

Untersuchung der Augenbewegung und visuellen Salienz in Bildausschnitten von Photographie

Dissertation

zur Erlangung des akademischen Grades
doctor medicinae (Dr. med.)

vorgelegt dem Rat der Medizinischen Fakultät der
Friedrich-Schiller-Universität Jena

von Jonas Abeln

geboren am 16.10.1991 in Seesen

Gutachter

1: Prof. Dr. Dr. Christoph Redies

2: Prof. Dr. Laszlo Kovacs

3: Prof. Dr. Achim Landwehr

Tag der öffentlichen Verteidigung: 30.06.2020

Verzeichnis der Abkürzungen

Abkürzung	Erklärung
bspw.	beispielsweise
bzw.	beziehungsweise
COM	<i>center-of-mass</i>
CGL	<i>Corpus geniculatum laterale</i>
d.h.	das heißt
et al.	und andere
GBVS	<i>Graph Based Visual Saliency Algorithmus</i>
°	Grad
ITTI-Algorithmus	Salienzalgorithmus
%	Prozent
RGB Farbraum	Rot-Grün-Blau Farbraum
s	Sekunde
sog.	sogenanntes
TDT	<i>total dwelling time</i>
u. a.	unter anderem
z.B.	zum Beispiel
z. Zt.	zur Zeit

Inhaltsverzeichnis

1	Zusammenfassung	6
2	Einleitung	7
2.1	Das Sehsystem	8
2.1.1	Aufbau des Sehsystems	8
2.1.1.1	Aufbau der Retina	9
2.1.1.2	Rezeptive Felder	11
2.1.1.3	Die Sehbahn	13
2.1.1.4	Der visuelle Cortex	15
2.1.2	Augenbewegung für die visuelle Wahrnehmung	17
2.2	Aufmerksamkeit	18
2.2.1	Salienz-Algorithmen	19
2.2.2	Itti und Koch Salienz-Algorithmus (ITTI)	20
2.2.3	<i>Graph-Based Visual Saliency</i> Algorithmus (GBVS)	21
2.3	Balance im Kontext visueller Ästhetik	22
2.4	<i>Computational Aesthetics</i>	24
2.5	<i>Cropping</i>	24
2.6	Fragestellung	26
3	Veröffentlichung	28
4	Diskussion	45
4.1	Kritische Betrachtung der Methoden	46
4.1.1	<i>Central-Tendency-Bias</i>	46
4.1.2	Auswahl der Salienzalgorithmen	47
4.2	Relevanz der Ergebnisse im Kontext von Balance	48
4.3	Relevanz der Ergebnisse im Kontext von <i>Computational Aesthetics</i>	50
4.4	Analyse der Augenbewegung in der Wahrnehmung von Kunst und in der Medizin	52

4.5 Schlussfolgerung	54
Literaturverzeichnis	55
Anhang	63
Danksagung	63
Ehrenwörtliche Erklärung	64

1 Zusammenfassung

Das Gebiet der Neuroästhetik befasst sich mit den Sinnesreizen, die ein ästhetisches Empfinden hervorrufen sowie ihrer Verarbeitung. In meiner Arbeit beschäftigte ich mich mit dem Zuschneiden von Photographien, um sie ästhetisch ansprechender zu gestalten (sog. *Cropping*). Der von mir analysierte Datensatz stammt aus vorangegangenen Studien von McManus et al. (2011a, 2011b), die zeigen konnten, dass *Cropping* wiederholbar und zuverlässig von Probanden durchgeführt wird. Es war den Autoren nicht möglich, objektive Parameter zu isolieren, die während des *Cropping*-Vorganges eine Rolle spielen.

In meiner Arbeit konnte ich zeigen, dass sowohl Salienz als auch Augenbewegungen dafür von zentraler Bedeutung sind. Probanden versuchen, so viel wie möglich von der salienten Struktur in das zugeschnittene Bild zu integrieren. Dabei wird die Salienz im zugeschnittenen Bild nicht zufällig, sondern balanciert verteilt. Dies bedeutet, dass der Bildschwerpunkt der Salienz einen geringen Abstand zum geometrischen Mittelpunkt besitzt.

Die Ergebnisse meiner Untersuchung zeigen zum einen eine Verbindung zwischen *Cropping* und Salienz auf. Zum anderen leistet diese Arbeit auch einen Beitrag zum Verständnis der *bottom-up* Prozesse des visuellen Systems und dem Konzept der *pictorial balance* bei der ästhetischen Bewertung von Photographien.

2 Einleitung

In der Photographie ist die Auswahl des richtigen Blickwinkels von zentraler Bedeutung. Hierbei kann zu Beginn durch die richtigen Perspektive, aber auch später durch Zuschneiden (sog. *Cropping*) das Bild ästhetisch ansprechender gestaltet werden. Meine Arbeit befasst sich mit einer Grundlagen, die der Mensch nutzt um Bilder zu *croppen*. Dabei ist die Wahrnehmung im menschliche Auge und dem damit verbundenem Sehsystem von Bedeutung. Hierbei gibt es zwei miteinander agierende Systeme (sog. *bottom-up* und *top-down* Prozesse) der Wahrnehmung, die für die Steuerung der Aufmerksamkeit entscheidend sind (Hudspeth et al. 2013). Daher wird in dem ersten Teil meiner Arbeit zunächst auf das Sehsystem und anschließend auf Wahrnehmung der Aufmerksamkeit eingegangen.

Durch Computeralgorithmen wird versucht, die Strukturen, die Aufmerksamkeit erregen (in der Psychologie als salient bezeichnet), darzustellen. Da dies in einem engem Zusammenhang mit den Augenbewegungen steht (Hogg und Vaughan 2010), wird in meiner Arbeit auf zwei Salienzalgorithmen näher eingegangen. Einmal der ITTI-Algorithmus (Itti et al. 1998), der einer der ältesten dieser Algorithmen ist, und der GBVS (Harel et al. 2007). Dieser nutzt identische Parameter, unterscheidet sich aber durch eine Herangehensweise über Wahrscheinlichkeiten.

Da das Konzept der Balance in der visuellen Ästhetik eine entscheidene Rolle bei der Wahrnehmung spielt (Arnheim 1954) und dies durch den Auswahl des Blickwinkels einer Photographie oder das Zuschneiden eines Bildes verändern lässt, wird dieser Aspekt im folgenden näher beleuchtet.

Im Kontext dieser Grundlagen werden die vorangegangenen Studien von McManus et al. (2011a, 2011b) betrachtet, die die Grundlage meiner Arbeit bilden und weitere Arbeiten aus diesem Feld beschrieben, bevor auf die Fragestellung meiner Arbeit eingegangen werden kann.

2.1 Das Sehsystem

"Das Auge war vor allen anderen das Organ, womit ich die Welt faßte."

Johann Wolfgang von Goethe (1812)

Wie auch schon Johann Wolfgang von Goethe aufzeigte, spielt das menschliche Sehsystem eine wichtige Rolle in der Wahrnehmung unserer Umgebung. Tomasello et al. (2007) fanden dazu einen Entwicklungsunterschied im Betrachtungsverhalten von Menschenaffen zu menschlichen Säuglingen. Im Unterschied zu Menschenaffen verfolgten die Säuglinge bei Kopfbewegungen des zu beobachtenden Menschen ohne Blickrichtungsveränderung weiterhin die Augen, wohingegen die Menschenaffen der Kopfbewegung folgten. Aber auch bei Erkrankungen aus dem Autismus-Spektrum lassen sich durch Analyse der Augenbewegungen Unterschiede finden. Jones et al. (2008) zeigten in ihrer Studie, dass sich die Betrachtung von Gesichtern zwischen zweijährigen Kindern mit und ohne Autismus, sowie zweijährigen Kindern, die in der Entwicklung verspätet sind, unterscheidet. Hierbei zeigte sich, dass die Kinder mit Autismus seltener auf die Augen der zu beobachtenden Person blicken als die anderen beiden Gruppen. Dies korreliert mit dem Schweregrad des Autismus.

Es lassen sich nicht nur bei der menschlichen Entwicklung oder Erkrankungen Unterschiede bei der Wahrnehmung durch die Augen finden, aber auch durch Erlerntes wird dies beeinflusst. Hierfür zeigten Vogt und Magnussen (2007) bspw., dass es Unterschiede der Augenbewegungen und der damit verbundenen Betrachtungsweise zwischen Künstlern und Nichtkünstlern bei Bildern gibt. Hierbei achten Künstler unter anderem mehr auf die Komposition oder Balance.

Da das Sehsystem eine wichtige physiologische Rolle bei der Wahrnehmung spielt, wird in den nächsten Abschnitt zunächst auf den Aufbau des Sehsystems und die Wahrnehmung von Farbe und Kontrasten eingegangen.

2.1.1 Aufbau des Sehsystems

Wie von Hudspeth et al. (2013) beschrieben, beginnt die Wahrnehmung durch das Sehsystem mit dem Eintreffen von Licht auf den Sehapparat (siehe Abb. 1). Das Licht wird zunächst an der Cornea und der Linse gebrochen und fokussiert. Anschließend tritt es durch den Glaskörper, um auf die Retina zu stoßen. Hier reagieren die Photorezeptorzellen auf das eintreffende Lichtsignal. Anschließend wird es vom Pigmentepithel absorbiert, um Streulicht zu vermeiden. Die optische Achse fokussiert hierbei auf die *Fovea centralis*, den Punkt des schärfsten Sehens.

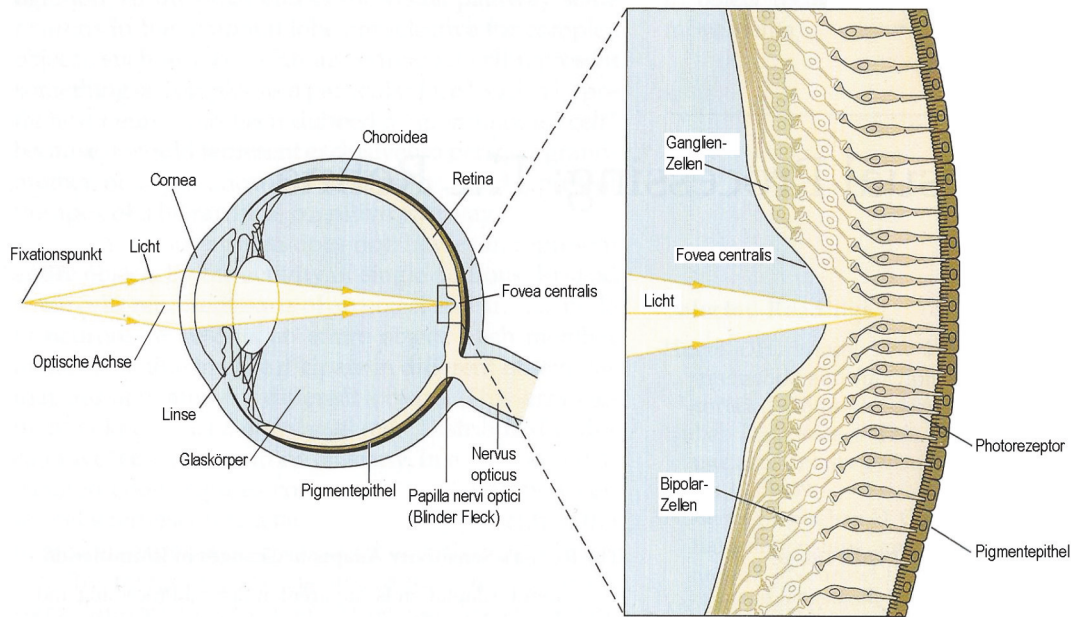


Abbildung 1: Schematische Darstellung des Querschnitts des Auges. Modifiziert nach Hudspeth et al. (2013)

Die Retina ist über dem gesamten Augenhintergrund nicht gleichmäßig strukturiert. Am Punkt des schärfsten Sehens sind die Perikarya der nachfolgenden Neurone an den Rand gedrängt und die Photorezeptorzellen eins zu eins mit den nachfolgenden Neuronen verschaltet (Prasad und Galetta 2011). Die *Papilla nervi optici* befindet sich 15° nasal der optischen Achse. Sie ist der Austrittspunkt des Sehnervs aus dem Auge. Hier sind keine Photorezeptorzellen vorhanden, da dort die Axone der Neurone, die aus dem Auge austreten, den *Nervus opticus* bilden. Die Retina liegt der Choroida an, welche unter anderem die Aufgabe hat, das Pigmentepithel und die Zellen der Retina zu versorgen (Welsch und Deller 2010).

2.1.1.1 Aufbau der Retina

Die Retina ist mehrschichtig aufgebaut (siehe Abb. 2) (Hudspeth et al. 2013). Die äußere Schicht bildet das Pigmentepithel, in das die Photorezeptorzellen hineinragen, von denen es zwei verschiedene Typen gibt: die Stäbchen, die Helligkeitsunterschiede detektieren, sowie die Zapfen. Letztere sind für die Wahrnehmung von Farbe zuständig. Sie werden in drei Untergruppen unterschieden, je nachdem, auf welchen Teil des sichtbaren

Lichtspektrums sie reagieren: S (*Short Wave*) für blaues, M (*Medium Wave*) für grünes und L (*Long Wave*) für rotes Licht (Schnapf et al. 1988).

Die Verteilung dieser zwei grundlegenden Typen ist unterschiedlich. Die farbwahnehmenden Zapfen befinden sich dicht gepackt fast ausschließlich in der *Fovea centralis*. Die Zapfen kommen aber auch sonst in der Retina vor. Die Stäbchen befinden sich um die *Fovea centralis* herum gruppieren. Die Rezeptorzellendichte nimmt zur Peripherie der Retina hin ab.

Das anschließende Signal geben die Photorezeptorzellen an die Bipolarzellen weiter. Innerhalb der *Fovea centralis* verschaltet sich je eine Rezeptorzelle auf eine Bipolarzelle. Außerhalb der *Fovea centralis* konvergieren Signale mehrerer Photorezeptorzellen auf eine Bipolarzelle. Diese wiederum leitet das Signal an die nachfolgenden Ganglienzellen weiter, wobei auch hier mehrere Bipolarzellen auf eine Ganglienzelle projizieren und so rezeptive Felder bilden (Hudspeth et al. 2013). Unterstützt wird dies zusätzlich durch laterale Inhibition, die von den Horizontal- und den amakrinen Zellen vermittelt wird.

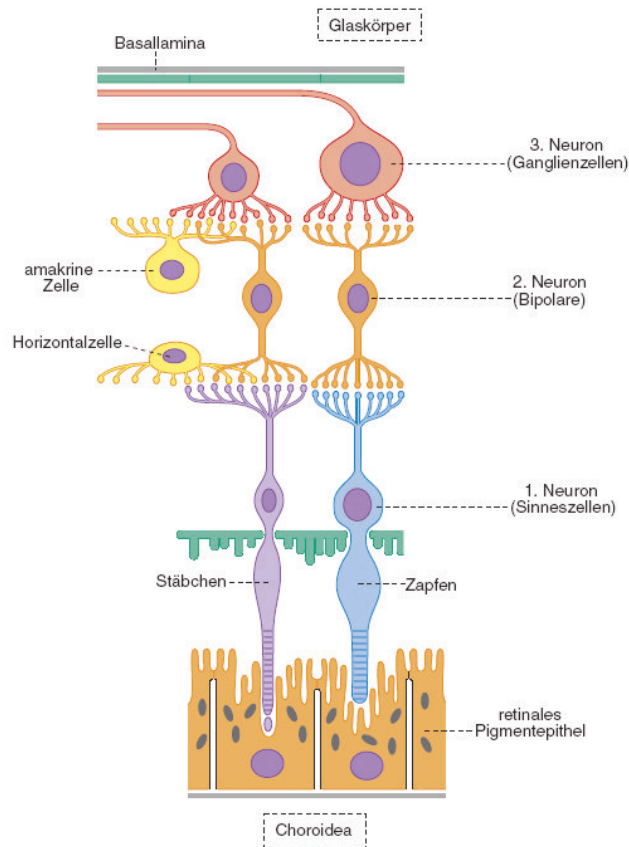


Abbildung 2: Schematische Darstellung der Retina. Modifiziert nach Welsch und Deller (2010)

2.1.1.2 Rezeptive Felder

Die Anzahl der Axone der Ganglienzellen, die als *Nervus opticus* das Auge verlassen, beträgt nur ungefähr 1 % der Anzahl der gesamten Photorezeptorzellen (Hudspeth et al. 2013). Dieses Verhältnis lässt sich durch die rezeptiven Felder erklären, die auf der Ebene der Bipolarzellen beginnen. Grundsätzlich gibt es zwei Arten von Bipolar- und Ganglienzellen: diejenigen, die ein Signal der Rezeptorzellen weitergeben (im Folgenden On-Zellen), und die, die ein Signal weitergeben, wenn sie nicht durch die Rezeptorzellen erregt werden (im Folgenden Off-Zellen). Dadurch, dass die Bipolarzellen ihr Signal in gleicher Weise an die Ganglienzellen weitergeben und es zur lateralen Inhibition durch Horizontalzellen und den amakrinen Zellen kommt, entstehen rezeptive Felder, wie in Abb. 3 schematisch gezeigt ist.

Hierbei entsteht zum Beispiel ein Signal, wenn in der Mitte des Areal einer On-Zelle ein Licht trifft (Fall 1). Um dieses mittige Areal befindet sich ein inhibitorisches Gebiet. Wenn hier ein Lichteinfall auftritt, sorgt dieser dafür, dass das weitergeleitete Signal stark unterdrückt wird (Fall 2). Trifft jetzt ein Licht sowohl auf das On-Areal als auch das Off-Areal, sorgt dies dafür, dass es zu einer Signalunterdrückung kommt. So entsteht ein abgeschwächtes Signal (Fall 3). Zuletzt gibt es die Möglichkeit, dass ein Lichteinfall nur im mittigen On-Feld und kein Licht im inhibitorischem Areal auftritt (Fall 4). Hierbei kommt es zu einer Verstärkung des weitergegebenen Signales. Dieses gilt analog für die Off-Zellen, nur das hier das inhibitorische Areal mittig ist (Hudspeth et al. 2013). Zusätzlich lassen sich auf der Ebene der Ganglienzellen zwei übergeordnete Systeme bei den rezeptiven Feldern unterscheiden, die sowohl aus On-, als auch aus Off-Zellen aufgebaut sind. Zum einen gibt es das magnozellige System, das über Parasol-Ganglienzellen, die große rezeptive Felder besitzen, projiziert und Eindrücke sowohl aus Stäbchen, als auch aus Zapfen aufnimmt. Es ist jedoch nicht farbselektiv (Hudspeth et al. 2013). Da große rezeptive Felder eine geringe Ortsauflösung besitzen, ist dieses System für die Wahrnehmung von Bewegung und grober räumlicher Auflösung zuständig. Dem gegenüber steht das parvozelluläre System, das Midget-Ganglienzellen nutzt und kleine rezeptive Felder besitzt. Es bezieht überwiegend die Informationen aus den Zapfen, die farbspezifisch organisiert sind. Dadurch wird Farbe und detaillierte räumliche Auflösung wahrgenommen. Das parvozelluläre System ist vor allem für die weitere Verarbeitung in der *Fovea centralis* zuständig. Das Prinzip der rezeptiven Felder wird über die vollständige Sehbahn hinweg weiterverwendet (Hudspeth et al. 2013).

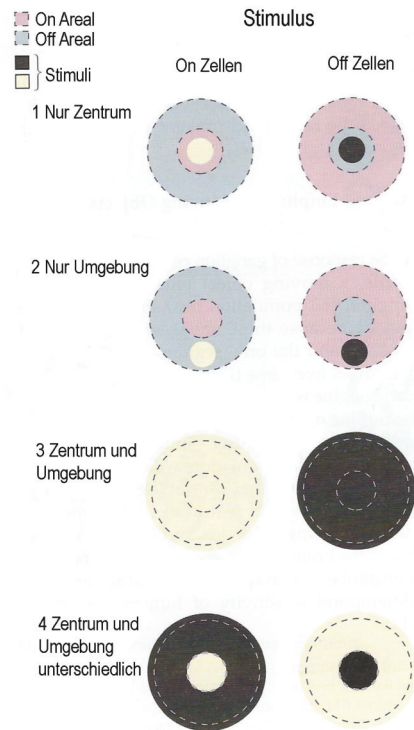


Abbildung 3: Schematische Darstellung der On-/Off-Bipolar- und Ganglienzellen. Modifiziert nach Hudspeth et al. (2013)

2.1.1.3 Die Sehbahn

Nachdem die Axone der Ganglienzellen aus dem Auge austreten, verlaufen sie als *Nervus opticus* weiter über das *Chiasma opticum*. Hier kreuzen die Axone der nasalen Hälften der Retina auf die kontralaterale Seite. Die Axone der temporalen Hälften verlaufen weiter ipsilateral und bilden anschließend gemeinsam mit den gekreuzten Hälften der Gegenseite den *Tractus opticus* (siehe Abb. 4). Im *Corpus geniculatum laterale* (CGL) enden die Axone und werden verschaltet.

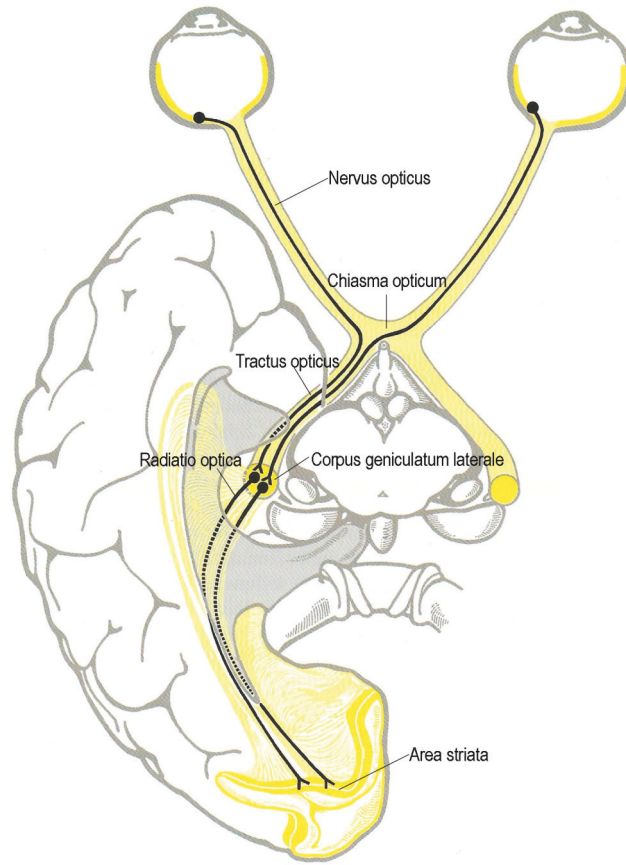


Abbildung 4: Schematische Darstellung der Sehbahn. Modifiziert nach Kretschmann und Weinrich (1996)

Hierbei ist das CGL in sechs Schichten organisiert (siehe Abb. 5). In den Schichten 1 und 2 wird das magnozelluläre und in den Schichten 3 bis 6 das parvozelluläre System umgeschaltet. Die Schichten 1, 4 und 6 sind dabei für Signale des kontralateralen und die Schichten 2, 3 und 5 für die des ipsilateralen Auges zuständig (Hudspeth et al. 2013). Im weiteren Verlauf der *Radiatio optica* projiziert das parvozelluläre System auf die Schicht IV β und das magnozelluläre System auf die Schicht IV α der primären Sehrinde (*Area striata*).

