeigen, wenn sich das beschriebene Objekt nicht im Bild befindet. Sobald das Objekt jedoch im Bild zu sehen ist, sollten alle Schlüsselpunkte gefunden und korrekt zugeordnet werden. Dieser optimale Verlauf ist nicht zu beobachten. Einerseits kann nicht sichergestellt werden, dass der Detektor alle Schlüsselpunkte korrekt wiederholt, andererseits besteht keine Garantie darüber, dass alle Schlüsselpunktumgebungen den exakt gleichen Deskriptorvektor wie im Referenzbild bewirken, zu stark sind zum Teil die genannten Transformationen. Während für jedes Referenzbild mindestens 240 mögliche Schlüsselpunkte existieren, liegt der maximale Ausschlag aller Kurven nie höher als 60, es wird also selbst im besten Fall nur ein Bruchteil der möglichen Punkte wiedergefunden.
Zunächst ist zu beobachten, dass der FREAK-Deskriptor im Schnitt die meisten Übereinstimmungen findet. Anhand der Ausschläge der schwarzen Kurven vom Beginn des Versuches bis Sekunde 100 lässt sich ablesen, dass eine Tafel im Bild ist. Ab Sekunde 100 finden sowohl FREAK als auch rgbFREAK bis zum Ende des Versuches nur noch vereinzelt Übereinstimmungen, hier befindet sich jedoch keine Tafel im Bild. Am höchsten sind die Ausschläge beider Kurven jeweils an den erwarteten Stellen, d. h. für Blau in den Sekunden 0-20, für Gelb von 20-40, für Grün von 40-60, für Lila von 60-80 und schließlich für Rot von 80-100.
Beide Verfahren finden im Experiment jedoch auch falsche Übereinstimmungen, zu Erkennen an Ausschlägen der Kurven in nicht durch einen Balken gekennzeichneten Bereichen, für Blau beispielsweise ab Sekunde 20 bis zum Ende des Versuches. Von Interesse

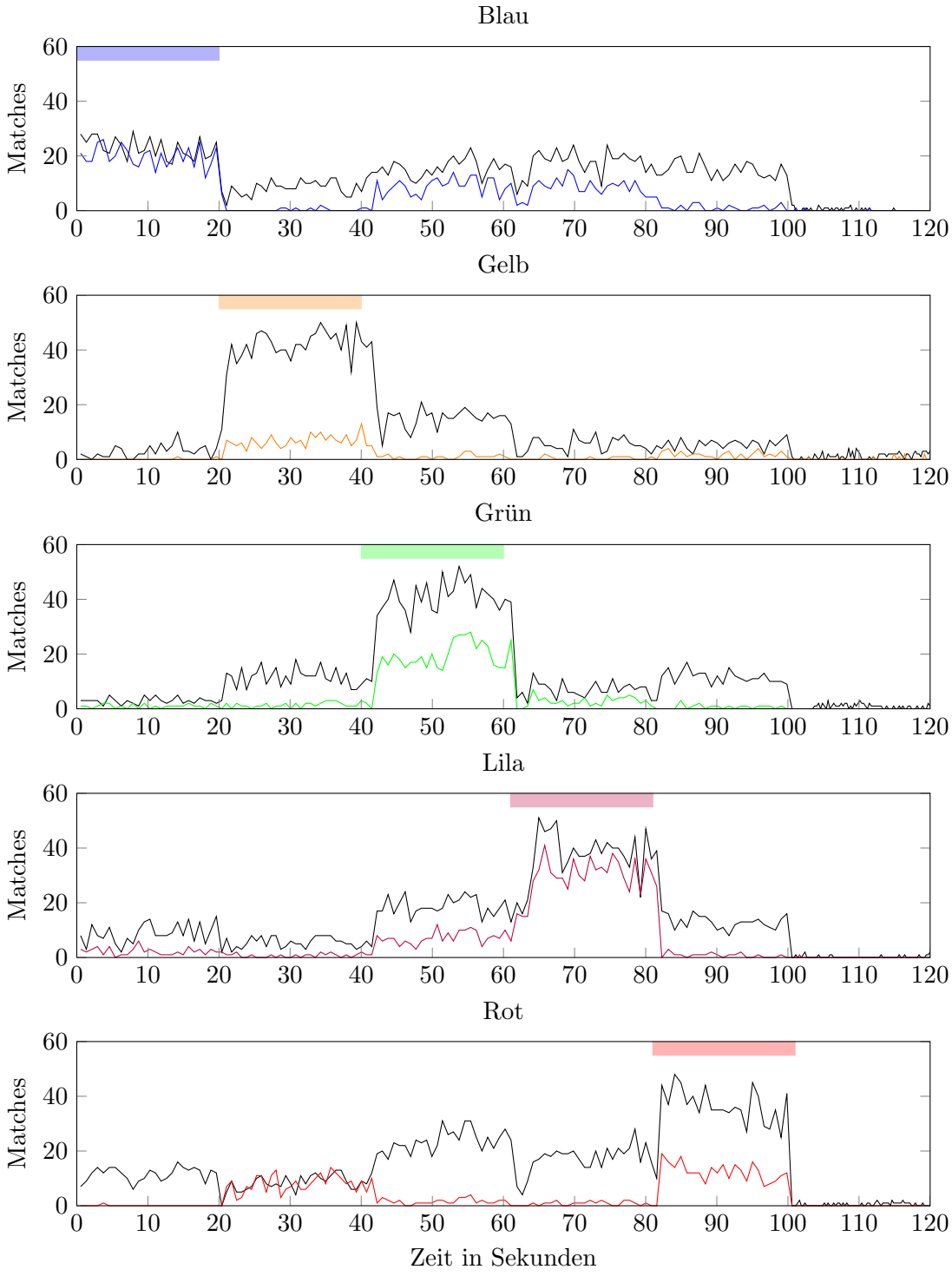


Abbildung 5.3: Darstellung des zeitlichen Verlaufes während des Experiments. Die Kurven zeigen die Zahl der gefundenen Übereinstimmungen, schwarz für FREAK, für rgbFREAK in Farbe. In hinterlegten Bereichen befindet sich das benannte Objekt (vgl. dazu Abb. 5.1) im Bild.

