

HSR HOCHSCHULE FÜR TECHNIK  
RAPPERSWIL

BACHELORARBEIT

---

# Tracking object recognition in the human brain

---

*Autoren:*

Jonathan STOLZ

Thomas JUTZI

*Betreuer:*

Prof. Dr. Ruedi STOOP

*Experte:*

Dr. Norbert STOOP

21. Dezember 2012

# Abstract

Für künstliche kognitive Systeme, wie sie etwa in der Robotik benötigt werden, stellt die visuelle Wahrnehmung von Objekten ein immer noch schlecht gelöstes Problem dar. Der Mensch hingegen kann Objekte sogar unter schwierigsten Verhältnissen korrekt erkennen.

In einer vielversprechenden Arbeit der ETH Zürich wurde in einem Pilot-Projekt die Natur der menschlichen visuellen Wahrnehmung mittels eines Eye-Trackers untersucht. Diese Pilotarbeit war aber auf einem sehr kleinen Untersuchungssample aufgebaut und die verwendeten Methoden und erzielten Ergebnisse waren unzureichend dokumentiert.

In unserer Arbeit ging es darum, 1) die Ergebnisse der Pilotarbeit zu verifizieren, indem die benötigten datenerfassenden und datenverarbeitenden Verfahren geeignet aufbereitet, teilweise neu entwickelt und für die weitere Verwendung geeignet abgelegt werden sollten, 2) die gemachten Experimente statistisch abzusichern, und 3) die gewonnene Basis an Datenmaterial dafür zu verwenden, die bestehenden Wahrnehmungsmodelle zu überprüfen und falls nötig weiter zu entwickeln.

Wir haben entsprechend in unserer Arbeit die verwendeten Programme der Pilotarbeit, so weit sie uns zugänglich waren, kritisch evaluiert, sie verbessert oder durch unsere eigenen Versionen ersetzt und alles gebrauchsfertig abgelegt. Die Methode der Datenerhebung wurde verbessert und die Datenbasis wurde erheblich erweitert, sodass die früheren Ergebnisse kritisch überprüft werden konnten. Dabei wurden stellenweise methodische Schwächen festgestellt, welche auch dazu führten, dass Ergebnisse der früheren Arbeit teilweise verworfen oder umformuliert werden mussten. Wir haben andererseits auch neue Erkenntnisse erhalten. Wir konnten beispielsweise zeigen, dass es unter den Probanden Gemeinsamkeiten der Betrachtungsreihenfolge gibt. Die gewonnenen Erkenntnisse weisen darauf hin, dass die Objektwahrnehmung durch den Menschen einem speziellen sequenziellen Ablauf unterliegen und dass es damit tiefer liegende Regeln für die Erkennung von visuellen Objekten geben muss. Die Erstellung eines speziellen Modells dafür bleibt jedoch Nachfolgearbeiten vorenthalten.

# Erklärung

Ich erkläre hiermit,

- dass ich die vorliegende Arbeit selber und ohne fremde Hilfe durchgeführt habe, ausser derjenigen, welche explizit in der Aufgabenstellung erwähnt ist oder mit dem Betreuer schriftlich vereinbart wurde,
- dass ich sämtliche verwendeten Quellen erwähnt und gemäss gängigen wissenschaftlichen Zitierregeln korrekt angegeben habe,
- dass ich keine durch Copyright geschützten Materialien (z.B. Bilder) in dieser Arbeit in unerlaubter Weise genutzt habe.

Ort, Datum:

Jonathan Stolz, Unterschrift:

Thomas Jutzi, Unterschrift:

# Danksagung

Für die Unterstützung während der Bachelorarbeit möchten wir folgenden Personen einen besonderen Dank ausrichten:

**Prof. Dr. Ruedi Stoop** für seine Unterstützung und sein wertvolles, konstruktives und motivierendes Feedback,

**Ladina Lechner** und **Michèle Fröhlich** für das Korrekturlesen der Arbeit,

**Prof. Dr. Peter Heinzmann** für seine wertvollen Tipps zur Präsentation und das Feedback zum Management Summary,

allen Testpersonen für Ihr Interesse, am Experiment teilzunehmen, und für die dafür aufgebrauchte Zeit.



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
<b>2</b>	<b>Aufgabenstellung</b>	<b>3</b>
<b>3</b>	<b>Experimenteller Ansatz</b>	<b>4</b>
3.1	Verwendete Hardware . . . . .	4
3.2	Verwendete Testbilder . . . . .	5
3.2.1	Akzeptanzgebiet eines Objektes . . . . .	6
3.2.2	Eigenschaften der Objekte . . . . .	9
3.3	Testaufbau . . . . .	9
3.3.1	Aufzeichnung . . . . .	10
3.4	Auswertung . . . . .	15
3.4.1	Fixationszeit des Auges . . . . .	15
3.4.2	Versuchsanalyse . . . . .	17
3.4.3	Kumulierte Analyse pro Testbild . . . . .	21
<b>4</b>	<b>Modelle der Objektwahrnehmung</b>	<b>26</b>
4.1	Features . . . . .	27
4.1.1	Fläche . . . . .	27
4.1.2	Goldener Schnitt . . . . .	27
4.1.3	Distanz zur Mitte . . . . .	28
4.1.4	Randkontrast . . . . .	28
4.1.5	SSC-Stabilität . . . . .	29
4.2	Modelle der ersten Objektwahl . . . . .	34
4.2.1	”Distanz zur Mitte”-Modell . . . . .	36
4.3	Modelle der zweiten Objektwahl . . . . .	37
4.3.1	Flächenmodell . . . . .	38
4.3.2	”Goldener Schnitt”-Modell . . . . .	38
4.3.3	”Distanz zur Mitte”-Modell . . . . .	39
4.3.4	Randkontrastmodell . . . . .	39
4.3.5	SSC Stabilitäts-Modell . . . . .	40

4.3.6	SSC Reihenfolge-Modell . . . . .	40
4.3.7	Ergebnisse zur zweiten Objektwahl und Interpretation . . . . .	41
4.4	Modelle von Mustern der Objektreihenfolge . . . . .	42
4.5	Vergleich mit Surrogate-Daten . . . . .	46
<b>5</b>	<b>Schlussfolgerung</b>	<b>50</b>
5.1	Vergleich mit der Pilotarbeit . . . . .	50
5.2	Neue Ergebnisse . . . . .	51
<b>6</b>	<b>Ausblick</b>	<b>54</b>
<b>7</b>	<b>Anhang</b>	<b>55</b>
7.1	Persönliche Berichte . . . . .	55
7.1.1	Jonathan Stolz . . . . .	55
7.1.2	Thomas Jutzi . . . . .	55
7.2	Wichtige Begriffe . . . . .	56
7.2.1	Sakkade . . . . .	56
7.2.2	Gyroskop . . . . .	56
7.2.3	Features . . . . .	57
7.3	Datenübersicht: Features . . . . .	57
7.4	Datenübersicht: Modelle der ersten Objektwahl . . . . .	62
7.4.1	Flächenmodell . . . . .	62
7.4.2	”Goldener Schnitt”-Modell . . . . .	63
7.4.3	”Distanz zur Mitte”-Modell . . . . .	64
7.4.4	Randkontrastmodell . . . . .	64
7.4.5	SSC Stabilitäts-Modell . . . . .	65
7.4.6	SSC Reihenfolge-Modell . . . . .	66
7.5	Verwendete Werkzeuge, Code . . . . .	67
7.5.1	Hardware, Software und Betriebssystem . . . . .	67
7.5.2	Umgebungsvariable . . . . .	67
7.5.3	Ordnerstruktur der DVD . . . . .	67
7.5.4	Beschreibung der entwickelten Tools . . . . .	68
7.5.5	Bildnummern in der Ordnerstruktur . . . . .	69
7.6	Arbeitsplanung . . . . .	70

# 1

## Einführung

Visuelle Wahrnehmung ist ein höchst wichtiger Bestandteil unseres täglichen Lebens. Wir leben in einer Flut von visuellen Reizen, sei es durch Zeitungen, Plakate oder durch eine schöne Aussicht vom Gipfel eines Berges. Man könnte denken, die visuelle Wahrnehmung sei eine gut verstandene Wissenschaft.

Schon Euklid (365–300 v. Chr.) setzte sich mit Problemen der Raumwahrnehmung auseinander. Erst später erkannte Leonardo da Vinci (1452–1519) als erster die optische Besonderheit des menschlichen Auges. Er schrieb: "Das Auge hat eine einzige Zentrallinie, und alle Dinge, welche durch diese Linie zum Auge gelangen, werden gut gesehen. Um diese Linie gibt es eine unendlich grosse Anzahl anderer Linien, die mit der Zentrallinie in Berührung kommen und die umso wirkungsloser sind, je weiter sie von besagter Zentrallinie entfernt sind." [1] In diesem Zusammenhang gibt es einen weiteren interessanten Kennwert, den *Goldenen Schnitt*, welcher beispielsweise in der Natur beim Efeublatt vorkommt. Auch in der Kunst, bei Albrecht Dürers Selbstbildnis von 1500 [2] und in Leonardo da Vinci's Abendmahl [3], ist er der Bildkomposition zu Grunde gelegt.

Ein Problem der visuellen Wahrnehmung ist, dass die Abläufe nicht nur im Auge stattfinden, sondern ein grosser Teil davon im Hirn abläuft. Trotz vieler Anstrengungen, die Wahrnehmung durch den Menschen zu untersuchen, ist es noch immer nicht gelungen, zu verstehen, wie der Mensch diese äusserst komplexe Aufgabe durchführt. Durch technische Entwicklungen ist es heute möglich, die Augenbewegungen eines Menschen mittels eines Augenverfolgers (Eye-Tracker) aufzuzeichnen. Dies ermöglicht es, den Menschen beim Prozess der visuellen Wahrnehmung ganz genau zu beobachten. In unserer Arbeit bedienen wir uns eines solchen Eye-Trackers, um Anhaltspunkte darüber zu erhalten, wie das menschliche Hirn während einer solchen Wahrnehmung vorgeht. Uns interessiert, ob es gewisse Qualitätsmerkmale der betrachteten Objekten gibt, für die sich der Mensch speziell interessiert.

Zu diesem Zweck untersuchen wir den zeitlichen Ablauf der Objektwahrnehmung. Aus diesen Daten versuchen wir dann, ein Modell des Objekterkennungsvorgangs zu erstellen.

## 2

# Aufgabenstellung

Sogar in stark verrauschten zweidimensionalen Bildern findet der Mensch mühelos korrekt versteckte Objekte. Für künstliche Systeme, wie sie etwa in der Robotik oder für Überwachungsaufgaben eingesetzt werden, ist dies eine bisher immer noch schlecht gelöste Aufgabe. Für technische Anwendungen ist es wichtig, zu wissen, wie das Hirn Objekte erfasst; dies kann zum Beispiel zur Aufstellung einfacher Regeln von optimalen Objektdarstellungen führen.

Das Auge und seine Bewegungen öffnet uns dabei einen Spalt, hineinzusehen, wie das Hirn solche Objekte wahrnimmt. Mit einem Augenverfolger (Eye-Tracker) werden die Blickbewegungen von Testpersonen während eines Experimentes aufgezeichnet und danach analysiert. Uns interessiert dabei, welche die für das Auge zielbestimmenden Qualitäten der Objekte sind und wie der Entdeckungsvorgang zeitlich vor sich geht. Dafür wollen wir die Übereinstimmung der Versuche mit verschiedenen Modellen, welche wir durch von uns entwickelte Algorithmen der künstlichen Intelligenz erhalten haben, vergleichen.

Vielversprechende Anfangsarbeiten [4] dazu sind bereits an der ETH Zürich / Universität Zürich durchgeführt worden. An dieser Stelle geht es jetzt darum, diese zu hinterfragen, sie technisch zu verbessern, die gemachten Folgerungen zu überprüfen und sie durch grössere Versuchszahlen statistisch abzusichern. Auf Grund der erhobenen Daten soll -dies der Wunsch- ein möglichst detailliertes Modell des Objekterkennungsvorgangs entwickelt werden, welches dann in einer Fachzeitschrift publiziert werden könnte.

Ort, Datum, Unterschrift:

# 3

## Experimenteller Ansatz

In diesem Kapitel wird das durchgeführte Experiment vorgestellt. Die dafür benötigten Testbilder werden aufgezeigt und erläutert. Die verwendete Hardware und Software wird vorgestellt. Die dabei angetroffenen Probleme und Schwierigkeiten werden beschrieben.

### 3.1 Verwendete Hardware

Für das Experiment stand uns ein Eye-Track System der CHRONOS VISION GmbH [5] zur Verfügung, welches in Abbildung 3.1 dargestellt ist. Diese Brille zeichnet die Augenbewegungen der Probanden auf. Die Brille kann für jeden Probanden individuell eingestellt werden, damit sie richtig sitzt und die Augen durch die Brille erfasst werden können.

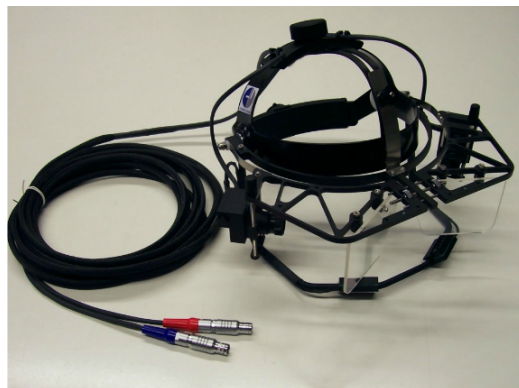


Abbildung 3.1: Der in unserem Experiment verwendete Eye-Tracker (CHRONOS VISION GmbH).

Bei dieser Brille handelt es sich um einen reinen Eye-Tracker, das heisst, es können keine Kopfbewegungen festgestellt werden. Da wir die Augenbewegungen auf Testbilder abbilden müssen, ist dies das gravierendste Problem an diesem System. Das heisst konkret, dass sich der Proband in unserem Experiment nicht bewegen darf, respektive der Kopf darf keinesfalls einer nennenswerten Bewegungen unterliegen. Dass dies gar nicht so leicht ist, konnten wir an unseren Selbstversuchen feststellen.

## 3.2 Verwendete Testbilder

Unsere Testdaten waren sehr einfach gehaltene Bilder mit einer Auflösung von jeweils  $30 * 30$  Pixel. Wie in der Abbildung 3.2 dargestellt ist, enthält jedes Bild sechs versteckte Objekte, welche unterschiedliche Grössen, Positionen und Grautonfärbungen aufweisen. Den Objekten ist ein Gaussches Rauschen im hellgrauen Bereich unterlegt. Dieses Rauschen dient dazu, dass der Proband die Objekte suchen muss und sie nicht gleich offensichtlich sieht. In einigen Bildern sind die Objekte leichter erkennbar als in anderen. Zu den erläuterten Bildern haben wir noch zwei weitere, spezielle Bilder angehängt, welche für weiterführende Analysen vorbereitet waren. Das zweitletzte farbige Bild (siehe Abbildung 3.2, 9) bricht aus den Grautonfärbungen der Objekte aus. Damit könnte man feststellen, ob die Farbwahl einen Einfluss auf die visuelle Wahrnehmung hat. Im letzten Bild (siehe Abbildung 3.2, 10) geht es darum, dass es in erster Linie keine versteckten Objekte gibt. Damit möchten wir sehen, wie die Probanden mit dieser Situation umgehen und ob man daraus mögliche Erkenntnisse für das Modell gewinnen kann.

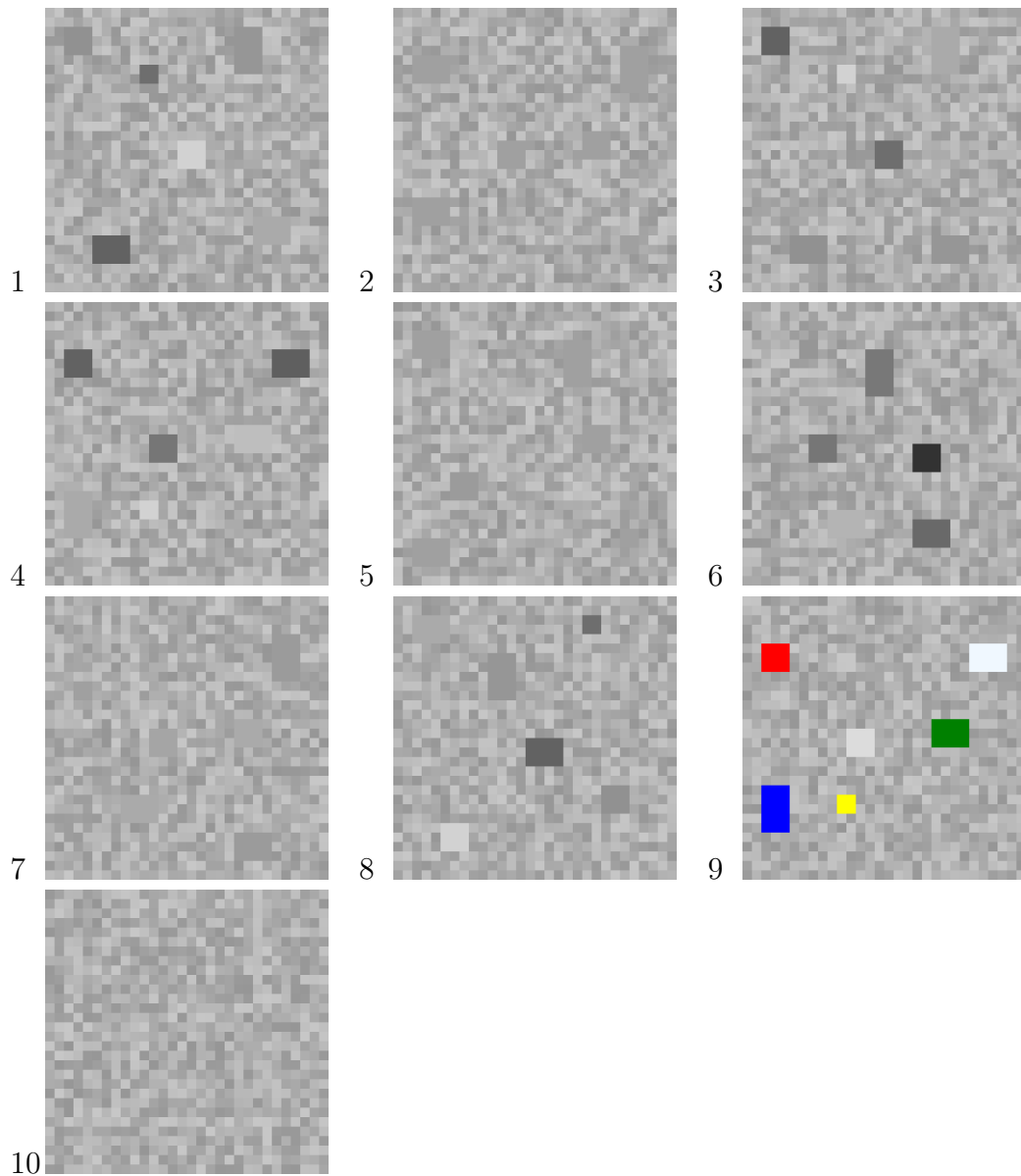


Abbildung 3.2: Die im Experiment verwendeten Testbilder: Die Testbilder 1 bis 8 sind die untersuchten Testbilder. Die beiden Testbilder 9 und 10 dienen weiterführenden Analysen. Beim Testbild 9 wird die Grautonfärbung der Objekte verlassen, hingegen fehlen die versteckten Objekte beim Testbild 10.

### 3.2.1 Akzeptanzgebiet eines Objektes

In den Testbildern haben wir speziell darauf geachtet, dass sich die Objekte niemals zu nahe kommen. Sonst kann nicht genau zugeordnet werden, welches



Objekt der Proband tatsächlich betrachtet hat. Dies ist wichtig, damit die Augenfixationen den Objekten zugeordnet werden können. Zu diesem Zweck muss rund um die Objekte herum ein Akzeptanzgebiet erstellt werden, innerhalb dessen ein Objekt als fixiert gilt. Wir haben uns für eine Toleranz von 2 Pixel entschieden. Wobei 1 Pixel dafür gedacht ist, dass der Proband den Rand des Objektes ansehen könnte und der andere Pixel ist ein Toleranzwert für minimale Kopfbewegungen.

