

ElectroEncephalographics: Eine neue Modalität für die Grafikkforschung¹

Maryam Mustafa²

Abstract: In dieser Arbeit präsentiere ich die Anwendung von Elektroenzephalografie (EEG) als eine neuartige Modalität zur Untersuchung von Wahrnehmungsfragen in der Computergraphik. Bisher wurde EEG vorwiegend für die klinische Diagnostik, in der Psychologie und in der BCI-Community verwendet. Ich erweitere den bisherigen Anwendungsbereich um die Untersuchung von perzeptueller Qualität bildgebender Verfahren auf Basis von neuronalem Feedback.

Da die Ergebnisse der meisten graphischen bildgebenden Verfahren für die Betrachtung durch Menschen bestimmt sind, ist bei der Bildsynthese neben der physikalischen Genauigkeit ebenso die durch den Betrachter tatsächlich wahrgenommene Qualität von großer Bedeutung. Um die tatsächliche wahrgenommene Qualität von Videos und Bildern zu ermitteln, setze ich in meiner Arbeit mit EEG gemessene Daten ein.

1 Einführung

Das Ziel der Computergraphik ist es, Bilder und Videos für menschliche Betrachter zu erzeugen. Der Erfolg jeder Grafik-Pipeline hängt davon ab, wie gut relevante Informationen an den menschlichen Betrachter übermittelt werden können. Die inhärente Komplexität der physikalischen Welt und die Grenzen der Hardware machen es jedoch unmöglich die reale Welt exakt zu replizieren. Die enorme Komplexität der Aufgabe physikalisch korrektes Bildmaterial zu erzeugen macht es notwendig sich auf die Erzeugung von korrekt wahrgenommenen Bildern zu beschränken. Um dieses Ziel zu erreichen ist es essentiell das menschliche Sehsystem (engl. Human Visual System (HVS)) zu analysieren und zu verstehen. Dieses Wissen erlaubt es Computergraphikern die Flexibilität und Robustheit der menschlichen Wahrnehmung auszunutzen, um so die Lücke zwischen Hardwareperformanz und gewünschter Performanz zu verkleinern. Die Integration von Forschung im Bereich der Wahrnehmung in die Computergraphik ist besonders wichtig, wenn realistisches Material für Filme, Spiele und immersive Umgebungen erzeugt werden soll.

Neuroimaging und Brainimaging-Techniken können wesentliche Einsichten und Erkenntnisse für Probleme der Computergraphik liefern. Bis vor kurzem wurde EEG hauptsächlich zur klinischen Diagnose, in der Psychologie und von der Brain-Computer-Interface (BCI) Community genutzt. In diese Arbeit wird der Anwendungsbereich so erweitert, dass EEG zur Verbesserung des Verständnis der Wahrnehmung von visuellen Ausgaben von computergraphischen Anwendungen genutzt werden kann. Außerdem erlauben die Erweiterungen neue Methoden basierend auf direktem neuronalen Feedback. Die in dieser Arbeit

¹ ElectroEncephalographics: A Novel Modality for Graphics Research

² Institut für Computergraphik, TU Braunschweig, mustafa@cg.cs.tu-bs.de

vorgestellte Methodik verwendet EEG Daten um die wahrgenommene Qualität von Videos und Bildern zu bestimmen. Dies ist besonders wichtig auf Grund des Unterschieds zwischen wahrgenommener Qualität und der physikalischen Präzision eines Bildes.

Die Untersuchung der Wahrnehmung ist schwierig, da sie weder direkt gemessen noch beobachtet werden kann. Die einzige Möglichkeit die menschliche Wahrnehmung zu untersuchen sind indirekte Messungen. Jeder Wahrnehmungsprozess beeinflusst das menschliche Verhalten bis zu einem gewissen Grad. Die Untersuchung dieses Verhaltens erlaubt die Konstruktion von Modellen des Wahrnehmungssystems. Eine der wichtigsten Methoden die verwendet werden um verdeckte Wahrnehmungsprozesse zu untersuchen sind psychophysikalische Experimente. Diese Experimente werden in stark kontrollierten Umgebungen durchgeführt in denen eine direkte Kontrolle über möglichst viele Faktoren der Umgebung möglich ist. Um diese absolute Kontrolle über möglichst zu machen werden diese Experimente mit einfachen abstrakten Stimuli durchgeführt. Dies erschwert die Durchführung solcher Experimente im Kontext von Computergraphik sowie die Modellierung der Ergebnisse. Eine weitere Technik zur Erforschung von Wahrnehmung sind Eye-Tracker um festzustellen auf welche Bereiche von Bildern, Videos und Visualisierungen sich die Betrachter konzentrieren. Diese Werkzeuge sind fester Bestandteil der Forschung in der Computergraphik geworden und haben einzigartige Einsichten ermöglicht.

