Ein deformationsinvarianter Point-of-Interest-Detektor

Arne Jacobs

Dissertation

zur Erlangung des Grades eines
Doktors der Ingenieurwissenschaften
- Dr.-Ing. -

Vorgelegt im Fachbereich 3 (Mathematik und Informatik)
der Universität Bremen
im Mai 2010

Hiermit möchte ich mich bei allen bedanken, die mir bei der Fertigstellung dieser Doktorarbeit mit inhaltlichem oder persönlichem Beistand geholfen haben, allen voran meinem Betreuer Prof. Dr. Otthein Herzog, meiner Freundin Linda Krannich und meinen Kollegen, speziell Daniel Möhlmann, der am Schreibtisch gegenüber jederzeit für einen Gedankenaustausch zur Verfügung stand. Ich danke auch Prof. Dr. Adalbert F.X. Wilhelm für seine Bereitschaft, als Zweitgutachter zur Verfügung zu stehen, und außerdem meinen Eltern und Freunden, die für Motivation gesorgt haben.

Inhaltsverzeichnis

1 Einleitung **7**
- 1.1 Motivation . . . 7
- 1.2 Ziel der Arbeit . . . 7
- 1.3 Aufbau der Arbeit . . . 8

2 Detektion und Beschreibung von Points-of-Interest **9**
- 2.1 Existierende Verfahren zur Detektion von Points-of-Interest . . . 9
  - 2.1.1 Autokorrelationsbasierte Ansätze . . . 10
  - 2.1.2 Blob-basierte Ansätze . . . 11
  - 2.1.3 Informationstheoretische Ansätze . . . 12
  - 2.1.4 Detektoren auf raum-zeitlichen und 3D-Volumendaten . . . 13
- 2.2 Existierende Verfahren zur Bestimmung lokaler Deskriptoren . . . 13
  - 2.2.1 Deskriptor-Codebooks . . . 15
- 2.3 Vergleichende Evaluationen . . . 16

3 Deformationsinvariante Bildbeschreibung mit Graphen **19**
- 3.1 Äquivalenz der Projektion von 3D-Transformationen und 2D-Deformationen 19
- 3.2 Ein deformationsinvariantes Abstandsmaß . . . 23
  - 3.2.1 Invarianz gegenüber Deformationen . . . 25
- 3.3 Repräsentation eines Bildes als Graph . . . 27
  - 3.3.1 Exakte Repräsentation . . . 27
  - 3.3.2 Graphmetrik . . . 33
  - 3.3.3 Reduzierte Repräsentation durch abstandsbasierte Approximation . 34
- 3.4 Abstände in Graphen unterschiedlicher Bildstrukturen . . . 43
  - 3.4.1 Kombinierte geodätisch-Euklidsche Graphmetrik . . . 49
  - 3.4.2 Einfluss von Rauschen . . . 56
  - 3.4.3 Markante Punkte/Points-of-Interest . . . 57
- 3.5 Point-of-Interest-Detektor . . . 60
  - 3.5.1 Algorithmus zur Bestimmung von Extrem- und Sattelpunkten . . . 60
  - 3.5.2 Rauschunterdrückung . . . 61
  - 3.5.3 Filterung nicht-lokaler Extrem- und Sattelpunkte . . . 61

4 Evaluation **63**
- 4.1 Transformationen . . . 64
  - 4.1.1 Skalierung . . . 65
  - 4.1.2 Rotation . . . 66
  - 4.1.3 Scherung . . . 67

4.1.4 Perspektivische Verzerrung 68
4.1.5 Zufällige Deformation 69
4.1.6 Hinzufügen von Gauß-Rauschen 70
4.2 Ergebnisse 70
4.2.1 Skalierung 72
4.2.2 Rotation 75
4.2.3 Scherung 78
4.2.4 Perspektivische Verzerrung 81
4.2.5 Zufällige Deformation 84
4.2.6 Hinzufügen von Gauß-Rauschen 87
4.2.7 Auswertung 90

5 Zusammenfassung und Ausblick **93**
5.1 Ausblick 94

Abbildungsverzeichnis **95**

Literaturverzeichnis **103**

1 Einleitung

Nach einer einleitenden Motivation zur Erläuterung der Bedeutung des Themengebietes der Point-of-Interest-Detektoren wird hier das Ziel dieser Arbeit dargelegt und der Aufbau der Arbeit zusammengefasst.

1.1 Motivation

Verfahren zur Detektion sogenannter Points-of-Interest, oder auch salienter oder markanter Punkte in Bildern haben sich im vergangenen Jahrzehnt als nützliche Werkzeuge im Bereich der grundlegenden (*Low-Level-*)Bildverarbeitung erwiesen. Ursprünglich als Vorstufe für Verfahren zum *Wide-Baseline Stereo* oder zur Bewegungsschätzung entwickelt, fanden sie Einsatz in Verfahren der Objekterkennung und später auch in allgemeineren Problemen des Bildverstehens. Während in kontrollierten Umgebung z.B. Segmentierungsverfahren eingesetzt werden können, um relevante Objekte vom Bildhintergrund zu trennen, um diese nach einer Merkmalsextraktion zu klassifizieren, ist das Segmentierungsproblem in unkontrollierter Umgebung bzw. auf allgemeinem Bildmaterial weitaus schwieriger. Point-of-Interest-Detektoren bieten die Möglichkeit, ohne Segmentierung und ohne Vorwissen Bildbereiche mit hohem Informationsgehalt und damit potentiell höherem Einfluss auf die *Bildbedeutung* zu erhalten. Aufgrund der lokalen Natur der Points-of-Interest zeigen auf deren Detektoren aufbauende Verfahren zudem günstige Eigenschaften in Bezug auf nicht vollständig sichtbare Objekte.

Eine wichtige Eigenschaft von Point-of-Interest-Detektoren ist die Invarianz z.B. gegenüber Änderungen des Blickwinkels oder anderer Transformationen, bzw. die Stabilität der detektierten Points-of-Interest unter solchen Transformationen. Die komplexeste Klasse von Transformationen, für die bisher invariante Detektoren vorgestellt wurden, ist die Klasse der zweidimensionalen affinen Transformationen. Von der Entwicklung von Detektoren, die invariant gegenüber komplexeren Transformationen sind, können alle aufbauenden Verfahren profitieren.

1.2 Ziel der Arbeit

Ziel der Arbeit ist die Entwicklung eines Algorithmus' zur Detektion von möglichst stabilen Points-of-Interest auf einzelnen zweidimensionalen Bildern beliebiger Szenen ohne jegliches Wissen über den Bildinhalt. Stabilität bedeutet hier, dass auf einem Bild detektierte Points-of-Interest möglichst auch auf transformierten Versionen dieses Bildes bzw. Bildern aus unterschiedlichen Blickwinkeln derselben Szene bzw. desselben Objektes detektiert werden. Da abgesehen von den konkreten Pixeldaten keine weiteren Infor-

mationen über ein Bild vorliegen, können keine Annahmen über die Aufnahmesituation getroffen werden, wie z.B. den Blickwinkel oder die Bildtiefe. Da der Bildinhalt nicht bekannt ist, können zudem keine weiteren Bilder derselben Szene oder desselben Objekts zum Vergleich herangezogen werden.

Eventuell vorliegende dreidimensionale Transformationen zwischen Bildern können daher nicht berücksichtigt werden. Wie in dieser Arbeit gezeigt werden wird, sind Deformationen die allgemeinste Klasse zweidimensionaler Transformationen (s. Abschnitt 3.1). Ziel der Arbeit ist daher die Entwicklung eines gegenüber Deformationen stabilen bzw. *invarianten* Point-of-Interest-Detektors.

1.3 Aufbau der Arbeit

Im folgenden Kapitel wird ein Überblick über existierende Verfahren zur Detektion von Points-of-Interest gegeben. Eine Prüfung bereits durchgeführter Evaluationen wird zeigen, dass das Problem für komplexe Transformationen und komplexe nicht-planare Szenen noch nicht zufriedenstellend gelöst ist. Kapitel 3 wird zunächst zeigen, dass allgemeine 2D-Deformationen die bestmögliche Transformationsklasse für das vorliegende Problem bilden. Es werden dann eine graphbasierte Bildbeschreibung und eine deformationsinvariante Graphmetrik definiert, die zur Identifikation bestimmer markanter Knoten im Bildgraphen führen (Sattelpunkte und Extrempunkte). Es wird ein Algorithmus zur automatischen Detektion dieser Punkte vorgestellt. In Kapitel 4 wird die Stabilität der von dem vorgestellten Algorithmus detektierten Points-of-Interest unter verschiedenen Transformationen getestet. Es folgen die Zusammenfassung und ein Ausblick in Kapitel 5.

2 Detektion und Beschreibung von Points-of-Interest

In diesem Kapitel wird ein Überblick über bereits existierende Verfahren zur Detektion von Points-of-Interest und Bestimmung lokaler Deskriptoren gegeben.

Da die Verfahren zur Detektion von Points-of-Interest und solche zur Bestimmung lokaler Deskriptoren überwiegend unabhängig voneinander sind, auch wenn sie häufig als Paar publiziert werden, wird dieser Überblick sie jeweils getrennt aufführen. Existierende Detektoren werden im folgenden Abschnitt gruppiert, die Deskriptoren im darauffolgenden Abschnitt. Falls ein Detektor-Deskriptor-Paar prinzipbedingt untrennbar ist, d.h., wenn der Deskriptor ohne den zugehörigen Detektor nicht anwendbar ist oder umgekehrt, so wird bei der Beschreibung des jeweiligen Ansatzes darauf explizit hingewiesen.

Bei verwandten Ansätzen wird zunächst das Grundprinzip erläutert. Danach wird jeweils auf die Unterschiede der einzelnen Ansätze eingegangen. Bei aufeinander aufbauenden Verfahren wird chronologisch vorgegangen.

Im Anschluss werden in einem gesonderten Abschnitt vergleichende Evaluationen mehrere Detektoren bzw. Deskriptoren behandelt. Auch wenn diese Evaluationen nicht erschöpfend sind hinsichtlich der gegenübergestellten Verfahren, so wird zumindest für die Detektoren klar werden, dass speziell bei zweidimensionalen Abbildungen dreidimensionaler Szenen noch Verbesserungsbedarf besteht.

2.1 Existierende Verfahren zur Detektion von Points-of-Interest

Die Menge der Points-of-Interest-Detektoren lässt sich grob in drei Klassen aufteilen, die in den folgenden Unterabschnitten separat behandelt werden. Die zuerst vorgestellte Klasse der auf der lokalen Autokorrelationsfunktion basierenden Ansätze [8, 27, 28, 30, 37] ist die kohärenteste Klasse. Es lässt sich eine eindeutige Weiterentwicklung bzgl. der Invarianz gegenüber verschiedenen Transformationen in dieser Klasse erkennen, von Rotationsinvarianz über Skalierungsinvarianz bis zur Invarianz gegenüber affinen Transformationen, wobei das Grundprinzip immer gleich bleibt. Dieser Fortschritt lässt sich auch chronologisch feststellen.

Die darauf folgend vorgestellten Blob-basierten Ansätze [1, 21, 23, 25] und die auf informationstheoretischen Maßen (z.B. der *Entropie*) basierenden Verfahren [15, 41] sind im Vergleich heterogener. Die ersten beiden Klassen vermischen sich in späteren Ansätzen.

Die am Ende dieses Abschnitts vorgestellten auf 3D-Daten (z.B. raum-zeitliche oder Volumen-Daten) basierenden Verfahren [18] werden aufgeführt, um den Unterschied zwi-

schen den dort zugrunde liegenden dreidimensionalen Daten und den in dieser Arbeit betrachteten zweidimensionalen Bildern dreidimensionaler Szenen hervorzuheben.

2.1.1 Autokorrelationsbasierte Ansätze

Die autokorrelationsbasierten Ansätze gehen auf Arbeiten aus den achtziger Jahren zurück, deren Ziel die Detektion von Ecken in Grauwertbildern war. Die wohl meistzitierte Arbeit beschreibt den sog. Harris-Corner-Detektor [8] von 1988. Ein fast identischer Detektor wurde 1994 unabhängig und mit einer gänzlich anderen Zielsetzung von Shi und Tomasi [37] hergeleitet. Auf dem Harris-Detektor aufbauende Arbeiten sind der Harris-Laplace-Detektor [27] und dessen Erweiterungen [28, 30].

Harris-Detektor

Die Haupterrungenschaft des Harris-Detektors [8] ist die Verbesserung eines von Moravec [32] auf der Basis lokaler Autokorrelation der Bildfunktion vorgestellten Ecken-Detektors und die damit einhergehende mathematische Umformulierung der Autokorrelationsfunktion hin zur Kovarianzmatrix der lokalen (gewichteten) Verteilung des Bildgradienten (der partiellen Ableitung der Bildfunktion) und deren Eigenschaften Spur und Determinante. Das Maß der *Cornerness* in einer lokalen Umgebung kann dann als Kombination der Determinante und Spur bestimmt werden, der sog. *Corner Response Function*. Lokale Maxima dieser Corner Response Function über einem bestimmten Schwellwert entsprechen den gesuchten Ecken.

Zur Berechnung des zweidimensionalen Bildgradienten werden lineare Ableitungsfilter in horizontaler und vertikaler Richtung verwendet. Im Gegensatz zum Ansatz von Moravec ist der Harris-Detektor rotationsinvariant, sofern die Gradientenmagnitude der verwendeten Gradientenfilter rotationsinvariant/isotropisch ist [37]. Die Originalimplementierung des Harris-Detektors nutzt ein anisotropische Filtermasken zur Bestimmung des Bildgradienten. Rotationsinvarianz kann aber sehr einfach durch Ersetzen der Filter durch isotropische Filter wie z.B. den Gauß'schen Ableitungsfilter erster Ordnung oder verwandte Filter wie z.B. dem sog. *Sobel*-Filter erreicht werden (s. [37] und auch Abschnitt 2.3).

Shi und Tomasi

Der dem Harris-Detektor zugrundeliegende Ansatz, die Kovarianzmatrix der lokalen Bildgradientenverteilung zu bestimmen, wurde unabhängig von Shi und Tomasi in [37] hergeleitet, mit dem Ziel, Punkte zu finden, die sich gut zur Schätzung von Bewegung z.B. für *block matching* eignen. Im Gegensatz zum Harris-Detektor werden dazu jedoch nicht Spur und Determinante herangezogen, sondern die Eigenwerte der Kovarianzmatrix.

Dem Ansatz liegt die Beobachtung zugrunde, dass lokale Umgebungen mit zwei niedrigen Eigenwerten eher unstrukturierten Flächen entsprechen (inhärent nulldimensional), Umgebungen mit einem hohen und einem niedrigen Eigenwert eher Kanten oder Verläufen (inhärent eindimensional), Umgebungen mit zwei hohen Eigenwerten dagegen aber inhärent zweidimensionalen Strukturen, an denen das sog. *Aperturproblem* der Bewe-

gungsschätzung minimal ist (wie z.B. Ecken). Dadurch kann die Qualität der Bewegungsschätzung erhöht werden [12].

Die Entscheidung, ob ein Punkt gewählt wird, richtet sich daher danach, ob der Kleinere der beiden Eigenwerte einen gegebenen Schwellwert überschreitet. Der Ansatz von Shi und Tomasi ist ebenso wie der Harris-Detektor rotationsinvariant [37].

Harris-Laplace-Detektor

Der Harris-Detektor ist nicht skalierungsinvariant [27], d.h. es werden nicht dieselben Punkte in einem Bild erkannt, wenn dieses skaliert wurde. Über ein Jahrzehnt nach Vorstellung des Harris-Detektors wurde von Mikolajczyk und Schmid ein erweiterter Ansatz vorgestellt, der Invarianz gegenüber Skalierungen bietet. Sie nennen ihren Ansatz Harris-Laplace-Detektor [27]. Die Skalierungsinvarianz des Ansatzes basiert auf Betrachtungen zum sog. *Scale Space* und einem Maß zur Schätzung der lokalen Skalierung, des sog. *Laplace-Operators* (s. auch Abschnitt 2.1.2). Über verschiedene Skalen werden mittels des Harris-Detektors Punkte detektiert und dann iterativ durch Schätzung der Skalierung an diesen Punkten verfeinert. Die Messung der lokalen Skalierung ist dabei stark verwandt mit den Blob-basierten Verfahren zur Detektion von Points-of-Interest, die im Abschnitt 2.1.2 vorgestellt werden, und die dem Harris-Laplace-Detektor chronologisch vorausgehen.

Mikolajczyk und Schmid stellen später in [28, 30] eine weitere Anpassung des Detektors vor, die invariant ist gegenüber affinen Transformationen. Erreicht wird die über Skalierung und Rotation hinausgehende affine Invarianz durch Schätzung der lokalen affinen Verzerrung, welche auf Basis der Hesse-Matrix der lokalen Gradientenverteilung vorgenommen werden kann. Der Detektor wird daher *Hessian-Laplace*-Detektor genannt. Da hier nicht mehr die Kovarianzmatrix und damit die lokale Autokorrelation die Basis für den Detektor ist, ist er eher den Blob-basierten Ansätzen zuzuordnen (s.u.). Er kann bereits annähernd perspektivische Transformationen des Bildmaterials verarbeiten, solange die gezeigte Szene hauptsächlich planar ist [30].

2.1.2 Blob-basierte Ansätze

Blobs bezeichnen zusammenhängende kompakte Bildregionen, die im Vergleich zu ihrer Umgebung sehr hell oder sehr dunkel sind.

Grundlage für einige der Blob-basierten Ansätze ist der *Laplace-Operator*, nicht zu verwechseln mit dem zwar verwandten, jedoch im Vergleich sehr stark vereinfachten *Laplace-Filter*. In seiner umfassenden Arbeit zur automatischen Skalenbestimmung in Bildern [21] führt Lindeberg den Laplace-Operator zur skalierungsinvarianten Detektion bestimmter Bildstrukturen ein, u.a. Blobs. Grundlage der Skalierungsinvarianz ist die automatische Bestimmung der lokalen Skala, gefolgt von einer Anpassung aller weiteren Operationen an diese Skala.

Der bekannteste auf diesem Ansatz zur Skalierungsschätzung basierende Points-of-Interest-Detektor ist der SIFT-Operator (SIFT: Scale-Invariant Feature Transform) [23, 24]. Dieser kombiniert die Skalenbestimmung mit einer Messung der intrinsischen lokalen

Dimensionalität, ähnlich den autokorrelationsbasierten Ansätzen. Anstelle der Kovarianzmatrix des lokalen Bildgradienten, die die erste partielle Ableitung der Bildfunktion beinhaltet, wird hier allerdings die Hesse-Matrix verwendet, welche aus den zweiten partiellen Ableitungen der Bildfunktion besteht.

Ebenfalls als Blob-basierter Detektor kann der Hessian-Laplace-Detektor [28, 30] angesehen werden, der allerdings eine Weiterentwicklung des autokorrelationsbasierten Harris-Laplace-Detektors ist (s. auch Abschnitt 2.1.1).

Eine Weiterentwicklung des SIFT-Operators im Hinblick auf die Laufzeit stellt der SURF-Operator dar (SURF: Speeded-Up Robust Features) [1]. Alle Teiloperationen des SIFT-Detektors werden auf einfache Operationen auf dem effizent zu berechnenden Integralbild des Eingabebildes abgebildet. Der resultierende Detektor ist nicht äquivalent zum SIFT-Detektor, aber von mindestens vergleichbarer Effektivität [1].

Ein weiterer Ansatz, der allerdings nicht auf dem Laplace-Operator basiert, ist der MSER-Detektor [25] (MSER: Maximally Stable Extremal Regions). Der Ansatz basiert auf einer Segmentierung des Eingabebildes ähnlich der Wasserscheidentransformation [44]. Einzelne Regionen, die besonders stabil sind in der Hinsicht, dass sich deren Form und Größe durch variierende Schwellwerte bei der Segmentierung nicht ändert, werden als Regions-of-Interest selektiert. Schwerpunkte dieser Regionen können dann als Points-of-Interest betrachtet werden.

2.1.3 Informationstheoretische Ansätze

Eine mögliche Herangehensweise an die Detektion von Points-of-Interests ist die Annahme, dass ihre Umgebung einen hohen Informationsgehalt besitzt. Ein informationstheoretisches Maß für den Informationsgehalt ist die *Entropie*. Bestimmt man die Verteilung z.B. der Helligkeitswerte der Bildpunkte in der Umgebung eines gegebenen Punktes, kann die Entropie dieser Verteilung berechnet und somit der Informationsgehalt der Umgebung gemessen werden. In [15] wird die lokale Entropie über verschiedene Skalen gemessen. Maxima der lokalen Entropie über Bildpunktposition und Skala werden als Kandidaten für Points-of-Interest selektiert. Der Ansatz ist dadurch skalierungsinvariant. In der Literatur wird dieser Ansatz auch als Salient-Regions-Detektor bezeichnet. Indem die Variation der Skalierung durch die Variation dreier Ellipsenparameter (Skalierung, Rotation und Verhältnis der beiden Hauptachsen) ersetzt wird, wird der Ansatz auf affine Invarianz erweitert [16].

Ein gänzlich anderer Ansatz, der die Entropie als informationstheoretisches Maß nutzt, wird in [41] vorgestellt. Points-of-Interest werden hier nicht als Punkte definiert, deren Umgebung eine hohe Entropie aufweist. Stattdessen wird mithilfe von *Evolutionären Algorithmen* ein Detektor auf einer Menge von Beispielbildern trainiert. Ein Maß für die Güte des trainierten Detektors ist die Entropie der globalen räumlichen Verteilung der Points-of-Interest im Bild.

2.1.4 Detektoren auf raum-zeitlichen und 3D-Volumendaten

Alle bisher vorgestellten Detektoren erwarten zweidimensionale Bilddaten als Eingabe und extrahieren zweidimensionale Bildpunktkoordinaten als Points-of-Interest. Die zugrundeliegenden Operatoren können jedoch auf mehrere Dimensionen verallgemeinert werden. In [18] wird eine Erweiterung des Harris-Detektors [8] auf drei Dimensionen vorgestellt. Die Eingabe des resultierenden Detektors sind dreidimensionale Bilddaten, wobei die dritte Dimension die Zeit repräsentiert. Als solche raum-zeitlichen Bilddaten können z.B. Videos gesehen werden. Die Ausgabe des raum-zeitlichen Detektors sind demnach raum-zeitliche Points-of-Interest. Da der Detektor auf dem Harris-Detektor basiert, entsprechen diese raum-zeitlichen Ecken. Eine raum-zeitliche Kante kann man sich als Ecke vorstellen, die sich mit konstanter Geschwindigkeit bewegt. Eine raum-zeitliche Ecke ist demnach eine Ecke, die abrupt ihre Bewegung ändert, also z.B. anfängt oder aufhört, sich zu bewegen, oder die Bewegungsrichtung wechselt. Durch diesen auf 3D-Daten erweiterten Harris-Detektor detektierte raum-zeitliche Points-of-Interest kann man demnach als bestimmte Arten von Ereignissen bezeichnen.

Da der Harris-Detektor nicht skalierungsinvariant ist, können nur Ereignisse einer bestimmten zeitlichen Granularität, also Dauer bzw. Geschwindigkeit detektiert werden. Der Detektor wird in [18] daher auf der Basis der in [21] vorgestellten Skalenschätzung erweitert und damit invariant gegenüber raum-zeitlichen Skalierungen gemacht. Anwendungen für den raum-zeitlichen Points-of-Interest-Detektor sind Aktions- und Gestenerkennung sowie die Erkennung von gehenden Personen bei Verdeckungen und in dynamischen Szenen.

Der Ansatz kann ohne weiteres auf 3D-Volumendaten, wie sie z.B. Kernspintomographen liefern, angewendet werden, um dreidimensionale Ecken zu finden. Die Anwendung auf zweidimensionale Abbildungen dreidimensionaler Szenen ist allerdings nicht möglich, da die Daten hier nur in zweidimensionaler Form vorliegen, und die dritte Dimension i.A. nicht aus den 2D-Daten rekonstruiert werden kann.

2.2 Existierende Verfahren zur Bestimmung lokaler Deskriptoren

Aufgabe eines lokalen Deskriptors ist es, die lokale Umgebung eines bereits detektierten Punktes mittels eines Merkmalsvektors fester, im Vergleich zur lokalen Umgebung reduzierter Dimension zu beschreiben. Da durch den Detektor bereits eine große Menge an Punkten und damit eine große Menge möglicher lokaler Grauwertverteilungen herausgefiltert werden, ist eine solche Dimensionsreduktion prinzipiell ohne Informationsverlust möglich. Eine weitere Dimensionsreduktion, die unabhängig vom verwendeten Deskriptor in der Praxis angewendet wird, ist die Erstellung sog. *Codebooks*. Sie wird in einem gesonderten Unterabschnitt im Anschluss an die Deskriptoren behandelt (s. Abschnitt 2.2.1).

Der bekannteste Deskriptor ist der sog. SIFT-Deskriptor, welcher von Lowe zusammen mit dem korrespondierenden SIFT-Detektor in [23, 24] beschrieben wird. Der Deskriptor basiert auf Histogrammen des Bildgradienten in kleinen Unterregionen der lokalen Umge-

bung in verschiedenen Richtungen und Abständen vom detektierten Punkt. Durch Hinzunahme von Informationen des Detektors über die lokale Skalierung und Rotation wird der Deskriptor wie der Detektor ebenfalls skalierungs- und rotationsinvariant. Durch Verwendung des Bildgradienten ist eine gewisse Robustheit gegenüber Beleuchtungs- bzw. Helligkeitsschwankungen gegeben. Der SIFT-Deskriptor liefert einen 128-dimensionalen Merkmalsvektor. Der dem SIFT-Operator nachempfundene SURF-Ansatz beinhaltet parallel zum Detektor ebenfalls einen korrespondierenden laufzeiteffizienten Deskriptor [1].

Ein anderer Ansatz zur Dimensionsreduktion wird von Ke und Sukthankar in [17] vorgestellt. Sie betrachten den Bildgradienten in einer Umgebung von 39×39 Bildpunkten um den detektierten Punkt, normalisiert durch die vom Detektor bestimmte Rotation und Skalierung. Dieser 3042-dimensionale Vektor wird durch eine Hauptachsentransformation (PCA - Principal Component Analysis) auf 36 Dimensionen reduziert. Die dazu nötigen 36 prominentesten Eigenvektoren werden zuvor durch Betrachtung sehr vieler lokaler Umgebungen gelernt. Als prominenteste Eigenvektoren werden hier diejenigen Eigenvektoren bezeichnet, die bei der Hauptachsentransformation der zugrundeliegenden Verteilung die höchsten Eigenwerte besitzen. Ke und Sukthankar nennen ihren Ansatz PCA-SIFT. Der Deskriptor basiert allerdings nicht auf dem SIFT-Deskriptor, sondern nutzt nur die vom SIFT-Detektor gelieferten Informationen.

Eine tatsächliche Weiterentwicklung des SIFT-Deskriptors wird dagegen von Mikolajczyk und Schmid in [31] vorgeschlagen. Sie basiert auf einem etwas komplexeren und damit höherdimensionalen SIFT-Deskriptor, für den abweichende Unterregionen der lokalen Umgebung herangezogen werden. Der resultierende 272-dimensionale Vektor wird anschließend durch eine Hauptachsentransformation auf 128 Dimensionen reduziert. Basis für die Hauptachsentransformation sind die auf der Basis von 47000 272-dimensionalen Beispieldeskriptoren bestimmten 128 prominentesten Eigenvektoren. Mikolajczyk und Schmid nennen ihren Deskriptor *Gradient Location and Orientation Histogram* (GLOH).

Anstelle einer Reduktion der Dimensionalität des Deskriptors wie in den oben genannten Arbeiten wird von Stommel et al. eine Diskretisierung der Deskriptordimensionen vorgeschlagen. Die Autoren zeigen in [39], dass eine Binarisierung des SIFT-Deskriptors, d.h. eine Reduktion von jeweils 256 möglichen Werten auf zwei mögliche Werte pro Dimension, keinen signifikanten Informationsverlust bedeutet.

Eine weitere Adaption des SIFT-Deskriptors zur Verwendung mit dem Blob-basierten Points-of-Interest-Detektor *Maximally Stable Extremal Regions* [25] stellen Forssen und Lowe in [5] vor. Er basiert auf einer Binarisierung der Umgebung des vom Detektor gelieferten Punktes und einer anschließenden Bestimmung des SIFT-Deskriptors auf dieser binarisierten Umgebung. Durch die Binarisierung wird eine höhere Robustheit gegenüber Helligkeitsänderungen erreicht.

Varma und Ray stellen in [43] einen gänzlich anderen Ansatz vor, der auf Kernel-basiertem Lernen mittels Support Vector Machines basiert. Der Deskriptor wird jeweils für eine konkrete Anwendung optimiert bzw. gelernt und benötigt eine auf die jeweilige Anwendung angepasste Trainingsmenge.

Ein weiterer, stark vom Konzept des SIFT-Deskriptors abweichender Ansatz wird von Ling und Jacobs in [22] vorgeschlagen, genannt *Geodesic-intensity Histogram* (GIH). Er basiert auf den sog. *Geodäten* des Bildes als *riemannsche Mannigfaltigkeit* über die Bildin-

tensitäten, d.h. den ausschließlich intensitätsbasierten Abständen benachbarter Punkte. Ausgehend von einem gegebenen Punkt werden Intensitäten aller Punkte mit demselben geodätischen Abstand zu diesem Punkt bestimmt. Dies wird für verschiedene Abstände getan, wodurch sich eine zweidimensionale Beschreibung der Umgebung des Punktes ergibt. Der Deskriptor ist invariant gegenüber Deformationen, deren Teilmenge u.a. die affinen Transformationen sind. Zudem ist keine Beschränkung auf ein festes Fenster um einen detektierten Punkt nötig, da die Größe der Umgebung vom Deskriptor selber automatisch bestimmt wird. Ein entsprechender deformationsinvarianter Detektor existiert allerdings noch nicht.

2.2.1 Deskriptor-Codebooks

Da lokale Deskriptoren Merkmalsvektoren fester Dimension darstellen, ist es möglich, den Deskriptorraum z.B. durch Clustering in eine begrenzte Anzahl diskreter Klassen zu zerteilen. Ähnliche Deskriptoren werden dabei derselben Klasse zugeordnet, es ergibt sich ein sog. *visuelles Vokabular*. Diese Analogie mit textuellen Daten ergibt sich durch die Verwendung in der Praxis. Ein Vorteil eines solchen Codebooks ist, dass nun statt Ähnlichkeiten zwischen Deskriptoren auf der Basis eines quantitativen Ähnlichkeitsmaßes wie z.B. dem euklidschen Abstand im Deskriptorraum eine Äquivalenz von Deskriptoren durch die Zugehörigkeit zu derselben Klasse im Codebook festgelegt werden kann. Jeder Deskriptor kann eindeutig einem *visuellen Wort* zugeordnet werden.

Sivic and Zisserman [38] nutzen ein solches Codebook, um innerhalb eines Videos effizient ähnliche Szenen finden zu können. Auf der Basis des SIFT-Deskriptors, berechnet auf vom SIFT-Deskriptor detektierten Punkten [23] und den Maximally Stable Extremal Regions aus [25], wird ein Teil des gegebenen Videos herangezogen, um das Codebook bzw. visuelle Vokabular zu erstellen. Der Film wird dann nach dem Prinzip der ursprünglich zur Textindizierung verwendeten sog. *Inverted Files* [45] indiziert: Jeder Frame wird als ein Vektor der in ihm vorkommenden Codebook-Einträge repräsentiert. Jedem Codebook-Eintrag werden alle Frames zugeordnet, die ihn enthalten. Durch die Mengenoperationen Schnitt und Vereinigung können nun effizient Anfragen auf der Basis einzelner Frames und sogar Teilen von Frames durchgeführt werden. Durch Berücksichtigung der räumlichen Lage der Codebook-Einträge zueinander im Ursprungs-Frame kann das Ergebnis weiter gefiltert werden.

Leibe und Schiele dagegen nutzen in [20] ein ähnliches Codebook zur Objekterkennung und Lokalisation. Mittels des Harris-Laplace-Detektors werden auf Trainingsbildern Punkte detektiert und deren umgebende Grauwerte werden direkt in das Codebook geclustert (der Deskriptor entspricht in diesem Fall einfach der normalisierten Grauwertumgebung eines detektierten Punktes). Zusätzlich wird jeweils die Position der Punkte in Relation zum zu erkennenden Objekt vermerkt. Durch eine verallgemeinerte Hough-Transformation können nur die entsprechenden Objekte in unbekannten Bildern detektiert werden. Jeder detektierte Punkt in einem unbekannten Bild wird einem Codebook-Eintrag zugeordnet. Anhand der gelernten erwarteten Position dieses Codebook-Eintrags zum Objekt „stimmt“ jeder Punkt für eine entsprechende Position im Bild. An und um die Position, an der sich ein Objekt befindet, werden viele solcher Stimmen erwartet. Es

ergibt sich bei Detektion eines Objekts dadurch automatisch auch seine Lokalisierung.

Ilies et al. nutzen in [11] ebenfalls ein visuelles Vokabular, welches aus einem Clustering von SIFT-Deskriptoren erstellt wird. Mittels eines Bag-of-(Visual-)Words-Ansatzes wird ein Klassifikator trainiert, der von Nachrichtenseiten stammende Bilder von 50 verschiedenen Personen des öffentlichen Lebens unterscheiden kann. Als Trainingsmenge für den Klassifikator werden die Ergebnisse eines Named-Entity-Recognizers auf den Bildunterschriften herangezogen.

Anstelle eines Clusterings wird von Stommel et al. in [40] eine schnellere Alternative zur Erstellung eines Codebooks vorgeschlagen. Durch eine Menge von zufälligen Deskriptoren als Cluster-Zentren wird das zeitaufwendige Clustering vermieden. Es wird gezeigt, dass die zufällige Generierung der Cluster keine negativen Auswirkungen auf eine nachfolgende Klassifizierung hat.

2.3 Vergleichende Evaluationen

Schmid et al. stellen in ihrer Studie [35] die Detektoren aus [2, 6, 8–10] gegenüber und vergleichen diese anhand ihrer sog. *Repeatability Rate*, d.h. der Stabilität bzgl. der detektierten Position unter verschiedenen Bildtransformationen. Die betrachteten Transformationen beinhalten Rotation, Skalierung, Beleuchtungsänderungen und Änderung des Blickwinkels. Die betrachteten Bilder zeigen planare Szenen. Zunächst wird die Originalversion des Harris-Detektors mit einer leicht angepassten Variante verglichen, welche isotropische Filter zur Berechnung der Bildgradienten verwendet, um die Rotationsinvarianz zu verbessern. Der angepasste Harris-Detektor schneidet bei dieser Gegenüberstellung erwartungsgemäß besser ab. In der anschließenden Evaluation wird nur noch die angepasste Version mit den anderen betrachteten Detektoren verglichen.

Obwohl der Harris-Detektor inhärent nur invariant gegenüber Rotation ist, weist er gegenüber den anderen Detektoren [2, 6, 9, 10] in allen vier Evaluationsteilen eine höhere Stabilität auf.

Ein Vergleich aktuellerer Detektoren wird in [29] vorgenommen. Verglichen werden der Hessian-Laplace-Detektor [28], der MSER-Detektor [25], der Salient-Regions-Detektor [16] sowie ein weiterer affin invarianter Detektor [42]. Als Evaluationskriterium wird ebenfalls die Stabilität gewählt. Die betrachteten Bilder zeigen planare Szenen. Der MSER-Detektor schneidet bei diesem Vergleich am besten ab.

Ein weiterer Vergleich aktuellerer Detektoren wird in [7] vorgenommen. Zusätzlich zu der Menge der in [29] verglichenen Detektoren werden der SIFT-Detektor [23] sowie der angepasste Harris-Detektor [8, 35] in die Evaluation aufgenommen. Das Evaluationskriterium ist wieder die Stabilität. Die praktische Relevanz der bisher zur Evaluation verwendeten Bilder planarer Szenen wird allerdings in dieser Evaluation hinterfragt. Daher werden hier Bilder komplexerer dreidimensionaler Szenen mit weniger planaren Flächen herangezogen. Die Evaluation zeigt, dass ein starker Unterschied in den Ergebnissen besteht, verglichen mit der Evaluation auf planaren Szenen. Der MSER-Detektor [25] zeigt auch hier die besten Ergebnisse, allerdings schneidet der Harris-Detektor [8] besser ab als komplexere affin invariante Detektoren. Dieses Ergebnis ist überraschend, da der

Harris-Detektor einfacher aufgebaut ist als die affin invarianten Detektoren und prinzipiell weniger mögliche Transformationen berücksichtigt.

In einer späteren Evaluation [33] wird diese Erkenntnis als Grundlage genommen, einen großen Datensatz von Bildern komplexer dreidimensionaler Objekte anzulegen und für einen weiteren Vergleich von Points-of-Interest-Detektoren zu nutzen. Das Hauptaugenmerk liegt auf dem Bilddatensatz, welcher aus 100 Objekten besteht, die von zwei Kameras jeweils aus 72 Blickwinkeln unter 3 verschiedenen Beleuchtungsbedingungen aufgenommen wurden. Pro Objekt liegen demnach 432 Bilder vor. Es wird außerdem ein Versuchsaufbau beschrieben, wie diese große Bildmenge ohne die manuelle Erstellung einer Ground Truth zur Messung der Stabilität von Detektoren genutzt werden kann. Im Vergleich der bereits in [7] betrachteten Detektoren schneidet hier der SIFT-Detektor am besten ab. Das wichtigste Ergebnis der Evaluation ist allerdings, dass keiner der aktuellen Detektoren unter einer Blickwinkelvariation von über 25 bis 30 Grad gute Ergebnisse erzielt. Die Schlussfolgerung der Autoren ist, dass bisherige Detektoren vor allem auf planare Szenen optimiert sind. Für Projektionen 3D-rigider Transformationen oder anderer über 2D-affine Transformationen hinausgehende Deformationen müssten neue Detektoren entwickelt werden [33, S. 7].

