

Bachelorarbeit

Gewinnung scharfer Bilder und
Tiefeninformation aus
fokusvariablen Bilderserien

Henrik Rosenberg

16. November 2018

Eingereicht bei Prof. Dr. Thorsten Grosch

Abteilung für Graphische Datenverarbeitung und Multimedia
Institut für Informatik
Technische Universität Clausthal

Bachelorarbeit

Titel: Gewinnung scharfer Bilder und Tiefeninformation
aus fokusvariablen Bilderserien

Autor: Henrik Rosenberg

Matrikelnummer: 466354

Abgabedatum: 16. November 2018

Erstgutachter: Prof. Dr. Thorsten Grosch

Zweitgutachter: Prof. Dr.-Ing. Michael Prilla

Betreuender Mitarbeiter: M.Sc. Johannes Jendersie

Versicherung der selbstständigen Arbeit

Hiermit versichere ich, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass alle Stellen dieser Arbeit, die wörtlich oder sinngemäß aus anderen Quellen übernommen wurden, als solche kenntlich gemacht wurden und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsstelle vorgelegt wurde.

Clausthal, den 16. November 2018

Unterschrift:

Veröffentlichungseinverständnis

Hiermit erkläre ich, dass ich mit der öffentlichen Bereitstellung meiner Abschlussarbeit in der Instituts- und/oder Universitätsbibliothek nicht einverstanden bin.

Mit der elektronischen Bereitstellung meiner Abschlussarbeit bin ich einverstanden.

Clausthal, den 16. November 2018

Unterschrift:

Kurzfassung

Bei der Fotografie gibt es einen bestimmten Tiefenbereich in dem die dort befindlichen Objekte scharf abgebildet werden. In dieser Arbeit soll es darum gehen dieses Phänomen auszunutzen, um aus einer Bilderserie ein komplett scharfes Bild sowie Tiefeninformationen der Szene zu gewinnen.

Dazu wird eine Folge von Bildern aus dem selben Blickwinkel, aber mit unterschiedlicher Fokuseinstellung aufgenommen. Für jedes Pixel der Bilder wird die Schärfe mithilfe des Gradienten des Bildes bestimmt. So kann pixelweise entschieden werden welches der Ausgangsbilder am schärfsten ist und es wird eine Indexmap generiert. In Bereichen, in denen keine oder kaum Textur vorhanden ist, ist es mitunter schwer einen passenden Index zu finden. Daher wird eine Bewertung der Pixelgüte vorgenommen und nur Pixel, deren Güte über einem Schwellwert liegt, werden in der Indexmap aufgenommen. Um für die übrigen Pixel ebenfalls einen Index vorweisen zu können, wird diese mithilfe eines Interpolationsalgorithmus gefüllt. Auf Basis der Indexmap und den Ausgangsbildern kann ein komplett scharfes Bild generiert werden. Indem die Indizes in Entfernungen umgerechnet werden, wird aus der Indexmap die Tiefenmap gewonnen. Diese wiederum dient als Ausgangspunkt zur Bestimmung der Positionen der zu den Pixeln gehörenden Punkte im Raum.

In dieser Arbeit werden im wesentlichen die beiden Fragen 'Wie lege ich auf Basis der Schärfewertung einen sinnvollen Index fest?' und 'Welches Kriterium eignet sich am besten zur Bewertung der Pixelgüte?' diskutiert. Dazu werden jeweils verschiedene Möglichkeiten getestet und ihre Ergebnisse verglichen.

Mithilfe der jeweils vielversprechendsten Methode werden einige Testszenen analysiert und die Ergebnisse dargestellt. Die Erstellung eines komplett scharfen Bildes der Szene ist fast immer möglich. Die Tiefeninformationen sind hingegen nur unter bestimmten Umständen von brauchbarer Qualität.

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen der Fotografie	3
3	Verwandte Arbeiten	7
4	Erstellung der Bilderserien	10
4.1	Einschränkungen durch die verwendete Hard- und Software	10
4.2	Aufnahme der Bilder	11
5	Verarbeitung der Bilder	13
5.1	Schärfemaß	13
5.2	Indexbestimmung	17
5.2.1	Maximaler Gradient	18
5.2.2	Gewichtete Indizes	19
5.2.3	Gewichtete Indizes mit Schwellwert	20
5.2.4	Vergleich der Methoden zur Indexbestimmung	23
5.3	Pixelgüte	25
5.3.1	Maximaler Gradient	26
5.3.2	Quotient aus maximalem und durchschnittlichem Gradienten	28
5.3.3	Quotient aus maximalem und durchschnittlichem Gradienten mit Generalisierung	30
5.3.4	Interpolationsalgorithmus	33
5.4	Generierung des scharfen Bildes	36
5.5	Generierung der Tiefenmap	37
5.6	Generierung der Punktwolke	41
6	Ergebnisse	44
7	Ausblick	52
8	Zusammenfassung	54
	Referenzen	57



Abbildung 1: Bokeh bei einem herbstlichen Bild [Lou11]

1 Einleitung

Eine zum Teil gewollte, zum Teil ungewollte Eigenschaft von Fotografien ist, dass nur bestimmte Teile des Fotos scharf sind. Dieser Effekt wird gerne genutzt, um Objekte in einem Motiv hervorzuheben. Gegenstände weiter vorne in der Szene und der Hintergrund sind dann nur verschwommen abgebildet. Im Japanischen wird dieses Stilmittel Bokeh genannt [BSS17]. Ein Beispiel ist die Herbst-Szene in Abbildung 1.

In anderen Anwendungsfällen ist es wünschenswert möglichst den gesamten Bildbereich so scharf wie möglich abzubilden. Es gibt mehrere Einstellungsmöglichkeiten, welche die Tiefe des scharfen Bereichs, die Schärfentiefe, beeinflussen. Im allgemeinen ist es aber nicht möglich mit einer einfachen Aufnahme Objekte in allen Entfernungen scharf abzubilden.

Durch die Kombination mehrerer Bilder aus einer Perspektive, in denen unterschiedliche Bereiche der Szene im Fokus liegen, ist die Erzeugung eines komplett scharfen Bildes möglich. In diesem sind alle Objekte, die einen bestimmten Mindestabstand zur Kamera haben, scharf abgebildet.

Dies geschieht über die Entscheidung in welchem, der Bilder jedes einzelne Pixel am schärfsten ist. Die Information welches Pixel in welchem der Ausgangsbilder im Fokus liegt lässt außerdem Rückschlüsse auf die Tiefe zu. So lassen sich letztendlich die Koordinaten der abgebildeten Punkte im Raum berechnen.

In dieser Arbeit soll diese Methodik implementiert und ausgetestet werden. Dabei werden verschiedene Variationen des Algorithmus miteinander verglichen.

Zu Beginn dieser Arbeit sollen zunächst einige Fachbegriffe aus der Fotografie eingeführt werden, die für das Verständnis notwendig sind. Dabei wird vor allem der Zusammenhang zwischen den Kameraeinstellungen und der Schärfentiefe erläutert. Die Einführung in weitere benötigte Grundlagen wird an den entspre-

chenden Stellen vorgenommen.

Einige verwandte Arbeiten werden vorgestellt. Die wesentlichen Gemeinsamkeiten und Unterschiede in der Vorgehensweise zwischen diesen und meiner Arbeit werden dabei hervorgehoben.

Für die Erstellung der Bilderserien stand eine digitale Spiegelreflexkamera mit entsprechenden Objektiven zur Verfügung. Die verwendete Hard- und Software wird kurz beschrieben und insbesondere auf die sich daraus ergebenden Einschränkungen eingegangen. Anschließend wird der Ablauf der Bilderaufnahme geschildert.

Das verwendete Maß für die Schärfe eines Pixels und die dafür nötigen Grundlagen werden erläutert. Um die Frage zu klären, wie auf Basis dieses Maßes sinnvoll ein Index der Bilderserie gewählt werden kann, werden mehrere Methoden gezeigt. Sie werden miteinander verglichen, indem sie auf eine der aufgenommenen Test-Szenen angewandt werden.

In Bereichen in denen das Motiv kaum Textur hat, können den Pixeln keine sinnvollen Indizes zugeordnet werden. Die vorgeschlagene Lösung ist, diese 'schlechten' Pixel zunächst auszulassen und lediglich eine löchrige Indexmap zu erstellen. Diese wird anschließend mittels eines Interpolationsalgorithmus, der hier vorgestellt wird, gefüllt.

Um zu entscheiden welche Pixel 'gut' und welche 'schlecht' sind, wurden verschiedene Heuristiken getestet. Auf eine Auswahl davon wird in dieser Arbeit genauer eingegangen. Die so erzielten Ergebnisse werden zur Beurteilung der Heuristiken verwendet.

Die gefüllte Indexmap ist nur ein Zwischenschritt zur Erzeugung der eigentlichen Resultate. Es wird erläutert, wie aus ihr unter Zuhilfenahme der Ausgangsbilder ein komplett scharfes Bild der Szene generiert wird.

Ebenfalls aus der Indexmap wird die Tiefenmap generiert. Diese enthält die echten Entfernungen der abgebildeten Punkte zur Kamera, wofür eine Kalibrierung notwendig ist.

Nachdem so bereits die Z-Koordinate der Punkte bestimmt wurde, wird erklärt wie die übrigen beiden Koordinaten zu berechnen sind. Die so erstellte Punktwolke wird ebenfalls beurteilt.

Die bestmöglichen Ergebnisse der Anwendung auf die Testszenen werden vorgestellt und die Qualitätsunterschiede zwischen scharfen Bildern und Entfernungsbestimmungen betrachtet. Die Erstellung der scharfen Bilder funktioniert sogar bei Motiven mit Transparenz relativ gut. Die Tiefenrekonstruktion liefert in diesem Fall allerdings keine guten Werte. Auch bei Motiven, die keine oder nur wenig harte Kanten aufweisen, funktioniert die Rekonstruktion der Tiefe sehr schlecht. Auf besser geeigneten Szenen liefert der Algorithmus deutlich plausiblere Ergebnisse.

Im Ausblick werden einige Ansätze vorgeschlagen, wie das System noch verbessert werden kann und welche weiteren Untersuchungen für sinnvoll gehalten werden.

2 Grundlagen der Fotografie

Da in dieser Arbeit einige Begrifflichkeiten und Sachzusammenhänge aus der Fotografie Verwendung finden, soll an dieser Stelle kurz in die Grundlagen eingeführt werden.

Für diese Arbeit wurden die Bilder mit einer digitalen Spiegelreflexkamera (Nikon D 5500) aufgenommen. Abbildung 2 zeigt die wichtigsten Komponenten und den Strahlengang einer solchen Kamera.

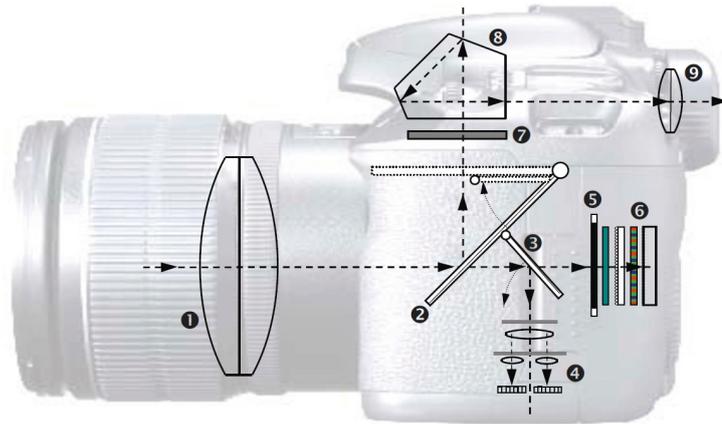


Abbildung 2: Strahlengang einer modernen digitalen Spiegelreflexkamera [Goc12]

Die Lichtstrahlen treten durch das Objektiv in die Kamera ein. Hier werden sie durch mehrere Linsen (1) gebrochen. Im Inneren des Kameragehäuses treffen sie zunächst auf den Umlenkspiegel (2). Dessen Position entscheidet, ob die Lichtstrahlen weiter geradeaus in Richtung Sensor (6) gelangen oder in Richtung Okular (9) umgelenkt werden. In der auf dem Bild eingezeichneten Position werden sie umgelenkt und treffen auf die Mattscheibe (7). Hier wird das Bild erzeugt, welches man sieht, wenn man durch das Okular blickt. Das Prisma (8) lenkt den Strahlengang entsprechend von der Mattscheibe zum Okular um. Ist der Spiegel nach oben geklappt, kann das Licht bei geöffnetem Verschluss (5) auf den Sensor treffen. Der Verschluss dient dabei zur Regelung der Belichtungszeit. Der Autofokus-Hilfsspiegel (3) lenkt einen Teil des Lichts in Richtung der Autofokus-sensoren (4). Diese bestimmen die Schärfe eines Bildes oder Bildausschnitts über Kantendetektion. Ist der Spiegel weg geklappt, können sie nicht verwendet werden und es wird stattdessen eine Kontrastmessung durchgeführt. [Goc12]

Der Begriff *Fokuspunkt* bezeichnet den Punkt des Motivs auf den scharf gestellt ist. Man sagt auch der Punkt liegt im Fokus. Die Ebene in der dieser Punkt liegt heißt *Fokusebene* oder *Einstellungsebene*. Der Abstand des Fokuspunktes zur Kamera wird hier als *FokUSDistanz* bezeichnet. [Goc12]

Von besonderer Bedeutung für diese Arbeit ist die *Schärfentiefe*. Sie bezeichnet die Ausdehnung des Bereichs vor und hinter dem Fokuspunkt in dem die dort befindlichen Objekte noch annehmbar scharf abgebildet werden. Der vorderste Punkt

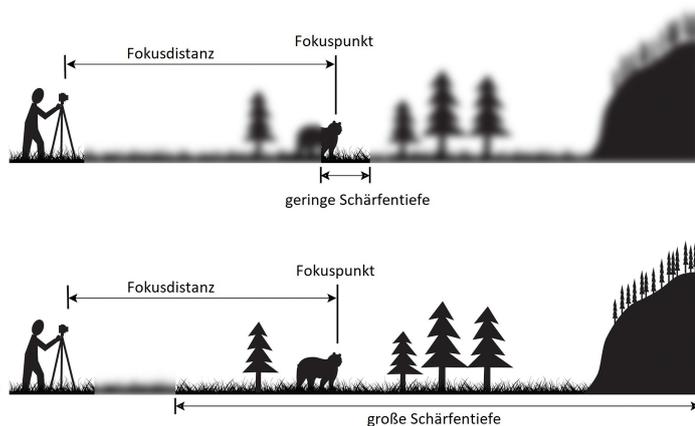
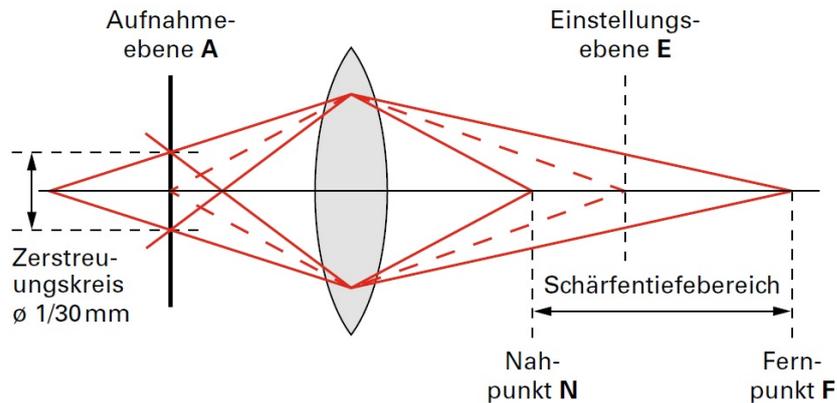


Abbildung 3: Geringe und große Schärfentiefe verdeutlicht nach [Gre18]

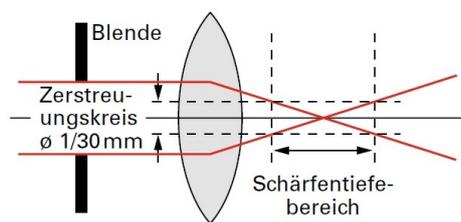
des scharfen Bereichs heißt dabei *Nahpunkt* und der am weitesten entfernte noch scharfe Punkt wird *Fernpunkt* genannt. Was genau annehmbar scharf bedeutet, ist über sogenannte *Zerstreuungskreise* definiert. Diese geben an, welche Ausdehnung die Abbildung eines einzelnen Punktes im Motiv auf dem Bild hat. Das menschliche Auge erkennt einen Punkt aus einer Entfernung von 25 bis 30 cm als scharf, wenn der Durchmesser des Zerstreuungskreises geringer als $1/30$ mm ist. [BSS17] In Abbildung 3 sind eine geringe und eine große Schärfentiefe gegenüber gestellt.

Um einem Punkt eine Entfernung auf Basis der Information bei welcher Fokusbildung er scharf ist zuzuordnen, ist eine geringe Schärfentiefe von Vorteil. Es gibt im wesentlichen drei Größen, die sich auf die Schärfentiefe auswirken. Zum einen ist die Brennweite des verwendeten Objektivs entscheidend. Hier gilt, je länger die Brennweite, desto geringer die Schärfentiefe. Die aktuelle Fokusbildung ist von Bedeutung. Im Nahbereich ist die Schärfentiefe eher gering, während sie mit wachsender Entfernung zunimmt. Letztendlich nimmt auch noch die Größe der Blendenöffnung Einfluss auf die Schärfentiefe. Die Blende lässt nur eine gewisse Menge Licht in die Kamera. Ihre Öffnung wird dabei durch die Blendenzahl angegeben, die der Kehrwert der relativen Öffnung ist. Eine kleine Blendenzahl bedeutet also eine weite Öffnung der Blende und viel Lichteinlass, während eine große Blendenzahl eine enge Blendenöffnung beschreibt, die nur wenig Licht durchlässt. [BSS17]

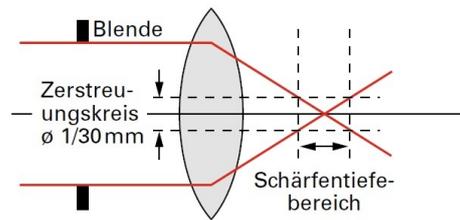
Abbildung 4 behandelt die Auswirkung der Blendeneinstellung auf die Schärfentiefe. Zunächst zeigt 4(a) wie die Schärfentiefe konstruiert wird. Lichtstrahlen von Punkten aus der Einstellungsebene werden auf genau einem Punkt des Sensors, hier allgemein mit *Aufnahmeebene* bezeichnet, gebündelt. Die vom Fernpunkt ausgehenden Strahlen sind, wenn sie die Aufnahmeebene schneiden schon so nah beieinander, dass sie innerhalb des Zerstreuungskreises liegen. Der Nahpunkt liegt so nah an dem Objektiv, dass seine Lichtstrahlen sich schon vor der Aufnahmeebene



(a) Konstruktion der Schärfentiefe



(b) Kleine Blendenöffnung
große Blendenzahl



(c) Weite Blendenöffnung
kleine Blendenzahl

Abbildung 4: Auswirkung der Blendenöffnung auf die Schärfentiefe [BSS17]

kreuzen. In der Aufnahmeebene liegen sie noch nah genug beieinander, um im Zerstreuungskreis zu liegen. Werden nun wie in 4(b) die äußeren Wege, die das Licht nehmen könnte von der Blende blockiert, so landen auch noch Strahlen von weiter entfernten Punkten als dem ehemaligen Fernpunkt innerhalb des Zerstreuungskreises. Analog wandert auch der Nahpunkt weiter nach vorne. Die Schärfentiefe ist also größer als bei einer weit geöffneten Blende (4(c)), bei der Lichtstrahlen in einem größeren Winkel auf das Objektiv treffen können. [BSS17]

Für eine möglichst geringe Schärfentiefe sollte also ein Objektiv mit weiter Brennweite gewählt werden. Die Blendenzahl sollte so gering wie möglich (möglichst weite Blendenöffnung) eingestellt werden und die Kamera nah an der entsprechenden Szene platziert werden.

Eine weitere Einstellmöglichkeit bietet die *Lichtempfindlichkeit* des Sensors. Diese kann man über den ISO-Wert¹ einstellen. Typische ISO-Angaben sind zum Beispiel 100, 200 und 400. Je höher der Wert, desto empfindlicher ist der Sensor eingestellt und desto weniger Licht wird zum Fotografieren benötigt. Ein hoher ISO-Wert ist in diesem Fall nicht erforderlich, daher wird er möglichst gering gewählt. So wird das Rauschen im Bild gering gehalten. [BSS17]

¹Sie wird nach einem Standard der internationalen Organisation für Normung bestimmt

Die letzte wichtige Einstellmöglichkeit ist die *Belichtungszeit*. Diese gibt an, wie lange der Verschluss geöffnet bleibt und der Sensor dem Licht ausgesetzt ist. Eine kurze Belichtungszeit ist dann von Vorteil, wenn sich das Motiv oder die Kamera bewegt. Eine lange Belichtungszeit ist hingegen bei sehr dunklen Szenen erforderlich, um genug Licht einzufangen. [Goc12] Da alle anderen Größen bereits festgelegt sind, muss diese den Beleuchtungsverhältnissen angepasst werden, um zu verhindern, dass das Bild unter- oder überbelichtet wird.