Neuroimaging Technologien erlauben es Forschern den Prozess der Verarbeitung von Information im Hirn direkt zu visualisieren. Die verbreitetste Brainimaging Methode ist Elektroenzephalografie (EEG), welche die Aktivität größerer Gruppen von Neuronen in der Nähe der Hirnoberfläche misst. Obwohl psychophysische Experimente und Eye-Tracker in der Computergraphik häufig verwendet werden, wurden Neuroimaging Methoden bisher nicht im Bereich der Computergraphik eingesetzt. Die meisten Neuroimaging Verfahren wurden überwiegend für medizinische Diagnostik, in der Wahrnehmungspsychologie und von der Brain-Computer Interface (BCI) Community verwendet, um die menschlichen kognitiven Fähigkeiten oder die Beweglichkeit zu unterstützen oder zu erweitern. Brainimaging und Brainmapping Techniken sind gut geeignet die digitale Repräsentation von Multimedia mit der Wahrnehmung und dem Verständnis der Inhalte zu verknüpfen. Diese Arbeit erweitert den Anwendungsbereich des EEG's um speziell die Wahrnehmung von visuellen Ausgaben von Computergraphik-Anwendungen zu erforschen. Es wird eine neue Methode entwickelt, die basierend auf direktem neuronalen Feedback einen Satz Parameter optimiert um realistischer wahrgenommenen Ausgaben zu erhalten.

In dieser Arbeit wird die Theorie aufgestellt, dass EEG als nicht-invasive und kostengünstige Methode eine praktikable Herangehensweise für die Untersuchung von Fragestellungen in der High-Level-Wahrnehmung im Bereich der Computergraphik ist [Mu15]. Es wurden die folgenden Hypothesen getestet: Die neuronale Reaktion auf verschiedene Bild- und Videoqualitäten, sind messbar, können mit Single-Trial EEG-Daten klassifiziert werden und können als Feedback verwendet werden, um die Bild-/Video-Parameter zu optimieren.

2 Bild- und Videoqualitätsbestimmung mit EEG

Bild- und Videobasierte Rendering Techniken erlauben die Erzeugung von realistischem Bildmaterial aus wenigen Bildern. Beim Renderingvorgang können Artefakte erzeugt werden, die besonders intensiv wahrgenommen werden und zur Ablehnung des Bildes führen, so dass die meisten Renderverfahren nur eingeschränkt anwendbar sind. Zu sehr häufig vorkommenden Artefakten gehören Ghosting, Blurring und Popping. Während andere Arbeiten ihren Fokus auf die expliziten Reaktionen (overt) des menschlichen Sehsystems (HVS) legen, verwendet diese Arbeit die impliziten Verarbeitung (covert) des HVS um die Wahrnehmbarkeit von Artefakten in Videos zu bestimmen. Sowohl die explizite als auch die implizite Verarbeitung im menschlichen Gehirn wird in der Literatur umfassend untersucht. In diesem Kapitel wird ein erster Schritt zur Nutzung aufgenommener neuronaler Reaktionen zur Erfassung der visuellen Qualität aufgezeigt. Da Bewegung in der menschlichen Wahrnehmung eine wichtige Rolle spielt, werden insbesondere Artefakte im Zusammenhang mit Bewegung in gerenderten Szenen untersucht.

Diese Studie wurde mit 8 gesunden Teilnehmern mit einem Durchschnittsalter von 25 und normaler oder zu normal korrigierter Sehkraft durchgeführt. Alle Teilnehmer hatten durchschnittliche Erfahrung mit digitalen Aufnahmen aber keine Erfahrung mit professioneller Bilderzeugung oder Verarbeitung. Ein 5,6 Sekunden langes Video (Auflösung 1440x1024 Pixel, 30fps) einer auf einem Parkweg gehenden Person (von links nach rechts) diente als Basisstimulus. Das Auftreten der Artefakte wurde um ± 4 Frames (± 132 ms) verzögert, um zu vermeiden, dass die sich die Teilnehmer an eine feste Zeit gewöhnen. Fünf verschiedene Arten von Artefakten wurden in die Szene eingebaut. Diese Artefakte beinhalten sowohl zeitliche als auch räumliche Komponenten.

