



UNIVERSITÄT
KOBLENZ · LANDAU

Fachbereich 4: Informatik



Arbeitsgruppe Aktives Sehen

Interaktion von Bild und Ton – „BildKlang“

Studienarbeit im Studiengang Computervisualistik

vorgelegt von

Anabell Heiß

Betreuer: Dipl.-Inf. Detlev Droege, Institut für Computervisualistik, Fachbereich Informatik

Koblenz, im März 2007

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Richtlinien der Arbeitsgruppe für Studien- und Diplomarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden. ja nein

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu. ja nein

Koblenz, den

Unterschrift

Inhaltsverzeichnis

1	Einleitung	7
1.1	Motivation	7
1.2	Gliederung der Arbeit	8
2	Stand der Wissenschaft	9
2.1	Farbsegmentierung zur Bildanalyse	9
2.1.1	Color Structure Code	10
2.2	Kunst und Medien	12
2.2.1	Medienkunst	12
2.2.2	Farb- und Klangwahrnehmung	15
2.2.3	Farbenlehre nach Goethe	15
3	Eigener Ansatz	17
3.1	Einleitung	17
3.2	Voraussetzung	21
3.3	Segmentierung von Farbbildern	21
3.4	Zuordnung von Bildelement zu Ton	22
3.5	Technisches Design	25

3.5.1	Übersicht	25
3.5.2	Module	25
3.6	Die GUI	28
3.7	Eingabegeräte	29
3.8	Die Audiobibliothek	29
3.8.1	Simple DirectMedia Layer	29
3.8.2	Klangbibliothek	31
4	Entwicklung und Erkenntnisse	35
4.1	Entwicklung	35
4.2	Spielaufbau	35
4.3	Erkenntnisse: Segmentierung von Farbbildern	36
4.3.1	Verifikation	36
4.4	Erkenntnisse: Zuordnung von Bildelement und Ton	39
4.4.1	Extremfälle	41
4.4.2	Verifikation	42
5	Zusammenfassung	43
5.1	Fazit	43
5.2	Ausblick	44
A	Anhang A	45

Kapitel 1

Einleitung

1.1 Motivation

Ziel der Arbeit war es ein audiovisuelles System aufzubauen, das die Interaktion mit Bild und Ton ermöglicht und die Natur der synästhetischen Wahrnehmung nutzt. Seit der Antike beschäftigen sich Menschen mit der Frage nach dem Zusammenhang und der Verwandtschaft von Tönen und visuellen Elementen wie Farben oder geometrischen Formen. Welches Bild baut sich beim Hören von Musik mit Hilfe der synästhetischen Wahrnehmung in unserer Vorstellung auf? Ein Musikstück kann “weich“ oder “kantig“ klingen, ein Klang wird als “warm“ oder “kühl“ empfunden. Die Umkehr dieser Frage stellt die Grundlage dieser Studienarbeit dar:

Welchen Klang erzeugt ein Bild?

Wie klingen kleine Objekte im Gegensatz zu großen, runde Objekte im Gegensatz zu eckigen und welchen Klängen entsprechen verschiedene Farben.¹ Wie sind die Erwartungen an den Klang eines Bildes und werden diese erfüllt?

Das Gesamtsystem besteht aus mehreren Komponenten. Zum einen bildet die Segmentierung des Farbbildes einen großen Teil des Ganzen, sie schafft die Basis. Darüber hinaus

¹siehe dazu Kapitel 3.4

spielt die harmonische Zuordnung und Verknüpfung von Bildeigenschaft und Ton eine wichtige Rolle in dieser Arbeit. Es wurde ein Verfahren implementiert, welches ein Eingabebild und die Objekte darin segmentiert, analysiert und Elementen eines Tonsystems zuordnet. Die Objekte im Bild, verknüpft mit Tönen einer Audiobibliothek, sollen ein harmonisches Klangbild ergeben. Das Bild soll wie ein Instrument bespielbar sein, so dass eine Interaktion durch den Benutzer möglich wird.

Der Benutzer soll die Objekte im Bild wie Noten in einem Tonsystem begreifen und den individuellen Klang des Bildes durch sein Spiel darauf entdecken können.

1.2 Gliederung der Arbeit

Nach der Einleitung in Kapitel 1 wird in Kapitel 2 zunächst der in die Arbeit integrierte `Color Structure Code` beschrieben und erklärt, weshalb er in seiner Funktionalität für die Segmentierung ausgewählt wurde. Ebenso werden die Grundlagen der Farbenlehre von Goethe in Bezug auf die relevanten Teile wiedergegeben. In Kapitel 3 wird der eigene Ansatz und die Umsetzung der Problemstellung beschrieben. Kapitel 4 befasst sich mit dem Entstehungs- und Entwicklungsprozess des audiovisuellen Systems und den Erkenntnissen, die während der Arbeit daran erlangt wurden. Das abschließende Kapitel 5 bildet das Fazit der vorliegenden Studienarbeit und weist Möglichkeiten zur Verbesserung und Erweiterung des Systems auf.

Kapitel 2

Stand der Wissenschaft

2.1 Farbsegmentierung zur Bildanalyse

Das im Rahmen dieser Studienarbeit entwickelte audiovisuelle System soll als Eingabe vor allem Farbbilder akzeptieren und verarbeiten. Es sollen allerdings alle Bilder, auch schwarz-weiße, beispielbar sein. Daher wurde nach einem Segmentierungsverfahren sowohl für Farbbilder als auch für Grauwertbilder gesucht. In der Vorbereitungs- und Literaturrecherchephase stellte sich heraus, dass sich der Blick in aktuellen Arbeiten und Projekten verstärkt auf die Verarbeitung und Analyse im 3D-Bereich richtet. Im Falle der Zielsetzung von *BildKlang* ist die Analyse im 2D-Bereich erforderlich. Daher waren die 3D-Verfahren, die in jüngster Zeit entwickelt wurden, nicht von weiterem Interesse für die Belange dieser Arbeit. Im 2D-Bereich behandelten Arbeiten der letzten Jahre vor allem die Analyse von Grauwertbildern. Es sollte nun ein Verfahren für Farbbilder implementiert werden. Gesucht wurden mehrere zum Gesamtsystem führende Komponente, insbesondere im Bereich der 2D-Form- und Bildanalyse von Farbbildern.

Regionenorientierte Verfahren, die zu den lokal beschränkten Modellen gehören, untersuchen zusammenhängende Flächen homogener Farbe und/oder Textur¹. Ebenfalls diesem Modell zugeordnet werden die kantenbasierten Methoden, welche hauptsächlich die Kon-

¹*Regiongrowing*

turen der Objekte extrahieren². Zu den globalen Modellen gehören die modellbasierten Verfahren. Häufig werden auch kantenbasierte Methoden angewandt, z.B. benutzt man Fourier-Deskriptoren zur Erhaltung der Formkanteninformation.

Neben der Bereitstellung eines audiovisuellen Gesamtsystems zur Interaktion mit Bild und Ton, besteht ein weiteres Ziel dieser Arbeit darin, die menschliche Wahrnehmung von Formen miteinzubeziehen. Zur Segmentierung von Farbbildern kann man sich an Punkten, Linien/Kanten und Regionen orientieren, die sich anhand ihrer Farbe ähneln. Für die Aufgabenstellung dieser Studienarbeit war das Ziel eine Regionensegmentierung, da die Objekte im Bild gefunden und mit Eigenschaften verknüpft werden sollten. Für die Analyse im Hinblick auf die Fragestellung dieser Arbeit reicht eine Farbsegmentierung als Grundlage aus, um deren Ergebnis mit dem Audioteil zu verknüpfen und erste Klänge zu erzeugen.

2.1.1 Color Structure Code

Der `Color Structure Code` wurde bereits 1994 implementiert [Reh94]. Dieser Algorithmus zur Farbsegmentierung wurde anschließend 2003 von Lutz Prieße und Patrick Sturm aus der Arbeitsgruppe Bildverarbeitung der Universität Koblenz-Landau überarbeitet. Er bildet die Grundlage für die Bildanalyse in dieser Studienarbeit.

Segmentiert werden Grauwertbilder sowie Bilder im HSV- und RGB-Farbraum. Bei der Segmentierung von Farbbildern richtet sich der Blick auf die Verteilung der Farbwerte von Regionen, aber auch von Linien und Punkten. Im Falle der Aufgabenstellung dieser Arbeit ist die Regionensegmentierung von entscheidender Bedeutung, weshalb der CSC Verwendung findet. Der CSC arbeitet schnell, einfach, robust und ermöglicht ein qualitativ hochwertiges Ergebnis. Er verbindet die Vorteile lokaler Verfahren, wie das angewandte Regionwachstum mit den Vorteilen der Verfahren, welche eine globale Sicht auf die Daten haben, wie zum Beispiel dem Histogramm-Algorithmus [PS03]. Der Algorithmus segmentiert parallel und hierarchisch in überlappenden, hexagonalen Inselstrukturen. Die kleinsten Strukturelemente sind die Pixel selbst. Sieben Pixel bilden eine Insel. Diese

²nur erwähnt und nicht weiter erklärt werden an dieser Stelle „aktive Konturen“ und „Fourier-Deskriptoren“

neu entstandenen Inseln werden „Inseln der Ebene 0“ genannt. Weiterführend bilden im nächsten Schritt jeweils sieben Inseln eine neue „Insel der Ebene 1“ und so weiter, bis am Ende nach N Schritten das Bild von einer „Insel der Ebene N “ bedeckt wird. Diese Insel auf Ebene N beinhaltet schließlich sieben „Inseln der Ebene $N - 1$ “. Sobald $N = 0$, besteht jede Insel der Ebene nicht mehr aus Inseln, sondern aus sieben atomaren Pixeln. Die Inseln überlappen in genau einer gemeinsamen Subinsel. Im Falle der untersten Ebene ($N = 0$) überlappen sie in einem gemeinsamen Pixel. Jede Region im Bild wird von einem so genannten Codeelement repräsentiert. Dieses Codeelement hat die Datenstruktur eines Wortes der Länge 32 Bit und enthält unter anderem Informationen über den gemittelten Farbwert der Region, die weitere Verschmelzbarkeit derselben und darüber welche Form diese hat. Eine Insel kann höchstens sieben Codeelemente beinhalten, welches die Verbindung von Subinseln bedeutet. Eine Verbindung und somit die Kennzeichnung mit demselben Codeelement entsteht dann, wenn es eine Ähnlichkeit in den Farbwerten gibt. Nach Abschluss der Segmentierung wird das ganze Bild, d.h. die Insel der höchsten Ebene³ von Codeelementen bedeckt. Die Inseln und deren Subinseln werden rekursiv auf Farbähnlichkeit geprüft. Wenn sie verbunden sind, werden zwei Subcodeelemente einer Insel zu einem Codeelement zusammengefasst. Ob zwei Elemente verbunden sind, sieht man daran, ob der Schlüssel zu der Liste ihrer Subcodeelemente derselbe ist. Diese Phase nennt man *Linking Phase* [PDA]. Auf der Ebene 0 unterscheidet sich die Zuordnung dergestalt, dass zwei Pixel demselben Codeelement zugeordnet werden sobald sie eine bestimmte Ähnlichkeit der Farbe vorweisen. In der darauffolgenden *Splitting Phase* wird das Verfahren um die schon zuvor angesprochene globale Sicht erweitert. Bevor zwei verbundene Subcodeelemente verbunden werden, wird die Ähnlichkeit ihres gemittelten Farbwertes verglichen. Sie müssen deshalb durch diese Methode getrennt werden, da sie auf der Ebene zuvor eine gemeinsame Region besitzen. Die Unterregionen dieser gemeinsamen Region werden derjenigen zugeordnet, welche an die eigene Farbe näher herankommt. Um tiefer in den Ebenen zu splitten, geht der Algorithmus rekursiv weiter in den Subinseln vor. Eine Darstellung der Inselstrukturen des CSC ist in Bild 2.1 zu sehen. Jedes Element besitzt ein sogenanntes Wurzelement⁴. Als Ergebnis der Segmentierung resultieren zwei Bilder: ein Regionenbild, in dem jede Region durch ihren gemittelten Farbwert repräsentiert wird und

³*top level island*

⁴*root element*

ein Labelbild, in dem jedes Objekt von einem andere Grauwert repräsentiert wird (16-Bit-Grauwertbild). Dieses Kommandozeilen-Tool wurde so angepasst, dass die Funktionalität des Systems voll genutzt werden kann.