Um den Fokus zu wählen, wird an einem Ring am Objektivgehäuse gedreht, was eine Bewegung der Linsen im Inneren des Objektivs bewirkt. Alternativ kann die Fokuseinstellung auch durch einen kleinen Elektromotor vorgenommen werden. Dies ist zum Beispiel bei der Verwendung der Autofokusfunktion der Fall.

3 Verwandte Arbeiten

Im Bereich der Tiefenrekonstruktion aus einer einzigen Perspektive lassen sich im wesentlichen zwei Arten von Systemen unterscheiden. Bei *aktiven Systemen* wird in irgendeiner Form Energie ausgesendet. Einfache Beispiele hierfür wären Sonar, Radar oder Lidar. Sie senden Schall- beziehungsweise Radiowellen oder Laserstrahlen aus und messen die Zeit nach der sie von einem Objekt zurückgeworfen und vom System detektiert worden sind. Aus der Laufzeit und der jeweiligen Ausbreitungsgeschwindigkeit lässt sich der Abstand zum Objekt ausrechnen. [98]

Ein Beispiel für ein eher visuelles aktives System ist das Beleuchten der Szene mit einem bekanntem Muster, häufig im infraroten Bereich. Durch die unterschiedlich weit entfernten Objekte wird das Muster entsprechend verzerrt, woraus die Tiefe rekonstruiert werden kann. [FSV04]

Um Tiefeninformationen aus Schärfe beziehungsweise Unschärfe zu gewinnen gibt es verschiedene Ansätze, wovon die meisten zu den passiven Systemen zählen. Das bedeutet sie nehmen keinen Einfluss auf die Szene, sondern nutzen lediglich das von ihr ausgesandte Licht.

Schon 1985 hat P. Grossmann einen Algorithmus beschrieben, um aus nur einem einzigen Bild mit geringer Schärfentiefe Tiefeninformationen zu gewinnen.[Gro87] Dafür findet er zunächst Kanten im gegebenen Bild mittels des Marr-Hildreth-Operators. Anschließend misst er deren Verschwommenheit mithilfe der ersten Ableitung des Bildbereichs orthogonal zur Kante. Je verschwommener die Kante ist, desto weiter muss sie von der Fokusebene entfernt sein. Aus dem so gewonnenen Abstand zur Fokusebene lässt sich mit der Fokusdistanz die Entfernung zur Kamera ausrechnen. Um zu bestimmen auf welcher Seite der Fokusebene sich die untersuchte Kante befindet, wird entweder ein zweites Bild mit anderer Fokusdistanz benötigt oder es muss darauf geachtet werden, dass bei der Aufnahme des Bildes alle Objekte näher oder weiter entfernt sind als die Fokusdistanz.

Ebenfalls mit nur einem Bild kommt die *Dual Aperture Photography* aus. [Mar+15] Hierbei handelt es sich um eine besondere Kamera für Mobilgeräte, die über eine zusätzliche kleinere Blende für infrarotes Licht und einen entsprechenden Bildsensor verfügt. Dadurch ist die Schärfentiefe im Infrarotkanal deutlich größer als im Farbraum. Über die unterschiedlich starke Verschwommenheit in Farb- und Infrarotbild lässt sich hier ähnlich wie bei Grossmann die Tiefeninformation ableiten. Die Erstellung eines komplett scharfen Bildes ist ebenfalls möglich.

Im Gegensatz zu den Methoden, die nur ein oder zwei Bilder verwenden stehen solche, die komplette Serien von Bildern verwenden. Eine solche ist die von Martel et al. vorgestellte Methode.[Mar+17] Hierbei wird ebenfalls spezielle Hardware eingesetzt. Ein programmierbarer Bildsensor führt Berechnungen auf dem gerade aufgenommenem Bild hoch parallelisiert aus und gibt nur das Ergebnis an die CPU weiter, wodurch die limitierte Bandbreite auf diesem Weg umgangen wird. Außerdem wird eine Linse verwendet, deren Brennweite durch entsprechend einstellbare Verformung sehr schnell verändert werden kann. Die Fokusdistanz ist so deutlich schneller änderbar als es bei einem normalen Objektiv der Fall wäre. Die Fokusebe-

ne kann mit einer Frequenz von bis zu einem kHz verschoben werden. Während die Fokusdistanz verändert wird, werden fortlaufend Bilder aufgenommen und direkt die Schärfe pixelweise über lokale Kontrastmessung bestimmt. Dabei wird sich der Zeitpunkt gemerkt zu dem ein bestimmtes Pixel am schärfsten ist. Über die Fokusdistanz zu diesem Zeitpunkt kann die Objektdistanz bestimmt werden. Hier wird ein Schwellwert eingesetzt über dem die Schärfe liegen muss, damit das Pixel überhaupt verarbeitet wird. So entsteht eine lückenhafte Tiefenmap mit bis zu 128 Tiefenebenen. Wenn die z-Auflösung auf 32 Ebenen reduziert wird ist eine Generierung der Tiefenmap mit über 25 Bildern die Sekunde möglich.

Kontrastmessung wird auch häufig für die Autofokusfunktion von Digitalkameras als Fokusmaß verwendet. [BSS17]

Ohne besondere Hardware kommen Safae-Rad und Aleksic aus.[SA08] Sie verwenden herkömmliche Handy-Kameras, um eine niedrig aufgelöste Tiefenmap zu erstellen. Ihr Hauptziel ist die Beschleunigung des Autofokusprozesses. Es wurden acht verschiedene Linsenpositionen im Vorfeld festgelegt. Um ein Bild zu fokussieren, werden diese so durchsucht, dass möglichst schnell die optimale gefunden werden kann. Die verwendeten Linsenpositionen haben dabei nicht alle den selben Abstand, sondern sind so gewählt, dass die entsprechenden Fokusebenen äquidistant sind. Zur Generierung der Tiefenmap werden mit allen Linsenpositionen Bilder aufgenommen und die Schärfe jeweils regionenweise bewertet. Das Bild wird dazu in 5x5 Regionen aufgeteilt. Die Tiefenmap kann bei der Entscheidung herangezogen werden, ob der Blitz eingeschaltet wird oder nicht. Wenn in einer Region zu wenig Kanten liegen, um die Tiefe sinnvoll zu bestimmen, so wird dieser Region keine Tiefe zugeordnet und die entstehende Tiefenmap ist löchrig. Hierbei wird das **durchschnittliche Fokusmaß** der Region zur Bewertung genutzt.

Einen besonderen Wert auf die Robustheit gegenüber Rauschen legen Pertuz et al. bei der Erstellung eines komplett scharfen Bildes aus einer Serie von Bildern mit geringer Schärfentiefe.[Per+12] Zur pixelweisen Fokusbewertung setzen sie die Varianz des Graustufenbildes in der Umgebung des untersuchten Pixels ein. Sie betrachten die Funktion des Fokus über den Index des Bildes als eine Überlagerung aus Gaußfunktion und Rauschen. Es wird daher eine Gaußkurve an die Fokuskurve angepasst. Die Differenz der beiden Kurven wird anschließend zur Bewertung der Qualität dieses Pixels herangezogen. Um ein Pixel des komplett scharfen Bildes zu bestimmen wird der gewichtete Durchschnitt dieses Pixels in den originalen Bildern bestimmt. Die Gewichte sind dabei von der angepassten Gaußfunktion und dem Qualitätsmaß abhängig. Bei geringer Qualität wird weniger auf die Erhaltung von Details, sondern mehr auf die Unterdrückung von Rauschen geachtet.

Die in dieser Arbeit verwendete Methode zählt ebenfalls zu den passiven Systemen und verwendet, wie Martel et al., Safae-Rad und Aleksic sowie Pertuz et al. eine Serie aus mehreren Bildern. Da eine möglichst gute Tiefenauflösung erreicht werden soll, werden hier sehr viele Bilder verwendet. Während bei der Verwendung des programmierbaren Bildsensors aus Hardwaregründen nur maximal 128 Tiefenebenen möglich sind und sich die Tiefenmap der Handykamera aus Laufzeitgründen auf 8 Bilder beschränkt, werden hier bis zu 1000 Bilder verarbeitet.

Um die Auflösung in X- und Y-Richtung möglichst hoch zu halten, wird wie bei Martel et al. und Pertuz et al. der Fokus pixelweise bestimmt und nicht wie bei Salfaee-Rad und Aleksic die Eingangsbilder in Regionen aufgeteilt.

Die größere Menge an insgesamt zu betrachtenden Pixeln (Auflösung * Bilderanzahl) bedeutet einen entsprechend höheren Rechenaufwand. Während die vorgestellten Arbeiten lediglich Sekundenbruchteile (Handykameras und programmierbarer Bildsensor) oder Sekunden (rausch-resistenter Algorithmus) benötigen, liegt die Berechnungszeit hier bei circa zwei Minuten².

Als Fokusmaß wird in dieser Arbeit der von Pertuz et al. vorgeschlagene Gradient des Bildes verwendet. Er wird mithilfe des Sobel-Filters bestimmt.

Wie in den vorgestellten Arbeiten, wird auch hier eine Bewertung der Qualität der untersuchten Pixel beziehungsweise Regionen vorgenommen. So wohl die von Martel et al. genutzte Variante über das Maximum der Fokusbewertung als auch die von Salfaee-Red und Aleksic vorgeschlagene Variante über die durchschnittliche Fokusbewertung wurden getestet.

²Auf einem Intel i5-7200U bei 2.5 GHz

4 Erstellung der Bilderserien

Zur Aufnahme der Bilder stand eine Nikon D 5500 mit zwei verschiedenen Objektiven zur Verfügung. Das AF-S DX NIKKOR 35 mm Objektiv und das AF-S DX NIKKOR 18–55 mm Objektiv standen zur Auswahl. Die große Brennweite spricht zwar eher für das 18-55 mm Objektiv, allerdings lässt sich bei dem 35 mm Objektiv die Blende weiter öffnen ($f1.8$ im Gegensatz zu $f3.5$). Dadurch können letztlich geringere Schärfentiefen erzeugt werden. Die Kamera wird also mit dem 35 mm Objektiv auf einem Stativ möglichst dicht an der zu fotografierenden Szene aufgestellt. Über eine USB-Verbindung mit einem Computer lässt sich die Kamera mittels des Nikon SDKs steuern.

4.1 Einschränkungen durch die verwendete Hard- und Software

Die Kamera lässt sich in zwei Modi betreiben, zwischen denen frei per Software umgeschaltet werden kann. Im normalen Modus ist der Spiegel herunter geklappt und das Bild ist im Okular sichtbar. Im Live-View-Modus ist der Spiegel dauerhaft nach oben geklappt, sodass das Licht auf den Sensor fallen kann. Hier kann man sich das aktuelle Bild auf dem Display der Kamera anzeigen lassen.

Um automatisch ein Foto aufzunehmen, gibt es zwei Möglichkeiten. Im Live-View-Modus lässt sich ein Vorschau-Bild mit relativ geringer Auflösung von 640×424 Pixeln direkt auf dem angeschlossenen PC abspeichern. Alternativ kann der normale Auslöseprozess aktiviert werden. Dabei klappt der Spiegel kurz hoch, um den Sensor zu belichten und anschließend wieder herunter. Befindet sich die Kamera dabei im Live-View-Modus, so wird davor der Spiegel herunter und nach der Aufnahme wieder nach oben geklappt. Somit entfallen vier Spiegelbewegungen auf eine Aufnahme. Die so aufgenommenen Fotos werden standardmäßig auf der SD-Karte abgespeichert.

Der Fokus lässt sich elektronisch schrittweise verschieben. Die mögliche Schrittweite kann zwischen 1 und 32767 gewählt werden. Allerdings zeigt eine Schrittweite von weniger als 10 keine Wirkung. Daher wird eine Schrittweite von 15 gewählt. Größere Schrittweiten werden nicht verwendet, da die Fokusebenen der einzelnen Bilder nah beieinander liegen sollen um möglichst genaue Tiefeninformationen zu erhalten. Um bei einer Schrittweite von 15 den Fokus von der nächsten Einstellung auf Unendlich zu verschieben sind knapp 1000 Schritte nötig. Zwischen jedem Schritt soll ein Foto aufgenommen werden. Somit liegen, wenn man von einer linearen Schrittweite ausgeht, circa 15000 Einzelschritte zwischen den beiden Anschlägen. Eventuell werden die deutlich größeren Schrittweiten für andere Objektive verwendet. Die Fokusdistanz kann per Software nur im Live-View-Modus und bei richtiger Autofokus-Einstellung verändert werden. Der Fokusmodus darf nicht MF (Manueller Fokus) oder AF-F (Permanenter Autofokus) sein. Da die Veränderung der Fokusdistanz nur im Live-View-Modus möglich ist, werden bei normaler Auslösung ungefähr 4000 Spiegelbewegungen verursacht. Dies hätte eine starke Abnutzung der Mechanik zur Folge. Daher wird nur die Möglich-

keit der Vorschau-Bilder genutzt. Dies hat neben der bereits erwähnten niedrigen Auflösung den Nachteil, dass die Fotos im JPEG-Format gespeichert werden. Die entsprechende Kompression ist nicht verlustfrei und es entstehen Artefakte, die später in dieser Arbeit noch sichtbar werden.

4.2 Aufnahme der Bilder

Nachdem die Kamera positioniert, der Fokusmodus richtig eingestellt, die Blende komplett geöffnet und die Belichtungszeit angepasst wurde, kann das in Algorithmus 1 beschriebene Programm gestartet werden.

Algorithmus 1 Berechnung des durchschnittlichen Gradienten

```
1: select Device
2: select Outputfolder
3: turn on LiveView
4: drive focus close
5: driveStep = 15
6: for i = 1 to 1000 do
7:   take PreviewImage
8:   save as "Image%04d.jpg", i
9:   drive focus
10:  wait
11: end for
```

Zunächst muss das richtige Gerät ausgewählt werden, für den Fall, dass mehr als nur die USB-Schnittstelle der Kamera erkannt wurde. Unter dem anschließend eingegebenen Pfad werden die aufgenommenen Bilder gespeichert. Nachdem die Kamera in den Live-View-Modus versetzt wurde, wird die Fokusdistanz so gering wie möglich eingestellt und die Schrittweite auf 15 gesetzt. In der Schleife werden nun 1000 Fotos aufgenommen, indem zunächst das Live-View-Bild gespeichert wird. Anschließend wird der Fokus um die eingestellte Schrittweite verstellt. Da dies eine mechanische Bewegung ist, dauert es einige Zeit bis die Veränderung abgeschlossen ist. Darauf muss im Programmfluss Rücksicht genommen und gewartet werden.

Die Erstellung der 1000 Bilder dauert ungefähr ein bis zwei Minuten. Es ist darauf zu achten, dass sich in dieser Zeit das Motiv nicht verändert. Ansonsten kann die Auswertung der Bilder nicht korrekt funktionieren.

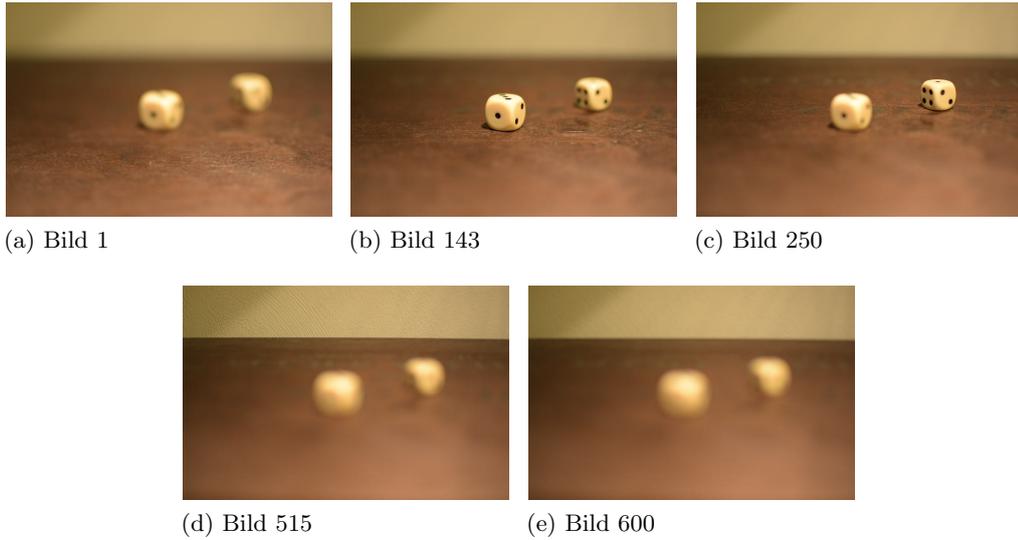


Abbildung 5: Einige Bilder der Würfelserie

Beispielhaft ist in Abbildung 5 eine Auswahl der Bilderserie mit zwei Würfeln dargestellt. Im ersten Bild (5(a)) ist ein Streifen vor dem linken Würfel scharf. Der Bereich vor diesem ist in jedem Bild unscharf, da die Kamera so nah nicht fokussieren kann. Der vordere der beiden Würfel ist im 143. Bild (5(b)) am schärfsten und der hintere in Bild 250 (5(c)). Die Tapete im Hintergrund ist um Bild 515 (5(d)) im Fokus. Spätestens ab Bild 600 (5(e)) ist die komplette Szene unscharf, da hier die Fokusdistanz größer ist als der Abstand zur Wand. Die darauffolgenden Bilder 601 bis 1000 können für diese Szene problemlos gelöscht werden. Für keinen Punkt der Szene wird eines von ihnen am schärfsten sein.

5 Verarbeitung der Bilder

Die Verarbeitung der Bilder ist fast komplett von der Aufnahme entkoppelt. So können theoretisch auch Bilderserien aus anderen Quellen analysiert werden. Der einzige Punkt, der bei der Verarbeitung angepasst werden muss, ist die Berechnung des Abstandes aus dem Index des Bildes, wie sie in Kapitel 5.5 erklärt ist. Eine entsprechende Anpassung wäre auch bei einer anderen Schrittweise des Fokus nötig.

Für die Auflösung der entsprechenden Bilderserien gibt es keine Anforderungen. Lediglich die Dateinamen müssen dem Muster folgen, in dem sie auch bei der Erstellung abgespeichert wurden: "Image0001.jpg", "Image0002.jpg", ... Die so in den Dateinamen gespeicherten Indizes werden für die Verarbeitung benötigt.

Es wird die CImg-Library [Tsc18] verwendet. Hierbei handelt es sich um eine Template-Bildverarbeitungs-Bibliothek für C++. Bilder mit bis zu drei Dimensionen und mehreren Farbkanälen können hier sehr einfach verarbeitet werden. Dabei werden sie wie Matrizen behandelt und es sind einige hilfreiche Funktionen auf ihnen definiert. Benutzt werden in dieser Arbeit unter anderem das punktweise Addieren, Dividieren, Quadrieren und die Wurzelberechnung. Die Koordinaten der Bilder starten jeweils in der oberen linken Ecke mit (0, 0) und werden nach unten rechts hin größer bis (Breite - 1, Höhe - 1).

5.1 Schärfemaß

Um zu bestimmen in welchem der Bilder ein bestimmtes Pixel am schärfsten ist, wird ein Maß für die Schärfe benötigt. Dafür wird der Gradienten jedes Bildes berechnet. Der Gradienten eines Bildes gibt einfach ausgedrückt an, wie sehr sich das Bild an einer bestimmten Stelle verändert. Die Gradientenberechnung ist mittels Sobel-Filtern implementiert.