3 Deformationsinvariante Bildbeschreibung mit Graphen

Dieses Kapitel beschreibt den eigenen Ansatz eines deformationsinvarianten Points-of-Interest-Detektors und des dazugehörigen lokalen Deskriptors. Zunächst wird gezeigt, dass die Deformationsinvarianz die höchste Form der Invarianz darstellt, die man auf zweidimensionalen Bildern von dreidimensionalen Szenen erreichen kann (s. Abschnitt 3.1). Im darauffolgenden Abschnitt 3.2 wird ein deformationsinvariantes Abstandsmaß eingeführt.

Basierend auf diesem Abstandsmaß wird in Abschnitt 3.3 eine Beschreibung eines Bildes als Graph definiert. Da der Aufwand von Algorithmen auf Graphen vor allem von der Anzahl der Knoten eines Graphen abhängt, ist es aus praktischen Gründen bei größeren Bildern notwendig, den entstehenden Graphen vor der Weiterverarbeitung zu reduzieren. Es wird daher in einem entsprechenden Unterabschnitt 3.3.3 die Approximation von Graphen behandelt.

Grundlage des in Abschnitt 3.5 eingeführten deformationsinvarianten Detektors ist die graphbasierte Bildbeschreibung und insbesondere das zuvor eingeführte Abstandsmaß. Ein gegebenes Bild wird reduziert auf die Abstände seiner Punkte unter diesem Abstandsmaß. Durch diese Reduktion wird eine Bildbeschreibung erzeugt, die wie das zugrundeliegende Abstandsmaß invariant gegenüber Deformationen ist. Alle weiteren aus dieser reduzierten Beschreibung extrahierten Informationen sind dann ebenfalls deformationsinvariant. Der Detektor nutzt diese Beschreibung, um die Points-of-Interest zu bestimmen und ist so automatisch invariant gegenüber Deformationen.

Die Evaluation des Detektors wird im darauf folgenden Kapitel 4 vorgenommen.

3.1 Äquivalenz der Projektion von 3D-Transformationen und 2D-Deformationen

Anstelle der ursprünglich angestrebten Invarianz gegenüber 3D-rigiden Transformationen der betrachteten Szene wird zunächst eine Invarianz gegenüber zweidimensionalen Deformationen des Graubildes angestrebt. In der Tat sind beide Transformationsklassen äquivalent, sofern der Grauwert eines Punktes in der Szene unter der Transformation konstant bleibt, und sofern keine Verdeckungen vorkommen. Ein Detektor, der invariant gegenüber Deformationen ist, kommt dem angestrebten Ziel also sehr nahe. Diese Behauptung soll im folgenden bewiesen werden. Zunächst wird die Behauptung umgangssprachlich formuliert:

Die Projektion einer 3D-rigiden Transformation ist eine 2D-Deformation, wenn keine Verdeckungen vorkommen.

Grundlage der Betrachtung sind eine dreidimensionale Szene und deren zweidimensionale Projektionen bzw. Bilder der Szene. Da hierfür nur der sichtbare Teil der Szene von Belang ist, wird eine Szene auf ihre sichtbaren Oberflächen reduziert. Transparenz und z.B. Nebel werden nicht berücksichtig. Es wird das Lambertsche Beleuchtungsmodell angenommen.

Jeder Punkt der betrachteten Szene, der auf ihrer sichtbaren Oberfläche liegt, hat eine Position im dreidimensionalen Raum sowie eine Farbe. Ein sichtbarer Punkt der Szene entspricht dann einem dreidimensionalen Punkt (x, y, z). Sichtbar heißt in diesem Fall sichtbar aus einer beliebigen Perspektive. Da die Oberflächen einer Szene intrinsisch zweidimensional sind, aber im dreidimensionalen Raum eingebettet, eignet sich das Konzept der *Mannigfaltigkeit* zur Definition einer Szene:

Definition 1. Eine *Szene* $S \subset \mathbb{R}3$ sei eine topologische Mannigfaltigkeit der Dimension 2 (eine 2-*Mannigfaltigkeit*). Ein Punkt (x, y, z) im dreidimensionalen Raum ist genau dann sichtbarer Punkt der Szene, wenn gilt $(x, y, z) \in S$.

Korollar 1. Alle topologischen Mannigfaltigkeiten sind per Definition topologische Räume [19, S. 33].

Zunächst wird die Projektion eines einzelnen Punktes einer Szene definiert:

Definition 2. Sei $\mathbf{P} \in \mathbb{R}^{3\times4}$ eine Projektionsmatrix. Die Projektion eines Punktes einer Szene $p = (x, y, z) \in S$ unter $\mathbf{P}$ ist gegeben durch die Abbildung $P : S \times \mathbb{R}^{3\times4} \longrightarrow \mathbb{R}^3$:

$$\begin{aligned} P(p, \mathbf{P}) &= \left(\frac{x'}{w'}, \frac{y'}{w'}, w'\right) \text{ mit} \\ \begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} &= \mathbf{P} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \end{aligned}$$

Dabei enthält w' die Entfernung von der Projektionsebene.

Die Projektion einer Szene wird nun definiert als die Abbildung einer Teilmenge des $\mathbb{R}^2$ (der Projektionsebene) auf die Szene. Dabei wird w' verwendet, um für die Projektion einer Szene den Punkt der Szene zu wählen, der näher an der Projektionsebene liegt, falls mehr als ein Punkt der Szene auf denselben Punkt in der Projektionsebene projiziert wird:

Definition 3. Sei $\mathbf{P} \in \mathbb{R}^{3\times4}$ eine Projektionsmatrix. Sei $S_\mathbf{P}$ die Menge aller Punkte des $\mathbb{R}^2$, die Projektionen von Punkten der Szene S sind:

$$S_\mathbf{P} = \left\{ \left(\frac{x'}{w'}, \frac{y'}{w'}\right) \,\middle|\, \exists p : \left(\frac{x'}{w'}, \frac{y'}{w'}, w'\right) = P(p, \mathbf{P}) \right\}$$

Dann ist die Projektion der Szene unter $\mathbf{P}$ gegeben durch die Abbildung $P_{\mathbf{P}} : S_{\mathbf{P}} \longrightarrow S$:

$$P_{\mathbf{P}}(x,y) \quad = \quad \min_{w} \left\{ (x,y) = \left(\frac{x'}{w'}, \frac{y'}{w'} \right), (x',y',w') = P(p,\mathbf{P}) \middle| p \in S \right\}$$

Korollar 2. Eine aus S durch $\mathbf{P}$ projezierte Punktmenge $S_{\mathbf{P}}$ ist ebenfalls ein topologischer Raum und die Abbildung P ist stetig.

Es kann nun definiert werden, wann die Projektion einer Szene Verdeckungen enthält:

Definition 4. Sei $P_{\mathbf{P}} : S_{\mathbf{P}} \longrightarrow S$ die Projektion einer Szene S. Die Projektion enthält Verdeckungen genau dann, wenn $P_{\mathbf{P}}$ bijektiv ist, d.h. wenn es keine zwei Punkte $p, q \in S$ gibt, die auf denselben Punkt $(x,y) \in S \subseteq \mathbb{R}^2$ projiziert werden. Eine Projektion heißt verdeckungsfrei, wenn sie keine Verdeckungen enthält.

Korollar 3. Wenn $P_{\mathbf{P}}$ verdeckungsfrei, d.h. bijektiv und damit umkehrbar ist, dann ist P ebenfalls umkehrbar.

Indem der Begriff des *Homöomorphismus* zur Beschreibung von Deformationen herangezogen wird, kann die oben umgangssprachlich formulierte Behauptung nun als Satz formalisiert werden:

Satz 1. *Sei S eine* Szene. *$P_{\mathbf{P}} : S_{\mathbf{P}} \longrightarrow S$ und $P_{\mathbf{P}'} : S_{\mathbf{P}'} \longrightarrow S$ seien Projektionen der Szene unter $\mathbf{P}$ bzw. $\mathbf{P}$. Dann existiert ein Homöomorphismus h zwischen $S_{\mathbf{P}}$ und $S_{\mathbf{P}'}$, wenn weder $P_{\mathbf{P}}$ noch $P_{\mathbf{P}'}$ Verdeckungen enthalten:*

$$P_{\mathbf{P}}, P_{\mathbf{P}'} \text{ sind verdeckungsfrei} \implies \exists h : S_{\mathbf{P}} \to S_{\mathbf{P}'} \text{ homöomorph} \tag{3.1}$$

Beweis 1. *Nach der Definition des Homöomorphismus-Begriffs gilt: h ist ein Homöomorphismus genau dann, wenn h bijektiv und stetig ist, und wenn die Umkehrabbildung h^{-1} ebenfalls stetig ist.*

Sei $P : S \times \mathbb{R}^{3\times4} \longrightarrow \mathbb{R}^3$ die Projektion der Szene S unter der Projektionsmatrix $\mathbf{P}$ und P^{-1} ihre Umkehrung. Außerdem sei $P' : S \times \mathbb{R}^{3\times4} \longrightarrow \mathbb{R}^3$ die Projektion der Szene S unter der Projektionsmatrix $\mathbf{P}'$. Es wird nun eine Abbildung $g : S_{\mathbf{P}} \to S_{\mathbf{P}'}$ wie folgt definiert:

$$g(x,y) = P_{\mathbf{P}'}^{-1} \left(P' \left(P^{-1} \left(P_{\mathbf{P}}(x,y) \right) \right) \right)$$

Da $P_{\mathbf{P}'}$ nach Voraussetzung verdeckungsfrei und damit bijektiv ist, existiert die Umkehrabbildung $P_{\mathbf{P}'}^{-1}$. Die Umkehrabbildung g^{-1} existiert und ist wie folgt definiert:

$$g^{-1}(x,y) = P_{\mathbf{P}}^{-1} \left(P \left(P'^{-1} \left(P_{\mathbf{P}'}(x,y) \right) \right) \right)$$

Demnach ist g bijektiv. Außerdem ist g stetig, da sowohl P und P' als auch $P_{\mathbf{P}}$ und $P_{\mathbf{P}'}$ und deren Umkehrungen stetig sind, und eine Komposition stetiger Abbildungen ebenfalls stetig ist [13, S. 17]. Die definierte Abbildung g entspricht demnach dem gesuchten Homöomorphismus h. □

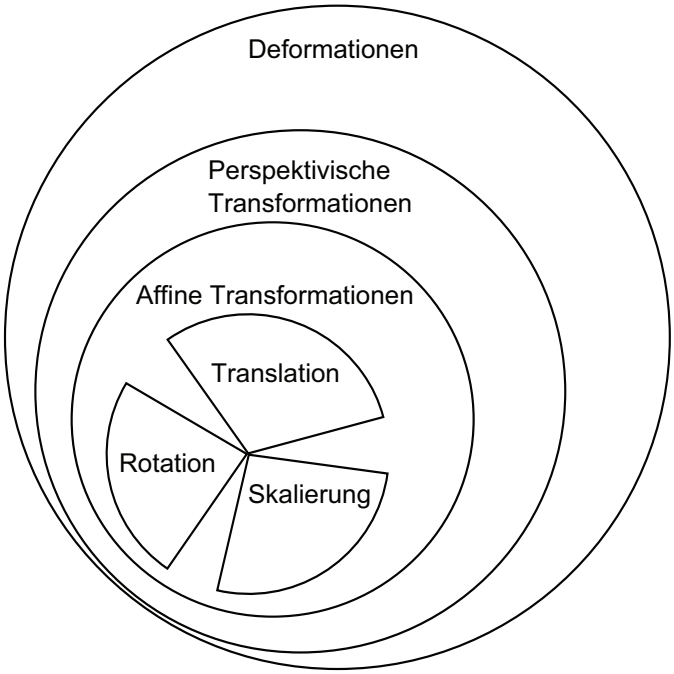


Abbildung 3.1: Mengentheoretische Beziehungen zwischen Klassen zweidimensionaler Transformationen

Mit dem Beweis ist nicht gezeigt, dass es keine Projektionen von 3D-rigiden Transformationen gibt, die Deformationen sind. D.h., es mag durchaus 3D-rigide Transformationen geben, deren Projektionen ebenfalls durch Deformationen beschrieben werden können. Für diese Arbeit ist es allerdings nur von Interesse, zu zeigen, dass die Projektion jeder 3D-rigiden Transformation, die keine Verdeckungen enthält, durch eine zweidimensionale Deformation beschrieben werden kann.

Die Menge der zweidimensionalen Deformationen beinhaltet außerdem alle (nicht singulären) perspektivischen Transformationen [13] und damit auch die in existierenden Verfahren betrachteten affinen Transformationen. Abbildung 3.1 illustriert die mengentheoretischen Beziehungen zwischen verschiedenen zweidimensionalen Transformationen.

3.2 Ein deformationsinvariantes Abstandsmaß

Zunächst werden die Begriffe *Pfad, Pfadlänge, monotoner Pfad, Abstand, Pfadextremum* und *Sattelpunkt* eingeführt, wie sie in dieser Arbeit verwendet werden. Es folgen einige Sätze und entsprechende Beweise, die zu einer Repräsentation des Bildes als Graph von *Sattelpunkten* als Knoten und *Pfaden* als Kanten führen, die invariant gegenüber Deformationen ist.

Definition 5. Sei $d(p,q)$ die Manhattan Distance zwischen p und q. Dann ist ein *Pfad* Q zwischen zwei Punkten p_1 und p_2 eines Bildes, auch bezeichnet mit $Q(p_1,p_2)$, eine Folge benachbarter Punkte $q_1,\ldots,q_n$ mit $q_1 = p_1$ und $q_n = p_2$ und $d(q_i,q_{i+1}) = 1$ für alle $1 \leq i < n$.

Definition 6. Sei $i(p)$ die Intensität des Bildes am Punkt p. Dann ist die *Länge* $\|Q\|$ eines Pfads $Q = q_1,\ldots,q_n$ gegeben als die Summe der Intensitätsdifferenzen der benachbarten Punkte:

$$\|Q\| = \sum_{1\leq i<n} |i(q_i) - i(q_{i+1})|$$

Definition 7. Ein Pfad $Q = q_1,\ldots,q_n$ heißt *monoton*, wenn die Folge der Intensitäten seiner Punkte monoton ist, d.h. wenn $i(q_1),\ldots,i(q_n)$ monoton ist.

Korollar 4. Die Länge eines monotonen Pfades entspricht immer der Intensitätsdifferenz der beiden Endpunkte des Pfades. Diese Bedingung ist sogar ausreichend, was eine alternative Formulierung unserer *Monotonie* erlaubt:

Ein Pfad $Q = q_1,\ldots,q_n$ ist genau dann *monoton*, wenn gilt:

$$\|Q\| = |i(q_1) - i(q_n)|$$

Definition 8. Der *Abstand* $\|p_1,p_2\|$ zwischen zwei Punkten p_1 und p_2 sei definiert als die Länge des kürzesten Pfades zwischen p_1 und p_2:

$$\|p_1,p_2\| = \min_Q \|Q(p_1,p_2)\|$$

Definition 9. Sei p ein Punkt auf einem Pfad $Q = q_1,\ldots,q_n$ mit $i(p) \geq i(q_j)$ für alle $1 \leq j \leq n$ und $i(p) > i(q_1)$ und $i(p) > i(q_n)$. Dann ist p ein *lokales Pfadmaximum* von Q. Analog gilt p als *lokales Pfadminimum* von Q, wenn $i(p) \leq i(q_j)$ für alle $1 \leq j \leq n$ und $i(p) < i(q_1)$ und $i(p) < i(q_n)$. Es sei p ein *lokales Pfadextremum* von Q, wenn p entweder lokales Pfadmaximum oder lokales Pfadminimum von Q ist. Die Menge der Pfadextrema eines Pfades Q wird bezeichnet mit $E(Q)$.

Korollar 5. Pfade sind genau dann monoton, wenn sie keine lokalen Pfadextrema besitzen.

Definition 10. Ein Punkt s ist ein *Sattelpunkt*, wenn zwei Punkte p_1 und p_2 existieren, so dass s ein lokales Pfadextremum des kürzesten Pfades zwischen p_1 und p_2 ist:

$$\|Q(p_1, p_2)\| = \|p_1, p_2\| \wedge s \in E(Q) \Rightarrow s \,\text{ist Sattelpunkt}$$

Satz 2. *Sei Q der kürzeste Pfad zwischen zwei Punkten p_1 und p_2 und seien $s_i, 1 \leq i \leq n$ alle lokalen Pfadextrema von Q, also Sattelpunkte, geordnet nach ihrer Reihenfolge im Pfad.*

Dann ist der Teilpfad $Q_0 = p_1, \ldots, s_1 \subset Q$ kürzester Pfad zwischen p_1 und s_1, $Q_i = s_i, \ldots, s_{i+1} \subset Q$ sind kürzeste Pfade zwischen s_i und s_{i+1} für $1 \leq i < n$, und $Q_n = s_n, \ldots, p_2 \subset Q$ ist kürzester Pfad zwischen s_n und p_2. Anmerkung: Hieraus folgt alle Teilpfade $Q_j, 0 \leq j \leq n$ sind monoton.

Beweis 2. *Beweis durch Widerspruch: Sei Q_i einer der oben genannten Teilpfade von Q und nicht der kürzeste Pfad zwischen seinen Endpunkten. Daraus folgt es gibt einen kürzeren Pfad Q_i' zwischen seinen Endpunkten. Wir ersetzen Q_i durch Q_i' und erhalten so einen modifizierten Pfad $Q' = \bigcup Q_0, \ldots, Q_{i-1}, Q_i', \ldots, Q_n$. Dieser Pfad ist ein Pfad zwischen p_1 und p_2 und ist kürzer als Q, woraus folgt Q ist nicht der kürzeste Pfad zwischen p_1 und $p_2 \rightarrow$ Widerspruch.* □

Satz 3. *Zu jedem Punkt p gibt es einen Sattelpunkt s, so dass der kürzeste Pfad zwischen p und s monoton ist, d.h. $\|p, s\| = |i(p) - i(s)|$.*

Beweis 3. *Sei s ein beliebiger Sattelpunkt und sei Q der kürzeste Pfad zwischen s und p. Dann sind zwei Fälle zu unterscheiden:*

1. *Q ist monoton und s ist der gesuchte Sattelpunkt.*

2. *Q ist nicht monoton. Daraus folgt nach Definition 9, dass Q lokale Pfadextrema besitzt. Nach Definition 10 sind diese Pfadextrema Sattelpunkte, da Q der kürzeste Pfad zwischen zwei Punkten ist.*

   Nach Satz 2 gilt nun: Der Pfad Q kann aufgeteilt werden in $n+1$ monotone Teilpfade $Q_i, 0 \leq i \leq n$, wobei n die Anzahl der lokalen Pfadextrema (und damit Sattelpunkten) von Q ist. Diese Teilpfade sind die kürzesten Pfade zwischen ihren Endpunkten. Der letzte Teilpfad Q_n hat die Endpunkte s_n und p, wobei s_n ein Sattelpunkt ist. Q_n ist monoton und s_n ist der gesuchte Sattelpunkt.

□

Korollar 6. Aus Satz 2 und 3 folgt, dass die Suche des kürzesten Pfades zwischen zwei beliebigen Punkten eines Bildes zurückführbar ist auf die Suche des kürzesten Weges in einem Graphen, welcher als Knoten die Sattelpunkte des Bildes und die beiden betrachteten Punkte sowie als Kanten die kürzesten monotonen Pfade zwischen den Sattelpunkten sowie den Sattelpunkten und den beiden betrachteten Punkten enthält.

Der Graph der Sattelpunkte und der monotonen kürzesten Pfade zwischen ihnen beinhaltet also *fast* die gesamte Information hinsichtlich unseres in Definition 8 formulierten

Abstandsmaßes. Insbesondere ist der Graph, sofern das Abstandsmaß invariant ist gegenüber Deformationen, diesen gegenüber ebenfalls invariant. Nach Hinzunahme lokaler Extremwerte als weitere Knoten und dem jeweils kürzesten Pfad zum nächsten Sattelpunkt als weitere Kanten enthält der Graph alle Informationen bzgl. des definierten Abstandsmaßes. Die Deformationsinvarianz geht dadurch nicht verloren.

Im folgenden Abschnitt wird das Abstandsmaß für den kontinuierlichen Fall definiert und die Invarianz gegenüber Deformationen bewiesen. Ein Bild wird dabei als riemannsche Mannigfaltigkeit gesehen. Zweidimensionale Bilddeformationen können dann als Homöomorphismen, also als Transformationen, die die Topologie der Mannigfaltigkeit erhalten, betrachtet werden.

3.2.1 Invarianz gegenüber Deformationen

Für den kontinuierlichen Fall kann gezeigt werden, dass das intensitätsbasierte Abstandsmaß invariant gegenüber Deformationen ist. Ein Bild mit den Bildkoordinaten (x, y) und den Grauwerten $i(x, y)$ kann als zweidimensionale riemannsche Mannigfaltigkeit M betrachtet werden, mit $((1-\alpha)x, (1-\alpha)y, \alpha i(x, y)) \in M$. Der Skalierungsparameter α dient zur Stauchung der Mannigfaltigkeit, um später ein Abstandsmaß nur basierend auf dem Grauwert für $\alpha \to 1$ definieren zu können.

Auf dieser Mannigfaltigkeit sei $f : [0, 1] \to M$ mit $f(0) = a, f(1) = b$ und $a, b \in M$ ein Weg, der die Punkte a und b verbindet.

Nach der riemannschen Metrik kann die Länge L des Weges f bestimmt werden durch

$$L = \int_0^1 \sqrt{f'(t) \cdot f'(t)} dt$$

Dabei ist f' die Ableitung von f. Definiert man

$$f(t) = ((1-\alpha)x(t), (1-\alpha)y(t), \alpha i(x(t), y(t))$$

und somit

$$f'(t) = ((1-\alpha)\frac{\partial x(t)}{\partial t}, (1-\alpha)\frac{\partial y(t)}{\partial t}, \alpha \frac{\partial i(x(t), y(t))}{\partial t})$$

erhält man dadurch

$$L = \int_0^1 \sqrt{(1-\alpha)^2 \left(\frac{\partial x(t)}{\partial t}\right)^2 + (1-\alpha)^2 \left(\frac{\partial y(t)}{\partial t}\right)^2 + \alpha^2 \left(\frac{\partial i(x(t), y(t))}{\partial t}\right)^2} dt$$

Sei $d : M \to \bar{M}$ ein *Homöomorphismus*, der die Mannigfaltigkeit M wie folgt deformiert:

$$\begin{aligned} x &\mapsto d(x) \\ y &\mapsto d(y) \\ i(x, y) &\mapsto i(x, y) \end{aligned}$$

Die Deformation d erhält somit den Grauwert eines Bildpunktes. Da d als Homöomorphismus stetig ist, ergibt sich die Länge $\bar{L}$ des deformierten Weges $d(f)$ durch

$$\bar{L} = \int_0^1 \sqrt{(1-\alpha)^2 \left(\frac{\partial d(x(t))}{\partial t}\right)^2 + (1-\alpha)^2 \left(\frac{\partial d(y(t))}{\partial t}\right)^2 + \alpha^2 \left(\frac{\partial d(i(x(t),y(t)))}{\partial t}\right)^2} dt$$

Ignoriert man bei der Längenberechnung die Bildkoordinaten x und y durch $\alpha \to 1$ und berechnet so die Länge allein basierend auf dem Grauwert $i(x,y)$, so werden die entsprechend modifzierten L und $\bar{L}$ gleich, da die ersten beiden Terme im Integral durch $\alpha \to 1$ verschwinden, und $\alpha^2 d(i(x(t),y(t))) = \alpha^2 i(x(t),y(t))$.

Das Abstandsmaß ist daher invariant gegenüber homöomorphen Deformationen, die den Grauwert erhalten, darunter insbesondere für alle affinen Transformationen des Bildes.

3.3 Repräsentation eines Bildes als Graph

Betrachtet man ein Bild als einen ungerichteten Graphen mit den Bildpunkten als Knoten und Kanten zwischen benachbarten Bildpunkten, deren Gewichte dem intensitätsbasierten Abstandsmaß entsprechen, kann man einige graphentheoretische Formulierungen und Ansätze auf das vorliegende Problem anwenden. Im Folgenden wird zunächst die Beschreibung eines gesamten Bildes als Graph eingeführt. Im darauffolgenden Abschnitt werden verschiedene Algorithmen diskutiert, die einen existierenden Graphen approximieren, um die Anzahl der Knoten zu reduzieren. Dies ist wegen der algorithmischen Komplexität u.U. nötig, um eine effiziente Bearbeitung größerer Bilder mit vielen Bildpunkten (und dadurch auch vielen Knoten im korrespondierenden Graphen) zu ermöglichen.

3.3.1 Exakte Repräsentation

In diesem Abschnitt wird die Konvertierung eines (aus Pixeln bestehenden) Rasterbildes in einen Graphen beschrieben, der aus Knoten und gewichteten Kanten besteht. Dies wird anhand eines einfachen Beispielbildes der Größe 8×8 illustriert. Abbildung 3.2 zeigt dieses Beispielbild in unterschiedlichen Repräsentationen. Abbildung 3.2 (a) zeigt das vergrößerte Graubild, (b) eine dreidimensionale Darstellung unter Verwendung der Intensität als z-Koordinate, und (c) die 8×8-Matrix der Intensitäten.

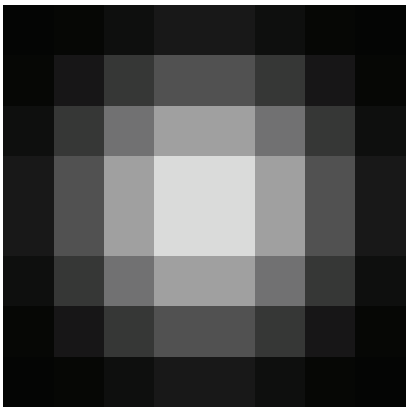
Zunächst wird ein Bild $I = (P, i)$ definiert als ein Paar, bestehend aus einer Menge von Pixeln $P \subset \mathbb{R} \times \mathbb{R}$ und einer Abbildung von Pixeln auf ihre Intensität $i : P \longrightarrow \mathbb{R}$. Obwohl Pixelkoordinaten typischerweise nur natürlichzahlige Werte annehmen, werden die Pixelkoordinaten und auch die Intensitäten reellwertig dargestellt, um u.U. auch durch Interpolation erhaltene zusätzliche Pixel zuzulassen. Dies wird weiter unten in diesem Abschnitt relevant.

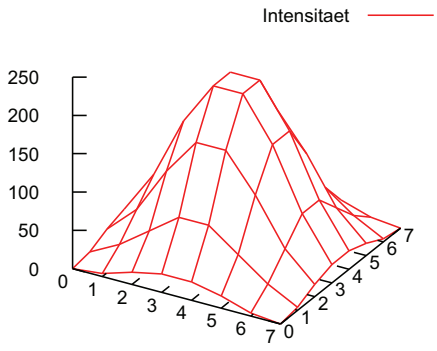
Ein Graph $G = (V, E, w)$ ist definiert als eine Menge von Knoten V, eine Menge von Kanten $E \subseteq V \times V$ und eine Gewichtungsfunktion $w : E \longrightarrow \mathbb{R}$, die jeder Kante ein reellwertiges Gewicht zuweist.

Um ein Bild in einen Graphen zu konvertieren, werden zusätzlich eine Nachbarschaftsrelation $N \subseteq P \times P$ und eine Distanzfunktion zwischen zwei Pixeln definiert. Der Graph $G = (V, E, w)$ eines Bildes $I = (P, i)$ ergibt sich mit N und d wie folgt:

$$
\begin{aligned}
V &= P && (3.2)\\
E &= N \\
w((x,y),(x',y')) &= \sqrt{(1-\alpha)^2(x-x')^2 + (1-\alpha)^2(y-y')^2 + \alpha^2(i(x,y)-i(x',y'))^2} \\
\text{mit} \quad & ((x,y),(x',y')) \in E, (x,y),(x',y') \in V \\
\text{und} \quad & (x,y),(x',y') \in P && (3.3)
\end{aligned}
$$

Nach [22] definieren wir dabei die Distanzfunktion zwischen zwei Pixeln und damit das Kantengewicht w als den euklidschen Abstand der Kombination aus Koordinaten $(x, y) \in V$ bzw. $(x', y') \in V$ und Intensität $i(x, y)$ bzw. $i(x', y')$, gewichtet mit einem Parameter α (s. Abschnitt 3.2.1).



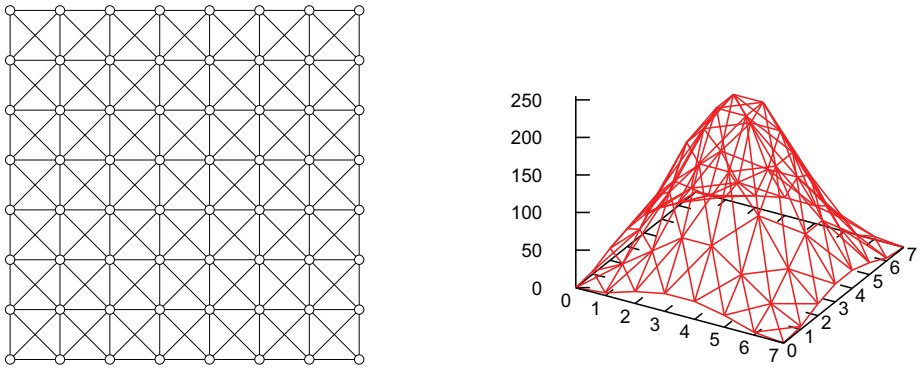


(a) (b)

0	4	16	24	24	16	4	0
4	24	52	80	80	52	24	4
16	52	112	160	160	112	52	16
24	80	160	216	216	160	80	24
24	80	160	216	216	160	80	24
16	52	112	160	160	112	52	16
4	24	52	80	80	52	24	4
0	4	16	24	24	16	4	0

(c)

Abbildung 3.2: Beispielbild, bestehend aus einem hellen Fleck auf dunklem Hintergrund: (a) Originalbild, (b) 3D-Darstellung unter Verwendung der Intensität als z-Koordinate, (c) Matrix der Intensitäten



(a) (b)

Abbildung 3.3: Graph des Beispielbildes aus Abbildung 3.2, unter Verwendung der Achternachbarschaft: (a) Knoten (als Kreis dargestellt) und Kanten ohne Kantengewichte, (b) 3D-Darstellung der Knoten und Kanten, basierend auf den Koordinaten und der Intensität der korrespondierenden Pixel

Als Nachbarschaftsrelation wird die Achternachbarschaft verwendet:

$$N = \{((x, y), (x', y')) \in P \times P | max(|x - x'|, |y - y'|) = 1\}$$

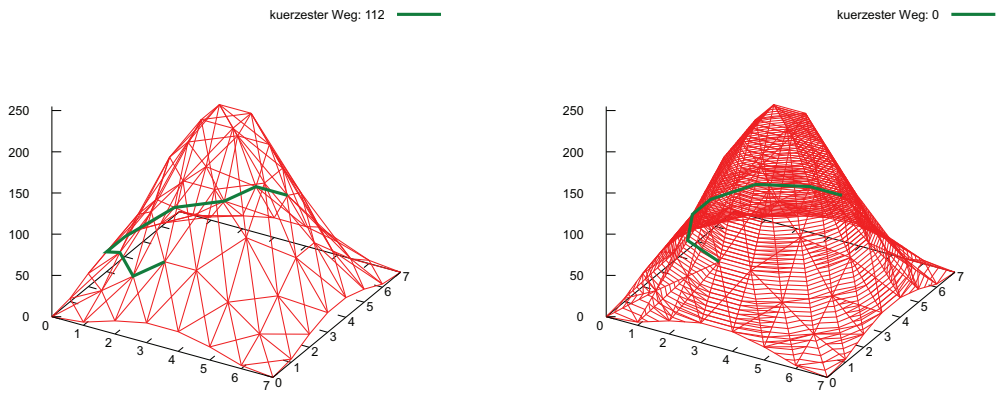
Da die Nachbarschaftsrelation symmetrisch ist, ist der resultierende Graph ungerichtet.

Abbildung 3.3 zeigt den resultierenden Graphen bei Konvertierung des Beispielbildes. Dabei sind in Abbildung 3.3 (a) die 8×8 Knoten V (als Kreise angeordnet nach den (x, y)-Koordinaten der Knoten) und Kanten E (als Linien zwischen Knoten) des Graphen dargestellt, (b) zeigt eine dreidimensionale Darstellung der Kanten unter Verwendung der Intensität der Knoten als z-Koordinate. Die Kantengewichte w sind in der Abbildung nicht dargestellt.

Wie bereits in Abschnitt 3.2 bewiesen, ist das oben angegebene Abstandsmaß für $\alpha \to 1$ im kontinuierlichen Fall invariant gegenüber Deformationen. Da die betrachteten Bilder und auch die daraus entstehenden Graphen diskreter Natur sind, gilt der Beweis im Allgemeinen nicht. Durch Erhöhung der Abtastrate kann allerdings der kontinuierliche Fall approximiert werden. Durch eine Erweiterung des Graphen um zusätzliche Knoten und Kanten kann bei fester Abtastrate die Deformationsinvarianz wie im Folgenden beschrieben erhöht werden.

Erweiterte Repräsentation

Das Diskretisierungsproblem fällt vor allem dort auf, wo hohe Intensitätssprünge auftreten. Der Grund hierfür ist, dass die Knoten des Graphen durch eine regelmäßige



(a) (b)

Abbildung 3.4: Kürzester Weg (in grün) zwischen zwei Knoten des Beispielgraphen: (a) Exakte Graphrepräsentation, (b) Graphrepräsentation erweitert um zusätzliche Knoten und Kanten an Intensitätssprüngen

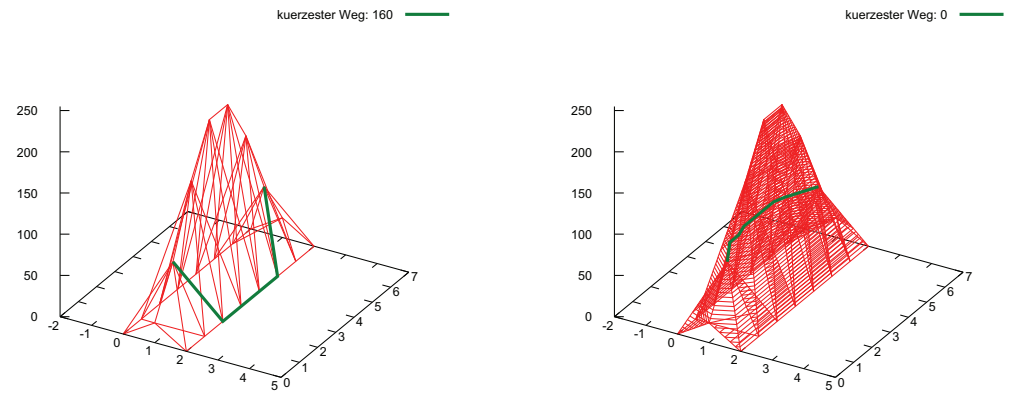
Abtastung der Bildkoordinaten erstellt werden, die Intensitäten aber nicht regelmäßig abgetastet werden. Durch Einfügen zusätzlicher Knoten zwischen benachbarten Pixeln mit hoher Intensitätsdifferenz lässt sich dieses Problem umgehen.

Abbildung 3.4 zeigt das Diskretisierungsproblem an einem Beispiel. Abbildung 3.4 (a) zeigt die oben beschriebene Graphrepräsentation und den kürzesten Weg zwischen zweier seiner Knoten. Obwohl beide Knoten dieselbe Intensität haben, ist der Abstand größer als 0, da zwischen den beiden Knoten Knoten mit abweichender Intensität liegen. Die kontinuierliche Repräsentation des Beispiels (eine zweidimensonale Gaußkurve) hat dagegen sehr wohl einen kürzesten Weg der Länge 0 zwischen den beiden Punkten, welche den Knoten des Graphen entsprechen.

Abbildung 3.4 (b) zeigt eine erweiterte Graphrepräsentation, die durch regelmäßige Abtastung der Intensitäten entstanden ist, zusammen mit dem kürzesten Weg zwischen denselben Knoten des ersten Beispiels. Die Länge des kürzesten Weges entspricht hier dem kontinuierlichen Fall.

Zur Illustration der Unterschiede der beiden Graphrepräsentationen zeigt Abbildung 3.5 einen Graph einer deformierten Version des Beispielbildes. Die Längen der kürzesten Wege weichen hier noch stärker voneinander ab, was nahelegt, dass die Deformationsinvarianz ohne Erweiterung des Graphen stark leidet. Im erweiterten Graph wiederum entspricht die Länge des kürzesten Weges exakt den Erwartungen.

