

# Evaluierung eines nichtlinearen Disparitätsmapping Ansatzes zur Anpassung von Stereo 3D Inhalten

DIPLOMARBEIT

zur Erlangung des akademischen Grades

**Diplom-Ingenieur/in**

im Rahmen des Studiums

**Medieninformatik**

eingereicht von

**Sabine Stransky**

Matrikelnummer 0525589

an der  
Fakultät für Informatik der Technischen Universität Wien

Betreuung  
Betreuerin: Mag. DI Dr. Margrit Gelautz, ao. Univ. Prof.  
Mitwirkung: Dipl.-Ing. Florian Seitner

Wien, 16.09.2012

\_\_\_\_\_  
(Unterschrift Verfasserin)

\_\_\_\_\_  
(Unterschrift Betreuerin)

# Erklärung zur Verfassung der Arbeit

Sabine Stransky  
Kleine Stadtgutgasse 11-13/2/8  
1020 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 16.09.2012, \_\_\_\_\_

## Kurzfassung

Bei der Produktion von hochqualitativen 3D-Filmen mittels Stereokameras stehen Filmschaffende oft vor der Herausforderung, eine Veränderung des 3D-Tiefeneindrucks in der Post-Produktion durchführen zu müssen, um ein angenehmes Tiefenerlebnis auf unterschiedlich großen Bildschirmen zu erhalten. Ziel dieser Arbeit ist es, ein Verfahren zur nachträglichen Korrektur des Tiefeneindrucks mittels des in „Nonlinear Disparity Mapping“ [Lang et al. 2010] vorgestellten Ansatzes zu untersuchen.

Im ersten Schritt erfolgt eine Literaturrecherche über verschiedene Methoden zur Anpassung der Tiefenparameter für stereoskopische Filminhalte. Dabei wird auf Methoden zur tiefenbasierten Generierung von stereoskopischen 3D-Ansichten und auf die damit verbundenen Probleme eingegangen. Die Methode für „Nonlinear Disparity Mapping“ wird im Detail vorgestellt. Im Weiteren wird ein Datensatz an Testbildern zusammengestellt und beschrieben. Dieser Datensatz wird anschließend zur Evaluierung der Methode mittels einer Benutzerstudie verwendet. In dieser subjektiven Evaluierung werden unterschiedliche Fragestellungen, bezüglich Tiefeneindruck und Sichtbarkeit von Artefakten, bei verschiedenen Kamerapositionen und Tiefeneinstellungen untersucht.

Die Arbeit hat nach Auswertung der durchgeführten Usertests ergeben, dass eine Anpassung der Tiefenparameter mit der untersuchten Methode möglich ist, dass dies aber für externe Kamerapositionen in Bezug auf das Auftreten von unbeabsichtigten Artefakten in der Szene eine Verschlechterung darstellt.

## **Abstract**

When producing a high quality 3D film via stereo cameras, producers often have to face the challenge to make changes of the depth perception in post production, in order to get a more comfortable depth experience on different screen sizes. The goal of this master thesis is to analyze a method for the post correction of the depth perception by „Nonlinear Disparity Mapping“ [Lang et al. 2010].

In the first step, a literature review of the different methods for the adaption of depth parameters for stereoscopic films is performed. Here methods for a depth based generation of stereoscopic 3D views and the problems associated with it are discussed. The method for „Nonlinear Disparity Mapping“ [Lang et al. 2010] is described in detail. Furthermore, a dataset of test images is compiled and described. This dataset is used for the evaluation of the method by conducting a user study. In this subjective evaluation, different questions with various camera positions and depth options are investigated.

After an analysis of the performed user tests, the results have revealed that an adjustment of the depth parameters with the explored method is possible. For external camera positions, however, the occurrence of unintended artifacts was observed.



# Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Stereo sehen und aufnehmen.....	1
1.2	Anpassung des Tiefeneindrucks mittels linearem Disparitätswarping.....	4
1.3	Motivation und Zielsetzung der Arbeit.....	7
1.4	Aufbau der Arbeit.....	8
2	Nichtlineare Bildtransformation basierend auf Saliency Maps.....	9
2.1	Einsatz von Saliency Maps bei der Skalierung von 2D Inhalten.....	10
2.2	Einsatz von Saliency Maps bei der Segmentierung von 2D Inhalten.....	14
2.3	Einsatz von Saliency Maps bei der Modellierung von 3D Inhalten.....	15
3	Systembeschreibung.....	17
3.1	Berechnung der Saliency Map.....	17
3.1.1	Lokale Kantenkarte.....	17
3.1.2	Globale Texturkarte.....	18
3.1.3	Disparitätsbasierende Karte.....	20
3.1.4	Saliencyberechnung.....	21
3.2	Warping.....	25
3.3	Fusion des linken und rechten Stereobildes.....	27
3.4	Inpainting.....	30
4	Usertests und Ergebnisse.....	33
4.1	Bilder und Eigenschaften.....	33
4.2	Kamerasetup und generierte Ansichten.....	35
4.3	Disparitätskarten.....	37
4.4	Darstellung und Umsetzung.....	38
4.5	Validierung.....	40
4.5.1	Gesamtauswertung für Frage (a).....	41
4.5.2	Auswertung der einzelnen Testbilder für Frage (a).....	43
4.5.3	Gesamtauswertung für Frage (b).....	62
4.5.4	Auswertung der einzelnen Testbilder für Frage (b).....	65
5	Zusammenfassung und Ausblick.....	75
	Referenzen.....	76
	Abbildungsverzeichnis.....	80

Tabellenverzeichnis.....	82
Anhang I: Auszug - Regressionsmatrizen für den zusätzlichen Faktor Sehhilfe bei der Frage nach dem Tiefeneindruck.....	83

# 1 Einleitung

Stereoskopische 3D Inhalte werden sowohl für Filme im Kino, Fernsehen und auf Blue-Rays, als auch für sonstige Fernsehsendungen und Sportübertragungen in Echtzeit immer populärer. Es ist aber noch immer eine Herausforderung, die stereoskopischen Inhalte derart aufzubereiten, dass sie für den Zuseher ein komfortables Seherlebnis liefern.

## 1.1 Stereo sehen und aufnehmen

Bei der Generierung von 3D Inhalten mit zwei Kameras (Stereoaufnahme) wird jede Szene von zwei Kameras aus unterschiedlichen Betrachtungswinkeln aufgenommen, um daraus Tiefeninformationen zu erhalten. Wie in **Abbildung 1** gezeigt, entspricht das linke Stereobild dabei dem Bildeindruck des linken Auges und das rechte Stereobild entspricht dem des rechten Auges.

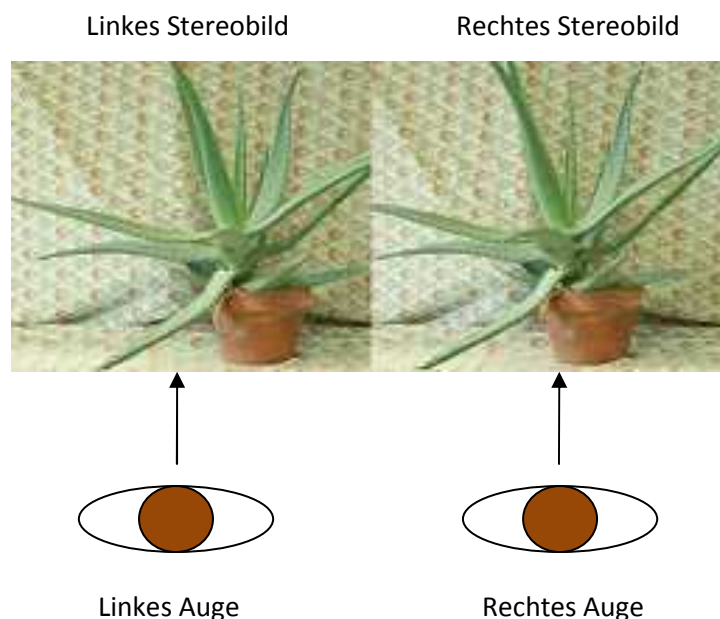


Abbildung 1: Stereo sehen: Das linke Stereobild entspricht dem Bildeindruck des linken Auges und das rechte Stereobild entspricht dem Bildeindruck des rechten Auges. (Stereobilder übernommen von [Middleburry 2006].)

Der Abstand zwischen den beiden Kameras ist variabel und wird Kamera-Baseline genannt. Die Tiefe der Objekte im Bildraum wird in der sogenannten Tiefenkarte gespeichert (siehe **Abbildung 2**). Dabei handelt es sich um ein Grauwertbild, in dem die helleren Objekte näher beim Betrachter liegen. Je dunkler der Grauwert ist, desto weiter im Hintergrund liegt das betreffende Objekt. [Schu 2010]

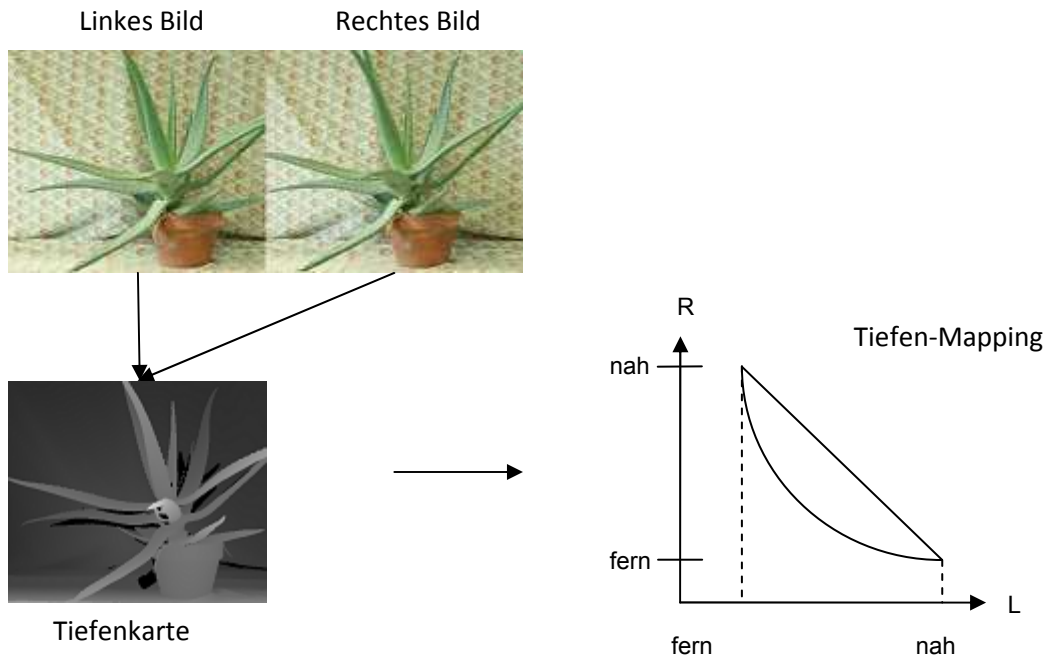


Abbildung 2: Von den aus unterschiedlichen Winkeln aufgenommenen Kameraaufnahmen für das rechte und linke Auge kann eine Tiefenkarte generiert werden. Das Ausmaß der Positionsänderung eines Objekts im Bild gibt an, wie nah ein Objekt beim Betrachter liegt. Helle Objekte in der Tiefenkarte zeigen die nahen Objekte an, wobei die Helligkeit des Grautons nachlässt, je tiefer ein Objekt liegt [Schu 2010]. (Stereobilder übernommen von [Middlebury 2006].)

Der Abstand zwischen korrespondierenden Punkten im linken und rechten Stereo Bild wird als Disparität bezeichnet (siehe **Abbildung 3**). Je näher ein Punkt dabei beim Auge des Betrachters liegt, desto größer ist die Disparität, das heißt, umso mehr verschiebt sich der Punkt im generierten Bild. Objekte im Vordergrund verschieben sich daher über größere Distanzen als Objekte im Hintergrund. Verdeckt man ein Auge und anschließend das andere, dann lässt sich dieser Effekt leicht nachstellen. Objekte, die im 3D Bild vor der Bildschirmenebene liegen, weisen eine negative Disparität auf, während die Disparität der Objekte hinter der Bildschirmenebene positive Werte umfasst. Durch diese Verschiebungen der Objekte für das linke bzw. rechte Auge entsteht räumliches Sehen. [Sonka et al. 1999]

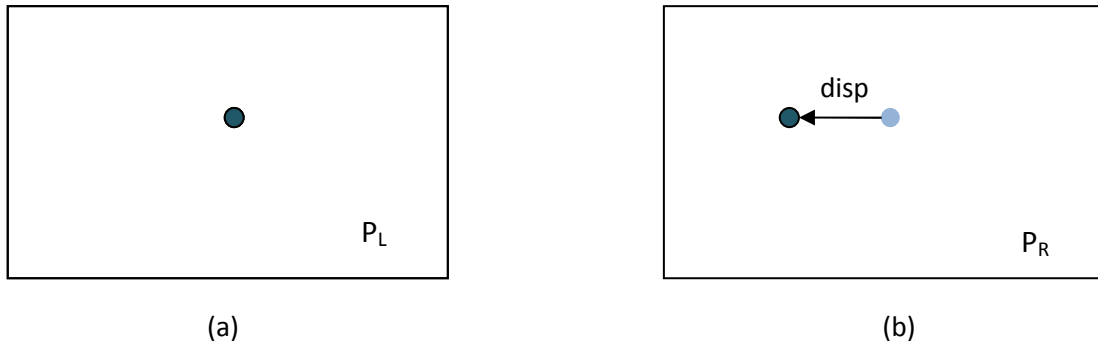


Abbildung 3: (a) Punkt in linker Stereoaufnahme ( $P_L$ ). (b) Verschiebung des Punktes um einen Disparitätswert (disp) im rechten Stereobild ( $P_R$ ).

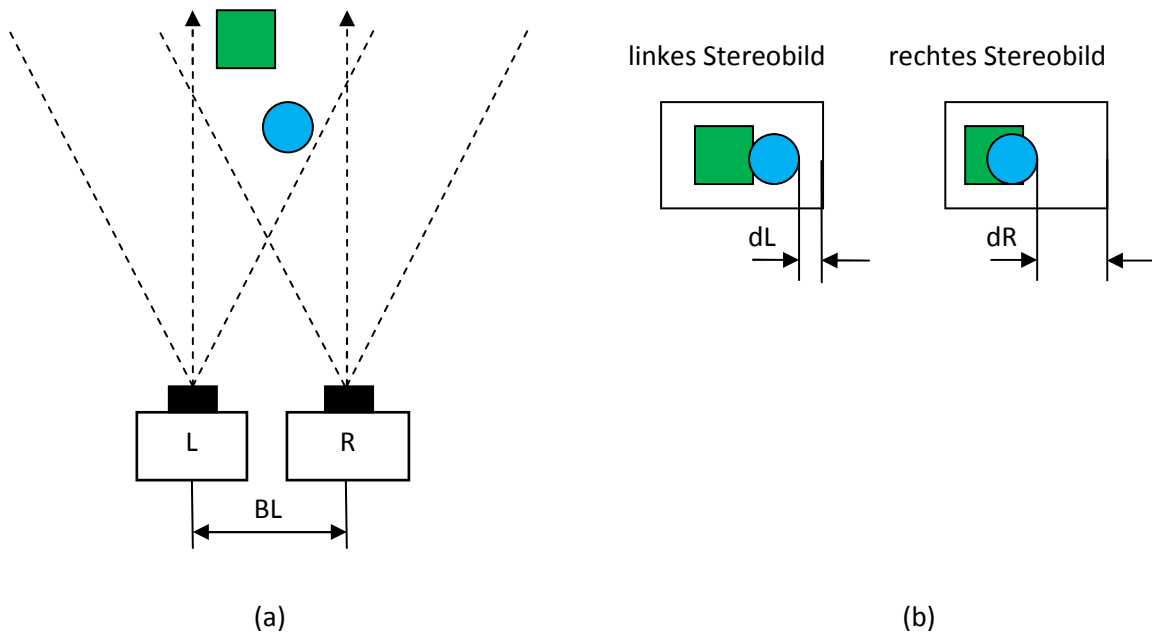


Abbildung 4: Aufnahme eines Stereobildpaares; (a) Aufteilung der Objekte und Kameras im Raum. (BL...Baseline) (b) Resultierendes Stereobildpaar. [Fraunhofer 2011]

**Abbildung 4** zeigt, wie sich nahe und ferne Objekte in Stereobildpaaren verhalten. Die Disparität  $d$  ergibt sich hier durch folgende Formel [Fraunhofer 2011]:

$$d = |dL - dR|$$

Bei der Stereobildaufnahme wird außerdem die Baseline vergrößert oder verkleinert, um den gewünschten Tiefeneindruck zu erreichen. Je nachdem, ob eine Szene näher oder weiter weg sein soll, muss der Winkel zwischen den beiden Kameras größer oder kleiner sein. Dabei wird die Funktion des Auges simuliert. Während die Blickrichtungen des linken und rechten Auges bei der Betrachtung von fernen Objekten annähernd parallel sind, wird der Winkel zwischen den zwei Blickrichtungen immer größer, je näher ein Objekt am Betrachter ist (siehe **Abbildung 5**). Der Abstand zwischen den Augen ist allerdings nicht bei allen Menschen gleich, daher kann hier nur ein Durchschnittsmaß verwendet werden, welches nicht für alle Menschen passt. 3D Inhalte werden dadurch unterschiedlich wahrgenommen. [Schu 2010]

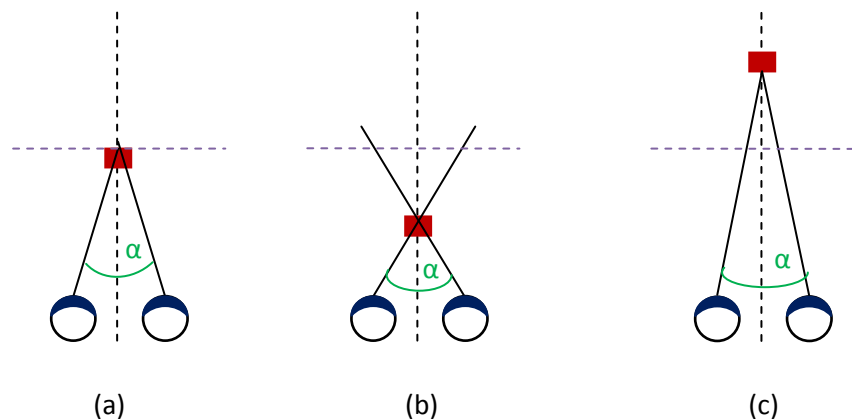


Abbildung 5: Lage eines betrachteten Objekts in Relation zur Bildebene. (a) Das Objekt (rot eingezeichnet) befindet sich auf der Bildebene (grün eingezeichnet). (b) Das Objekt befindet sich vor der Bildebene. (c) Das Objekt befindet sich hinter der Bildebene. Der Winkel  $\alpha$  der Blickrichtung wird größer, je näher das Objekt beim Betrachter ist.

## 1.2 Anpassung des Tiefeneindrucks mittels linearem Disparitätswarping

Basierend auf der Disparität, kann das linke bzw. das rechte Bild auf neue Positionen gewarpt werden (siehe **Abbildung 6**). Beim Warping werden neue Ansichten aus verschiedenen Betrachtungswinkeln (Novel Views) erzeugt, wodurch die Baseline angepasst und das Bild für unterschiedliche Bildschirmgrößen optimiert werden kann. [Lang et al. 2010]

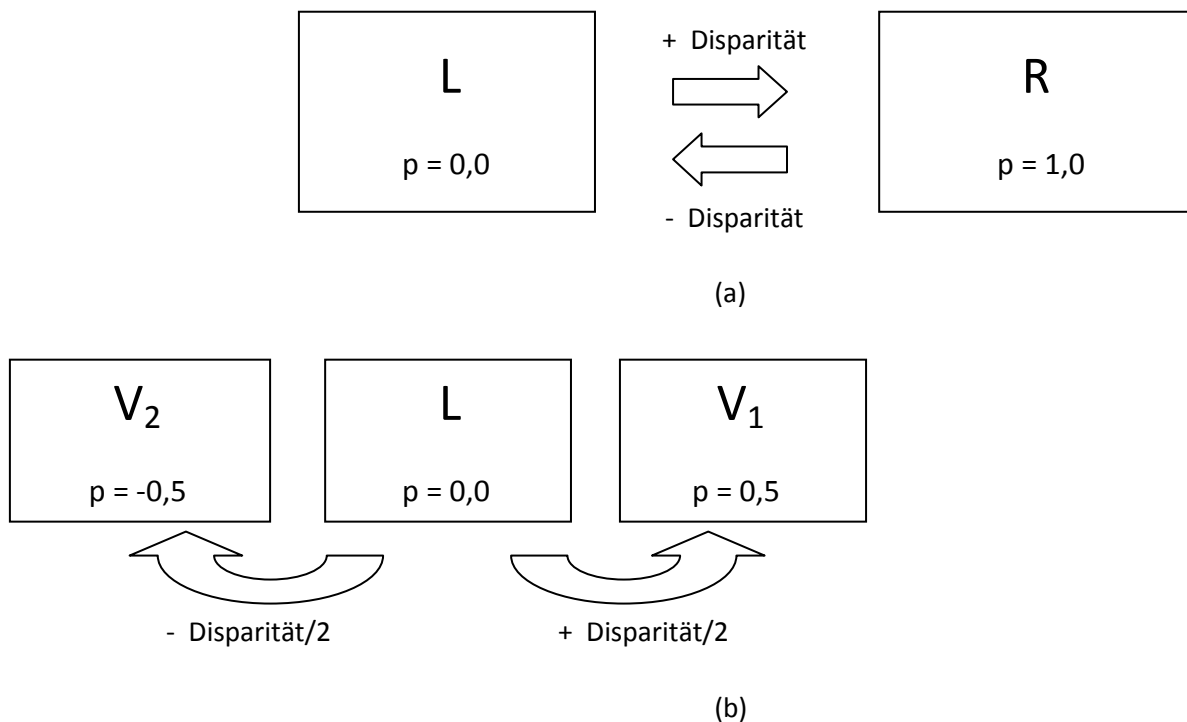


Abbildung 6: (a) Durch Addition der Disparität auf die jeweiligen Werte in der Tiefenkarte kann vom linken Bild (L, Position 0,0) ausgehend das rechte Bild (R, Position 1,0) gewarpt werden. Umgekehrt kann durch Subtraktion der Disparität auf die jeweiligen Werte in der Tiefenkarte des rechten Bildes das linke Bild gewarpt werden. (b) Durch Modifikation der Disparität können weitere virtuelle Ansichten ( $V_1$ ,  $V_2$ ) generiert werden.

Aus einem Stereobildpaar können beliebig viele neue Stereobildpaare erzeugt werden. Diese werden benötigt, damit Betrachter an unterschiedlichen Positionen im Raum das gleiche optimale 3D-Bild sehen können. Durch die Anpassung der Baseline (und damit der Disparität) können Stereobildpaare auf unterschiedliche Bildschirmgrößen oder unterschiedliche Betrachtungsumgebungen angepasst werden. [Lang et al. 2010]

Durch Warping der Bilder können externe Kamerapositionen (ausgehend von den Kamerapositionen 0,0 und 1,0: Position -2,0; Position -1,5; Position 1,5, Position 2,0; und andere) extrapoliert und interne Kamerapositionen (ausgehend von den Kamerapositionen 0,0 und 1,0: Position 0,5 oder andere Positionen zwischen 0 und 1) interpoliert werden (siehe **Abbildung 7**). Die entsprechenden Disparitätswerte in der Tiefenkarte der Ausgangsbilder (hier Position 0,0 und Position 1,0) werden mit dem Interpolationsfaktor multipliziert, um die Tiefenwerte der neuen Position zu erhalten. Mit diesen neuen Werten wird an die neue Position gewarpt, indem die Pixel im Ausgangsbild um die neuen Disparitätswerte verschoben werden. Die Transformationsfunktion ist linear. Damit ändert sich jeder Bildpunkt durch das Warping gleichermaßen [Lang et al. 2010].

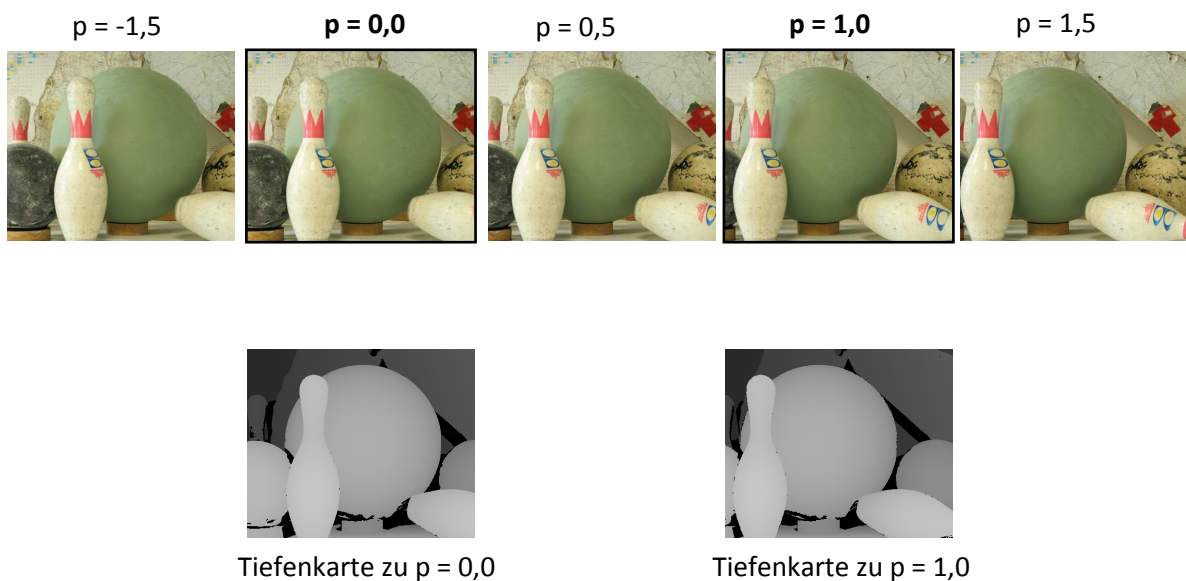


Abbildung 7: Interne und externe Kamerapositionen ausgehend von den Originalbildern an Position  $p = 0,0$  (links) und Position  $p = 1,0$  (rechts)

[DeCarlo et al. 2002] beschreiben auf Grund einer empirischen Studie, dass das menschliche Auge vor allem auf Kontrast, räumliche Frequenz und Herausstechendes reagiert. Derartige Parameter werden beim linearen Disparitätswarping allerdings außer Acht gelassen.