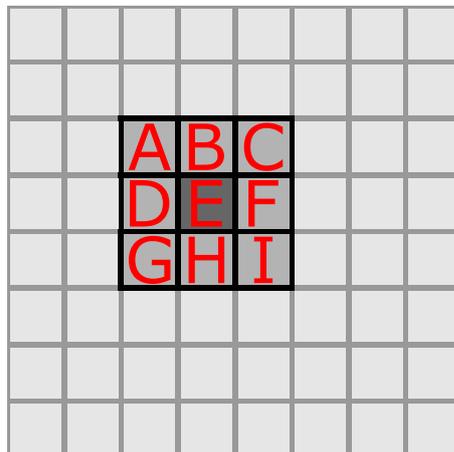


Abbildung 6: Allgemeiner Filter nach [Gro18]

Ein Filter berechnet für jedes Pixel den gewichteten Mittelwert seiner Nachbarn. Die Anwendung eines allgemeinen 3x3-Filter ist in Abbildung 6 dargestellt. [Gro18]

Dem Pixel E in der Mitte wird die gewichtete Summe der Pixel A bis I zugewiesen. Die Gewichte hängen von dem jeweiligem Filter ab und können für diesen frei gewählt werden. Um einen Filter zu definieren, wird eine Matrix mit den entsprechenden Gewichten angegeben. Für den Sobel-Filter sind diese Matrizen in Abbildung 7 dargestellt. [Gro18]

-1	0	+1
-2	0	+2
-1	0	+1

-1	-2	-1
0	0	0
+1	+2	+1

(a) für vertikale Kanten

(b) für horizontale Kanten

Abbildung 7: Sobel-Filter nach [Gro18]

Der Filter auf der linken Seite detektiert vertikale Kanten. Seine Ausgabe entspricht dem Gradienten entlang der X-Richtung. Für die Y-Richtung wird die Matrix auf der rechten Seite verwendet. Pixel am Rand des Bildes werden separat behandelt, da sie nicht an jeder Seite Nachbarn haben. [Gro18] Bei der Implementierung in der CImg-Library wird die Neumann-Randbedingung verwendet. Dabei wird für Pixel außerhalb des Bildes der Wert des nächstgelegenen tatsächlich existierenden Pixels angenommen. [Tsc18]

Bei der Anwendung auf ein Farbbild B werden die Gradienten auf dem Rot-, Grün und Blau-Kanal unabhängig voneinander berechnet. Anschließend werden die X- und Y-Gradienten miteinander kombiniert, indem sie zunächst quadriert werden, um anschließend die Wurzel aus der Summe zu ziehen:

$$G_{\text{coloured}} = \sqrt{\left(\frac{\delta B}{\delta x}\right)^2 + \left(\frac{\delta B}{\delta y}\right)^2} \quad (1)$$

Dies entspricht der Berechnung der Länge des Gradientenvektors. Die Richtung geht dabei verloren. Die Stärke des Gradienten bleibt erhalten. Bei einer einfachen Addition der beiden Gradienten könnten sich ein positiver Gradient in X-Richtung und ein negativer Gradient in Y-Richtung gegenseitig auslöschen.

Um die drei Farbkanäle zu einem Graustufenbild zusammenzufassen, werden diese addiert und das Ergebnis durch drei geteilt:

$$G = \frac{G_{\text{Red}} + G_{\text{Green}} + G_{\text{Blue}}}{3} \quad (2)$$

Der Ablauf der Gradientenberechnung ist im Algorithmus 2 beschrieben. Alle Operationen sind pixelweise zu verstehen.

Algorithmus 2 Gradientenberechnung

```
1: function COMPUTEGRADIENT(image)
2:   gradient_x = image.get_gradient("x")
3:   gradient_y = image.get_gradient("y")
4:   G_coloured = (gradient_x.sqr() + gradient_y.sqr()).sqr()
5:   G = (G_coloured.red() + G_coloured.green() + G_coloured.blue()) / 3
6:   return G
7: end function
```

Die Anwendung ist in Abbildung 8 beispielhaft am Würfelbild dargestellt. Der Gradient in X- und Y-Richtung des Originalbildes 8(a) sind in 8(b) beziehungsweise 8(c) dargestellt. Grau bedeutet hier Null. Hellere Werte stehen für einen positiven und dunklere Werte für einen negativen Gradienten. Die beiden Gradienten werden dann zu dem Farbbild $G_coloured$ (8(d)) wie oben beschrieben zusammengeführt. Da sich die Kanten in allen drei Farbkanälen ähnlich stark wiederfinden, ist in 8(d) sehr wenig Farbe zu erkennen. Das Bild darf für die weitere Verarbeitung nur einen Farbkanal haben, daher wird es in das in 8(e) zu sehende Graustufenbild G umgewandelt. Die Bilder der kombinierten Gradienten sind dabei in der gesamten Arbeit zur besseren Darstellung invertiert.

Die Schärfentiefe im Originalbild war so eingestellt, dass der linke Würfel scharf ist. Daher sind seine Kanten im Ergebnis auch am stärksten zu sehen. Ebenfalls wurden Kanten in der Struktur des Bodens in der scharfen Ebene detektiert. Der rechte Würfel ist im Originalbild unscharf, somit sind seine Kanten verwischt und im Resultat deutlich schwächer als die des linken Würfels.

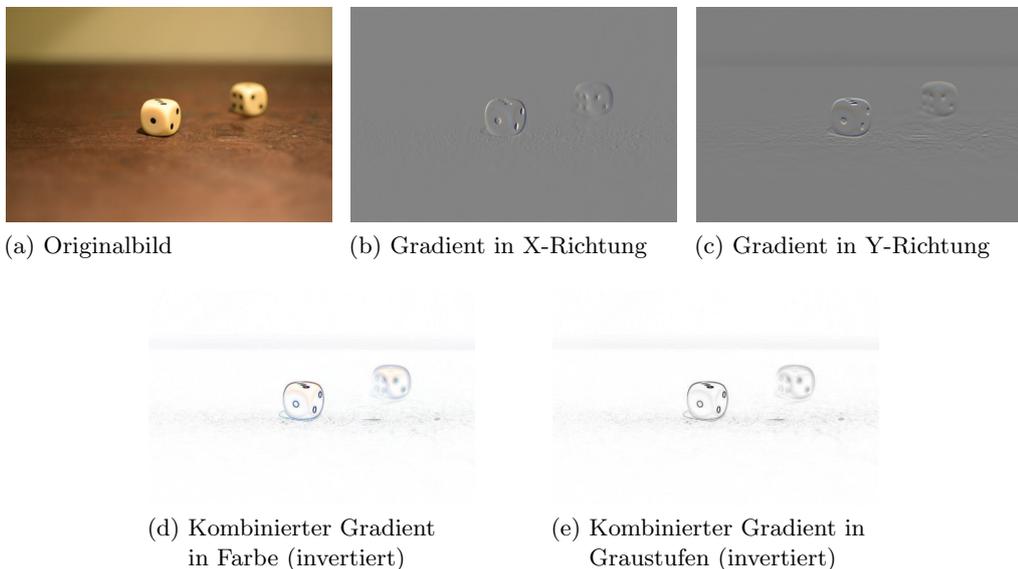


Abbildung 8: Gradientenberechnung am Beispiel des Würfelbildes

Im weiteren Verlauf der Arbeit ist mit dem Gradienten eines Bildes das Ergebnis des eben beschriebenen Algorithmus gemeint. Technisch korrekt wäre dies allerdings die Stärke des Gradienten.

In Abbildung 9 ist die Schärfentiefe anders gewählt. Hier ist der rechte Würfel klar zu erkennen und der linke unscharf. Im Gradienten ist daher auch der rechte Würfel deutlicher vertreten als der linke. Den Pixeln des linken Würfels sollte also die Entfernung des Bildes aus Abbildung 8 zugeordnet werden und die Pixel, die den rechten Würfel ausmachen, sollten die zu Abbildung 9 gehörende Entfernung erhalten.



(a) Originalbild



(b) Kombiniertes Gradienten in Graustufen (invertiert)

Abbildung 9: Gradient zu einem Bild mit anderer Schärfentiefe

Alternativ zum Sobel-Filter lässt sich ein Gradient mit der CImg-Library auch auf andere Weise bestimmen. Es gibt die Möglichkeiten der Vorwärts-, Rückwärts- und Zentraldifferenzen. Die jeweiligen Filter für vertikale Kanten sind in Abbildung 10 dargestellt. Für horizontale Kanten ist eine Drehung um 90° erforderlich. Der Sobel-Filter führt im Gegensatz zu diesen einfachen Differenzen noch eine Glättung in der Richtung durch, in der er nicht die Kanten erkennt. Die Vorwärts- und Rückwärtsdifferenzen bewirken zudem eine Verschiebung um je ein Pixel nach links und oben beziehungsweise rechts und unten, sodass die Kante im Gradienten nicht exakt dort angezeigt wird, wo sie im Ausgangsbild vorhanden war.

0	-1	1
---	----	---

(a) Vorwärtsdifferenzen für vertikale Kanten

-1	1	0
----	---	---

(b) Rückwärtsdifferenzen für vertikale Kanten

-1	0	1
----	---	---

(c) Zentraldifferenzen für vertikale Kanten

Abbildung 10: Differenzen-Filter

Des weiteren gibt es in der CImg-Library noch die Möglichkeit rotationsinvariante Kerne, den Deriche-Kantendetektor oder Van Vliet-Filter einzusetzen. Ihre Ergebnisse sind jeweils sehr ähnlich zu denen des Sobel-Filters. Sie sind allerdings etwas komplexer als dieser und haben daher eine längere Laufzeit. Während die

rotationsinvarianten Kerne nur etwas mehr Zeit benötigen, dauert die Berechnung mithilfe des Deriche-Kantendetektors anderthalb mal und mithilfe der Van Vliet-Filter sogar doppelt so lange. Da die Gradientenberechnung den größten Teil der gesamten Laufzeit ausmacht, scheint die Verwendung des Sobel-Filters am sinnvollsten. [Gro18] [Tsc18]

5.2 Indexbestimmung

Wenn für jedes Bild in der Serie der Gradient bestimmt wurde, so kann man sich für jedes Pixel anzeigen lassen, wie stark der Gradient an dieser Stelle in den Bildern ist. In Abbildung 11 ist beispielhaft der Gradient über den Index für drei Pixel aufgetragen. Alle drei Pixel befinden sich auf dem linken Würfel. In 11(d) bis 11(e) ist die Position jeweils rot eingekreist. In den Graphen sind oben rechts die Koordinaten des gewählten Pixels zu sehen.

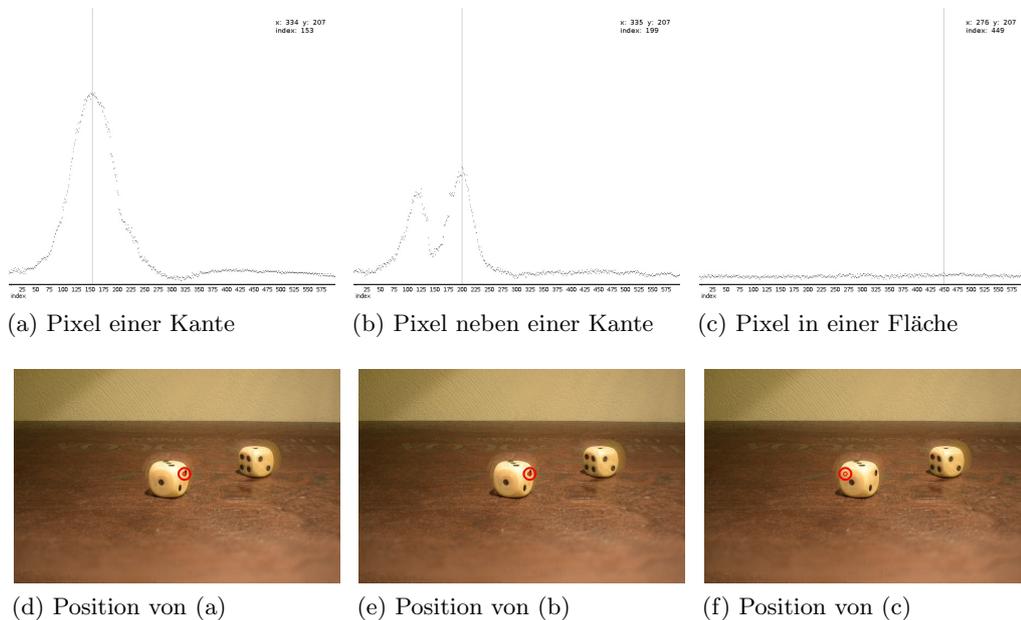


Abbildung 11: Stärke des Gradienten über den Index

In 11(a) wurde ein Pixel gewählt, das zu einer Kante gehört. An dieser Stelle war der Gradient in Bild Nummer 153 am stärksten. Davor und dahinter ist das Pixel weniger scharf, da die Fokus-Ebene vor beziehungsweise hinter der Stelle auf dem Würfel liegt. Der senkrechte Strich markiert das Maximum des Gradienten.

Das Pixel rechts daneben (11(b)) befindet sich neben der Kante. Hier tritt ein Effekt ein, der in Abbildung 12 zu sehen ist. Wenn der Fokus vor oder hinter dem Würfel liegt, so ist das Auge unscharf. Dadurch ist der schwarze Punkt in 12(a) und 12(c) kleiner und in diesen Bildern gibt es eine Kante, die etwas weiter rechts liegt als in 12(b). Bei Betrachtung des Pixels knapp innerhalb des Auges am linken



Abbildung 12: Ausschnitt aus der Bilderserie

Rand in Abbildung 11(b) stellt man somit fest, dass der Gradient an dieser Stelle bei einem scharfen Bild (12(b)) geringer ist, als bei einem unscharfen Bild (12(a) und 12(c)).

Das dritte beispielhaft ausgewählte Pixel in Abbildung 11 liegt inmitten einer einfarbigen Fläche. In 11(c) ist sehr gut zu erkennen, dass hier in keinem Bild ein nennenswerter Gradient vorhanden ist. Der Gradient unterliegt lediglich einem gewissen Rauschen. Aus diesen Informationen lässt sich keine sinnvolle Aussage über die Entfernung des entsprechenden Punktes machen. In Kapitel 5.3 Pixelgüte wird die Möglichkeit besprochen Pixel dieser Art von der Auswertung auszuschließen.

5.2.1 Maximaler Gradient

Die einfachste und offensichtlichste Möglichkeit für ein Pixel einen Index aus der Bilderreihe auszuwählen, ist das Bild zu wählen, in dem der Gradient für dieses Pixel am größten ist. Diese Methode wurde auch in Abbildung 11 angewendet. Der senkrechte Strich markiert jeweils den gewählten Index. Zusätzlich steht er auch in der oberen rechten Ecke.

Für Pixel, die zu einer Kante gehören, funktioniert dieser Ansatz sehr gut. Dem Pixel aus Abbildung 11(a) wurde Bild Nummer 153 (siehe Abbildung 12(b)) zugeordnet. In diesem Bild ist das Auge des Würfels auch scharf gestellt.

Wie der senkrechte Strich in Abbildung 11 erkennen lässt, liefert hier die Methode des maximalen Gradienten den falschen Index. Der Graph weist aufgrund der eben besprochenen Verschwommenheit zwei lokale Maxima auf. Das erste befindet sich bei Index Nummer 120 und das zweite, welches gleichzeitig das globale Maximum ist, bei Index Nummer 199. Die entsprechenden Würfelausschnitte sind in Abbildung 12(a) und 12(c) abgebildet. In beiden Bildern ist der Würfel unscharf. Eine mögliche Lösung dieses Problems wird in Abschnitt 5.2.2 Gewichtete Indizes besprochen.

Für Pixel inmitten einer Fläche, wie den in Abbildung 11 gewählten, erhält man über das Maximum des Gradienten einen eher zufälligen Index. Dieser ist offensichtlich falsch und daher nicht für die Tiefenrekonstruktion zu gebrauchen.

5.2.2 Gewichtete Indizes

Um für Pixel neben einer Kante ein besseres Ergebnis zu erhalten wurde eine weitere Methode der Indexbestimmung ausprobiert. Hierbei wird für ein Pixel der Index jedes Bildes mit dem jeweiligen Gradienten gewichtet und so ein gewichteter Durchschnitt ermittelt. Dafür wird beim Iterieren über die Gradienten für jedes Pixel der Wert von `sum` um `index·gradient` und `divider` um den Gradienten erhöht. So kann am Ende mit `sum/divider` der entsprechende Index errechnet und in `indexmap` gespeichert werden. Dieser Prozess ist in Algorithmus 3 beschrieben. Die Hoffnung ist, dass dadurch der gewählte Index zwischen den beiden Maxima liegt.

Algorithmus 3 Indexberechnung über gewichtete Indizes

```
1: for i = 0 to numberOfImages do
2:   image = images[i]
3:   gradient = calculateGradient(image)
4:   for each pixel (x, y) do
5:     sum(x, y) += i * gradient(x, y);
6:     divider(x, y) += gradient(x, y);
7:   end for
8: end for
9: for each pixel (x, y) do
10:  if divider(x, y) > 0 then
11:    indexmap(x, y) = sum(x, y) / divider(x, y);
12:  end if
13: end for
```

So ergeben sich für die oben gewählten Pixel die in Abbildung 13 dargestellten Indizes. Wie in 13(b) zu sehen, ist der gewünschte Effekt nicht eingetreten. Der gewählte Index liegt sogar weiter vom Optimum entfernt als bei Verwendung des Auswahlkriteriums des maximalen Gradienten. Der Grund dafür liegt in den Bildern mit höheren Indizes. In diesen ist der Gradient für das gewählte Pixel nicht so stark, weswegen ein Index relativ schwach gewichtet wird. Dafür gibt es allerdings mehr Bilder hinter dem Optimum als davor, sodass der Einfluss durch die Anzahl wieder steigt. Im nächsten Abschnitt wird daher eine Möglichkeit vorgestellt den Einfluss der weit vom Optimum entfernten Indizes zu verringern.

Für Pixel auf einer Kante ist der ausgewählte Index aus dem selben Grund nun nicht optimal. Abbildung 13(a) zeigt, dass er ebenfalls weiter in Richtung Mitte wandert.

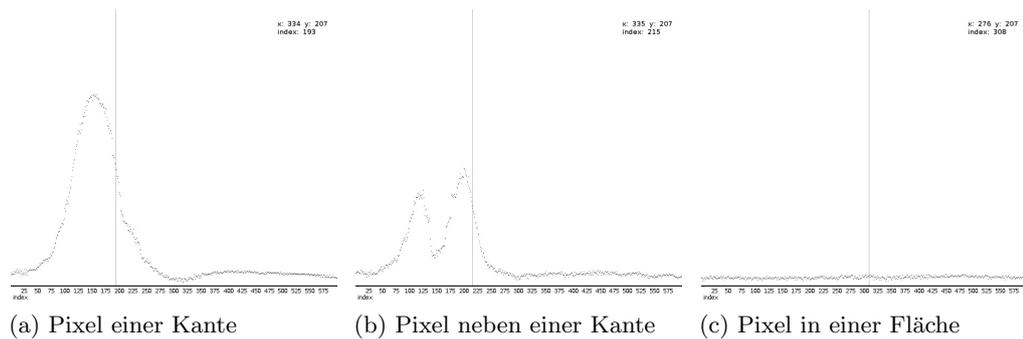


Abbildung 13: Indexbestimmung über gewichtete Indizes

Wenn der Gradient bei jedem Index ungefähr gleich stark ist, wie in Abbildung 13(c), so wird erwartet, dass sich der über diese Methode bestimmte Index in der Nähe der Hälfte des größten Indexes befindet.

5.2.3 Gewichtete Indizes mit Schwellwert

Nun soll der Einfluss von Indizes, die das Ergebnis hauptsächlich durch ihre Existenz beeinflussen und nicht weil sie besonders stark gewichtet werden, verringert werden. Um das zu erreichen werden für den gewichteten Durchschnitt nur noch Indizes berücksichtigt, bei denen der Gradient über einem gewissen Schwellwert liegt. Dafür wird bei Algorithmus 3 in Zeile 5 eine If-Abfrage eingefügt:

If gradient(x, y) > threshold **then**

...

Wenn für ein Pixel in keinem Bild der Gradient über dem Schwellwert liegt, so kann für dieses Pixel kein Index berechnet werden. Die so entstehenden Löcher in der Indexmap, die Pixelkoordinaten auf Indizes abbildet, werden später interpoliert. Die Details werden in Kapitel 5.3.4 besprochen.

Die Güte der Ergebnisse dieser Methode hängt von dem gewählten Grenzwert ab. In Abbildung 14 sind für die Pixel auf und neben einer Kante wieder der Gradient über dem Index aufgetragen. Der senkrechte Strich markiert jeweils den bestimmten Index für einen kleinen, mittleren und großen Schwellwert (30, 50 und 80).

Der Schwellwert, der in Abbildung 14(a) und 14(d) gewählt wurde, ist zu gering. Für das Pixel neben einer Kante (14(d)) gibt es keine Verbesserung im Vergleich zur Methode des maximalen Gradienten. In 14(a) kann man sogar eine Verschlechterung erkennen.

