

# Die Bedeutung von Alpha-Matting für die Darstellung von Bildern auf 3D-Bildschirmen: eine Benutzerstudie

DIPLOMARBEIT

zur Erlangung des akademischen Grades

**Diplom-Ingenieur/in**

im Rahmen des Studiums

**Medieninformatik**

eingereicht von

**Simone Arnberger**

Matrikelnummer 0627276

an der  
Fakultät für Informatik der Technischen Universität Wien

Betreuung  
Betreuer/in: ao. Univ. Prof. Mag. DI Dr. Margrit Gelautz  
Mitwirkung: Univ.-Ass. Mag. DI Dr. Christoph Rhemann

Wien, 20.09.2011

\_\_\_\_\_  
(Unterschrift Verfasser/in)

\_\_\_\_\_  
(Unterschrift Betreuer/in)

Simone Arnberger  
Costenoble­gasse 2/110  
1130 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.

.....

Ort, Datum

.....

Unterschrift

## Kurzfassung

Stereoskopische 3D-Bildschirme simulieren die menschliche visuelle Tiefenwahrnehmung, indem sie dem linken und dem rechten Auge gleichzeitig leicht verschiedene Bilder, die sich durch ihre horizontale Abweichung voneinander unterscheiden, präsentieren. Eine Schlüsselstelle in der 3D-Video-Kette ist daher das Erzeugen von Stereobildpaaren für die Darstellung auf 3D-Bildschirmen. Während neues stereoskopisches Videomaterial mit zwei horizontal leicht verschobenen Kameras erstellt werden kann, muss für bestehendes Videomaterial eine zweite synthetische Ansicht durch pixelweises Verschieben der Objekte entsprechend ihrer räumlichen Tiefe erzeugt werden. In Bildregionen mit transparenten Bereichen gibt es jedoch keine eindeutige Tiefenzuordnung, da diese Bildbereiche eine Überblendung von Vorder- und Hintergrund darstellen. Die daraus resultierenden Bildfehler können durch das Einbeziehen von Alpha-Matting-Information – also Kenntnis über Einfluss von Vorder- und Hintergrund an jedem Pixel – vermieden werden.

Ziel dieser Arbeit ist es zu untersuchen, inwieweit Alpha-Matting-Information die Qualität der Bilddarstellung auf 3D Bildschirmen beeinflusst. Um diese Frage zu beantworten, wurde im Zuge dieser Arbeit eine Benutzerstudie durchgeführt. In der Studie wurden mehrere Stereobildpaare, welche sowohl mit als auch ohne Alpha-Matting erstellt wurden, auf einem 3D-Monitor dargestellt. Die Betrachter wurden befragt, ob sichtbare Unterschiede in der Bildqualität zwischen den 3D-Bildern erkennbar sind. Ein weiteres Ziel der Untersuchung war die Feststellung, ob die wahrgenommenen Unterschiede mit speziellen Hintergründen oder einzelnen Motiven in Verbindung gebracht werden können. Die Analyse der Ergebnisse hat gezeigt, dass die Probanden eine bessere Bildqualität bei den mit Alpha-Matting erstellten Bildpaaren gegenüber den ohne Alpha-Matting erstellten Stereobildpaaren feststellen konnten. Das Ergebnis der Studie spricht somit eindeutig für einen Einsatz von Alpha-Matting bei der Generierung von synthetischen Ansichten, da eine korrekte Darstellung der transparenten Regionen Grundvoraussetzung für qualitativ hochwertige 3D-Bilder ist.

## **Abstract**

Stereoscopic 3D displays simulate the human visual depth perception by simultaneously showing two horizontally displaced images of the same scene to the left and the right eye. Therefore, a key factor in the 3D video chain is the generation of stereo image pairs that are fed into 3D monitors. While new stereoscopic video footage can be generated by recording a scene with two slightly displaced cameras, for existing footage a second view has to be generated synthetically. This synthetic view can be generated by displacing pixels in the existing view according to the scene depth. However, for pixels that lie on transparent foreground objects, no unique depth value can be estimated, because these image regions are a mixture of the foreground and the background. As a consequence, artifacts can occur when displaying the stereo image pair on a 3D monitor. These artifacts can be alleviated by the use of alpha matting methods, which infer the transparency of foreground at each pixel.

The main goal of this thesis is to examine to which extent transparency information influences the quality of images displayed on 3D monitors. To answer this question, a user study was carried out. During the study, stereo image pairs which were generated both with and without alpha matting were displayed on a 3D monitor. The subjects participating in the study were asked whether they can find visible differences in the quality of the displayed 3D images. A further goal of the study was to find out whether differences in 3D image quality are linked to certain backgrounds or individual images. The results of the study show that there is a substantial improvement of 3D image quality when using stereo images that were generated with alpha matting over stereo images generated without matting information. The results of the study motivate the use of alpha matting when generating synthetic views for 3D monitors, because a correct representation of transparent regions is necessary for high-quality 3D images.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b> .....	<b>7</b>
1.1	Motivation und Zielsetzung .....	7
1.2	Aufbau und Organisation .....	8
<b>2</b>	<b>Visuelle Wahrnehmung</b> .....	<b>9</b>
2.1	Aufbau und Funktion des visuellen Systems .....	9
2.2	Organisation der visuellen Information.....	11
2.2.1	Gestaltprinzipien .....	11
2.2.2	Perzeptuelle Gliederung .....	13
2.3	Objektwahrnehmung .....	13
<b>3</b>	<b>Tiefenwahrnehmung</b> .....	<b>16</b>
3.1	Okulomotorische Tiefeninformation.....	16
3.2	Monokulare Tiefeninformation .....	17
3.2.1	Tiefeninformation durch bildbezogene Hinweise .....	17
3.2.2	Tiefeninformation durch Bewegung .....	21
3.3	Binokulare Tiefeninformation.....	22
3.3.1	Querdisparität .....	23
3.3.2	Zufallspunkstereogramm.....	25
3.3.3	Halb-Verdeckung .....	26
3.3.4	Pseudoskopisches Sehen .....	28
3.3.5	Binokulare Rivalität .....	28
3.3.6	Okulare Dominanz .....	31
<b>4</b>	<b>Stereoskopie</b> .....	<b>32</b>
4.1	Stereosehen.....	32
4.1.1	Stereoblick.....	32
4.1.2	Stereoskop .....	33
4.2	3D-Bildschirmsysteme .....	35
4.2.1	Stereoskopische Bildschirmsysteme .....	36
4.2.2	Autostereoskopische Bildschirmsysteme .....	38
4.2.3	Volumetrische Systeme.....	42
4.3	3D-Video Formate.....	44
4.3.1	Basis 3D-Video Formate.....	44
4.3.2	Erweiterte 3D-Video Formate.....	47

<b>5</b>	<b>Alpha-Matting .....</b>	<b>51</b>
5.1	Das Matting-Problem .....	51
5.2	Methoden.....	52
5.2.1	Trimap .....	52
5.2.2	Closed-form Matting.....	53
5.3	Alpha Matting und Stereo Vision .....	54
<b>6</b>	<b>Benutzerstudie .....</b>	<b>58</b>
6.1	Planung.....	59
6.1.1	Auswahl der Bilder.....	60
6.1.2	Erstellung der 3D-Bilder .....	61
6.2	Vorstudie .....	63
6.2.1	Vorbereitung.....	63
6.2.2	Durchführung .....	65
6.2.3	Ergebnisse .....	66
6.3	Studie.....	67
6.3.1	Vorbereitung.....	67
6.3.2	Durchführung .....	69
<b>7</b>	<b>Ergebnisse und Auswertung.....</b>	<b>70</b>
7.1	Analyse.....	70
7.1.1	Kategorie: Bewegungsunschärfe.....	71
7.1.2	Kategorie: Tiefenunschärfe .....	76
7.1.3	Kategorie: Struktur.....	82
7.1.4	Kategorie: Transparenz .....	88
7.1.5	Kategorie: „normale“ Objekte.....	94
7.2	Schlussfolgerungen .....	99
<b>8</b>	<b>Zusammenfassung.....</b>	<b>101</b>
	<b>Literaturverzeichnis .....</b>	<b>103</b>
	<b>Abbildungsverzeichnis .....</b>	<b>106</b>

# 1 Einleitung

*The commercial success and acceptability of 3D technology will critically depend on the overall visual quality of the rendered images. [K10]*

Der menschlichen visuellen Wahrnehmung nachempfunden, werden bei der Stereoskopie dem linken und dem rechten Auge gleichzeitig leicht verschiedene Bilder präsentiert, die sich durch eine horizontale Abweichung voneinander unterscheiden. Ist das Stereobildpaar korrekt erzeugt, kann das Gehirn daraus die Tiefeninformation berechnen und es entsteht ein dreidimensionaler Eindruck. Verwendete der britische Erfinder Charles Wheatstone seinerzeit Spiegel, um eine Trennung von rechter und linker Ansicht sicher zu stellen, wird dies heute mit Hilfe spezieller Brillen erreicht. Für die nächste Generation von 3D-Technologien, die sogenannten autostereoskopischen Bildschirmsysteme, sind keine Brillen mehr notwendig. Außerdem ermöglichen sie das Betrachten einer Szene aus mehreren Blickwinkeln.

## 1.1 Motivation und Zielsetzung

Für einen dreidimensionalen Eindruck müssen zwei Bilder gleichzeitig am Bildschirm dargestellt werden. Das Übertragen von zwei oder mehr Ansichten erfordert eine größere Bandbreite als herkömmliche Fernseh- oder Videosignale. Aus der Vielzahl an unterschiedlichen 3D-Video-Formaten haben sich daher diejenigen durchgesetzt, die die erforderlichen Ansichten mittels Basisansicht und Tiefenkarte synthetisch erstellen. Zur Erstellung der zweiten Ansicht werden die Pixel der Basisansicht einzeln auf Grund der entsprechenden Werte in der Tiefenkarte verschoben. Da sich Vorder- und Hintergrund unterschiedlich stark verschieben, entstehen in der synthetischen Ansicht für gewöhnlich sichtbare Artefakte in Bildregionen mit transparenten Bereichen, da es dort zu einer Überlagerung von Vorder- und Hintergrund kommt. Transparenz entsteht nicht nur durch transparente Objekte sondern auch durch Unschärfe (Tiefenunschärfe, Bewegungsunschärfe) oder feine Struktur (Haare, Fell).

Eine mögliche Lösung des Problems ist die Verwendung mehrerer Ebenen für Farbe, Transparenz und Tiefe. Die korrekten Farb- und Transparenzwerte können mittels Alpha-Matting ermittelt werden. Ziel dieser Arbeit ist zu untersuchen, ob und in welchem Aus-

maß das Einbeziehen von Alpha-Matting-Information bei der Generierung von synthetischen Ansichten eine sichtbare Verbesserung der 3D-Bildqualität bewirkt. Im Zuge der Arbeit soll mit Hilfe einer Benutzerstudie herausgefunden werden, ob Personen einen qualitativen Unterschied zwischen 3D-Bildern wahrnehmen, wenn eines der beiden mit und das andere ohne Alpha-Matting erstellt wurde. Des Weiteren stellt sich die Frage nach Alternativen zu Alpha-Matting in Form von geschickt gewählten Grenzwerten zur Trennung von Vordergrund und Hintergrund in der Basisansicht.

## **1.2 Aufbau und Organisation**

Diese Arbeit setzt sich aus einer Literaturrecherche und einer Benutzerstudie zusammen. Die Literaturrecherche befasst sich zunächst mit der visuellen Wahrnehmungspsychologie – im Speziellen mit der Tiefenwahrnehmung beim Menschen. Danach folgt ein Überblick über gängige 3D-Technologien und verfügbare 3D-Video Formate. Abschließend wird auf ausgewählte Themen im Bereich Alpha-Matting näher eingegangen. Die Benutzerstudie zur Evaluierung der Bedeutung von Alpha-Matting für die Darstellung von Bildern auf 3D-Bildschirmen wird in Vorstudie und Hauptstudie unterteilt. Die Hauptstudie umfasst eine Befragung von Probanden zu qualitativen Unterschieden zwischen den ihnen gezeigten 3D-Bildern, die mit beziehungsweise ohne Alpha-Matting-Information generiert wurden. Die Vorstudie dient mit der Festlegung des bestmöglichen Verfahrens und der Ermittlung geeigneter Bilder dazu, die Qualität der anschließenden Hauptstudie zu optimieren. Den Abschluss der Arbeit bilden die detaillierte Analyse der gewonnenen Information und die Auswertung der Ergebnisse.

Kapitel 2 gibt einen Überblick über das menschliche visuelle System. In Kapitel 3 wird die Funktion der menschlichen Tiefenwahrnehmung erörtert. Kapitel 4 befasst sich mit den Prinzipien der Stereoskopie. Kapitel 5 beschäftigt sich mit Alpha-Matting und seiner Bedeutung für den Bereich „Stereo Vision“. In Kapitel 6 erfolgt die Planung und Durchführung der Benutzerstudie. Kapitel 7 beinhaltet die Auswertung und Präsentation der gewonnenen Erkenntnisse. In Kapitel 8 werden die Arbeit zusammengefasst und Vorschläge für weiterführende Studien gebracht.

## 2 Visuelle Wahrnehmung

Über die Sinnesorgane kann der Mensch physikalische Reize aus seiner Umwelt aufnehmen, die anschließend im Gehirn verarbeitet werden und gegebenenfalls eine Reaktion hervorrufen. Diesen Vorgang bezeichnet man als Wahrnehmung. Gemeinsam mit Gehör-, Geruch-, Geschmack- und Tastsinn erzeugt das Sehen ein Abbild der Außenwelt im Gehirn, das dem Menschen ermöglicht, sich in seiner Umwelt zu Recht zu finden. Die visuelle Wahrnehmung stellt mit der Aufnahme und Auswertung von visuellen Reizen, wie beispielsweise Helligkeit, Formen, Farbe, Bewegung oder Räumlichkeit, die wichtigste Form der Wahrnehmung für den Menschen dar.

### 2.1 Aufbau und Funktion des visuellen Systems

Grundlage des Sehens ist sichtbares Licht, das heißt elektromagnetische Energie mit einer Wellenlänge von 400nm bis 700nm, die von natürlichen oder künstlichen Quellen ausgestrahlt wird. Mit dem Eintreten eines von Objekten in der Umwelt reflektierten Lichtstrahles in das Auge beginnt der Sehprozess. Wie in Abb. 1 veranschaulicht, durchdringt das Licht die durchsichtige Hornhaut und gelangt durch die Pupille – ein kleines Loch in der Regenbogenhaut („Iris“) – auf die Linse. Die Pupille übernimmt die Funktion einer Blende, die sich je nach Lichteinfall verkleinert oder erweitert. Sowohl Hornhaut als auch Linse brechen die eintreffenden Strahlen. Zum Fokussieren auf verschiedene Entfernungen wird die Brechkraft der Linse mit Hilfe des Ziliarmuskels verändert. Die Linse ist flach, wenn sie auf ferne Objekte scharf stellt und wird runder bei nahen Objekten, die sich weniger als sechs Meter entfernt befinden. Diesen Prozess nennt man Akkommodation. So entsteht auf der Netzhaut („Retina“), die sich an der hinteren inneren Oberfläche des Auges befindet, ein auf den Kopf gestelltes, scharfes Abbild. [D06, G08]

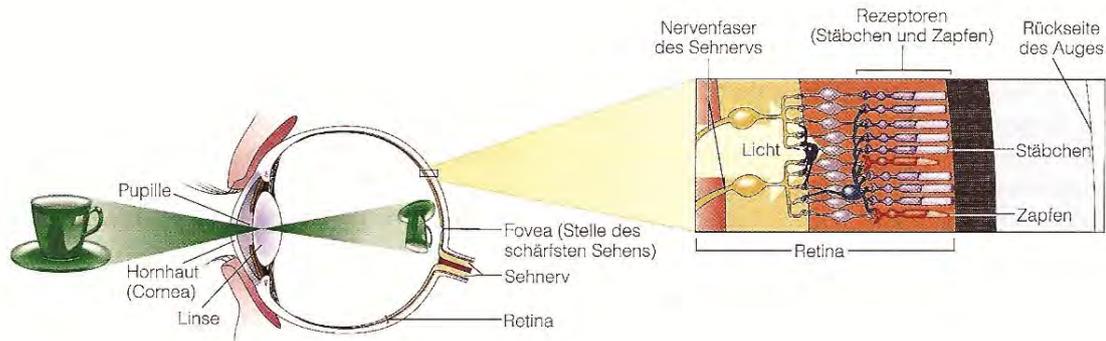


Abbildung 1: Das Auge. Abbildungsprozess einer Tasse auf die Retina (links). Vergrößerter Ausschnitt der Retina mit Rezeptoren und Neuronen (rechts). Abbildung aus [G08]

Die verkleinerte Repräsentation der gesehenen Umwelt auf der Retina wird durch lichtempfindliche Sehzellen in elektrische Energie umgewandelt. Wie der in Abb. 1 dargestellte Querschnitt der Retina zeigt, liegen diese visuellen Rezeptoren in jener Schicht der Netzhaut, die am weitesten von der Linse entfernt ist, unmittelbar vor einer absorbierenden Pigmentschicht, die interne Reflexionen reduziert. Die Sehzellen werden auf Grund ihrer Form in Zapfen und Stäbchen unterteilt. Zapfen sind tagsüber bei starkem Licht aktiv. Ihre hohe Detailauflösung ermöglicht ein sehr scharfes und farbiges Sehen. Im Gegensatz dazu arbeiten die Stäbchen ausschließlich bei schlechten Lichtverhältnissen. Sie erfassen weder Details noch Farbe, wodurch bei Dämmerlicht oder nachts nur Grautöne wahrgenommen werden können. Während sich der größte Anteil der Zapfen auf engstem Raum in der sogenannten Sehgrube („Fovea centralis“) befindet, sind die Stäbchen gleichmäßig auf der Netzhaut – mit Ausnahme der Fovea – verteilt. Die elektrischen Impulse von Zapfen und Stäbchen werden an ein Netzwerk aus Neuronen (Bipolarzellen, Horizontalzellen, Amakrinen und Ganglienzellen) weitergeleitet. Hier findet eine Vorverarbeitung der Sehinformation statt. [B04, D06, G08]

Die Nervenfasern der Ganglienzellen verlaufen an der Oberfläche der Retina zum blinden Fleck, wo sie zum Sehnerv zusammengefasst werden. An dieser Stelle gibt es keine Sehzellen. Die fehlende Sehinformation wird auf Grund der umliegenden Sehinformation ergänzt. Über den Sehnerv gelangt die visuelle Information zum Corpus geniculatum laterale im Thalamus und weiter zur Sehrinde im Hinterhauptslappen („Okzipitallappen“) des Gehirns. In beiden Stationen findet eine Verarbeitung der Sehinformation statt. Der Weg der neuronalen Signale nach Verlassen des Auges ist in Abb. 2 schematisch dargestellt. [D06, D04, G08, M75]

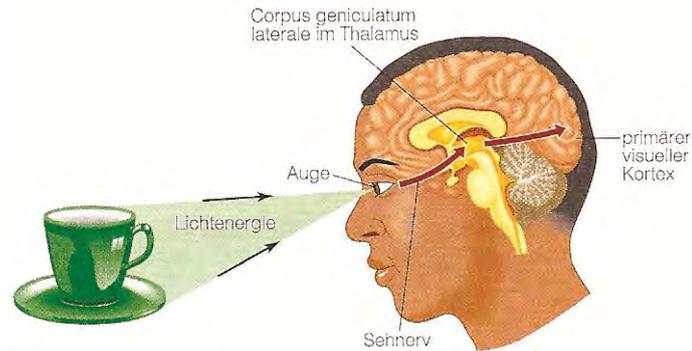


Abbildung 2: Überblick über das visuelle System. Abbildung aus [G08]

## 2.2 Organisation der visuellen Information

Die Verarbeitung des Netzhautbildes beginnt mit der Erfassung der vorkommenden Helligkeitsstufen und deren Grenzen. Die Welt wird jedoch nicht als Ansammlung von Kanten und Flecken wahrgenommen, sondern als eine Anordnung von Oberflächen und Objekten. Wie diese Ordnung erreicht wird und zusammengehörige Teile als solche erkannt werden, wird im folgenden Abschnitt beschrieben. [B04]

### 2.2.1 Gestaltprinzipien

Die Organisation der menschlichen Wahrnehmung wird mit Hilfe der Gestalttheorie beschrieben. Diese besagt, dass die Wahrnehmung des Ganzen mehr als eine Summe oder Folge von Teilwahrnehmungen ist. Die Gestaltpsychologen Max Wertheimer, Kurt Koffka und Ivo Köhler haben daraus eine Reihe von Gesetzen der Wahrnehmungsorganisation abgeleitet, die beschreiben, warum manche Wahrnehmungen wahrscheinlicher auftreten als andere. Heute spricht man in der Wahrnehmungspsychologie eher von Gestaltprinzipien, da man der Meinung ist, dass die aufgestellten Regeln zu unpräzise Voraussagen treffen, um sie als Gesetze anzusehen. [B04, G08]

**Prinzip der Prägnanz:** Das Prinzip der Prägnanz wird als zentrales Prinzip der Gestaltpsychologie gesehen. Es besagt, dass aus einer Vielzahl an Möglichkeiten jene geometrische Anordnung erscheinen wird, die am besten, einfachsten und stabilsten ist. Das Prägnanzprinzip wird auch Prinzip der Einfachheit oder Prinzip der guten Gestalt genannt. Der Begriff „gut“ umfasst dabei Eigenschaften wie beispielsweise Symmetrie, Regelmäßigkeit und Geschlossenheit. Daher werden die vier Punkte in Abb. 3a als Quadrat gesehen und nicht als „X“ oder Dreieck mit einem zusätzlichen Punkt.

**Prinzip der Nähe:** Der Faktor der Nähe ist für die Wahrnehmungsorganisation sehr bedeutend. So werden Objekte, die nahe beieinander sind, zusammen gefasst. In der Abb. 3b werden Spalten wahrgenommen, weil die Punkte horizontal weiter auseinander liegen als vertikal. Nähe ist bei der Tiefenwahrnehmung ein wichtiger Organisationsfaktor. Das Quadrat in der Mitte eines Zufallspunkstereogramms (siehe Kapitel 3) ist erst sichtbar, wenn das Stereobildpaar mit einem Stereoskop betrachtet wird. Punkte mit denselben Disparitätswerten werden gruppiert und das Quadrat erscheint vom Hintergrund abgehoben.

**Prinzip der Ähnlichkeit:** Das Prinzip der Ähnlichkeit sagt, dass ähnliche Objekte als zusammengehörig gesehen werden. Die einzelnen Elemente können sich in Form, Farbe, Größe oder Orientierung ähnlich sein. Ein Beispiel für Farbe als Ähnlichkeitsfaktor sind die Kreise in Abb. 3c, die in Spalten gruppiert werden.

**Prinzip des guten Verlaufs:** Abb. 3d verdeutlicht das Prinzip des guten Verlaufs. Die Figur besteht eher aus zwei durchgängigen Linien, die sich bei der Markierung kreuzen, als zwei „V-Formen“ die sich bei der Markierung berühren. Das Prinzip wird deshalb auch das Prinzip der Fortsetzung oder Prinzip der fortgesetzt durchgehenden Linie genannt.

**Prinzip des gemeinsamen Schicksals:** Das Prinzip des gemeinsamen Schicksals besagt, dass Dinge, die sich gleich schnell und in die gleiche Richtung bewegen, zu einem Ganzen zusammengefasst werden. Am besten wird dieses Prinzip an Hand von Vogel- oder Fischschwärmen verdeutlicht, wo hunderte Tiere als eine Einheit wahrgenommen werden. Abb. 3e zeigt, wie sich ein Teil der Vögel vom Schwarm trennt und so eine neue Gruppe bildet.

**Prinzip der Erfahrung:** Auf Grund von Vorwissen und Erfahrung werden die zwei gebrochenen Linien in Abb. 3f zu einem schattenwerfenden „T“ ergänzt. Ein Analphabet könnte den Umriss nicht auf diese Weise vervollständigen, da ihm die Kenntnis des Buchstaben fehlt. [B04, D06, G08, M75]

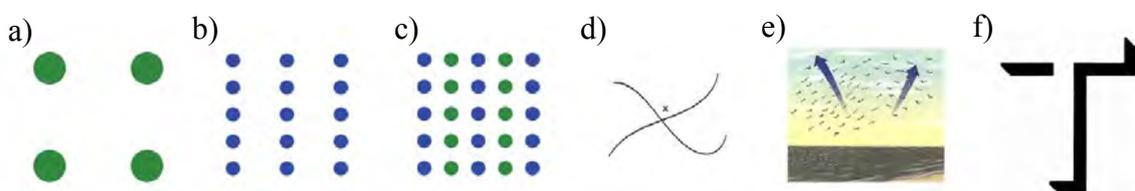


Abbildung 3: Gestaltprinzipien. a) Prägnanz. Abbildung nach [B04]; b) Nähe. Abbildung nach [B04]; c) Ähnlichkeit. Abbildung nach [B04]; d) Guter Verlauf. Abbildung aus [B04]; e) Gemeinsames Schicksal. Abbildung aus [G08]; f) Erfahrung. Abbildung aus [D06]

### **2.2.2 Perzeptuelle Gliederung**

Während die Gestaltprinzipien beschreiben, wie einzelne Teile zu einem Größeren zusammengefasst werden, befasst sich die perzeptuelle Gliederung mit dem Trennen einzelner Objekte von der gesamten Szene. Es ist vollkommen selbstverständlich die Umgebung als Zusammensetzung mehrerer, unterschiedlich großer Objekte vor einem Hintergrund zu sehen. Im Normalfall können die Objektgrenzen ohne Schwierigkeiten erkannt werden und es besteht kein Zweifel darüber, welche Regionen die Vordergrundflächen, die sogenannten Figuren, darstellen und welche Bereiche dem Hintergrund angehören. Diese Problemstellung wird daher auch Figur-Grund-Trennung genannt. Dieses Teilgebiet der Gestaltpsychologie befasst sich mit der Frage, welche Eigenschaften eine Figur oder der Grund aufweisen muss, um als solches wahrgenommen zu werden. Untersuchungen haben gezeigt, dass Areale im unteren Bereich einer Darstellung häufiger als Figur gesehen werden, als Regionen im oberen Bereich. Im Gegensatz dazu ist bei links und rechts Trennung keine Präferenz festzustellen. Bei ambivalenten, also mehrdeutigen Bildern, wie beispielsweise der Rubin-Kelch oder die „Ente-Hase“-Grafik von Jastrow, spielt diese Erkenntnis eine wesentliche Rolle. Weitere vier Faktoren – Symmetrie, Größe, Ausrichtung und Bedeutung – erhöhen die Wahrscheinlichkeit einer Entscheidung zu Gunsten der Figur. Als Vordergrund werden bevorzugt symmetrische, kleinflächigere, horizontal oder vertikal ausgerichtete und bedeutungshaltige Regionen wahrgenommen. [B04, D06, G08, M75]

## **2.3 Objektwahrnehmung**

Die Wahrnehmung von Objekten ist sowohl im Bereich der Wahrnehmungspsychologie als auch im Bereich maschinelles Sehen („Computer Vision“) ein zentrales Thema. Während Menschen ohne viel nachzudenken entscheiden können, um welches Objekt es sich handelt, hat ein Computer mit dieser Aufgabe wesentliche Schwierigkeiten. Grundsätzlich besitzt der Mensch zusätzlich zu den Rezeptoren und neuronalen Schaltkreisen auch Wissen über die Eigenschaften der Umwelt. Diese Intelligenz verschafft ihm bei der Objektwahrnehmung einen großen Vorteil gegenüber dem Computer. Menschen sind beispielsweise in der Lage unscharfe oder teilweise verdeckte Objekte in einer Szene einfach und ohne bewusste Anstrengung zu erkennen, während Computer in vielen Fällen an dieser Aufgabe scheitern. [G08]

Ein weiteres perzeptuelles Problem ist die Mehrdeutigkeit. Mehrdeutige Information an den Rezeptoren entsteht, wenn ein Objekt nur aus einer Perspektive wahrgenommen wird. Abb. 4 verdeutlicht das sogenannte Problem der inversen Projektion. Die Abbildung des Quadrats auf der Retina, die von dem kleinen Quadrat erzeugt wird, kann sowohl von dem abgebildeten gedrehten Trapez oder dem weiter entfernten größeren Quadrat als auch von unendlich vielen weiteren Objekten stammen. [G08]

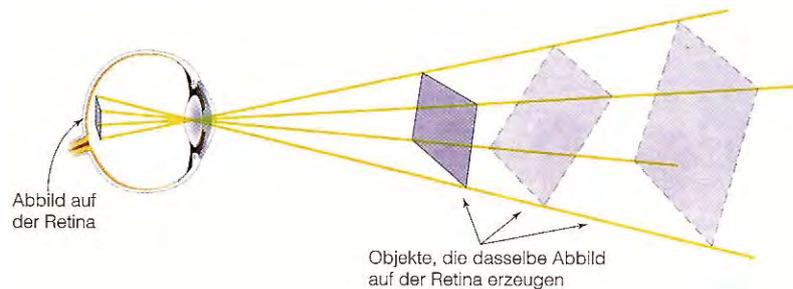


Abbildung 4: Mehrdeutigkeit der Abbildung eines Objekts auf der Retina. Abbildung aus [G08]

Das menschliche Gehirn wählt üblicherweise die nahegelegenste, einfachste Deutung aus. Diese unbewusste Entscheidung kann durch Vorwissen oder Prägung beeinflusst werden. Die mehrdeutige, weiße Umrisslinie in Abb. 5 wird deshalb in Verbindung mit einem wasserähnlichen Muster als Delfin und mit einer Steppenlandschaft im Hintergrund als Känguru wahrgenommen. [D06]

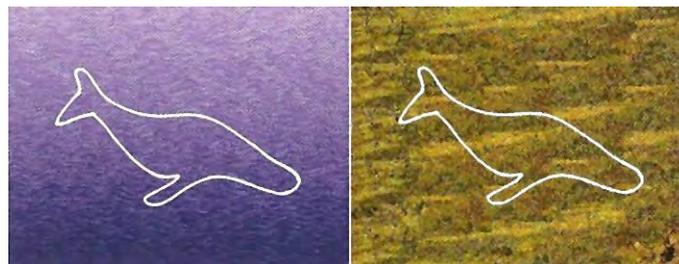


Abbildung 5: Prägung. Delfin oder Känguru? Abbildung aus [D06]

Die Blickwinkelinvarianz bezeichnet die menschliche Fähigkeit, ein Objekt aus einer Vielzahl unterschiedlicher Blickwinkel als dasselbe zu identifizieren. Im Gegensatz zum Menschen ist das Wahrnehmen von Objekten aus verschiedenen Perspektiven für den Computer nicht offensichtlich, da sich die Abbilder der Objekte stets verändern. Es gibt zwei unterschiedliche Ansätze, wie Menschen dreidimensionale Objekte erkennen. Der Objekterkennungsprozess geschieht gemäß der Theorie der kanonischen Ansichten durch den Ver-

gleich der aktuellen Wahrnehmung mit zweidimensionalen Bildern des Objekts aus verschiedenen Blickwinkeln, die im Gedächtnis gespeichert sind. Der andere Ansatz beschreibt das Erkennen dreidimensionaler Objekte als grobe Ansammlung elementarer Teilkörper. Laut David Marr sind diese volumetrischen Merkmale einfache geometrische Körper wie beispielsweise Zylinder. Ein Beispiel für die Repräsentation einer Person ist in Abb. 6 zu sehen.

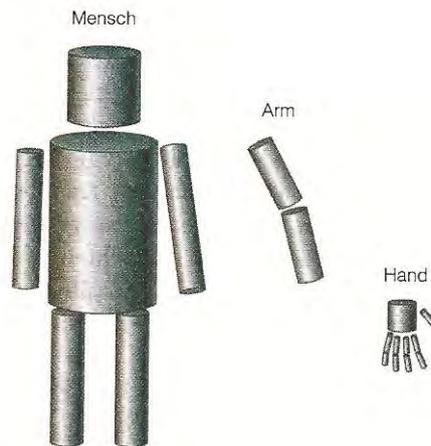


Abbildung 6: Modell nach David Marr. Abbildung aus [G08]

Irving Biedermann erweitert dieses System indem er nicht nur Zylinder sondern auch andere Formen, wie unter anderem Quader oder Pyramide, als Grundbausteine zulässt. Diese Grundbausteine nennt er Geons – einer Verknüpfung der Worte „geometric icons“. Einige Geons und daraus konstruierte Objekte sind in Abb. 7 zu sehen. [B04, G08]

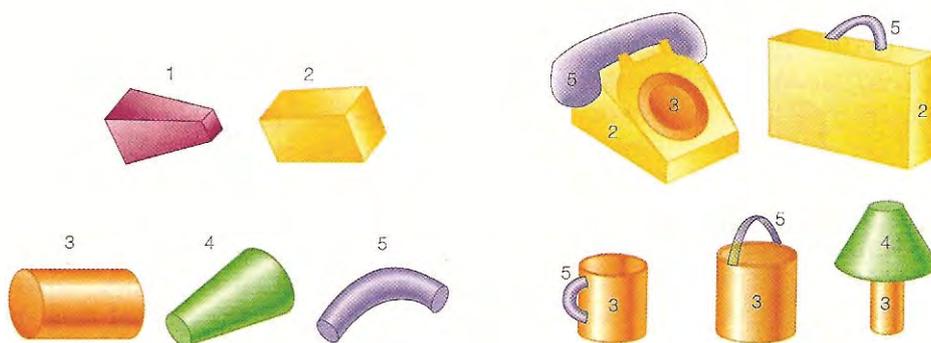


Abbildung 7: Geons nach Irving Biedermann. Abbildung aus [G08]

### 3 Tiefenwahrnehmung

Obwohl der mit den Augen wahrgenommene Ausschnitt der Umwelt als zweidimensionales Bild auf der Netzhaut dargestellt wird, nimmt man eine dreidimensionale Welt wahr. Sowohl für Menschen als auch Tiere ist die Fähigkeit sagen zu können, wie weit Dinge entfernt sind, wichtig. Den Menschen steht zum Erkennen ihrer Lage im Raum eine Vielzahl an Möglichkeiten zur Verfügung. Diese Strategien basieren auf der Identifikation von Hinweisen im Abbild auf der Netzhaut, die Aufschluss über die räumliche Anordnung geben. Des Weiteren ermöglicht die Schätzung der relativen Tiefe die Bestimmung der Form eines dreidimensionalen Objekts. So erzeugen beispielsweise kreisförmige Objekte immer elliptische Abbilder, die jedoch problemlos als kreisförmig interpretiert werden. [B04]

#### 3.1 Okulomotorische Tiefeninformation

Okulomotorische Tiefeninformation basiert auf Konvergenz, Divergenz und Akkommodation der Augen. Das Ausmaß des nach innen Drehens („Konvergenz“) beziehungsweise des nach außen Drehens („Divergenz“) der Augen und der Verformung der Linse beim Fokussieren („Akkommodation“) auf sehr nahe Objekte kann mit Tiefe verbunden werden. Die Position der Augen sowie die Spannung der Ziliarmuskeln geben dabei Rückmeldung über die Distanz des betrachteten Objekts. Die unterschiedliche Winkelstellung der Augen beim Fokussieren auf nahe beziehungsweise ferne Objekte ist in Abb. 8 zu sehen.

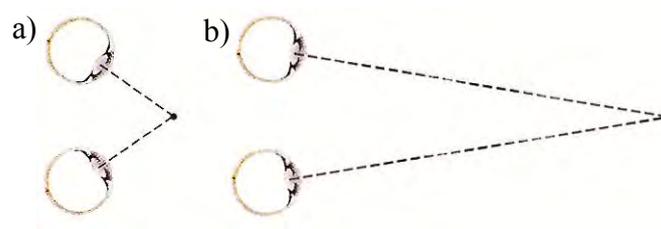


Abbildung 8: Konvergenz. a) nahes Objekt; b) fernes Objekt. Abbildung aus [G08] nach [B04]

Da der Grad von Konvergenz beziehungsweise Divergenz und Akkommodation nur ein einzelner Wert ist, können diese Tiefenhinweise zu jedem Zeitpunkt Information über nur eine Entfernung bereitstellen. Des Weiteren ist der Nutzen der beiden Tiefenreize auf eine Distanz von ungefähr einer Armlänge begrenzt. Wegen dieser Einschränkungen spielen

Konvergenz, Divergenz und Akkommodation nur eine untergeordnete Rolle für die Gewinnung der Tiefeninformation. [B04, D06, G08, H95]

## 3.2 Monokulare Tiefeninformation

Monokulare Tiefenhinweise ermöglichen das Erfassen von räumlicher Tiefe mit nur einem Auge. Diese Strategie stellt sicher, dass sich auch einäugige Menschen in ihrem dreidimensionalen Umfeld zurechtfinden können. Monokulare Tiefenreize sind außerdem für die räumliche Wahrnehmung bei größeren Distanzen ausschlaggebend. [B04, D06]

### 3.2.1 Tiefeninformation durch bildbezogene Hinweise

Bildbezogene Tiefenhinweise stellen Information über die dreidimensionale Anordnung in zweidimensionalen Bildern, wie beispielsweise Standardfernsehbildern oder dem Abbild auf der Netzhaut, zur Verfügung. Bereits seit der Renaissance ist man sich dieser Art der Tiefenreize bewusst, denn sie wurden damals bereits in der Malerei zur Vermittlung von Tiefeneindrücken verwendet. [B04]

#### 3.2.1.1 Verdeckung

Die meisten natürlichen Szenen sind überhäuft von einer Vielzahl an Objekten, die sich in unterschiedlichen Distanzen befinden. Die Sicht auf ein Objekt ist oft durch ein näher gelegenes Objekt teilweise oder gänzlich verdeckt. Das menschliche Wahrnehmungssystem erhält die Tiefeninformation durch das Vergleichen, welches Objekt welches andere überdeckt. Die verdeckten Objekte werden gemäß der Gestaltprinzipien und der Erfahrung ergänzt. Abb. 9 zeigt, dass Verdeckung bei Entfernungen von wenigen Zentimetern bis zu mehreren Kilometern genutzt werden kann.