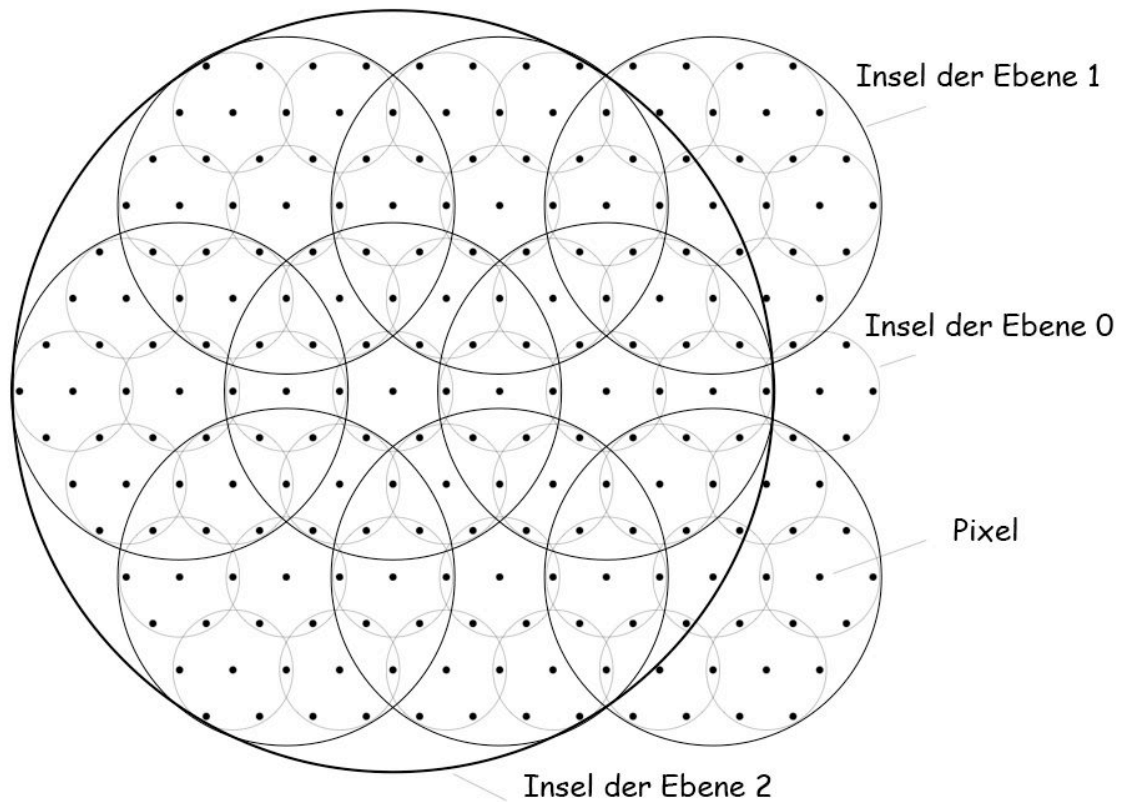


Bild 2.1: Inselstruktur des Color Structure Code

2.2 Kunst und Medien

2.2.1 Medienkunst

Im Zusammenhang mit dieser Studienarbeit wird der Blick besonders auf die Medienkunst gerichtet. Dies liegt nahe, da die Grundlage für Arbeiten dieser Kunstrichtung darin besteht, computergestützte Verfahren zu verwenden und digitale Daten zu verarbeiten. Die

Medienkunst beschäftigt sich hauptsächlich mit den neuen Medien. Sie vereinbart Bild, Klang, Code, Sprache und Text⁵. Das älteste Medium dieser neuen Medien ist die Fotografie. Sie hat ihre zentrale Stellung nicht verloren und bildet auch heute noch einen wichtigen Gesichtspunkt der zeitgenössischen Kunst. Im Speziellen spricht man in dieser Kunstrichtung von Videokunst, digitaler Kunst, Computerkunst und elektronischer Kunst. Verbreitet wird sie zum Beispiel durch Film, Video, Internet oder Mobiltelefonie. Der augenscheinlichste Aspekt dieser Richtung ist das Verstehen der Medien als Träger und Vermittler der Kunst. Ferner sind Computernetze, Programmiersprachen oder Mathematik abstrakte Träger der Kunst. Die Medienkunst arbeitet eng zusammen mit elektronischer Musik und der Entwicklung fotografischer Techniken. Als Gegenbewegung der technologisch hochwertigen Techniken haben sich Low-Tech-Strategien hervorgetan, die das Hauptaugenmerk auf Einfachheit legen. Außerdem versuchte man mit der Verwendung von Open-Source-Programmierung dem Microsoft-Monopol entgegenzuwirken.

Anfang des 20. Jahrhunderts entwarfen einige Künstler Farbenklaviere oder versuchten es zumindest. Diese Versuche mit der Kombination von Ton und Farbe und deren sinnvollem Zusammenspiel ein Gesamtkunstwerk zu schaffen gehen zurück bis ins 17. Jahrhundert, wo sich unter anderem Louis-Bertrand Castel⁶ und später um 1910 Alexander Skrjabin⁷ mit dem so genannten „Augenklavier“ beschäftigten. Kandinsky versuchte 1912 mit seiner Bühnenkomposition „Der Gelbe Klang“, Musik mit Farbe zu verbinden. Im Bereich „Szenografie“ gab es die „Bilder einer Ausstellung“ der Gruppe *Bildklang* im Jahre 2005⁸. Zum Klavierzyklus von Modest Mussorgski erlebt der Zuschauer hier live die Bühnenkomposition „Bilder einer Ausstellung“ von Wassily Kandinsky.

Die Werke im Bereich der Medienkunst sind prozessbasierte künstlerische Arbeiten, die den Betrachter zur Erstellung des Werkes miteinbeziehen wollen. Sie rufen zu (inter)aktiver Gestaltung auf, wobei der Interaktionsprozess intuitiv sein soll. Es gibt weniger abgeschlossene Werke wie es bei Gesamtkunstwerken der Fall ist, sondern mehr offene Prozesse. Medienkunstwerke können in Verbindung mit der Forschung auch explorativ sein.

⁵http://de.wikipedia.org/wiki/Neue_Medien [Stand: Juni 2007]

⁶http://de.wikipedia.org/wiki/Louis-Bertrand_Castel [Stand: Juni 2007]

⁷http://de.wikipedia.org/wiki/Alexander_Nikolajewitsch_Skrjabin [Stand: Juni 2007]

⁸<http://www.bildklang.de/> [Stand: Februar 2007]

Ein Projekt im Bereich „Interaktion“ wurde von Janoschek/Schlömer 2002 vorgestellt und trug den Namen „Actionist Respoke“⁹. Hier wurden Soundcollagen mit im Computer erzeugten visuellen Elementen verbunden. Am Ende kommt ein interaktiver Musikclip mit Sounds der Band *Mouse on Mars* heraus, bei dem die Position der Soundbruchstücke die Lautstärke und Balance regeln.

Bei den Installationen stellten Robert und Tellinga¹⁰ 2005 die „Three Spaces“ vor, bei der Visuals und Sound integriert wurden. Auf die umgebenden Wände einer Ausstellung wurde durch „musikalisierte“ Morsecodes eine computergenerierte Wandverkleidung projiziert.

Selbst im Pädagogikbereich nimmt die Medienkunst eine immer größere Rolle ein. Am Beispiel einer Berliner Schule gab es 2004 unter dem Titel „Kinder machen Kunst mit Medien“ ein Projekt in welchem Goethes „Zauberlehrling“ als Stop-Motion-Film mit Sprechgesang vorbereitet und aufgeführt wurde¹¹. Ziel des Einsatzes der Medienkunst war es, den Kindern klassische Texte leichter nahezubringen und sie in den Umgang mit neuen Medien zielorientiert einzuführen. Generell kann man als Ziele und Anforderungen der Medienkunst erwähnen, dass sie generativ, interaktiv, intuitiv und performativ sein sollte. Zudem soll sie außerdem noch technisch brilliant, klar, einfach, mit philosophischen und pädagogischen Zügen, ästhetisch, spielerisch im Umgang mit Technik und ohne Grenzen sein (alles soll frei sein und offen). Medienkunstwerke haben den Anspruch Assoziationen hervorzurufen und „Denkbilder zu erzeugen“¹².

Das im Laufe der Studienarbeit entwickelte audiovisuelle System ist als Mittel zur Kunst zu verstehen, nicht als Kunst selbst. Als Grundlage des *BildKlanges* dient ein Foto, kein Bewegtbild. Eine interessante Frage, die an diese Studienarbeit gestellt wurde, ist die, ob ein Bild mit komplexer, anspruchsvoller Komposition ein ebenso komplexes und anspruchsvolles, musikalisches Spiel darauf entstehen lässt und ob umgekehrt ein einfacher Bildaufbau ein begrenztes, einfaches tonales Spektrum bereitstellt. Einen Schritt in Rich-

⁹<http://www.stromgasse.de/actionist/> [Stand: Februar 2007]

¹⁰<http://martijntellinga.nl> [Stand: Juni 2007]

¹¹<http://netzspannung.org/learning/kids-arts-media/sorcerer/> [Stand: Juni 2007]

¹²Jurystimme beim Medienkunstpreis 2005 *denkbilder*, <http://www.swr.de/medienkunstpreis/> [Stand: Juni 2007]

tung einer Antwort geht Kapitel 4.

2.2.2 Farb- und Klangwahrnehmung

Im Bereich der Wahrnehmung wurden viele Schriften verfasst und es sind viele Texte im Internet zu finden. Allerdings kann man sich nicht auf wissenschaftliche Statistiken stützen oder auf die Ergebnisse langfristiger Versuchsreihen. Die Wahrnehmung, gerade im Farb- und Tonbereich, ist so individuell unterschiedlich, dass keine allgemeingültige audiovisuelle Studie existiert, nach der man Töne bestimmten Farben zuordnen kann. Der Mensch nimmt Farbe wahr, nachdem das Licht, das auf Objekte der Umgebung fällt, von diesen abgestrahlt wird und so ins Auge gelangt. Im Gehirn wird diese Information (die elektromagnetische Strahlung) in Nervenimpulse umgewandelt. Es folgen Impulse in den Hirnstrukturen, die zur Empfindung einer „Farbe“ führen¹³. Farbe hängt ab von Beschaffenheit von Oberflächen und deren Helligkeit. Einheitliche statistische Ergebnisse psychologischer Wahrnehmungstests existieren nicht. Im Falle der Wahrnehmung im tonalen Bereich verhält es sich ähnlich. Die Wahrnehmung jedes einzelnen Menschen ist im Bereich des tonalen Hörens derart unterschiedlich ausgeprägt, dass es kein generell akzeptables Schema dafür gibt. Eine Schallwelle oder eine Schwingung (in der Luft) verursacht Druckänderungen im Ohr, die zu neuronalen Signalen führen. Das heißt, im Ohr wird Schallinformation in einen Nervenimpuls umgewandelt und über 30 000 Nervenfasern vermitteln durch elektrische Impulse unzählige Tonhöhenunterschiede und Stärkestufen über den Hörnerv zum Gehirn[Mic97]. Dies wiederum führt zu einem Höreindruck. Die Zuordnung von Farbe und Ton muss willkürlich nach eigenem Belieben erfolgen. Die Entwicklung der Zuordnung in dieser Arbeit ist aufgrund dieser Umstände subjektiv, entspricht allerdings vielleicht der Wahrnehmung vieler anderer oder kommt dieser nahe.