Die folgenden sechs Testfälle wurden gezeigt:

1. Popping on Person: Ein kleiner rechteckiger Bereich, der Teile der gehende Person enthält, wird für einen Frame eingefroren.
2. Popping: Ein statischer rechteckiger Bereich, wird für ein Frame eingefroren.
3. Blurring on Person: ein kleiner rechteckiger Bereich, der Teile der gehende Person enthält, wird mit einem Gauß-Kernel mit einer Größe von 15 Pixeln in 10 aufeinander folgenden Frames unscharf gemacht. Der Unschärfe-Bereich bewegt sich mit der Bewegung der Person.
4. Blurring: Ein statischer rechteckiger Bereich in der Mitte der Szene wird mit einem unscharfen Gauß-Kernel von einer Größe von 15 Pixeln in 10 aufeinander folgenden Frames unscharf gemacht.
5. Ghosting on Person: Eine teilweise transparente Silhouette der Person bleibt für 10 Frames zurück und wird über die letzten 5 Frames unsichtbar.
6. No Artifacts: Original ohne Artefakte.

Ein Trial bestand aus einer Warteansicht gefolgt von einem Video mit Artefakten. Im direkten Anschluss wurde eine Qualitätsbewertung gefordert. Die Teilnehmer wurden auf-

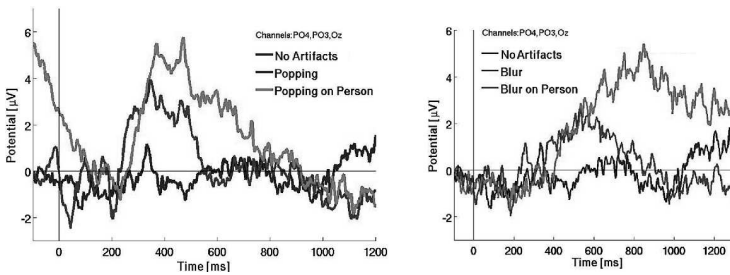


Abb. 1: Die ERPs für Popping und Popping on Person, Blurring, Blurring on Person und No Artifacts gemittelt über alle Trials und Teilnehmer.

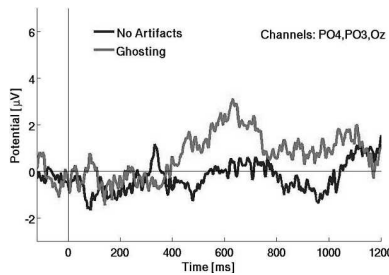


Abb. 2: Die ERPs für Ghosting und No Artifacts, gemittelt über alle Trials und Teilnehmer.

gefordert der bewegten Person mit dem Blick zu folgen und die Qualität jedes Testfalls mit einer ganzen Zahl zwischen 1 (sehr schlecht) und 5 (sehr gut) auf einer Mean Opinion Score (MOS) Skala zu bewerten. Die Teilnehmer wurden nicht auf das Vorhandensein von Artefakten in den Videos hingewiesen.

Abb. 1 und 2 zeigen die unterschiedlichen ERPs gemittelt über alle Teilnehmer der Studie, alle Trials und über alle Elektroden P04, P03 und Oz. Zum Vergleich zeigt “No Artifacts” das ERP der gleichen Sequenz ab dem gleichen Startzeitpunkt ohne Artefakte. Die Gültigkeit der Kurven wurde mit einem zweiseitigen t-Test bestätigt. Der Test liefert eine Wahrscheinlichkeit für die Annahme, dass zwei Mengen von Werten aus unterschiedlichen Gruppen stammen. Traditionell wird ein Testergebnis als Signifikant betrachtet wenn $p < 0.05$, was bedeutet, dass die Chance das beide Wertemengen aus der gleichen Gruppe stammen kleiner als 5% ist. In unserem Fall wurde der zweiseitige T-Test für alle Fälle von Artefakten gegen die Groundtruth Daten der exakt gleichen Sequenz vorgenommen. In allen Fällen wurde die Null-Hypothese nicht bestätigt. Für die P-Werte für Popping, Popping on Person, Blurring on Person und Ghosting gilt $p < 0.0001$ während für Blurring $p < 0.004$ gilt. Da die Null-Hypothese in allen Fällen nicht bestätigt wurde und die Wahrscheinlichkeit kleiner als 0.05 (5%) war, kann dies als ausreichender Hinweis für die statistische Signifikanz der Ergebnisse betrachtet werden. Der zweiseitige T-Test wurde zusätzlich auch zwischen Popping und Popping auf Person sowie zwischen Blurring und Blurring auf Person berechnet. In beiden Fällen wurde die Null-Hypothese nicht bestätigt (p -Werte kleiner 0.0001). Mit diesen Werten kann angenommen werden, dass ein signifikanter Unterschied zwischen den unterschiedlichen Arten von Artefakten vorliegt.