ist, ob rgbFreak im Experiment diskriminativer ist als Freak in seiner ursprünglichen Form. Dies ist messbar, indem der relative Abstand des höchsten Ausschlages zum zweithöchsten in Beziehung gesetzt wird. Tabelle 5.1 zeigt die durchschnittlichen Ausschläge für Zeiträume, in denen sich ein Objekt im Bild befindet. Die Maxima m_1 einer jeden Zeile befinden sich alle auf der Diagonale, deutlich gemacht durch farbige Hinterlegung, was zeigt, dass sich anhand des Verfahrens alle Objekte korrekt zuordnen ließen. Durch einen Rahmen hervorgehoben sind die zweithöchsten Ausschläge m_2 einer Zeile. Der Quotient m_2/m_1 lässt nun ein Urteil darüber zu, wie diskriminativ das Verfahren im Vergleich ist. Der Wert bewegt sich zwischen 0 und 1, wobei ein hoher Quotient auf geringe Diskriminativität deutet, ein geringer Quotient hingegen auf eine hohe Diskriminativität. Im optimalen Fall ist der Quotient 0, was genau dann der Fall ist, wenn Schlüsselpunkte einzig im korrekten Fall gefunden werden. In Tabelle 5.1 ist neben den Durchschnittswerten auch diese Maß gelistet.

		5–15s	25–35s	45–55s	65–75s	85–95s	105–115s	m_2/m_1
Blau	Freak	22.7	9.4	15.2	18.5	14.9	0.4	0.81
	rgbFreak	19.5	0.4	9.1	9.5	0.9	0.0	0.48
Gelb	Freak	3.6	43.0	15.1	5.3	5.3	1.2	0.35
	rgbFreak	0.1	6.6	0.7	0.3	1.3	0.1	0.20
Grün	Freak	2.9	12.8	41.2	6.7	12.0	1.1	0.31
	rgbFreak	0.8	1.2	19.2	2.4	0.7	0.0	0.13
Lila	Freak	8.5	5.9	19.3	40.3	12.5	0.1	0.48
	rgbFreak	1.8	0.5	7.0	31.8	0.9	0.0	0.22
Rot	Freak	11.3	9.0	24.0	17.9	35.6	0.6	0.67
	rgbFreak	0.0	8.4	1.4	1.1	12.0	0.0	0.70

Tabelle 5.1: Die Tabelle listet die durchschnittliche Zahl wiedergefundener Schlüsselpunkte für verschiedene Zeitabschnitte. Farblich hervorgehoben ist jeweils der Bereich, in welchem sich das entsprechende Objekt tatsächlich im Bild befindet, hier zeigt dies gleichzeitig den maximalen Ausschlag m_1 einer Zeile. Der zweithöchste Ausschlag m_2 ist mit einem Rahmen versehen. Je geringer der Quotient m_2/m_1, desto höher ist die Diskriminativität des Verfahrens.

Zwar ist die Gesamtzahl wiedergefundener Schlüsselpunkte für rgbFREAK stets geringer als für FREAK, positiv lässt sich jedoch feststellen, dass die Diskriminativität, so wie sie eingeführt wurde, in 4 von 5 Fällen höher ist, als für FREAK in seiner ursprünglichen Form. Im letzten Fall der roten Tafel zeigt rgbFREAK fälschlicherweise sehr viele Ausschläge für die gelbe Tafel. Eine Erklärung hierfür könnte der hohe Rotanteil des Bildes der gelben Tafel (siehe Abbildung 5.1(b)) sein. Die Diskriminativität von rgbFREAK ist in diesem Fall schlechter als die von FREAK in seiner ursprünglichen Form.

6 Diskussion der Ergebnisse

Die Experimente der vorliegenden Arbeit gliederten sich in einen theoretischen Teil und einen praktischen, in welcher eine Anwendung demonstriert wurde. Im theoretischen Teil wurden zunächst zwei verschiedene Verfahren anhand eines in der Forschung etablierten Datensatzes getestet, den FREAK-Deskriptor um Farbe zu erweitern. Auf diesen Ergebnissen aufbauend wurde anschließend auch ein Praxisversuch durchgeführt, wobei das Ziel war, farblich verschiedene Objekte in einem Kamerabild zu erkennen. Dieses Kapitel soll die Ergebnisse zusammenführen und bewerten, um abschließend eine Beurteilung zuzulassen, inwieweit die Nutzung von Farbinformation gewinnbringend sein kann.

6.1 Vergleich verschiedener Farbdarstellungen

Der FREAK-Deskriptor beschreibt Schlüsselpunktumgebungen durch einen Bitstring, welcher das Ergebnis von Intensitätsvergleichen verschieden großer, rezeptiver Felder ist. Unterscheiden sich zwei dieser Bitstrings an bestimmten Stellen, so zeigt das dies ein voneinander verschiedenes Ergebnis bestimmter Intensitätsvergleiche an. Die Reihenfolge der Bits ist dabei von Bedeutung, da auch die Reihenfolge der Vergleiche festgesetzt ist. Auf dieser Einsicht aufbauend soll zunächst die Eignung bestimmter Farbdarstellungen besprochen werden, welche im ersten Experiment in Abschnitt 4.2 genutzt wurden. Die Integration der Farbe erfolgte in diesem durch die Konkatenation der Deskriptorvektoren verschiedener Farbkanäle.