Die Kopfbewegungen haben wir experimentell mittels eines Gyroskopes sowie eines Laser-Pointers vermessen (mehr Informationen über die verwendete Hardware befindet sich im Anhang im Kapitel 7.2.2). Dabei haben wir zwei grundsätzliche Messungen durchgeführt. Einerseits eine normale Aufzeichnung, bei der der Proband versucht, möglichst nicht den Kopf zu bewegen (Versuch  $M_{ruhig}$ ) und andererseits eine Messung, bei der der Kopf von oben links nach oben rechts um die Distanz von 28 Pixel gedreht wird (Versuch  $M_{beweg}$ ). Der Laser-Pointer diente dabei als Hilfsmittel, um die Ausrichtung des Kopfes den Pixel auf dem Bildschirm zuordnen zu können. Das Gyroskop liefert Koordinaten  $(X, Y, Z)$  aus dem 3-dimensionalen Raum. Um das Vorgehen zu vereinfachen, werden anstelle der  $X, Y, Z$ -Koordinaten ihr Summenwert  $x = |X| + |Y| + |Z|$  verwendet. Die Unterschiede zwischen  $M_{ruhig}$  und  $M_{beweg}$  sind in der Abbildung 3.3 illustriert.

Damit kann man feststellen, ob man beim Stillhalten des Kopfes sich tatsächlich nicht mehr als einen Pixel auf dem Testbild bewegt. Die Startpunkte aus unseren Mengen  $(M_{ruhig}, M_{beweg})$  halten wir mit  $S_{ruhig}$  = erster Messwert aus  $M_{ruhig}$  und  $S_{beweg}$  = erster Messwert aus  $M_{beweg}$  fest. Die Intervalle, die in den Versuchen  $M_{ruhig}$  und  $M_{beweg}$  überstrichen werden, sind:

$$Max_{ruhig} = Max[|S_{ruhig} - x|] \forall x \in M_{ruhig}, \quad (3.1)$$

$$Max_{beweg} = Max[|S_{beweg} - x|] \forall x \in M_{beweg}. \quad (3.2)$$

Damit kann man feststellen, dass der Wert, welcher bei  $Max_{beweg}$  entsteht, genau die Länge von 28 Pixel widerspiegelt. Das heisst, wir können diesen Wert durch 28 teilen, um eine Abweichung von der Länge eines Pixels zu bekommen. Da wir aber diesen Pixel verlassen können, nehmen wir an, dass man den doppelten Weg (2 Pixel) zurücklegen darf. Dieser Tatsache wird in der Formel 3.3 Rechnung getragen, indem der Wert  $Max_{beweg}$  mit dem doppelten Weg (2 Pixel) multipliziert wird. Dieser berechnete Wert darf dann  $Max_{ruhig}$  nicht überschreiten. Aufgrund von Messungenauigkeiten in unserem Experiment, wie auch unter Berücksichtigung der Schwankungen des Gyroskopes, haben wir eine Fehlertoleranz von 15% definiert:

$$Max_{ruhig} \leq \frac{Max_{beweg}}{28} * 2 \pm 15\%. \quad (3.3)$$

Unsere Messungen haben ergeben:

$$108.133 \leq \frac{1317.03}{28} * 2 \approx 99.807 \pm 15\%. \quad (3.4)$$

Als Ergebnis kann man festhalten, dass die gemessene Kopfbewegung tatsächlich nur etwa 1 Pixel auf der Grafik ausmacht, sofern man den Kopf möglichst still hält. Wir konnten dies auch mit Hilfe eines an das Gyroskop gekoppelten Laser-Pointers bestätigen.

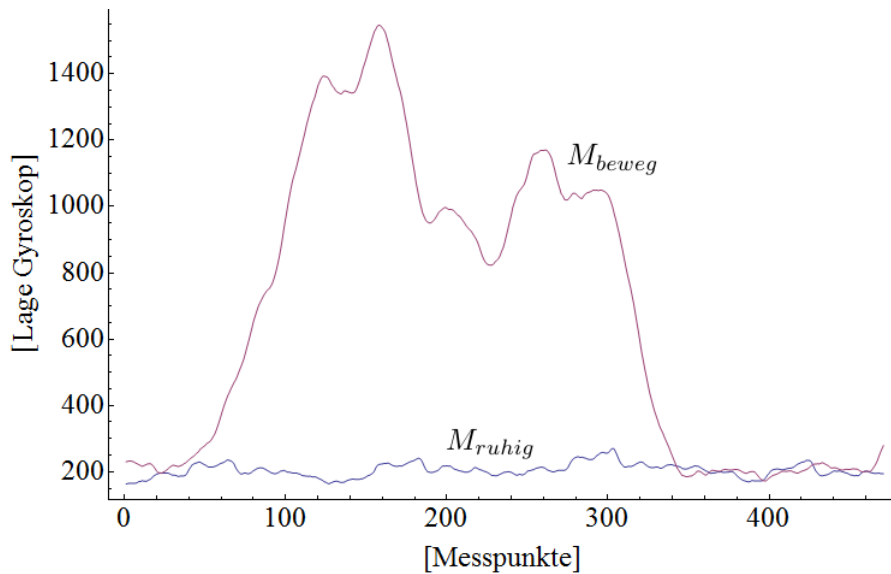


Abbildung 3.3: Experimentelle Messung der Kopfbewegung mittels Gyroskop, wobei  $M_{ruhig}$  die Ausgabe für einen stillgehaltenen Kopf zeigt. Bei der Messung  $M_{beweg}$  wird der Kopf absichtlich um 28 Pixel gedreht.

### 3.2.2 Eigenschaften der Objekte

Die sechs Objekte, welche wir in den verschiedenen Grafiken eingebaut haben, sind auf jeder Grafik dieselben. Lediglich ihre Farbe und Position ändert sich von Bild zu Bild. Dies ist für spätere Auswertungen wichtig, um eine Aussage über den Einfluss von Objektposition und Farbe machen zu können.

## 3.3 Testaufbau

Für das Experiment wurde den Probanden eine spezielle PowerPoint Präsentation gezeigt, welche die Testbilder automatisch in bestimmten Zeitintervallen anzeigt.

Um später die Messwerte trotz unterschiedlichen Koordinatensystemen auf unsere Testbilder zuordnen zu können, haben wir eine spezielle Kalibrierung konzipiert. Diese erlaubt uns später, aus der Datenmenge die individuellen Koordinatenbezugspunkte der Probanden abzuleiten. Damit den Testpersonen die Kalibrierung einfacher fällt, haben wir uns für eine Animation entschieden. In dieser fährt ein Punkt sämtliche Eckpunkte der Testbilder ab, welchem die Probanden folgen müssen. Später können wir anhand dem charakteristischen Muster, welches dabei entsteht, die Daten zurechtschneiden. In Abbildung 3.4 wird die Animation dargestellt. Gestartet wird, wie links abgebildet, beim grünen Punkt (in der Mitte) und man folgt dann automatisch den Pfeilrichtungen. Ist man beim Quadrat unten rechts angekommen, geht es gemäss dem rechten Bild weiter. Damit die Probanden das Experiment möglichst entspannt durchführen können, ist die Präsentation in vier Teile gegliedert. Damit haben wir bewusst der Tatsache Rechnung getragen, dass es nicht leicht ist, sehr lange den Kopf absolut still zu halten.

Der erste Teil ist die Einleitung, in welcher Informationen über das Experiment stehen und die Verhaltensregeln erklärt werden. Dann folgen die drei eigentlichen Experiment-Sequenzen, welche wie folgt gegliedert sind:

**Erste Sequenz** : Kalibrierungsanimation, gefolgt von den Testbildern 1 - 4 im Abstand von 10s

**Zweite Sequenz** : Kalibrierungsanimation, gefolgt von den Testbildern 5 - 7 im Abstand von 10s

**Dritte Sequenz** : Kalibrierungsanimation, gefolgt vom Testbild 8 sowie den beiden Zusatzbildern im Abstand von 10s

Die Unterteilung des Experiments in drei Testsequenzen hat noch einen weiteren Vorteil: Die Datenqualität kann merklich erhöht werden, da -wie bereits

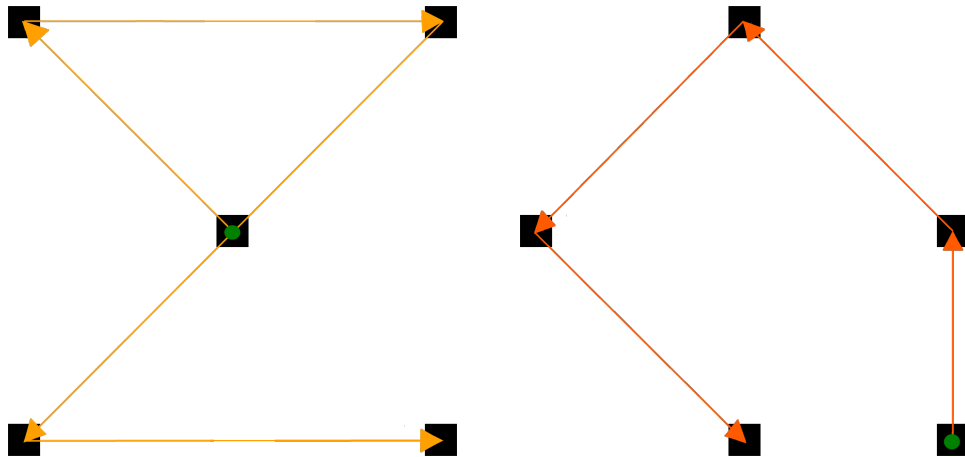


Abbildung 3.4: Kalibrierungsanimation, welche für die Zuordnung der Aufzeichnung zu den gezeigten Testbildern verwendet wird. Gestartet wird links in der Mitte, weiter geht es gemäss den Pfeilrichtungen. Das rechte Bild stellt den zweiten Teil der Animation dar, wobei der Start dem Ende der vorangegangenen Grafik entspricht.

beschrieben- der Kopf viel weniger lang still gehalten werden muss. Bei Eigenversuchen mit mehr Testbildern haben wir festgestellt, dass spätestens nach dem fünften Testbild der Kopf soweit wegbewegt wurde, dass man eigentlich keine Treffer mehr auf den Bildern ausmachen konnte.

### 3.3.1 Aufzeichnung

Für unser Experiment konnten wir 22 freiwillige Versuchspersonen gewinnen. Der Anteil der Frauen ist mit zwei Teilnehmerinnen leider ein wenig tief. Damit es unseren Probanden etwas leichter fiel, den Kopf nicht zu bewegen, haben wir eine improvisierte Kinnstütze aus Kartonschachteln zur Verfügung gestellt. In der Abbildung 3.5 ist einer der Autoren (J. Stolz) während der Aufnahme zu sehen. Die Kinnstütze weist noch einen weiteren Vorteil auf. Die Probanden konnten so auch die Distanz zwischen den Augen und dem Bildschirm nicht verändern. Anderenfalls würden sich die Drehwege der Augen verlängern, respektive verkürzen, was die Auswertung erheblich erschweren würde.

Die mit dem Eye-Tracker aufgezeichneten separaten Rohdaten für das linke und rechte Auge müssen wie folgt weiter aufbereitet werden. Als erstes wird das Chronos-Tool "Iris Tracker" verwendet, welches die Hough-Transformation [6] auf die Daten anwendet. Um diese Transformation durch-

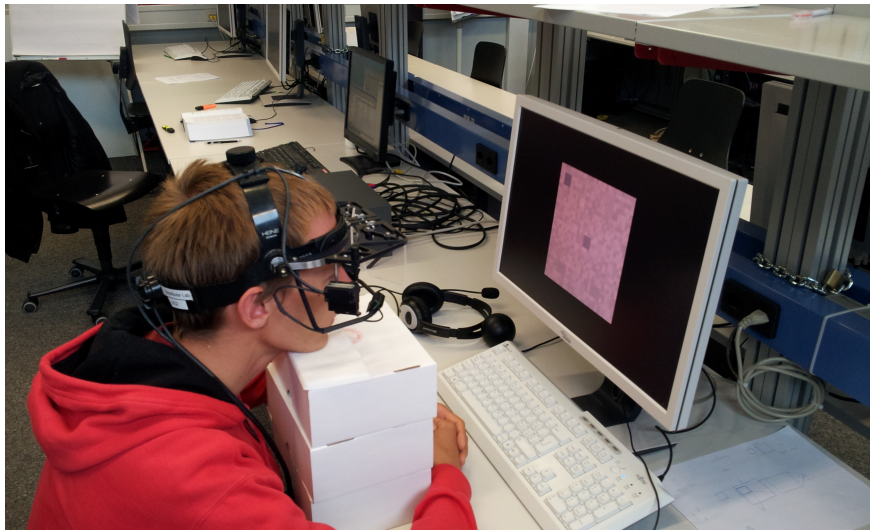


Abbildung 3.5: Bei der Durchführung einer Experimentes mit dem Eye-Tracker.

zuführen, definiert man ein Referenzbild, bei welchem der Algorithmus die Iris und Pupille korrekt erkennt. Abbildung 3.6 zeigt die Parametrisierung des Algorithmus', wobei eigentlich nur mittels dem "Pupil Detection Threshold" gearbeitet werden muss. Falls "geblinzelt" wurde, muss ein darauf folgendes Bild gewählt werden, bei dem das nicht der Fall ist. Nachfolgend wird die gesamte Verarbeitung automatisch durch das Tool prozessiert. Das Ergebnis ist eine Liste von Koordinaten, welche der Proband angesehen hat.

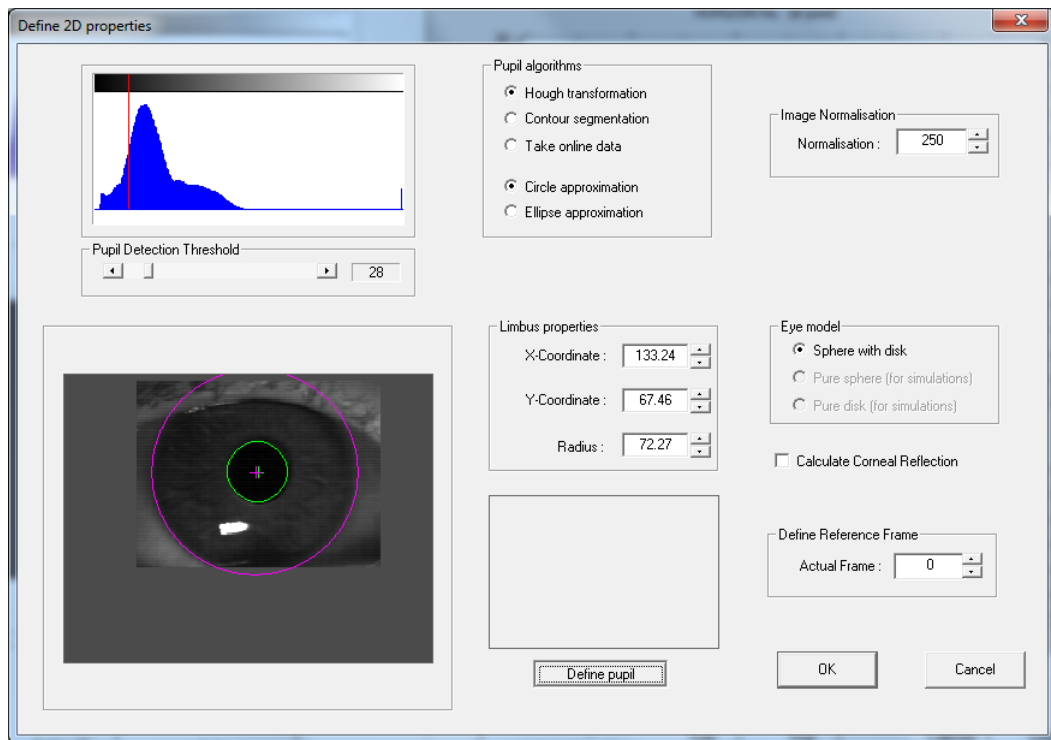


Abbildung 3.6: Das Chronos-Tool "Iris Tracker" mit der Auswahl der optimalen Parameter für die Hough-Transformation.

Die Koordinaten, welche das "Iris Tracker"-Tool uns liefert, liegen in der Form einer Datei vor. In der Datei befindet sich in der Kopfzeile unter anderem folgende Information:

```
# Hor/Vert Reference      : 109.030432 74.954892
```

Dies ist der Bezugspunkt für alle nachfolgenden Koordinaten, welche nur noch relativ zu diesem Punkt abgelegt sind. Die Koordinaten sind ebenfalls in einem kartesischen Koordinatensystem mit vier Quadranten abgelegt. Nachfolgend sind vier Beispiele, welche die relativen Bezüge aufzeigen:

SYNC-OUT [/]	Horizontal [pix]	Vertical [pix]	PupilDiam. [pix]
0	0.000000	0.000000	38.841881
1	0.038150	0.029461	38.948540
6	0.179108	-0.018655	39.464809
16	-0.028466	-0.064688	39.644997
849	-5.573748	13.236625	68.269722

Wie aus dem Handbuch des Eye-Trackers ersichtlich ist, können mitunter an Stelle der X-Y Koordinaten verschiedene Fehlermeldungen stehen:

- -777: Ein Bild fehlt (z.B. aufgrund langsamer Hardware)
- -888: Ein Bild ging während der Aufnahme verloren
- -999: Fehler in der Pupillendetektion (i.A. wegen Blinzeln)

Für die Weiterverarbeitung müssen diese Fehler entfernt werden.

Um diese Umwandlung möglichst einfach vornehmen zu können, haben wir ein grafisches Tool entwickelt, welches diese Arbeiten übernimmt. Der Schwerpunkt bei dieser Software war weniger die programmiertechnische Eleganz des Quellcodes, als viel mehr das Mittel zum Zweck. Einerseits dient das Tool dazu, diese aufbereiteten Dateien auf unsere Testbilder abzubilden, andererseits sind darin auch Fehler, wie zum Beispiel starke Kopfbewegungen, zu erkennen. In der Abbildung 3.7 ist oben eine perfekte Aufnahme ersichtlich und unten ein Bild mit drastischen Kopfbewegungen.

Bei der perfekten Aufnahme ist sehr schön zu sehen, welches Gewicht der Kalibrierungsanimation zukommt: Damit die nachfolgenden Bilder korrekt auf die 30\*30 Pixel der Testbilder zugeordnet werden können, müssen die Kalibrationsdaten korrekt auf die Referenzpunkte fallen.