Um die Abweichung einzelner Teilnehmer vom Durchschnitt zu bestimmen wurde ein zusätzlicher zweiseitiger T-Test zwischen einem zufällig ausgewählten Teilnehmer und dem Durchschnitt der Teilnehmer berechnet. Für alle Stimuli wurde die Null-Hypothese angenommen (p -Werte waren für Popping $p = 0.14$, Popping auf Person $p = 0.86$, Blurring $p = 0.85$, Blurring auf Person $p = 0.97$ und Ghosting $p = 0.74$). Damit ist klar, dass die Reaktion jedes Teilnehmers nahe am Durchschnitt liegt und es somit keine statistisch signifikanten Unterschiede gibt.

Die Auswertung dieser ERPs bietet einige sehr interessante Erkenntnisse. Zunächst wurden alle Artefakte, wenn auch mit unterschiedlicher Stärke, vom Gehirn erfasst. Der Artefakttyp mit der größten Reaktion im ERP war Popping on Person mit einer Latenz von 264ms und einer maximalen Amplitude von 5.758 μ V, dicht gefolgt von Blurring on Person. Statisches Popping und Blurring riefen jeweils kleinere Reaktionen hervor. Offenbar ist Popping ein einfacher wahrzunehmendes Artefakt, das eine schnellere Reaktion zu Folge hat. Ghosting dagegen scheint eine Verarbeitung der wahrgenommenen Verzerrung zu erfordern bevor es zu einer Reaktion kommt. Diese erhöhte Latenz kann man auch beim weniger offensichtlichen Blurring beobachten.

Die Ergebnisse zeigen, dass das Gehirn nicht nur unterschiedliche auf verschiedene Arten von Artefakten reagiert sondern insbesondere auch auf die Verknüpfung der Artefakte mit Bewegung. Diese Analyse liefert Informationen über die Wahrnehmung von Videos die so bisher nicht modelliert werden konnten. Zusätzlich erlauben die Erkenntnisse eine Reduktion der Renderzeiten indem Berechnungen für Bild-Features ausgelassen werden, die keine starke Reaktion im Hirn hervorrufen. Zusätzlich ist die Reaktion des Gehirns auf Artefakte auch von entscheidender Bedeutung bei der Modellierung von Maskierungsalgorithmen für gerenderte Sequenzen. Diese ersten Experimente zeigen die Möglichkeit auf, mittels EEG Daten die Qualität von Video- oder Bildmaterial zu bestimmen.

3 Single Trial Analyse von EEG-Daten

In diesem Kapitel der Arbeit werden die Daten des vorangegangenen Kapitels verwendet um den Analyseprozess von EEG Daten zu verkürzen. Statt mehreren Versuchen über mehrere Versuchspersonen durchführen zu müssen, kann ein Versuch für eine Versuchsperson vorgenommen werden. Eine der größten Einschränkungen bei der Verwendung eines EEGs für die Bewertung der Bildqualität ist das niedrige Signal-Rausch Verhältnis (SNR), das es zunehmend schwierig macht Nervenreaktionen von Rauschen zu unterscheiden. Traditionell verwendet man für die Auswertung von EEG Daten ereignisbezogene Potentiale (ERP) verwendet. Da diese jedoch auf Mittelwertbildungen basieren, benötigen sie eine große Anzahl von Versuchspersonen und Versuchen um aussagekräftige Daten zu erhalten. Aufgrund des niedrigen SNR, sind ERPs nicht ohne weiteres für die single-trial Klassifizierung geeignet.

Die hier vorgestellte Methode verwendet einen neuen Wavelet-basierten Ansatz zur Auswertung von EEG Signalen, der eine Vorhersage der wahrgenommenen Bildqualität aus nur einem einzigen Versuch ermöglicht. Der Wavelet-basierte Algorithmus ist in der Lage die EEG Daten zu filtern und das Rauschen zu entfernen, wodurch die Notwendigkeit

vieler Versuchspersonen und Versuche entfällt. So ist es möglich, Daten aus nur 10 Elektrodenkanälen für single-trial Klassifizierung zu verwenden und das Vorhandensein von Artefakten mit einer Genauigkeit von 85% vorherzusagen. Die Ergebnisse zeigen außerdem, dass es möglich ist, Versuche anhand des gezeigten Artefakttypen zu unterscheiden und zu klassifizieren. Diese Arbeit ist besonders nützlich um zu verstehen wie das menschliche Wahrnehmungssystem auf verschiedene Arten von Artefakten in Bildern und Videos reagiert. Ein Verständnis der Wahrnehmung typischer bildbasierter Rendering-Artefakte bildet die Basis für die Optimierung von Rendering- und Maskierungsalgorithmen.