2.2.3 Farbenlehre nach Goethe

Auch wenn sich viele Maler, Forscher und Pädagogen in den vergangenen Jahren mit dem Wirken von Farben beschäftigt haben und es einige Schriften zu diesem Thema gibt, war

¹³<http://www.allpsych.uni-giessen.de/karl/teach/farbe.html> [Stand: Juni 2007]

der Wegweiser in dieser Arbeit in die Richtung der Farbwahrnehmung vor Beginn der Implementierung in der Vorbereitungsphase die Farbenlehre von Goethe [Goe]. Seine Beschreibungen der einzelnen Farbnuancen und deren Wirkung auf den Menschen gelten seit Erscheinen als fundiert und werden vor allem unter Künstlern häufig in diesem Zusammenhang erwähnt. Noch heute können sie als gültig angesehen werden und wurden deshalb als Grundlage gewählt. Im Internet¹⁴ existiert ein so genannter Tonfarbenkreis mit Bezug auf Goethes Farbenkreis, welcher Toneigenschaften den Farben des Farbenkreises zuordnet. Diese Zuordnung geschah allerdings willkürlich und wurde daher nur als Vorschlag betrachtet.

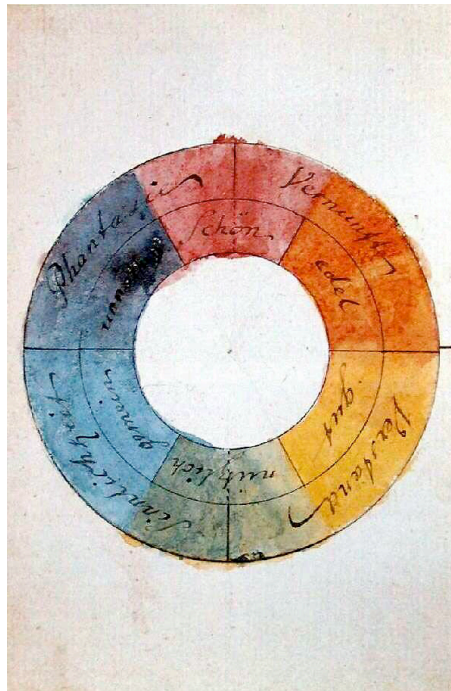


Bild 2.2: Der Farbenkreis nach Goethe zur Symbolisierung des menschlichen Geistes- und Seelenlebens, 1809 [Goe]

¹⁴<http://www.oberton.info/obertonbrevier18.htm> [Stand: Juni 2007]

Kapitel 3

Eigener Ansatz

3.1 Einleitung

Zu Beginn stand die Überlegung, wie die Zuordnung von Bildobjekten und deren Eigenschaften zu Tönen vor sich gehen sollte. Sollte man Farben den Tonhöhen zuordnen oder sollte doch eher die Größe des Objektes die Tonhöhe bestimmen? Das Hauptaugenmerk dieser Arbeit lag von Anfang an auf dem Wohlklang und dem harmonischen Zusammenspiel von Farbe und Klang. Die Zuordnungsarten werden in Kapitel 3.4 beschrieben.

Zum Wirken der Farben und ersten eigenen Assoziationen mit Instrumenten folgen nun einige Stellen aus [Goe]. Im nachfolgenden Zitat beschreibt Goethe die Farbe *Rot*.

„Die Wirkung dieser Farbe ist so einzig wie ihre Natur. Sie gibt einen Eindruck sowohl von Ernst und Würde, als von Huld und Anmut. Jenes leistet sie in ihrem dunklen, verdichteten, dieses in ihrem hellen verdünnten Zustande. Und so kann sich die Würde des Alters und die Liebeswürdigkeit der Jugend in eine Farbe kleiden.“¹

Warmen Farbtönen im Bereich Rot, Braun bis hin zu Orange wurden in dieser Arbeit der Klang eines dunklen, warmen Kontrabasses zugeordnet.

¹dieses und alle folgenden Zitate in Kapitel 3.1 stammen aus Goethes Werk *Zur Farbenlehre*[Goe]

Die klaren, reinen Klänge des Klaviers sollten zu hören sein, wenn helle, freundliche Farben wie Gelb oder Weiß im Bild vorhanden sind. Über Gelb schrieb Goethe Folgendes:

„Es ist die nächste Farbe am Licht [...]. Sie führt in ihrer höchsten Reinheit immer die Natur des Hellen mit sich, und besitzt eine heitere, muntere, sanft reizende Eigenschaft [...]. So ist es der Erfahrung gemäß, daß das Gelbe einen durchaus warmen und behaglichen Eindruck mache.“

Die kühlen Farben im Bereich um die Farbe Blau bestimmte ich dem harten, metallischen Klang von Bläsern. In der Farbenlehre wird dieser Farbton wie folgt beschrieben:

„So wie Gelb immer ein Licht mit sich führt, so kann man sagen, daß Blau immer etwas Dunkles mit sich führe [...]. Das Blaue gibt uns ein Gefühl von Kälte, so wie es uns auch an Schatten erinnert.“

Bei den grünlichen Nuancen entschied ich mich für den Klang der Gitarre, weil sie einen beruhigenden, verträumten Klang hat.

„Unser Auge findet in derselben eine reale Befriedigung. Wenn beide Mutterfarben² sich in der Mischung genau das Gleichgewicht halten, dergestalt, daß keine vor der anderen bemerklich ist, so ruht das Auge und das Gemüt auf diesem Gemischten wie auf einem Einfachen. Man will nicht weiter und man kann nicht weiter.“

Aufgrund der Farbtonverteilung auf einem leicht einteilbaren Farbkreis und der Tatsache, dass in Bezug auf das Farbsehen, dieses Farbraummodell am ehesten der Wahrnehmung des Menschen entspricht, fiel die Wahl auf die Verwendung des HSV-Farbraummodells (H = Hue, S = Saturation, V = Value). In diesem Farbkreis interpolieren alle Farbtöne im Bereich zwischen Rot um 0 Grad, Grün um 120 Grad und Blau um 240 Grad, wie in Bild 3.1 zu sehen ist. Durch eine Unterteilung in Bereiche mit jeweils 30 Grad Öffnung entstehen so 12 Farbsektoren, die anschließend den 12 Tonhöhen zugeordnet werden können.

²gemeint sind Gelb und Blau

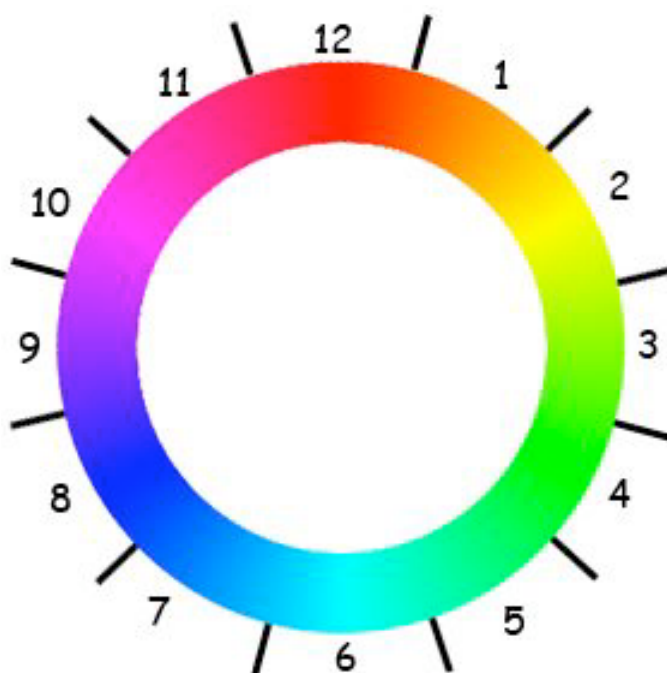


Bild 3.1: Farbkreis eingeteilt in zwölf Sektoren

Die erste Überlegung der Schritte zu *BildKlang* war es, die Tonart des Bildes zu bestimmen. Hinsichtlich der im Bild vorkommenden Farben sollte mit Hilfe eines Tonfarbenkreises (siehe Bild 3.2) das Spektrum bestimmt werden. Die Idee war, das Vorkommen und die Häufigkeit der einzelnen Farben zu bestimmen und daraus eine Tonleiter zu konstruieren. Sollte es im Bild zum Beispiel nur die Farben entsprechend der Töne *cis*, *fis*, *a* und *h* geben, würde daraus die Tonart D-Dur bestimmt (*d-e-fis-g-a-h-cis*). Schon nach einem ersten Gespräch mit Ansprechpartnern des musikwissenschaftlichen Instituts der Universität stellte sich heraus, dass dieser Schritt nicht von Nöten ist, da man beim Spielen auf dem Bild die Tonart nicht heraushören würde. Die Konsequenz dieser Feststellung war es schließlich, Tonarten außer Acht zu lassen.

Als Einteilung des Tonsystems sollte zuerst eine Pentatonik verwendet werden, da der Wohlklang des Ergebnisses als eines der zu erreichenden Ziele galt. Eine Pentatonik besteht aus einer 5tönigen Tonleiter ohne Halbtonschritte mit drei Ganztönen und zwei kleinen Terzen[Mic97]. Verwendung findet sie unter dem Begriff der *Fünftonmusik*. Der Vor-

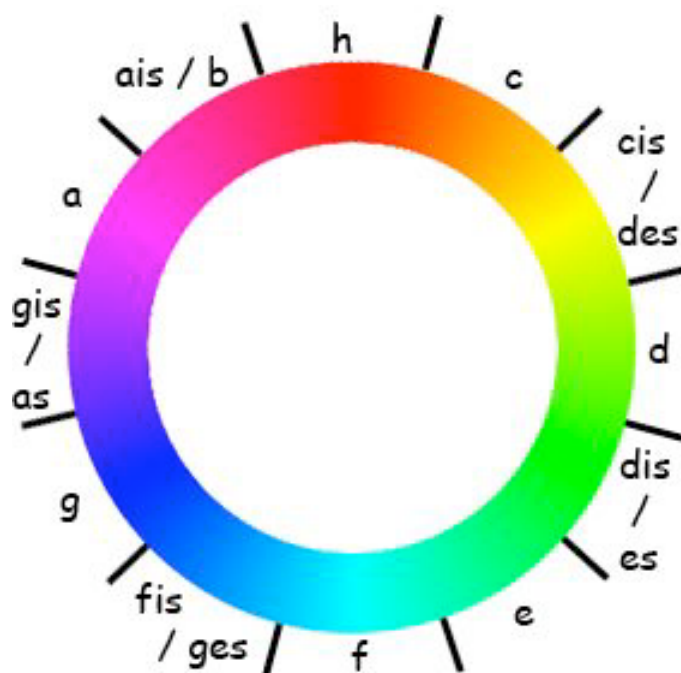


Bild 3.2: Tonfarbenkreis nach eigener Zuordnung

teil ihrer Anwendung liegt darin, dass sie zum Beispiel im Falle einer Quintenpentatonik immer wohl klingt und dem Benutzer nicht viel musikalisches Fingerspitzengefühl abverlangt. Der Nachteil besteht darin, dass durch die Spektrumsbeschränkung kein hoher musikalischer Anspruch möglich ist. Zumal es nicht garantiert ist, dass ein Bild das volle Spektrum, sprich alle fünf Töne der Pentatonik, beinhaltet. Dieser Umstand ermöglicht leider nur ein sehr unabwechslungsreiches Spiel. Da die Musikalität nicht von vorneherein schon beschränkt sein sollte, sah die weitere Arbeit die Verwendung eines chromatischen Systems vor, bestehend aus einer 12tönigen Leiter mit einer Halbtonfolge. Als Zwischenlösung hätte man nur einzelne Akkorde der Tonart einer Diatonik (7tönige Leiter mit fünf Ganz- und zwei Halbtönen) bereitstellen können, so dass in einem Bild nur Akkorde aufbauend auf die Tonika (1.Stufe), die Subdominante (4.Stufe) und die Dominante (5.Stufe) klingen. Doch dieser Gedanke wurde schnell von der Überlegung abgelöst, dass es auch hier nicht immer alle Töne im Bild geben muss und dass die Klangerwartung nicht aufgehen würde. Es werden im System *BildKlang* nun alle Töne bereitgestellt (sofern sie im

Bild vorhanden sind) und können ohne Einschränkung benutzt werden.