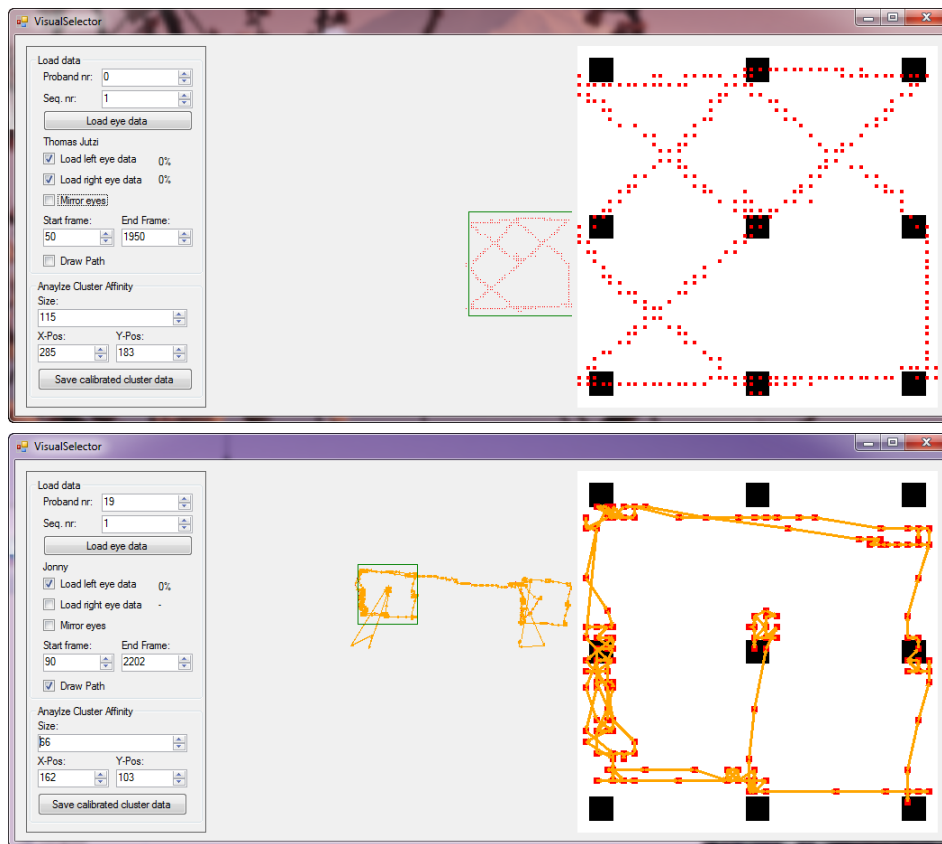


Abbildung 3.7: Koordinaten werden mit dem selbst entwickelten grafischen Tool ausgewählt. Oben eine perfekte Aufnahme, in der das Kalibrierungsbild ideal auf die Referenzpunkte gelegt werden kann. Unten eine unbrauchbare Aufnahme, bei der der Proband den Kopf während der Kalibrierung gedreht hat.

Mit dem von uns entwickelten Tool kann weiterführend ebenfalls das Auseinanderschneiden der Aufzeichnung entsprechend der jeweiligen Testbilder erledigt werden. Dies ist nötig, weil eine Aufnahmesequenz mit dem Eye-Tracker immer mehrere Testbilder umfasst. Da einerseits der Eye-Tracker eine definierbare Sample-Rate hat und andererseits unsere Präsentation automatisch nach gewissen Zeitintervallen umschaltet, kann dies relativ einfach bewerkstelligt werden. Für sämtliche Aufnahmen mit dem Eye-Tracker haben wir eine Sample-Rate von 100Hz gewählt. Das steht ebenfalls in der Kopfzeile der Ausgabedatei des "Iris-Tracker"-Tools, unter folgendem Punkt:

```
# Sample rate           : 100
```

Dadurch können wir relativ einfach die Daten zurechtschneiden, da jedes einzelne Sample eine eigene aufsteigende Nummer erhält. Diese Nummer und



die Information, dass die Sample-Rate 100Hz beträgt, lässt nun auf die relative Zeit des Einzelbildes schliessen. Anhand dieser Zeitinformation und der Zeitinformation über die Dauer der einzelnen Bildern in der Präsentation, lassen sich nun die einzelnen Ausschnitte aus den Daten herausschneiden. Die einzelnen Ausschnitte sind dann allerdings noch nicht auf die 30x30 Pixel grossen Testbilder eingepasst und müssen entsprechend skaliert werden. Dies ist ebenfalls relativ einfach machbar, indem die Grösse des Auswahlquadrates (in der Abbildung 3.7 jeweils der grüne Rahmen im Mittelbereich) proportional gestaucht oder gestreckt wird.

Nach diesen Verarbeitungsschritten sind die Daten nun für weitere Analysen bereit.

## **3.4 Auswertung**

In diesem Kapitel werden die aufgenommenen Aufzeichnungen analysiert und ausgewertet. Zuerst gehen wir auf die probandenbezogenen Auswertungen näher ein. Danach folgt die kumulierte Analyse für jedes Testbild, wobei wir ausgewählte Ergebnisse der Objektwahl genauer präsentieren. Weitere Ausführungen zur Modellbildung und zu den entsprechenden Datenauswertungen folgen im Kapitel 4.

### **3.4.1 Fixationszeit des Auges**

Um die Ergebnisse der Objektzuordnung weiter zu verbessern, haben wir die Augenbewegungen so stark im Detail analysiert, dass wir zwischen einer Sakkade (schnelle Abfolge von Augenbewegungen, siehe Anhang: Kapitel 7.2.1) und einer Fixation unterscheiden können. Diese Unterscheidung hilft uns, zufällige Objekttreffer, welche aufgrund eines Überfluges während einer Sakkade zustande kommen, zu eliminieren. Abbildung 3.8 zeigt das Problem anhand eines Testbildes sehr gut. Auf der linken Seite sind die Punkte, die während der Sakkade durch den Eye-Tracker aufgenommen wurden, nicht aus dem Datenmaterial entfernt worden. Dies bedeutet, dass während dem eigentlichen Wechsel vom Objekt F (unten links, grüner Punkt) zum Objekt C (oben rechts, roter Punkt) auch noch das Objekt D (mitte, blaue Punkte) überflogen und als Treffer markiert wird. Da dieses Objekt bei einem solchen Wechsel aber eigentlich gar nicht betrachtet wurde, sollte er nicht als Treffer gewertet werden. Dieses Problem ist im rechten Bild korrigiert, womit das Ergebnis dem entspricht, was tatsächlich angesehen wurde.

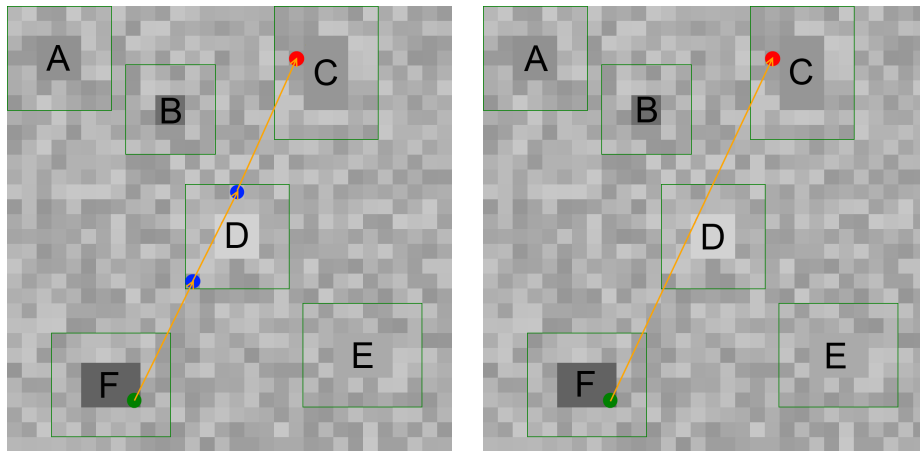


Abbildung 3.8: "Überflugproblem" bei einer Sakkade. Links: Wenn die Sakkaden nicht gefiltert werden zählt Objekt D (Mitte) ebenfalls als Treffer. Rechts: Wenn die Sakkade herausgefiltert wird (korrekt).

Daher beschränken wir uns für die Objektzuordnung auf diejenigen Pixel, die für eine Mindestdauer von 30ms angeschaut wurden. Diese 30ms wurden gewählt, weil sie eine charakteristische Zeitgrösse im neuronalen Bereich darstellen. Viele Neuronen lassen im Mittel diese Zeit verstreichen, bis sie wieder feuern [7].

Zusätzlich haben wir ermittelt, wie viele Punkte für verschiedene Zeitspannen eliminiert werden. Dabei zeigt sich, wie in Abbildung 3.9 dargestellt, dass der grösste Knick bei 10ms entsteht, dann fällt es mehr oder weniger linear ab. Dies zeigt, dass wir mit einer Zeit von 30ms auf der sicheren Seite sind und nicht zu viele Punkte verlieren werden. Berücksichtigt man auch die Latenz der Muskulatur, welche 30ms - 70ms beträgt [8], stellt ein Wert von 30ms ein gutes Minimum der Fixationszeit dar.

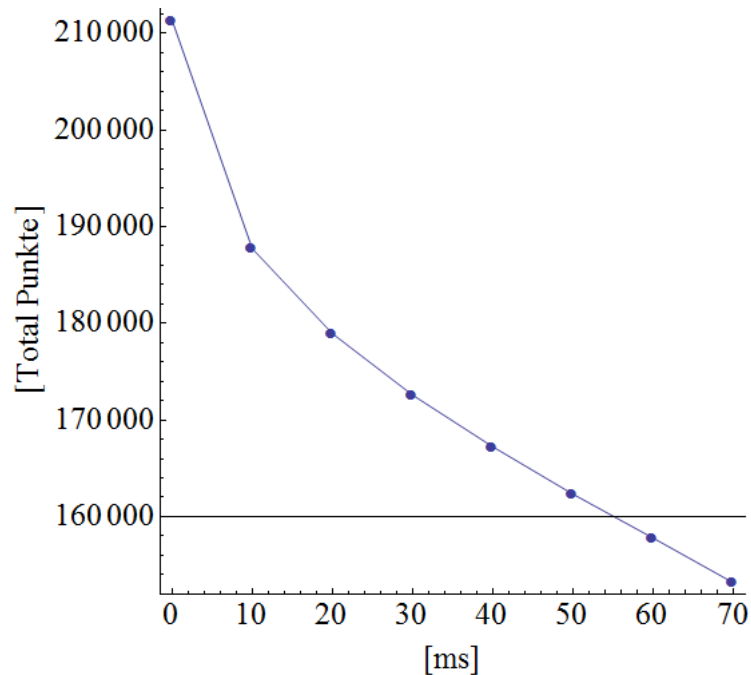


Abbildung 3.9: Eliminierung der anvisierten Punkte über die Fixationszeit in ms.  
Ein guter Wert ist 30ms.

### 3.4.2 Versuchsanalyse

Die korrekt zugeschnittenen Daten nach der Verarbeitung der Aufzeichnung (Kapitel 3.3.1) werden nun für jeden Proband pro Bild einzeln ausgewertet. Diese Daten haben das folgende Format:

X	Y
13	12
12	11
9	10
7	9
7	8

Die einzelnen  $x$  und  $y$  Werte der Koordinaten liegen zwischen 0 und 29. Diese Abfolge von Koordinaten widerspiegeln die betrachteten Punkte im zeitlichen Verlauf des einzelnen Probanden. Aus diesen Daten wird eine Grafik erstellt, welche die Augenbewegungen des Probanden anhand von Pfeilen auf die Grafik mit den Objekten und deren Akzeptanzgebiet zeichnet. Eine solche Grafik eines Probanden für das Testbild 1 ist in Abbildung 3.10 dargestellt.

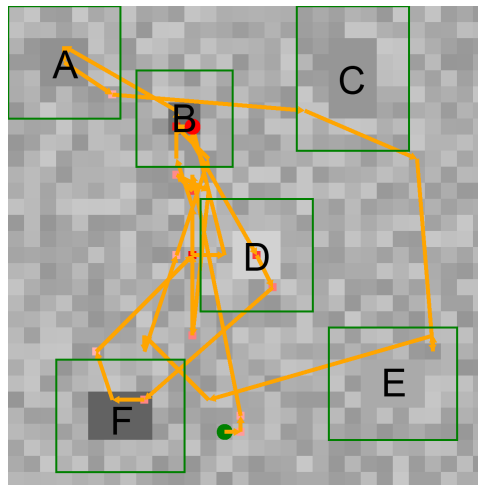


Abbildung 3.10: Augenbewegungen eines Probanden (Testbild 1). Der grüne Punkt ist der Startpunkt, der rote Punkt der Endpunkt der Aufzeichnung.

Diese Grafik dient lediglich zur manuellen Kontrolle der Testdaten und als Veranschaulichung der abstrakten Daten.

Als zweites wird nun eine Sequenz erstellt, welche die betrachteten Objekte enthält. Der folgende Code veranschaulicht, wie die Sequenz erstellt und die Zeit für die Betrachtung aller Objekte berechnet wird. Die Methode `Intersect()` entscheidet, ob die Koordinate innerhalb des Akzeptanzgebiets des Objektes liegt oder nicht.

```
string objectSequence = "";
int totalTimeToVisitAllClusters = 0;

int framesCount = 0;
bool visitedA = false;
bool visitedB = false;
...
foreach(point in coordinates)
{
    //sequence generation
    if(ObjectA.Intersect(point))
    {
        objectSequence += "A";
        visitedA = true;
    }
    else if (ObjectB.Intersect(point))
```

```

{
    objectSequence += "B";
    visitedB = true;
}
//and so on for all objects C-F
...

//time calculation
framesCount++;
if(visitedA && visitedB && ...)
{
    totalTimeToVisitAllClusters = framesCount * 10;
    break;
}
}

```

Die für den Besuch aller Objekte benötigte Zeit beträgt in diesem Beispiel 8250ms. Die erzeugte Sequenz kann etwa wie folgt aussehen:

```
BBBBBBBAAACCCCCCCCCCCCCCEEEEEEEEEEBB.....
```

Die Sequenz beinhaltet in diesem Format auch Informationen zur verweilten Zeit auf einem Objekt. Da die Aufzeichnung, wie beschrieben in Kapitel 3.3.1, mit 100 Frames pro Sekunde durchgeführt wird, kann nun daraus gelesen werden, wie lange der Proband auf dem jeweiligen Objekt verblieb. Ein Frame hat daraus folgend eine Dauer von 10ms. Für unsere Modelle benötigen wir nur die Information, von welchem Objekt auf welches andere gewechselt wird und welches zuerst angeschaut wird. Aus diesem Grund wird die erhaltene Sequenz nochmals vereinfacht, indem die Zeitinformation entfernt wird. Die dauerbereinigte Sequenz sieht dann wie folgt aus:

```
BACEB.....
```

Diese Sequenz enthält jetzt nur noch die Übergänge. Für die Erstellung einer Markov-Übergangsmatrix und eines Automatenbildes, ohne Übergänge auf sich selbst, ist dies die gewünschte Form. Auf dieser Datenbasis werden nun Markov-Übergangsmatrizen für jeden Proband pro Bild erstellt. Ein Beispiel einer solchen Auswertung für einen Probanden ist in der Abbildung 3.11 dargestellt.

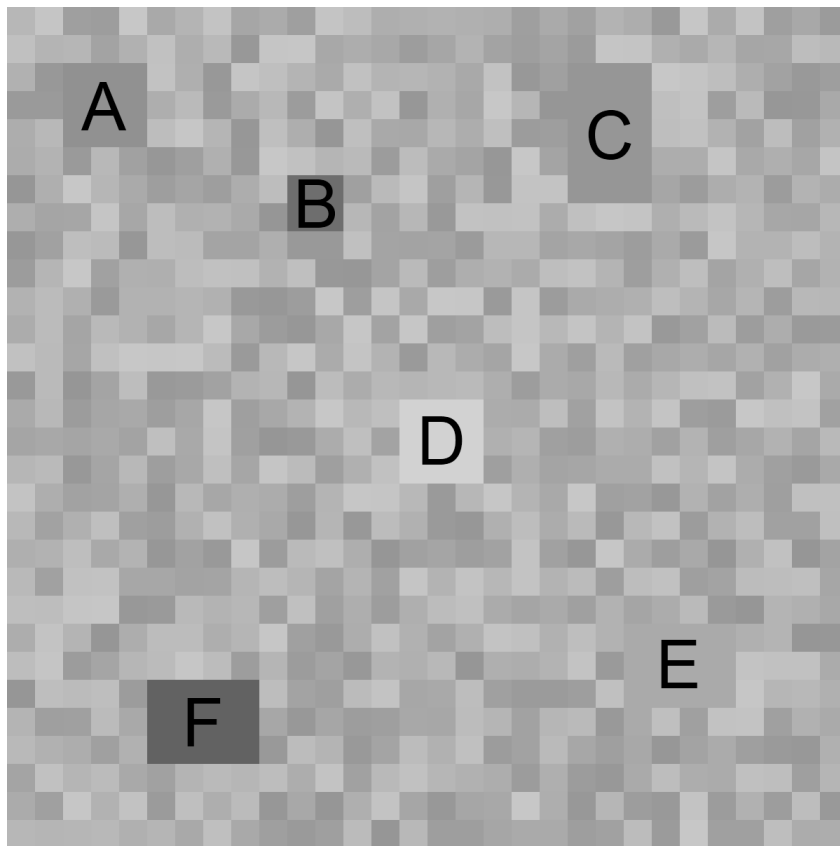
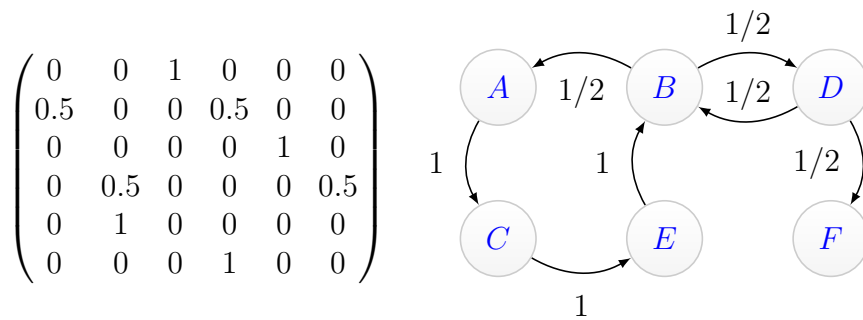


Abbildung 3.11: Obere Zeile: Markov-Matrix (Übergangswahrscheinlichkeiten), daneben das dazugehörige Automatenbild. Untere Zeile: Zugehöriges Testbild 1.

Diese hier berechneten Daten bilden die Grundlage für weitergehende Analysen, auf welche in den folgenden Kapiteln genauer eingegangen wird.

### 3.4.3 Kumulierte Analyse pro Testbild

Die Daten, welche pro Proband und pro Bild erstellt wurden, werden in diesem Schritt weiter verarbeitet. Es wird für jedes Bild ein Zusammenzug aus allen Probanden generiert. Für jedes Bild wird nun eine Grafik mit allen Start- und Stopp-Punkten erstellt.

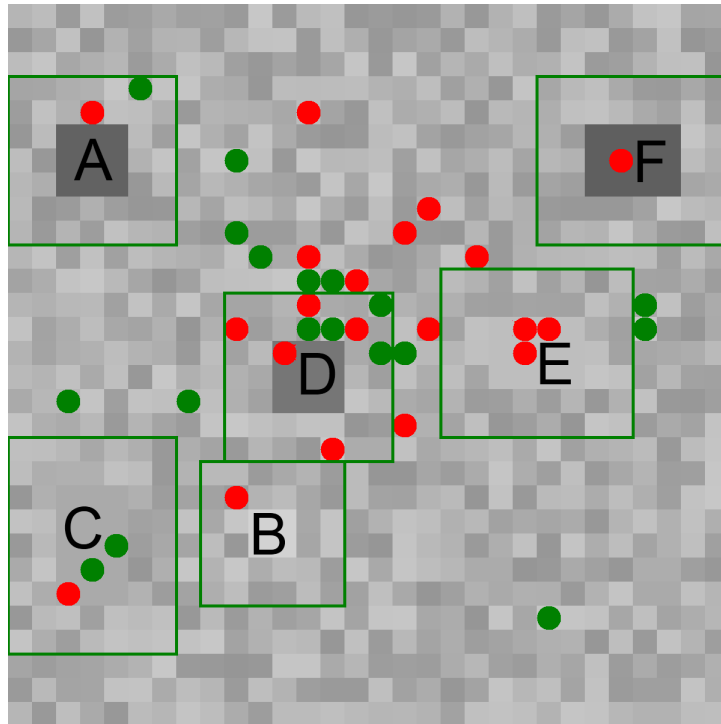


Abbildung 3.12: Grün: Startpunkte, rot: Stopp-Punkte (Testbild 4).