Die EEG-Daten werden zunächst mit Hilfe einer komplexen Wavelet-Transformation verarbeitet. Sie werden dann mit einer standard Support Vector Machine (SVM) für alle Klassifizierungsaufgaben klassifiziert. Für die Statistiken wird ein standard 5-fold Cross-Correlation Test vorgenommen. The Daten werden zufällig in 5 Gruppen von 288 Versuchen aufgeteilt. Mit Hilfe einer C-SVM mit einem radialen Basisfunktionen (RBF) $e^{-g|x_i-x_j|^2}$ Klassifikator und einer Menge von festen Parametern, wird die SVM mit Daten von 4 Gruppen (976 Versuche) angelernt und gegen die Versuche der verbleibenden Gruppe (288 Versuche) getestet. Dieser Vorgang wird wiederholt, bis alle Versuche klassifiziert worden sind. Der Vorgang wird wiederholt bis die beste Menge an Parametern gefunden worden ist.

Die Ergebnisse zeigen, dass alle Artefakte vom Geherin erkannt wurden. Das Artefakt, das die größte Reaktion hervorgerufen hat, war 'Popping on Person' (popP). Dies ist dicht gefolgt von 'Popping'. Popping ist ein deutlicher wahrgenommenes Artefakt und ruft eine schnellere und stärkere Reaktion hervor, als popping das nicht mit Bewegung zusammenhängt. Wie man sehen kann, zeigt 'Ghosting' die geringste Reaktion im Hinblick auf Latenz und relative Stärke. Offenbar muss das Gehirn die wahrgenommene Verzerrung zunächst verarbeiten, ehe eine EEG-Reaktion messbar wird. Diese Latenz durch der Verarbeitung der wahrgenommenen Stimuli, kann auch beim 'blurring' gefunden werden, bei dem es sich ebenfalls um ein weniger offensichtliches Artefakt handelt. Es ist jedoch interessant zu beobachten, dass blurring in Verbindung mit Bewegung eine längere Latenz hat, aber eine höhere Reaktion im Hinblick auf die relative Steigung der Stärke.

Angesichts der statistischen Signifikanz zwischen Ground Truth und eines Artefaktes in den EEG-Daten, befasst sich die Studie mit insgesamt drei Klassifizierungsaufgaben: Die Klassifizierung von Versuchen in eine von zwei Kategorien, Versuche mit Artefakten vs Versuche ohne Artefakte; die Klassifizierung von Versuchen nach der Stärke des Artefakts mit dem Ziel schwere Artefakte zu erkennen; und die Klassifizierung jedes Versuches im Hinblick auf den spezifischen Typ des Artefakts.

Die Klassifikation der single-trial Daten der frontalen Elektroden die Wavelet-Transformiert wurden, besitzt eine Treffsicherheit von 85%. Betrachtet man die Wavelet-Transformierten Daten der frontalen Elektroden beträgt die Treffsicherheit der Klassifikation 93%. Dies bedeutet, dass die Erkennung von starken Artefakten am einfachsten erkannt werden kann. Eine zufällige Auswahl und Zuordnung eines Artefakttyps würde zu einer erwarteten Präzision von 16% führen. Die Klassifizierung muss also eine relevant bessere Genauigkeit aufweisen um als erfolgreich betrachtet werden zu können. Unter Verwendung der Daten der

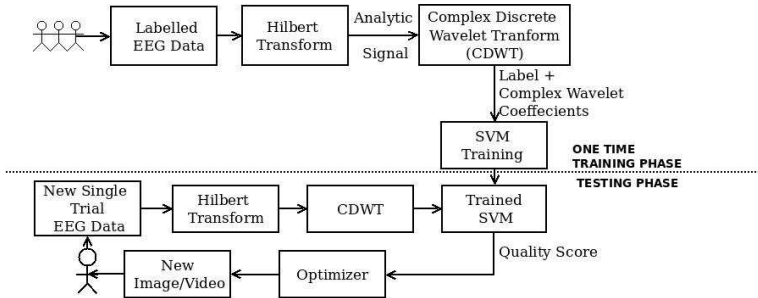


Abb. 3: Das Framework der Optimierungsschleife benötigt eine Trainingsphase. Diese wird nur ein einmal und mit einer disjunkten Menge von Menschen, die nicht Bestandteil der Prüfung sind, durchgeführt. Danach kann das Verfahren Bilder und Videos für beliebige Personen optimieren.

frontalen Elektroden liegt die Genauigkeit der Klassifizierung bei 64%. Somit ist nicht nur die Bestimmung der wahrgenommenen Qualität von gerendertem Material möglich, sondern auch die Bestimmung der Art von Problemen die wahrgenommen wurden.