### 1.3 Motivation und Zielsetzung der Arbeit

Bei der Erstellung von hochqualitativen 3D Filmen mittels Stereokameras stellt die Einbeziehung des Tiefeneindrucks die Produktion und die Post-Produktion vor neue Herausforderungen. Der Tiefeneindruck, der beim Betrachter entsteht, wird hier durch eine Vielzahl an Faktoren bestimmt. Auf Seiten der Produktion bestimmt zum Beispiel die Distanz zwischen den beiden Kameras oder der Abstand der Objekte von der Kamera diesen Tiefeneindruck. Bei der Darstellung des 3D Films wird der Tiefeneindruck durch die Bildschirmgröße oder den Abstand des Betrachters zur Leinwand definiert.

Oft können diese Faktoren bei der Produktion nur unzulänglich berücksichtigt werden, was schlussendlich zu einem unangenehmen Empfinden beim Betrachten dieser stereoskopischen 3D Inhalte führt. Um dieses Problem zu lösen, wird in der Post-Produktion eine Anpassung des 3D Filminhalts an die gewünschte Darstellungsmöglichkeit durchgeführt. Dies ermöglicht es, die Tiefenparameter eines 3D Films nachträglich zu korrigieren. In „Nonlinear Disparity Mapping for Stereoscopic 3D“ [Lang et al. 2010] wird eine neue Methode beschrieben, die diese Anpassung des 3D Filminhalts ermöglicht. Im Gegensatz zum linearen Disparitätsmapping, wo der gesamte Inhalt eines Bildes mittels Tiefeninformation verändert wird, werden beim nichtlinearen Disparitätsmapping zusätzliche Faktoren wie Texturiertheit oder Kanteninformation bei der Korrektur des Tiefeneindrucks eingesetzt.

Im Zuge dieser Diplomarbeit soll der in [Lang et al. 2010] beschriebene Algorithmus implementiert und mit Hilfe von Usertests die Qualität dieser Methode evaluiert werden. Ziel ist es zu analysieren, ob mittels Saliency-basiertem Mapping der Tiefeneindruck verbessert werden kann, und festzustellen, in wie weit diese Methode rein tiefenbasierten Methoden überlegen ist.

## **1.4 Aufbau der Arbeit**

Die Arbeit ist in fünf Kapitel gegliedert. In Kapitel 1 wird, ausgehend von den Grundlagen der 3D Inhaltserstellung, auf lineares Disparitätswarping eingegangen. In Kapitel 2 werden die aktuellen Methoden für nichtlineares Disparitätswarping beschrieben. In Kapitel 3 ist die Implementierung der speziellen Methode von [Lang et al. 2010] beschrieben und wie diese Methode im Detail funktioniert. Zudem werden die weiteren getesteten Methoden zur Verbesserung der Tiefenkarte bei der Generierung neuer Ansichten, sowie zusätzliche notwendige Schritte bei der Erstellung der 3D-Bilder erklärt. Um die beschriebene Methode von [Lang et al. 2010] zu prüfen, wurden Usertests durchgeführt, welche in Kapitel 4 beschrieben sind. Darin wird zuerst das Testsetup betrachtet und anschließend die Testergebnisse im Detail behandelt. Bei den Schlussbetrachtungen in Kapitel 5 werden die Ergebnisse der Arbeit zusammengefasst und Ausblicke für zukünftige Arbeiten gegeben.

## 2 Nichtlineare Bildtransformation basierend auf Saliency Maps

In sogenannten Saliency Maps werden jene Regionen im Bild hervorgehoben, die beim Betrachten hervorstechen, d.h. auf welche das menschliche Auge primär reagiert. Auffallende Bildregionen können durch verschiedene Faktoren identifiziert werden:

- Räumliche Bildeigenschaften (Textur oder ähnliche Bildeigenschaften)
- Zeitliche Bildeigenschaften (zum Beispiel der Disparitätsbereich in vorangegangenen Frames)
- Benutzereingaben (Segmentierung)

Die damit gefundenen Regionen werden als Erweiterung der Tiefenkarte in der Saliency Map gespeichert.

Im Gegensatz zum linearen Disparitätswarping wird beim nichtlinearen Disparitätswarping nicht uniform mittels Disparitätsbereich des Bildes gewarpt. Stattdessen werden, als zusätzliche Faktoren für ein optimiertes Ergebnis, die Werte in der Saliency Map mit einberechnet. Durch diese Art des Warpings kann eine Anpassung der Tiefenparameter (Verschieben/Verändern des Tiefeneindrucks von einzelnen Objekten oder anderer Tiefenparameter) und eine zeitlich kontinuierliche Tiefenadaption erzielt werden. [Lang et al. 2010]

Im Bereich der Bild- und Videoverarbeitung wird oft von Saliency Maps Gebrauch gemacht. In dieser Karte werden die wichtigen Elemente eines Bildes oder Videoframes hervorgehoben. Um diese zu bestimmen, können verschiedene Parameter herangezogen werden. In **Tabelle 1** wird ein Überblick über unterschiedliche verwendete Parameter je nach Verwendungszweck gegeben. Die hier vorgestellten Methoden werden in den folgenden Kapiteln näher beschrieben.

## Parameterübersicht in der Bildverarbeitung

Autor	Anwendung	Saliency Map Parameter
[Avidan et al. 2007]	Video Retargeting	Kanten, Harris Eckenerkennung, Gradienten, Gesichtserkennung, Blickmessung
[Krähenbühl et al. 2009]	Video Retargeting	Textur, Bewegung, Sobel Kantendetektor, Bilateralfilter
[Liu und Gleicher 2006]	Bild- und Video Retargeting	Kontrast, Fuzzy Growing, Gesichtserkennung, Bewegung
[Ren et al. 2010]	Bild Retargeting	Mean Shift, Gaussian Mixture Model
[Setlur et al. 2005]	Bild Retargeting	Segmentierung durch Objekterkennung und Mean-Shift, Farbhistogramm
[Rubinstein et al. 2008]	Seam Carving für Video Retargeting	Kanten, Harris Eckenerkennung, Gradienten, Gesichtserkennung, Blickmessung, Betrachtung der gestörten Kontinuität zu den Nachbarpixeln
[Shamir et al. 2009]	Skalierung von virtuellen Medien durch Seam Carving	Kantenbild, Bewegung
[Wang et al. 2008]	Skalierung von Bildern durch Cropping	Farbe, Intensität, Orientierung, Gauß Pyramiden, Gradientenkarte
[Wolf et al. 2007]	Skalierung von Videos unter Berücksichtigung von Inhalten	Gradientenausprägung
[Guo et al. 2008]	Objektsegmentierung	Textur, Bewegung
[Shokoufandeh et al. 1998]	Objektrepräsentation	Wavelet-Transformation
[Zhang et al. 2009]	Objektsegmentierung	abgeändertes LASW (Locally Adaptive Support Weight)
[Ma et al. 2003]	Bildsegmentierung	Kontrast und Fuzzy Growing
[Longhurst et al. 2006]	Erstellung von High-Fidelity Bildern für die Bildverarbeitung	Objektkanten, abrupte Farbänderungen, Bewegung

Tabelle 1: Übersicht über die verwendeten Parameter in Saliency Maps in der Bildverarbeitung

### 2.1 Einsatz von Saliency Maps bei der Skalierung von 2D Inhalten

Bei der Skalierung von Bildern und Videos für Seam Carving, wie bei [Shamir et al. 2009] beschrieben, wird beispielsweise eine Saliency Map benötigt, um die Größe von Bildern oder Videos zu ändern, ohne Bildinformation zu verlieren. Der erste Schritt ist die Definition einer Saliency Map, welche die wichtigen Eigenschaften eines Bildes, die auch nach der Skalierung des Bildes beibehalten werden sollen, beschreibt. Für jedes Pixel gibt es, wie bei einer Tiefenkarte, ein auf seiner Wichtigkeit basierendes Ranking.

[Shamir et al. 2009] unterscheidet zwischen zwei Arten von Eigenschaften:

- Einfache Bildeigenschaften: Kantenbild (wichtige Konturen von wichtigen Objekten im Bild), beispielsweise durch:
  - Canny Edge Detection (Canny Kantendetektion)
  - Multiresolution Pyramide des Bildes, wo nach signifikanten Intensitäten und Farbänderungen auf allen Levels gesucht wird (verhindert Störungen)
  - Harris-Corners Detection (Harris Eckenerkennung)
  - Histogram of Gradients (HoG)
  - Kombination aus Gradientenkarte und Saliency Map
- Komplexe Bildeigenschaften: Strukturen (Textur), Symmetrien, beispielsweise durch:
  - Automatische Erkennung von Gesichtern
  - Temporäre Änderungen (Schnitte bei Szenenübergängen)
  - Kamera/Objekt Bewegung
  - Direkter Benutzereingriff auf das Bild

Im zweiten Schritt wird die Größe des Bildes, immer in Bezug auf die Saliency Map und die dort definierten Eigenschaften, geändert.

Für das Warping von Pixeln mit niedriger Wichtigkeit wird eine lineare Transformation verwendet [Shamir et al. 2009], deren Output dem Input in veränderter Größe entspricht. Während sich die Regionen mit niedriger Wichtigkeit ändern dürfen, sollen sich die Eigenschaften der als wichtig deklarierten Regionen in der Saliency Map nicht verändern und werden daher nicht-linear gewarpt. Das Warping wird für jedes Pixel einzeln durchgeführt um nach Wichtigkeit unterscheiden zu können.

Eine weitere Methode zum Skalieren von Bildern, mittels derer wichtige Objekteigenschaften erhalten bleiben sollen, beschreiben [Wang et al. 2008]. Hier ist kein zwingendes Beibehalten der Größe von wichtigen Objekten im Fokus, sondern es wird die Größe jedes Objekts in einem optimalen Verhältnis verändert. [Wang et al. 2008] verwenden die Saliency Map für Änderungen des Abbildungsverhältnisses (Aspect Ratio). Dies wird mittels Cropping, einem Verkleinern des Bildes mit anschließendem Abschneiden der Ränder, erreicht. Anders als bei anderen Cropping Methoden sollen bei [Wang et al. 2008] auch wichtige Objekte am Bildrand erhalten bleiben, wofür eine Saliency Map benötigt wird. Wichtige Objekte, für die bei [Wang et al. 2008] verwendete Saliency Map sind jene, die attraktiv für das menschliche Auge sind und die eine bestimmte Struktur aufweisen, die im Folgenden näher beschrieben wird.

[Wang et al. 2008] berechnen die Saliency Map zuerst wie bei [Itty et al. 1998], wobei nach Farbe, Intensität und Orientierung gefiltert wird. Jene Objekte, die andere Eigenschaften als ihre Umgebung haben, werden damit gefunden. Das Ergebnis soll die Attraktivität einer Region anzeigen. Die Saliency Map wird von [Wang et al. 2008] um eine Gradientenkarte erweitert, welche unterschiedliche Strukturen im Bild aufzeigt. Verzerrungen durch die Größenänderungen werden auf homogene Regionen des Bildes ohne herausragende Objekte, wie beispielsweise Regionen mit sich wiederholenden Mustern, umgelagert, wo diese nicht auffallen. Durch die Kombination der beiden Methoden werden sich wiederholende Strukturen, die keine speziellen Objekte sind, sowie scheinbar attraktive Regionen, die keine Struktur haben, ausgeschlossen. Im Gegensatz dazu verwenden [Wolf et al. 2007] bei der Skalierung von Bildern und Videos nur die Gradientenausprägung als Indiz für die Wichtigkeit eines Pixels. Die übrigen unwichtigen Pixel werden in der zu reduzierenden Richtung gemischt.

[Rubinstein et al. 2008] und [Rubinstein et al. 2009] verwenden eine verbesserte Schichtzerlegung (Seam Carving) nach [Avidan et al. 2007] für die Videoskalierung. Es wird dabei eine durchgehende Linie gezeichnet, die je ein Pixel je Zeile/Spalte beinhaltet. Diese Linie wird so eingezeichnet, dass sie keine wichtigen Regionen bzw. Regionen, die beim Ändern der Größe gelöscht werden dürfen, betrifft. Es wird dabei die Energiefunktion eines Bildes betrachtet. Pixelreihen mit niedriger Energie werden als unwichtig erachtet und können entfernt werden.

Um wichtige Regionen erkennen zu können, wird eine Saliency Map, bestehend aus einem Kantenbild, einem Punktbild und einer Gradientenkarte erzeugt. Weiters werden durch Gesichtserkennung und Blickmessung gefundene Regionen ebenfalls in der Saliency Map gespeichert. Außerdem ergibt sich die Pixelwichtigkeit durch die Betrachtung der gestörten Kontinuität zu den Nachbarpixeln, wenn das betrachtete Pixel entfernt wird. Bei dieser Methode können leicht sichtbare Artefakte entstehen, was bei einem Video weniger auffällt als bei einem einzelnen Bild. Trotzdem können diese Artefakte bei der Auswahl der zu entfernenden Pixelreihen vermindert werden, wenn man statt einzelner Frames alle Frames betrachtet. Das kann durch ein Miteinbeziehen von Bewegung (Motion) für die Saliency Map erreicht werden. [Rubinstein et al. 2009]

[Setlur et al. 2005] entwickelten einen ähnlichen Algorithmus zur Verkleinerung von Bildern, indem noch vor der Erstellung einer Saliency Map, an den Bildern eine Segmentierung in homogene (gleich beschaffenen) Regionen nach [Meer et al. 2001] und eine Mean-Shift Segmentierung vorgenommen werden. Bei der Mean-Shift Segmentierung wird nach Veränderungen der Farbmittelwerte der Pixel gefiltert. Weiters werden für die Saliency Map wiedererkennbare Objekte (Gesichter oder andere markante Objekte) und Objekte, die das visuelle System auf niedrigem Level ansprechen, also jene, die sofort auffallen, als wichtig

erachtet. Die Methode von [Itti et al. 1998] und [Niblack et al. 1993] wird dabei für die Gesichtserkennung eingesetzt.

[Liu et al. 2005] bilden ihre Saliency Map zur Bildskalierung (zum Beispiel für Handys, PDAs, Bilddatenbanken) ebenfalls auf Basis wiedererkennbarer Objekte und Regionen des Bildes, die das visuelle System auf niedrigem Level ansprechen. Liegt ein Pixel im Bild innerhalb eines Gesichtes, dann wird für dieses Pixel der Wert 1 gespeichert, außerhalb eines Gesichtes der Wert 0. Um zu beschreiben, wie stark jedes Pixel oder jede Region das visuelle System anspricht, wird die Methode von [Ma et al. 2003] verwendet. Basierend auf einem Drei-Stufen Modell werden in hierarchischer Reihenfolge folgende Ebenen betrachtet:

1. Hauptteil des Bildes, welcher die meiste Information enthält
2. Regionen mit wichtigen Inhalten, die die menschliche Aufmerksamkeit am meisten erregen
3. Stellen im Bild, von denen maximale Stimulation ausgeht

Laut [Ma et al. 2003] ist ein Schlüsselfaktor bei der menschlichen Wahrnehmung der Kontrast. Regionen mit hohem Kontrast enthalten die meisten Informationen und erregen die menschliche Aufmerksamkeit am meisten. Daher wird eine kontrastbasierende Saliency Map verwendet, um die ersten beiden Ebenen abzudecken. Weiters wird für die dritte Ebene Fuzzy Growing angewendet, wobei die wichtigen Regionen ausgeweitet werden, sodass die Übergänge zum Hintergrund nicht mehr hart, sondern verwischt sind.

[Krähenbühl et al. 2009] verwenden für Video Retargeting die Strategie von [Guo et al. 2008], welche für die Erstellung einer Saliency Map eine schnelle 2D Fourier Transformation mit Quanternisierung zur Texturerkennung verwenden, sowie Optical Flow, um auch Bewegung mit einzubeziehen. Optical Flow beschreibt ein Vektorfeld, welches die Bewegungsrichtung und -geschwindigkeit für jedes Pixel angibt. Die Saliency Map wird um Objektkanten erweitert, welche mittels des Sobel Kantendetektor errechnet werden.

[Ren et al. 2010] beschreiben eine Methode für Image Retargeting. Die Erkennung von wichtigen Regionen basiert auf Superpixeln, also Gruppen von Pixeln. Zuerst werden Superpixel im Eingabebild mittels des Mean-Shift Algorithmus erkannt. Anschließend wird ein GMM (Gaussian Mixture Model) zum Bündeln solcher Superpixel, basierend auf deren Farbähnlichkeit, angewendet. Im letzten Schritt werden die Wichtigkeitswerte jedes Bündels, mit einer Dichtheitsmetrik kombiniert, und mit RageRank (Algorithmus zur Erkennung von zusammengehörenden Bündeln) in einer Saliency Map gespeichert.

Für Video Retargeting existieren weitere Methoden zur Saliency Map Findung, die zusätzliche zeitliche Faktoren miteinbeziehen. Die Saliency Map bei [Wang et al. 2009] wird im räumlichen und zeitlichen Kontext erstellt. Es wird nicht nur ein

einzelner Frame dafür betrachtet, sondern zusätzlich eine definierte Menge an Vorgängern und Nachfolgern. Wichtige Objekte werden erkannt und in einer Bewegungs-Saliency Map gespeichert. Bewegung, die nur durch Kamerabewegung erzeugt wurde, wird ignoriert. Warping wird dabei mittels simplem Mesh-Warp durchgeführt. Die Grundlage des Mesh-Warp-Verfahrens bildet ein Gitter. Dieses Gitter wird aus Kontrollpunkten aufgebaut und wird sowohl für das Eingangs- als auch für das Ausgangsbild erstellt. Solche Kontrollpunkte können bei Gesichtern zum Beispiel Augen, Mund und Nase sein. Das Gitternetz wird manuell in einem graphischen Interface erstellt. Im Zuge dessen werden für jede Masche sowohl vier Punkte im Quellbild als auch vier Punkte im Zielbild spezifiziert, die als Korrespondenzpunkte für diese Masche gelten, so dass nun aus dem Verhalten der Korrespondenzpunkte das Verhalten der gesamten Masche berechnet werden kann. Diese Gitter werden einander durch Interpolation angenähert. [Wang et al. 2009]

Bei [Liu et al. 2006] und [Liu et al. 2009] werden zusätzliche Komponenten für die Saliency Map beim Anpassen von Videos auf unterschiedliche Bildschirmgrößen betrachtet. Der Bewegungskontrast beschreibt dabei die Bewegung jedes Pixels. Die Bewegung der Pixel wird von der Hintergrundbewegung subtrahiert, um wichtige Objekte zu identifizieren. Durch die Berechnung des Optical Flows ist leicht zu erkennen, welche Objekte sich bewegen. Es wird angenommen, dass Objekte, die sich bewegen, auch wichtig sind. [Harman et al. 2002] haben im Gegensatz dazu einen MLA (Machine Learning Algorithmus) entwickelt, der darauf trainiert wird, die Saliency Map zu errechnen. Der MLA berechnet die Saliency Map für Schlüsselframes (Frames mit wichtigen Eigenschaften, Farbänderungen oder Änderungen von Objektpositionen) im Video. Die übrigen Saliency Maps zwischen diesen Frames werden von der MLA interpoliert.

## **2.2 Einsatz von Saliency Maps bei der Segmentierung von 2D Inhalten**

[Shokoufandeh et al. 1998] erstellen neue Ansichten (Novel-Views) für die Repräsentation von 3D Objekten. Die Saliency Map wird dabei für die Objektsegmentierung generiert. Dazu wird zuerst eine Wavelet Transformation auf das Bild angewendet um das Zeit-Frequenz Spektrum zu erhalten, wo weiters die lokalen Maxima gesucht werden. Lokale Maxima weisen auf das Auftreten wichtiger Objekte hin. Jede dieser Regionen, an der sich ein wichtiges Objekt befindet, wird in der Saliency Map gespeichert.

[Zhang et al. 2009] beschreiben in ihrem Paper eine Objektsegmentierungsmethode für die Bildverarbeitung. Modifiziertes LASW (Locally Adaptive Support Weight) wird für die Erkennung der Saliency Map verwendet. Bei der traditionellen LASW Methode werden die Gewichte der betrachteten Pixel durch die geometrischen Beziehungen der Pixel zu einander berechnet. Die Disparitäten werden dann durch die WTA-Strategie (Winner-Takes-



All) festgelegt. Diese Methode ist aber nicht sehr genau und dadurch fehleranfällig bei der Segmentierung. Daher empfehlen [Zhang et al. 2009] im ersten Schritt der LASW, die Anzahl der fehleranfälligen Regionen zu vermindern, um die Details zu behalten.

### **2.3 Einsatz von Saliency Maps bei der Modellierung von 3D Inhalten**

[Longhurst et al. 2006] bieten einen Ansatz für die Erstellung von qualitativ hochwertigen 3D-Bildern für die Bildverarbeitung. Für die Saliency Map kommt die Methode von [Itti and Koch 2000] zum Einsatz, die auf der Erkennung von Objektkanten und abrupten Farbänderungen beruht. Diese wird um die Methode von [Yee et al. 2001] erweitert. [Yee et al. 2001] verwenden den Algorithmus von [Itti et al. 2000], fügen aber Bewegung als zusätzliches Merkmal für die Saliency Map dazu (**Abbildung 8**).

Ausgehend von den drei Kanälen rot vs. grün, blau vs. gelb und Helligkeit (dunkel vs. hell) und einer Kantenrepräsentation des Bildes (ersetzt den Richtungskanal bei [Itti and Koch 2000]) wird die Saliency Map berechnet. Die Unterschiede der Kanäle werden wie bei [Itti and Koch 2000] mit einer Gauß'schen Pyramide errechnet. Zusätzlich werden die Faktoren Bewegung, Tiefe und Gewöhnung einbezogen.

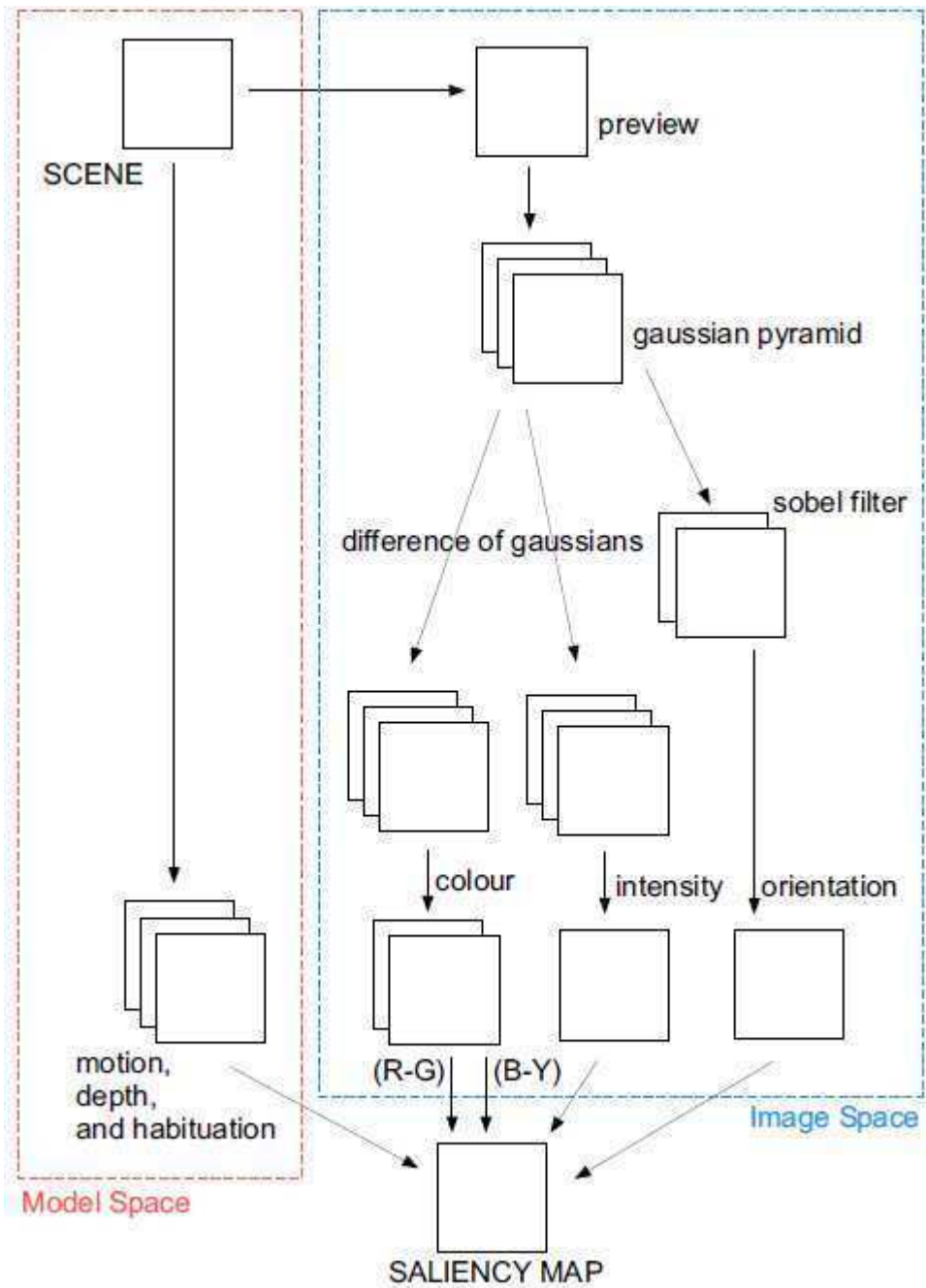


Abbildung 8: Modell zu Erstellung einer Saliency Map nach [Longhurst et al. 2006]: Im Bildraum beeinflussen Kanten, Farbe, Intensität und Orientierung die Saliency Map. Weiters werden Bewegung, Tiefe und Gleichheit aus dem Modellraum als zusätzliche Einflussfaktoren der Saliency Map herangezogen.