Um den erweiterten Graph zu erhalten, wird das Bild $I = (P, i)$ als eine Abtastung der zugrundeliegenden kontinuierlichen Bildmannigfaltigkeit M angesehen. Die Pixel P zusammen mit ihrer Intensität i bilden dann eine Teilmenge der Mannigfaltigkeit $(x, y) \in P \Rightarrow (x, y, i(x, y)) \in M$ (vgl. Abschnitt 3.2.1). Abbildung 3.6 (a) zeigt



(a) (b)

Abbildung 3.5: Kürzester Weg zwischen zwei Knoten eines deformierten Beispielgraphen: (a) Oben beschriebene Graphrepräsentation, (b) Graphrepräsentation erweitert um zusätzliche Knoten und Kanten an großen Intensitätssprüngen

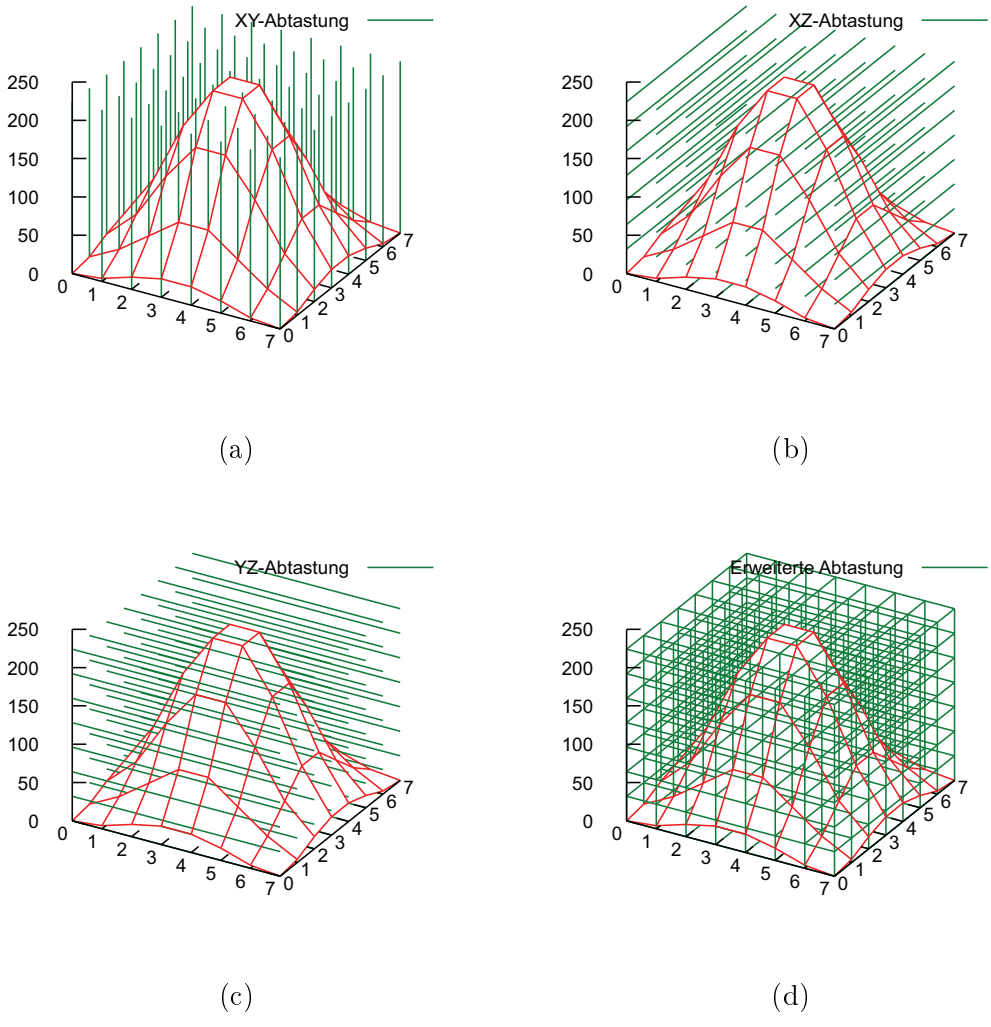
diese Abtastung, die in dem ersten (nicht erweiterten) Graphen resultiert. Die Punkte $(x, y, i(x, y)), (x, y) \in P$ entsprechen Schnittpunkten der Bildmannigfaltigkeit mit den gezeigten Geraden.

Erweitert man die Abtastung der Mannigfaltigkeit um die Intensität (die z-Koordinate), erhält man weitere Knoten im Graphen. Abbildung 3.6 (b) und (c) zeigen die zwei zusätzlichen Abtastungen bei Hinzunahme der Intensität. Der Anschaulichkeit halber ist die Auflösung der Intensitätsabtastung in der Abbildung niedrig gehalten (alle 32 Graustufen). Für größte Genauigkeit würde die Intensitätsabtastung der Auflösung der Darstellung der Intensitätswerte entsprechen, bei 8Bit-Graubildern also typischerweise 1 bei 256 Abtaststufen.

Die Knoten des erweiterten Graphen entsprechen nun nicht mehr Punkten (x, y) in der Ebene, sondern Punkten (x, y, z) im dreidimensionalen Raum bzw. in der Mannigfaltigkeit M. Diese Punkte entsprechen Schnittpunkten der Bildmannigfaltigkeit mit den Abtastgeraden. Die Kanten des erweiterten Graphen ergeben sich nicht mehr durch die (zweidimensionale) Nachbarschaft der Pixel, sondern durch eine Nachbarschaft auf der Bildmannigfaltigkeit. Semiformal ist der erweiterte Graph $G_e = (V_e, E_e, w_e)$ definiert durch:

$$\begin{aligned} V_e &= \{(x, y, z) \in M | (x, y, z) \text{ liegt auf einer Abtastgeraden}\} \\ E_e &= \{(v_1, v_2) \in M \times M | v_1 \text{ und } v_2 \text{ sind benachbart in } M\} \\ w_e((x, y, z), (\bar{x}, \bar{y}, \bar{z})) &= \sqrt{(1-\alpha)^2(x-\bar{x})^2 + (1-\alpha)^2(y-\bar{y})^2 + \alpha^2(z-\bar{z})^2} \end{aligned} \quad (3.4)$$

Da die abgetastete kontinuierliche Mannigfaltigkeit im praktischen Fall nicht vorliegt,



(a) (b)

(c) (d)

Abbildung 3.6: (a) Gleichmäßige Abtastung des Bildes über die x- und y-Koordinaten, (b) gleichmäßige Abtastung des Bildes über die x- und z-Koordinaten (Intensität), (c) gleichmäßige Abtastung des Bildes über die y- und z-Koordinaten (Intensität), (d) kombinierte erweiterte Abtastung

muss sie angenähert werden. Dies kann z.B. durch lineare Interpolation zwischen den Pixeln geschehen (die Darstellung der Kanten in Abbildung 3.3 (b) veranschaulicht bereits eine solche Interpolation).

3.3.2 Graphmetrik

Durch die Beschreibung eines Bildes als Graph lassen sich gängige Graphalgorithmen auf ein Bild anwenden, z.B. der Algorithmus von Dijkstra [4] zur Bestimmung kürzester Wege in Graphen. Dies ermöglicht die Definition einer Metrik $d : V \times V \longrightarrow \mathbb{R}$ über den Knoten eines Graphen, basierend auf den Längen der kürzesten Wege zwischen Knoten. Wir definieren $d(a, b)$ als die Länge des kürzesten Weges zwischen den Knoten a und b.

Satz 4. *$d : V \times V \longrightarrow \mathbb{R}$ ist eine Metrik über den Knoten eines Graphs $G = (V, E, w)$ genau dann wenn G keine Kanten der Länge 0 enthält. Anders formuliert:*

$$G \text{ enthält keine Kanten der Länge } 0 \iff d \text{ ist eine Metrik}$$

Beweis 4. *Die Bedingungen an eine Metrik sind:*

$$d(a, a) = 0 \tag{3.5}$$
$$d(a, b) = 0 \Rightarrow a = b \tag{3.6}$$
$$d(a, b) = d(b, a) \tag{3.7}$$
$$d(a, b) \leq d(a, c) + d(c, b) \tag{3.8}$$

Bedingung 3.5 (der kürzeste Weg zwischen einem Knoten und sich selbst hat die Länge 0) ist offensichtlich erfüllt. Da die betrachteten Graphen ungerichtet sind, ist auch die Symmetrie-Bedingung 3.7 erfüllt. Bedingung 3.8 (die Dreiecksungleichung) lässt sich einfach durch einen Widerspruchsbeweis zeigen:

Angenommen, es existierte ein c, so dass $d(a, c) + d(c, b) < d(a, b)$. Dann existiert offensichtlich ein Weg zwischen den Knoten a und b (nämlich über c), der kürzer ist als $d(a, b)$. Dies ist ein Widerspruch, da $d(a, b)$ definiert ist als die Länge des kürzesten Weges zwischen a und b.

Bedingung 3.6 ist offensichtlich erfüllt, wenn G keine Kanten der Länge 0 enthält. Damit ist die eine Richtung der Äquivalenz in Satz 4 bewiesen.

Enhält G Kanten der Länge 0, dann ist Bedingung 3.6 nicht erfüllt, denn dann existiert ein kürzester Weg zwischen den beiden Knoten a und b der Kante mit $d(a, b) = 0$ und $a \neq b$. Daraus folgt die andere Richtung der Äquivalenz. □

Um eine Metrik zu erhalten, muss ein gegebener Graph demnach zunächst so modifiziert werden, dass er keine Kanten der Länge 0 enthält. Dies kann wie folgt geschehen:

1. Wähle eine Kante (a, b) mit $w(a, b) = 0$. Falls keine derartige Kante existiert, ist die Modifikation vollständig.
2. Ersetze die Knoten a und b durch einen neuen Knoten ab
3. Lösche die Kante (a, b)

4. Ersetze alle Kanten (a, c) (für beliebige c) durch Kanten (ab, c) mit $w(ab, c) = w(a, c)$, und ersetze analog alle Kanten (b, c) (für beliebige c) durch Kanten (ab, c) mit $w(ab, c) = w(b, c)$. Existieren zu einem gegebenen c beide Kanten (a, c) und (b, c), so wird definiert $w(ab, c) = \min(w(a, c), w(b, c))$.

5. Fahre fort mit Schritt 1

Im Folgenden werden nur noch derart modifizierte Graphen betrachtet, auf denen d eine Metrik darstellt. Ein Graph G mit der Metrik d wird als Quadrupel $G = (V, E, w, d)$ bezeichnet.

3.3.3 Reduzierte Repräsentation durch abstandsbasierte Approximation

Da die Graphbeschreibung eines Bildes (vor allem des erweiterten Graphen) bei typischen Bildgrößen von mehreren Millionen Pixeln sehr groß wird, ist es in der Praxis u.U. nötig, den Graphen zu verkleinern. Der Algorithmus von Dijkstra [4] z.B. besitzt für die Berechnung der kürzesten Wege ausgehend von einem Knoten eine Zeitkomplexität von $\mathcal{O}(|V| \log |V|)$, wobei $|V|$ der Anzahl der Knoten im Graph entspricht. Das All-Pairs-Shortest-Paths-Problem hingegen benötigt nur zum Ausgeben der Ergebnisse bereits Speicherplatz der Größenordnung $\mathcal{O}(|V|^2)$, was bei einem n von mehreren Millionen selbst im Arbeitsspeicher sehr moderner Rechner nicht mehr unterzubringen ist. Die Zeitkomplexität des Algorithmus' von Johnson [14] zur Lösung des All-Pairs-Shortest-Paths-Problems liegt bei $\mathcal{O}(|V|^2 \log |V|)$, wenn die Anzahl der Kanten wie in unserem Fall proportional zur Anzahl der Knoten ist. Darüberhinaus ist die Anzahl der Knoten quadratisch abhängig von der räumlichen Auflösung des zugrundeliegenden Bildes: Die Anzahl der Pixel eines Bildes ist typischerweise proportional zum Quadrat seiner Breite. Abbildung 3.7 veranschaulicht die Zeitkomplexität der beiden oben genannten Algorithmen. Man beachte, dass moderne Digitalkameras bereits Bilder weit jenseits der in der Abbildung dargestellten Auflösung generieren können.

Die Approximation eines Graphen besteht aus zwei Schritten: Der Auswahl der Knoten des Originalgraphs, die in der Approximation enthalten sein sollen, und dem Einfügen von Kanten zwischen den ausgewählten Knoten.

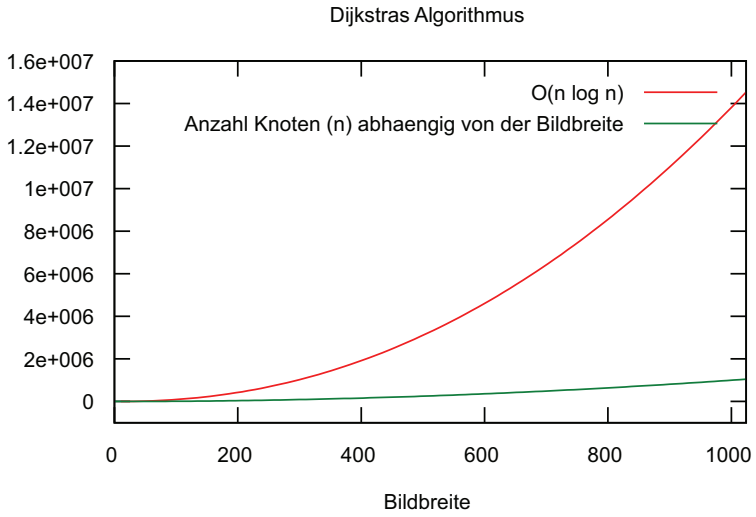
Auswahl von Knoten

Zur Reduzierung der Knotenanzahl wird der Begriff der ε-Umgebung verwendet. Die ε-Umgebung U_ε auf einer Metrik d ist wie folgt definiert:

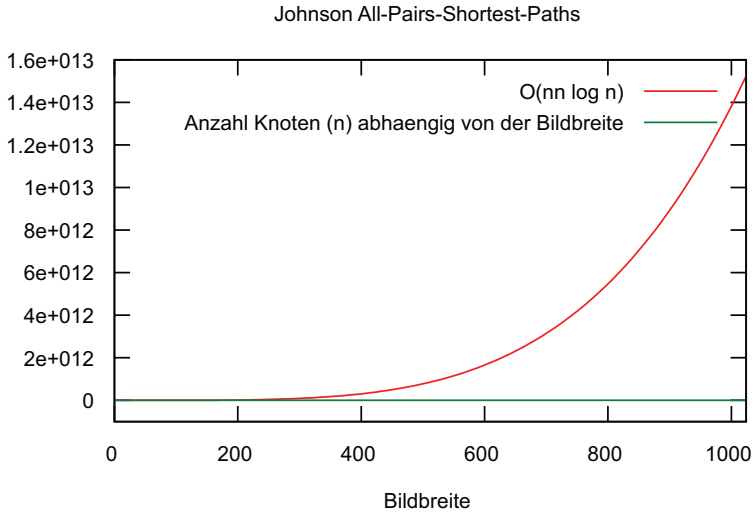
$$U_\varepsilon(a) = \{b \in V | d(a, b) < \varepsilon\} \tag{3.9}$$

Es kann nun eine Knotenauswahl V_ε definiert werden als eine Teilmenge von V, die folgende Bedingungen erfüllt:

$$\begin{aligned} \forall a, b \in V_\varepsilon, a \neq b &: b \notin U_\varepsilon(a) \\ \forall a \in V &: \exists b \in V_\varepsilon(a) \\ V_\varepsilon &\subset V \end{aligned} \tag{3.10}$$



(a)



(b)

Abbildung 3.7: Zeitkomplexität von Graphalgorithmen in Abhängigkeit von der Bildbreite: (a) Dijkstras Algorithmus zur Bestimmung kürzester Wege, (b) Johnsons Algorithmus zur Lösung des All-Pairs-Shortest-Paths-Problems

Die Bedingungen fordern, dass kein Knoten in V_ε im ursprünglichen Graphen in der Umgebung eines anderen Knoten aus V_ε liegen darf, und dass jeder Knoten aus V in der Umgebung eines Knoten in V_ε liegen muss.

Es gibt u.U. viele solcher Knotenmengen. Interessant sind jedoch vor allem Mengen von Knoten, die möglichst nah aneinanderliegen und dennoch die Bedingung in Gleichung 3.10 erfüllen. Um diese zusätzliche Bedingung, dass die Knoten möglichst nah beeinanderliegen sollen, formulieren zu können, wird zunächst der Rand R_ε der Umgebung U_ε wie folgt definiert:

$$R_\varepsilon(a) = \{b \in V | \exists c : (b, c) \in E \wedge c \in U_\varepsilon(a)\} \tag{3.11}$$

Der Rand einer Umgebung enthält alle Knoten, die mit einem Knoten innerhalb der Umgebung durch eine Kante verbunden sind. Die zweite Bedingung wird nun wie folgt definiert:

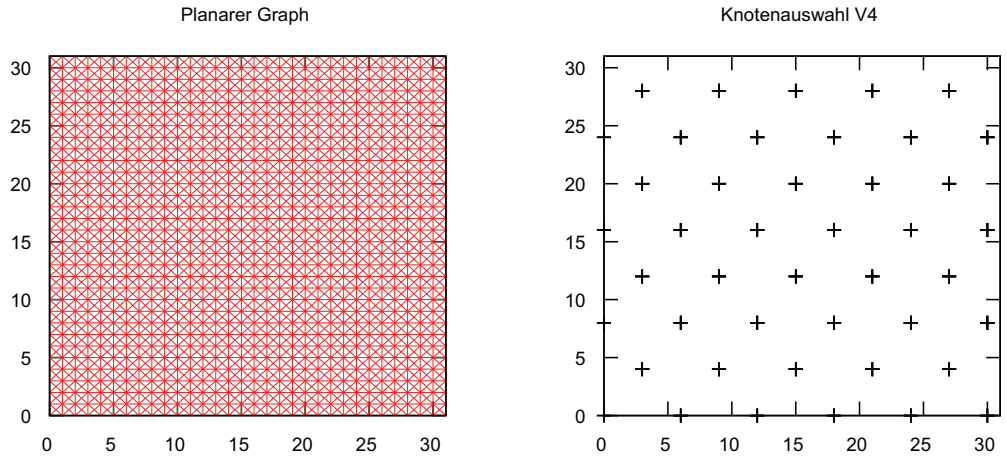
$$\begin{aligned} R(a,b) &= \begin{cases} 1 \text{ falls } b \in R_\varepsilon(a) \vee a \in R_\varepsilon(b), \\ 1 \text{ falls } \exists c : R(a,c) = 1 \wedge R(c,b) = 1, \\ 0 \text{ sonst} \end{cases} \\ \forall a, b \in V_\varepsilon, a \neq b &: \quad R(a,b) = 1 \end{aligned} \tag{3.12}$$

Im Folgenden wird ein Algorithmus vorgestellt, der Knotenmengen V_ε erzeugt, die den Bedingungen 3.10 und 3.12 bei einem gegebenen ε genügen:

1. Wähle einen Knoten $a \in V$ und füge ihn zu V_ε hinzu
2. Wähle einen noch nicht ausgewählten Knoten b auf dem Rand der Umgebung von a: $b \in R_\varepsilon(a) \wedge b \notin V_\varepsilon$. Brich ab, falls kein solcher Knoten existiert
3. Füge dem Graphen eine Kante der Länge 0 zwischen a und b hinzu
4. Fahre fort mit Schritt 2

Die Erfüllung der Bedingungen ergibt sich direkt aus der Konstruktion. Die Umgebung eines Knotens und der Rand können effizient mittels des Algorithmus' von Dijkstra bestimmt werden. Der Algorithmus von Dijkstra kann dabei nach Erreichen des ersten Knotens außerhalb der Umgebung des Startknotens gestoppt werden. Die Anzahl seiner Aufrufe entspricht der Größe der entstehenden Knotenmenge, welche der Anzahl aller Knoten in V geteilt durch die durchschnittliche Anzahl von Knoten in der ε-Umgebung eines Knotens entspricht. Bereits einmal besuchte Knoten müssen nicht ein zweites Mal besucht werden. Daher entspricht der Aufwand des oben angegebenen Algorithmus' insgesamt dem Aufwand des Algorithmus' von Dijkstra. Abbildung 3.8 zeigt ein Beispiel einer Knotenauswahl V_4 ($\varepsilon = 4$) auf einem planaren Graphen mit 32×32 Knoten.

Im Folgenden werden verschiedene Ansätze vorgestellt, Kanten zwischen den ausgewählten Knoten einzufügen, um den approximierten Graph zu erhalten. Die verschiedenen Approximationen werden anhand eines Beispielgraphen (s. Abbildung 3.9 (a)) illustriert. Der Anschaulichkeit halber wurde im Beispiel der α-Parameter für die Kantengewichte auf einen Wert kleiner 1 gesetzt, so dass die Abstände der Knoten in etwa den wahrgenommenen Abständen in der dreidimensionalen Darstellung entsprechen.



(a) (b)

Abbildung 3.8: (a) Planarer Graph mit 32×32 Knoten. Horizontale und vertikale Kanten haben ein Gewicht von 1, diagonale Kanten ein Gewicht von $\sqrt{2}$ (b) Knotenauswahl für $\varepsilon = 4$

Delaunay-Graph

Die Delaunay-Triangulierung, benannt nach Boris Delaunay [3] ist ursprünglich eine Triangulierung auf einer gegebenen Menge von Punkten in der Ebene. Nutzt man die Äquivalenz der Delaunay-Triangulierung mit dem dualen Graphen der Voronoi-Zerlegung, so kann man das Konzept auf Graphen übertragen. Die Zerlegung eines Graphen $G = (V, E, w, d)$ in Voronoi-Regionen VR, gegeben eine Menge von Knoten V_d, kann wie folgt definiert werden:

$$\begin{aligned} VR : V_d &\longrightarrow \mathcal{P}(V) \\ VR(a) &= \{b \in V | \forall c \in V_d : d(a,b) < d(c,b)\} \end{aligned} \tag{3.13}$$

Die Voronoi-Region eines Knotens a beinhaltet demnach alle Knoten aus V, die näher an a liegen als an allen anderen Knoten aus V_d. Der Delaunay-Graph $G_d = (V_d, E_d, w_d)$ wird nun wie folgt definiert:

$$\begin{aligned} E_d &= \{(a,b) \in V_d \times V_d \,|\, (\exists (c,d) \in E : c \in VR(a) \wedge d \in VR(b))\} \\ w_d(a,b) &= d(a,b) \end{aligned} \tag{3.14}$$

Der resultierende Graph enthält demnach Kanten zwischen denjenigen Knoten, deren Voronoi-Regionen im Originalgraphen benachbart sind. Abbildung 3.9 (b) zeigt den Delaunay-Graphen des Beispiels. Die Voronoi-Zerlegung eines Graphen lässt sich sehr einfach algorithmisch bestimmen, indem die kürzesten Wege ausgehend von allen Knoten in V_d berechnet werden. Dies entspricht einem einzelnen Aufruf des Dijkstra-Algorithmus, wenn man zuvor einen zusätzlichen Knoten in den Graphen einfügt sowie Kanten mit

einem Gewicht von 0 zwischen diesem neuen Knoten und allen Knoten in V_d. Die Kanten E_d lassen sich dann mit linearem Aufwand abhängig von der Anzahl der Kanten im Originalgraphen bestimmen. Der Aufwand zur Erstellung des Delaunay-Graphs entspricht daher der Größenordnung $\mathcal{O}(|V|\log|V|+|E|)$. Da die Anzahl der Kanten der vorliegenden Graphen proportional ist zur Anzahl seiner Knoten, entspricht dies $\mathcal{O}(|V|\log|V|)$.

Delaunay-Graph von Graphfärbungen

Die Genauigkeit der Approximation des Delaunay-Graphen kann erhöht werden, indem zusätzliche Kanten zwischen Knoten eingefügt werden, deren Voronoi-Regionen nicht direkt benachbart sind. Durch Auswahl von Teilmengen V'_d von V_d und Bestimmung des Delaunay-Graphen von V'_d erhält man weitere solcher Kanten. An die Auswahl solcher Teilmengen stellen sich die folgenden Anforderung: Sie sollen disjunkt sein. Ihre Vereinigung soll wieder V_d entsprechen. Die Knoten in einer Teilmenge sollten im Delaunay-Graphen G_d nicht benachbart sein. Außerdem sollte die Anzahl der Teilmengen möglichst gering sein. Dies lässt sich auf das Graphfärbungsproblem zurückführen:

Sei $f : V \longrightarrow \{1,\ldots,k\}$ eine gültige k-Knotenfärbung des Graphen $G = (V,E,w)$. Dann gilt: $\forall(a,b) \in E : f(a) \neq f(b)$. Definiert man die Knotenmengen $V_{di} = \{a \in V_d | f(a) = i\}, 1 \leq i \leq k$ als die Menge derjenigen Knoten, deren Farbe i ist, so erfüllt die Lösung des Graphfärbungsproblems die oben angegebenen Anforderungen.

Die Lösung des Graphfärbungsproblems ist NP-vollständig. Es existieren allerdings effiziente Heuristiken, z.B. der Algorithmus von Matula [26]. Wenn eine Aufteilung der Knoten V_d in die verschiedenen Farben $V_d i$ und die korrespondierenden Delaunay-Graphen $G_{di} = (V_{di}, E_{di}, w_{di})$ gegeben ist, ergibt sich der Delaunay-Graph von Graphfärbungen $G_f = (V_f, E_f, w_f)$ wie folgt:

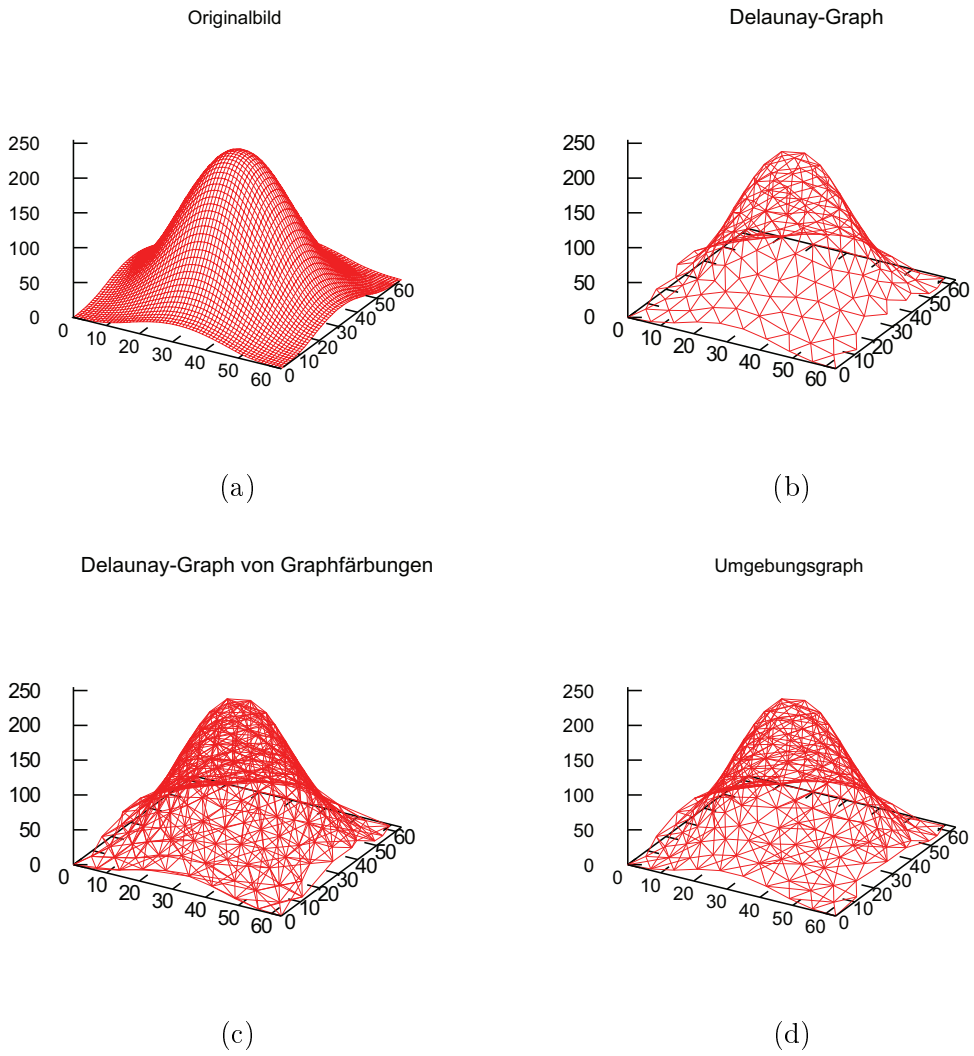
$$\begin{aligned} V_f &= V_d \\ E_f &= \{(a,b) \in V_d \times V_d | (a,b) \in E_d \vee \exists i : (a,b) \in E_{di}\} \\ w_f(a,b) &= d(a,b) \end{aligned} \tag{3.15}$$

Der Delaunay-Graph von Graphfärbungen enthält so alle Kanten des Delaunay-Graphen und zusätzlich die Kanten der Delaunay-Graphen der durch die Graphfärbung erhaltenen Knotenteilmengen V_{di}. Abbildung 3.9 (c) zeigt den Delaunay-Graph von Graphfärbungen des Beispiels.

$2 \cdot \varepsilon$-Umgebungsgraph

Der $2 \cdot \varepsilon$-Umgebungsgraph $G_{2\varepsilon} = (V_{2\varepsilon}, E_{2\varepsilon}, w_{2\varepsilon})$, gegeben einen Originalgraph $G = (V,E,w,d)$ und eine Knotenmenge $V_{2\varepsilon} \subset V$, wird definiert als derjenige Graph, der Kanten zwischen zweien seiner Knoten enthält, wenn die Distanz zwischen diesen weniger als $2 \cdot \varepsilon$ beträgt:

$$\begin{aligned} E_{2\varepsilon} &= \{(a,b) \in V_{2\varepsilon} \times V_{2\varepsilon} | d(a,b) < 2 \cdot \varepsilon\} \\ w_{2\varepsilon}(a,b) &= d(a,b) \end{aligned} \tag{3.16}$$



(a) (b)

(c) (d)

Abbildung 3.9: Beispielgraph und dessen Approximationen

Messung des Approximationsfehlers

Zur Messung der Güte einer Approximation wird die Abweichung der Distanzen d' im approximierten Graph von der Distanz d im Originalgraph betrachtet. Für zwei gegebene Knoten v' und u' im approximierten Graph gilt dann nach der Dreiecksungleichung:

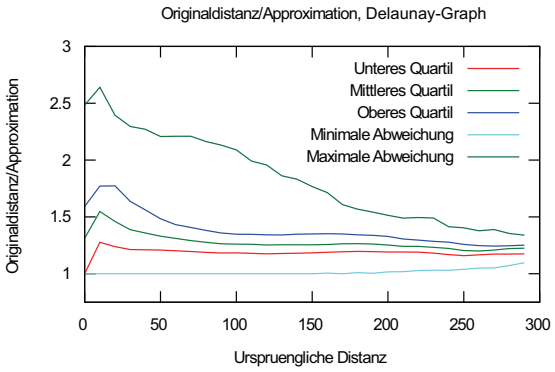
$$d'(v', u') \geq d(v', u')$$

Wir definieren die Distanzabweichung Δd als das Verhältnis zwischen Distanz im approximierten Graph und der Originaldistanz:

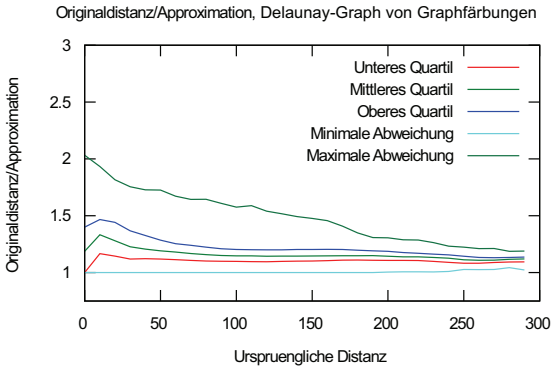
$$\Delta d(v', u') = d'(v', u')/d(v', u'),\ \Delta d(v', u') \geq 1 \tag{3.17}$$

Bei einer perfekten Approximation gilt $\forall v', u' \in V' : \Delta d(v', u') = 1$. Wir messen die Distanzabweichung auf einer Menge von Bildern durch Berechnung der kürzesten Wege ausgehend von zufällig gewählten Punkten, jeweils im Original- und im approximierten Graph. Die entstehenden Verteilungen sind in den Abbildungen 3.10 (einfache Graphrepräsentation wie in Abschnitt 3.3.1 beschrieben) und 3.11 (erweiterte Graphrepräsentation aus Abschnitt 3.3.1) zu sehen.

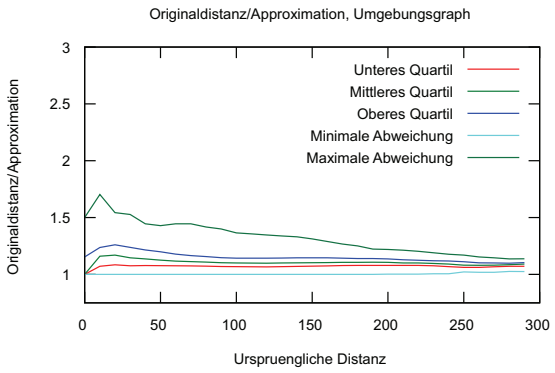
Aus den Messungen wird deutlich, dass der $2 \cdot \varepsilon$-Umgebungsgraph eine bessere Approximation bietet als die beiden anderen Approximationen, wobei der Delaunay-Graph von Graphfärbungen erwartungsgemäß genauer ist als der normale Delaunay-Graph.



(a)

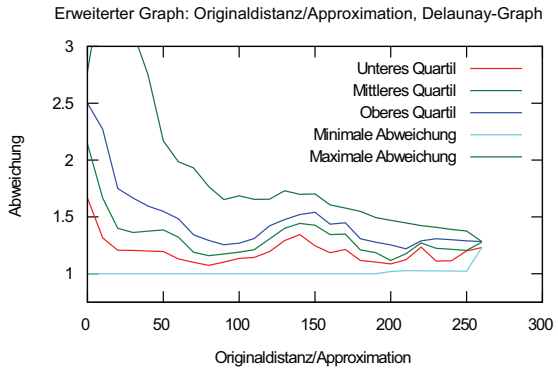


(b)

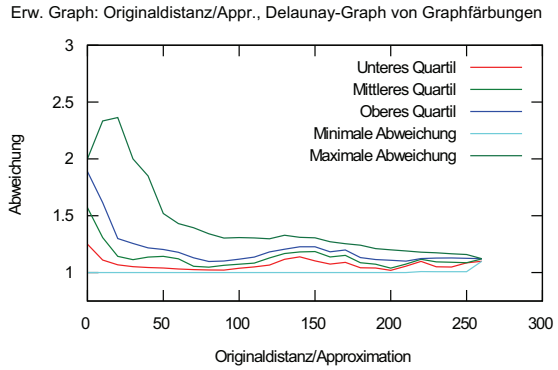


(c)

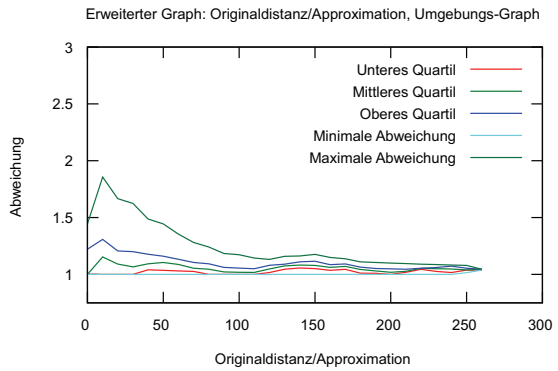
Abbildung 3.10: Verteilung der Distanzabweichung Δd, aufgetragen über der Originaldistanz: (a) Delaunay-Graph, (b) Delaunay-Graph von Graphfärbungen, (c) $2 \cdot \varepsilon$-Umgebungsgraph



(a)



(b)



(c)

Abbildung 3.11: Verteilung der Distanzabweichung Δd im erweiterten Graph, aufgetragen über der Originaldistanz: (a) Delaunay-Graph, (b) Delaunay-Graph von Graphfärbungen, (c) $2 \cdot \varepsilon$-Umgebungsgraph

3.4 Abstände in Graphen unterschiedlicher Bildstrukturen

In diesem Abschnitt werden einige einfache Bildstrukturen in ihrer Graphrepräsentation betrachtet. Untersucht werden sollen vor allem die Ausprägungen der Abstände zwischen den Knoten einer gegebenen Bildstruktur, sowie Unterschiede dieser Abstände zwischen unterschiedlichen Bildstrukturen.