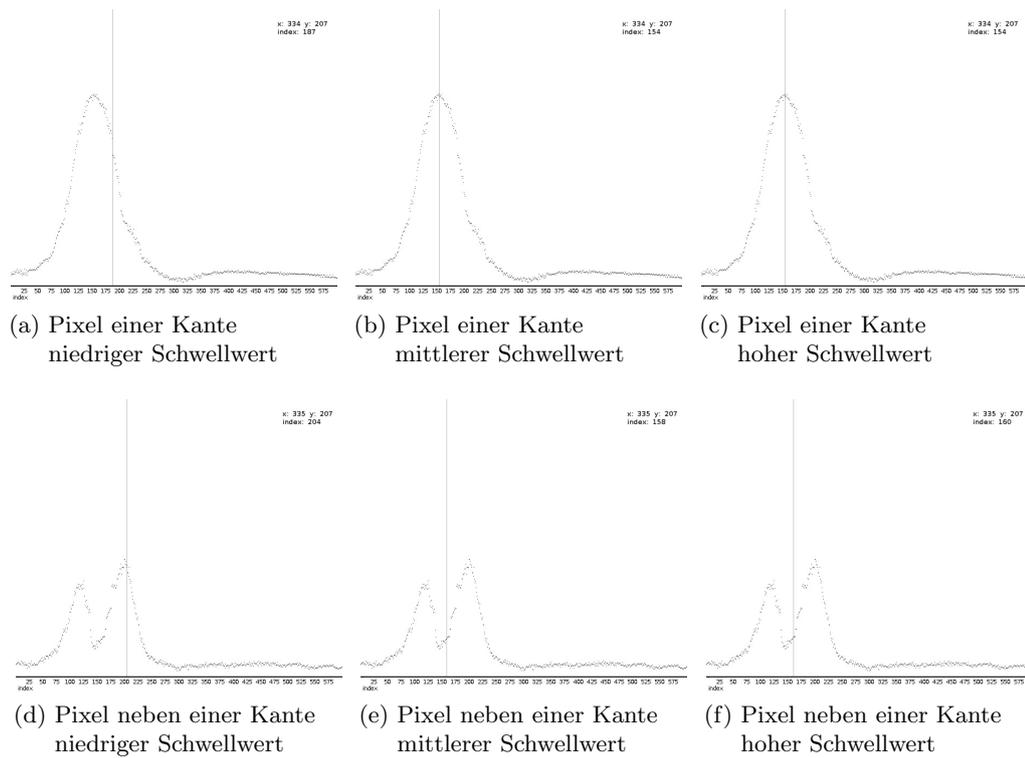


Abbildung 14: Indexbestimmung über gewichtete Indizes mit verschiedenen Schwellwerten

Ein Schwellwert von 50, wie er in 14(b) und 14(e) gewählt wurde, liefert für Pixel auf einer Kante genauso gute Ergebnisse wie bei der Auswahl nach dem stärksten Gradienten. Für Pixel neben einer Kante liefert die Methode sogar bessere Ergebnisse. In 14(e) liegt der berechnete Index sehr nah an der Mitte der beiden Peaks.

Eine weitere Erhöhung des Schwellwertes auf 80 liefert keine Verbesserung gegenüber dem mittleren Schwellwert. Allerdings sorgt der höhere Schwellwert dafür, dass mehr Pixel keinen Index erhalten, da bei ihnen der Gradient in jedem Bild unterhalb der Grenze liegt. Abbildung 15 zeigt das jeweils rekonstruierte scharfe Bild (mehr zur Rekonstruktion von scharfen Bildern in Kapitel 5.4). Die Löcher werden mit zunehmendem Grenzwert größer. Daher sollte er so klein wie möglich und gleichzeitig so groß wie nötig gehalten werden.



(a) niedriger Schwellwert (b) mittlerer Schwellwert (c) hoher Schwellwert

Abbildung 15: Rekonstruierte scharfe Bilder bei verschiedenen hohen Schwellwerten

In allen drei Bildern in Abbildung 15 kann man um die Würfel eine Art Halo erkennen. Noch deutlicher ist er im interpolierten Bild in Abbildung 17(f) zu erkennen. Dessen Ursache liegt in den zwei Peaks. Der Grund, warum es trotz der Indexbestimmung über gewichtete Indizes einen Halo gibt, ist der folgende:

Je weiter das betrachtete Pixel von der Kante entfernt liegt, desto größer ist der Abstand der beiden Gradientenpeaks. Das erste Bild in Abbildung 16 zeigt den Gradientenverlauf an der Kante, die oben rechts vom linken Würfel liegt. Die weiteren Bilder zeigen den Verlauf für Pixel, die immer weiter oben und somit weiter von der Kante entfernt liegen. In 16(b) liegt der bestimmte Index noch in der Nähe der für den Würfel richtigen Indizes, wobei dieses Pixel eigentlich schon zur Fläche hinter dem Würfel gehört und daher einen größeren Index erhalten sollte.

Mit größerem Abstand zur Kante wandern die Peaks weiter auseinander und der linke wird abgeschnitten, da es keine Bilder gibt, in denen der Fokus näher an der Kamera liegt. So ergibt sich eine Verschiebung der berechneten Indizes nach hinten. In 16(e) ist der linke Peak schon fast komplett verschwunden. Der gewählte Index wird somit nur noch vom rechten Peak beeinflusst und liegt mit 476 hinter dem für diese Fläche richtigem Index (~ 330). Noch weiter von der Kante entfernt werden die Werte wieder normal, da der Blur des Würfels nicht so weit reicht.

Wenn der vordere Peak abgeschnitten wird, wie beim linken Würfel, gibt es einen hellen Halo. Wird der hintere Peak abgeschnitten, wie es bei Objekten weiter hinten in der Szene ist, so werden die bestimmten Indizes zu klein und es gibt einen dunklen Halo wie beim rechten Würfel.

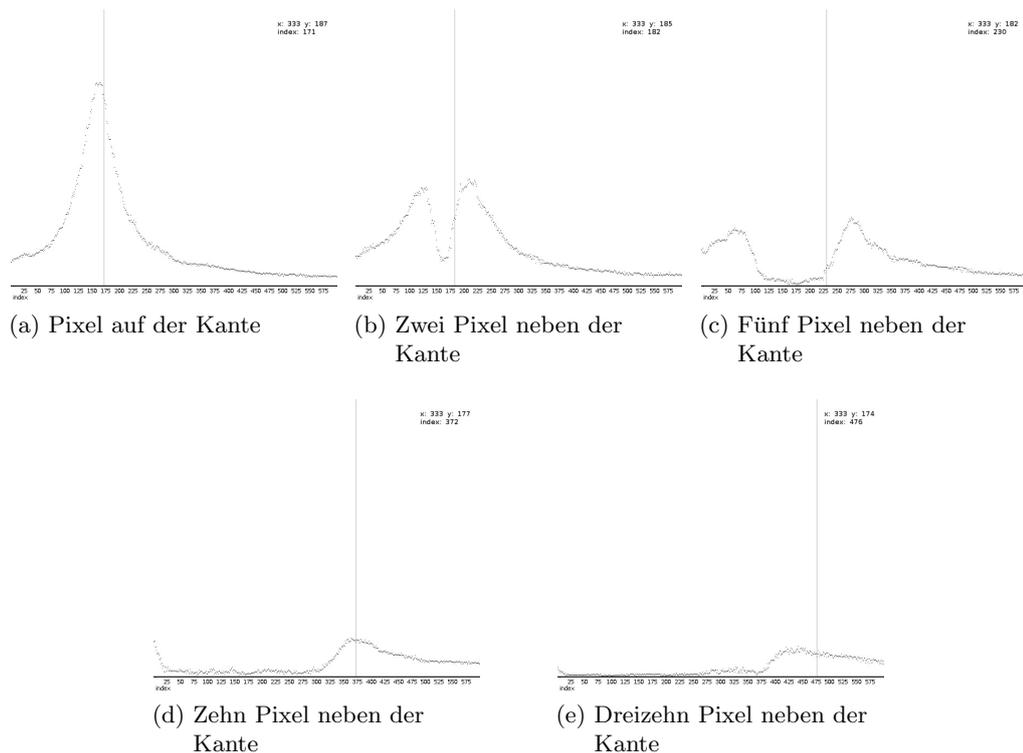


Abbildung 16: Gradient über Index mit Indexbestimmung über gewichtete Indizes mit Schwellwert bei verschiedenen Abständen zur Kante

5.2.4 Vergleich der Methoden zur Indexbestimmung

Abbildung 17 zeigt die Ergebnisse der drei Methoden zur Indexbestimmung. In 17(a) und 17(b) ist jeweils die Indexmap dargestellt, welche die Pixelkoordinaten auf Indizes abbildet. Je höher der gewählte Index ist, desto heller ist das Pixel eingefärbt. Für die in 17(c) abgebildete Indexmap wurden die bei der Generierung entstandenen Löcher bereits gefüllt, so wie es in Kapitel 5.3.4 beschrieben wird. Die Helligkeitsunterschiede sind in der Normalisierung der Bilder begründet. Der kleinste in einer Indexmap vorkommende Index wird schwarz und der größte Index weiß eingefärbt. Alle dazwischenliegenden Indizes erhalten den entsprechenden Grauwert. Wenn sich bei zwei Indexmaps Minimum oder Maximum unterscheiden, so werden die selben Indizes in den beiden Indexmaps nicht auf die selbe Helligkeit abgebildet. Abbildungen 17(d) bis 17(f) zeigen die aus den Indexmaps rekonstruierten scharfen Bilder.

Auf der Indexmap, die mit gewichteten Gradienten erstellt wurde (17(b)), fallen sofort regelmäßige gerade Linien auf. Hierbei handelt es sich vermutlich um Artefakte aus der JPEG-Komprimierung. Auch wenn man diese ignoriert, ist die Fläche auf der die Würfel liegen eher "fleckig". In 17(a) und 17(c) findet man

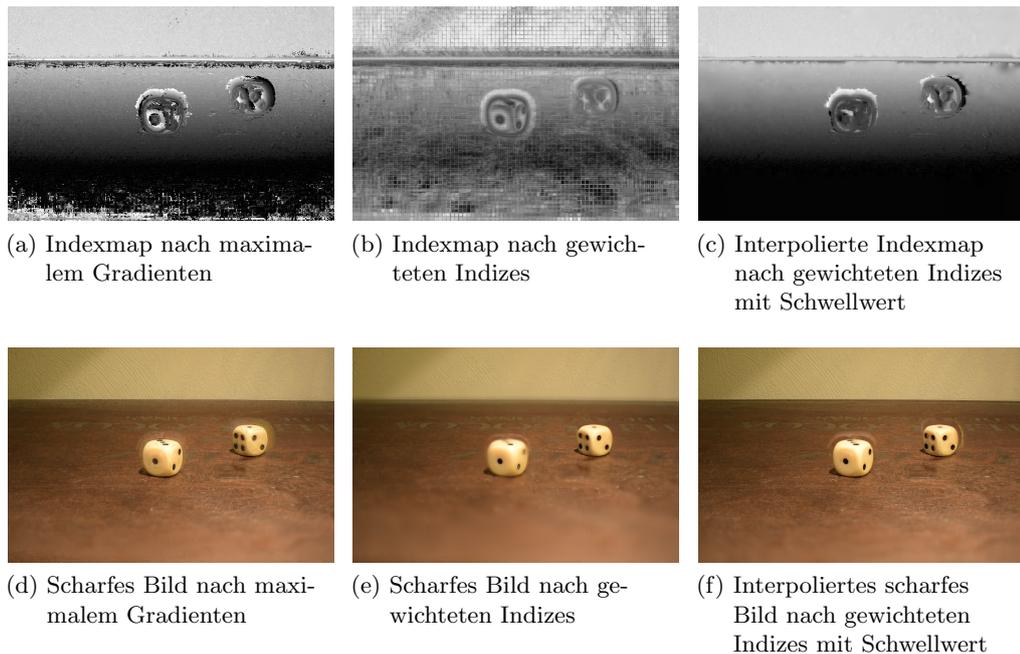


Abbildung 17: Indexmaps und die daraus bestimmten scharfen Bilder nach den drei verschiedenen Methoden zur Indexbestimmung

hingegen einen gleichmäßigen Verlauf von dunkel zu hell. Dies spiegelt sich auch in den scharfen Bildern wieder. Während in 17(d) und 17(e) der Boden durchweg scharf ist (abgesehen vom unteren Rand der Bilder - dieser ist in keinem der Ausgangsbilder scharf), finden sich in 17(e) einige unscharfe Stellen.

Insgesamt sieht die Indexmap nach gewichteten Indizes mit Schwellwerten 17(c) plausibler aus als die nach maximalem Gradienten 17(a). Der Farbverlauf ist glatter und im komplett unscharfen Bereich vorne im Bild gibt es keine extremen Fehler wie in 17(a). Auch am Übergang zwischen Wand und Boden sind die Fehler in 17(c) weniger stark. Eigentlich sollte die Wand durchweg einen Grauton haben und der Boden hinten mit diesem beginnen und nach vorne hin dunkler werden. In 17(c) ist die Kante zur Wand etwas dunkler, sie hat also einen zu geringen Index zugewiesen bekommen. Dies ist in 17(a) ebenfalls der Fall, wobei hier die Indizes noch kleiner sind. Außerdem kommen hier noch zu helle Pixel, also zu große Indizes, dazu.

Betrachtet man die Würfel, so erscheint auch hier das Ergebnis über gewichtete Indizes besser. Die Helligkeitsunterschiede innerhalb eines Würfels sind in Abbildung 17(a) größer als in 17(c). Das heißt die berechnete Tiefe des Würfels hat in 17(a) eine größere Ausdehnung. Vor allem den Pixeln der Fläche um das Auge der Eins wurden zu hohe Indizes zugewiesen. Dies liegt an den zwei Peaks im Gradienten bei Pixeln in der Nähe von Kanten, die in Abschnitt 5.2.2 besprochen

wurden. Bei den Würfflächen wirkt sich dies allerdings nicht sehr stark in den scharfen Bildern aus. Hier lassen sich zwischen 17(d) und 17(f) keine Unterschiede erkennen. Lediglich um die Würfel herum gibt es einen Unterschied. Der Halo ist in 17(f) weniger stark ausgeprägt als in 17(d).

Zur Anfälligkeit bezüglich verrauschter Eingabedaten wurden zwar keine Versuche durchgeführt, aber es ist zu vermuten, dass sich hier die Durchschnittsbildung (mit oder ohne Generalisierung) am besten schlägt. Wenn der Index des maximalen Gradienten gewählt wird, genügt ein Ausreißer, bei dem der Gradient besonders stark ist, um das Ergebnis zu kompromittieren. Der Durchschnitt über mehrere Indizes ist wesentlich stabiler. Pertuz et al. bestimmen zwar keinen Index, aber ihr Vorgehen ist relativ ähnlich zu meiner Methode der gewichteten Indizes. [Per+12]

Die Methode über gewichtete Indizes mit Schwellwert liefert also insgesamt die besten Ergebnisse, aber hat den Nachteil, dass bereits bei der Indexbestimmung einige Pixel verworfen werden. Als nächstes sollen verschiedene Möglichkeiten betrachtet werden bestimmte Pixel von der weiteren Verarbeitung auszuschließen. Um diese am besten beurteilen zu können werden daher im Folgendem die Indizes über den maximalen Gradienten bestimmt.

5.3 Pixelgüte

Wie bereits erwähnt sind manche Pixel nicht sehr gut dafür geeignet, ihnen auf Basis ihres Gradientenverlaufs in der Bilderserie eine Tiefe zuzuordnen. Zum Beispiel gibt es für Pixel in schwach texturierten Flächen, wie das in Abbildung 11(c) gewählte Pixel, keine sinnvolle Wahlmöglichkeit für den Index. Die Tiefe dieser Art von Pixel soll nun über die Tiefe der benachbarten Pixel bestimmt werden. Dafür wird eine Heuristik für die Güte eines Pixels eingeführt. Wenn die Güte eines Pixels unter einem bestimmten Grenzwert liegt, wird der für ihn berechnete Index wieder verworfen. Die so entstehenden Löcher in der Indexmap werden anschließend mit einem Interpolationsalgorithmus gefüllt.

Es gibt diverse Heuristiken, die man zur Bewertung der Pixelgüte verwenden kann. Getestet wurden die folgenden:

- Maximaler Gradient
- Arithmetischer Durchschnitt des Gradienten
- Geometrischer Durchschnitt des Gradienten
- Varianz des Gradienten
- Differenz zwischen maximalem und durchschnittlichem Gradienten
- Quotient aus maximalem und durchschnittlichem Gradienten
- Quotient aus maximalem und durchschnittlichem Gradienten mit Generalisierung

Hier soll nicht auf alle Heuristiken eingegangen werden, sondern nur auf eine Auswahl der am besten funktionierenden. Zunächst wird die einfachste Heuristik, bei der der Wert des stärksten Gradienten verwendet wird, vorgestellt. Außerdem wird der Quotient aus maximalem und durchschnittlichem Gradienten sowie eine Verbesserung auf dieser Basis betrachtet. Zur Beurteilung der Heuristiken werden die rekonstruierten scharfen Bilder verwendet, da man auf ihnen mehr erkennen kann als auf der Indexmap.

5.3.1 Maximaler Gradient

Bei dieser Heuristik werden nur die Pixel berücksichtigt, bei denen der höchste Peak des Gradienten über einem festzulegendem Grenzwert liegen.

$$\max\{G(x, y)\} \geq \text{Schwellwert} \quad (3)$$

Da die Maxima bereits für die Bestimmung der Indizes gespeichert werden müssen, ist diese Heuristik sehr einfach zu implementieren und benötigt kaum weiteren Rechenaufwand. Nachdem alle Indizes bestimmt wurden, wird für jedes Pixel überprüft, ob es Ungleichung 3 erfüllt. Sollte dies nicht der Fall sein, wird der berechnete Index verworfen.

Für die Beurteilung wird die Bilderserie der vier Würfel auf kariertem Papier herangezogen. Die Striche auf dem Papier sollten einen ausreichend starken Gradienten erzeugen, um an diesen Stellen einen sinnvollen Index bestimmen zu können. Die weißen Flächen dazwischen hingegen haben quasi keine Kanten. Sie sollten durch die Heuristik herausgefiltert werden. In Abbildung 18 sind die Ergebnisse zu verschiedenen Schwellwerten abgebildet. 18(d) zeigt das Ergebnis bei einem Grenzwert von Null, was einem Weglassen der Heuristik entspricht, da die Stärke des Gradienten immer größer oder gleich Null ist. Dies ist also das Vergleichsbild bei dem alle Pixel verwendet wurden.

Bei einem Schwellwert von 10 (18(a)) werden kaum Pixel herausgefiltert. Lediglich im vorderen Bereich, der in jedem Bild unscharf ist, werden wenig Pixel verwendet. In diesem Bereich liegt aber eigentlich gar kein verwertbares Pixel. Die regelmäßigen rechtwinkligen Strukturen sind wahrscheinlich in der JPEG-Komprimierung begründet.

Ein höherer Schwellwert von 25 (18(b)) liefert eine deutlich bessere Pixelauswahl. Der komplett unscharfe Bereich wurde verworfen und die Striche des karierten Papiers werden noch berücksichtigt.

Bei einem Grenzwert von 30 würde man bereits einige der Karos verlieren. In 18(c) ist das Ergebnis bei einem Schwellwert von 40 abgebildet. Hier sieht man schon sehr deutlich, dass am Rand fast alle Pixel unterhalb des Grenzwertes liegen.

Der Grenzwert, der gerade so noch genug Pixel für eine ordentliche Interpolation zurücklässt, produziert die am besten aussehenden scharfen Bilder. Bei dieser Bilderserie ist das bei einem Grenzwert von 25 der Fall. Das interpolierte Ergebnis ist in Abbildung 18(e) zu sehen. Der Unterschied zu 18(d) ist nicht sehr groß. Im vorderen Bereich gibt es keine Rückstände der JPEG-Artefakte, sondern die

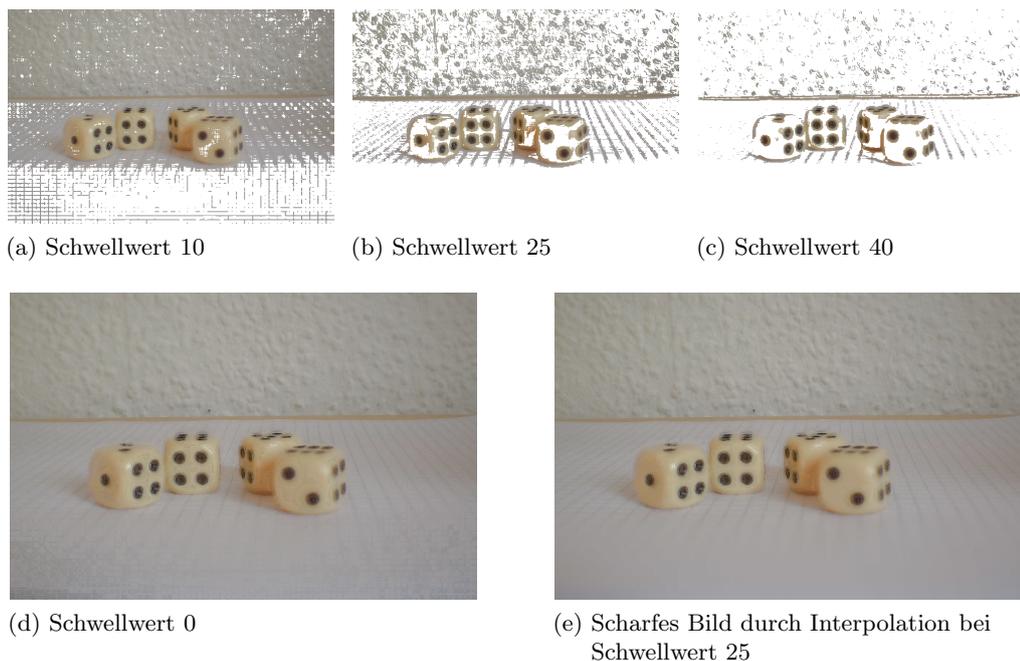


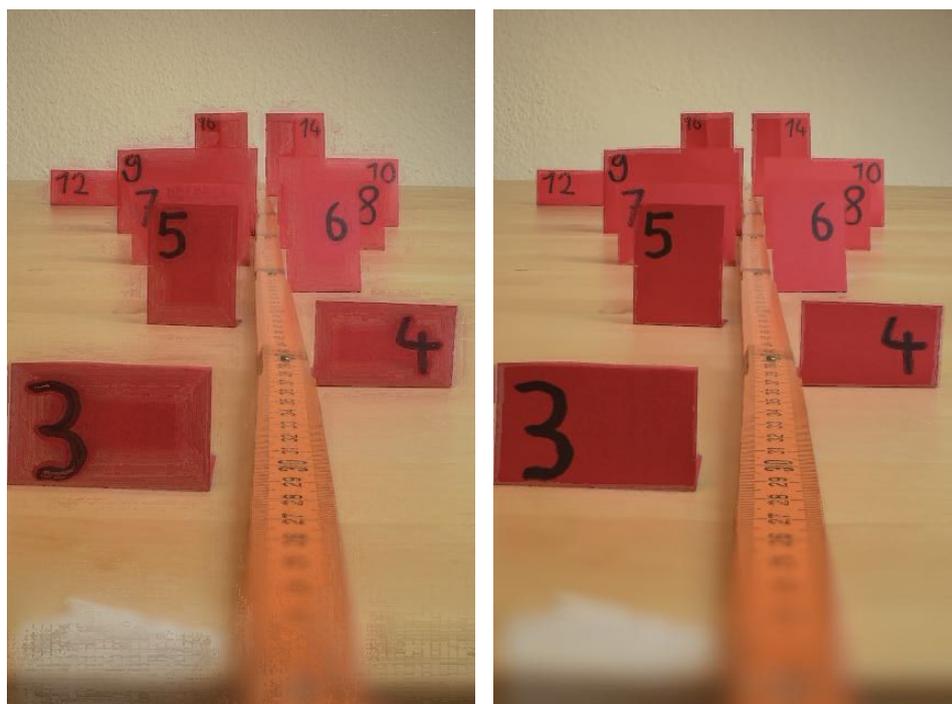
Abbildung 18: Ergebnisse nach Anwendung der Heuristik des maximalen Gradienten bei verschiedenen Schwellwerten

Fläche ist gleichmäßig unscharf. Die Würfel selbst sehen etwas 'sauberer' aus. In 18(d) kann man bei genauem Hinschauen ein paar dunkle Flecken erkennen.