### 3 Systembeschreibung

[Lang et al. 2010] beziehen sich in ihrem Paper bezüglich Warping auf nichtlineares Disparitätswarping von [Shamir et al 2009], um ein optimales Ergebnis zu erzielen. Das bedeutet, dass nicht für jedes Pixel gleich gewarpt wird, sondern, dass je nach Wichtigkeit eines Pixels, andere Parameter für das Warping verwendet werden. Die Wichtigkeit ergibt sich aus der Saliency Map. Die wichtigen Pixel sollen soweit wie möglich beibehalten werden, während die niedrig gewichteten Pixel homogen verändert werden können. Objekte (Umrisse) sollen damit bestehen bleiben und nicht zerrissen werden.

#### 3.1 Berechnung der Saliency Map

Die Saliency Map von [Lang et al. 2010] setzt sich folgendermaßen zusammen (siehe **Abbildung 9**):

- Lokale Kantenkarte
- Globale Texturkarte
- Disparitätsbasierende Karte

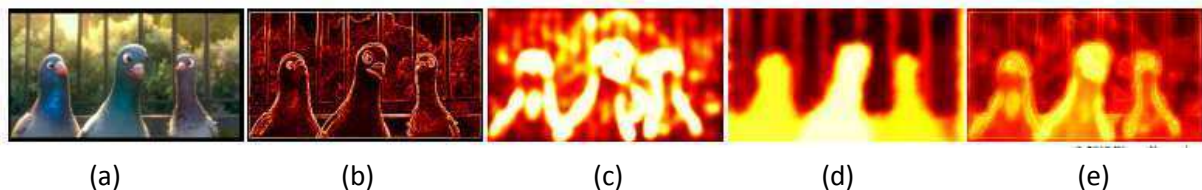


Abbildung 9: Saliency Measures nach [Lang et al. 2010] (a) originales linkes Bild einer Stereoaufnahme, (b) lokale Kantenkarte, (c) globale Texturkarte, (d) Disparitätskarte, (e) Kombinierte Saliency Map aus (b)-(d); © 2010 Disney Enterprises [Lang et al. 2010]

##### 3.1.1 Lokale Kantenkarte

Als lokale Kantenkarte wurde für die Testsets der Canny Kantendetektor verwendet. Dieser Kantenoperator ist eine Standardmethode zur Kantenfindung in Bildern und wird in der Bildverarbeitung häufig angewendet [Shamir et al 2009]. Einzelne Kanten sollen dabei nicht doppelt vorkommen und möglichst wenige Fehler sollen im Bild auftreten. Fehler entstehen durch fehlerhaftes Erkennen von Objekten im Bild. Das Ergebnis der Anwendung des Canny Kantendetektor auf die Testbilder ist in **Abbildung 10** zu sehen. Die hellen Pixel stellen hier die erkannten Kanten dar.



(a)

(b)

Abbildung 10: Lokale Kantenkarte mit Canny Kantendetektor auf a) Aloe aus dem Middlebury Datenset [Middlebury 2006] b) Usertestbild3. Die hellen Pixel kennzeichnen die gefundenen Kanten.

### 3.1.2 Globale Texturkarte

Um die globale Texturkarte zu berechnen, wurde wie bei [Lang et al. 2010] beschrieben, die Methode von [Guo et al. 2008] herangezogen. Dabei wird die Saliency Map durch eine Quaternion Repräsentation dargestellt. Berechnet wird diese durch das Phasenspektrum einer Quaternion Fourier Transformation (QFT), was eine Verbesserung des Phasenspektrums einer diskreten Fourier Transformation (DFT) darstellt. Eine DFT bildet die Zerlegung einer zeitbezogenen Schwingung auf ein diskretes periodisches Frequenzspektrum ab, welches auch als Bildbereich bezeichnet wird [Sonka et al. 1999].

Repräsentation eines Quaternion Bildes:  $q(t) = f_1(t) + f_2(t)\mu_2$  (1)

$f_1(t) = M(t) + RG(t)\mu_1$  (2)

$f_2(t) = BY(t) + I(t)\mu_1$  (3)

$\mu_i, i = 1, 2, 3$  und erfüllt  $\mu_i^2 = -1, \mu_1 \perp \mu_2, \mu_2 \perp \mu_3, \mu_1 \perp \mu_3, \mu_3 = \mu_1\mu_2$ .  $RG(t)$  und  $BY(t)$  stellen zusammengefasste Farbkanäle dar (rot/grün, blau/gelb).  $M(t)$  gibt die Bewegung im Bild an und  $I(t)$  entspricht der Intensität eines Bildes.

Die QFT eines Quaternion Bildes  $q(n, m, t)$  kann folgendermaßen angeschrieben werden:

$$Q[u, v] = F_1[u, v] + F_2[u, v]\mu_2 \quad (4)$$

$$F_i[u, v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\mu_1 2\pi((mv/M)+(nu/N))} f_i(n, m) \quad (5)$$

$(n, m)$  und  $(u, v)$  geben die entsprechende Position jedes Pixels im Zeit- und Frequenzbereich an.  $M$  und  $N$  stehen für die Höhe und Breite des Bildes.  $f_i$ ,  $i \in \{1, 2\}$  werden wie oben berechnet.

$Q(t)$  in Polarform:

$$Q(t) = ||Q(t)||e^{\mu\Phi(t)} \quad (6)$$

$\Phi(t)$  ist das Phasenspektrum von  $Q(t)$  und  $\mu$  ist eine Quaternion. Setzt man  $||Q(t)|| = 1$  dann bleibt  $Q(t)$  nur noch das Phasenspektrum im Frequenzbereich. Aus Formel (5) kann  $Q(t)$  als  $q(t)$  rekonstruiert werden:

$$q(t) = a(t) + b(t)\mu_1 + c(t)\mu_2 + d(t)\mu_3 \quad (7)$$

Daraus ergibt sich die spatio temporale Saliency Map:

$$sM(t) = g * ||q'(t)||^2 \quad (8)$$

wobei  $g$  einen 2D Gaussfilter darstellt mit Standardabweichung  $\sigma = 8$ . [Guo et al. 2008] **Abbildung 11** zeigt die mittels QFT berechnete Saliency Map des verwendeten Testbildes.

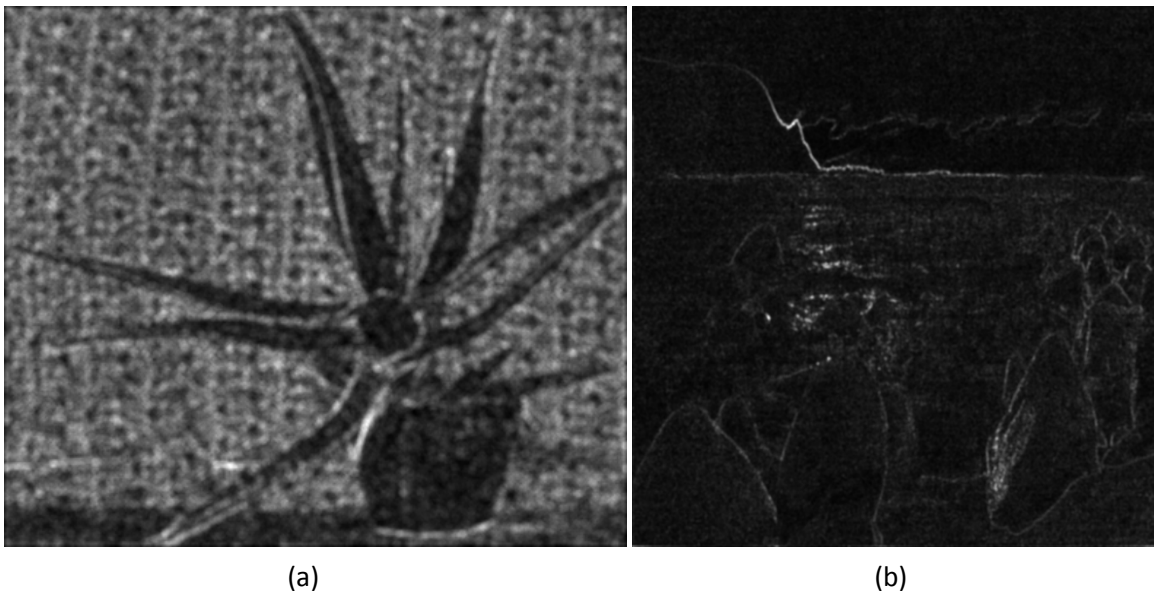


Abbildung 11: Globale Texturkarte auf a) Aloe aus dem Middlebury Datenset [Middlebury 2006] b) Usertestbild3. Die hellen Pixel kennzeichnen die gefundenen Texturen.

### 3.1.3 Disparitätsbasierende Karte

Für die Saliency Map wird eine Disparitätskarte verwendet, welche nach dem Stereo Matching Algorithmus nach [Rhemann et al. 2011] berechnet wurde. Stereo Matching ist der Prozess des Findens korrespondierender Punkte zwischen zwei versetzten Bildern der gleichen Szene. Das Auge des Menschen macht das, ohne auch nur darüber nachzudenken. Wenn wir ein Objekt betrachten, sehen wir davon mit unserem linken und rechten Auge zwei Bilder aus unterschiedlichen Blickwinkeln. Wir erkennen allerdings, dass es sich um das gleiche Objekt handelt und nutzen die Unterschiede, um die Tiefe abzuschätzen. Stereo Matching Algorithmen definieren also die Tiefe für jedes Pixel in korrespondierenden Stereoaufnahmen.

Im Paper von [Rhemann et al. 2011] wird ein Suchalgorithmus präsentiert, welcher eine qualitativ hochwertige und schnelle Lösung für Stereo Matching, basierend auf Guided Filtering [He et al. 2010], bietet. Für das Eingangsbild wird ein Kostendiagramm erstellt, in welchem für jede Pixelzeile die Farbwerte betrachtet werden. Die Pixelzeilen werden mit einem Guided Filter, welcher ein Bild nach Wahl (kann auch das Eingabebild sein) als Referenz heranzieht, geglättet. Dadurch bilden sich durch die Ähnlichkeit der gefilterten Werte eindeutige zusammengehörige Pixelgruppen heraus. Diese werden mit einem entsprechenden Label versehen. Das Ergebnis der Anwendung des Frameworks auf unser Testbild ist in **Abbildung 12** zu sehen.

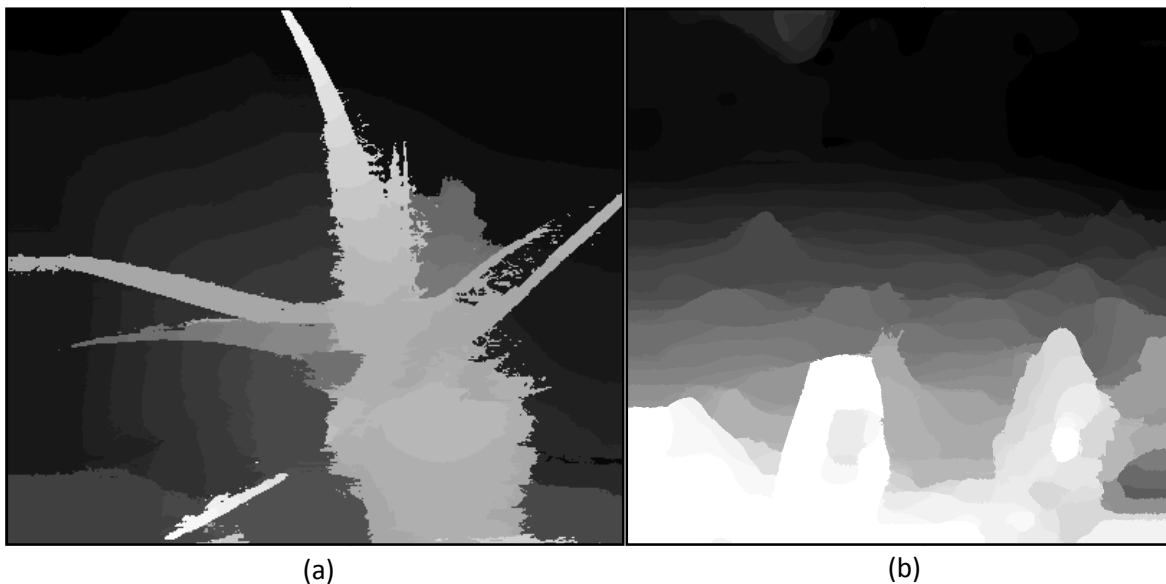


Abbildung 12: Generierte Disparitätskarte auf a) Aloe aus dem Middlebury Datenset [Middlebury 2006] b) Disparitätskarte für User-Testbild3 nach der Methode von [Rhemann et al. 2011]. Die unterschiedlichen Grautöne kennzeichnen die gefundenen Disparitäten.

Zum Vergleich wurde eine zweite Disparitätskarte nach der Methode von [Bleyer et al. 2008] erstellt. [Bleyer et al. 2008] verwenden für das Stereo Matching die „Simple Tree Method“. Dabei wird für jedes Pixel individuell die Disparität berechnet, indem ausgehend vom betrachteten Pixel ein Baum aufgebaut wird, in welchem die Pixel aus dem Referenzbild gespeichert sind. Anschließend wird mittels „dynamic programming“ der optimale Pfad gesucht und damit die Disparität für dieses Pixel festgelegt. Diese Methode liefert exaktere Ergebnisse, weil die Disparitätserkennung nicht auf Bildsegmentierung beruht, sondern für jedes Pixel einzeln berechnet wird (siehe Ergebnis in **Abbildung 13**).

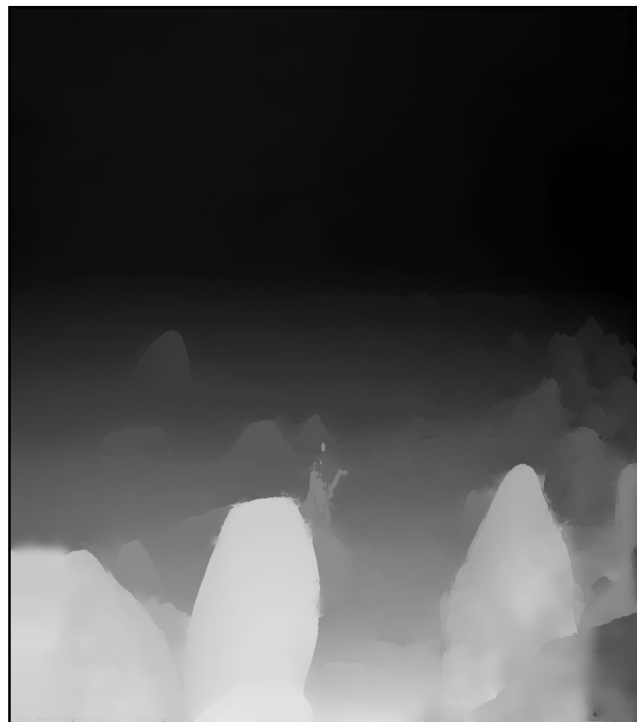


Abbildung 13: Disparitätskarte für Usertestbild3 nach der Methode von [Bleyer et al. 2008]. Die unterschiedlichen Grautöne kennzeichnen die gefundenen Disparitäten.

### 3.1.4 Saliencyberechnung

Die Gewichtung für die einzelnen Bestandteile der Saliency Map ist nach [Lang et al. 2010] für alle Pixel  $x \in$  Eingabebild folgendermaßen vorgesehen:

$$S_i(x) = \frac{S_{lokal}(x) + S_{global}(x)}{2} \quad (9)$$

$S_i$  repräsentiert die bildbasierende Saliency und setzt sich zum gleichen Prozentsatz aus der lokalen Kantenkarte und der globalen Texturkarte zusammen.

$$S(x) = \lambda S_i(x) + (1 - \lambda)S_d(x) \quad (10)$$

$S_d(x)$  gibt die auf Disparität basierende Saliency an. Als Gewichtung  $\lambda$  für  $S_i$  und  $S_d$  wird für die Usertests der Wert  $\lambda = 0,5$  wie bei [Lang et al. 2010] angenommen. Neben der bei [Lang et al. 2010] beschriebenen Parameter für die Saliency Map wurden weitere Methoden zur Erstellung der bildbasierenden Saliency zu Testzwecken implementiert. Außer dem Canny Kantendetektor wurden auch verschiedene andere Kantenoperatoren getestet. Sobel, Prewitt, Laplace und Roberts lieferten allerdings, wie in **Abbildung 14** deutlich zu sehen ist, ein ungenaueres Ergebnis, weshalb diese Methoden nicht weiter betrachtet wurden. Auch der Harris Eckenerkennung, der markante Punkte im Bild finden soll, brachte kein eindeutigeres Ergebnis als der Canny Kantendetektor.

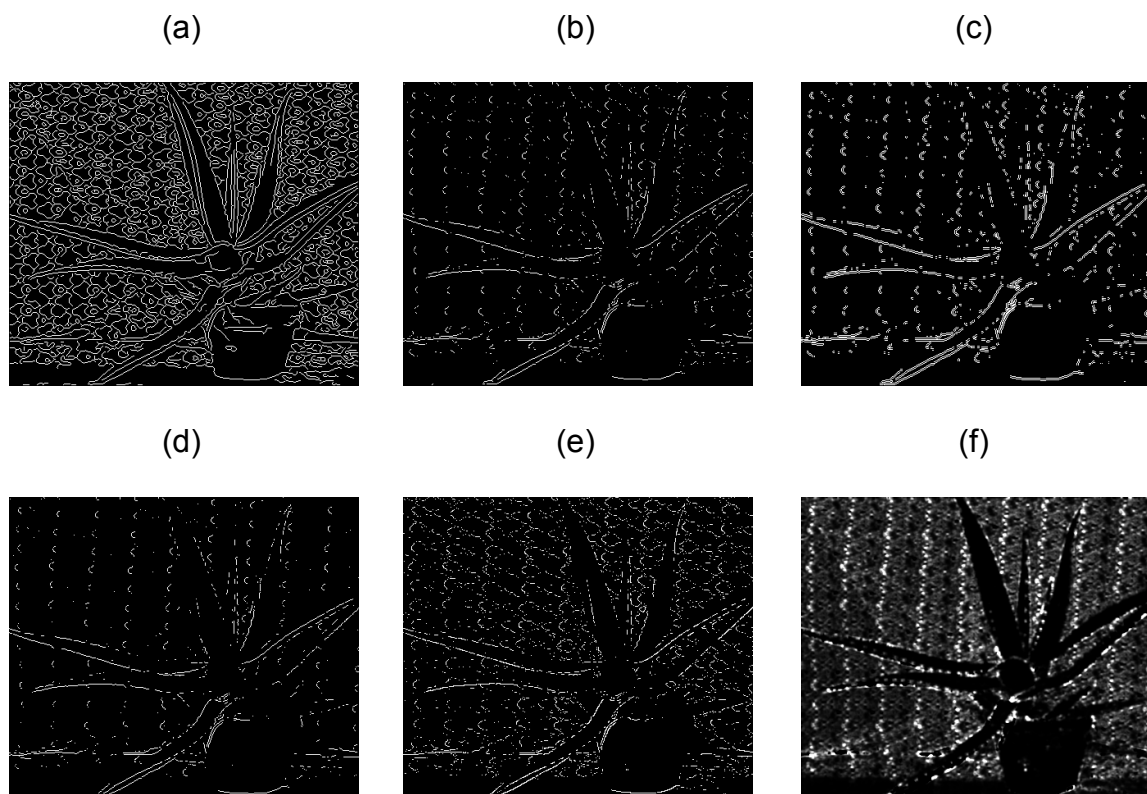


Abbildung 14: Kantenerkennung auf Bildern aus dem Middlebury Datenset [Middlebury 2006]. (a) Canny (b) Sobel (c) Laplace (d) Prewitt (e) Roberts (f) Harris Eckenerkennung.

[Ren et al. 2010] beschreiben in ihrer Methode zur Generierung einer Saliency Map die Verwendung eines Mean-Shift Algorithmus. Diese Methode zur Erkennung von Superpixel wurde implementiert und die Werte in der Disparitätskarte dem Ergebnis entsprechend gewichtet. Daraus ergibt sich, welche erkannten Objekte wie tief (weit entfernt vom Betrachter) liegen. Die Objekterkennung liefert gute und detaillierte Ergebnisse. Aufgrund dieser Detailliertheit kommt es allerdings bei der Gewichtung nach der Tiefe zu vielen Fehlern (siehe **Abbildung 15** und **Abbildung 16**).





Abbildung 15: Erkennung von Superpixel für Image Retargeting im Bild nach Anwendung des Mean-Shift Algorithmus.



Abbildung 16: Saliency Map aus Mean-Shift als Segmentierungsalgorithmus und anschließender Gewichtung nach Tiefe der Objekte.

[Liu et al. 2006] und [Krähenbühl et al. 2009] berechnen für die Saliency Map einen Optical Flow, um Bewegungskontrast (Motion Contrast) im Bild anzuzeigen. Die Berechnung des Optical Flow soll dabei die sich bewegenden Objekte anzeigen. Eine Methode von Lucas Kanade zur Ermittlung des Optical Flows wurde implementiert. Diese konnte allerdings nur bei einer Videosequenz mit mehreren Frames sinnvoll eingesetzt werden.

[Krähenbühl et al. 2009] verwenden die bilaterale Filterung zur Erstellung einer Saliency Map. Dieser Algorithmus dient dazu, Störungen im Bild durch Glätten zu verringern, wobei die Objektkanten trotzdem erhalten bleiben [Tomasi et al. 1998]. Eine Implementierung von Douglas Lanman (2006) dieses Algorithmus wurde im Framework integriert und auf Testbilder angewendet (siehe **Abbildung 17**). Das Ergebnis ist, in Kombination mit anderen Methoden wie Mean-Shift, für eine Saliency Map gut anwendbar, wurde aber für die Benutzerstudie nicht verwendet, da die Methode von [Lang et al. 2010] unverändert evaluiert werden sollte.



Abbildung 17: Mit einem bilateralen Filter generierte Saliency Map

### 3.2 Warping

Beim Warping nach [Lang et al. 2010] wird von einem Stereobildpaar ausgegangen. Von der ursprünglichen Position dieses Stereobildpaares wird jeweils in die Mitte des Bildpaares gewarpt, indem die halbe Disparität aus der Tiefenkarte addiert bzw. subtrahiert wird. Von dort aus sollen die Bilder wieder an eine Position für das linke und rechte Auge gewarpt werden, allerdings um den durch die Saliency Map veränderten Wert (siehe **Abbildung 18**).

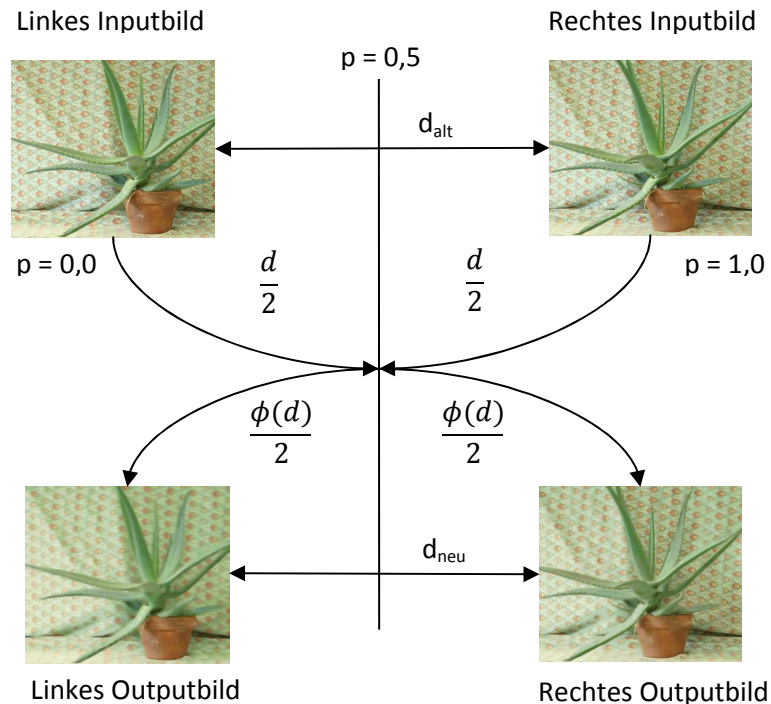


Abbildung 18: Stereoskopisches Warping: Es wird um die halbe Disparität  $\frac{d}{2}$  an Position  $p = 0,5$  gewarpt. Anschließend wird die Disparität  $d$  durch Operatoren verändert ( $\phi(d)$ ) und das linke und rechte Stereobild werden durch Addition/Subtraktion von  $\frac{\phi(d)}{2}$  an die neue Position gewarpt. (Stereobilder übernommen von [Middlebury 2006].)

Die Tiefenwerte sind in der Tiefenkarte und in der Saliency Map gespeichert. Weiters werden Operatoren für das Disparitätswarping definiert:

**Lineare Operatoren** ermöglichen eine lineare Adaption des Disparitätsbereichs  $\Omega$  für die Disparität  $d \in \Omega = [d_{min}, d_{max}]$  zu einem Zielbereich  $\Omega' = [d'_{min}, d'_{max}]$ . Daraus ergibt sich folgende Funktion:

$$\phi_l(d) = \frac{d'_{max} - d'_{min}}{d_{max} - d_{min}} (d - d_{min}) + d'_{min} \quad (11)$$

Durch Ändern des Intervalls  $\Omega$  kann der Disparitätsbereich verändert und damit neue Bilder mit Tiefenwerten im Bereich  $\Omega'$  erstellt werden. Durch diese Adaptierung können Inhalte auf unterschiedlich große Bildschirme angepasst werden.