Um die Abstände innerhalb einer Bildstruktur visualisieren zu können, wird das Verfahren der multidimensionalen Skalierung [34, S. 504ff.], ein klassisches statistisches Visualisierungsverfahren, angewendet. Die multidimensionale Skalierung verarbeitet eine Abstandsmatrix der Größe $n \times n$ und bestimmt daraus eine Konfiguration von Punkten im n-dimensionalen Euklidschen Raum $\mathbb{R}^n$. Sei $(d_{ab}), a, b \in 1, \ldots, n$ die Matrix der im Graphen gemessenen Abstände zwischen den Knoten a und b. Sei δ_{ab} die Matrix der Euklidschen Abstände der durch die multidimensionale Skalierung bestimmten Punkte im n-dimensionalen Raum. Dann findet das Verfahren der multidimensionalen Skalierung eine Punktekonfiguration im $\mathbb{R}^n$, die die folgende Funktion S^2, auch Stress genannt, minimiert:

$$S^2 = \frac{\sum (d_{ab} - \delta_{ab})^2}{\sum d_{ab}} \tag{3.18}$$

Eine wichtige Rolle in diesem Verfahren spielt die folgende Matrix $\mathbf{B}$:

$$\begin{aligned} \mathbf{B} &= \left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\left(-\frac{1}{2}d_{ab}^2\right)\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right), \text{ mit} \\ \mathbf{I} &\quad \text{Identität der Größe } n \times n \\ \mathbf{J} &= \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix} \text{ der Größe } n \times n \end{aligned} \tag{3.19}$$

Ist $\mathbf{B}$ positiv-semidefinit, so existiert eine perfekte Lösung mit $S^2 = 0$, d.h., das Verfahren liefert eine Punktkonfiguration im $\mathbb{R}^n$, deren Abstände denen in der gegebenen Abstandsmatrix entsprechen. Die Eigenwerte $(\lambda_1, \ldots, \lambda_n)$ der Matrix $\mathbf{B}$ geben dann Aufschluss über die Bedeutung der n Dimensionen der resultierenden Punktkonfiguration. Sind nur die ersten q Eigenwerte ungleich 0, so existiert eine Punktkonfiguration im $\mathbb{R}^q$, die die gegebenen Abstände perfekt beschreibt. Damit lässt sich die Dimensionalität eines Graphen bzw. der entsprechenden Bildstruktur definieren.

Im Folgenden werden die Graphrepräsentationen verschiedener Bildstrukturen per All-Pairs-Shortest-Paths in eine Abstandsmatrix überführt. Mittels multidimensionaler Skalierung wird dann die Dimensionalität q einer Bildstruktur bestimmt. Die resultierende Punktkonfiguration im $\mathbb{R}^q$ wird betrachtet. Abbildung 3.12 (a) zeigt einen hellen Fleck auf dunklem Grund, ähnlich dem bereits in Abschnitt 3.3.1 betrachteten Beispielbild (s. Abbildung 3.2). Abbildung 3.12 (c) zeigt die ersten beiden Dimensionen der resultierenden Punktkonfiguration nach multidimensionaler Skalierung, (d) zeigt die ersten zehn Eigenwerte der Matrix $\mathbf{B}$. Die Dimensionalität dieser Bildstruktur entspricht demnach

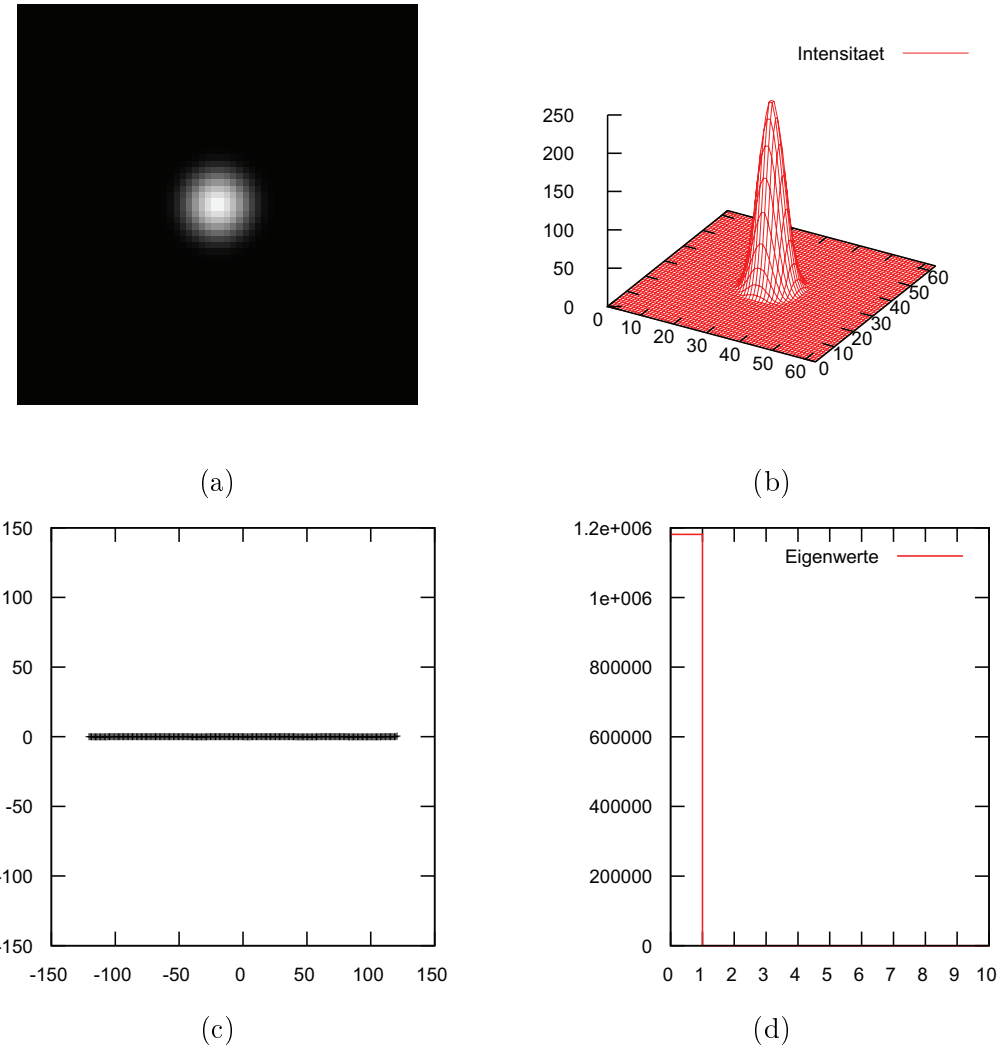


Abbildung 3.12: Beispielbild, bestehend aus einem hellen Fleck auf dunklem Hintergrund: (a) Originalbild, (b) 3D-Darstellung unter Verwendung der Intensität als z-Koordinate, (c) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$, (d) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$

1. Die ermittelte Punktkonfiguration bildet die gegebenen Abstände perfekt, d.h. mit $S^2 = 0$, ab.

Abbildung 3.13 zeigt dasselbe für eine Bildstruktur mit zwei Leuchtpunkten. Die Dimensionalität beider Strukturen ist gleich. Die Punktkonfiguration ist ebenfalls sehr ähnlich, nur die Ausdehnung ist größer, da die maximalen Abstände in der zweiten Bildstruktur größer sind.

Bei Untersuchung einer Bildstruktur mit drei Leuchtpunkten, wie in Abbildung 3.14 zu sehen, zeigen sich plötzlich große Unterschiede: Der Stress S^2 ist nun größer null, d.h. die gegebenen Abstände können im Euklidschen Raum (zumindest im $\mathbb{R}^n$) nicht mehr genau abgebildet werden. Die Dimensionalität der Struktur ist drei. Abbildung 3.14 (c) zeigt die resultierende dreidimensionale Punktkonfiguration (bzw. eine Projektion davon). Es sind vier markante Punkte zu erkennen: Der Mittelpunkt, an dem die drei „Ausleger“ sich treffen, sowie deren Endpunkte. Die Endpunkte entsprechen im Bild den hellsten Punkten der drei Leuchtpunkte. Der Mittelpunkt entspricht dem dunklen Hintergrund. Betrachtet man die Abstandsmatrix $\mathbf{A}$ dieser vier Punkte in 3.20, wird klar, dass es keine Punktkonfiguration im Euklidschen Raum (gleich welcher Dimensionalität) geben kann:

$$\mathbf{A} = \begin{pmatrix} 0 & 240 & 240 & 240 \\ 240 & 0 & 480 & 480 \\ 240 & 480 & 0 & 480 \\ 240 & 480 & 480 & 0 \end{pmatrix} \tag{3.20}$$

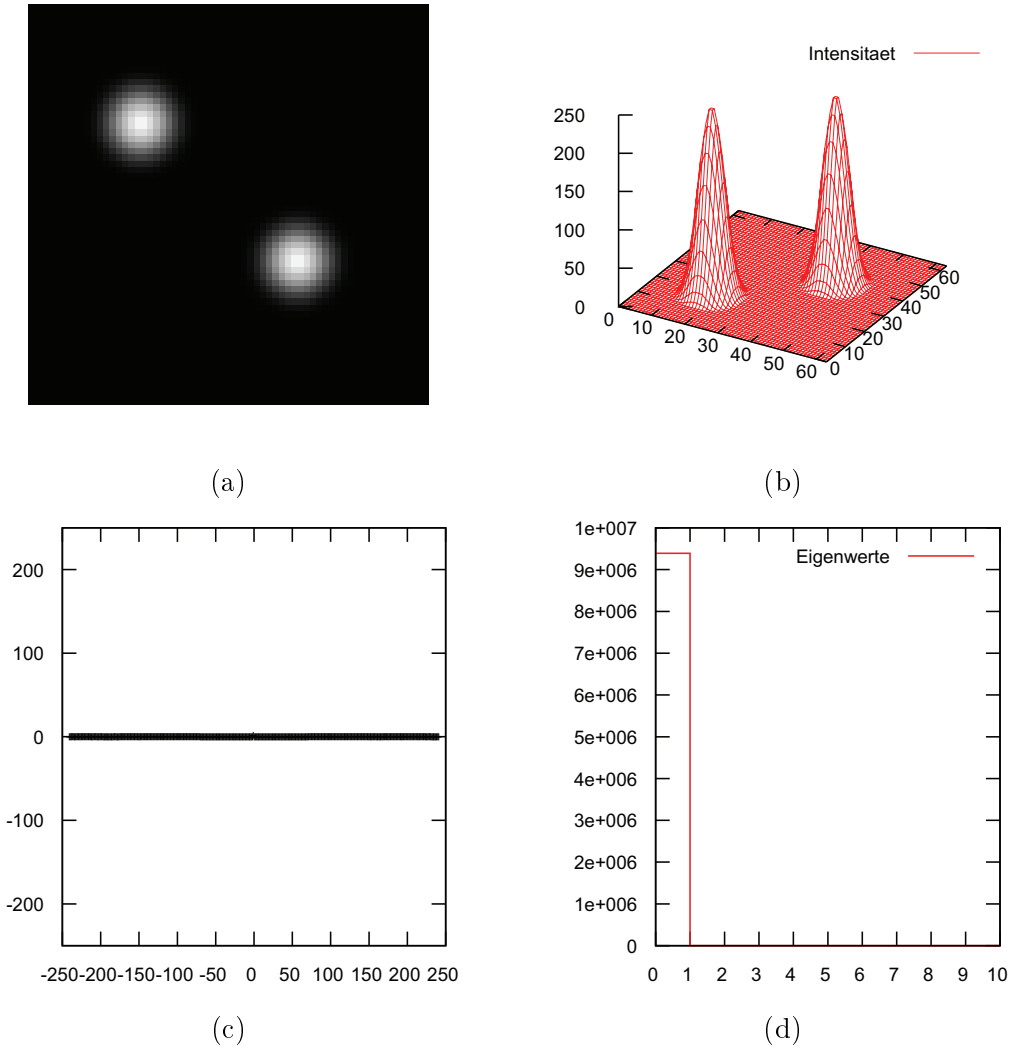
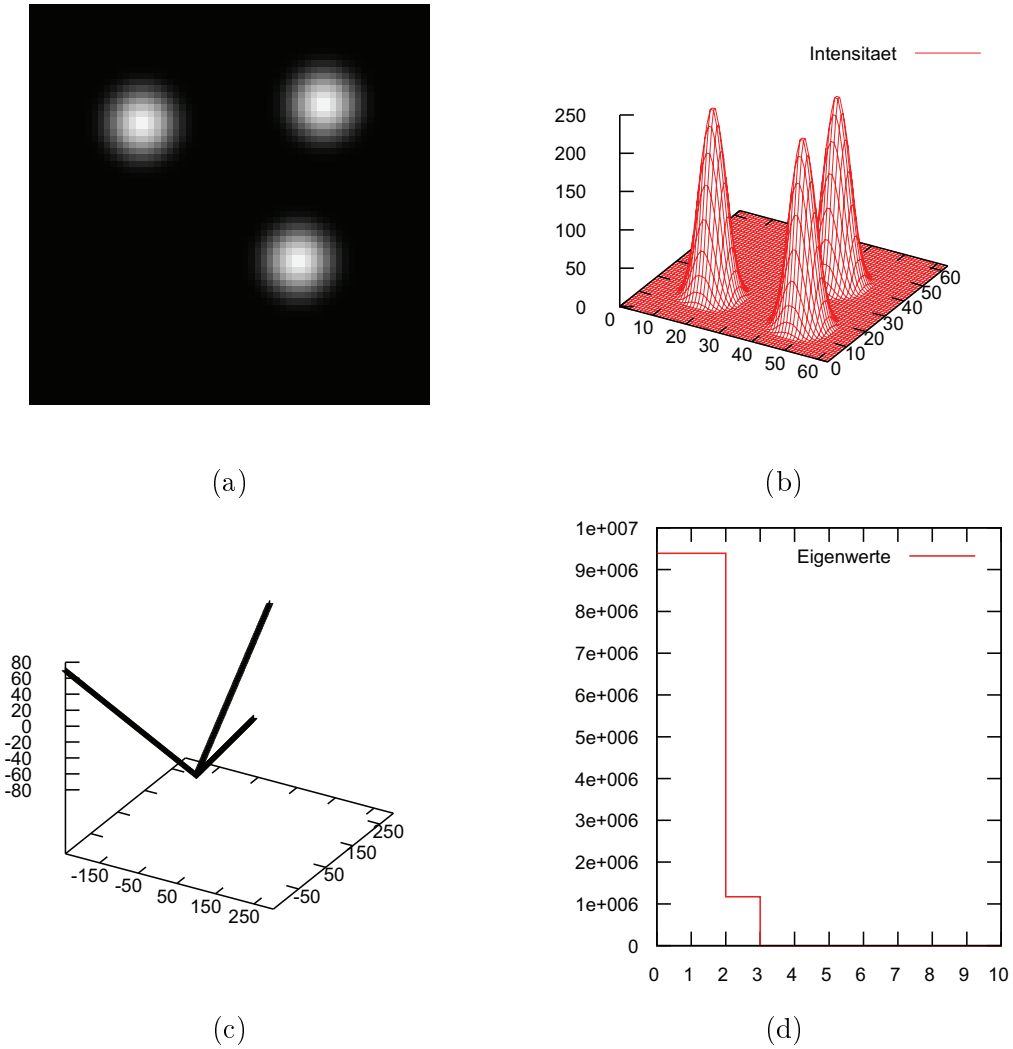


Abbildung 3.13: Beispielbild, bestehend aus zwei Leuchtpunkten auf dunklem Hintergrund: (a) Originalbild, (b) 3D-Darstellung unter Verwendung der Intensität als z-Koordinate, (c) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$, (d) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$



(a) (b)

(c) (d)

Abbildung 3.14: Beispielbild, bestehend aus drei Leuchtpunkten auf dunklem Hintergrund: (a) Originalbild, (b) 3D-Darstellung unter Verwendung der Intensität als z-Koordinate, (c) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 \approx 0.011$), eingebettet im $\mathbb{R}^3$, (d) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$

Satz 5. *Für die in Gleichung 3.20 angegebene Abstandsmatrix* $\mathbf{A}$ *gibt es keine Punktkonfiguration im Euklidschen Raum, die dieselben Abstände aufweist.*

Beweis 5. *Nach Rencher [34, S. 505] (vgl. Schoenberg [36]) existiert eine solche Punktkonfiguration genau dann, wenn die Matrix* $\mathbf{B} = (\mathbf{I} - 1/n\mathbf{J}) - 1/2\mathbf{A}^2(\mathbf{I} - 1/n\mathbf{J})$ *positiv-semidefinit ist. Für die gegebene Abstandsmatrix ist*

$$\mathbf{B} = \begin{pmatrix} -21600 & 7200 & 7200 & 7200 \\ 7200 & 151200 & -79200 & -79200 \\ 7200 & -79200 & 151200 & -79200 \\ 7200 & -79200 & -79200 & 151200 \end{pmatrix}$$

Eine notwendige Bedingung für eine positiv-semidefinite Matrix $\mathbf{B}$ *ist, dass* $\mathbf{B}_{ii} \geq 0$*, d.h. alle Elemente in der Diagonalen der Matrix müssen größer oder gleich Null sein. Das ist in diesem Fall offensichtlich nicht gegeben.* □

Für diesen einfachen Fall lässt sich die Behauptung allerdings alternativ auch anschaulich zeigen: Der Mittelpunkt entspricht der ersten Zeile bzw. Spalte in der Matrix. Da Rotation und Translation im Euklidschen Raum die Abstände erhalten, kann man die Positionen der ersten beiden Punkte frei wählen. Setzt man den Mittelpunkt p_1 auf den Ursprung, muss der zweite Punkt p_2 irgendwo auf der (Hyper-)Kugel mit Radius 240 um den Ursprung liegen (der Abstand von 240 ist durch die Abstandsmatrix vorgegeben). Wählt man einen Punkt auf dieser Kugel, so ergeben sich für den dritten Punkt p_3 bereits zwei Einschränkungen: Er muss ebenfalls auf der Kugel mit Radius 240 um den Ursprung liegen (da er denselben Abstand vom Mittelpunkt hat). Außerdem muss er auf einer Kugel mir Radius 480 um den zweiten Punkt liegen. Die beiden Kugeln haben nur genau einen Schnittpunkt: Dieser liegt auf der Gerade, die durch p_1 und p_2 läuft, von p_1 gesehen genau gegenüber von p_2. Für den vierten Punkt p_4 ergeben sich dieselben Einschränkungen, d.h., $p_4 = p_3$. Da durch die Abstandsmatrix allerdings ein Abstand zwischen p_3 und p_4 von 480 vorgegeben ist, ergibt sich ein Widerspruch.

3.4.1 Kombinierte geodätisch-Euklidsche Graphmetrik

Dass die ersten beiden betrachteten Bildstrukturen dieselbe Dimensionalität und nach multidimensionaler Skalierung fast dieselbe Repräsentation im Euklidschen Raum besitzen, ist eher unerwartet, da die Bildstrukturen sich stark unterscheiden. Durch multidimensionale Skalierung und Betrachtung der resultierenden Punktkonfiguration lassen sie sich dagegen nicht unterscheiden. Es stellt sich die Frage, ob ein Abstandsmaß existiert, welches eine solche Unterscheidung ermöglicht. Dazu betrachten wir jeweils die beiden entferntesten Punkte in beiden Bildstrukturen: Im ersten Bild, welches einen Leuchtpunkt vor dunklem Hintergrund zeigt, sind dies der hellste Punkte des Bildes und der dunkle Hintergrund. Im zweiten Bild, welches zwei Leuchtpunkte vor dunklem Hintergrund zeigt, sind dies dagegen jeweils die beiden hellsten Punkte. Die beiden Punktepaare unterscheiden sich offensichtlich durch ihre Intensitätsdifferenz. Im ersten Bild unterscheiden sich beide Punkte stark in ihrer Helligkeit, im zweiten Bild haben beide Punkte dieselbe Helligkeit.

Im Folgenden wird eine Kombination aus dem Abstand im Graphen und dem Euklidschen Abstand zweier Punkte als neues Abstandsmaß betrachtet. Wir bezeichnen einen kürzesten Weg zwischen zwei Punkten im Graphen im Folgenden als *Geodäte*. Der *geodätische Abstand* d zwischen zwei Punkten wird definiert als die Länge des kürzesten Weges zwischen diesen Punkten. Er entspricht dem bisher verwendeten Abstandsmaß. Der *Euklidsche* Abstand e zwischen zwei Punkten (x, y, z) und (x', y', z') im Graph ist gegeben:

$$e((x,y,z),(\bar{x},\bar{y},\bar{z})) = \sqrt{(1-\alpha)^2(x-\bar{x})^2 + (1-\alpha)^2(y-\bar{y})^2 + \alpha^2(z-\bar{z})^2} \tag{3.21}$$

Dies entspricht dem Gewicht einer Kante zwischen zwei benachbarten Knoten (s. Abschnitt 3.3.1, Gleichung 3.4), mit dem Unterschied, dass e auf allen Knotenpaaren im Graph definiert ist. Dabei kann durch Wahl des Parameters α zwischen Bildkoordinaten und Intensität gewichtet werden. Bei $\alpha = 1$ entspricht e der absoluten Intensitätsdifferenz. Offensichtlich ist e eine Metrik (die Euklidsche Metrik).

Wir definieren nun ein neues Abstandsmaß de als Kombination der Metriken d und e:

$$\begin{aligned} de &: V \times V \longrightarrow \mathbb{R} \\ de(a,b) &= \sqrt{d(a,b)^2 + e(a,b)^2} \end{aligned} \tag{3.22}$$

Es stellt sich die Frage, ob de ebenfalls eine Metrik ist. Im Folgenden soll bewiesen werden, dass die angegebene Kombination für jedes Paar von Metriken wieder eine Metrik darstellt.

Satz 6. *Sind d und e Metriken, so ist die Kombination $de(a,b) = \sqrt{d(a,b)^2 + e(a,b)^2}$ ebenfalls eine Metrik.*

Beweis 6. *Die Bedingungen für eine Metrik wurden bereits in Beweis 4 in Abschnitt 3.3.2 aufgeführt. Die Erfüllung der einzelnen Bedingungen wird im Folgenden gezeigt.*

Bedingung 3.5 ($de(a,a) = 0$) ergibt sich aus der Definition von de und der Tatsache, dass d und e Metriken sind: $de(a,a) = \sqrt{d(a,a)^2 + e(a,a)^2} = \sqrt{0^2 + 0^2} = 0$.

Aus $de(a,b) = 0$ folgt $d(a,b) = 0$ und $e(a,b) = 0$ und damit $a = b$. Damit ist Bedingung 3.6 ($de(a,b) = 0 \Rightarrow a = b$) erfüllt.

Da die Metriken d und e symmetrisch sind, ist dies auch de, womit Bedingung 3.7 ($de(a,b) = de(b,a)$) erfüllt ist.

Die Erfüllung der Bedingung 3.8, der Dreiecksungleichung ($de(a,b) \leq de(a,c)+de(c,b)$), lässt sich wie folgt zeigen. Zu zeigen ist:

$$\begin{array}{rcl}
de(a,c) + de(c,b) & \geq & de(a,b) \\
\Updownarrow & & \\
(de(a,c) + de(c,b))^2 & \geq & de(a,b)^2 \\
\Updownarrow & & \\
de(a,c)^2 + de(c,b)^2 + 2de(a,c)de(c,b) & \geq & de(a,b)^2 \\
\Updownarrow & & \\
\sqrt{d(a,c)^2 + e(a,c)^2}^2 + \sqrt{d(c,b)^2 + e(c,b)^2}^2 + 2de(a,c)de(c,b) & \geq & \sqrt{d(a,b)^2 + e(a,b)^2}^2 \\
\Updownarrow & & \\
d(a,c)^2 + e(a,c)^2 + d(c,b)^2 + e(c,b)^2 + 2de(a,c)de(c,b) & \geq & d(a,b)^2 + e(a,b)^2
\end{array} \tag{3.23}$$

Die Äquivalenzen der Ungleichungen gelten, da alle Seiten immer positiv sind. Die Dreiecksungleichung gilt für d und e, da beides Metriken sind. Es gilt demnach

$$\begin{array}{rcl}
d(a,c) + d(c,b) & \geq & d(a,b) \\
\Downarrow & & \\
(d(a,c) + d(c,b))^2 & \geq & d(a,b)^2 \\
\Downarrow & & \\
d(a,c)^2 + d(c,b)^2 + 2d(a,c)d(c,b) & \geq & d(a,b)^2
\end{array}$$

und

$$\begin{array}{rcl}
e(a,c) + e(c,b) & \geq & e(a,b) \\
\Downarrow & & \\
(e(a,c) + e(c,b))^2 & \geq & e(a,b)^2 \\
\Downarrow & & \\
e(a,c)^2 + e(c,b)^2 + 2e(a,c)e(c,b) & \geq & e(a,b)^2
\end{array}$$

Aus der Addition der beiden Ungleichungen folgt, da alle Seiten positiv sind:

$$d(a,c)^2 + e(a,c)^2 + d(c,b)^2 + e(c,b)^2 + 2d(a,c)d(c,b) + 2e(a,c)e(c,b) \geq d(a,b)^2 + e(a,b)^2 \tag{3.24}$$

Nach Kombination der Ungleichungen 3.23 und 3.24 bleibt noch zu zeigen:

$$\begin{array}{rcl}
de(a,c)de(c,b) & \geq & d(a,c)d(c,b) + e(a,c)e(c,b) \\
& \Updownarrow & \\
\sqrt{d(a,c)^2 + e(a,c)^2}\sqrt{d(c,b)^2 + e(c,b)^2} & \geq & d(a,c)d(c,b) + e(a,c)e(c,b) \\
& \Updownarrow & \\
(d(a,c)^2 + e(a,c)^2)(d(c,b)^2 + e(c,b)^2) & \geq & (d(a,c)d(c,b) + e(a,c)e(c,b))^2 \\
& \Updownarrow & \\
d(a,c)^2d(c,b)^2 + e(a,c)^2e(c,b)^2 & + & d(a,c)^2e(c,b)^2 + e(a,c)^2d(c,b)^2 \geq \\
d(a,c)^2d(c,b)^2 + e(a,c)^2e(c,b)^2 & + & 2d(a,c)d(c,b)e(a,c)e(c,b) \\
& \Updownarrow & \\
d(a,c)^2e(c,b)^2 + e(a,c)^2d(c,b)^2 & \geq & 2d(a,c)d(c,b)e(a,c)e(c,b) \\
& \Updownarrow & \\
(d(a,c)e(c,b))^2 + (e(a,c)d(c,b))^2 & \geq & 2d(a,c)e(c,b)e(a,c)d(c,b)
\end{array}$$

Substituiert man mit $p = d(a,c) \cdot e(c,b)$ und $q = e(a,c) \cdot d(c,b)$, bleibt zu zeigen:

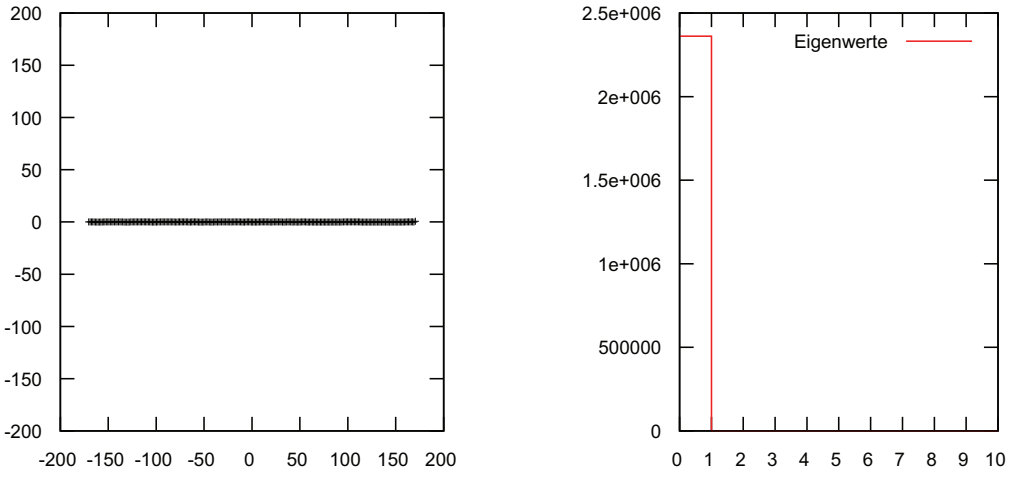
$$p^2 + q^2 \geq 2pq$$

Durch Ausmultiplizieren von $(p-q)^2$ erhält man:

$$\begin{array}{rcl}
(p-q)^2 & = & p^2 + q^2 + 2pq \text{ bzw.} \\
& \Updownarrow & \\
p^2 + q^2 & = & 2pq + (p-q)^2 \\
& \Downarrow & \text{da } (p-q)^2 > 0 \\
p^2 + q^2 & \geq & 2pq
\end{array}$$

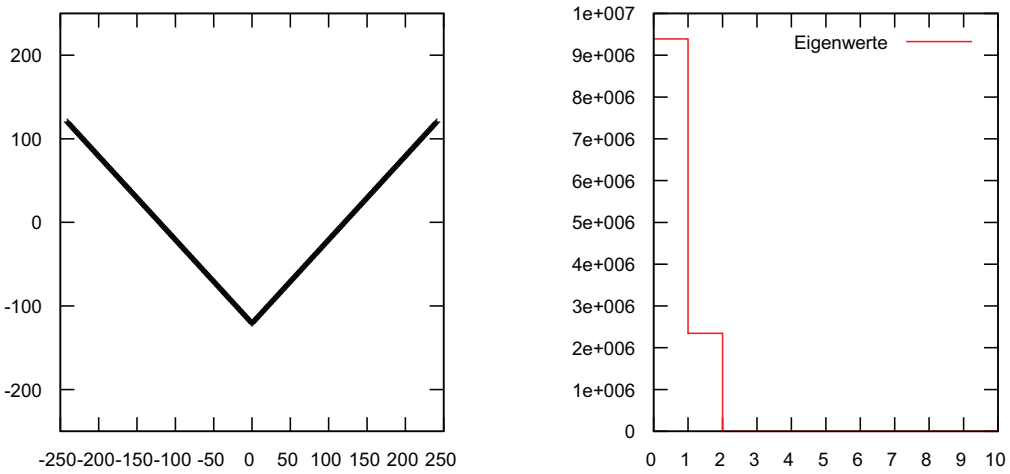
Die Dreiecksungleichung für de, und damit die letzte Bedingung, ist demnach erfüllt. Die angegebene Kombination zweier Metriken ist folglich ebenfalls eine Metrik. □

Die Abbildungen 3.15 bis 3.17 zeigen jeweils die Einbettung der drei Beispielbilder mittels der neuen Metrik *de*. Die Einbettungen ähneln für die Bilder mit einem und drei Leuchtpunkten sehr den Einbettungen unter Verwendung der Metrik d, insbesondere die Dimensionalität ist dieselbe. Der Stress ist allerdings beim Bild mit drei Leuchtpunkten auf 0 gesunken. In Abbildung 3.16 sieht man die Einbettung des Bildes mit zwei Leuchtpunkten. Diese unterscheidet sich signifikant von der Einbettung unter Verwendung der Metrik d, da sie zweidimensional ist. Mit der kombinierten Metrik *de* kann nun also zwischen allen drei Beispielen unterschieden werden.



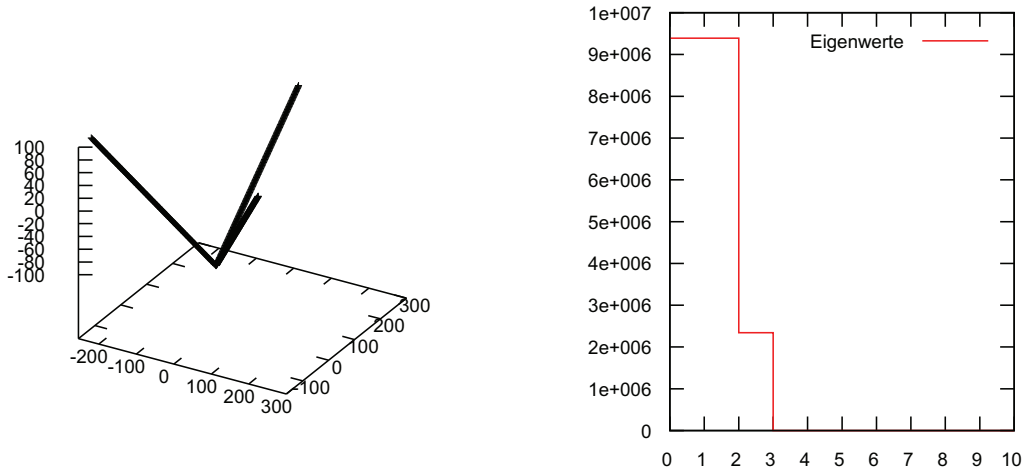
(a) (b)

Abbildung 3.15: Kombinierter Abstand auf dem Beispielbild, bestehend aus einem Leuchtpunkt auf dunklem Hintergrund: (a) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$, (b) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$



(a) (b)

Abbildung 3.16: Kombinierter Abstand auf dem Beispielbild, bestehend aus zwei Leuchtpunkten auf dunklem Hintergrund: (a) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$, (b) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$



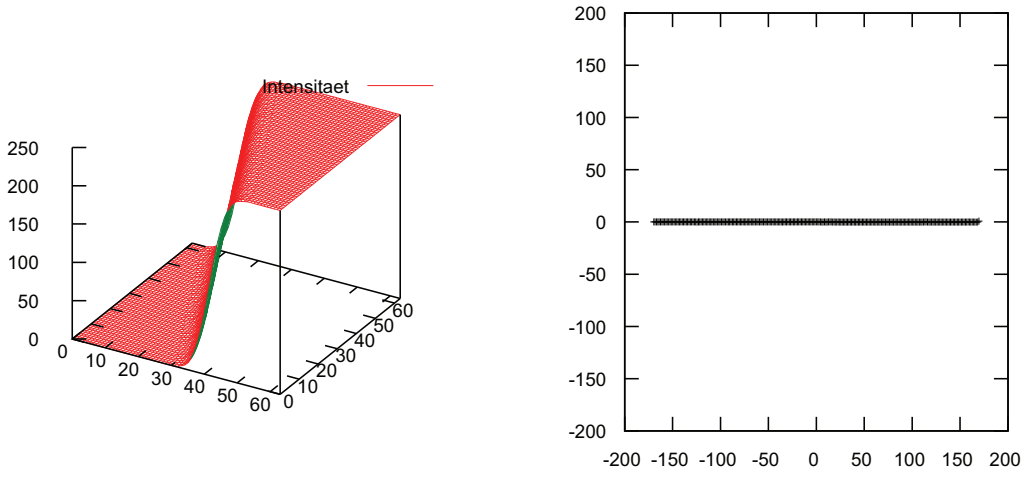
(a) (b)

Abbildung 3.17: Kombinierter Abstand auf dem Beispielbild, bestehend aus drei Leuchtpunkten auf dunklem Hintergrund: (a) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^3$, (b) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$

Eine Kante ist nach Einbettung in den Euklidschen Raum mittels multidimensionaler Skalierung bei $\alpha = 1$ äquivalent zu einem einzelnen Leuchtpunkt. Abbildung 3.18 (a) zeigt ein Bild einer Kante in dreidimensionaler Darstellung, Abbildung 3.18 (b) zeigt die Einbettung der Knoten im Euklidschen Raum. Die Dimensionalität einer Kante ist 1.

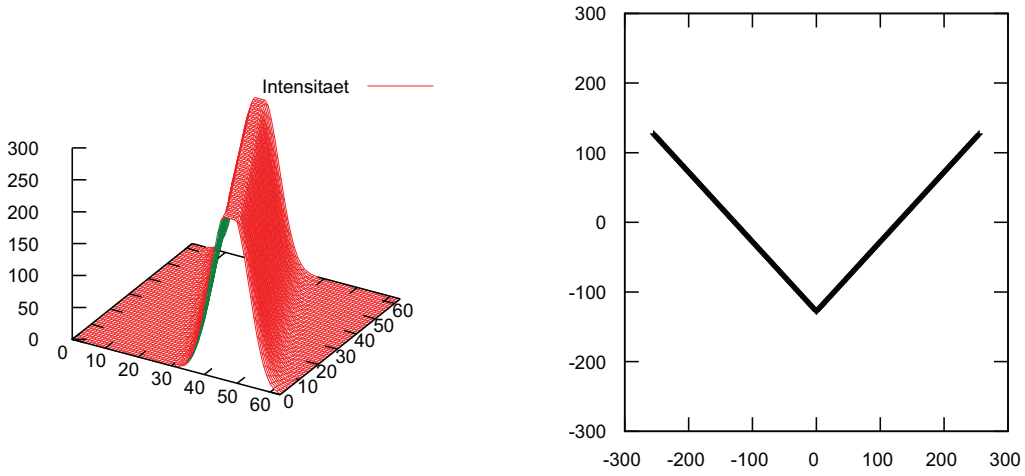
Der Graph einer Linienkante, dreidimensional dargestellt in Abbildung 3.19 (a), ist nach Einbettung in den Euklidschen Raum mittels multidimensionaler Skalierung äquivalent zu zwei Leuchtpunkten. Abbildung 3.19 (b) zeigt die Einbettung der Knoten. Die Dimensionalität einer Linienkante ist 2.

Es ist zu beobachten, dass die bisher untersuchten Graphen mit dem kombinierten Abstandsmaß *de* mit einem Stress von 0 in einen Euklidschen Raum eingebettet werden können. Auch bei Variation des α-Parameters bleibt diese Eigenschaft erhalten. Abbildung 3.20 zeigt anhand des Beispielbildes mit zwei Leuchtpunkten auf dunklem Hintergrund die Auswirkungen des α-Parameters auf die Einbettung des kürzesten Weges zwischen den beiden Leuchtpunkten in den Euklidschen Raum. Abbildung 3.20 (a) zeigt den Graphen des Beispielbildes mit dem kürzesten Weg zwischen den beiden Leuchtpunkten für $\alpha = 1$. Abbildung 3.20 (b) zeigt jeweils die ersten zehn Eigenwerte der Matrix $\mathbf{B}$ bei multidimensionaler Skalierung und Variation des α-Parameters. Die Eigenwerte werden logarithmiert dargestellt, damit auch die kleineren Eigenwerte besser erkennbar sind. Abbildung 3.20 (c) zeigt die ersten beiden Dimensionen der Einbettung derselben kürzesten Wege in den Euklidschen Raum. Der Stress bleibt jeweils bei 0, nur die Dimensionalität der Einbettung variiert. Sie ist 2 bei $\alpha = 0$ und $\alpha = 1$ und nimmt zu mittleren Werten von α zu, bis zu einer Dimensionalität von 4.