Die vorhandenen Sequenzen werden pro Proband und Bild weiter verarbeitet. Aus diesen Sequenzen wird das erste und zweite besuchte Objekt herausgelesen. Dies ermöglicht eine Aussage, welches Objekt wie häufig als erstes und wie häufig als zweites angeschaut wurde. Bei der manuellen Kontrolle dieser zusammengezogenen Daten zeigt sich für das Testbild 1 eine Auffälligkeit bei den ersten besuchten Objekten. Die meisten Probanden schauten sich zuerst das Objekt F an, welches unten auf dem Bild platziert ist. Zu beachten ist, dass sich die Probanden vor diesem Testbild die Kalibrierungsanimation betrachteten. Nach dieser Sequenz ist der Fokus der Probanden im unteren Bereich des Bildes, wie in der Abbildung 3.13 links zu erkennen ist.

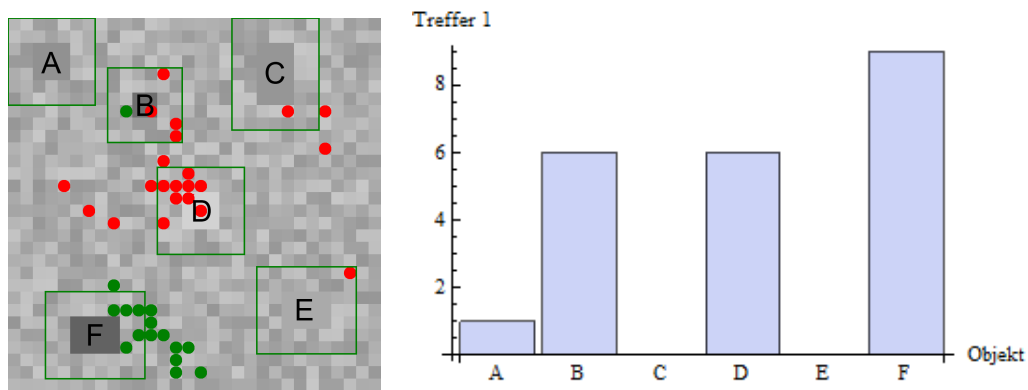


Abbildung 3.13: Links: Startposition der Probanden (grüne Punkte). Rechts: Die Verteilung der zuerst betrachteten Objekte. In der Latenzzeit von 30ms (aufgrund des vorausgegangenen Bildwechsels) verließen aber offenbar viele Probanden das Objekt F.

Dies verleitete die meisten Probanden dazu, zuerst auf das Objekt F zu schauen. Aus diesem Grund wird bei dem ersten Bild für alle Probanden das erste erkannte Objekt aus der Sequenz herausgelöscht. Das zweite Objekt wird in diesem Fall als erstes Objekt genommen. Alle anderen Bilder haben keine solche Auffälligkeit gezeigt und wurden deshalb nicht korrigiert.



Um die wichtigsten Objekte in den einzelnen Bildern zu erkennen, haben wir für jedes Bild eine Heatmap [9] erstellt. Diese Heatmap färbt die Punkte heller, welche häufiger betrachtet werden.

In der Abbildung 3.14 ist die Heatmap für die Kalibrierungsanimation dargestellt. Hier ist klar zu erkennen, dass die Probanden sehr präzise nur den Kalibrierungspfad fokussiert haben. Bei dem Testbild 4 wurden die Objekte leicht hervorgehoben. Das Testbild 10 enthält nur ein Gaussssches Rauschen ohne Objekte. Bei allen Testbildern ist die Mitte der wärmste Punkt auf der gesamten Grafik. Sogar bei dem Testbild 10 wurde die Mitte am stärksten fixiert.

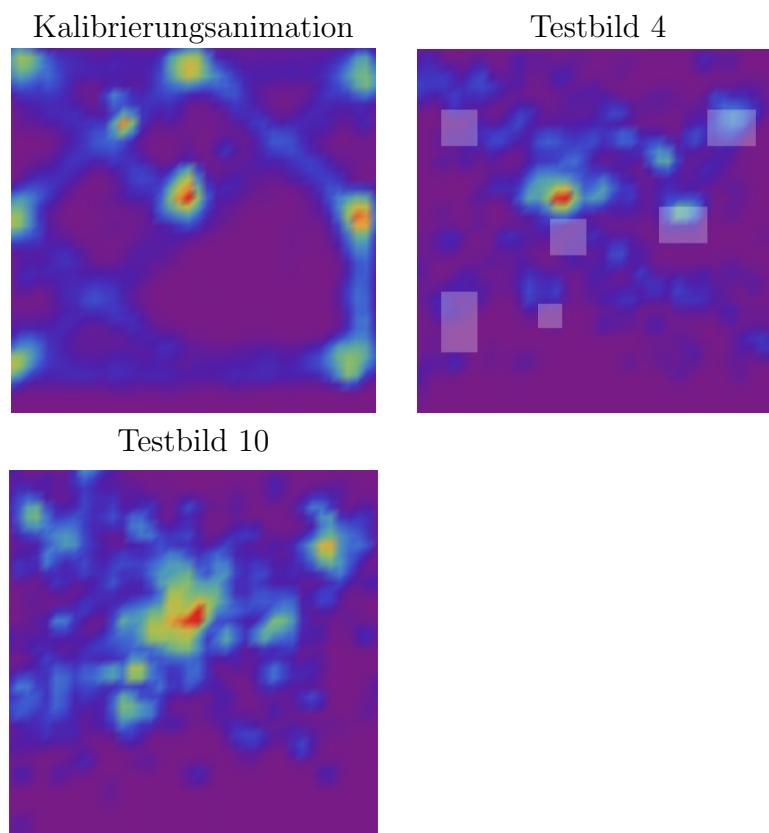


Abbildung 3.14: Die Heatmap der Kalibrierungsanimation zeigt klar den fokussierten Pfad auf. Im Testbild 4 haben wir zusätzlich zur Heatmap die zugrunde liegenden Objekte grafisch herausgehoben. Testbild 10 enthält keine Objekte. In beiden Testbildern sticht die Fokussierung auf die Bildmitte klar hervor.

Diese Heatmaps zeigen die für die Probanden wichtigsten Punkte in den Bildern über die gesamte Zeitdauer auf. Uns interessierte auch, wie lange die Probanden im Schnitt benötigten, um alle Objekte in den Bildern zu erkennen. Diese Zahlen helfen auch allfällige Ausreisser zu erkennen. Die durchschnittliche Zeitdauer wird nur über diejenigen Probanden ermittelt, welche auch alle Objekte erkannt haben. Auch wird in der folgenden Tabelle 3.15 aufgelistet, wie viele Probanden alle Objekte erkannt haben.

Testbild	⊙ Dauer [s]	Anz. Probanden	% Probanden
1	7.9	6	29%
2	8.1	4	19%
3	7.5	6	29%
4	7.8	6	29%
5	8.4	4	19%
6	8.1	8	38%
7	8.3	7	33%
8	8.0	9	43%
9	7.6	14	67%

Abbildung 3.15: Tabelle mit Daten für jedes Testbild, wie lange die angegebene Anzahl Probanden im Schnitt benötigte, um alle Objekte zu erkennen.

Das Testbild 9 sticht in dieser Tabelle mit der Anzahl der Probanden stark heraus. Dieses Bild ist das Bild mit den farbigen Objekten, alle anderen Bilder haben nur Objekte mit Grautonfärbungen. Es stellt sich die Frage, weshalb bei diesem Bild viel mehr Probanden alle Objekte erkannt haben. Dies mag vielleicht an den schwierigeren Graustufen-Testbildern wie zum Beispiel Testbild 2 und Testbild 5 liegen.

Die im Kapitel 3.4.2 erstellten Markov-Übergangsmatrizen werden hier nun pro Testbild zusammengerechnet, um für jedes Testbild eine einzige Markov-Übergangsmatrix zu erhalten. Die Abbildung 3.16 zeigt die berechneten Markov-Übergangsmatrizen für alle Bilder.

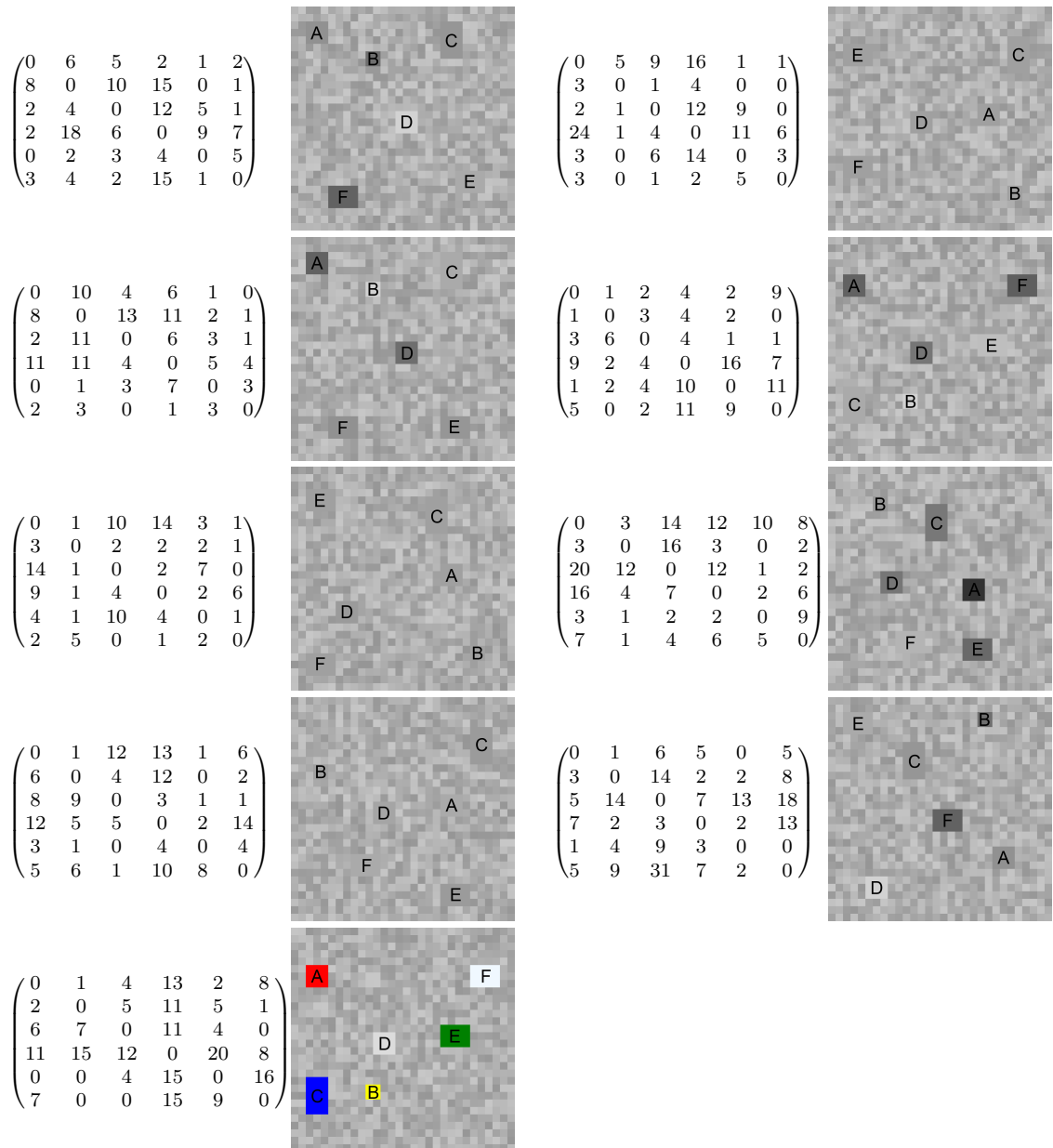


Abbildung 3.16: Markov-Übergangsmatrizen für alle Bilder. Das erste Matrix ist oben links angeordnet, das dazugehörige Testbild ist immer direkt rechts neben der Matrix. Die weiteren Matrizen und deren Bilder sind in der Lesefolge angeordnet. Das letzte Bild ist das spezielle Testbild mit den farbigen Objekten. Bei allen Matrizen sind die Objekte immer in der Reihenfolge 'ABCDEF' aufgebaut. Um z.B. die Anzahl Übergänge von Objekt B auf D auszulesen, muss in der zweiten Zeile (B) die vierte Spalte (D) angeschaut werden.

## 4

# Modelle der Objektwahrnehmung

In diesem Kapitel werden Modelle zur Vorhersage der Wahrnehmung erstellt. Diese Vorhersagemodelle werden anhand der Qualitätsmerkmale der einzelnen Bilder erstellt. Die Korrektheit wird mit den erhobenen Probandendaten kontrolliert.

Als erstes werden die verwendeten Qualitätsmerkmale -die Features- genauer beschrieben. Danach werden Modelle zur Vorhersage der ersten und zweiten Objektwahl erstellt. Das Ziel ist es, hier die Vorhersagegenauigkeit der einzelnen Modellen zu analysieren. Die Abbildung 4.1 zeigt eine Auflistung der verwendeten Modelle, welche genauer betrachtet werden. Der letzte Teil befasst sich mit der Erstellung von Modellen zur Vorhersage, in welcher Reihenfolge die Objekte betrachtet werden.

Modell	Beschreibung
Flächenmodell	Bei diesem Modell werden die Objekte der Grösse nach ausgewählt.
”Distanz zur Mitte”-Modell	Das Objekt, welches am nächsten in der Mitte ist, wird zuerst gewählt.
Randkontrastmodell	Objekte mit hohem Randkontrast werden vor denjenigen mit einem tiefen ausgewählt.
SSC Stabilitätsmodell	Dieses Modell benützt die Stabilitäten aus dem SSC Clustering absteigend.
SSC Ordnungsmodell	Die Vorhersage-Reihenfolge des Modells entspricht der SSC Clustering Reihenfolge.

Abbildung 4.1: Tabelle mit den Modellen, welche ihre Voraussagen auf den berechneten Features aufbauen.

## 4.1 Features

”Features” sind unabhängig von den experimentell erhobenen Daten, da die Berechnung auf den Gegebenheiten des jeweiligen Testbildes basiert. Nachfolgend beschreiben wir sämtliche verwendeten Features. Dabei erläutern wir ebenfalls die Beweggründe zur Wahl des jeweiligen Features.

### 4.1.1 Fläche

Wir stellen uns vor, dass die Fläche eines Objektes einen Einfluss auf die Objektwahl hat, weil man grössere Objekte früher erkennt.

Dieses Feature stellt die geometrischen Fläche eines Objektes dar. Bei allen Testbildern wurde ein Pixel als Masseinheit gewählt, wobei  $l$  die Länge und  $b$  die Breite des Objektes repräsentiert:

$$F = l * b. \quad (4.1)$$

### 4.1.2 Goldener Schnitt

Liest man den Artikel zum goldenen Schnitt [10], so wird darin gezeigt, dass in der Natur, wie auch in der Darstellung durch den Menschen, dem goldenen Schnitt eine besondere Bedeutung zukommt. Dieses Feature könnte auch für diese Arbeit eine Rolle spielen. Der goldene Schnitt stellt dabei das Verhältnis von zwei Strecken dar. Berechnet wird er nach folgender Formel, wobei zu beachten ist, dass die längere Seite  $l$  als Zähler des Bruches zu nehmen ist und die kürzere  $b$  als Nenner:

$$\phi = l/b, \text{ wobei } l : b = (l + b) : l \text{ gilt.} \quad (4.2)$$

### 4.1.3 Distanz zur Mitte

Für dieses Feature nehmen wir an, dass die Mitte des Testbildes eine wesentliche Rolle für die Objektwahl darstellt und dass die Mitte eine gewisse Anziehungskraft auf die Augen ausübt.

Dieses Feature stellt die Distanz vom Mittelpunkt des Objektes zum Mittelpunkt des Testbildes dar.  $O_x$  und  $O_y$  sind dabei die Mittelpunkt-Koordinaten des Objektes, und  $TB_x$  und  $TB_y$  die Mittelpunkt-Koordinaten des Testbildes. Aufgrund der Wahl von 30x30 Pixel als Grösse der Testbilder sind  $TB_x$  und  $TB_y$  Konstanten vom Wert 15:

$$D_{mitte} = \sqrt{(O_x - TB_x)^2 + (O_y - TB_y)^2}. \quad (4.3)$$

Mit den gegebenen Konstanten ausgedrückt:

$$D_{mitte} = \sqrt{(O_x - 15)^2 + (O_y - 15)^2}. \quad (4.4)$$

### 4.1.4 Randkontrast

Hier gehen wir davon aus, dass der Kontrast zwischen einem Objekt und seinem Hintergrund eine wesentliche Rolle spielen muss. Betrachtet man als Beispiel in einer wolkenlosen Nacht die Sterne, so fällt dabei rasch auf, dass sich die helleren Sterne leichter und schneller finden lassen, als schwächer leuchtende. Der Randkontrast zeigt das Intensitätsverhältnis zwischen der Objektfarbe und den Farben der direkt an das Objekt angrenzenden Punkte auf. Auf der Abbildung 4.2 ist genauer abgebildet, wie der Rand ausgewählt wird.



Abbildung 4.2: Als Rand (r) wurde genau ein Ring mit einer Breite von einem Pixel gewählt.

Dabei ist  $r$  die Menge der angrenzenden Randpunkte. Das Objekt  $c$  ist auch eine Menge, aber es handelt sich dabei immer um dieselbe Farbe, somit

kann die Menge der Objektpunkte  $c$  vereinfacht als Objektfarbe  $O_c$  betrachtet werden. Für die Berechnung braucht es zwei Werte:  $B_{rand}$  stellt die durchschnittliche Helligkeit aller Punkte aus  $r$  dar und  $B_{objekt}$  die Helligkeit des Objekts. Damit erhält man:

$$K_{rand} = |B_{objekt} - B_{rand}|. \quad (4.5)$$

Für die Berechnung der Helligkeit haben wir die eingebaute Funktion `GetBrightness()` von C# [11] verwendet. Diese liefert Werte zwischen 0 und 1, wobei 0 = schwarz und 1 = weiss bedeutet.

#### 4.1.5 SSC-Stabilität

Der Sequenzielle-Superparamagnetische-Clustering Algorithmus (SSC) [12] ist ein besonders mächtiger und autonomer Clustering-Algorithmus. Dessen Funktionsweise kann anhand eines abstrakten Beispiels erklärt werden.

Man stelle sich einen Gletscher vor, der ins Meer "kalbt". Die Gletschermasse bricht dabei unter dem Einfluss der höheren Temperatur ins Meer längs Bruchflächen auf, bei der der Zusammenhalt der Moleküle besonders schwach ist. In unserem Fall sind die Pixel die Moleküle; ihr Zusammenhalt wird durch eine Distanzfunktion, welche beschreibt, wie weit sie in Lage und Farbe voneinander entfernt sind. Die Pixel in einem Objekt haben dementsprechend einen grossen Zusammenhalt und werden deshalb unter der Wirkung einer erhöhten Temperatur nur schwer aufbrechen. Über den Objektrand und innerhalb der umgebenden Menge ist der Zusammenhalt hingegen klein, sodass der "Objektblock" bestehen bleibt, wenn die Temperatur ansteigt, während sich die umgebende Menge in Einzelpixel auflöst. Der SSC-Algorithmus bildet diesen physikalischen Prozess nach. Die Besonderheit ist, dass die Objekt dabei sequenziell identifiziert werden: In einem ersten Durchlauf wird das stabilste Objekt festgestellt und aus der Menge entfernt. Mit der Restmenge wird dann der Prozess wiederholt. Damit stellt man sicher, dass die Erkennung von "schwächeren" Clustern nicht durch die Anwesenheit von "stärkeren" Clustern verunmöglicht wird. Der Algorithmus liefert dabei ein Mass für die "Stärke" eines Clusters, "Stabilität" genannt.

Mit diesem Stabilität-Feature erhoffen wir uns, dass die in den Testbildern versteckten Objekte gewisse Stabilitätsmerkmale aufweisen, die für die Reihenfolge, in der die Probanden die Objekte erkennen, massgebend sind. Dieses Feature ist mitunter eines der spannendsten überhaupt, da es darauf hinweisen würde, dass der Mensch sequenziell Bilder ähnlich wie der SSC "clustert".