**Nichtlineare Operatoren** bewirken eine Skalierung der Disparität. Global erfolgt dies durch folgende Funktion mit gewünschtem Skalierungsfaktor  $s$ .

$$\phi_n(d) = \log(1 + sd) \quad (12)$$

Die Disparitäten können bei Bedarf auch lokal mit individuellen Skalierungsfaktoren, beispielsweise basierend auf dem Histogramm der Disparitätswerte oder einer Saliency Map, skaliert werden.

$$\phi_n(d) = \int_0^d S(x) dx \quad (13)$$

Durch ein Warpen der Disparitätsgradienten ist zusätzlich ein Verschieben im Bildraum möglich. Dazu wird der **Gradient Domain Operator** eingesetzt. Folgende Funktion wird dabei von [Lang et al. 2010] verwendet:

$$\phi_\nabla(\nabla d(x), S(x)) = S(x)\phi_l(\nabla d(x)) + (1 - S(x))\phi_n(\nabla d(x)) \quad (14)$$

$S(x)$  stellt dabei die errechnete Saliency Map dar.

Um glatte Szenenübergänge oder visuelle Effekte in Videos zu erzeugen, kann auch eine zeitliche Adaption mit **zeitlichen Operatoren** vorgenommen werden.

$$\phi_t(d, t) = \sum_i w_i(t) \phi_i(d) \quad (15)$$

Zwei oder mehrere Frames können mit jeweils geeignetem Gewichtungsfaktor  $w_i(t)$ , wobei  $t$  die Zeit angibt, den gewünschten Szenenübergang erzeugen. [Lang et al. 2010]

Durch zusätzliche Pixelinformationen aus der Saliency Map, sowie durch die Beeinflussung der linearen (Tiefenbereich), nichtlinearen (Skalierung) und zeitlichen Operatoren (um weichere Szenenübergänge zu schaffen), soll laut [Lang et al. 2010] sichergestellt werden, dass im Ergebnis weniger Artefakte auftreten, sowie ein verbesserter Tiefeneindruck geschaffen wird.

Artefakte sind Fehler im 3D-Bild, meistens an Objektkanten, welche durch das Warping auf neue Positionen entstehen. Für neue Positionen (besonders bei externen Kamerapositionen) liegen keine Pixelinformationen für Stellen vor, die im Ausgangsbild von Objekten verdeckt wurden. Wurde zusätzlich ein Objekt im Bild bei der Generierung der Disparitätskarte schlecht erkannt, können Teile von

Objekten zum Hintergrund gezählt werden, wodurch sich dieser nicht erkannte Teil nach dem Warping an einer falschen Position befindet. Durch Warping entstehen außerdem Löcher, wenn für zuvor verdeckte Bereiche keine Farbwerte vorhanden sind. Da den betroffenen Pixel im Ergebnis keinen Farbwert zugewiesen werden kann, werden diese schwarz dargestellt [Bertalmio et al. 2000]. Mittels Inpainting Methoden können diese Löcher gefüllt werden. Inpainting wird in Kapitel 3.4 näher erläutert.

### **3.3 Fusion des linken und rechten Stereobildes**

Um die Anzahl der Löcher im generierten Bild nach dem Warpen zu dezimieren, kann vor dem Inpainting eine Fusion erfolgen. Dabei wird, wie in **Abbildung 19** gezeigt, das rechte Bild an die gleiche Position gewarpt wie das linke Bild und umgekehrt. Aus dem daraus entstandenen Bild können Farbinformationen für die Löcher im ursprünglich gewarpten Bild übernommen werden. [Raskar et al. 2004]

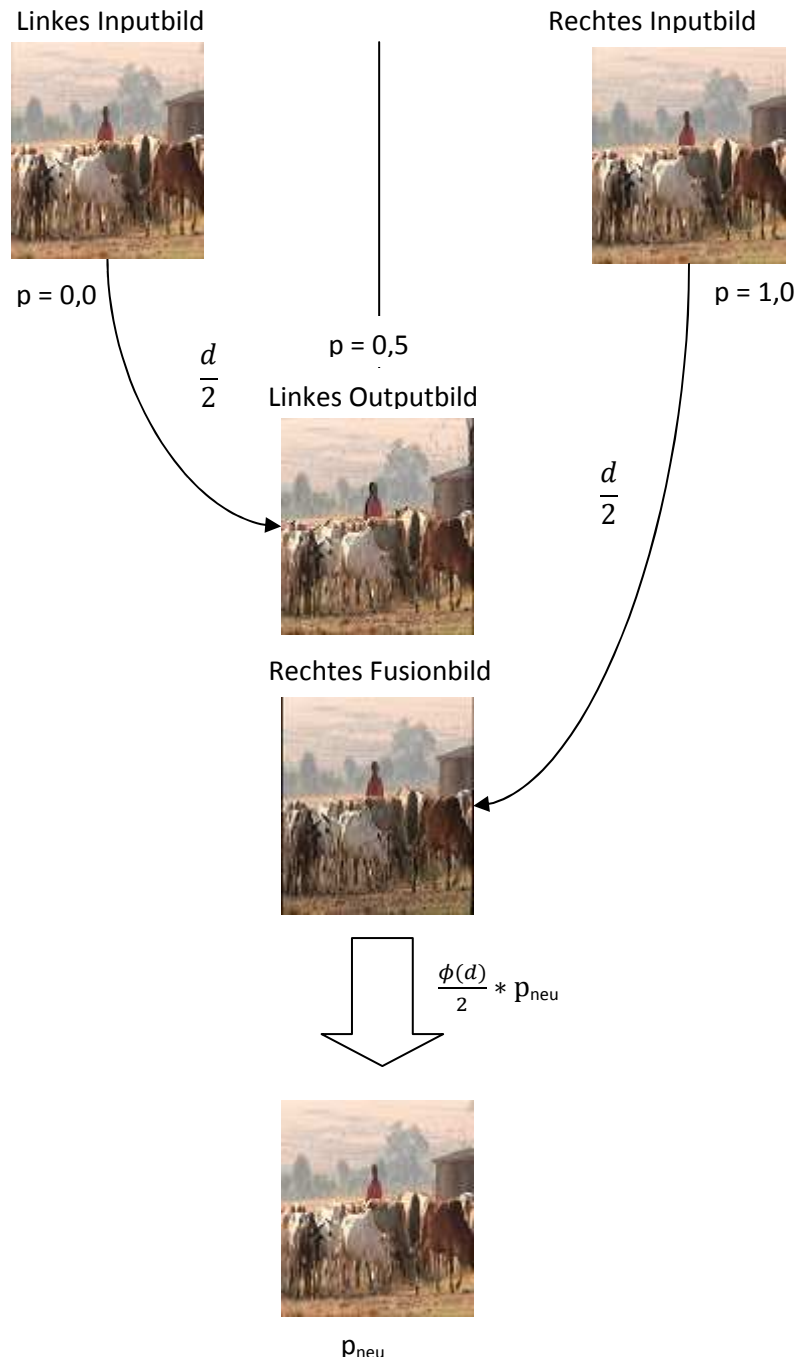


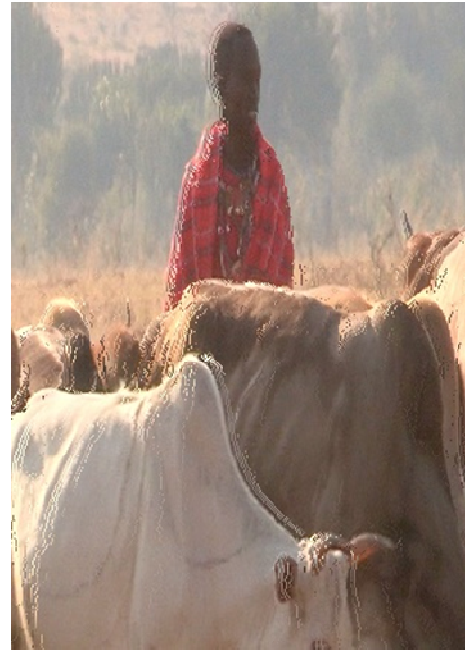
Abbildung 19: Fusion - linkes und rechtes Stereobild werden an die gleiche Position (hier  $p = 0,5$ ) gewarpt und die L cher durch Werte des jeweils anderen Bildes erg nzt. Erst dann wird an die Zielposition gewarpt [Bildmaterial bereitgestellt von 3D Safari LLC Production].

Diese Vorbearbeitung f hrte allerdings bei der Generierung von externen Ansichten zu vielen Artefakten in den Testbildern. In **Abbildung 20 (a)** ist das Ergebnis eines Bildes zu sehen, welches durch stereoskopisches Warping von Position  $0,0$  auf Position  $-2,0$  gewarpt wurde. Die entstandenen L cher wurden zuerst durch Fusion der gewarpten Ansicht von Position  $1,0$  auf  $-2,0$  gef llt. Die  brigen schwarzen Stellen wurden durch Inpainting (Erkl rung in Kapitel 3.4) beseitigt. Wie man in dem vergr o erten Bildausschnitt in **Abbildung 20 (b)** sieht,

sind in den Objekten selbst eindeutige Artefakte in Form von unpassenden Farbwerten zu sehen. Besonders an den Rändern der Objekte, wo durch Warping am häufigsten leere Stellen entstehen, ist die Füllung mit falschen Farbwerten deutlich zu erkennen.



(a)



(b)

Abbildung 20: Löcher im Bild wurden mit Fusion und Inpainting gefüllt. (a) zeigt das komplette Bild und (b) einen vergrößerten Ausschnitt. Hier sind durch Fusion und anschließendes Inpainting entstandene Artefakte im Bild zu erkennen [Bildmaterial bereitgestellt von 3D Safari LLC Production].

Je weiter die Position, an die ein Bild gewarpt werden soll, von der ursprünglichen Ansicht entfernt ist, desto mehr Löcher entstehen, weil mehr Objekte dabei zum Vorschein kommen, die in der Originalansicht vom Hintergrund verdeckt sind (siehe **Abbildung 4**). Durch die Betrachtung aus einem anderen Winkel überlappen sich die Objektkanten bei unterschiedlich gewarpten Positionen daher nicht mehr. Das macht sich durch ausgefranzte Doppelkanten, wie in der Vergrößerung in **Abbildung 20 (b)** erkennbar, deutlich. Fusion bei internen Ansichten, also bei Ansichten innerhalb der beiden Ausgangspositionen, liefert ein besseres Ergebnis. Da das Hauptaugenmerk bei den Testbildern auf externen Kamerapositionen lag, wurde hierbei auf Fusion verzichtet und nur Inpainting angewendet.

### 3.4 Inpainting

Mit dem veränderten Blickwinkel auf eine neu gewarppte Position können auch mehrere Farbwerte auf ein Pixel fallen, wenn sich der Vordergrund durch den neuen Betrachtungswinkel über den Hintergrund schiebt. Dabei wird jener Wert aus dem Originalbild übernommen, welcher einen niedrigeren Wert in der Tiefenkarte eingetragen hat und damit weiter im Vordergrund ist. [Bertalmio et al. 2000]

Andererseits können Stellen im Hintergrund sichtbar werden, die vorher durch Objekte im Vordergrund verdeckt waren. In der neuen Ansicht haben diese Stellen keine Farbwerte, da kein Disparitätswert auf jene Stellen verweist. Diese Stellen erscheinen als schwarze Löcher im Bild. Durch Füllmethoden (sogenannte Inpainting Methoden) können diese Löcher mit Farbwerten versehen werden. Inpainting verwendet dazu benachbarte Punkte im Bild, welche zum Hintergrund gehören. Die nächsten Farbwerte neben einem Loch werden interpoliert, damit keine Artefakte entstehen [Bertalmio et al. 2000]. **Abbildung 21** zeigt ein Stereobildpaar, welches nach dem Waring Löcher aufweist.

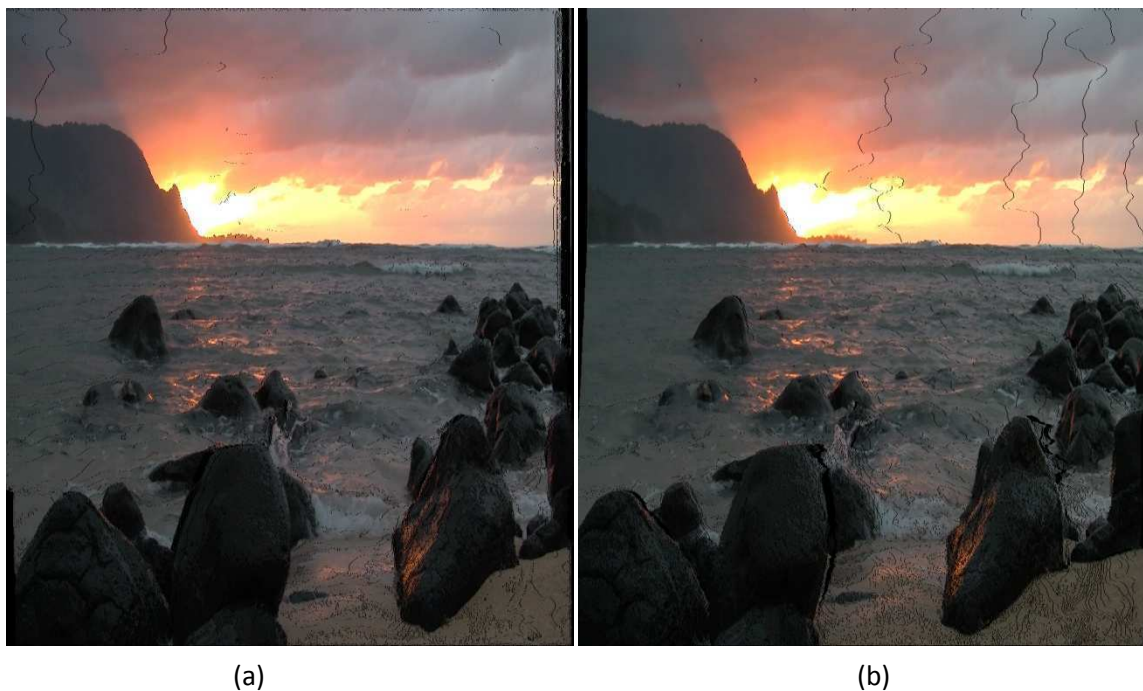


Abbildung 21: Linkes (a) und rechtes (b) Stereobild nach dem Waring ohne Fusion und vor dem Inpainting [Bildmaterial bereitgestellt von 3D Safari LLC Production].



Für die Testbilder wurde eine einfache Inpainting Methode implementiert. Dabei wird das Bild Pixelreihe für Pixelreihe nach Löchern (Fehlen von Farbwerten) gescannt und diese mit den Farbwerten benachbarter Pixel gefüllt. Dabei werden bevorzugt jene benachbarten Pixel mit der höheren Disparität (Hintergrund) herangezogen, da davon ausgegangen wird, dass beim Warping Objekte im Vordergrund verschoben und dadurch der Hintergrund sichtbar wird (siehe **Abbildung 22**).

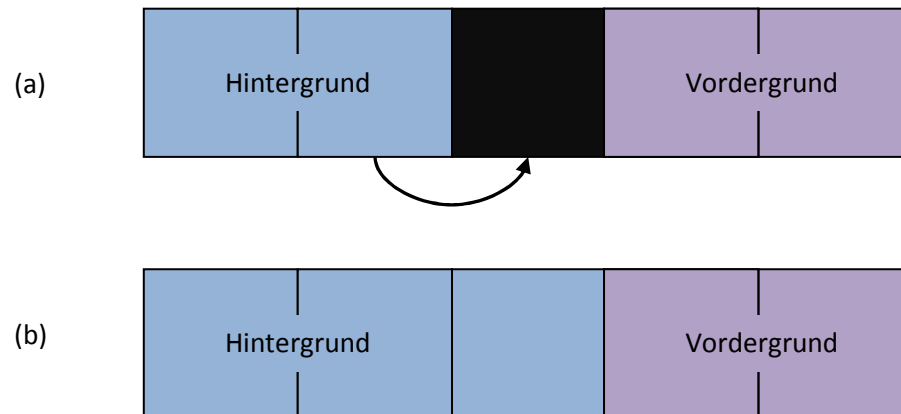


Abbildung 22: Fehlende Farbwerte füllen - Es wird überprüft, ob das links oder rechts vom betrachteten Pixel (schwarz) stehende Pixel zum Hintergrund gehört (höhere Disparität). (a) Links vom Pixel mit dem fehlendes Farbwert befinden sich Pixel, die zum Hintergrund gehören; rechts Pixel des Vordergrunds. Daher wird der Farbwert des Pixels links vom betrachteten Pixel für dieses übernommen (b).

In **Abbildung 23** ist dargestellt, wie der jeweils benachbarte Pixelwert (grün) den Wert des schwarzen Pixels (gelb) ersetzen soll. **Abbildung 24** zeigt das Ergebnis der Farbwerte des gezeigten Bildausschnitts nach diesem Vorgang.

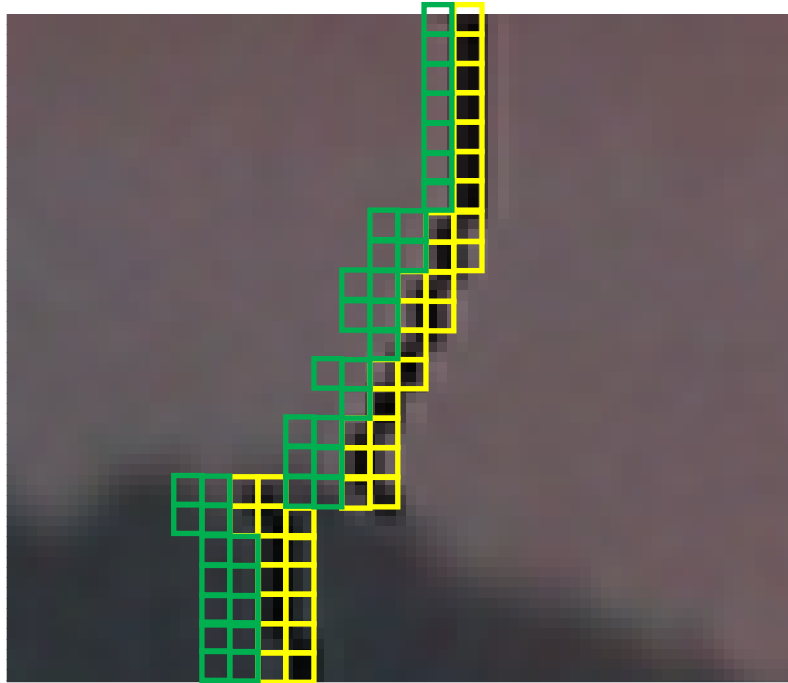


Abbildung 23: Inpainting auf Bildausschnitt. Schrittweise wird jeweils der Farbwert des Pixels (grün), welches links des jeweils ungefüllten Pixels (gelb) liegt, in das ungefüllte Pixel eingesetzt. Sind mehrere Pixel nebeneinander ungefüllt, dann werden die Farbwerte der entsprechenden Anzahl an Pixel von links eingesetzt.

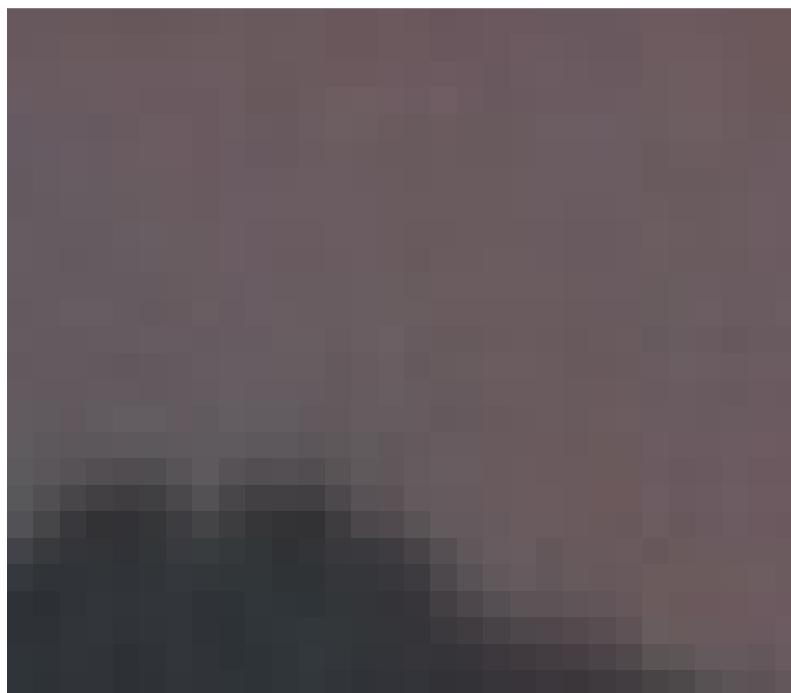


Abbildung 24: Bildausschnitt aus Abbildung 23 nach dem Inpainting.

## 4 Usertests und Ergebnisse

Im folgenden Kapitel wird zunächst das Testsetup und damit die Testbilder mit ihren spezifischen Eigenschaften und deren Ansichten im Kamerasetup, sowie die Generierung der verwendeten Disparitätskarten und die letztliche Darstellung im Usertest beschrieben. Anschließend wird die Validierungsmethode erläutert und die Ergebnisse der Studie im Detail erklärt.

Folgende Fragen sollten durch die Usertests geklärt werden:

- Kann die Qualität der resultierenden 3D-Bilder durch die Verwendung von Saliency Maps verbessert werden?
- Verändert das Warming mit Saliency Map den Tiefeneindruck positiv?

### 4.1 Bilder und Eigenschaften

Um die Testzeiten zu reduzieren, wurden Bilder und keine Videosequenzen herangezogen, da die Betrachtung von Bildern schneller erfolgt. Die Testpersonen können damit innerhalb einer angemessenen Zeit trotzdem sechs Testbilder zu je zwölf Einzelvergleichen vornehmen. Die Testzeit betrug durchschnittlich eine halbe Stunde. Die sechs Testbilder unterschieden sich in ihrem Disparitätsbereich (minimale/maximale Disparität), um auszuschließen, dass das Ergebnis durch eine zu einseitige Betrachtungsweise verfälscht wird. Das Bildmaterial wurde bereitgestellt von 3D Safari LLC Production, einer Partnerschaft zwischen Digital Revolution Studios und Wild Eyes Productions. Bei Bildern, welche von vorneherein einen geringen Disparitätsbereich umfassen, ist es schwieriger, Unterschiede in der Tiefe zu erkennen, als bei jenen, die einen größeren Disparitätsbereich besitzen. Genauso macht es einen Unterschied, ob das Bild im Hintergrund Objekte als Anhaltspunkt hat oder ob sich der Betrachter an einer beinahe ebenen Oberfläche orientieren muss. Weiters spielt es eine wichtige Rolle, ob Objekte hinter der Bildebene anfangen und bis tief in den Hintergrund reichen, oder ob sie bereits vor der Bildebene anfangen und dafür weniger tief reichen. Diese Aspekte wurden mit der Wahl der unterschiedlichen Bilder abgedeckt. In **Abbildung 25** sind die jeweils mit einer Disparitätskarte von niedriger Qualität (siehe Kapitel 4.3) an Position 1,0 gewarpten Ansichten der verwendeten Testbilder abgebildet.



Testbild 1



Testbild 2



Testbild 3



Testbild 4



Testbild 5



Testbild 6

Abbildung 25: Im Usertest verwendete Testbilder. Jeweils linkes Stereobild (Position 1,0) mit Disparitätswerte niedriger Qualität ohne Saliency Map erzeugt [Bildmaterial bereitgestellt von 3D Safari LLC Production].

Testbild 1 enthält sowohl Objekte im Vordergrund als auch im Hintergrund, an welchen sich die Testpersonen bezüglich der Tiefe orientieren können. Die Disparitätswerte, welche sich auf die Bildbreite von 960 Pixel bei allen Testbildern beziehen, liegen zwischen +18 (hinter der Bildebene) und -7 (vor der Bildebene), wodurch der Eindruck entsteht, dass das Bild weiter in den Hintergrund reicht, als aus dem Bildschirm heraus.

Testbild 2 hat einen sehr einheitlichen Hintergrund ohne ausgeprägte Struktur, was zu mehr Problemen bei der Tiefenbeurteilung führt. Die Disparitätswerte liegen zwischen +12 und -11, was ein ausgewogenes Verhältnis darstellt.

In Testbild 3 weist das Wasser durch die Wellen viel Struktur auf. Dadurch und durch die Steine im Vordergrund ist der Tiefeneindruck leicht zu beurteilen. Die Disparitätswerte liegen zwischen +38 und -3. Dadurch kommt dem Betrachter das 3D-Bild sehr tief vor und wurde im Vergleich mit Testbild 2 auch tiefer bewertet.

Testbild 4 kennzeichnen vor allem die Personen im Vordergrund und der Baum, welcher von der Seite ins Bild hineinragt. Für den Tiefeneindruck kann man sich aber auch gut am Graben im Hintergrund orientieren. Die Disparitätswerte liegen zwischen +21 und -19, was ein sehr großes, aber auch ausgeglichenes Verhältnis darstellt.

Testbild 5 weist vor allem im Vordergrund viele Objekte auf, anhand derer man sich bei der Erkennung des Tiefeneindrucks orientieren kann. Im Hintergrund fällt das aber schwerer, weil das Wasser keine Struktur aufweist und als gleichmäßige Fläche wahrgenommen wird. Trotzdem liegen die Disparitätswerte zwischen +39 und -17, was darauf hinweist, dass das 3D-Bild sehr tief ist.

Testbild 6 weist sehr viele Strukturen im Vordergrund auf, welche sich positiv auf die Tiefenwahrnehmung auswirken, und auch anhand der Schattierungen an der Wand im Hintergrund kann Tiefe erkannt werden. In diesem Bild sind die Objekte im Vordergrund für den Tiefeneindruck allerdings wichtiger, weil die Disparitätswerte zwischen +16 und -30 liegen, was darauf hinweist, dass das Bild zum größeren Teil vor der Bildebene liegt und damit zum Betrachter hinragt.

## 4.2 Kamerasetup und generierte Ansichten

Bei [Lang et al. 2010] werden die Positionen 8,0 und 10,0 als Originalpositionen in drei verschiedenen Datasets herangezogen. Daraus wird die Zwischenposition 9,0 interpoliert und die externen Positionen 7,0 und 11,0 extrapoliert. Damit sollte verglichen werden, ob die angewendete Methode sowohl bei der Generierung von Zwischenansichten, als auch von Ansichten für externe Kameras geeignet ist.