Der hueFREAK-Deskriptor, welcher Farbe anhand des Farbwinkels kodiert, schneidet in allen Fällen am schlechtesten ab. Einzig der Farbwinkel scheint Farbe im Kontext des FREAK-Deskriptors also sehr schlecht zu beschreiben. Für die Begründung sei noch einmal auf Grafik 3.6 (siehe “Hue”) verwiesen, welche zeigt, was der Farbwinkel konkret bedeutet. Anhand eines Beispiels soll das auftretende Problem verdeutlicht werden. Gegeben seien zwei rezeptive Felder, in einem dominiert die Farbe Blau ($\approx 225°$), in dem anderen Grün ($\approx 135°$). Ein Vergleich würde also ergeben, dass Blau “größer” ist als Grün, was in einem Bit festgehalten wird. Genauso kann Blau jedoch auch mit Cyan, mit Gelb oder auch bestimmten Tönen von Rot verglichen werden. Das Ergebnis wäre als einzelnes Bit ausgedrückt in jedem dieser Fälle gleich, denn Blau ist ausgedrückt als

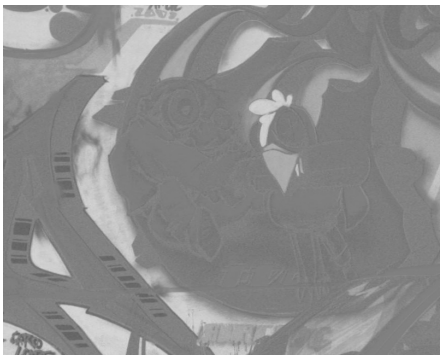
Winkel “größer” als die genannten Farben, welche sich, ebenfalls als Farbwinkel ausgedrückt, zwischen 0° und 225° bewegen (vergleiche Abbildung 3.4). Der Deskriptor würde also einen Farbwechsel von Cyan nach Gelb nicht beschreiben können. Es gibt jedoch noch drastischere Auswirkungen dieser Darstellung. Angenommen, die verglichenen rezeptiven Felder eines Schlüsselpunktes würden sich kaum unterscheiden und sehr nahe beieinander liegen. Beide könnten durch Blau dominiert sein, eines durch 225°, das andere durch 224°, letzteres also kaum merkbar Richtung Cyan gelagert sein. Durch eine perspektivische Verzerrung könnte letztgenanntes dann leicht Richtung Magenta verschoben werden und in der Folge bei einem Farbton von ungefähr 226° liegen – der Intensitätsvergleich würde das entsprechende Bit kippen, obwohl sich der Farbwinkel kaum verändert hat. An der Stelle sei auch noch einmal auf die Instabilität des Farbtons nahe der Grauwertachse verwiesen (siehe Abbschnitt 3.2), welche das Problem zusätzlich verstärkt, da der Farbton bei geringer Sättigung schwer festzustellen ist und entsprechend schwanken kann.

Ebenfalls schlechte Ergebnisse im Experiment zeigt oppFREAK, welcher Farben durch zwei Kanäle anhand des Gegenfarbmodells beschreibt. Ein Darstellung, welches beispielhaft die Gegenfarbkanäle eines Bildes aufführt, ist in Abbildung 6.1 gegeben. Zunächst fällt auf, dass diese verglichen mit dem Intensitätsbild, also dem hell–dunkel Kontrast, nach Grau verschoben scheinen. Extreme Bereiche, also sehr hell oder sehr dunkel, treten hingegen selten auf. Der Grund hierfür liegt darin, dass der Wertebereich gestaucht wird: Sowohl Schwarz als auch Weiß werden in beiden Kanälen auf 128 abgebildet und sind nicht weiter unterscheidbar. Gewissermaßen wird Information über die Intensität zugunsten der Farbe gelöscht. Im Kontext des FREAK-Deskriptors ist dies ein Nachteil, da binäre Deskriptoren auf genau diesen Intensitätskontrasten beruhen und eine hohe Varianz derselben voraussetzen. Optimal durch oppFREAK beschreibbare Bilder dürften kein Weiß oder Schwarz enthalten und müssten durch deutliche Übergänge von Blau nach Gelb und Rot nach Grün gekennzeichnet sein, denn auch hier gilt, was bereits für hueFREAK bemerkt wurde: liegen Werte zu nahe beieinander, kann das bei geringfügigen Änderungen ein unerwünschtes Kippen der Bits zur Folge haben. Hier tritt dieses Problem im Bereich der Werte um 128 auf. Nimmt man hingegen die Intensität als weiteren Kanal hinzu, wie geschehen für oppFREAK in Form von oppiFREAK oder für hueFREAK durch hsvFREAK, erhöht sich die Leistung der Deskriptoren im Ver-

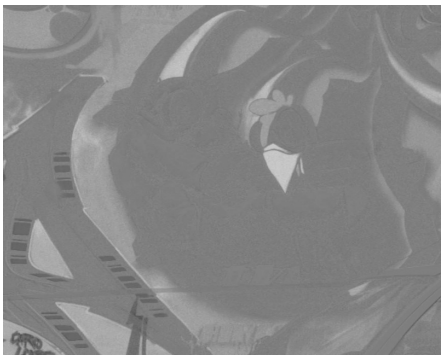
gleich. Auf dem Motiv `wall` übertreffen diese sogar sowohl FREAK, als auch rgbFREAK. Wie sich das im Detail begründet, müsste in weiteren Experimenten erforscht werden. Da hsvFREAK und oppiFREAK jedoch auf den anderen Datensätzen nicht besser als rgbFREAK abschneiden, kann hier nicht ohne Weiteres verallgemeinert werden.
Die besten Ergebnisse erzielt rgbFREAK, welcher den Wertebereich unverändert lässt, und dadurch auch schwarze und weiße Bereiche eines Bildes gut beschreibt (siehe Abbildung 6.2). Für die Motive `leuven` und `ubc` kann gegenüber dem unveränderten FREAK-Deskriptor mit keinem der Farbdeskriptoren im Test eine Verbesserung erzielt werden. Dieser Umstand soll in einem eigenen Abschnitt besprochen werden.