3.2 Voraussetzung

Um auf die Eigenschaften von Objekten im Bild zugreifen und eine Zuordnung treffen zu können, ist ein funktionierender Bildsegmentierungsalgorithmus von Nöten. Dieser muss das Eingabebild, egal ob Farb- oder Grauwertbild, lesen, bearbeiten und analysieren können. Außerdem muss eine Tonbibliothek zur Verfügung gestellt werden, aus der Töne verwendet werden können.

3.3 Segmentierung von Farbbildern

Die bereits erfolgreiche Implementierung des `Color Structure Code` stellt die wichtigste Arbeit im Verfahren dar. Das Bild wird eingelesen und die Segmentierung wird gestartet. Die Pixel im Bild werden, wie bereits in Kapitel 2 erwähnt, nach der hierarchischen Inselstruktur mit Codeelementen versehen und mit den Pixeln in der Nachbarschaft in Verbindung gesetzt. Hieraus ergibt sich eine Regionensegmentierung und ein dazugehöriges Regionenbild, das unter dem Namen `region.ppm` gespeichert wird. Dieses Bild besteht aus den durch die Anwendung der CSC-Segmentierung erhaltenen Regionen. Diese werden durch ihre gemittelten Farbwerte farbig dargestellt. Des Weiteren wird ein sogenanntes Labelbild unter dem Namen `mapImage.ppm` gespeichert. Hierbei werden die segmentierten Regionen in einem 16-bit-Graustufenbild anhand eines eindeutigen Grauwertes repräsentiert.

Während der Arbeit an dieser Studienarbeit wurden diese Bilder zur Überprüfung und als Anhaltspunkt verwendet, um zu entscheiden, welche Parameter für ein optimales Ergebnis zu wählen sind und ob der Algorithmus richtig arbeitet. Das Labelbild wird zur erfolgreichen Verwendung von *BildKlang* am Ende nicht mehr benötigt. Das Regionenbild wird in der GUI in Teilen (oder vollständig) über dem Originalbild angezeigt, um dem Spieler einen Anhaltspunkt zu geben, in welchen Bereichen des Bildes sich welche Instrumente spielen lassen. Hierzu in Kapitel 3.6 später mehr.

3.4 Zuordnung von Bildelement zu Ton

Das Bild wird eingelesen und nach der Segmentierung existieren darin für jedes Pixel Informationen über dessen RGB-Farbwert, den HSV-Farbwert, den Farbsektor sowie über die Zugehörigkeit dieses Pixels zu einer Region. All diese Informationen lassen sich anhand der Pixelkoordinaten abfragen.

Im Rahmen dieser Studienarbeit wurden nun zwei relevante Zuordnungsarten erarbeitet:

1. **Zuordnung von zwölf Farbsektoren zu zwölf Tönen und von Objektgröße zu Oktave.** Hierbei erfolgt die Zuordnung nach dem einmal willkürlich festgelegten Tonfarbenkreis aus Bild 3.2. Die Objektgröße spielt bei dieser Zuordnung insofern eine Rolle, als dass sie entscheidet, in welcher Oktave sich der Ton befindet. Das heißt, es gibt im Bildbereich drei Objektgrößen *klein*, *mittel* und *groß*, die übertragen in den Tonbereich für die eingestrichene, zweigestrichene und dreigestrichene Oktave stehen³. Wird nun ein Pixel in einem Objekt angeklickt, überprüft das Programm den Farbsektor und ermittelt eine Tonhöhe zwischen *c* und *h*. Daraufhin wird die Objektgröße ermittelt und so eine der drei Oktaven bestimmt, in welcher der Ton liegen soll.

Als Beispiel wird ein rotes Objekt ausgewählt, dessen Größe als *mittel* zu bezeichnen ist. Nun wissen wir anhand des Tonfarbenkreises, dass der zwölfte Farbsektor (rot) für den Ton *h* steht. Die Größeninformation *mittel* entscheidet, dass dieser Ton in der zweigestrichenen Oktave liegt. Deshalb erklingt beim Klicken dieses Objektes das zweigestrichene *h*.

2. **Zuordnung von Farbgruppen zu Instrumentgruppen und von Objektgröße zu Tönen (innerhalb von drei Oktaven).** Hierbei sind die zwölf Farbsektoren in vier Untergruppen gegliedert. Die grünlichen Farbsektoren bilden eine Gruppe, wie auch die rötlichen Nuancen, die Gelbnuancen und die bläulichen. Diese Einteilung erfolgte willkürlich, bleibt aber innerhalb des Systems konsistent. Nun existieren zu diesen vier Farbnuancengruppen in *BildKlang* vier Instrumente mit jeweils einem Tonrepertoire von 36 Tönen⁴. Es gibt in je drei Oktaven Töne der Gitarre, des Pianos,

³siehe S.66 in [Mic97]

⁴entsprechend drei Oktaven zu je zwölf Tönen

der Trompete und des Basses. Die Zuordnung erfolgte auch hier willkürlich nach eigenem Ermessen, orientiert sich allerdings an den Theorien der Farbwahrnehmung und der eigenen Wahrnehmung.

Für den Zusammenhang von Farbe und Instrument gilt innerhalb dieser Studienarbeit folgende Zuordnung:

Gitarre: = Menge der Farben mit grünlicher Nuance; freundliche, ruhige Farben siehe Bild 3.3.



Bild 3.3: Grünnuancen und die Gitarre

Bass: = Menge der Farben mit rötlicher sowie bräunlicher Nuance; warme, dunkle Farben, siehe 3.4.



Bild 3.4: Rotnuancen und der Bass

Piano: = Menge der Farben mit gelblicher sowie weißlicher Nuance; fröhliche, helle Farben, siehe 3.5.



Bild 3.5: Gelbnuancen und das Piano

Trompete:= Menge der Farben mit bläulicher Nuance; kühle, metallische Farben, siehe 3.6.



Bild 3.6: Blaunuancen und die Trompete

Die Farbsektoren (siehe Bild 3.1) verteilen sich wie folgt auf die Instrumente:

$$M_{Gitarre} := \{3, 4, 5, 6, 9\}$$

$$M_{Bass} := \{0, 10, 11, 12\}$$

$$M_{Piano} := \{1, 2\}$$

$$M_{Trompete} := \{7, 8\}$$

Die Zuordnung erfolgt dergestalt, dass ein Objekt angeklickt wird und erneut zuerst der Farbsektor ermittelt wird. Dieser entscheidet über das Instrument. Ist dies festgelegt, wird die Objektgröße erfragt. In diesem Zuordnungsverfahren gibt es 36 Größeneinteilungen. Jeder Größe dieser Einteilung ist ein Ton innerhalb von drei

Oktaven zugeordnet. Es gibt also zwölf *kleinere* Objektbereiche, zwölf *mittlere* Objektbereiche und zwölf *große* Objektbereiche.

Als Beispiel wird ein Objekt ausgewählt, das in den Bereich der kleineren Objekte gehört und in den bläulichen Farbsektor fällt. Der achte Farbsektor im Bildbereich gehört zum Instrumentbereich der Trompete. Wird nun also ein blaues Objekt ausgewählt, erklingt die Trompete. Die Objektgröße beträgt circa 500 Pixel. Diese Information führt zur Zuordnung dieser Region zum Ton *d* der eingestrichenen Oktave. Aus diesem Grund erklingt beim Klicken eines Pixels dieses Objektes ein eingestrichenes *d* der Trompete.

Beobachtungen und Ergebnisse dieser beiden Zuordnungsarten finden sich in Kapitel 4.

3.5 Technisches Design

3.5.1 Übersicht

Als Sprache zur Erzeugung und Darstellung der GUI, Anzeige der Bilder darin und generell als Verwaltungsstruktur ist die Hauptklasse `AHController.mm` in Objective-C⁵ geschrieben. Die Klassen, welche mit dem Controller kommunizieren und mit der Visualisierung zu tun haben, sind in Objective-C programmiert. Alle Klassen, die sich mit der Bildverarbeitung und Audioverwendung befassen, sind in C++ geschrieben, da sich für diese Aufgabengebiete die Methoden und Funktionen von Objective-C als sehr langsam herausstellten.

3.5.2 Module

Das Bild 3.7 zeigt die Klassen in übersichtlicher Form eines Klassendiagrammes. Es folgen die Auflistung aller selbst implementierten Klassen sowie eine kurze Beschreibung ihrer Funktionalität.

⁵aufgrund der zusätzlichen Verwendung von C++-Aufrufen Objective-C++ genannt

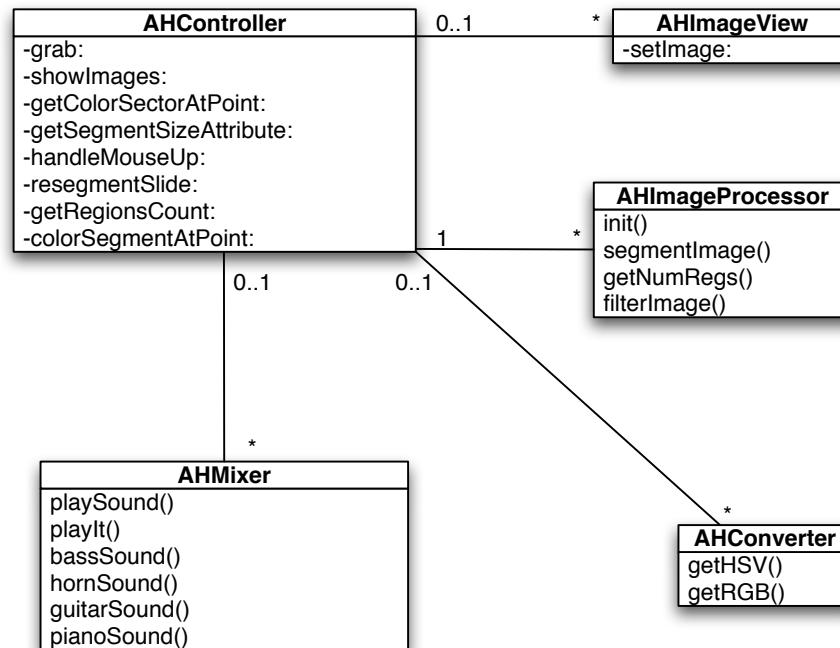


Bild 3.7: Klassendiagramm des Verfahrens. **AHController** ist die Hauptklasse, die alle Klassen aufruft. Die Klasse **AHMixer** öffnet die Audioeinheit und spielt die Töne ab. Mausklicks werden in der Klasse **AHImageView** abgefangen. Der **AHImageProcessor** startet die Segmentierung. In **AHConverter** können gegebenenfalls die Farbräume umgewandelt werden.

AHController

Klassenname **AHController**

Dateiname **AHController.mm**

Beschreibung Diese Klasse stellt die Verwaltungsstruktur der Anwendung bereit. Das Bild wird hier eingelesen und in der GUI angezeigt. Hier findet die Kommunikation sowohl mit dem Benutzer als auch mit den anderen Modulen statt. Von hier aus wird die Farbsegmentierung des CSC gestartet, die Ergebnisse werden an den **AHMixer**

weitergeleitet, wo Sounds erzeugt werden. Es handelt sich um eine Objective-C++-Klasse.