4 Human in the Loop

Die Idee ist neue interaktive Ansätze von EEG als direkte Eingabe in Rendering Algorithmen zu erforschen. Da die wahrgenommene Bildqualität nicht immer ein objektives Maß ist und auch bei Nachfrage üblicherweise weder reflektiert noch genauer erklärt werden, stellt dies eine besondere Herausforderung dar.

Die Arbeit untersucht die Hypothese, dass sowohl die allgemeine visuelle Qualität, wie auch die individuellen Präferenzen zuverlässig und reproduzierbar anhand von EEG Messungen bewertet werden können. Hierfür werden die Gehirnströme von Testpersonen zu gerenderten Bildern gemessen und mittels einer vorher trainierten Support Vector Machine (SVM) wird unmittelbar ein “visual appeal” Wert berechnet. Dieser Wert wird von einem numerischen Optimierungsverfahren verwendet um die Parameter eines Rendering-Algorithmus entsprechend den Vorlieben der Testpersonen zu verändern. Durch diese stetige Optimierung wird die gerenderte Ausgabe auf das ästhetische Optimum der einzelnen Personen angepasst.

Bild 3 zeigt die wesentlichen Bestandteile der EEG gesteuerten Optimierungsschleife. Die Optimierungsschleife besteht aus einer einmaligen Trainingsphase, die benötigt wird um dem Klassifikator den Unterschied zwischen neuronalen Reaktionen zu guten und schlechten Stimuli lernen zu lassen. Die Trainingsphase wird mit einer Menge von Testpersonen durchgeführt, die nicht Teil der Test- und Optimierungsphase sind. Sobald ein Klassifikator antrainiert wurde, kann ein visueller Stimulus, unabhängig vom Inhalt des Bildes, von einer einzigen EEG Messung optimiert werden. Das EEG eines Benutzers zu einem Bild oder Video wird aufgezeichnet und dem SVM-Klassifikator übergeben. Der Klassifikator berechnet den visual appeal Wert auf Basis der EEG Daten, welche die Vorliebe des Benutzers für das Bild oder Video reflektiert. Das verwendete Optimierungsverfahren ist

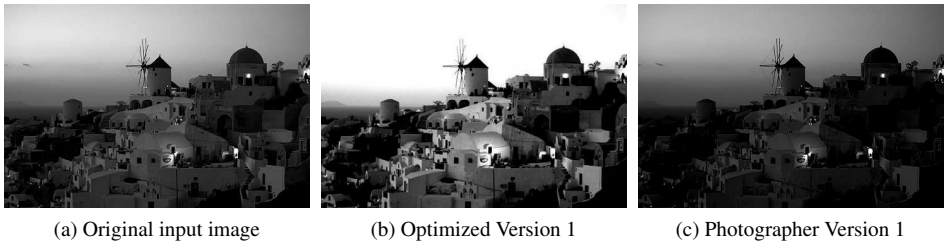


Abb. 4: EEG-optimierte Ergebnisse sind einzigartig für jede Personen und können mit von Fotografen verbesserten Versionen mithalten.

ein Downhill-Simplex-Verfahren, welcher eine vordefinierte Anzahl von Parametern des Rendering-Algorithmus in Abhängigkeit des visual appeal Wertes variiert. Das Bild oder Video wird mit Hilfe dieser Daten neu gerendert und danach wieder dem Benutzer gezeigt. Das Optimierungs-Framework ist auf jeden Rendering-Algorithmus anwendbar, der Bilder oder animierte Sequenzen erzeugt, und dessen Ausgang von gewissen Mengen von Parameterwerten abhängig ist. Der modulare Aufbau ermöglicht den einfachen Austausch von Rendering-Komponenten um die verschiedenen Parameter zu optimieren. Die Schleife wird mit zwei verschiedenen Anwendungsszenarien ausgewertet. Im ersten Szenario, sind statische Bilder von realen Szenen in Bezug auf Sättigung, Helligkeit und Kontrast je nach individuellem Geschmack EEG-optimiert. Im zweiten Szenario sind drei Rendering-Parameter variiert wurden um die allgemeine Attraktivität einer Animationssequenz zu optimieren. In dem ersten Optimierungs-Szenario wurden Sättigung, Helligkeit und Kontrast eines Fotos variiert, um ästhetisch ansprechende Versionen des Originalbildes zu erhalten. Die grundlegenden Impulse für die Trainingsphase bestanden aus 23 zufällig ausgewählten Bilder von der MIT-Adobe FiveK Datenbank. Die Datenbank verfügt über insgesamt sechs Varianten für jedes Bild: das Originalfoto plus fünf verschiedenen, von professionellen Fotografen modifizierte Versionen. Da alle Bilder in der Datenbank bis zu einem gewissen Punkt ästhetisch ansprechend sind, wurden zwei zusätzliche Versionen entweder durch Übersättigung oder Überbelichtung geschaffen.