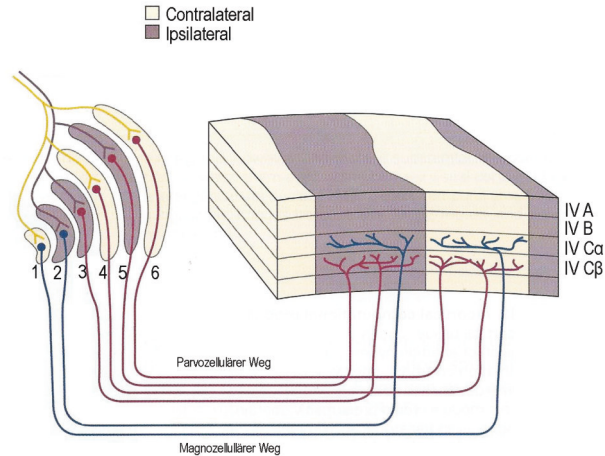


Abbildung 5: Schematische Darstellung des CGL. Modifiziert nach Hudspeth et al. (2013)

2.1.1.4 Der visuelle Cortex

Auf die Neurone der primären Sehrinde (V1) konvergieren wiederum mehrere Neurone des CGL mit ihren rezeptiven Feldern. Durch deren Überlagerung entstehen in der primären Sehrinde rezeptive Felder, die nicht mehr konzentrisch, sondern längs orientiert sind (Hubel und Wiesel 1962) (siehe Abb. 6). Dadurch wird auf kortikaler Ebene ermöglicht, dass nicht nur Kontraste, sondern auch Konturen wahrgenommen werden.

Innerhalb der primären Sehrinde ist die Verarbeitung in Orientierungssäulen und sog. Blobs gegliedert (siehe Abb. 7). Die visuellen Eindrücke vom linken bzw. rechtem Auge zum gleichen optischen Areal werden in getrennten nebeneinanderstehenden okulären Dominanzsäulen analysiert. Jede dieser Dominanzsäulen besteht aus mehreren Orientierungssäulen, die jeweils für eine Orientierung von Konturen zuständig sind. Die Blobs sind für die Farbwahrnehmung zuständig. Sie besitzen kaum Zellen für die Wahrnehmung von Konturen (daher ohne Orientierungspräferenz), aber haben eine hohe Dichte an farbselektiven Zellen. Die Blobs ermöglichen es auch, die Beschaffenheit von Oberflächen zu analysieren (Hudspeth et al. 2013).

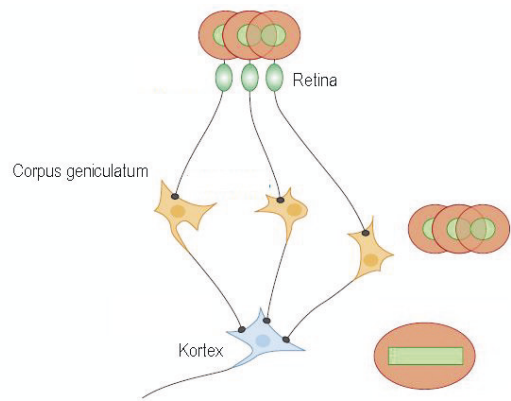


Abbildung 6: Schematische Darstellung der Konvergenz der rezeptiven Felder im Verlauf der Sehbahn. Modifiziert nach Behrends et al. (2010)

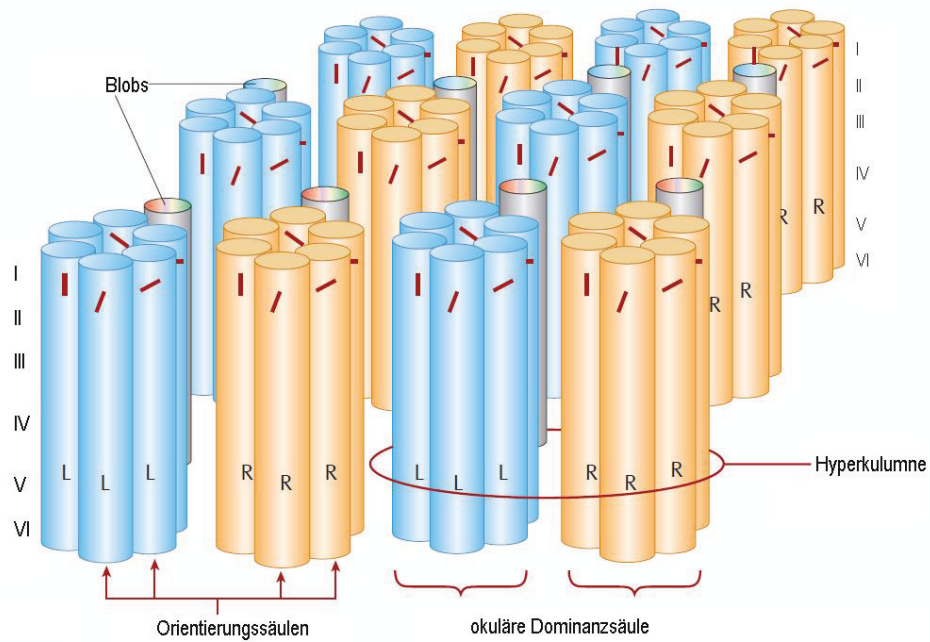


Abbildung 7: Schematische Darstellung der Orientierungssäulen des primären visuellen Kortex. Modifiziert nach Behrends et al. (2010)

Die okulären Dominanzsäulen eines bestimmten optischen Areals werden wiederum zu einer Hyperkolumne zusammengefasst. Hierdurch entsteht eine retinotopie Wahrnehmung, wobei die *Fovea centralis* überproportional im Vergleich zur restlichen Retina repräsen-

tiert ist (Hudspeth et al. 2013).

Die weitere Analyse folgt auf höheren kortikalen Ebenen, wo die einzelnen Bildteile nicht mehr strikt separat betrachtet und so Informationen global zusammengeführt werden. In der sekundären Sehrinde (V2) wird bei dem sich bis dahin ergebenen Bild nach Scheinkonturen gesucht und führt die Informationen aus dem Bild global zusammen. V4 führt die Informationen über Farbe und Objektgestalt zusammen und V5 analysiert Bewegungen.

Die höhere kognitive Verarbeitung findet im inferotemporalen Cortex statt. Dieser ist unter anderem für die Gesichts- und Objekterkennung zuständig. Er steht zusätzlich in Verbindung mit dem Arbeitsgedächtnis im präfrontalem Kortex und dem Langzeitgedächtnis im medialem Temporallappen (Hudspeth et al. 2013). Dazu lassen sich verschiedene Areale im Kortex isolieren, die auf bestimmte Kategorien des Gesehenem reagieren und daher für ihre Verarbeitung zuständig sind (Troiani et al. 2012). Es spielen die gesamten Informationen und globalen Eigenschaften aus dem Blickfeld eine Rolle. Es lassen sich separate Areale für Gesichter (Kanwisher et al. 1997), Körper (Downing et al. 2001) und Szenen (Epstein und Kanwisher 1998) finden. Vor allem das Areal für Szenen ist aktiv bei der Betrachtung von Objekten, die im Kontext eines passenden Hintergrundes gezeigt werden (Troiani et al. 2012).

2.1.2 Augenbewegung für die visuelle Wahrnehmung

Augenbewegungen machen es dem Sehsystem möglich, ohne eine Kopfbewegung die hochauflösende *Fovea centralis* auf verschiedene Abschnitte des zu Betrachtenden zu lenken und so einen Fokus für die Wahrnehmung zu setzen (Hudspeth et al. 2013). Eine besondere Bedeutung kommt dabei den Sakkaden zu (Yarbus 1967). Hierbei werden durch schnelle Augenbewegungen neue Fixationspunkte gesetzt und so der Blick auf die Strukturen gelenkt, die von besonderem Interesse sind und daher unsere Aufmerksamkeit erregen. Eine Sakkade kann dabei von 0,01 s bis 0,07 s dauern und maximal 20° Sehwinkel unter normalen Bedingungen überwinden. Das Auge kann während einer Sakkade eine Geschwindigkeit von $450^\circ/\text{s}$ erreichen. Obwohl das visuelle Feld durch Sakkaden unterbrochen wird, bleibt der Seheindruck stabil (Wurtz 2008). Als Grund dafür werden drei Ursachen angenommen: Die Erste ist die visuelle *Reafference* (von Holst und Mittelstaedt 1950). Dies bedeutet, dass durch die Veränderung des visuell Gezeigten die Bewegung der Augen wahrgenommen wird. Die Zweite ist das Propriozeptionssignal der Augenmuskeln. Die Dritte ist die Kopie des Signals, das für die Augenmuskeln bestimmt ist. Durch diese Informationen gelingt es dem visuellen System, einen sta-

bilen Bildeindruck ohne Verschwimmen zwischen den Sakkaden zu erzeugen. So ist es dem menschlichen Sehapparat möglich die Umgebung effizient zu betrachten. Die genau stattfindenden neuronalen Prozesse sind Grundlage aktueller Forschung.

2.2 Aufmerksamkeit

Aufmerksamkeit wird im Sehsystem über zwei parallel ablaufende und miteinander interagierende Systeme vermittelt (Hudspeth et al. 2013, Treisman und Gelade 1980). Das erste System beruht auf *top-down* Verfahren, die die Aufmerksamkeit lenken. Auf Grundlage persönlicher Erfahrungen wird das visuelle Signal wahrgenommen und im Anschluss bewusst ein neuer Fixationspunkt gesetzt. Hierbei wird Aufmerksamkeit vom Sehapparat über Sakkaden vermittelt. Ein Beispiel dafür ist das schnelle Betrachten menschlicher Gesichter (Yarbus 1967). Dies wird in Abb. 8 dargestellt. Die schwarzen Punkte stehen für die Fixationspunkte und die schwarze Linie für die Sakkadenbewegung.

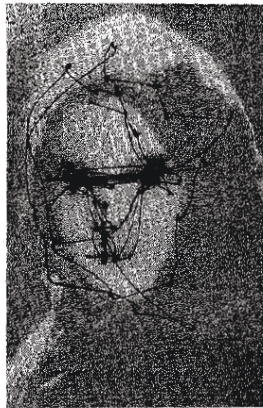


Abbildung 8: Beispiel für Sakkaden bei Betrachtung eines menschlichen Gesichtes. Modifiziert nach Yarbus (1967)

Das zweite System nutzt *bottom-up* Verfahren, um schnell die Aufmerksamkeit auf neue Eindrücke zu lenken. Es nutzt grundlegende Bildparameter wie Helligkeit, Farbe und Orientierung von Linien (siehe Abb. 9) (auch *low-level* Bildeigenschaften genannt), um den Fokus vor dem eigentlichen Wahrnehmungsprozess zu setzen (Treisman 1986). Dieses wird durch eine Karte von grundlegenden Bildeigenschaften (Farbe, Orientierung, Größe,

räumliche Erkennung) erreicht, die die Basis für weitere *bottom-up* Verfahren aus dem Erkennungsnetzwerk darstellt.

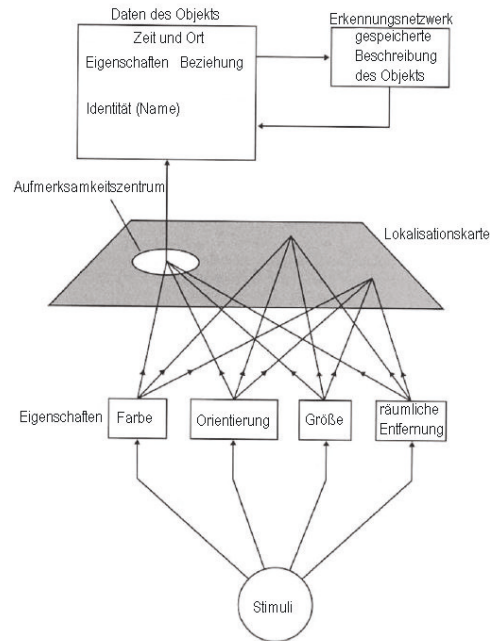


Abbildung 9: *Bottom-up* Aufmerksamkeitsmodell. Modifiziert nach Treisman (1986)

Beide Verfahren interagieren dauerhaft miteinander (Connor et al. 2004). Sobald ein visueller Reiz präsentiert wird, übernehmen zunächst die *bottom-up* Verfahren, da sie nicht auf höheren kognitiven Prozessen, sondern auf grundlegende Bildeigenschaften beruhen. Die *top-down* Verfahren fangen 100ms später an, den Aufmerksamkeitsprozess zu modulieren.

2.2.1 Salienz-Algorithmen

Die Bildparameter, die die menschliche Aufmerksamkeit erregen, werden in der Sehforschung - ähnlich wie in der Psychologie (Hogg und Vaughan 2010) - auch als salient bezeichnet. Dies steht daher für Strukturen, die unsere Aufmerksamkeit auf sich ziehen und so im Verhältnis zu anderen Stimuli herausstechen. Daneben werden noch weitere Eigenschaften mit Salienz in Verbindung gebracht. Hierunter fallen zum einen kognitive Prozesse, Emotionen und Ziele (Henderson et al. 2007, Frintrop et al. 2010, Niu et al. 2012), aber auch die im Abschnitt 2.2 beschriebenen *low-level* Charakteristika (Treisman 1986).

Über diese *low-level* Charakteristika versuchen Salienzalgorithmen die salienten und damit verbundenen Aufmerksamkeit erregenden Strukturen zu markieren (Borji et al. 2013). Der Anwendungsbereich für solche Algorithmen ist sehr groß und reicht u. a. von Objekterkennung (Walther und Koch 2006), graphischer Anwendungen wie *rendern* von Photographie (DeCarlo und Santella 2002) oder automatisiertes Zuschneiden von Bildern (Ardizzone et al. 2013), bis zu Retina-Implantaten als Sehprothese (Parikh et al. 2010).

Insgesamt wurden über die Jahre sehr viele verschiedene Algorithmen entwickelt, wobei diese unter anderem aufgabenspezifisch sind. Deshalb wählten wir für meine Arbeit zwei anerkannte Salienzalgorithmen aus (Borji et al. 2013, Ardizzone et al. 2013). Unsere Wahl fiel auf den Algorithmus nach Itti und Koch (Itti et al. 1998), wobei es sich um einen der ersten Salienzalgorithmen überhaupt handelt (siehe Abschnitt 2.2.2). Der zweiten Algorithmus, den wir auswählten, war der *Graph-Based Visual Saliency (GBVS)* (Harel et al. 2007) (siehe Abschnitt 2.2.3), der die besten Ergebnisse bei natürlichen Szenen in einer Metaanalyse von 35 Salienzalgorithmen lieferte (Borji et al. 2013). Jedoch tritt bei beiden Algorithmen ein *central tendency bias* auf, da die Salienzwerte zum Rand eines Bildes hin abnehmen (Harel et al. 2007, Borji et al. 2013). Dieser lässt sich außerdem bei Augenbewegungen finden, wenn ein Bild betrachtet wird (Bindemann 2010). Um diesen systematischen Fehler zu umgehen, wurde die in Abschnitt 4.1.1 dargestellte Methodik angewendet.

2.2.2 Itti und Koch Salienz-Algorithmus (ITTI)

Itti et al. (1998) simulieren mit ihrem Salienzalgorithmus (ITTI) die *bottom-up* Verfahren der visuellen Aufmerksamkeit. In Anlehnung an Triesman (1986) wird das Bild zunächst in drei verschiedene Eigenschaften zerlegt (siehe Abb. 10). Der erste Parameter, der bei dem Algorithmus eine Rolle spielt, ist die Farbe. Im Bezug auf das menschliche Sehsystem wird hierfür der RGB-Farbraum benutzt, wobei über eine Normalisierung die Eigenschaft Farbe von der Intensität getrennt wird. Um hierbei sicherzustellen, dass dunkle Areale als nicht salient gekennzeichnet werden, wird eine Grenze im Verhältnis zum gesamten Bild festgelegt.

Intensität ist der zweite Parameter. Dieser wird über Kontraste generiert, wobei die Umgebung eines Bildpunktes mit ihm verglichen wird.

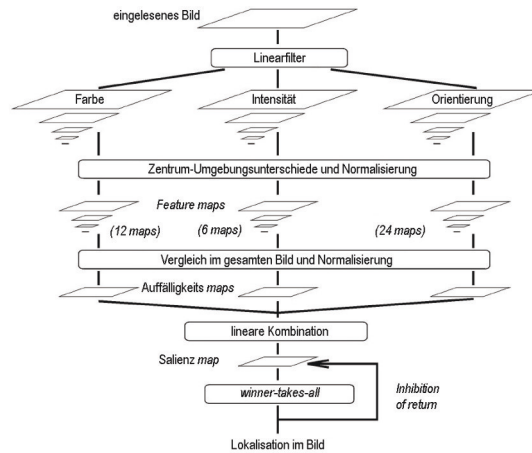


Abbildung 10: Der generelle Aufbau des Itti und Koch Salienz Modells modifiziert nach Itti et al. (1998)

Der dritte Parameter nutzt ebenfalls die Intensität des zweiten Parameters, wobei durch einen Linearfilter die Orientierung ermittelt wird.

Der Algorithmus vergleicht, angelehnt an die rezeptiven Felder (siehe Abschnitt 2.1.1.2), die Bildpixel in der Mitte des zu analysierenden Areals mit denen in der Umgebung. Im nächsten Schritt wird das Bild „normalisiert“, wodurch bis dahin entstandene lokale Maxima verstärkt und Bildareale, die viele Maxima nebeneinander besitzen, geschwächt werden. Dies dient dazu, die laterale Inhibition (siehe Abschnitt 2.1.1.1) nachzuempfinden. Ein „winner-takes-all“-Prinzip verteilt im letzten Schritt die Salienzmaxima im Bild.

Beispiele für Salienzkarten des ITTI-Algorithmus sind in Abb. 11 (Bilder C und D) gezeigt. Hierbei stehen die blauen Areale für die nicht salienten Bildabschnitte und die roten für die Salienten.

2.2.3 Graph-Based Visual Saliency Algorithmus (GBVS)

Harel et al. (2007) wählten im Vergleich zu dem von Itti et al. (1998) vorgeschlagenen Salienzalgorithmus eine andere Herangehensweise. Zunächst werden auch hier die grundlegenden Parameter definiert und die damit verbundenen *feature maps* (Farbe, Intensität und Orientierung) erzeugt. Im Unterschied zum ITTI-Algorithmus werden die rezeptiven Felder über Markow-Ketten (Behrends 2000) verglichen. So entstehen im Bereich des Gleichgewichts dieser Ketten Werte. Anschließend erfolgt auch hier eine Normalisierung, wobei im Vergleich zu ITTI-Algorithmus wiederum Markow-Ketten und kein „winner-

takes-all“- Prinzip verwendet werden. Dadurch sollen die Interaktionen zwischen den Neuronen der Sehbahn nachempfunden werden, wobei die Knoten einer Markow-Kette die Interaktion repräsentieren. Beispiele für Ergebnisse des GBVS-Algorithmus ist in Abb. 11 (Bilder E und F) gezeigt.

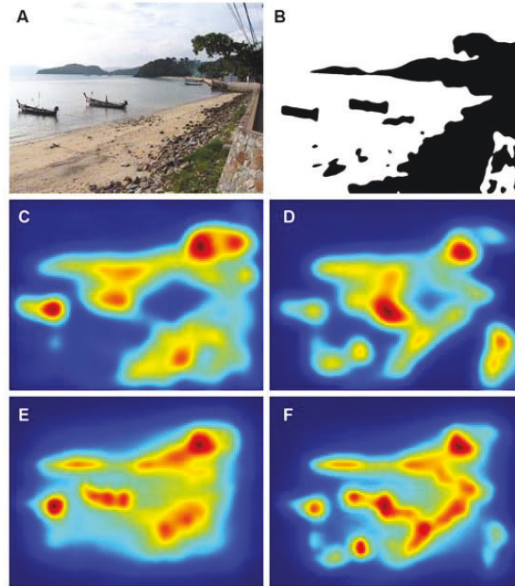


Abbildung 11: Beispiele für Ergebnisse der Salienz-Algorithmen. Werte dargestellt als *heatmap*. A und B Originalbilder. C und D, Salienzkarten nach dem ITTI-Algorithmus. E und F, Salienzkarten nach dem GBVS-Algorithmus. (aus: Abeln et al. 2016)

2.3 Balance im Kontext visueller Ästhetik

Die Wahrnehmung der Umgebung erfolgt in verschiedenen kortikalen Arealen, die die verschiedenen Eigenschaften des Blickfeldes einbeziehen (Troiani et al. 2012). Da sich über das *Croppen* eines Bildes die Komposition und damit Ästhetik verändern lässt, wird in diesem Abschnitt darauf eingegangen, wie sich die Anordnung der sich im Blickfeld befindenden Reize auf die ästhetische Wahrnehmung auswirkt.

Die Komposition eines Bildes spielt eine entscheidene Rolle bei der ästhetischen Wahrnehmung (Arnheim 1954, Palmer et al. 2013). Nach Arnheim (1988) ist hierbei vor allem die Mitte des Bildes entscheidend, jedoch untermauerte er diesen Hypothese nicht mit

Experimenten. Bei gerahmten Bildern zeigten Palmer et al. (2008), dass vor allem Elemente, die auf den Betrachter zeigen, in der Mitte des Bildes platziert werden. Außerdem erfolgt bei Elementen, die seitlich ins Bild gerichtet sind, bevorzugt eine Anordnung, die mehr rechts bzw. links von der Bildmitte ist. Zusätzlich bewerteten Beobachter bei der Platzierung eines Kreises diesen in der Mitte am ästhetischsten (Palmer und Guidi 2011). Weitere hohe Bewertungen erzielten Kreise, die sich entlang der vertikalen bzw. horizontalen Achse befanden, wobei die Kreise, die sich am Bildrand befanden, als weniger ästhetisch empfunden wurden.

Locher et al. (1996) untersuchte bei seinen Studien über Reproduktionen von Malerei aus dem zwanzigsten Jahrhundert, wie Balance von Probanden wahrgenommen wird. Die Autoren zeigten, dass die Bildelemente balanciert um einen Mittelpunkt angeordnet sind, sodass sich dort ein Schwerpunkt bildet. Hierbei ist die gesamte Bildkomposition von Bedeutung, somit würden bei Bildveränderungen wie Zuschneiden neue Mittelpunkte entstehen (McManus et al. 1985). Außerdem steht die Balance mit dem geometrischen Bildmittelpunkt in Zusammenhang (Locher et al. 1998, Locher et al. 2001). Probanden ordneten in der Arbeit von Locher et al. (1998) Strukturen so an, dass sie mit dem geometrischen Mittelpunkt in Verbindung stehen. Weder das Bildformat, noch die Art der Bildelemente hatten einen Einfluss auf die Anordnung der Balance mit dem geometrischen Mittelpunkt. Auch die Wahl der Farbe und die Größe der dazugehörigen Fläche wirkt sich auf die Balance aus. Locher et al. (2005) zeigten dies anhand von Piet Mondrians abstrakten Bildern. Im Vergleich zu den farblich veränderten Varianten bezeichneten die Probanden die Originale als mehr um den Mittelpunkt balanciert. Dazu konnte gezeigt werden, dass es Probanden möglich ist, Balance sowohl in schwarz-weißen Bildern als auch Farbmalerie in Zeiträumen von 100 *ms* zu evaluieren (Locher und Nagy 1996). Daher können Probanden schon mit *bottom-up* (siehe Abschnitt 2.2) Prozessen Balance beurteilen.