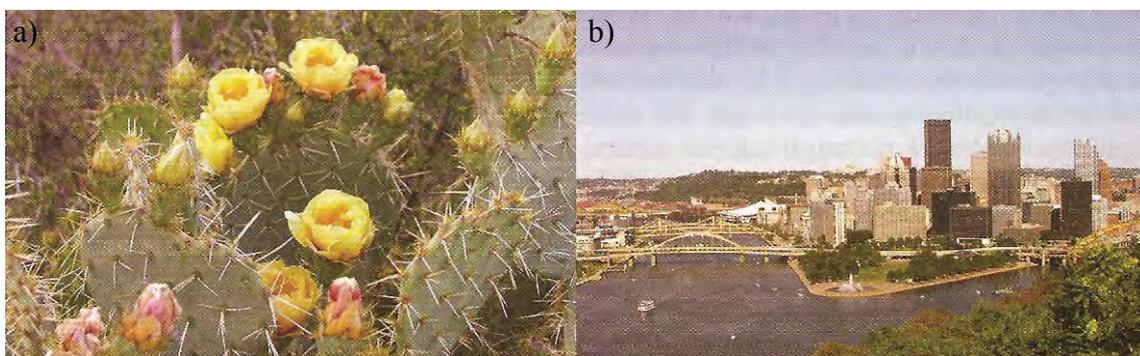


Abbildung 9: Verdeckung. a) kleine Entfernung; b) große Entfernung. Abbildung aus [G08]

Eine Sonderstellung nimmt das Erkennen von Transparenz ein, da das durchsichtige Objekt logischerweise immer vor den durchschimmernden Objekten liegen muss. Auf Grund dieser Erkenntnis kann man beispielsweise einer Fensterscheibe, einem Aquarium oder einem Trinkglas eine Tiefe zuordnen. Allgemein ist jedoch zu beachten, dass Verdeckung zwar Information über die Tiefenanordnung von Objekten enthält, aber kein Maß an relativer oder absoluter Entfernung darstellt. [B04, D06, G08]

### 3.2.1.2 Perspektive

Der bekannteste Tiefenreiz dieser Art ist die lineare Perspektive. Zwei parallele Linien, wie zum Beispiel Eisenbahnschienen oder Straßenmarkierungslinien, nähern sich mit zunehmender Entfernung an, bis sie sich schließlich scheinbar im Unendlichen treffen. Diese perzeptuelle Annäherung bezeichnet man auch als perspektivische Konvergenz. Eine weitere perspektivische Quelle der Tiefenwahrnehmung ist der Texturgradient. Ähnliche, gleich große Elemente, die sich in gleichen Abständen zueinander befinden, erscheinen mit zunehmender Distanz kleiner und dichter beieinander, wie beispielsweise der ausgetrocknete Erdboden in Abb. 10. [B04, G08]



Abbildung 10: Texturgradient. Abbildung aus [G08]

Ein ebenso perspektivischer Tiefenhinweis ist der Größenvergleich. Die Abbildung eines Objekts ist kleiner, wenn sich der Gegenstand weit vom Gesichtspunkt entfernt befindet, und wird größer, wenn er sich dem Auge nähert. Die Motorräder oder Telefonmasten sind als Beispiele dafür in Abb. 11 zu finden. Voraussetzung für das Abschätzen der Tiefe ist dabei das Wissen über die Größe der einzelnen vorkommenden Objekte und das Stimmen der Proportionen. Bereits einfache, kleine Abweichungen von der Alltagserfahrung bereiten dem visuellen System Schwierigkeiten bei der Gewinnung der Tiefeninformation. [B04, D06, G08]

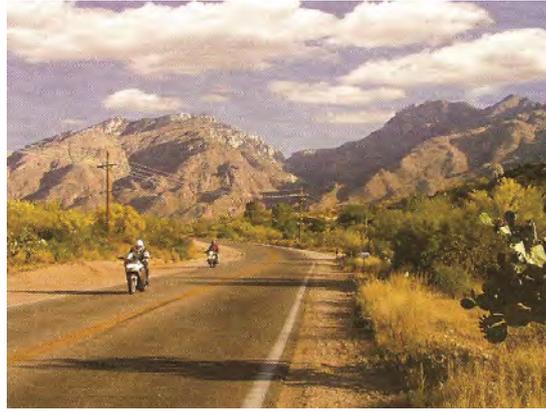


Abbildung 11: Größenvergleich. Abbildung aus [G08]

### 3.2.1.3 Atmosphärische Perspektive

Auf Grund der atmosphärischen Perspektive werden Objekte mit zunehmender Distanz nicht nur kleiner, sondern auch weniger klar und leicht bläulich wahrgenommen. Ursache dafür ist die Absorption und Streuung des Lichts an Partikeln, wie zum Beispiel Staub oder Wassertröpfchen, in der Luft. Die Hügel im Vordergrund in Abb. 12 erscheinen detailreich und scharf, während die entfernteren Hügel immer verwuschener wirken bis sie am Ende der Bucht fast mit dem Meer verschmelzen. [B04, G08]



Abbildung 12: Atmosphärische Perspektive. Abbildung aus [G08]

### 3.2.1.4 Schattierung

Schattierung entsteht durch unterschiedliche Ausleuchtung einer Oberfläche in Abhängigkeit von der Position der Lichtquelle. Die Grundannahme über die visuelle Welt ist, dass das Licht (Sonne) von oben kommt. Da dies auch auf die meisten künstlichen Lichtquellen zutrifft, kann das menschliche visuelle System von dem Beleuchtungsmuster auf die drei-

dimensionale Gestalt eines Objekts schließen. Abb. 13 zeigt den Einfluss der Schattierung auf die Wahrnehmung der Form.

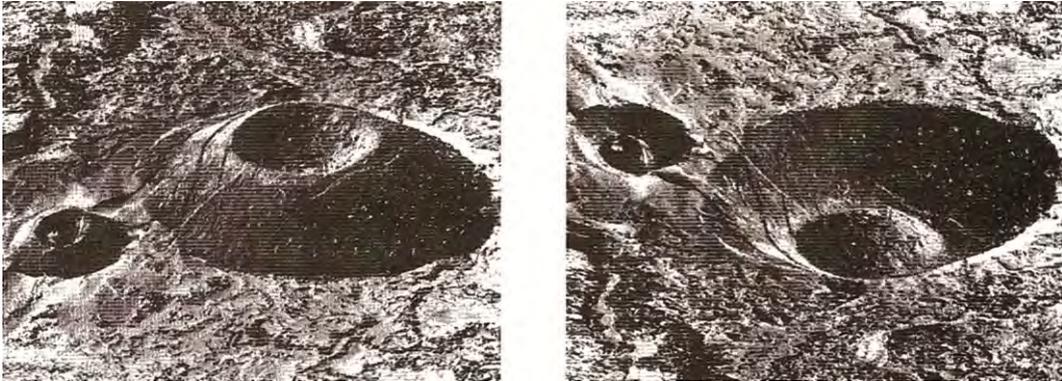


Abbildung 13: Schattierung. Abbildung aus [W95]

Das linke Bild zeigt zwei Lava-Kegel mit Kratern. Das rechte Bild scheint Krater mit Erdhügeln zu enthalten. Dennoch sind beide Bilder gleich abgesehen von einer horizontalen Spiegelung. Dies kann überprüft werden, indem man diese Arbeit um  $180^\circ$  dreht. [B04, G08, W95]

### 3.2.1.5 Schatten

Schatten werden von einem Objekt auf ein anderes oder auf den Boden geworfen, wenn ein Objekt die Lichtstrahlen blockiert. Schatten sind dunkel, transparent, manchmal verschwommen. Sowohl Form als auch Bewegung sind vom schattenwerfenden Objekt abhängig. Obwohl Schatten keine Objekte sind und für gewöhnlich kaum Aufmerksamkeit auf sich ziehen, spielen sie eine wesentliche Rolle bei der Tiefenwahrnehmung. Abb. 14 verdeutlicht den Zusammenhang zwischen Schatten und Tiefe.

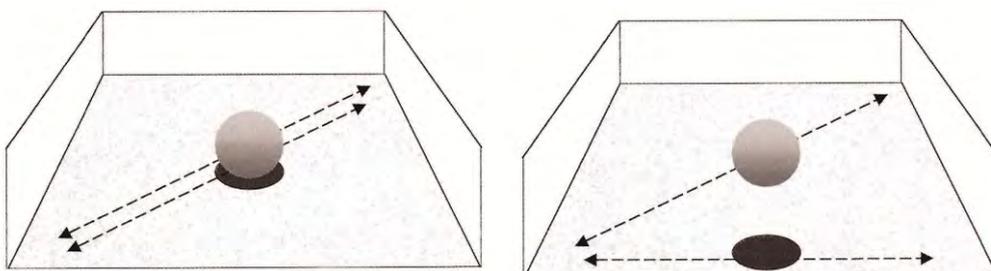


Abbildung 14: Schatten. Abbildung aus [B04]

Der graue Ball und die dunkle Ellipse bewegen sich entlang der eingezeichneten Pfeile. Obwohl sich der Ball immer gleich bewegt, nimmt der Betrachter unterschiedliche Bewegungen des Balles im Raum wahr. [B04, D06, G08]

### 3.2.2 Tiefeninformation durch Bewegung

Die Bewegung eines Beobachters oder Objekts unterstützt die Wahrnehmung der zuvor beschriebenen, statischen Tiefenhinweise. Information über die relative Distanz wird mittels Bewegungsparallaxe erzeugt. Bei Trickfilmen und Videospielen wird fast ausschließlich mit diesem Tiefenreiz gearbeitet, um einen Eindruck von räumlicher Tiefe zu vermitteln.

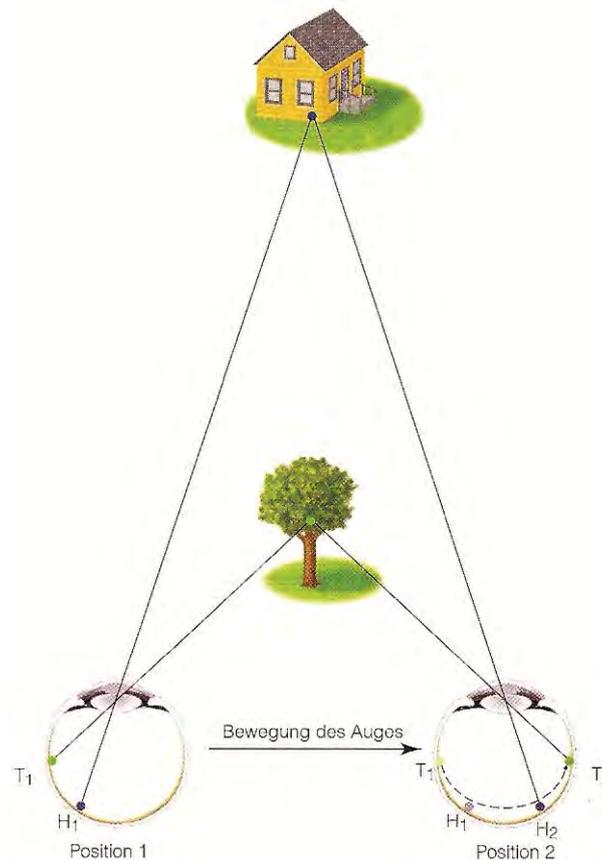


Abbildung 15: Bewegung. Abbildung aus [G08]

Ein Auge sieht zwei Objekte in unterschiedlicher Entfernung, wie beispielsweise einen Baum und ein Haus in Abb. 15. Bewegt sich das Auge nun von Position 1 zu Position 2, wandert gemäß der Bewegungsparallaxe die Abbildung des näheren Objekts (Baum) weiter über die Netzhaut, als die des weiter entfernten Objekts (Haus). Der Baum, dessen Abbild eine weitere Strecke auf der Retina zurücklegt, zieht daher scheinbar rascher am Betrachter vorbei, während sich das Haus – gleich wie sein Abbild auf der Retina – viel weniger bewegt.

Ein weiterer Tiefenreiz, der auf Grund von Bewegung entsteht, ist das Auf- und Zudecken von Flächen eines (teilweise) verdeckten Objekts. Im vorherigen Beispiel würde der Baum bei Bewegung des Betrachters immer andere Bereiche des Hauses freigeben beziehungsweise verdecken. Die Kombination aus Bewegungsparallaxe und Überlappung – also der scheinbar unterschiedliche Bewegungsumfang sich überlappender Oberflächen – ist sehr effektiv beim Erkennen von Tiefe anhand von Kanten. [G08, B04]

### **3.3 Binokulare Tiefeninformation**

Mit Ausnahme von Konvergenz und Divergenz, erfordern die zuvor beschriebenen Tiefenhinweise nur die Beteiligung von einem Auge. Eine weit wichtigere und zuverlässigere Quelle ist das Stereosehen, für das beide Augen notwendig sind. Binokulare Tiefeninformation basiert auf der Tatsache, dass jedes Auge auf Grund des horizontalen Abstands voneinander einen leicht verschiedenen Ausschnitt der Welt sieht. Bei einem durchschnittlichen Erwachsenen beträgt dieser Abstand ungefähr sechs Zentimeter. Durch den Vergleich des linken und rechten Netzhautbildes kann das Gehirn die relative Tiefe mit großer Genauigkeit schätzen. Die Frage, wie das visuelle System die Abbilder des linken und des rechten Auge richtig überlagert, wird als Korrespondenzproblem bezeichnet und kann bis jetzt nicht ausreichend beantwortet werden.

Grundvoraussetzung für das Stereosehen ist eine gekoppelte Bewegung der beiden Sehachsen, also das Einstellen der Augen auf denselben Beobachtungspunkt. Nicht alle Menschen verfügen über diese Abstimmung der Augen. Personen, die schielen („Strabismus“), können die beiden Netzhautbilder nicht zu einer Tiefenwahrnehmung zusammenfügen, weil diese zusätzlich eine unnatürliche, vertikale Verschiebung aufweisen. Neben einigen normalsichtigen Personen können weder Menschen, die auf einem Auge blind sind, noch Personen mit sehr unterschiedlicher Sehstärke der Augen dreidimensional sehen. [H95, B04]

### 3.3.1 Querdisparität

*Die Querdisparität (auch: binokulare Disparität) ist der Unterschied zwischen den Abbildern in linkem und rechtem Auge. [G08]*

Ein simples Experiment, für das man einen Finger und ein etwas weiter entferntes Objekt benötigt, veranschaulicht diesen Unterschied. Der Finger wird nun so vor das Gesicht gehalten, dass er bei geschlossenem rechtem Auge das entfernte Objekt (teilweise) verdeckt. Schließt man nun das linke Auge, verändert sich die Position des Fingers relativ zu dem entfernten Objekt. Abb. 16 veranschaulicht, wie es zu dieser scheinbaren Bewegung des Fingers kommt.

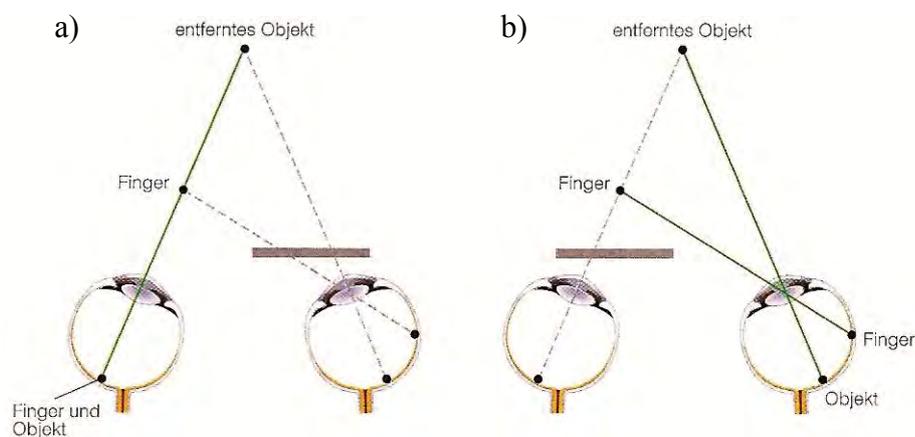


Abbildung 16: Das „Finger-Experiment“. Abbildung aus [G08]

Durch das direkte Ansehen beider Objekte bei geschlossenem rechtem Auge, fallen ihre Abbilder auf die Fovea. Dies ist durch die grüne Linie in Abb. 16a hervorgehoben. Das Abbild des betrachteten Objekts fällt bei geschlossenem linkem Auge weiterhin auf die Fovea, das Abbild des Fingers jedoch an einen seitlich versetzten Ort, wie an den grünen Linien in Abb. 16b zu sehen ist.

Unter korrespondierenden Netzhautpunkten versteht man Orte auf der Retina, die sich überlagern würden, wenn man die beiden Netzhäute auf einander legen könnte. Angenommen der Bademeister in Abb. 17 sieht Frieda direkt an, so fällt ihr Abbild auf die Fovea (F), die korrespondierende Netzhautpunkte darstellen. Ebenso fallen die Abbilder von Harry (H) und Susan (S) auf korrespondierende Netzhautpunkte. Daraus lässt sich schließen, dass die Abbilder aller Objekte, die sich in derselben Entfernung wie das betrachtete Objekt befinden, auf korrespondierende Netzhautpunkte fallen. Dieser gedachte Kreis, dessen Mittelpunkt genau zwischen den beiden Augen des Beobachters liegt, wird als Horopter bezeichnet.

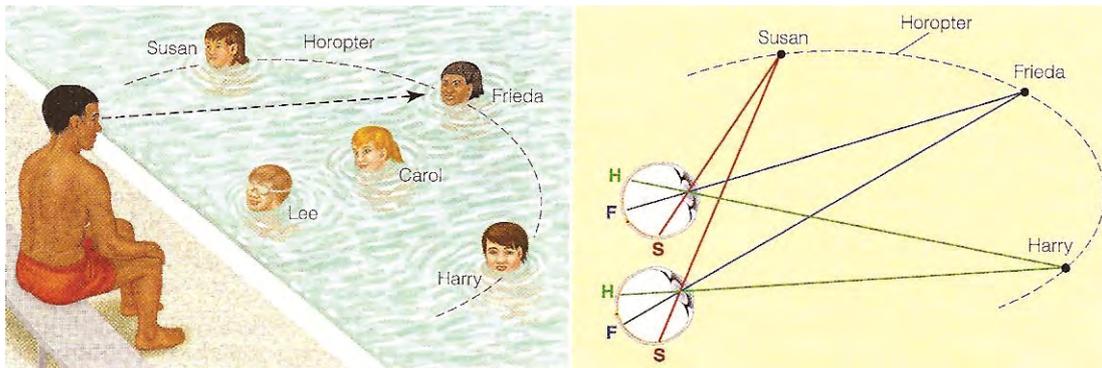


Abbildung 17: Horopter. Abbildung aus [G08]

Die in Abb. 18 mit C und L markierten Abbilder von Carol und Lee im linken Auge fallen nicht auf die korrespondierenden Netzhautpunkte C' beziehungsweise L' im rechten Auge, sondern auf die Punkte X und Y, da sich Carol und Lee nicht auf dem Horopter befinden. Die Abbilder von Carol und Lee fallen somit auf nichtkorrespondierende Netzhautpunkte. Der Abstand zwischen den korrespondierenden Netzhautpunkten C' und X beziehungsweise L' und Y wird als Querdisparität bezeichnet.

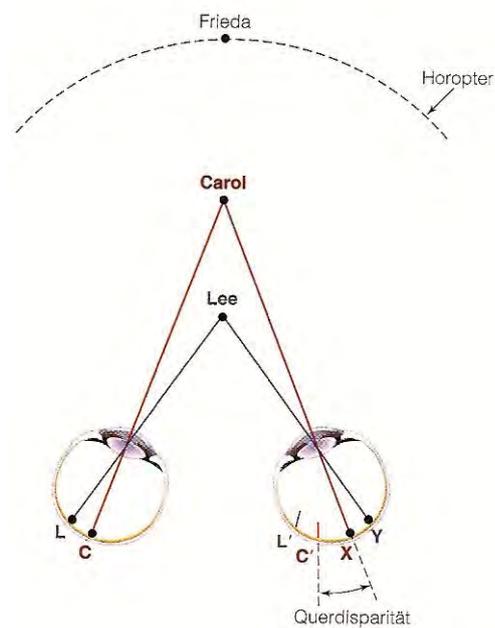


Abbildung 18: Querdisparität. Abbildung aus [G08]

Der Fixationspunkt bestimmt die Position des Horopters und hat daher immer die Querdisparität 0. Die Querdisparität wird mit zunehmendem Abstand eines Objekts vom Horopter größer. Das ist auch an Hand des vorherigen Beispiels von Carol und Lee ersichtlich. Querdisparität besitzt allerdings nicht nur eine Größe, sondern auch ein Vorzeichen. Ist ein Objekt vom Betrachter aus gesehen hinter dem Horopter, so befinden sich ihre Abbilder weiter innen auf der Retina. Man spricht dabei von einer ungekreuzten Querdisparität. Im Gegensatz dazu liegen bei der gekreuzten Querdisparität die Objekte vor dem Horopter und ihre Abbilder weiter außen auf der Netzhaut. Das Prinzip von ungekreuzter und gekreuzter Querdisparität ist in Abb. 19 schematisch dargestellt. [G08]

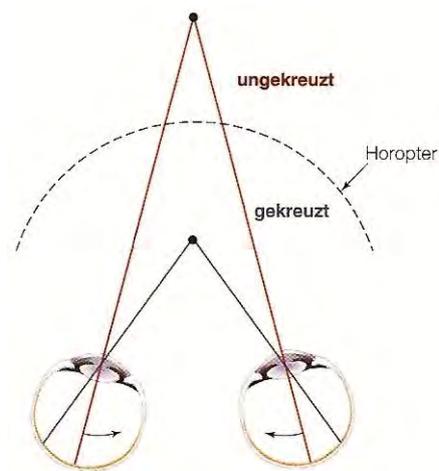


Abbildung 19: Ungekreuzte und gekreuzte Querdisparität. Abbildung aus [G08]

### 3.3.2 Zufallspunktstereogramm

Lange Zeit wurde vermutet, dass auf Grund der geschätzten Oberflächen- und Kanteninformation in den monokularen Bildern die Querdisparität und somit die Stereotiefe berechnet wird. Um diese Annahme zu widerlegen entwickelte Bela Julesz das Zufallspunktstereogramm. Ein Zufallspunktstereogramm besteht zunächst aus zwei identischen Zufallsmustern. In dem rechten Bild wird dann beispielsweise ein quadratischer Ausschnitt der Punkte horizontal verschoben und die dadurch entstandene Lücke mit Zufallspunkten aufgefüllt. Abb. 20 zeigt ein Beispiel für ein Zufallspunktstereogramm mit Herstellungsmuster.

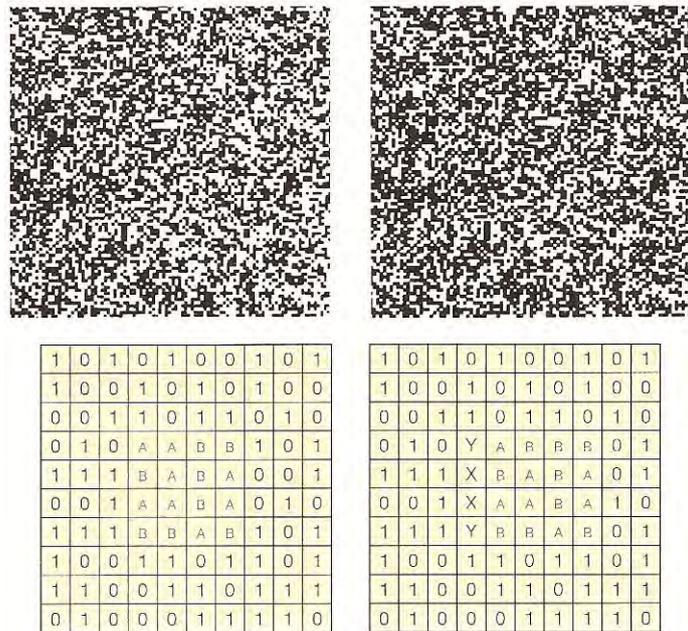


Abbildung 20: Zufallspunktstereogramm. Abbildung aus [G08]

Präsentiert man nun dem linken Auge nur das linke Bild und dem rechten Auge nur das Rechte, so erscheint das Quadrat in einer anderen Tiefenebene als die Grundfläche. Objekte können also in Stereotiefe gesehen werden, obwohl in den monokularen Bildern keine Objektkanten zu sehen sind. Zum einen kann daraus geschlossen werden, dass es für das Stereosehen nicht notwendig ist in den Netzhautbildern Strukturen und Formen zu erkennen. Zum anderen wird damit gezeigt, dass das menschliche Wahrnehmungssystem räumliche Tiefe vor Gestalt und Form erkennen kann. [G08, D06, B04]

### 3.3.3 Halb-Verdeckung

Die Tatsache, dass ein Beobachter in Zufallspunktstereogrammen eine Raumtiefe wahrnimmt, bedeutet jedoch nicht, dass andere Tiefenhinweise keinen Einfluss auf das Stereosehen haben. Speziell Verdeckung und Querdisparität werden als sich ergänzende Tiefeninformationsquellen gesehen. Die Analyse von Halb-Verdeckung spielt deshalb auch beim binokularen Tiefensehen eine wichtige Rolle.

Betrachtet man ein Objekt in einer natürlichen Szene, so gibt es Regionen, die von beiden Augen, von einem Auge oder gar nicht gesehen werden. Jene Bereiche, die nur von einem Auge erfasst werden, nennt man halb-verdeckte Regionen. Außerdem gibt es hinter einem kleinen Objekt sowohl einen kleinen vollständig verdeckten Bereich, als auch eine andere Region, die für beide Augen sichtbar ist. Fixiert man einen Punkt in dieser weit entfernten

Region, so erscheint das kleine Objekt im Vordergrund effektiv durchsichtig. Die Beziehung zwischen Verdeckung, Halb-Verdeckung und Transparenz ist in Abb. 21 dargestellt.

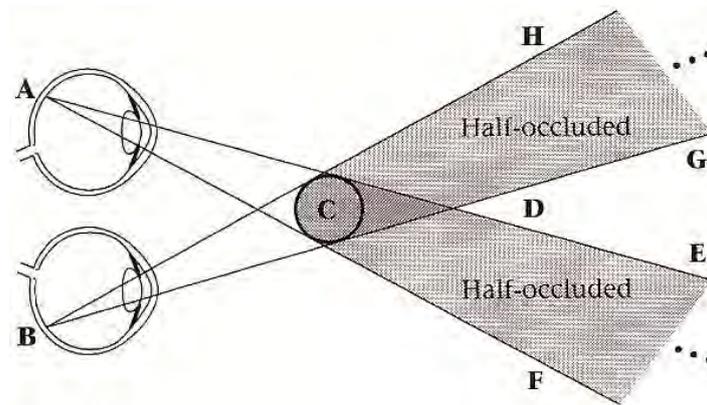


Abbildung 21: Halb-Verdeckung. Abbildung aus [W95]

Auch Zufallspunktstereogramme beinhalten genau genommen halb-verdeckte Regionen. Durch das Verschieben des Ausschnitts wird ein Teil des originalen Zufallsmusters überschrieben. Diese überschriebenen Punkte sind halb-verdeckt, da sie nur von dem Auge erfasst werden können, das das Originalbild sieht. Die zur Ergänzung nach der Verschiebung hinzugefügten Punkte sind ebenfalls halb-verdeckt, weil sie lediglich vom anderen Auge wahrgenommen werden können.

Halb-verdeckte Bereiche haben zwei grundsätzliche Eigenschaften. Zum einen können halb-verdeckte Regionen, die vom linken Auge gesehen werden, ausschließlich an der linken Seite des nahen Objekts gesehen werden. Dies gilt analog auch für das rechte Auge. Die anderen beiden Möglichkeiten, also dass das linke Auge eine Verdeckung an der rechten Seite sieht oder umgekehrt, sind physikalisch unmöglich. Zum anderen ist die Größe der halb-verdeckten Region von der Distanz zwischen Objekt und Betrachter abhängig. Je näher sich das Objekt beim Betrachter befindet, desto größer wird der halb-verdeckte Bereich. Auf Grund der Information von Position und Größe der Halb-Verdeckung kann auf die Tiefe geschlossen werden. [B04, W95]

### 3.3.4 Pseudoskopisches Sehen

Es gibt jedoch nicht nur Hinweise darauf, dass Querdisparität und andere Tiefeninformationsquellen mit einander kooperieren, sondern sich auch gegenseitig widersprechen können, wie es häufig bei pseudoskopischen Darstellungen vorkommt. Beim pseudoskopischen Sehen wird die Querdisparität invertiert, indem man dem rechten Auge das Bild des linken Auges darbietet und umgekehrt. Die daraus abgeleitete Annahme, dass sich so auch die Tiefenwahrnehmung umkehrt, ist jedoch nur für Zufallspunktstereogramme und einfache Strichzeichnungen wahr. In fotografischen Aufnahmen natürlicher Szenen, die pseudoskopisch gezeigt werden, sieht man hingegen kaum hohle Gebäude, Pflanzen oder Tiere, da eine Tiefenumkehrung des Bildes der gewohnten Wahrnehmung stark widersprechen würde. In diesem Fall wird die Stereoinformation sowohl durch Tiefenhinweise, wie Verdeckung, Perspektive und Schattierung, als auch durch Erfahrung und Vorwissen überschrieben. Ein bekanntes Beispiel dafür sind die sogenannten „Hohlmasken“. Hohlmasken sind pseudoskopisch dargestellte menschliche Gesichter, die eigentlich räumlich nach innen gewölbt erscheinen sollten. Da hohle Gesichter zu unwahrscheinlich sind, werden sie vom Gehirn nicht akzeptiert und auf Grund der gelernten Erfahrung nach außen gewölbt wahrgenommen. Einzig und alleine aus nächster Nähe überwiegt die stereoskopische Information und lässt ein Gesicht hohl erscheinen. [D06]

### 3.3.5 Binokulare Rivalität

Binokulare Rivalität entsteht, wenn das visuelle System kein vernünftiges Ergebnis bei der Verschmelzung zweier sehr unterschiedlicher Netzhautbilder erzielen kann. Die Folgen sind Unterdrückung oder instabile Wahrnehmungen ohne Tiefeneindruck. Ein Beispiel dafür ist in Abb. 22 dargestellt. Es ist so gut wie unmöglich beide Gruppen von orthogonalen Streifen gleichzeitig zu sehen. Man sieht entweder nur alle vertikalen, nur alle horizontalen Streifen oder ein Patchwork an vertikal und horizontal gestreiften Regionen, deren Grenzen sich ständig verändern. Neben rivalisierenden Strukturen, können auch konträre Farben binokulare Rivalität hervorrufen. [H95]

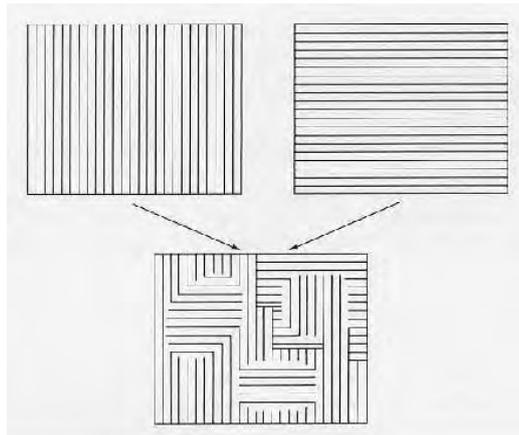


Abbildung 22: Binokulare Rivalität. Abbildung aus [H95]

### 3.3.5.1 Rivalität von Strukturen

Das Merkmal einer strukturellen Rivalität ist eine Wahrnehmung, in der Teile beider Bilder nacheinander in verschiedener Weise kombiniert und wieder verworfen werden. Ein Beispiel dafür sind die beiden Balken in Abb. 23, die starke strukturelle Unterschiede aufweisen. Die Verschmelzung beider Bilder führt nicht zu einer stabilen Wahrnehmung, sondern zu sich zeitlich abwechselnd überdeckenden Balken. [D06, G01]

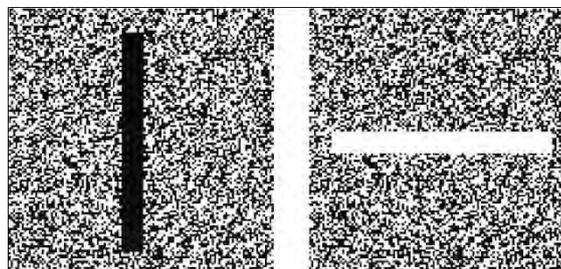


Abbildung 23: Rivalität von Strukturen. Abbildung aus [D06]

Unterscheiden sich die rivalisierenden Bilder jedoch in ihrer Klarheit oder Helligkeit, bevorzugt das visuelle System die Information eines Auges und unterdrückt jene des anderen Auges. Die Gestaltprinzipien spielen bei der Entscheidung eine wichtige Rolle. In dem Stereobildpaar in Abb. 24 dominiert daher das rechte Auge auf Grund der stärkeren strukturellen Gliederung. Das gleichmäßigere Bild des linken Auges wird unterdrückt. Dieses Phänomen bezeichnet man in der Literatur als binokulare Unterdrückung.

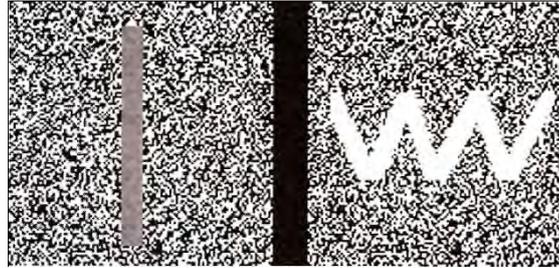


Abbildung 24: Binokulare Unterdrückung. Das rechte Bild dominiert über das linke Bild. Abbildung aus [D06]

Bela Julesz hat in diesem Zusammenhang festgehalten, dass Personen, die nicht auf beiden Augen scharf sehen, sehr wohl einen scharfen, dreidimensionalen Eindruck bekommen können, da bei der endgültigen 3D-Wahrnehmung das schärfere Abbild dominiert. Das unschärfere Bild dient lediglich der Gewinnung der Tiefeninformation. [S00, SV09]

### 3.3.5.2 Rivalität von Farben

Das Auftreten von Farbrivalitäten steht im engen Zusammenhang mit der zu Grunde liegenden Struktur. Bei Flächen mit rivalisierenden Farben ohne Struktur kommt es zu keiner Vereinigung, sondern zu einer abwechselnden Dominanz der Farben. Verschiedenfarbige, strukturierte Flächen werden jedoch vermischt, da die vorhandene Struktur verschmolzen werden kann und so die Rivalität der Farben in den Hintergrund drängt. Abb. 25 veranschaulicht die Rivalität von Farben in Abhängigkeit von Struktur. Bei den beiden kaum strukturierten Halbkreisen kommt es nach stereoskopischer Vereinigung des Stereobildpaars zu einer Rivalität der Farben Rot und Blau, während der stark strukturierte Hintergrund in der Mischfarbe Violett erscheint. [D06, H95]

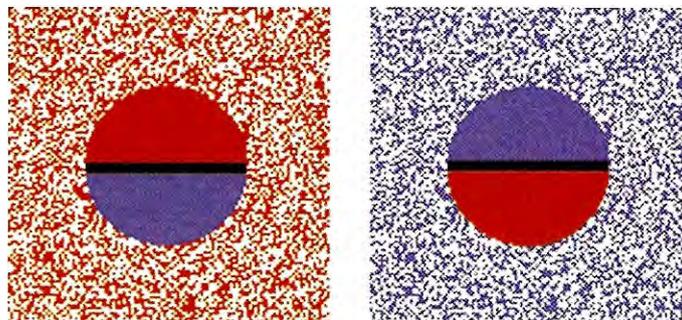


Abbildung 25: Rivalität von Farbe. Abbildung aus [D06]

### 3.3.6 Okulare Dominanz

*Ocular dominance, sometimes called eye dominance or eyedness, is the tendency to prefer visual input from one eye to the other. [SS]*

Die Bevorzugung eines Auges ist vergleichbar mit der Rechts- oder Linkshändigkeit eines Menschen, wobei die Seite des dominanten Auges nicht zwingend mit der Seite der dominanten Hand übereinstimmen muss. Personen, bei denen das dominante Auge und die dominante Hand auf unterschiedlichen Körperseiten liegen, bezeichnet man als „kreuzdominant“. Das dominante Auge nimmt bei der visuellen Wahrnehmung eine bestimmende Rolle ein – es ist das „sehende“ Auge. Das bedeutet, dass es besser auf ein Objekt fokussieren kann und somit primär für eine präzise Tiefeninformation herangezogen wird.

Bei etwa zwei Drittel der Bevölkerung dominiert das rechte Auge. Einige wenige Menschen weisen kein dominantes Auge auf. Zur Bestimmung der Augendominanz gibt eine Vielzahl an Verfahren. Eine Möglichkeit ist die Durchführung des „Miles-Tests“. Dafür streckt man beide Arme vor dem Gesicht aus und formt mit den Händen eine kleine Öffnung, wie in Abb. 26 dargestellt. Mit beiden Augen betrachtet man ein fernes Objekt durch die kleine Öffnung hindurch. Ohne den Blickkontakt zu dem Objekt zu unterbrechen, führt man beide Hände langsam näher zum Gesicht bis eindeutig nur mehr ein Auge durch die Öffnung schauen kann. Dieses ist das dominante Auge. [SS, G08]



Abbildung 26: Miles Test. Abbildung aus [SS].

## 4 Stereoskopie

*It is possible to create strong depth impressions from pictures by sending to each eye separately the view that the eye would see if an actual object in depth were presented. [B04]*

Alle Techniken und Verfahren, die mit Hilfe eines Stereobildpaars einen räumlichen Eindruck hervorrufen, werden unter dem Begriff Stereoskopie zusammengefasst. Für die Darstellung eines Bildpaares gibt es eine Vielzahl an Möglichkeiten, die alle auf einer Trennung von linker und rechter Ansicht beruhen. Diese Ansichten werden anschließend getrennt an jedes Auge gesandt und auf Grund der Querdisparität zu einer dreidimensionalen Wahrnehmung verknüpft. Das Ergebnis für den Betrachter ist ein „solides“ Objekt mit räumlicher Tiefe.