Den Unterschied im scharfen Bild zwischen dem optimalen Grenzwert und keiner Heuristik kann man besonders deutlich bei größeren kantenfreien Flächen erkennen. Diese gibt es auf den roten Kärtchen in der Meterstab-Bilderserie. In Abbildung 19(a) ist das rekonstruierte scharfe Bild ohne Verwendung einer Heuristik für die Pixelgüte abgebildet. In 19(b) wurde als Schwellwert 100 angesetzt. Hier gibt es am Rand der Karten einen deutlich schmaleren Bereich in dem etwas Hellbraun vom Tisch auf der Karte ist.

An dieser Stelle gilt es anzumerken, dass der optimale Grenzwert nicht nur von der gewählten Heuristik abhängt, sondern auch von der betrachteten Szene. Während er bei den vier Würfeln auf kariertem Papier bei 25 liegt, ist für die Bilderserie mit zwei Würfeln ein Schwellwert von ungefähr 50 besser geeignet.

Eine automatische Festlegung des Grenzwertes wäre wünschenswert, ist allerdings nicht trivial umzusetzen. Den Grenzwert so hoch zu setzen, dass nur noch ein fester Prozentsatz der Pixel darüber liegt, wird vermutlich nicht funktionieren, da in verschiedenen Szenen durchaus unterschiedlich viele Kanten vorhanden sind. Die einfachste Lösung ist die Festlegung dieses Parameters durch den Nutzer. Wenn dies in der Implementierung entsprechend berücksichtigt wurde, kann der Schwellwert auch noch im Nachhinein verändert werden. Dazu muss die komplette



(a) Ohne Pixelfilterung

(b) Schwellwert 100 für die Heuristik des maximalen Gradienten

Abbildung 19: Scharfe Bilder des Meterstabes bei Indexwahl über maximalen Gradienten

Indexmap und die Gütebewertung jedes Pixels gespeichert werden. So kann relativ schnell die Indexmap angepasst und interpoliert werden. Lediglich die Berechnung des scharfen Bildes aus der interpolierten Indexmap würde etwas länger dauern.

5.3.2 Quotient aus maximalem und durchschnittlichem Gradienten

Für diese Heuristik ist es nötig den durchschnittlichen Gradienten für jedes Pixel zu bestimmen. Dafür wird während alle Bilder verarbeitet werden, bei jedem Bild für jedes Pixel `meanGradient` um `gradient/numberOfImages` erhöht:

Algorithmus 4 Berechnung des durchschnittlichen Gradienten

```

1: for i = 0 to numberOfImages do
2:   [...]
3:   for each pixel (x, y) do
4:     meanGradient(x, y) += gradient(x, y) / numberOfImages;
5:   end for
6: end for

```

Der maximale Gradient jedes Pixels wird bereits aufgrund der Indexbestimmung gespeichert. Um nun für jedes Pixel einen Wert zu erhalten, der seine Güte beschreibt, wird das Maximum durch den Durchschnitt geteilt, wobei in diesem Fall 0/0 als 1 ausgewertet wird. Es werden also nur diejenigen Pixel berücksichtigt, die Ungleichung 4 erfüllen.

$$\frac{\max\{G(x, y)\}}{\text{meanGradient}(x, y)} \geq \text{Schwellwert} \quad (4)$$

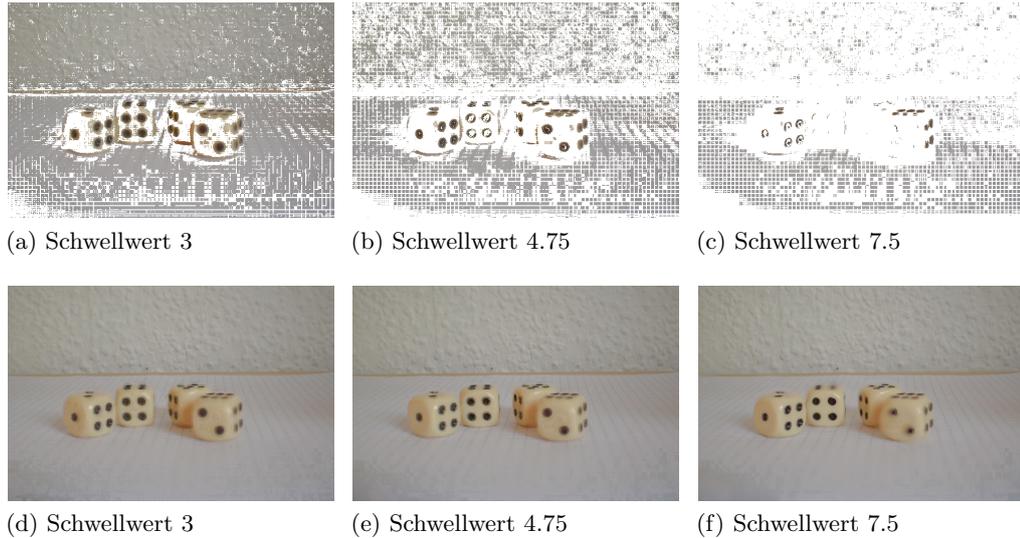


Abbildung 20: Ergebnisse nach Anwendung der Heuristik des Quotienten aus maximalem und durchschnittlichem Gradienten bei verschiedenen Schwellwerten

Die Ergebnisse für verschiedene Schwellwerte sind in Abbildung 20 dargestellt. In 20(a) bis 20(c) sind die scharfen Bilder mit den durch den Grenzwert herausgefilterten Pixeln zu sehen. 20(d) bis 20(f) zeigen die entsprechenden interpolierten scharfen Bilder. Auch hier gibt es wieder sehr regelmäßige Strukturen, die vermutlich auf die JPEG-Komprimierung zurückzuführen sind.

Bei einem relativ kleinem Grenzwert von 3 (20(a)) werden nur die Flächen der Würfel herausgefiltert. Eine Erhöhung des Schwellwertes auf 4.75 (20(b)) bewirkt, dass die Würfelflächen großzügiger verworfen werden. Hier bleiben nur Pixel übrig, die zu einer ausgeprägten Kante gehören. Dies liefert das beste interpolierte scharfe Bild (20(e)). Die Filterung der Pixel des Blattes hat sich jedoch nur ein bisschen verstärkt.

Auch eine weitere Erhöhung des Grenzwertes (20(c)) führt nur zu einer minimalen Abnahme der ausgewerteten Pixel auf dem Papier. Allerdings gehen hier auch schon einige wichtige Kanten der Würfel verloren. Dies äußert sich im interpolier-

ten scharfen Bild (20(f)). Hier sind die Augen der Zwei etwas verschwommen. Die Oberseite des zweiten Würfels von links ist ebenfalls unschärfer geworden.

In Flächen mit geringem durchschnittlichen Gradienten gibt diese Heuristik relativ hohe Werte aus. Auch wenn der maximale Gradient eines Pixels klein ist, kann durch einen noch kleineren durchschnittlichen Gradienten an dieser Stelle die Gütebewertung hoch ausfallen. Besonders extrem ist diese bei Punkten, die so nah an der Kamera liegen, dass sie in keinem Bild scharf sind.

In Abbildung 21(a) ist der Verlauf des Gradienten eines solchen Pixels dargestellt. Ihm wurde eine Güte von 9.27 zugewiesen. Pixel B, welcher eindeutig besser ist, hat hingegen nur einen Gütewert von 3.6 erhalten. Um dieses Problem zu beheben wird eine Generalisierung vorgenommen.

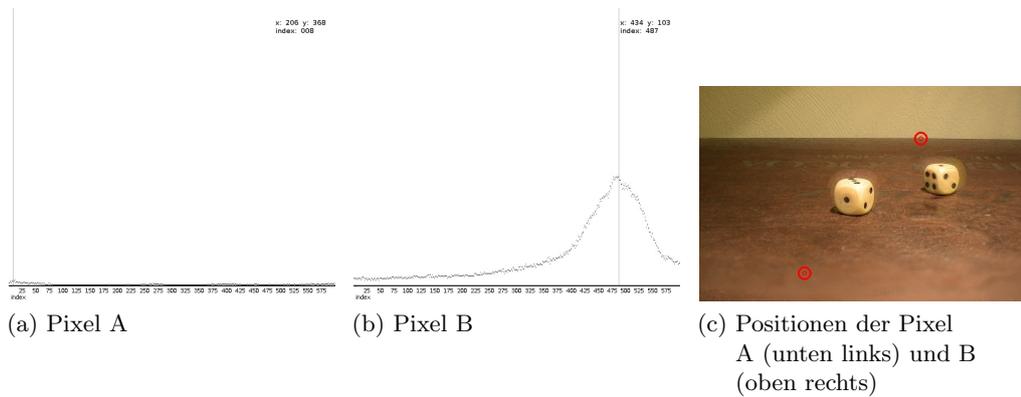


Abbildung 21: Gradientenverlauf für zwei Pixel

5.3.3 Quotient aus maximalem und durchschnittlichem Gradienten mit Generalisierung

Durch die Generalisierung soll ausgeglichen werden, dass der Gütewert durch eine sehr kleine Zahl im Nenner zu stark anwächst. Dementsprechend wird zum Nenner eine Zahl α hinzu addiert. Die neue zu erfüllende Ungleichung lautet:

$$\frac{\max\{G(x, y)\}}{\text{meanGradient}(x, y) + \alpha} \geq \text{Schwellwert} \quad (5)$$

Ein höheres Alpha verringert den Wert der Heuristik für alle Pixel. Allerdings wirkt es sich bei Pixeln, die nicht 'gut' sind, deutlich stärker aus. Tabelle 1 listet die Güte der oben gewählten Pixel A und B bei verschiedenen Werten für α .

Ab einem α -Wert von 3 wird Pixel B nicht mehr schlechter bewertet als Pixel A. Da der berechnete Index für Pixel B deutlich verlässlicher ist als der für Pixel A, wird α noch weiter erhöht, damit sich diese Beziehung in der Bewertung wieder findet. Es wird für α letztendlich ein Wert von 5 festgelegt. Ein größerer Wert ist nicht nötig. Er würde lediglich dafür sorgen, dass die 'guten' Pixel schlechter

α	0	1	2	3	4	5	6	7
Pixel A	9.27	5.96	4.39	3.47	2.87	2.45	2.14	1.89
Pixel B	3.60	3.56	3.52	3.49	3.45	3.41	3.38	3.35

Tabelle 1: Auswirkung verschiedenerer α -Werte auf die Güte der Pixel A und B

bewertet werden. Ein zu großes α würde die Bedeutung des durchschnittlichen Gradienten in der Heuristik abschwächen, wodurch die Heuristik wieder näher an der des maximalen Gradienten läge.

Da Pixel A eine Gütebewertung von 2.45 hat und dieses Pixel nicht verarbeitet werden soll, muss der Grenzwert über 2.45 liegen. Für drei verschiedene Schwellwerte sind die Ergebnisse in Abbildung 22 dargestellt.



(a) Schwellwert 2.5

(b) Schwellwert 3.5

(c) Schwellwert 4.5

Abbildung 22: Ergebnisse nach Anwendung der Heuristik des Quotienten aus maximalem und durchschnittlichem Gradienten mit Generalisierung bei verschiedenen Schwellwerten

Bei einem Grenzwert von 2.5 (22(a)) ist das Ergebnis bereits zufriedenstellend. Der unscharfe Bereich vorne im Bild wird komplett herausgefiltert, auf dem kariertem Papier werden nur die Linien ausgewertet und bei den Würfeln liegen fast nur die Kanten und Augen über dem Schwellwert.

Ein höherer Grenzwert von 3.5 (22(b)) sorgt für eine leichte Reduzierung der einbezogenen Würfelflächen. Die Pixelstreifen um die Linien des Papiers werden ebenfalls dünner. Es sind gerade noch genug Pixel vorhanden, um eine sinnvolle Interpolation durchzuführen.

In 22(c) liegt der Grenzwert so hoch, dass nicht mehr alle Linien mit einbezogen werden. Dementsprechend sind Teile dieser im rekonstruierten Gesamtbild unscharf. Der Unterschied zwischen den interpolierten scharfen Bildern ist bei diesen drei Grenzwerten jedoch so gering, dass man kaum einen Unterschied erkennen kann. Der optimale Schwellwert scheint hier allerdings um 3.5 zu liegen.

Im Vergleich zur Heuristik des maximalen Gradienten (siehe Abbildung 23) fällt auf, dass hier insgesamt weniger Pixel gewählt werden. Zwischen den aus der Interpolation rekonstruierten scharfen Bildern ist der Unterschied so gering, dass er hier kaum sichtbar ist. In 23(d) sind auf den Würfeln weniger dunkle Flecken als in 23(c).

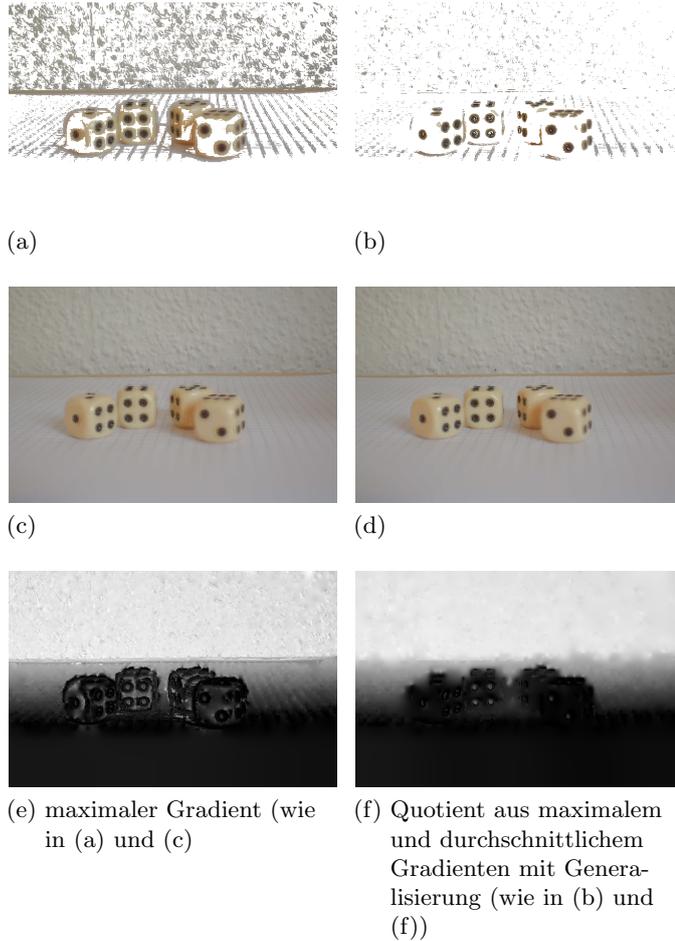


Abbildung 23: Vergleich der Heuristiken zur Pixelgüte am lückenhaften scharfen Bild, am aus der Interpolation rekonstruierten scharfen Bild und an der interpolierten Indexmap

Einen größeren Unterschied sieht man zwischen den entsprechenden interpolierten Indexmaps 23(e) und 23(f). Bei der Heuristik des maximalen Gradienten kann man auch in der Indexmap die Linien des karierten Papiers erahnen. Da die Linien genauso weit entfernt sind wie der weiße Zwischenraum, sollte dies eigentlich nicht der Fall sein. Auch das Rauschen der Wand ist in 23(f) geringer. Die Augen auf den Würfeln, welche zumindest ungefähr soweit entfernt sind wie Flächen in denen sie liegen, lassen sich in 23(e) besser erkennen. In 23(f) haben sie einen Index deutlich näher an dem ihrer Umgebung erhalten. Dies ist korrekt. Lediglich bei den hinteren Kanten der Würfel könnte man sagen, dass diese bei maximalem Gradienten als Heuristik plausibler sind. Vor allem beim linken Würfel ist diese in 23(f) etwas verwischt.

5.3.4 Interpolationsalgorithmus

Um die L cher in der Indexmap zu f llen, in denen keine zuverl ssige Tiefeninformation vorliegt, wird ein pyramidenbasierter Interpolationsalgorithmus eingesetzt. Dieser arbeitet in zwei Phasen: Einer Pull-Phase und einer Push-Phase. Die unterste Ebene der Pyramide stellt das Ausgangsbild, das interpoliert werden soll, dar. In der Pull-Phase wird die Pyramide von unten nach oben gef llt wie in Abbildung 24 dargestellt. Dabei wird zu jeder weiteren Ebene die Kantenl nge halbiert und somit vier Pixel zu einem zusammengefasst. Jedes der vier Pixel wird gleichstark gewichtet. Wenn nicht alle vier Pixel valide sind, so werden die Gewichte entsprechend angepasst. [VM]

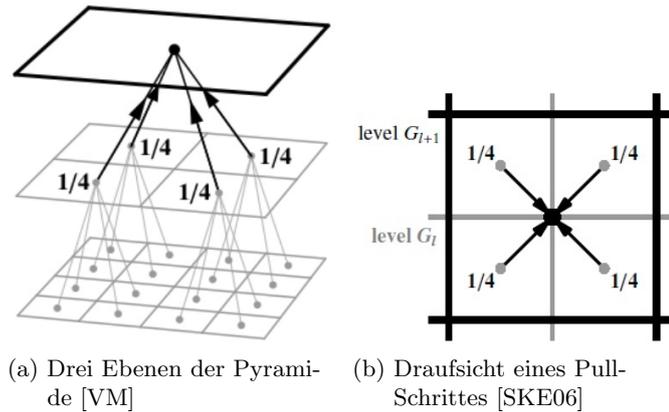


Abbildung 24: Pull-Phase des Interpolationsalgorithmus

In dem Fall, dass keines der vier Pixel gef llt ist, wird das neue Pixel in der dar berliegenden Ebene ebenfalls nicht gef llt. In der Pyramidenspitze ist das Bild dann nur noch 1×1 Pixel gro  und dieses ist garantiert gef llt, da ansonsten keines der Pixel im Ausgangsbild gef llt gewesen w re. [VM]

Als n chstes kommt die Push-Phase in der die Information von oben nach unten "gepusht" wird. Hat ein Pixel bereits einen Wert, so beh lt es diesen. Ansonsten nimmt es den Wert des dar berliegenden Pixels an. Dieses ist garantiert gef llt, da in der obersten Ebene mit einem validen Pixel gestartet wurde und so die darunterliegenden ebenfalls komplett gef llt werden konnten. Abbildung 25 illustriert diese Phase. [SKE06]

In Abbildung 26 wurden aus dem Originalbild (26(a)) zun chst einige Pixel entfernt (26(b)). Dieses l ckenhafte Bild wurde anschlie end mithilfe des pyramidenbasierten Algorithmus interpoliert (26(c)). Es sind sehr deutlich quadratische K stchen zu sehen, die von den Pixeln in h heren Ebenen der Pyramide herr hren.