Dieses Konzept von Balance konnten Gershoni und Hochstein (2011) bestätigen. Sie untersuchten japanische Kalligraphie und ließen Probanden nach kurzen Betrachtungszeiten (200 *ms*) diese nach Balance bewerten. Insbesondere Linien und Elemente, die sich auf den horizontalen und vertikalen Achsen befanden, hatten auf die empfundene Balance einen besonderen Einfluss.

2.4 Computational Aesthetics

Das Feld der *computational aesthetics* versucht unter zu Hilfenahme von computergestützten Analysen, Grundlagen der ästhetische Wahrnehmung zu beschreiben. Im Fall von Photographie werden sich *Community* Internetseiten wie *Flickr.com*, wo z. Zt. täglich 6,5 Millionen Fotos hochgeladen werden (Bhattacharya et al. 2010), zunutze gemacht, um große Datensätze zu untersuchen und unter Berücksichtigung der Bewertung der jeweiligen *Community* weitere nützliche Parameter zu identifizieren. Auch hier wird schon lange davon ausgegangen, dass *low-level* Bildeigenschaften, wie bspw. Farbe, Orientierung und Kontraste, bei der ästhetischen Gestaltung eine Rolle spielen. Es konnte gezeigt werden, dass schon wenige Bildeigenschaften wie Farbe und Kontrast ausreichen um zwischen Photographie von Laien und Photographen zu unterscheiden (Tong et al. 2004). Aber auch Bildeigenschaften wie Farbsättigung und Seitenverhältnisse reichen aus, um einen Hinweis darüber zu erhalten, ob ein Bild in der jeweiligen *Community* hoch oder niedrig bewertet wurde (Datta et al. 2006). Aus den Bewertungen der *Community* lässt sich schließen, dass manche Menschen besser darin sind, den richtigen Ausschnitt für ein Foto zu wählen als andere. Außerdem war es durch ITTI-Algorithmus möglich, zwischen Fotos von Photographen und *snap-shots*, d.h. nicht professionelle Fotos, zu differenzieren (Wong und Low 2009). Wong und Low (2011) zeigten zudem, dass es unter Berücksichtigung der Salienz möglich ist, Fotos durch Zuschneiden ansprechender zu gestalten.

2.5 Cropping

In der Arbeit von McManus et al. (2011b), die auch die Grundlage für unsere Arbeit darstellt, wurde das manuelle Zuschneiden von Photographie (Englisch: *cropping*) in mehreren Experimenten untersucht.

Das erste Experiment in dieser Arbeit diente dazu, herauszufinden, ob das Zuschneiden zuverlässig von den Probanden durchgeführt wurde. Hierfür musste jeder Proband zweimal die gleichen 40 Photographien zuschneiden, um sie so schön wie möglich aussehen zu lassen. Die originalen Photographien entstammten aus der Sammlung der Autoren, dem Internet und aus der Arbeit von Santella et al. (2006). Durch dieses Experiment konnte gezeigt werden, dass sowohl innerhalb der Gruppen, als auch für jeden einzelnen Probanden, die medianen Mittelpunkte der zugeschnittenen Bilder in beiden Fällen ähnlich waren, da sich die Mittelpunkte der Zuschnitte der gleichen Photographie ähnelten. Im zweiten Experiment wurde überprüft, ob einige Probanden besser im Zuschneiden

sind als andere. Hierfür wurden die zugeschnittenen Bilder von sechs Probanden aus dem ersten Experiment ausgewählt, sodass es für jedes originale Bild 6 zugeschnittene Varianten gab. Anschließend wurden diese sechs zugeschnittenen Bilder aus einer Photographie einem anderen Probanden zugeteilt mit der Aufgabe, eine Rangliste nach den besten Zuschnitten zu erstellen. Es zeigte sich, dass einige Probanden besser bewertete Zuschnitte erstellten als andere.

Das dritte Experiment diente der Überprüfung der Ergebnisse des vorherigen Experimentes. Dafür wurde eine neue Gruppe aus Probanden ausgewählt, die jeweils zwei zugeschnittene Bilder pro Photographie direkt verglichen. Diese Gruppe von Probanden bestand aus 41 Nichtexperten (Studenten am University College London) und zehn Experten (Studenten, die einen Masterabschluss in Photographie vom Royal College of Art in London anstreben). Die Nichtexperten konnten die Ergebnisse aus dem zweiten Experiment, dass es Probanden gibt, die besser Fotos zuschneiden, bestätigen, die Experten jedoch nicht. Die Autoren stellen die These auf, dass dies wahrscheinlich damit zusammenhängt, dass Experten weniger bereit sind, Aussagen über Schönheit zu treffen (McManus et al. 1993).

Da die vorherigen Experimente ausschließlich farbige Photographien nutzten, galt es im vierten Experiment diese mit Graustufen- und Mooney-Bildern, d.h. Bilder mit ausschließlich schwarzen und weißen Flächen (Mooney 1956) (siehe Abb. 11, Bild B), zu vergleichen. Dafür wurden aus den originalen Photographien aus dem ersten Experiment 16 Photographien ausgewählt und die korrespondierenden Graustufen- und Mooney-Bilder erstellt. Die Aufgabenstellung blieb identisch zum ersten Experiment. Als Probanden dienten die Nichtexperten aus dem dritten Experiment. Hierbei ließ sich bestätigen, dass die Probanden die Fotos zuverlässig zuschneiden und dass die Graustufenbilder identisch zu den Farbfotos behandelt werden. Die Mooney-Version hingegen unterschieden sich zu dessen Farbversion. Jedoch spielte hierbei die Reihenfolge eine Bedeutung, in der die Fotos und Bilder zu bearbeiten waren. Wenn die Farbphotographien vor den dazugehörenden Mooney-Bildern dargestellt worden waren, ähnelten sich die zugeschnittenen Bilder der Farb- und Mooney-Version mehr. Wenn zuerst die Mooney-Versionen gezeigt wurden, ähnelten sich die Zuschnitte nicht. Hieraus lässt sich schlussfolgern, dass sowohl Farbe als auch das Dargestellte das Zuschneiden beeinflusst, da bei Reduzierung des Bildes auf weiße und schwarze Flächen sich das Zuschneiden ändert.

Im fünften Experiment wurden schließlich die Experten aus dem dritten Experiment gebeten, die Farbfotos aus dem vierten Experiment zuzuschneiden. Es zeigte sich, dass Experten die Aufgabe etwas anders lösen als die Nichtexperten. Die Mittelpunkte der

zugeschnittenen Bilder variierten mehr auf der Horizontal- und besonders der Vertikalachsen im Vergleich zu den Bildern der Nichtexperten aus dem vierten Experiment.

Im sechsten Experiment wurden von 38 neuen Probanden die farbigen zugeschnittenen Bilder der Nichtexperten aus dem vierten Experiment mit denen der Experten aus dem vorherigen verglichen. Dieses erfolgte in einem direkten Vergleich, wobei immer ein Bild eines Experten und das eines Laiens gegenübergestellt wurde. Eine signifikante Präferenz für eine der beiden Gruppen ließ sich hierbei nicht finden.

Somit lässt sich zusammenfassen, dass das *Cropping* von Probanden wiederholbar durchgeführt wird. Weiterhin scheint es einigen Personen zu gelingen, die Aufgabe besser zu lösen als anderen. Aus dem Vergleich mit den Mooney-Versionen lässt sich ableiten, dass sowohl Eigenschaften wie der Inhalt des Bildes, als auch Farbe und Form eine Rolle beim Zuschneiden zu spielen.

In einer anschließenden Studie (McManus et al. 2011a) wurde u. a. versucht, eine Verbindung zwischen dem Zuschneiden und der Balance-Theorie nach Arnheim (siehe Abschnitt 2.3) zu finden. Hierfür wurden die farbigen zugeschnittenen Bilder der Experten und Nichtexperten aus dem dritten Experiment der vorherigen Arbeit genutzt. Die Mittelpunkte der zugeschnittenen Bilder wurden dann anschließend mit den vertikalen und horizontalen Achsen, sowie den Diagonalen, verglichen, an denen das zugeschnittene Bild balanciert wäre. Es konnte bei Nichtexperten gezeigt werden, dass die Mittelpunkte näher an den Achsen sind als bei zufällig zugeschnittenen Bildern. Interessanterweise produzierten die Experten keine zugeschnittenen Bilder, die sich näher an den Achsen befanden. Die Tendenz zwischen den Experten und Nichtexperten war jedoch nicht signifikant.

2.6 Fragestellung

McManus et al. (2011b, 2011a) war es in ihren Arbeiten zwar möglich zu zeigen, dass es Bildparameter geben muss, die das Zuschneiden der Probanden beeinflussen, sie konnten diese aber nicht identifizieren. Auch der Versuch, das Modell der Balance auf Grundlage der Koordinaten des geometrischen Mittelpunktes (siehe Abschnitt 2.3) anzuwenden, blieb ohne zufriedenstellende Ergebnisse. Somit war es den Autoren nicht möglich Parameter zu identifizieren, die beim ästhetischen Zuschneiden von Photographie eine Rolle spielen.

Aus dem Feld der *computational aesthetics* lässt sich ein weiterer Ansatz ableiten. Wong und Low (2011) zeigten bspw., dass Salienz als ein Parameter genutzt werden kann,

um ästhetisch ansprechendere Bilder zu erzeugen. In der computergestützten Bildbearbeitung findet Salienz dazu schon länger Verwendung (Ardizzone et al. 2013, Suh et al. 2003, Ciocca et al. 2007) und es wurde ebenfalls versucht, unter zusätzlichem Einbezug von formellen Bildkompositionsregeln wie Balance eine Messmethode für Ästhetik zu entwickeln (Liu et al. 2010).

Trotz dieser Anwendungen wurde bisher noch nicht gezeigt, dass Salienz eine Rolle beim Zuschneiden von Photographie spielt und ob sie im Zusammenhang mit Balance steht. Da es sich bei Salienzalgorithmen um den Versuch handelt, das menschliche Sehsystem nachzuahmen, ist es ebenso interessant, wie sich die Algorithmen im Vergleich zu Augenbewegungsdaten verhalten - da die Salienzalgorithmen auf *low-level* Bildeigenschaften beruhen und Augenbewegungen sowohl von *bottom-up*, als auch von *top-down* Prozessen gesteuert werden.

So lassen sich für meine Arbeit folgende Fragen für die Experimente ableiten:

- Wählen Probanden eher die Bildausschnitte aus, die einen höheren Salienzwert und längere Augenfixationsdauern besitzen?
- Sind Salienz und Bildfixationen in den Bildausschnitten, die von den Probanden erstellt worden waren, balanciert verteilt?
- Ist die Auswahl des Bildausschnittes abhängig vom Inhalt des Bildes? Um dies zu überprüfen, wurden Mooneys-Bilder (Mooney 1956) verwendet.
- Unterscheiden sich Experten und Nichtexperten in der Auswahl von Bildausschnitten?

3 Veröffentlichung



Preference for Well-Balanced Saliency in Details Cropped from Photographs

Jonas Abeln¹, Leonie Fresz¹, Seyed Ali Amirshahi^{1,2}, I. Chris McManus³, Michael Koch^{1,2}, Helene Kreysa⁴ and Christoph Redies^{1*}

¹ Experimental Aesthetics Group, Institute of Anatomy I, University of Jena School of Medicine, Jena University Hospital, Jena, Germany, ² Computer Vision Group, Institute of Computer Science, Friedrich Schiller University, Jena, Germany, ³ Division of Psychology and Language Sciences, University College London, London, UK, ⁴ General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University, Jena, Germany

Photographic cropping is the act of selecting part of a photograph to enhance its aesthetic appearance or visual impact. It is common practice with both professional (expert) and amateur (non-expert) photographers. In a psychometric study, McManus et al. (2011b) showed that participants cropped photographs confidently and reliably. Experts tended to select details from a wider range of positions than non-experts, but other croppers did not generally prefer details that were selected by experts. It remained unclear, however, on what grounds participants selected particular details from a photograph while avoiding other details. One of the factors contributing to cropping decision may be visual saliency. Indeed, various saliency-based computer algorithms are available for the automatic cropping of photographs. However, careful experimental studies on the relation between saliency and cropping are lacking to date. In the present study, we re-analyzed the data from the studies by McManus et al. (2011a,b), focusing on statistical image properties. We calculated saliency-based measures for details selected and details avoided during cropping. As expected, we found that selected details contain regions of higher saliency than avoided details on average. Moreover, the saliency center-of-mass was closer to the geometrical center in selected details than in avoided details. Results were confirmed in an eye tracking study with the same dataset of images. Interestingly, the observed regularities in cropping behavior were less pronounced for experts than for non-experts. In summary, our results suggest that, during cropping, participants tend to select salient regions and place them in an image composition that is well-balanced with respect to the distribution of saliency. Our study contributes to the knowledge of perceptual bottom-up features that are germane to aesthetic decisions in photography and their variability in non-experts and experts.

Keywords: experimental aesthetics, photography, visual saliency, eye movements, visual balance, statistical image properties, bottom-up processing

OPEN ACCESS

Edited by:

Mark E. McCourt,
North Dakota State University, USA

Reviewed by:

Mike Nicholls,
Flinders University, Australia
Gina M. Grimshaw,
Victoria University of Wellington,
New Zealand

*Correspondence:

Christoph Redies
christoph.redies@med.uni-jena.de

Received: 09 October 2015

Accepted: 14 December 2015

Published: 11 January 2016

Citation:

Abeln J, Fresz L, Amirshahi SA, McManus IC, Koch M, Kreysa H and Redies C (2016) Preference for Well-Balanced Saliency in Details Cropped from Photographs. *Front. Hum. Neurosci.* 9:704. doi: 10.3389/fnhum.2015.00704

INTRODUCTION

Since the invention of photography by Nicéphore Niépce and Louis Daguerre in the early 19th century, a central issue in photography has been how to choose a suitable viewpoint and a proper viewing window, through which a small part of the real world is captured on a photograph in a visually pleasing way. With the recent rise of affordable digital cameras, this question has become of interest not only to professional photographers but also to a wide audience of laypersons who take photographs casually on many occasions.

A task closely related to choosing a good photographic detail is the act of selecting part of a photograph that already exists (photographic “cropping”), to enhance its visual impact or its aesthetics. Cropping is not only common practice for professional photographers, but also non-expert photographers crop their photographs at home with the help of widely available computer programs, such as Adobe Photoshop. In a conceptual sense, the taking of the original photograph can be seen as “cropping the visual world” through the viewfinder or screen of the camera.

In a first experimental study on the psychometrics of photographic cropping, McManus et al. (2011b) studied how reliably people crop photographs, how much variability there is between individual croppers in the quality of cropping decisions, and whether expert (professional) photographers crop differently from non-experts. In their study, participants cropped everyday photographs, which were displayed on a computer screen in a laboratory setting, to half of their linear size (1/4 of the area). The study revealed that both experts and non-experts cropped photographs confidently and reliably. Additionally, independent observers preferred aesthetically the crops of some participants over those of other croppers, suggesting individual differences in expertise. Experts tended to select details from a wider range of possible cropping positions than non-experts, but judges did not generally prefer the expert crops.

Although subjectively, most people feel rather confident about how to crop a photograph (McManus et al., 2011b), the criteria, on which experts and non-experts ground their cropping decisions, are not well understood. Without any doubt, one major issue for cropping a photograph is which key objects or parts of a scene people regard as essential to be included in the photograph. Besides such content-based cropping criteria, some professional photographers and psychologists claim that other criteria, which relate to formal rules of image composition, should also be followed (Arnheim, 1954, 1982; Palmer et al., 2008; Liu et al., 2010). In photography, several compositional rules were made explicit (Hicks, 2005) and some of these rules have been scrutinized recently at the experimental level.

For example, Bruno et al. (2014) studied three well-known principles of photographic composition (the rule of thirds, the golden ratio rule and the eye centering principle) in self-portraits taken with a hand-held smartphone camera (“selfies”). The rule of thirds postulates that images are aesthetically more pleasing if important compositional elements are placed close to one of the third lines of the image. The golden ratio rule says that

rectangular images are preferred aesthetically if the ratio of the larger side over the smaller side (a/b) equals the ratio of the sum of the two sides over the larger side ($a/[a+b]$), i.e., if the ratio is about 1.618. The eye centering principle claims that, in portrait images, one of the eyes of a depicted person should be centered horizontally. Bruno et al. (2014) found that non-professional photographers do not follow any of these rules. In another recent study, Amirshahi et al. (2014) investigated the rule of thirds in a large set of photographs and paintings and did not detect any preference for images that followed the rule of thirds, compared to those that did not.

Another principle commonly used to assess photographic quality is visual saliency (Frintrop et al., 2010). The term saliency denotes any number of properties that make an object or feature stand out from its background; such properties can therefore attract visual attention and direct the observer's gaze to a particular (salient) region of an image. The properties that confer saliency to image regions can be tied to cognitive factors, emotional value, or goals (Henderson et al., 2007; Frintrop et al., 2010; Niu et al., 2012), but they can also be visual characteristics, such as color, luminance, contrast, and spatial frequency. In the present work, we use computer-based algorithms that are based on low-level visual characteristics to measure visual saliency and to determine how well such characteristics determine cropping behavior. There are many different algorithms for calculating visual saliency and each of them uses a different set of low-level visual features to predict eye movements. In an exhaustive survey of these algorithms, Borji et al. (2013) found that the performance of 35 state-of-the-art models differed between each other and depended on the respective task. Because the magnitude of the bottom-up contribution to active gaze control is still controversial, we assess the performance of two saliency models by comparing computed saliency maps with experimentally measured eye movements in the present study.

Finally, the century-old concept of pictorial balance plays a role in the subjective evaluation of visual stimuli, including photographs and artworks (for example, see Ross, 1907; Howard, 1914; Arnheim, 1954). Pictorial balance is thought to unify picture elements into a cohesive composition. In a recent overview on pictorial balance, Gershoni and Hochstein (2011; p. 509) stated that “balance is achieved by structural properties working like mechanical weights with a fulcrum at the picture's center, on which an imaginary lever is poised, so that heavy weights can be counterbalanced by lighter ones located further from the center”. In the Arnheim-Ross model of visual balance, the framework of these levers is set on the major geometrical axes (vertical, horizontal and diagonal) that intersect in the image center (Ross, 1907; Arnheim, 1954). For example, the image is thought to be balanced on the horizontal axis if as much force is present on the right side as on the left side of the axis (image center).

Relatively few studies have linked the subjective impression of pictorial balance to objective physical properties. In a follow-up to their cropping study, McManus et al. (2011a) did not find any evidence for a correlation of cropping selections with the

Arnheim-Ross model of visual balance. As an objective measure of balance for a given physical property, such as luminance, the center-of-mass has been used. It represents the unique point where the weighted distributed mass of the property is equal on either side of an image axis; an image is more balanced if this point is located closer to the geometrical image center. Jahanian et al. (2015) studied pictorial balance by modeling visual weight with the saliency of low-level visual features. For a large set of aesthetically pleasing photographs, they obtained results compatible with Arnheim (1954) concept of major axes of composition, including the relevance of the image center. Besides the structural properties mentioned above, there may be additional formal rules that are followed intuitively and await a description in physical or perceptual terms.

Despite the overall difficulties in explaining the quality of photographic composition by unique and simple rules, there is evidence that combinations of multiple low-level image properties can be used to predict the aesthetic outcome of photographic cropping. This evidence stems mostly from the field of computational aesthetics, where researchers used large datasets of photographs posted on websites, such as *Photo.net* or *Flickr.com*, which have been rated by the photographic community. For example, Tong et al. (2004) were among the first to successfully use a set of low-level image features in an image classification task that distinguished between professional photographs and photographs by home users. Their classifier also predicted the quality ratings by human observers with relatively high accuracy. Datta et al. (2006) trained an automated classifier to distinguish between high- and low-rating photographs, based on 56 low-level image features that related to rules of thumb and common intuition in photography, such as colorfulness, the rule of thirds, various texture and shape measures, size and aspect ratio, and depth of field. With a short-list of 15 of these measures, they achieved an accuracy rate of around 70% in predicting highest/lowest ratings of a large dataset of 1664 photographs. Wong and Low (2009) used a saliency-enhanced approach for the classification of professional photographs and snap-shots. With this higher-level approach, they achieved classification rates of up to 79%. Sun et al. (2009) integrated top-down supervision and personalized parameters into a bottom-up attentional model to predict image quality. Low-level statistical image properties have also been studied in artworks (for reviews, see Graham and Redies, 2010; Redies, 2015).

Based on saliency calculations of low-level image properties, a large number of computational tools have been devised to carry out cropping decisions automatically (also called *image retargeting*; for reviews, see Vaquero et al., 2010; Ardizzone et al., 2013). For example, Ardizzone et al. (2013) compared five different saliency algorithms for automatic cropping and obtained favorable results with all of them. Two other studies on saliency-based cropping applications were published by Suh et al. (2003) who shrank original images to produce easily recognized thumbnails for image retrieval, and by Ciocca et al. (2007) who redesigned large images for small screens with an adaptive visual attention

model that incorporated semantic information. Santella et al. (2006) introduced an interactive method by which saliency-based cropping was combined with information about important image content that was obtained from eye tracking data. They demonstrated experimentally that viewers prefer gaze-based crops to fully automated crops. Liu et al. (2010) combined several compositional rules to derive an aesthetic measure for evaluating cropped photographs. Other studies that have used similar approaches are too numerous to be discussed here in detail.

Although saliency-based cropping methods are used widely, relatively few psychological studies have investigated the role of saliency in cropping decisions. In the present work, we re-analyzed the experimental data that were obtained by McManus et al. (2011a,b). To assess the relevance of the saliency maps for gaze control, we obtained maps of eye fixations for comparison.

In the studies by McManus et al. (2011a,b), participants selected a large number of rectangular details from a series of every-day photographs (see above). In addition to the two sets of details that were selected by non-experts and experts (here called *selected* details), we systematically examined all details that were not selected during the cropping of the same photographs (here called *avoided* details). We asked the following questions:

1. Did participants select details during cropping that contained higher overall saliency and attracted longer eye fixations, compared to the avoided details? We anticipated that this would be the case because participants are more likely to select image regions that attract their attention during the cropping procedure (Ardizzone et al., 2013).
2. Did participants select details that had a higher pictorial balance of saliency and of total dwelling (fixation) time than avoided details (as assumed, for example, by Liu et al., 2010; Chen et al., 2013; Wang et al., 2015)? As a measure of pictorial balance, we determined how close the center-of-mass values for visual saliency and for total dwelling times were to the geometrical image center of the details.
3. Do regularities in cropping decisions depend on the recognition of image content? To answer this question, we analyzed versions of the original photographs that had been transformed to binarized, monochrome images, in which the brighter parts of the image were rendered white and the darker parts black (Mooney, 1956; here called *Mooney images*; for details of their generation, see McManus et al., 2011b). This manipulation made it very difficult to recognize image content. McManus et al. (2011b) reported that the Mooney versions of the photographs showed dramatically altered cropping positions, but were still cropped consistently. The authors suggested that image structure, perhaps in the form of low-level image properties may explain this consistency (McManus et al., 2011b). We therefore asked whether cropping decisions for Mooney images were based on a similar pattern of low-level image properties and eye movements as the cropping of the original color photographs.
4. Did the selected details differ between non-expert participants and expert (professional) participants in any of these measures? We expected to find differences because expert

photographers were previously found to crop photographs differently from non-experts (McManus et al., 2011b).