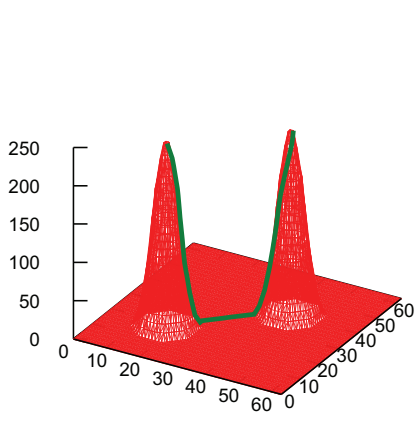
(a) (b)

Abbildung 3.18: Kombinierter Abstand auf dem Bild einer Kante: (a) dreidimensionale Darstellung unter Verwendung der Intensität als z-Koordinate, (b) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$

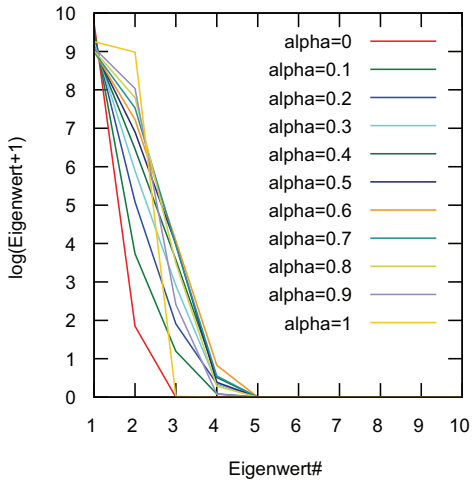


(a) (b)

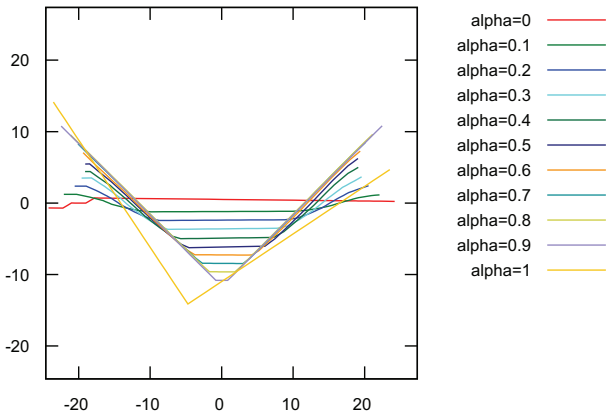
Abbildung 3.19: Kombinierter Abstand auf dem Bild einer Kante: (a) dreidimensionale Darstellung unter Verwendung der Intensität als z-Koordinate, (b) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$



(a)

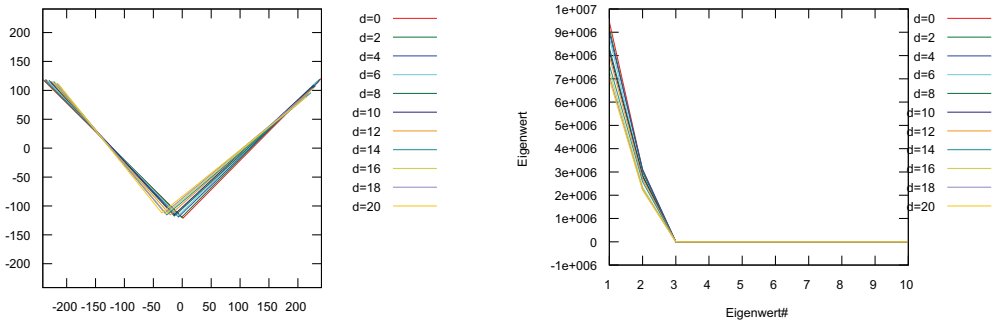


(b)



(c)

Abbildung 3.20: Graph des Bildes mit zwei Leuchtpunkten auf dunklem Hintergrund: (a) dreidimensionale Darstellung des Graphen unter Verwendung der Intensität z-Koordinate (in rot) mit kürzestem Weg zwischen den beiden Leuchtpunkten für $\alpha = 1$ (in grün), (b) Logarithmus der ersten 10 Eigenwerte e_i ($\log(e_i + 1)$) bei Variation des α-Parameters und Einbettung des kürzesten Weges zwischen beiden Leuchtpunkten in den Euklidschen Raum durch multidimensionale Skalierung, (c) die ersten beiden Dimensionen der Einbettung des kürzesten Weges in den Euklidschen Raum bei Variation des α-Parameters



(a) (b)

Abbildung 3.21: Graph des Bildes mit zwei Leuchtpunkten auf dunklem Hintergrund: (a) Die ersten 10 Eigenwerte bei Variation der Standardabweichung d des additiven Rauschens und Einbettung des kürzesten Weges zwischen beiden Leuchtpunkten in den Euklidschen Raum durch multidimensionale Skalierung, (b) die ersten beiden Dimensionen der Einbettung des kürzesten Weges in den Euklidschen Raum bei Variation der Standardabweichung d des additiven Rauschens

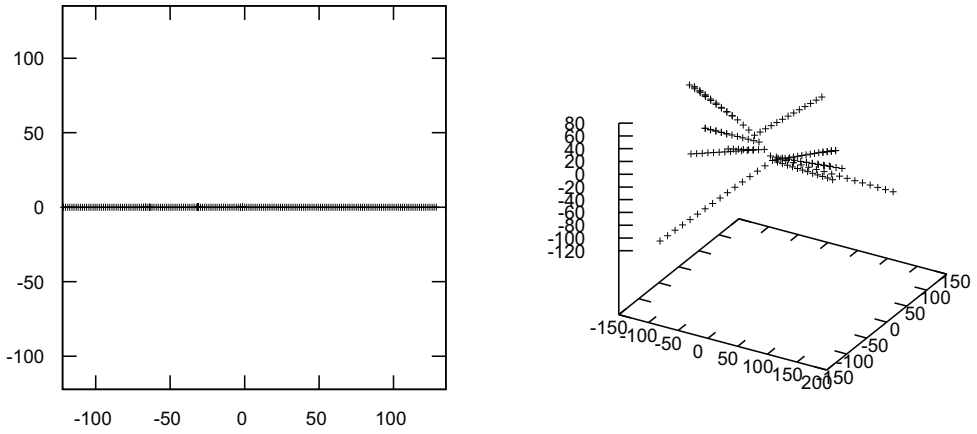
In dieser Hinsicht ist die oben definierte Metrik de besser ist als die zuvor definierte Metrik d, da sie eine bessere Einbettung in den Euklidschen Raum ermöglicht. Es stellt sich die Frage, ob dies für alle möglichen Graphen gilt, d.h., ermöglicht die Metrik de für jeden beliebigen Graphen eine perfekte (mit $S^2 = 0$) Einbettung in den Euklidschen Raum? Nach Rencher [34, S. 505] ist dies äquivalent zu der folgenden Frage: Sei (de_{ab}) die Matrix der Abstände zwischen Paaren von Knoten des Graphen $(a, b \in V)$. Ist dann die Matrix $\mathbf{B} = \left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\left(-\frac{1}{2}de_{ab}^2\right)\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)$ mit $n = |V|$ (s. auch Gleichung 3.20) positiv semidefinit? Um diese Frage zu beantworten, werden im Folgenden die Auswirkungen von Rauschen auf die Einbettungen untersucht.

3.4.2 Einfluss von Rauschen

Betrachtet wird der Einfluss von additivem Gaußschem Rauschen auf die Abstände im Graphen und damit die Einbettung in den Euklidschen Raum mittels multidimensionaler Skalierung. Zu den Beispielbildern wird jeweils ein zufällig erzeugtes Rauschbild mit dem Mittelwert 0 und gegebener Standardabweichung addiert. Dabei wird die Standardabweichung, also die Stärke des Rauschens, schrittweise erhöht.

Zunächst wird das zuvor verwendete Beispielbild betrachtet, bestehend aus zwei Leuchtpunkten auf dunklem Hintergrund. Abbildung 3.21 zeigt den Einfluss von additivem Rauschen auf den kürzesten Weg zwischen beiden Leuchtpunkten. Die Standardabweichung (d in der Abbildung) variiert hier von 0 (kein Rauschen) bis 20 (starkes Rauschen). Der Einfluss des Rauschens auf den kürzesten Weg zwischen den beiden Leuchtpunkten ist demnach sehr gering. Der Stress S^2 ist ebenfalls jeweils 0.

Zum Vergleich wird nun ein Rauschbild betrachtet. Betrachtet man einen kürzesten



(a) (b)

Abbildung 3.22: Einbettungen eines Rauschbildes in den Euklidschen Raum: (a) zweidimensionale Darstellung der Einbettung des kürzesten Weges zwischen zwei Punkten des Rauschbildes, Stress $S^2 = 0$, (b) dreidimensionale Darstellung der Einbettung des gesamten Rauschbildes, Stress $S^2 = 0.02$

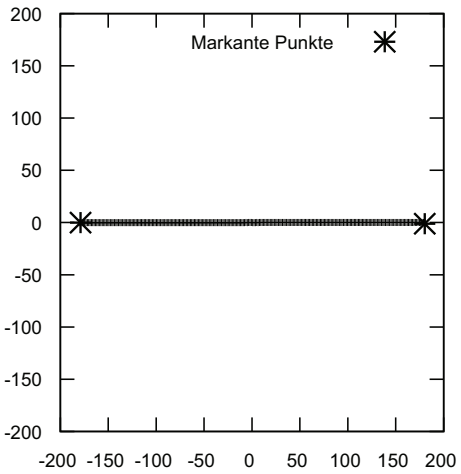
Weg zwischen zwei Punkten eines Rauschbildes, so ergibt sich eine eindimensionale Struktur. Der resultierende Stress liegt bei 0. Abbildung 3.22 (a) zeigt diese Einbettung. Betrachtet man allerdings alle Punkte des Rauschbildes und versucht, sie gemeinsam in einem Euklidschen Raum einzubetten, so ist dies nicht mehr perfekt möglich. Der Stress ist größer 0 und liegt in diesem Beispiel bei $S^2 = 0.02$. Abbildung 3.22 (b) zeigt die ersten drei Dimensionen der resultierenden Einbettung.

Während also Rauschen die kürzesten Wege zwischen Bildstrukturen nicht stark beeinträchtigen, können die Rauschpunkte selbst nicht alle perfekt in den Euklidschen Raum eingebettet werden. Weiterhin folgt, dass die Matrix B der Abstände unter der Metrik *de* (s. Gleichung 3.20) i.A. nicht positiv-semidefinit ist (für die untersuchten rauschfreien Bildstrukturen ist sie dies allerdings).

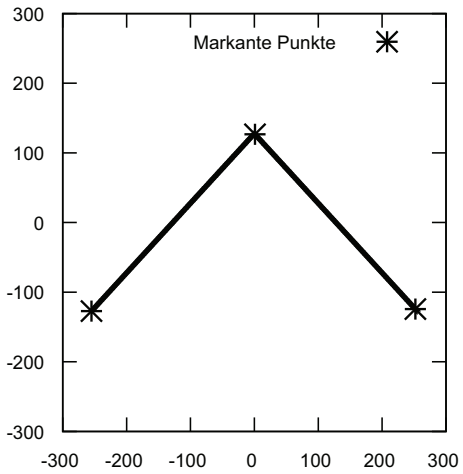
3.4.3 Markante Punkte/Points-of-Interest

Die untersuchten Bilder, bzw. deren Einbettungen in den Euklidschen Raum unter der Graphmetrik *de*, zeigen einige markante Punkte. Abbildung 3.23 zeigt jeweils die markanten Punkte in den Einbettungen. Im Bild entsprechen diese den Leuchtpunkten und dem dunklen Hintergrund. In den Einbettungen entsprechen die Leuchtpunkte immer *Extrempunkten*, während der dunkle Hintergrund in der Einbettung des ersten Bildes einem Extrempunkt, im zweiten Bild einem 90°-Knick und im dritten Bild einer Kreuzung (ebenfalls im 90 °-Winkel) entspricht. Diese Punkte werden im Folgenden *Sattelpunkte* genannt.

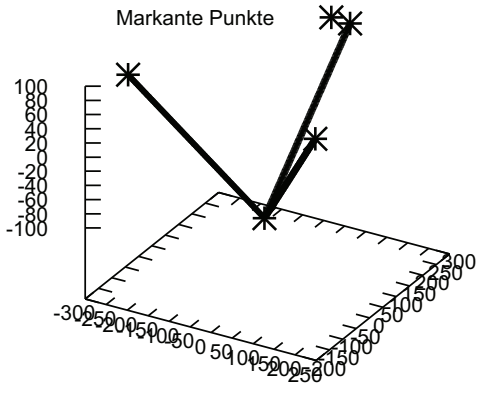
Im folgenden Abschnitt wird ein Algorithmus vorgestellt, wie Extrempunkte und Sat-



(a)



(b)



(c)

Abbildung 3.23: Markante Punkte in den drei Beispielbildern bzw. deren Einbettungen im Euklidschen Raum. Die markanten Punkte entsprechen jeweils dem/den Leuchtpunkte(n) und dem Hintergrund: (a) Ein Leuchtpunkt (hervorgehoben sind zwei Extrempunkte), (b) zwei Leuchtpunkte (hervorgehoben sind zwei Extrempunkte und ein Sattelpunkt), (c) drei Leuchtpunkte (hervorgehoben sind drei Extrempunkte und ein Sattelpunkt)

telpunkte effizient gefunden werden können, ohne jeweils den gesamten Graph mittels multidimensionaler Skalierung in den Euklidschen Raum einbetten zu müssen. Dieser Algorithmus stellt den in dieser Arbeit eingeführten Points-of-Interest-Detektor dar.

3.5 Point-of-Interest-Detektor

Ziel dieses Abschnitts ist es, einen Algorithmus zu definieren, der die in den vorangegangenen Abschnitten vorgestellten markanten Punkte bzw. Points-of-Interests in Graphen bestimmt. Es wird ein effizienter Algorithmus für $\alpha = 1$ vorgestellt. Dieser teilt sich in zwei Teile: Einen Teil zur Bestimmung der Extrempunkte und einen Teil zur Bestimmung der Sattelpunkte.

3.5.1 Algorithmus zur Bestimmung von Extrem- und Sattelpunkten

Zunächst werden die Begriffe des Extrem- und Sattelpunkts definiert:

Definition 11. Ein Extrempunkt wird definiert als ein Knoten, der nur einen direkten Nachbarn hat.

Ein Sattelpunkt wird definiert als ein Knoten, der zwei direkte Nachbarn besitzt, deren Euklidsche und geodätische Distanz sich unterscheiden.

Sei $G = (V, E, w, de)$ ein Graph mit den Knoten V, den Kanten E, den Kantengewichten $w : E \longrightarrow \mathbb{R}$ und der in Abschnitt 3.4.1 vorgestellten Metrik $de : V \times V \longrightarrow \mathbb{R}$ mit $\alpha = 1$. Die Abtastung der Intensitäten (s. Abschnitt 3.3.1) entspreche außerdem genau der Intensitätsauflösung. Wie oben definiert, entsprechen die Extrempunkte X genau den Knoten im Graph, die nur einen Nachbarknoten, d.h. einen Grad von 1, haben. Die Bestimmung der Extrempunkten kann also durch einfaches Zählen der Nachbarn eines Knoten geschehen.

Die Sattelpunkte S sind Knoten, die entweder

1. genau zwei Nachbarknoten haben, deren geodätische und Euklidsche Distanz sich unterscheiden (d.h. die drei Knoten liegen eingebettet im Euklidschen Raum nicht auf einer Geraden), oder

2. mehr als zwei Nachbarknoten haben.

Satz 7. *Der oben genannte Algorithmus findet genau die Extrem- und Sattelpunkte des gegebenen Graphen G.*

Beweis 7. *Der Beweis, dass der Algorithmus die Extrempunkte findet, ergibt sich aus der Definition.*

Für die Sattelpunkte ist zu zeigen: Wenn ein Knoten v mehr als zwei direkte Nachbarknoten hat, dann gibt es zwei Nachbarknoten, deren Euklidsche und geodätische Distanz sich unterscheiden. Seien n_1 und n_2 zwei direkte Nachbarknoten, deren geodätische und Euklidsche Distanz sich nicht unterscheiden. Da die Abtastung der Intensitäten der Intensitätsauflösung entspricht und $\alpha = 1$ heißt dies, dass einer der beiden eine höhere Intensität und der andere eine niedrigere Intensität hat als v. Sei (ohne Beschränkung der Allgemeinheit) n_1 der Knoten mit der geringeren Intensität. Sei n_3 ein beliebiger anderer Nachbarknoten, dessen geodätische und Euklidsche Distanz von n_1 sich unterscheiden. Daraus folgt, dass die Intensität von n_3 höher ist als die von v und damit gleich der Intensität von n_2. Die Euklidsche Distanz von n_2 und n_3 ist demnach 0. Da

$n_3 \neq n_2$ folgt, dass die geodätische Distanz zwischen n_2 und n_3 größer ist als 0 und sich demnach von der Euklidschen Distanz unterscheidet. □

3.5.2 Rauschunterdrückung

Wie in Abschnitt 3.4.2 gezeigt, beeinflusst Rauschen die kürzesten Wege innerhalb eines Graphen, und damit die Sattelpunkte, nur in geringem Maße. Einzelne Rauschpunkte werden jedoch vom oben angegebenen als Extrempunkte identifiert. In einem weiteren Schritt werden daher Extrempunkte, die Rauschen entsprechen könnten, mittels eines Rauschschwellwerts r entfernt:

1. Bestimme die Extrempunkte X und die Sattelpunkte S nach dem o.a. Algorithmus
2. Suche alle Extrempunkte $v_i \in X$, deren geodätische Distanz zum jeweils nächsten Sattelpunkt $v_i' \in S$ kleiner oder gleich dem Schwellwert r ist:

$$\{v_i\} = \{v \in X | \exists v' \in S : de(v, v') \leq r\}$$

3. Entferne alle v_i aus dem Graphen G
4. Wenn $|\{v_i\}| > 0$, d.h., wenn Extrempunkte entfernt wurden, fahre mit Punkt 1 fort, unter Verwendung des modifizierten Graphen

3.5.3 Filterung nicht-lokaler Extrem- und Sattelpunkte

Da die Extrem- und Sattelpunkte des Graphen G jeweils mehreren Pixeln im Ursprungsbild entsprechen können (s. Abschnitt 3.3.2), sind sie u.U. nicht mehr lokaler Natur und damit nicht als Point-of-Interest geeignet. Es werden daher in einem letzten Schritt alle Extrem- und Sattelpunkte herausgefiltert, deren korrespondierende Pixelmenge einen geringen Anteil der Gesamtanzahl an Pixeln des Ursprungsbildes nicht überschreitet. So wird sichergestellt, dass ein Extrem- bzw. Sattelpunkt immer einer lokal begrenzten Region des Ursprungsbildes zugeordnet werden kann.

4 Evaluation

In diesem Kapitel wird die Deformationsinvarianz des vorgestellten Detektors evaluiert. Dies geschieht durch Anwendung auf eine Menge von Testbildern bzw. Bildpaaren und Messung der Stabilität des Detektors. Die Bilder werden dabei mittels einer Menge von Deformationen variierender Stärke und additivem Rauschen transformiert.

Durch die Evaluation soll die Hypothese, dass der vorgestellte Detektor deformationsinvariant ist, getestet werden. Dabei ist zu beachten, dass der in Abschnitt 3.2.1 durchgeführte Beweis der Deformationsinvarianz nur im kontinuierlichen Fall gilt.

Für den bei der praktischen Anwendung vorliegenden diskreten Fall kann ein solcher Beweis nicht geführt werden. Dies wird klar, wenn man ein einfaches Gegenbeispiel betrachtet: Ein Bild bestehend aus zwei Pixeln unterschiedlicher Grauwerte enthält zwei Extremwerte. Verkleinert man dieses Bild auf eine Größe von einem Pixel, reduziert sich dies auf einen (Extrem-)Wert. Die Verkleinerung ist eine spezielle Deformation. Da sich die Extremwerte unter dieser ändern, ist dies offensichtlich ein Gegenbeispiel für die Deformationsinvarianz im diskreten Fall.

Da ein formaler Beweis also nicht geführt werden kann, soll nun durch empirische Tests festgestellt werden, inwiefern die Deformationsinvarianz auf den diskreten Fall übertragen werden kann. Durch das oben angegebene Gegenbeispiel wird klar, dass die Deformationsinvarianz für den diskreten Fall i.A. nicht gilt. Durch die Tests soll gemessen werden, inwieweit dies auch Auswirkungen in der praktischen Anwendung hat.

Auf jedem Testbild werden mittels des im vorangegangenen Kapitel vorgestellten Detektors Points-of-Interest detektiert. Das Testbild wird dann einer Reihe von Transformationen unterzogen. Auf dem Ergebnis einer Transformation werden ebenfalls Points-of-Interest detektiert. Die resultierenden Points-of-Interest werden anschließend verglichen. Da die zugrundeliegende Transformation jeweils bekannt ist, kann jeder Point-of-Interest des transformierten Testbildes auf den erwarteten Point-of-Interest des Originalbildes abgebildet werden. Zu jeder Transformation werden drei Stabilitätswerte gemessen: *Precision* (Genauigkeit), *Recall* (Trefferquote) und das *F-Maß*. Diese Maße liegen jeweils im Intervall $[0, 1]$ und sind definiert durch:

$$\text{Recall} = \frac{|R \cap P|}{|R|} \tag{4.1}$$

$$\text{Precision} = \frac{|R \cap P|}{|P|} \tag{4.2}$$

$$\text{F-Maß} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.3}$$

Dabei entspricht R der Menge der Pixel der (transformierten) Points-of-Interest des Testbildes, also den erwarteten Points-of-Interest. Die Menge P entspricht der Menge der

Pixel der auf dem transformierten Testbild detektierten Points-of-Interest. Die Precision entspricht also dem Anteil der auf dem transformierten Testbild als Points-of-Interest identifizierten Pixel, der auch im Testbild selbst gefunden wurde. Der Recall dagegen entspricht dem Anteil der auf dem Testbild als Points-of-Interest identifizierten Pixel, der auch auf dem transformierten Testbild gefunden wurde. Das F-Maß entspricht dem harmonischen Mittel von Precision und Recall. Der Maximalwert von 1 entspricht jeweils dem Optimum.

Die Transformationen werden jeweils in mehreren Varianten durchgeführt: Von geringer Stärke, und damit geringer Änderung des Eingabebildes, bis zu hoher Stärke und starker Veränderung des Eingabebildes. Durch Messung der Stabilitätswerte kann beurteilt werden, inwiefern der Detektor auch im diskreten Fall deformationsinvariant ist. Ist er dies, so erwarten wir, dass die gemessenen Werte für Precision, Recall und F-Maß bei Erhöhung der Transformationsstärke konstant bleiben. Ist der Detektor dagegen nicht deformationsinvariant, wird eine starke Abnahme der Stabilitätswerte erwartet.

Die Arbeit von Moreels und Perona [33] beinhaltet abgesehen von den Testbildern, welche auch in dieser Arbeit verwendet werden, eine Evaluation bestehender Detektoren. Ein direkter Vergleich mit ihren Ergebnissen ist in dieser Arbeit jedoch nicht möglich, und zwar aus den folgenden Gründen: Die Voraussetzung des Lambertschen Beleuchtungsmodells, die dem hier vorgestellten Detektor zugrundeliegt, ist in der Evaluation von Moreels und Perona nicht gegeben. Zudem unterliegen die Objekte dort starken Verdeckungen, was die zweite grundlegende Annahme in dieser Arbeit verletzt. Diese beiden Faktoren würden die Evaluation beeinflussen und verhindern, dass die aufgestellte Hypothese anhand des oben beschriebenen Vorgehens geprüft werden kann. Es werden daher in dieser Arbeit dieselben Testbilder, jedoch nicht dieselben Transformationen verwendet.

Die hier durchgeführte Evaluation ist also keine vergleichende Evaluation. Es wird stattdessen versucht, die absolute Aussage zu treffen, dass der vorgestellte Detektor im diskreten Fall deformationsinvariant ist. Diese Aussage gilt als bestätigt, wenn die gemessenen Stabilitätswerte nicht von der Stärke der Deformationen abhängen, also bei unterschiedlichen Deformationen relativ konstant bleiben.

Die durchgeführten Transformationen werden im folgenden Abschnitt beschrieben. Die wichtigste Transformationsklasse ist dabei die der zufälligen Deformationen. Vor allem von den Messergebnissen dieser Klasse hängt die Bestätigung der aufgestellten Hypothese ab.

4.1 Transformationen

Die Transformationen, denen die Testbilder unterzogen werden, sind im Einzelnen: Skalierung, Rotation, Scherung, perspektivische Verzerrung und Hinzufügen von Gauß-Rauschen. Die ersten drei Transformationen bilden zusammen die Familie der affinen 2D-Transformationen. Die verwendete perspektivische Verzerrung dagegen simuliert eine 3D-Rotation der Bildebene. Die Parameter der Transformationen werden dabei jeweils innerhalb eines vorgegebenen Intervalls variiert. Die folgenden Abschnitte zeigen die Auswirkungen der verwendeten Transformationsparameter anhand eines Beispielgitters und eines Beispielbildes.

4.1.1 Skalierung

Die Testbilder werden jeweils in 10-Prozent-Schritten vergrößert bis zu einem Faktor von 2. Die Skalierung wird um den Mittelpunkt des Bildes herum durchgeführt, d.h. der Mittelpunkt bleibt stets an derselben Position. Abbildung 4.1 illustriert die Transformation anhand des Beispielgitters und des Beispielbildes.

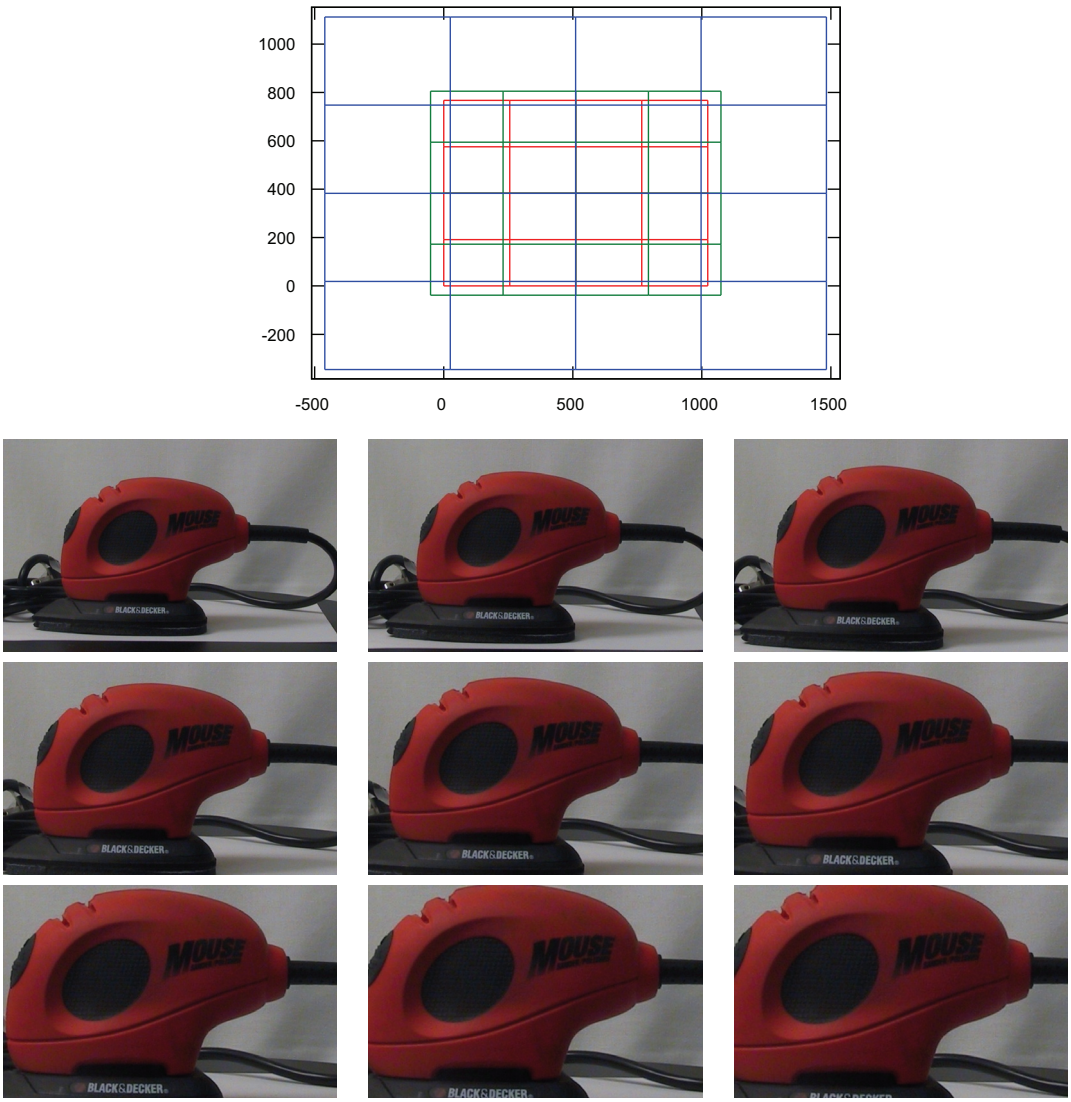


Abbildung 4.1: Skalierung des Beispielgitters (oben): Original (in rot), um 10 Prozent vergrößert (in grün), um den Faktor 1.9 (90 Prozent) vergrößert (in blau), Skalierung des Beispielbildes (unten) in 10-Prozent-Schritten von 20 bis 100 Prozent

4.1.2 Rotation

Die Testbilder werden in Schritten von 4.5 ° um den Bildmittelpunkt gedreht, bis zu einer Drehung von 90 °. Abbildung 4.2 zeigt die Rotationen anhand der Beispiele.

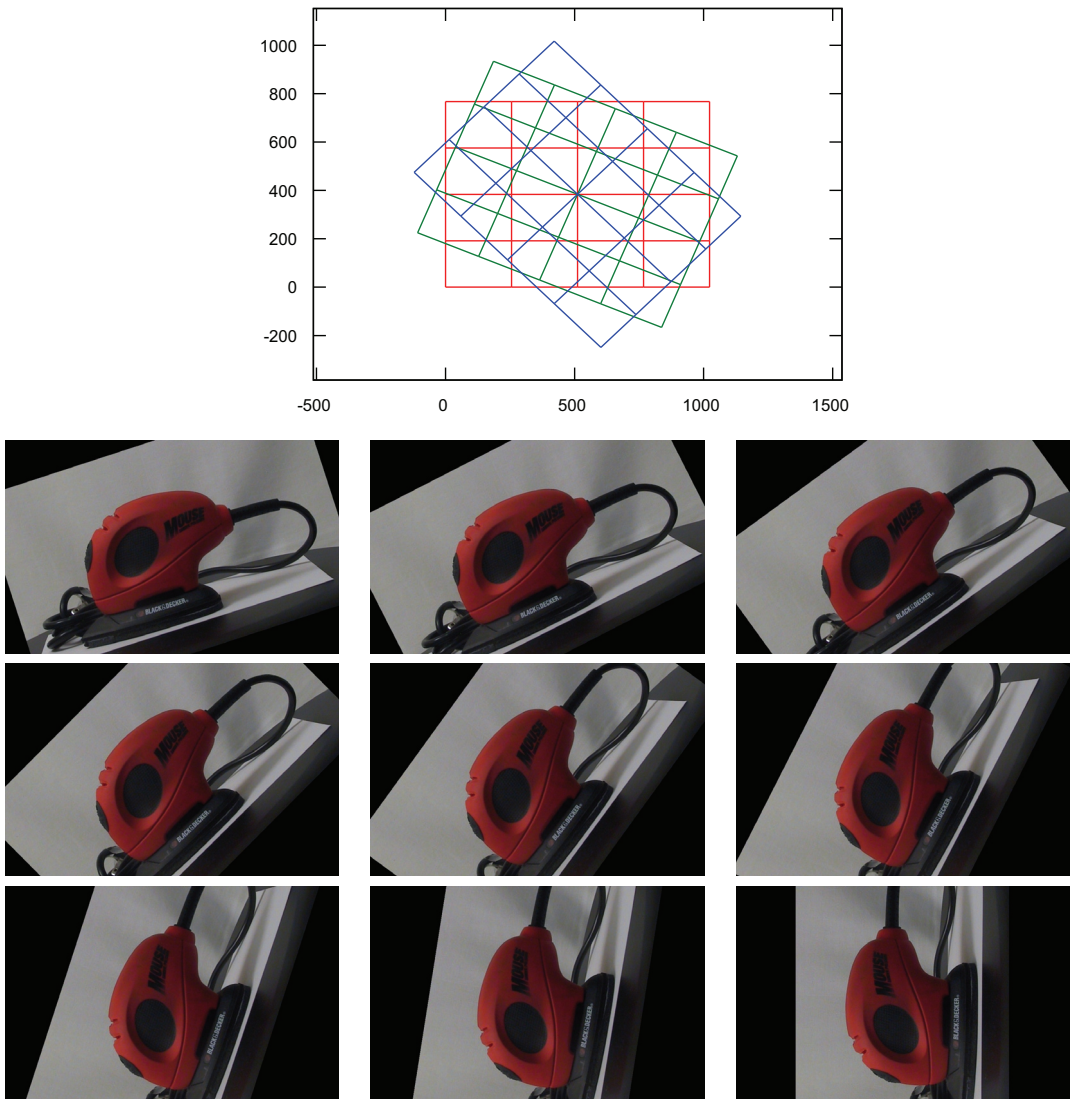


Abbildung 4.2: Rotation des Beispielgitters (oben): Original (in rot), um 22.5 ° gedreht (in grün), um 45 ° gedreht (in blau), Rotation des Beispielbildes (unten) in 9 °-Schritten von 18 ° bis 90 °

4.1.3 Scherung

Die Testbilder werden jeweils in 10-Prozent-Schritten nach rechts geneigt bis zu einer Neigung von 100 Prozent (dies entspricht 45 °). Abbildung 4.3 zeigt die Scherung anhand der Beispiele.

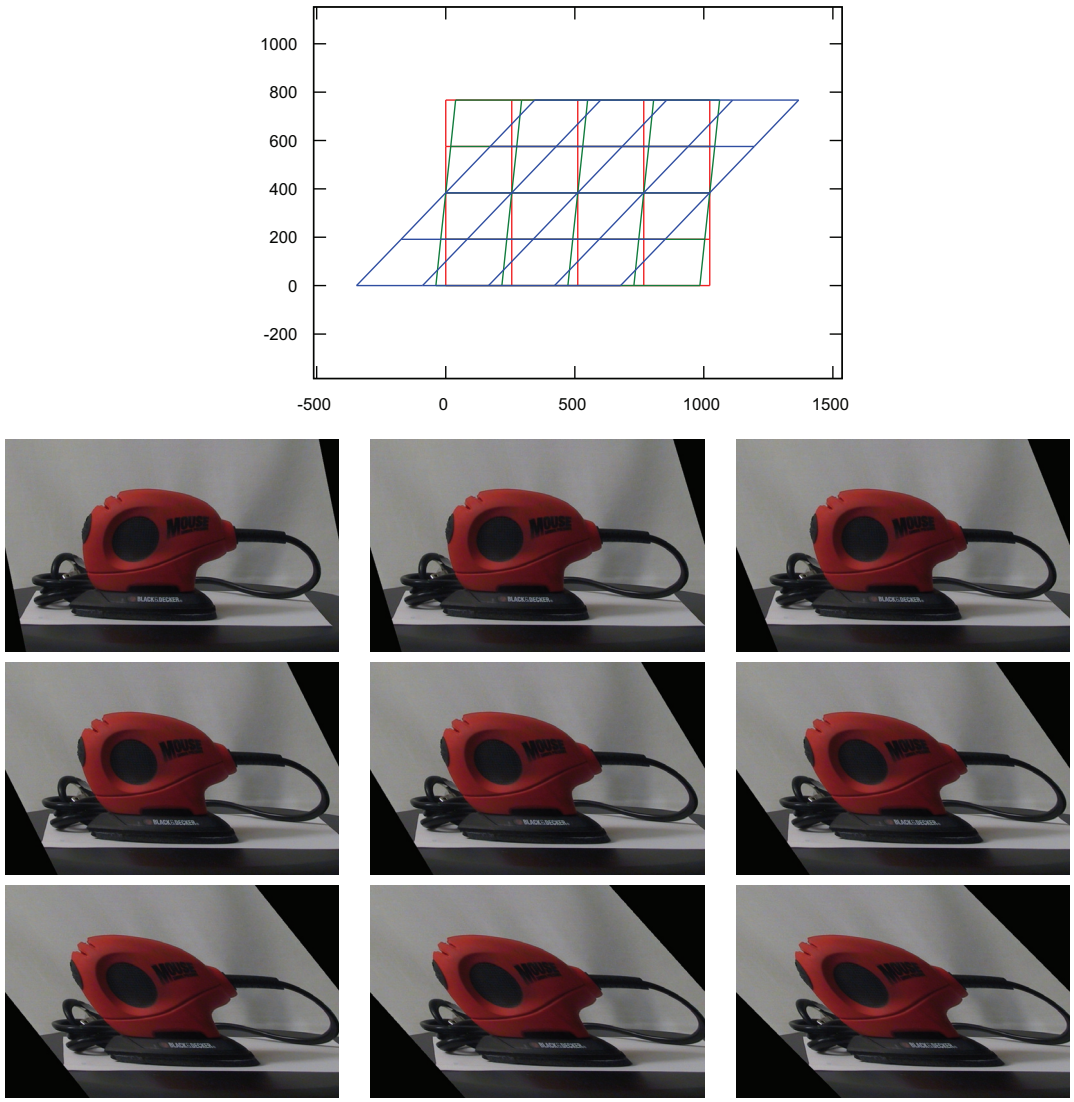


Abbildung 4.3: Scherung des Beispielgitters (oben): Original (in rot), um 10 Prozent geneigt (in grün), um 100 Prozenz (45 °) geneigt (in blau), Scherung des Beispielbildes (unten) in 10-Prozent-Schritten von 20 Prozent bis 100 Prozent

4.1.4 Perspektivische Verzerrung

Die Bildebene des Testbildes wird jeweils um die y-Koordinate des Bildes gedreht und gezerrt. Dabei wird die rechte Bildseite in die Tiefe vom Betrachter weg und die linke Bildseite in Richtung des Betrachters verschoben. Die rechte Seite des Bildes wird dabei kleiner, die linke Seite größer. Dies wird in 9 Schritten getan, bis die rechte Bildseite nur noch der Hälfte und die linke Bildseite dem Doppelten der ursprünglichen Größe entspricht. Abbildung 4.4 zeigt die perspektivische Verzerrung anhand der Beispiele.