### 4.1 Stereosehen

Das „normale“ räumliche Sehen, das der Mensch seit seiner Kindheit zur Erkennung von allen Objekten in seiner Umwelt verwendet, erfolgt durch die Kopplung von Akkommodation und Konvergenz. Ein beliebiger Punkt im Raum wird fixiert, indem sich die beiden Sehachsen in diesem Punkt schneiden und die Linsen automatisch scharf stellen. Zur Verschmelzung eines Stereobildpaars zu einem räumlichen Bild ist es jedoch notwendig, diesen Mechanismus des Sehens zu umgehen. Für die Trennung von Sehwinkel und Scharfstellung stehen zwei Sehtechniken zur Verfügung: der Stereoblick und das Stereoskop.

#### 4.1.1 Stereoblick

Beim Stereoblick stellen die Augen weiterhin auf die Bildebene scharf, der Blick wird aber nicht auf das Blatt gerichtet. Die Sehachsen treffen einander somit außerhalb der Blattebene. Man spricht von Starren, wenn der Schnittpunkt der Sehachsen hinter der Blattebene liegt und von Schielen, wenn er sich davor befindet. Die beiden Sehstrategien sind in Abb. 27 schematisch dargestellt.

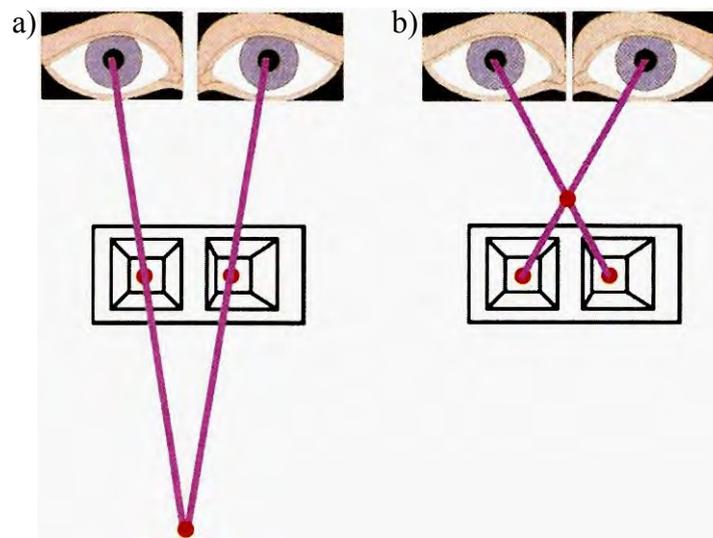


Abbildung 27: Stereoblick. a) Starren; b) Schielen. Abbildung aus [D06]

Die Blickweise des Starrens wird als angenehmer empfunden und kann über einen längeren Zeitraum aufrechterhalten werden. Der Nachteil ist jedoch, dass nur jene Bildpunkte stereoskopisch verbunden werden können, die weniger als 6,5cm (Augenabstand) voneinander entfernt sind. Beim Schielen gibt es keine solche Grenze. Bei dieser Stereoblicktechnik ist jedoch wesentlich mehr Anstrengung notwendig, um ein Stereobildpaar zur Deckung zu bringen. Außerdem kommt es zu einer Umkehr der Tiefeninformation auf Grund des vor der Bildebene liegenden Schnittpunkts der beiden Sehachsen. [D06]

#### 4.1.2 Stereoskop

Während das Stereosehen mit bloßem Auge einer gewissen Übung bedarf, ist die Verschmelzung eines Stereobildpaares mit einem Stereoskop wesentlich einfacher und bequemer. Außerdem muss bei der Betrachtung eines Stereobildes mit einem Stereoskop keine spezielle Sehtechnik erlernt werden, da die Sehstrahlen mit Hilfe optischer Elemente korrekt auf die Netzhaut auftreffen. Dies wird entweder durch die Anpassung des Strahlenganges oder der Schärfereinstellung erreicht.

#### 4.1.2.1 Spiegelstereoskop

Die Erfindung des Stereoskops wird Charles Wheatstone zugeschrieben. Im Jahr 1838 entwickelte er mit seinem Spiegelstereoskop ein optisches Instrument, das die Augen unterstützt, um aus einem 2D-Bildpaar einen dreidimensionalen Eindruck zu erzielen.

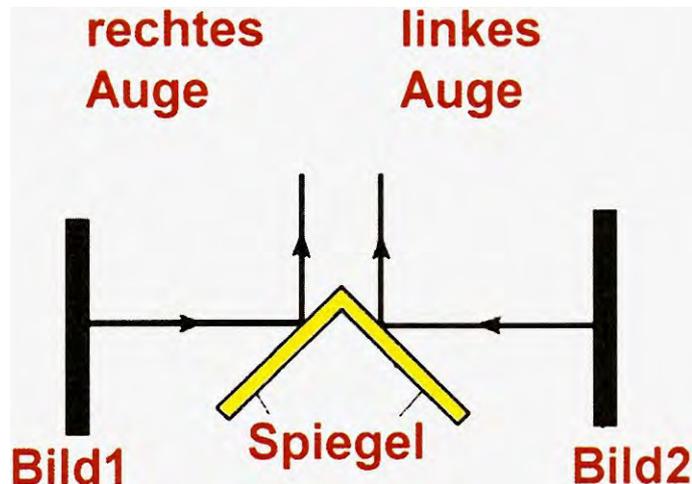


Abbildung 28: Spiegelstereoskop. Abbildung aus [D06]

Das in Abb. 28 schematisch dargestellte Spiegelstereoskop besteht aus zwei rechtwinkelig angeordneten Spiegeln. Der Betrachter sieht den linken Spiegel mit dem linken Auge und den rechten Spiegel mit dem rechten Auge. Seitlich der beiden Spiegel sind künstlich erzeugte Zeichnungen angebracht, die sich nur durch ihre horizontale Abweichung voneinander unterscheiden. Die Spiegel werden so eingestellt, dass die von den entsprechenden Punkten der beiden Zeichnungen ausgehenden Strahlen so abgelenkt werden, dass sie von einem einzigen Punkt hinter den Spiegeln zu kommen scheinen. Unter der Voraussetzung, dass Spiegel und Bilder korrekt justiert sind, entsteht eine dreidimensionale Wahrnehmung. [D06]

#### 4.1.2.2 Linsenstereoskop

Indem David Brewster die Spiegel durch ein spezielles Linsensystem ersetzte, machte er das Gerät wesentlich handlicher und für eine kommerzielle Nutzung interessant. Das Linsenstereoskop ist heute noch allgemein in Gebrauch.

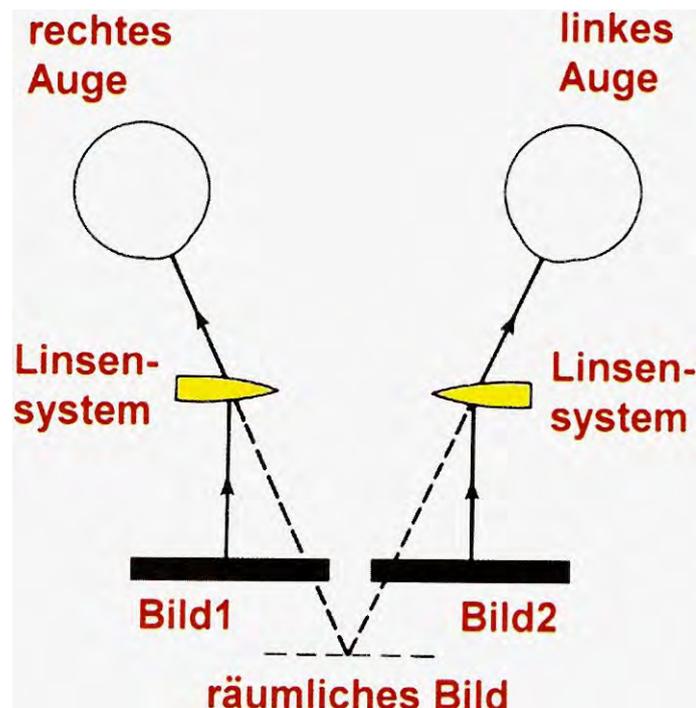


Abbildung 29: Linsenstereoskop. Abbildung aus [D06]

Abb. 29 zeigt den Aufbau eines Linsenstereoskops. Zwischen Augen und Stereobildpaar befinden sich zwei Linsen. Durch die Brechkraft dieser Linsen werden die Sehstrahlen so abgelenkt, dass sie einander erst hinter der Bildebene treffen. Da beide Einzelbilder in einer Ebene nebeneinander liegen, eignet sich diese Methode gut für das Abdrucken in Büchern. Sowohl das Linsenstereoskop als auch das Spiegelstereoskop haben den großen Nachteil, dass immer nur eine Person das 3D-Bild sehen kann. [D06]

## 4.2 3D-Bildschirmssysteme

Im Gegensatz zu den bisherigen Stereobildbetrachtungssystemen bieten die folgenden Technologien die Möglichkeit, dass der 3D-Inhalt von mehreren Personen gleichzeitig wahrgenommen werden kann. Die Einzelbilder sind nicht mehr räumlich getrennt voneinander angeordnet, sondern befinden sich auf derselben Bildschirmfläche. Um einen Stereoeindruck zu bekommen, müssen die überlagerten Bilder von beiden Augen separat betrachtet werden. Die einzelnen Bildschirmssysteme verwenden verschiedene Technologien zur Trennung von linker und rechter Ansicht. Die wichtigsten sind stereoskopische und autostereoskopische Bildschirme, da sie auf bereits bestehenden Systemen aufbauen. Weniger verbreitet sind hingegen volumetrische Bildschirmssysteme, die ein Bild mit echter Tiefe erzeugen und sowohl Augen als auch Gehirn auf natürliche Art und Weise arbeiten lassen.

## 4.2.1 Stereoskopische Bildschirmsysteme

Für das Betrachten von 3D-Inhalten mit stereoskopischen Bildschirmen benötigt man eine Spezialbrille. Die Brillengläser fungieren als Filter, die entscheiden, welches der beiden Bilder dem linken beziehungsweise dem rechten Auge präsentiert wird. Wenn die korrekten Bilder den entsprechenden Augen zugeordnet werden, entsteht eine räumliche Wahrnehmung. Die bei stereoskopischen Bildschirmen verwendeten Techniken – Anaglyphen, Polarisation und Shutter-Brille – haben ihren Ursprung im Bereich des Kinos. Sie kommen inzwischen auch zunehmend bei Fernsehgeräten und Computer-Monitoren zum Einsatz.

### 4.2.1.1 Anaglyphentechnik

Die beiden Einzelbilder werden bei der Anaglyphentechnik mit Hilfe des sichtbaren Lichtspektrums, also Farbe, voneinander getrennt. Das Stereobildpaar wird in Komplementärfarben, wie zum Beispiel Rot-Grün, eingefärbt. Das grüne Bild und das rote Bild werden mit richtiger Querdiparität überlagert, so dass ein Doppelbild - ähnlich wie bei einem Fehldruck – entsteht. Sieht ein Betrachter durch eine Brille mit einem grünen und einem roten Glas hindurch, so gelangt immer nur ein Bild zu jedem Auge, da die jeweils andere Ansicht von der Rot-Grün Brille absorbiert wird. Bei richtiger Abstimmung der spektralen und intensitätsmäßigen Durchlässigkeit der Filter, entsteht so auf einfache Weise ein räumlicher Eindruck.

Neben der einfachen Durchführung des Verfahrens und der geringen Anschaffungskosten, ist der Anwendungsbereich ein wesentlicher Vorteil. Die Anaglyphentechnik kann praktisch auf jedem Medium, das im Stande ist Farbe wiederzugeben, verwendet werden, unabhängig davon ob es sich um elektronische Geräte, Film oder Papier handelt. Nachteile sind, dass die Tiefenbilder nicht farbig sind und es zu visuellen Unannehmlichkeiten auf Grund des unterschiedlichen Farbinhalts für jedes Auge kommt (Rivalität von Farbe). Außerdem kann es bei schlechter Abstimmung zu starken Helligkeitsverlusten oder Doppelkonturen im Stereobild kommen. [D06, V65]

#### 4.2.1.2 Polarisationsfiltertechnik

Zur Trennung der beiden Ansichten wird bei der Polarisationsfiltertechnik die Eigenschaft der Polarisation von Lichtwellen genutzt. Man verwendet üblicherweise zwei Projektoren mit unterschiedlichen Polarisationsfiltern, die das Stereobildpaar auf eine nicht-depolarisierende Leinwand projizieren. Diese spezielle Leinwand soll sicherstellen, dass die Polarisation erhalten bleibt. Entscheidet man sich für ein System mit linearer Polarisation, wird ein Projektor mit einem horizontal-polarisierenden Filter und der andere mit einem vertikal-polarisierenden Filter ausgestattet. Jedes Auge erfasst nur das Licht von einer Ansicht, wenn der Betrachter dementsprechend polarisierte Brillen trägt. Lineare Polarisation ist jedoch anfällig für optisches Übersprechen zwischen den beiden Ansichten bei Rotation, wie beispielsweise einer Kopfbewegung des Betrachters. Dieses Problem kann durch die Verwendung von zirkular-polarisierenden Filtern (links- und rechtsdrehend) eingedämmt werden. Ein Rest an gegenseitiger Beeinflussung bleibt dennoch auf Grund unzureichender Auslöschungseigenschaften der Filter und Depolarisation auf der Leinwand. Außerdem sind Zirkular-Polarisatoren teurer als Linear-Polarisatoren.

Das System mit zwei Projektoren hat zwei weitere Nachteile. Zum einen blockieren die Filter einen beträchtlichen Teil des Lichtes. Dies kann mit lichtstärkeren Projektoren kompensiert werden. Zum anderen müssen die beiden Projektoren sorgfältig eingerichtet werden, um Augenschmerzen zu verhindern. Eine Alternative zu den zwei Projektoren mit statischen Polarisationsfiltern ist die Verwendung eines einzelnen Projektors mit einem dynamischen Polarisationsmodulator. Bei zeitlicher Verschränkung beider Einzelbilder wechselt der Modulator in Abhängigkeit der Bilder die Polarisationsrichtung. Das Ergebnis sind unterschiedlich polarisierte Einzelbilder, die einfach mit der vom Benutzer getragenen Brille getrennt werden können. Trotz der eben genannten Schwierigkeiten hat das Polarisationsfiltersystem – vor allem gegenüber dem Anaglyphenverfahren – einen entscheidenden Vorteil, nämlich die vollständige Wiedergabe von Farbe. [D06, V65]

#### 4.2.1.3 Shutter-Technik

Bei der Shutter-Technik erscheinen linke und rechte Ansicht abwechselnd am Bildschirm. Die Gläser der Shutter-Brille enthalten Flüssigkristalle, die mit Hilfe einer speziellen Elektronik lichtundurchlässig gemacht werden können. Die schnell schaltenden Linsen der Brille, die mit dem Bildschirm synchron arbeiten, ermöglichen durch wechselnden Verschluss einer Seite, dass jedes Auge nur die passende Ansicht sieht. Die Bildwiederholrate

muss bei diesem Verfahren das Doppelte der normalen Bildfrequenz betragen, um ein flimmerfreies Betrachten des Stereoinhaltes sicherzustellen.

Das Shutter-System bietet eine hohe Qualität in Bezug auf stereoskopische Visualisierung. Der entscheidende Vorteil gegenüber den Polarisationsverfahren besteht darin, dass es zu keiner störenden Überlagerung der Einzelbilder kommen kann, da die Ansichten beider Augen nie gleichzeitig am Bildschirm zu sehen sind. Als Nachteil der Shutter-Technik sind der erhebliche technische Aufwand, sowie die damit verbundenen hohen Kosten zu nennen. [D06]

## **4.2.2 Autostereoskopische Bildschirmsysteme**

Das Tragen von speziellen Brillen zur Wahrnehmung von stereoskopischen Inhalten erweist sich in vielen Fällen, nicht nur wegen der physischen Unbequemlichkeit sondern auch wegen der Kosten und Pflege der Brillen, als Unannehmlichkeit. Außerdem beeinträchtigt das Tragen von Shutter- und Polarisationsbrillen das gleichzeitige Betrachten von herkömmlichen 2D-Monitoren. Diese Nachteile fallen bei autostereoskopischen Bildschirmsystemen weg, da für das stereoskopische Sehen mit dieser alternativen Technologie keine (Spezial-)Brillen notwendig sind. Die gängigsten Systeme – Lentikular-System („lenticular display“ oder „microlens display“) und Parallaxe-Barriere-System („parallax-barrier display“) – basieren auf der Flachbildschirmtechnologie. Das Licht, das Pixel der jeweiligen Ansicht ausstrahlen, wird mit Hilfe von optischen Effekten, wie Brechung oder Verdeckung zum entsprechenden Auge gelenkt. [K07, DE05, D05]

### **4.2.2.1 Multiplex-Systeme**

Bei Multiplex-Systemen wird die horizontale Auflösung eines Bildschirms auf linke und rechte Ansicht spaltenweise aufgeteilt. Eines der sichtbaren Bilder besteht aus jeder zweiten Pixelspalte. Die anderen Spalten werden dem zweiten Bild zugeordnet. Zur getrennten Darstellung des Stereobildpaars werden entweder eine Reihe zylindrischer Linsen (Lentikular-Bildschirm) oder eine Abdeckungsmaske (Parallaxe-Barriere-Bildschirm) verwendet.

Bei Parallaxe-Barriere-Systemen, wie in Abb. 30a schematisch dargestellt, wird eine undurchsichtige Ebene mit schmalen, gleichmäßig verteilten Schlitzen, sehr nahe an einem pixelweise-adressierbaren Bildschirm, wie beispielsweise Plasma-, LCD- oder LED-Bildschirm, angebracht. Jeder dieser Schlitze fungiert als eine Art „Pinhole-Projektor“.

Voraussetzung für eine korrekte räumliche Wahrnehmung des dargebotenen Inhalts ist die genaue Abstimmung von Pixelspalten, Größe und Ausrichtung der Schlitze sowie der Abstand der Sichtbarriere. Ein Betrachter, der sich im richtigen Abstand vor dem Bildschirm aufhält, sieht mit jedem Auge einen unterschiedlichen Satz an Pixel entsprechend der linken und rechten Ansicht. Dieser Stereobildbereich („viewing zone“) umfasst jene Stellen vor dem Bildschirm, an denen stimmige 3D-Wahrnehmungen erlebt werden können.

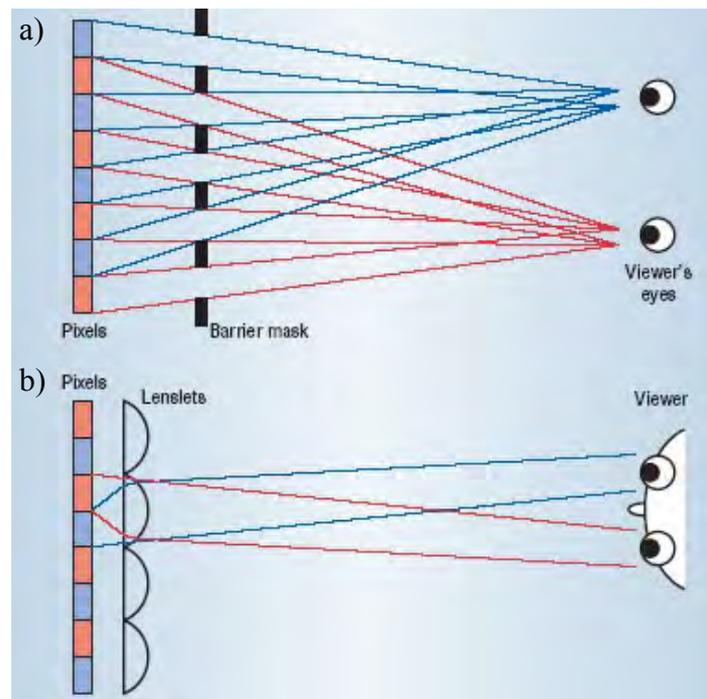


Abbildung 30: Autostereoskopische Bildschirmsysteme. a) Parallaxe-Barriere-System; b) Lentikular-System. Abbildung aus [D05]

Wie in Abb. 30b veranschaulicht, befindet sich bei Lentikular-Systemen eine Scheibe mit schmalen, dünnen, zylindrischen Mikrolinsen im Abstand der Brennweite vor dem Bildschirm. Das Licht der angrenzenden Pixel wird durch die Linsen so abgelenkt, dass jedes Auge nur die Spalten mit dem passenden Stereobild sieht, wenn Pixel- und Mikrolinsenraster präzise ausgerichtet sind und sich der Betrachter im vorgesehenen Stereobildbereich befindet.

Ein Nachteil der Multiplex-Systeme ist die optische Überschneidung der einzelnen Ansichten, wenn ein Spalt oder eine Linse Teile der benachbarten Daten mitverwendet. Eine mögliche Lösung für dieses Problem ist das Einführen von Sicherheitsbändern für jeden Schlitz beziehungsweise für jede Mikrolinse. Diese Bänder liegen zwischen den sichtbaren Daten und enthalten keine Information, wodurch ein oder beide Ansichten auf Kosten der Größe

des Stereobildbereichs verdunkelt und nicht vermischt dargestellt werden. Auf Grund der Pixelstruktur ist dieser Ansatz in Praxis eher unpraktisch. Man versucht stattdessen eine Balance zwischen störenden Überlagerungen, Größe der Stereobildzone und Lichtertrag zu finden.

Ein Parallaxe-Barriere oder Lentikular-Bildschirm erzeugt mehrere Stereobildbereiche, in denen ein Auge entweder das linke oder das rechte Bild sieht. Die Möglichkeit, dass linke und rechte Ansicht vertauscht betrachtet werden und so ein falsches, pseudoskopisches Bild entsteht, liegt also bei 50 Prozent. Abb. 31 verdeutlicht dieses Problem. Nur an den sogenannten „sweet spots“ kommt eine richtige Tiefenwahrnehmung zustande. Jedes Wegbewegen des Betrachters von der idealen Distanz zum Bildschirm verringert zusätzlich die Chance ein korrektes Stereobild wahrzunehmen.

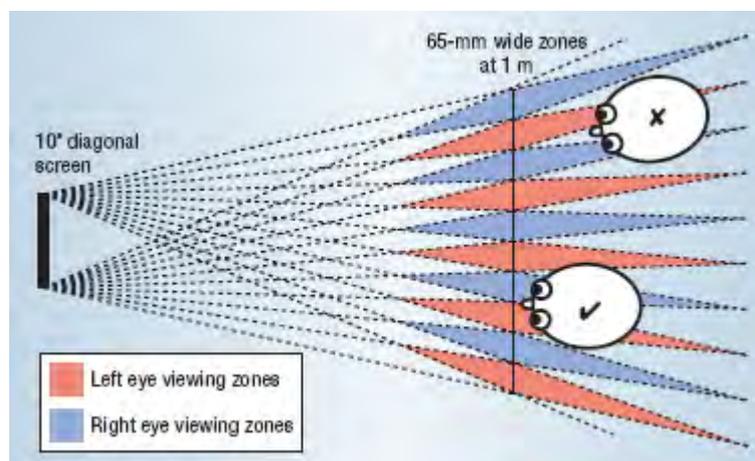


Abbildung 31: „Sweet Spots“. Abbildung aus [D05]

Das unmittelbare Nebeneinanderliegen von korrekten und umgekehrten Stereobildzonen kann beim Betrachter zu visuellen Verwirrungen und Unannehmlichkeiten führen. Mittels Head-Tracking, also dem Erfassen der aktuellen Kopfposition eines Benutzers, können die beiden Einzelbilder zu jeder Zeit dem jeweils richtigen Auge zugeordnet und pseudoskopische Bilder vermieden werden, vorausgesetzt der Augenabstand des Betrachters unterscheiden sich nicht signifikant vom jenem Wert, den das Bildschirmsystem erwartet. Es kommen zwei unterschiedliche Head-Tracking Systeme zum Einsatz. Eine Methode ist das Verschieben der Zonen bei Bewegung des Betrachters. Bei der anderen Methode werden nur zwei Zonen erzeugt, aber das Gerät entscheidet, wo die beiden Ansichten gezeigt werden. Diese Systeme funktionieren jedoch immer nur für eine einzelne Person. Beide Head-Tracking Mechanismen sind in Abb. 32 dargestellt. [K07, D05]

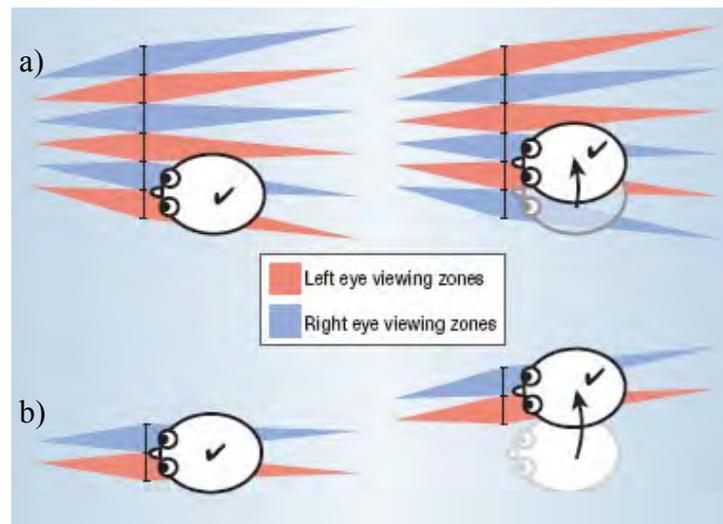


Abbildung 32: Head-Tracking Mechanismen. a) System mit allen Zonen; b) System mit nur zwei Zonen. Abbildung aus [D05]

#### 4.2.2.2 Multiview -Systeme

Bei der Darstellung von nur zwei Ansichten wird unabhängig von der Position des Betrachters immer dasselbe räumliche Bild angezeigt. Dies kann zu Verwirrungen über die wahre Gestalt eines Objektes führen, da in der Realität um ein Objekt herum gesehen werden kann. Zur Nachahmung der Tiefeninformation durch Bewegung sind daher mehrere positionsabhängige Ansichten notwendig. Mehr-Ansichten-Systeme („multiview display“) können Bewegungsparallaxe zusätzlich zu stereoskopischer Tiefe bereitstellen.

Mittels Head-Tracking wird die Position des Betrachters erfasst und die entsprechende Ansicht am Bildschirm dargestellt. Im Idealfall kann diese Methode, die auch als „active mutiview“ bezeichnet wird, dieselben Bewegungs- und Stereotiefenhinweise wie ein reales Objekt präsentieren. Der größte Mangel an diesem Multiview-Bildschirmssystem ist die Tatsache, dass immer nur die eine getrackte Person die korrekte Ansicht sehen kann. Im Gegensatz dazu werden bei passiven Multiview-Bildschirmen mehrere Ansichten gleichzeitig dargestellt. Der Benutzer wechselt mit seiner Position auch die Sicht auf die Szene. Der Vorteil dieser Methode ist, dass jeder Betrachter ein dreidimensionales Bild sieht, wenn sich beide seiner Augen im Stereobildbereich befinden. Abb. 33 zeigt schematisch das Konzept eines sogenannten automultiskopischen Bildschirms.

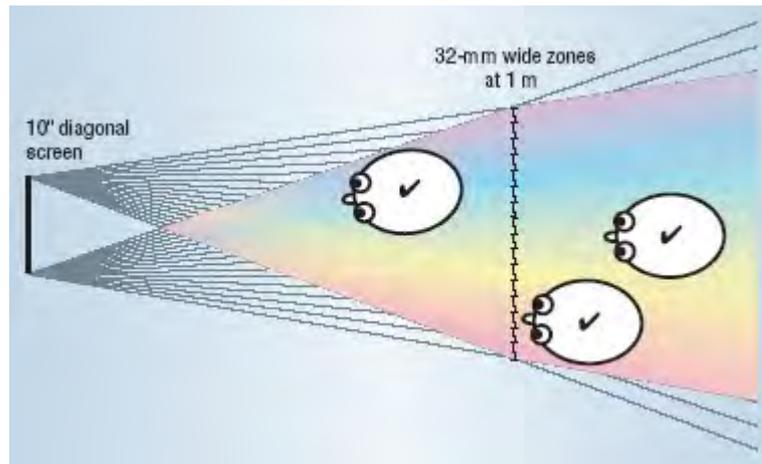


Abbildung 33: Automultiskopischer Bildschirm. Abbildung aus [D05]

Ein wesentliches Problem ist die Herstellung von Bildschirmen mit vielen Ansichten, da beispielsweise bei zwei Ansichten die horizontale Auflösung halbiert wird, bei drei gedrittelt und so weiter. Da die derzeitige Auflösung von Plasma-, LCD- und LED-Bildschirmen noch zu gering für eine vollständige Bewegungsparallaxe ist, kommt nur die horizontale Parallaxe zum Einsatz. Das bedeutet, dass keine Auf- und Abwärtsbewegungen sondern nur seitliche Bewegungen des Kopfes berücksichtigt werden. Eine weitere Herausforderung ist die gleichzeitige Darstellung aller Ansichten, ob sie nun gerade von einem Betrachter gesehen werden oder nicht. [K07, D05]

#### 4.2.3 Volumetrische Systeme

Im Gegensatz zu den autostereoskopischen Flachbildschirmsystemen, die mittels Ablenkung des ausgesendeten Lichts einen dreidimensionalen Eindruck erzeugen, scheint das emittierte Licht bei volumetrischen Bildschirmen von einem Rauminhalt zu kommen. Die darzustellende 3D-Szene wird in das Bildschirmvolumen gezeichnet. Da die Lage jedes Punktes im Volumen mit seiner Lage in der Szene korrespondiert, entsteht ein Abbild auf der Retina des Betrachter, das dem eines realen Objekts entspricht. Diese Eigenschaft bedeutet, dass bei volumetrischen Bildschirmssystemen Akkommodation möglich ist, weil die Augen auf das volumetrische Bild in gleicher Weise scharfstellen können, wie auf die reale Szene. Neben der horizontalen Parallaxe kann bei diesen Systemen auch die vertikale Parallaxe realisiert werden, wodurch sich in Kombination mit nicht-diskretisierten Stereobildbereichen eine Vielzahl an Betrachtungspositionen ergibt. Volumetrische Bildschirme können Verdeckung jedoch nicht für alle Betrachter fehlerfrei darstellen. Da Verdeckung

aber einer der stärksten Tiefenhinweise für das menschliche visuelle System ist, schränkt dies die Szenen, in denen das Bildschirmsystem überzeugen kann, erheblich ein.

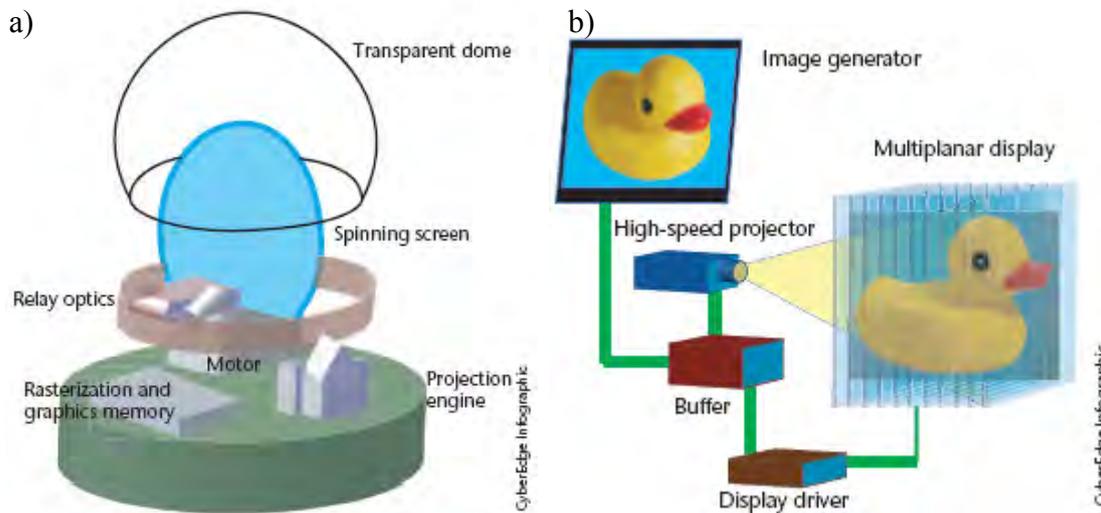


Abbildung 34: Volumetrische Bildschirmsysteme. a) „Perspecta“; b) „DepthCube“. Abbildung aus [DE05]

Die beiden grundlegenden Arten von volumetrischen Bildschirmsystemen sind in Abb. 33 dargestellt. Das in Abb. 34a abgebildete System „Perspecta“ von Actuality Systems [AS] erzeugt 3D-Bilder, indem eine Bildfolge auf einen schnell rotierenden, flachen, beidseitigen Bildschirm projiziert wird. Durch das Ausfüllen des Bildraumes mit der Projektionsfläche entsteht bei ausreichend hoher Aktualisierungsrate ein dreidimensionales Bild. Die Rotation führt im Vergleich zu den herkömmlichen 2D Bildschirmen zu einer relativ komplexen Rasteranordnung der volumetrischen Bilddaten. Diese Problematik fällt bei dem in Abb. 34b dargestellten System „DepthCube“ von LightSpace Technologies [LS] weg. Die verschiedenen Tiefenebenen eines Bildes („slices“) werden auf diskreten Bildschirmflächen gezeigt. Die Bildschirmflächen können einzeln mittels elektronischen Impulses zwischen transparent und lichtstreuend umgeschaltet werden. Zu jedem Zeitpunkt in einem Bildschirmdurchlauf ist immer nur ein Shutter aktiv, der als Projektionsfläche für das dahinter liegende Bild dient. Synchron mit der Darstellung eines neuen projizierten Bildes werden die Bildschirmflächen nacheinander aktiviert und beschreiben so ein gesamtes Bildschirmvolumen. Die sichtbaren Unterbrechungen in einem Bild, die auf Grund der Abstände zwischen benachbarten Ebenen entstehen, werden mit Hilfe von Interpolation verwischt. [DE05, F05]

## 4.3 3D-Video Formate

*3D video is commonly understood as a type of visual media that provides depth perception of the observed scenery. [SW09]*

Die Vielzahl an 3D-Bildschirmsysteme hat unterschiedliche 3D-Video Formate zur Folge, da diese meist mit der verwendeten Bildschirmtechnologie zusammenhängen. Das bedeutet, dass die gesamte Produktionskette sehr unflexibel ist, da sie bereits von der Produktion an durch das endgültige Darstellungsformat bestimmt wird. Um der Vielzahl an Anwendungen von 3D-Video gerecht zu werden, ist ein einheitliches Medienformat für 3D-Inhalte notwendig. Die Anforderungen an den allgemeinen Standard sind in Abb. 35 grafisch dargestellt.

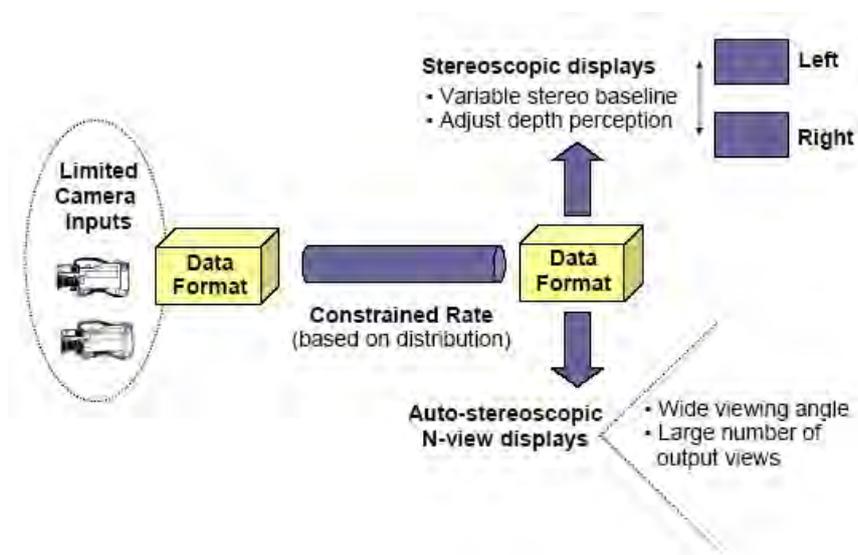


Abbildung 35: Anforderungen an einen 3D-Video Standard. Abbildung aus [SV09]

### 4.3.1 Basis 3D-Video Formate

Die Basis 3D-Video Formate werden bereits zur Markteinführung von 3D-Video verwendet. Sie erfüllen jedoch bei weitem nicht die Anforderungen an einen universal einsetzbaren Standard. Zu den Basis 3D-Video Formaten zählen „Conventional Stereo Video“ und „Video plus Depth“.

#### 4.3.1.1 Conventional Stereo Video

„Conventional Stereo Video“ (CSV) ist die einfachste Form zur Darstellung von 3D-Inhalten. Zwei von unterschiedlichen Kameras aufgezeichnete Video-Signale, die den Ansichten für linkes und rechtes Auge entsprechen, werden entweder zeitlich oder räumlich

überlagert übertragen und können weitgehend direkt am Bildschirm dargestellt werden. Bei der räumlichen Überlagerung enthält ein Frame immer beide Ansichten. Die Ansichten können links-rechts, oben-unten oder im Schachbrettmuster angeordnet sein. In jedem Fall kommt es zu einer Reduktion der Auflösung des Gesamtbildes. Dieser Nachteil entfällt bei der zeitlichen Überlagerung, da linke und rechte Ansicht alternierend am Bildschirm zu sehen sind. Um ein flimmerfreies Betrachten zu ermöglichen, ist allerdings eine höhere Bildfrequenz notwendig.

Da die Übertragung von Stereobildern für gewöhnlich eine Verdoppelung der benötigten Bandbreite bedeutet, gibt es bereits Ansätze zur Datenreduktion. Eine Möglichkeit ist die Verwendung einer Kombination von zeitlich- und ansichtenübergreifenden Differenzbildern. Dieses Prinzip kann auch auf mehr als zwei Ansichten angewandt werden. Man spricht dann von „Multiview Video Coding“ (MVC). Ein Beispiel für eine Stereo-Video-Kodierung von zwei Ansichten ist in Abb. 36 veranschaulicht.