Da die Generierung externer Kameraansichten aufgrund häufig fehlender Farbinformationen eine Herausforderung darstellt, wurden für das Testsetup zwei externe Kamerapositionen gewählt. Als Originalposition wird von einer Position 0,0 für das Bild des linken Auges und Position 1,0 für das Bild des rechten Auges ausgegangen. Im ersten Testsetup sollte die neue asymmetrische errechnete Position 0,0 und 2,0 abgebildet werden (siehe **Abbildung 26**).

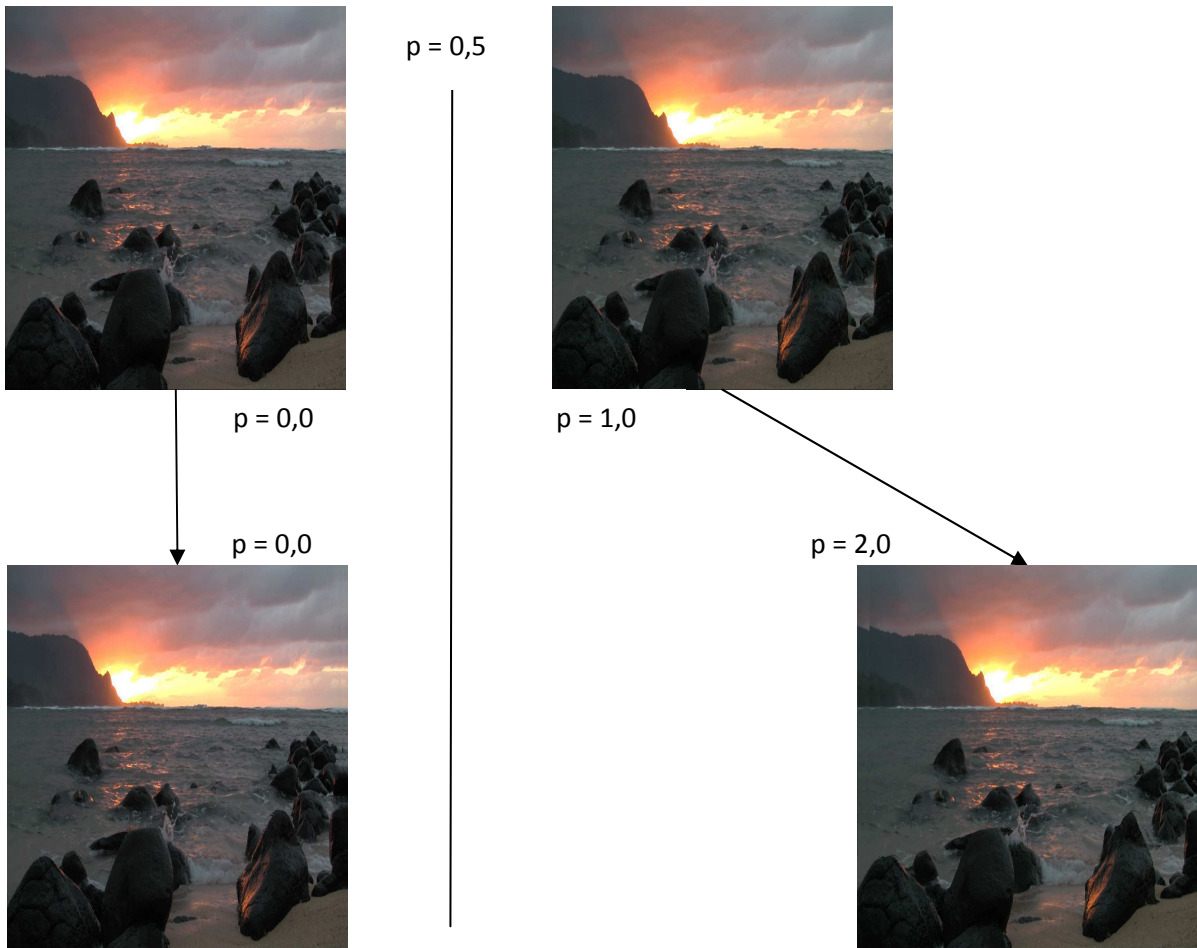


Abbildung 26: Testset 1 mit neu generierten (gewarpten) asymmetrischen Positionen  $p = 0,0$  und  $p = 2,0$ . Aus  $p = 0,0$  wird  $p = 0,0$  gewarpt und aus  $p = 1,0$  wird  $p = 2,0$  gewarpt [Bildmaterial bereitgestellt von 3D Safari LLC Production].

Als zweites Testset wurde eine neue symmetrisch errechnete externe Kameraposition gewählt: Position -0,5 und 1,5 (siehe **Abbildung 27**).

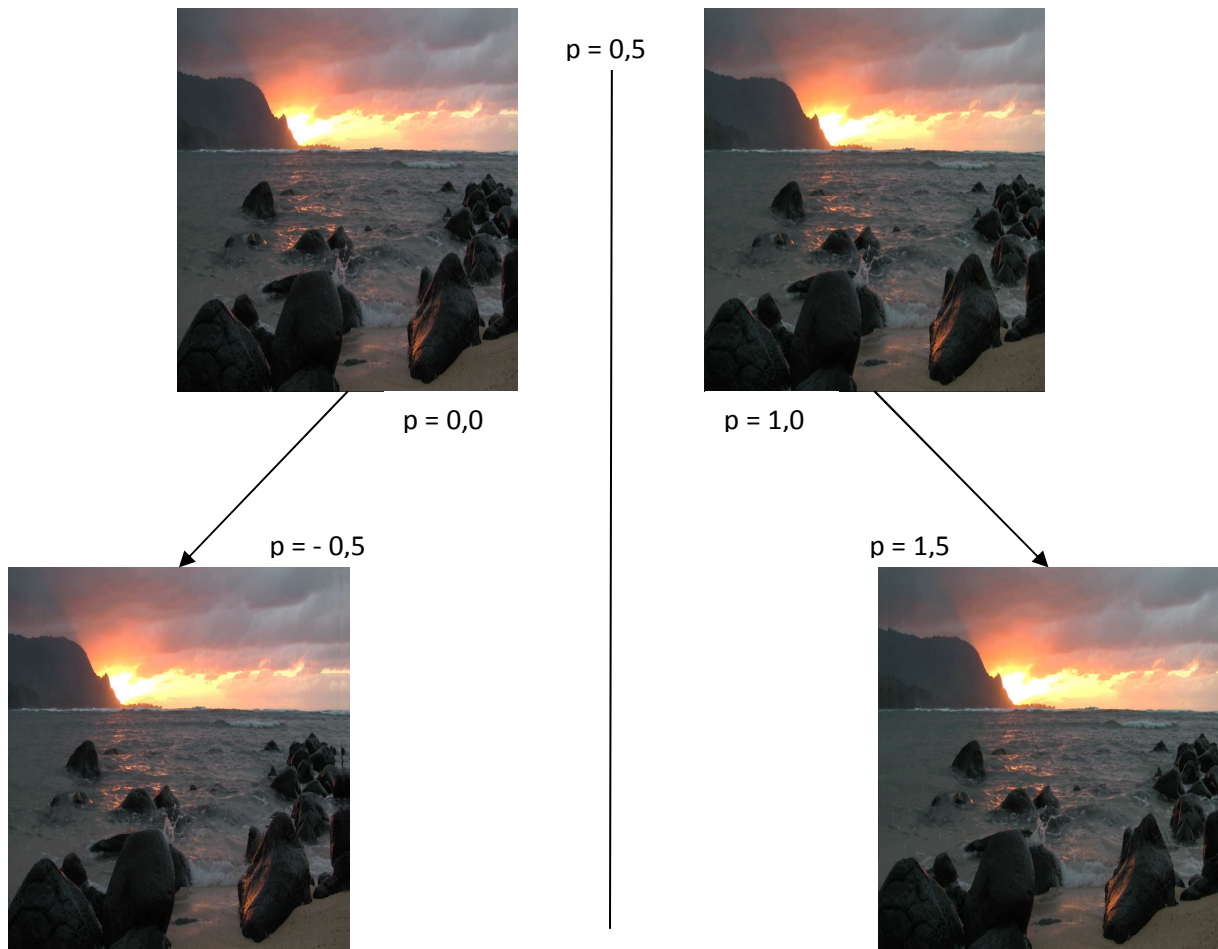


Abbildung 27: Testset 2 mit neu generierten (gewarpten) symmetrischen Positionen  $p = -0,5$  und  $p = 1,5$ . Aus  $p = 0,0$  wird  $p = -0,5$  gewarpt und aus  $p = 1,0$  wird  $p = 1,5$  gewarpt [Bildmaterial bereitgestellt von 3D Safari LLC Production].

Diese beiden generierten Kamerapositionen wurden auf sechs unterschiedliche Testbilder angewendet und 24 Testpersonen (davon sechs Frauen und 18 Männer) im Alter zwischen 22 und 37 Jahren im Zuge eines Usertests gezeigt.

### 4.3 Disparitätskarten

Um die Wirkung auf unterschiedlich genaue Disparitätskarten zu untersuchen, wurde neben der nach [Rhemann et al. 2011], durch schnelles Stereo Matching erzeugten Disparitätskarte (DK NQ, Disparitätskarte mit niedriger Qualität, **Abbildung 12 b**) weiters eine sehr genaue, auf den gleichen Testbildern erstellte Disparitätskarte nach [Bleyer et al. 2008] (DK HQ, Disparitätskarte mit hoher Qualität, **Abbildung 13**) für die Testsets verwendet. Das Warming wurde jeweils für die beiden unterschiedlichen Kamerapositionen und unterschiedlichen Disparitätskarten sowohl mit, als auch ohne Verbesserung durch den Einsatz einer



Saliency Map durchgeführt, wobei vier unterschiedliche 3D-Bilder pro Kameraposition und Testbild entstanden sind. Bei der Berechnung der Operatoren für das Disparitätswarping (siehe Kapitel 3.2) wurden die Werte für den Disparitätsbereich nicht verändert, um die Ergebnisse mit und ohne Einsatz der Saliency Map optimal vergleichen zu können.

#### 4.4 Darstellung und Umsetzung

Die Usertests wurden auf einem 24 Zoll TFT stereoskopischem Bildschirm der Marke Acer GD245HQ mit voller HD Auflösung durchgeführt und mittels NVIDIA 3D Vision Photo Viewer angezeigt. Mit Hilfe einer Active Shutter Brille von NVIDIA wurde das jeweilige Stereobild dreidimensional angezeigt. Active Shutter Brillen verwenden ein Synchronisationssignal vom Bildschirmsystem, damit der LCD vor dem jeweiligen Auge im richtigen Moment schließt, abgestimmt auf den Frame für das linke oder rechte Auge am Bildschirm. Da Active Shutter Brillen sehr anfällig auf künstliches Licht sind, wurden die Usertests nur bei natürlicher Sonneneinstrahlung bzw. ohne künstlich erzeugtes Licht durchgeführt.

Es wurden jeweils zwei generierte 3D-Bilder nebeneinander angezeigt, wobei für jedes der vier erzeugten Bilder Bildvergleiche durchgeführt wurden, damit jedes Bild der jeweiligen Kameraposition miteinander verglichen wird. Die 3D-Bilder wurden aus den generierten linken und rechten Bildern mittels Stereophotomaker erstellt. In **Abbildung 28** sieht man den Vergleich der Stereobilder, welche mit der Disparitätskarte niedriger Qualität ohne Saliency Map und mit der Disparitätskarte niedriger Qualität mit Saliency Map an der asymmetrischen Position 0,0/2,0 erzeugt wurden. Diese Bilddarstellung wird durch NVIDIA 3D Vision Photo Viewer als ein durch Active Shutter Brillen erfassbares 3D-Bild angezeigt.

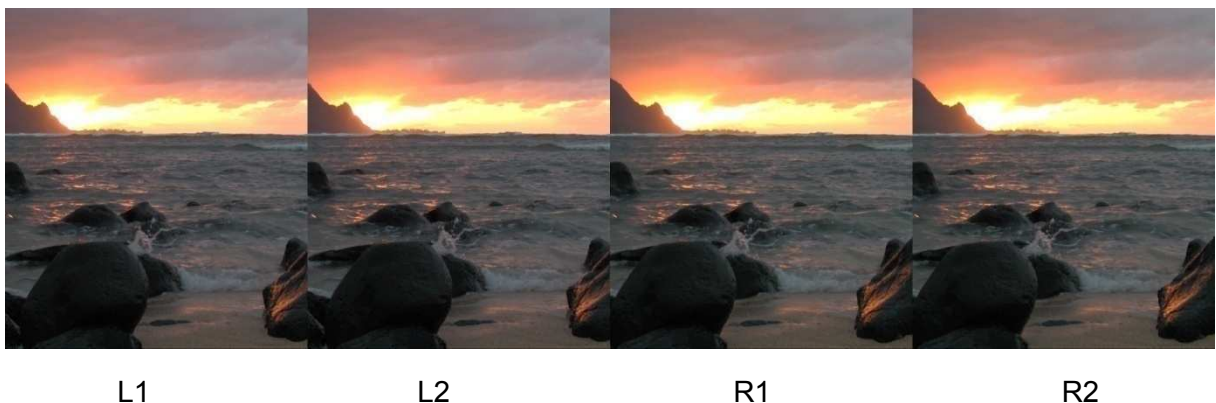


Abbildung 28: Gegenüberstellung Testbild 3 mit DK NQ ohne Saliency Map - Bild mit DK NQ mit Saliency Map. Die jeweils linken (L1, L2) und rechten (R1, R2) Bilder der zu vergleichenden 3D-Bilder liegen nebeneinander in der Generierung der 3D-Bilder in Stereophotomaker [Bildmaterial bereitgestellt von 3D Safari LLC Production].



Bei den Usertests wurden die generierten Ansichten wie in **Tabelle 2** für jedes Testbild gegenübergestellt. Jene Bilder, die mit 1 in der Tabelle vermerkt sind, wurden verglichen. Die Reihenfolge der Gegenüberstellungen und die Position der Bilder links oder rechts in der Gegenüberstellung variierten je nach Testbild, blieben aber bei jeder Testperson gleich.

Position	DK NQ		DK HQ	
	ohne SM (NQ)	mit SM (NQ')	ohne SM (HQ)	mit SM (HQ')
0,0 / 2,0	1	1	0	0
	1	0	1	0
	1	0	0	1
	0	1	1	0
	0	1	0	1
	0	0	1	1
-0,5 / 1,5	1	1	0	0
	1	0	1	0
	1	0	0	1
	0	1	1	0
	0	1	0	1
	0	0	1	1

Tabelle 2: Beispiel für eine User-Test-Vergleichsmatrix (SM = Saliency Map, DK = Disparitätskarte).

Den Testpersonen wurde vor Beginn des Tests ein 3D-Bild gezeigt und es wurde gefragt, ob sie das Bild dreidimensional sehen. Alle 24 Testpersonen waren fähig, Tiefe in einem Stereobild zu erkennen und wurden damit zum Test zugelassen. Zu jedem 3D-Bildpaar wurden den Testpersonen anschließend zwei konkrete Fragen gestellt:

(a) In welchem Bild sind mehr Artefakte zu sehen?

- Links oder Rechts

(b) In welchem Bild ist der Tiefeneindruck höher?

- Links oder Rechts

## 4.5 Validierung

Die Auswertung erfolgte nach dem Thurstone Modell [Woods et al. 2010], wobei zu Beginn zusätzliche Faktoren wie eine Sehschwäche der Testpersonen sowie das Geschlecht mit einbezogen wurden [Rajae-Joordens et al. 2005]. Jedes erstellte Bild wird hier jedem anderen Bild einzeln gegenübergestellt, um herauszufinden, welches entsprechend der jeweiligen Frage den besseren Tiefeneindruck bzw. mehr Artefakte aufweist. **Tabelle 3** zeigt den Aufbau der Thurstone Matrix, aufgrund welcher die Bilder verglichen wurden.

	a	b	c	$Z_i$
a	-	$Z_{a>b}$	$Z_{a>c}$	$Z_a$
b	$Z_{b>a}$	-	$Z_{b>c}$	$Z_b$
c	$Z_{c>a}$	$Z_{c>b}$	-	$Z_c$

Tabelle 3: Komponenten einer Thurstone Matrix [Woods et al. 2010].

Hierbei gilt:

$$Z_i = \sum Z_{i>k} / n$$

Stimmen also eine prozentual höhere Anzahl an Testpersonen bei einer der Fragen für ein Bild, dann wird hier eine 1 in der Regressionsmatrix vermerkt und bei dem gegenübergestellten Bild eine -1. Also wenn  $X_{ij} > X_{ik}$  dann  $X_{ij} = +1$  und  $X_{ik} = -1$  und umgekehrt, wenn  $X_{ij} > X_{ik}$  falsch ist, dann drehen sich die Vorzeichen um:  $X_{ij} = -1$  und  $X_{ik} = +1$ . Alle nicht betrachteten Felder werden mit 0 angenommen [Woods et al. 2010]. Daraus lässt sich ablesen, welches Testobjekt am besten ist.

Bei jeder Testperson wurde - neben den Antworten auf die beiden oben genannten Fragen - das Geschlecht notiert, und ob die Testperson eine Sehhilfe (in Form von Brille oder Kontaktlinsen) trägt, um ausschließen zu können, dass das Ergebnis durch einen dieser Faktoren beeinflusst wird. Für die Auswertung wurde jeweils eine Regressionsmatrix für alle Testbilder und Kamerapositionen erstellt, zuerst getrennt nach Geschlecht, dann nach dem Kriterium, ob die Personen eine Sehhilfe benötigen oder nicht. Wie sich im Zuge der Validierung herausstellte, waren keine signifikanten Unterschiede im Ergebnis zwischen Testpersonen mit und ohne Sehhilfe, sowie männlichen und weiblichen Testpersonen festzustellen. Im weiteren Verlauf dieser Studie wurden diese Faktoren daher nicht mit einbezogen (siehe Anhang I: Auszug aus den Regressionsmatrizen für den zusätzlichen Faktor Sehhilfe bei der Frage nach dem Tiefeneindruck).

Die Testpersonen wurden also in einer Regressionsmatrix für jedes Bild und für jede Kameraposition zusammengefasst. Für die beiden gestellten Fragen wurde jeweils getrennt validiert. Stimmenthaltungen wurden abgerechnet, indem die Anzahl der Testpersonen N um die Anzahl der Stimmenthaltungen dezimiert wurde.

#### 4.5.1 Gesamtauswertung für Frage (a)

**Tabelle 4** und **Tabelle 5** zeigen das Ergebnis der Usertests für Frage (a). Daraus ist erkennbar, dass der Einfluss von Saliency Maps bei unterschiedlich generierten Disparitätskarten die im Bild auftretenden Artefakte nicht vermindert. Für die Berechnung der Saliency Map wurden Kanten- und Texturkarten verwendet. Jede dieser zusätzlichen Segmentierungen im Bild bringt eine weitere Fehleranfälligkeit, in deren Folge Objekte ungenau erkannt werden. Dies führt im Endergebnis zum vermehrten Auftreten von Artefakten. Das Zustandekommen dieser Artefakte bei den einzelnen Testbildern ist in den Einzelergebnissen in 4.5.2 beschrieben.

Ergebnis zu Frage (a): In welchem Bild sind mehr Artefakte zu sehen?

Bild	NQ	NQ'	HQ	HQ'
Summe B1-B6 (gewichtet)	-12,302	7,959	-4,777	10,120
NQ < HQ < NQ' < HQ'				

Tabelle 4: Frage (a) Gesamtergebnis für die Bilder B1 – B6, Position 0,0/2,0; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' entspricht den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' entspricht den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Die Summe ist das Ergebnis der gewichteten Punktevergabe aller Bilder.

Bild	NQ	NQ'	HQ	HQ'
Summe B1-B6 (gewichtet)	-13,054	8,005	-5,802	10,851
NQ < HQ < NQ' < HQ'				

Tabelle 5: Frage (a) Gesamtergebnis für die Bilder B1 – B6, Position -0,5/1,5; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' entspricht den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' entspricht den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Die Summe ist das Ergebnis der gewichteten Punktevergabe aller Bilder.

Bilder, die mit der DK NQ generiert wurden, weisen weniger Artefakte auf. Diese Methode erkennt die Disparitäten im Bild besser. Bei der Summe der Bilder der symmetrischen Kameraposition  $-0,5/0,5$  sowie bei der asymmetrischen Kameraposition  $0,0/2,0$ , kommt das gleiche Ergebnis heraus (siehe **Tabelle 6**).

Reihung nach Artefakten (1 = höchste Anzahl)	Methode
1	HQ'
2	NQ'
3	HQ
4	NQ

Tabelle 6: Reihung der Methode, welche die meisten Artefakte hervorruft, bei Kameraposition  $0,0/2,0$  und  $-0,5/0,5$  beginnend mit jener Methode, die die meisten Artefakte hervorruft.

**Abbildung 29** zeigt deutlich, wie viel Prozent der befragten Personen jeweils für welche Methode gestimmt haben. Die Verteilung ist bei symmetrischem und asymmetrischem Kamerasetup annähernd gleich. Bilder mit symmetrischen Kamerapositionen und Bilder mit asymmetrischen Kamerapositionen wurden nicht direkt miteinander verglichen, da beim asymmetrischen Warming auf eine Position gewarpt wurde, die weiter von der Originalposition entfernt ist (Position  $2,0$ ) als beim symmetrischen Warming (beide Positionen wurden um  $0,5$  verschoben). Beim asymmetrischen Warming sind durch diesen Schritt mehr Artefakte entstanden als beim symmetrischen Warming. Bei Bildern, die auf die gleiche Kameraposition gewarpt wurden, ergab sich für jedes Kamerasetup das gleiche Ergebnis.

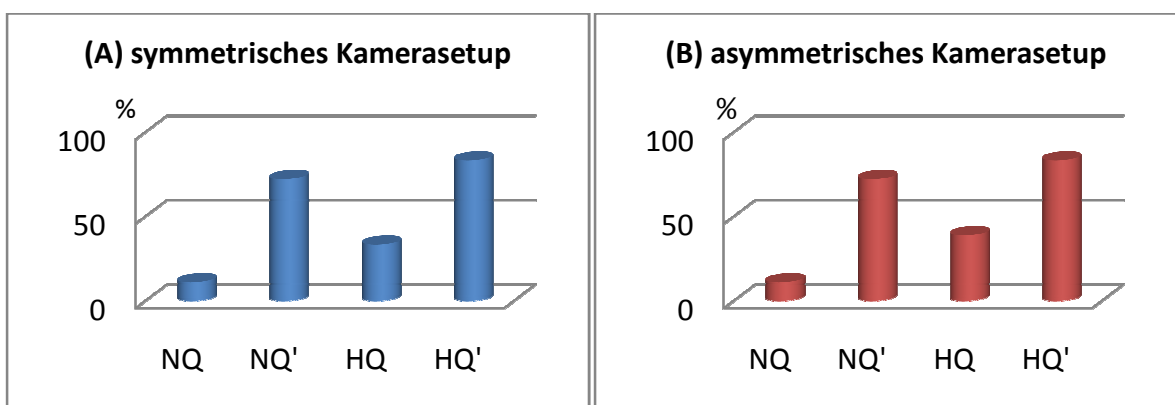


Abbildung 29: Auswertung der Ergebnisse zu Frage (a); (A) zeigt, wie viele der befragten Personen beim symmetrischen Kamerasetup ( $-0,5/1,5$ ) für welche Methode gestimmt haben. (B) zeigt, wie viele der befragten Personen beim asymmetrischen Kamerasetup ( $0,0/2,0$ ) für welche Methode gestimmt haben.

#### 4.5.2 Auswertung der einzelnen Testbilder für Frage (a)

##### Testbild 1



Abbildung 30: Testbild 1, linkes Stereobild, DK NQ ohne Saliency Map, Position 1,0.

Das Ergebnis von Frage (a) für beide Kamerapositionen  $0,0/2,0$  und  $-0,5/0,5$  ergibt sich laut **Tabelle 7** und **Tabelle 8** wie folgt: Methode NQ' weist die meisten Artefakte auf, gefolgt von Methode HQ'. Bei Kameraposition  $0,0/2,0$  stimmten im direkten Vergleich allerdings gleich viele Personen für NQ' und HQ' als jene Methoden mit den meisten Artefakten. Die Methoden ohne Einsatz der Saliency Map weisen bessere Ergebnisse auf. Methode HQ ist auf Platz 2 und Methode NQ auf Platz 1 in der Bewertung der am wenigsten auftretenden Artefakte. Laut Usertests weisen in diesem Fall jene Bilder, welche mit zusätzlichem Einsatz einer Saliency Map gewarpt wurden, bei beiden Kamerapositionen mehr Artefakte auf. Eine Verbesserung ist somit nicht gegeben.

Resultierende Regressionsmatrix für Frage (a):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	-1	1			0,875	NQ'
20	-1		1		0,700	HQ
24	-1			1	0,917	HQ'
20		1	-1		0,950	NQ'
16		1		1	0,500	NQ' / HQ'
22			-1	1	0,773	HQ'
Summe	-3	3	-1	3	NQ < HQ < HQ' < NQ'	
Summe (gewichtet)	-2,492	2,325	-1,023	2,189		

Tabelle 7: Bild 1, Kameraposition 0,0/2,0; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	-1	1			1,000	NQ'
21	-1		1		0,762	HQ
23	-1			1	0,870	HQ'
24		1	-1		0,917	NQ'
20		1		-1	0,850	NQ'
21			-1	1	0,952	HQ'
Summe	-3	3	-1	1	NQ < HQ < HQ' < NQ'	
Summe (gewichtet)	-2,631	2,767	-1,107	0,972		

Tabelle 8: Bild 1, Kameraposition -0,5/1,5; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

**Abbildung 31** und **Abbildung 32** zeigen die unterschiedlichen verwendeten Disparitätskarten. Die DK NQ ohne Saliency Map, welche das beste Ergebnis in Bezug auf das Artefakt-Auftreten erzielt, hat die Objekte am besten erkannt. Auch das Fell der abgebildeten Affen wurde erkannt, was an den ausgefransten Kanten erkennbar ist, und daran, dass die Objekte hier insgesamt breiter sind. Im Gegensatz dazu bringt der Einsatz der Saliency Map auf der DK NQ das schlechteste Ergebnis und damit die meisten auftretenden Artefakte. Das kann damit begründet werden, dass die Kanten vom Kantenoperator ungenau erkannt wurden. Dieser Widerspruch führt zu Artefakten. Die insgesamt unvollständige Erkennung der Objektkanten, die auch durch eine unvollständige Erkennung in der Saliency Map nicht verbessert wurde, führt ebenfalls zu Artefakten.