To compute visual saliency in the present study, we used two algorithms, the Itti-Koch model (Itti et al., 1998) and the graph-based visual saliency model (Harel et al., 2007; for details, see “Materials and Methods” Section). The two methods will be referred to as ITTI and GBVS, respectively, in the remainder of this work. Moreover, selected and avoided details were carefully matched to avoid possible artifacts due to a central tendency, which is inherent in both eye gaze patterns and saliency calculations (Harel et al., 2007; Tatler, 2007; Foulsham and Underwood, 2008; Bindemann, 2010). There is a central tendency in gaze patterns because observers tend to fixate the central regions of an image more than its peripheral regions, irrespective of the distribution of image content. Likewise, in the saliency maps calculated with ITTI and GBVS, saliency falls off towards the edges of the image. Both central tendencies were controlled for by analysing pairs of selected and avoided details with matched distances between their centers and the geometrical image center of the original photographs.

MATERIALS AND METHODS

Datasets of Images

The present work re-analyzed a series of four datasets of cropped photographs; some but not all of the data were reported in the studies by McManus et al. (2011a,b). Specifically, the first dataset corresponded to the 20 croppers described in Study 1 of McManus et al. (2011b). The second dataset was obtained from the 41 non-expert and 10 expert participants described in Studies 3, 4 and 5 of McManus et al. (2011b), with the cropping data corresponding to Study 4 of that publication. The third dataset consisted of 36 participants who took part in Study 5 of McManus et al. (2011a), although not all of their data were reported there. Finally, a fourth dataset was from 38 participants, who took part both in Study 6 of McManus et al. (2011b), and study 3 of McManus et al. (2011a), their cropping data being analyzed only in the latter. The four groups carried out 80 separate croppings, although not always of the same images, as different hypotheses were being investigated. All subjects carried out some repeat croppings of some images to assess reliability (see below).

In the following, we will briefly summarize those parts of the studies by McManus et al. (2011a,b) that relate to the cropping procedure. In brief, non-expert participants (mostly college students) and expert participants (master students of photography at an art college) viewed up to 44 color photographs from various private sources and the internet. The color photographs contained ordinary subject matters, which people would normally photograph, such as family, landscapes and townscapes. They were displayed at their original resolution (1024 × 768 pixels) in landscape format at a viewing distance of about 70 cm (field of view of 24 degrees width) and filled the entire screen. Examples are shown in **Figure 1**. In an initial viewing phase, participant saw the entire image with a small yellow rectangle (arrows in **Figures 1A,B**) indicating the focus or subject of the image that had to be included in the subsequent

crop. During the cropping phase, participants saw a rectangular part of the image of one quarter of the size of the original image (512 × 384 pixels; large red rectangles in **Figures 1A,B**), the rest of the screen being black. They could freely move the cropping window around the screen using a computer mouse until they felt that they had achieved a satisfactory crop of the original photograph; this was indicated with a mouse click. A computer program recorded the central positions of the details selected during the cropping. About 71% of the croppings were carried out at first-time presentations and 29% represented repetitions to assess cropping reliability. The same procedure was used for the Mooney versions of a subset of 14 of the original images (see “Introduction” Section; only first-time presentations).

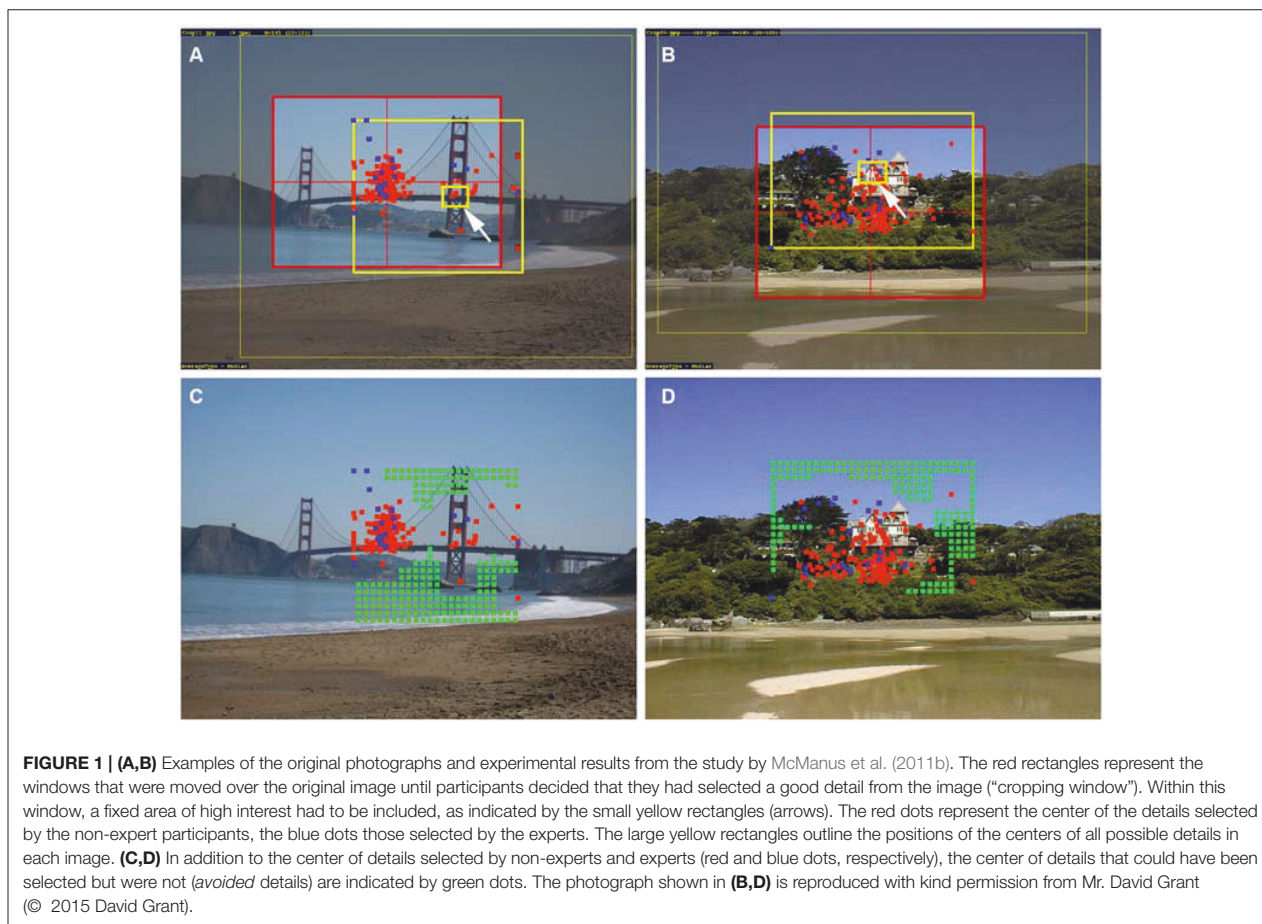
The resulting data formed the basis of our own study, which analyzed a total of 5748 details originally selected by non-experts, and 807 details by experts (**Table 1**). In order to study whether content played a role in the cropping decisions, we also asked whether the Mooney versions differed with respect to the corresponding saliency maps and eye movement patterns. From the study of McManus et al. (2011b), we analyzed a total of 457 details originally selected by non-experts, and 126 details by experts (**Table 1**).

To compare salience and gaze control data between details selected and details that were not selected during cropping (*avoided* details) in the studies by McManus et al. (2011a,b), we mapped all possible cropped details selected by non-experts and experts onto each image. On the original photographs (**Figure 1**), red dots indicate the geometrical centers of the details selected by non-experts, and blue dots those selected by experts. The large yellow rectangles in **Figures 1A,B** outline the areas of the central positions of all possible details in each image. To obtain the details that were avoided by non-experts and experts, we discretized the large yellow rectangles in **Figures 1A,B** by a 64 × 48 grid. As geometrical centers of avoided details, we took all positions that were lying outside the area of squares of 5 × 5 grid points placed on top of the geometrical centers of all selected details in each image. The central positions of the avoided details are indicated by the green dots in **Figures 1C,D**. A total of 5102 avoided details were analyzed for 45 color photographs and 1459 avoided details for 14 Mooney versions (**Table 1**). Note that the avoided details were not obtained separately for non-experts and experts because the differences in the number of selected crops between the two groups may have confounded results.

Eye Tracking

Eye movements were recorded from 34 students, mostly of psychology and medicine (19–38 years old, mean: 23.0, 7 male) who were paid 8 Euros for participating. The Ethics Committee of the University of Jena approved the experiment. Before the experiment, participants were informed about the experimental procedure and agreed to participate by signing a consent form.

The stimuli were 45 original color photographs and 14 Mooney versions that had been selected by both non-experts and experts in the study by McManus et al. (2011b). They were displayed on a BenQ G2200W monitor (resolution 1680 × 1050 pixel) by using the software E-Prime



2.0.822. Before the series of experiments, the monitor was luminance-calibrated (i1Display Pro, X-Rite) and luminance density was set to 200 cd/m². The monitor was positioned at a distance of 74 cm from the eyes in a shaded room with moderate ambient illumination. Presentation of the stimulus was preceded by a fixation cross displayed for 1000 ms. To reduce the influence of the position of the fixation cross on measured eye movements, the crosses were presented in a randomized fashion on either the left, right, top or bottom side of the screen outside the stimulus

area. After the fixation cross disappeared, one of the stimuli (1265 × 949 pixel, extending over 25.7 × 20.0 degrees) was shown in a randomized order at the center of the screen for 5000 ms. Then, the next trial was initiated. The participants had no special task and were simply instructed to look at the images. In order to minimize the recognition of objects in the Mooney images, they were presented as the first block, followed by a block with the original color photographs. Each participant viewed every stimulus image once only.

TABLE 1 | Deviation from the geometrical image center and Euclidean distance to the image center for all (unmatched) details (in percent of normalized width and height, respectively).

		<i>i</i>	<i>n</i>	Width	Height	Euclidean distance
Color photographs	Non-experts	45	5747	0.096 ± 0.072 ^{1,2}	0.090 ± 0.064 ³	0.147 ± 0.072 ⁴
	Experts	35	807	0.114 ± 0.079 ¹	0.108 ± 0.088 ³	0.174 ± 0.093 ⁴
	Avoided	45	5102	0.114 ± 0.069 ²	0.137 ± 0.069 ³	0.193 ± 0.064 ⁴
Mooney images	Non-experts	14	457	0.120 ± 0.079	0.100 ± 0.070 ^{5,6}	0.170 ± 0.079 ^{7,8}
	Experts	14	126	0.125 ± 0.085	0.124 ± 0.079 ⁵	0.194 ± 0.084 ⁷
	Avoided	14	1459	0.111 ± 0.060	0.128 ± 0.070 ⁶	0.183 ± 0.062 ⁸

Values indicate Mean ± SD. *i*, number of original images; *n*, number of details. Significantly different from each other (Kruskal-Wallis test with Dunn's post-test; ^{1,2,3,4,6}*p* < 0.0001; ^{5,7,8}*p* < 0.001).

During the presentation of the stimuli, movements of the right eye were recorded using a monocular eye tracker column with chin and forehead rest (SMI iView × HiSpeed 1250). The Person Perception Research Unit, Institute of Psychology, University of Jena, kindly provided the equipment, which uses video-based dark-pupil and corneal reflection tracking (infrared illumination) at a sampling rate of 500 Hz. The eye tracker was 12-point-calibrated for every participant at the beginning of the experiment. If necessary, calibration was repeated to reach an accuracy of $<1^\circ$ average error. The resulting data were prepared for further analysis with the program SMI BeGaze 3.2.28.

Analysis

To obtain gaze maps, we summed the fixation times for all fixations of an image point for all participants during the entire 5 s viewing periods. The total dwelling times thus obtained were plotted at the respective image co-ordinates onto the images. **Figures 2A,B** shows examples for the color photographs, and **Figures 2C,D** for the Mooney images. From these two-dimensional dwelling time plots, we calculated the total dwelling time for regions that corresponded to each of the selected and avoided details in their size and positions.

Saliency Maps

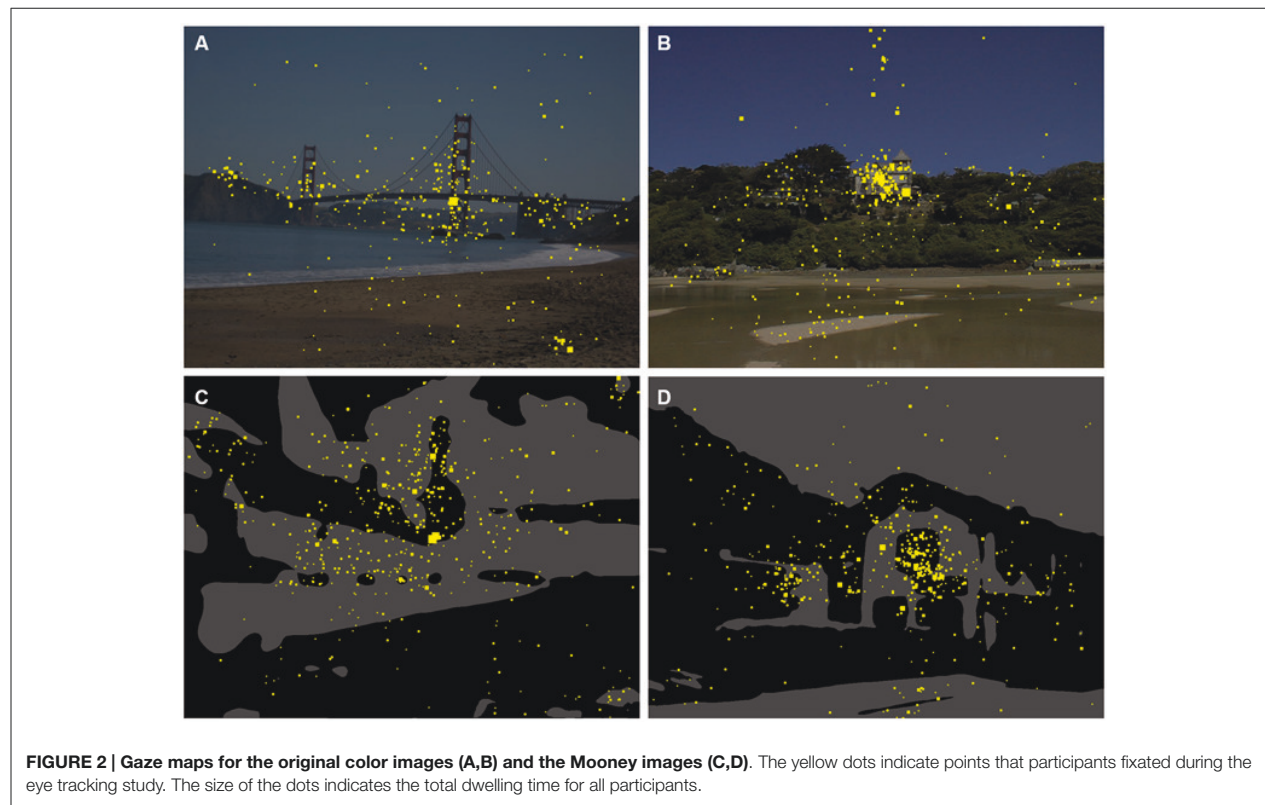
Saliency maps were calculated for each of the original images (color photographs and Mooney images) using the ITTI and GBVS methods.

ITTI is based on the use of the Gaussian blur filter in a pyramid and employs center-surround operations to highlight local gradients in 42 intensity, color and orientation feature maps in conjunction with a dynamic neural network (Itti et al., 1998). GBVS exploits graph algorithms and defines Markov chains over various image feature maps, treating the equilibrium distribution over map locations as saliency values (Harel et al., 2007). The two methods were chosen because they are well-established and have previously been applied to a wide variety of problems both in vision research in general and for photographic cropping in particular (Suh et al., 2003; Sun et al., 2009; Wong and Low, 2009; Vaquero et al., 2010; Ardizzone et al., 2013; Borji et al., 2013; Amirshahi et al., 2014; Koide et al., 2015).

From the saliency map of each photograph, we cut out details with positions and size that corresponded exactly to those of the selected and avoided details for this photograph. For each of the details, the sum of saliency was calculated.

Matching Details for Euclidean Distance to the Geometrical Image Center

When looking at images displayed on computer monitors, human observers tend to fixate the center of the screen (Bindemann, 2010), regardless of the distribution of the image features (Tatler, 2007). Such a central bias is also incorporated in some of the saliency models (Borji et al., 2013). For example, Harel et al. (2007) consider it as an explicit emergent property of



their GBVS method. Saliency maps generated by both the GBVS and ITTI methods show a fall-off of values towards the border of the images. It is therefore likely that, during the cropping procedure, participants tended to select details that are closer to the image center while avoiding details from the periphery of the photographs. Indeed, such a central cropping bias is observed in our data. The details that non-experts and experts selected from color photographs are closer to the geometrical center of the images than the avoided details (Table 1). For Mooney images, a similar difference was found for non-experts but not for expert croppers (Table 1).

Because of the central bias for the cropping decisions, the avoided details, which are from more peripheral regions in the photographs, were expected to have lower total dwelling times and lower saliency values. To counteract this tendency, we matched selected and avoided details with respect to their (Euclidean) distance to the geometrical center. For each photograph, an image was randomly picked from the set of selected details. For this selected detail, an image was randomly drawn from the pool of avoided details until a sample was found that matched its Euclidean distance within a 5% tolerance limit for divergence. The two details were entered into the further analysis as a matched pair and deleted from the set of selected and avoided details, respectively. If no matching avoided detail was found, the selected detail was eliminated from the analysis. The procedure was repeated until all selected details were matched to an avoided detail. The matching procedure was carried out independently for the non-expert details and expert details, respectively. It was also carried out for a pairwise comparison of details selected by non-experts and experts, respectively. The number of images analyzed is listed in Table 2.

Statistical Analysis

The mean values for all details that were selected or avoided by non-expert and expert participants, respectively,

were calculated for each photograph and Mooney image. Subsequently, mean values were calculated across images. The D'Agostino-Smirnov omnibus normality test was used to assess the normal distribution of the means. If datasets did not pass the normality test, a non-parametric statistical test was used for multiple comparisons (Kruskal-Wallis test with Dunn's multiple comparison post-test). For paired datasets, a paired *t*-test was carried out for normally distributed data and a Mann-Whitney test for datasets that were not normally distributed. *P*-values smaller than 0.05 were considered significant.

RESULTS

In this section, we will first address some methodological issues. Second, we will describe differences in total dwelling times and calculated saliency values between details selected during the cropping of photographs (here called *selected* details) and details that were not selected (*avoided* details). Third, we will show that both the dwelling times and saliency values are more balanced for selected than for avoided details. In each case, we will compare the results between the original color photographs and the thresholded (Mooney) versions, and between non-experts and experts.

Methodological Considerations

To minimize artifacts that are introduced by the central tendency for eye fixations and saliency calculations, we matched the different categories of photographic details (selected by *non-experts*, selected by *experts*, and *avoided* by both) for their Euclidean distance to the geometrical image center in a pairwise fashion (see "Materials and Methods" Section). The resulting matched details did not differ significantly in their Euclidean distance (Mann-Whitney test; Table 2; compare to unmatched

TABLE 2 | Deviation from the geometrical image center and Euclidean distance to the image center for pairwise matched details (in percent of normalized width and height, respectively).

		<i>i</i>	<i>n</i>	Width	Height	Euclidean distance
Color photographs						
Non-expert/avoided	Non-expert	45	1977	0.127 ± 0.073 ³	0.099 ± 0.063 ³	0.178 ± 0.061
	Avoided	45	1977	0.102 ± 0.064 ³	0.131 ± 0.064 ³	0.180 ± 0.059
Expert/avoided	Expert	35	517	0.126 ± 0.078	0.112 ± 0.069 ²	0.185 ± 0.072
	Avoided	35	517	0.116 ± 0.065	0.128 ± 0.070 ²	0.186 ± 0.067
Non-expert/expert	Non-expert	35	747	0.114 ± 0.080	0.100 ± 0.067	0.161 ± 0.172
	Expert	35	747	0.112 ± 0.078	0.103 ± 0.068	0.162 ± 0.173
Mooney images						
Non-expert/avoided	Non-expert	14	336	0.118 ± 0.075 ¹	0.097 ± 0.068 ¹	0.169 ± 0.071
	Avoided	14	336	0.102 ± 0.058 ¹	0.121 ± 0.070 ¹	0.169 ± 0.069
Expert/avoided	Expert	14	114	0.121 ± 0.085	0.126 ± 0.079	0.193 ± 0.081
	Avoided	14	114	0.116 ± 0.061	0.135 ± 0.081	0.190 ± 0.075
Non-expert/expert	Non-expert	14	121	0.136 ± 0.081	0.107 ± 0.078	0.190 ± 0.081
	Expert	14	121	0.121 ± 0.085	0.120 ± 0.078	0.189 ± 0.082

Values indicate Mean ± SD. *i*, number of original images; *n*, number of details.^{1,2,3} Significantly different with pairwise comparison (Mann-Whitney test; ¹*p* < 0.05, ²*p* < 0.001, ³*p* < 0.0001).

original values in **Table 1**). Differences were, however, observed for the deviations along the x -axis and y -axis. The mean deviation decreased along the x -axis and increased along the y -axis.

The same pairwise matching was also carried out for the details selected by non-experts and experts because experts tended to chose details from more peripheral regions of color photographs and Mooney images than non-experts (**Table 1**). After matching, none of the distances differed significantly (**Table 2**).

One of the issues in evaluating methods to calculate visual saliency is whether the obtained values can indeed predict eye movements. Here, we address this question by comparing the measures for fixation times and the ITTI and GBVS saliency measures. Specifically, for each detail used in this study, we correlated total dwelling times with saliency values, as well as their respective center-of-mass positions along the x -axis and y -axis (**Table 3**). The correlation of total dwelling times with the total saliency values was moderate (range of Spearman coefficients: 0.42–0.48) whereas the correlations between the center-of-mean values for eye movements and saliency were high (range: 0.60–0.86). Generally, high correlations were also observed for the comparison of data obtained by the ITTI and GBVS methods (range: 0.55–0.90; all p -values < 0.0001, **Table 3**).

In **Figure 3**, the saliency maps (ITTI, **Figures 3C,D**; GBVS, **Figures 3E,F**) and the eye fixation data (**Figures 3G,H**) are compared for an original color photograph (**Figure 3A**) and the corresponding Mooney image (**Figure 3B**). Although the saliency maps roughly resemble each other, there are distinct differences for some details in the image. For example, the boat on the left hand side is more prominently represented in all three types of map for the color image. The inverse is true for the wall in the lower right corner of the image. In the present analysis, we therefore calculated values separately for the original color photographs and the Mooney images.

Total Dwelling Times and Saliency Values are Higher for Selected Details

Figure 2 shows eye fixation maps for representative color photographs (**Figures 2A,B**) and thresholded binarized

(Mooney) images (**Figures 2C,D**). **Figure 4** depicts one of the color photographs (**Figure 4A**) and the corresponding saliency maps (**Figures 4E,I**) together with representative details that were selected by a non-expert (**Figure 4B**) and an expert (**Figure 4C**), respectively, and an avoided detail (**Figure 4D**). **Figures 4F,G,H,J,K,L** show the corresponding details from the saliency map. A similar set of images is shown in **Figure 5** for a representative Mooney image.