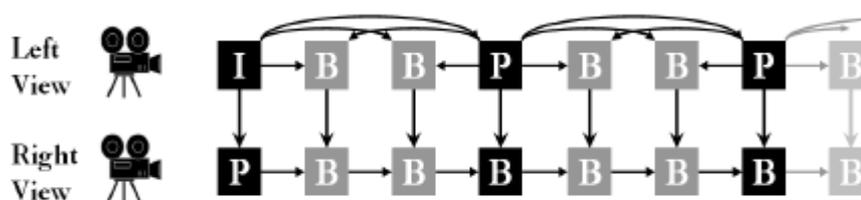


Abbildung 36: „Multiview Video Coding“-Konzept. Abbildung aus [SW09]

Ein anderer Ansatz basiert auf der in Kapitel 3 beschriebenen Theorie zur binokularen Unterdrückung. Eine der beiden Ansichten wird mit voller Auflösung und die andere mit herunter gerechneter, geringerer Auflösung übertragen. Diese räumliche Filterung hat keine Auswirkung auf die Tiefenwahrnehmung und bis zu einem gewissen Grad der Datenreduktion wird die Gesamtqualität und Schärfe des 3D-Videos durch das Bild mit höherer Qualität bestimmt. Ein Beispiel für ein Stereobildpaar, bei dem die rechte Ansicht mittel Tiefpassfilterung erzeugt wurde, ist in Abb. 37 zu sehen.



Abbildung 37: Stereobildpaar mit unterschiedlicher Auflösung. Abbildung aus [SW09]

Mittels der zusätzlich übertragenen Stereo SEI („Supplemental Enhancement Information“) Daten kann das Empfängergerät die überlagerten Ansichten trennen und als Stereobild am Bildschirm darstellen. Für alte Fernsehgeräte, die nicht in der Lage sind, das Videosignal richtig zu interpretieren, ist dieses Format unbrauchbar. Ein weiterer genereller Nachteil von CSV ist, dass die 3D-Wahrnehmung nicht verändert werden kann. Der Betrachtungsabstand ist bereits durch die Aufnahme vorgegeben und die Tiefenwahrnehmung kann auch nicht an unterschiedliche Bildschirmarten und Größen angepasst werden. Außerdem wird bei CSV keine Bewegungsparallaxe unterstützt, da die Anzahl der Ansichten begrenzt ist und keine synthetischen Ansichten erzeugt werden können. [SW09, S00]

#### 4.3.1.2 Video plus Depth

Das in Abb. 38 dargestellte „Video plus Depth“ (V+D) Format setzt sich aus einem Videosignal und einer pixelweisen Tiefenkarte („Depth Map“) zusammen. Der Tiefenbereich wird durch Grauwerte repräsentiert. Höhere Werte (Weiß) entsprechen dabei Punkten, die näher zur Kamera sind, während niedrige Grauwerte (Schwarz) die fernen Punkte darstellen. Die Tiefenkarte kann daher als reines Helligkeitssignal („Luminanzsignal“) übertragen werden. Erst beim Empfänger wird aus Video- und Tiefeninformation das Stereobildpaar errechnet, in dem die Grauwerte als Maß für die Verschiebung der einzelnen Objekte in der zweiten Ansicht herangezogen werden.



Abbildung 38: Video plus Depth. Abbildung aus [SW09]

Die Vorteile dieses 3D-Video-Signals sind zum einen die Abwärtskompatibilität und zum anderen die Möglichkeit, den Stereoeindruck nach der Übertragung an die jeweiligen Gegebenheiten anzupassen. Dies beinhaltet auch die Unterstützung von Bildschirmsystemen, die mehr als zwei Ansichten gleichzeitig darstellen können. Die Generierung des Stereobildpaars aus V+D bringt allerdings auch eine erhöhte Komplexität mit sich. Neben der Bildsynthese für die zweite Ansicht beim Empfänger, muss vor der Übertragung die Tiefeninformation (Querdisparität) ermittelt werden. Die Berechnung der Querdisparität aus einem aufgenommenen Stereobildpaar ist jedoch auf Grund der Komplexität fehleranfällig. [SW09]

#### 4.3.2 Erweiterte 3D-Video Formate

Die Hauptschwachstelle der Basis 3D-Video Formate liegt bei der Bereitstellung einer Vielzahl an unterschiedlichen Ansichten einer Szene, wie es für Mehr-Ansichten-Bildschirmsysteme erforderlich ist. Sowohl bei MVC als auch bei V+D kann nur eine sehr begrenzte Anzahl an synthetischen Ansichten in der Nähe des verfügbaren Originals erzeugt werden, da mit der Distanz die Häufigkeit von Artefakten steigt. Mit Hilfe der erweiterten 3D-Video Formate, wie „Multiview Video plus Depth“, „Layered Depth Video“ und „Depth Enhanced Stereo“, die als Erweiterung und Zusammensetzung der Basis 3D-Video Formate angesehen werden, wird eine Vielzahl an verschiedenen Ansichten zur Verfügung gestellt, um Mehr-Ansichten-Bildschirmsysteme zu unterstützen. Die Grundidee dieser 3D-Video Formate ist in Abb. 39 dargestellt.

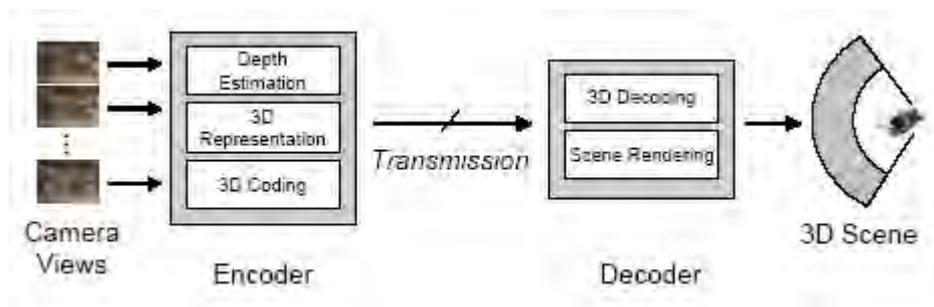


Abbildung 39: Grundidee der erweiterten 3D-Video-Formate. Abbildung aus [SV09]

Nach der Aufnahme der unterschiedlichen Ansichten erfolgt entweder eine Tiefenberechnung oder 3D-Rekonstruktion und die Erzeugung des 3D-Datenstroms. Diese 3D-Daten werden nach der Übertragung vom Empfängergerät entschlüsselt und bilden das Ausgangsmaterial zur Generierung der restlichen Ansichten. Diese erweiterte Videoverarbeitungskette beinhaltet jedoch fehleranfällige Schritte, wie beispielsweise die Berechnung der Tiefe oder die Generierung der synthetischen Ansichten. [SV09]

#### 4.3.2.1 Multiview Video plus Depth

„Multiview Video plus Depth“ (MVD) ist eine Erweiterung zu V+D. Aus Gründen der Effektivität werden bei MVD nur ein paar Originalvideosignale mit zugehöriger Tiefeninformation übertragen. Die fehlenden Ansichten werden auf Seite des Empfängers synthetisch interpoliert. Für Geräte, die das MVD-Format nicht interpretieren können, sind Basisdaten, wie zum Beispiel ein Farbvideo und eine Tiefenkarte mit eingeschränkter Auflösung, verfügbar.

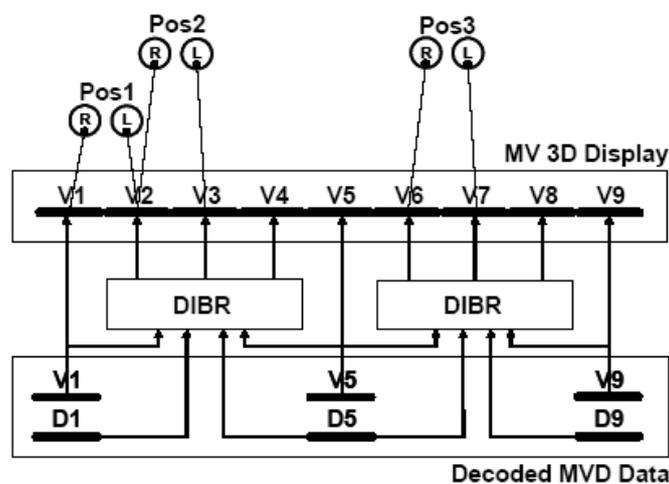


Abbildung 40: „Multiview Video plus Depth“-Konzept. Abbildung aus [S08]

Das Beispiel in Abb. 40 geht von einem Bildschirmsystem mit neun unterschiedlichen Ansichten aus, von denen nur drei Videosignale (V1, V5 und V9) mit entsprechender Tiefenkarten (D1, D5 und D9) übertragen werden. Dieses erweiterte Format ermöglicht nicht nur mehreren Personen (Pos1 und Pos3) das gleichzeitige Betrachten einer 3D-Video Szene, sondern unterstützt auch Bewegungsparrallaxe, beispielsweise wenn sich die Person von Pos1 nach Pos2 bewegt. Der räumliche Bewegungseindruck ist wegen der beschränkten Anzahl an möglichen Positionen jedoch nicht kontinuierlich. [S08]

#### 4.3.2.2 Layered Depth Video

„Layered Depth Video“ (LDV) ist eine Alternative zu MVD. Mit dem LDV-Format wird eine noch geringere Datenrate erzielt, indem nur eine zentrale Ansicht mit Tiefenkarte übertragen wird. Zur Generierung der übrigen Ansichten benötigt man jedoch noch Zusatzinformation für jene Bereiche, die in der Hauptansicht durch Vordergrundobjekte verdeckt sind. Daher wird zu der einen zentralen Ansicht entweder der Hintergrund separat mit eigener Tiefenkarte oder sogenannte Restvideosignale („residual layer“) für die Seitenansichten übermittelt. Abb. 41 veranschaulicht das LDV-Konzept mit residualen Videosignalen. Das LDV-Datenformat enthält neben der zentralen Ansicht mit entsprechender Tiefenkarte, farbige Differenzbilder mit zugehöriger Tiefeninformation.

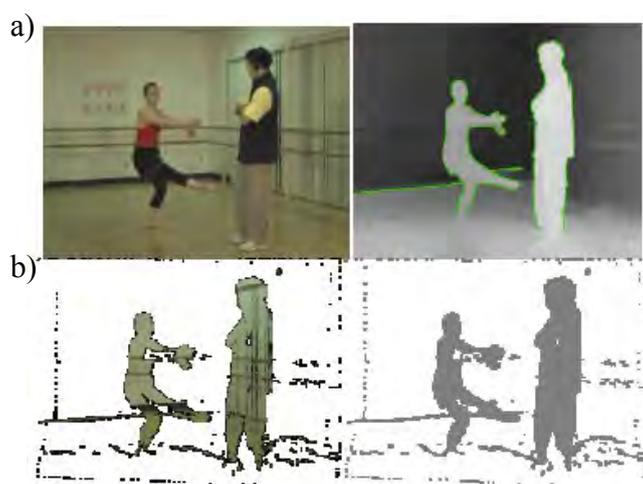


Abbildung 41: Layered Depth Video. a) Zentrale Ansicht und Tiefenkarte; b) Differenzbild mit Tiefenkarte. Abbildung aus [SV09]

Vorteil des LDV gegenüber MVD ist die geringere Menge an Daten, die übertragen werden muss. Auf der anderen Seite benötigt man zur Generierung der synthetischen Ansichten mehr fehleranfällige Berechnungsprozesse. Damit steigt auch die Zahl der Artefakte. [SV09, M08]

#### 4.3.2.3 Depth Enhanced Stereo

„Depth Enhanced Stereo“ (DES) ist ein Konzept für ein 3D-Video Format, das direkt von CVS-Systemen genutzt werden kann. Diese Kompatibilität mit CVS ist sehr wichtig, da derzeit alle 3D-Produktionen für Kino oder Fernsehen in diesem Format erstellt werden. Durch zusätzliche Information über Tiefe oder mögliche Verdeckungen kann DES auch als erweitertes 3D-Video Format eingesetzt werden und alle entsprechenden Funktionalitäten, wie Anpassung des Betrachtungsabstandes oder Generierung von synthetischen Ansichten, zur Verfügung stellen. Abb. 42 zeigt die Grundidee dieses allgemeinen 3D-Videoformats, das als Basis für einen flexiblen und effektiven Standard denkbar ist. [SV09]

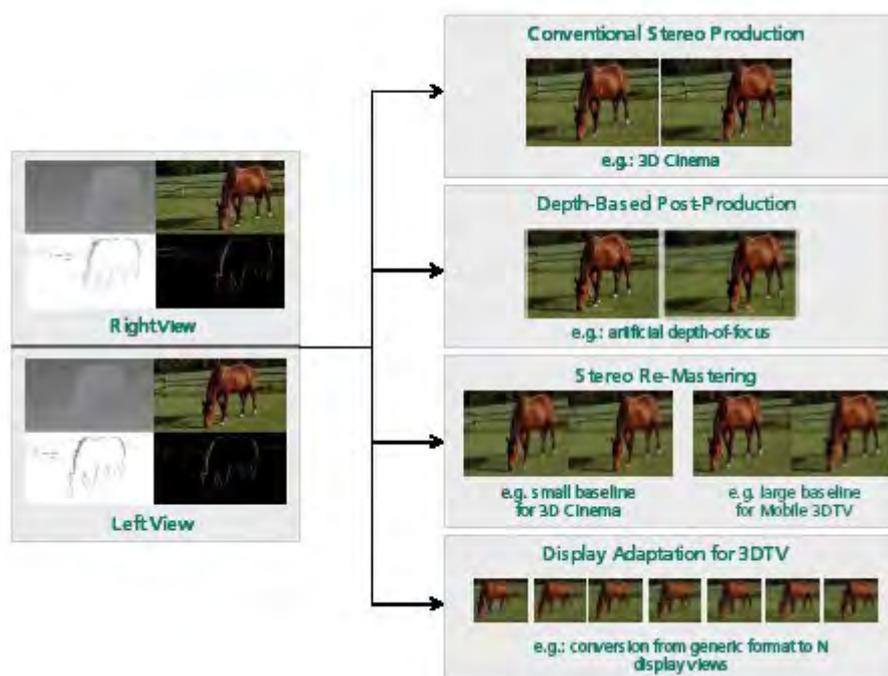


Abbildung 42: „Depth Enhanced Stereo“-Konzept. Abbildung aus [SV09]

## 5 Alpha-Matting

*Matting refers to the problem of accurate foreground estimation in images and video. [W07]*

Das Herauslösen eines Vordergrundobjektes aus einem (natürlichen) Bild, um das Objekt anschließend vor einen neuen Hintergrund zu stellen, ist ein wichtiges Element bei der Bild- und Videobearbeitung. Die Voraussetzung für qualitativ hochwertige Ergebnisse ist eine präzise Trennung von Vordergrund und Hintergrund.

### 5.1 Das Matting-Problem

Mit der Einführung eines Alpha-Kanals zur linearen Interpolation zwischen Vordergrund- und Hintergrundfarbe zur Vermeidung von Artefakten an Objektgrenzen im Zuge des Einsetzens eines Vordergrundes vor einen neuen Hintergrund, wurde das sogenannte Matting-Problem erstmals mathematisch erfasst und durch folgende Gleichung definiert:

$$I_z = \alpha_z F_z + (1 - \alpha_z) B_z$$

Dadurch wird ein Bild  $I$  als Kombination von Vordergrundbild  $F$  und Hintergrundbild  $B$  mit der Vordergrundopazität („Alpha-Matte“)  $\alpha$  beschrieben, wobei  $\alpha$  Werte aus dem Intervall  $[0,1]$  annimmt. Jedes Pixel  $z$  kann so entweder dem Vordergrund ( $\alpha_z = 1$ ) beziehungsweise dem Hintergrund ( $\alpha_z = 0$ ) zugeordnet werden oder der gemischten Region ( $0 < \alpha_z < 1$ ) angehören. Pixel der gemischten Region machen üblicherweise nur einen relativ geringen Teil eines natürlichen Bildes aus, da die meisten Pixel eindeutig Vordergrund oder Hintergrund sind. Für die vollständige Trennung ist eine genaue Berechnung der Alpha-Werte im gemischten Bereich essentiell. Die Schwierigkeit die Matting-Gleichung zu lösen besteht darin, dass von jedem Pixel des Ausgangsbildes nur ein dreidimensionaler Farbvektor  $I_z$  zur Verfügung steht und somit die Farbvektoren  $F_z$  und  $B_z$ , sowie das Skalar  $\alpha_z$  unbekannt sind. Um den Vordergrund gänzlich vom Hintergrund zu trennen, müssen jedoch die Werte an jeder Pixelstelle bestimmt werden. Nur wenn die korrekten Werte für  $F, B$  und  $\alpha$  vorliegen, kann der originale Hintergrund  $B$  durch einen neuen Hintergrund  $B'$  fehlerfrei ersetzt werden. [P84, W07, L08]

## 5.2 Methoden

Es wurden viele verschiedene Ansätze entwickelt, um das Problem der starken Unterbestimmung der Variablen zu lösen – mit drei Eingangswerten sollten sieben Unbekannte berechnet werden. Bei früheren Matting-Systemen wurde ein Objekt vor einem einfarbigen Hintergrund – meist blau oder grün – aufgenommen. Dieses Verfahren bezeichnet man als „Blue/Green Screen Matting“. Auf Grund der bekannten Hintergrundfarbe müssen nur mehr die beiden anderen Werte  $F$  und  $\alpha$  bestimmt werden. Da bei diesen Algorithmen eine Farbe, nämlich die Hintergrundfarbe, gänzlich herausgefiltert wird, darf der Vordergrund diese selbstverständlich nicht enthalten. Diese Methode ist somit nicht für die Segmentierung natürlicher Bilder geeignet. Um gute Resultate zu erzielen, arbeiten daher die meisten derzeit gängigen Verfahren mit einem (natürlichen) Originalbild und einer zusätzlichen Benutzereingabe – üblicherweise in Form einer sogenannten „Trimap“. [W07, S96]

### 5.2.1 Trimap

Der Benutzer unterteilt vorab das ursprüngliche Bild in drei Bereiche: Vordergrund, Hintergrund und unbekannt. Die Karte dieser drei markierten Regionen wird als „Trimap“ bezeichnet. Ein Beispiel für eine Trimap ist in Abb. 43 zu sehen.



Abbildung 43: Ursprüngliches Bild (links) und Trimap (rechts). Abbildung aus [WC07]

Das Matting-Problem wird dadurch auf die Bestimmung der Werte  $B$ ,  $F$  und  $\alpha$  für Pixel in der unbekanntenen Region reduziert, wobei Vordergrund- und Hintergrundpixel bekannt sind. Dabei gilt, je kleiner der als „unbekannt“ markierte Bereich ist, desto bessere Ergebnisse kann der Matting-Algorithmus liefern. Im Idealfall sollten nur die tatsächlich gemischten Pixel als unbekanntene Region gekennzeichnet sein. Daraus lässt sich ableiten, dass Ansätze, die eine Trimap verwenden, üblicherweise Schwierigkeiten mit Bildern haben, die einen großen Anteil an gemischten Pixel haben – wie zum Beispiel bei großen halbtransparenten Regionen – oder bei denen das Vordergrundobjekt (viele) Löcher aufweist. In diesen Fällen ist sowohl viel Erfahrung als auch Aufwand auf Seiten des Benutzers erforderlich, um gute Resultate zu erzielen. Das Hauptproblem liegt darin, dass der Benutzer

die Alpha-Matte an den wichtigsten Stellen – den gemischten Pixeln – nicht direkt beeinflussen kann. [W07, L08, R08]

### 5.2.2 Closed-form Matting

Eine mögliche Lösung des Matting-Problems mit Hilfe einer Trimap ist der „Closed-form Solution“-Ansatz von A. Levin und anderen [L08]. Diese Methode benötigt neben dem (natürlichen) Originalbild eine grobe, handgezeichnete Trimap („Scribbles“) als Eingangsdaten. Die Berechnung der Alpha-Matte erfolgt mittels Optimierungsfunktion, die auf der Annahme beruht, dass in einem kleinen Bereich  $w$  (von zum Beispiel  $5 \times 5$  Pixel), die Vordergrundfarben  $F$  und Hintergrundfarben  $B$  auf einer Linie im RGB-Farbraum liegen – also, dass jedes Pixel von  $F$  beziehungsweise  $B$  eine lineare Mischung von nur zwei Farben ist. Man bezeichnet diese Annahme auch als „Color Line Model“. Die Alpha-Werte sind in einem kleinen Fenster  $w$  daher definiert als:

$$\alpha_i \approx \sum_c a^c I_i^c + b \quad \forall i \in w,$$

wobei  $c$  auf den Farbkanal verweist und  $a^c$  und  $b$  innerhalb eines Fensters konstant sind. Daraus leitet sich folgende Matting-Optimierungsfunktion für RGB-Bilder ab:

$$J(\alpha, a, b) = \sum_{j \in I} \left( \sum_{i \in w_j} \left( \alpha_i - \sum_c a_j^c I_i^c - b_j \right)^2 + \epsilon \sum_c a_j^{c^2} \right).$$

Schließlich können die beiden Konstanten  $a^c$  und  $b$  aus der Optimierungsfunktion entfernt werden. Dies liefert eine quadratische Optimierungsfunktion in  $\alpha$ :

$$J(\alpha) = \alpha^T L \alpha.$$

Die als „matting Laplacian“ bezeichnete Matrix  $L$  ist eine  $N \times N$  Matrix, deren Element  $(i, j)$  durch:

$$\sum_{k | (i,j) \in w_k} \left( \delta_{ij} - \frac{1}{|w_k|} \left( 1 + (I_i - \mu_k) \left( \Sigma_k + \frac{\epsilon}{|w_k|} I_3 \right)^{-1} (I_j - \mu_k) \right) \right)$$

definiert ist.  $\Sigma_k$  ist eine  $3 \times 3$  Kovarianzmatrix,  $\mu_k$  ist ein  $3 \times 1$  Mittelwertvektor der Farben in einem Fenster  $w_k$  und  $I_3$  ist eine  $3 \times 3$  Einheitsmatrix.

Die optimalen Alpha-Werte werden mittels:

$$\alpha = \arg \min \alpha^T L \alpha, s. t. \alpha_i = 1 \text{ oder } 0, \forall i \in \partial \Omega$$

ermittelt. Wie anhand der Herleitung zu sehen ist, werden die Alpha-Werte direkt berechnet. Das bedeutet, dass keine verlässliche Schätzung der Vordergrund- und Hintergrundfarben zur Bestimmung der Alpha-Matte notwendig ist. Erst nach Berechnung der Alpha-Matte werden der Vordergrund  $F$  und in manchen Fällen auch der Hintergrund  $B$  des Originalbildes rekonstruiert. Um möglich Artefakte zu vermeiden, werden bevorzugte Glättkriterien definiert. Ein Beispiel für Berechnung der Alpha-Matte sowie für die Rekonstruktion von Vorder- und Hintergrund ist in Abb. 44 zu sehen. [L08, W07]



Abbildung 44: Closed-form Matting. a) Trimap; b) Alpha-Matte; c) Rekonstruktion des Vordergrundes; d) Rekonstruktion des Hintergrundes; e) Vordergrund mit neuem Hintergrund. Abbildung aus [L08]

Neben dem „Closed-form Solution“-Ansatz gibt es zahlreiche weitere Verfahren zur Lösung der Matting-Gleichung, auf die in dieser Arbeit nicht näher eingegangen wird. Für Beschreibungen anderer Matting-Methoden wird auf „Image and Video Matting: A Survey“ von J. Wang und M. Cohen [W07] verwiesen.

### 5.3 Alpha Matting und Stereo Vision

„Stereo Matching“ ist ein zentrales Forschungsthema im Bereich des maschinellen Stereosehens („Stereo Vision“). Es beschäftigt sich mit der Problematik in einem Stereobildpaar korrespondierende Punkte zu finden. Diese Algorithmen werden einerseits zur maschinellen Objekterkennung und andererseits zur Erstellung von Tiefenkarten verwendet. Letztere Anwendung hat durch die Einführung der 3D-TV Technologie an Bedeutung gewonnen. Sowohl bei den 3D-Video-Formaten der stereoskopischen als auch der autostereoskopischen Bildschirmsysteme werden meist nicht alle Ansichten übertragen, um die Datenraten möglichst gering zu halten. Die für eine vollständige 3D-Darstellung fehlenden Ansichten werden mit Hilfe der Tiefenkarte beim Empfänger berechnet und anschließend am Bild-

schirm dargestellt. Zur Gewährleistung hochwertiger, synthetischer Ansichten ist somit eine möglichst exakte Tiefenkarte unumgänglich.

Stereo-Matching Algorithmen erhalten ein Stereobildpaar, dessen zwei Einzelbilder aus leicht unterschiedlichen Blickwinkeln aufgenommen wurden, als Eingangsdaten. In Anlehnung an das menschliche visuelle System wird das aufnehmende Kamerasystem so kalibriert, dass die beiden Bilder keine vertikale Verschiebung aufweisen. Die Aufgabe der Stereo-Matching Algorithmen ist das Lösen des Korrespondenzproblems, also das Auffinden von Punkten in beiden Bildern, die denselben Punkt in der Szene repräsentieren, wobei korrespondierende Punkte auf einer horizontalen Linie liegen. Der Abstand zwischen einem Pixel in der linken Ansicht und dem korrespondierenden Pixel in der rechten Ansicht bezeichnet man als Disparität. Auf Grund der berechneten Disparitätswerte können die räumlichen Koordinaten eines Punktes bestimmt werden. Ein Beispiel ist in Abb. 45 zu sehen.

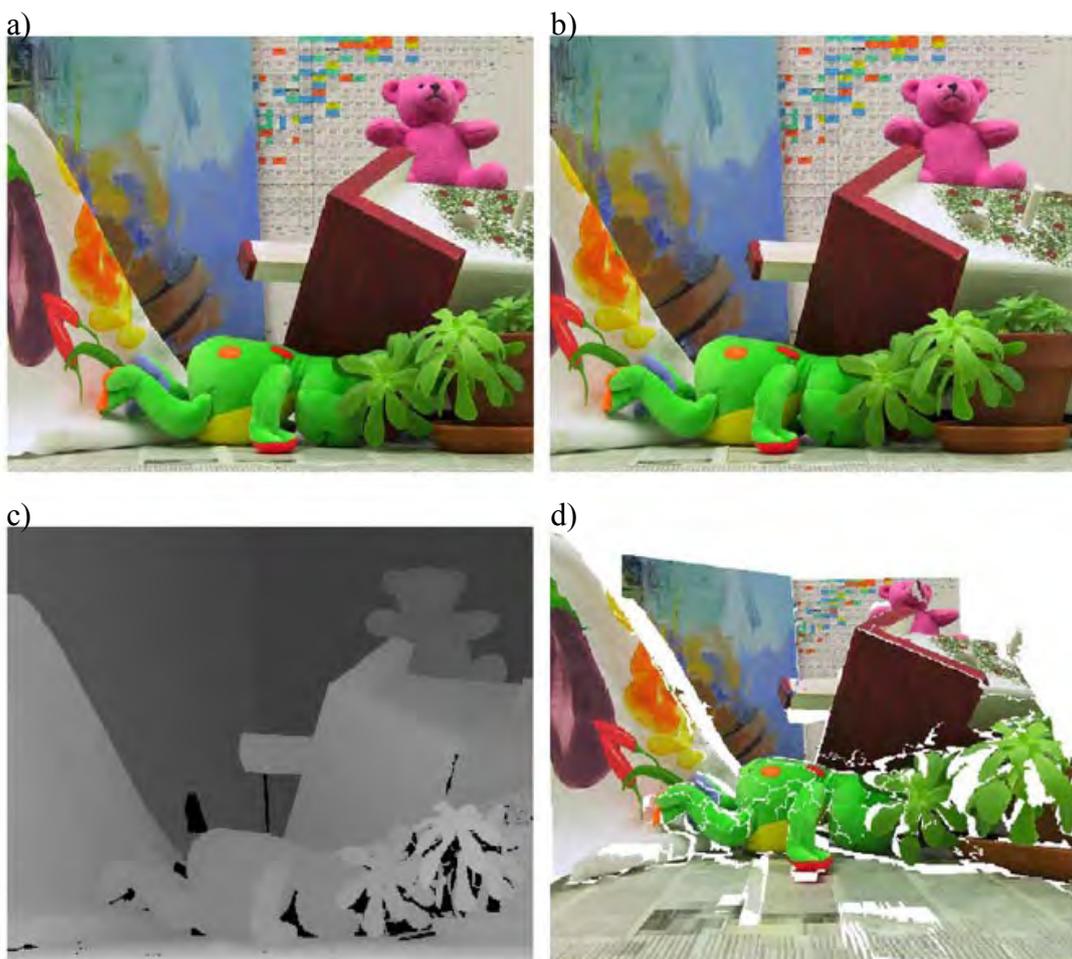


Abbildung 45: Stereo-Matching. a) linke Ansicht; b) rechte Ansicht; c) Disparität/ Tiefenkarte; d) Rekonstruktion der Szene. Abbildung aus [B05]

Die Ermittlung korrekter Disparitätswerte erweist sich häufig als schwierig. Beeinträchtigende Faktoren sind unter anderem Bildrauschen, untexturierte Flächen und Verdeckung. Zur Vereinfachung des Problems müssen zwei Annahmen getroffen werden: jene der Eindeutigkeit und jene der Stetigkeit. Konventionelle Stereo-Matching Algorithmen setzen also voraus, dass für jedes Pixel eine eindeutige Übereinstimmung existiert, sowie, dass die Disparitätswerte innerhalb einer Region mit homogener Farbe geringe Schwankungen aufweisen und Tiefenunstetigkeiten mit Objektgrenzen zusammenfallen. [B05]

Bei der Suche nach zusammengehörigen Punkten geht man davon aus, dass die Farbe von korrespondierenden Pixeln in den Eingabebildern konstant ist. Diese Annahme der Farbkonstanz ist jedoch bei Pixeln, die nahe oder auf der Objektgrenze liegen, nicht immer korrekt, da diese Pixel üblicherweise eine Mischung aus Vordergrund- und Hintergrundfarbe sind. Auf Grund der Parallaxe – also der unterschiedlichen Verschiebung von Vordergrund und Hintergrund zueinander – entstehen an diesen Stellen bei korrespondierenden Pixeln andere Mischfarben. Abb. 46 verdeutlicht das Problem der gemischten Pixel in einem Stereobildpaar. [S98, B09, X07]

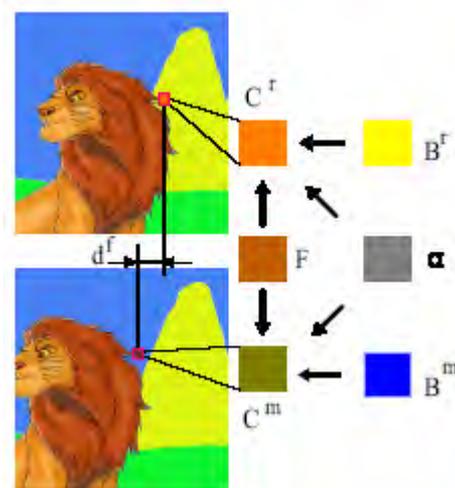


Abbildung 46: Farbkonstanz in einem Stereobildpaar. Abbildung aus [X07]

An Stellen mit gemischten Pixel kommt es häufig zu falschen Zuordnungen und in weiterer Folge zu Artefakten, wie Schattenbilder („halo“) oder Farbübertragungen („color bleeding“), in neu zusammengesetzten Bildern. Das Problem kann gelöst werden, indem mit Hilfe eines Stereobildpaars nicht nur die Querdistanz, sondern auch die wahre Farbe und Deckkraft jeder sichtbaren Oberfläche bestimmt wird. Zur Ermittlung der zuletzt genannten Werte ist eine Trennung von Vorder- und Hintergrund notwendig – also das Lösen der Matting-Gleichung.

Das Zusammenspiel von Stereo-Matching und Alpha-Matting, das als „Stereo-Matting“ bezeichnet wird, bringt beiden Bereichen Vorteile für das Lösen der jeweiligen Probleme. Der Nutzen auf dem Gebiet Stereo-Matching ist die Zerlegung von gemischten Pixeln in Vorder- und Hintergrundfarbe, um so die Annahme der Farbkonstanz auch in den Regionen am Rand eines Objektes aufrecht zu erhalten. Alpha-Matting profitiert von der Existenz eines zweiten Bildes, das zum einen keine Benutzereingabe notwendig macht und zum anderen Zweideutigkeiten vermeidet. [B09, S98, X07]

## 6 Benutzerstudie

Die Qualität von 3D-Fernsehen ist – wie in Kapitel 4.3 erörtert – eng mit einer fehlerfreien Generierung von synthetischen Ansichten („view synthesis“) verbunden. Um allen Grundanforderungen gerecht zu werden, arbeiten die erweiterten 3D-Videoformate mit mehreren Ebenen von Farbe, Transparenz und Tiefe. H. Kannan und andere [K10] haben zur Bestimmung der Farb- und Transparenzwerte den Einsatz von Alpha-Matting vorgeschlagen, um Artefakte in Regionen mit gemischten Pixeln zu reduzieren. Abb. 47 zeigt, dass die Bildqualität an Objektkanten verbessert werden kann, wenn Alpha-Matting zur Berechnung synthetischer Ansichten herangezogen wird. [K10]

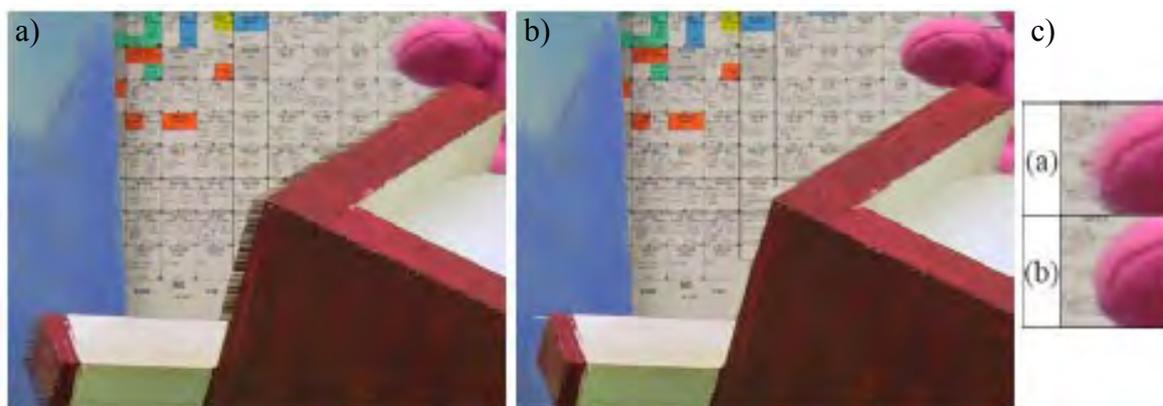


Abbildung 47: Ausschnitt einer synthetischen Ansicht. a) ohne spezielle Berechnung der gemischte Regionen; b) mit Alpha-Matting („Closed-form Matting“); c) Vergrößerung einer Region. Abbildung aus [K10]

Sowohl das in Kapitel 5.3 beschriebene Stereo-Matting-Verfahren als auch das eben gebrachte Beispiel sprechen für einen standardmäßigen Einsatz von Alpha-Matting bei der Berechnung von neuen Ansichten beim 3D-Fernsehen. Dem gegenüber stehen Kenntnisse über die menschliche Wahrnehmung, die vermuten lassen, dass bei der Betrachtung von dreidimensionalen Bildern Artefakte in den Einzelbildern bis zu einem gewissen Grad durch das visuelle System des Menschen ausgeglichen werden. In diesem Zusammenhang spielen wahrnehmungspsychologische Faktoren, wie beispielsweise die Organisation der visuellen Information (siehe Kapitel 2.2), die Objektwahrnehmung (siehe Kapitel 2.3), die binokulare Unterdrückung (siehe Kapitel 3.3.5) und okulare Dominanz (siehe Kapitel 3.3.6), ebenso wie die Erfahrung des Menschen eine wichtige Rolle.

Ziel der Benutzerstudie ist es festzustellen, ob Probanden einen Unterschied in der Bildqualität wahrnehmen, wenn eines der 3D-Bilder mit und das andere ohne Alpha-Matting erstellt wurde. Im Zuge der Befragung soll zusätzlich herausgefunden werden, ob die sichtbaren Unterschiede mit speziellen Hintergründen, bestimmten Grenzwerten oder einzelnen Motiven in Verbindung gebracht werden können. Die Vorbereitungen zur Benutzerstudie beinhalten die Auswahl und Erstellung der Testbilder, sowie eine Vorstudie zur Optimierung der Hauptstudie. Nach der Durchführung der Benutzerstudie werden die Ergebnisse im Detail ausgewertet und analysiert.

## 6.1 Planung

Die Benutzerstudie wird mit Hilfe der 3D-Technologie von NVIDIA „3D Vision“ durchgeführt. Das 3D-System besteht aus Shutter-Brille und LCD-Bildschirm mit 120 Hz Bildwiederholungsrate. Eine ausführliche Beschreibung der Shutter-Technologie ist unter Kapitel 4.2.1.3 zu finden. Die verwendeten Komponenten sind beispielhaft in Abb. 48 dargestellt. [NV]



Abbildung 48: NVIDIA „3D Vision“-System. NVIDIA 3D Vision Kit + 120 Hz Monitor. Abbildung aus [NV].

Als Eingangsdaten benötigt das „3D Vision“-System ein Bild, das sowohl eine rechte als auch eine linke Ansicht beinhaltet. Vor der Generierung der künstlichen 3D-Testbilder, erfolgt eine Zusammenstellung von Bildern mit geeigneten Motiven und die Wahl eines zweckmäßigen Bildhintergrundes.

### 6.1.1 Auswahl der Bilder

Im Zuge der Literaturrecherche hat sich Transparenz in natürlichen Bildern als Herausforderung für Algorithmen zur Berechnung von synthetischen Ansichten herauskristallisiert. Transparenz entsteht sowohl durch transparente Objekte als auch durch Unschärfe (Tiefenunschärfe, Bewegungsunschärfe) oder feine Struktur (Haare, Fell, Federn). In all diesen Fällen kommt es zu einer Überblendung von Vorder- und Hintergrund – also zu Regionen mit gemischten Pixeln. Aber auch scharf abgegrenzte Objekte weisen naturgemäß auf Grund der Halb-Verdeckung (siehe Kapitel 3.3.3) ein gewisses Maß an Transparenz an ihren Rändern auf. Jedes der insgesamt zwanzig ausgewählten Motive wird eindeutig einer der vier Kategorien Bewegungsunschärfe, Tiefenunschärfe, Struktur und „normale“ Objekte zugeordnet. Drei Beispiele sind in Abb. 49 zu sehen.