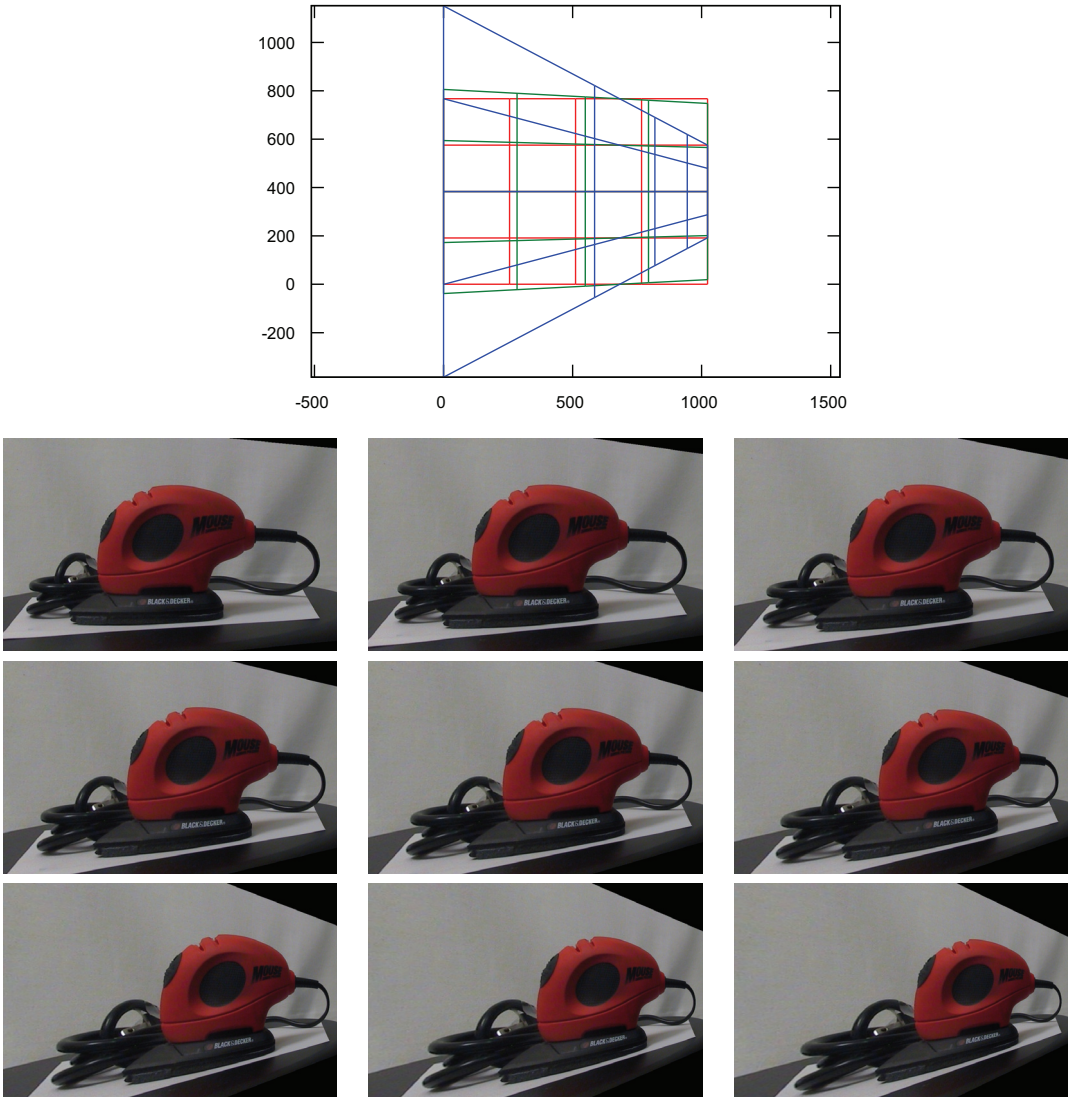


Abbildung 4.4: Perspektivische Verzerrung des Beispielgitters (oben): Original (in rot), geringste Verzerrung (in grün), maximale Verzerrung (in blau), perspektivische Verzerrung des Beispielbildes (unten), von minimaler Verzerrung bis maximaler Verzerrung

4.1.5 Zufällige Deformation

Um beliebige dreidimensionale Transformationen ohne Verdeckungen zu simulieren, werden die Testbilder zufälligen Deformationen unterzogen. Diese Deformationen werden anhand der Stützpunkte eines über das Bild gelegten Gitters durchgeführt. Die Stützpunkte werden mit gleichverteiltem Rauschen zunehmender Intensität verschoben. Die dazwischenliegenden Bildkoordinaten werden interpoliert. Abbildung 4.5 zeigt die resultierende Deformation anhand der Beispiele.

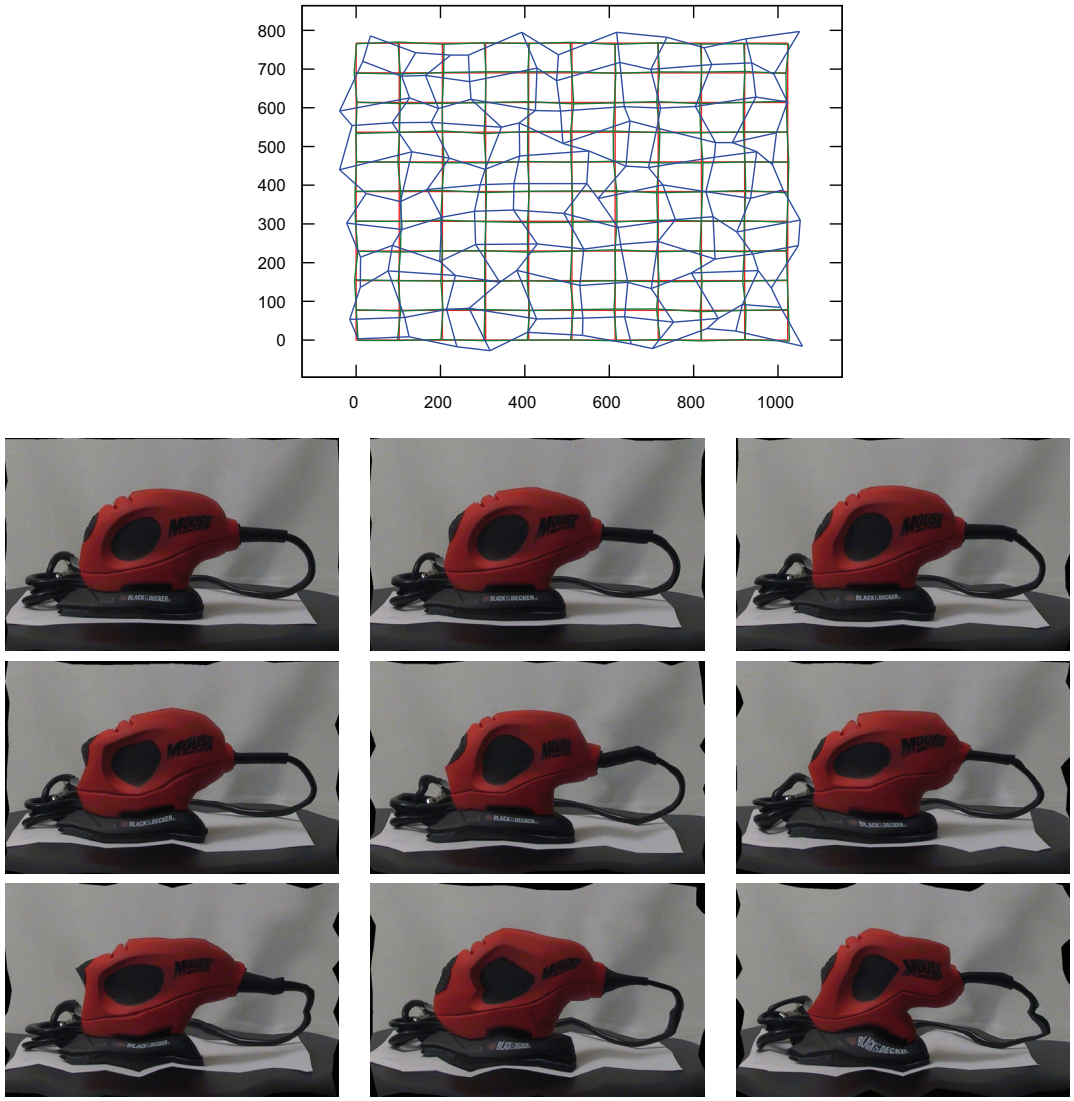


Abbildung 4.5: Zufällige Deformation des Beispielgitters (oben): Original (in rot), geringste Deformationsstufe (in grün), maximale Deformation (in blau), Deformation des Beispielbildes (unten), von minimaler Deformation bis maximaler Deformation

4.1.6 Hinzufügen von Gauß-Rauschen

Dem Testbild wird jeweils Gauß-Rauschen mit Mittelwert 0 und gegebener Standardabweichung hinzugefügt. Die Standardabweichung wird dabei in 10 Schritten von 1 bis 10 variiert. Abbildung 4.6 zeigt die Auswirkung des Rauschens an einem Detail des Beispielbildes.

(a) (b)

Abbildung 4.6: Hinzufügen von Gauß-Rauschen: (a) Detail des Beispielbildes (b) verrauschtes Detail des Beispielbildes, Standardabweichung des Gauß-Rauschens: 10

Jedes Testbild wird also einer Menge von 55 Transformationen unterzogen. Die Ergebnisse der Evaluierung werden im folgenden Abschnitt präsentiert.

4.2 Ergebnisse

In diesem Abschnitt werden die unter den oben angegebenen Transformationen gemessenen Werte von Precision, Recall und F-Maß angegeben und interpretiert. Es werden dabei jeweils die Mittelwerte und die Standardabweichung dieser Maße über alle Testbilder betrachtet. Außerdem werden jeweils die Testbilder mit dem besten und schlechtesten Ergebnis einer Transformation im Einzelnen betrachtet.

Die Points-of-Interest werden jeweils mit einem Rausch-Schwellwert von $r = 12$ detektiert. Abbildung 4.7 zeigt die Menge der Testbilder in Verkleinerung. Die Testbilder entstammen der Arbeit von Moreels und Perona [33], in der anhand dieser Bilder existierende Verfahren evaluiert wurden (s. Abschnitt 2.3).

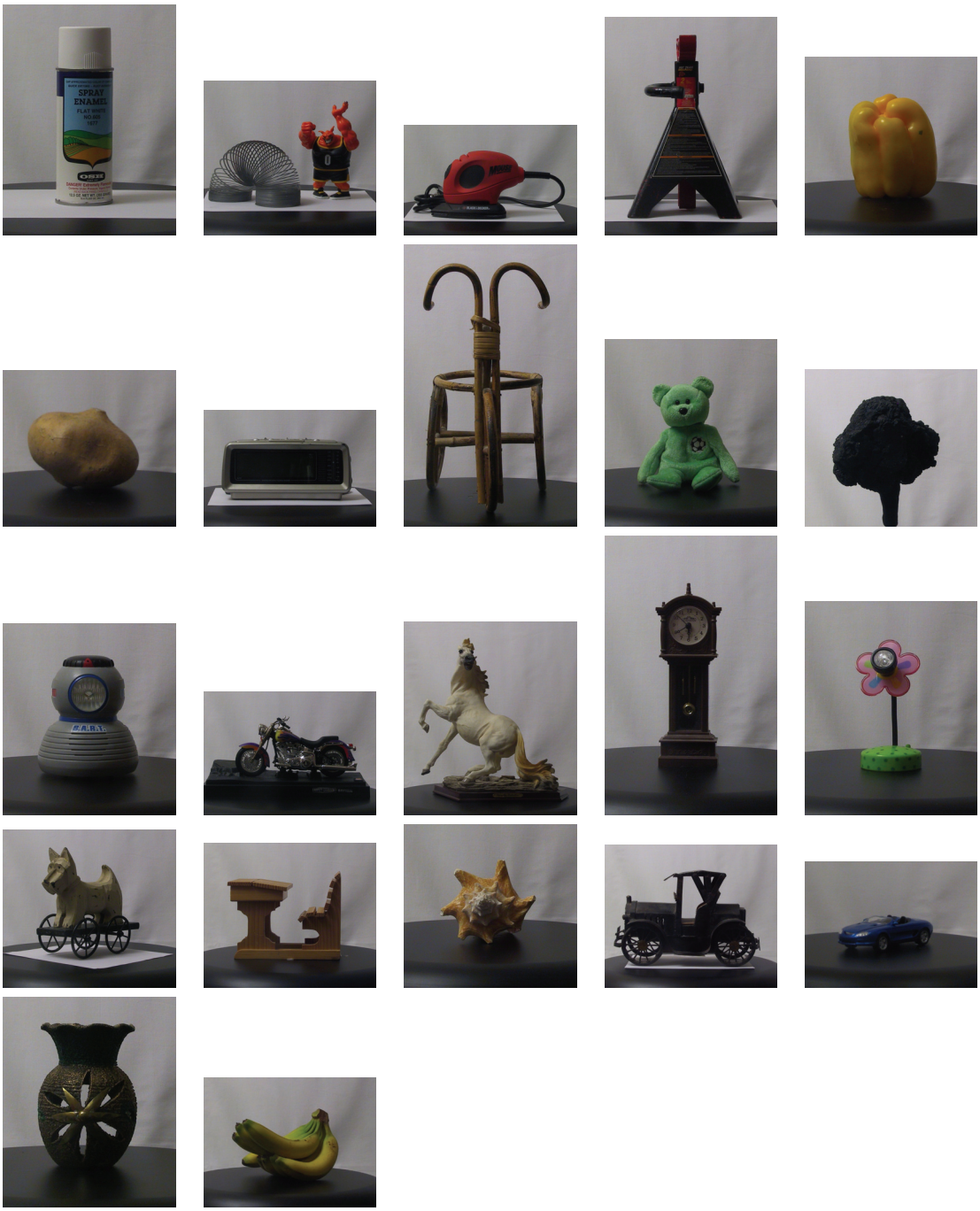


Abbildung 4.7: Testbilder (in Leserichtung): Spraydose, Monster und Spirale, Schleifgerät, Ständer, Paprika, Knolle, Radiowecker, Dreirad, Teddybär, Felsbrocken, Roboter, Motorrad, Pferd, Standuhr, Tischlampe, Holzhund, Schreibtisch, Muschel, Oldtimer, Modellauto, Bodenvase, Banane

4.2.1 Skalierung

Abbildung 4.8 zeigt die über alle Testbilder gemittelten Ergebnisse für die Skalierung. Die auf den skalierten Testbildern gefundenen Points-of-Interest sind zu großem Anteil korrekt (die Precision liegt durchgängig bei ca. 90 Prozent). Es werden allerdings unabhängig vom Skalierung jeweils nur ca. 70 Prozent aller auf dem Ursprungsbild detektierten Points-of-Interest wiedergefunden. Das gewichtete Mittel von Precision und Recall, das F-Maß, liegt relativ konstant bei 80 Prozent.

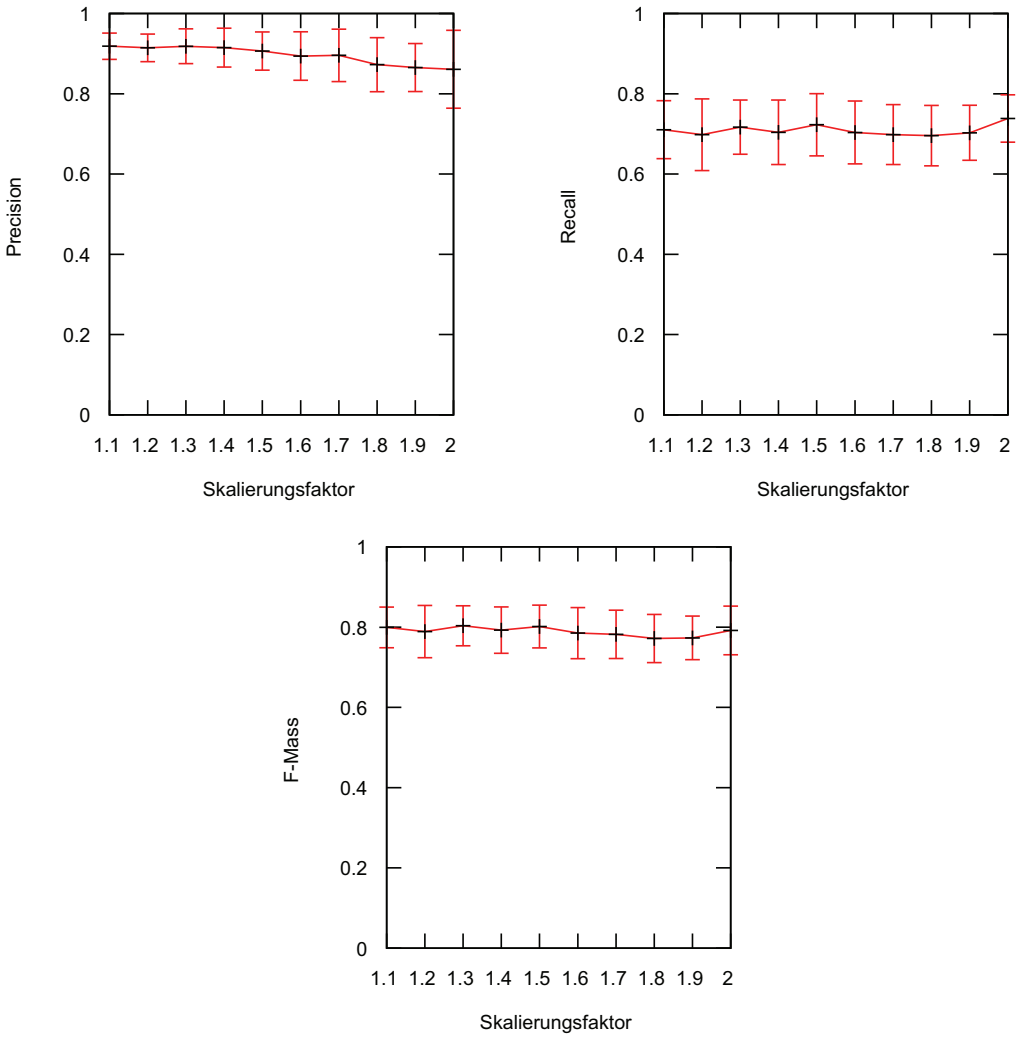


Abbildung 4.8: Stabilität des Detektors bei Skalierung: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über dem Skalierungsfaktor sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen

Abbildung 4.9 zeigt das F-Maß für alle Testbilder (aufgeteilt in zwei Gruppen zu je 11 Bildern). Die Ergebnisse sind auch für die einzelnen Bilder jeweils relativ unabhängig vom Skalierungsfaktor. Das mit Abstand schlechteste Ergebnis zeigt das Bild der Knolle (s. Abbildung 4.7 linke Spalte, zweites Bild von oben). Das beste Ergebnis zeigt das Bild des Modellautos (s. Abbildung 4.7 rechte Spalte unten).

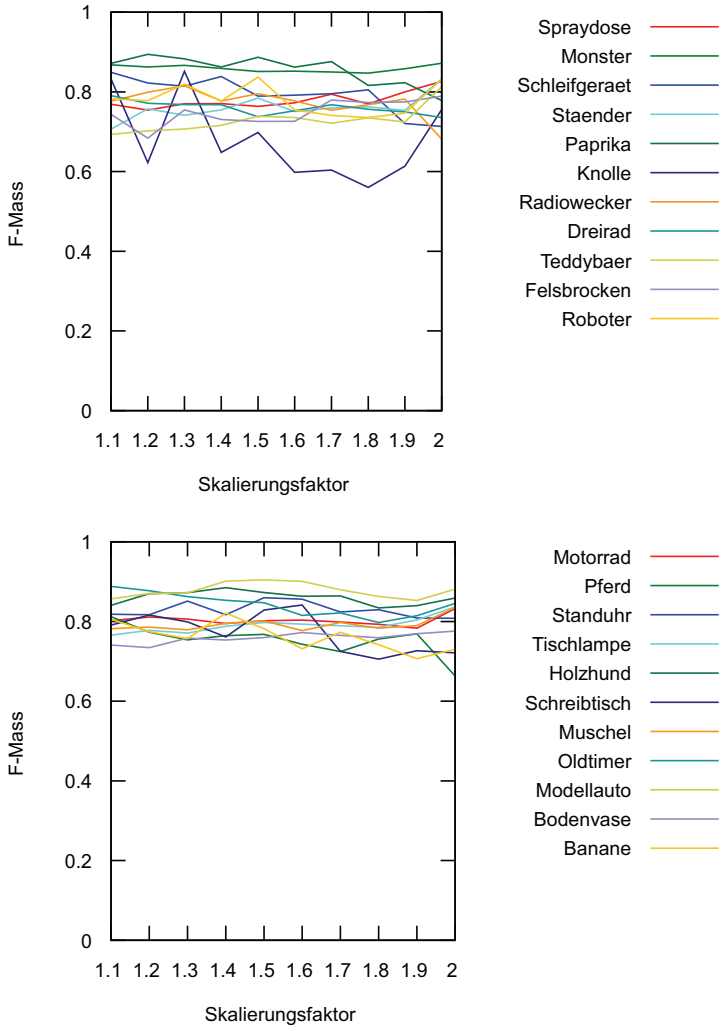


Abbildung 4.9: F-Maß der Stabilität des Detektors bei Skalierung für alle Testbilder

Für die Bilder mit dem besten bzw. schlechtesten Ergebnis sind in Abbildung 4.10 die Ergebnisse des Detektors für den Skalierungsfaktor 1.6 bzw. 1.2 dargestellt. Auf dem Bild der Knolle wurden im Ursprungsbild insgesamt nur sechs Points-of-Interest detektiert. Von diesen sechs wurden im skalierten Bild nur vier wiedergefunden. Außerdem wurde ein Point-of-Interest zu klein detektiert. Auf dem Bild des Modellautos wurden 32 Points-of-

Interest detektiert, von denen alle bis auf einen im skalierten Bild wiedergefunden wurden. Es wurde außerdem ein zusätzlicher Point-of-Interest gefunden, der im Ursprungsbild nicht detektiert wurde.

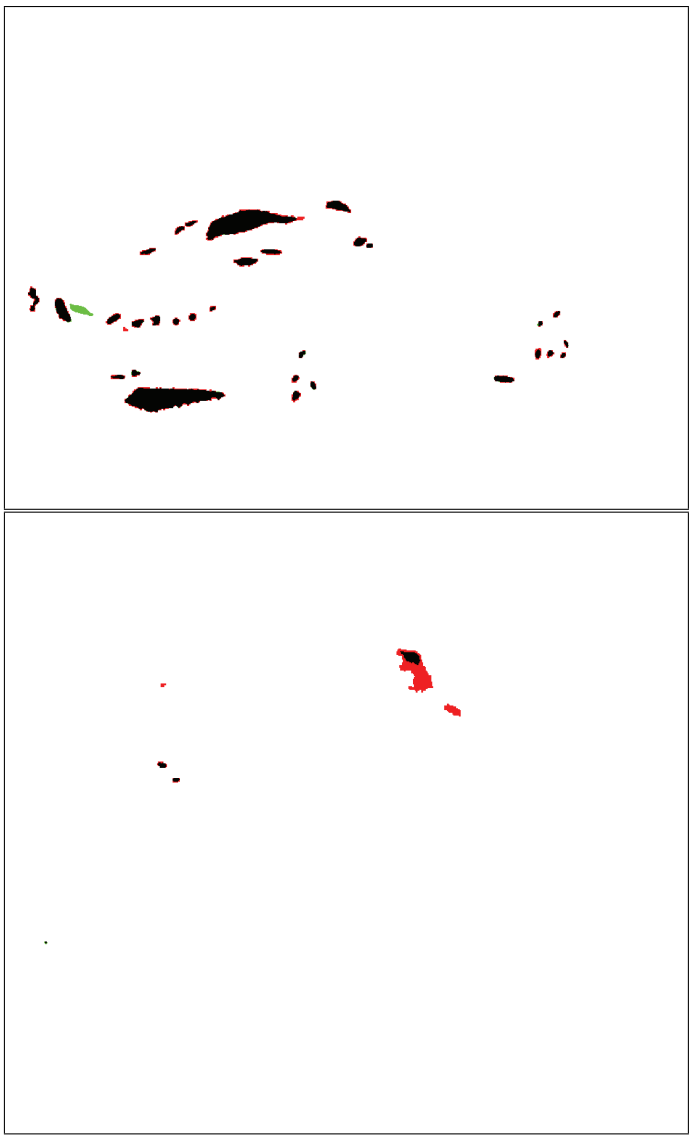


Abbildung 4.10: Ergebnis des Detektors für das Bild des Modellautos bei einem Skalierungsfaktor von 1.6 (oben) und für das Bild der Knolle bei einem Skalierungsfaktor von 1.2 (unten). Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im skalierten Bild, sind rot dargestellt. Punkte, die im skalierten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt.

4.2.2 Rotation

Abbildung 4.11 zeigt die über alle Testbilder gemittelten Ergebnisse für die Rotation. Auch hier zeigt sich eine höhere Precision als Recall und eine relativ Unabhängigkeit der Performanz vom Rotationswinkel. Dies trifft allerdings nicht für den 90 °-Winkel zu, für den deutlich bessere Werte gemessen werden als für alle anderen Winkel. Dies ist dadurch zu erklären, dass die Rotation um 90 ° äquivalent ist zu einer Vertauschung von x- und y-Achse und Spiegelung des Bildes, Operationen, durch die sich die Graphrepräsentation des Bildes nicht wesentlich verändert.

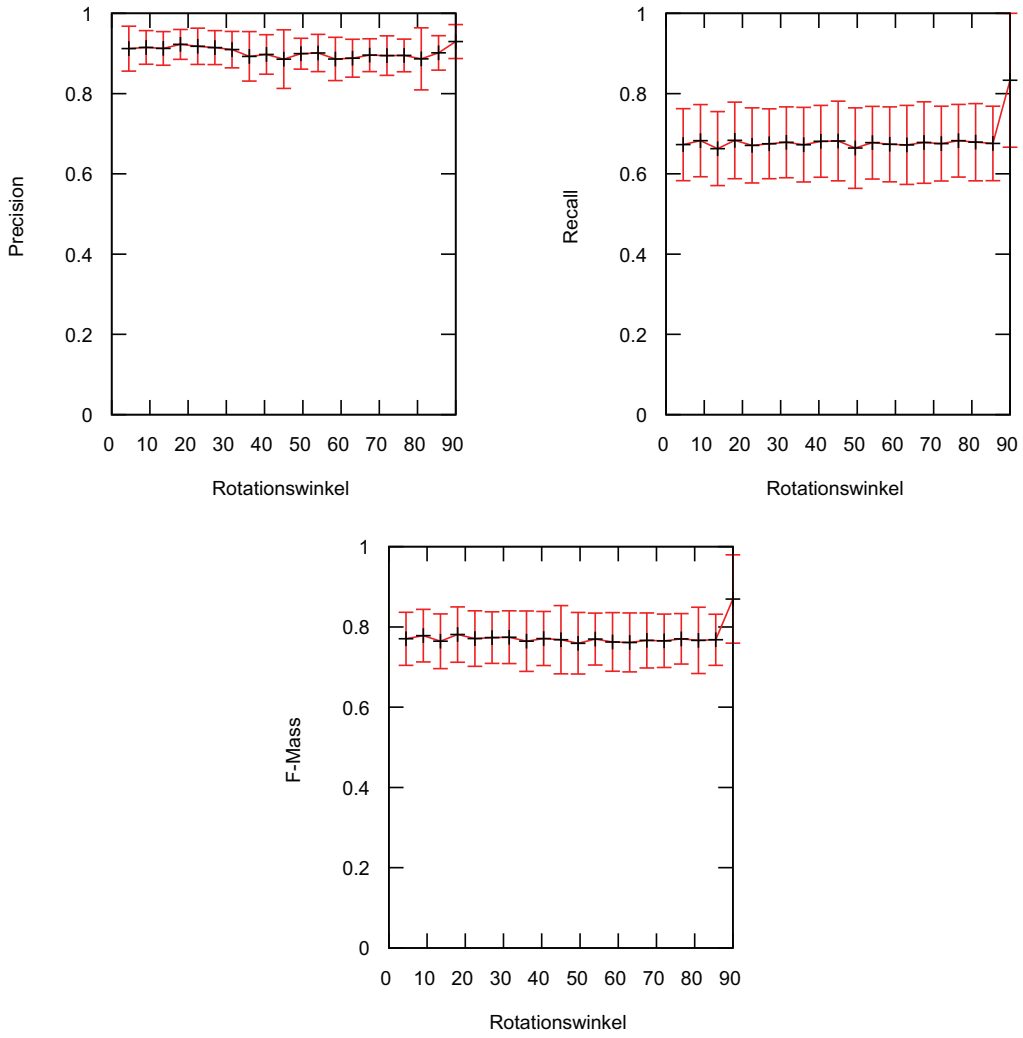


Abbildung 4.11: Stabilität des Detektors bei Rotation: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über dem Rotationswinkel sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen

Abbildung 4.12 zeigt das F-Maß der einzelnen Testbilder. Das Bild der Knolle zeigt hier ebenfalls mit Abstand das schlechteste Ergebnis, während das Monster mit Spirale (s. Abbildung 4.7 zweite Spalte, oberstes Bild) das beste Ergebnis zeigt. Wie bei der Skalierung sind die Ergebnisse auch bei den einzelnen Bildern relativ unabhängig von der Stärke der Transformation (hier dem Rotationswinkel).

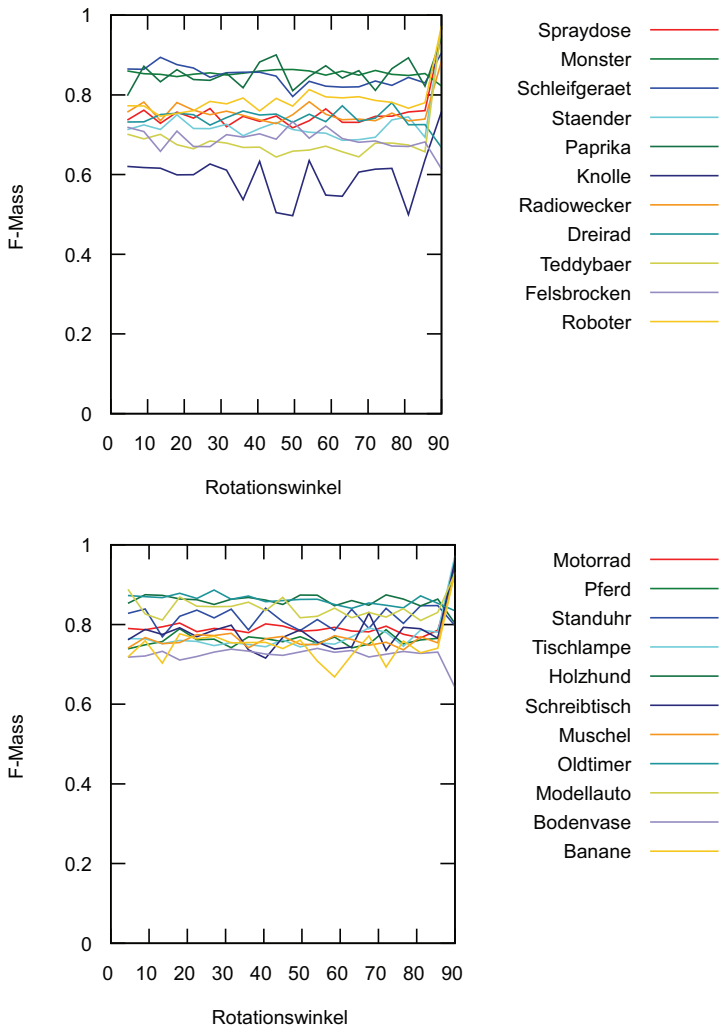


Abbildung 4.12: F-Maß der Stabilität des Detektors bei Rotation für alle Testbilder

Für die Bilder mit dem besten bzw. schlechtesten Ergebnis sind in Abbildung 4.13 die Ergebnisse des Detektors für den Rotationswinkel 45 ° dargestellt. Von den sechs im Ursprungsbild der Knolle detektierten Points-of-Interest wurden im rotierten Bild wieder nur vier wiedergefunden. Es wurden außerdem zwei weitere Points-of-Interest gefunden, die im Ursprungsbild nicht detektiert wurden. Auf dem Bild des Monsters mit Spirale

wurden aufgrund der großen Menge an Bildstrukturen viele Points-of-Interest gefunden, von denen die meisten auch im rotierten Bild detektiert wurden.

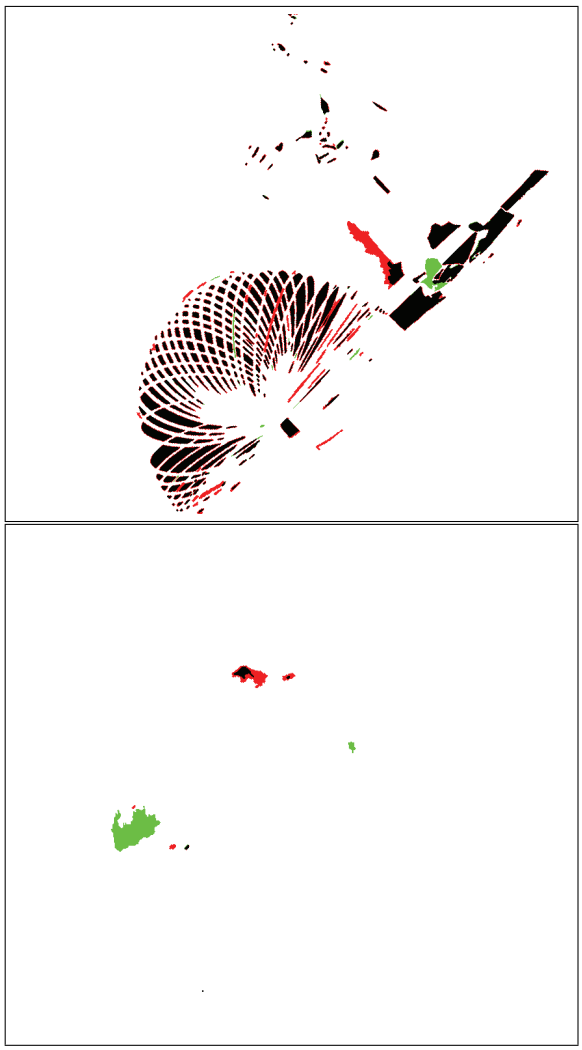


Abbildung 4.13: Ergebnis des Detektors für das Bild des Monsters mit Spirale (oben) und für das Bild der Knolle (unten) bei einem Rotationswinkel von 45 °. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im rotierten Bild, sind rot dargestellt. Punkte, die im rotierten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt.

4.2.3 Scherung

Wie bei der Skalierung zeigen die Ergebnisse bei Scherung eine höhere Precision als Recall. Abbildung 4.14 zeigt die über alle Testbilder gemittelten Ergebnisse für Precision, Recall und F-Maß. Die Ergebnisse zeigen keine große Varianz bei Variation der Scherung. Das F-Maß liegt unabhängig von der Stärke der Scherung bei über 80 Prozent.