Figure 6 summarizes the results for all photographs (red) and all Mooney images (blue). **Figure 6A** shows results for the ITTI method, **Figure 6B** for the GBVS method, and **Figure 6C** for the total dwelling time.

First, we compared results for the original color photographs with the Mooney images. The total dwelling times were systematically higher for the Mooney images compared to the photographs, both for details selected by non-experts and experts participants, as well as for the avoided details (**Figure 6**; two-tailed t -test, $p < 0.0025$). No such differences were obtained for the ITTI saliency values (p -values between 0.08 and 0.17). GBVS saliency values were lower for the photographs (*Mean*: 93.3 ± 18.8 *SD*) than for the Mooney images (*Mean*: 113.1 ± 28.0 *SD*; two-tailed t -test, df : 13, $p = 0.032$), but only for the details selected by experts. The GBVS values of the details avoided by both non-experts and experts were also higher for the Mooney images than for the photographs (df : 13; $p = 0.034$ and $p < 0.019$, respectively).

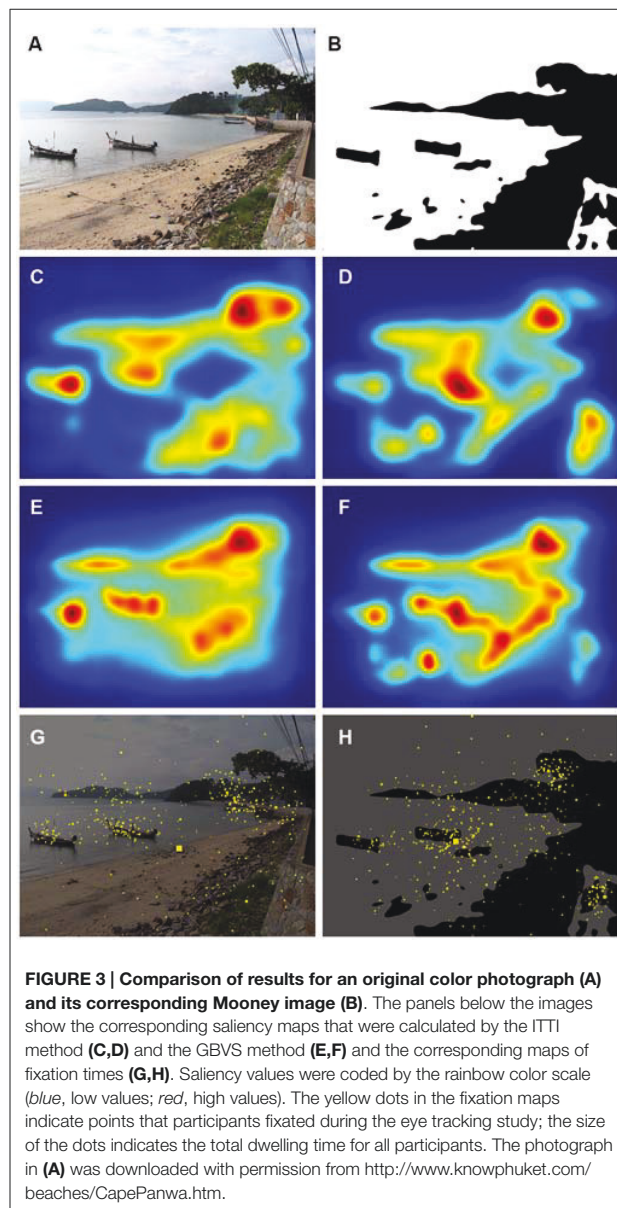
Second, the selected and avoided details from the original color photographs were compared. For the color photographs, saliency values calculated according to the ITTI method (**Figure 6A**) were significantly higher for details selected by non-experts (*Mean*: 88.4 ± 20.4 *SD*) than for the avoided details (*Mean*: 80.4 ± 23.0 *SD*; two-tailed t -test, df : 44, $p = 0.0006$). The same holds true for details selected by experts (*Mean*: 86.8 ± 22.5 *SD*; compared to *Mean*: 80.2 ± 23.2 *SD* for avoided details; df : 34, $p = 0.015$).

Saliency calculated according to the GBVS method (**Figure 6B**) yielded similar differences for the comparison of details selected and avoided by non-experts (*Mean*: 95.6 ± 20.9 *SD* for selected details vs. *Mean*: 86.9 ± 21.6 *SD* for avoided details; df : 44, $p < 0.0001$) and details selected and avoided by experts (*Mean*: 93.3 ± 18.8 *SD* for selected details vs. *Mean*: 86.6 ± 19.4 *SD* for avoided details; df : 34, $p = 0.0002$).

TABLE 3 | Correlations between eye fixation data and saliency measures (ITTI and GBVS metrics).

	Total dwelling time or saliency	CoM-Position (x axis)	CoM-Position (y axis)
Color photographs (n =11555)			
Eye fixation/ITTI	0.48 (0.46–0.49)	0.797 (0.790–0.804)	0.858 (0.853–0.863)
Eye fixation/GBVS	0.45 (0.43–0.46)	0.622 (0.611–0.634)	0.829 (0.823–0.835)
ITTI/GBVS	0.78 (0.77–0.79)	0.550 (0.537–0.563)	0.768 (0.760–0.776)
Mooney images (n =2042)			
Eye fixation/ITTI	0.42 (0.39–0.46)	0.674 (0.648–0.697)	0.698 (0.674–0.720)
Eye fixation/GBVS	0.42 (0.38–0.46)	0.603 (0.574–0.631)	0.717 (0.695–0.738)
ITTI/GBVS	0.78 (0.77–0.80)	0.871 (0.860–0.881)	0.903 (0.894–0.911)

The values represent Spearman coefficients r with 95% confidence intervals in parentheses. The correlations for the center-of-mean (CoM) positions are based on their normalized distances to the geometrical image center. All correlations are significant ($p < 0.0001$). n , number of details.



Results for the total dwelling time (Figure 6C) were similar to those obtained by calculating saliency. Dwelling times were higher for details selected by non-experts ($Mean: 1.71 \pm 0.32 SD$) than for avoided details ($Mean 1.52 \pm 0.32 SD$; $df: 44$, $p < 0.0001$), and details selected by experts ($Mean 1.70 \pm 0.34 SD$) than for avoided details ($Mean 1.52 \pm 0.32 SD$; $df: 34$, $p < 0.0001$).

Next, we compared values for the selected and avoided details from the Mooney images. The only significant difference was that Mooney images showed higher ITTI values for the details selected by non-experts ($Mean: 99.3 \pm 21.4 SD$) than the avoided details ($Mean: 91.9 \pm 18.0 SD$; $df: 13$, $p = 0.011$), and higher dwelling times for the details selected by non-experts ($Mean 2.27$

$\pm 0.29 SD$) than the avoided details ($Mean 2.09 \pm 0.29 SD$; $df: 12$, $p < 0.0001$).

Finally, we compared the values for non-expert and expert participants. Because the details selected by experts were from more peripheral regions than those selected by non-experts on average (Table 1; see “Materials and Methods” Section), we also analyzed details that were matched pairwise for their Euclidean distance to minimize the effect of central bias (see above). GBVS saliency values were higher for details selected by non-experts ($Mean 96.8 \pm 18.9 SD$) than by experts ($Mean 95.7 \pm 18.8 SD$; $df: 34$, $p = 0.024$; Figure 5). No other comparisons yielded significant differences.

In summary, participants tended to select details that contained higher visual saliency and were fixated longer than details that were avoided during cropping.

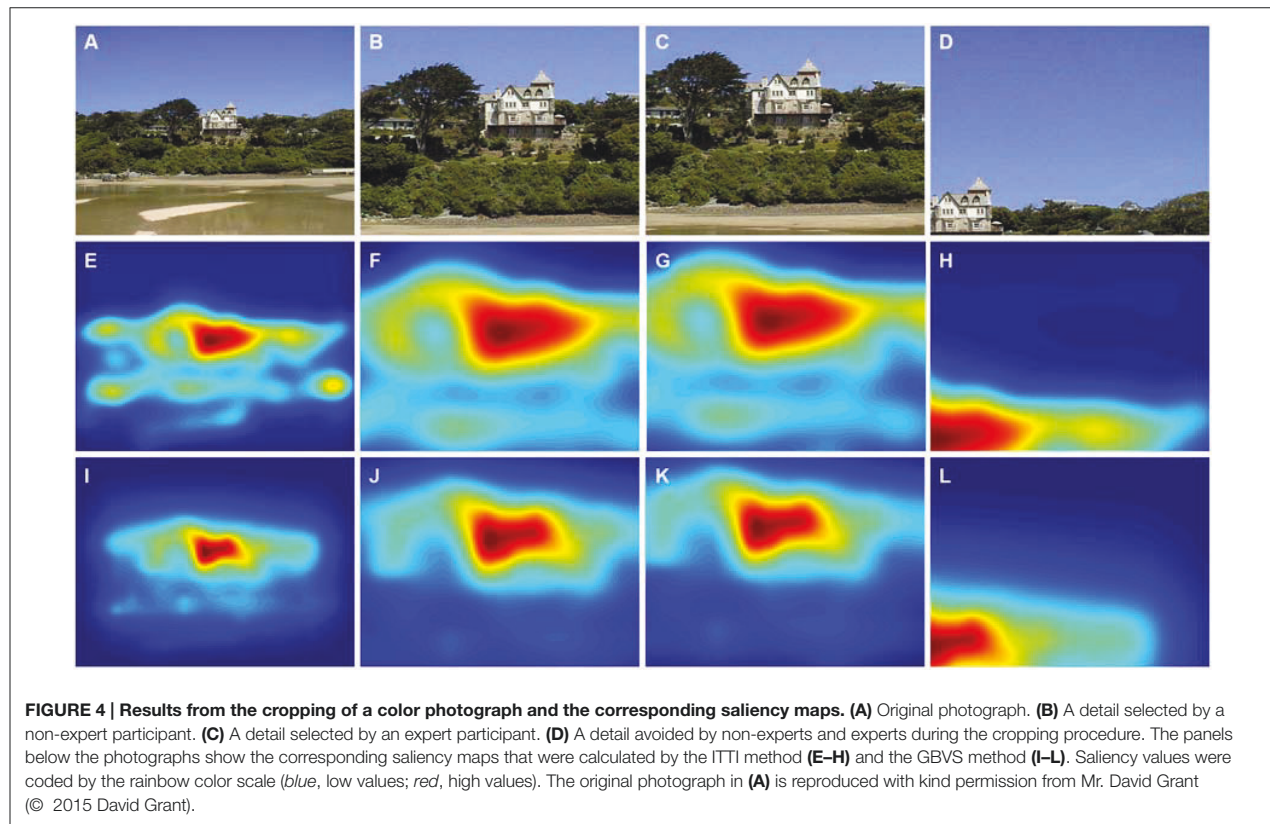
Dwelling Time and Saliency Show a More Balanced Distribution in Details Selected During Cropping

As a measure of visual balance, we chose to analyse the distribution of dwelling (fixation) times on each detail because, as a behavioral measure, it may relate to the subjective feeling of visual balance in images more closely than simpler physical measures, such as the distribution of luminance, which has been analyzed previously (McManus et al., 2011b). Our intuition was that a well-balanced image would attract eye movements equally strongly to its left and right halves as well as to its upper and lower halves. In parallel, the distribution of saliency was calculated by the ITTI and the GBVS methods to assess how well the calculated saliency measures can predict the behavioral data. To quantify the balance, the center-of-mean of the distribution of these measures was determined for each detail. Consequently, the measures were distributed in a well-balanced fashion if the center-of-mean was close to the geometrical center of the details. The more the center-of-mean deviated from the geometrical image center, the more the distribution was out of balance.

Figure 7 shows the average deviations of the center-of-means for the distributions of the ITTI saliency measure (Figures 7A,B), the GBVS measure (Figures 7C,D) and the total dwelling times (Figures 7E,F). Data for non-experts are shown on the left-hand side (Figures 7A,C,E) and for experts on the right-hand side (Figures 7B,D,F).

In general, the three measures showed similar deviations for non-experts and experts. For the color photographs, deviations were smaller for the selected details than for the avoided details along the y -axis. The deviation for ITTI (Figures 7A,B) was smaller for the details selected by non-experts ($Mean: 0.066 \pm 0.042 SD$) than for avoided details ($Mean 0.088 \pm 0.063 SD$; $df: 44$, $p = 0.0027$). The same was found for details selected by experts ($Mean 0.067 \pm 0.045 SD$) when compared to avoided details ($Mean 0.089 \pm 0.063 SD$; $df: 34$, $p = 0.0092$). No significant difference was found for deviations along the x -axis.

Similar results were obtained for GBVS (Figures 7C,D). Again, non-experts selected details that deviated less from the



geometrical image center (*Mean*: 0.041 ± 0.026 *SD*) than the avoided details (*Mean* 0.062 ± 0.048 *SD*; *df*: 44, $p = 0.0006$), as did the experts (*Mean*: 0.044 ± 0.029 *SD* vs. *Mean*: 0.065 ± 0.048 *SD*; *df*: 34, $p = 0.0004$).

Deviations for eye fixations (**Figures 7E,F**) were different not only along the *y*-axis, but also along the *x*-axis. For non-experts, deviations were smaller for the selected details than for avoided details along the *y*-axis (*Mean*: 0.050 ± 0.039 *SD* for selected details vs. *Mean*: 0.123 ± 0.083 *SD* for avoided details; *df*: 44, $p < 0.0001$) and along the *x*-axis (*Mean*: 0.054 ± 0.041 *SD* vs. *Mean*: 0.077 ± 0.054 *SD*; *df*: 44, $p = 0.032$). For experts, results were similar (for *y*-axis: *Mean*: 0.060 ± 0.040 *SD* vs. *Mean*: 0.125 ± 0.088 *SD*; *df*: 34, $p < 0.0001$; and for *x*-axis: *Mean*: 0.054 ± 0.047 *SD* vs. *Mean*: 0.078 ± 0.058 *SD*; *df*: 34, $p = 0.036$).

The averaged saliency maps for the color photographs (**Figure 8**) confirmed these observations. Compared to the details selected by non-experts (**Figures 8A,E**) and experts (**Figures 8C,G**), high saliency values were distributed more widely along the *y*-axis for the avoided details (**Figures 8B,F,D,H**).

For the Mooney images, fewer differences were observed, and only for the *y*-axis. Deviations for the ITTI method (**Figures 7A,B**) were smaller for details selected by non-experts (*Mean*: 0.054 ± 0.049 *SD*) than for the avoided details (*Mean* 0.073 ± 0.053 *SD*; *df*: 12, $p = 0.0086$). Deviations for fixation times (**Figures 7E,F**) were smaller for details

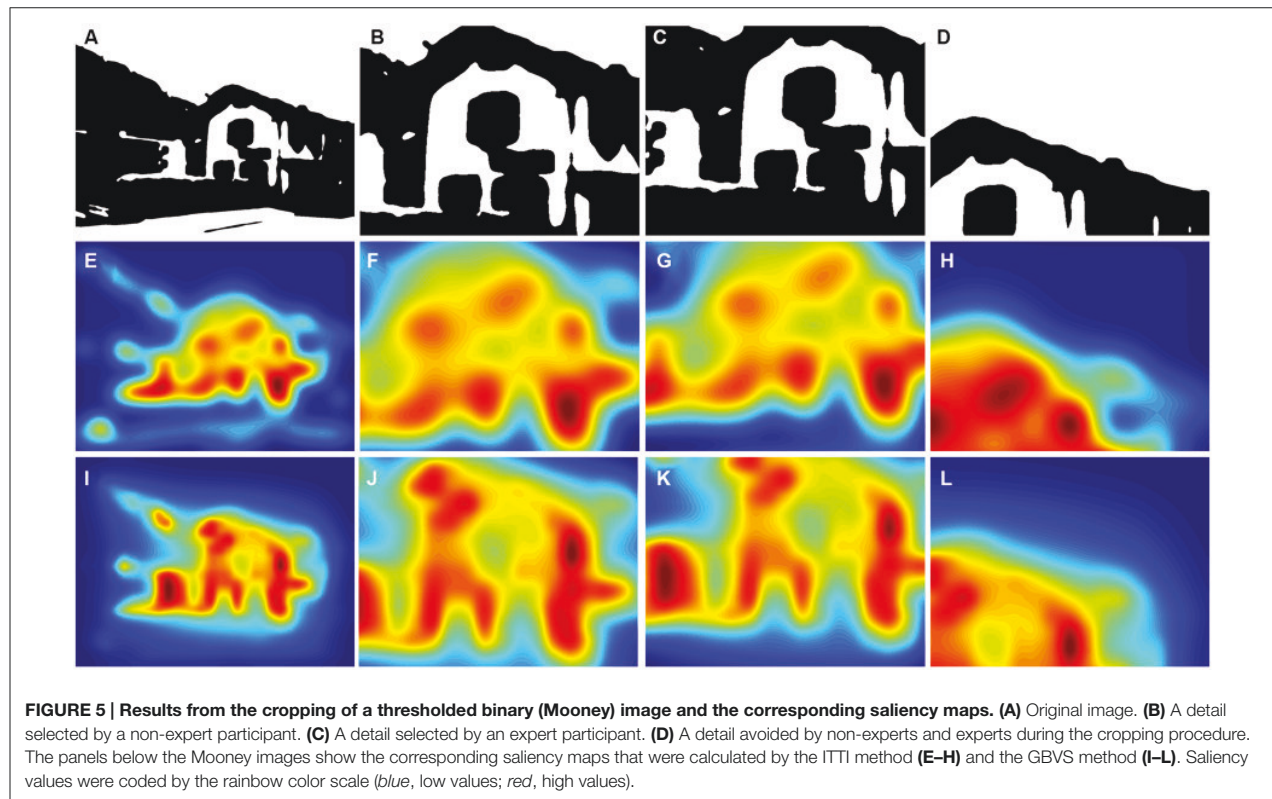
selected by both non-experts and experts (for non-experts: *Mean*: 0.043 ± 0.025 *SD* vs. *Mean*: 0.085 ± 0.055 *SD*; *df*: 12, $p = 0.0075$; and for experts: *Mean*: 0.045 ± 0.025 *SD* vs. *Mean*: 0.079 ± 0.053 *SD*; *df*: 13, $p = 0.043$). In the averaged saliency maps for the Mooney images (**Figure 9**), differences were less prominent than for the color photographs (**Figure 8**).

The comparison between data from non-experts and experts did not result in significant differences after pairwise matching for Euclidean distance (data not shown).

In summary, for both calculated visual saliency and eye fixation times, the selected details tended to be more balanced than the avoided details, in particular along the *y*-axis.

DISCUSSION

In the present study, we identified two perceptual features that played a role when participants cropped a photograph. First, participants generally selected regions during cropping that displayed a relatively high degree of visual saliency and longer dwelling times. This result was expected because salient regions are, by definition, those that attract visual attention and therefore, participants tend to include them in the selected detail in most cases (Ardizzone et al., 2013). Second, calculated visual saliency and dwelling times are distributed in a well-balanced manner in the selected details. This result relates to previous studies on



pictorial balance. Following the discussion of this relation in the next section, we will review the factors that influence pictorial balance and how our results can be used in automated cropping methods.

Pictorial Balance, Saliency and Gaze Behavior

In the present work, we analyzed the balance of visual saliency, a relatively complex measure that combines several low-level visual features (Itti et al., 1998; Harel et al., 2007). Reanalysing the image material from the studies by McManus et al. (2011a,b), we demonstrate that participants selected photographs during cropping so that the center-of-mean for saliency was close to the geometrical image center (Figure 7). This finding is mirrored by the behavioral data. For the dwelling times, the center-of-means was even more consistently centered around the geometrical image center than calculated saliency. A good correlation between eye movement and saliency data was also found for total saliency and total dwelling times and for the y- and x-positions of the center-of-means (Table 3). These results underline the general usefulness of saliency measures in predicting specific aspects of gaze behavior (Borji et al., 2013).

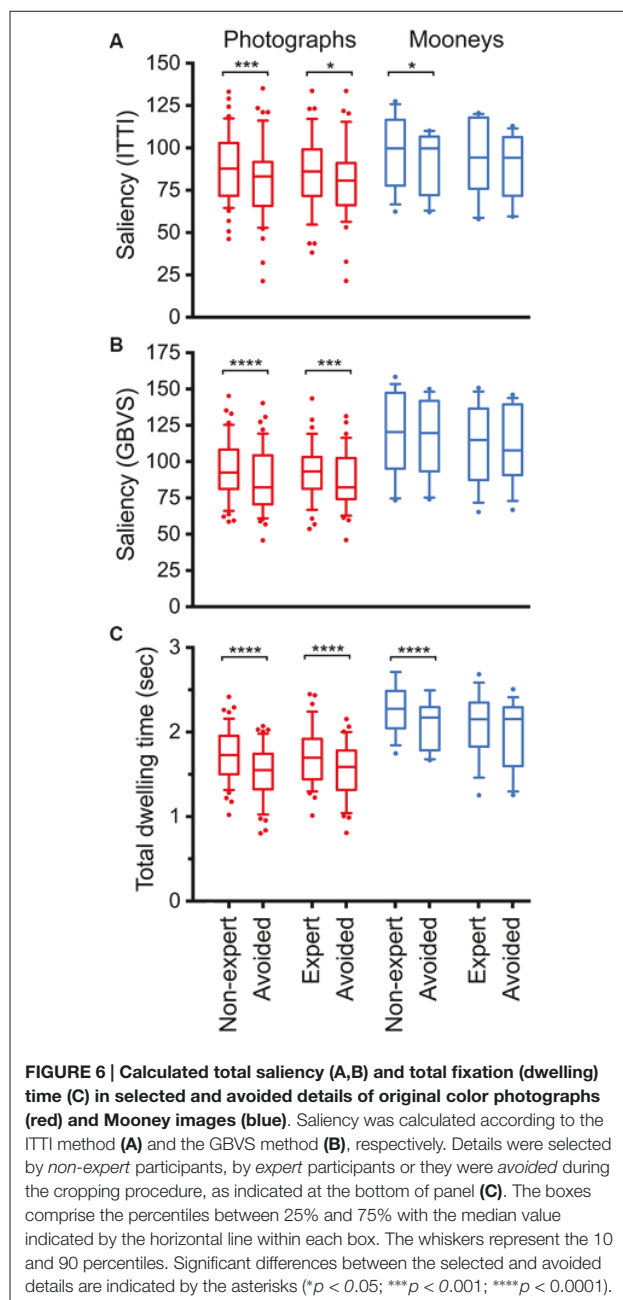
According to Locher et al. (1996), pictorial balance is achieved when the elements of a painting are positioned about a balance center so that the elements seem to be anchored and stable. Such balance judgments are interpreted to be the result of a

global integration of information across the entire picture field (McManus et al., 1985), especially from its central region (Locher et al., 1996). Moreover, there is evidence that the center of subjective balance tends to be closely aligned with the geometric center (midpoint) of the image (Locher et al., 1998), regardless of element type, format, or phase of construction of a visual display created by participants (Locher et al., 2001). This applies also for color, as shown for paintings by Mondrian (Locher et al., 2005). Our present findings are compatible with this notion and suggest saliency as an objective measure that contributes to pictorial balance. Whether this measure is useful also for other types of images (for example, artworks) remains to be studied.