(a) Original

(b) Intensität



(c) Grün–Rot



(d) Blau–Gelb

Abbildung 6.1: Die Konvertierung eines Farbbildes in das vorgestellte Gegenfarbmodell löscht Intensitätsinformation und verschiebt den Wertebereich dahingehend, dass sehr helle und sehr dunkle Farben in den Bereich von Grau abgebildet werden. Wenn Grün, Rot, Gelb und Blau selten gesättigt auftreten, hat das eine schlechte Ausschöpfung des Wertebereichs zur Folge.

(a) Rot (b) Grün (c) Blau

Abbildung 6.2: Dargestellt sind die RGB-Kanäle des Bildes aus Abbildung 6.1(a). Die Nutzung der einzelnen Kanäle des RGB-Farbraumes behält wichtige Intensitätsinformationen bei und schneidet im Vergleich besser ab als die anderen Farbdeskriptoren.

Darüber hinaus ist rgbFREAK homogen bezüglich der Bedeutung einzelner Bits. Als Abstandsmaß wird im Experiment der integrierten Farbe der Hamming-Abstand verschiedener Deskriptorvektoren genutzt. Im Kontext von hsv- oder oppiFREAK aber kodieren die einzelnen Bits teilweise verschiedene Eigenschaften eines Schlüsselpunktes, was bei der Abstandsberechnung hingegen keine Beachtung findet. Anhand eines konstruierten Beispiels soll die sich daraus ergebende Problematik verdeutlicht werden: Im Grunde ist es denkbar, bedeutende Änderungen durch unbedeutende zu relativieren. Es wurde bereits festgestellt, dass der Intensität eine Schlüsselrolle bei der Berechnung des Deskriptorvektors zukommt. Eine Änderung der Intensität, welche sich in den entsprechenden Bits niederschlägt, ist also eine wichtigere Aussage zu Unterscheidung, als eine etwaige Änderung der Bits, welche den Farbwinkel kodieren. Gegeben sei der Deskriptorvektor D_A, welcher mit D_B und D_C verglichen werden soll:

$$
\begin{aligned}
D_A &= \overbrace{0\ \ 0\ \ 0\ \ 0}^{\text{Intensität}}\ \overbrace{0\ \ 0\ \ 0\ \ 0}^{\text{Farbwinkel}} \\
D_B &= 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 1\ \ 1\ \ 1, \qquad d(D_A, D_B) = 4 \\
D_C &= 1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0, \qquad d(D_A, D_C) = 3
\end{aligned}
$$

Obwohl die Schlüsselpunktumgebung aufgrund der Intensität als gleich zu betrachten wäre, der Farbwinkel jedoch aufgrund der beschriebenen Änderung die Bits von D_B kippen lässt, wird dies nicht berücksichtigt. In diesem Fall wird ein geringerer Abstand

zu D_C ermittelt, der sich bezüglich der Intensität jedoch stark von D_A unterscheidet. Semantisch wäre hier eine Trennung von Farbinformation und Intensität von Vorteil. Der rgbFREAK Deskriptor kodiert in jedem Bit Intensität, die Bits können also diesbezüglich als gleichwertig betrachtet werden.

6.2 Die Eignung von Farbabständen zum Vergleich

Eine klare Trennung der Farb- und Intensitätsinformation nimmt der zweite Ansatz in Abschnitt 4.3 vor. Dieser trennt die Entscheidung über die Übereinstimmung von Deskriptorvektoren und die der Farbvektoren in der Umgebung eines Schlüsselpunktes. In der Konsequenz entstehen zwei Klassifikatoren, deren Urteil dann miteinander verknüpft wird. Bei der Erhebung der Farbe wird sich am Muster des FREAK-Deskriptors orientiert, insofern ist auch der Farbvektor durch den FREAK-Deskriptor motiviert.
Andererseits wird im zweiten Experiment das zentrale Element der paarweisen Intensitätsvergleiche binärer Deskriptoren verworfen. Die Farbinformation wird in Form reellwertiger Vektoren verschiedener Länge beschrieben, womit auch der Hamming-Abstand nicht zur Anwendung gebracht werden kann und durch den euklidischen Abstand ersetzt wird. Dies kann als deutlicher Bruch mit der Methodik des FREAK-Deskriptors gesehen werden.
Die Ergebnisse zeigen eine leichte Verbesserung verglichen mit FREAK, welche wie im ersten Experiment auf dem Datensatz `graf` am deutlichsten ist. Der Grund für die Verbesserung könnte darin liegen, dass das Graffiti-Motiv im Gegensatz zu den anderen Motiven einen hohen Anteil gesättigter Farben aufweist. Wie auch im ersten Experiment kann keine Verbesserung für die Motive `leuven` und `ubc` erreicht werden, vielmehr ist die Betrachtung der Farbe auch hier nachteilig.

Für das praktische Experiment wurde der Ansatz, Farben in Form eines Farbvektors vom FREAK-Deskriptor zu trennen, nicht weiter verfolgt. Dieser bringt verglichen mit rgbFREAK einen schwerwiegenden Nachteil mit sich: Im Experiment hatte sich gezeigt, dass eine Ermittlung des maximalen Farbanstandes nötig ist, um tatsächlich eine Verbesserung gegenüber FREAK zu erzielen. Dieser unterschiedet sich außerdem von Bildreihe zu Bildreihe, im praktischen Einsatz wäre jedoch wünschenswert, diesen nicht für verschiedene Einsätze gezielt ermitteln zu müssen.