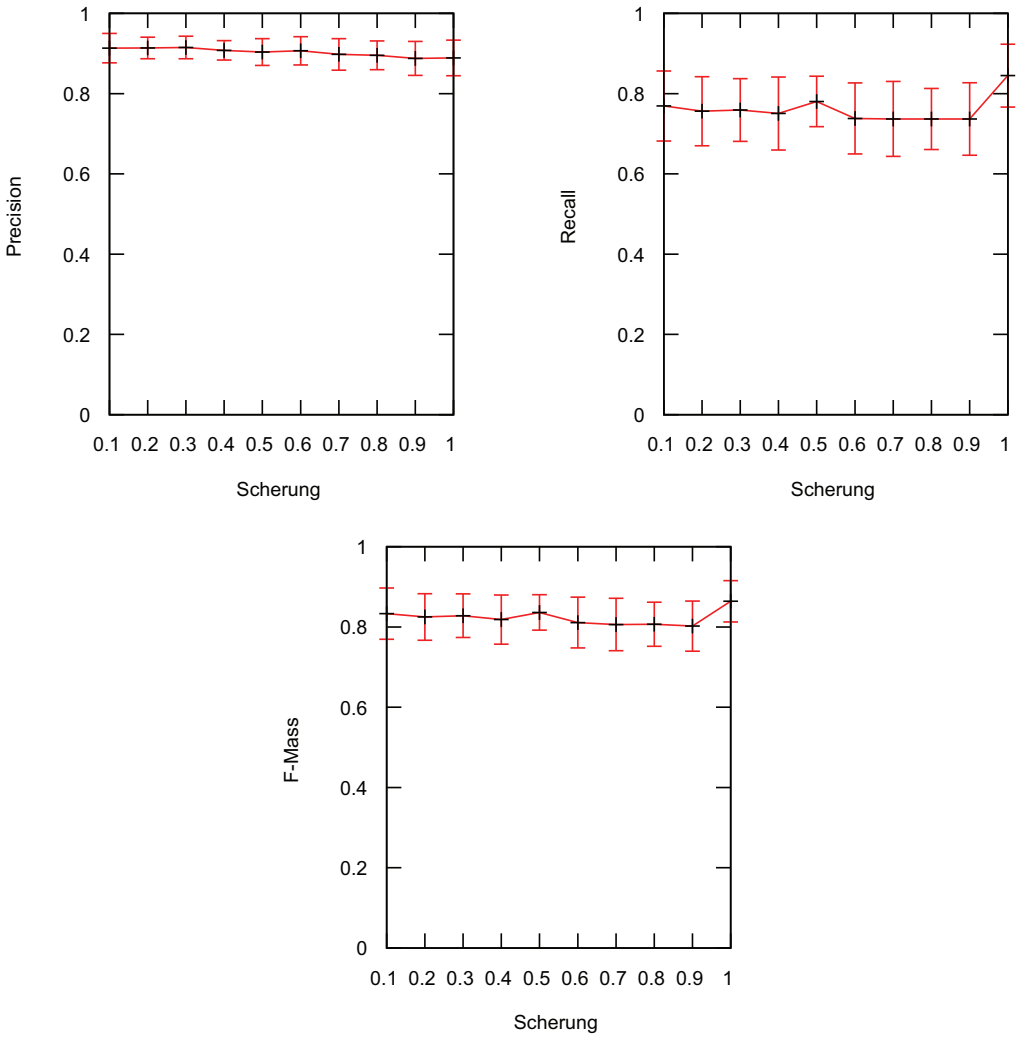


Abbildung 4.14: Stabilität des Detektors bei Scherung: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über der Stärke der Scherung (0: keine Scherung, 1: maximale Scherung) sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen

Abbildung 4.15 zeigt das F-Maß für alle Testbilder. Die Ergebnisse sind auch hier relativ unabhängig von der Stärke der Scherung. Das schlechteste Ergebnis zeigt wieder mit Abstand das Bild der Knolle (s. Abbildung 4.7 linke Spalte, zweites Bild von oben). Das Monster mit Spirale zeigt auch hier das beste Ergebnis.

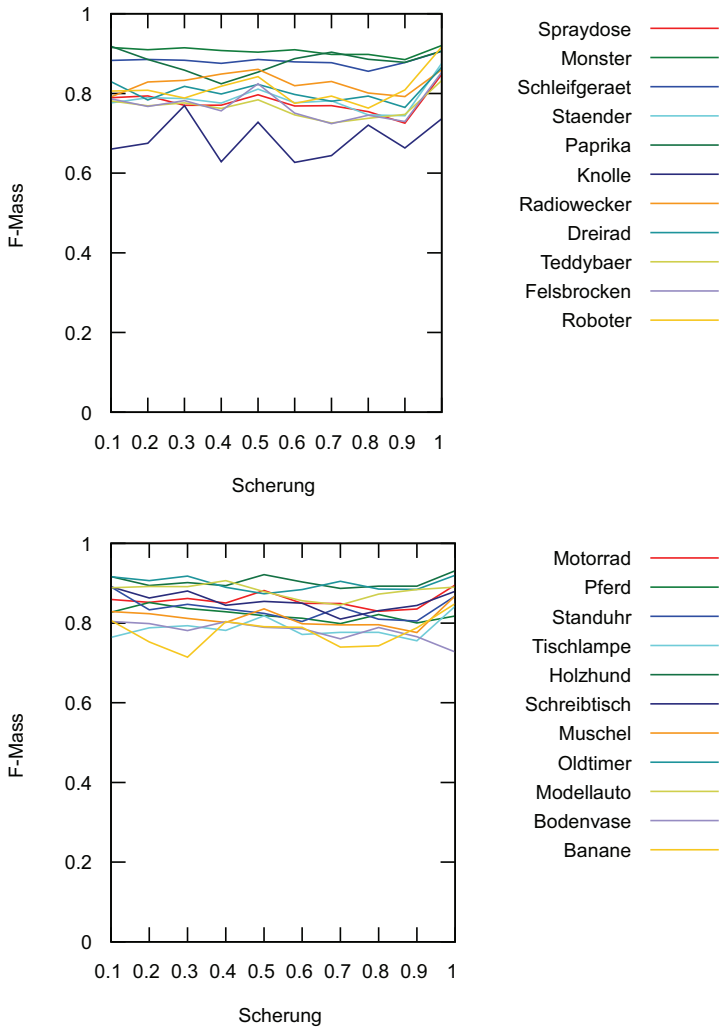


Abbildung 4.15: F-Maß der Stabilität des Detektors bei Scherung für alle Testbilder

Abbildung 4.16 zeigt die Ergebnisse des Detektors bei maximaler Scherung auf den Bildern der Knolle und des Monsters mit Spirale. Im Gegensatz zu den anderen Transformationen wurden auf dem transformierten Bild der Knolle alle sechs Points-of-Interest des Ursprungsbildes detektiert, allerdings weicht ihre Größe ab.

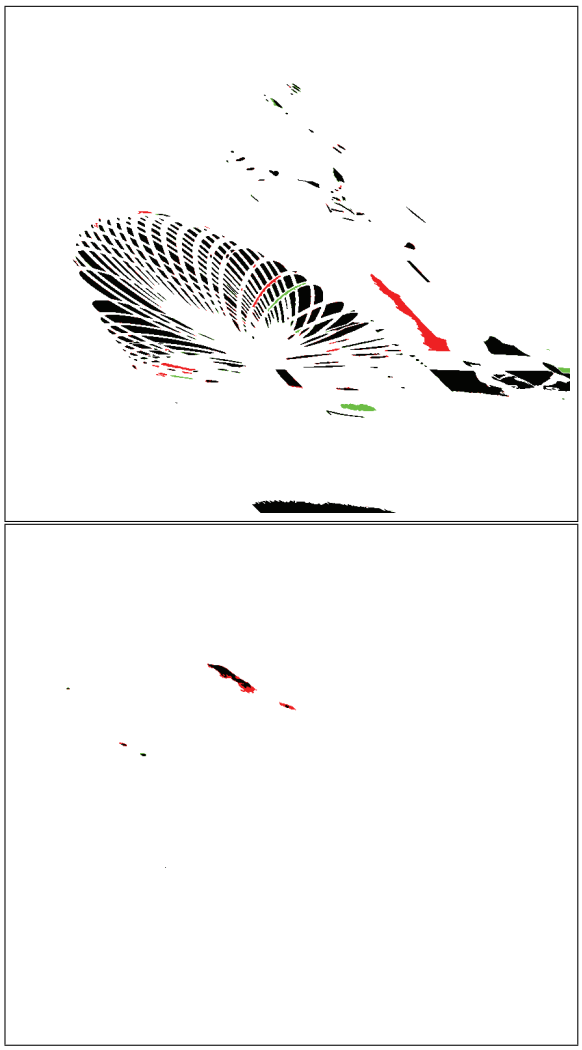


Abbildung 4.16: Ergebnis des Detektors für das Bild des Monsters mit Spirale (oben) und für das Bild der Knolle (unten) bei maximaler Scherung. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im gescherten Bild, sind rot dargestellt. Punkte, die im gescherten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt.

4.2.4 Perspektivische Verzerrung

Die gemittelten Ergebnisse der perspektivischen Verzerrung (s. Abbildung 4.17) gleichen denen der Skalierung und Scherung. Die Precision-Werte liegen deutlich höher als die Recall-Werte bei ca. 90 Prozent, während letztere bei knapp unter 70 Prozent liegen. Die Werte sind auch hier relativ unabhängig gegenüber der Stärke der Transformation.

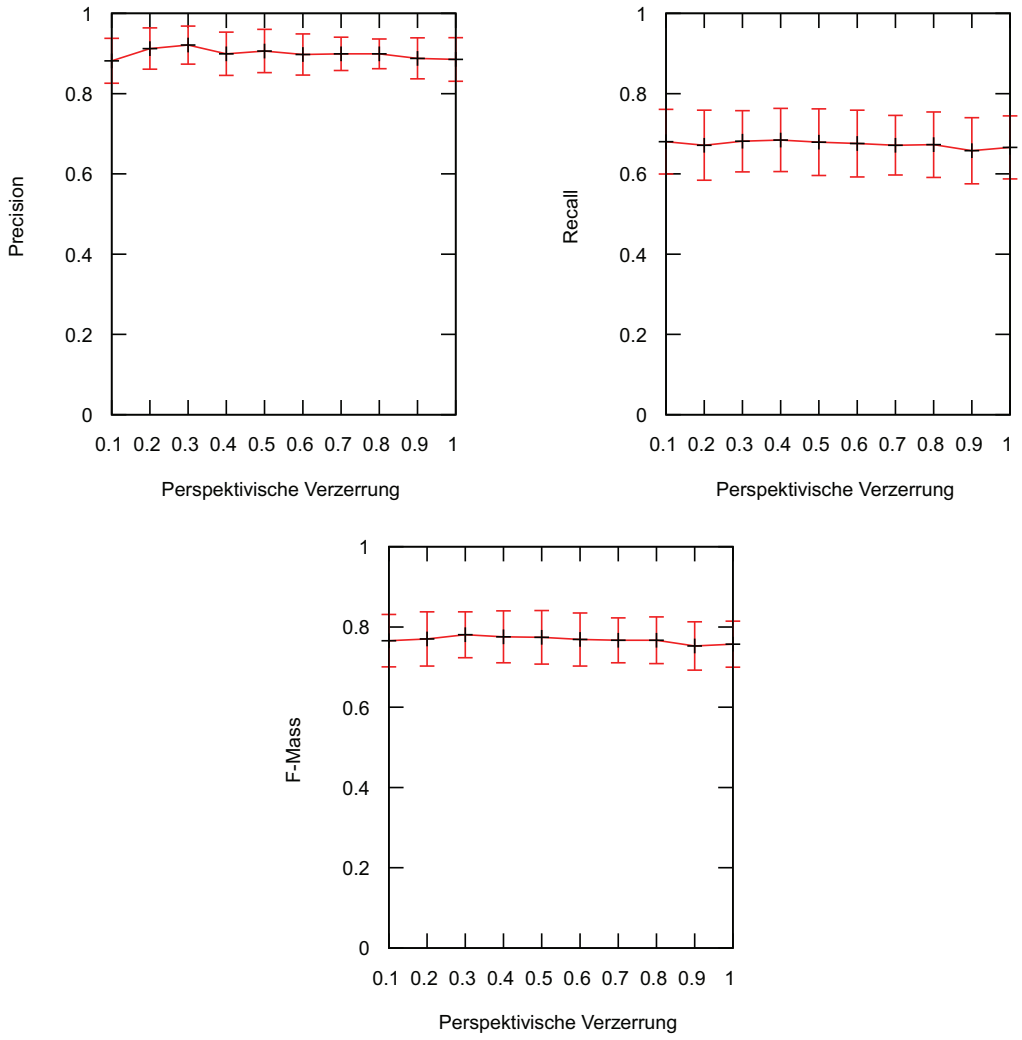


Abbildung 4.17: Stabilität des Detektors bei perspektivischer Verzerrung: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über der Stärke der perspektivischen Verzerrung (0: keine Verzerrung, 1: maximale Verzerrung) sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen

Abbildung 4.18 zeigt das F-Maß für alle Testbilder. Auch hier sind die Ergebnisse wieder unabhängig von der Stärke der Transformation. Bei dem Testbild mit dem schlechtesten Ergebnis (wieder die Knolle) steigt das F-Maß sogar bei zunehmender Verzerrung an. Das Bild des Monsters mit Spirale zeigt wieder eines der besten Ergebnisse.

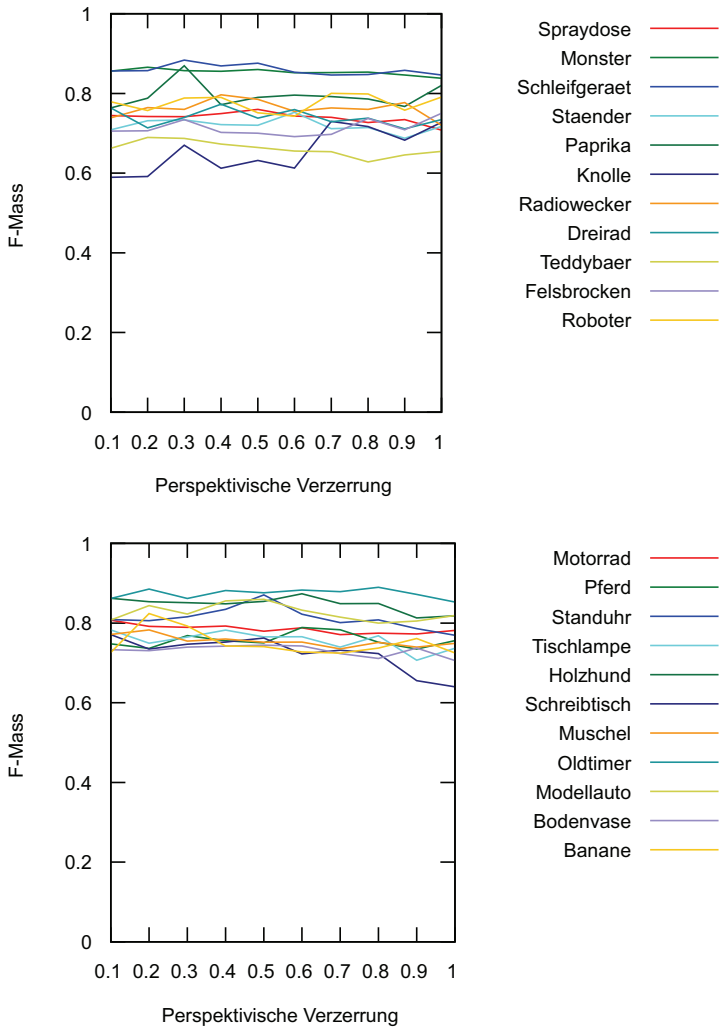


Abbildung 4.18: F-Maß der Stabilität des Detektors bei perspektivischer Verzerrung für alle Testbilder

Abbildung 4.19 zeigt die Ergebnisse des Detektors auf den perspektivisch verzerrten Bildern der Knolle und des Monsters mit Spirale. Auf dem verzerrten Bild der Knolle wurden fünf der sechs Points-of-Interest des Ursprungsbildes detektiert. Die Größe der richtig detektierten Points-of-Interest weicht auch hier von den Points-of-Interest des Ursprungsbildes ab.

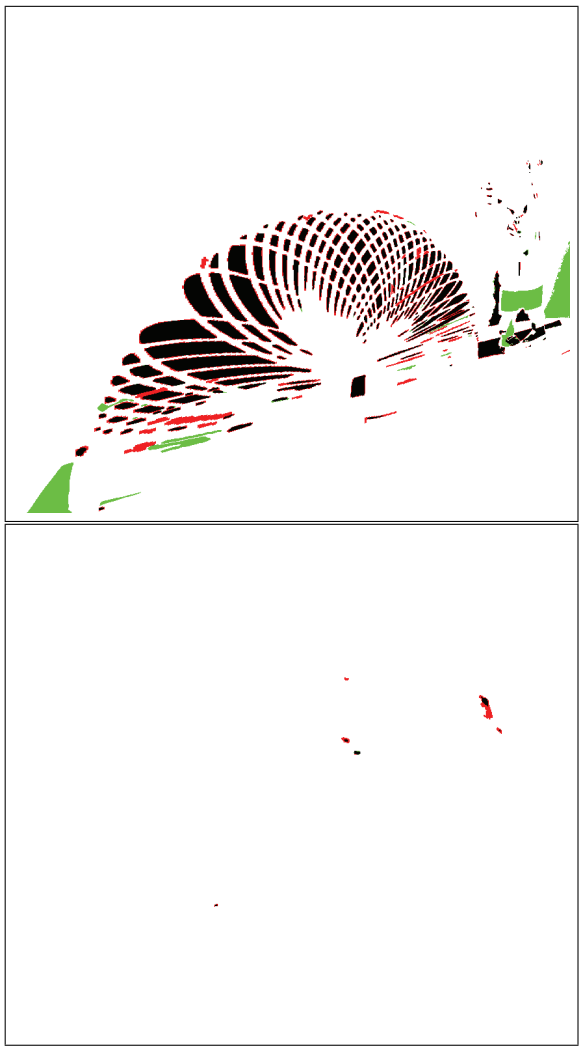


Abbildung 4.19: Ergebnis des Detektors für das Bild des Monsters mit Spirale (oben) und für das Bild der Knolle (unten) bei maximaler perspektivischer Verzerrung. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im verzerrten Bild, sind rot dargestellt. Punkte, die im verzerrten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt.

4.2.5 Zufällige Deformation

Precision und Recall bei Deformation (s. Abbildung 4.20) liegen jeweils bei ca. 80 Prozenz, die Varianz der Precision ist jedoch weitaus größer als bei den anderen Transformationen. Dies deutet darauf hin, dass die Ergebnisse bei einzelnen Bilder schlechter sind. Alle gemessenen Werte sind wieder relativ unabhängig gegenüber der Stärke der Transformation.

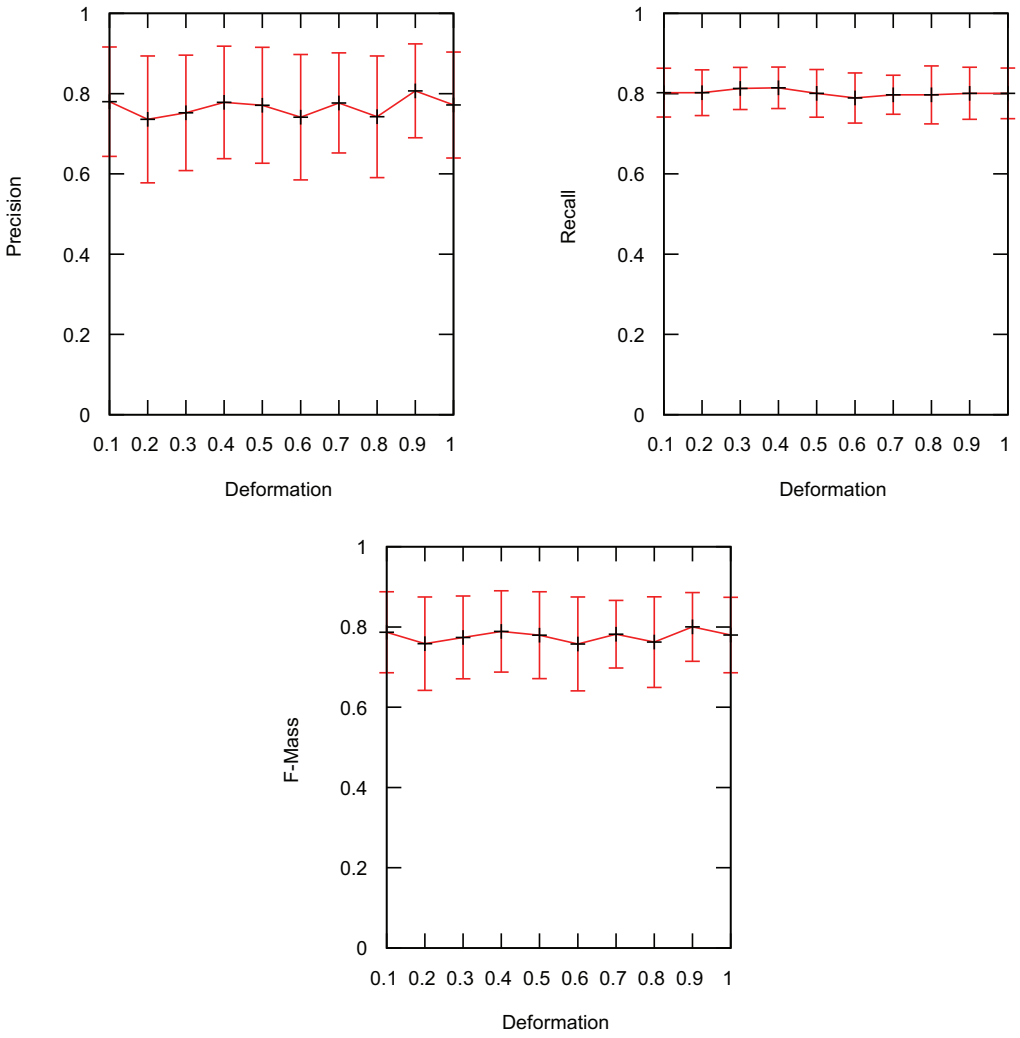


Abbildung 4.20: Stabilität des Detektors bei zufälliger Deformation: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über der Stärke der Deformation (0: keine Deformation, 1: maximale Deformation) sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen

Abbildung 4.18 zeigt das F-Maß für alle Testbilder. Hier sieht man, dass die gemessenen Werte für die meisten Bilder bei ca. 80 Prozent und darüber liegen. Nur vier Bilder zeigen schlechtere Werte, drei bei ca. 60 bis 70 Prozenz, eines bei nur ca. 40 Prozent. Dieses schlechteste Ergebnis liefert wieder das Bild der Knolle. Von diesem Bild rührt die hohe Varianz der Precision-Werte. Wie bei Skalierung, Scherung und perspektivischer Verzerrung scheinen auch hier die Ergebnisse relativ unabhängig von der Stärke der Transformation. Das Bild des Monsters mit Spirale liefert auch hier das beste Ergebnis.

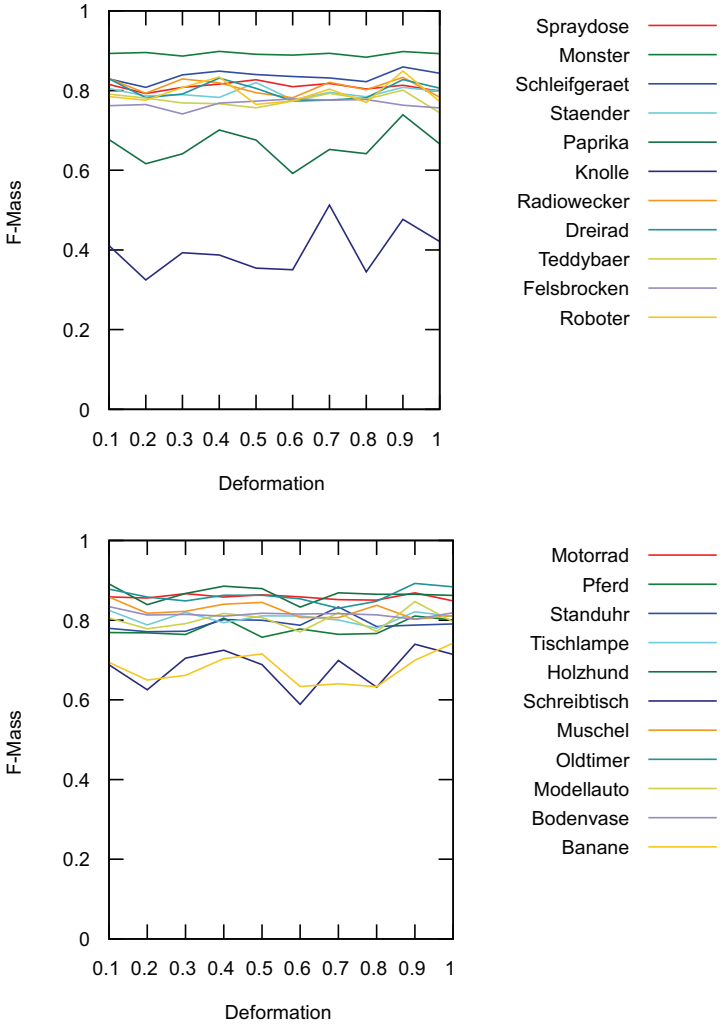


Abbildung 4.21: F-Maß der Stabilität des Detektors bei zufälliger Deformation für alle Testbilder

Abbildung 4.22 zeigt die Ergebnisse des Detektors auf den maximal deformierten Bildern der Knolle und des Monsters mit Spirale. Die durch die Deformation am Rand

der Bilder entstandenen schwarzen Flecken wurden jeweils fälschlicherweise als Points-of-Interest detektiert. Bei dem Bild der Knolle sorgt dies aufgrund der geringen Anzahl an ursprünglichen Points-of-Interests für einen sehr niedrigen Precision-Wert. Bei dem Bild des Monsters mit Spirale fallen diese zusätzlichen Detektionen aufgrund der hohen Anzahl an ursprünglichen Points-of-Interest nicht stark ins Gewicht.

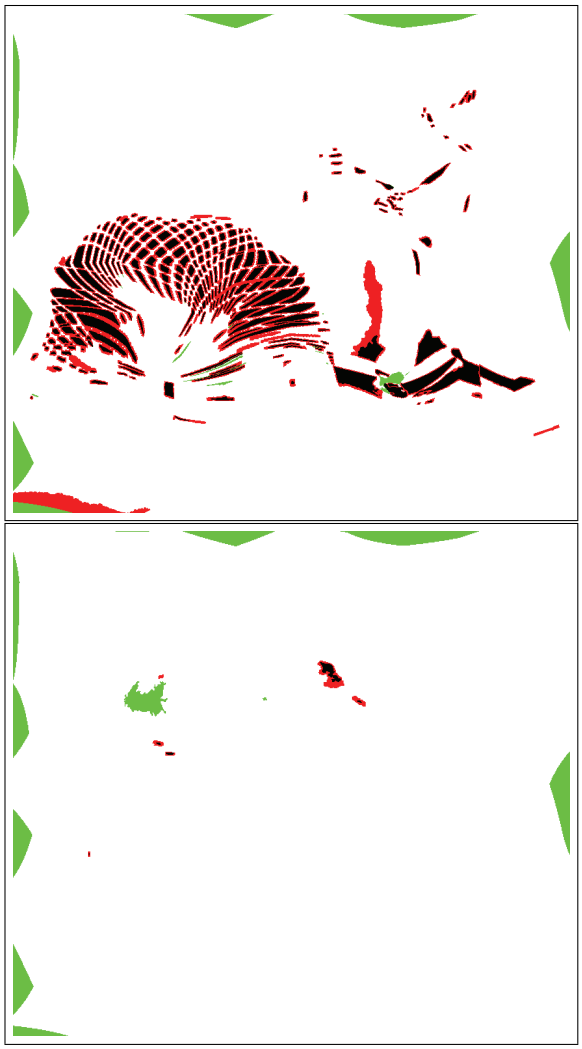


Abbildung 4.22: Ergebnis des Detektors für das Bild des Monsters mit Spirale (oben) und für das Bild der Knolle (unten) bei maximaler zufälliger Deformation. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im deformierten Bild, sind rot dargestellt. Punkte, die im deformierten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt.

4.2.6 Hinzufügen von Gauß-Rauschen

Abbildung 4.23 zeigt die über alle Testbilder gemittelten Ergebnisse für das Hinzufügen von Gauß-Rauschen variierender Intensität. Im Unterschied zu den Ergebnissen der anderen Transformationen zeigt sich hier ein monotoner Abfall der Performanz des Detektors bei steigender Intensität des Rauschens. Auch unterscheiden sich hier im Gegensatz zu den anderen Transformationen Precision und Recall kaum.

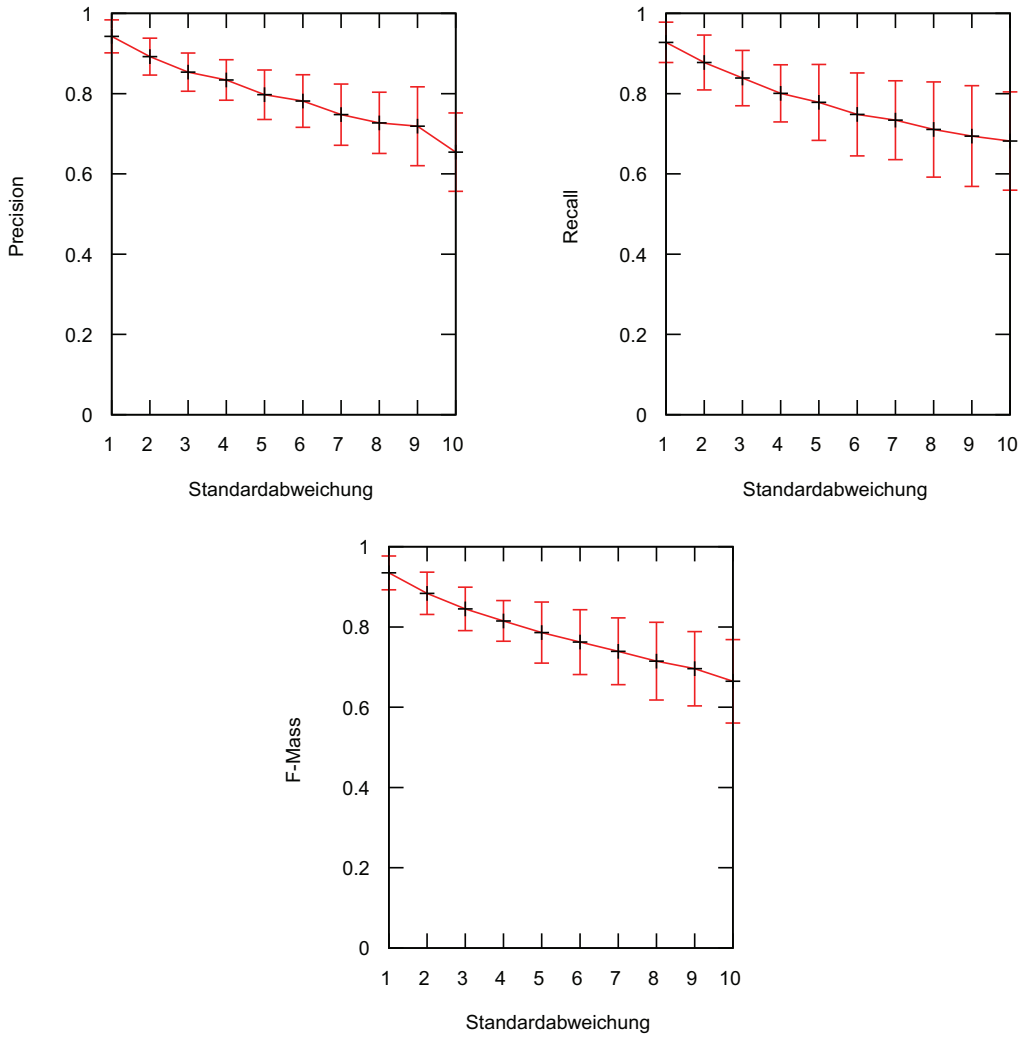


Abbildung 4.23: Stabilität des Detektors bei Hinzufügen von Gauß-Rauschen: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über der Standardabweichung des hinzugefügten Rauschens sind jeweils der Mittelwert (in schwarz) und die Standardabweichung des F-Maßes (als rote Balken über und unter dem Mittelwert) aufgetragen

Abbildung 4.24 zeigt das F-Maß für die einzelnen Testbilder. Im Unterschied zu den anderen Transformationen zeigt das Bild der Knolle nicht das alleinig schlechteste Ergebnis. Die schlechtesten Ergebnisse zeigen die Bilder der Knolle, des Teddybärs, des Felsbrockens und des Pferdes. Das Monster mit Spirale ist allerdings auch hier unter den Bildern mit dem besten Ergebnis. Dies scheinen Bilder mit einem niedrigeren Kontrast innerhalb der Objektregion zu sein. So können tatsächliche Extrempunkte häufiger nicht von Rauschpunkten unterschieden werden. Rauschpunkte werden so u.U. fälschlicherweise als Points-of-Interest erkannt, und tatsächliche Extrempunkte geringerer Stärke werden durch die Rauschunterdrückung entfernt.

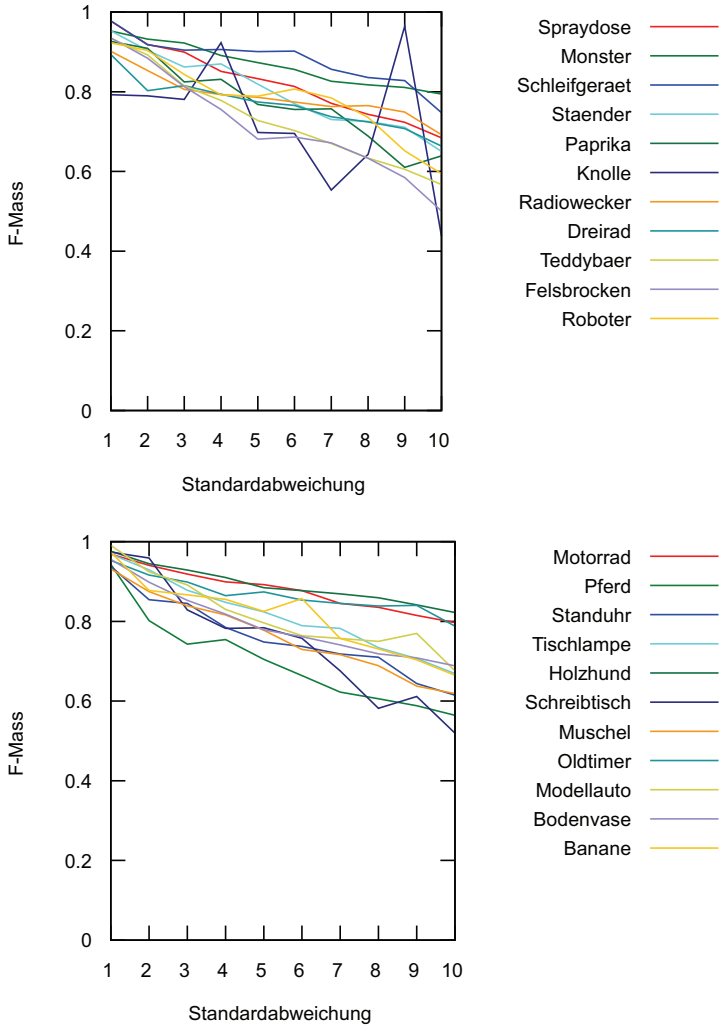


Abbildung 4.24: F-Maß der Stabilität des Detektors bei Hinzufügen von Gauß-Rauschen für alle Testbilder

Abbildung 4.25 zeigt die Ergebnisse des Detektors bei Hinzufügen von Rauschen mit maximaler Standardabweichung auf den Bildern der Knolle und des Monsters mit Spirale. Da die meisten Points-of-Interest im Bild des Monsters mit Spirale in kontrastreichen Bildbereichen liegen, werden die Ergebnisse von additivem Rauschen nicht stark beeinträchtigt. Auf dem kontrastärmeren Bild der Knolle werden dagegen nur zwei der ursprünglich sechs Points-of-Interest wiedergefunden. Zwei Rauschpunkte werden fälschlicherweise als zusätzliche Points-of-Interest detektiert.

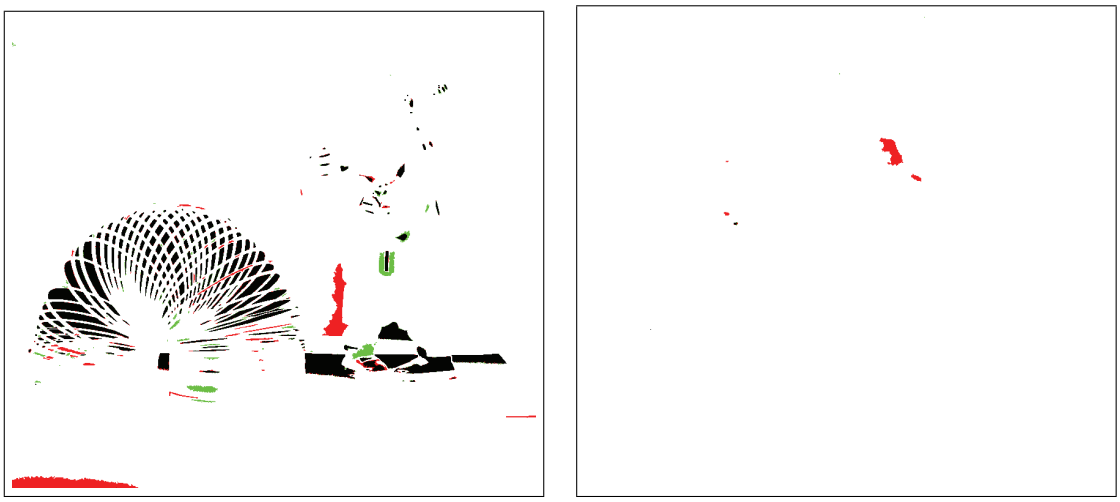


Abbildung 4.25: Ergebnis des Detektors für das Bild des Monsters mit Spirale (links) und für das Bild der Knolle (rechts) bei maximalem Rauschen. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im verrauschten Bild, sind rot dargestellt. Punkte, die im verrauschten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt.