Specific Image Content and Perceptual Factors Affect Pictorial Balance

Two different types of information may influence the subjective assessment of visual balance: Specific content displayed in the images, and the visual structure of the images, i.e., low- or mid-level perceptual features.

The dependence of balance on perceptual features has been assessed in studies that compare experts on art or photography and naïve (non-expert) participants. For example, Locher (2003) suggested that the visually right structure of artworks (i.e., the “good” composition) can be recognized not only by art experts, but also by viewers who lack formal training in the



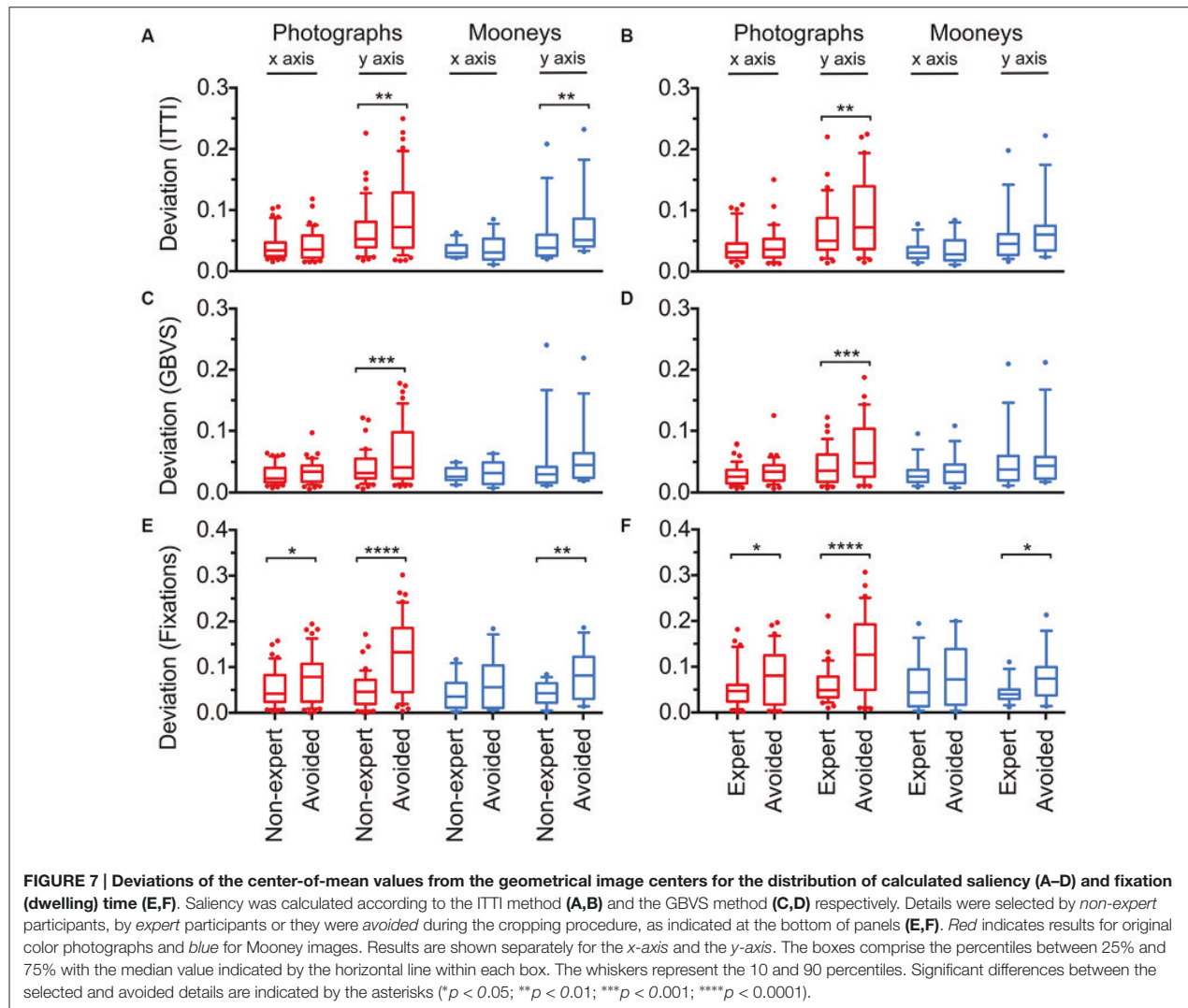
visual arts. Accordingly, pictorial balance assessments by art or design experts are in good agreement with those of non-expert viewers (Locher et al., 1996, 1999). Supposedly, non-expert participants are less prone to take into account specific image content in their judgments and as a consequence, they may rely more on formal compositional properties (Koide et al., 2015). This notion is supported by results from the present study where the cropping decisions of non-experts were more clearly related to the saliency and eye movement measures than those of the experts (Figures 6, 7). This result is also

consistent with the observation by McManus et al. (2011b) that the crops of experts were from positions that scattered more widely across the original photographs. The same study also carried out a qualitative analysis of the reasons participants gave for their cropping decisions. Non-expert participants more often mentioned image content, whereas experts mentioned formal compositional properties. Balance in artworks can also be recognized rapidly and effortlessly (McManus et al., 1985), even “at a glance” with exposure times of as short as 100 ms duration (Locher and Nagy, 1996), which may be too fast for extensive top-down processing based on semantic information. Together, these and previous findings (Koide et al., 2015) indicate that experts base their cropping decisions on semantic information to a larger extent than non-experts. Alternatively, experts may use different perceptual features that are not appreciated by non-experts. This explanation, however, seems less likely because the general results for saliency and eye movements were similar for experts and non-experts.

Of interest is our finding that the differences between selected and avoided details are more pronounced along the y -axis than along the x -axis (Figure 7). In other words, in the study by McManus et al. (2011b), moving the cropping window to the left side or to the right side changed visual balance less than moving the cropping window up and down. Possibly, this is because the horizon extends horizontally (for example, see photographs in Figure 1), so that the visual structure changes less horizontally than vertically in photographs of natural scenes. Alternatively, there may be intrinsic perceptual differences between the left/right and upper/lower halves of our perceptual visual field. For example, confirming earlier studies, Niekamp (1981) found that the upper half of a visual field has inherently greater visual weight than the lower half of stimuli that consisted of simple geometrical forms; he did not obtain similar differences between the left and right sides of the images. McManus et al. (1985) showed that objects in the upper part of an image have greater weight than those in the lower part, and objects on the right part have greater weight than on the left part.

The effect of content and perceptual features can be assessed also by comparing the original photographs and their corresponding Mooney images. Because participants viewed the Mooney images before the color photographs, it was hard for them to recognize content in the Mooney images, which represented rather abstract patterns to them. We did observe differences for saliency and gaze behavior (Figure 3), but the general pattern of cropping decisions was similar: For both types of image, participants preferred details that included highly salient regions in a balanced fashion. This result indicates that participants used similar criteria for cropping images with content and largely devoid of content. A detailed analysis of the differences between original and Mooney versions of the photographs is beyond the scope of the present study.

Besides perceptual features, semantic information can affect gaze behavior and visual behavior (see “Introduction” Section). For example, in the case of paintings, information about the title of a painting can have an influence both



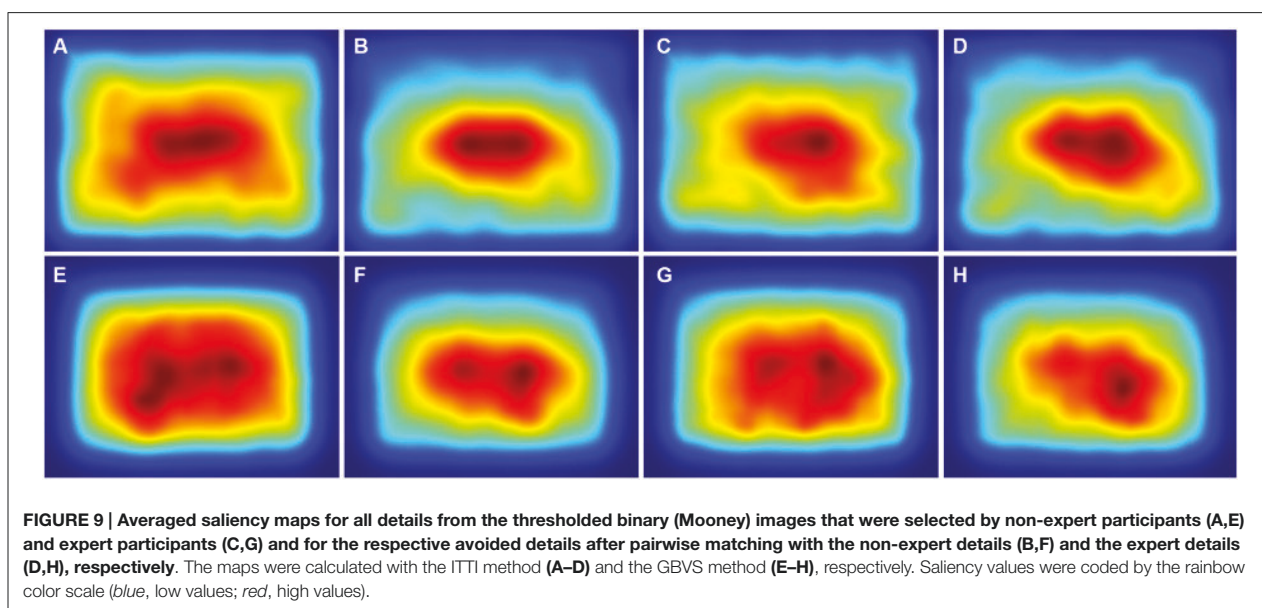
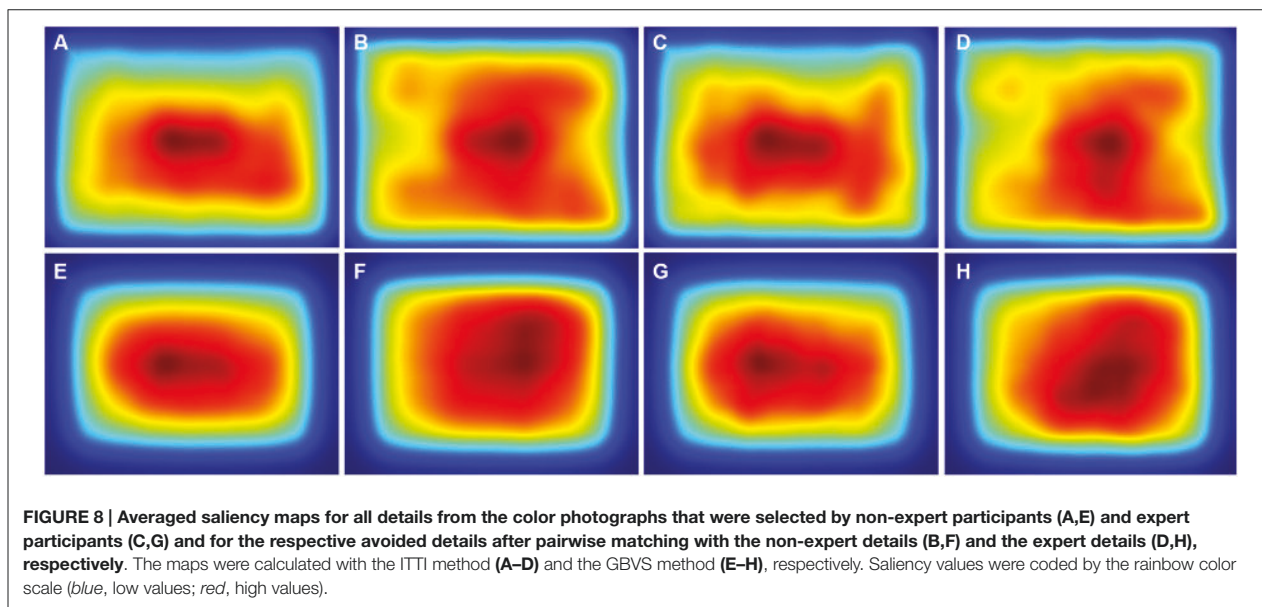
on eye movements and on the distribution of fixation times over different regions of the painting (Kapoula et al., 2009). Balance on its own does not suffice to distinguish between masterworks of art and paintings of lesser artistic quality (Vartanian et al., 2005). Moreover, for the field of architecture, Hasse and Weber (2012) reported that visual balance does not affect beauty judgments of facades. For artworks, it has been suggested that both perceptual and cognitive processing contribute to aesthetic experience (Redies, 2015).

Relevance of Results for Automated Cropping Methods

Our study provides experimental support for the notion widely held in computer vision that saliency calculations can be employed in automated cropping procedures to improve the aesthetic outcome (Santella et al., 2006; Liu

et al., 2010; Jahanian et al., 2015; Wang et al., 2015). In a well-controlled psychological experiment, we demonstrate that the centering of saliency mass onto the geometric image center results in images that are preferred by viewers compared to more unbalanced images from the same photograph. This principle has already been used in some cropping procedures as *a priori* knowledge. For example, Liu et al. (2010) used saliency maps and applied several principles of image composition, including pictorial balance, to automatically select photographic details that were aesthetically more pleasing.

The usefulness of saliency calculations, however, depends on the metric chosen for a given application. While the two methods used in the present study (ITTI, Itti et al., 1998; GBVS, Harel et al., 2007) did not differ much in their predictive power, we obtained less satisfactory results with a third method (frequency-tuned salient region detection; Achanta et al., 2009; data not shown). Despite



the general usefulness of saliency calculations in automated cropping procedures, viewers still prefer crops that are based on actual gaze behavior (Santella et al., 2006), possibly because actual eye movement data more closely relate to subjective preferences than calculated predictions of gaze behavior.

AUTHOR CONTRIBUTIONS

JA, LF, HK and CR conceived the experiments; ICM provided the dataset of photographs and unpublished experimental data; SAA,

MK and CR developed the computer programs for the analysis; JA and LF carried out the experiments; JA, LF and CR analyzed the data; CR wrote the first version of the manuscript; and SAA, MK, ICM, HK and CR contributed to the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank Mr. David Grant for permission to reproduce the photograph shown in **Figure 3A**. This work was supported by funds from the Institute of Anatomy, Jena University Hospital.

REFERENCES

- Achanta, R., Hemami, S., Estrada, F., and Sussstrunk, S. (2009). "Frequency-tuned salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL: IEEE), 1597–1604.
- Amirshahi, S. A., Hayn-Leichsenring, G. U., Denzler, J., and Redies, C. (2014). Evaluating the rule of thirds in photographs and paintings. *Art Percept.* 2, 163–182. doi: 10.1163/22134913-00002024
- Ardizzone, E., Bruno, A., and Mazzola, G. (2013). Saliency based image cropping. *Lect. Notes Comput. Sci.* 8156, 773–782. doi: 10.1007/978-3-642-41181-6_78
- Arnheim, R. (1954). *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA: University of California Press.
- Arnheim, R. (1982). *The Power of the Center: A Study of Composition in the Visual Arts*. Berkeley, CA: University of California Press.
- Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Res.* 50, 2577–2587. doi: 10.1016/j.visres.2010.08.016
- Borji, A., Sihite, D. N., and Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Trans. Image Process.* 22, 55–69. doi: 10.1109/tip.2012.2210727
- Bruno, N., Gabriele, V., Tasso, T., and Bertamini, M. (2014). 'Selfies' reveal systematic deviations from known principles of photographic composition. *Art Percept.* 2, 45–58. doi: 10.1163/22134913-00002027
- Chen, J., Zhao, H., Han, Y., and Cao, X. (2013). "Visual saliency detection based on photographic composition," in *Fifth International Conference on Internet Multimedia Computing and Service* (New York, NY: ACM), 13–16.
- Ciocca, G., Cusano, C., Gasparini, F., and Schettini, R. (2007). Self-adaptive image cropping for small display. *IEEE Trans. Consum. Electron.* 53, 1622–1627. doi: 10.1109/icece.2007.341331
- Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. *Lect. Notes Comput. Sci.* 3953, 288–301. doi: 10.1007/11744078_23
- Foulsham, T., and Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *J. Vis.* 8:6. doi: 10.1167/8.2.6
- Frintrop, S., Rome, E., and Christensen, H. I. (2010). "Computational visual attention systems and their cognitive foundations: a survey," in *ACM Transactions on Applied Perception (TAP)* (New York, NY: ACM), 6.
- Gershoni, S., and Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *Iperception* 2, 508–527. doi: 10.1068/i0472
- Graham, D. J., and Redies, C. (2010). Statistical regularities in art: relations with visual coding and perception. *Vision Res.* 50, 1503–1509. doi: 10.1016/j.visres.2010.05.002
- Harel, J., Koch, C., and Perona, P. (2007). Graph-based visual saliency. *Adv. Neural Inf. Process. Syst.* 19, 545–552.
- Hasse, C., and Weber, R. (2012). Eye movements on facades: the subjective perception of balance in architecture and its link to aesthetic judgment. *Empir. Stud. Arts* 30, 7–22. doi: 10.2190/em.30.1.c
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., and Mack, M. (2007). "Visual saliency does not account for eye movements during visual search in real-world scenes," in *Eye Movements: A Window on Mind and Brain*, eds R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (Amsterdam: Elsevier), 537–562.
- Hicks, R. W. (2005). "Composition," in *The Oxford Companion to the Photograph*, ed. R. Lenman (Oxford: Oxford University Press), 135–136.
- Howard, R. (1914). A note on pictorial balance. *Br. J. Psychol.* 7, 56–73. doi: 10.1111/j.2044-8295.1914.tb00243.x
- Itti, L., Koch, C., and Niebur, E. (1998). "A model of saliency-based visual attention for rapid scene analysis," in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Washington, DC: IEEE), 1254–1259.
- Jahaniyan, A., Vishwanathan, S. V. N., and Allebach, J. P. (2015). "Learning visual balance from large-scale datasets of aesthetically highly rated images," in *Human Vision and Electronic Imaging XX*, eds B. E. Rogowitz, T. N. Pappas, and H. De Ridder (San Francisco, CA: SPIE), 1–9.
- Kapoula, Z., Daunys, G., Herbez, O., and Yang, Q. (2009). Effect of title on eye-movement exploration of cubist paintings by Fernand Leger. *Perception* 38, 479–491. doi: 10.1068/p6080
- Koide, N., Kubo, T., Nishida, S., Shibata, T., and Ikeda, K. (2015). Art expertise reduces influence of visual saliency on fixation in viewing abstract-paintings. *PLoS One* 10:e0117696. doi: 10.1371/journal.pone.0117696
- Liu, L., Chen, R., Wolf, L., and Cohen-Or, D. (2010). Optimizing photo composition. *Computer Graphics Forum* 29, 469–478. doi: 10.1111/j.1467-8659.2009.01616.x
- Locher, P. J. (2003). An empirical investigation of the visual rightness theory of picture perception. *Acta Psychol.* 114, 147–164. doi: 10.1016/j.actpsy.2003.07.001
- Locher, P., Cornelis, E., Wagemans, J., and Stappers, P. J. (2001). Artists' use of compositional balance for creating visual displays. *Empir. Stud. Arts* 19, 213–227. doi: 10.2190/ekmd-ymn5-njug-34bk
- Locher, P., Gray, S., and Nodine, C. (1996). The structural framework of pictorial balance. *Perception* 25, 1419–1436. doi: 10.1068/p251419
- Locher, P., and Nagy, Y. (1996). Vision spontaneously establishes the percept of pictorial balance. *Empir. Stud. Arts* 14, 17–31. doi: 10.2190/x8u3-ctq6-a7j1-8jq8
- Locher, P., Overbeeke, K., and Stappers, P. J. (2005). Spatial balance of color triads in the abstract art of Piet Mondrian. *Perception* 34, 169–189. doi: 10.1068/p5033
- Locher, P. J., Stappers, P. J., and Overbeeke, K. (1998). The role of balance as an organizing design principle underlying adults' compositional strategies for creating visual displays. *Acta Psychol. (Amst.)* 99, 141–161. doi: 10.1016/s0001-6918(98)00008-0
- Locher, P. J., Stappers, P. J., and Overbeeke, K. (1999). An empirical evaluation of the visual rightness theory of pictorial composition. *Acta Psychol. (Amst.)* 103, 261–280. doi: 10.1016/s0001-6918(99)00044-x
- McManus, I. C., Edmondson, D., and Rodger, J. (1985). Balance in pictures. *Brit. J. Psychol.* 76, 311–324. doi: 10.1111/j.2044-8295.1985.tb01955.x
- McManus, I. C., Stöver, K., and Kim, D. (2011a). Arnheim's Gestalt theory of visual balance: examining the compositional structure of art photographs and abstract images. *Iperception* 2, 615–647. doi: 10.1068/i0445aap
- McManus, I. C., Zhou, F. A., L'Anson, S., Waterfield, L., Stöver, K., and Cook, R. (2011b). The psychometrics of photographic cropping: the influence of colour, meaning and expertise. *Perception* 40, 332–357. doi: 10.1068/p6700
- Mooney, C. M. (1956). Closure with negative after-images under flickering light. *Can. J. Psychol.* 10, 191–199. doi: 10.1037/h0083671
- Niekamp, W. (1981). An exploratory investigation into factors affecting visual balance. *Educ. Technol. Res. Dev.* 29, 37–48.
- Niu, Y., Todd, R. M., Kyan, M., and Anderson, A. K. (2012). Visual and emotional salience influence eye movements. *ACM Trans. Appl. Percept.* 9:13. doi: 10.1145/2325722.2325726
- Palmer, S. E., Gardner, J. S., and Wickens, T. D. (2008). Aesthetic issues in spatial composition: effects of position and direction on framing single objects. *Spat. Vis.* 21, 421–449. doi: 10.1163/156856808784532662
- Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Front. Hum. Neurosci.* 9:219. doi: 10.3389/fnhum.2015.00218
- Ross, D. W. (1907). *A Theory of Pure Design: Harmony, Balance, Rhythm*. Boston, MA: Houghton Mifflin.
- Santella, A., Agrawala, M., Decarlo, D., Salesin, D., and Cohen, M. (2006). "Gaze-based interaction for semi-automatic photo cropping," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)* (New York, NY: ACM), 771–780.
- Suh, B., Ling, H., Bederson, B. B., and Jacobs, D. W. (2003). "Automated thumbnail cropping and its effectiveness," in *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)* (New York, NY: ACM).
- Sun, X., Yao, H., Ji, R., and Liu, S. (2009). "Photo assessment based on computational visual attention model," in *Proceedings of the 17th ACM International Conference on Multimedia (MM '09)* (New York, NY: ACM).
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *J. Vis.* 7, 4.1–4.17. doi: 10.1167/7.14.4
- Tong, H., Li, M., Zhang, H.-J., He, J., and Zhang, C. (2004). Classification of digital photos taken by photographers or home users. *Lect. Notes Comput. Sci.* 3331, 198–205. doi: 10.1007/978-3-540-30541-5_25
- Vaquero, D., Turk, M., Pulli, K., Tico, M., and Gelfand, N. (2010). "A survey of image retargeting techniques," in *Applications of Digital Image Processing XXXIII*, ed. A. G. Tescher (San Francisco, CA: SPIE), 1–15.