AHImageProcessor

Klassenname AHImageProcessor

Dateiname AHImageProcessor.cpp

Beschreibung Diese Klasse startet die Segmentierung des Bildes mit dem CSC. Sie bildet die Schnittstelle zwischen den eigenen Modulen und den verwendeten CSC-Klassen. Es gibt zusätzlich eine Funktion zur Vorverarbeitung des Bildes, die einen Gaussfilter anwendet. Sie ist optional und zur eigentlichen Segmentierung nicht notwendig.

AHMixer

Klassenname AHMixer

Dateiname AHMixer.cpp

Beschreibung Diese Klasse bindet die Bibliothek von SDL ein. Sie initialisiert den Audioplayer sowie den Audiomixer. Sie wird vom AHController aufgerufen. Hier werden die Informationen über das Segment im Bild verwendet, das vom Benutzer ausgewählt wurde und mit dem entsprechenden Soundfile verknüpft.

AHImageView

Klassenname AHImageView

Dateiname AHImageView.m

Beschreibung In dieser reinen Objective-C-Klasse werden die Koordinaten des Mausclicks abgefangen. Hier erfolgt die Umrechnung der Bildschirmkoordinaten in Bildkoordinaten. Das Eingabebild wird hier entsprechend der GUI skaliert. Es werden zwei Notifications gesendet, auf die im `AHController` reagiert wird.

AHConverter

Klassenname `AHConverter`

Dateiname `AHConverter.cpp`

Beschreibung In dieser Klasse werden die RGB-Werte in HSV-Werte umgerechnet und umgekehrt. Außerdem wird darin der Farbsektor eines HSV-/RGB-Wertes ermittelt. Diese Klasse nimmt Anfragen vom `AHController` entgegen und gibt Informationen an ihn zurück. Der Rückgabewert kann aus dem Wert für den Farbton, die Sättigung, dem Grauwert oder dem Farbsektor einer Pixelfarbe bestehen.

3.6 Die GUI

Der Bildanzeige-/Spielebereich nimmt in der GUI der Anwendung die Hauptfläche ein. Wie in Bild 3.8 abgebildet, befindet sich am unteren Rand des Fensters ein Textfeld zur Eingabe des Bildnamens und einem Button zur Anzeige dieses Bildes. Zusätzlich zu dieser Grundfunktionalität befinden sich am rechten Rand der GUI vier Checkboxes, beschriftet mit *Guitar*, *Bass*, *Horns* und *Piano*. Der Benutzer hat hierdurch die Möglichkeit, sich die Bereiche im Bild anzeigen zu lassen, in denen sich Töne des entsprechenden Instrumentes befinden. Diese Bereiche werden dann nach Auswählen des Kästchens und erneutem Drücken des *Grab Image!*-Buttons im Bild farbig angezeigt (siehe Bild 3.9). Zu sehen ist an dieser Stelle das bereits zuvor erwähnte Regionenbild. Der Benutzer hat jederzeit die Möglichkeit ein anderes Bild auszuwählen, so wie ihm auch die Wahl frei steht, welche Instrumentgruppen er sich anzeigen lassen möchte⁶.

⁶zur Aktualisierung dient der *Grab Image!*-Button

3.7 Eingabegeräte

Als Eingabemedium der *BildKlang*-Interaktion dient zum Zeitpunkt der Abnahme die Maus. Der Benutzer klickt auf das Objekt im Bild, das er zum Klingen bringen möchte und löst damit den entsprechenden Ton dieser Region aus. Überlegungen zu weiteren Eingabegeräten finden sich im Ausblick von Kapitel 5.

3.8 Die Audiobibliothek

Um die Sounddateien zu verwalten und in das *BildKlang*-System einzubringen, benötigt man eine Audiobibliothek, die sich am Besten leicht und unkompliziert in das bestehende System eingliedern lässt. Hierfür kamen zwei Frameworks in Frage: *Marsyas* (Music Analysis, Retrieval and Synthesis for Audio)⁷ und *Simple DirectMedia Layer*⁸, kurz: *SDL*.

Marsyas ist eine kleine Audiobibliothek und in ihrer Verwendung noch wenig verbreitet. *SDL* fand bereits in anderen Arbeiten der Arbeitsgruppe Aktives Sehen Verwendung⁹ und überzeugt durch seine schnelle Einbindung und einfache Handhabung. Wenn man schnell und einfach nur Audiodateien abspielen möchte, kommt man mit der ausführlichen Dokumentation von *SDL* eher ans Ziel. Daher fiel die Wahl zur Verwendung in der vorliegenden Arbeit auf diese komplexere Multimedia-Bibliothek.

3.8.1 Simple DirectMedia Layer

SDL wurde als Multimedia-Software speziell für Spiele und Emulatoren entwickelt und ist in C geschrieben, unterstützt allerdings über Wrapperklassen sämtliche andere Programmiersprachen (von Ada, C++ über Objective-C, Java bis hin zu Smalltalk) und ist plattformübergreifend. Die Einbindung der Audiodateien ist sehr einfach und mit weni-

⁷<http://opihi.cs.uvic.ca/marsyas/> [Stand: Juni 2007]

⁸<http://www.libsdl.org/> [Stand: Juni 2007]

⁹“cam2dance“, die “Virtuelle Tanzmatte“, die im Sommer 2006 auf der Nintendo-Ausstellung auf der Festung Ehrenbreitstein in Koblenz zu sehen war

gen Zeilen Code lässt sich SDL sowie der für Audio notwendige SDL-Mixer initialisieren und die Soundfunktionalität steht bereit. Der Mixer verwaltet die Audiohardware und wenige Befehle führen zu einem schnellen Klangergebnis. Für die Arbeit an dieser Studienarbeit wurde nur von der Funktionalität der Audioausgabe Gebrauch gemacht. SDL bietet darüber hinaus noch andere Funktionalitäten, die für die Arbeit an der vorliegenden Studienarbeit allerdings nicht von Belang waren und keine Verwendung fanden.

Die wichtigsten Funktionen sind:

```
SDL_Init(SDL_INIT_AUDIO);
```

`SDL_Init()` lädt die Bibliothek dynamisch und initialisiert diese. `SDL_INIT_AUDIO` ist ein Flag, das übergeben wird, um den Audioteil zu aktivieren¹⁰.

```
Mix_OpenAudio(audio_rate, audio_format, audio_channels, audio_buffers);
```

`Mix_OpenAudio` initialisiert die Mixer API mit den Parametern `audio_rate` (int *frequency*), `audio_format` (Uint16 *format*), `audio_channels` (int *channels*) und `audio_buffers` (int *chunksize*)¹¹.

Zur Verwendung von mp3-Dateien müssen noch folgende Funktionen aufgerufen werden:

```
Mix_LoadMUS("music.mp3");  
.  
.  
Mix_PlayMUS(music, -1);
```

`Mix_LoadMUS()` lädt die Musikdatei, die als Parameter übergeben wird und `Mix_PlayMUS()` spielt diese (welche als Variable *music* übergeben wurde) in *loops* (wobei *-1* eine unendliche Wiederholung bedeutet).

¹⁰Dokumentation zu SDL: <http://www.libsdl.org/intro.de/toc.html>

¹¹Dokumentation zu SDL_mixer: http://jcatki.no-ip.org/SDL_mixer/

```
Mix_FreeMusic("music.mp3");  
.  
.  
Mix_CloseAudio();
```

Schließlich wird die Musikdatei durch `Mix_FreeMusic()` nach ihrer Verwendung wieder freigegeben und die Audioeinheit mit `Mix_CloseAudio()` geschlossen.

3.8.2 Klangbibliothek

Generiert wurden jeweils drei Oktaven Töne der Instrumente Piano, Gitarre, Bass und Trompete mit der Software *Garageband 3* von Apple¹². Entstanden ist so eine Klangbibliothek mit 144 Dateien im mp3-Format, die zur Verwendung im System bereitstehen.

¹²<http://www.apple.com/ilife/garageband/> [Stand: Juni 2007]