Der SSC steht in unserer Arbeit als fertiges Programm (unter Linux) zur Verfügung. Um aber damit unsere Testbilder clustern zu können, mussten die Bilder in ein spezielles Format umgewandelt werden. Das Programm verlangt als Eingabe eine Matrix, in welcher jeder Punkt mit jedem über ein Distanzmass verbunden ist. Für unsere Testbilder von 30x30 Pixel bedeutet das eine Matrix von 900x900 Werten.

Zur Vereinfachung wird hier die Aufbereitung der Eingabe mit einem Bild von nur vier Bildpunkten erläutert. Jeder Bildpunkt dieses Bildes wird als Buchstabe (a-d) dargestellt.

a	b
c	d

Abbildung 4.3: Ausgangsdaten Bild.

Aus der der Tabelle 4.3 wird nun eine Matrix wie in Abbildung 4.4 aufgebaut, welche die Beziehung eines jeden Punktes mit allen anderen aufzeigt.

d(a,a)	d(a,b)	d(a,c)	d(a,d)
d(b,a)	d(b,b)	d(b,c)	d(b,d)
d(c,a)	d(c,b)	d(c,c)	d(c,d)
d(d,a)	d(d,b)	d(d,c)	d(d,d)

Abbildung 4.4: Matrix-Eingabe für den SSC in aufbereiteter Form.

Wie die Abbildung 4.4 zeigt, ist die Distanzmassfunktion  $d(p_1, p_2)$  noch anzuwenden. Für unsere Bilder müssen dabei folgende Werte einfließen:

1. X-Position
2. Y-Position
3. Rot-Wert (0-255)
4. Grün-Wert (0-255)
5. Blau-Wert (0-255)



Die Berechnung der Distanz besteht aus drei Schritten. Zuerst wird die euklidische Distanz  $D_{eukl}$  zwischen den beiden Punkten ermittelt. Dann benötigen wir den Farbwertunterschied  $U_{farbe}$ , der beiden Punkte. Hier fließen alle drei Farbwerte (in C#: *RGB*-Werte für die Farbtensitäten von rot, grün und blau) in die Berechnung ein, da eines unserer Testbilder farbige Objekte enthält. Bei Graustufen kann eine Vereinfachung in der Berechnung gemacht werden, indem einer dieser drei Farbwerte stellvertretend als Intensität für alle gewählt wird, weil die drei Farbtensitäten bei Grautönen die gleichen *RGB*-Werte aufweisen. Zuletzt benötigen wir noch ein frei wählbares  $\alpha$ , um  $D_{eukl}$  mit  $U_{farbe}$  gewichten zu können:

$$D_{eukl}(a, b) = \sqrt{(a.x - b.x)^2 + (a.y - b.y)^2}, \quad (4.6)$$

$$U_{farbe}(a, b) = \frac{|a.rot - b.rot| + |a.grün - b.grün| + |a.blau - b.blau|}{3}, \quad (4.7)$$

$$d(a, b) = D_{eukl}(a, b) + \alpha * U_{farbe}(a, b). \quad (4.8)$$

Nach eingehenden Tests mit dem SSC mussten wir feststellen, dass wir so nicht zu den gewünschten Ergebnissen kamen. Deshalb schien es sinnvoller, nur die  $k$ -nächsten Nachbarn rund um einen Punkt herum eine Distanz  $d$  zu geben und alle weiter entfernten Punkte mit  $\infty$  (respektive in C# mit `double.MaxValue`) zu gewichten. In der Abbildung 4.5 ist das Problem grafisch dargestellt. Man kann sich das anhand einer Zwiebel und ihren Schalen vorstellen. Der Vorteil ist, dass der SSC eine schnellere Distanz-Abschwächung erhält, und somit zusammenhängende Objekte besser erkennt.

In unseren Versuchen stellte sich heraus, dass  $k = 4$  die optimale Anzahl Ringe ist. Dadurch liessen sich die Objekte am besten finden.

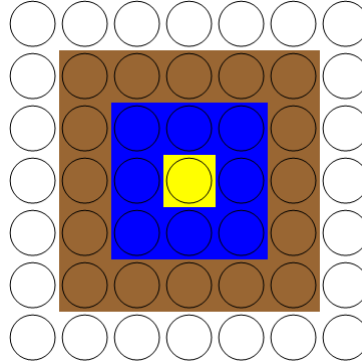


Abbildung 4.5: Zwiebelprinzip: Nur die  $k$ -nächsten Zwiebelschalen (Ringe) werden mit einer Distanz  $d$  versehen. Hier der blaue Ring ( $k = 1$ ) und der braune ( $k = 2$ ) um den gewählten Punkt (gelb). Die übrigen Punkte (weiss) bekommen als Wert  $\infty$ .

Das Tool des SSC liefert als Ergebnis des Clusterings immer ein Baum mit gefundenen Clustern, sowie eine Liste von Zeilennummern für zusammenhängende Cluster. Der Baum ist in der Abbildung 4.6 dargestellt. Das Testbild, welches geclustert wurde, ist in der der Abbildung 4.7 dargestellt. In diesem Baum ist klar zu erkennen, dass die stabilen Cluster (blau) sich von den instabilen Datenpunkten des Restes (grau) abheben. In der Abbildung 4.7 sind die erkannten Objekte direkt in das Testbild eingezeichnet.

Die Liste mit den Zeilennummern muss noch zurück transformiert werden, um wieder die Koordinaten  $x$  und  $y$  im Bild zu erhalten. Dies kann über die folgenden Formeln (in Mathematica) erreicht werden:

$$Coord(x) = If[Mod[x, 30] == 0, 30, Mod[x, 30]]. \quad (4.9)$$

$$Coord(y) = Ceiling[\frac{y}{30}]. \quad (4.10)$$

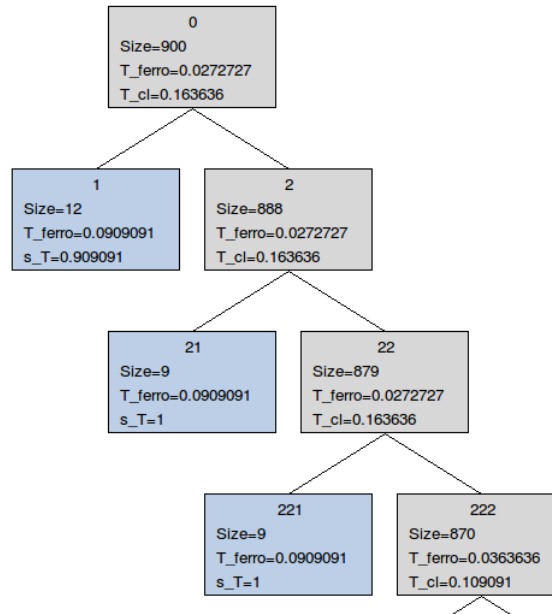


Abbildung 4.6: Ausschnitt aus dem Clustering-Baum des SSC Tools für das Testbild aus Abbildung 4.7. Die blauen Rechtecke markieren stabile Cluster, die grauen sind die noch instabilen Datenmengen.

### Featurewerte für ein Testbild

In den Tabellen der Abbildung 4.8 sind alle berechneten Features für das Testbild aus Abbildung 4.7 dargestellt. Im Anhang (Kapitel 7.2.3) sind die berechneten Features für alle Bilder angefügt.

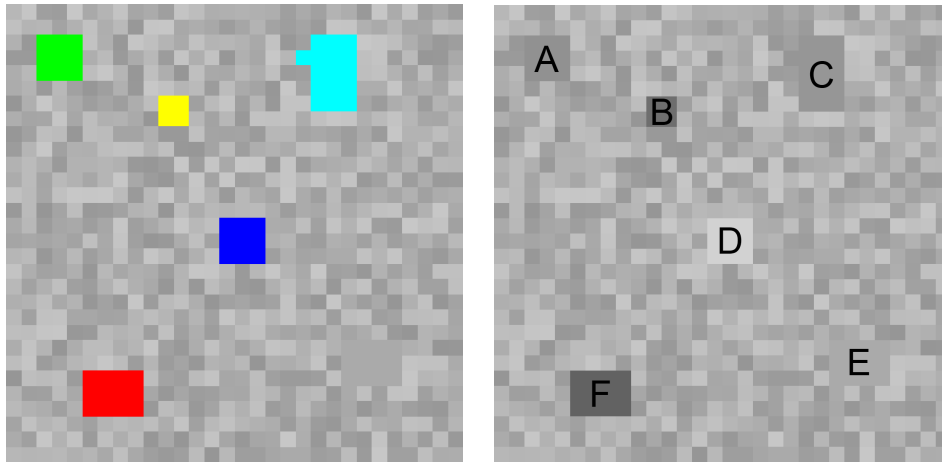


Abbildung 4.7: Links: gefundene Cluster für Testbild farbig markiert. Rechts: Das originale Testbild. Es wurden 5/6 aller Objekte gefunden, lediglich das schlechteste zu erkennende Objekt wurde nicht gefunden (Objekt E).

Objekt	$F$	$\phi$	$D_{mitte}$	$K_{rand}$	SSC Reihenfolge	SSC Stabilität
A	9	1	16.26	0.14	2	0.09
B	4	1	8.94	0.35	4	0.19
C	15	1.67	12.35	0.05	5	0.06
D	9	1	0.71	0.10	3	0.09
E	12	1.33	12.38	0.03	-	-
F	12	1.33	13.20	0.33	1	0.09

Abbildung 4.8: Berechnete Features für das Testbild aus Abbildung 4.7.

## 4.2 Modelle der ersten Objektwahl

In den Experimenten wird eine Sequenz von besuchten Objekten aufgezeichnet. Aus diesen soll das jetzt betrachtete Modell eine Vorhersage für das zuerst besuchte Objekt liefern. Aus diesem Grund müssen für jedes Testbild nur die zuerst besuchten Objekte betrachtet werden.

Die einzelnen Modelle werden nun mit den experimentell erhobenen Daten verglichen. Diese Daten sind in der Tabelle 4.9 abgebildet. Es wird für jedes Testbild das am häufigsten fixierte erste Objekt aufgelistet. Insgesamt fließen die Daten von 22 Probanden in diese Tabelle ein.

Testbild	häufigstes erstes Objekt	Anzahl Probanden
1	D	6
2	D	12
3	D	17
4	D	13
5	E	7
6	C	8
7	D	8
8	F	15

Abbildung 4.9: Tabelle mit den experimentell erhobenen Daten über das erste analysierte Objekt und die Anzahl der Probanden, welche dieses Objekt betrachtet haben.

Für die Vorhersage des ersten Objektes werden sieben Modelle betrachtet, welche mit ihren Vorhersagegenauigkeiten in Abbildung 4.10 aufgelistet sind. Es ist gut zu erkennen, dass sich das Modell mit der Distanz zur Mitte mit seiner Vorhersagequalität stark von den anderen abhebt.

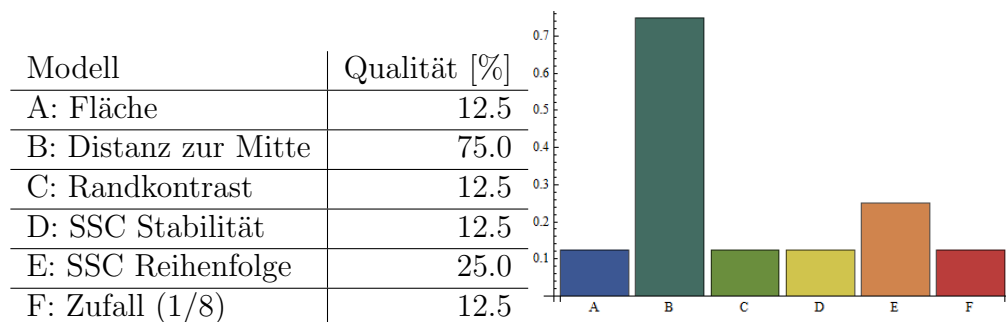


Abbildung 4.10: Links: Qualitäten aller Modelle zusammengefasst. Das "Distanz zur Mitte"-Modell hebt sich mit 75% klar von den anderen ab. Die meisten anderen Modelle haben eine Vorhersage-Qualität, welche gleich der des Zufallmodells F ist. Rechts: Grafische Darstellung der Tabelle.

### 4.2.1 "Distanz zur Mitte"-Modell

Dieses Modell begründet seine Vorhersage auf der Distanz zwischen den Objektmittelpunkten und der Bildmitte. Die Vorhersage lautet: Das Objekt, welches sich am nächsten in der Bildmitte befindet, wird von den meisten Probanden zuerst fixiert. Die Vorhersage dieses Modells ist mit 75% bedeutend höher als die der anderen Modelle, welche maximal 25% erreichen. Dies lässt darauf schliessen, dass die meisten Probanden bei jedem Testbild zuerst in die Mitte blicken, um ein neues Testbild nach interessanten Objekten zu durchsuchen. *Aus diesem Grund wird in diesem Kapitel nur dieses Modell betrachtet. Auf alle anderen Modelle wird im Anhang (Kapitel 7.4 "Datenübersicht: Modelle der ersten Objektwahl") genauer eingegangen.* In der folgenden Abbildung 4.11 ist die Auswertung für das Modell über alle Bilder aufgestellt.

Testbild	Vorhersage	Treffer
1	D	Ja
2	D	Ja
3	D	Ja
4	D	Ja
5	A	Nein
6	A	Nein
7	D	Ja
8	F	Ja
	<b>Quote:</b>	<b>75</b>

Abbildung 4.11: Vorhersagekraft des "Distanz zur Mitte"-Modells.

### 4.3 Modelle der zweiten Objektwahl

Diese Modelle benützen die selben Daten wie die Modelle im vorhergehenden Kapitel 4.2. Auch hier wird bei dem ersten Testbild das zuerst betrachtete Objekt für alle Probanden gelöscht. Nun untersuchen wir aber das zweite angeschaute Objekt. Dazu werden alle Modelle mit den experimentell erhobenen Daten verglichen. Diese Daten sind in der Abbildung 4.12 dargestellt. Auch hier sind die Daten von 22 Probanden eingeflossen. Alle Modelle aus der Abbildung 4.1 werden hier betrachtet. Am Schluss wird eine Zusammenfassung und unsere erstellte Interpretation über alle Modelle aufgezeigt.

Testbild	häufigstes zweites Objekt	Anzahl Probanden
1	B	7
2	E	8
3	A	9
4	F	6
5	A/C	7
6	A	7
7	D	6
8	C	11

Abbildung 4.12: Tabelle mit den Daten aus den Experimenten. Das zweite betrachtete Objekt sowie die Anzahl der Probanden, welche sich dieses Objekt angeschaut haben.

### 4.3.1 Flächenmodell

Das grösste Objekt wird laut Vorhersage dieses Modells als zweites betrachtet. Die Vorhersage-Richtigkeit für dieses Modell ist 25%. Die Abbildung 4.13 zeigt alle Vorhersagen auf.

Testbild	Vorhersage	Treffer
1	C	Nein
2	C	Nein
3	C	Nein
4	C	Nein
5	C	Ja
6	C	Nein
7	C	Nein
8	C	Ja
	<b>Quote:</b>	<b>25</b>

Abbildung 4.13: Vorhersagekraft des Flächenmodells.

### 4.3.2 "Goldener Schnitt"-Modell

Hier wird laut Modell das Objekt als zweites erkannt, bei welchem das Verhältnis der Seiten dem goldenen Schnitt am nächsten ist. Dieses Modell hat eine Vorhersagekraft von 25%. Die einzelnen Ergebnisse sind in der Abbildung 4.14 aufgelistet.

Testbild	Vorhersage	Treffer
1	C	Nein
2	C	Nein
3	C	Nein
4	C	Nein
5	C	Ja
6	C	Nein
7	C	Nein
8	C	Ja
	<b>Quote:</b>	<b>25</b>

Abbildung 4.14: Vorhersagekraft des "Goldenen Schnitt"-Modells.



### 4.3.3 "Distanz zur Mitte"-Modell

Dieses Modell prognostiziert, dass das Objekt mit dem kleinsten Abstand zur Bildmitte als zweites angeschaut wird. Die Quote dieses Modells ist 37.5%. In der folgenden Abbildung 4.15 ist die Auswertung für alle Bilder aufgestellt.

Testbild	Vorhersage	Treffer
1	D	Nein
2	D	Nein
3	D	Nein
4	D	Nein
5	C	Ja
6	A	Ja
7	D	Ja
8	F	Nein
	<b>Quote:</b>	<b>37.5</b>

Abbildung 4.15: Vorhersagekraft des "Distanz zur Mitte"-Modells.

### 4.3.4 Randkontrastmodell

Das Objekt, welches den stärksten Randkontrast hat, wird laut Aussage dieses Modells, als zweites betrachtet. Dieses Modell hat eine Qualität von 50% erreicht. Die Abbildung 4.16 zeigt die Vorhersagekraft dieses Modells.

Testbild	Vorhersage	Treffer
1	B	Ja
2	D	Nein
3	A	Ja
4	F	Ja
5	F	Nein
6	A	Ja
7	C	Nein
8	B	Nein
	<b>Quote:</b>	<b>50</b>

Abbildung 4.16: Vorhersagekraft des Randkontrastmodells.

### 4.3.5 SSC Stabilitäts-Modell

Für dieses Modell wird das Objekt mit der stärksten SSC-Stabilität als Kandidat vorgeschlagen. Dieses Modell hat eine Vorhersagekraft von 25% erreicht, welche in der Abbildung 4.17 ersichtlich ist.

Testbild	Vorhersage	Treffer
1	B	Ja
2	B	Nein
3	B	Nein
4	B	Nein
5	C	Ja
6	C	Nein
7	E	Nein
8	B	Nein
	<b>Quote:</b>	<b>25</b>

Abbildung 4.17: Vorhersagekraft des SSC Stabilität-Modells.

### 4.3.6 SSC Reihenfolge-Modell

Bei diesem Modell wird die Reihenfolge des SSC Clustering Prozesses zur Vorhersage verwendet. Auch hier wird das Objekt welches zuerst vom Clustering erkannt wird, von den meisten Probanden als zweites angesehen. Die Vorhersagekraft dieses Modells ist mit 25% genau gleich, wie bei der SSC Stabilität. In der Abbildung 4.18 sind die Vorhersagen dieses Modells dargestellt.

Testbild	Vorhersage	Treffer
1	F	Nein
2	E	Ja
3	D	Nein
4	A	Nein
5	C	Ja
6	C	Nein
7	E	Nein
8	D	Nein
	<b>Quote:</b>	<b>25</b>

Abbildung 4.18: Vorhersagekraft mittels der Reihenfolge des SSC Clusterings.

### 4.3.7 Ergebnisse zur zweiten Objektwahl und Interpretation

In der Abbildung 4.19 sind alle Ergebnisse für die einzelnen Modelle nochmals aufgelistet. Hier ist gut zu erkennen, dass es keinen Ausreisser nach oben gibt, wie bei der Vorhersage des ersten Objektes. Bei diesen Modellen ist jedoch spannend, dass alle eine höhere Vorhersage als der pure Zufall ( $1/8$ , 12.5%) liefern.