Insgesamt wurden im zweiten theoretischen Ansatz vier verschiedene Farbvektoren getestet, welche sich in der Zahl der erhobenen Farben unterscheiden. Eine Rangfolge der Leistung bezüglich der Länge des Vektors ist dabei nicht erkennbar. In drei von acht Fällen zeigt d13-FREAK das beste Ergebnis, gefolgt von d19- und d7-FREAK mit dem besten Ergebnis in zwei von acht Fällen. Der d1-FREAK Deskriptor zeigt nur auf dem Datensatz `leuven` das beste Ergebnis, bleibt hier aber auch hinter FREAK als Urverfahren zurück. Die Robustheit gegenüber Beleuchtungsänderungen stellt sich somit in allen Experimenten als wichtiger Punkt heraus, weswegen dies im folgenden Abschnitt separat besprochen werden soll.

6.3 Die Bedeutung von Beleuchtung und Aufnahmequalität

In keinem der Experimente in Abschnitt 4 konnte eine Verbesserung durch die Hinzunahme von Farbinformation für die Motive `leuven` und `ubc` erzielt werden. Darauf folgend fiel in der praktischen Anwendung auf, dass rgbFREAK im Schnitt weniger Schlüsselpunkte findet als FREAK. Diese Umstände machen einen großen Nachteil sichtbar, welcher die zusätzliche Unterscheidung von Farbe mit sich bringt: Gegenüber dem FREAK-Deskriptor geht die Robustheit gegenüber *Beleuchtungsänderungen* und *Farbartefakten* verloren. Die Invarianz gegenüber Beleuchtungsänderungen kann anhand der Bildreihe `leuven` untersucht werden, in welcher das Bild zunehmend verdunkelt wird. Die Robustheit gegenüber Farbartefakten ist anhand der Bildreihe `ubc` untersuchbar. Durch starke Komprimierung erscheinen auf den Vergleichsbildern der Bildreihe `ubc` zusätzliche Farben, welche im Originalbild nicht vorhanden, und den Kompressionsalgorithmen geschuldet sind. Im praktischen Experiment des Abschnitts 5 treten diese Probleme gleichzeitig auf: Einerseits kann eine gleichmäßige Beleuchtung, wie sie beim Scannen der Referenzbilder möglich war, für Kamerabilder nicht sichergestellt werden. Andererseits kann es durch die Verwendung einfacher Kameras mit schlechter Farbauflösung dazu kommen, dass Farben nicht zuverlässig erfasst werden. Bereits in Abschnitt 5.2 wurde darauf hingewiesen, dass dieses Problem in der praktischen Anwendung beispielsweise für die Farbe Weiß beobachtbar ist.
Dennoch konnte unter Verwendung von rgbFREAK eine Leistungssteigerung gegenüber FREAK auf den von `ubc` und `leuven` verschiedenen Bildreihen beobachtet werden. An dieser Stelle muss also im Anwendungskontext abgewogen werden, ob es lohnt, diese

Leistungssteigerung auf Kosten eines Verlustes der Robustheit gegenüber Beleuchtungsänderung oder Bildartefakten erreichen zu wollen. Ist sichergestellt, dass die genannten Transformationen nicht auftreten, kann dies durchaus in Erwägung gezogen werden. Eine Anwendung könnte beispielsweise sein, die Verwendung geschützter Bildern in Büchern festzustellen, Voraussetzung ist dann allerdings, dass die einzelnen Seiten mit einer gleichbleibend hohen Aufnahmequalität bei gleichbleibender Beleuchtung erfasst werden können.

7 Schlussbetrachtung

In der vorliegenden Arbeit wurden verschiedene Ansätze geprüft, den FREAK-Deskriptor um Farbinformation zu erweitern. Zunächst erfolgte eine Darstellung der Grundlagen, in welcher dargelegt wurde, wie die Bildbeschreibung anhand lokaler Merkmale funktioniert, wobei binäre Deskriptoren von den etablierten reellwertigen Deskriptoren abgegrenzt wurden. Auch eine Darstellung der Methodik, anhand derer Deskriptoren miteinander verglichen werden können, wurde vorgenommen. In einem zweiten Grundlagenkapitel erfolgte eine Einführung zu Farben, wobei einerseits erklärt wurde, was unter dem Phänomen Farbe zu verstehen ist und darauf folgend, wie Farbe durch verschiedene Farbräume dargestellt werden kann. Das Farbkapitel abgeschlossen hat eine Darstellung verschiedener Versuche, das etablierte Verfahren der SIFT um Farbinformation zu erweitern.

In den Experimenten der Arbeit wurden zwei Ansätze geprüft, die Farbinformationen eines Schlüsselpunktes bei dessen Beschreibung einzubeziehen. Dies geschah im ersten Versuch, indem der FREAK-Deskriptor der einzelnen Kanäle verschiedener Farbräume konkateniert wurde. In der Folge konnten wichtige Eigenschaften des Deskriptors beibehalten werden, beispielsweise dessen binäre Form und damit auch der Hamming-Abstand als Vergleichskriterium. Dabei hat sich rgbFREAK, mit dem RGB-Farbraum als Grundlage, als am geeignetsten herausgestellt, welcher auf sechs von acht Bildreihen zur Evaluation den FREAK-Deskriptor übertraf. Auf Bildreihen mit auftretender Beleuchtungsänderung und durch Kompression eingebrachten Farbartefakten blieb FREAK hingegen unübertroffen.
Im zweiten Experiment wurde der FREAK-Deskriptor selbst unverändert übernommen, die Farben eines Schlüsselpunktes hingegen in Form eines Farbvektors als zusätzliches Attribut verarbeitet. Im Grundlagenkapitel eingeführte Farbabstände ermöglichten so die Beurteilung der Ähnlichkeit von Schlüsselpunkten, was mit dem Urteil des FREAK-Deskriptors verknüpft wurde. Auch hier konnte eine Verbesserung gegenüber FREAK in den selben sechs von acht Fällen erzielt werden. Allerdings setzte dies für jede Bildreihe die Feststellungen des maximal auftretenden Farbabstandes als Parameter voraus. Das wurde als erheblicher Nachteil gewertet, weswegen für diesen Ansatz keine Empfehlung ausgesprochen werden kann. Die wichtigste Erkenntnis der Experimente ist die Tatsa-

che, dass die Betrachtung von Farbe durchaus eine Verbesserung bringen kann, diese bei Beleuchtungsänderungen und schlechter Bildqualität jedoch zum Nachteil wird, da viele Schlüsselpunkte im Gegensatz zu FREAK nun falsch verworfen werden.