Abbildung 31: Bild 1, linke DK HQ; links: ohne Saliency Map, rechts: mit Saliency Map.

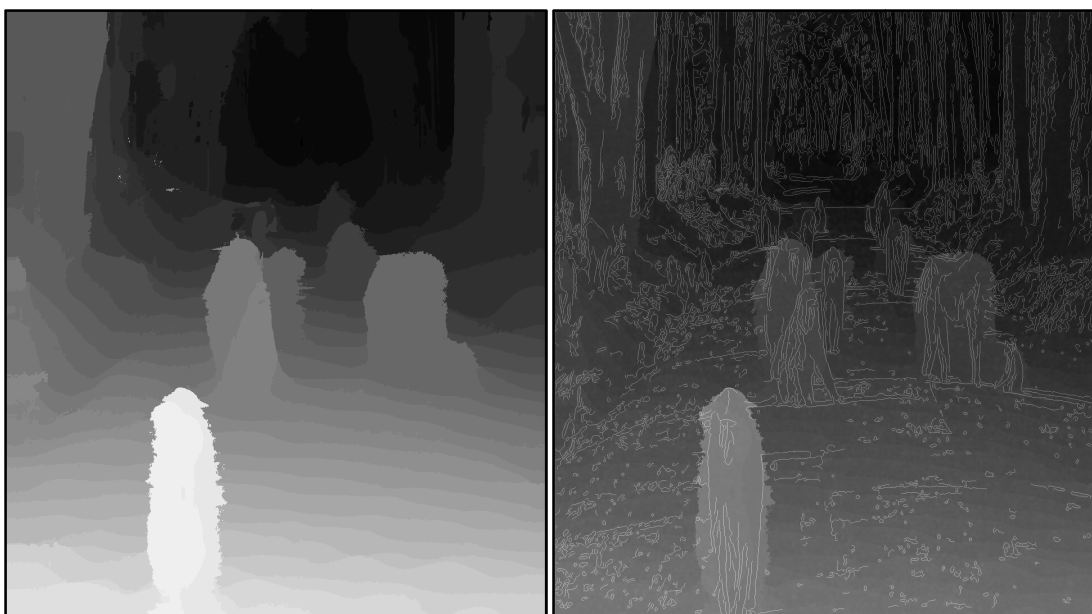


Abbildung 32: Bild 1, linke DK NQ; links: ohne Saliency Map, rechts: mit Saliency Map.

## Testbild 2



Abbildung 33: Testbild 2, linkes Stereobild, DK NQ ohne Saliency Map, Position 1,0.

Das Ergebnis von Frage (b) ist laut **Tabelle 9** und **Tabelle 10** für beide Kamerapositionen  $0,0/2,0$  und  $-0,5/0,5$  gleich und ergibt sich wie folgt: Die beiden Methoden, welche eine Saliency Map verwenden, weisen die meisten Artefakte auf, wobei bei Methode NQ' die meisten Artefakte entstehen. Die wenigsten Artefakte finden sich in dem Bild aus Methode NQ, gefolgt von jenem aus Methode HQ. Wie in Testbild 1 hat auch hier die Berücksichtigung der Saliency Map zu einem vermehrten Auftreten von Artefakten geführt.



Resultierende Regressionsmatrix für Frage (a):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	-1	1			1,000	NQ'
22	-1		1		0,682	HQ
24	-1			1	0,917	HQ'
24		1	-1		1,000	NQ'
18		1		-1	0,778	NQ'
18			-1	1	0,778	HQ'
Summe	-3	3	-1	1	NQ < HQ < HQ' < NQ'	
Summe (gewichtet)	-2,598	2,778	-1,096	0,917		

Tabelle 9: Bild 2, Kameraposition 0,0/2,0; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	-1	1			1,000	NQ'
23	-1		1		0,957	HQ
24	-1			1	0,958	HQ'
24		1	-1		1,000	NQ'
24		1		-1	0,917	NQ'
20			-1	1	0,900	HQ'
Summe	-3	3	-1	1	NQ < HQ < HQ' < NQ'	
Summe (gewichtet)	-2,915	2,917	-0,943	0,942		

Tabelle 10: Bild 2, Kameraposition -0,5/1,5; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

Die Objekte wurden bei der DK HQ ähnlich wie bei der Kombination aus DK HQ und Saliency Map segmentiert (siehe **Abbildung 34**). Trotzdem sind Ungenauigkeiten in der Erkennung entstanden, welche zu mehr Artefakten bei Bildern aus DK HQ mit Saliency Map führen. In **Abbildung 35** sieht man, dass die DK NQ und die DK NQ mit Saliency Map Unterschiede in der Erkennung der Objektkanten aufweisen. Der Kantenoperator hat andere Objektgrenzen erkannt, als in der Disparitätskarte vorgegeben sind. Das führt sowohl bei der Generierung neuer asymmetrischer als auch symmetrischer Kamerapositionen zu erhöhtem Artefakt-Vorkommen.



Abbildung 34: Bild 2, linke DK HQ; links: ohne Saliency Map, rechts: mit Saliency Map.



Abbildung 35: Bild 2, linke DK NQ; links: ohne Saliency Map, rechts: mit Saliency Map.

### Testbild 3



Abbildung 36: Testbild 3, linkes Stereobild, DK NQ ohne Saliency Map, Position 1,0.

Alle generierten Bilder dieses Testsets hatten tendenziell wenige bzw. unauffällige Artefakte, weshalb es bei der Bestimmung der Artefakte verhältnismäßig oft zu Stimmenthaltungen kam. Laut Usertests für Kameraposition 0,0/2,0 hat Methode HQ die meisten Artefakte vor Methode HQ', Methode NQ und Methode NQ' bei Frage (a) (siehe **Tabelle 11**). Für Kameraposition -0,5/0,5 kam es bei Frage (a) zu dem Ergebnis, dass Methode HQ' mehr Artefakte aufweist als Methode HQ. Methode HQ weist aber auch hier mehr Artefakte auf als Methode NQ und Methode NQ', und Methode NQ' hat die wenigsten Artefakte (siehe **Tabelle 12**).

Daraus lässt sich ablesen, dass eine zusätzliche Verwendung von Saliency Maps bei der Generierung von Stereoansichten das Auftreten von Artefakten vermindern kann. Speziell bei der Erstellung von asymmetrischen Kameraansichten ließ sich zeigen, dass beide Disparitätskarten in Kombination mit einer Saliency Map weniger Artefakte aufweisen, als die Disparitätskarte ohne zusätzliche Verbesserungen.

Resultierende Regressionsmatrix für Frage (a):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
9	1	-1			0,667	NQ
19	-1		1		0,684	HQ
22	-1			1	0,954	HQ'
17		-1	1		0,588	HQ
21		-1		1	0,952	HQ'
21			1	-1	0,904	HQ
Summe	-1	-3	3	1	NQ' < NQ < HQ' < HQ	
Summe (gewichtet)	-0,971	-2,207	2,176	1,002		

Tabelle 11: Bild 3, Kameraposition 0,0/2,0; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
12	1	-1			0,583	NQ
21	-1		1		0,857	HQ
23	-1			1	0,956	HQ'
18		-1	1		0,889	HQ
20		-1		1	0,950	HQ'
22			-1	1	0,863	HQ'
Summe	-1	-3	1	3	NQ' < NQ < HQ < HQ'	
Summe (gewichtet)	-1,23	-2,422	0,883	2,769		

Tabelle 12: Bild 3, Kameraposition -0,5/1,5; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

In **Abbildung 37** und **Abbildung 38** scheinen die Unterschiede der verschiedenen Disparitätskarten auf, die zu der Generierung der 3D-Bilder in Testset 3 verwendet wurden. Die Objekte wurden größtenteils klar erkannt und liegen innerhalb der Objektgrenzen, die durch den Canny Kantenoperator zur Kantendetektion im jeweils rechten Bild mit kombinierter Saliency Map eingezeichnet sind. Dadurch ergeben sich im resultierenden Bild wenige Artefakte in Form von doppelten oder ausgefranzten Kanten und unvollständigen Objekten.

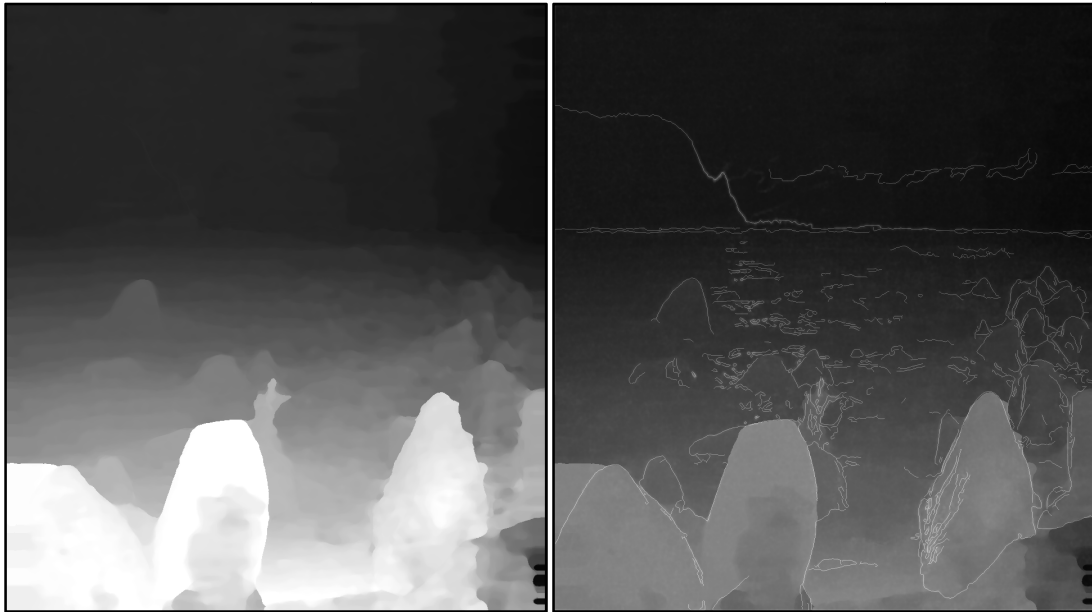


Abbildung 37: Bild 3, linke DK HQ; links: ohne Saliency Map, rechts: mit Saliency Map.

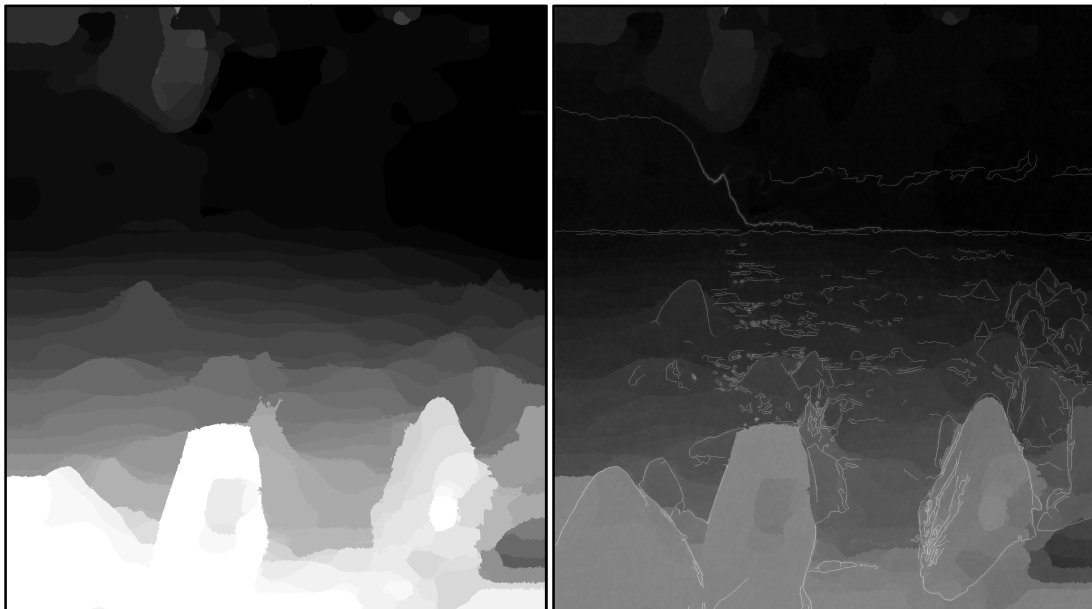


Abbildung 38: Bild 3, linke DK NQ, links: ohne Saliency Map, rechts: mit Saliency Map.

## Testbild 4



Abbildung 39: Testbild 4, linkes Stereobild, DK NQ ohne Saliency Map, Position 1,0.

Frage (a) nach dem häufigeren Artefakt-Vorkommen in den gezeigten Bildern brachte für beide Kamerapositionen  $0,0/2,0$  und  $-0,5/0,5$  das gleiche eindeutige Ergebnis (siehe **Tabelle 13** und **Tabelle 14**). Methode NQ' verursachte die meisten Artefakte gefolgt von Methode HQ'. Die Methoden NQ und HQ ohne Saliency Map führten zu weniger Artefakten im Bild. Demnach konnte die Saliency Map die Disparitätskarte nicht verbessern, sondern verschlechterte das Ergebnis in diesem Fall deutlich.

Resultierende Regressionsmatrix für Frage (a):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	-1	1			0,958	NQ'
22	1		-1		0,727	NQ
22	-1			1	0,636	HQ'
24		1	-1		1,000	NQ'
23		1		-1	1,000	NQ'
24			-1	1	1,000	HQ'
Summe	-1	3	-3	1	HQ < NQ < HQ' < NQ'	
Summe (gewichtet)	-0,867	2,958	-2,727	0,636		

Tabelle 13: Bild 4, Kameraposition 0,0/2,0; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	-1	1			0,917	NQ'
24	1		-1		0,875	NQ
23	-1			1	0,783	HQ'
24		1	-1		1,000	NQ'
24		1		-1	1,000	NQ'
24			-1	1	1,000	HQ'
Summe	-1	3	-3	1	HQ < NQ < HQ' < NQ'	
Summe (gewichtet)	-0,824	2,917	-2,875	0,783		

Tabelle 14: Bild 4, Kameraposition -0,5/1,5; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

In **Abbildung 40** bis **Abbildung 43** sind jeweils die DK HQ ohne und mit Beeinflussung durch die Saliency Map bzw. die DK NQ ohne und mit Beeinflussung durch die Saliency Map dargestellt.



Abbildung 40: Bild 4, linke DK HQ, links: ohne Saliency Map, rechts: mit Saliency Map.

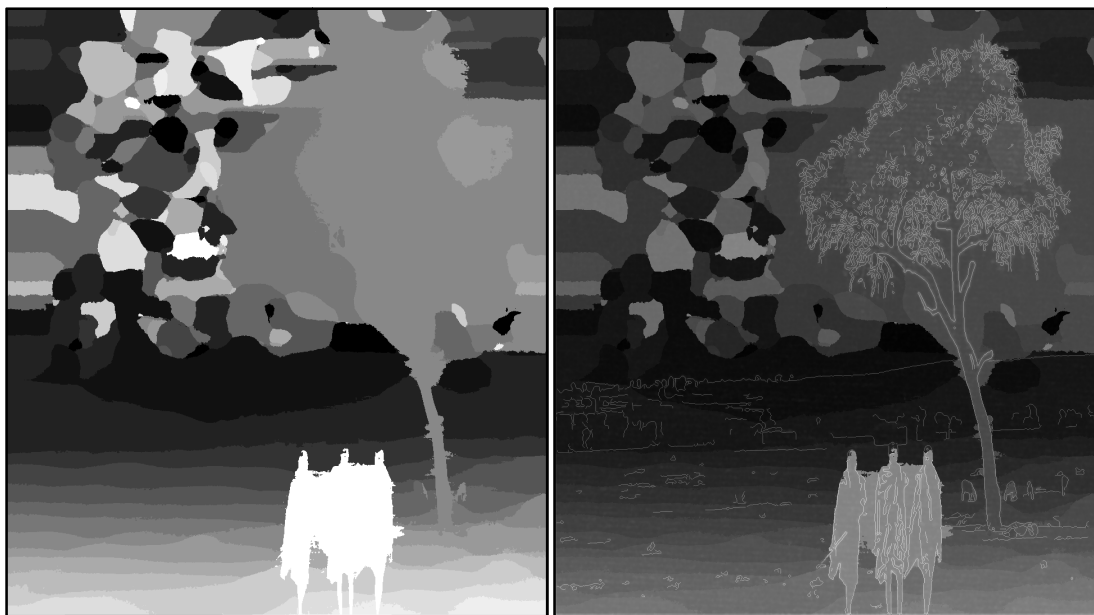


Abbildung 41: Bild 4, linke DK NQ, links: ohne Saliency Map, rechts: mit Saliency Map.



In der Vergrößerung in **Abbildung 42** und **Abbildung 43** der jeweiligen Disparitätskarten ist anhand der in der Saliency Map enthaltenen Kantenkarte, welche aus dem realen Bild erstellt wurde, zu erkennen, dass sich die Objekte nicht korrekt innerhalb der erkannten Kanten befinden. Das führt gerade beim Warpen auf neue externe Positionen, wo weniger Pixelinformationen für neu aufgedeckte Bildteile vorhanden sind, zu mehr Artefakten. Bei diesem Testbild treten die Artefakte vor allem bei den Personen im Vordergrund auf. Die ungenau erkannten Figuren führen im gewarpten Bild zu abgeschnittenen Köpfen und scheinbar zerrissenen Körpern der Figuren.



Abbildung 42: Ausschnitt aus linker DK HQ mit Saliency Map.



Abbildung 43: Ausschnitt aus linker DK NQ mit Saliency Map.

## Testbild 5



Abbildung 44: Testbild 5, linkes Stereobild, DK NQ ohne Saliency Map, Position 1,0.

Die Häufigkeit von Artefakten nach Frage (a) in den Usertests an den Kamerapositionen 0,0/2,0 und -0,5/0,5 ist wie folgt gereiht (siehe **Tabelle 15** und **Tabelle 16**). Methode HQ' führt zu den meisten Artefakten, gefolgt von Methode NQ'. Methode HQ auf Platz 3 und Methode NQ auf Platz 4 führen zu den wenigsten Artefakten.

Die Bilder, welche unter Berücksichtigung der Saliency Map gewarpt wurden, weisen mehr Artefakte auf und trugen in diesem Fall nicht zur Verbesserung bei. Diese Bilder wurden bei Frage (b) auch mit einem schlechteren Tiefeneindruck bewertet, was darauf hinweist, dass sich die Testpersonen bei diesem Testset von den Artefakten beeinflussen ließen.

Resultierende Regressionsmatrix für Frage (a):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
23	-1	1			1,000	NQ'
18	-1		1		0,667	HQ
24	-1			1	1,000	HQ'
23		1	-1		0,739	NQ'
22		-1		1	0,682	HQ'
22			-1	1	0,909	HQ'
Summe	-3	1	-1	3	NQ < HQ < NQ' < HQ'	
Summe (gewichtet)	-2,667	1,057	-0,982	2,591		

Tabelle 15: Bild 5, Kameraposition 0,0/2,0; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	-1	1			1,000	NQ'
21	-1		1		0,810	HQ
24	-1			1	1,000	HQ'
23		1	-1		0,826	NQ'
21		-1		1	0,810	HQ'
24			-1	1	0,917	HQ'
Summe	-3	1	-1	3	NQ < HQ < NQ' < HQ'	
Summe (gewichtet)	-2,810	1,017	-0,933	2,726		

Tabelle 16: Bild 5, Kameraposition -0,5/1,5; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

In **Abbildung 45** ist allerdings auch zu erkennen, dass der Hintergrund wie bei Bild 2 sehr unstrukturiert ist, was die Bewertung des Tiefeneindrucks erschwert. Die DK NQ ohne Saliency Map in **Abbildung 46** zeigt wieder, dass hier die Objekte am besten erkannt wurden, was auch durch die Usertests bestätigt wurde. Die Kantenerkennung in der Saliency Map entspricht eher der Erkennung der Objektkanten der DK HQ, welche die Objektkanten an sich schlechter erkannt hat. Eine Überlagerung mit dieser Saliency Map führt dadurch zu mehr Artefakten.

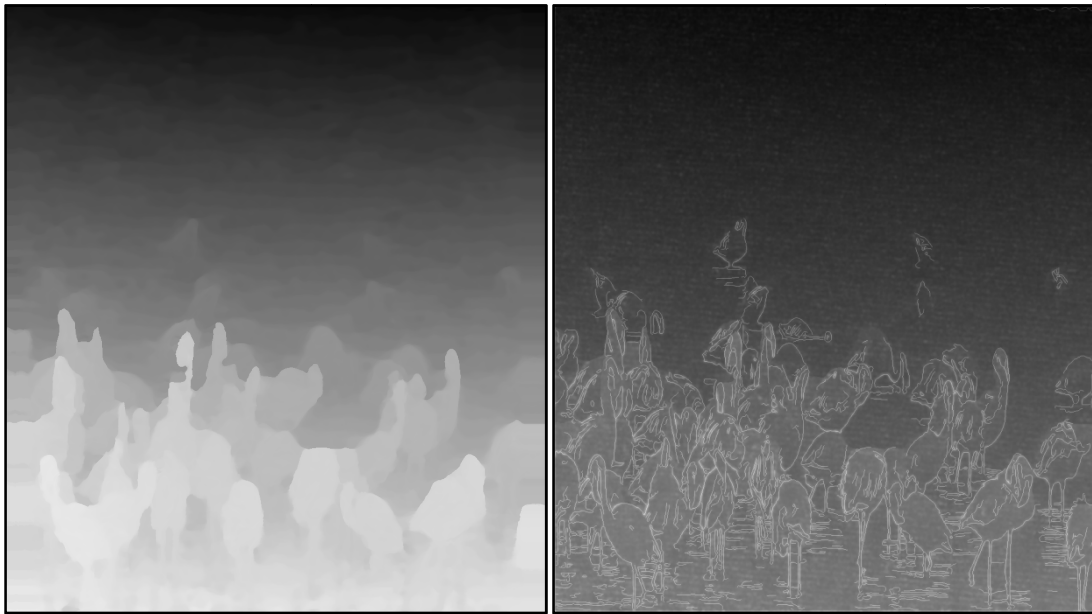


Abbildung 45: Bild 5, linke DK HQ, links: ohne Saliency Map, rechts: mit Saliency Map.

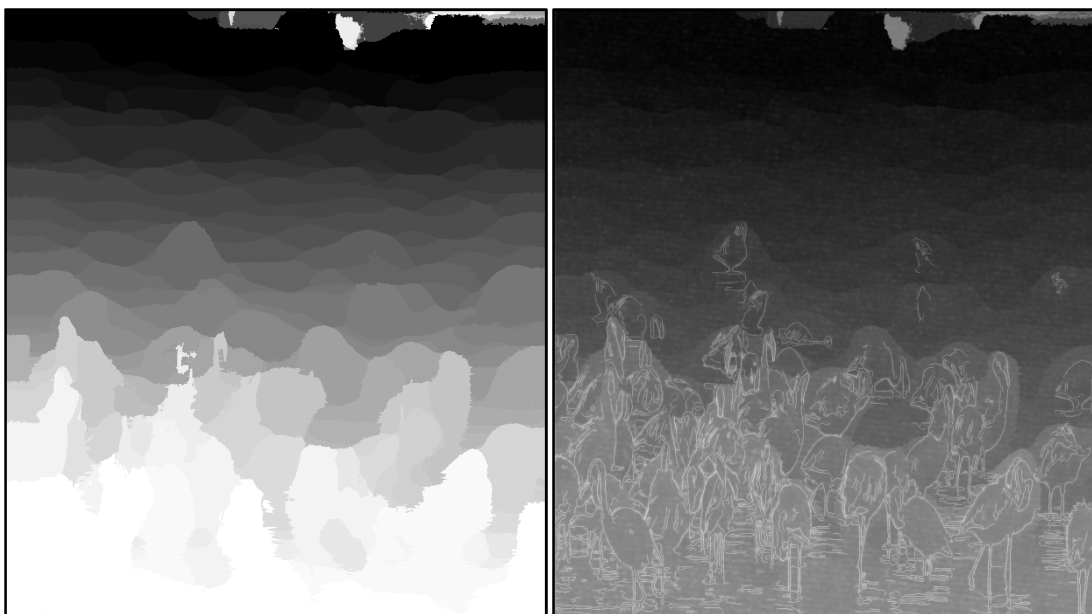


Abbildung 46: Bild 5, linke DK NQ, links: ohne Saliency Map, rechts: mit Saliency Map.

## Testbild 6



Abbildung 47: Testbild 6, linkes Stereobild, DK NQ ohne Saliency Map, Position 1,0.

Beide Kamerapositionen  $0,0/2,0$  und  $-0,5/0,5$  führten im Test zu der gleichen Reihung bei Frage (a) nach der Häufigkeit des Artefakt-Vorkommens (siehe **Tabelle 17** und **Tabelle 18**). Methode HQ', gefolgt von Methode NQ', erzeugen die meisten Artefakte, während Methode NQ sowie Methode HQ die wenigsten Artefakte erzeugen. Für beide Kamerapositionen und beide Disparitätskarten wurden die Artefakte durch Verwendung von Saliency Maps in diesem Testset vermehrt.