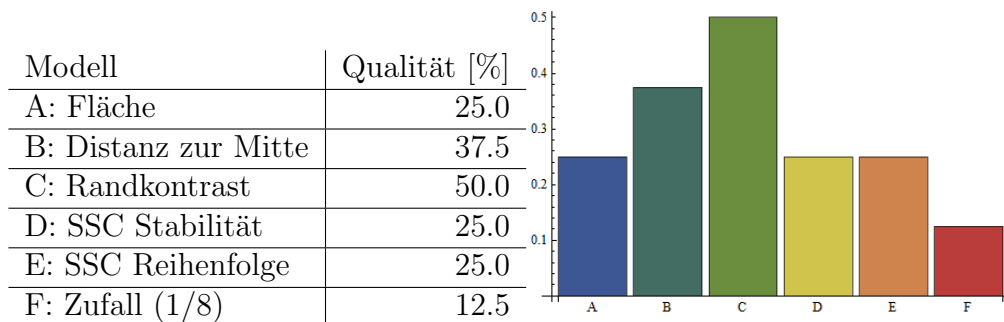


Abbildung 4.19: Die Qualitäten aller Modelle werden hier zusammengefasst. Das Randkontrast-Modell hebt sich mit 50% ein wenig von den anderen ab. Links: Qualitäten aller Modelle zusammengefasst. Das Randkontrast-Modell hebt sich mit 50% ein wenig von den anderen ab. Rechts: Als Diagramm dargestellt. Alle Modelle haben eine Vorhersagekraft, die grösser als die des Zufallsmodells F ist.

Es scheint plausibel, dass wir durch Kombination dieser einzelnen Kriterien zu einem besseren Modell kommen können. Wir haben diese Möglichkeit aus Zeitgründen jedoch nicht weiter verfolgt. Wir haben uns stattdessen auf das Auffinden von möglichen Mustern in der Objektwahrnehmungsreihenfolge konzentriert.

## 4.4 Modelle von Mustern der Objektreihenfolge

In diesem Unterkapitel werden alle Sequenzen der besuchten Objekte untersucht. Das Ziel dieser Untersuchungen ist es, wiederkehrende Muster in den Sequenzen zu finden. Wenn solche wiederkehrende Sequenzen gefunden werden, weist dies darauf hin, dass die Beobachtungssequenzen nicht rein zufällig sind, sondern einem bestimmten Vorgehen entsprechen.

Zuerst werden die Sequenzen auf gewählte Muster hin untersucht. Danach werden die Sequenzen auf Gemeinsamkeiten untereinander durchsucht.

In der Abbildung 4.20 werden alle gesuchten Muster aufgelistet und kurz beschrieben. Die Muster geben darüber Auskunft, ob bestimmte Sequenzen in der gleichen Reihenfolge wieder vorkommen.

Muster	Beschreibung
AX...X	Das erste Objekt kommt in der Sequenz nie mehr vor.
AXA	Das erste Objekt ist auch das dritte Objekt.
AXYA	Das erste Objekt ist auch das vierte Objekt.
...AXA...	Das erste Objekt ist auch das dritte Objekt, diesmal irgendwo in der Sequenz.
...AXYA...	Das erste Objekt ist auch das vierte Objekt, diesmal irgendwo in der Sequenz.
AB...AB...	Die ersten beiden Objekte kommen irgendwo in der Sequenz wieder vor.
...AB...AB...	Zwei Objekte kommen irgendwo in der Sequenz wieder vor.

Abbildung 4.20: Gesuchte Muster, wobei A und B bestimmte Objekte sind und X und Y Platzhalter für andere Objekte.

Die Muster werden mittels einfachen regulären Ausdrücken in den Sequenzen gesucht. In der Abbildung 4.21 sind alle regulären Ausdrücke für ein Beispiel aufgelistet. Zu beachten ist, dass diese Ausdrücke bei der Erkennung allgemein gehalten werden und dieses Beispiel nur zur Veranschaulichung dient.

Muster	Regex	Beschreibung
AX...X	$^A[^A]*\$$	Das Objekt A kommt in der Sequenz als erstes vor und danach nie mehr.
AXA	$^B[^B]\{1\}B$	Das Objekt B ist das erste sowie das dritte Objekt.
AXYA	$^C[^C]\{2\}C$	Das Objekt C ist das erste und das vierte Objekt.
...AXA...	$D[^D]\{1\}D$	Das Objekt D, gefolgt von einem anderen Objekt, dann wieder das Objekt D; irgendwo in der Sequenz.
...AXYA...	$E[^E]\{2\}E$	Das Objekt E, gefolgt von zwei anderen Objekten, dann wieder das Objekt E; irgendwo in der Sequenz.
AB...AB...	$^FA.*FA$	Die am Anfang der Sequenz stehende Objektfolge FA kommt irgendwo wieder vor.
...AB...AB...	$FB.*FB$	Die Objektfolge FB kommt irgendwo wieder vor.

Abbildung 4.21: Hier werden die regulären Ausdrücke, um ein bestimmtes Muster zu suchen, abgebildet. Um die Übersicht zu bewahren wird der Ausdruck immer durch ein konkretes Beispiel erläutert.

Alle vorhandenen Sequenzen werden mit diesen regulären Ausdrücken untersucht. Untersucht werden nur die Sequenzen, welche mindestens drei besuchte Objekte beinhalten. Alle anderen Sequenzen werden von dieser Untersuchung ausgeschlossen. Dies ergibt total 189 zu untersuchende Sequenzen. In der Abbildung 4.22 werden die Anzahl Übereinstimmungen, sowie der prozentuale Anteil des Musters an der Summe aller Treffer, aufgelistet. Interessant bei diesen Ergebnissen ist, dass erstaunlich viele solche Muster in den Sequenzen auftreten. Auch kehren die Probanden vermehrt wieder zu schon gesehenen Objekten zurück.

Muster	Treffer	Prozent
AX...X	38	7.1%
AXA	45	8.4%
AXYA	24	4.5%
...AXA...	195	36.4%
...AXYA...	90	16.8%
AB...AB...	44	8.2%
...AB...AB...	99	18.5%

Abbildung 4.22: Anzahl Treffer für ein Muster und deren prozentualer Anteil an der Summe aller Treffer.

Das Zurückkehren der Probanden ist auch in der kumulierten Markov-Übergangsmatrix pro Bild sehr gut zu erkennen. In der Abbildung 4.23 ist die Markov-Übergangsmatrix für das Testbild 1 dargestellt. Hier ist gut zu erkennen, dass die meisten Probanden vom Objekt B, mit der grössten Häufigkeit, auf das Objekt D schauen (siehe 2. Zeile, 4. Spalte). Beim Objekt D wechseln die meisten Probanden wieder zurück zum Objekt B (siehe 4. Zeile, 2. Spalte). Solche einfachen Schleifen sind in vielen Matrizen zu erkennen.

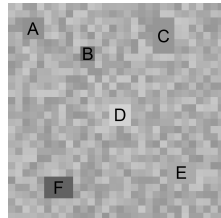
$$\begin{pmatrix} 0 & 6 & 5 & 2 & 1 & 2 \\ 8 & 0 & 10 & \mathbf{15} & 0 & 1 \\ 2 & 4 & 0 & 12 & 5 & 1 \\ 2 & \mathbf{18} & 6 & 0 & 9 & 7 \\ 0 & 2 & 3 & 4 & 0 & 5 \\ 3 & 4 & 2 & 15 & 1 & 0 \end{pmatrix}$$

Abbildung 4.23: Kumulierte Markov-Übergangsmatrix über alle Probanden für das Testbild 1. Eine Schleife von B nach D und zurück ist in dieser Markov-Übergangsmatrix fett hervorgehoben.

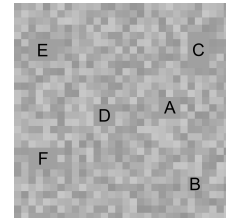
Durch diese Untersuchung zeigte sich, dass die Probanden nach bestimmten Mustern bei der Objekterkennung vorgehen. Sie kehren vermehrt zu den ihnen bekannten Objekten zurück. Um nun weitere Muster der einzelnen Probanden zu entdecken, werden die Sequenzen im kommenden Teil auf Gemeinsamkeiten untersucht. Die Gleichheit der Sequenzen wird mit der `SmithWatermanSimilarity[]` - kurz SW - [13] aus Mathematica verglichen. Des Weiteren wird die längste gemeinsame Sequenz - kurz LS - zwischen zwei Sequenzen gesucht. Die Ergebnisse sind in der Tabelle 4.24 abgebildet. Auch bei diesen Ergebnissen zeigt sich, dass die verschiedenen Probanden mehrere Objekte in der gleichen Reihenfolge betrachtet haben. Diese gleichen Sequenzen sind teilweise erstaunlich lange, was die Wahrscheinlichkeit, dass

dies purer Zufall ist, sehr stark reduziert. Auch bei diesen Sequenzen sind einige Schleifen zu erkennen, bei denen die Probanden wieder zu einem bereits betrachteten Objekt wechseln. Bei diesen Sequenzen ist auch zu erkennen, dass immer wieder das Objekt in der Mitte fixiert wird.

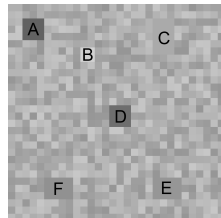
Testbild 1  
SW max: 6  
SW  $\phi$ : 2.3  
LS: FDBDBD  
LS  $\phi$ : 2.4



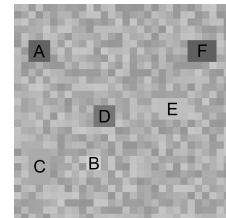
Testbild 2  
SW max: 8  
SW  $\phi$ : 2.7  
LS: DADECD  
LS  $\phi$ : 2.8



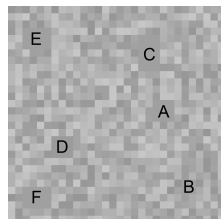
Testbild 3  
SW max: 8  
SW  $\phi$ : 2.2  
LS: DACDEFBC  
LS  $\phi$ : 2.3



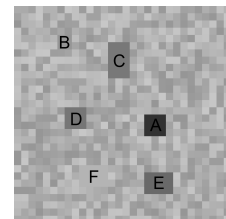
Testbild 4  
SW max: 6  
SW  $\phi$ : 2.2  
LS: DEF AFE  
LS  $\phi$ : 2.4



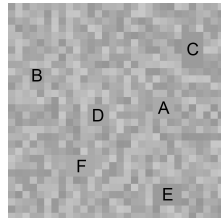
Testbild 5  
SW max: 4  
SW  $\phi$ : 2.0  
LS: ACED  
LS  $\phi$ : 2.3



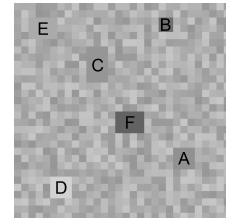
Testbild 6  
SW max: 8  
SW  $\phi$ : 2.6  
LS: DBCBCA  
LS  $\phi$ : 2.6



Testbild 7  
SW max: 6  
SW  $\phi$ : 2.4  
LS: FACBDA  
LS  $\phi$ : 2.5



Testbild 8  
SW max: 7  
SW  $\phi$ : 2.8  
LS: FCECB  
LS  $\phi$ : 2.7



Testbild 9  
SW max: 9  
SW  $\phi$ : 2.8  
LS: ADCBDEFE  
LS  $\phi$ : 3.0

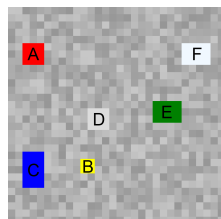


Abbildung 4.24: In dieser Abbildung wird die Smith-Waterman Ähnlichkeit (SW) und die längste gemeinsame Sequenz (LS) zwischen zwei Sequenzen. Die Durchschnittswerte für SW und LS über alle Sequenzen sind für jedes Bild aufgelistet.

## 4.5 Vergleich mit Surrogate-Daten

Damit wir unsere Modelle auf deren statistische Relevanz überprüfen können, haben wir Ersatzdaten - sogenannte Surrogate-Daten [14] - erzeugt. Zu diesem Zweck haben wir sämtliche erhobene Sequenzen aus dem Experiment, für jedes Testbild und jeden Probanden, 40 mal permutiert. Aus diesen Surrogate-Sequenzen erstellten wir neue Markov-Übergangsmatrizen. Diese können mit den aus den Experimentdaten erstellten Markov-Übergangsmatrizen (siehe Kapitel 3.4.3) verglichen werden. In der Abbildung 4.25 sind die Übergangswahrscheinlichkeiten aus den Surrogate-Daten und aus den Experimentdaten für das Testbild 1 dargestellt.

$$\begin{array}{lcl}
 \text{Experimentdaten:} & & \begin{pmatrix} 0. & 0.375 & 0.313 & 0.125 & 0.063 & 0.125 \\ \mathbf{0.236} & 0. & 0.294 & 0.441 & 0. & 0.029 \\ 0.083 & 0.167 & 0. & 0.5 & 0.208 & 0.042 \\ 0.048 & 0.429 & 0.143 & 0. & 0.214 & 0.167 \\ 0. & 0.143 & 0.214 & 0.286 & 0. & \mathbf{0.357} \\ 0.12 & 0.16 & 0.08 & 0.6 & 0.04 & 0. \end{pmatrix} \\
 \\ 
 \text{Surrogate-Daten:} & & \begin{pmatrix} 0. & 0.264 & 0.182 & 0.285 & 0.090 & 0.179 \\ \mathbf{0.123} & 0. & 0.188 & 0.409 & 0.119 & 0.159 \\ 0.127 & 0.246 & 0. & 0.320 & 0.114 & 0.193 \\ 0.097 & 0.352 & 0.203 & 0. & 0.134 & 0.214 \\ 0.113 & 0.236 & 0.194 & 0.305 & 0. & \mathbf{0.152} \\ 0.112 & 0.235 & 0.182 & 0.381 & 0.091 & 0. \end{pmatrix}
 \end{array}$$

Abbildung 4.25: Oben: Markov-Matrix des Experimentes, unten: Markov-Matrix der Surrogate-Daten (Übergangswahrscheinlichkeiten). Die beiden Matrizen haben markante Unterschiede (fett hervorgehoben Einträge der Matrizen). Die durchschnittliche Abweichung beträgt 79.55%.

Wir haben das Verhältnis der beiden Markov-Matrizen zueinander als  $Markov_{\text{verhältnis}} = \frac{Markov_{\text{experiment}}}{Markov_{\text{surrogate}}}$  definiert. In der Abbildung 4.26 sind die Verhältnisse der Matrizen als Matrix  $Markov_{\text{verhältnis}}$  sowie auch grafisch dargestellt. Dabei zeigt sich, dass die beiden Matrizen sich minimal um 0% und maximal um 235.81% unterscheiden (fett markiert).



$$Markov_{\text{verhältnis}} : \begin{pmatrix} 1 & 1.421 & 1.713 & 0.439 & 0.698 & 0.698 \\ 1.907 & 1 & 1.560 & 1.077 & 0 & 0.185 \\ 0.659 & 0.678 & 1 & 1.563 & 1.824 & 0.215 \\ 0.490 & 1.217 & 0.704 & 1 & 1.605 & 0.778 \\ \mathbf{0} & 0.606 & 1.105 & 0.936 & 1 & \mathbf{2.3518} \\ 1.072 & 0.682 & 0.439 & 1.576 & 0.441 & 1 \end{pmatrix}$$

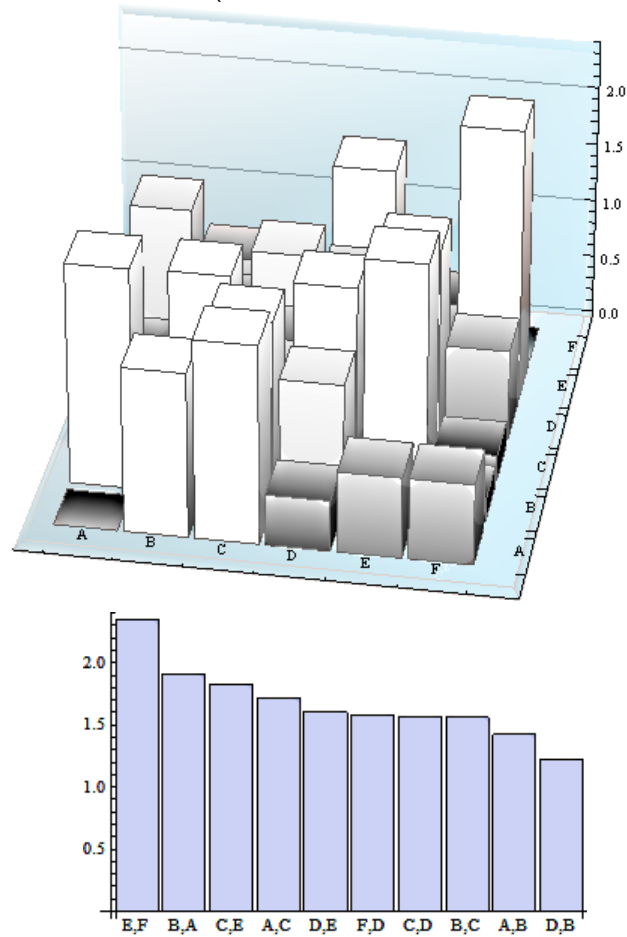


Abbildung 4.26: Oben: Die Verhältnismatrix  $Markov_{\text{verhältnis}}$ ; Mitte: Die Verhältnismatrix grafisch dargestellt (die Diagonalelemente wurden auf 0 gesetzt; Graufärbung ohne Belang). Unten: Die Verteilung der Unterschiede absteigend sortiert. Die Verteilung ist auf die zehn unterschiedlichsten Punkte beschränkt. Die Surrogate-Methode zeigt klar auf, dass es sich bei den Experimenten nicht um rein zufällige Daten handelt.

Im Kapitel 4.4 haben wir verschiedene Muster in den Sequenzen untersucht. Nach diesen Mustern haben wir ebenfalls in unseren Surrogate-Daten gesucht, um die Unterschiede zu zeigen. In der Abbildung 4.27 sind die Ergeb-

nisse aufgelistet. Dabei zeigen die beiden Muster  $AX...X$  und  $...AB...AB...$  die grössten Abweichungen zu den experimentell gemessenen Daten auf. Diese beiden Muster sind Kandidaten für weiterführende Modelle, welche aber nicht Bestandteil dieser Arbeit sind. Aufgrund der hohen Differenzen zwischen den Treffern in den Experiment- und Surrogate-Daten unterliegen diese Muster nicht dem reinen Zufall. Dies könnte auf ein sequentielles Abarbeiten bei der Wahrnehmung von Objekten hindeuten.

Muster	Experiment		Surrogate	
	Treffer	Prozent	Treffer	Prozent
$AX...X$	38	<b>7.1%</b>	2853	<b>15.3%</b>
$AXA$	45	8.4%	1330	7.2%
$AXYA$	24	4.5%	1041	5.6%
$...AXA...$	195	36.4%	5864	31.5%
$...AXYA...$	90	16.8%	3944	21.2%
$AB...AB...$	44	8.2%	991	5.3%
$...AB...AB...$	99	<b>18.5%</b>	2576	<b>13.9%</b>

Abbildung 4.27: Anzahl Treffer und deren prozentualer Anteil an der Summe aller Treffer für die Experiment-Daten sowie die Surrogate-Daten. Die grössten Unterschiede zwischen den Experiment- und Surrogate-Messungen sind fett markiert.

Ebenfalls im Kapitel 4.4 haben wir die Sequenzen untereinander auf Übereinstimmungen verglichen. Dasselbe haben wir auch mit den Surrogate-Daten durchgeführt. Dabei zeigt sich deutlich, dass auch in den Surrogate-Daten längere gemeinsame Sequenzen entstehen und nicht nur in den Experimentdaten. In der Abbildung 4.28 ist eine Aufstellung mit Kennzahlen aus den Experiment- sowie aus den Surrogate-Daten ersichtlich. Dabei ist der durchschnittliche Smith-Waterman-Wert der experimentellen Daten um 0.488 Zeichen länger als jener der Surrogate-Daten (fett markiert). Dasselbe ist auch bei den längsten gemeinsamen Sequenzen zu beobachten. Bei den experimentellen Daten liegt der Durchschnitt um 0.456 Zeichen höher als derjenige der Surrogate-Daten (kursiv markiert). Dieser grosse Unterschied zwischen den experimentellen- und den Surrogate-Daten zeigt auf, dass die Testpersonen ein gemeinsames Muster bei der Reihenfolge der Objektwahrnehmung folgen, und diese gefundenen Reihenfolgen nicht durch Zufall entstanden sind.