Aus diesem Grund wird eine andere Variante des Algorithmus verwendet, die sich in der Push-Phase unterscheidet. Hier wird f r ein leeres Pixel nicht einfach der Wert des dar berliegenden Pixels verwendet, sondern zwischen dem direkt dar berliegenden und den anderen drei an dieser Ecke angrenzenden Pixeln

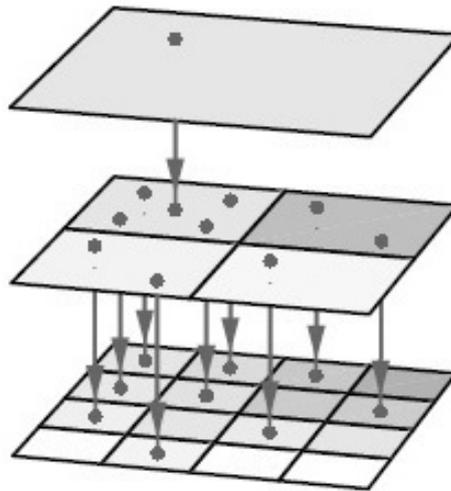


Abbildung 25: Push-Phase des Interpolationsalgorithmus [SKE06]

gemittelt. Dabei werden die Pixel der darüberliegenden Ebene entsprechend ihres Abstandes zum zufüllenden Pixel gewichtet. Das Pixel unmittelbar über dem zufüllenden ist am nächsten und erhält mit $9/16$ die stärkste Gewichtung. Die direkt benachbarten Pixel gehen je zu $3/16$ ein und das am weitesten entfernte Pixel erhält mit $1/16$ die kleinste Gewichtung. [SKE06]

Das Ergebnis der zweiten Variante ist in Abbildung 26(d) abgebildet. Dieses ist sichtlich näher am Originalbild als das Ergebnis von Variante Eins 26(c). Daher kommt die zweite Variante des Algorithmus zum Einsatz.



(a) Originalbild [Ray03]



(b) Lückenhaftes Bild

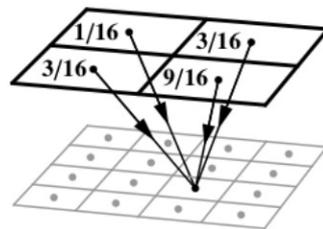


(c) Lücken mit Variante 1 gefüllt

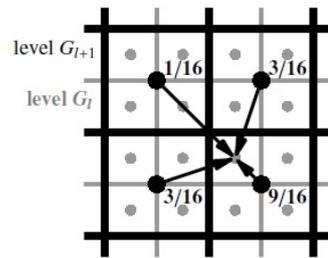


(d) Lücken mit Variante 2 gefüllt

Abbildung 26: Vergleich der beiden Varianten des Interpolationsalgorithmus



(a) Zwei Ebenen der Pyramide [VM]



(b) Draufsicht eines Push-Schrittes [SKE06]

Abbildung 27: Push-Phase der zweiten Variante des Interpolationsalgorithmus

5.4 Generierung des scharfen Bildes

In der simpelsten Form des Algorithmus, bei dem der Index durch den maximalen Gradienten bestimmt wird und keine Heuristik zur Bestimmung der Pixelgüte angewandt wird, lässt sich das komplett scharfe Bild am einfachsten generieren. Hier muss lediglich bei der Iteration durch die Bilder bei einem neuem Maximum neben der Indexmap auch ein scharfes Bild aktualisiert werden. Dieser Prozess ist in Algorithmus 5 dargestellt. Die entscheidende Aktualisierung von *sharpImage* findet in Zeile 9 statt.

Algorithmus 5 Simple Generierung des scharfen Bildes

```
1: maxima.fill(-1)
2: sharpImage.fill(0)
3: for i = 0 to numberOfImages do
4:   image = images[i]
5:   gradient = calculateGradient(image)
6:   for each pixel (x, y) do
7:     if gradient(x, y) > maxima(x, y) then
8:       indexmap(x, y) = i;
9:       sharpImage(x, y) = image(x, y)
10:    end if
11:  end for
12: end for
```

Wenn jedoch die Indizes erst feststehen, nachdem durch alle Bilder iteriert wurde, wie es zum Beispiel bei der Indexbestimmung über gewichtete Indizes der Fall ist, ist die Erstellung des scharfen Bildes nicht so direkt möglich. Auch wenn für einige Pixel der korrekte Index erst durch die Interpolation der Indexmap festgelegt wird, ist es nötig das scharfe Bild im Nachhinein aus der Indexmap zu rekonstruieren.

Dafür wird für jedes Pixel des scharfen Bildes in der Indexmap nachgeschaut aus welchem Ausgangsbild die Farbwerte entnommen werden müssen. Entweder wird dieser Vorgang durch das erneute Laden der entsprechenden Bilder stark verlangsamt oder alle Bilder der Originalreihe verbleiben so lange im Hauptspeicher. Dies führt allerdings zu einer hohen Arbeitsspeicherauslastung. Für die hier verwendeten Bilder mit einer Auflösung von 640 x 424 Pixeln ist es noch gut möglich das erneute Einlesen zu vermeiden. In Algorithmus 6 ist dieser Vorgang dargestellt. Das Konvertieren zu einem Integer in Zeile 2 ist notwendig, da durch die Interpolation durchaus nicht-ganzzahlige Werte entstehen können.

Algorithmus 6 Generierung des scharfen Bildes

```
1: for each pixel (x, y) do
2:   sharpImage(x, y) = images[(int)indexmap(x, y)](x, y)
3: end for
```



Abbildung 28: Direkte Anwendung des Interpolationsalgorithmus auf das lückenhafte scharfe Bild

Den Interpolationsalgorithmus direkt auf das lückenhafte scharfe Bild anzuwenden ist nicht sehr zielführend, wie Abbildung 28 sehr deutlich zeigt.

5.5 Generierung der Tiefenmap

Bisher liegt an gewonnenen Tiefeninformationen lediglich die Indexmap vor, die jedem Pixelkoordinatenpaar einen Index aus der Bilderreihe zuordnet. Im direkten Vergleich von zwei Punkten kann so ermittelt werden, welcher weiter von der Kamera entfernt liegt. Der Zusammenhang zwischen Bildindex und Abstand ist allerdings nicht linear. So bedeutet ein doppelt so hoher Index nicht einen doppelt so großen Abstand zur Kamera. Es soll nun eine Tiefenmap erstellt werden, die jedem Pixel einen Abstand zuordnet.

Um dies umzusetzen muss eine Funktion gefunden werden, die einen Index in einen Abstand umrechnet. Diese ist logischerweise von der verwendeten Kamera, dem Objektiv und der Schrittweite abhängig. Für die hier verwendete Hardware und Schrittweite wird die Kalibrierung mittels der Meterstab-Bilderserie vorgenommen. Das scharfe Bild dieser Serie ist in Abbildung 31(c) zu sehen. Der Meterstab ist dabei so platziert, dass der Nullpunkt auf Höhe des Bildsensors der Kamera liegt. Die Position des Sensors ist durch entsprechende Markierungen am Gehäuse der Kamera gekennzeichnet.

Da sich die Kamera etwa 10 cm oberhalb der Nullmarkierung des Meterstabs befindet und leicht nach unten geneigt ist, stellen die Angaben auf diesem weder den Abstand zur Kamera noch den Abstand zur Kameraebene dar, sondern müssen noch umgerechnet werden. Dieser Sachverhalt wird in der Skizze in Abbildung 29 dargestellt. Der betrachtete Punkt auf dem Meterstab ist dabei x cm von dessen Nullmarke entfernt. Sein Abstand d zur Kamera lässt sich mithilfe des Satz des Pythagoras einfach berechnen:

$$d = \sqrt{10^2 + x^2} \quad (6)$$

Es haben jedoch nicht alle in einem Bild scharf gestellten Punkte den selben

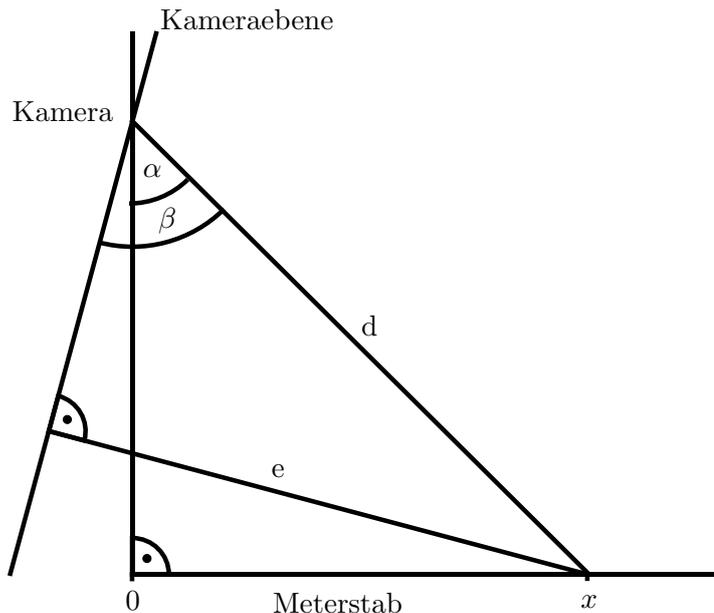


Abbildung 29: Skizze der Aufnahme der Meterstab-Serie

Abstand zur Kamera, denn die Fokusebene ist eine Ebene und kein Ausschnitt einer Kugeloberfläche. Daher muss die Distanz e zwischen der Ebene, in der die Kamera liegt, und dem betrachteten Punkt berechnet werden. Dafür wird zunächst der Winkel α über das rechtwinklige Dreieck *Kamera*, *Null*, x berechnet:

$$\alpha = \arctan\left(\frac{x}{10}\right) \quad (7)$$

Die Kamera ist für das Kalibrierungssetup auf die Zahl 40 des Meterstabes ausgerichtet und befindet sich 10 cm über diesem. Darüber kann der Neigungswinkel der Kamera bestimmt werden:

$$\beta - \alpha = \arctan\left(\frac{10}{40}\right) \approx 14^\circ \quad (8)$$

Da sich so nun β berechnen lässt kann letztlich auch der Abstand von x zur Kameraebene e bestimmt werden:

$$e = d \cdot \sin(\beta) \quad (9)$$

In einer Gleichung zusammengefasst lässt sich e berechnen zu

$$e = \sqrt{10^2 + x^2} \cdot \sin\left(\arctan\left(\frac{x}{10}\right) + \arctan\left(\frac{10}{40}\right)\right) \quad (10)$$

Für Werte von x , die größer als 40 sind ergibt sich die leicht andere Formel

$$e = \sqrt{10^2 + x^2} \cdot \sin\left(\pi - \arctan\left(\frac{x}{10}\right) - \arctan\left(\frac{10}{40}\right)\right) \quad (11)$$

Aufgrund der Symmetrie des Sinus sind die Formeln 10 und 11 jedoch äquivalent, sodass keine Fallunterscheidung vorgenommen werden muss.

Für 17 Punkte auf dem Meterstab wurden die entsprechenden Indizes der Bilderreihe bestimmt, bei denen diese Punkte im Fokus liegen. Eine Auswahl ist in Tabelle 2 eingetragen. Dazu wurde jeweils mit Gleichung 6 und 10 der Abstand zur Kamera und zur Kameraebene ausgerechnet.

Index	66	127	244	512	674	755	819	862
Wert auf Meterstab	28	30	34	50	70	90	120	160
Abstand zur Kamera [cm]	29.7	31.6	35.4	51	70.7	90.6	120.4	160.3
Abstand zur Ebene [cm]	29.6	31.5	35.4	50.9	70.3	98.7	118.8	157.6

Tabelle 2: Indizes und Abstände der zur Kalibrierung gewählten Punkte (Auswahl)

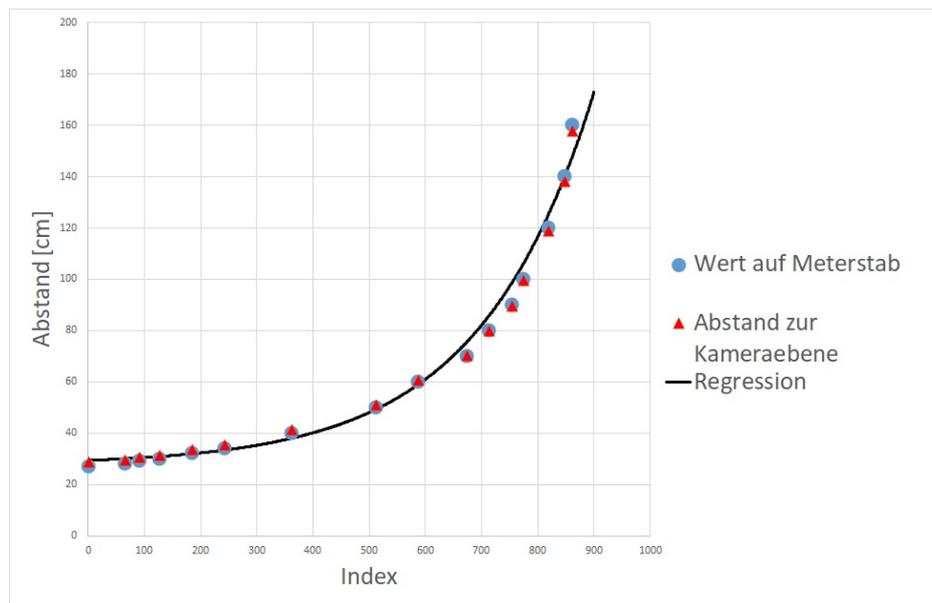


Abbildung 30: Regression des Abstands in Abhängigkeit des Index

Das Diagramm in Abbildung 30 zeigt zwei der Datenreihen. Der Unterschied zwischen den vom Meterstab abgelesenen Werten und den echten Abständen zur Kameraebene ist nicht sehr groß. Zusätzlich ist hier noch eine Regression aufgetragen, die mithilfe von Excel erstellt wurde. So ergibt sich die Umrechnung vom Bildindex zum Abstand zur Kameraebene wie folgt:

$$d = 1.7658 \cdot e^{0.0049 \cdot index} + 27.6 \quad (12)$$

Diese Formel auf jedes Pixel der Indexmap angewandt liefert die Tiefenmap. Die hier eingetragenen Distanzen sind in Zentimetern zu verstehen. Um die Tiefenmap nicht interaktiv darstellen zu können muss sie normalisiert werden. Dafür werden alle Abstände linear in den Bereich von 0 bis 255 transformiert. Ein Unterschied zur ebenfalls normalisierten Indexmap ist so trotzdem zu erkennen. Abbildung 31 zeigt die beiden für die Meterstab-Szene zum Vergleich. In 31(c) ist auch noch das scharfe Bild dargestellt.

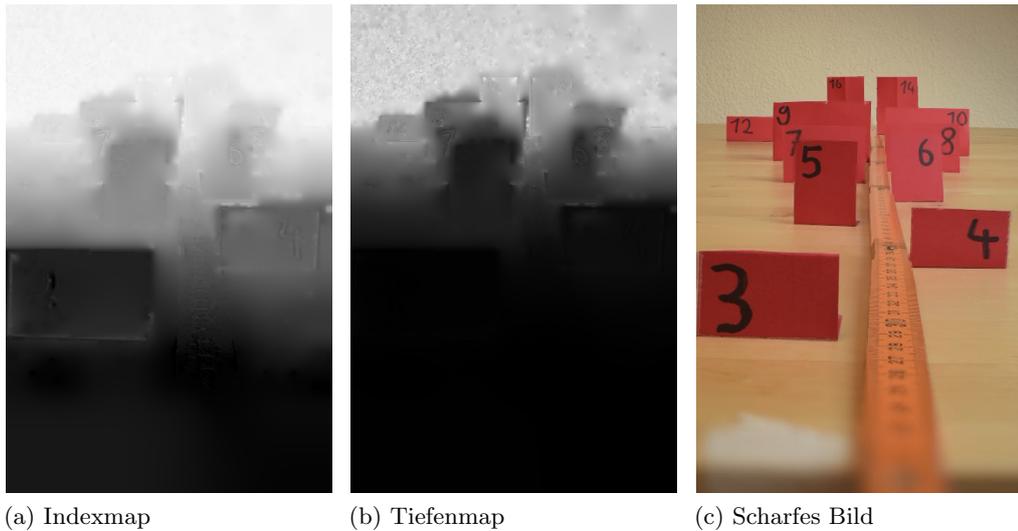


Abbildung 31: Die verschiedenen Ergebnisse im Vergleich

In der Tiefenmap sollten die Helligkeitsunterschiede proportional zum Abstand sein. In der Indexmap hingegen bewirkt eine Änderung des Abstandes von 10 cm im vorderen Bereich eine deutlich stärkere Helligkeitsveränderung als die selbe Änderung weiter hinten im Bild. Dies liegt daran, dass die Fokusebenen am Anfang der Bilderserie deutlich dichter beieinander liegen als an ihrem Ende.

5.6 Generierung der Punktwolke

Die Koordinaten eines Pixels x_{Kamera} und y_{Kamera} zusammen mit dem entsprechenden Tiefenwert $z_{\text{Welt}} := \text{Tiefenmap}(x_{\text{Kamera}}, y_{\text{Kamera}})$ können nicht als Koordinaten eines kartesischen Koordinatensystems interpretiert werden. Da die Punkte der Szene bei der Aufnahme der Bilder nicht orthografisch, sondern perspektivisch auf das Bild projiziert wurden, lässt sich von den Bildkoordinaten nur auf den Winkel schließen in dem sich der entsprechende Punkt zur Kamera befunden hat.

Dieser Sachverhalt ist in Abbildung 32 dargestellt. Hier ist nur der Mittelpunktstrahl, also der Lichtstrahl, der vom aufgenommenen Objekt durch die Mitte der Linse verläuft, eingezeichnet. Die Strahlen, welche die Linse weiter außen treffen, werden, vorausgesetzt das Objekt liegt im Fokus, so gebrochen, dass sie den Bildsensor im selben Punkt schneiden. Sie wurden hier zur Übersichtlichkeit ausgelassen. Vom Mittelpunktstrahl kann angenommen werden, dass er geradlinig durch die Linse verläuft. [BSS17]

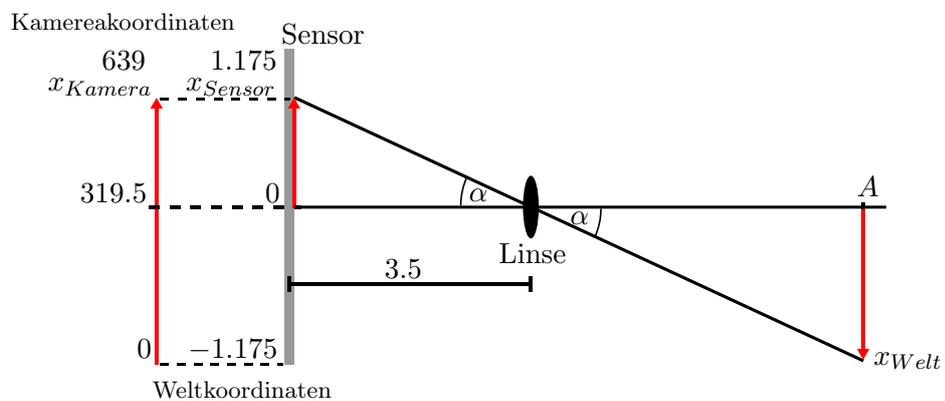


Abbildung 32: Umrechnung von Bildkoordinaten zu Weltkoordinaten

Die Zeichnung in Abbildung 32 stellt die Ansicht auf die Kamera dar. Links befindet sich der in diesem Fall 2.35 cm breite Bildsensor [Nik15]. Die Pixel der 640 Pixel breiten Bilder sind gleichmäßig auf dem Sensor verteilt, sodass aus den Kamerakoordinaten die Position des Pixels auf dem Bildsensor errechnet werden kann. Dabei wird der Nullpunkt des verwendeten Weltkoordinatensystems genau in die Mitte des Sensors gelegt.

x_{Kamera} ist die X-Koordinate eines Punktes im aufgenommenen Bild in Pixeln.

x_{Sensor} beschreibt die X-Position des entsprechenden Pixels auf dem Bildsensor in Weltkoordinaten.

x_{Welt} beschreibt die X-Position des entsprechenden Punktes des Motivs in Weltkoordinaten.