Abbildung 49: Testbilder (Auswahl). a) Rose (Tiefenunschärfe); b) Auto (Bewegungsunschärfe); c) Hase (Struktur). Abbildung nach [R09].

Als Bildhintergründe wurden zwei gemusterte Hintergründe ausgewählt, da Hintergründe in einer konstanten Farbe in der Natur kaum vorkommen. Der eine Hintergrund enthält ein verwaschenes, eher einfarbiges Muster und der andere eine klare, starke Struktur. Beide Hintergründe weisen eine leichte Unschärfe auf, weil andernfalls die neu zusammengesetzten Bilder übertrieben künstlich und unnatürlich wirken könnten. Die beiden verwendeten Hintergründe sind in Abb. 50 dargestellt.

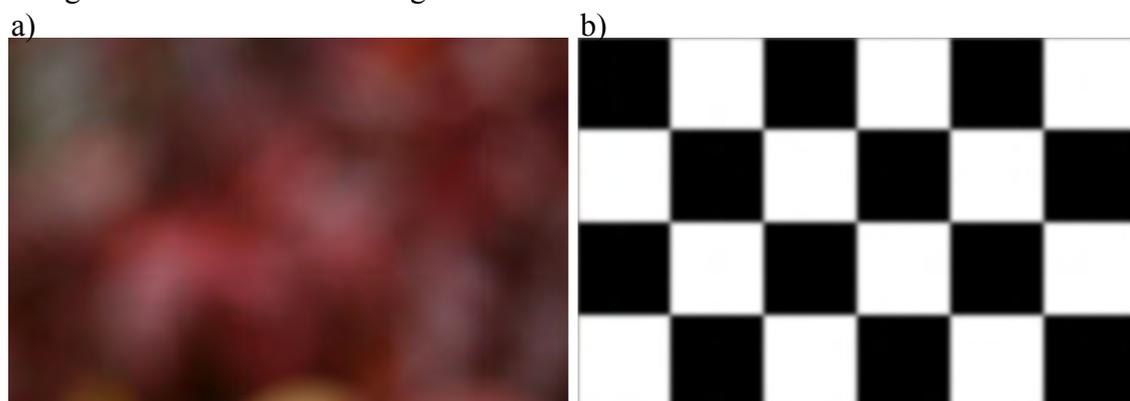


Abbildung 50: Hintergrundbilder. a) Muster; b) Struktur.

### 6.1.2 Erstellung der 3D-Bilder

Jedes der zwanzig ausgewählten Motive wird vor die beiden in Kapitel 6.1.1 abgebildeten Hintergründe platziert. Die Zusammensetzung mit einem neuen Hintergrund ist notwendig, da der Hintergrund des Originalbilds nicht vollständig vorhanden ist. Dies hätte Löcher im Hintergrund der synthetischen, zweiten Ansicht zur Folge. Des Weiteren ermöglicht ein einheitlicher Hintergrund ein kontrollierbares Vergleichen der Bilder. Um die Objekte vor einem neuen Hintergrund einzufügen, müssen diese zuerst mittels Alpha-Matte aus dem Originalbild herausgelöst werden.

Ein Teil der Bilder mit bekannter Alpha-Matte (Ground Truth) stammt aus der Datenbank des Instituts für Softwaretechnik und Interaktive Systeme, Arbeitsgruppe für Interaktive Multimediale Systeme, der Technischen Universität Wien. Für alle anderen Bilder wird die Alpha-Matte mit Hilfe des Closed-Form-Matting Algorithmus, dessen Funktionsweise in Kapitel 5.2.2 beschrieben wird, berechnet. Mittels der Alpha-Matting-Information werden die Motive mit jedem der beiden Hintergründe zusammengesetzt. Abb. 51 zeigt den Entstehungsprozess eines Bildes an Hand eines Beispiels.

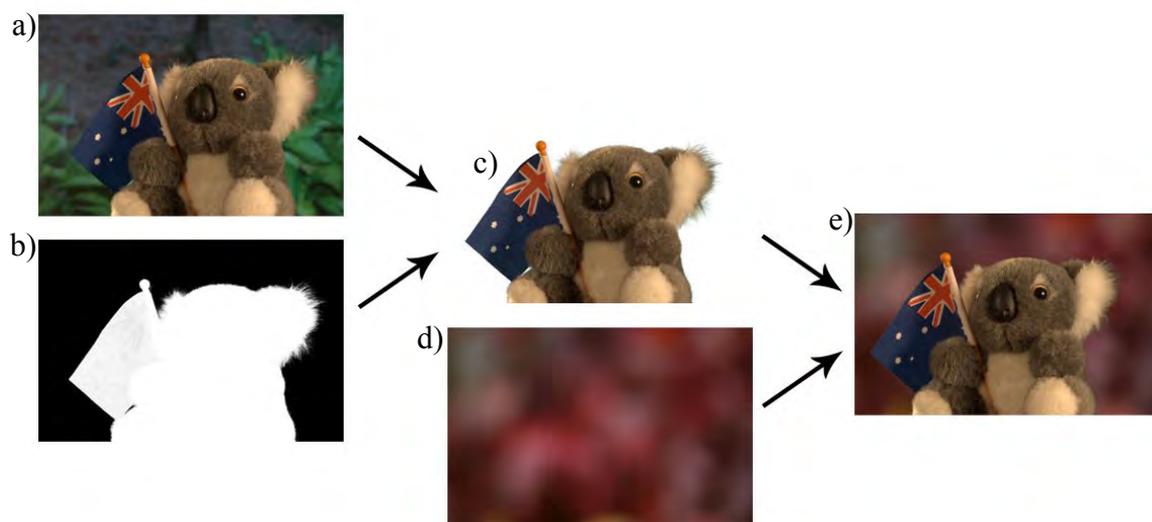


Abbildung 51: Erstellung eines Bildes. a) Originalbild. Abbildung nach [R09]; b) Alpha-Matte. Abbildung nach [R09]; c) freigestelltes Motiv; d) Hintergrund; e) Motiv vor neuem Hintergrund.

Die auf diese Weise erstellten Bilder werden als Basisansichten eines Stereobildpaars herangezogen. Die jeweils zweiten Ansichten, die für die 3D-Wahrnehmung notwendig sind, entstehen durch Verschiebung des Vordergrundobjektes. Zur Bestimmung, welche Pixel zum Vordergrund gehören, wurden bei einer Gruppe von Testbildern die Alpha-Matte und bei der anderen Gruppe drei unterschiedliche Grenzwerte („Threshold“) herangezogen. Diese Schwellenwerte sind so gewählt, dass das Motiv entweder knapp („Threshold 200“), moderat („Threshold 128“) oder grob („Threshold 60“) ausgeschnitten wird. Während die Testbilder mit Alpha-Matting fehlerfreie Ergebnisse enthalten, kann es in der anderen Gruppe zu Artefakten, wie beispielsweise fehlende Struktur, oder zur Mitnahme des Hintergrundes kommen. Abb. 52 verdeutlicht die unterschiedlichen Resultate.

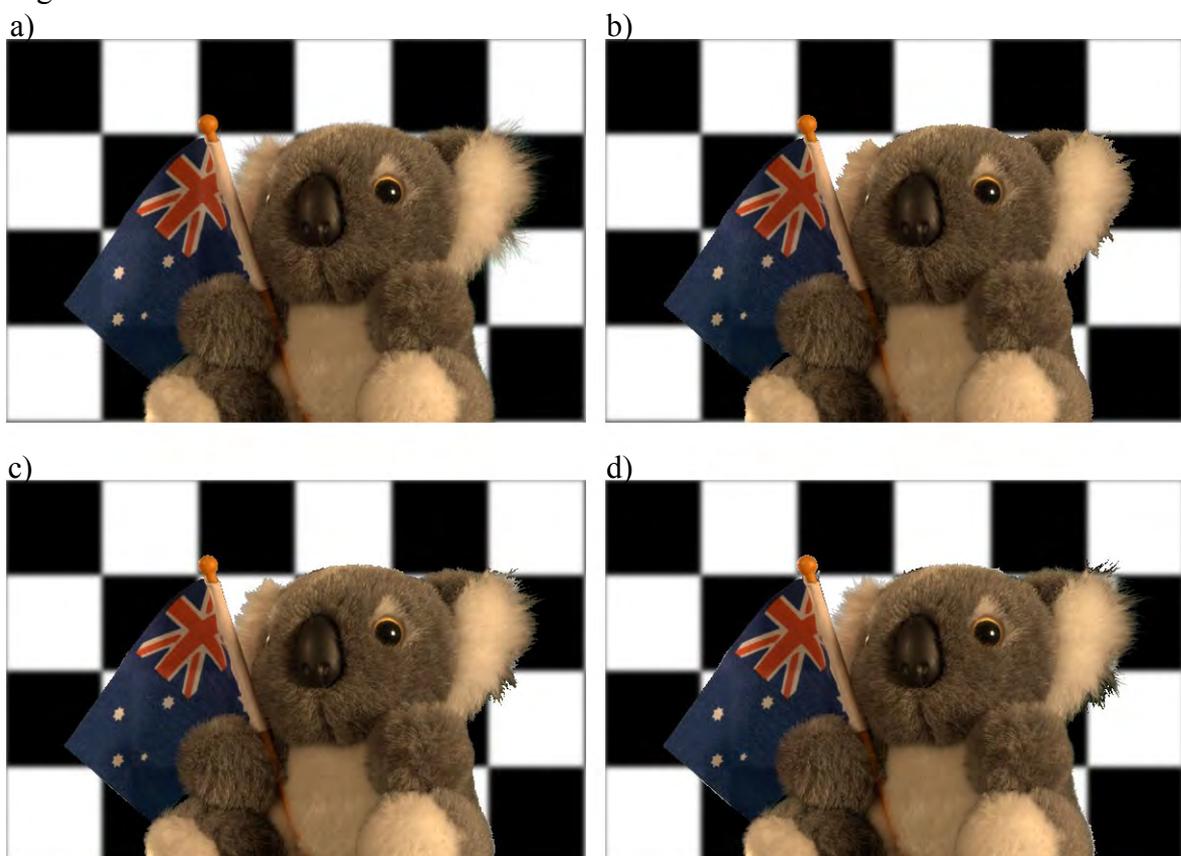


Abbildung 52: Synthetische Ansicht. a) mit Alpha-Matting; b) – d) ohne Alpha Matting: b) „Threshold 200“; c) „Threshold 128“; d) „Threshold 60“.

Für jedes der zwanzig Motive werden vier Stereobildpaare erzeugt. Diese bestehen aus der Basisansicht und jeweils einer der zuvor abgebildeten vier synthetischen Ansichten. Ein Beispiel für ein solches Stereobildpaar ist in Abb. 53 dargestellt.



Abbildung 53: Stereobildpaar. Basisansicht (links) und synthetische Ansicht ohne Alpha-Matting (rechts).

Auf Grund der in Kapitel 3.3.6 beschriebenen okularen Dominanz werden zwei Datensätze an Stereobildpaaren erzeugt. Bei dem Datensatz für Probanden mit dominantem, rechtem Auge ist die linke Ansicht die Basisansicht und die rechte die künstlich erzeugte Ansicht. Für Personen mit dominantem, linkem Auge ist es genau umgekehrt. Dies soll sicherstellen, dass die synthetischen Ansichten vom dominanten Auge wahrgenommen werden. Eine Unterdrückung der variierenden Ansicht durch das menschliche visuelle System kann die Entscheidung des Probanden und somit das Ergebnis der Studie beeinflussen. Im Zuge der Vorstudie gilt es daher zu klären, ob die Augendominanz tatsächlich berücksichtigt werden muss oder für den weiteren Verlauf der Studie vernachlässigbar ist.

## 6.2 Vorstudie

Die Vorstudie wird mit wenigen Personen durchgeführt und dient als Probelauf für die eigentliche Benutzerstudie. Das Ziel der Vorstudie ist die Optimierung der Hauptstudie in Bezug auf Inhalt, Durchführung und Qualitätssicherung. Im Rahmen der Vorstudie werden unter anderem die Auswahl und Darstellung der Bilder, das geplante Durchführungskonzept, der zeitliche Rahmen und der Dokumentationsprozess, sowie die Rolle der Augendominanz getestet.

### 6.2.1 Vorbereitung

Die zwanzig Motive werden bei der Vorstudie nur vor einem Hintergrund gezeigt, da die Vielzahl der Testbilder anderenfalls den zeitlichen Rahmen bei Weitem übersteigen würde. Die insgesamt sechzig dreidimensionalen Testbilder, die jeder Proband am Bildschirm zu sehen bekommt, bestehen immer aus zwei 3D-Bildern. Bei einem der beiden 3D-Bilder ist

die synthetische Ansicht mit Alpha-Matting berechnet worden, bei dem anderen mit einem der drei Grenzwerte. Es wird darauf hingewiesen, dass das Stereobild mit Alpha-Matting das optimale Ergebnis darstellt und immer als Referenz dient. Für die Anordnung der beiden Stereobildpaare ergeben sich zwei Möglichkeiten: nebeneinander oder untereinander. Die Entscheidung fiel eindeutig auf nebeneinander, da auf diese Weise der Breitbild-Monitor besser ausgenutzt werden kann. Abb. 54 zeigt die Anordnung der Stereobildpaare A und B. Auf der linken Seite befinden sich die linken Ansichten der Stereobildpaare, die mit Hilfe der Shutter-Brille an das linke Auge geleitet werden. Analog dazu verhält es sich mit der rechten Seite und den rechten Ansichten. Durch diese Anordnung sieht der Proband zwei nebeneinander liegende 3D-Bilder.

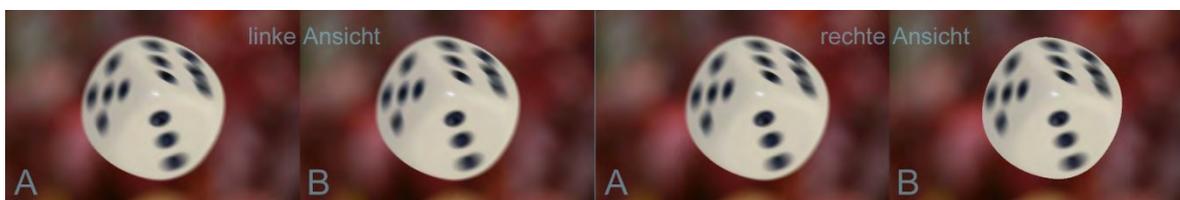


Abbildung 54: Anordnung der Stereobildpaare.

Der Einfluss der Augendominanz wird mit Hilfe von fünf speziell ausgewählten Stereobildpaaren untersucht. Jedes Stereobildpaar besteht aus der Basisansicht und einer synthetischen Ansicht, die eindeutig Artefakte aufweist. Die daraus erstellten Testbilder beinhalten jeweils zweimal dasselbe Stereobildpaar, jedoch ist das eine aus dem Datensatz für Probanden mit rechtem, dominantem Auge und das andere aus jenem für Probanden mit linkem, dominantem Auge. In Abb. 55 zeigt eines der Augendominanz-Testbilder.

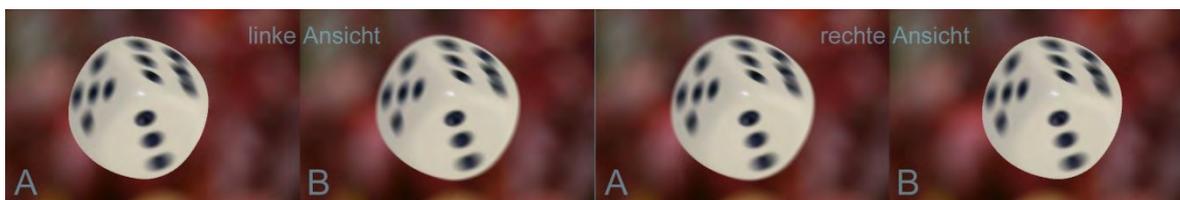


Abbildung 55: Augendominanz-Testbild.

Die Aufzeichnung der Antworten und Kommentare der Probanden erfolgt mittels eigens angefertigten Formulars durch den Studienleiter. Diese Vorgehensweise ist notwendig, da die 3D-Software keine Interaktion zulässt und zugleich die Testperson durch die 3D-Brille stark eingeschränkt ist. Aus diesem Grund können sich die Probanden nicht selbst das nächste Testbild aufrufen.

## 6.2.2 Durchführung

An der Vorstudie haben vier Personen – zwei Männer und zwei Frauen im Alter von 24 bis 28 Jahren – teilgenommen. Die Probanden bekamen zu Beginn einen Überblick über den Inhalt, die Aufgabenstellung und den Ablauf der Studie. Außerdem wurden sie aufgefordert während der gesamten Befragungsdauer „laut“ zu denken – also Bemerkungen im Zusammenhang mit der Studie auszusprechen.

Die Vorstudie ist in zwei Teile unterteilt. Der eine Teil befasst sich mit der okularen Dominanz, der zweite Teil mit der Wahrnehmung von 3D-Bildern. Da die meisten Menschen – mit sehr wenigen Ausnahmen – sich ihres dominanten Auges nicht bewusst sind, wurde dies mit Hilfe des in Kapitel 3.3.6 beschriebenen „Miles Test“ ermittelt. Die Bestimmung des dominanten Auges ergab, dass bei drei Personen das rechte, bei einer das linke Auge dominant ist. Anschließend setzten die Probanden die Shutter-Brille auf und bekamen die fünf Testbilder zur okularen Dominanz am Bildschirm zu sehen. Diese wurden in beliebiger Reihenfolge gezeigt. Diese Maßnahme beugt einer Verzerrung der Ergebnisse auf Grund einer bestimmten Bildfolge vor. Ebenso wurde die Position der beiden Stereobildpaare zufällig festgelegt. Das heißt, dass das 3D-Bild aus dem Datensatz für Personen mit dominantem, rechtem Auge sich entweder auf der linken oder auf der rechten Seite befinden kann. Die Frage an die Probanden war, ob sie einen Unterschied zwischen dem rechten und dem linken 3D-Bild wahrnehmen. Die Antworten sollten Auskunft über den möglichen Einfluss des dominanten Auges geben.

Im zweiten Teil der Befragung wurden die sechzig Testbilder zur Wahrnehmung von 3D-Bildern mit und ohne Alpha-Matting gezeigt. Probanden mit dominantem linkem Auge erhielten den Datensatz, dessen linke Ansicht synthetisch erstellt wurde, Probanden mit dominantem rechtem Auge jeweils den anderen. Die Aufeinanderfolge der Motive, die Position des 3D-Bildes mit beziehungsweise ohne Alpha-Matting, sowie die Reihenfolge der Grenzwerte innerhalb einer Dreier-Motiv-Gruppe, wurden zufällig gewählt. Wie bereits beim ersten Teil der Vorstudie hatten die Probanden die Aufgabe festzustellen, ob es sichtbare Unterschiede zwischen den beiden dargestellten 3D-Bildern gibt. Bei Beantwortung dieser Frage mit „Ja“ sollte zusätzlich eine Präferenz für jenes der beiden Bilder abgegeben werden, das echter, realistischer, natürlicher erscheint. Die Probanden wurden aufgefordert anzugeben, in welchem Bereich des Bildes sie einen Unterschied wahrnehmen und welches Kriterium zu ihrer Entscheidung geführt hat. Neben einem ersten Trend, soll auf diese Weise auch die Eignung der Bilder sowie deren Darstellung, der zeitliche

Umfang, die Klarheit der Aufgabenstellung und die Vollständigkeit der notwendigen Vorinformation getestet werden. Die laufende Beobachtung der Probanden wurde durch abschließende Fragen ergänzt.

### 6.2.3 Ergebnisse

Die Auswertung der Vorstudie lieferte dank der Gewissenhaftigkeit der Probanden wertvolle Erkenntnisse und Hinweise für Änderungen in Bezug auf die Hauptstudie. Alle Probanden befanden die Einleitung für gut. Ebenso wurden die Aufgabenstellungen als klar und verständlich empfunden. Die Art der Durchführung konnte ebenfalls bestätigt werden, lediglich die Dauer der Befragung von rund einer Stunde haben die Testpersonen mit „gegen Ende hin anstrengend, aber noch akzeptabel“ bewertet.

Die Rolle der Augendominanz konnte nicht eindeutig definiert werden. Sowohl Probanden mit dominantem linkem also auch mit dominantem rechtem Auge nahmen bei dem einen oder anderen Bild einen leichten Unterschied wahr, der auf einen Einfluss des dominanten Auges deutet. Das Ausmaß ist jedoch nicht ausreichend signifikant, um die okuläre Dominanz in der Hauptstudie in diesem Umfang zu berücksichtigen.

Mit Hilfe des zweiten Teils der Vorstudie konnte generell ein Trend Richtung Alpha-Matting festgestellt werden. Bei den Kategorien Bewegungsunschärfe und Struktur wurde für das 3D-Bild mit Alpha-Matting auf Grund von sichtbaren, störenden Artefakten in den Bildern ohne Alpha-Matting entschieden. Im Gegensatz dazu wurde bei den Kategorien Tiefenunschärfe und „normale“ Objekte auch immer wieder aus dem Bauch heraus bewertet. Das Stereobild ohne Alpha-Matting wurde dabei oft als unpassend und komisch empfunden. Die Auswertung der Bilder hat außerdem ergeben, dass Artefakte in Bildregionen, die zu transparenten Objekten gehören, mehrmals für die Entscheidung ausschlaggebend waren und andere Fehler in dem 3D-Bild unbeachtet blieben. Dies legt „Transparenz“ als eigene Kategorie bei der Hauptstudie nahe.

In einigen Fällen gingen die Antworten auf die Frage, welches Bild ihrer Meinung nach realistischer ist, weit auseinander. Bei näherer Betrachtung der Begründungen, die die Probanden für ihre Entscheidungen angaben, ist erkennbar, dass bei diesen Bildern jeweils unterschiedliche Regionen betrachtet und bewertet wurden. Öfters wurde auch Farbe oder Helligkeit als Unterschied zwischen den Bildern genannt. Diese entstehen jedoch durch die Shutter-Brille und den Betrachtungswinkel. Dies sollte bei der Hauptstudie eingangs erwähnt werden, um hier falsche Resultate zu vermeiden.

Eine nicht unwesentliche Anzahl der zu vergleichenden 3D-Bilder wurden von den Probanden als gleich erachtet. Es handelt sich dabei vorwiegend um Bilder der Kategorie „normale“ Objekte und um Bilder, bei denen die zweite Ansicht ohne Alpha-Matting grob („Threshold 60“) ausgeschnitten wurde. Eine mögliche Erklärung für letzteres ist der recht einfarbige, monotone Hintergrund. Zur Überprüfung dieser Vermutung werden in der Hauptstudie die Motive auch vor dem Schachbrettmuster-Hintergrund präsentiert.

### 6.3 Studie

Die Vorstudie hat einige Punkte aufgezeigt, die im Zuge der Vorbereitungen für die Hauptstudie angepasst werden müssen. Die Änderungen umfassen die Adaption der einleitenden Information, die Einbeziehung des zweiten Hintergrundes und eine Unterstützung zur Blickführung der Probanden. Außerdem ist zu entscheiden, ob und in welchem Ausmaß die okulare Dominanz bei der anschließenden Befragung berücksichtigt werden muss.

#### 6.3.1 Vorbereitung

Eine wichtige Erweiterung zur Vorstudie ist die Verwendung des zweiten Hintergrundes, der in Kapitel 6.1.1 bereits abgebildet ist. Die Struktur des Schachbrettmusters führt zu unübersehbaren Fehlern bei der Berechnung der zweiten Ansicht ohne Alpha-Matting, die bei relativ einheitlichem Hintergrund untergehen. Diese Artefakte entstehen hauptsächlich bei jenen Stereobildpaaren, deren synthetische Ansicht durch grobes Ausschneiden des Objektes erzeugt wurde. Abb. 56 zeigt die synthetische Ansicht eines Motives vor unterschiedlichen Hintergründen. Während die künstlich erzeugte Ansicht in Abb. 56a keine sichtbaren Hintergrundfehler aufweist, sind diese in Abb. 56b kaum zu übersehen.

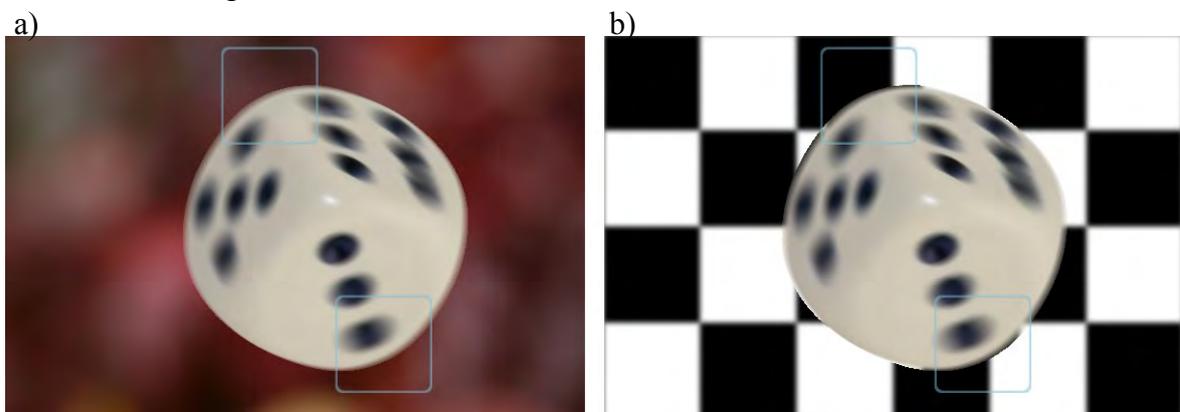


Abbildung 56: Auswirkung unterschiedlicher Hintergründe. a) Würfel („Threshold 60“) mit Muster-Hintergrund; b) Würfel („Threshold 60“) mit Struktur-Hintergrund.

Diese Erweiterung des Umfangs der Testbilder bedeutet jedoch nicht nur einen Informationsgewinn, sondern auch eine notwendige Reduktion der Motive, da sich die Durchführungsdauer verdoppeln würde. Um das zeitliche Maximum von einer Stunde nicht zu überschreiten, werden für die Hauptstudie zehn Motive ausgewählt. Diese zehn Motive werden jeweils einer der fünf Kategorien (Bewegungsunschärfe, Tiefenunschärfe, Struktur, Transparenz und „normale“ Objekte) zugeordnet.

Ein weiterer wesentlicher Punkt ist die Blickführung der Probanden. Diese Maßnahme soll sicherstellen, dass alle Probanden bestimmte Regionen betrachten und auf Grund derselben visuellen Information entscheiden. Ohne die Blickführung könnte es zu widersprüchlichen Antworten der Probanden kommen, wie es in der Vorstudie der Fall war. Die weitere Folge wäre, dass die Auswertung der Daten wenig Aussagekraft hat. Um dies zu verhindern, standen zwei Möglichkeiten zur Auswahl: die Verwendung von Bildausschnitten oder das Kennzeichnen der zu beachtenden Regionen. Die Präsentation von Bildausschnitten anstelle des gesamten Bildes hat den Vorteil, dass die Entscheidung des Probanden nur auf Grund des speziellen Fehlers in der gezeigten Region getroffen wird. Der entscheidende Nachteil bei dieser Methode ist der Verlust des Kontexts, wie zum Beispiel, dass ein Objekt in Bewegung ist. Mit einem zusätzlichen Abbilden des ganzen Bildes in einer Bildschirmecke könnte das Problem umgangen werden. Dies hätte jedoch ein sehr unübersichtliches Gesamtbild zur Folge. Da Information bezüglich des Kontexts für den Probanden essentiell ist, fiel die Entscheidung auf die Kennzeichnung bestimmter Regionen durch ein hellblaues Quadrat. Die Wahl der Farbe „Hellblau“ ergab sich aus der Anforderung, dass das Quadrat in jedem Bild leicht zu erkennen sein muss. Zusätzlich scheint das Quadrat im Vordergrund zu schweben, um sich deutlich vom Inhalt des Bildes abzuheben. Jedes Motiv erhält einheitlich zwei dieser blauen Rahmen, wie das Beispiel in Abb. 57 zeigt. Die Positionen der beiden Quadrate sind vom jeweiligen Motiv abhängig.

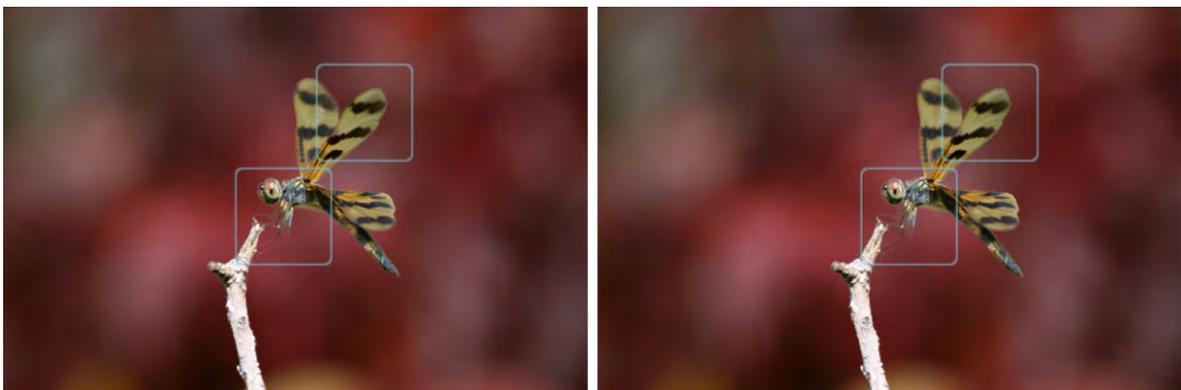


Abbildung 57: Stereobildpaar mit gekennzeichneten Bildregionen.

Da in der Vorstudie keine klare Aussage über die Rolle der okularen Dominanz getroffen werden konnte, gibt es in der Hauptstudie keinen eigenen Datensatz für Probanden mit dominantem, linkem Auge. Diese Entscheidung wurde durch die Tatsache unterstützt, dass bei zwei Drittel der Menschen das rechte Auge dominiert [SS]. Zu Beginn der Studie wird dennoch für jeden Probanden das dominante Auge mittels „Miles-Test“ festgestellt. Dies soll sicherstellen, dass in der Auswertung mögliche Zusammenhänge zwischen den Antworten der Probanden und der Augendominanz erkannt werden.

### **6.3.2 Durchführung**

An der Studie nahmen vierzehn Frauen und zehn Männer im Alter von 21 bis 57 Jahren teil. Der Altersdurchschnitt der insgesamt 24 Personen liegt bei 31 Jahren. Die Bestimmung der Augendominanz hat ergeben, dass bei 13 Probanden das rechte Auge und bei 11 Probanden das linke Auge dominiert. Die Befragung dauerte im Schnitt eine halbe Stunde – also deutlich weniger als die, in der Vorstudie ermittelte, zeitliche Obergrenze von einer Stunde. Keiner der Probanden wies auf Grund seiner täglichen Tätigkeit ein besonderes Vorwissen in Bezug auf Stereoskopie, 3D-Technologie oder Bilderwahrnehmung auf. Alle Testpersonen haben Erfahrung mit künstlichen 3D-Inhalten durch Kinofilme und Autostereogramme („single-image stereogram“).

Zu Beginn der Studie bekamen die Probanden eine kurze Einleitung über Inhalt, Ablauf und Aufgabenstellung. Da sich das Konzept des „laut Denkens“ bereits in der Vorstudie bewährt hat, wurde es für die Hauptstudie beibehalten. Die Probanden wurden angehalten, jede Überlegung beziehungsweise jeden Gedanken während der Entscheidungsfindung auszusprechen, weil auf diese Weise wertvolle Information gewonnen werden kann. Nach Ermittlung der okularen Dominanz mit Hilfe des in Kapitel 6.2.2 beschriebenen „Miles-Tests“, hatten die Probanden die Aufgabe, sechzig Testbilder, die sie mittels Brille und Monitor in 3D sahen, auf wahrnehmbare Unterschiede zwischen linkem und rechtem Bild zu begutachten. Die Aufeinanderfolge der Motive, die Position des 3D-Bildes mit beziehungsweise ohne Alpha-Matting, sowie die Reihenfolge der Grenzwerte innerhalb einer Dreier-Motiv-Gruppe wurden, wie auch in der Vorstudie, zufällig gewählt. Bei festgestellten Unterschieden sollten die Probanden angeben, welches der beiden gezeigten 3D-Bilder realistischer, echter, natürlicher wirkt. Die Testpersonen wurden aufgefordert, bei der Beurteilung der Bilder, den hellblau umrahmten Bereichen besondere Aufmerksamkeit zu widmen. Außerdem wurde darauf hingewiesen, dass es keine Farb- und Helligkeitsunterschiede zwischen den beiden Bildern gibt.

## 7 Ergebnisse und Auswertung

Nach Erhebung der Daten werden die Antworten der Probanden nach Hintergrund, Kategorie und Grenzwert sortiert und zusammengefasst. Eine Entscheidung für das Bild mit Alpha-Matting wird mit dem Wert 1 belegt und ohne Alpha-Matting mit -1. Hat die Testperson keinen Unterschied zwischen den beiden gezeigten Bildern feststellen können, wird dies mit 0 vermerkt. Aus einer ersten Analyse mit Hilfe von Mittelwerts- und Abweichungsberechnungen geht bereits hervor, dass es einen deutlichen Trend zu Alpha-Matting gibt und, abgesehen von einigen, wenigen Ausnahmen, weitgehende Einigkeit unter den Probanden herrscht. Es konnten keine signifikanten Unterschiede zwischen Männern und Frauen festgestellt werden. Die Auswertung der Daten in Bezug auf die okulare Dominanz der Probanden hat gelegentliche Abweichungen zwischen den Antworten von Probanden mit linker und rechter Augendominanz aufgezeigt. Auf diese wird im Zuge der detaillierten Analyse der einzelnen Bilder in Kapitel 7.1 näher eingegangen.

### 7.1 Analyse

Die Ergebnisse der einzelnen 3D-Bilder werden aufgeschlüsselt nach den fünf Kategorien Bewegungsunschärfe, Tiefenunschärfe, Struktur, Transparenz und „normale“ Objekte ausgewertet. Jedes der zehn Motive ist einer Kategorie zugeordnet. Die Analyse eines Motives umfasst eine Beschreibung der auftretenden Artefakte, einen Vergleich der beiden Hintergründe sowie eine Gegenüberstellung der drei Grenzwerte. Die Probanden der Studie haben alle Bilder zur Gänze in Farbe mit zwei hellblau umrahmten Bereichen gesehen. Um die 3D-Bilder für den Druck geeignet zu machen, wurden sie in schwarz-weiße Anaglyphenbilder umgewandelt. Die im Zuge der Analyse gezeigten Bilder können mit einer Rot-Blau-Brille dreidimensional betrachtet werden.

### 7.1.1 Kategorie: Bewegungsunschärfe

Die 3D-Bilder dieser Kategorie zeigen Objekte, die in Bewegung sind und dadurch unscharfe Regionen enthalten. In diesen Bereichen kommt es zu einer Überlagerung mit dem jeweiligen Hintergrund. Wird bei der Generierung von synthetischen Ansichten diese Verschmelzung nicht berücksichtigt, treten an diesen Stellen abhängig vom gewählten Schwellenwert unterschiedliche, sichtbare Fehler in den Einzelbildern auf. Die Beispielbilder der Kategorie Bewegungsunschärfe sind ein rotierendes Windrad und ein geworfener Würfel, der sowohl Translationsbewegung als auch Rotationsbewegung aufweist.

#### 7.1.1.1 Bild: Würfel

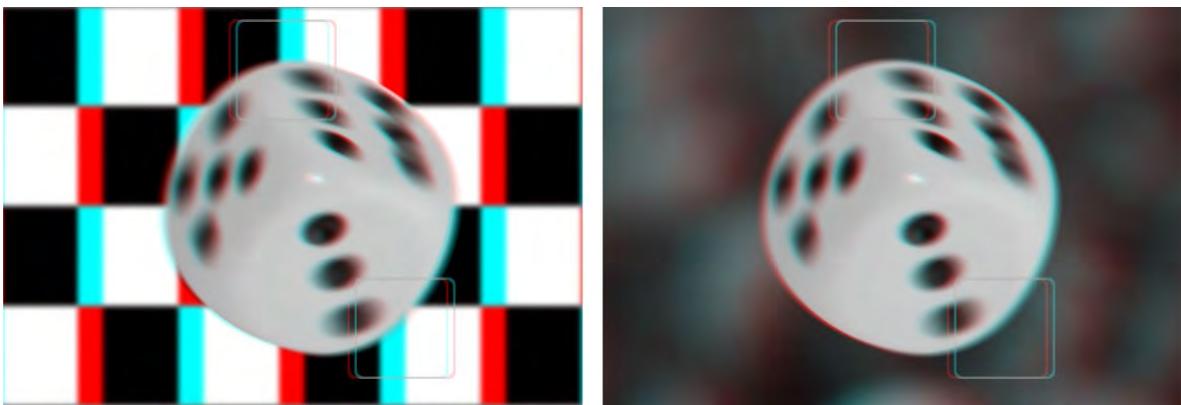


Abbildung 58: Würfel („Alpha-Matting“).

Sowohl die Augen als auch der Rand des rotierenden, weißen Würfels in Abb. 58 weisen ein gewisses Maß an Unschärfe auf. Diese Unschärfe ist für das visuelle System des Menschen ein eindeutiger Indikator für Bewegung. Durch enges Ausschneiden des Würfels („Threshold 200“) zwecks Erstellung einer neuen Ansicht geht dieses Merkmal verloren. Bei konservativem Zuschneiden des Objektes („Threshold 60“) bleibt die Eigenschaft der Bewegung zwar erhalten, jedoch werden Teile des Hintergrundes mitgenommen. Tabelle 1 beinhaltet einen Überblick der Ergebnisse.