- Vartanian, O., Martindale, C., Podsiadlo, J., Overbay, S., and Borkum, J. (2005). The link between composition and balance in masterworks vs. paintings of lower artistic quality. *Br. J. Psychol.* 96, 493–503. doi: 10.1348/000712605x47927
- Wang, P., Lin, Z., and Mech, R. (2015). “Learning an aesthetic photo cropping cascade,” in *IEEE Winter Conference on Applications of Computer Vision: IEEE* (Waikoloa, HI: IEEE), 448–455.
- Wong, L.-K., and Low, K.-L. (2009). “Saliency-enhanced image aesthetics class prediction,” in *16th IEEE International Conference on Image Processing (ICIP): IEEE* (Cairo: IEEE), 997–1000.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Abeln, Fesz, Amirshahi, McManus, Koch, Kreysa and Redies. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

4 Diskussion

McManus et al. (2011a, 2011b) stellten in ihren Arbeiten fest, dass das Zuschneiden von Probanden wiederholbar und zuverlässig durchgeführt wird. Weiterhin konnten sie zeigen, dass einige Probanden die Aufgabe, ästhetisch ansprechendere Zuschnitte zu gestalten, besser erfüllen als andere. Objektive Parameter, die während des Entscheidungsprozess des ausgewählten Zuschnitts von Photographie eine wichtige Rolle spielen, konnten sie jedoch nicht identifizieren.

In unserer Arbeit (siehe Kapitel 3) analysierten wir die Augenbewegungen von 34 Probanden, die den Datensatz von McManus et al. (2011a, 2011b) betrachteten. Anschließend wurden die Augenbewegungskarten mit den Salienzkarten des ITTI und GBVS Algorithmusses der Bilder verglichen. Hierbei zeigte sich eine deutliche Übereinstimmung zwischen den Daten der Augenbewegungen und beider Algorithmen in ihrer Verteilung in den zugeschnittenen Details. Dieses Ergebniss unterstreicht den Nutzen der Salienzalgorithmen, um spezifische Aspekte der Wahrnehmung vorherzusagen (Borji et al. 2013). Probanden wählen die Regionen während des Zuschneidens aus, die höhere Salienzwerte und eine höhere *total dwelling time* (TDT) besitzen. Das Ergebnis stimmt mit der Hypothese überein, dass saliente Regionen per definitionem die sind, die visuelle Aufmerksamkeit auf sich ziehen.

Für die Hypothese, dass Balance bei der Auswahl der Zuschnitte eine Rolle spielt, konnte ich zeigen, dass sowohl die berechnete Salienz, als auch die TDT gut balanciert ist. Dies zeigte sich darin, dass Probanden die Details während des Zuschneidens so auswählen, dass das *center-of-mass* (COM) der Salienz und der TDT sich nahe dem geometrischem Mittelpunkt des jeweiligen zugeschnittenen Details befinden. Zudem zeigte sich für die TDT, dass das COM noch stärker zentriert um den geometrischen Mittelpunkt angeordnet war als bei der Salienz. Dieses steht in Übereinkunft mit anderen Studien zur Balance, die zeigten, dass die Bildmitte einen besonderen Einfluss auf die wahrgenommene Balance hat (Locher et al. 1996, Locher et al. 1998, McManus et al. 1985).

Den Einfluss des Inhaltes untersuchten wir mithilfe des Vergleich der originalen Fotos mit den korrespondierenden Mooney Bildern (Mooney 1956). Den Probanden in unserer

Studie wurden zuerst die Mooney Bilder präsentiert, um das Erkennen des Inhaltes zu erschweren. Obwohl es Unterschiede in den salienten Regionen zwischen den originalen Fotos und den dazu gehörenden Mooney Bildern gab, veränderte sich das Ergebnis des auf Salienz und TDT basierendem Zuschneiden nicht. Dieses Ergebnis lässt vermuten, dass Probanden identische Kriterien für das Zuschneiden von Bildern mit weitestgehend leeren Inhalt, wie bei Mooneys, wie für die Originalbilder nutzen.

Obwohl Unterschiede beim Betrachten von Kunst zwischen den Experten und Nichtexperten von bspw. Vogt und Magnussen (2007) beschrieben werden, zeigten sich in unseren Analysen keine Unterschiede im *Cropping*-Verhalten von Bildern zwischen Künstlern und Laien im direkten Vergleich der beiden Gruppen (Experten zu Nichtexperten). Jedoch war der Einfluss der Salienz und der TDT beim Zuschneiden der Nichtexperten stärker ausgeprägt.

Im folgenden Teil werden zunächst die Methoden kritisch betrachtet und anschließend unsere Ergebnisse im Kontext der *pictorial balance* diskutiert. Im Anschluss folgen einige Anwendungsbeispiele aus dem Bereich des automatisierten Zuschneidens, bevor die Rolle von Augenbewegungen in der Wahrnehmung von Kunst und in der Medizin näher betrachtet wird.

4.1 Kritische Betrachtung der Methoden

4.1.1 *Central-Tendency-Bias*

Beim Betrachten eines Bildschirmes fixieren menschliche Beobachter vermehrt die Mitte (Bindemann 2010). Die Verteilung des Inhaltes ist hierbei nach Tatler (2007) von keiner Bedeutung. Ebenso findet sich bei einigen Salienzmodellen ein ähnlicher *central tendency bias* (Borji et al. 2013). Beim GBVS-Algorithmus wird dies zum Beispiel als ein unvermeidlicher Prozess beschrieben (Harel et al. 2007) und auch der ITTI Algorithmus ist hiervon betroffen. Bei beiden nehmen algorithmusbedingt die Salienzwerte zur Bildperipherie hin ab. Auch in dem von uns ausgewählten Datensatz ließ sich dies vor allem für die Nichtexperten eine zentrale Tendenz beobachten, da die ausgewählten Details mit ihrem Mittelpunkt näher am geometrischen Mittelpunkt des originalen Bildes waren.

Auf Grundlage dessen ist zu erwarten, dass bei den Details, die nicht ausgewählt wurden und einen größeren Abstand ihres Mittelpunktes zum geometrischen Mittelpunktes haben, kleinere Werte für Salienz und TDT messbar sind. Damit dieser Bias sich nicht auf die Ergebnissen auswirkt, machten wir uns diesen Abstand zu nutze. Hierfür berechneten

wir für jedes Detail des Datensatzes den euklidischen Abstand zwischen den Koordinaten des Mittelpunktes des jeweiligen Details und des Mittelpunktes des originalen Bildes. Im Anschluss erfolgte eine zufällige Paarbildung der Details aus einem Original, wobei der euklidische Abstand bei den Paaren identisch sein musste. Anschließend erfolgte die Analyse in diesen Paaren.

4.1.2 Auswahl der Salienzalgorithmen

Salienz beschreibt die Strukturen, die die Aufmerksamkeit auf sich ziehen und so herausstechen. Sehr viele Eigenschaften werden mit diesem Begriff in Verbindung gebracht. Darunter fallen zum einen kognitive Prozesse, Emotionen und Ziele (Henderson et al. 2007, Frintrop et al. 2010, Niu et al. 2012), aber auch Bildeigenschaften, wie Farbe, Helligkeit, Kontrast und Ortsfrequenzen (Treisman 1986). Wir nutzten zwei computerbasierte Salienzalgorithmen, die auf *low-level* Charakteristika beruhen. Den ITTI-Algorithmus (Itti et al. 1998) wählten wir aus, da er einer der ersten und am längsten etablierten Salienzalgorithmen ist, sowie den GBVS-Algorithmus (Harel et al. 2007), da er einen innovativeren Ansatz über die Markow-Ketten nutzt (siehe Abschnitt 2.2.2 und 2.2.3). Anfangs schlossen wir zusätzlich den *frequency-tuned* Algorithmus von Achanta et al. (2009) ein, der auf einer Analyse des Fourier-Spektrums von Bildern beruht. Da dieser Farbbilder in die Analyse einschließt und so keine weiterführenden Ergebnisse liefern konnte, wurde davon abgesehen, ihn für die weitere Analyse zu verwenden.

Zusätzlich sind Salienzalgorithmen aufgabenspezifisch und haben als Grundlage verschiedenste *low-level* Charakteristika (Borji et al. 2013), wobei die Größe des Einflusses der *bottom-up* Prozesse ebenfalls kontrovers diskutiert wird. Um dem entgegen zu wirken, nutzten wir Augenbewegungen als Vergleichsparameter, da diese von den Salienzalgorithmen vorhergesagt werden sollen.

In unserer Arbeit zeigte sich eine deutliche Korrelation zwischen der Salienz der von uns ausgewählten Algorithmen und den Augenbewegungen. In jedem Fall waren die Ergebnisse der Augenbewegungen stärker als die der Salienz. Dies ließe sich damit in Zusammenhang bringen, dass die originalen Bilder in unserem Experiment von den Probanden für 5000ms zu sehen waren und bereits ab 100ms auch *high-level* Prozesse beginnen, die Aufmerksamkeit zu modellieren (Connor et al. 2004). Daher beruhten die Augenbewegungen nicht ausschließlich auf *bottom-up* Prozessen und damit *low-level* Charakteristika.

4.2 Relevanz der Ergebnisse im Kontext von Balance

Nach Locher et al. (1996) wird Balance in einem Bild erreicht, wenn die Bildelemente um ein Zentrum balanciert platziert sind. So wird ein Eindruck von Bildstabilität vermittelt. Bei der Beurteilung von Balance spielen die Informationen des gesamten Bildes eine Rolle (McManus et al. 1985). Der Bildmitte kommt eine besondere Rolle zu (Locher et al. 1996), da der geometrische Bildmittelpunkt in enger Verbindung mit dem Zentrum der Balance stehen soll (Locher et al. 1998). Dies ist unabhängig von verschiedenen Objekten im Bild oder Bildformaten (Locher et al. 2001) und es gilt ebenso für Farbe, wie an Piet Mondrians Bilder gezeigt werden konnte (Locher et al. 2005).

Nach heutigem Wissenstand gibt es wahrscheinlich zwei Arten von Informationen, die wichtig für die Wahrnehmung visueller Balance sind: Zum einen der Inhalt des Gezeigten und zum anderen die visuelle Struktur des Bildes, also *low-level* Faktoren der Wahrnehmung. In Vergleichsstudien zwischen Laien und professionellen Künstlern wurde die Abhängigkeit der Balance von Eigenschaften der Wahrnehmung untersucht. Locher et al. (2003) stellten für Kunstwerke mit guter Komposition und damit verbundener guter visueller Charakteristika die These auf, dass nicht nur Künstler oder Photographen, sondern auch Laien die Qualität der Komposition in Kunstwerken erkennen können. Passend hierzu fand Locher et al. (1996, 1999), dass die Beurteilung von Balance zwischen Fachleuten und Laien übereinstimmt. Jedoch lassen sich Unterschiede in der Betrachtungsweise zwischen Photographen und Laien feststellen (Vogt und Magnussen 2007). Photographen *scannen* mehr die Bilder und verwenden mehr Betrachtungszeit für abstrakte Objekte, als Laien, die mehr Betrachtungszeit für bekannte Objekte wie Gesichter verwenden. Koide et al. (2015) bestätigten dies in ihrer Studie und stellten weiterhin fest, dass Laien den Bildinhalt weniger wahrnehmen und sich beim Betrachten von Kunstwerken mehr auf formelle *low-level* Bildeigenschaften (Salienz) verlassen. Es gelang uns, diese Ergebnisse mit unseren Daten zu bestätigen, da die Entscheidung des Zuschnittes bei den Laien mehr auf Salienz beruht. Auch McManus et al. (2011b) beobachteten dies, da die Zuschnitte der Experten sich mehr auf dem originalen Bild verteilten. Im qualitativen Teil ihrer Analyse, in dem sie durch Mikrophone Begründungen für die jeweilige Auswahl des Zuschnitts aufzeichneten, stellten sie fest, dass die Gruppe der Experten eher formelle Bildkompositionsregeln, wie z. B. den goldenen Schnitt, aber auch den Bildinhalt angaben als die Laien.

Die Wahrnehmung der Balance in der Kunst wird zudem ohne Anstrengung erreicht (McManus et al. 1985). Es reichen kurze Darbietungszeiten, wie etwa 100 *ms*, um die

wahrgenommene Balance sicher zu evaluieren (Locher und Nagy 1996). Solche Zeiten sind höchstwahrscheinlich zu kurz für einen starken Einfluss der *top-down* Prozesse des visuellen Systems (Connor et al. 2004). Zusammenfassend lässt sich behaupten, dass die Entscheidung des Zuschneidens von Laien weniger auf dem Inhalt des Bildes beruht. Eine Erklärung dafür könnte sein, dass Experten eine andere Art der Wahrnehmung gelernt haben, die Laien nicht zur Verfügung steht. Diese Alternative ist jedoch als eher unwahrscheinlich anzunehmen, da die TDT bei den Zuschnitten der Experten und Laien sehr ähnlich waren.

Um die Balance zu bewerten, nutzten wir die Abweichung entlang der X- bzw. Y-Achse der Balancezentren zum geometrischem Zentrum des Zuschnittes. Es zeigte sich hierbei eine unterschiedliche Einflussnahme der Achsen. Abweichungen entlang der Y-Achse waren deutlich prominenter bei den vermiedenen zugeschnittenen Bildern, als die Abweichungen entlang der X-Achse. Daraus lässt sich schlussfolgern, dass bei der vorangegangenen Studie von McManus et al. (2011b) die Bewegung des Fensters für das Zuschneiden entlang der Y-Achse einen größeren Einfluss auf das später Balanceempfinden hatte als eine Verschiebung entlang der X-Achse. Da einige der Photographien natürliche Szenen darstellten und einen Horizont besaßen, hat eine Verschiebung entlang dieses Horizontes einen geringeren Einfluss als die Vertikalverschiebung. Eine andere mögliche Erklärung ist ein Unterschied in der intrinsischen Wahrnehmung zwischen dem unteren und oberen sowie linken und rechten visuellen Feld. Untersuchungen von Niekamp et al. (1981) zeigten, dass geometrische Formen in der oberen Hälfte des visuellen Feldes ein größeres subjektives Gewicht repräsentieren als in der unteren Hälfte. McManus et al. (1985) bestätigten dies und konnten zeigen, dass Objekte, die sich im rechten Teil des Bildes befanden, ein größeres wahrgenommenes Gewicht im Verhältnis zu denen im linken besitzen.

Außer der von uns durch die Algorithmen (Itti et al. 1998, Harel et al. 2007) betrachteten visuellen salienten *low-level* Eigenschaften, die wir für die Analyse von Balance nutzten, gibt es noch weitere Faktoren, die Aufmerksamkeit modulieren. Bestandteile der *top-down* Prozesse, die damit in Verbindung stehen, sind unter anderem kognitive Prozesse, Emotionen und Ziele (Henderson et al. 2007, Frintrop et al. 2010, Niu et al. 2012). Im Kontext der Malerei würde dies zum Beispiel beeinflusst werden durch Informationen über den Künstler oder der Titel des Bildes. Kapoula et al. (2009) zeigten, dass dies Einfluss auf die Verteilung der Fixationen in den verschiedenen Regionen des Bildes hat. Außerdem reicht Balance als Kriterium nicht aus, um zwischen Meisterwerken der Kunst und Malerei mit geringerer künstlerischer Qualität zu differenzieren (Vartanian

et al. 2005). Auch auf die Bewertung der Schönheit von Fassaden hat Balance keinen Einfluss (Hasse und Weber 2012).

Es lässt sich resümieren, dass unsere Ergebnisse das Konzept von *pictorial balance* unterstützen und Salienz ein möglicher Parameter ist, um dies zu quantifizieren. Da unsere Ergebnisse sich auf Photographien beziehen und bei Kunstwerken die Vermutung besteht, dass sowohl Wahrnehmung als auch kognitive Prozesse eine Rolle beim ästhetischen Empfinden spielen (Redies 2015), bleibt es zukünftigen Untersuchungen überlassen, herauszuarbeiten, inwieweit dieses Konzept auch für Bereiche der bildenden Kunst gilt.

4.3 Relevanz der Ergebnisse im Kontext von *Computational Aesthetics*

Im Feld der *computational aesthetics* geht man schon seit längerem davon aus, dass *low-level* Bildeigenschaften bei der ästhetischen Gestaltung von zugeschnittenen Bildern eine wichtige Rolle spielen. Tong et al. (2004) erstellten einen Klassifikator, der auf *low-level* Bildeigenschaften wie Farbe und Kontrast beruhte, um zwischen Fotos von Laien und Photographen unterscheiden zu können. Außerdem konnte dieser Klassifikator in einem weiteren Experiment mit einer Genauigkeit von 84,7 % die ästhetische Bewertung der Probanden vorhersagen. Datta et al. (2006) trainierten einen automatisierten Klassifikator mit 56 *low-level* Bildeigenschaften, wie Farbsättigung, Drittelregel oder Seitenverhältnisse. Dadurch wollten sie zwischen Fotos unterscheiden, die hoch bzw. niedrig ästhetisch bewertet wurden. Es genügten für den Klassifikator nach dem Training 15 der *low-level* Bildeigenschaften, um mit einer Treffsicherheit von 70 % in einem Datensatz von 1664 Fotos eine Vorhersage über die Bewertung des einzelnen Fotos zu treffen. Einen weiteren Klassifikator setzten Wong und Low (2009) ein, um zwischen Fotos von Photographen und *snap-shots* unterscheiden zu können. Dies geschah mit einer Erfolgsrate von 79 %, wobei als Parameter die Salienz basierend auf dem ITTI Algorithmus (Itti et al. 1998) gewählt wurde. Eine Erweiterung dieses Ansatzes, in dem *top-down* Elemente und personalisierte Parameter zu dem *bottom-up* Modell hinzugefügt wurden, lieferten Sun et al. (2009).

Außerdem dient Salienz als Parameter bei einer großen Anzahl an Computerprogrammen, die Zuschneiden automatisch ausführen. Ardizzone et al. (2013) verglichen hierfür fünf verschiedene Salienzalgorithmen, wobei unter anderem auch GBVS- (Harel et al. 2007) und ITTI-Algorithmus (Itti et al. 1998) untersucht wurden. Es zeigten sich für

jeden dieser Algorithmen zufriedenstellende Ergebnisse. Suh et al. (2003) wandten dieses Verfahren an, um leichter erkennbare Vorschaubilder (*thumbnails*) zu generieren. Sie verglichen normal verkleinerte Bilder mit *thumbnails*, die durch die Auswahl unter zur Hilfenahme des ITTI-Algorithmus (Itti et al. 1998) entstanden. Außerdem nutzten sie als weiteren Vergleich *thumbnails* die durch *face detection*, d.h. durch erkennen und Auswählen von Gesichtern, erstellt wurden. Es zeigte sich, dass es für Probanden einfacher ist, Bilder wiederzuerkennen, wenn der saliente Bereich in den *thumbnails* gezeigt wird. Bei Bildern mit Gesichtern war das *face detection* Modell besser geeignet als der Algorithmus. Das Salienzmodell war aber weiterhin den einfach verkleinerten Bildern überlegen. Ciocca et al. (2007) erweiterten dieses Verfahren für das Zuschneiden von Bildern für kleinere Bildschirme. Sie setzten ein adaptives visuelles Aufmerksamkeitsmodell ein, dass die salienten Regionen nach ITTI (Itti et al. 1998) berücksichtigt und zusätzlich auf Gesichter prüft. So wurde erreicht, dass die Botschaft des Bildes weiterhin Beachtung findet und 53 % der so entstandenen Bilder in einer kleinen Anzahl an Probanden besser bewertet wurden als die herunterskalierten Bilder. Liu et al. (2010) entwickelte durch diese Erfahrungen unter Einbezug der Salienz und verschiedenen formellen Bildkompositionsregeln wie Balance eine Messmethode für Ästhetik bei zugeschnittenen Bildern. Inwieweit Salienz und Balance hierbei interagieren, war bisher noch nicht gezeigt worden. Wir konnten mit den Ergebnissen unserer Studie den direkten Nachweis erbringen, dass die Salienz in Bildern, die von Probanden bevorzugt werden, zum geometrischen Mittelpunkt hin zentriert ist und so Salienz als einen entscheidenden Parameter für die Balance bestätigen.

Santella et al. (2006) kombinierten in ihrem Ansatz zum automatisierten Zuschneiden Salienz mit *Eye-Tracking*-Daten. Es stellte sich heraus, dass Beobachter die automatisch generierten Zuschnitte, die auf *Eye-Tracking*-Daten beruhten, gegenüber reiner salienzbasierter Zuschnitten bevorzugten. In unserer Arbeit fanden wir eine signifikante Korrelation zwischen Salienz und den Augenbewegungsdaten. Im späteren Vergleich mit den vermiedenen Bildausschnitten waren die Ergebnisse für die Augenbewegungsdaten signifikanter als die Salienzalgorithmen. Daher ist anzunehmen, dass Augenbewegungen die subjektiven Präferenzen besser vorhersagen, als die durch Salienzalgorithmen berechneten Präferenzen.

Auch außerhalb von Photographie werden zum Beispiel Kunstwerke nach *low-level* Bildeigenschaften untersucht (Graham und Redies 2010, Redies 2015). Es bleibt zukünftigen Untersuchungen überlassen, inwiefern sich unsere Ergebnisse für die Photographie auf Kunstwerke übertragen lassen.

4.4 Analyse der Augenbewegung in der Wahrnehmung von Kunst und in der Medizin

Die Untersuchung von Augenbewegungen ist für das Verständnis der Wahrnehmung von Kunst von großer Bedeutung (Locher 2006). Vogt und Magnussen (2007) konnten Unterschiede der Betrachtungsweise zwischen Künstlern und Laien zeigen. Künstler achten unter anderem mehr auf Bildkomposition und Balance. Miall und Tchalenko (2001) verglichen die Fixationen des Künstlers Humphrey Ocean während des Zeichnens von Portraits mit denen eines Laiens. Es zeigte sich, dass die Fixationen des Künstlers länger waren und sich ausschließlich auf einzelne Details beschränkten. Damit verglichen waren die der Laien halb so lang und auf mehrere Positionen im Bild verteilt. Für Rembrandts Portraits wurde zudem festgestellt, dass das Auge des Betrachters durch die Detaildichte der Texturen gelenkt wird (DiPaola et al. 2010).

Uusitalo et al. (2009) fragten sich, ob abstrakte und realistische Kunst unterschiedlich betrachtet wird. Außerdem wurde die Vorliebe zwischen diesen beiden Kategorien an Bildern bei Laien untersucht. Sie zeigten, dass Laien in ihrer Auswahl an Bildern die realistische Kunst bevorzugten. Interessanterweise unterschied sich die allgemeine Betrachtung nicht zwischen der abstrakten und der realistischen Kunst. Um den Einfluss von *top-down* und *bottom-up* Prozessen auf das visuelle Verhalten von Laien zu untersuchen, verglichen Massaro et al. (2012) das Blickverhalten für Farb- und Schwarzweißgemälde. In diesen Gemälden wurden sowohl Landschaften als auch Menschen dargestellt. Es zeigte sich, dass bei Landschaftsmalerei die Blicke eher nach *bottom-up* Prozessen angeordnet sind. Sobald aber ein Mensch im Gemälde abgebildet war, galt die Aufmerksamkeit ihm. Es wird angenommen, dass für diese Modulation der Wahrnehmung die *top-down* Prozesse verantwortlich sind. Dies steht im Einklang mit dem *eye contact effect* (Senju und Johnson 2009). Er besagt, dass bei Augenkontakt die Netzwerke für soziales Verhalten miterregt werden.