Die X-Koordinate eines Pixels auf dem Sensor berechnet sich somit zu

$$x_{\text{Sensor}} = (x_{\text{Kamera}} - 319.5) \cdot \frac{2.35}{640} \quad (13)$$

. Die Einheit des Ergebnisses ist Zentimeter. Da die Dreiecke x_{Kamera} , 0, *Linse* und x_{Welt} , *A*, *Linse* zueinander ähnlich sind, ist das Verhältnis von x_{Sensor} zum Abstand zwischen *Linse* und *Sensor* (entspricht der Brennweite von 3.5 cm) gleich dem Verhältnis von x_{Kamera} zum Abstand von *A* zur *Linse*. Für die Bestimmung des Abstands zwischen der *Linse* und *A* wird die Z-Koordinate benötigt. Diese ist direkt aus der Tiefenmap zu entnehmen:

$$z_{\text{Welt}} = -\text{Tiefenmap}(x_{\text{Kamera}}, y_{\text{Kamera}}) \quad (14)$$

Im Weltkoordinatensystem hat der Punkt also die X-Koordinate

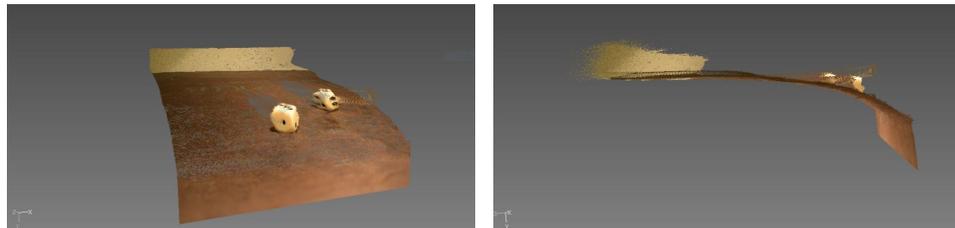
$$\begin{aligned} x_{\text{Welt}} &= (-z_{\text{Welt}} - 3.5) \cdot \frac{x_{\text{Sensor}}}{3.5} \\ &= (-z_{\text{Welt}} - 3.5) \cdot \frac{(x_{\text{Kamera}} - 319.5) \cdot \frac{2.35}{640}}{3.5} \end{aligned} \quad (15)$$

Analoge Überlegungen für die Y-Richtung ergeben mit einer Bildhöhe von 424 Pixeln und einer Sensorhöhe von 1.56 cm [Nik15] für die Y-Koordinate:

$$\begin{aligned} y_{\text{Welt}} &= (-z_{\text{Welt}} - 3.5) \cdot \frac{y_{\text{Sensor}}}{3.5} \\ &= (-z_{\text{Welt}} - 3.5) \cdot \frac{(211.5 - y_{\text{Kamera}}) \cdot \frac{1.56}{424}}{3.5} \end{aligned} \quad (16)$$

Da in der CImg-Library die Y-Koordinaten nach unten hin größer werden, die Y-Achse im Welt-Koordinatensystem jedoch nach oben zeigen soll, kommt hier noch ein Vorzeichen hinzu.

So lassen sich also nun für alle Punkte in der Tiefenmap die X- und Y-Weltkoordinaten berechnen. Die Z-Koordinate ist bereits in der Tiefenmap enthalten. Aus dem komplett scharfen Bild erhält man zusätzlich die Farbinformationen der Punkte.



(a) Von schräg vorn

(b) Von der Seite

Abbildung 33: Punktwolke der zwei Würfel

In Abbildung 33 ist die Punktwolke der Szene mit zwei Würfeln abgebildet. In 33(a) ist gut zu erkennen, wie hinter den Würfeln keine Punkte liegen, da diese bei der Aufnahme durch die Würfel verdeckt waren. Die senkrechte Fläche ganz vorne ist dadurch entstanden, dass dieser Bereich so nah ist, dass er in keinem der Bilder scharf ist. Ihm wurde der geringst mögliche Abstand zugewiesen.

Von der Seite betrachtet (33(b)) fällt auf, dass der eigentlich ebene Untergrund nicht komplett gerade dargestellt ist. Rechts beziehungsweise vorn ist eine Krümmung zu sehen. Dies könnte mehrere Ursachen haben. Es kann bei der Fotografie zu sogenannter *Verzeichnung* kommen. Dabei ist die Abbildung der Szene nicht im gesamten Bild gleich. So kommt es zu Verzerrungen. Diese wären allerdings schon in den Ausgangsbildern enthalten. Dort sind jedoch keine Verzerrungen aufgefallen. Daher ist die zweite mögliche Ursache wahrscheinlicher.

Diese besagt, dass das Modell, welches zur Umrechnung der Kamerakoordinaten in Weltkoordinaten verwendet wurde, zu sehr vereinfacht. Zum einen besteht das verwendete Objektiv aus mehreren Linsen und nicht nur aus einer. Zum anderen wurde der Abstand der Linse zum Sensor als konstant angenommen. Dies ist jedoch nicht richtig. Bei der Fokusdistanz 'Unendlich' ist der Abstand gleich der Brennweite. Denn so werden die parallel einfallenden Lichtstrahlen auf dem Sensor gebündelt. Um die Fokusdistanz zu verringern, wird dieser Abstand jedoch erhöht. Dies wurde in den Formeln zur Koordinatentransformation nicht berücksichtigt.

6 Ergebnisse

Es wurde festgestellt, dass die Indexbestimmung über gewichtete Indizes mit Schwellwert den sinnvollsten Index liefert und die Pixelgüte am besten über den Quotienten aus maximalem und durchschnittlichem Gradienten mit Generalisierung bewertet wird. Daher werden diese nun kombiniert, um für jede der aufgenommenen Bilderserien die bestmöglichen Ergebnisse zu erzielen. Für die Indexbestimmung wird der in Kapitel 5.2.3 bestimmte Grenzwert von 50 verwendet. In den folgenden Abbildungen sind jeweils das rekonstruierte scharfe Bild, die Indexmap und die daraus errechnete Tiefenmap sowie die entsprechende Punktwolke dargestellt. Der Schwellwert für die Heuristik zur Bewertung der Pixelgüte wurde jeweils so gewählt, dass das scharfe Bild gut aussieht und ist in Klammern dahinter angegeben.



Abbildung 34: Gartenstuhl (5.5)

Generell fällt auf, dass die Qualität der Tiefeninformation deutlich schlechter ist als die der scharfen Bilder. So ist zum Beispiel die Stützstrebe zwischen den Beinen des Gartenstuhls in Abbildung 34(c) und 34(d) nicht zu erkennen. Auch bei der Hecke im Hintergrund gibt es einige Probleme. Hier reichen die bestimmten Entfernungen von 80 bis 180 cm. Da einige Punkte der Hecke sogar näher an der Kamera liegen als der Stuhl selbst, ist von diesem in der Punktwolke nicht sehr viel zu erkennen. Dieser starke Fehler scheint in der Struktur der Hecke begründet zu sein. Bei der Meterstab-Szene in Abbildung 35 ist die Wand ähnlich weit entfernt. Die bestimmten Tiefen liegen allerdings in dem engeren Intervall von

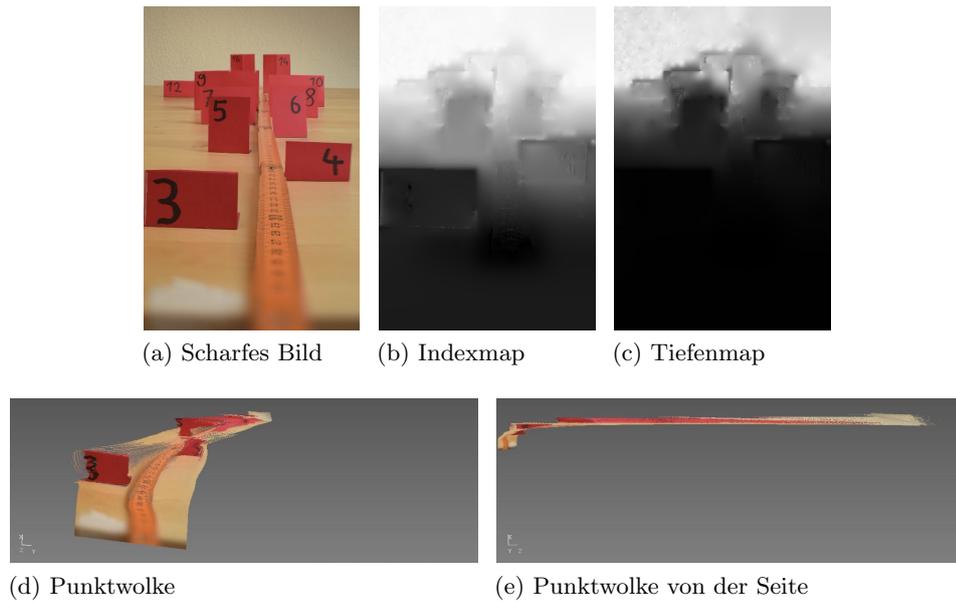
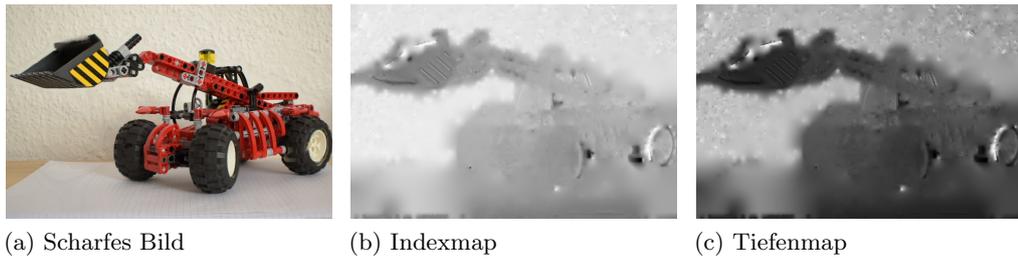


Abbildung 35: Meterstab (4.2)

150 bis 178 cm. Bei geringeren Abständen zum Hintergrund wird auch die Entfernungsbestimmung für diesen entsprechend genauer. Bei den übrigen Szenen sind sie, abgesehen von der Glas-Szene, beim Legotechnik Radlader (Abbildung 36) am schlechtesten. Hier variieren die Werte der Wand zwischen 60 und 79 cm.

In der Punktwolke der Meterstabserie (Abbildung 35(d)), sind je nach Blickwinkel die Schilder bis einschließlich Schild 8 relativ gut zu lesen. Die Zahlen, der weiter entfernten Schilder sind nur bei senkrechtem Blickwinkel zu erkennen, welcher quasi dem scharfen Bild gleicht. In der Seitenansicht (35(e)) ist auch hier die Krümmung im nahen Bereich zu erkennen. Sie sollte jedoch nicht mit dem senkrechten Bereich verwechselt werden, der für die Rekonstruktion zu nah an der Kamera lag. Die ersten drei bis vier Zentimeter ab dem scharfen Bereich, also eine Entfernung von 30 bis 31 cm zur Kamera (gemessen vom Sensor aus), sind quasi nicht zu gebrauchen. Danach nimmt die Krümmung sehr schnell ab. Dies ist bei allen Szenen zu beobachten, bei denen sich im entsprechenden Bereich eine gerade Fläche befunden hat.

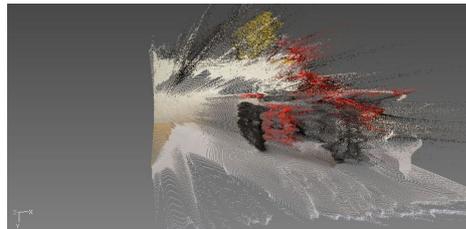
Die Tiefeninformationen des Radladers selbst sind nicht von sehr guter Qualität. In der Punktwolke kann man ihn gerade noch so erkennen. Das scharfe Bild hingegen hat lediglich am hinteren Rad und an der Schaufel merkliche Fehler.



(a) Scharfes Bild

(b) Indexmap

(c) Tiefenmap



(d) Punktwolke

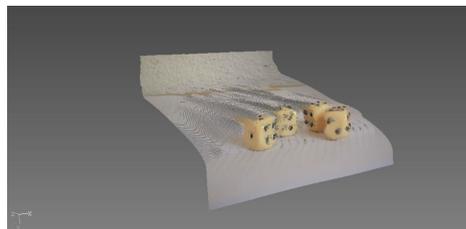
Abbildung 36: Legotechnik Radlader (3.75)



(a) Scharfes Bild

(b) Indexmap

(c) Tiefenmap



(d) Punktwolke

Abbildung 37: Vier Würfel (4.1)

Bei der Szene mit vier Würfeln (Abbildung 37) ist die Rekonstruktion sehr gut. Im scharfen Bild (37(a)) sind lediglich ein paar Unreinheiten in den Augen des linken Würfels und es fehlen ein paar Linien des Karomusters hinter den linken Würfeln. In der Punktwolke (37(d)) ist das Papier etwas mehr gewölbt als der Untergrund der zwei Würfeln. Die Textur des Holzes scheint besser für diese Anwendung geeignet als kariertes Papier.

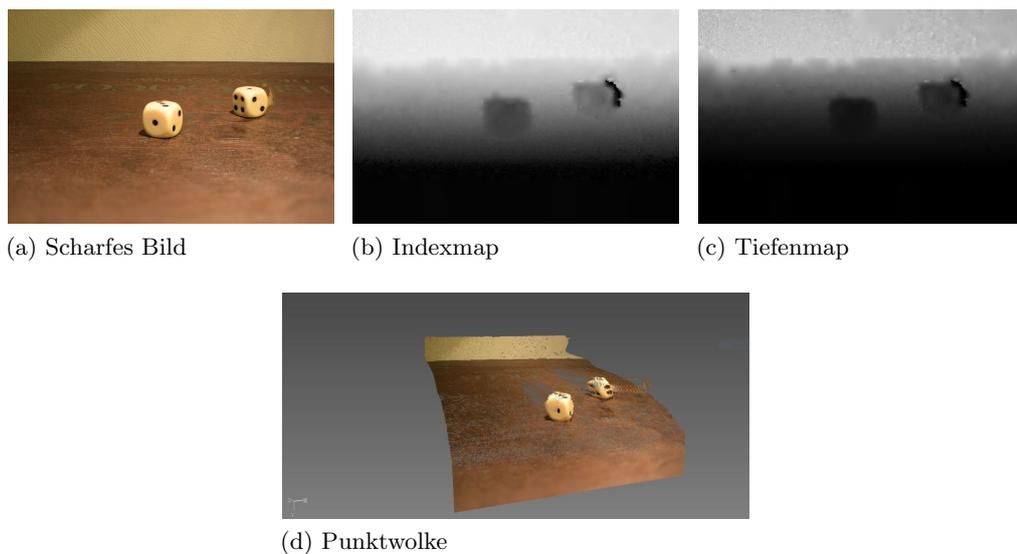
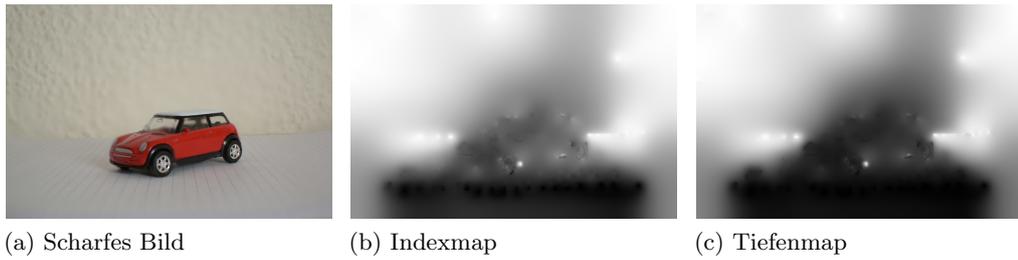


Abbildung 38: Zwei Würfel (4)

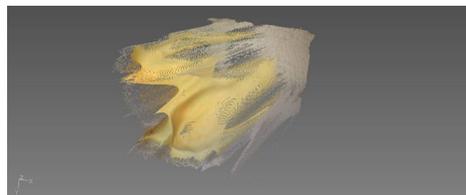
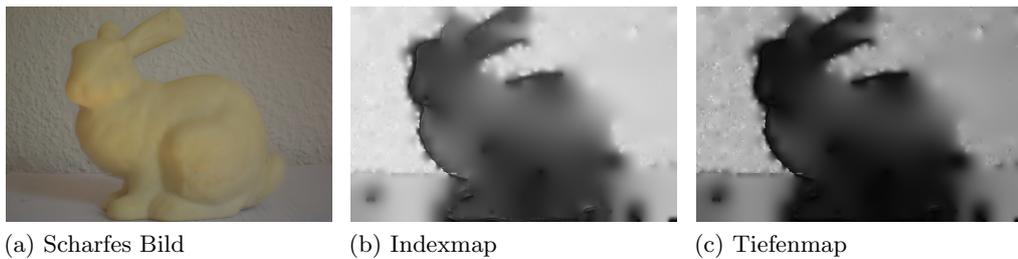
In der Index- beziehungsweise Tiefenmap mit zwei Würfeln (Abbildung 38) fällt beim rechten Würfel sehr stark ein Halo auf. An dieser Stelle waren die Kanten des stark verschwommenen Würfels bei geringer Fokusdistanz stärker als die Kanten des strukturierten Holzes bei passender Fokusdistanz. Dementsprechend wurden diesen Pixeln geringere Indizes und somit Entfernungen zugeordnet. Auch im scharfen Bild kann man die Auswirkung erkennen. Rechts des Würfels ist ein Bereich ebenfalls würfelfarben, der eigentlich zum Untergrund gehört. In der Punktwolke (38(d)) ist die Auswirkung ebenfalls zu erkennen. Am Rand des rechten Würfels liegen einige Punkte sehr deutlich vor diesem.

Ein Beispiel bei dem die Tiefenrekonstruktion so schlecht funktioniert, dass man in der Tiefenmap kaum etwas erkennen kann ist die Modellauto-Szene in Abbildung 39. Zusätzlich zu den sehr verschwommenen Konturen des Autos gibt es hier einige starke Fehler am Übergang vom Papier zur Wand. Erstaunlicherweise ist in der Punktwolke das Auto problemlos zu erkennen. Vermutlich liegt dies an der im Vergleich zur Tiefenmap hinzugekommenen Farbe. Aufgrund der Verzerrung ist deutlich zu erkennen, dass die rekonstruierten Tiefen nicht korrekt sind.



(d) Punktwolke

Abbildung 39: Modellauto (3.5)



(d) Punktwolke

Abbildung 40: Modellhase (0)

Besonders schwer tut sich der Algorithmus mit Objekten, die kaum harte Kanten haben, sondern eher abgerundet mit weichen Konturen sind. Die Einfarbigkeit des Hasen in Abbildung 40 ist hier auch nicht von Vorteil. Durch den Schwellwert bei der Indexbestimmung sind hier bereits so viele Pixel ausgeschieden, dass durch ein weiteres Aussortieren von Pixeln über die Güteheuristik keine Verbesserung der Ergebnisse erzielt werden konnte. In der Index- beziehungsweise Tiefenmap lässt sich lediglich die Form des Hasen erkennen. Tiefenunterschiede innerhalb

des Hasen sind zwar im Ergebnis vorhanden, allerdings entsprechen diese nicht unbedingt der Realität. Dass der Kopf näher ist als der Hals ist korrekt, aber bei Brust und Schwanz ist dies nicht der Fall. Die Punktwolke 40(d) bestätigt dieses Bild. Hier lässt sich nicht einmal erkennen, um welches Tier es sich handelt.

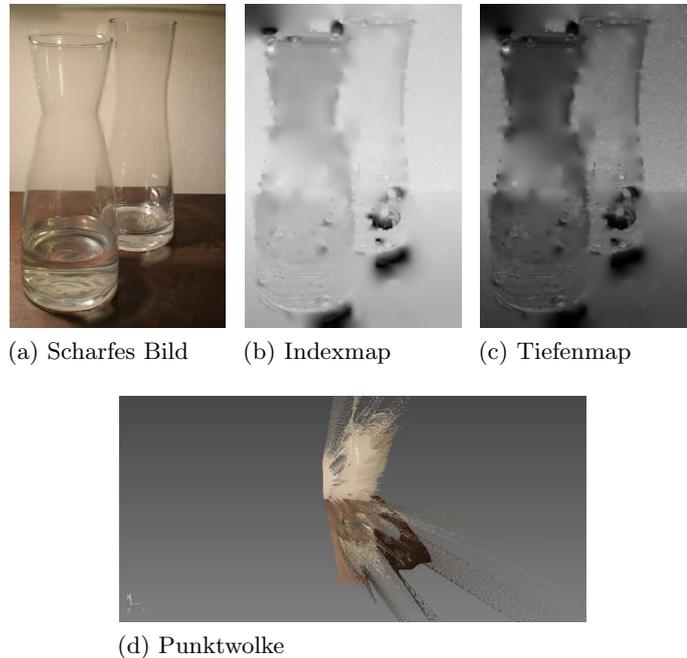
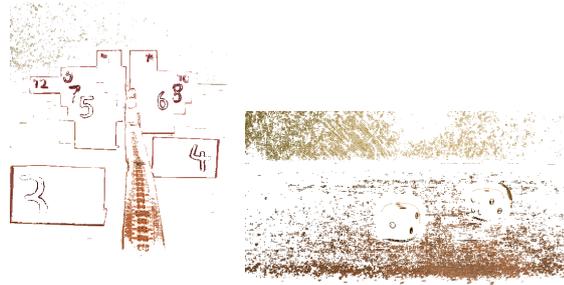


Abbildung 41: Glaskaraffen (3.75)

Die Tiefenergebnisse bei der Anwendung auf transparente Gegenstände sind nicht wirklich brauchbar. In Abbildung 41 ist eine Szene mit zwei Karaffen, von denen eine zum Teil mit Wasser gefüllt ist, dargestellt. Hier wird jeweils die Tiefe von dem Objekt zugeordnet, welches an der entsprechenden Stelle vornehmlich zu sehen ist. Auf mittlerer Höhe der linken Karaffe liegt die Kante zwischen Untergrund und Wand. Den Pixeln an dieser Stelle wurde die Entfernung zu dieser Kante und nicht die zur Karaffe zugeordnet. Dementsprechend ist die Z-Koordinate der meisten Punkte in der Punktwolke 41(d) falsch.