Tabelle 1: Ergebnis „Würfel“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	6	6	12	4	2	18	3	1	20
Muster	5	7	12	7	9	8	3	20	1

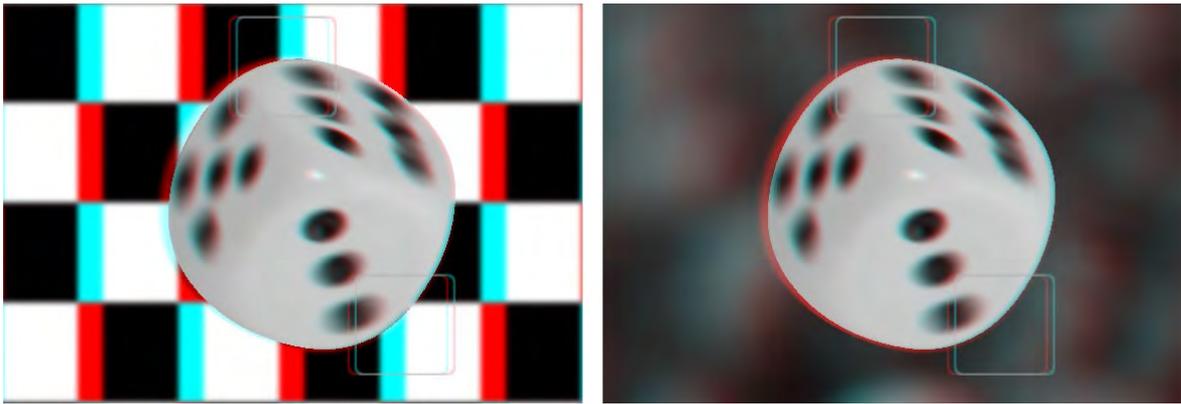


Abbildung 59: Würfel („Threshold 200“).

Bei „Threshold 200“ (siehe Abb. 59) entschied sich die Hälfte der Probanden unabhängig vom Hintergrund für das Stereobild mit Alpha-Matting. Die zu harten, scharfen Konturen des Würfels im 3D-Bild ohne Alpha-Matting fanden sie unlogisch und unrealistisch. Es erweckt den Eindruck, als würde der Rand stehengeblieben sein, während sich der Rest des Würfels weiterdreht. Zwei Probanden erschien der Würfel schwierig anzusehen, da die klare Mitte und der scharfe Rand nicht mit der Unschärfe im Bereich der Augen zusammenpassen. Sechs Testpersonen sind der Meinung, dass durch den genaueren Rand die Bewegung besser zur Geltung kommt. Für eine Person war die eckigere Form des Würfels der Beweggrund, sich für das Bild ohne Alpha-Matting zu entscheiden. Die restlichen sechs Probanden nahmen keinen Unterschied zwischen den beiden 3D-Bildern wahr. Die Analyse hat gezeigt, dass es sich dabei fast ausschließlich um Personen mit dominantem linkem Auge handelt.

Der Vergleich zwischen dem 3D-Bild mit Alpha-Matting und dem 3D-Bild mit „Threshold 128“ weisen unterschiedliche Resultate in Abhängigkeit des Hintergrundes auf. Während das Ergebnis bei strukturiertem Hintergrund eindeutig für Alpha-Matting spricht, herrscht beim Muster-Hintergrund starke Uneinigkeit unter den Probanden. Beim Würfel mit Schachbretthintergrund hielten nur vier Personen das Bild mit „Grenzwert 128“ für realistischer. 18 Probanden stimmten für das Bild mit Alpha-Matting. Die Entscheidungsgründe waren ähnlich denen bei „Threshold 200“. Einigen fiel der unpassende schwarze oder weiße Rand im 3D-Bild ohne Alpha Matting auf, der auf die Mitnahme des Hintergrundes von der Originalposition in der Basisansicht zurückzuführen ist. Da diese Art von Fehler auf Grund des verschwommenen, roten Musters beim anderen Hintergrund untergeht, entstand hier ein Verhältnis von acht zu sieben für Alpha-Matting. Die sieben Testpersonen argumentierten, dass im Bild mit „Threshold 128“ die weicheren Ränder in Kombination mit

einem gewissen Maß an Schärfe die Bewegung schöner sichtbar macht. Die meisten Probanden – nämlich neun – nahmen jedoch keinen Unterschied wahr.

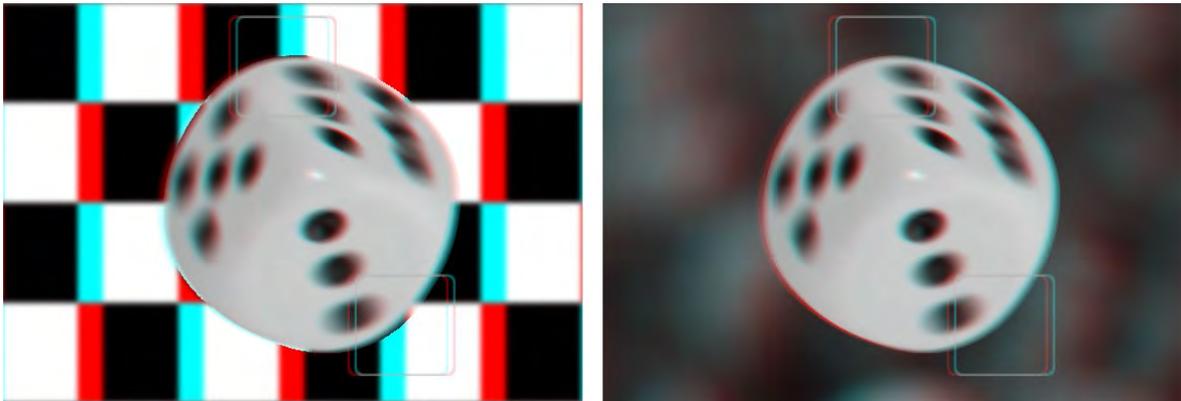


Abbildung 60: Würfel („Threshold 60“).

Der bereits bei „Threshold 128“ beginnende Trend verstärkt sich bei „Threshold 60“ (siehe Abb. 60) deutlich. 20 Personen entschieden sich bei strukturiertem Hintergrund für Alpha-Matting, da die Hintergrundfehler im Bild mit „Schwellenwert 60“ nun kaum mehr zu übersehen sind. Drei Probanden empfanden die schwarzen und weißen Flecken um den Würfel herum als nicht störend. Der gemusterte Hintergrund absorbiert diese Artefakte. In Verbindung mit dem Erhalt der Bewegungsunschärfe konnten daher 20 der 24 Probanden keinen Unterschied feststellen. Drei Testpersonen fanden das Bild mit Alpha-Matting zu verschwommen und lehnten es ab.

#### 7.1.1.2 Bild: Windrad

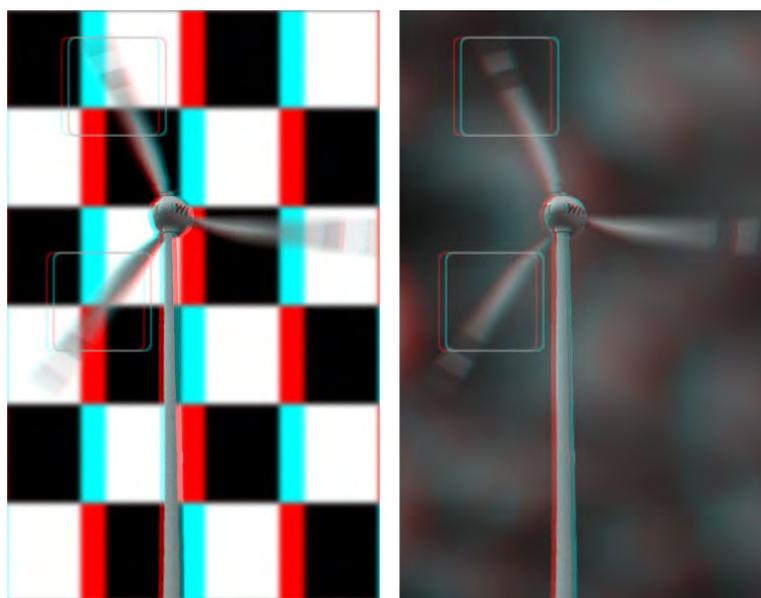


Abbildung 61: Windrad („Alpha-Matting“).

Durch die bewegten Rotorblätter des Windrades, das in Abb. 61 zu sehen ist, entstehen unscharfe und somit transparente Bereiche. Auf Grund der Größe und der Form dieser Region, geht im Zuge der Berechnung einer neuen Ansicht bei engem beziehungsweise moderatem Ausschneiden („Threshold 200“ und „Threshold 128“) nicht nur die Bewegungsinformation, sondern auch Teile der Struktur verloren. Bei großzügigem Herauslösen des Objekts aus der Basisansicht („Threshold 60“), bilden sich Artefakte durch die Übertragung des ursprünglichen Hintergrundes. Die Ergebnisse sind in Tabelle 2 zusammengefasst.

Tabelle 2: Ergebnis „Windrad“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	1	0	23	1	0	23	1	0	23
Muster	2	0	22	3	0	21	4	6	14

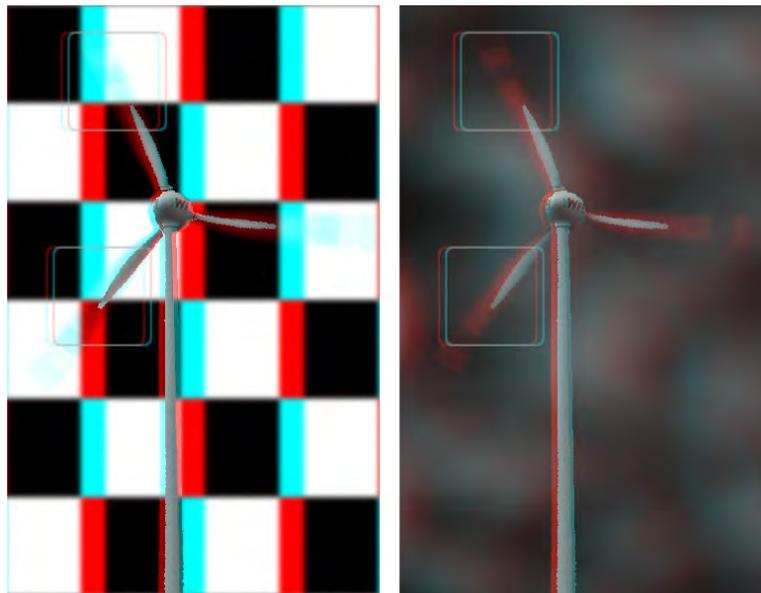


Abbildung 62: Windrad („Threshold 200“).

Auf Grund der offensichtlichen Artefakte bei „Threshold 200“ (siehe Abb. 62) stimmte eine klare Mehrheit von 23 Probanden beim Hintergrund mit Struktur und 22 Probanden beim Hintergrund mit Muster für Alpha-Matting. Die Unterschiede der beiden Einzelbilder waren so groß, dass viele Testpersonen angaben, Doppelbilder, ähnlich wie beim Schielen, zu sehen. Dieser Effekt ist eine Auswirkung der binokularen Rivalität, bei der das dargebotene Stereobildpaar nicht zu einem räumlichen Bild verschmolzen werden kann (siehe Ka-

pitel 3.3.5). Bei einigen Probanden wurde die „unschärfere“ Basisansicht auf Grund der binokularen Unterdrückung ausgeblendet. Sie bemängelten an dem Bild nicht nur die fehlende Struktur sondern auch die fehlende Bewegungsinformation. Die insgesamt drei Gegenstimmen wurden ausschließlich von Personen mit dominantem, linkem Augen abgegeben. Die komplett fehlenden Teile der Rotorblätter wurden als schnellere Drehbewegung gedeutet.

Eine ähnliche Verteilung der Stimmen gab es auch bei den Bildern mit „Threshold 128“. Das Hauptargument der 23 Probanden (Struktur-Hintergrund „S“) beziehungsweise der 21 Probanden (Muster-Hintergrund „M“), die das 3D-Bild mit Alpha-Matting bevorzugten, war, dass jenes Bild ohne Alpha-Matting zu extreme Übergänge innerhalb der Rotorblätter aufweist. Dieser unzusammenhängende Übergang zwischen der Mitte und den Spitzen des Rotorblattes bewirkte, dass das Stereobild einen sehr komischen und unharmonischen Eindruck machte. Bei den drei Testpersonen, die das Bild mit „Grenzwert 128“ und gemustertem Hintergrund wegen der klareren Konturen besser fanden, handelt es sich wiederum um Personen mit dominantem linkem Auge. Sowohl bei „Grenzwert 200“ als auch bei „Grenzwert 128“ hat keine der Testperson die gezeigten 3D-Bilder mit gleich bewertet.

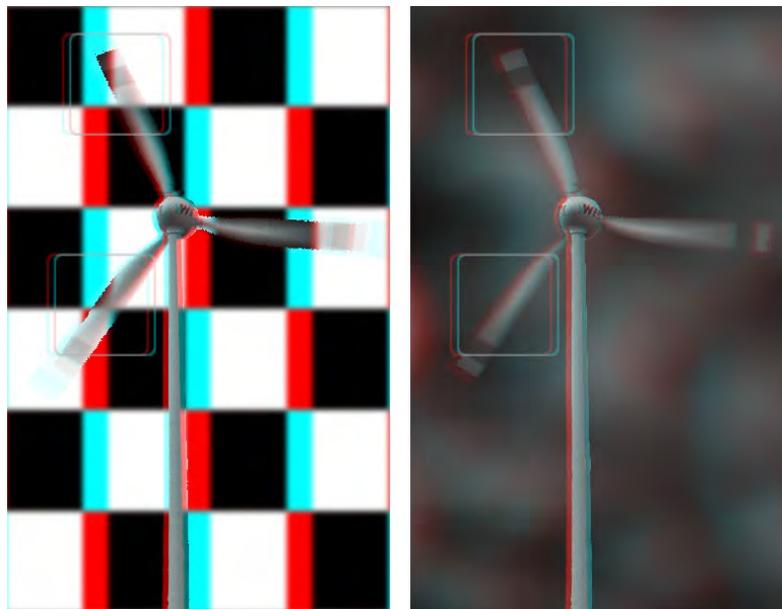


Abbildung 63: Windrad („Threshold 60“).

Bei „Threshold 60“ (siehe Abb. 63) fällt das Ergebnis bei strukturiertem Hintergrund deutlich für Alpha-Matting aus, während bei den 3D-Bildern mit Muster-Hintergrund alle drei Meinungen vertreten sind. Als Hauptgrund für die unterschiedliche Verteilung der Antworten ist der nur beim Schachbretthintergrund auftretende Hintergrundfehler, der durch den

eher homogenen Muster-Hintergrund neutralisiert wird. Sechs Probanden gaben an, keinen Unterschied zwischen den 3D-Bildern mit gemusterten Hintergrund zu erkennen. Vier der Befragten entschieden zu Gunsten des „Grenzwert 60“-Bildes, da in diesem die Form des Windrades besser zu Geltung kommt.

Allgemein empfanden die Probanden bei den Bildern mit Alpha-Matting ein echteres Bewegungsgefühl. Das Fehlen der Unschärfe am Rand eines Objektes ist mit Stillstand verknüpft. Diese Wahrnehmung konnte jedoch nicht mit der verschwommenen Struktur innerhalb eines Objektes in Einklang gebracht werden, weshalb die Bilder ohne Alpha-Matting verwirrend und unnatürlich erschienen.

### 7.1.2 Kategorie: Tiefenunschärfe

Diese Kategorie beinhaltet Bilder, die aus der Nähe aufgenommen wurden. Die auffälligste Eigenschaft dieser Bilder ist, dass jene Teile des Objektes, die sich außerhalb des Fokus befinden, ein bestimmtes Maß an Unschärfe aufweisen. Analog zur Bewegungsunschärfe hat die Tiefenunschärfe eine Mischung aus Vorder- und Hintergrund zur Folge. Artefakte, die bei der Berechnung synthetischer Ansichten entstehen, wirken sich störend auf die Tiefenwahrnehmung aus. Die für diese Kategorie ausgewählten Motive sind Nahaufnahmen einer Rosenknospe und einer Libelle.

#### 7.1.2.1 Bild: Knospe

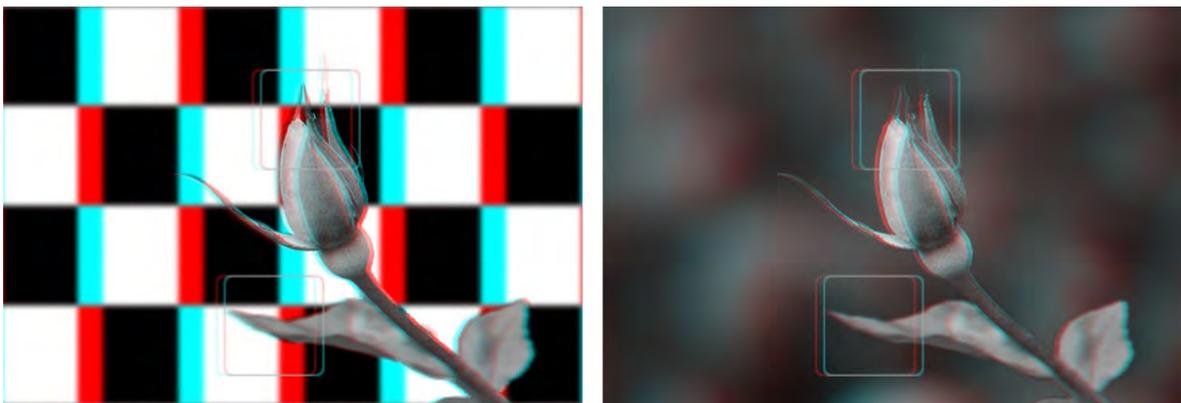


Abbildung 64: Knospe („Alpha-Matting“).

Die Knospe einer Rose, die in Abb. 64 zu sehen ist, hat scharfe Konturen, da sie sich im Bildfokus befindet. Die Blätter liegen in einer anderen Bildebene und sind daher unscharf. Bei der Berechnung der zweiten Ansicht ohne Alpha-Matting kommt es in diesen Regionen zu Artefakten, die mit jenen aus der Kategorie Bewegungsunschärfe vergleichbar sind.

Eine weitere Fehlerquelle bei diesem Motiv ist die feine Struktur an den Spitzen von Blüte und Blättern. Diese Strukturelemente bleiben ohne Alpha-Matting nur durch konservatives Ausschneiden („Threshold 60“) erhalten und weisen gleichzeitig Hintergrundfehler auf. Die Antworten der Probanden sind überblicksmäßig in Tabelle 3 dargestellt.

Tabelle 3: Ergebnis „Knospe“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	0	4	20	6	6	12	5	4	15
Muster	0	1	23	3	8	13	3	15	6

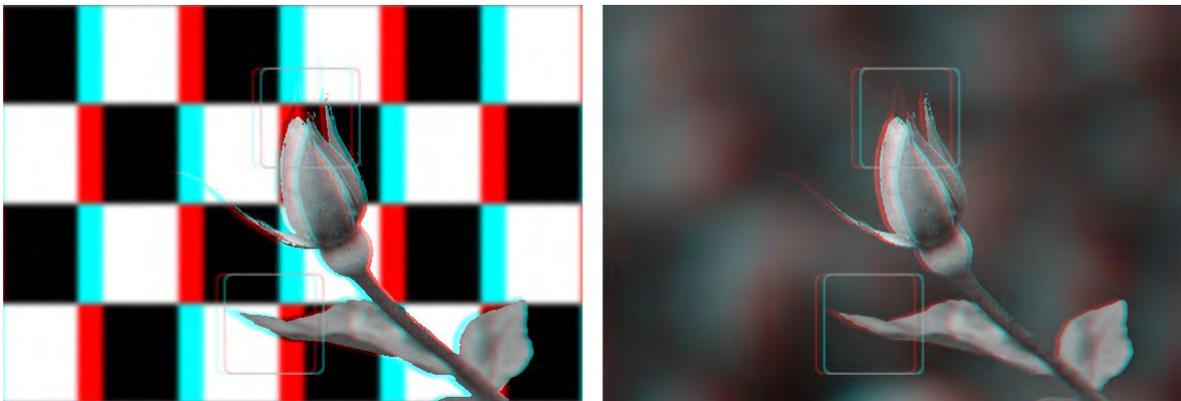


Abbildung 65: Knospe („Threshold 200“).

Bei „Threshold 200“ (siehe Abb. 65) waren 20 (S) beziehungsweise 23 (M) Probanden der Meinung, dass die Knospe mit Alpha-Matting realistischer aussieht. Einige Probanden konnten nicht genau erklären, was an dem Bild mit „Grenzwert 200“ nicht passte. Sie bezeichneten es als „irgendwie falsch“ oder „seltsam“. Andere sagten, dass es zu keiner kompletten Verschmelzung der Einzelbilder kommt. Das von den Testpersonen dabei beobachtete Schillern oder Flackern, das in manchen Regionen auftrat, ist ein eindeutiger Hinweis auf eine strukturelle, binokulare Rivalität. Ein kleiner Teil der Befragten argumentierte, dass die fehlenden Teile bei der Blüte und beim Blatt störend wirken. Die restlichen Probanden nahmen keinen Unterschied zwischen den gezeigten 3D-Bildern wahr.

Bei „Threshold 128“ gingen die Ansichten der Probanden über das Wahrgenommene auseinander. Beim strukturierten Hintergrund bevorzugten zwölf und beim gemusterten Hintergrund 13 Probanden das Bild mit Alpha-Matting. Viele begründeten ihre Entscheidung damit, dass sie das Gefühl haben, dass etwas fehlt oder etwas nicht stimmt. Ein paar konn-

ten konkretere Antworten geben. Diese beinhalteten unter Anderem, dass die Struktur bei der Blüte fehlt, dass der scharfe Rand des Blattes nicht zum restlichen Bild passt oder, dass wieder nur ein Doppelbild zu sehen war. Sechs (S) beziehungsweise drei (M) Probanden stimmten für das Bild mit „Threshold 128“, weil ihnen das Bild generell schärfer erschien. Eine Testperson bemerkte einen weißen Rand um das Objekt, den diese aber nicht als störend empfand. Die Probanden, die „keinen Unterschied“ angaben, nahmen die fehlende Struktur zwar wahr, argumentierten aber, dass dies ohne Vergleichsbild nicht auffallen würde, weil die Knospe vollständig als solche erkennbar ist.

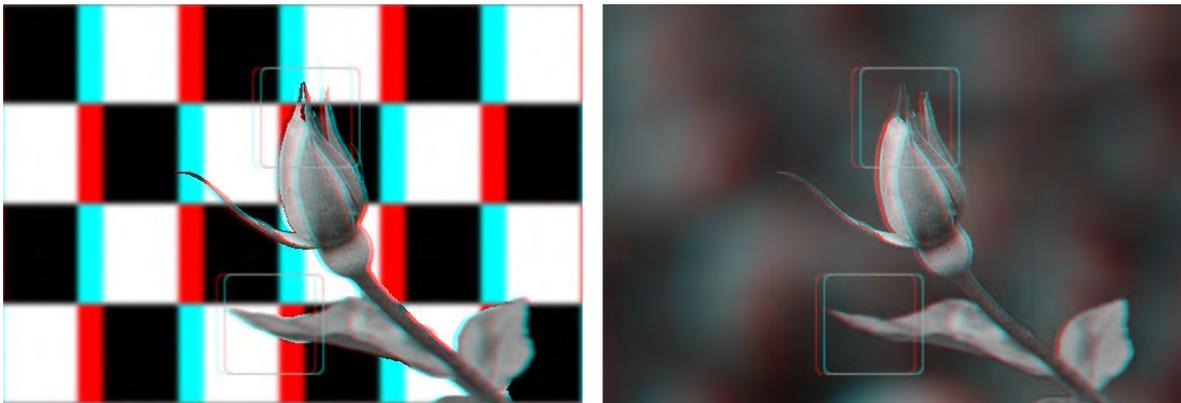


Abbildung 66: Knospe („Threshold 60“).

Die Sichtbarkeit der Hintergrundartefakte im 3D-Bild mit Struktur-Hintergrund spielt für das Ergebnis bei „Threshold 60“ (siehe Abb. 66) eine entscheidende Rolle. 15 Probanden entschieden sich daher für Alpha-Matting. Neben den störenden schwarzen oder weißen Rändern, wurden auch fehlende Teile, unpassende Transparenz bei der Blüte und zu scharfe Blätter als Entscheidungsgrundlage genannt. Vier Probanden nahmen keinen Unterschied wahr, und fünf der Befragten fanden das Bild ohne Alpha-Matting natürlicher. Beim Muster-Hintergrund sieht die Verteilung der Antworten anders aus. Nur neun der 24 Testpersonen sprachen sich dezidiert für Alpha-Matting aus. Sie gaben an, dass die fehlende oder zu transparent wirkende Struktur eine unangenehme, visuelle Wahrnehmung verursachte. Für eine Mehrheit von 15 Probanden waren die beiden gezeigten 3D-Bilder gleich.

## 7.1.2.2 Bild: Libelle

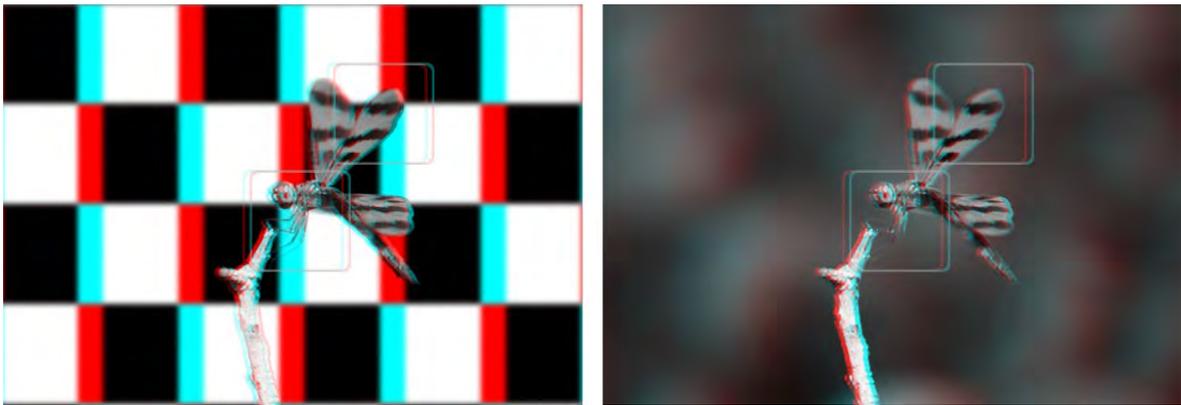


Abbildung 67: Libelle (,Alpha-Matting').

In der Aufnahme einer Libelle, die in Abb. 67 abgebildet ist, wurde auf den Kopf des Tieres fokussiert. Im Unterschied zum Körper sind die Flügel der Libelle unscharf, da sie in einer anderen Tiefenebene liegen. Die durch diese Bildeigenschaft entstehenden Artefakte bei der Berechnung einer synthetischen Ansicht ohne Alpha-Matting reichen, wie auch zuvor bei dem Bild mit der Rosenknospe, von Hintergrundfehlern über Verlust der richtigen Tiefenwahrnehmung bis zu strukturellen Bildfehlern. Im Gegensatz zur Knospe, deren Blätter durch knappes oder moderates Ausschneiden (,Threshold 200“ beziehungsweise ,Threshold 128“) verkürzt erscheinen, ist die fehlende Struktur der Beine für die korrekte Wahrnehmung der Libelle entscheidend. Tabelle 4 zeigt einen Überblick der Ergebnisse.

Tabelle 4: Ergebnis ,Libelle“.

	,Threshold 200“			,Threshold 128“			,Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	1	1	22	2	4	18	9	1	14
Muster	5	1	18	3	6	15	2	13	9

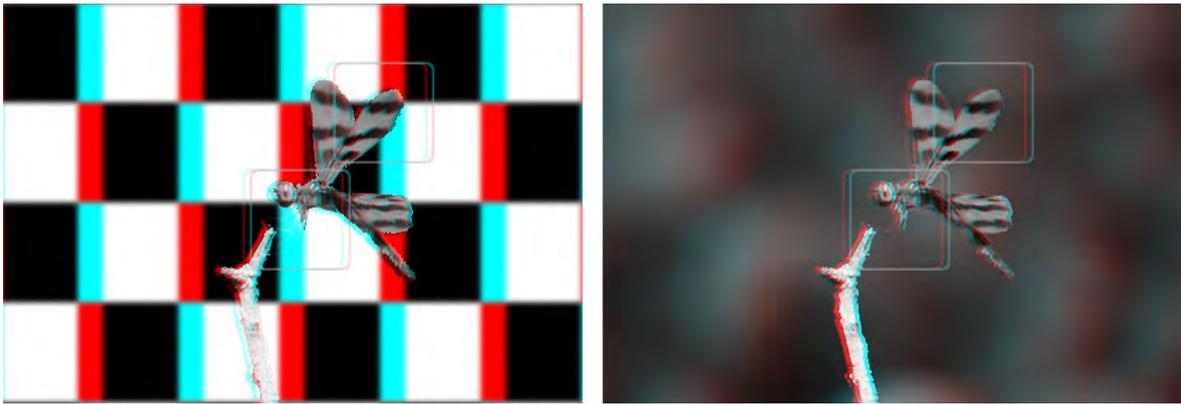


Abbildung 68: Libelle („Threshold 200“).

Das Fehlen der Beine war bei „Threshold 200“ (siehe Abb. 68) das Hauptargument der 22 Probanden, die beim Schachbretthintergrund das 3D-Bild mit Alpha-Matting bevorzugten. Ähnlich deutlich fiel das Ergebnis beim gemusterten Hintergrund aus. Die 18 Testpersonen meinten, dass das Bild ohne Alpha-Matting irritierend wirkt, weil die Libelle ihre Beine braucht, um auf dem Ast zu sitzen. Das wahrgenommene Bild stimmt somit nicht mit der Wahrnehmungserfahrung der Probanden überein und wurde daher als unecht abgelehnt. Bei beiden Hintergründen gab jeweils eine Testperson an, keinen Unterschied erkannt zu haben. Während beim Strukturhintergrund ein Befragter für das 3D-Bild mit „Grenzwert 200“ stimmte, waren es beim Musterhintergrund fünf. Es handelt sich dabei fast ausschließlich um Personen mit dominantem linkem Auge.

Bei „Threshold 128“ wurden die nur teilweise vorhandenen Beine ebenfalls als störend empfunden. Zusätzlich argumentierten die 18 (S) beziehungsweise 15 (M) Probanden, die sich für das Bild mit Alpha-Matting aussprachen, dass die zu scharfen Flügeln das Bild ohne Alpha-Matting unrealistisch erscheinen lassen. Eine Testperson nahm im Bereich der Beine ein „Schillern“ wahr, das auf eine Form der binokularen Rivalität hindeutet. Vier (S) beziehungsweise sechs (M) Testpersonen erachteten die beiden gezeigten 3D-Bilder für gleich. Die restlichen Probanden, die für das 3D-Bild mit „Grenzwert 128“ stimmten, gaben an, dass dieses Bild einen besseren Tiefeneindruck vermittelt und das Objekt plastischer erscheinen lässt. Eine Testperson nannte den echter wirkenden Ast als Grundlage für ihre Entscheidung.

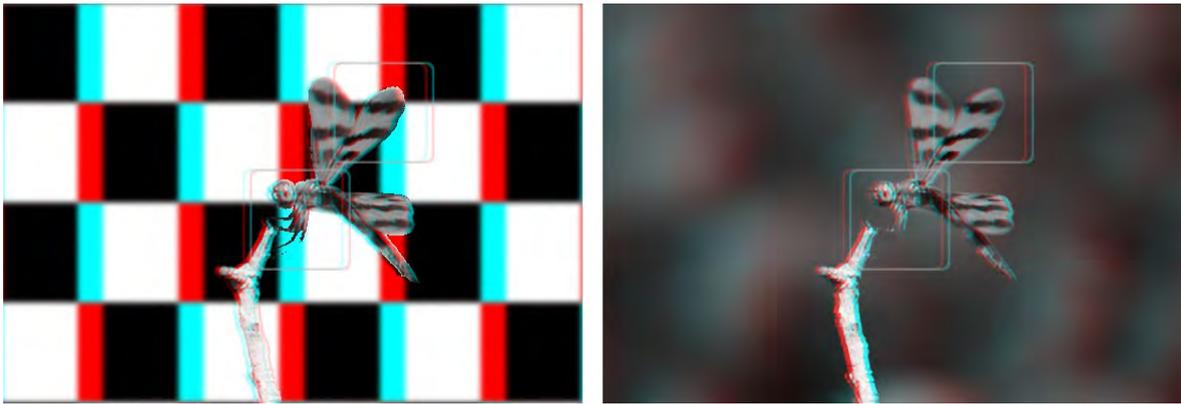


Abbildung 69: Libelle („Threshold 60“).

Bei „Threshold 60“ (siehe Abb. 69) herrschte bei beiden Hintergründen weitgehende Uneinigkeit unter den Befragten. Der strukturierte Hintergrund brachte vor allem im Bereich der Beine der Libelle zwei stark konträre Gruppen hervor. Die neun Befürworter des Bildes mit „Grenzwert 60“ fanden die Beine angemessen und detailgetreuer, während die 14 Probanden, die für Alpha-Matting stimmten, die Beine als stark überzeichnet, dunkel und unheimlich wahrnahmen. Eine Testperson befand beide 3D-Bilder für gleich annehmbar. Beim gemusterten Hintergrund kam es zu einer gänzlich anderen Verteilung. Neun Probanden waren der Meinung, das Bild mit Alpha-Matting sieht realistischer aus und zwei Probanden entschieden zu Gunsten des Bildes ohne Alpha-Matting. Auch hier spielten die Beine der Libelle eine ausschlaggebende Rolle. Die Mehrheit – nämlich 13 Testpersonen – sahen jedoch keinen Unterschied zwischen den beiden gezeigten 3D-Bildern.

Sowohl beim Motiv „Knospe“ also auch beim Motiv „Libelle“ wurde das Bild mit Alpha-Matting als weich und sympathisch bezeichnet. Die feinen Ränder der Objekte und die bessere Sichtbarkeit der Details wurden ebenfalls von den Befragten positiv hervorgehoben. Des Weiteren hat sich herauskristallisiert, dass unvollständige Strukturelemente, wie beispielsweise Beine, auffälliger und störender wirken, als eine unpassende Tiefenschärfe. Die Entscheidung für eines der beiden 3D-Bilder wurde jedoch bei dieser Kategorie häufig aus dem Bauch heraus getroffen.

### 7.1.3 Kategorie: Struktur

Das Merkmal der Motive dieser Kategorie sind die enthaltenen (sehr) feinen Strukturen. Diese Bildelemente sind wichtig, um ein korrektes Abbild des Objektes zu erhalten und um dessen Eigenschaften zu erkennen. Bei der Generierung synthetischer Ansichten entstehen in Bildregionen mit feinen Strukturelementen, wie Haare, Fell oder Federn, aus zwei Gründen schwerwiegende Artefakte. Zum einen führt die teilweise transparente Struktur des Vordergrundobjektes zu Hintergrundfehlern und zum anderen gehen die feinen Strukturelemente bei einfachem Ausschneiden leicht verloren. Die für diese Kategorie ausgewählten Bilder zeigen einen Stoffkoala mit Fahne und einen Rad schlagenden Pfau.

#### 7.1.3.1 Bild: Koala

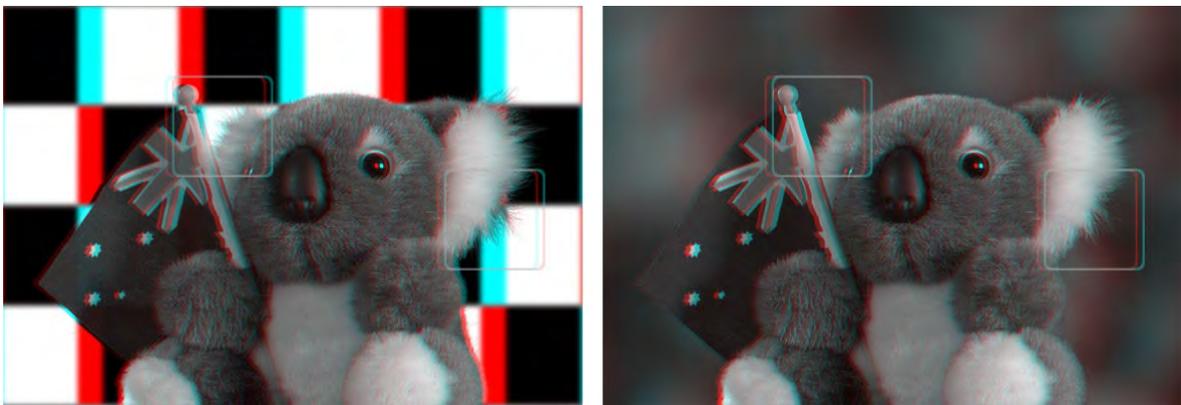


Abbildung 70: Koala (,Alpha-Matting‘).

Der Koala in Abb. 70 wird auf Grund des plüschartigen Fells als Stofftier wiedererkannt. Das Gefühl der Weichheit wird vor allem durch die feinen Haare am Rand vermittelt. Bei der Berechnung neuer Ansichten, bei der das Objekt eng (,Threshold 200‘) oder moderat (,Threshold 128‘) ausgeschnitten wird, geht mit dem Fehlen der detailreichen Struktur ein wesentliches Merkmal des Stoffkoalas verloren. Bei konservativem Herauslösen des Objektes (,Threshold 60‘) kommt es zu Artefakten, weil der ursprüngliche Hintergrund auf das Vordergrundobjekt übertragen und mit verschoben wurde. Eine Zusammenfassung der Ergebnisse ist in Tabelle 5 zu sehen.

Tabelle 5: Ergebnis ,Koala‘.

	,Threshold 200‘			,Threshold 128‘			,Threshold 60‘		
	-1	0	1	-1	0	1	-1	0	1
Struktur	0	1	23	0	0	24	1	0	23
Muster	0	1	23	3	8	13	3	19	2

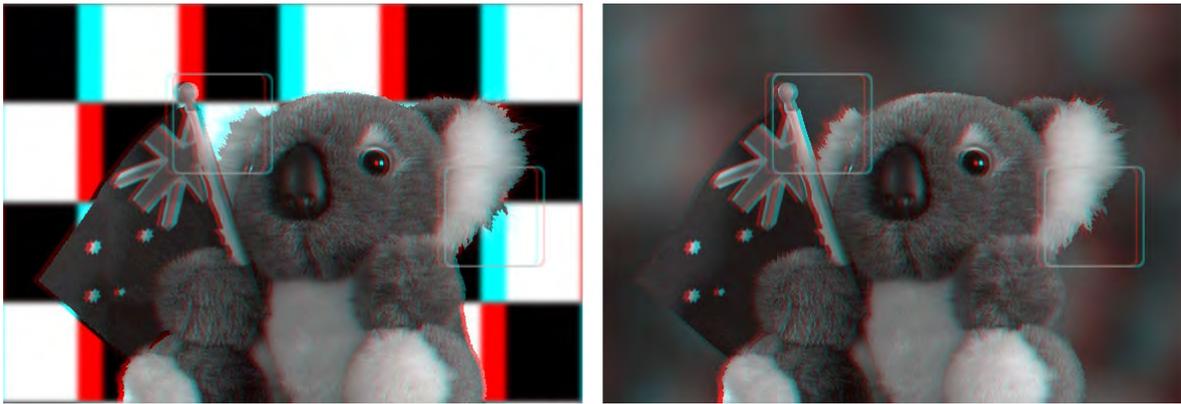


Abbildung 71: Koala („Threshold 200“).