Die SVM-Trainingsdaten wurden gesammelt, in dem den Teilnehmern ein Originalfoto, zwei der Experten-retuschierten ästhetischen Versionen, sowie die übersättigten und überbelichteten Versionen, gezeigt wurden. Insgesamt waren es fünf Versionen pro Foto und 115 verschiedene Bilder. Die resultierenden EEG-Daten wurden mit Hilfe von komplexen Wavelet-Transformationen und einem Standard SVM verarbeitet. Das trainiert SVM wurde dann zum Testen der Optimierungsschleife verwendet. Das Verfahren wurde mit 15 Benutzern, die nicht in der SVM-Trainingsphase teilgenommen haben, evaluiert. Das Auswertungs-Experiment wurde mit 12 zufällig ausgewählten Fotos von der MIT-Adobe FiveK Datenbank durchgeführt. Jedem Teilnehmer wurde eines dieser Bilder gezeigt. Dieses wurde dann mit Hilfe der single-trial EEG-Daten automatisch verbessert. Die Ergebnisse der mit Hilfe der Optimierung erstellten verschiedenen Version der Originalbilder (Fig.4). Die jeweiligen Versionen waren für jeden Probanden unterschiedlich und waren sehr verschieden von den professionell bearbeiteten Versionen. Nach jeder

Bildoptimierung, hatten wir ausführliche Gespräche mit den Teilnehmern bezüglich der optimierten Bilder. Alle Teilnehmer bevorzugten ihre eigene optimierte Version über das entsprechende Originalbild.

Interessanterweise waren den Probanden nicht interessiert an der Genauigkeit des Bildes in Bezug auf die Farbe oder die Details. Viel mehr basierte ihre Vorliebe für das verbesserte Bild auf dem Gefühl was es ihnen vermittelte. Nachdem das Framework mit 15 Nutzern, die 12 Fotos optimierten, getestet wurde, führten wir eine Wahrnehmungsstudie durch, um unsere Ergebnisse mit den Fotografen-erweiterte Versionen der Bilder zu vergleichen. Die Studie wurde darüber hinaus auch online durchgeführt, um eine größere Stichprobennmenge für die Bewertung der die optimierten Versionen zu erhalten. Es wurden 8 Originalbilder mit den zugehörigen von Fotografen verbesserten Versionen, sowie einige EEG-Optimierte Bilder gezeigt. Die Teilnehmer sollten jeder Bildversion einen Wert von Eins bis Fünf zuweisen. Insgesamt gab es 90 Teilnehmer. Die Ergebnisse aus der Wahrnehmungsstudie unterstützen die Schlussfolgerung, dass die optimierten Versionen mehr bevorzugt werden als die von Fotografen verbesserten Versionen, da die durchschnittliche Punktzahl der EEG-optimierten Version bei ungefähr 3.1 lag, wohingegen die Fotografen-verbesserte Versionen nur einen Wert von 2.8 erreichten.

Im zweiten Optimierungsszenario evaluiert die Arbeit die Leistung der Optimierungsschleife, wenn die visuelle Qualität einer animierten guided image-filtered Raytracing Sequenz optimiert werden soll. Die Filterung von verrauschten Bildern ist eine grundlegende Bildverbesserungsoperation in der Video- und Bildbearbeitung, aber auch in Global Illumination Methoden die auf Monte Carlo Sampling basieren. Der Filter verfügt über drei Parameter, den Filterradius und zwei epsilon-Werte, die gewählt werden müssen, deren optimale Werte abhängig von den Szeneneigenschaften sind. Alle drei Parameter beeinflussen das Rendering-Ergebnis in unterschiedlicher Weise, indem sie sowohl die Oberflächenglätte, wie auch die Menge an Details kontrollieren. Um Trainingsdaten zu erhalten, wurde die populäre Crytek Sponza Szene und die Sibenik Kathedrale graytraced um mehrere Versionen in unterschiedlichen Qualitäten zu erhalten. Um die EEG-Feedbackschleife zu testen wurde in die Szene des Blenderfilms Sintel verwendet. Die optimierten Filterparameter waren Radius, Epsilon (Normale) und Epsilon (Tiefe). Um die Präferenz der EEG-optimierten Videos für ein allgemeines Publikum zu bewerten, wurde eine Wahrnehmungsstudie durchgeführt, die Teilnehmer aufforderte, Videos zu bewerten, die vom Framework und Fachleuten optimiert wurden. Die Daten wurden mit einem zweiseitiger t -Test analysiert, bei dem die Null-Hypothese für den t -Test war, dass die optimierten und manuellen Versionen derselben Population angehören. Die t -Test-Wahrscheinlichkeit für die Null-Hypothese $P(H_0)$ zeigte, dass die Ratings der optimierten Versionen nicht signifikant anders waren als die der manuellen Version (Nullhypothese wurde nicht abgelehnt).