4.2.7 Auswertung

Für alle Transformationen außer dem Hinzufügen von Rauschen ist die Performanz des Detektors unabhängig von der Stärke der jeweiligen Transformation.

Insbesondere nehmen die Stabilitätswerte bei Zunahme der Transformationsstärke im Mittel nicht ab. Dies gilt für alle Transformationsklassen ausgenommen des Rauschens, speziell auch für die zufällige Deformation (s. die Abbildungen 4.20 und 4.21). Die Messungen unterstützen demnach die Hypothese, dass der vorgestellte Detektor auch im diskreten Fall deformationsinvariant ist.

Bei Zunahme des Rauschens nimmt die Performanz des Detektors dagegen ab. Der im Detektor verwendete Ansatz zur Rauschunterdrückung kann die Auswirkungen starken Rauschens also nicht vollständig neutralisieren. Dies ist allerdings zu erwarten, da ein Rauschpunkt nach der vorgestellten Graphmetrik nicht von einer hellen oder dunklen Bildstruktur zu unterscheiden ist. Die Invarianz des Detektors gegenüber den anderen Transformationen wird durch die Ergebnisse bestätigt.

Für das Bild der Knolle liefert der Detektor in den meisten Fällen das schlechteste Ergebnis. Auf diesem Bild werden vom Detektor aufgrund der sehr einfachen Struktur und des niedrigen Kontrastes nur sehr wenige Points-of-Interest detektiert. So sinkt der Recall bereits stark, wenn nur wenige der sechs Points-of-Interest nicht im transformierten Bild detektiert werden. Die sechs detektierten Points-of-Interest sind außerdem nicht sehr stabil, da sie nur knapp über dem Rauschschwellwert r liegen. Dies führt zu den schlechten Ergebnissen bei dem Bild der Knolle.

Interessant ist die Beobachtung, dass die Performanz jeweils bereits bei der geringsten Stärke der Transformationen unter dem Optimum liegt und dann nicht weiter fällt. Bei allen Transformationen, die Interpolation beinhalten (dies sind alle außer dem Hinzufügen von Rauschen), zeigt sich dieses Verhalten. Zudem liegen die Recall-Werte bei fast allen Transformationen, die Interpolation beinhalten, unter den Precision-Werten. Dies ist ein Indiz dafür, dass einige der detektierten Points-of-Interest nicht stabil gegenüber einer Abnahme der Bildschärfe bzw. gegenüber Glättungseffekten sind, wie sie lokal bei Interpolation vorkommen. Z.B. findet der Detektor auch Points-of-Interest sehr geringer räumlicher Ausdehnung, die bei Interpolation verschwinden können.

Diese Vermutung erklärt die bereits bei geringer Transformationsstärke unter dem Optimum liegenden Stabilitätswerte, da solche Glättungseffekte bereits bei geringfügigen Verzerrungen eines Bildes auftreten.

Um die Vermutung zu überprüfen, werden die Testbilder in einem weiteren Experiment mittels eines Gauß-Filters zunehmend unschärfer gemacht und danach jeweils denselben zufälligen Deformationen wie im vorangegangenen Abschnitt ausgesetzt. Trifft die Vermutung zu, verschwinden die instabilen Points-of-Interest durch die Gauß-Filterung, und die Recall-Werte werden dadurch besser.

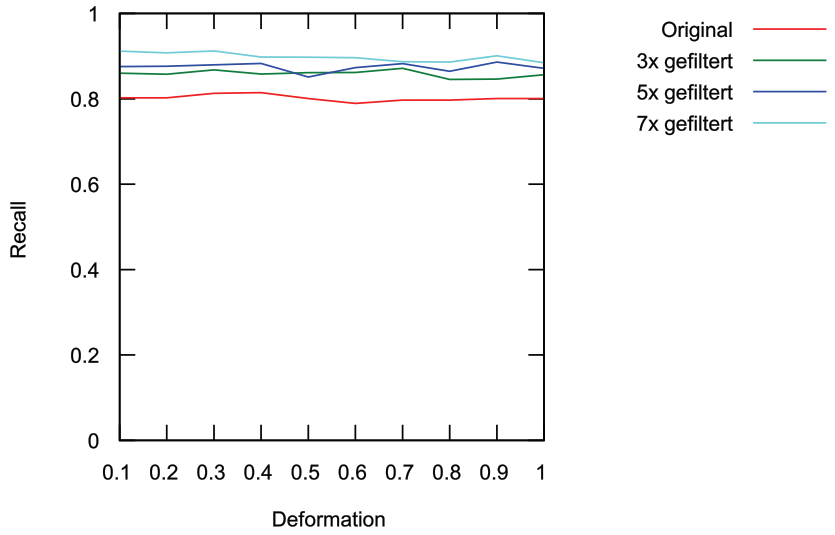


Abbildung 4.26: Recall der Stabilität des Detektors unter zufälliger Deformation, gemittelt über alle Bilder, bei zunehmender Unschärfe durch mehrfache Anwendung eines Gauß-Filters

Abbildung 4.26 zeigt die Ergebnisse des Experiments, die die Vermutung bestätigen. Je häufiger die Bilder mit einem Gauß-Filter geglättet werden, desto höher steigt der Recall-Wert des Detektors.

Durch Filterung mit einem Gauß-Filter vor der Anwendung des Detektors können demnach die Ergebnisse verbessert werden, allerdings auf Kosten der Anzahl der detektierten Points-of-Interest. Es ist außerdem zu beachten, dass die Gauß-Filterung nicht deformationsinvariant ist. Sie basiert auf linearer Faltung mit einem symmetrischen Faltungskernel und ist daher nur rotationsinvariant.

Die Betrachtung des Beweises zur Deformationsinvarianz im kontinuierlichen Fall (s. Abschnitt 3.2.1) legt nahe, dass die instabilen Punkte bei Glättung und Diskretisierung die Monotoniebedingung im Beweis verletzen. Die stetige Zunahme der Stabilität bei mehrfacher Anwendung des Gauß-Filters lässt vermuten, dass die betrachteten instabilen Punkte die einzigen Punkte sind, die die Deformationsinvarianz im diskreten Fall beeinträchtigen. Sie stellen somit die kritische Differenz zwischen dem kontinuierlichen und dem diskreten Fall dar.

Weiterhin konnte beobachtet werden, dass die gefundenen Sattelpunkte meist einer großen Menge von Bildpunkten im Ursprungsbild entsprechen und somit als nicht geeignet herausgefiltert werden (s. Abschnitt 3.5.3). Im Ausblick (s. Abschnitt 5.1) wird darauf eingegangen, wie die räumliche Ausdehnung der Sattelpunkte durch einen Algorithmus für $\alpha < 1$ verringert werden kann.

5 Zusammenfassung und Ausblick

In den vorangegangenen Kapiteln wurde anhand von existierenden Evaluationen aufgezeigt, dass bestehende Verfahren zur Detektion von Points-of-Interest für 2D-Projektionen von komplexeren Szenen, die 3D-rigiden Transformationen unterliegen, nicht zufriedenstellend funktionieren (s. Abschnitt 2.3). Es wurde gezeigt, dass die Eigenschaft der 2D-Deformationsinvarianz der Invarianz gegenüber solchen 3D-Transformationen mindestens ebenbürtig ist, solange keine Verdeckungen vorkommen (s. Abschnitt 3.1).

Insbesondere wurde damit gezeigt, dass die Klasse der Deformationen alle in bisherigen Verfahren betrachteten zweidimensionalen Transformationen beinhaltet (u.A. Rotation, Skalierung und affine Transformationen, s. Abschnitt 2.1).

Es wurde daraufhin eine Beschreibung von Bildern mittels Graphen und einer deformationsinvarianten Graphmetrik vorgeschlagen und die Einbettung des entstehenden Bildgraphen in den Euklidschen Raum mittels multdimensionaler Skalierung untersucht (s. Abschnitte 3.3 und 3.4). Dabei wurde eine neue Graphmetrik entwickelt, die eine bessere Einbettung in den Euklidschen Raum ermöglicht (s. Abschnitt 3.4.1). Auf der Basis dieser Metrik und der Untersuchung bestimmter Bildstrukturen in ihrer Einbettung im Euklidschen Raum wurden deformationsinvariante markante Knoten des Graphen identifiziert und ein Algorithmus zur automatischen Bestimmung dieser Knoten und der entsprechenden Pixel im Bild entwickelt (s. Abschnitt 3.5). Dem damit vorgestellten Verfahren zur Detektion von Points-of-Interest liegt die Annahme des Lambertschen Beleuchtungsmodells zugrunde.

Der Hauptbeitrag dieser Arbeit liegt in der Vorstellung der neuen Graphmetrik *de*, welche eine Kombination aus Euklidscher und geodätischer Distanz im Bildgraphen darstellt, und in der Entwicklung eines Algorithmus' zur Bestimmung deformationsinvarianter Points-of-Interest. Die Evaluation dieses Algorithmus' zeigt die Stabilität der detektierten Points-of-Interest gegenüber Deformationen variierender Stärke und die Robustheit gegenüber additivem Gauß-Rauschen.

Die Deformationsinvarianz der Metrik wurde für den kontinuierlichen Fall bewiesen (s. Abschnitt 3.2.1). Für den diskreten Fall in der Anwendung als Graphmetrik ist ein solcher Beweis nicht zu führen, wie das Gegenbeispiel aus Kapitel 4 zeigt. Der Detektor wurde daher abschließend anhand einer Reihe von Deformationen anhand eines Testdatensatzes [33] empirisch evaluiert (s. Kapitel 4). Die Hypothese, der entwickelte Detektor sei auch im diskreten Fall hinreichend deformationsinvariant, konnte durch die Evaluation bestätigt werden. Damit wurde das Ziel der Arbeit erreicht (s. Abschnitt 1.2).

Die einzigen der vom Algorithmus detektierten Points-of-Interest, die nicht deformationsinvariant sind, sind instabil gegenüber geringen Bildänderungen, die z.B. durch Interpolation und Glättung entstehen. Dies wurde durch ein weiteres Experiment bestätigt (s. Abschnitt 4.2.7). Solche instabilen Punkte können demnach durch Gauß-Glättung her-

ausgefiltert werden. Hierbei ist zu beachten, dass die Gauß-Filterung wie jede Faltung nicht deformationsinvariant ist. Bei Anwendung einer solchen Filterung muss daher jeweils ein Kompromiss zwischen Deformationsinvarianz der Points-of-Interest und ihrer Stabilität gegenüber Glättungseffekten gefunden werden.

5.1 Ausblick

Nicht in dieser Arbeit untersucht wurde das Verhalten des Algorithmus' bei Verletzung der Annahme des Lambertschen Beleuchtungsmodells und im Falle von Verdeckungen. Da die vorgestellte Graphmetrik auf der Arbeit von Ling und Jacobs [22] aufbaut, kann die Deformationsinvarianz der graphbasierten Beschreibung durch Variation eines α-Parameters gesteuert werden. In dieser Arbeit wurde ein effizienter Algorithmus für den Fall $\alpha = 1$ entwickelt (d.h. das Abstandsmaß basiert nur auf der Intensität der Bildpunkte und ihrer Topologie, nicht auf der räumlichen Entfernung der Bildpunktkoordinaten voneinander). Zukünftige Arbeit besteht in weiterer Komplexitätsreduktion, um einen effizienten Algorithmus zur Detektion der vorgestellten *Sattel-* und *Extrempunkte* für den Fall $\alpha < 1$ entwickeln zu können. Mittels eines solchen Algorithmus' kann der Einfluss des α-Parameters auf die Robustheit des Detektors, speziell unter Verdeckungen, untersucht werden. Nach Ling und Jacobs [22] senkt die Verminderung des α-Parameters zwar die Deformationsinvarianz, erhöht dafür aber die Möglichkeit zur Unterscheidung zwischen mehreren Points-of-Interest. Da die räumlichen Pixelkoordinaten bei $\alpha < 1$ in das Abstandsmaß miteinbezogen werden, wird außerdem die räumliche Ausdehnung der detektierten Points-of-Interest verringert, was die Anzahl an nicht-lokalen Punkten senkt, die durch die anschließende Filterung entfernt werden. Dies betrifft vor allem die Sattelpunkte, die bei dem vorgestellten Algorithmus für $\alpha = 1$ in den meisten Fällen eine zu hohe räumliche Ausdehnung haben, um als Point-of-Interest gewählt zu werden.

Für den Fall der Verletzung des Lambertschen Beleuchtungsmodells und allgemein im Falle von Beleuchtungsschwankungen ist es u.U. lohnenswert, anstatt der Intensität der Pixel andere Merkmal zu betrachten, z.B. den Intensitätsgradienten oder Farbmerkmale. Die Verwendung von Farbe anstelle von Intensitäten sollte eine bessere Unterscheidung mehrerer Points-of-Interest ermöglichen. Die Verwendung von Gradienten wiederum würde zu einer höheren Robustheit gegenüber Beleuchtungsänderungen führen, da der Intensitätsgradient nicht von der absoluten Helligkeit abhängt. Hierbei ist allerdings zu beachten, dass der Intensitätsgradient sich im Allgemeinen unter Deformationen ändert, also nicht deformationsinvariant ist. Ein auf den Gradienten basierender Detektor wäre daher weniger deformationsinvariant als der auf Intensitäten basierende Detektor.

Der Bildgradient besteht typischerweise aus den beiden partiellen Ableitungen der Bildintensität, ist also zweidimensional. Auch höherdimensionale Merkmale sind vorstellbar. Zukünftige Arbeit besteht daher in der Erweiterung des Algorithmus' auf mehrdimensionale Merkmale. Hier besteht ein Zusammenhang zur Entwicklung eines Algorithmus' für $\alpha < 1$: Während im Fall $\alpha = 1$ nur die Bildintensität berücksichtigt wird, werden bei $\alpha < 1$ zusätzlich die beiden Pixelkoordinaten in das Abstandsmaß miteinbezogen. Dies entspricht in gewisser Weise der Verwendung eines dreidimensionalen Pixelmerkmals. Um

einen effizienten Algorithmus für mehrdimensionale Merkmale entwickeln zu können, ist also wie für den Fall $\alpha < 1$ eine Reduktion der algorithmischen Komplexität notwendig. Die Verwendung anderer eindimensionaler Merkmale, wie z.B. des Gradientenbetrags, ist dagegen ohne Anpassung des entwickelten Algorithmus' möglich. Hier besteht die zukünftige Arbeit eher darin, festzustellen, welche Auswirkungen die Verwendung des Gradientenbetrags auf die Deformationsinvarianz hat.

Die Menge der bekannten Verfahren konnte in dieser Arbeit um einen deformationsinvarianten Detektor erweitert werden, der es erlaubt, eine Obermenge der bisher möglichen Transformationen in Bildern zu verarbeiten. Damit können neue Anwendungsgebiete für Point-of-Interest-Detektoren erschlossen werden.

Abbildungsverzeichnis

3.1 Mengentheoretische Beziehungen zwischen Klassen zweidimensionaler Transformationen 22

3.2 Beispielbild, bestehend aus einem hellen Fleck auf dunklem Hintergrund: (a) Originalbild, (b) 3D-Darstellung unter Verwendung der Intensität als z-Koordinate, (c) Matrix der Intensitäten 28

3.3 Graph des Beispielbildes aus Abbildung 3.2, unter Verwendung der Achternachbarschaft: (a) Knoten (als Kreis dargestellt) und Kanten ohne Kantengewichte, (b) 3D-Darstellung der Knoten und Kanten, basierend auf den Koordinaten und der Intensität der korrespondierenden Pixel 29

3.4 Kürzester Weg (in grün) zwischen zwei Knoten des Beispielgraphen: (a) Exakte Graphrepräsentation, (b) Graphrepräsentation erweitert um zusätzliche Knoten und Kanten an Intensitätssprüngen 30

3.5 Kürzester Weg zwischen zwei Knoten eines deformierten Beispielgraphen: (a) Oben beschriebene Graphrepräsentation, (b) Graphrepräsentation erweitert um zusätzliche Knoten und Kanten an großen Intensitätssprüngen 31

3.6 (a) Gleichmäßige Abtastung des Bildes über die x- und y-Koordinaten, (b) gleichmäßige Abtastung des Bildes über die x- und z-Koordinaten (Intensität), (c) gleichmäßige Abtastung des Bildes über die y- und z-Koordinaten (Intensität), (d) kombinierte erweiterte Abtastung 32

3.7 Zeitkomplexität von Graphalgorithmen in Abhängigkeit von der Bildbreite: (a) Dijkstras Algorithmus zur Bestimmung kürzester Wege, (b) Johnsons Algorithmus zur Lösung des All-Pairs-Shortest-Paths-Problems . . . 35

3.8 (a) Planarer Graph mit 32×32 Knoten. Horizontale und vertikale Kanten haben ein Gewicht von 1, diagonale Kanten ein Gewicht von $\sqrt{2}$ (b) Knotenauswahl für $\varepsilon = 4$ 37

3.9 Beispielgraph und dessen Approximationen 39

3.10 Verteilung der Distanzabweichung Δd, aufgetragen über der Originaldistanz: (a) Delaunay-Graph, (b) Delaunay-Graph von Graphfärbungen, (c) $2 \cdot \varepsilon$-Umgebungsgraph 41

3.11 Verteilung der Distanzabweichung Δd im erweiterten Graph, aufgetragen über der Originaldistanz: (a) Delaunay-Graph, (b) Delaunay-Graph von Graphfärbungen, (c) $2 \cdot \varepsilon$-Umgebungsgraph 42

3.12 Beispielbild, bestehend aus einem hellen Fleck auf dunklem Hintergrund: (a) Originalbild, (b) 3D-Darstellung unter Verwendung der Intensität als z-Koordinate, (c) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$, (d) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$ 44
3.13 Beispielbild, bestehend aus zwei Leuchtpunkten auf dunklem Hintergrund: (a) Originalbild, (b) 3D-Darstellung unter Verwendung der Intensität als z-Koordinate, (c) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$, (d) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$ 46
3.14 Beispielbild, bestehend aus drei Leuchtpunkten auf dunklem Hintergrund: (a) Originalbild, (b) 3D-Darstellung unter Verwendung der Intensität als z-Koordinate, (c) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 \approx 0.011$), eingebettet im $\mathbb{R}^3$, (d) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$ 47
3.15 Kombinierter Abstand auf dem Beispielbild, bestehend aus einem Leuchtpunkt auf dunklem Hintergrund: (a) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$, (b) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$ 52
3.16 Kombinierter Abstand auf dem Beispielbild, bestehend aus zwei Leuchtpunkten auf dunklem Hintergrund: (a) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$, (b) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$ 52
3.17 Kombinierter Abstand auf dem Beispielbild, bestehend aus drei Leuchtpunkten auf dunklem Hintergrund: (a) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^3$, (b) die ersten zehn Eigenwerte der Matrix $\mathbf{B}$ 53
3.18 Kombinierter Abstand auf dem Bild einer Kante: (a) dreidimensionale Darstellung unter Verwendung der Intensität als z-Koordinate, (b) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$ 54
3.19 Kombinierter Abstand auf dem Bild einer Kante: (a) dreidimensionale Darstellung unter Verwendung der Intensität als z-Koordinate, (b) räumliche Struktur nach multidimensionaler Skalierung (Stress $S^2 = 0$), eingebettet im $\mathbb{R}^2$ 54
3.20 Graph des Bildes mit zwei Leuchtpunkten auf dunklem Hintergrund: (a) dreidimensionale Darstellung des Graphen unter Verwendung der Intensität z-Koordinate (in rot) mit kürzestem Weg zwischen den beiden Leuchtpunkten für $\alpha = 1$ (in grün), (b) Logarithmus der ersten 10 Eigenwerte e_i ($\log(e_i + 1)$) bei Variation des α-Parameters und Einbettung des kürzesten Weges zwischen beiden Leuchtpunkten in den Euklidschen Raum durch multidimensionale Skalierung, (c) die ersten beiden Dimensionen der Einbettung des kürzesten Weges in den Euklidschen Raum bei Variation des α-Parameters 55

3.21 Graph des Bildes mit zwei Leuchtpunkten auf dunklem Hintergrund: (a) Die ersten 10 Eigenwerte bei Variation der Standardabweichung d des additiven Rauschens und Einbettung des kürzesten Weges zwischen beiden Leuchtpunkten in den Euklidschen Raum durch multidimensionale Skalierung, (b) die ersten beiden Dimensionen der Einbettung des kürzesten Weges in den Euklidschen Raum bei Variation der Standardabweichung d des additiven Rauschens 56
3.22 Einbettungen eines Rauschbildes in den Euklidschen Raum: (a) zweidimensionale Darstellung der Einbettung des kürzesten Weges zwischen zwei Punkten des Rauschbildes, Stress $S^2 = 0$, (b) dreidimensionale Darstellung der Einbettung des gesamten Rauschbildes, Stress $S^2 = 0.02$ 57
3.23 Markante Punkte in den drei Beispielbildern bzw. deren Einbettungen im Euklidschen Raum. Die markanten Punkte entsprechen jeweils dem/den Leuchtpunkte(n) und dem Hintergrund: (a) Ein Leuchtpunkt (hervorgehoben sind zwei Extrempunkte), (b) zwei Leuchtpunkte (hervorgehoben sind zwei Extrempunkte und ein Sattelpunkt), (c) drei Leuchtpunkte (hervorgehoben sind drei Extrempunkte und ein Sattelpunkt) 58

4.1 Skalierung des Beispielgitters (oben): Original (in rot), um 10 Prozent vergrößert (in grün), um den Faktor 1.9 (90 Prozent) vergrößert (in blau), Skalierung des Beispielbildes (unten) in 10-Prozent-Schritten von 20 bis 100 Prozent 65
4.2 Rotation des Beispielgitters (oben): Original (in rot), um 22.5 ° gedreht (in grün), um 45 ° gedreht (in blau), Rotation des Beispielbildes (unten) in 9 °-Schritten von 18 ° bis 90 ° 66
4.3 Scherung des Beispielgitters (oben): Original (in rot), um 10 Prozent geneigt (in grün), um 100 Prozenz (45 °) geneigt (in blau), Scherung des Beispielbildes (unten) in 10-Prozent-Schritten von 20 Prozent bis 100 Prozent 67
4.4 Perspektivische Verzerrung des Beispielgitters (oben): Original (in rot), geringste Verzerrung (in grün), maximale Verzerrung (in blau), perspektivische Verzerrung des Beispielbildes (unten), von minimaler Verzerrung bis maximaler Verzerrung 68
4.5 Zufällige Deformation des Beispielgitters (oben): Original (in rot), geringste Deformationsstufe (in grün), maximale Deformation (in blau), Deformation des Beispielbildes (unten), von minimaler Deformation bis maximaler Deformation 69
4.6 Hinzufügen von Gauß-Rauschen: (a) Detail des Beispielbildes (b) verrauschtes Detail des Beispielbildes, Standardabweichung des Gauß-Rauschens: 10 70
4.7 Testbilder (in Leserichtung): Spraydose, Monster und Spirale, Schleifgerät, Ständer, Paprika, Knolle, Radiowecker, Dreirad, Teddybär, Felsbrocken, Roboter, Motorrad, Pferd, Standuhr, Tischlampe, Holzhund, Schreibtisch, Muschel, Oldtimer, Modellauto, Bodenvase, Banane 71

4.8 Stabilität des Detektors bei Skalierung: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über dem Skalierungsfaktor sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen 72

4.9 F-Maß der Stabilität des Detektors bei Skalierung für alle Testbilder . . . 73

4.10 Ergebnis des Detektors für das Bild des Modellautos bei einem Skalierungsfaktor von 1.6 (oben) und für das Bild der Knolle bei einem Skalierungsfaktor von 1.2 (unten). Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im skalierten Bild, sind rot dargestellt. Punkte, die im skalierten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt. . . . 74

4.11 Stabilität des Detektors bei Rotation: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über dem Rotationswinkel sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen 75

4.12 F-Maß der Stabilität des Detektors bei Rotation für alle Testbilder 76

4.13 Ergebnis des Detektors für das Bild des Monsters mit Spirale (oben) und für das Bild der Knolle (unten) bei einem Rotationswinkel von 45 °. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im rotierten Bild, sind rot dargestellt. Punkte, die im rotierten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt. 77

4.14 Stabilität des Detektors bei Scherung: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über der Stärke der Scherung (0: keine Scherung, 1: maximale Scherung) sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen 78

4.15 F-Maß der Stabilität des Detektors bei Scherung für alle Testbilder 79

4.16 Ergebnis des Detektors für das Bild des Monsters mit Spirale (oben) und für das Bild der Knolle (unten) bei maximaler Scherung. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im gescherten Bild, sind rot dargestellt. Punkte, die im gescherten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt. 80

4.17 Stabilität des Detektors bei perspektivischer Verzerrung: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über der Stärke der perspektivischen Verzerrung (0: keine Verzerrung, 1: maximale Verzerrung) sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen 81

4.18 F-Maß der Stabilität des Detektors bei perspektivischer Verzerrung für alle Testbilder 82

4.19 Ergebnis des Detektors für das Bild des Monsters mit Spirale (oben) und für das Bild der Knolle (unten) bei maximaler perspektivischer Verzerrung. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im verzerrten Bild, sind rot dargestellt. Punkte, die im verzerrten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt. 83
4.20 Stabilität des Detektors bei zufälliger Deformation: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über der Stärke der Deformation (0: keine Deformation, 1: maximale Deformation) sind jeweils der Mittelwert (in schwarz) und die Standardabweichung (als rote Balken über und unter dem Mittelwert) aufgetragen 84
4.21 F-Maß der Stabilität des Detektors bei zufälliger Deformation für alle Testbilder 85
4.22 Ergebnis des Detektors für das Bild des Monsters mit Spirale (oben) und für das Bild der Knolle (unten) bei maximaler zufälliger Deformation. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im deformierten Bild, sind rot dargestellt. Punkte, die im deformierten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt. 86
4.23 Stabilität des Detektors bei Hinzufügen von Gauß-Rauschen: (a) Precision, (b) Recall, (c) kombiniertes F-Maß. Über der Standardabweichung des hinzugefügten Rauschens sind jeweils der Mittelwert (in schwarz) und die Standardabweichung des F-Maßes (als rote Balken über und unter dem Mittelwert) aufgetragen 87
4.24 F-Maß der Stabilität des Detektors bei Hinzufügen von Gauß-Rauschen für alle Testbilder 88
4.25 Ergebnis des Detektors für das Bild des Monsters mit Spirale (links) und für das Bild der Knolle (rechts) bei maximalem Rauschen. Korrekt detektierte Punkte sind schwarz dargestellt. Punkte, die im Ursprungsbild detektiert wurden, aber nicht im verrauschten Bild, sind rot dargestellt. Punkte, die im verrauschten Bild detektiert wurden, aber nicht im Ursprungsbild, sind grün dargestellt. 89
4.26 Recall der Stabilität des Detektors unter zufälliger Deformation, gemittelt über alle Bilder, bei zunehmender Unschärfe durch mehrfache Anwendung eines Gauß-Filters 91

Literaturverzeichnis

[1] Bay, Herbert ; Ess, Andreas ; Tuytelaars, Tinne ; Gool, Luc V.: SURF: Speeded Up Robust Features. In: *Computer Vision and Image Understanding (CVIU)* 110 (2008), Nr. 3, S. 346–359

[2] Cottier, J. C.: Extraction et appariements robustes des points d'intérêt de deux images non étalonnées / LIFIA-IMAG-INRIA Rhône-Alpes. 1994. – Forschungsbericht

[3] Delaunay, Boris N.: Sur la sphère vide. In: *Bulletin of Academy of Sciences of the USSR* (1934), Nr. 6, S. 793–800

[4] Dijkstra, E. W.: A Note on Two Problems in Connexion with Graphs. In: *Numerische Mathematik* 1 (1959), S. 269–271

[5] Forssén, Per-Erik ; Lowe, David: Shape Descriptors for Maximally Stable Extremal Regions. In: Lew, Michael S. (Hrsg.) ; Sebe, Nicu (Hrsg.) ; Huang, Thomas S. (Hrsg.) ; Bakker, Erwin M. (Hrsg.): *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV)*. Washington, DC, USA : IEEE Computer Society, Oktober 2007, S. 1–8

[6] Förstner, Wolfgang: A Framework for Low Level Feature Extraction. In: Eklundh, Jan-Olof (Hrsg.): *Proceedings of the Third European Conference on Computer Vision (ECCV)* Bd. 2. London, UK : Springer-Verlag, 1994, S. 383–394

[7] Fraundorfer, Friedrich ; Bischof, Horst: Evaluation of local detectors on non-planar scenes. In: *Proceedings of the 28th Workshop AAPR*, 2004, S. 125–132

[8] Harris, Chris ; Stephens, Mike: A Combined Corner and Edge Detector. In: *Proceedings of the Fourth Alvey Vision Conference*, 1988, S. 147–151

[9] Heitger, Friedrich ; Heydt, Rüdiger von der ; Peterhans, Esther ; Rosenthaler, Lukas ; Kübler, Olaf: Simulation of neural contour mechanisms: representing anomalous contours. In: *Image and Vision Computing* 16 (1998), Nr. 6-7, S. 407–421

[10] Horaud, R. ; Veillon, F.: Finding geometric and relational structures in an image. In: Faugeras, Olivier D. (Hrsg.): *Proceedings of the First European Conference on Computer Vision (ECCV)*. London, UK : Springer-Verlag, 1990, S. 374–384

[11] Ilies, Iulian ; Jacobs, Arne ; Wilhelm, Adalbert F. X. ; Herzog, Otthein: Classification of News Images Using Captions and a Visual Vocabulary / Universität Bremen, TZI. 2009. – Forschungsbericht

[12] Jacobs, Arne: *Mosaicing auf Szenen mit bewegten Objekten*, Universität Bremen, Diplomarbeit, 2003

[13] Jänich, Klaus: *Topologie*. Springer, 2005

[14] Johnson, Donald B.: Efficient Algorithms for Shortest Paths in Sparse Networks. In: *J. ACM* 24 (1977), Nr. 1, S. 1–13

[15] Kadir, Timor ; Brady, Michael: Scale, Saliency and Image Description. In: *International Journal of Computer Vision* 45 (2001), Nr. 2, S. 83–105

[16] Kadir, Timor ; Zisserman, Andrew ; Brady, Michael: An affine invariant salient region detector. In: Pajdla, Tomás (Hrsg.) ; Matas, Jiri (Hrsg.): *Proceedings of the 8th European Conference on Computer Vision (ECCV)*, 2004, S. 228–241

[17] Ke, Yan ; Sukthankar, Rahul: PCA-SIFT: A more distinctive representation for local image descriptors. In: Guler, Sadiye (Hrsg.) ; Hauptmann, Alexander G. (Hrsg.) ; Henrich, Andreas (Hrsg.): *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004, S. 506–513

[18] Laptev, Ivan: On Space-Time Interest Points. In: *International Journal of Computer Vision* 64 (2005), Nr. 2-3, S. 107–123

[19] Lee, John M. ; Axler, S. (Hrsg.) ; Gehring, F. W. (Hrsg.) ; Ribet, K. A. (Hrsg.): *Introduction to Topological Manifolds*. Springer, May 2000

[20] Leibe, Bastian ; Schiele, Bernt: Scale invariant object categorization using a scale-adaptive mean-shift search. In: *DAGM'04 Annual Pattern Recognition Symposium*, August 2004, S. 145–153

[21] Lindeberg, Tony: Feature detection with automatic scale selection. In: *International Journal of Computer Vision* 30 (1998), S. 79–116

[22] Ling, Haibin ; Jacobs, David W.: Deformation Invariant Image Matching. In: Sebe, Nicu (Hrsg.) ; Lew, Michael S. (Hrsg.) ; Huang, Thomas S. (Hrsg.): *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV)*. Washington, DC, USA : IEEE Computer Society, 2005, S. 1466–1473

[23] Lowe, David G.: Object Recognition from Local Scale-Invariant Features. In: *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV)*, 1999, S. 1150–1157

[24] Lowe, David G.: Distinctive image features from scale-invariant keypoints. In: *International Journal of Computer Vision* 60 (2004), S. 91–110

[25] Matas, J. ; Chum, O. ; Urban, M. ; Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Rosin, Paul L. (Hrsg.) ; Marshall, A. D.

(Hrsg.): *Proceedings of the British Conference on Machine Vision (BMVC)*, 2002, S. 384–393

[26] Matula, David W. ; Beck, Leland L.: Smallest-last ordering and clustering and graph coloring algorithms. In: *J. ACM* 30 (1983), Nr. 3, S. 417–427

[27] Mikolajczyk, Krystian ; Schmid, Cordelia: Indexing based on scale invariant interest points. In: *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV)*, 2001, S. 525–531

[28] Mikolajczyk, Krystian ; Schmid, Cordelia: An affine invariant interest point detector. In: Heyden, Anders (Hrsg.) ; Sparr, Gunnar (Hrsg.) ; Nielsen, Mads (Hrsg.) ; Johansen, Peter (Hrsg.): *Proceedings of the 7th European Conference on Computer Vision (ECCV)*. London, UK : Springer-Verlag, 2002, S. 128–142

[29] Mikolajczyk, Krystian ; Schmid, Cordelia: Comparison of affine-invariant local detectors and descriptors. In: *Proceedings of the 12th European Signal Processing Conference*, 2004, S. 1729–1732

[30] Mikolajczyk, Krystian ; Schmid, Cordelia: Scale & Affine Invariant Interest Point Detectors. In: *International Journal of Computer Vision* 60 (2004), Oktober, Nr. 1, S. 63–86

[31] Mikolajczyk, Krystian ; Schmid, Cordelia: A performance evaluation of local descriptors. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 27 (2005), Nr. 10, S. 1615–1630

[32] Moravec, Hans P.: *Obstacle avoidance and navigation in the real world by a seeing robot rover*. Stanford, CA, USA, Stanford University, Dissertation, 1980

[33] Moreels, P. ; Perona, P.: Evaluation of Features Detectors and Descriptors Based on 3D Objects. In: *International Journal of Computer Vision* 73 (2007), Juli, Nr. 3, S. 263–284

[34] Rencher, A. C.: *Methods of multivariate analysis*. New York : John Wiley & Sons, 2002

[35] Schmid, Cordelia ; Mohr, Roger ; Bauckhage, Christian: Evaluation of Interest Point Detectors. In: *International Journal of Computer Vision* 37 (2000), Nr. 2, S. 151–172

[36] Schoenberg, I. J.: Remarks to Maurice Frechet's Article "Sur La Definition Axiomatique D'Une Classe D'Espace Distances Vectoriellement Applicable Sur L'Espace De Hilbert. In: *The Annals of Mathematics* 36 (1935), Nr. 3, S. 724–732

[37] Shi, Jianbo ; Tomasi, Carlo: Good features to track. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Juni 1994, S. 593–600

[38] Sivic, Josef ; Zisserman, Andrew: Video google: A text retrieval approach to object matching in videos. In: *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV)*, 2003, S. 1470–1477

[39] Stommel, M. ; Herzog, O.: Binarising SIFT-Descriptors to Reduce the Curse of Dimensionality in Histogram-Based Object Recognition. In: Slezak, Dominik (Hrsg.) ; Pal, Sankar K. (Hrsg.) ; Kang, Byeong-Ho (Hrsg.) ; Gu, Junzhong (Hrsg.) ; Kurada, Hideo (Hrsg.) ; Kim, Tai hoon (Hrsg.): *Proceedings of the International Symposium on Signal Processing, Image Processing and Pattern Recognition (SIP)*, Dezember 2009, S. 320–327

[40] Stommel, M. ; Herzog, O.: SIFT-Based Object Recognition With Fast Alphabet Creation and Reduced Curse of Dimensionality. In: Bailey, Donald (Hrsg.): *Proceedings of the International Conference on Image and Vision Computing New Zealand (IVCNZ)*, November 2009, S. 136–141

[41] Trujillo, Leonardo ; Olague, Gustavo: Synthesis of interest point detectors through genetic programming. In: Keijzer, Maarten (Hrsg.): *Proceedings of the 8th annual Conference on Genetic and Evolutionary Computation (GECCO)* Bd. 1. New York, NY, 10286-1405, USA : ACM Press, Juli 2006, S. 887–894

[42] Tuytelaars, Tinne ; Gool, Luc J. V.: Content-Based Image Retrieval Based on Local Affinely Invariant Regions. In: *Proceedings of the Third International Conference on Visual Information and Information Systems (VISUAL)*. London, UK : Springer-Verlag, 1999, S. 493–500

[43] Varma, M. ; Ray, D.: Learning The Discriminative Power-Invariance Trade-Off. In: Lew, Michael S. (Hrsg.) ; Sebe, Nicu (Hrsg.) ; Huang, Thomas S. (Hrsg.) ; Bakker, Erwin M. (Hrsg.): *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV)*, 2007, S. 1–8

[44] Vincent, L. ; Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (1991), Nr. 6, S. 583–598

[45] Witten, Ian H. ; Moffat, Alistair ; Bell, Timothy C.: *Managing gigabytes : compressing and indexing documents and images*. New York; South Melbourne, Vic. : Van Nostrand Reinhold ; Thomas Nelson Australia, 1994. – 429 S. – ISBN 0-442-01863-0