Das „Threshold 200“-Bild (siehe Abb. 71) wurde wegen des Fehlens der feinen Struktur – vor allem bei den Ohren des Koalas – sowohl beim strukturierten als auch beim gemusterten Hintergrund von 23 Probanden abgelehnt. Der Großteil der Befragten empfand die abgeschnittenen oder zusammengedrückten Haare als sehr störend und unpassend. Einige Testpersonen fanden den groben Übergang beziehungsweise die scharfen Kanten für ein Stofftier inakzeptabel. Drei Probanden nahmen die beiden Einzelbilder nicht vollständig verschmolzen wahr. Sie gaben an, das auf binokulare Rivalität deutende „Flackern“ wahrzunehmen. Eine Person beobachtete „geisterhaft verschwommene“ Ränder. Dies könnte als Hinweis auf eine unvollständige, binokulare Unterdrückung interpretiert werden. Bei beiden Hintergründen entdeckte jeweils ein Proband keinen Unterschied zwischen den zu bewertenden, dreidimensionalen Bildern.

Bei „Threshold 128“ in Kombination mit dem Schachbretthintergrund gab es das einzige einstimmige Ergebnis. Alle 24 Probanden entschieden, dass das Bild mit Alpha-Matting realistischer wirkt. Einige Testpersonen versuchten, die Hintergrundfehler im Bild ohne Alpha-Matting logisch zu erklären. Die dunklen Flecken bei den Fransen am Ohr waren ihrer Meinung nach dadurch entstanden, dass die Haare nass, gefärbt oder schmutzig sind. Diese möglichen Ursachen überzeugten die Befragten schlussendlich nicht. Der Großteil der Probanden empfand die übertriebene Betonung der Fransen durch die schwarze Farbe unnatürlich. Zwei Personen gab an, dass die abgeschnittenen Haare nicht mit der Vorstellung eines Stofftieres zusammenpassen. Der gemusterte Hintergrund brachte hingegen eine Streuung der Antworten hervor. 13 Probanden stimmten für das mit Alpha-Matting berechnete Bild, während acht Personen angaben, keinen Unterschied zwischen den beiden 3D-Bildern wahrzunehmen. Der Grund für die Ablehnung des Bildes mit „Grenzwert 128“ reichte von auftretender binokularer Rivalität in Form von „Flackern“ bis hin zu den feh-

lenden Haarspitzen. Die drei Testpersonen, die das Bild ohne Alpha-Matting bevorzugten, fanden das Objekt in diesem Bild plastischer. Viele Probanden vervollständigten ihre Entscheidung, dass beide Bilder durchaus akzeptabel sind, aber in jenem mit Alpha-Matting der Koala flauschiger aussieht.

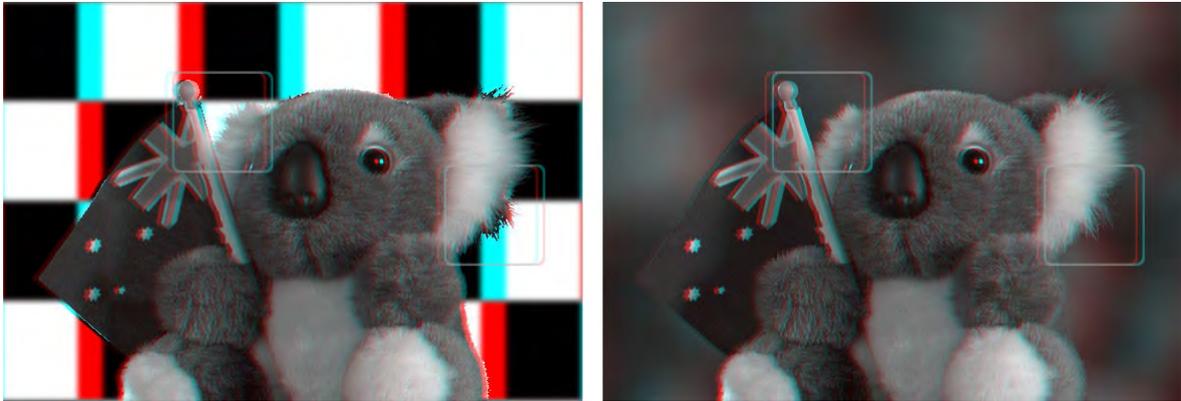


Abbildung 72: Koala („Threshold 60“).

Bei „Threshold 60“ (siehe Abb. 72) ist einmal mehr die Sichtbarkeit des Hintergrundfehlers das entscheidende Kriterium. 23 Probanden bevorzugten das 3D-Bild mit strukturiertem Hintergrund und Alpha-Matting. Der Hauptkritikpunkt beim Bild ohne Alpha-Matting war der schwarze beziehungsweise weiße Rand, der den Koala weitgehend umgibt. Diese Artefakte, die durch die Übertragung des Hintergrundes auf das Vordergrundobjekt entstanden sind, wurden bei den feinen Fransen am Ohr als besonders störend empfunden. Eine Testperson war der Meinung, dass das 3D-Bild mit „Grenzwert 60“ wegen den dunklen Stellen beim Ohr die Haare besser sichtbar machen. Beim gemusterten Hintergrund, bei dem der Hintergrundfehler nicht auffällt, nahmen 19 Probanden keinen erkennbaren Unterschied zwischen den gezeigten 3D-Bildern wahr. Zwei der Befragten entschieden sich für das Bild mit Alpha-Matting. Sie ergänzten ihre Entscheidung allerdings mit der Aussage, dass das andere Bild ebenfalls akzeptabel ist. Die drei Testpersonen, die das Bild ohne Alpha-Matting für realistischer hielten, meinten, dass dieses ein wenig schärfer und plastischer wirkt.

## 7.1.3.2 Bild: Pfau

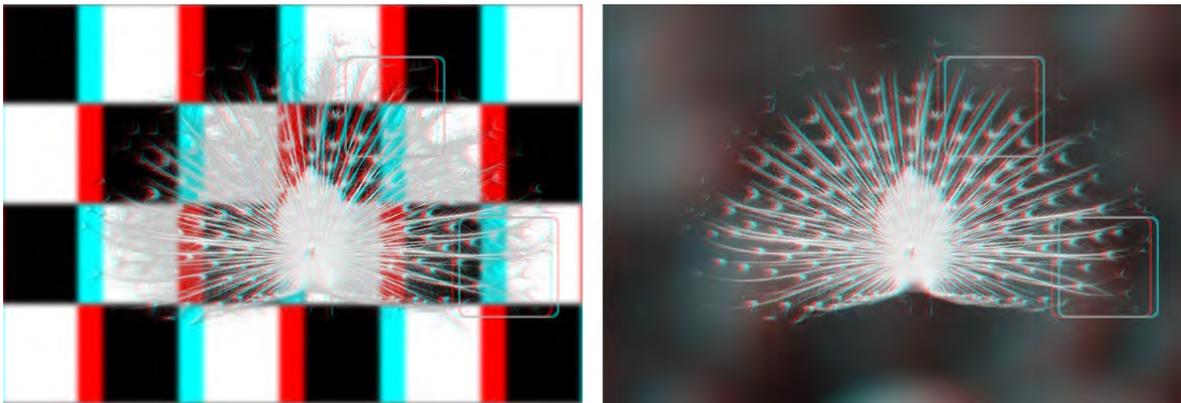


Abbildung 73: Pfau („Alpha-Matting“).

Das Erkennungsmerkmal des Rad schlagenden Pfau in Abb. 73 ist sein typisches, dichtes Federkleid. Die filigrane Struktur der einzelnen Federn bietet viel Angriffsfläche für Bildfehler. Die Artefakte, die während der Generierung von synthetischen Ansichten ohne Alpha-Matting entstehen, umfassen, wie zuvor beim Stoffkoala, den Verlust der Haupteigenschaft beziehungsweise eine teilweise Wiedergabe des Originalhintergrundes am Objekt. Die auftretenden Fehler sind bei diesem Motiv besonders störend, da das Federkleid des Pfau einen großen Teil des Bildes ausmacht. Die Antworten der Probanden sind in Tabelle 6 zusammengefasst.

Tabelle 6: Ergebnis „Pfau“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	3	0	21	1	0	23	2	0	22
Muster	1	0	23	4	2	18	4	18	2

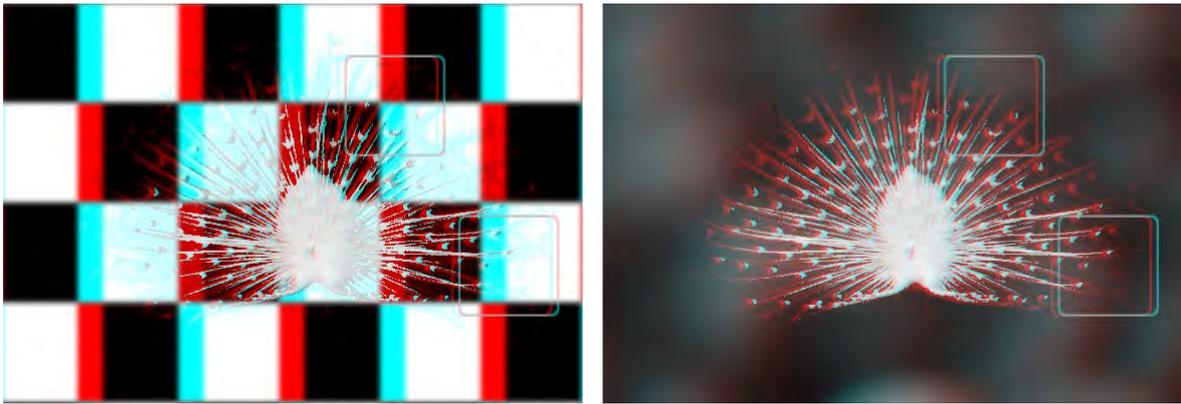


Abbildung 74: Pfau („Threshold 200“).

Bei „Threshold 200“ (siehe Abb. 74) haben alle Probanden unabhängig vom Hintergrund einen Unterschied zum Referenzbild mit Alpha-Matting gesehen. Beim strukturierten Hintergrund stimmten 21 Testpersonen für Alpha-Matting. Viele Probanden fanden das Bild mit „Grenzwert 200“ aus unterschiedlichen Gründen deutlich schlechter. Einige der Befragten fanden das Bild sehr anstrengend zu betrachten, da sie keine klaren Konturen wahrnehmen konnten. Als Ursache dafür nannten sie einen seltsamen „Schimmer-Effekt“, der auf binokulare Rivalität hindeutet. Andere gaben an, dass der Pfau nackt wirkt, weil sein Federkleid fast nicht mehr vorhanden ist. Eine Testperson meinte sogar, das abgebildete Tier sieht wegen der stacheligen Federstummeln eher einem Igel als einem Pfau ähnlich. Drei Probanden bevorzugten das Bild ohne Alpha-Matting, weil in diesem der Pfau schärfer und besser erkennbar ist. Beim gemusterten Hintergrund bevorzugten 23 Personen das 3D-Bild mit Alpha-Matting und eine jenes ohne. Der Grund für die Ablehnung des Bildes mit „Threshold 200“ war in den meisten Fällen eine Form von binokularer Rivalität. Die am häufigsten genannten Ausprägungen waren Doppelbilder, „Schimmern“ und „Verschwommenheit“. Mehrere Probanden beanstandeten, dass durch das Verschwinden des äußeren Federkranzes sehr viel Information verloren gegangen ist.

Bei „Threshold 128“ wurde das Bild mit Alpha-Matting von 23 (S) beziehungsweise 18 (M) Probanden favorisiert. Die meisten gesehenen Artefakte bezogen sich auf das Fehlen des Federkleides. Bei beiden Hintergründen kam es im 3D-Bild ohne Alpha-Matting zu binokularer Rivalität, die sich in Form von „Flimmern“ äußerte. Drei Testpersonen sahen nur die zweidimensionalen Einzelbilder an Stelle eines 3D-Bildes. Eine kleine Gruppe von Personen empfand die Konturen im Bild mit „Grenzwert 128“ als unpassend scharf. Ein Proband beschrieb das Bild als generell sehr unruhig. Zwei der Befragten gaben beim Muster-Hintergrund an, dass die beiden 3D-Bilder qualitativ gleich aussahen. Sie fügten aller-

dings hinzu, dass der Pfau im Bild mit Alpha-Matting näher wirkt. Vier Personen sprachen sich gegen Alpha-Matting aus. Ihrer Meinung nach sind die Federn des Pfaus im schärferen „Threshold 128“-Bild besser erkennbar.

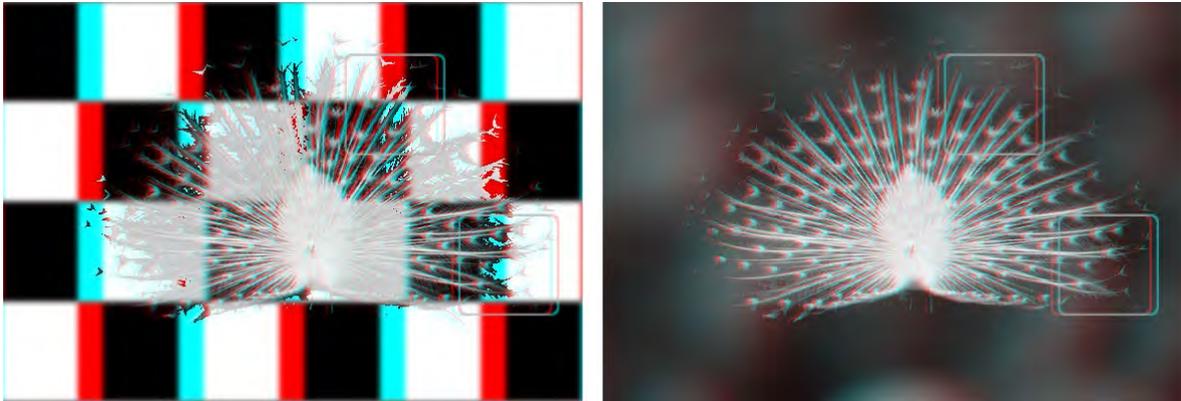


Abbildung 75: Pfau („Threshold 60“).

Bei „Threshold 60“ (siehe Abb. 75) gibt es unterschiedliche Ergebnisse für die beiden Hintergründe. 22 Probanden entschieden sich beim Schachbretthintergrund für Alpha-Matting. Dem gegenüber stehen nur zwei Stimmen für Alpha-Matting beim gemusterten Hintergrund. Die Ursache dafür sind die Artefakte, die nur beim strukturierten Hintergrund wahrgenommen wurden. Wie bereits bei den anderen Bildern ohne Alpha-Matting wurde von manchen Probanden ein „Schimmern“, das auf binokulare Rivalität zurückzuführen ist, beobachtet. Als besonders unangenehm und störend bezeichneten viele der Befragten die fehlende Transparenz im Bereich der feinen Struktur. Eine Person nahm herumfliegende Einzelteile wahr. Diese stellten sich als Federspitzen heraus, die durch den Hintergrund in der Basisansicht fälschlicherweise schwarz eingefärbt wurden. Während beim Strukturhintergrund alle Testpersonen zwei unterschiedliche 3D-Bilder ausmachten, gaben beim Muster-Hintergrund 18 Personen an, dass die vorliegenden Bilder für sie ident sind. Die zwei (S) beziehungsweise vier (M) Probanden, die für das 3D-Bilder mit „Grenzwert 60“ stimmten, bezeichneten dieses als schärfer und detailreicher.

Obwohl es beim Pfau einige Spekulationen gab, was das dargestellte Objekt tatsächlich ist, geht bei beiden Motiven der Kategorie „Struktur“ klar hervor, dass fehlende Strukturelemente schwerwiegende Artefakte sind. Die Bilder mit Alpha-Matting wurden daher nicht nur als realistischer sondern auch als angenehmer und sympathischer wahrgenommen. Das flauschige Gefühl des weichen Stofffells beziehungsweise des buschigen Federkleids wird durch die dichten und feinen Strukturen verstärkt. Außerdem konnten nach Angaben der Probanden, trotz leichter Unschärfe, mehr Details erkannt werden.

### 7.1.4 Kategorie: Transparenz

Die abgebildeten Objekte dieser Kategorie bestehen aus transparenten Materialien, wie beispielsweise Glas. Neben dem Durchscheinen des Hintergrundes sind auch Reflexionen am Objekt ein wesentliches Merkmal dieser Materialien. Bei der Berechnung synthetischer Ansichten muss darauf geachtet werden, dass der Hintergrund herausgefiltert wird und gleichzeitig die Eigenschaft des Objekts erhalten bleibt. Bei einfachem Ausschneiden des Objektes geht entweder die Struktur des Objektes verloren oder Teile des Hintergrundes werden dem Vordergrund zugeordnet. Die Bilder, die dieser Kategorie zugeordnet wurden, sind ein vorbeifahrendes Auto und ein mit Münzen gefülltes Glas.

#### 7.1.4.1 Bild: Auto

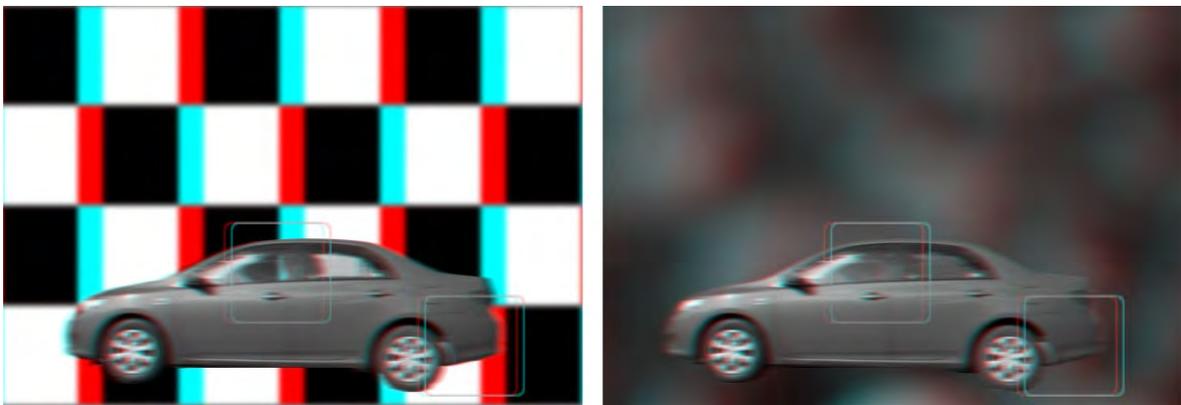


Abbildung 76: Auto („Alpha-Matting“).

Das in Abb. 76 dargestellte Motiv des vorbeifahrenden, roten Autos enthält im Bereich der Fenster transparente Stellen. Obwohl es sich um eine relativ kleine Region handelt, entstehen auf Grund der Transparenz beim einfachen Herauslösen des Objektes aus der Basisansicht sehr auffällige Fehler. Bei engem Ausschneiden der undurchsichtigen Teile des Autos („Threshold 200“), um den Hintergrund gänzlich zu entfernen, geht die Scheibe verloren. Bei konservativem Ausschneiden („Threshold 60“) zur Erhaltung der gesamten Struktur wird der Originalhintergrund auf die Glasfläche übertragen. Tabelle 7 gibt einen Überblick über die Ergebnisse.

Tabelle 7: Ergebnis „Auto“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	2	1	21	4	1	19	3	2	19
Muster	1	0	23	10	10	4	6	12	6

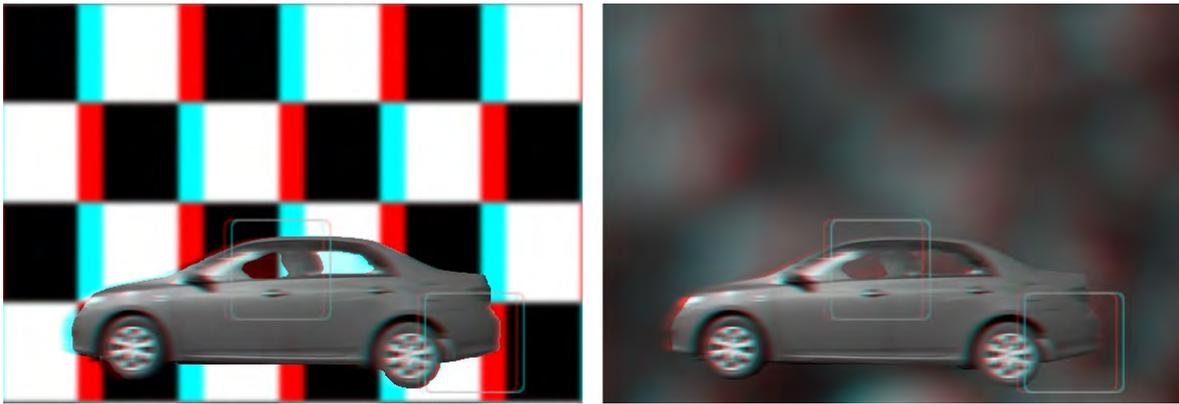


Abbildung 77: Auto („Threshold 200“).

Bei „Threshold 200“ (siehe Abb. 77) entschieden sich 21 (S) beziehungsweise 23 (M) Probanden für das Bild mit Alpha-Matting. Unabhängig vom Hintergrund wurde das Fehlen der Fensterscheibe beim Bild mit „Grenzwert 200“ als häufigste Entscheidungsgrundlage genannt. Das Loch in der Scheibe wurde von einer Testperson als zu extreme Spiegelung wahrgenommen. Eine andere Person meinte, dass das Fenster künstlich eingefügt wirkt. Ein paar der Befragten fanden die Schärfe am Rand des Fensters unrealistisch. Ein Proband mit dominantem linkem Auge nahm beim Schachbretthintergrund keinen Unterschied zwischen den gezeigten 3D-Bildern wahr. Insgesamt gab es drei Stimmen gegen das Bild mit Alpha-Matting. Sie fanden das „Threshold 200“-Bild dynamischer und nicht so verschwommen.

Die beiden Hintergründe bringen bei „Threshold 128“ unterschiedliche Ergebnisse hervor. Während sich beim Struktur-Hintergrund 19 Probanden für Alpha-Matting aussprachen, waren es beim Muster-Hintergrund nur vier. Beim strukturierten Hintergrund spielten die Hintergrundartefakte im Bild mit „Grenzwert 128“ eine entscheidende Rolle. Die Testpersonen nahmen diese Bildfehler in Form weißen Schatten beziehungsweise Rändern am Heck wahr. Einige der Befragten fanden die Fensterscheiben zu undurchsichtig. Eine Person gab an, dass Bewegungsunschärfe nicht so gut zur Geltung kommt. Beim Schachbretthintergrund sah eine Person keinen Unterschied zwischen den zu vergleichenden Bildern. Vier Testpersonen bevorzugten das 3D-Bild ohne Alpha-Matting, weil es auf sie weniger verschwommen wirkte. Beim gemusterten Hintergrund nahmen zehn Probanden die beiden 3D-Bilder als gleich wahr und weitere zehn stimmten für das „Threshold 128“-Bild, das ihnen eine Spur schärfer vorkam. Eine Person beobachtete einen „Schiller-Effekt“, der sich nach Angaben des Probanden mit Lichtreflexionen am fahrenden Auto erklären lässt und die Bewegung dadurch realistischer macht.

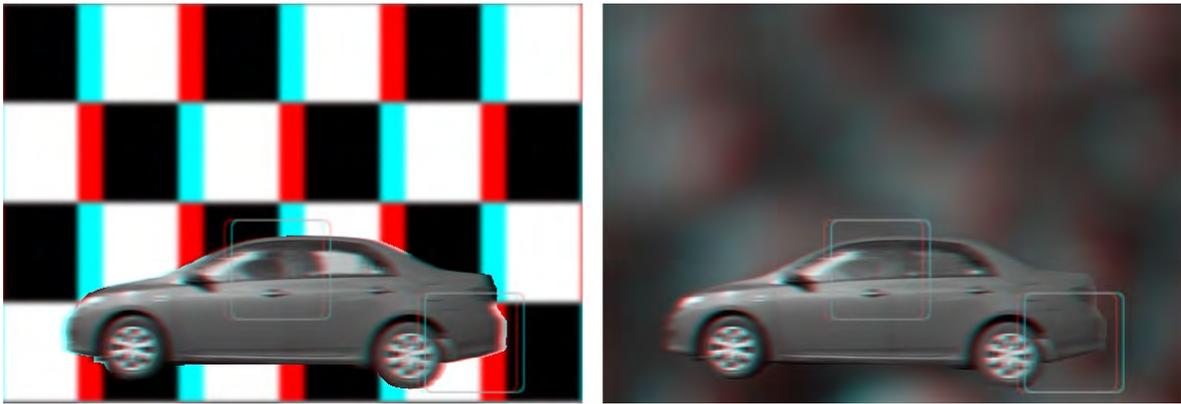


Abbildung 78: Auto („Threshold 60“).

Eine Mehrheit von 19 Probanden lehnte das 3D-Bild mit „Threshold 60“ (siehe Abb. 78) und strukturiertem Hintergrund ab. Besonders negativ empfanden sie die weiße Umrahmung des Autos. Einige Testpersonen konnten nicht genau beschreiben, was sie an dem Bild stört. Sie umschrieben den zuvor genannten Fehler mit unrealistischen Konturen und komischen Kanten am Heck des Wagens. Eine Person nahm ein unpassendes „Blinken“ wahr. Eine genauere Erklärung durch den Probanden hat ergeben, dass es sich um eine Form von binokularer Rivalität handelt. Mehrere Testpersonen bemerkten die Undurchsichtigkeit der Scheibe auf Grund des in ihr abgebildeten, ursprünglichen Hintergrundes. Zwei der Befragten konnten keinen Unterschied zwischen den gezeigten 3D-Bildern feststellen. Die drei Personen, die für das Bild mit „Grenzwert 60“ stimmten, antworteten rein intuitiv. Das Ergebnis des gemusterten Hintergrundes zeigt einen Trend zu „keinen Unterschied“. Die Hälfte der Probanden empfand die beiden zu vergleichenden 3D-Bilder als gleich. Für einen Probanden ist das Auto im Bild mit Alpha-Matting auf der Autobahn unterwegs, während es sich im anderen Bild nur mit Schritttempo bewegt. Jeweils sechs Testpersonen entschieden sich für beziehungsweise gegen das Bild mit Alpha-Matting. Das 3D-Bild mit „Schwellenwert 60“ wurde wegen der zu scharfen Konturen und zu starken, unrealistischen Lichtreflexionen am Heck abgelehnt. Das Hauptargument für das Bild ohne Alpha-Matting war der bessere 3D-Eindruck.

## 7.1.4.2 Bild: Glas



Abbildung 79: Glas („Alpha-Matting“).

Im Gegensatz zur relativ kleinen Fläche der Autofensterscheibe besitzt das Glas in Abb. 79 einen sehr großen, transparenten Bereich. Die bereits beim Auto beschriebenen Fehler, wie der Verlust der Glasstruktur und die Wiedergabe des Originalhintergrundes in den transparenten Regionen, sind noch gravierender und deutlicher sichtbar. Bei knappem Ausschneiden der undurchsichtigen Bereiche („Threshold 200“) verschwindet beinahe die gesamte Glasfläche, während bei grobem Herauslösen des Objektes („Threshold 60“) die Transparenz des Glases verloren geht. Eine Übersicht der Ergebnisse ist in Tabelle 8 zu sehen.

Tabelle 8: Ergebnis „Glas“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	2	0	22	3	0	21	5	2	17
Muster	1	0	23	5	0	19	3	16	5

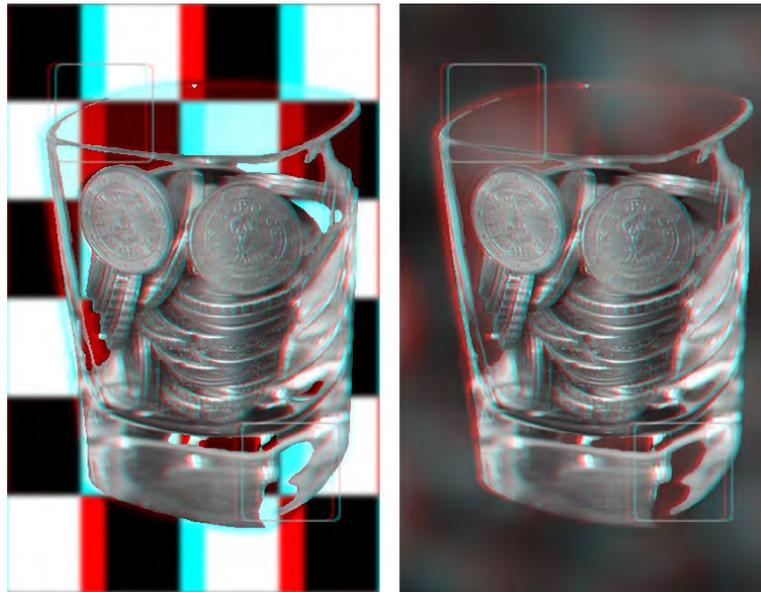


Abbildung 80: Glas („Threshold 200“).

Bei „Threshold 200“ (siehe Abb. 80) fiel das Ergebnis unabhängig vom Hintergrund eindeutig für Alpha-Matting aus. 22 (S) beziehungsweise 23 (M) der Befragten waren der Meinung, dass das unvollständige Glas im Bild ohne Alpha-Matting inakzeptabel ist. Die meisten Testpersonen gaben an, dass die Kanten zu hart, die Übergänge zu scharf und die Glasflächen zu transparent wirken. Eine kleine Gruppe fand den oberen Rand des Glases durchaus realistisch, jedoch das Loch im unteren Teil komisch und unpassend. Zwei Probanden argumentierten, dass die fehlende Lichtreflexion am Glas das Bild unnatürlich erscheinen lässt. Bei einer anderen Testperson kam es zur binokularen Rivalität in Form von „Flackern“. Insgesamt drei Personen bevorzugten das Bild mit „Grenzwert 200“. Keiner der Befragten hielt die beiden gezeigten 3D-Bilder für gleich.

Bei „Threshold 128“ entstand eine ähnliche Verteilung der Antworten. 21 (S) beziehungsweise 19 (M) Probanden stimmten für Alpha-Matting. Die Argumente gegen das Bild mit „Grenzwert 128“ reichten von ausgeschlagenen Kanten über stellenweise zu starker Transparenz bis hin zu seltsamen Bildeindrücken. Ein Proband meinte, dass das Objekt schlecht ausgeschnitten aussieht und ein anderer kritisierte den unpassenden Übergang zwischen dem Glas und den Münzen. Auch die meisten der Kritikpunkte, die bereits bei „Threshold 200“ genannte wurden, wie beispielsweise das Loch im Glas oder das „Flackern“, kamen hier erneut zur Sprache. Niemand gab an, keinen Unterschied zwischen den dreidimensionalen Vergleichsbildern wahrzunehmen. Von den drei (S) beziehungsweise fünf (M) Befragten entschieden sich fast alle intuitiv für das Bild mit „Schwellenwert 128“. Eine Testperson war der Meinung, dass das Glas einheitlicher und daher realistischer wirkte.

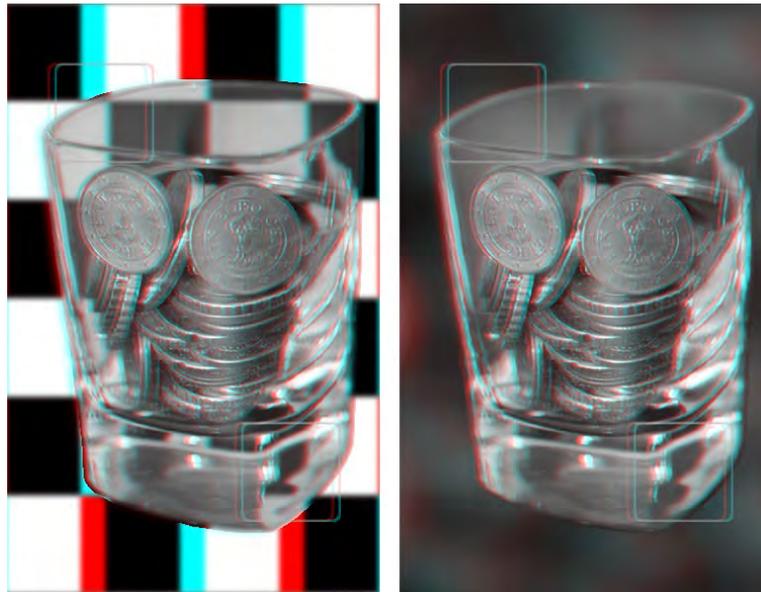


Abbildung 81: Glas („Threshold 60“).

Bei „Threshold 60“ (siehe Abb. 81) ergaben sich für strukturierten und gemusterten Hintergrund unterschiedliche Ergebnisse. Beim Schachbretthintergrund wählten 17 Probanden das Bild mit Alpha-Matting. Hauptursache dafür ist die fehlende Transparenzeigenschaft, die lediglich beim Struktur-Hintergrund sichtbar ist. Die Testpersonen erkannten, dass ein Durchschauen nicht möglich ist. Das Glas scheint mit dem Hintergrund bemalt oder beklebt worden zu sein. Ein paar der Befragten fanden den Übergang zwischen Rand und Rest vom Glas zu hart. Zwei Probanden nahmen keinen Unterschied zwischen den gezeigten 3D-Bildern wahr und fünf stimmten für das Bild mit „Grenzwert 60“. Sie empfanden es echter, da die Kanten klarer definiert sind. Fünf Personen bevorzugten beim gemusterten Hintergrund das Bild mit Alpha-Matting, wegen der harmonischeren Übergänge. 16 gaben an, keinen Unterschied zu bemerken und drei bewerteten das „Threshold 60“-Bild mit realistischer, da es schärfer, besser durchsichtig und sauberer wirkt.

Beim Bild „Auto“ kritisierten manche, dass es selbst für ein Objekt in Bewegung zu unscharf aussieht. Das Hauptaugenmerk lag dennoch bei der Vollständigkeit der Fensterscheibe und bei harmonischen Übergängen zum Hintergrund. Obwohl es unterschiedliche Auffassungen gab, wie viel Transparenz bei einem Glas natürlich ist, hatten die Probanden eine einheitliche Vorstellung von dem Gefäß. Beim Bild „Glas“ mit Alpha-Matting wurde öfters angemerkt, dass das Glas zu milchig erscheint. Trotzdem wirkte es wegen der vorhandener Transparenz und den passenden Spiegelungen vertrauter. Eine falsche Darstellung einer transparenten Fläche erzeugt unabhängig von Größe und Hintergrund einen auffallend unrealistischen Eindruck.

### 7.1.5 Kategorie: „normale“ Objekte

Zu dieser Kategorie werden alle jene Motive gezählt, die Objekte mit scharfen und klar abgrenzbaren Rändern enthalten. Die Abbildungen dieser Objekte weisen trotzdem harmonische Übergänge auf. Die Bereiche, in denen es bei der Generierung synthetischer Ansichten auf Grund gemischter Pixel zu Artefakten kommt, sind relativ klein. Dennoch entstehen durch einfaches Ausschneiden unnatürlich harte Kanten. Die Motive dieser Kategorie sind ein Holzbrett mit aufgezeichnetem Wildschwein und ein Kürbis.

#### 7.1.5.1 Bild: Holzbrett

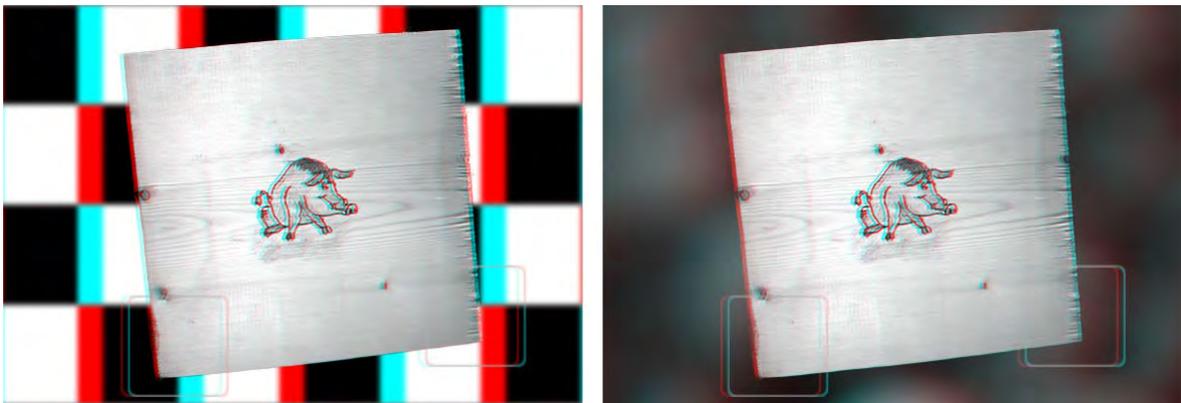


Abbildung 82: Holzbrett („Alpha-Matting“).

Das in Abb. 82 gezeigte Holzbrett weist grundsätzlich klare, geradlinige Kanten auf. Durch die Holzfasern kommt es zu weichen Übergängen zwischen dem Objekt und dem Hintergrund. Bei knappem Ausschneiden („Threshold 200“) geht die Information über die Holzstruktur am Rand verloren. Durch konservatives Herauslösen des Objektes („Threshold 60“) entstehen Hintergrundartefakte. Die Fehler sind im Vergleich zu den vorherigen Motiven weit weniger sichtbar, da nur sehr kleine Regionen betroffen sind. Die Ergebnisse sind in Tabelle 9 zusammengefasst.