Resultierende Regressionsmatrix für Frage (a):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
23	-1	1			0,957	NQ'
16	-1		1		0,750	HQ
24	-1			1	1,000	HQ'
24		1	-1		0,917	NQ'
23		-1		1	0,826	HQ'
24			-1	1	0,958	HQ'
Summe	-3	1	-1	3	NQ < HQ < NQ' < HQ'	
Summe (gewichtet)	-2,707	1,047	-1,125	2,784		

Tabelle 17: Bild 6, Kameraposition 0,0/2,0; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
22	-1	1			1,000	NQ'
22	-1		1		0,727	HQ
24	-1			1	0,917	HQ'
22		1	-1		0,636	NQ'
23		-1		1	0,826	HQ'
24			-1	1	0,917	HQ'
Summe	-3	1	-1	3	NQ < HQ < NQ' < HQ'	
Summe (gewichtet)	-2,644	0,810	-0,826	2,659		

Tabelle 18: Bild 6, Kameraposition -0,5/1,5; Evaluierung der häufigsten Artefakt-Vorkommen; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

Dieses Ergebnis ist ähnlich wie bei den vorhergehenden Testsets. Wenn man die unterschiedlichen verwendeten DKs in **Abbildung 48** und **Abbildung 49** betrachtet, ergeben sich auch hier die gleichen Rückschlüsse: Die DK NQ hat die Objekte am besten erkannt, während sowohl die Methoden für die DK HQ, als auch die Methoden mit zusätzlicher Saliency Map vermehrt Artefakte bei Objekten zulassen. Eine Überlagerung von ausgefransten Objektkanten aus der Disparitätskarte mit scharfen Objektkanten aus der Saliency Map erklärt die auftretenden Artefakte an den Objekträndern.

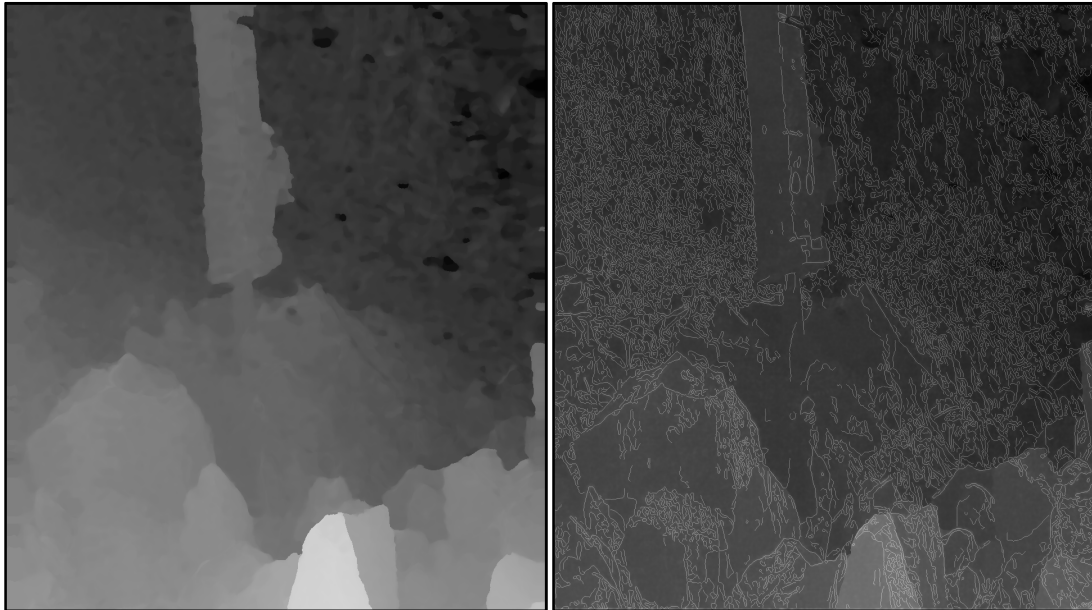


Abbildung 48: Bild 6, linke DK HQ, links: ohne Saliency Map, rechts: mit Saliency Map.

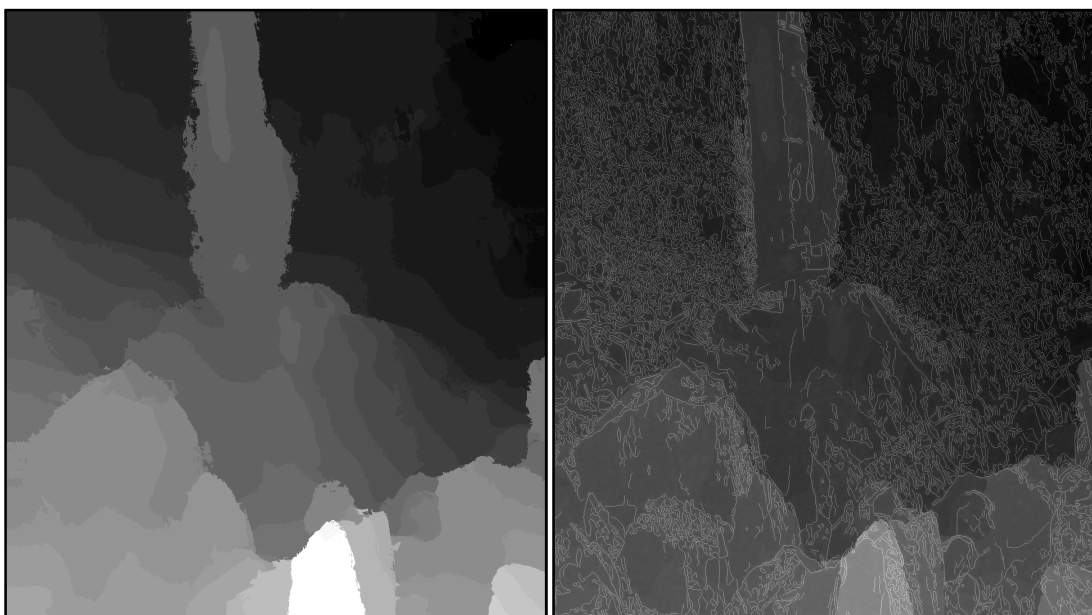


Abbildung 49: Bild 6, linke DK NQ, links: ohne Saliency Map, rechts: mit Saliency Map.

### 4.5.3 Gesamtauswertung für Frage (b)

In **Tabelle 19** ist das Ergebnis der Usertests für Frage (b), nämlich in welchem Bild der Tiefeneindruck subjektiv gesehen höher ist, für alle Testbilder mit Kameraposition 0,0/2,0 zusammengefasst. Bei Bildern, die mit der DK NQ erstellt wurden, ist der Tiefeneindruck niedriger als bei jenen, die mit der DK HQ erstellt wurden. Der Tiefeneindruck hat sich durch den Einsatz der Saliency Map bei der DK HQ erhöht, während er bei der DK NQ niedriger wurde.

Ergebnis zu Frage (b): In welchem Bild ist der Tiefeneindruck höher?

Bild	NQ	NQ'	HQ	HQ'
Summe B1-B6 (gewichtet)	0,201	-6,962	5,314	6,447
NQ' < NQ < HQ < HQ'				

Tabelle 19: Frage (b) Gesamtergebnis für die Bilder B1 – B6, Position 0,0/2,0; Evaluierung des Tiefeneindruckes; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Die Summe ist das Ergebnis der gewichteten Punktevergabe aller Bilder.

**Tabelle 20** zeigt, gereiht nach dem höchsten bis zum niedrigsten wahrgenommenen Tiefeneindruck, die dazugehörigen Bildererstellungsmethoden für Kameraposition 0/2.

Reihung nach Tiefeneindruck (1 = stärkster Tiefeneindruck)	Methode
1	HQ'
2	HQ
3	NQ
4	NQ'

Tabelle 20: Reihung der Methoden nach dem am höchsten wahrgenommenen Tiefeneindruck bei Kameraposition 0,0/2,0, beginnend mit jener Methode, die den höchsten Tiefeneindruck aufweist.

Betrachtet man die Disparitätskarte neben den Disparitätskarten mit Saliency Map (siehe **Abbildung 50** und Abbildungen in Kapitel 4.5.2), dann ist leicht zu erkennen, dass die mit Saliency Map berechnete Disparitätskarte dunkler erscheint. Die unterschiedlichen Grauwerte beschreiben die im Bild auftretenden Disparitätswerte. Durch die Kombination der Kantenerkennung, Texturerkennung und Disparitätskarte ist das rechte Bild dunkler geworden, was eine Verminderung des Disparitätsbereichs darstellt.



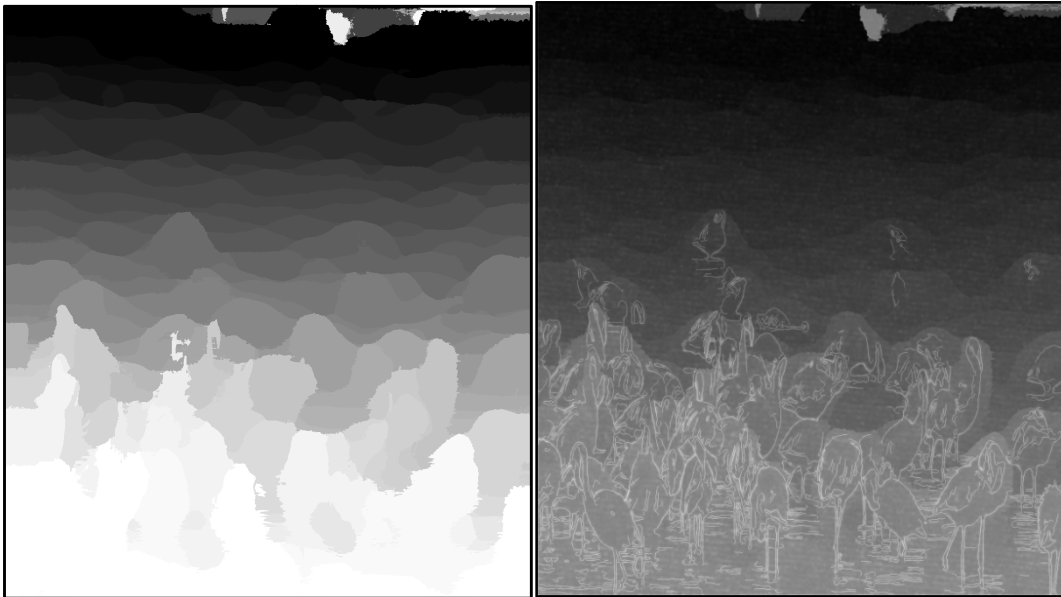


Abbildung 50: Bild 5, linke DK NQ, links: ohne Saliency Map, rechts: mit Saliency Map.

Damit erklärt sich, warum Bilder, die mit DK NQ und Saliency Map erstellt wurden, laut Testpersonen einen niedrigeren Tiefeneindruck aufweisen als Bilder, die mit DK NQ ohne Saliency Map erstellt wurden. Eine knappe Mehrheit hat entschieden, dass Bilder, die mit DK HQ und Saliency Map erstellt wurden, einen besseren Tiefeneindruck aufweisen, als Bilder die nur mit DK HQ generiert wurden. Nachdem diese Studie auf der subjektiven Wahrnehmung der Testpersonen beruht, ist davon auszugehen, dass der Tiefeneindruck bei Bildern, die mit DK HQ erstellt wurden, bereits sehr hoch wahrgenommen wurde und beim Vergleich zwischen Bildern, die mit DK HQ ohne und mit Saliency Map erstellt wurden, keine klare Entscheidung bezüglich der Tiefe getroffen werden konnte - das erklärt auch das knappe Ergebnis. Weiters macht die asymmetrische Kameraposition das Abschätzen der Tiefe im Bild schwieriger, wenn kein offensichtlicher Unterschied vorhanden ist.

Bild	NQ	NQ'	HQ	HQ'
Summe B1-B6 (gewichtet)	-4,343	-9,243	9,325	5,261
<b><math>NQ' &lt; NQ &lt; HQ' &lt; HQ</math></b>				

Tabelle 21: Frage (b) Gesamtergebnis für die Bilder B1 – B6, Position -0,5/1,5; Evaluierung des Tiefeneindruckes; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Die Summe ist das Ergebnis der gewichteten Punktevergabe aller Bilder.

In **Tabelle 21** und **Tabelle 22** ist das Ergebnis der Usertests für Frage (b), in welchem Bild der Tiefeneindruck höher ist, für alle Testbilder der Kameraposition  $-0,5/1,5$  zu sehen. Hier weisen jene Bilder, die mit der DK HQ erstellt wurden, den höchsten Tiefeneffekt auf. Jene Bilder, die zusätzlich durch eine Saliency Map beeinflusst wurden, weisen eindeutig einen etwas schlechteren Tiefeneindruck auf, als jene die ohne Saliency Map erstellt wurden. Wie oben beschrieben, lässt sich dies dadurch erklären, dass der Einsatz der Saliency Maps den Disparitätsbereich vermindert.

Reihung nach Tiefeneindruck (1 = höchster Tiefeneindruck)	Methode
1	HQ
2	HQ'
3	NQ
4	NQ'

Tabelle 22: Reihung der Methoden nach dem am höchsten wahrgenommenen Tiefeneindruck bei Kameraposition  $-0,5/1,5$ , beginnend mit jener Methode, die den höchsten Tiefeneindruck aufweist.

**Abbildung 51** zeigt, wie viel Prozent der befragten Personen für welche Methode gestimmt haben. In der Gegenüberstellung der Ergebnisse von symmetrischem Kamerasetup und asymmetrischem Kamerasetup zeigen sich deutlich die Unterschiede in den Ergebnissen. Für die Ergebnisse der einzelnen Testbilder siehe Kapitel 4.5.4.

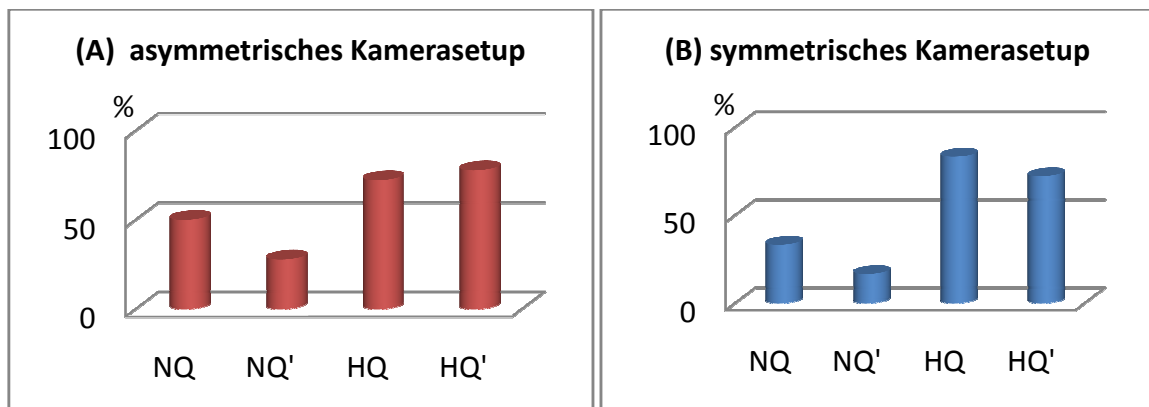


Abbildung 51: Auswertung der Ergebnisse zu Frage (b); (A) zeigt, wie viele der befragten Personen beim symmetrischen Kamerasetup ( $-0,5/1,5$ ) für welche Methode gestimmt haben. (B) zeigt, wie viele der befragten Personen beim asymmetrischen Kamerasetup ( $0,0/2,0$ ) für welche Methode gestimmt haben.

#### 4.5.4 Auswertung der einzelnen Testbilder für Frage (b)

##### Testbild 1

In Bild 1 wurden Unterschiede im Tiefeneindruck gut erkannt, da sich die Testpersonen hier sowohl an Objekten im Vordergrund- und Hintergrund sowie an Unebenheiten am Boden orientieren konnten.



Die Ergebnisse für Frage (b) der Usertests für Bild 1 waren für beide Kamerapositionen gleich (siehe **Tabelle 23** und **Tabelle 24**). Der Tiefeneindruck bei Methode HQ war am besten, gefolgt von Methode HQ'. Beide Methoden erzielten einen besseren Tiefeneindruck als NQ, und NQ' wies den schlechtesten Tiefeneindruck auf. Die Entscheidung für den besseren Tiefeneindruck zwischen HQ und HQ' im direkten Vergleich fiel allerdings für beide Kamerapositionen knapp aus.

In diesem Fall trug das Warping mit Saliency Map nicht zur Verbesserung des Tiefeneindrucks bei. Der Tiefeneindruck ist wie erwartet bei jenen Bildern größer, die einen höheren Tiefenbereich aufweisen. Ansonsten verschlechtert das Warping mit Saliency Map das Ergebnis des Tiefeneindrucks, wenn, wie in den betrachteten Testsets, der Tiefenbereich nicht verändert wird.

Resultierende Regressionsmatrix für Frage (b):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
19	1	-1			0,684	NQ
21	-1		1		0,619	HQ
24	-1			1	0,667	HQ'
23		-1	1		0,913	HQ
24		-1		1	0,834	HQ'
24			1	1	0,500	HQ / HQ'
Summe	-1	-3	3	3	NQ' < NQ < HQ' < HQ	
Summe (gewichtet)	-0,602	-2,431	2,032	2,001		

Tabelle 23: Bild 1, Kameraposition 0,0/2,0; Evaluierung des Tiefeneindrucks; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
19	1	-1			0,631	NQ
22	-1		1		0,863	HQ
24	-1			1	0,708	HQ'
23		-1	1		0,956	HQ
23		-1		1	0,782	HQ'
23			1	-1	0,521	HQ
Summe	-1	-3	3	1	NQ' < NQ < HQ' < HQ	
Summe (gewichtet)	-0,94	-2,369	2,34	0,969		

Tabelle 24: Bild 1, Kameraposition -0,5/1,5; Evaluierung des Tiefeneindrucks; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

## Testbild 2

In Bild 2 war es für die Testpersonen schwierig, den Tiefeneindruck zu vergleichen, weil der Wald im Hintergrund relativ flach ist und auch die Wiese nicht viel Tiefenstruktur aufweist. Aufgrund dessen ließen sich die Testpersonen teilweise von Artefakten in der Wiese oder an den Objekten irritieren. In **Abbildung 34** erkennt man, wie wenig Struktur der Hintergrund in Bild 2 aufweist, was die Orientierung für den Tiefeneindruck erschwert.



Die Reihung des besten Tiefeneindrucks nach Frage (b) für die Kameraposition 0,0/2,0 in Bild 2 ist nach **Tabelle 25** folgendermaßen: Methode HQ' wies den besten Tiefeneindruck auf, gefolgt von Methode HQ. Den drittenbesten Tiefeneindruck hatte die Methode NQ vor der Methode NQ'. Die Reihung für Kameraposition -0,5/0,5 nach Frage (b) ergibt sich nach **Tabelle 26** wie folgt: In diesem Fall wies die Methode HQ den besten Tiefeneindruck auf, vor der Methode HQ'. Die Methode NQ und die Methode NQ' sind auch hier auf Platz 3 und Platz 4.

Auch in diesem Testbild konnte die Verwendung der Saliency Map bei der Erstellung der 3D-Bilder den Tiefeneindruck nicht erhöhen. Nur bei Bildern mit asymmetrischer Kameraposition und DK HQ wurde mit der Saliency Map eine Erhöhung des Tiefeneindrucks erzielt.

Resultierende Regressionsmatrix für Frage (b):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
21	1	-1			0,809	NQ
22	-1		1		0,545	HQ
24	-1			1	0,541	HQ'
23		-1	1		0,782	HQ
22		-1		1	0,818	HQ'
23			-1	1	0,608	HQ'
Summe	-1	-3	1	3	NQ' < NQ < HQ < HQ'	
Summe (gewichtet)	-0,277	-2,409	0,719	1,967		

Tabelle 25: Bild 2, Kameraposition 0,0/2,0; Evaluierung des Tiefeneindrucks; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
23	1	-1			0,652	NQ
21	-1		1		0,714	HQ
24	-1			1	0,625	HQ'
23		-1	1		0,826	HQ
23		-1		1	0,782	HQ'
22			1	-1	0,681	HQ
Summe	-1	-3	3	1	NQ' < NQ < HQ' < HQ	
Summe (gewichtet)	-0,687	-2,26	2,221	0,726		

Tabelle 26: Bild 2, Kameraposition -0,5/1,5; Evaluierung des Tiefeneindrucks; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

### Testbild 3

In Bild 3 konnte der Tiefeneindruck gut bewertet und nachvollzogen werden, da sowohl im Vordergrund als auch im Hintergrund Objekte sichtbar sind. Weiters sind die Wellen strukturiert, was einen guten Anhaltspunkt für den Tiefeneindruck darstellt. Sowohl für die Kameraposition 0,0/2,0, als auch für die Kameraposition -0,5/0,5 fiel das Ergebnis der Usertests ähnlich aus (siehe auch **Tabelle 27** und **Tabelle 28**).



Resultierende Regressionsmatrix für Frage (b):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
23	-1	1			0,739	NQ'
22	1		1		0,500	NQ / HQ
24	1			1	0,500	NQ / HQ'
24		-1	1		0,625	HQ
23		1		-1	0,521	NQ'
24			1	-1	0,541	HQ
Summe	1	1	3	-1	HQ' < NQ < NQ' < HQ	
Summe (gewichtet)	0,261	0,635	1,666	-0,562		

Tabelle 27: Bild 3, Kameraposition 0,0/2,0; Evaluierung des Tiefeneindrucks; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
16	-1	1			0,562	NQ'
23	-1		1		0,782	HQ
24	-1			1	0,583	NQ'
23		-1	1		0,608	HQ
24		1		1	0,500	NQ' / HQ'
24			1	-1	0,667	HQ
Summe	-3	1	3	1	NQ < HQ' < NQ' < HQ	
Summe (gewichtet)	-1,927	0,454	2,057	0,416		

Tabelle 28: Bild 3, Kameraposition -0,5/1,5; Evaluierung des Tiefeneindruckes; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

Die Testpersonen bevorzugten das 3D-Bild, welches mit DK HQ und ohne Saliency Map erzeugt wurde, vor dem 3D-Bild, welches mit DK NQ und mit Saliency Map generiert wurde. Letzteres lag bei Kameraposition 0,0/2,0 in der Präferenz vor dem Bild, welches mit DK NQ und ohne Saliency Map gewarpt wurde, das den Tiefeneindruck betreffend jenem Bild bevorzugt wird, welches mit DK HQ und mit Saliency Map erstellt wurde. Bei Kameraposition -0,5/0,5 ist die Präferenz von NQ und HQ' umgekehrt. In Bildern, die mit DK HQ erstellt wurden, hat sich der Tiefeneindruck durch die Verwendung einer Saliency Map nicht verbessert. Bei jenen Bildern, die aufgrund von der DK NQ generiert wurden, konnte durch das Miteinbeziehen der Saliency Map eine Verbesserung erzielt werden, obwohl dadurch ein geringerer Disparitätsbereich einbezogen wurde.

Insgesamt ist das Ergebnis hier nicht immer eindeutig. So stimmten beispielsweise beim Vergleich zwischen NQ und HQ bei Frage (b) Kameraposition 0/2 gleich viele Personen für NQ und für HQ. Betrachtet man allgemein den Prozentsatz der Personen, die jeweils für ein linkes oder rechtes Bild gestimmt haben, dann ist der Prozentsatz eher ausgeglichen, was darauf zurückzuführen ist, dass der Tiefeneindruck in Testbild 3 insgesamt bei jeder getesteten Methode sehr gut war und damit nicht immer eine eindeutige Entscheidung für eine bestimmte Antwort getroffen werden konnte.



## Testbild 4

Der Tiefeneindruck konnte in Bild 4 gut unterschieden werden, da viele Anhaltspunkte in Form von Objekten oder Oberflächenstrukturen im Bild vorhanden sind. Aber sowohl die DK NQ als auch DK HQ konnte die Personen im Vordergrund nicht korrekt erkennen. Die Testpersonen wurden allerdings dazu angehalten, diesen Umstand für die Bewertung des Tiefeneindrucks außer Acht zu lassen.



Die Antwort auf Frage (b) für die Kamerapositionen 0,0/2,0 in Bild 4 ist laut **Tabelle 29** wie folgt: Die Methoden, die eine Saliency Map verwenden, haben den höheren Tiefeneindruck, wobei Methode HQ' am besten bewertet wird, gefolgt von Methode NQ'. Methode HQ hat den niedrigsten Tiefeneindruck ergeben, gefolgt von Methode NQ.

In diesem Fall kann davon ausgegangen werden, dass sich die Testpersonen trotzdem von den Artefakten beeinflussen ließen und durch diese Artefakte, die hier in Form von verschwommenen und zerrissenen Objekten auftreten, einen verbesserten Tiefeneindruck wahrgenommen haben. Bei Testbildern aus Testset 4 mit asymmetrischer Kameraposition traten Artefakte am häufigsten auf.

Für die Kameraposition -0,5/0,5 verhält sich der Tiefeneindruck anders (siehe **Tabelle 30**). Auch hier wurde Methode HQ' im Bezug auf den Tiefeneindruck am besten bewertet, die Methode NQ' hat in diesem Fall allerdings den schlechtesten Tiefeneindruck ergeben. Methode HQ liegt auf Platz 2 und Methode NQ auf Platz 3 des besten Tiefeneindrucks bei Kameraposition -0,5/0,5.

Resultierende Regressionsmatrix für Frage (b):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
23	-1	1			0,652	NQ'
23	1		-1		0,652	NQ
24	1			1	0,500	NQ / HQ'
24		1	-1		0,541	NQ'
24		-1		1	0,625	HQ'
24			-1	1	0,791	HQ'
Summe	1	1	-3	3		HQ < NQ < NQ' < HQ'
Summe (gewichtet)	0,5	0,568	-1,984	1,916		

Tabelle 29: Bild 4, Kameraposition 0,0/2,0; Evaluierung des Tiefeneindrucks; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um



die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	1	-1			0,667	NQ
24	-1		1		0,625	HQ
24	-1			1	0,667	HQ'
24		1	-1		0,583	NQ'
24		-1		1	0,583	HQ'
23			-1	1	0,608	HQ'
Summe	-1	-1	-1	3	NQ' < NQ < HQ < HQ'	
Summe (gewichtet)	-0,625	-0,667	-0,566	1,858		

Tabelle 30: Bild 4, Kameraposition -0,5/1,5; Evaluierung des Tiefeneindrucks; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

Aus den Usertests ergibt sich, dass in diesem Fall die Saliency Map den Tiefeneindruck - besonders bei der asymmetrischen Kameraposition - positiv beeinflusst hat. Allerdings ist das Ergebnis wiederum knapp ausgefallen, da sich bei beiden Kamerapositionen und bei allen Vergleichen jeweils nur geringfügig viele Testpersonen mehr für das Bild mit der letztlich favorisierten Methode entschieden haben.