Testbild	Experiment				Surrogate			
	SW		LS		SW		LS	
	max	⊘	max	⊘	max	⊘	max	⊘
1	6	<b>2.3</b>	6	<i>2.4</i>	7	<b>1.8</b>	8	<i>2.0</i>
2	8	<b>2.7</b>	8	<i>2.8</i>	8	<b>2.1</b>	8	<i>2.2</i>
3	8	<b>2.2</b>	8	<i>2.4</i>	7	<b>1.8</b>	7	<i>2.0</i>
4	6	<b>2.2</b>	6	<i>2.4</i>	9	<b>1.9</b>	9	<i>2.0</i>
5	4	<b>2.0</b>	4	<i>2.3</i>	7	<b>1.7</b>	7	<i>1.9</i>
6	8	<b>2.6</b>	6	<i>2.6</i>	9	<b>2.0</b>	8	<i>2.2</i>
7	6	<b>2.4</b>	6	<i>2.5</i>	11	<b>1.9</b>	11	<i>2.1</i>
8	7	<b>2.8</b>	5	<i>2.7</i>	11	<b>2.2</b>	10	<i>2.3</i>
9	9	<b>2.8</b>	8	<i>3.0</i>	9	<b>2.2</b>	9	<i>2.3</i>

Abbildung 4.28: In dieser Tabelle wird die Smith-Waterman Ähnlichkeit (SW) und die längste gemeinsame Sequenz (LS) zwischen zwei Sequenzen sowie deren Durchschnittswerte über alle Sequenzen für jedes Bild aufgelistet. Diese Auflistung ist für die Experiment- und Surrogate-Daten erstellt.

Mit diesem ausführlichen Zahlenmaterial haben wir die Basis für ein detailliertes Modell der Objekterfassung gelegt. Die Ausarbeitung eines solchen Modelles ist nicht Gegenstand dieser Arbeit.

# 5

## Schlussfolgerung

### 5.1 Vergleich mit der Pilotarbeit

Die spannende Pilotarbeit, welche an der ETH entstanden ist, zeigt einige interessante Ergebnisse. Wir haben versucht, diese möglichst gut nachzuvollziehen. Eine Hauptschwierigkeit, auf die wir dabei gestossen sind, war, dass oftmals präzise Herleitungen fehlten, sodass wir viele mögliche Varianten zuerst überprüfen mussten, um eine klare Aussage zu erhalten. Beispielsweise enthielten vereinzelte Testbilder versteckte Objekte, welche sich berührten. Wenn man ein Akzeptanzgebiet von zwei Pixel rund um ein Objekt erlaubt, dürfte dies nicht gemacht werden, da sonst Objekttreffer nicht eindeutig identifiziert werden können. In der Abbildung 5.1 ist ein solches Testbild aus der Pilotarbeit ersichtlich.

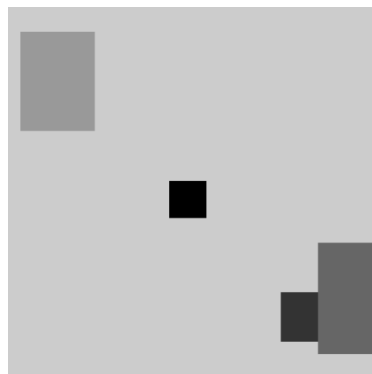


Abbildung 5.1: Eines der verwendeten Testbilder aus der Pilotarbeit der ETH Zürich. Unten rechts im Bild sind die problematischen Objekte dargestellt. Diese berühren sich, was aber für eine korrekte Objekttreffer-Zuordnung nicht gültig ist.

Auch bei den untersuchten Features fehlten Angaben über deren Berechnung. Dies führte dazu, dass wir nicht immer sicher sein konnten, exakt dasselbe untersucht zu haben.

## 5.2 Neue Ergebnisse

Wir konnten zeigen, dass die Probanden zuerst die Mitte der Testbilder bevorzugen. Dies liegt möglicherweise der ganz normalen Gegebenheit zugrunde, dass man sich an der Mitte eines Bildes orientiert. Der SSC war neben der Distanz zur Mitte das einzige Feature, welches eine höhere Trefferquote als der Zufall aufwies. Die Abbildung 5.2 zeigt eine 3D-Heatmap über alle Testbilder zusammengefasst, darin erkennt man die Bildmitte als heissesten -höchsten-Punkt.

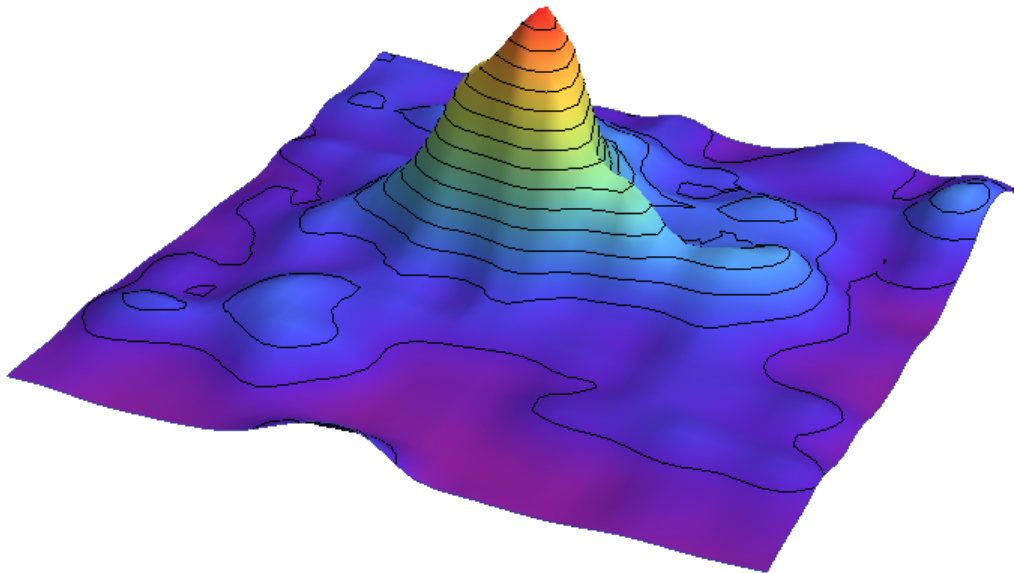


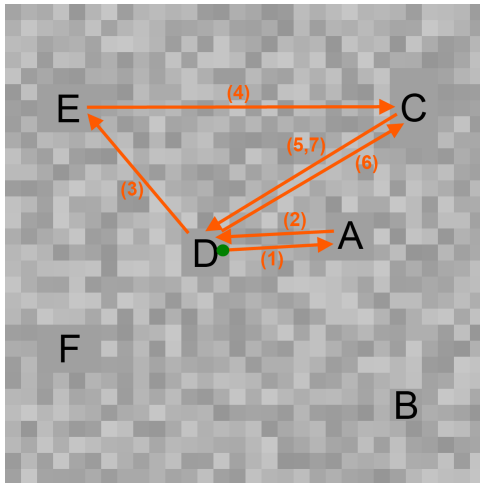
Abbildung 5.2: 3D-Heatmap über alle Testbilder hinweg. Die Bildmitte ist mit Abstand der längste betrachtete Bereich.

Bei der zweiten Objektwahl gibt es kein Modell, welches sich klar von den anderen abhebt. Sämtliche Modelle haben eine stärkere Vorhersagekraft als der Zufall. Das Randkontrastmodell lieferte mit einer Vorhersage von 50% das beste Ergebnis. Trotzdem bleibt die Distanz zur Mitte auch hier noch ein relevantes Feature. Eine Kombination der Features müsste hier zu einem besseren Modell führen können.

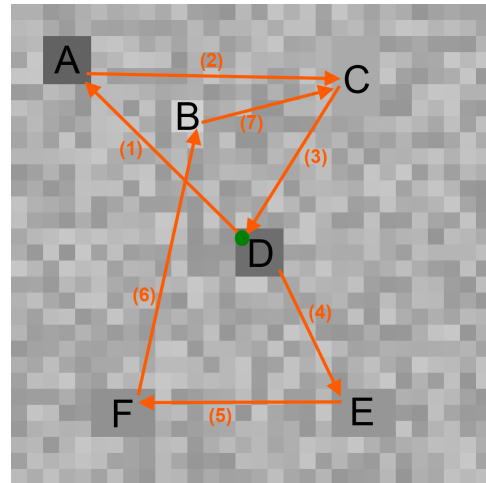
Das Muster ...AXA... kommt in unseren Untersuchungen am häufigsten vor. Der Proband kehrt vermehrt auf bereits besuchte Objekte zurück. Die Surrogate-Daten untermauern diese Beobachtung ebenfalls. Dies deutet auf ein spezielles sequenzielles Abarbeiten der Objektwahrnehmung hin, bei der der Proband nach bestimmten Regeln in der Erkennung von Objekten vorgeht. Die Untersuchung der längsten gemeinsamen Objektreihenfolgen zeigt, dass die längsten gemeinsamen Objektreihenfolgen beachtliche Längen aufweisen. Diese sind mit einer Länge von acht gleichen Objekten auf den Testbildern 2, 3 und 9 vorhanden. Das spannende an diesen Gemeinsamkeiten ist, dass bei jedem Bild zwei unterschiedliche Probanden die längste gemeinsame Sequenz aufwiesen. Die Reihenfolge für Testbild 2 ist DADECDCD, für das Testbild 3 DACDEFBC und für das Testbild 9 ist sie ADCBDEFE. Alle Testbilder und deren Sequenzen sind in der Abbildung 5.3 dargestellt. In den Surrogate-Daten lassen sich ähnlich lange Sequenzen finden, jedoch konnten wir zeigen, dass deren durchschnittliche Sequenzlänge bedeutend kleiner ist, als diejenige der experimentellen Daten. Dies bedeutet, dass es unter den Probanden längere Gemeinsamkeiten der Betrachtungsreihenfolge gibt, als zufällig zu erwarten wäre.

Experimentelle Markov-Übergangsmatrizen besitzen eine stark verschiedene Struktur im Vergleich zu Übergangsmatrizen, welche aus Surrogate-Daten erhalten werden. Die durchschnittliche Abweichung beider Matrizen ist mit  $\approx 80\%$  beachtlich hoch, wobei die stärkste Diskrepanz den Faktor 2.35 aufweist.

Testbild 2: DADECD



Testbild 3: DACDEFBC



Testbild 9: ADCBDEFE

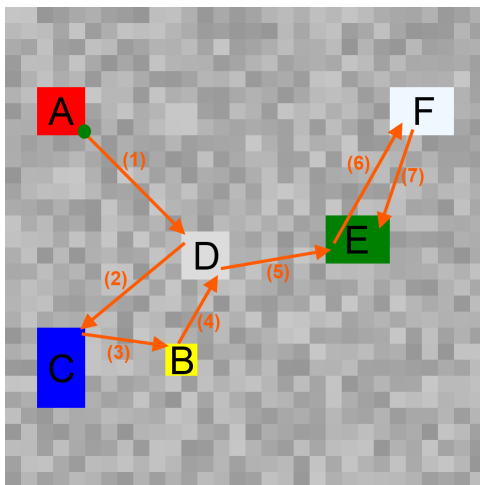


Abbildung 5.3: Testbilder mit den längsten besuchten gemeinsamen Objektsequenzen, welche von zwei Probanden betrachtet wurde. Der Start ist mit einem grünen Punkt markiert.

## 6

# Ausblick

Unsere Modelle zur Objektwahl beschränkten sich auf die ersten und zweiten betrachteten Objekte und das Auftreten von Mustern in der Objekt-Wahrnehmungsreihenfolge. Die Dauer des Verweilens auf einem Objekt wurde nicht genauer analysiert. Dies könnte aber durchaus sehr spannend sein, da eventuell gewisse Features, wie der goldene Schnitt, relevant werden.

Es gibt noch keine spezifischen Analysen, welche den Unterschied der Objektwahlmodelle für Testbilder in Grautonfärbungen mit dem farbigen Testbild 9 vergleichen. Vielleicht stellen sich dann gewisse Features als wichtiger oder weniger wichtig heraus. Insbesondere wäre hier das Feature Randkontrast interessant, da beim farbigen Testbild die Anzahl der Probanden, welche alle Objekte erkannt haben, am höchsten war. Dazu müsste beachtet werden, dass die Farbwahrnehmung durch den Menschen nicht für alle Farben gleich gut ist. Es ist zum Beispiel bekannt, dass die Farbe blau, aufgrund physikalischer Eigenschaften des Auges [15], nicht scharf gesehen werden kann.

Ein Modell zur Vorhersage der Reihenfolge, anhand der Markov-Übergangsmatrix, fehlt noch gänzlich. Ein solches Modell wäre sicher einer der nächsten Schritte, die es zu tun gilt. Dadurch könnte ein möglichst präzises Modell über den Verlauf der Objektbetrachtungsreihenfolge gewonnen werden. Ein solches Modell könnte ebenfalls um die Erkenntnisse, aus der Untersuchung der wiederkehrenden Muster, erweitert werden.



# 7

## Anhang

### 7.1 Persönliche Berichte

#### 7.1.1 Jonathan Stolz

Die Herausforderung, Objekte zu erkennen, erscheint für einen Menschen als nicht wirklich komplexe Aufgabe. Bei kognitiven Systemen der künstliche Intelligenz ist dies jedoch eine Problemstellung, welche bisher nicht erfolgreich gelöst werden konnte. Mit dieser Arbeit wurde mit bewusst, dass das Erkennen von Objekten eines einzelnen Menschen nicht rein zufällig ist, sondern bestimmten Regeln folgt. Durch die verschiedenen Inputs von Herrn Prof. Dr. Stoop wurden wir immer wieder motiviert, diese Prozesse noch genauer zu untersuchen. Es war eine sehr lehrreiche und spannende Arbeit.

#### 7.1.2 Thomas Jutzi

Als Bachelorarbeit habe mir immer gewünscht, etwas nicht ganz alltägliches zu untersuchen. Genau dieser Wunsch wurde mir mit dieser Arbeit erfüllt. Das menschliche Hirn ist und bleibt äusserst faszinierend. Nun konnte ich einen kleinen Einblick auf dessen Verarbeitung von visueller Wahrnehmung erhalten, das war grandios. Anfänglich konnte ich mir die Komplexität unseres Experiments und dessen Ergebnisse kaum vorstellen. Genau das machte dies aber spannend, mitunter auch aufgrund der fortwährend motivierenden Art von Herrn Prof. Dr. Stoop.

## 7.2 Wichtige Begriffe

### 7.2.1 Sakkade

Bei Sakkaden [16] handelt es sich um schnelle, ruckartige Rückwärtsbewegungen der Augäpfel nach einer Augenbewegung, bei der ein Gegenstand fixiert wird. Man spricht auch von einem "Abtastsprung im Blickverlauf". Während einer Sakkade ist die visuelle Wahrnehmung stark eingeschränkt. Im erweiterten Sinne bezeichnet der Begriff "Sakkade" jede Form der ruckartigen Folge von Muskelbewegungen, also beispielsweise auch der Extremitätenmuskulatur.

Beim Fahren in einem Zug wird ein Haus fixiert. Sobald man dieses aus den Augen verliert, schnellen die Augen in Fahrtrichtung, um einen neuen Gegenstand zu fixieren. Diese Rückwärtsbewegung stellt die Sakkade dar.

### 7.2.2 Gyroskop

Für das Experiment mit dem Gyroskop haben wir ein Arduino Mega 2560 [17] Mikroprozessor-Entwicklerboard erstanden. Auf dieses haben wir über den  $I^2C$ -Bus das MPU-6050 Gyroskop [18] von Sparkfun angeschlossen. Nach intensiver Recherche im Internet haben wir eine passende Bibliothek -die  $I^2C$  Device Library [19]- für den  $I^2C$ -Bus sowie auch für das Gyroskop gefunden, welche als Open-Source unter der MIT-Lizenz zur Verfügung steht. In der Abbildung 7.1 ist das verwendete, noch nicht verkabelte, Gyroskop abgebildet. Die Betriebsspannung ist mit 3.3V angegeben. In der Abbildung 7.2 ist die vollständig verkabelte Elektronik, inklusive Laser-Pointer, dargestellt.

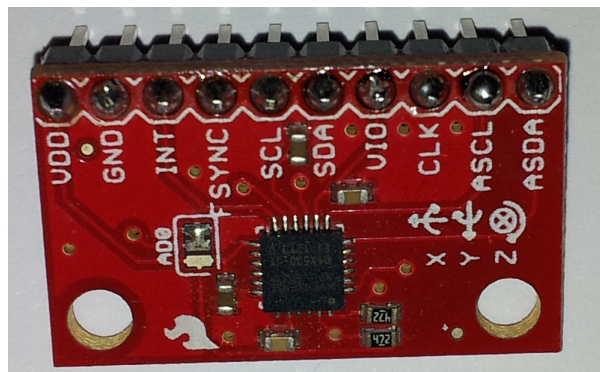


Abbildung 7.1: Das im Experiment verwendete Gyroskop MPU-6050 von Sparkfun.

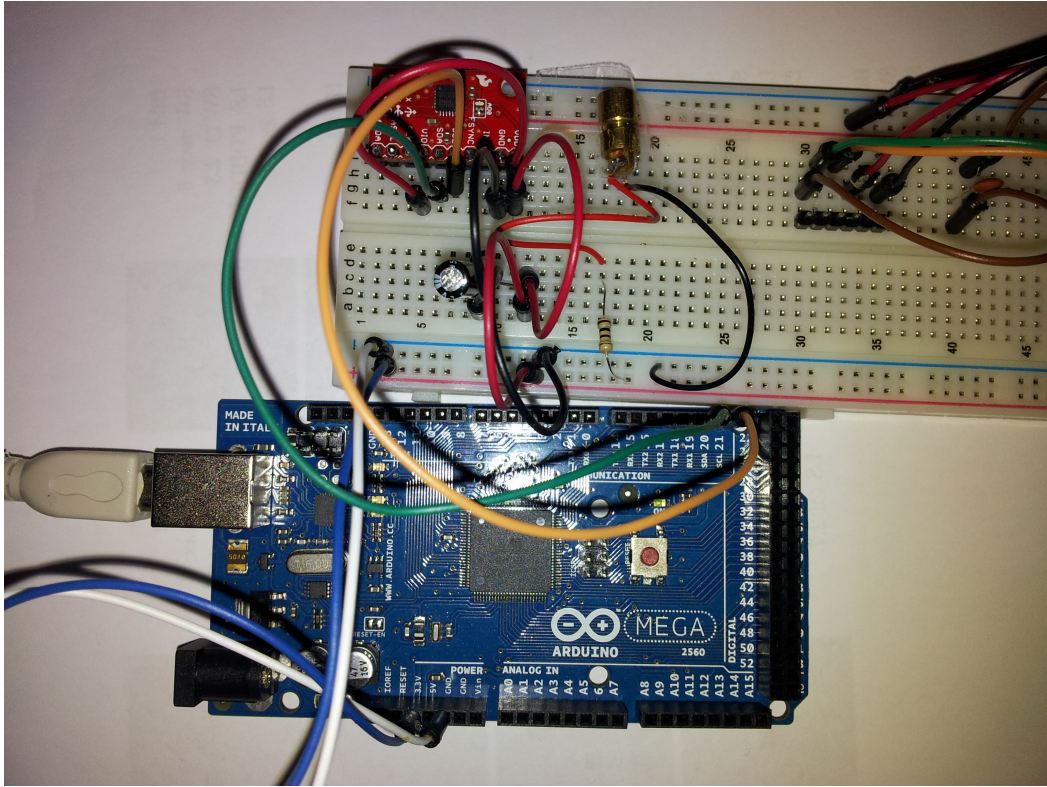


Abbildung 7.2: Das verkabelte Arduino-Entwicklerboard mit dem an den  $I^2C$ -Bus angeschlossenen Gyroskop. Rechts neben dem Gyroskop befindet sich der kleine, rot leuchtende Laser-Pointer.