Im praktischen Experiment wurde eine konkrete Anwendung des Verfahrens der Bildbeschreibung lokaler Merkmale demonstriert. Dabei sollten sehr ähnliche Objekte, welche sich jedoch farblich stark unterschieden, identifiziert werden. Das Ziel war eine Überprüfung, inwiefern die in den Experimenten gewonnenen Erkenntnisse in der Praxis von Relevanz sind. Erneut zeigte sich hier, dass bezogen auf die Diskriminativität bei sehr ähnlichen Objekten eine Verbesserung erzielt werden kann, welche aber für den praktischen Einsatz nicht als vorrangig zu bewerten ist. Selbst auf einem ausgesuchten Datensatz, welcher gezielt Farbe als Merkmal von Objekte betont, konnte durch FREAK in seiner ursprünglichen Form eine korrekte Zuordnung aller Objekte erfolgen. Die in den Experimenten festgestellten Nachteile der fehlenden Robustheit gegenüber Beleuchtungsänderung und schlechter Bildqualität von rgbFREAK traten auch im praktischen Experiment als nachteilig hervor.

Als abschließendes Urteil kann festgehalten werden, dass eine Betrachtung der Farbe durchaus lohnenswert sein kann. Dafür muss jedoch sichergestellt sein, dass Beleuchtungsänderungen und Farbartefakte ausgeschlossen sind. Denkbar ist das in Szenarien, in denen hochqualitative Bilder unter gleichbleibender Lichtsituation aufgenommen werden können, etwa beim Scannen von Buchseiten. Gegenüber allen getesteten Verfahren hat sich hier rgbFREAK als am geeignetsten herausgestellt. In Anwendungskontexten wie beispielsweise der Robotik, wo einfache Kameras in verschiedenen Beleuchtungssituationen Objekte identifizieren sollen, ist von einer Unterscheidung der Farben hingegen abzuraten, da besonders hier die genannten Schwierigkeiten von Farbartefakten und Beleuchtungsänderungen besonders hervortreten. An dieser Stelle kann der FREAK-Deskriptor in seiner ursprünglichen Form gute Ergebnisse bringen.

Offene Fragen und Ausblick

In der Arbeit wurden zwei Ansätze der Farberweiterung verfolgt. Dabei wurden verschiedene Farbräume für die Kodierung der Farben getestet, und deren Eignung untersucht.

Offen bleibt dabei allerdings, ob nicht eine andere Möglichkeit der Farberweiterung ein besseres Ergebnis erzielen kann, welche hier nicht getestet wurde. Beispielsweise könnte das anhand eines hier nicht eingesetzten Farbraumes geschehen, welcher im Kontext der Intensitätsvergleiche binärer Deskriptoren mögliche Vorteile besitzt. Entsprechend müsste ein solcher Ansatz jedoch auch die aufgezeigten Probleme lösen, welche eine Farberweiterung mit sich bringt.

Alle vorgestellten binären Deskriptoren sind zum Zeitpunkt der Anfertigung dieser Arbeit sehr neue Verfahren. Auf dem Gebiet der Beschreibung lokaler Bildmerkmale sind diese gegenüber den ebenfalls vorgestellten etablierten Ansätzen reellwertiger Deskriptoren eine wichtige Entwicklung: Sowohl in Leistung, als auch in Geschwindigkeit sind die binären den reellwertigen Deskriptoren überlegen. Im Spannungsverhältnis einfacher Lösungen gegenüber den detailverliebten Verfahren sprechen binäre Deskriptoren für die erste Gruppe und zeigen, dass einfache Ansätze trotz, oder gerade durch ihre Einfachheit, sehr gute Ergebnisse erzielen können. Binäre Deskriptoren werfen gleichzeitig die Frage auf, wie weit diese Vereinfachung getrieben werden kann, prägnant ausgedrückt in der Frage “How low can you go?”. Spannend bleibt dabei, wie sich diese Entwicklung im Kontext der Bildbeschreibung anhand lokaler Merkmale fortsetzen wird.

Literatur

Alexandre Alahi, Raphael Ortiz, und Pierre Vandergheynst. FREAK: Fast Retina Keypoint. In *Conference on Computer Vision and Pattern Recognition*, pages 510–517. IEEE, 2012. ISBN 978-1-4673-1226-4.

Herbert Bay, Andreas Ess, Tinne Tuytelaars, und Luc Van Gool. Speeded-Up Robust Features SURF. *Computer Vision and Image Understanding*, 110(3):346–359, June 2008.

Paul Beaudet. Rotationally Invariant Image Operators. In *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pages 579–583, 1978.

Anna Bosch, Andrew Zisserman, und Xavier Munoz. Scene Classification Using a Hybrid Generative/Discriminative Approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(4):712–727, April 2008.

Matthew Brown und David Lowe. Invariant Features from Interest Point Groups. In *British Machine Vision Conference*, page Poster Session, 2002.

Michael Calonder, Vincent Lepetit, Christoph Strecha, und Pascal Fua. BRIEF: Binary Robust Independent Elementary Features. In *Computer Vision - ECCV 2010, 11th European Conference on Computer Vision*, volume 6314 of *Lecture Notes in Computer Science*, pages 778–792. Springer, 2010. ISBN 978-3-642-15560-4.

Franklin Crow. Summed-Area Tables for Texture Mapping. In *Computer Graphics (Special Interest Group on Graphics and Interactive Technique '84 Proceedings)*, volume 18, pages 207–212, July 1984.