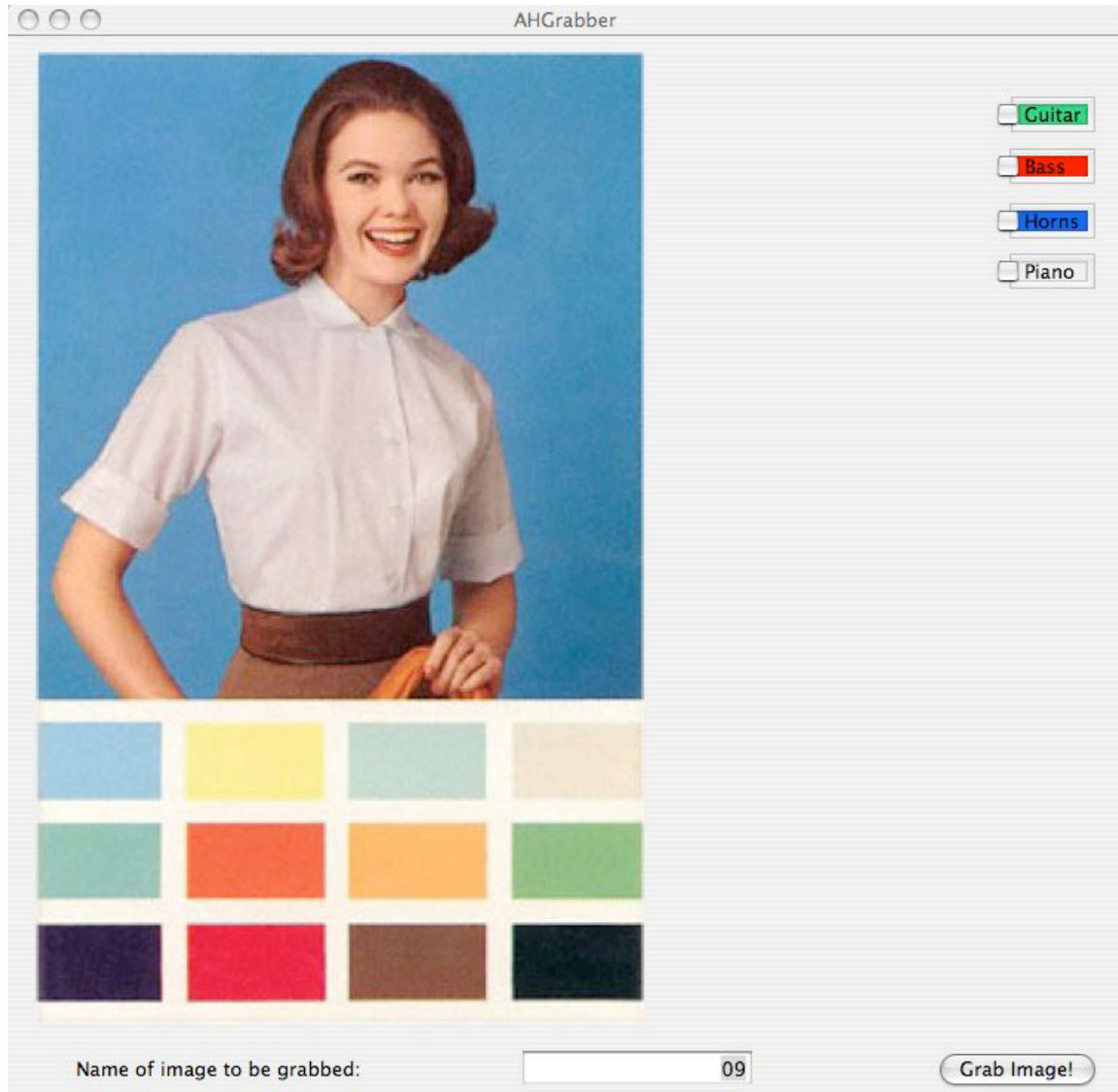


Bild 3.8: Die GUI ohne Anzeige der Bereiche einzelner Instrumente

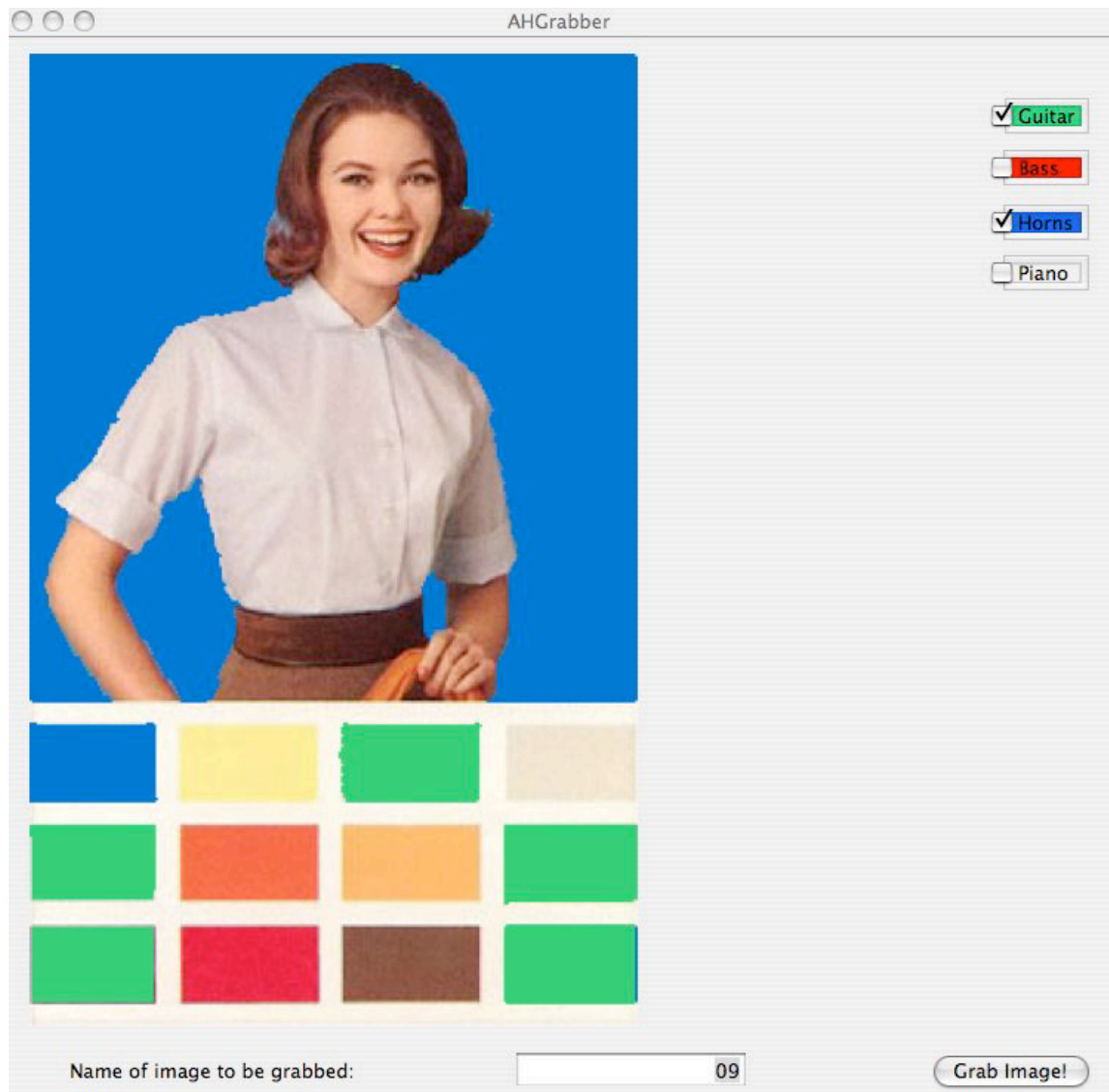


Bild 3.9: Die GUI mit zusätzlicher Anzeige der Bereiche mit Tönen der Gitarre und der Trompete

Kapitel 4

Entwicklung und Erkenntnisse

4.1 Entwicklung

Die in Kapitel 2 und Kapitel 3 eingeführten Algorithmen, Zuordnungsarten und der Audioteil wurden im Rahmen dieser Studienarbeit zu einem Gesamtsystem zusammengeführt und auf einer Reihe von Farbbildern angewandt. Die Idee zu dieser Arbeit war neu, daher gab es keine Vorlagen. Aus diesem Grund war die Arbeit an *BildKlang* ein sich ständig verändernder Prozess. Dieses Kapitel soll einen Einblick bieten und Ergebnisse darstellen.

4.2 Spielaufbau

Der Benutzer setzt sich vor den Computer und entscheidet sich zuerst für ein Bild auf dem er spielen möchte. Die aus der GUI wählbaren Bilder finden sich im Kapitel A. Er gibt den Namen des Bildes im Textfeld der GUI ein und betätigt den *Grab Image!* - Button. Das Bild wird nun geladen und nach wenigen Sekunden angezeigt. Die Schritte, die in der Zeit passieren sind die folgenden: Das Bild wird segmentiert und die Regionen darin werden mit Tonhöhen und verschiedenen Instrumenten verknüpft, wie zuvor in Kapitel 3.4 beschrieben. Das Bild ist nun, sobald es am Bildschirm angezeigt wird, bespielbar. Der Benutzer klickt auf verschiedene Stellen im Bildbereich und löst dadurch einzelne Töne

aus. Er findet sich in den ersten Minuten zurecht. Er exploriert, wo im Bild sich welche Klänge befinden und lernt, welche Farben und Größen welche Klänge erzeugen. Nach wenigen Minuten sollte er in der Lage sein, Töne nach seinen Erwartungen auszulösen und eine kurze Melodie zu spielen (sofern das Tonspektrum dies zulässt). Vor Beginn des Versuches bekam der Testspieler die Aufgabe, auf dem angezeigten Bild die Töne zur Melodie des Kinderliedes „Alle meine Entchen“ zu finden und diese kurze Melodie spielen.¹

4.3 Erkenntnisse: Segmentierung von Farbbildern

Es folgen nun Bilder, die während der verschiedenen Segmentierungsphasen entstanden sind. Sie zeigen zum einen das Originalbild (siehe Bild 4.1), das der Benutzer vor sich auf dem Bildschirm sieht und auf dem er (vermeintlich) spielt². Zum Anderen zeigen sie die Segmentierungsergebnisse mit unterschiedlichen Parametern. Verändert wurde hierbei nur der Wert t . Dieser Wert, der bei der Anwendung des CSC-Algorithmus zu Beginn der Segmentierung festgelegt wird, beschreibt den minimalen euklidischen Abstand zwischen zwei ähnlichen RGB-Farbvektoren. Bild 4.2 zeigt die Segmentierungsergebnisse nach einer Verarbeitung mit Standardparametern. Bild 4.3 veranschaulicht das Ergebnis nach Eingabe einer niedrigen Schwelle und Bild 4.4 zeigt das Ergebnis einer Segmentierung mit einem hohen Wert für t .

4.3.1 Verifikation

Die Wahl geeigneter Parameter ist für ein gutes Segmentierungsergebnis und somit für ein optimales Spielerlebnis sehr wichtig. Als Anhaltspunkt zur Anpassung der Parameter diente die Anzahl der ähnlichen Regionen im Bild. Das Problem hierbei ist, dass dieser

¹zu bemerken ist hier, dass es zum Zeitpunkt der Abnahme zu diesem Spielaufbau keine großangelegten, einheitlichen Testreihen gab. Die Versuche fanden während des Prozesses mit verschiedenen Testpersonen unter unterschiedlichen Systemvoraussetzungen (d.h. verschiedenen Entwicklungsstadien von *BildKlang*) statt. Es wurde ein Testszenario herausgegriffen, das aufgrund seiner Einfachheit die konstantesten Ergebnisse lieferte.

²tatsächlich bilden allerdings die Informationen des Regionenbildes die Grundlage für sein Spiel.



Bild 4.1: Beispiel eines Eingabebildes

Wert dem System erst nach der Segmentierung bekannt ist. Die Werte können von Bild zu Bild stark variieren. Dieser Umstand macht eine Verallgemeinerung der Bilder und einen generalisierten Startwert für t , der für alle Bilder gut ist, nicht möglich. Die Vorverarbeitung und Analyse des Bildes ist vor dem Spielen unerlässlich. Eine Automatisierung hierfür existiert noch nicht. Vorverarbeitung bedeutete während der ersten Experimente, die Anwendung des Segmentierungsalgorithmus und die anschließende manuelle Analyse der Ergebnisbilder. Dazu wurde die GUI dahingehend umgestaltet, dass zum Originalbild sowohl das entstandene Regionenbild als auch das Labelbild angezeigt wurden (siehe Bild 4.5). Desweiteren wurde ein Slider³ eingebaut, mit dem die Schwelle t manuell verändert

³der so genannte *resegmentSlider*.

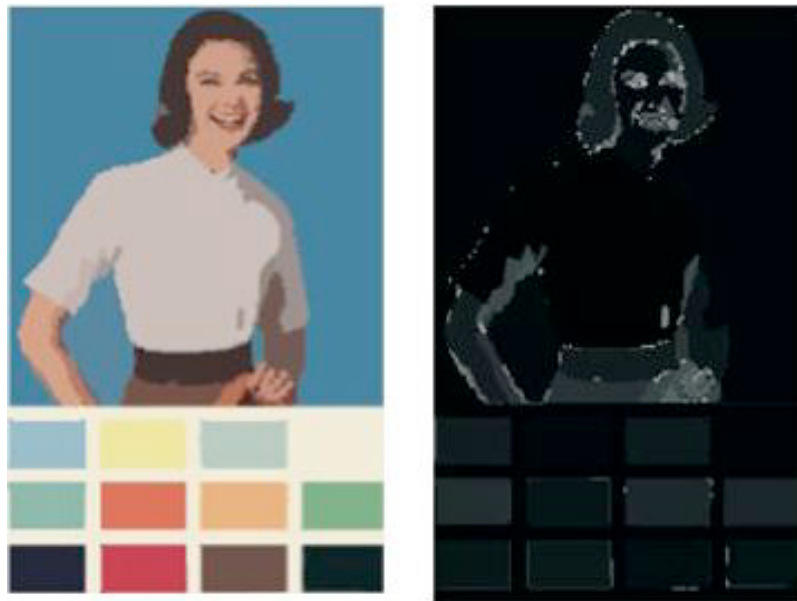


Bild 4.2: Segmentierung mit Standardparametern ($t = 50$). Das linke Bild zeigt das entstandene Regionenbild (mit 240 Regionen). Es ist die Grundlage, auf dem der Benutzer tatsächlich spielt. Im rechten sind die Label zu sehen.

werden konnte. Es stellte sich heraus, dass für verschieden gestaltete Bilder extrem unterschiedliche Schwellen gesetzt werden müssen. Bei Bildern mit sehr viel Information und vielen Objekten ist es für ein gut beispielbares Ergebnis wichtig, die Schwelle niedrig zu halten. So ist zum Beispiel bei einem Bild, das nach der Startsegmentierung circa 240 Regionen aufweist (wie bei Beispielbild 4.1), eine Schwelle von $t = 50$ optimal, damit auch noch kleine Objekte im Bild beispielbar bleiben und nicht heraussegmentiert werden. Das Ergebnis der Segmentierung ist als Regionenbild und Labelbild in Abbildung 4.2 zu sehen. Setzt man für das Beispielbild nun die Schwelle für den euklidischen Abstand herunter (auf circa $t = 10$), so erhält man ein schwächer segmentiertes, weniger abstrahiertes Ergebnis, siehe Bild 4.3. Umgekehrt erhält man bei einer hohen Schwelle wie $t = 80$ ein stark abstrahiertes Endbild, das einer starken Segmentierung entspricht und sehr wenige Regionen im Bild als Ergebnis hat. Die Details im Bild gehen verloren und das beispielbare Bild wirkt sehr grob, wie in Bild 4.4 zu beobachten. Nachdem das zu bespielende Bild manuell analysiert und ein individueller Wert für die Schwelle t gefunden wurde, ist ein