Tabelle 9: Ergebnis „Holzbrett“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	3	8	13	3	11	10	2	14	8
Muster	3	6	15	5	15	4	1	20	3

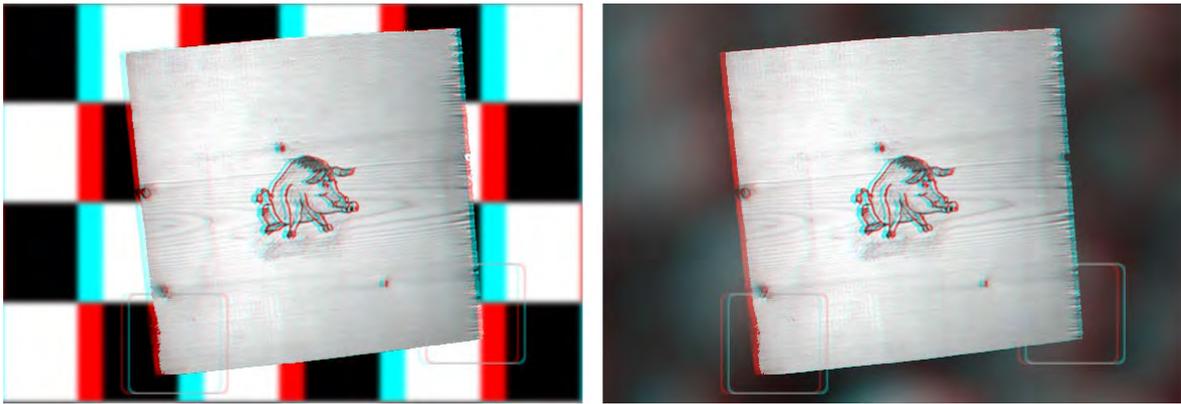


Abbildung 83: Holzbrett („Threshold 200“).

Bei „Threshold 200“ (siehe Abb. 83) zeigen die Ergebnisse einen leichten Trend zu Alpha-Matting. Die 13 Probanden, die beim Struktur-Hintergrund für das 3D-Bild mit Alpha-Matting stimmten, beobachteten im Vergleichsbild ohne Alpha-Matting unangenehme Effekte, wie „Schillern“ oder „Flimmern“. Diese instabile, visuelle Wahrnehmung ist auf binokulare Rivalität zurückzuführen. Beim gemusterten Hintergrund bevorzugten 15 Testpersonen das Bild mit Alpha-Matting. Das Auftreten von binokularer Rivalität ist hier nur ein Grund von vielen. Andere genannte Kritikpunkte in Bezug auf das Bild mit „Grenzwert 200“ waren zu deutliche Ränder, fehlende Kantenteile und unharmonische Übergänge, die nicht zu der Vorstellung von Holz passen. Acht (S) beziehungsweise sechs (M) Probanden nahmen die beiden 3D-Bilder als ident wahr. Bei jedem der beiden Hintergründe gab es drei Stimmen für das Bild ohne Alpha-Matting. Es handelt sich dabei ausschließlich um Personen mit dominantem linkem Auge.

Der Schachbretthintergrund im Bild mit „Threshold 128“ brachte eine knappe Entscheidung. Zehn Probanden sprachen sich für Alpha-Matting aus und elf gaben an, keinen Unterschied zwischen den beiden 3D-Bildern zu sehen. Viele Personen merkten an, dass ihre Entscheidung intuitiv war. Ein paar der Befragten argumentierten, dass im Bild ohne Alpha-Matting Teile vom Rand fehlen beziehungsweise die Ecken verschwimmen. Drei Probanden entschieden sich ohne nähere Angabe von Gründen für das 3D-Bild mit „Grenzwert 128“. Die Verteilung der Antworten beim Bild mit gemustertem Hintergrund fiel deutlicher aus, denn 15 Testpersonen nahmen keine sichtbaren Unterschiede zwischen den gezeigten Bildern wahr. Fehlende Stücke, binokulare Rivalität in Form von „Flackern“ und unnatürliche Kanten im Bild mit „Schwellenwert 128“ gaben vier Probanden als Entscheidungsgrundlage für Alpha-Matting an. Fünf der Befragten bevorzugten das Bild ohne Alpha-Matting, da ihnen die Struktur des Objektes gefühlsmäßig besser gefiel.

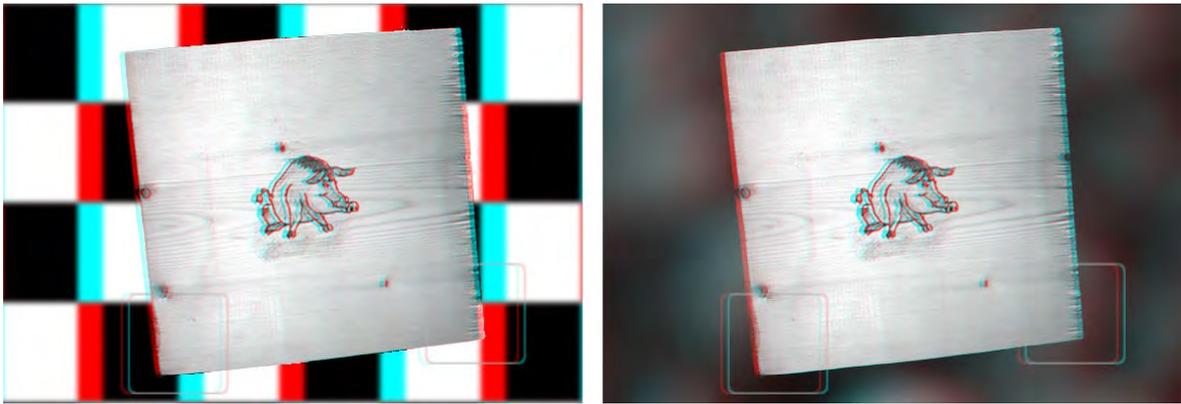


Abbildung 84: Holz Brett („Threshold 60“).

Bei „Threshold 60“ (siehe Abb. 84) bestätigten 14 (S) beziehungsweise 20 (M) Probanden, dass es sich um identische Bilder handelt. Acht Personen stimmten beim strukturierten Hintergrund und drei Personen beim gemusterten Hintergrund für Alpha-Matting. Ihre Beweggründe dafür waren nicht nur der natürlicher wirkende Rand, das plastischere Erscheinen sondern oftmals auch Sympathie. Die Personen mit dominantem, linkem Auge, die ihre Stimme für das Bild mit „Grenzwert 60“ abgaben, nannten Bauchgefühl als Basis für ihre Entscheidung.

#### 7.1.5.2 Bild: Kürbis

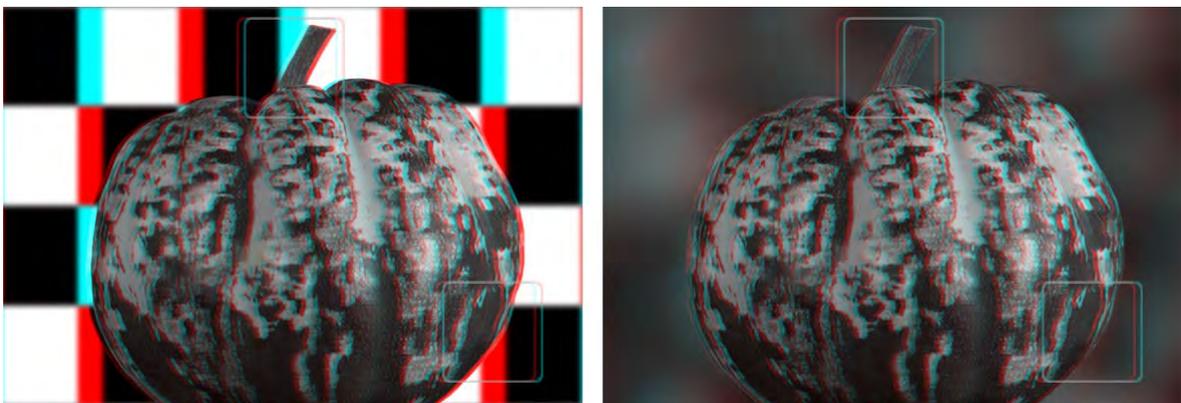


Abbildung 85: Kürbis („Alpha-Matting“).

Der in Abb. 85 dargestellte Kürbis hat im Vergleich zum Holz Brett auf Grund der Materialeigenschaft glattere Ränder. Auch die leicht glänzende Oberfläche trägt dazu bei, dass eventuell auftretende Artefakte als Reflexion gesehen werden können. Bei knappem Ausschneiden des Objektes („Threshold 200“) entstehen dadurch kaum wahrnehmbare Bildfehler. Bei grobem Herauslösen („Threshold 60“) kommt es sehr wohl zu kleinen, aber

sichtbaren Hintergrundfehlern. Tabelle 10 zeigt einen Überblick über die Antworten der Probanden.

Tabelle 10: Ergebnis „Kürbis“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	2	17	5	1	14	9	2	12	10
Muster	1	20	3	2	21	1	1	20	3

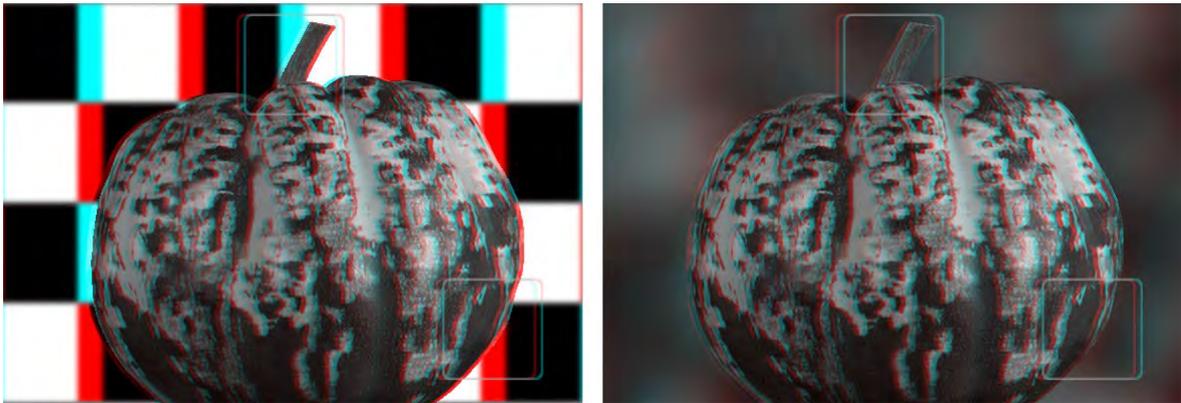


Abbildung 86: Kürbis („Threshold 200“).

Bei „Threshold 200“ (siehe Abb. 86) waren 17 (S) beziehungsweise 20 (M) Probanden der Meinung, dass sich die beiden 3D-Bilder nicht voneinander unterscheiden. Fünf Testpersonen bemerkten beim Schachbretthintergrund unpassende, harte Konturen im Bild mit „Grenzwert 200“. Ebenso empfanden drei Personen das Bild mit gemustertem Hintergrund weniger weich und etwas unschärfer. Insgesamt gab es drei Stimmen für das „Threshold 200“-Bild. Alle drei Entscheidungen wurden intuitiv getroffen.

Auch bei „Threshold 128“ fiel das Ergebnis ähnlich aus. Beim Struktur-Hintergrund nahmen 14 und beim Muster-Hintergrund 21 Probanden zwei idente Kürbisse wahr. Eine Person gab an, dass die beiden 3D-Bilder irgendwie anders aussahen, konnte die Verschiedenheit aber nicht benennen und entschied sich schlussendlich für „kein Unterschied“. Komische Kanten und unpassende Konturen im Bild mit Schachbretthintergrund veranlassten neun der Befragten für Alpha-Matting zu stimmen. Beim gemusterten Hintergrund beobachtete nur eine Testperson Ungereimtheiten an den Kanten im Bild ohne Alpha-Matting. Ein Proband (S) beziehungsweise zwei Probanden (M) fanden den Kürbis im „Threshold 128“-Bild realistischer abgebildet, da er ihnen plastischer erschien.

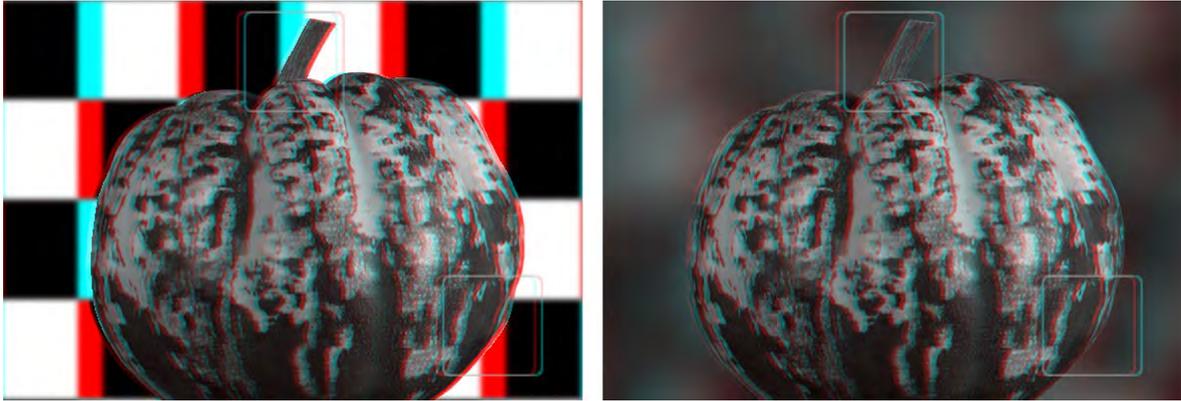


Abbildung 87: Kürbis („Threshold 60“).

Bei „Threshold 60“ (siehe Abb. 87) zeigt sich der Einfluss des Hintergrundes auf die Entscheidungen der Probanden. Der gemusterte Hintergrund brachte einmal mehr die bereits bekannte Verteilung der Antworten hervor. 20 Testpersonen empfanden die beiden zu vergleichenden Bilder als gleich. Drei der Befragten lehnte das Bild ohne Alpha-Matting ab, da es zu ausgeschnitten wirkt und eine Person bevorzugte das „Threshold 60“-Bild wegen der besseren Schärfe. Beim strukturierten Hintergrund entschieden sich zehn Probanden für Alpha-Matting, da das Bild mit „Grenzwert 60“ seltsam ausgefranste Kanten, zu starke Konturen und stellenweise weiße Ränder aufweist. Die Hälfte der Befragten gab an, keinen Unterschied zwischen den gezeigten 3D-Bildern wahrzunehmen. Lediglich zwei Testpersonen bevorzugten intuitiv das Bild ohne Alpha-Matting.

Bei dieser Kategorie wurde sehr häufig nach Gefühl entschieden, welches das realistischere 3D-Bild von beiden ist. Ein wesentliches Entscheidungskriterium sowohl beim Holzbrett als auch beim Kürbis waren dennoch die zu den Materialeigenschaften des Objektes passenden Kanten. Das Bild mit Alpha-Matting, das allgemein glattere Ränder aufweist, wurde aus diesem Grund oft als natürlicher, harmonischer oder sympathischer bezeichnet.

## 7.2 Schlussfolgerungen

Die Vielzahl an unterschiedlichen Antworten zeigt die Subjektivität der menschlichen visuellen Wahrnehmung. Aus der detaillierten Auswertung der Studie in Kapitel 7.1 können dennoch einige wesentliche Aspekte für die Generierung synthetischer Ansichten abgeleitet werden. Eine der wichtigsten Erkenntnisse der Benutzerstudie ist die Tatsache, dass sichtbare Artefakte in den Einzelbildern eines Stereobildpaars auch in den dreidimensionalen Bildern wahrgenommen werden, wie die Entscheidungen der Probanden belegen. Bei fast allen getesteten Kombinationen aus Kategorie, Grenzwert und Hintergrund stellten die Testpersonen einen Unterschied zwischen den 3D-Bildern mit beziehungsweise ohne Alpha-Matting fest. Der Trend geht dabei eindeutig Richtung Alpha-Matting. Insgesamt entschieden sich die Probanden 882-mal für Alpha-Matting. Das entspricht einem Prozentsatz von 61,25 Prozent. Jede Testperson hat im Durchschnitt rund 37-mal für das Bild mit Alpha-Matting gestimmt. Eine mittlere Abweichung von 6,56 Prozent bestätigt die weitgehende Einigkeit der Testpersonen.

Tabelle 11: Ergebnis „Hintergrund“.

	„Threshold 200“			„Threshold 128“			„Threshold 60“		
	-1	0	1	-1	0	1	-1	0	1
Struktur	20	38	182	25	38	177	33	36	171
Muster	19	36	185	45	79	116	30	159	51

Die Annahme, dass die Art des Hintergrundes eine Rolle spielt, konnte durch die Studie bestätigt werden, wie die Summe über alle Motive in Tabelle 11 zeigt. Beim Struktur-Hintergrund wurde öfters für Alpha-Matting gestimmt als beim Muster-Hintergrund. Dies lässt sich damit erklären, dass Hintergrundartefakte beim klar strukturierten Schachbrett zu sehen sind, während sie im unregelmäßig gemusterten Hintergrund untergehen. Allerdings blieben Fehler im Vordergrundobjekt bei beiden Hintergründen nicht unbemerkt. Dieses Erkenntnis spiegelt sich auch bei der Auswertung getrennt nach Grenzwerten wieder. Mit Ausnahme beim Motiv „Kürbis“ bevorzugten die Probanden – unabhängig vom Hintergrund – das 3D-Bild mit Alpha-Matting gegenüber dem 3D-Bild mit „Threshold 200“. Obwohl der Informationsverlust bei „Threshold 128“ nicht ganz so groß ist, wurden die 3D-Bilder ohne Alpha-Matting weitgehend als inakzeptabel abgelehnt. Bei einem strukturierten Hintergrund treten in kleineren Regionen zusätzlich sichtbare Hintergrundartefakte auf. Dies hat zur Folge, dass das Ergebnis beim Struktur-Hintergrund klarer für Alpha-

Matting ausfällt, als es beim Muster-Hintergrund der Fall ist. Noch deutlicher zeichnete sich der Unterschied zwischen den beiden Hintergründen bei „Threshold 60“ ab. Die Bilder mit Schachbrettmuster wurden auf Grund der unübersehbaren Hintergrundartefakte abgelehnt, während die Testpersonen beim verschwommenen Muster keine Unterschiede wahrgenommen haben. Daraus lässt sich ableiten, dass ein konservatives Ausschneiden der Objekte bei unscharfen, verwaschenen Hintergründen eine qualitativ gleichwertige Alternative zur Berechnung mit Alpha-Matting darstellt.

Die Analyse der einzelnen Kategorien ergab, dass eine korrekte Darstellung der transparenten Regionen Grundvoraussetzung für qualitativ hochwertige 3D-Bilder ist. Die Studie hat gezeigt, dass bei allen Motiven mit transparenten Regionen die 3D-Bilder mit Alpha-Matting als realistischer und glaubwürdiger bezeichnet wurden. Artefakte, die bei der Generierung von synthetischen Ansichten ohne Alpha-Matting entstehen, entfremden den eigentlichen Bildinhalt. Ein geworfener Würfel, der in der Luft stillzustehen scheint, ein Tier ohne Beine oder löchrige Gläser wirken auf einen Betrachter irritierend und falsch. Aus den Antworten der Probanden geht hervor, dass Artefakte in der Struktur des Vordergrundobjektes besonders störend wirken. Häufig entstehen dadurch große Unterschiede zwischen den beiden Ansichten, die zu unangenehmen und instabilen Wahrnehmungen in Folge von binokularer Rivalität führen. Lediglich bei der Kategorie „normale Objekte“ konnten die Probanden keinen bewussten Unterschied zwischen dem 3D-Bild mit beziehungsweise ohne Alpha-Matting feststellen. Dennoch wurde das Bild mit Alpha-Matting immer wieder mit „irgendwie sympathischer“ beschrieben, was auf den natürlichen Effekt der „Halbverdeckung“ zurückzuführen ist.

Die Kommentare der Probanden haben verdeutlicht, dass das visuelle System des Menschen die Fehler, die bei der Berechnung von neuen Ansichten entstehen, nur in einigen wenigen Fällen ausgleichen kann. Ein wichtiger Aspekt ist daher die Reduktion der Artefakte in den Einzelansichten, um die 3D-Bildqualität zu verbessern. Auf diese Weise kann dem Benutzer die Wahrnehmung von 3D-Inhalten einfacher und angenehmer gestaltet werden. Dies stellt eine wesentliche Grundlage für die Akzeptanz der 3D-Technologie dar. Das Ergebnis der Studie spricht somit eindeutig für einen standardmäßigen Einsatz von Alpha-Matting bei der Generierung von synthetischen Ansichten.

## 8 Zusammenfassung

Die meisten modernen 3D-Systeme basieren auf den Prinzipien der Stereoskopie. Der menschlichen visuellen Wahrnehmung nachempfunden, werden dem linken und dem rechten Auge getrennt, leicht unterschiedliche Ansichten einer Szene dargeboten. Auf Grund der enthaltenen Querdparität werden die beiden Einzelbilder im Gehirn des Menschen zu einem räumlichen Bild verschmolzen. Entsprechen die beiden Einzelbilder nicht den natürlichen Kriterien des visuellen Systems, kommt es zu unangenehmen Wahrnehmungserlebnissen.

Im Zuge dieser Arbeit wurde mit Hilfe einer Benutzerstudie die Frage, welchen Einfluss die Verwendung von Alpha-Matting bei der Generierung von synthetischen Ansichten auf die 3D-Bildqualität hat, beantwortet. Vor der Durchführung der Studie wurden geeignete Motive ausgewählt und in die fünf Kategorien „Bewegungsunschärfe“, „Tiefenunschärfe“, „Struktur“, „Transparenz“ und „normale Objekte“ unterteilt. Jedes Motiv wurde vor zwei unterschiedliche Hintergründe platziert. Der erste Hintergrund weist eine klare, regelmäßige Struktur auf, während der andere ein verschwommenes Muster zeigt. Die zweite Ansicht, die für ein dreidimensionales Bild notwendig ist, wurde entweder mit Alpha-Matting oder mit Schwellenwerten berechnet. Ziel der Benutzerstudie war die Feststellung, ob es sichtbare Unterschiede zwischen dem 3D-Bild mit oder ohne Alpha-Matting gibt und ob dies mit speziellen Hintergründen, bestimmten Grenzwerten oder einzelnen Motiven zusammenhängt. Im Zuge einer Vorstudie wurde auch der Einfluss der okularen Dominanz untersucht. Dies brachte bei der Vorstudie jedoch kein aussagekräftiges Ergebnis hervor und wurde daher in der Hauptstudie nicht umfangreich berücksichtigt.

An der Studie nahmen 24 Personen im Alter von 21 bis 57 Jahren teil. Mit Hilfe von Shutter-Brille und 3D-Monitor bekamen sie insgesamt sechzig Testbilder zu begutachten, die immer aus einem 3D-Bild mit und einem ohne Alpha-Matting bestanden. Die Probanden hatten die Aufgabe festzustellen, ob sie Unterschiede zwischen den beiden gezeigten 3D-Bildern wahrnehmen konnten. Bei sichtbaren Unterschieden sollten sie zusätzlich angeben, welches der beiden 3D-Bilder echter, realistischer, natürlicher wirkte.

Bereits in der Vorstudie war ein eindeutiger Trend Richtung Alpha-Matting erkennbar, der sich in der Hauptstudie bestätigte. Außerdem zeigte die Benutzerstudie, dass sichtbare Fehler in den Einzelbildern auch im dreidimensionalen Bild wahrgenommen wurden und dass

– mit sehr wenigen Ausnahmen – immer Unterschiede zwischen den 3D-Bildern mit beziehungsweise ohne Alpha-Matting entdeckt wurden. Des Weiteren konnte festgestellt werden, dass die Art des Hintergrundes und die Wahl des Grenzwertes eine Rolle bei der Entscheidung spielen. Das Ergebnis der Benutzerstudie spricht für einen Einsatz von Alpha-Matting bei der Generierung von synthetischen Ansichten, weil dadurch Artefakte in Bildregionen mit gemischten Pixeln, die das menschliche, visuelle System nicht ausgleichen kann, vermieden werden.

Mögliche, weiterführende Themen zu dieser Arbeit sind neben der Erweiterung der Untersuchung auf bewegte Bilder (3D-Video) auch die Erforschung, welche Rolle die okulare Dominanz bei der Betrachtung von 3D-Inhalten spielt. Im Zuge der Benutzerstudie gab es immer wieder Hinweise, die vermuten lassen, dass die Augendominanz bei der Darstellung von 3D-Bildern in manchen Fällen berücksichtigt werden sollte. Eine interessante Frage in diesem Zusammenhang ist, ob das Wissen über okulare Dominanz ähnlich wie jenes über binokulare Unterdrückung zur Datenreduktion ohne qualitative Verluste genutzt werden kann.

## Literaturverzeichnis

- [AS] Actuality Systems: *Perspecta*, Website: <http://actuality-medical.com/indexAS.html>, Seitenaufruf: Dezember 2010.
- [B04] V. Bruce, P.R. Green, M.A. Georgeson: *Visual Perception – Physiology, Psychology and Ecology*, 4th Edition, Psychology Press, Hove/New York, 2004.
- [B05] M. Bleyer, M. Gelautz: *A Layered Stereo Matching Algorithm Using Image Segmentation and Global Visibility Constraints*, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 59, No. 3, pp. 128-150, April 2005.
- [B07] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, C. von Kopylow: *A Survey of 3DTV Displays: Techniques and Technologies*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 11, pp. 1647-1658, November 2007.
- [B09] M. Bleyer, M. Gelautz, C. Rother, C. Rhemann: *A Stereo Approach that Handles the Matting Problem via Image Warping*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 501-508, Juni 2009.
- [D05] N.A. Dodgson: *Autostereoscopic 3D Displays*, Computer, Vol. 38, No. 8, pp. 31-36, August 2005.
- [DE05] B. Delaney: *Forget the Funny Glasses [autostereoscopic display systems]*, IEEE Computer Graphics and Applications, Vol. 25, No. 3, pp. 14-19, Mai/Juni 2005.
- [D06] T. Ditzinger: *Illusionen des Sehens – Eine Reise in die Welt der visuellen Wahrnehmung*, 1. Auflage, Elsevier Spektrum Akademischer Verlag, München, 2006.
- [F05] G.E. Favalora: *Volumetric 3D Displays and Application Infrastructure*, Computer, Vol. 38, No. 8, pp. 37-44, August 2005.
- [G01] R.L. Gregory: *Auge und Gehirn – Psychologie des Sehens*, Deutsche Erstausgabe, Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg, 2001.
- [G08] E. B. Goldstein, H. Irtel [Hrsg.]: *Wahrnehmungspsychologie – Der Grundkurs*, 7. Auflage, Spektrum Akademischer Verlag, Berlin/Heidelberg, 2008.
- [H95] D.H. Hubel: *Eye, Brain, and Vision*, 1995.  
Website: <http://hubel.med.harvard.edu>, Seitenaufruf: August 2010.

- [K07] J. Konrad, M. Halle: *3-D Displays and Signal Processing – An Answer to 3-D Ills?*, IEEE Signal Processing Magazine, Vol. 24, No. 6, pp. 97-111, November 2007.
- [K10] H. Kannan, K.N. Iyer, K. Maiti, D. Purbiya, A. Bopardikar, A. Sharma: *Alpha Model Based Mixed Pixel Processing for View Synthesis*, IEEE 3DTV-Conference: The True Vision – Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1-4, Juni 2010.
- [L08] A. Levin, D. Lischinski, Y. Weiss: *A Closed-Form Solution to Natural Image Matting*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 2, pp. 228-242, Februar 2008.
- [LS] LightSpace Technologies: *DepthCube*, Website: <http://www.lightspacetech.com/index.html>, Seitenaufruf: Dezember 2010.
- [M75] W. Metzger: *Gesetze des Sehens*, 3. Auflage, Kramer, Frankfurt am Main, 1975.
- [M07] P. Merkle, A. Smolic, K. Müller, T. Wiegand: *Efficient Prediction Structures for Multiview Video Coding*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 11, pp. 1461-1473, November 2007.
- [M08] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, T. Wiegand: *Reliability-based Generation and View Synthesis in Layered Depth Video*, IEEE International Workshop on Multimedia Signal Processing (MMSP), pp. 24-39, Oktober 2008.
- [MW08] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, T. Wiegand: *View Synthesis for Advanced 3D Video Systems*, EURASIP Journal on Image and Video Processing, Vol. 2008, pp. 1-11, November 2008.
- [NV] NVIDIA: *3D Vision*, Website: <http://www.nvidia.de/object/3d-vision-main-de.html>, Seitenaufruf: April 2011.
- [P84] T. Porter, T. Duff: *Compositing Digital Images*, Computer Graphics (ACM SIG-GRAPH), Vol. 18, No. 3, pp. 253-259, Juli 1984.
- [P10] S. Pastoor, M. Wöpking: *3-D Displays: A Review of Current Technologies*, Website: <http://tinyurl.com/4lrn8>, Seitenaufruf: August 2010.
- [R08] C. Rhemann, C. Rother, A. Rav-Acha, T. Sharp: *High Resolution Matting via Interactive Trimap Segmentation*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, Juni 2008.

- [R09] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, P. Rott: *A Perceptually Motivated Online Benchmark for Image Matting*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1826-1833, Juni 2009.
- [S96] A.T. Smith, J.F. Blinn: *Blue Screen Matting*, Proc. of ACM SIGGRAPH, pp. 259-268, August 1996.
- [S98] R. Szelinski, P. Golland: *Stereo Matching with Transparency and Matting*, 6th International Conference on Computer Vision (ICCV), pp. 517-524, Januar 1998.
- [S00] L. Stelmach, W.J. Tam, D. Meegan, A. Vincent: *Stereo Image Quality: Effects of Mixed Spatio-Temporal Resolution*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 2, pp. 188-193, March 2000.
- [S08] A. Smolic, K. Müller, K. Dix, P. Merkle, P. Kauff, T. Wiegand: *Intermediate View Interpolation Based on Multiview Video Plus Depth for Advanced 3D Video Systems*, IEEE International Conference on Image Processing (ICIP), pp. 2448-2451, Oktober 2008.
- [SS] Sharp-Sighted: *Ocular Dominance*, Website: <http://www.sharp-sighted.org>, Seitenaufruf: Mai 2011.
- [SV09] A. Smolic, K. Müller, P. Merkle, A. Vetro: *Development of a New MPEG Standard for Advanced 3D Video Applications*, Proc. of the 6th International Image and Signal Processing and Analysis (ISPA), pp. 400-407, September 2009.
- [SW09] A. Smolic, K. Müller, P. Merkle, P. Kauff, T. Wiegand: *An Overview of Available and Emerging 3D Video Formats and Depth Enhanced Stereo as Efficient Generic Solution*, Proc. Picture Coding Symposium (PCS), pp. 1-4, Mai 2009.
- [V65] O. Vierling: *Die Stereoskopie in der Photographie und Kinematographie*, Wissenschaftliche Verlagsgesellschaft, Stuttgart, 1965.
- [W95] B. A. Wandell: *Foundations of Vision*, Sinauer Associates, Sunderland, 1995.
- [W07] J. Wang, M. Cohen: *Image and Video Matting: A Survey*, Foundations and Trends in Computer Graphics and Vision, Vol. 3, No. 2, pp.97-175, Januar 2007.
- [X07] W. Xiong, J. Jia: *Stereo Matching on Objects with Fractional Boundary*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, Juni 2007.
- [Z04] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, R. Szeliski: *High-Quality Video View Interpolation Using a Layered Representation*, ACM SIGGRAPH and ACM Trans. on Graphics, Vol. 23, No. 3, pp. 600-608, August 2004.

## Abbildungsverzeichnis

Abbildung 1: Das Auge. Abbildung aus [G08] .....	10
Abbildung 2: Überblick über das visuelle System. Abbildung aus [G08] .....	11
Abbildung 3: Gestaltprinzipien. Abbildung aus [B04]; [G08]; [D06] .....	12
Abbildung 4: Mehrdeutigkeit der Abbildung eines Objekts. Abbildung aus [G08] .....	14
Abbildung 5: Prägung. Delfin oder Känguru? Abbildung aus [D06] .....	14
Abbildung 6: Modell nach David Marr. Abbildung aus [G08] .....	15
Abbildung 7: Geons nach Irving Biedermann. Abbildung aus [G08] .....	15
Abbildung 8: Konvergenz. Abbildung aus [G08] nach [B04] .....	16
Abbildung 9: Verdeckung. Abbildung aus [G08] .....	17
Abbildung 10: Texturgradient. Abbildung aus [G08] .....	18
Abbildung 11: Größenvergleich. Abbildung aus [G08] .....	19
Abbildung 12: Atmosphärische Perspektive. Abbildung aus [G08] .....	19
Abbildung 13: Schattierung. Abbildung aus [W95] .....	20
Abbildung 14: Schatten. Abbildung aus [B04] .....	20
Abbildung 15: Bewegung. Abbildung aus [G08] .....	21
Abbildung 16: Das „Finger-Experiment“. Abbildung aus [G08] .....	23
Abbildung 17: Horopter. Abbildung aus [G08] .....	24
Abbildung 18: Querdisparität. Abbildung aus [G08] .....	24
Abbildung 19: Ungekreuzte und gekreuzte Querdisparität. Abbildung aus [G08] .....	25
Abbildung 20: Zufallspunktstereogramm. Abbildung aus [G08] .....	26
Abbildung 21: Halb-Verdeckung. Abbildung aus [W95] .....	27
Abbildung 22: Binokulare Rivalität. Abbildung aus [H95] .....	29
Abbildung 23: Rivalität von Strukturen. Abbildung aus [D06] .....	29
Abbildung 24: Binokulare Unterdrückung. Abbildung aus [D06] .....	30
Abbildung 25: Rivalität von Farbe. Abbildung aus [D06] .....	30
Abbildung 26: Miles Test. Abbildung aus [SS]. .....	31
Abbildung 27: Stereoblick. Abbildung aus [D06] .....	33
Abbildung 28: Spiegelstereoskop. Abbildung aus [D06] .....	34
Abbildung 29: Linsenstereoskop. Abbildung aus [D06] .....	35
Abbildung 30: Autostereoskopische Bildschirmsysteme. Abbildung aus [D05] .....	39
Abbildung 31: „Sweet Spots“. Abbildung aus [D05] .....	40
Abbildung 32: Head-Tracking Mechanismen. Abbildung aus [D05] .....	41
Abbildung 33: Automultiskopischer Bildschirm. Abbildung aus [D05] .....	42
Abbildung 34: Volumetrische Bildschirm Systeme. Abbildung aus [DE05] .....	43
Abbildung 35: Anforderungen an einen 3D-Video Standard. Abbildung aus [SV09] .....	44
Abbildung 36: „Multiview Video Coding“-Konzept. Abbildung aus [SW09] .....	45
Abbildung 37: Bildpaar mit unterschiedlicher Auflösung. Abbildung aus [SW09] .....	46
Abbildung 38: Video plus Depth. Abbildung aus [SW09] .....	47

Abbildung 39: Grundidee der erweiterten 3D-Video Formate. Abbildung aus [SV09]	48
Abbildung 40: „Multiview Video plus Depth“-Konzept. Abbildung aus [S08]	48
Abbildung 41: Layered Depth Video. Abbildung aus [SV09]	49
Abbildung 42: „Depth Enhanced Stereo“-Konzept. Abbildung aus [SV09]	50
Abbildung 43: Ursprüngliches Bild und Trimap. Abbildung aus [WC07]	52
Abbildung 44: Closed-form Matting. Abbildung aus [L08]	54
Abbildung 45: Stereo-Matching. Abbildung aus [B05]	55
Abbildung 46: Farbkonstanz in einem Stereobildpaar. Abbildung aus [X07]	56
Abbildung 47: Ausschnitt einer synthetischen Ansicht. Abbildung aus [K10]	58
Abbildung 48: NVIDIA „3D Vision“-System. Abbildung aus [NV]	59
Abbildung 49: Testbilder (Auswahl). Abbildung nach [R09]	60
Abbildung 50: Hintergrundbilder.	60
Abbildung 51: Erstellung eines Bildes. Abbildung nach [R09]	61
Abbildung 52: Synthetische Ansicht.	62
Abbildung 53: Stereobildpaar.	63
Abbildung 54: Anordnung der Stereobildpaare.	64
Abbildung 55: Augendominanz-Testbild.	64
Abbildung 56: Auswirkung unterschiedlicher Hintergründe.	67
Abbildung 57: Stereobildpaar mit gekennzeichneten Bildregionen.	68
Abbildung 58: Würfel („Alpha-Matting“).	71
Abbildung 59: Würfel („Threshold 200“).	72
Abbildung 60: Würfel („Threshold 60“).	73
Abbildung 61: Windrad („Alpha-Matting“).	73
Abbildung 62: Windrad („Threshold 200“).	74
Abbildung 63: Windrad („Threshold 60“).	75
Abbildung 64: Knospe („Alpha-Matting“).	76
Abbildung 65: Knospe („Threshold 200“).	77
Abbildung 66: Knospe („Threshold 60“).	78
Abbildung 67: Libelle („Alpha-Matting“).	79
Abbildung 68: Libelle („Threshold 200“).	80
Abbildung 69: Libelle („Threshold 60“).	81
Abbildung 70: Koala („Alpha-Matting“).	82
Abbildung 71: Koala („Threshold 200“).	83
Abbildung 72: Koala („Threshold 60“).	84
Abbildung 73: Pfau („Alpha-Matting“).	85
Abbildung 74: Pfau („Threshold 200“).	86
Abbildung 75: Pfau („Threshold 60“).	87
Abbildung 76: Auto („Alpha-Matting“).	88
Abbildung 77: Auto („Threshold 200“).	89
Abbildung 78: Auto („Threshold 60“).	90
Abbildung 79: Glas („Alpha-Matting“).	91
Abbildung 80: Glas („Threshold 200“).	92

---

Abbildung 81: Glas („Threshold 60“). .....	93
Abbildung 82: Holzbrett („Alpha-Matting“). .....	94
Abbildung 83: Holzbrett („Threshold 200“). .....	95
Abbildung 84: Holzbrett („Threshold 60“). .....	96
Abbildung 85: Kürbis („Alpha-Matting“). .....	96
Abbildung 86: Kürbis („Threshold 200“). .....	97
Abbildung 87: Kürbis („Threshold 60“). .....	98