Navneet Dalal und Bill Triggs. Histograms of oriented gradients for human detection. In *Conference on Computer Vision and Pattern Recognition*, pages 886–893, 2005.

Christopher Evans. Notes on the OpenSURF Library. Technical Report CSTR-09-001, University of Bristol, January 2009.

Martin Fischler und Robert Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

Gerd Gigerenzer und Peter M. Todd. *Simple Heuristics That Make Us Smart*. Oxford University Press, New York, 1999. ISBN 978-0195143812.

Kristen Grauman und Bastian Leibe. *Visual Object Recognition*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2011.

Richard Hamming. Error Detecting and Error Correcting Codes. *Bell System Technical Journal*, 29:147, April 1950.

Chris Harris und Mike Stephens. A combined corner and edge detector. *Alvey Vision Conference*, pages 147–151, 1988.

Ewald Hering. *Zur Lehre vom Lichtsinne. Zweiter, unveränderter Abdruck*. Geronds Sohn, Wien, 1878. URL `http://www.deutschestextarchiv.de/book/view/hering_lichtsinn_1878`.

Ron Kimmel. Demosaicing: Image Reconstruction from Color CCD Samples. *IEEE Trans. Image Processing*, 8(9):1221–1228, September 1999.

Rainer Klinke, Hans-Christian Pape, und Stefan Silbernagl. *Physiologie, 5.Auflage*. Thieme, 2005.

Stefan Leutenegger, Margarita Chli, und Roland Siegwart. BRISK: Binary Robust Invariant Scalable Keypoints. In *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, pages 2548–2555. IEEE, 2011.

David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

Elmar Mair, Gregory D. Hager, Darius Burschka, Michael Suppa, und Gerd Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. In *Computer Vision - ECCV 2010, 11th European Conference on Computer Vision*, volume 6312 of *Lecture Notes in Computer Science*, pages 183–196. Springer, 2010. ISBN 978-3-642-15551-2.

Krystian Mikolajczyk und Cordelia Schmid. A Performance Evaluation of Local Descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.

Konstantinos Plataniotis und Anastasios Venetsanopoulos. *Color image processing and applications*. Springer-Verlag New York, Inc., New York, NY, USA, 2000. ISBN 3-540-66953-1.

Edward Rosten und Tom Drummond. Machine Learning for High-Speed Corner Detection. In *European Conference on Computer Vision*, pages 430–443, 2006.

Ethan Rublee, Vincent Rabaud, Kurt Konolige, und Gary R. Bradski. ORB: An efficient alternative to SIFT or SURF. In *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, pages 2564–2571. IEEE, 2011.

Chris Solomon. *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley, 2011. ISBN 0-470-84472-8.

John Swets. Effectiveness of information retrieval methods. *American Documentation*, 20(1):72–89, 1969.

Richard Szeliski. *Computer Vision Algorithms and Applications*. Springer, 2011. ISBN 978-1-84882-934-3.

Matthew Turk und Alex Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuro Science*, 3(1):71–86, 1991.

Tinne Tuytelaars und Krystian Mikolajczyk. Local Invariant Feature Detectors: A Survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2007.

Koen van de Sande, Theo Gevers, und Cees Snoek. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010.

Joost van de Weijer und Cordelia Schmid. Coloring Local Feature Extraction. In *European Conference on Computer Vision*, pages II: 334–348, 2006.

Paul Viola und Michael Jones. Rapid object detection using a boosted cascade of simple features. *Proc. Conference on Computer Vision and Pattern Recognition*, 1:511–518, 2001.

A Anhang

A.1 Ableitungen der Gaußfunktion

Die Gaußfunktion, welche in der Bildverarbeitung beispielsweise zur Glättung von Bildern verwendet wird, ist für zwei Veränderliche wie folgt definiert:

$$L(x,y) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

Der Parameter $\sigma > 0$ bestimmt dabei die Breite der entstehenden Glocke. Höhere Werte für σ haben bei Anwendung des Filters eine stärkere Glättung zur Folge. Für die ersten Ableitungen von L ergibt sich:

$$L_x(x,y) = -\frac{x}{\sigma^3\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$
$$L_y(x,y) = -\frac{y}{\sigma^3\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

Durch erneutes Differenzieren erhält man die zweiten Ableitungen, für die sich folgende Funktionsvorschriften ergeben:

$$L_{xx}(x,y) = \frac{x^2-\sigma^2}{\sigma^5\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$
$$L_{yy}(x,y) = \frac{y^2-\sigma^2}{\sigma^5\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$
$$L_{xy}(x,y) = L_{yx}(x,y) = \frac{xy}{\sigma^5\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

Eine grafische Darstellung der Gaußfunktion sowie der zweiten Ableitungen findet sich in Abbildung 2.4.

A.2 Umrechnung von Farbräumen

RGB nach Grauwert

Die Konvertierung eines RGB-Bildes in ein Grauwertbild ist anhand der folgenden Formel möglich:

$$I = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$$

Da die Grün-empfindlichen Zapfen des menschliches Auges einen höheren Anteil beim Kontrastsehen haben als Rot und Blau, erzeugt dies aus Sicht des Menschen ein besseres Ergebnis als etwa eine einfache Mittelung der drei Farbkanäle.

RGB nach HSV

Die Umrechnung von RGB-kodierten Farben $R, G, B \in [0, 1]$ nach HSV ist nach folgender Vorschrift möglich:

$$\mathrm{H} = \begin{cases} 0°, & \text{falls } max = min \\ 60° \cdot \left(0 + \frac{G-B}{max-min}\right), & \text{falls } max = R \\ 60° \cdot \left(2 + \frac{B-R}{max-min}\right), & \text{falls } max = G \\ 60° \cdot \left(4 + \frac{R-G}{max-min}\right), & \text{falls } max = B \end{cases}$$

$$\mathrm{S} = \begin{cases} 0, & \text{falls } max = 0 \\ \frac{max-min}{max}, & \text{sonst} \end{cases}$$

$$\mathrm{V} = max$$

Es gilt dabei $max := \max(R, G, B)$ und $min := \min(R, G, B)$. Sollte $\mathrm{H} < 0°$ zutreffen, so muss der Wert 360° zu H addiert werden.