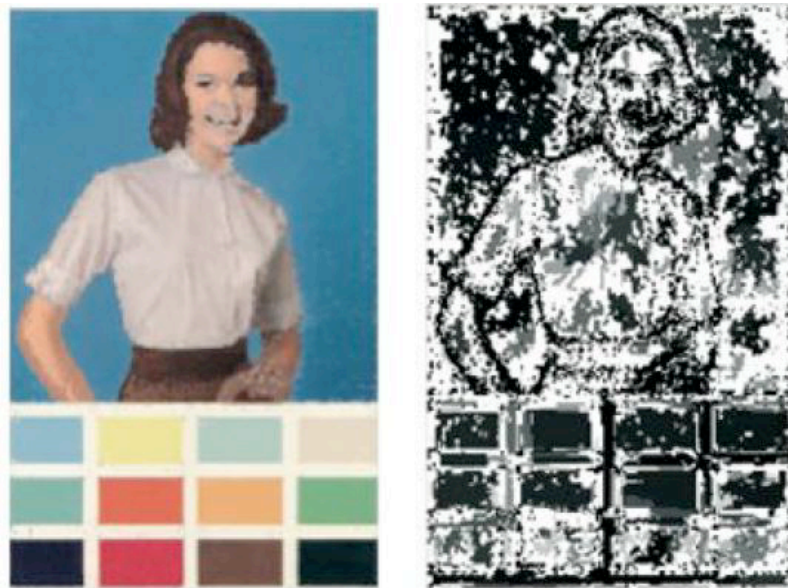


Bild 4.3: Segmentierung mit niedriger Schwelle, $t = 10$. Im resultierenden Regionenbild gibt es 9059 Regionen.

optimales Spielergebnis möglich.

4.4 Erkenntnisse: Zuordnung von Bildelement und Ton

Im Folgenden werden die beiden Zuordnungsarten noch einmal zusammengefasst. Es wird dargelegt in wiefern sie anwendbar sind.

1. Zuordnung von zwölf Farbsektoren zu zwölf Tönen und von Objektgröße zu Oktave

Die erste Zuordnung sah vor, die einzelnen Farbsektoren jeweils einem Ton der Tonleiter zuzuordnen. Die Einteilung erfolgte im Farbkreis ausgehend vom Farbton Gelb und dem Ton c (siehe Bild 3.2). Dem Uhrzeigersinn folgend erhielt jede Farbnuance einen Ton bis hin zu den Rottönen, mit denen h assoziiert wurde. Wurde nun im Bild ein Objekt angeklickt, wurde ausgehend vom Pixel an der Mausposition der Farbwert und daraufhin der Farbsektor bestimmt. Diesem Sektor entsprechend

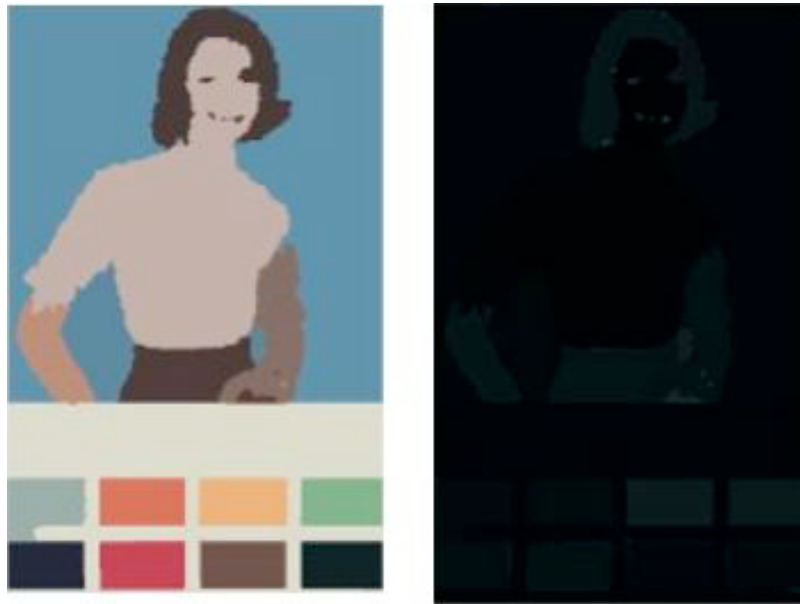


Bild 4.4: Segmentierung mit hoher Schwelle, $t = 80$. Im resultierenden Regionenbild gibt es 54 Regionen.

erklang schließlich der dazugehörige Ton. Zur genaueren Beschreibung siehe Kapitel 3.4

2. Zuordnung von Farbgruppen zu Instrumentgruppen und von Objektgröße zu Tönen (innerhalb von drei Oktaven)

Die zweite Zuordnung beinhaltet sowohl den Farbsektor des ausgewählten Pixels als auch die Größe des angeklickten Objektes. Ausgehend von den Koordinaten des Mausclicks wurde die Anzahl der Pixel mit demselben Farbwert ermittelt und zusammen mit dem Farbsektor wurde ein Instrument und eine Tonhöhe bestimmt. Zuerst erfolgt die Bestimmung des Instrumentes. Zur genaueren Beschreibung siehe auch hier Kapitel 3.4

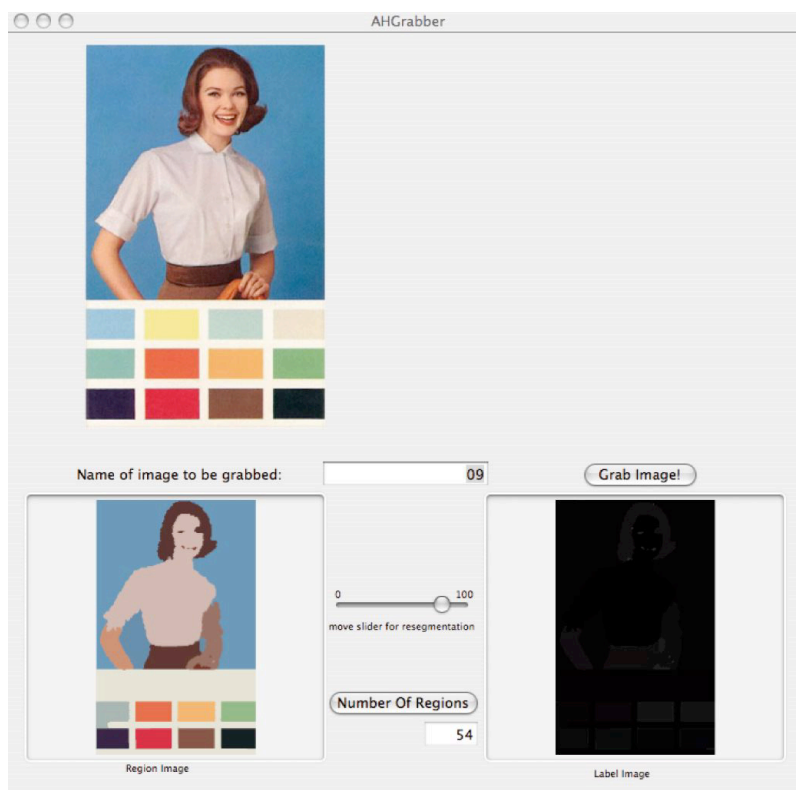


Bild 4.5: Test-GUI zum manuellen Einstellen der Schwelle t und zur anschließenden manuellen Bildanalyse.

4.4.1 Extremfälle

Als Extremfälle stellten sich sehr helle Farben heraus. Ein helles Grün (wie auch jede andere Farbe, die eine große Helligkeit besitzt), das vom Auge als Weiß oder als eine Grauschattierung angenommen wird, bekommt beim Zuordnungsverfahren die Instrumentgruppe *grün* zugewiesen. In diesem Fall erklingt die Gitarre, obwohl der Spieler das Piano für die Instrumentgruppe *weiß* erwartet. Dieses Problem kann abgefangen werden, indem eine zusätzliche Abfrage des Grades der Helligkeit implementiert wird. Diese bestimmt, ob die Farbe schon als Weiß oder noch als die spezifische Farbe anerkannt werden soll. Da es für *BildKlang* wichtig ist, menschlichem Farbsehen nahezukommen, sollte diese Ergänzung vorgenommen werden.

4.4.2 Verifikation

Schon nach den ersten Tests stellte sich heraus, dass die **Zuordnung 1** für das Endsystem nicht funktionierte. Ähnlich der Feststellung zu Beginn von Kapitel 3, dass es nicht möglich ist, Tonarten herauszuhören, konnte hier kein Zusammenhang einzelner Töne und Farbe erkannt werden⁴. Das musikalische Ergebnis war zwar nicht disharmonisch oder klang schlecht, sondern es war innerhalb des Spieles für die Testspieler schwierig, Töne auszulösen, die sie hören wollten. Umgekehrt entsprachen die Töne, die zu hören waren, oft nicht den Erwartungen der Benutzer. Es gestaltete sich für sie schwer, sich im Bild- und Tonbereich zurecht zu finden. Bis zur Melodie von „Alle meine Entchen“ (oder zumindest einem erkennbaren Anfang davon) dauerte es lange, sofern die Spieler überhaupt soweit kamen. Die Zuordnung von Instrumentgruppe zu Farbe und von Objektgröße zu Tonhöhe hingegen, wie in **Zuordnung 2**, ergab von Anfang an ein harmonisches Klangbild. Die Assoziation von Farbe und Klang (genauer gesagt, der Klangfarbe) vollzog sich schnell und es fiel den Spielern leicht, sich zurecht zu finden. Der Zusammenhang *je kleiner das Objekt, desto höher der Ton*⁵ wurde schnell erkannt und intuitiv gut von den Spielern angenommen und umgesetzt. Nach kurzer Zeit waren die meisten Benutzer in der Lage eine Tonleiter zu spielen oder zumindest eine Folge immer höher (oder immer tiefer) klingender Töne. Die Bild-Ton-Zuordnung entsprach den Erwartungen der meisten Spieler. Daher war es für die Benutzer einfacher, Töne auszulösen, die sie hören wollten und die Anfangsmelodie von „Alle meine Entchen“ war deutlich erkennbar.

Bei der farblichen Zuordnung zu den Instrumenten fiel auf, dass sich diesbezüglich klangliche Präferenzen bei den Spielern zeigten und sich diese auf ihr Spiel auswirkten. Nachdem sie spielerisch entdeckt hatten, welche Farbe zu welchem Instrument gehört, mieden einige Testspieler zum Beispiel Objekte der Farbgruppe *blau*, da sie den Klang der Trompete als „nicht angenehm“ empfanden.

⁴Es ist an dieser Stelle schwer zu erklären, wieso das Ergebnis nicht zufriedenstellend war, denn es ist, wie die Wahrnehmung selbst, subjektiv. Allerdings war anhand der Spieler zu beobachten, dass sie ähnlich empfanden.

⁵oder umgekehrt *je größer das Objekt, desto tiefer der Ton*

Kapitel 5

Zusammenfassung

5.1 Fazit

Das Ergebnis dieser Arbeit ist eine interaktive Anwendung, die es ermöglicht, digitale Bilder als eine Art virtuelles Instrument zu benutzen.

Das Einbinden der erfolgreichen Implementierung des CSC erwies sich als Bereicherung und große Zeitersparnis für das System *BildKlang*. Durch die individuelle Anpassung der Segmentierungsschwellen für jedes einzelne Bild kann mithilfe des CSC eine sehr gute Segmentierung erzielt werden, auf der sich solide arbeiten lässt. Es bietet eine gute Ausgangssituation zur weiteren Verknüpfung mit einem Tonspektrum.