Die Ergebnisse zeigen, dass die EEG-optimierten Video-Sequenzen genauso visuell ansprechend sind wie Experten optimierte Sequenzen. Darüber hinazs zeigen die Ergebnisse, dass es Möglich ist die Wahrnehmungsqualität eines Video von einer Single-Trial EEG Messung zu bestimmen, ohne das ein Referenzvideo benötigt wird. Dies ist besonders von Bedeutung, da Videos dafür bekannt sind extrem schwer mit Hilfe von EEG Techniken analysierbar zu sein.

5 Conclusion

Es werden experimentelle Daten und eine Analyse vorgestellt, die zeigt dass sich die covert (implizite) and overt (explizite) output des menschlichen Wahrnehmungssystems unterscheiden und dass dieser Unterschied in einigen Fällen sehr stark ist. Die Arbeit zeigt außerdem dass das Gehirn sehr unterschiedlich, nicht nur auf verschiedene Arten von Artefakten reagiert, sondern im Besonderen auf Artefakte die mit Bewegung zusammenhängen. Die Experimente zeigen außerdem, dass es möglich ist Artefakte nach Wahrnehmung zu kategorisieren. Dies erlaubt Einsichten in die Wahrnehmung von Videos die zuvor nicht modelliert worden sind. Diese Arbeit ermöglicht außerdem eine Verkürzung von Renderzeiten durch die Entfernung von Berechnungen für Bildmerkmale die keine starke Reaktion des Gehirns auslösen. Diese Arbeit stellt eine neue Methode für single-trial Klassifizierung von typischen IBR Artefakten. Es ist gezeigt worden, dass Wavelets und eine SVM ein effektiver Weg sind um das Problem des niedrigen Signal-Rausch-Verhältnisses in EEG Daten zu umgehen. Eine neue Methodik wird vorgestellt um zwischen verschiedenen Arten von Artefakten zu unterscheiden die in Videostimuli auftreten, anhand eines single-trial von neuronalen Daten. Diese Arbeit stellt auch einen neuen Ansatz für die Optimierung von Rendering-Parametern vor, die auf single-trial EEG Messungen besteht. Es wird außerdem die Benutzung neuronaler Daten für die Bewertung des visual appeal von gerendereten Bildern und Animationen eingeführt.

Die vorgestellten Projekte haben gezeigt dass es nicht nur möglich ist die Qualität eines Bildes oder Videos mittels EEG zu bestimmen, sondern dass es auch möglich ist diese Qualität in einer direkten Feedback-Schleife zu verbessern, bis eine optimale Lösung erreicht wurde. Neuronale Daten stellen ein praktikables Maß der Ästhetik von Bildern und Videos und für Biofeedback für Grafikanwendungen dar; insbesondere um die Wahrnehmung von gerendereten Bildsequenzen zu untersuchen.

Die Kontributionen dieser Arbeit ebnen den Weg für neue, spannende Kombinationen von Computergrafik und EEG. Es existieren viele Grafikprobleme, die von ihrer Nutzung profitieren können. Zum Beispiel, ist es ideal geeignet für die Qualitätsbewertung von 3D Bildern und Videos.

Literatur

[Mu15] Mustafa, Maryam: ElectroEncephaloGraphics: A Novel Modality For Graphics Research: Dissertation. BoD-Books on Demand, 2015.

Maryam Mustafa ist Postdoc am Institut für Computergraphik an der TU Braunschweig. Ihre Forschungsinteressen beinhalten Mensch-Maschine Interaktion, Wahrnehmung und Erkennung in der Computergraphik. Mustafa erhielt Ihren MIng in Computer Science von der Cornell Universität. Dort war sie auch ein Fulbright Scholar. Erreichbar ist sie unter: mustafa@cg.cs.tu-bs.de