Das Auslesen der Gyroskop-Werte erfolgt über eine Terminal-Session auf den seriellen Port des Arduino.

### 7.2.3 Features

Das Wort Feature wird aus dem Englischen mit "Merkmal" oder "Eigenschaft" übersetzt. Für unsere Arbeit bedeutet ein Feature, eine berechenbare Eigenschaft eines Objektes in einem Testbild.

## 7.3 Datenübersicht: Features

Hier werden für alle Testbilder die berechneten Feature-Werte aufgelistet. Diese Daten werden von den verschiedenen Modellen für ihre Vorhersage verwendet. Das Clustering mit dem SSC liefert nicht für jedes Bild alle sechs Objekte. Diese Fälle sind mit "-" markiert.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	16.26	0.14	0.0909	2
B	4	1.00	8.94	0.35	0.1909	4
C	15	1.67	12.35	0.05	0.0636	5
D	9	1.00	0.71	0.10	0.0909	3
E	12	1.33	12.38	0.03	0.0010	6
F	12	1.33	13.20	0.33	0.0909	1

Abbildung 7.3: Tabelle mit allen berechneten Features für das Testbild 1.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	6.52	0.07	0.0818	2
B	4	1.00	14.14	0.07	0.1182	3
C	15	1.67	13.51	0.00	0.0091	6
D	9	1.00	2.55	0.07	0.0091	4
E	12	1.33	13.90	0.01	0.0727	1
F	12	1.33	12.78	0.02	0.0909	5

Abbildung 7.4: Tabelle mit allen berechneten Features für das Testbild 2.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	16.26	0.36	0.0909	2
B	4	1.00	8.94	0.16	0.1909	4
C	15	1.67	12.35	0.04	-	-
D	9	1.00	0.71	0.21	0.0909	1
E	12	1.33	12.62	0.09	-	-
F	12	1.33	13.20	0.08	0.0909	3

Abbildung 7.5: Tabelle mit allen berechneten Features für das Testbild 3.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	14.30	0.37	0.0909	1
B	4	1.00	8.06	0.22	0.1909	4
C	15	1.67	13.73	0.02	0.0182	5
D	9	1.00	2.55	0.16	0.0909	2
E	12	1.33	7.02	0.11	-	-
F	12	1.33	13.90	0.38	0.0909	3

Abbildung 7.6: Tabelle mit allen berechneten Features für das Testbild 4.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	6.52	0.08	-	-
B	4	1.00	14.14	0.08	-	-
C	15	1.67	9.62	0.13	0.0091	1
D	9	1.00	8.75	0.13	-	-
E	12	1.33	15.21	0.07	-	-
F	12	1.33	15.91	0.15	-	-

Abbildung 7.7: Tabelle mit allen berechneten Features für das Testbild 5.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	4.74	0.53	0.0909	4
B	4	1.00	12.81	0.03	-	-
C	15	1.67	7.52	0.18	0.0727	1
D	9	1.00	6.52	0.30	0.0909	2
E	12	1.33	10.74	0.33	0.0909	3
F	12	1.33	9.39	0.04	-	-

Abbildung 7.8: Tabelle mit allen berechneten Features für das Testbild 6.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	6.52	0.03	-	-
B	4	1.00	12.08	0.06	-	-
C	15	1.67	13.51	0.17	-	-
D	9	1.00	2.55	0.11	0.0909	2
E	12	1.33	13.46	0.09	0.1636	1
F	12	1.33	9.01	0.01	-	-

Abbildung 7.9: Tabelle mit allen berechneten Features für das Testbild 7.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	10.70	0.11	0.0909	3
B	4	1.00	13.42	0.34	0.1909	4
C	15	1.67	7.38	0.10	-	-
D	9	1.00	13.51	0.18	0.0909	1
E	12	1.33	15.91	0.10	-	-
F	12	1.33	1.80	0.31	0.0909	2

Abbildung 7.10: Tabelle mit allen berechneten Features für das Testbild 8.

Name	Fläche	goldener Schnitt	Distanz	Randkontrast	SSC-Stabilität	SSC-Reihenfolge
A	9	1.00	14.30	0.19	0.0909	3
B	4	1.00	8.06	0.23	-	-
C	15	1.67	13.73	0.26	0.0727	4
D	9	1.00	2.55	0.26	0.0909	5
E	12	1.33	7.02	0.38	0.0909	1
F	12	1.33	13.90	0.25	0.0909	2

Abbildung 7.11: Tabelle mit allen berechneten Features für das Testbild 9.

## 7.4 Datenübersicht: Modelle der ersten Objektwahl

### 7.4.1 Flächenmodell

Dieses Modell stellt seine Vorhersage anhand der Fläche her. Die grossen Objekte werden vor den kleineren betrachtet. Die Vorhersagekraft für dieses Modell ist 12.5%. Die Abbildung 7.12 zeigt die Vorhersagen dieses Modells.

Testbild	Vorhersage	Treffer
1	C	Nein
2	C	Nein
3	C	Nein
4	C	Nein
5	C	Ja
6	C	Nein
7	C	Nein
8	C	Nein
	<b>Quote:</b>	<b>12.5</b>

Abbildung 7.12: Vorhersagekraft des Flächenmodells.



### 7.4.2 "Goldener Schnitt"-Modell

Die Vorhersage dieses Modells 7.13 basiert auf dem goldenen Schnitt. Die Objekte, deren Verhältnis am nächsten beim goldenen Schnitt liegt, werden zuerst fixiert. Die Vorhersage dieses Modells ist mit 12.5% nicht sehr stark. Zudem ist es für die Vorhersage der Objektwahl ungeeignet, da sich der goldene Schnitt nur auf das Objekt, nicht aber auf dessen Lage bezieht. Aus diesem Grund haben wir dieses Modell nicht weiter verfolgt.

Testbild	Vorhersage	Treffer
1	C	Nein
2	C	Nein
3	C	Nein
4	C	Nein
5	C	Ja
6	C	Nein
7	C	Nein
8	C	Nein
	<b>Quote:</b>	<b>12.5</b>

Abbildung 7.13: Vorhersagekraft des "Goldenen Schnitt"-Modells.

### 7.4.3 "Distanz zur Mitte"-Modell

Die Distanz zur Mitte des Bildes dient bei diesem Modell als Grundlage. Die Objekte, welche sich am nächsten in der Bildmitte befinden, werden zuerst betrachtet. Je näher ein Objekt der Mitte ist, desto wahrscheinlicher wird es zuerst fokussiert. Dieses Modell liefert eine sehr gute Vorhersage mit 75%. In der folgenden Abbildung 7.14 ist die Auswertung für alle Bilder aufgestellt.

Testbild	Vorhersage	Treffer
1	D	Ja
2	D	Ja
3	D	Ja
4	D	Ja
5	A	Nein
6	A	Nein
7	D	Ja
8	F	Ja
<b>Quote:</b>		<b>75</b>

Abbildung 7.14: Vorhersagekraft des "Distanz zur Mitte"-Modells.

### 7.4.4 Randkontrastmodell

Beim Randkontrast geht man davon aus, dass die Objekte mit dem höchsten Randkontrast zuerst betrachtet werden. Dieses Modell hat eine Qualität von 12.5% erreicht. Abbildung 7.15 zeigt die Vorhersagen dieses Modells.

Testbild	Vorhersage	Treffer
1	B	Nein
2	D	Ja
3	A	Nein
4	F	Nein
5	F	Nein
6	A	Nein
7	C	Nein
8	B	Nein
<b>Quote:</b>		<b>12.5</b>

Abbildung 7.15: Vorhersagekraft des Randkontrastmodell.

### 7.4.5 SSC Stabilitäts-Modell

Dieses Modell berücksichtigt die Stabilitäten, welche beim Clustering Prozess des SSC berechnet wurden. Die Objekte mit der stärksten Stabilität werden, laut diesem Modell, zuerst angeschaut. Dieses Modell hat eine Vorhersagequote von 12.5% erreicht. Abbildung 7.16 zeigt die Vorhersagen dieses Modells.

Testbild	Vorhersage	Treffer
1	B	Nein
2	B	Nein
3	B	Nein
4	B	Nein
5	C	Nein
6	C	Ja
7	E	Nein
8	B	Nein
	<b>Quote:</b>	<b>12.5</b>

Abbildung 7.16: Vorhersagekraft des SSC Stabilität-Modells.

### 7.4.6 SSC Reihenfolge-Modell

Dieses Modell berücksichtigt die Reihenfolge der vom Clustering Prozess gefundenen Objekte. Das Objekt, welches beim Clustering zuerst erkannt wird, wird am häufigsten von den Probanden zuerst betrachtet. Die Vorhersagegenauigkeit dieses Modells ist mit 25% besser als die Vorhersage mit der SSC Stabilität. In der Abbildung 7.17 sind die Vorhersagen dieses Modells dargestellt.

Testbild	Vorhersage	Treffer
1	F	Nein
2	E	Nein
3	D	Ja
4	A	Nein
5	C	Nein
6	C	Ja
7	E	Nein
8	D	Nein
	<b>Quote:</b>	<b>25</b>

Abbildung 7.17: Vorhersagekraft mittels der Reihenfolge des SSC Clusterings.

## 7.5 Verwendete Werkzeuge, Code

### 7.5.1 Hardware, Software und Betriebssystem

Für die Aufzeichnungen stand uns ein komplett konfigurierter Rechner mit-  
samt Eye-Tracking System von der ETH Zürich zur Verfügung. Darin einge-  
baut ist die spezielle Hardware, welche für das Eye-Tracking System benötigt  
wird. Das Betriebssystem dieses Rechners ist Windows 2000 Professional.

Die von der HSR zur Verfügung gestellten Rechner sind mit Windows 7  
x64 Enterprise als Betriebssystem ausgestattet. Die verwendete Software ist  
das .NET-Framework 4.0, Visual Studio 2010 Ultimate - mit C# als Sprache  
- sowie Wolfram Mathematica in der Version 8.0.4.

### 7.5.2 Umgebungsvariable

Um die Problematik mit verschiedenen Pfaden umgehen zu können, haben  
wir eine Umgebungsvariable eingeführt. Diese wird konsequent überall ver-  
wendet, sowohl im C#-Code wie auch in Mathematica. Die Umgebungsvara-  
ble heisst `VisualPerception`. Abbildung 7.18 zeigt die konfigurierte Variable  
unter Windows 7.

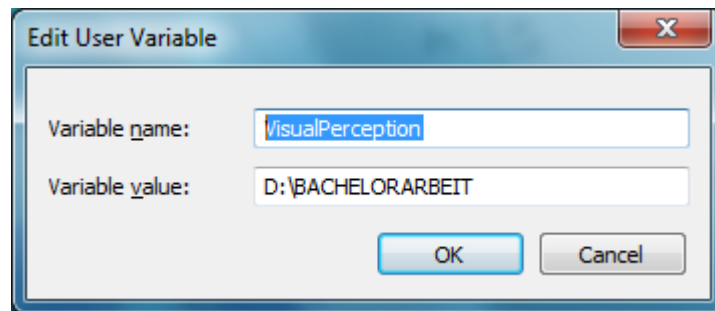


Abbildung 7.18: Die verwendete Umgebungsvariable für Mathematica und C#-Code.

### 7.5.3 Ordnerstruktur der DVD

Die Ordnerstruktur auf der DVD umfasst folgende Hauptordner:

**00 Doc** Dokumente und Unterlagen, Latex

**01 Source** C#-Quellcode der von uns entwickelten Tools

**02 Binaries** Enthält ausführbare Binärdateien. Darunter SSC für Linux und einige Chronos-Tools

**03 Model Generation** Sämtliche Daten von unserem Experiment, ausser die Rohdaten der Aufzeichnung

**99 EyeTracker Raw Data** Rohdaten der Aufzeichnungen

## 7.5.4 Beschreibung der entwickelten Tools

Im Ordner des C#-Quellcodes befindet sich eine Visual Studio Solution, welche alle drei von uns erstellten Tools umfasst. Eine kurze Erläuterung:

### Picture Generator

Der Picture Generator dient der automatisierten Erstellung der Testbilder. Die darin versteckten Cluster sind "Hard-Coded" in der dll `Data.dll`. Das Programm ist ein Konsolenprogramm, welches beim Start direkt alle Bilder erstellt und diese automatisch in die entsprechenden Unterordner von **03 Model Generation** abspeichert.

### Visual Selector

Der Visual Selector dient dazu, die Eye-Track-Daten entsprechend dem Kalibrationsbild einzupassen und direkt auf die entsprechenden Testbilder zuzuschneiden. Als Eingabe für einen Probanden benötigt das Tool die drei Aufnahmesequenzen, jeweils für das linke und rechte Auge, also sechs Dateien. Die Namensgebung ist auch fest im Code verdrahtet und sieht wie folgt aus:

```
LeftEye{SequenzNummer}  
RightEye{SequenzNummer}
```

Über das grafische UI kann auf die entsprechenden Sequenzen der Probanden zugegriffen werden. Mit einem Klick auf "Save calibrated cluster data" werden sämtliche - für das letzte Tool - benötigte Ordner und Dateien erstellt. Falls die Zeitintervalle in der Präsentation angepasst würden, müsste in diesem Tool der Zeitstrahl für die Zuschneidung geändert werden.

### Visual Analyze

Das letzte von uns erstellte Tool ist die Analyse für jeden einzelnen Probanden. Damit werden die betrachteten Punkte den Objekten zugeordnet, sowie Sequenzen und Markov-Übergangsmatrizen erstellt.

Das Tool erstellt die Eingabedateien für den SSC, die Datengrundlage für die Heatmaps und kumulierte Grafiken/Aussagen über alle Probanden für ein einzelnes Testbild.

Auch dieses Tool ist ein Konsolenprogramm, welches nach dem Start sämtliche vorgestellte Operationen automatisch durchführt und die Ergebnisse in den dafür vorgesehenen Ordnern speichert.

### 7.5.5 Bildnummern in der Ordnerstruktur

Es gibt einen Unterschied zwischen den in den Ordnerstrukturen und derjenigen in dieser Dokumentation verwendeten Bildnummern. Die Bildnummern in der Dokumentation sind der Einfachheit halber ohne Lücken durchnummeriert. In der Ordnerstruktur besitzen diese Bilder andere Nummern, weil zwischen den einzelnen Testbildern eine Kalibrierungsanimation angezeigt wird. Die dazugehörigen Kalibrierungsbilder sind in der Ordnerstruktur ebenfalls enthalten, was die Nummerierung der Testbilder in der Dokumentation etwas versetzt. In der Abbildung 7.19 ist eine Zuordnungstabelle zwischen den Bildnummern aus der Dokumentation und derjenigen aus der Ordnerstruktur aufgelistet.

Name Dokumentation	Name Ordnerstruktur	Kommentar
–	Picture00	Kalibrierungsbild
Testbild 1	Picture01	
Testbild 2	Picture02	
Testbild 3	Picture03	
Testbild 4	Picture04	
–	Picture05	Kalibrierungsbild
Testbild 5	Picture06	
Testbild 6	Picture07	
Testbild 7	Picture08	
–	Picture09	Kalibrierungsbild
Testbild 8	Picture10	
Testbild 9	Picture11	Farbige Objekte
Testbild 10	Picture12	Nur Rauschen

Abbildung 7.19: Zuordnungstabelle: Zwischen der Nummerierung in der Dokumentation und der Nummerierung in der Ordnerstruktur gibt es Unterschiede. Diese kommen aufgrund der Kalibrierungsbilder zustande.

## 7.6 Arbeitsplanung

Die für die Bachelorarbeit zur Verfügung stehenden 14 Wochen haben wir gemäss Abbildung 7.20 in die folgenden Bereiche aufgeteilt. Wir konnten den Arbeitsplan ziemlich gut einhalten, abgesehen von der Dokumentation, welche mehr Zeit in Anspruch nahm als geplant.

Woche	Tätigkeiten	Meilensteine
Woche 01 Woche 02	Kickoff: Vorbereitung, einlesen, einrichten	
Woche 03 Woche 04	System aufsetzen, Eye-Tracker in Betrieb nehmen	- "System running" - bereit für erste Tests
Woche 05 Woche 06	Testbilder erstellen, Experimentablauf definieren, Testen des Experiments, Datenkontrolle	- Experiment bereit - Features berechenbar
Woche 07 Woche 08	Experimente mit Probanden durchführen	- Daten gesammelt - Daten aufbereitet
Woche 09 Woche 10	Modelle erstellen, Varianten ausprobieren Experiment dokumentieren	- Modellvarianten klar - Features gewichtet
Woche 11 Woche 12	Dokumentation	- Dokumentation steht
Woche 13	Surrogate-Daten, Tests der Modelle	- Aussage über Modell
Woche 14	Abschluss Dokumentation	- Erfolgreiche Abgabe

Abbildung 7.20: Arbeitsplanung für die Bachelorarbeit.



# Literaturverzeichnis

- [1] S. Piantanida, C. Baroni, and K.K. Eberlein. *Leonardo da Vinci: das Lebensbild eines Genies*. Pennsylvania, E. Vollmer, 1955.
- [2] J. Koller. *Albrecht Dürer: Selbstbildnis um 1500*. 8-9, München, GRIN Verlag, 2011.
- [3] S. Steiner-Welz. *Ornamente. Band 1*. 16-17, Mannheim, Reinhard Welz Vermittler Verlag, 2004.
- [4] P. Wespi. *Statistical methods of human visual perception*. Master Thesis, Zürich, ETH, 2012.
- [5] Chronos-Vision Homepage. <http://www.chronos-vision.de>. Zuletzt besucht: 14.11.2012.
- [6] W. Burger and M.J. Burge. *Digitale Bildverarbeitung: Eine Einführung Mit Java und ImageJ*. X. media. press Series. 167-169, Berlin, Springer-Verlag, 2006.
- [7] D. Heeger. *Poisson Model of Spike Generation*. 5-8, New York, New York University, 2000.
- [8] R.S. Woodworth and H. Schlosberg. *Experimental Psychology*. Michigan, University of Michigan, 1954.
- [9] J.A. Jacko. *Human-Computer Interaction: New Trends*. Lecture Notes in Computer Science. 31-39, Berlin, Springer-Verlag, 2009.
- [10] H. Rapp. *Mathematik für die Fachschule Technik: Algebra, Geometrie, Differentialrechnung, Integralrechnung, Vektorrechnung, Komplexe Rechnung*. Viewegs Fachbücher der Technik. 311, Wiesbaden, Vieweg+Teubner Verlag, 2010.
- [11] A. Troelsen. *C# and the .NET Platform*. .Net Developer. 693, Berlin, Apress, 2003.

- [12] T. Ott T. Jasa, T. Lanz and R. Stoop. Sequential Superparamagnetic Clustering as a Predictor of Visual Fixations. *Conference Proceedings, NDES-Conference, IEEE, HSR Rapperswil*, 2009.
- [13] T.F. Smith and M.S. Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology* 147, 195-197, 1981.
- [14] A. Forrester, A. Sobester, and P.A. Keane. *Engineering Design via Surrogate Modelling: A Practical Guide*. 33-76, Hoboken (New Jersey), John Wiley & Sons, 2008.
- [15] H. J. A. Dartnall J. K. Bowmaker. Visual pigments of rods and cones in a human retina. *The Journal of Physiology* 298, 501–511, 1980.
- [16] H.O. Karnath and P. Thier. *Neuropsychologie: Psychologie, Neurobiologie und Störungen kognitiver Funktionen*. Springer-Lehrbuch Series. 313-319, Berlin, Springer-Verlag, 2003.
- [17] Arduino Mega 2560 Specification.  
<http://arduino.cc/en/Main/arduinoBoardMega2560>. Zuletzt besucht: 18.12.2012.
- [18] MPU-6050 Product Specification. <http://shop.boxtec.ch/pub/sparkfun/PS-MPU-6000A.pdf>. Zuletzt besucht: 18.12.2012.
- [19] I2C Device Library Homepage. <http://www.i2cdevlib.com/>. Zuletzt besucht: 18.12.2012.