Als gute Wahl stellte sich auch die Verwendung von SDL als Audiobibliothek heraus, da damit schnell und einfach hörbare Ergebnisse erzielt werden konnten und die Ansprüche der Studienarbeit an den Audioteil vollständig abgedeckt wurden.

Es wurden zwei Zuordnungsmöglichkeiten herausgearbeitet, wobei sich eine davon als intuitiver herausstellte und vertieft wurde (siehe **Zuordnung 2** in Kapitel 3.4). Im Laufe der Entwicklung wurde deutlich, dass es schwieriger ist, Farbe mit Tönen zu kombinieren und ein erwartungskonformes Ergebnis zu erhalten, als vor Beginn der Arbeit angenommen. Die **Zuordnung 1** aus Kapitel 3.4 stellte sich als unbefriedigend heraus. Töne werden weniger mit einer bestimmten Farbe assoziiert als umgekehrt, bestimmte Farben mit dem Klang eines Instrumentes (und zusätzlich der Größe eines Objektes mit der Höhe des To-

nes).

Daher steht am Ende dieser Studienarbeit die Erkenntnis, dass das Klangergebnis intuitiver ist, wenn Instrumente den Farbgruppen und die Töne der Tonleiter über die Größe der Objekte zugeordnet werden. Es wurde deutlich, dass Farben nicht nach bestimmten Tönen einer Tonleiter klingen. Es ist vielmehr die Klangfarbe eines Instrumentes, die mit Farbe assoziiert wird.

5.2 Ausblick

Die Segmentierung kann in weiteren Arbeiten verfeinert werden. Denkbar und hilfreich wäre eine Phase der Bildvorverarbeitung. Der CSC stellt hierfür Filter bereit, die angewandt werden können, wie unter Anderem *Gauß*, *Sobel*, *Symmetric Nearest Neighbor* und *Kuwahara/Nagao*.

Anstelle des Monitors als Anzeigemedium kann das Bild mit einem Beamer an eine Wand projiziert werden. Als Eingabemedium könnten dann (von einer Kamera) getrackte Stöcke verwendet werden, anhand welcher auf dem Bild wie auf einer Kombination aus Schlagzeug und Xylophon gespielt werden kann. Die Eingabe per getracktem Laserpointer ist bereits in einem anderen Projekt erfolgreich implementiert und könnte in Zusammenarbeit sicherlich schnell eingebunden werden. Bei der Verwendung eines Controllers (oder zweier Controller) der Nintendo Wii¹ kann auf ein Trackingverfahren verzichtet werden, da sie über ihre Bluetooth-Schnittstelle leicht mit dem Computer verbunden und verwendet werden können. Der Controller kann verwendet werden wie eine Maus².

Die Idee des interaktiven Spiels auf einem Bild war schnell umgesetzt. Zum optimalen Klangerlebnis bedarf es allerdings noch einiger Verfeinerungen, sowohl in der Anwendung des CSC, als auch in der Generierung des Klang-Repertoires. Individuelle Parameter zur Segmentierung der einzelnen Bilder (möglicherweise in naher Zukunft automatisiert) und eine Erweiterung der Instrumentgruppen machen aus *BildKlang* ein ausgewogenes audiovisuelles System zur Interaktion mit Bild und Ton.

¹Information siehe unter <http://de.wii.com/>

²Anleitung zur Programmierung der Controller z.B. unter http://carl.kenner.googlepages.com/glovepie_download

Anhang A

Anhang A

Dieser Anhang bezieht sich auf Kapitel 4.2. Es folgt die Sammlung der Bilder, welche der Spieler bei *BildKlang* in der GUI auswählen kann¹.

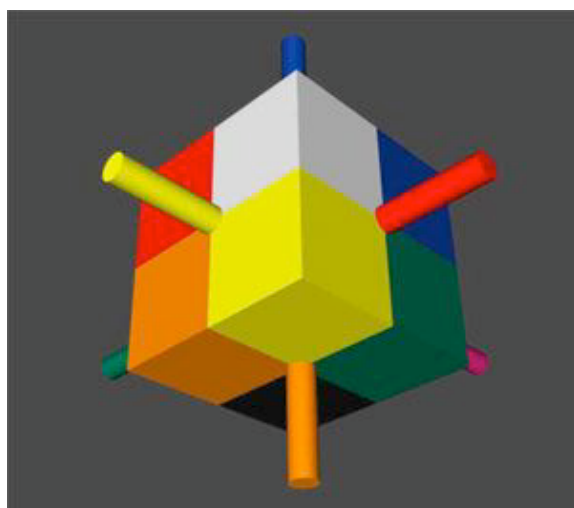


Bild A.1: Bildname in der GUI: 01

¹alle Bilder wurden im Rahmen dieser Studienarbeit mit Photoshop generiert oder stammen aus einer privaten Bildsammlung. Außer 01, 02, und 09, diese stammen aus dem Internet. 07 stammt aus dem Archiv der Arbeitsgruppe Aktivees Sehen.

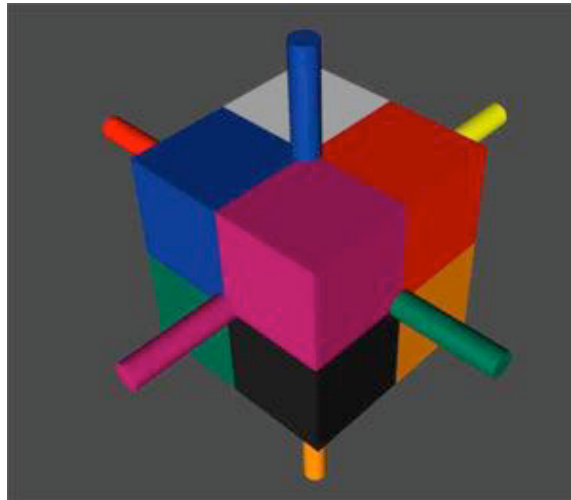


Bild A.2: Bildname in der GUI: 02



Bild A.3: Bildname in der GUI: 03



Bild A.4: Bildname in der GUI: 04



Bild A.5: Bildname in der GUI: 05

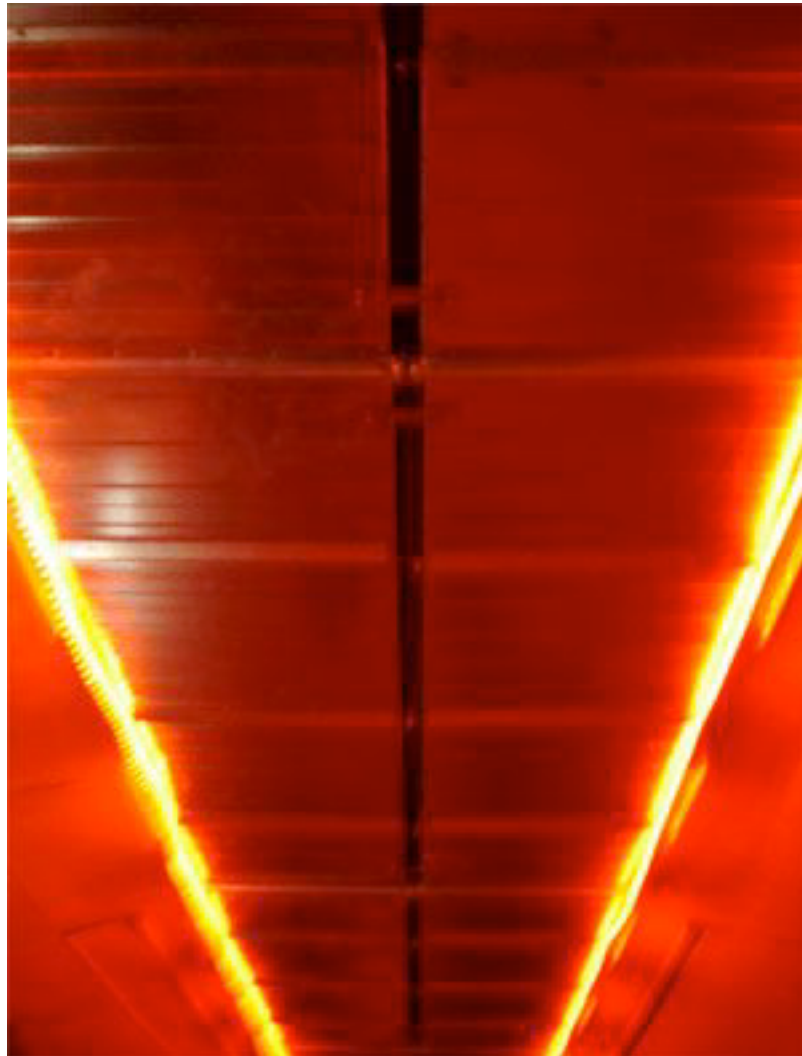


Bild A.6: Bildname in der GUI: 06

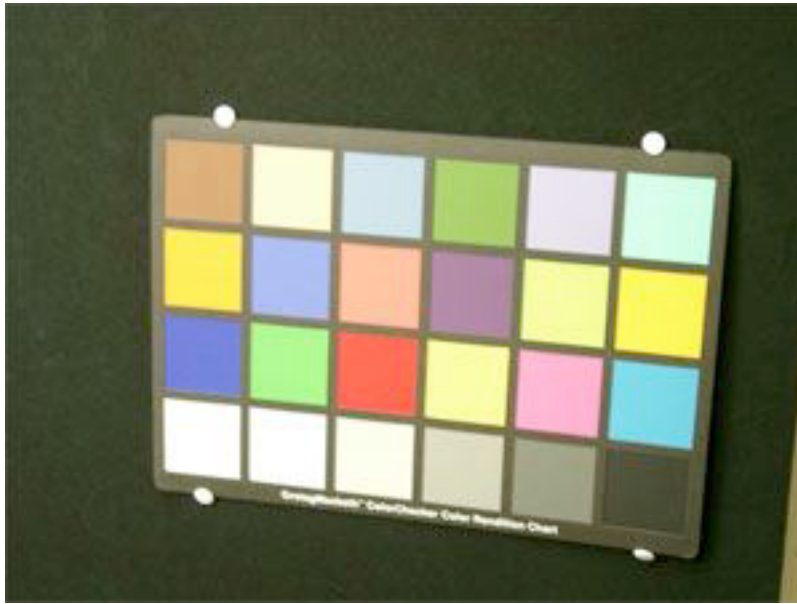


Bild A.7: Bildname in der GUI: 07



Bild A.8: Bildname in der GUI: 08



Bild A.9: Bildname in der GUI: *09*



Bild A.10: Bildname in der GUI: *rot*



Bild A.11: Bildname in der GUI: *gruen*



Bild A.12: Bildname in der GUI: *blau*

Literaturverzeichnis

- [Goe] GOETHE, Johann W.: *Sämtliche Werke 10. Münchner Ausgabe. Zur Farbenlehre.* Hanser, Carl GmbH & Co. KG
- [Mic97] MICHELS, Ulrich: *dtv-Atlas Musik.* Deutscher Taschenbuch Verlag & Co. KG, 1997
- [PDA] PAULUS, Dietrich ; DORKÓ, Gyuri ; AHLRICHS, Ulrike: Color segmentation for scene exploration, S. 13–20
- [PS03] PRIESE, Lutz ; STURM, Patrick: *Introduction to the Color Structure Code and its Implementation.* http://www.uni-koblenz.de/~lb/lb_downloads/download/csc.pdf. Version: 3 2003
- [Reh94] REHRMANN, Volker: *Stabile, echtzeitfähige Farbbildauswertung.* Fölbach Verlag, Koblenz, Universität Koblenz-Landau, Diss., 1994