## Testbild 5

Für die Testreihe mit Bild 5 war die Unterscheidung des Tiefeneindrucks für zwei nebeneinander gestellte 3D-Bilder erschwert, da die Wellen flach und unstrukturiert sind und sich im Hintergrund keine Objekte befinden. Trotzdem schafften es die Testpersonen, sich anhand der Flamingos im Vordergrund bzw. der Flamingogruppe zu orientieren.



Für die Kamerapositionen 0,0/2,0 und -0,5/0,5 in Testbild 5 ist die Reihenfolge für den besten Tiefeneindruck gleich (siehe **Tabelle 31** und **Tabelle 32**). Die Methoden, welche keine Saliency Map verwendeten, wurden hinsichtlich des Tiefeneindrucks besser bewertet, wobei Methode HQ vor Methode NQ auf Platz 1 gesetzt wurde. Methode NQ' mit Saliency Map hatte den schlechtesten Tiefeneindruck hinter Methode HQ'.

Der Einsatz der Saliency Map bei DK HQ hat den Tiefeneindruck bei Testbild 5 sehr verschlechtert, sodass auch Bilder, die mit DK NQ ohne Saliency Map erstellt wurden, einen höheren Tiefeneindruck aufwiesen. Daraus lässt sich schließen, dass die Saliency Map den Disparitätsbereich in diesem Bild sehr verringert hat.

Resultierende Regressionsmatrix für Frage (b):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
21	1	-1			0,826	NQ
23	-1		1		0,565	HQ
24	1			-1	0,541	NQ
24		-1	1		0,916	HQ
24		1		1	0,500	NQ' / HQ'
24			-1	1	0,541	HQ'
Summe	1	-3	1	1	NQ' < HQ' < NQ < HQ	
Summe (gewichtet)	0,802	-1,242	0,94	0,5		

Tabelle 31: Bild 5, Kameraposition 0,0/2,0; Evaluierung des Tiefeneindruckes; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
22	1	-1			0,636	NQ
23	-1		1		0,652	HQ
24	1			-1	0,541	NQ
24		-1	1		0,875	HQ
24		-1		1	0,583	HQ'
24			1	-1	0,625	HQ
Summe	1	-3	3	-1	NQ' < HQ' < NQ < HQ	
Summe (gewichtet)	0,525	-2,094	2,152	-0,583		

Tabelle 32: Bild 5, Kameraposition -0,5/1,5; Evaluierung des Tiefeneindruckes; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

## Testbild 6

Anhand der Disparitätswerte lässt sich ablesen, dass der minimale Disparitätswert mit -30 Pixel weiter im negativen Bereich ist als bei den anderen Testbildern. Daraus ergibt sich im 3D Raum der Effekt, dass die Objekte weiter vor der Bildebene und damit näher beim Betrachter erscheinen.



Die Reihung nach den Usertests für Testbild 6 gemäß Frage (b) über den besseren Tiefeneindruck ergibt sich für die Kameraposition 0,0/2,0 folgendermaßen (siehe **Tabelle 33**): Methode HQ wurde am besten bewertet, gefolgt von Methode HQ'. Methode NQ liegt auf Platz 3 und NQ' auf Platz 4 hinsichtlich des besseren Tiefeneindrucks. Die zusätzliche Verwendung von Saliency Maps hat den Tiefeneindruck in diesem Testset geringfügig verschlechtert, was sich auf die Änderung des Disparitätsbereichs durch Einsatz der Saliency Map zurückführen lässt.

Resultierende Regressionsmatrix für Frage (b):

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
24	1	-1			0,708	NQ
23	-1		1		0,608	HQ
24	-1			1	0,583	HQ'
24		-1	1		0,708	HQ
24		-1		1	0,667	HQ'
24			1	-1	0,625	HQ
Summe	-1	-3	3	1		NQ' < NQ < HQ' < HQ
Summe (gewichtet)	-0,483	-2,083	1,941	0,625		

Tabelle 33: Bild 6, Kameraposition 0,0/2,0; Evaluierung des Tiefeneindrucks; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

Für die Kameraposition -0,5/0,5 ergibt sich laut **Tabelle 34** folgende Reihung: Methode HQ' mit Saliency Map bewirkt einen besseren Tiefeneindruck als Methode HQ. Methode NQ liegt wieder auf Platz 3 und Methode NQ' auf Platz 4. Bei dieser symmetrischen Kameraposition konnte der Einsatz der Saliency Map auf eine DK HQ den Tiefeneindruck geringfügig verbessern.

N	NQ	NQ'	HQ	HQ'	Anteil	Favorit
22	1	-1			0,727	NQ
24	-1		1		0,791	HQ
24	-1			1	0,625	HQ'
23		-1	1		0,913	HQ
24		-1		1	0,667	HQ'
24			-1	1	0,583	HQ'
Summe	-1	-3	1	3	NQ' < NQ < HQ < HQ'	
Summe (gewichtet)	-0,689	-2,307	1,121	1,875		

Tabelle 34: Bild 6, Kameraposition -0,5/1,5; Evaluierung des Tiefeneindruckes; NQ entspricht den Ergebnissen mit Verwendung der DK NQ und NQ' den Ergebnissen mit Verwendung der DK NQ, um die Saliency Map erweitert. HQ entspricht den Ergebnissen mit Verwendung der DK HQ und HQ' den Ergebnissen mit Verwendung der DK HQ, um die Saliency Map erweitert. N ist die Anzahl der gültigen Stimmen. Der Anteil gibt an, wie viel Prozent der Testpersonen für dieses Bild als Antwort gestimmt haben.

## 5 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde der von [Lang et al. 2010] beschriebene Vorteil einer Generierung neuer 3D-Ansichten mit Beeinflussung durch Saliency Maps überprüft. Durch den Einsatz von Saliency Maps in Kombination mit Disparitätskarten sollte das Ergebnis der resultierenden 3D-Bilder dahingehend verändert werden, dass Artefakte in den resultierenden Bildern dezimiert werden, sowie ein besserer Tiefeneindruck geschaffen werden kann. Nach ausführlichen Usertests konnten in dem beschriebenen Testsetup für die Generierung externer Kamerapositionen die beschriebenen Vorteile nicht nachgewiesen werden.

Der Einfluss der Saliency Map von [Lang et al. 2010] auf den Tiefeneindruck und die Anzahl der Artefakte im Bild wurden im Zuge eines Usertests an zwei unterschiedlichen Kamerapositionen getestet. Für den Vergleich der Methode eignete sich am besten eine Disparitätskarte, welche nicht durch Saliency Maps verändert wurde. Um den von [Lang et al. 2010] beschriebenen Aspekt zu testen, dass sehr grob generierte Disparitätskarten durch den Einfluss einer Saliency Map gute Ergebnisse erzielen können, wurden die Usertests mit zwei unterschiedlichen Verfahren zur Erzeugung von Disparitätskarten durchgeführt.

Mit Hilfe der Usertests konnte gezeigt werden, dass Saliency Maps bei der Generierung externer Kamerapositionen nicht dazu beitragen, dass weniger Artefakte im 3D-Bild entstehen. [Lang et al. 2010] führen allerdings an, dass ihre Methode in einigen Bereichen Schwächen aufweist. In Randbereichen von Objekten im Vordergrund kann es trotz der Verbesserung durch Saliency Maps zu Artefakten kommen. Diese Schwäche hat sich in den Usertests nicht nur bestätigt, sondern es hat sich auch gezeigt, dass die Ergebnisse durch den Einsatz von Saliency Maps dahingehend verschlechtert werden. Bei einigen Bildern, vor allem bei asymmetrischer Kamerapositionierung und unter Verwendung von Disparitätskarten hoher Qualität, konnte eine subjektive Verbesserung durch den Einsatz von Saliency Maps beobachtet werden. Es bleibt zu testen, inwiefern andere Methoden zur Erstellung der Saliency Maps, die zusätzlich zur Disparität nicht nur Textur und Kanten im Bild berücksichtigen, die 3D-Bilder im Hinblick auf Tiefe und Artefakt-Vorkommen beeinflussen.

Bei [Lang et al. 2010] wird ergänzt, dass jene Artefakte in 3D Stereo Videos, welche durch die Erstellung mit ihrer Methode entstehen können, aufgrund der menschlichen Wahrnehmung im Gesamteindruck nicht weiter auffallen. Eine visuelle Studie in diesem Bereich überprüft per Eye-Tracking, wo und wie lange ein Zuseher während jeder Szene mit seinem Blick verweilt, um dies letztlich auf jedes einzelne Frame, in welchem Artefakte vorkommen, zurückzuführen. Dies wurde im Zuge dieser Studie nicht durchgeführt, wäre aber für zukünftige Untersuchungen interessant.

## Referenzen

- [Avidan et al. 2007] S. Avidan, A. Shamir, "Seam Carving for Content-Aware Image Resizing," ACM Transactions on Graphics, vol.26, no.3, article 10, 2007.
- [Bertalmio et al. 2000] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester, "Image Inpainting," Proceedings SIGGRAPH, pp. 417- 424, 2000.
- [Bleyer et al. 2008] M. Bleyer, M. Gelautz, "Simple but Effective Tree Structures for Dynamic Programming-Based Stereo Matching," International Conference on Computer Vision Theory and Applications, vol.2, pp.415-422, 2008.
- [DeCarlo et al. 2002] D. DeCarlo, A. Santella, "Stylization and Abstraction of Photographs," Proceedings SIGGRAPH, pp.769-776, 2002.
- [Fraunhofer 2011] Echtzeit-Tiefenmessung mit hochauflösender 3-D-Kamera. <http://www.vision.fraunhofer.de/de/projekte/579.html> (zuletzt besucht: 30.06.2012)
- [Gal et al. 2006] R. Gal, O. Sorkine, D. Cohen-Or, "Feature-Aware Texturing," Eurographics Symposium on Rendering, pp.297-303, 2006.
- [Guo et al. 2008] C. Guo, Q. Ma, L. Zhang, "Spatio-Temporal Saliency Detection using Phase Spectrum of Quaternion Fourier Transform," Conference on Computer Vision and Pattern Recognition, pp.1-8, 2008.
- [Harman et al. 2002] P. Harman, J. Flack, S. Fox, M. Dowley, "Rapid 2D to 3D Conversion," Conference on Stereoscopic Displays and Virtual Reality Systems IX, pp.78-86, 2002.
- [He et al. 2010] K. He, J. Sun, X. Tang, "Guided Image Filtering," Proceedings of the 11<sup>th</sup> European Conference on Computer Vision, pp.1-14, 2010.
- [Itti et al. 1998] L. Itti, C. Koch, E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.20, no.11, pp.1254-1259, 1998.
- [Itti et al. 2000] L. Itti, C. Koch, "A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention," Vision Research, vol.40, pp.1489-1506, 2000.
- [Krähenbühl et al. 2009] P. Krähenbühl, M. Lang, A. Hornung, M. Gross, "A System for Retargeting of Streaming Video," ACM Transactions on Graph, vol.28, no.5, article 126, 2009.

- [Lang et al. 2012] M. Lang, A. Hornung, O. Wang, S. Poulakas, A. Smolic, M. Gross, "Nonlinear Disparity Mapping for Stereoscopic 3D," ACM Transactions on Graphics, vol. 29, no.4, article 75, 2010.
- [Liu et al. 2005] F. Liu, M. Gleicher, "Automatic Image Retargeting with Fisheye-View Warping," Proceedings of the 18<sup>th</sup> Annual ACM Symposium on User Interface Software and Technology, pp.153-162, 2005.
- [Liu et al. 2006] F. Liu, M. Gleicher, "Video Retargeting: Automating Pan and Scan," Proceedings of the 14<sup>th</sup> annual ACM International Conference on Multimedia, pp. 241-250, 2006.
- [Liu et al. 2009] F. Liu, M. Gleicher, H. Jin, A. Agarwala, "Content-Preserving Warps for 3D Video Stabilization," ACM Transactions on Graphics, vol.28, no.3, article 44, 2009.
- [Longhurst et al. 2006] P. Longhurst, K. Debattista, A. Chalmers, "A GPU Based Saliency Map for High-Fidelity Selective Rendering," Proceedings of the 4<sup>th</sup> International Conference on Computer Graphics, Virtual Reality, Visualization and Interaction in Africa, pp.21-29, 2006.
- [Ma et al. 2003] Y. Ma, H. Zhang, "Contrast-Based Image Attention Analysis by using Fuzzy Growing," Proceedings of the 11<sup>th</sup> ACM International Conference on Multimedia, pp. 374-381, 2003.
- [Meer et al. 2001] P. Meer, B. Georgescu, "Edge Detection with Embedded Confidence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, no.12, pp.1351-1365, 2001.
- [Middlebury 2006] Middlebury Datasets. <http://vision.middlebury.edu/stereo> (Zuletzt besucht: 17.4.2012.)
- [Niblack et al. 1993] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, G. Taubin, "The QBIC Project: Querying Images by Content, using Color, Texture, and Shape," Proceedings of SPIE Storage and Retrieval for Image and Video Databases, vol.1908, pp.173-187, 1993.
- [Rajae-Joordens et al. 2005] R. Rajae-Joordens, J. Engel, "Paired Comparisons in Visual Perception Studies using Small Sample Sizes," an article from: Displays, vol.26, no. 1, pp.1-7, 2005.

- [Raskar et al. 2004] R. Raskar, A. Ilie, J. Yu, "Image Fusion for Context Enhancement and Video Surrealism", Proceedings of the 3<sup>rd</sup> International Symposium on Non-Photorealistic Animation and Rendering, pp.85-152, 2004.
- [Ren et al. 2010] Z. Ren, Y. Hu, L. Chia, D. Rajan, "Improved Saliency Detection Based on Superpixel Clustering and Saliency Propagation," Proceedings of the International Conference on Multimedia, pp.1099-1102, 2010.
- [Rhemann et al. 2011] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, M. Gelautz, "Fast Cost-Volume Filtering for Visual Correspondence and Beyond," Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, pp.3017-3024, 2011.
- [Rubinstein et al. 2008] M. Rubinstein, A. Shamir, A. Avidan, "Improved Seam Carving for Video Retargeting," ACM Transactions on Graphics, vol.27, no.3, article 16, 2008.
- [Rubinstein et al. 2009] M. Rubinstein, A. Shamir, S. Avidan, "Multi-operator Media Retargeting," ACM Transactions on Graphics, vol.28, no.3, article 23, 2009.
- [Schu 2010] M. Schu, "Video Processing Required for 3DTV Displays," Trident Microsystems, 2010.  
<http://www.3dathome.org/report-detail.aspx?item=2358&code=resources-white-papers&section=resources> (Zuletzt besucht: 15.7.2012)
- [Setlur et al. 2005] V. Setlur, S. Takagi, R. Raskar, M. Gleicher, B. Gooch, "Automatic Image Retargeting," Proceedings of the 4<sup>th</sup> International Conference on Mobile and Ubiquitous Multimedia, pp.59-68, 2005.
- [Shamir et al. 2009] A. Shamir, O. Sorkine, "Visual Media Retargeting," ACM SIGGRAPH Asia 2009 Courses, article 11, 2009.
- [Shokoufandeh et al. 1998] A. Shokoufandeh, I. Marsic, S. Dickinson, "View-based Object Recognition using Saliency Maps," Proceedings of the 6<sup>th</sup> International Conference on Computer Vision, pp. 445-460, 1998.
- [Sonka et al. 1999] M. Sonka, V. Hlavac, R. Boyle, "3D vision, Geometry and Radiometry" in Image Processing, Analysis, and Machine Vision, 2<sup>nd</sup> Edition, pp.457-462, 1999.
- [Tomasi et al. 1998] C. Tomasi, R. Manduchi, "Bilateral Filtering for Gray and Color Images," Proceedings of the 6<sup>th</sup> International Conference on Computer Vision, pp. 839- 846, 1998.



- [Wang et al. 2008] Y. Wang, C. Tai, O. Sorkine, T. Lee, "Optimized Scale-and-Strech for Image Resizing," in ACM Transactions on Graphics, vol.27, no.5, article 118, 2008.
- [Wang et al. 2009] Y. Wang, H. Fu, O. Sorkine, T. Lee, H. Seidel, "Motion-Aware Temporal Coherence for Video Resizing," ACM Transactions on Graphics, vol.28, no.5, article 127, 2009.
- [Woods et al. 2010] R. Woods, P. Satgunam, M. Bronstad, E. Pelli, "Statistical Analysis of Subjective Preferences for Video Enhancement," Proceedings of Human Vision and Electronic Imaging, vol.7527, pp. 75270E-75270E-10, 2010.
- [Wolf et al. 2007] L. Wolf, M. Guttman, D. Cohen-Or, "Non-Homogeneous Content-Driven Video-Retargeting," Proceedings of the 11<sup>th</sup> International Conference on Computer Vision, pp.1-6, 2007.
- [Yee et al. 2001] H. Yee, S. Pattanaik, D. Greenberg, "Spatiotemporal Sensitivity and Visual Attention for Efficient Rendering of Dynamic Environments," ACM Transactions on Graphics, vol.20, no.1, pp.39-65, 2001.
- [Zhang et al. 2009] Q. Zhang, S. Liu, P. An, Z. Zhang, "Object Segmentation Based on Disparity Estimation," Proceedings of the 1<sup>st</sup> ACM/SIGEVO Summit on Genetic and Evolutionary Computation, pp.1053-1056, 2009.

## Abbildungsverzeichnis

Abbildung 1: Stereo sehen .....	1
Abbildung 2: Parameter für 3D Signale .....	2
Abbildung 3: Disparität. ....	3
Abbildung 4: Aufnahme eines Stereobildpaars.....	3
Abbildung 5: Lage eines betrachteten Objekts in Relation zur Bildebene.....	4
Abbildung 6: Lineares Warping. ....	5
Abbildung 7: Interne und externe Kamerapositionen.....	6
Abbildung 8: Modell zu Erstellung einer Saliency Map.....	16
Abbildung 9: Saliency Measures .....	17
Abbildung 10: Lokale Kantenkarte mit Canny Kantendetektor .....	18
Abbildung 11: Globale Texturkarte .....	19
Abbildung 12: Generierte Disparitätskarte.....	20
Abbildung 13: Disparitätskarte für Usertestbild3.....	21
Abbildung 14: Kantenerkennung auf Bildern aus dem Middlebury Datenset .....	22
Abbildung 15: Erkennung von Superpixel für Image Retargeting .....	23
Abbildung 16: Saliency Map aus Meanshift als Segmentierungsalgorithmus .....	23
Abbildung 17: Mit Bilateral Filter generierte Saliency Map .....	24
Abbildung 18: Stereoskopisches Warping .....	25
Abbildung 19: Fusion.....	28
Abbildung 20: Löcher im Bild wurden mit Fusion und Inpainting gefüllt.....	29
Abbildung 21: Stereobild nach dem Warping ohne Fusion und vor Inpainting .....	30
Abbildung 22: Fehlende Farbwerte füllen.....	31
Abbildung 24: Bildausschnitt aus Abbildung 23 nach dem Inpainting.....	32
Abbildung 23: Inpainting auf Bildausschnitt.....	32
Abbildung 25: Im Usertest verwendete Testbilder.....	34
Abbildung 26: Testset 1.....	36
Abbildung 27: Testset 2.....	37
Abbildung 28: Gegenüberstellung Testbild 3.....	38
Abbildung 29: Auswertung der Ergebnisse zu Frage (a) .....	42
Abbildung 30: Testbild 1 .....	43
Abbildung 31: Bild 1, linke DK HQ.....	45
Abbildung 32: Bild 1, linke DK NQ.....	45
Abbildung 33: Testbild 2 .....	46
Abbildung 34: Bild 2, linke DK HQ.....	48
Abbildung 35: Bild 2, linke DK NQ.....	48
Abbildung 36: Testbild 3 .....	49
Abbildung 37: Bild 3, linke DK HQ.....	51
Abbildung 38: Bild 3, linke DK NQ.....	51
Abbildung 39: Testbild 4 .....	52
Abbildung 40: Bild 4, linke DK HQ.....	54
Abbildung 41: Bild 4, linke DK NQ.....	54
Abbildung 42: Ausschnitt aus linker DK HQ mit Saliency Map .....	55

Abbildung 43: Ausschnitt aus linker DK NQ mit Saliency Map .....	55
Abbildung 44: Testbild 5 .....	56
Abbildung 45: Bild 5, linke DK HQ.....	58
Abbildung 46: Bild 5, linke DK NQ.....	58
Abbildung 47: Testbild 6 .....	59
Abbildung 48: Bild 6, linke DK HQ.....	61
Abbildung 49: Bild 6, linke DK NQ.....	61
Abbildung 50: Bild 5, linke DK NQ.....	63
Abbildung 51: Auswertung der Ergebnisse zu Frage (b) .....	64

## Tabellenverzeichnis

Tabelle 1: Übersicht über die verwendeten Parameter in Saliency Maps .....	10
Tabelle 2: Beispiel für eine Usertest Vergleichsmatrix .....	39
Tabelle 3: Komponenten einer Thurstone Matrix.....	40
Tabelle 4: Frage (a) Gesamtergebnis für die Bilder B1 – B6, Position 0,0/2,0 .....	41
Tabelle 5: Frage (a) Gesamtergebnis für die Bilder B1 – B6, Position -0,5/1,5 .....	41
Tabelle 6: Reihung der Methode .....	42
Tabelle 7: Bild 1, Kameraposition 0,0/2,0.....	44
Tabelle 8: Bild 1, Kameraposition -0,5/1,5.....	44
Tabelle 9: Bild 2, Kameraposition 0,0/2,0.....	47
Tabelle 10: Bild 2, Kameraposition -0,5/1,5.....	47
Tabelle 11: Bild 3, Kameraposition 0,0/2,0.....	50
Tabelle 12: Bild 3, Kameraposition -0,5/1,5.....	50
Tabelle 13: Bild 4, Kameraposition 0,0/2,0.....	53
Tabelle 14: Bild 4, Kameraposition -0,5/1,5.....	53
Tabelle 15: Bild 5, Kameraposition 0,0/2,0.....	57
Tabelle 16: Bild 5, Kameraposition -0,5/1,5.....	57
Tabelle 17: Bild 6, Kameraposition 0,0/2,0.....	60
Tabelle 18: Bild 6, Kameraposition -0,5/1,5.....	60
Tabelle 19: Frage (b) Gesamtergebnis für die Bilder B1 – B6, Position 0,0/2,0. ....	62
Tabelle 20: Reihung der Methoden .....	62
Tabelle 21: Frage (b) Gesamtergebnis für die Bilder B1 – B6, Position -0,5/1,5. ...	63
Tabelle 22: Reihung der Methoden .....	64
Tabelle 23: Bild 1, Kameraposition 0,0/2,0.....	65
Tabelle 24: Bild 1, Kameraposition -0,5/1,5.....	66
Tabelle 25: Bild 2, Kameraposition 0,0/2,0.....	67
Tabelle 26: Bild 2, Kameraposition -0,5/1,5.....	67
Tabelle 27: Bild 3, Kameraposition 0,0/2,0.....	68
Tabelle 28: Bild 3, Kameraposition -0,5/1,5.....	69
Tabelle 29: Bild 4, Kameraposition 0,0/2,0.....	70
Tabelle 30: Bild 4, Kameraposition -0,5/1,5.....	71
Tabelle 31: Bild 5, Kameraposition 0,0/2,0.....	72
Tabelle 32: Bild 5, Kameraposition -0,5/1,5.....	72
Tabelle 33: Bild 6, Kameraposition 0,0/2,0.....	73
Tabelle 34: Bild 6, Kameraposition -0,5/1,5.....	74

## Anhang I: Auszug - Regressionsmatrizen für den zusätzlichen Faktor Sehhilfe bei der Frage nach dem Tiefeneindruck

Testbilder 3 mit zusätzlichem Faktor Sehschwäche (B: Brille, K: Kontaktlinsen, O: ohne Sehhilfe, N: Anzahl der gültigen Stimmen)

Sehhilfe	N	Generierte Disparitätskarte		Gute Disparitätskarte		Anteil
		ohne Saliency	mit Saliency	ohne Saliency	mit Saliency	
B	10	-1	1			0,600
B	10	1		-1		0,666
B	10	1			-1	0,700
B	10		-1	1		0,600
B	10		1		-1	0,555
B	10			1	-1	0,700
K	4	-1	1			1,000
K	4	-1		1		0,750
K	4	-1			1	0,750
K	4		-1	1		0,750
K	4		1		-1	0,750
K	4			-1	1	0,750
O	10	-1	1			0,700
O	10	-1		1		0,555
O	10	-1			1	0,600
O	10		-1	1		0,600
O	10		1		-1	0,600
O	10			1	-1	0,500

Testbilder 1 mit zusätzlichem Faktor Sehschwäche (B: Brille, K: Kontaktlinsen, O: ohne Sehhilfe, N: Anzahl der gültigen Stimmen)

Sehhilfe	N	Generierte Disparitätskarte		Gute Disparitätskarte		Anteil
		ohne Saliency	mit Saliency	ohne Saliency	mit Saliency	
B	9	1	-1			0,555
B	9	-1		1		0,555
B	10	-1			1	0,500
B	9		-1	1		0,888
B	10		-1		1	0,800
B	10			1	-1	0,500
K	2	1	-1			1,000
K	2	1		-1		1,000
K	4	-1			1	1,000
K	4		-1	1		1,000
K	4		-1		1	0,750
K	4			-1	1	0,750
O	8	1	-1			0,750
O	10	-1		1		0,700
O	10	-1			1	0,700
O	10		-1	1		0,900
O	10		-1		1	0,900
O	10			1	-1	0,600