Von Bedeutung ist auch die Textur des Hintergrundes. Wie in Abbildung 42(a) zu sehen, wurde die Tiefe der Tischplatte nie direkt bestimmt, sondern lediglich die Tiefen der Karten und des Meterstabs interpoliert. Der Untergrund der zwei Würfel hat hingegen eine so ausgeprägte Textur, dass der Gradient dieser ausreichend stark wird. Im lückenhaften scharfen Bild 42(b) ist zu erkennen, dass die Güte zahlreicher Pixel oberhalb des Schwellwertes liegt. So müssen keine großen Flächen gefüllt werden.



(a) Meterstab (b) Zwei Würfel

Abbildung 42: Lückenhaftes scharfes Bild

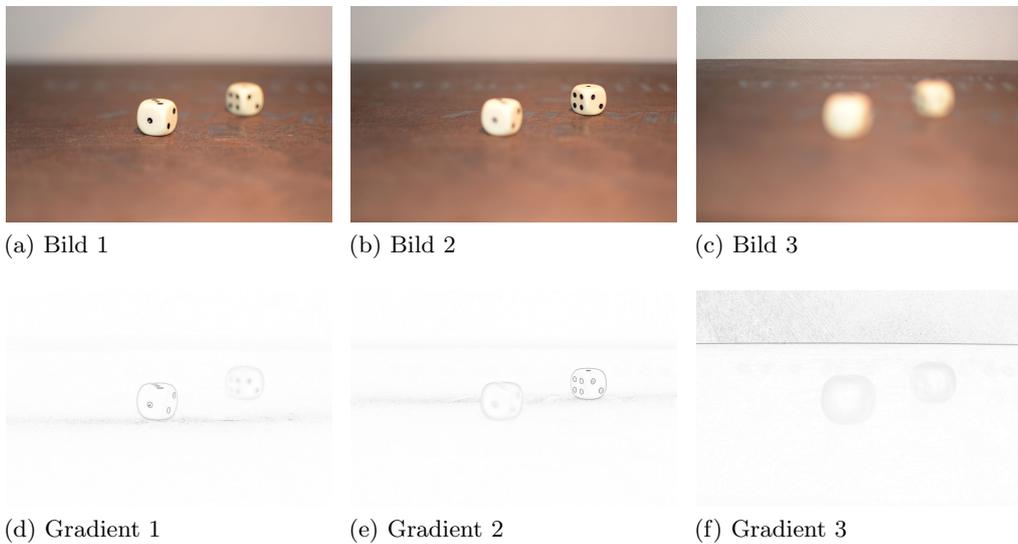


Abbildung 43: Hochauflösende kleine Bilderserie mit zwei Würfeln

Es wurde ebenfalls getestet eine Szene mit deutlich weniger Bildern (3), dafür mit höherer Auflösung (3000 x 2000 Pixel) aufzunehmen. In Abbildung 43(a) bis 43(c) sind die drei Ausgangsbilder hierfür abgebildet. Im ersten Bild (43(a)) ist der vordere Würfel scharf gestellt, im zweiten (43(b)) ist es der zweite und im dritten Bild (43(c)) liegt die Wand im Fokus. Darunter (43(d) bis 43(f)) sind die jeweiligen Gradienten zu sehen.

Abbildung 44 zeigt die Ergebnisse dieser Mini-Serie. Die Ebenen, in denen die Würfel liegen, und die Wand sind scharf (44(a)). In der Indexmap (44(b)) beziehungsweise Tiefenmap (44(c)) ist zu erkennen, dass eben diesen Bereichen der richtige Index beziehungsweise die richtige Entfernung zugeordnet wurden. Jedoch ist beim linken Würfel eine Art Halo zu erkennen. Hier wurden einigen Pixeln knapp neben dem Würfel der Index des dritten Bildes zugeordnet. Die Flächen,

die in keinem der Originalbilder im Fokus liegen, sind auch im Ergebnis unscharf. Die entsprechenden Stellen wurden in der Indexmap durch den Interpolationsalgorithmus gefüllt. Im scharfen Bild wurden dort die Pixel des durch die Interpolation bestimmten Indexes verwendet. Im Vergleich zur niedrigeren Auflösung mit mehr Bildern (Abbildung 38) ist hier der Halo um den rechten Würfel verschwunden. Es stellt sich nun die Frage, ob dies an der höheren Auflösung liegt oder daran, dass es weniger Bilder gab. Meine Vermutung ist die zweite Möglichkeit. Bei mehr Bildern gibt es in der Nähe der Kante immer ein Bild, bei dem der Würfel so weit verschwommen ist, dass an dieser Stelle eine Kante zwischen verschwommenem Würfel und Hintergrund gefunden werden kann.

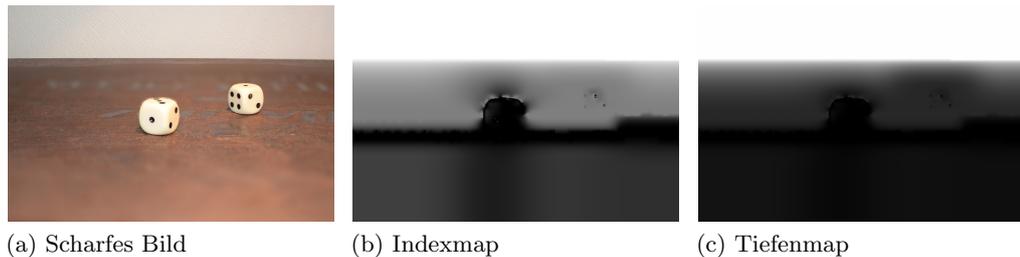


Abbildung 44: Zwei Würfel in hoher Auflösung mit kleiner Serie (2.1)

Insgesamt ist festzuhalten, dass die Rekonstruktion scharfer Bilder ziemlich gut funktioniert. An manchen Kanten kann es zu Halos kommen. Diese haben auch Pertuz et al., die sich besonders auf die Robustheit gegenüber Rauschen konzentriert haben, in ihren Ergebnissen festgestellt. [Per+12]

Die Qualität der bestimmten Tiefenwerte hingegen ist stark von der Szene anhängig. Sie reicht von absolut unbrauchbar (Gartenstuhl, Modellhase, Glaskaraffen) bis zu relativ gut (vier und zwei Würfel). Aber auch bei den besseren Tiefenrekonstruktionen gibt es noch zwei Probleme. Das eine sind die Halos, die lokal sehr starke Abweichungen der richtigen Tiefe bewirken. Das zweite Problem ist nicht direkt ein Problem der Tiefenwerte, sondern eher eines der X- und Y-Koordinaten. Durch die Krümmung können alle Punkte, die zu nah an der Kamera liegen nicht sinnvoll verwendet werden.

Wie in Kapitel 3 Verwandte Arbeiten schon erwähnt, ist die Laufzeit des hier verwendeten Algorithmus deutlich länger als die der anderen vorgestellten Ansätze. Auf einem Intel i5-7200U braucht die Berechnung einer Szene bei 2.5 GHz Taktrate im Schnitt zwei Minuten. Die meiste Zeit wird dabei auf die Berechnung der Gradienten verwendet. Für die Auswertung der höher aufgelösten Bilderserie werden ca. 1:20 Minuten benötigt.

Zusammenfassend lässt sich also sagen, dass bei gut texturiertem Hintergrund und Objekten sowie dem richtigen Mindestabstand zur Kamera durchaus zufriedenstellende Ergebnisse erzielt werden können.

7 Ausblick

Es gibt mehrere sinnvolle Ansätze die Ergebnisse weiter zu optimieren. Eine Möglichkeit wäre, die Eingabedaten zu verbessern. Mit der aktuell verwendeten Hardware sind zumindest theoretisch deutlich höher auflösende Aufnahmen möglich als die Vorschau-Bilder mit einer Auflösung von 640 x 424 Pixeln. Um zu vermeiden, dass die Mechanik der Kamera übermäßig belastet wird, könnte man die Videofunktion verwenden. Bei Filmaufnahmen ist der Spiegel dauerhaft hoch geklappt und die Auflösung der einzelnen Frames wäre trotzdem gut. Hier läge die Herausforderung darin, einem bestimmten Frame eine Fokus-Einstellung zuzuordnen. Sollte es möglich sein das Video direkt auf den angeschlossenen PC zu übertragen, so könnte nach jeder Fokusbewegung der aktuelle Frame abgespeichert werden.

Höher aufgelöste Eingangsbilder hätten den offensichtlichen Vorteil, dass die Auflösung der Ergebnisse auch entsprechend höher wäre. Als weiterer Vorteil käme noch hinzu, dass feinere Texturen besser hinsichtlich ihrer Schärfe beurteilt werden können. Bilder mit höherer Auflösung wurden bereits in Kapitel 6 Ergebnisse untersucht. Würde hier die Anzahl der Bilder erhöht werden, so ist neben der längeren Laufzeit des Algorithmus eigentlich nur eine Verbesserung der Tiefenauflösung zu erwarten.

Bei der Indexbestimmung über den maximalen Gradienten wäre es eine Überlegung wert, nicht den Index zu wählen bei dem der Gradient am höchsten ist, sondern den bei dem er am dritthöchsten ist. So kann Ausreißern entgegengewirkt werden. Allerdings sind bei allen bisherigen Versuchen keine Ausreißer aufgefallen. Für eine stärker rauschende Bildquelle wäre dieser Schritt aber sicherlich sinnvoll. Zumindest dann, wenn diese Methode der Indexbestimmung verwendet werden soll.

Ein weiterer Punkt, der verbessert werden könnte, wäre die Bewertung der Pixelgüte. Pixeln, die in der Nähe einer Kante liegen und somit zwei Spikes im Gradientenverlauf haben, wird bei der aktuellen Heuristik (Quotient aus maximalem und durchschnittlichem Gradienten mit Generalisierung) ein relativ hoher Gütewert zugeordnet. Auch bei der für diese Pixel am besten geeigneten Indexbestimmung über gewichtete Indizes mit Schwellwert sind die entsprechenden Indizes nicht sehr gut. Denkbar wäre einen statistischen Test oder ähnliches einzuführen, um bei guter Bewertung durch die Heuristik zu unterscheiden, ob der Gradientenverlauf einen oder zwei Peaks hat. Bei zwei Peaks könnte dann entweder der Mittelwert der beiden Indizes, die zu den Peaks gehören, gebildet werden oder das Pixel würde verworfen werden.

Ein deutliches Verbesserungspotential gibt es bei sehr kleinen Entfernungen. Hier sind die berechneten Punkte wegen der Krümmung unbrauchbar. Ein besseres Modell zur Umrechnung der Bildkoordinaten in Weltkoordinaten könnte entwickelt werden. Alternativ könnte versucht werden die entsprechenden Formeln experimentell zu bestimmen.

Es könnte außerdem noch untersucht werden, wie sich die Qualität in Abhängigkeit der Anzahl der Bilder verhält. Wenn die Fokusebenen in den Ausgangsbildern weiter auseinander lägen, würden weniger Bilder benötigt werden, um die gesamte Tiefe abzudecken. Dies würde sich in einer entsprechend schnelleren Berechnung positiv auswirken. Die Auflösung der Tiefe wäre allerdings niedriger. Hier gälte es also die Anzahl der Bilder soweit zu reduzieren, dass die Qualität der Ergebnisse nicht zu sehr leidet.

Für eine praktische Anwendung bei der die Tiefe einzelner Gegenstände untersucht werden soll, ist die Wahl des Hintergrundes vor dem diese Gegenstände aufgenommen werden von Bedeutung. Eine Untersuchung mit dem Ziel die bevorzugten Eigenschaften eben dieses Hintergrundes herauszufinden, wäre interessant.

Zur Beurteilung weiterer Forschungen wäre die Auswahl und Bereitstellung von geeigneten Testszenen mit echten Tiefenwerten, etwa über eine Tiefenkamera hilfreich. So könnten zukünftige Arbeiten in diesem Bereich besser bewertet und miteinander verglichen werden.

8 Zusammenfassung

Der gesamte Ablauf ist noch einmal in Algorithmus 7 zusammengefasst³. Die Ergebnisse der einzelnen Schritte sind in Abbildung 45 dargestellt.

Zunächst wird eine Bilderserie mit kleiner Schärfentiefe aufgenommen. Dabei wird die Fokusdistanz zwischen jedem Bild erhöht. Von den bis zu 1000 Bildern sind in Abbildung 45 repräsentativ nur zwei dargestellt.

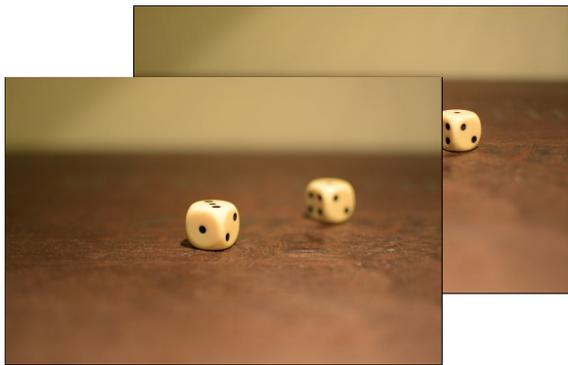
Algorithmus 7 Überblick über den gesamten Ablauf

```
1: Nehme Bilder auf
2: Bestimme Gradienten
3: for each pixel do
4:   Bestimme Index
5:   Bestimme Pixelgüte
6:   if Pixelgüte < Grenzwert then
7:     verwerfe Pixel
8:   end if
9: end for
10: Interpoliere Indexmap
11: Generiere scharfes Bild aus Indexmap
12: Generiere Tiefenmap aus Indexmap
13: Generiere Punktwolke aus Tiefenmap
```

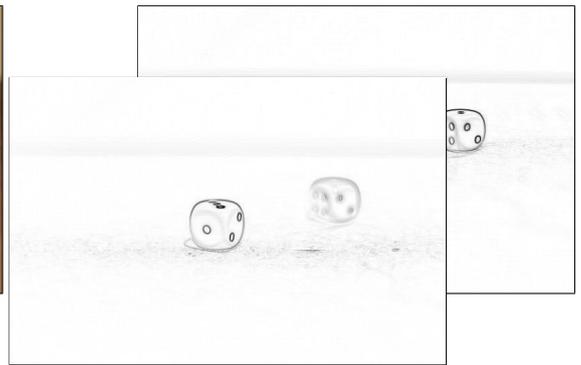
Anschließend wird mithilfe des Sobel-Filters von jedem der Ausgangsbilder der Gradient bestimmt 45(b). Dieser dient als Schärfemaß für den nächsten Schritt. Zur Bestimmung der korrekten Indizes wurden mehrere Methoden vorgestellt und verglichen. Letztendlich bewährt hat sich die Methode, bei der ein gewichteter Durchschnitt über die Indizes gebildet wird, deren Schärfe über einem gewissen Schwellwert liegt.

Nachdem so für alle Pixel, bei denen der Gradient in mindestens einem Bild über dem Schwellwert liegt, ein Index bestimmt wurde, wird die Güte der Pixel bewertet. Nur bei Pixeln, die auf einer ausreichend ausgeprägten Kante liegen, ist die Zuordnung des Index über den Verlauf des Gradienten sinnvoll. Der Index von Pixeln, die in einer Fläche ohne oder mit nur wenig Textur liegen, wird daher wieder verworfen. Es wurden mehrere Heuristiken für die Gütebewertung getestet. Eine Auswahl wurde hier vorgestellt und verglichen. Der beste Kandidat war eine Anpassung des Quotienten aus maximalem und durchschnittlichem Gradienten. An dieser Stelle im Ablauf liegt nun eine löchrige Indexmap (Abbildung 45(c)) vor.

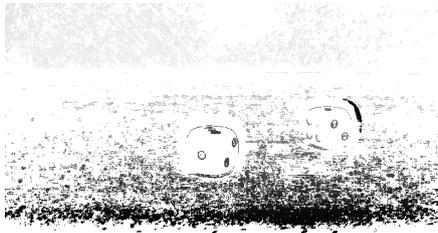
³Der hier dargestellte Ablauf entspricht nicht genau der Implementierung, sondern dient nur zur Verdeutlichung.



(a) Bilderserie



(b) Gradienten (invertiert)



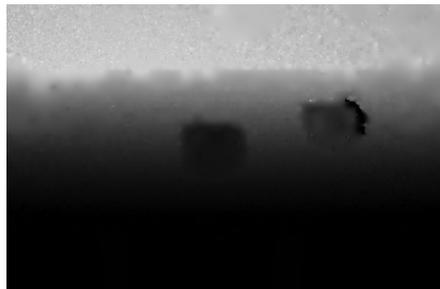
(c) Lückenhafte Indexmap



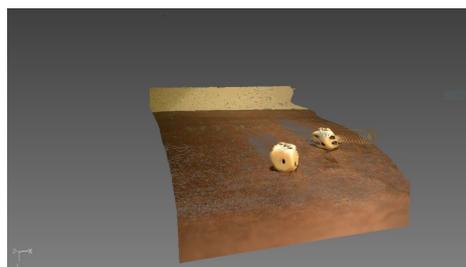
(d) Gefüllte Indexmap



(e) Scharfes Bild



(f) Tiefenmap



(g) Punktwolke

Abbildung 45: Die Ergebnisse der einzelnen Schritte

Um die Löcher in dieser zu füllen wird ein pyramidenbasierter Interpolationsalgorithmus eingesetzt. Das Ergebnis ist die komplett gefüllte Indexmap (45(d)).

Aus der gefüllten Indexmap kann mithilfe der Ausgangsbilder ein komplett scharfes Bild rekonstruiert werden (45(e)). Die Ergebnisse dieser Rekonstruktion sind relativ gut. Lediglich an sehr starken Kanten kommt es zu Halo-Effekten.

Wenn die Formel zur Umrechnung der Indizes zu Entfernungen, welche über eine Kalibrierung bestimmt werden kann, bekannt ist, kann aus der Indexmap problemlos eine Tiefenmap (45(f)) berechnet werden. Die so berechneten Tiefen unterliegen zum Teil relativ großen Schwankungen. Je detailreicher und näher das Objekt ist, desto besser sind die Ergebnisse.

Aus der Tiefenmap wiederum, können mit einigen Informationen über die Kamera und das Objektiv die Koordinaten der Punkte bestimmt werden. Die so erzeugte Punktwolke ist ebenfalls von abwechslungsreicher Qualität.

Referenzen

- [BSS17] P. Bühler, P. Schlaich, and D. Sinner. *Digitale Fotografie*. Springer Vieweg, 2017.
- [FSV04] D. Fofi, T. Sliwa, and Y. Voisin. “A comparative survey on invisible structured light”. In: *Electronic Imaging 2004*. 2004.
- [Goc12] T. Gockel. *Kompendium digitale Fotografie*. Springer, 2012.
- [Gre18] E. Grey. “Understanding Depth of Field – A Beginner’s Guide”. In: *photographylife.com* (2018).
- [Gro18] T. Grosch. *Computergraphik I Vorlesungsfolien*. 2018.
- [Gro87] P. Grossma. “Depth from focus”. In: *Pattern Recognition Letters* (1987).
- [98] *Lexikon der Physik*. Spektrum Akademischer Verlag, 1998.
- [Lou11] K. Louie. *Fall, sun, leaves and bokeh!* 2011. URL: [https://commons.wikimedia.org/wiki/File:Fall,_sun,_leaves,_and_bokeh!_\(6321333937\).jpg](https://commons.wikimedia.org/wiki/File:Fall,_sun,_leaves,_and_bokeh!_(6321333937).jpg).
- [Mar+17] J. N. P. Martel, L. K. Muller, S. J. Carey, J. Muller, Y. Sandamirskaya, and P. Dudek. “Real-Time Depth From Focus on a Programmable Focal Plane Processor”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* (2017), pp. 1–10.
- [Mar+15] M. Martinello, A. Wajs, S. Quan, H. Lee, C. Lim, T. Woo, W. Lee, S.-S. Kim, and D. Lee. “Dual Aperture Photography: Image and Depth from a Mobile Camera”. In: *ICCP 2015*. 2015.
- [Nik15] Nikon. *Digitalkamera D5500 Kompakthandbuch*. 2015.
- [Per+12] S. Pertuz, D. Puig, M. García, and A. Fusiello. “Generation of All-in-Focus Images by Noise-Robust Selective Fusion of Limited Depth-of-Field Images”. In: *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* (2012).
- [Ray03] E. Raymond. *Jargon File 4.4.7*. 2003. URL: <http://catb.org/jargon/html/L/lenna.html>.
- [SA08] R. Safaei-Rad and M. Aleksic. “Depth map from focus for cell-phone cameras”. In: *Proceedings of SPIE* (2008).
- [SKE06] M. Strengert, M. Kraus, and T. Ertl. “Pyramid Methods in GPU-Based Image Processing”. In: *Proceedings of VMV 2006*. 2006.
- [Tsc18] D. Tschumperlé. *The CImg Library*. 2018. URL: cimg.eu.
- [VM] A. Veensta and M. Mulder. *Efficient Point-Based Rendering Using Image Reconstruction*.