RGB zu Gegenfarben

Das Gegenfarbmodell stellt Farben als Rot–Grün Kontrast O_1, sowie als Blau–Gelb Kontrast O_2 dar. Zusätzlich wird ein Intensitätskanal O_3 angegeben, welcher hier durch einfache Mittelung der drei Farbkanäle gebildet wird.

Seien $R, G, B \in [0, 1]$, dann werden die Kanäle wie folgt berechnet:

$$
\begin{aligned}
O_1 &= \frac{R - G}{\sqrt{2}} \\
O_2 &= \frac{R + G - 2B}{\sqrt{6}} \\
O_3 &= \frac{R + G + B}{\sqrt{3}}
\end{aligned}
$$

Diese Definiton des Gegenfarbraumes folgt der Definition von van de Sande et al. (2010). Eine entsprechende Skalierung in den Wertebereich $[0, 1]$ erhält man anhand folgender Formeln:

$$
\begin{aligned}
O_1 &= \frac{R - G + 1}{2} \\
O_2 &= \frac{R + G - 2B + 2}{4} \\
O_3 &= \frac{R + G + B}{3}
\end{aligned}
$$

RGB nach L*a*b*

Um RGB-Farben in den L*a*b*-Farbraum zu konvertieren, muss als Ausgangspunkt zunächst eine Umrechnung der RGB-Farben in den XYZ-Farbraum vorgenommen werden. Dies erfolgt anhand folgender Lineartransformation:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

Anschließend ist die Konvertierung der XYZ-Farben in den L*a*b*-Farbraum unter Verwendung folgender Formeln möglich:

$$\begin{aligned} L^* &= 116 \cdot f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* &= 500 \cdot \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \\ b^* &= 200 \cdot \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \end{aligned}$$

wobei die Funktion $f(t)$ und die Konstanten wie folgt definiert sind:

$$\begin{aligned} f(t) &= \begin{cases} \sqrt[3]{t}, & \text{falls } t > \delta^3 \\ t/(3\delta^2) + 2\delta/3, & \text{sonst} \end{cases} \\ \delta &= 6/29 \approx 0.2068966 \\ X_n &= 0.950456 \\ Y_n &= 1.0 \\ Z_n &= 1.088754 \end{aligned}$$

A.3 Tabellen

Gelistet sind die gerundeten Flächeninhalte unter den Kurven der Experimente in Abschnitt 4.2 und 4.3. Die Maxima einer Zeile sind zum Zwecke der Lesbarkeit durch Hinterlegung hervorgehoben.

Experiment Abschnitt 4.2

	FREAK	rgbF.	oppF.	oppiF.	hueF.	hsvF.	rgbF./F.
bark	0.395	0.436	0.148	0.339	0.112	0.375	1.104
bikes	0.791	0.793	0.469	0.666	0.358	0.698	1.003
boat	0.611	0.611	0.0	0.609	0.0	0.609	1.000
graf	0.368	0.437	0.171	0.286	0.073	0.323	1.188
leuven	0.756	0.744	0.351	0.532	0.312	0.524	0.984
trees	0.420	0.435	0.283	0.382	0.130	0.386	1.036
ubc	0.905	0.887	0.029	0.405	0.009	0.333	0.980
wall	0.511	0.584	0.527	0.638	0.336	0.636	1.143

Tabelle A.1: Die Tabelle listet die Werte des Experimentes in Abschnitt 4.2.

Experiment Abschnitt 4.3

	FREAK	d1-F.	d7-F.	d13-F.	d19-F.	*max*/F.
bark	0.395	0.404	0.418	0.419	0.412	1.061
bikes	0.791	0.795	0.807	0.804	0.809	1.023
boat	0.611	0.616	0.628	0.638	0.637	1.044
graf	0.368	0.377	0.419	0.430	0.426	1.168
leuven	0.756	0.286	0.189	0.197	0.197	0.378
trees	0.420	0.426	0.439	0.443	0.444	1.057
ubc	0.905	0.881	0.894	0.892	0.892	0.988
wall	0.511	0.524	0.544	0.518	0.530	1.065

Tabelle A.2: Gelistet sind die Werte des Experimentes in Abschnitt 4.3.

A.4 Anmerkungen zur Implementierung

Die Implementierung der Experimente in Abschnitt 4 sowie der praktischen Anwendung in Abschnitt 5 erfolgte in C++ unter Verwendung der OpenCV Bibliothek in der Version 2.4.5. Für die statistische Auswertung der Experimente kam Python zum Einsatz. Der Quelltext kann unter `http://page.mi.fu-berlin.de/brachman/msc-freak/` heruntergeladen werden.

Für den Versuch zur praktischen Anwendung in Abschnitt 5 wurde ein MacBook (Ende 2008) verwendet und die darin verbaute Kamera eingesetzt. Die Aufzeichnung kann ebenfalls unter den genannten Adresse eingesehen werden.

Der in den Experimenten eingesetzte Datensatz von Mikolajczyk und Schmid (2005) ist verfügbar unter `http://www.robots.ox.ac.uk/~vgg/data/data-aff.html`.

Erklärung zur Urheberschaft

Ich versichere hiermit, dass diese Arbeit von niemand anderem als meiner Person verfasst worden ist. Alle verwendeten Hilfsmittel sind im Literaturverzeichnis angegeben, Zitate aus fremden Arbeiten sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungskommission vorgelegt.

Berlin, 11. Oktober 2013