Zusätzlich gibt es entwicklungsgeschichtliche und kulturelle Unterschiede. Menschenaffen verfolgen die Kopfbewegung eines Menschen, auch wenn sich dessen Fixation nicht ändert. Menschen hingegen achten ausschließlich auf die Fixationen des zu Beobachtenden (Tomasello et al. 2007). Außerdem fällt eine kulturelle Abhängigkeit von Augenbewegungen bei der Gesichtswahrnehmung auf. Blais et al. (2008) fanden in einer Vergleichsstudie zwischen 14 Ost-Asiaten und 14 West-Europäern, dass Ost-Asiaten für die Gesichtserkennung die Mitte des Gesichtes betrachten, hingegen schauen West-Europäer mehr auf die Augen und teilweise den Mund.

Villani et al. (2015) untersuchten den in einem Bild gezeigten sozialen Kontext und seine Auswirkungen auf die Augenbewegungen detaillierter. Sie nutzten Gemälde, die die Handlung eines Einzelnen oder eine soziale Interaktion zwischen mehreren Parteien zeigten und wiesen nach, dass bei sozialer Interaktion vermehrt Gesichter und bei einzelnen Handlungen mehr auf den Körper geachtet wird. Außerdem betrachten Probanden, die höhere Werte im *Interpersonal Reactivity Index* in der Kategorie *Empathic Concern*, die eine Aussage über die Empathie gibt, besitzen, schneller die Gesichter der gezeigten Personen. Es lässt sich somit zusammenfassen, dass viele Faktoren, wie *low-level* Bildeigenschaften, aber auch *high-level* visuelle Prozesse wie empathisches Empfinden und Erlerntes, wie der kultureller Hintergrund, unser Betrachten der Umgebung beeinflussen.

Eye Tracking wird auch in der medizinischen Forschung angewendet (Toh et al. 2011). Boraston und Blakemore (2007) fassen in ihrem Übersichtsartikel zur Anwendung von *Eye Tracking* bei Erkrankungen aus dem Autismusspektrum zusammen, dass Autisten seltener Gesichter und noch seltener die Augen betrachten. Diese Beobachtung beschränkt sich aber nicht nur auf die Wahrnehmung von Gesichtern, sondern lässt sich auch auf abstrakte Formen ausweiten. Hierfür zeigten Boraston et al. (2007), dass Autisten sog. emotionalen Dreiecke, d.h. abstrakte Formen die miteinander interagieren, anders bewerten. Dies konnten Jones et al. (2008) in ihrer Studie zur Betrachtung von Gesichtern nicht nur bestätigen, sondern sie entdeckten dabei eine direkte Korrelation zwischen dem Schweregrad des Autismus und der Veränderung des Blickverhaltens. Einen weiteren Beitrag leisteten Sterling et al. (2008), indem sie Unterschiede zwischen bekannten und unbekanntem Gesichtern zu finden suchten. Es bestätigte sich, dass Autismus die Fixationen der Augen des zu Betrachtenden verringert. Außerdem zeigen Autisten keinen Unterschied bei der Betrachtung von bekannten und unbekanntem Gesichtern. Die gesunden Probanden fixierten hingegen die bekannten Gesichter weniger im Vergleich zu den unbekanntem. Riby und Hancock (2009) wählten für ihre Studie nicht Gesichter, sondern Cartoons (Die Abenteuer von Tim und Struppi©), Filmausschnitte und einzelne realistische Bilder, die alle soziale Interaktionen von Menschen darstellten. Als Versuchsgruppen dienten Menschen mit Autismus und Williams-Beuren-Syndrom, wobei hier, anders als beim Autismus das empathische Verhalten gesteigert ist. Es zeigte sich, dass das Betrachtungsverhalten bei Autisten in jeder der gezeigten Kategorien verändert war. Interessanterweise galt dies nicht für das Williams-Beuren-Syndrom. Hier war das Betrachtungsverhalten bei menschlicher Interaktion verändert, nicht aber bei den Cartoons. Da bei der neurologisch degenerativen Erkrankung wie Alzheimer (Hal-

pern et al. 2008) gezeigt werden konnte, dass die ästhetische Wahrnehmung sich mit dem Fortschreiten der Erkrankung nicht ändert, wäre es für zukünftige Untersuchungen interessant, wie sich Erkrankungen des Nervensystems auf das *Cropping*-Verhalten auswirken und ob derartige Erkrankungen das ästhetische Empfinden beeinflussen. Man könnte davon ausgehen, dass sich bei neurodegenerativen Erkrankungen das *Cropping*-Verhalten nicht verändert, jedoch könnte es Unterschiede bei Erkrankungen aus dem Autismusspektrum geben. Diese Hypothesen sind zukünftigen Untersuchungen überlassen.

4.5 Schlussfolgerung

Durch meine Arbeit war es mir möglich, objektive Parameter zu identifizieren, die eine wichtige Rolle bei der Croppingentscheidung spielen. Eine zentrale Bedeutung nehmen hierbei die Salienz und die dadurch resultierenden Augenbewegungen ein. Ich konnte zeigen, dass die Verteilung dieser Parameter nicht zufällig, sondern balanciert erfolgt und so einen Beitrag zum Konzept der *pictorial balance* leisten. Ebenso konnte ich zeigen, dass dieses Konzept unabhängig vom Inhalt des Bildes ist. Leider war es mir nicht möglich, Unterschiede zwischen Experten und Laien zu finden, was noch genauerer Untersuchungen in zukünftigen Arbeiten bedarf, da wir aufgrund des *central-tendency-bias* nur gepaarte Details untersuchen konnten. Es war mir so möglich, einen Beitrag zum Verständnis der *bottom-up* Prozesse der visuellen Wahrnehmung zu leisten, die eine Rolle bei Fragestellungen zur Ästhetik in der Photographie spielen. Zum Schluss diskutierte ich, wie Erkrankungen des Nervensystems unser Sehempfinden beeinflussen. Für zukünftige Studien wäre es von Interesse zu untersuchen, wie sich Erkrankungen auf das *Cropping*-Verhalten auswirken.

Literaturverzeichnis

- Achanta R, Hemami S, Estrada F, Susstrunk S. 2009. Frequency-tuned salient region detection. In: Computer Vision and Pattern Recognition. Miami, USA: IEEE.
- Ardizzone E, Bruno A, Mazzola G. 2013. Saliency based image cropping. In: Petrosino A, Hrsg. International Conference on Image Analysis and Processing, Bd 8156 von Lecture Notes in Computer Science. Berlin, Heidelberg, Deutschland: Springer.
- Arnheim R. 1954. Art and Visual Perception: A Psychology of the Creative Eye. Berkeley, Los Angeles, USA: University of California Press.
- Arnheim R. 1988. The Power of the Center: A Theory of Composition in the Visual Arts. Berkeley, Los Angeles, USA: University of California Press.
- Behrends E. 2000. Introduction to Markov chains. Berlin, Heidelberg, Deutschland: Springer.
- Behrends J, Bischofsberger J, Deutzmann R, Ehmke H, Frings S, Grissmer S, Hoth M, Kurtz A, Leibziger J, Müller F, et al. 2010. Duale Reihe Physiologie. Stuttgart, Deutschland: Thieme.
- Bhattacharya S, Sukthankar R, Shah M. 2010. A Framework for Photo-quality Assessment and Enhancement Based on Visual Aesthetics. In: 18th ACM International Conference on Multimedia. New York, USA: ACM.
- Bindemann M. 2010. Scene and Screen Center Bias Early Eye Movements in Scene Viewing. Vision Research, 50(23):2577–2587.
- Blais C, Jack RE, Scheepers C, Fiset D, Caldara R. 2008. Culture Shapes how We Look at Faces. PloS One, 3(8):e3022.
- Boraston Z, Blakemore SJ. 2007. The Application of Eye-Tracking Technology in the Study of Autism. Journal of Physiology, 581(3):893–898.

- Boraston Z, Blakemore SJ, Chilvers R, Skuse D. 2007. Impaired Sadness Recognition is Linked to Social Interaction Deficit in Autism. *Neuropsychologia*, 45(7):1501–1510.
- Borji A, Sihite DN, Itti L. 2013. Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study. *IEEE Transactions on Image Processing*, 22(1):55–69.
- Ciocca G, Cusano C, Gasparini F, Schettini R. 2007. Self-Adaptive Image Cropping for Small Displays. *IEEE Transactions on Consumer Electronics*, 53(4):1622–1627.
- Connor CE, Egeth HE, Yantis S. 2004. Visual Attention: Bottom-Up Versus Top-Down. *Current Biology*, 14(19):850–852.
- Datta R, Joshi D, Li J, Wang JZ. 2006. Studying Aesthetics in Photographic Images Using a Computational Approach. In: A Leonardis AP H Bischof, Hrsg. *European Conference on Computer Vision*, Bd 3953 von *Lecture Notes in Computer Science*. Berlin, Heidelberg, Deutschland: Springer.
- DeCarlo D, Santella A. 2002. Stylization and Abstraction of Photographs. In: *29th Annual Conference on Computer Graphics and Interactive Techniques*. New York, USA: ACM.
- DiPaola S, Riebe C, Enns JT. 2010. Rembrandt’s Textural Agency: A Shared Perspective in Visual Art and Science. *Leonardo*, 43(2):145–151.
- Downing PE, Jiang Y, Shuman M, Kanwisher N. 2001. A Cortical Area Selective for Visual Processing of the Human Body. *Science*, 293(5539):2470–2473.
- Epstein R, Kanwisher N. 1998. A Cortical Representation of the Local Visual Environment. *Nature*, 392(6676):598.
- Frintrop S, Rome E, Christensen HI. 2010. Computational Visual Attention Systems and their Cognitive Foundations: A Survey. *ACM Transactions on Applied Perception (TAP)*, 7(1):6.
- Gershoni S, Hochstein S. 2011. Measuring Pictorial Balance Perception at First Glance Using Japanese Calligraphy. *i-Perception*, 2(6):508–527.
- von Goethe JW. 1812. *Aus meinem Leben. Dichtung und Wahrheit*. Tübingen, Deutschland: Cotta.

- Graham DJ, Redies C. 2010. Statistical Regularities in Art: Relations with Visual Coding and Perception. *Vision Research*, 50(16):1503–1509.
- Halpern AR, Ly J, Elkin-Frankston S, O'Connor MG. 2008. "I know what I like": Stability of Aesthetic Preference in Alzheimer's Patients. *Brain and Cognition*, 66(1):65–72.
- Harel J, Koch C, Perona P. 2007. Graph-Based Visual Saliency. In: 19th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press.
- Hasse C, Weber R. 2012. Eye Movements on Facades: the Subjective Perception of Balance in Architecture and its Link to Aesthetic Judgment. *Empirical Studies of the Arts*, 30(1):7–22.
- Henderson JM, Brockmole JR, Castelhana MS, Mack M. 2007. Visual Saliency does not Account for Eye Movements during Visual Search in Real-World Scenes. *Eye Movements: A Window on Mind and Brain*:537–562.
- Hogg M, Vaughan G. 2010. *Essentials of Social Psychology*. Essex, England: Pearson Education.
- von Holst E, Mittelstaedt H. 1950. Das Reafferenzprinzip. *Naturwissenschaften*, 37(20):464–476.
- Hubel DH, Wiesel TN. 1962. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *Journal of Physiology*, 160(1):106–154.
- Hudspeth AJ, Jessell TM, Kandel ER, Schwartz JH, Siegelbaum SA. 2013. *Principles of Neural Science*. New York, USA: McGraw-Hill, Health Professions Division.
- Itti L, Koch C, Niebur E. 1998. Computer Vision and Image Analysis-A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):12–1120.
- Jones W, Carr K, Klin A. 2008. Absence of Preferential Looking to the Eyes of Approaching Adults Predicts Level of Social Disability in 2-Year-Old Toddlers with Autism Spectrum Disorder. *Archives of General Psychiatry*, 65(8):946–954.

- Kanwisher N, McDermott J, Chun MM. 1997. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, 17(11):4302–4311.
- Kapoula Z, Daunys G, Herbez O, Yang Q. 2009. Effect of Title on Eye-Movement Exploration of Cubist Paintings by Fernand Léger. *Perception*, 38(4):479–491.
- Koide N, Kubo T, Nishida S, Shibata T, Ikeda K. 2015. Art Expertise Reduces Influence of Visual Salience on Fixation in Viewing Abstract-Paintings. *PloS One*, 10(2):e0117696.
- Kretschmann HJ, Weinrich W. 1996. Dreidimensionale Computergraphik neurofunktioneller Systeme: Grundlagen für die neurologisch-topische Diagnostik und die kraniale Bilddiagnostik. Stuttgart, Deutschland: Thieme.
- Liu L, Chen R, Wolf L, Cohen-Or D. 2010. Optimizing Photo Composition. *Computer Graphics Forum*, 29(2):469–478.
- Locher P. 2006. The Usefulness of Eye Movement Recordings to Subject an Aesthetic Episode with Visual Art to Empirical Scrutiny. *Psychology Science*, 48(2):106–114.
- Locher P, Cornelis E, Wagemans J, Stappers PJ. 2001. Artists' use of Compositional Balance for Creating Visual Displays. *Empirical Studies of the Arts*, 19(2):213–227.
- Locher P, Gray S, Nodine C. 1996. The Structural Framework of Pictorial Balance. *Perception*, 25(12):1419–1436.
- Locher P, Nagy Y. 1996. Vision Spontaneously Establishes the Percept of Pictorial Balance. *Empirical Studies of the Arts*, 14(1):17–31.
- Locher P, Overbeeke K, Stappers PJ. 2005. Spatial Balance of Color Triads in the Abstract Art of Piet Mondrian. *Perception*, 34(2):169–189.
- Locher PJ. 2003. An Empirical Investigation of the Visual Rightness Theory of Picture Perception. *Acta Psychologica*, 114(2):147–164.
- Locher PJ, Stappers PJ, Overbeeke K. 1998. The Role of Balance as an Organizing Design Principle Underlying Adults' Compositional Strategies for Creating Visual Displays. *Acta Psychologica*, 99(2):141–161.

- Locher PJ, Stappers PJ, Overbeeke K. 1999. An Empirical Evaluation of the Visual Rightness Theory of Pictorial Composition. *Acta Psychologica*, 103(3):261–280.
- Massaro D, Savazzi F, Di Dio C, Freedberg D, Gallese V, Gilli G, Marchetti A. 2012. When Art Moves the Eyes: A Behavioral and Eye-Tracking Study. *PloS One*, 7(5):e37285.
- McManus I, Edmondson D, Rodger J. 1985. Balance in Pictures. *British Journal of Psychology*, 76(3):311–324.
- McManus I, Stöver K, Kim D. 2011a. Arnheim’s Gestalt Theory of Visual Balance: Examining the Compositional Structure of Art Photographs and Abstract Images. *i-Perception*, 2(6):615–647.
- McManus IC, Cheema B, Stoker J. 1993. The Aesthetics of Composition: A Study of Mondrian. *Empirical Studies of the Arts*, 11(2):83–94.
- McManus IC, Zhou FA, l’Anson S, Waterfield L, Stöver K, Cook R. 2011b. The Psychometrics of Photographic Cropping: The Influence of Colour, Meaning, and Expertise. *Perception*, 40(3):332–357.
- Miall RC, Tchalenko J. 2001. A Painter’s Eye Movements: A Study of Eye and Hand Movement During Portrait Drawing. *Leonardo*, 34(1):35–40.
- Mooney CM. 1956. Closure with Negative After-Images under Flickering Light. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 10(4):191–199.
- Niekamp W. 1981. An Exploratory Investigation into Factors Affecting Visual Balance. *Educational Technology Research and Development*, 29(1):37–48.
- Niu Y, Todd RM, Kyan MJ, Anderson AK. 2012. Visual and Emotional Saliency Influence Eye Movements. *ACM Transactions on Applied Perception*, 9(3):13:1–13:18.
- Palmer SE, Gardner JS, Wickens TD. 2008. Aesthetic Issues in Spatial Composition: Effects of Position and Direction on Framing Single Objects. *Spatial Vision*, 21(3):421–449.
- Palmer SE, Guidi S. 2011. Mapping the Perceptual Structure of Rectangles Through Goodness-of-Fit Ratings. *Perception*, 40(12):1428–1446.

- Palmer SE, Schloss KB, Sammartino J. 2013. Visual Aesthetics and Human Preference. *Annual Review of Psychology*, 64(1):77–107.
- Parikh N, Itti L, Weiland J. 2010. Saliency-Based Image Processing for Retinal Prostheses. *Journal of Neural Engineering*, 7(1):016006.
- Prasad S, Galetta SL. 2011. Anatomy and Physiology of the Afferent Visual System. In: Kennard C, Leigh RJ, Hrsg. *Neuro-Ophthalmology*, Bd 102 von *Handbook of Clinical Neurology*. Kap 1. Elsevier, 3 – 19.
- Redies C. 2015. Combining Universal Beauty and Cultural Context in a Unifying Model of Visual Aesthetic Experience. *Frontiers in Human Neuroscience*, 9:218.
- Riby D, Hancock PJ. 2009. Looking at Movies and Cartoons: Eye-Tracking Evidence from Williams Syndrome and Autism. *Journal of Intellectual Disability Research*, 53(2):169–181.
- Santella A, Agrawala M, DeCarlo D, Salesin D, Cohen M. 2006. Gaze-based Interaction for Semi-automatic Photo Cropping. In: *SIGCHI Conference on Human Factors in Computing Systems*. New York, USA: ACM.
- Schnapf J, Kraft T, Nunn B, Baylor D. 1988. Spectral Sensitivity of Primate Photoreceptors. *Visual Neuroscience*, 1(3):255–261.
- Senju A, Johnson MH. 2009. The Eye Contact Effect: Mechanisms and Development. *Trends in Cognitive Sciences*, 13(3):127–134.
- Sterling L, Dawson G, Webb S, Murias M, Munson J, Panagiotides H, Aylward E. 2008. The Role of Face Familiarity in Eye Tracking of Faces by Individuals with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 38(9):1666–1675.
- Suh B, Ling H, Bederson BB, Jacobs DW. 2003. Automatic Thumbnail Cropping and Its Effectiveness. In: *16th Annual ACM Symposium on User Interface Software and Technology*. New York, USA: ACM.
- Sun X, Yao H, Ji R, Liu S. 2009. Photo Assessment Based on Computational Visual Attention Model. In: *17th ACM International Conference on Multimedia*. New York, USA: ACM.

- Tatler BW. 2007. The Central Fixation Bias In scene Viewing: Selecting an Optimal Viewing Position Independently of Motor Biases and Image Feature Distributions. *Journal of Vision*, 7(14):4.
- Toh WL, Rossell SL, Castle DJ. 2011. Current Visual Scanpath Research: A Review of Investigations into the Psychotic, Anxiety, and Mood Disorders. *Comprehensive Psychiatry*, 52(6):567–579.
- Tomasello M, Hare B, Lehmann H, Call J. 2007. Reliance on Head Versus Eyes in the Gaze Following of Great Apes and Human Infants: The Cooperative Eye Hypothesis. *Journal of Human Evolution*, 52(3):314–320.
- Tong H, Li M, Zhang HJ, He J, Zhang C. 2004. Classification of Digital Photos Taken by Photographers or Home Users. In: K Aizawa SS Y Nakamura, Hrsg. Pacific-Rim Conference on Multimedia, Bd 3331 von Lecture Notes in Computer Science. *Advances in Multimedia Information Processing*. Berlin, Heidelberg, Deutschland: Springer.
- Treisman A. 1986. Features and Objects in Visual Processing. *Scientific American*, 255(5):114–125.
- Treisman AM, Gelade G. 1980. A Feature-Integration Theory of Attention. *Cognitive Psychology*, 12(1):97–136.
- Troiani V, Stigliani A, Smith ME, Epstein RA. 2012. Multiple Object Properties Drive Scene-Selective Regions. *Cerebral Cortex*, 24(4):883–897.
- Uusitalo L, Simola J, Kuisma J. 2009. Perception of Abstract and Representative Visual Art. In: 10th Conference of the International Association of Arts and Cultural Management. Dallas, USA.
- Vartanian O, Martindale C, Podsiadlo J, Overbay S, Borkum J. 2005. The Link Between Composition and Balance in mMsterworks vs. Paintings of Lower Artistic Quality. *British Journal of Psychology*, 96(4):493–503.
- Villani D, Morganti F, Cipresso P, Ruggi S, Riva G, Gilli G. 2015. Visual Exploration Patterns of Human Figures in Action: An Eye Tracker Study with Art Paintings. *Frontiers in Psychology*, 6:1636.

- Vogt S, Magnussen S. 2007. Expertise in Pictorial Perception: Eye-Movement Patterns and Visual Memory in Artists and Laymen. *Perception*, 36(1):91–100.
- Walther D, Koch C. 2006. Modeling Attention to Salient Proto-Objects. *Neural Networks*, 19(9):1395–1407.
- Welsch U, Deller T. 2010. *Lehrbuch Histologie*. München, Deutschland: Elsevier.
- Wong LK, Low KL. 2009. Saliency-Enhanced Image Aesthetics Class Prediction. In: 16th IEEE International Conference on Image Processing (ICIP). Kairo, Ägypten: IEEE.
- Wong LK, Low KL. 2011. Saliency retargeting: An approach to enhance image aesthetics. In: Workshop on Applications of Computer Vision. Kona, USA: IEEE.
- Wurtz RH. 2008. Neuronal Mechanisms of Visual Stability. *Vision Research*, 48(20):2070–2089.
- Yarbus AL. 1967. *Eye Movement and Vision*. New York, USA: Plenum Press.

Anhang

Danksagung

Zuallererst gilt ein ganz besonderer Dank meinem Doktorvater Prof. Dr. med. Dr. rer. nat. Christoph Redies für die herausragende Betreuung und ständige Unterstützung, das stetige Einbinden im Prozess der Veröffentlichung der Daten und der Möglichkeit einer publikationsbasierten Doktorarbeit. Besonders dankbar bin ich ihm für das Heranführen an die wissenschaftliche Arbeitsweise und den Einblick in ein Forschungsgebiet aus Informatik und Psychologie.

Weiterer Dank gilt Dr. rer. nat. Seyed Ali Amirshahi für die Unterstützung bei der Arbeit mit den Salienzalgorithmen. Er war maßgeblich bei der Auswahl der Algorithmen und Programmierung der Analyseprogramme beteiligt.

Prof. Chris McManus möchte ich für die Bereitstellung und Erläuterung der von ihm erhobenen Daten danken, die die Grundlage unserer Versuche darstellen.

Dr. Helene Kreysa und der „Person Perception DFG Research Unit“ der Friedrich-Schiller-Universität Jena möchte ich für die Bereitstellung der Räumlichkeiten, des Eye-Trackers und Einweisung in den Eye-Tracker danken.

Leonie Fresz danke ich für die Hilfe bei der Durchführung und Erstellung der Eye-Tracking Experimente.

Sabrina Vanessa Schmitz danke ich für das Lektorieren meiner Arbeit.

Außerdem möchte ich meiner Mutter Kerstin und meinem verstorbenen Vater Dr. med. Hermann Abeln für ihre ständige Unterstützung, und dafür, dass sie mir das Studium der Humanmedizin ermöglicht haben, danken.

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass

- mir die Promotionsordnung der Medizinischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist,
- ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,
- mich folgende Personen bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt haben: Herrn Prof. Dr. med. Dr. rer. nat. Christoph Redies, Herrn Dr. rer. nat. Seyed Ali Amirshahi, Frau Leonie Fresz, Herrn Prof. Chris McManus MD
- die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde,
- Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,
- ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Münster, am 17. Juli 2020

Jonas Abeln, Verfasser