Informationen zu den Arbeiten der Preisträgerinnen des Wettbewerbs 2021

Weiteres Pressematerial finden Sie unter
http://www.ard-zdf-foerderpreis.de/presse/

Bei weitergehenden Fragen oder zur Kontaktvermittlung wenden Sie sich bitte an:

ARD/ZDF Förderpreis »Frauen + Medientechnologie«
Monika Gerber
Wallensteinstr. 121
D-90431 Nürnberg
Telefon + 49 911 9619 495
E-mail: info@ard-zdf-foerderpreis.de

ARD/ZDF FÖRDERPREIS
FRAUEN + MEDIENTECHNOLOGIE

1. Preis 2021

Jennifer Rasch

„Signal Adaptive Methods to Optimize Prediction Signals in Video Coding“

Dissertation

Technische Universität Berlin

Videokompression der Zukunft

Jennifer Rasch

Videos erobern den Markt, Plattformen wie Youtube oder Netflix erfreuen sich immer größerer Beliebtheit. Aber welche technischen Innovationen stecken hinter dem Boom? Neueste Technologien in der Videokompression ermöglichen die weltweite Verbreitung von Videos.

Eine Studie von Cisco, einer der größten Hersteller von Netzwerkkomponenten, schätzt, dass Videodaten seit diesem Jahr etwa 80 % des globalen, privat genutzten Internetverkehrs ausmachen. Laut dem US-amerikanischen Wirtschaftsmagazin Forbes generiert allein das soziale Netzwerk Facebook durchschnittlich 8 Milliarden Videoansichten pro Tag. Die Videoplattform Youtube dokumentiert eine jährliche Verdopplung des Konsums von mobilen Videos seit 2017. Aber auch abseits der Entertainment- und Social-Media-Industrie haben Videos Hochkonjunktur. Die US amerikanische Eliteuniversität MIT beispielsweise macht seit 2002 sukzessive seine Lehreinheiten übers Internet zugänglich – u.a. durch komplette per Video aufgezeichnete Kurse. Auch in der Medizin kommen Videos immer mehr zum Einsatz, z.B. bei Videoübertragung von Operationen. Angesichts dieser enormen Mengen an Videodaten steigt der Bedarf an effizienten Videokompressionsmethoden, d.h. Methoden, die Videodaten verkleinern, ohne große Qualitätseinbußen zu verzeichnen.

Typischerweise setzt sich eine solche Methodik aus zwei Teilen zusammen, dem Enkodierer (engl. „Encoder“) und dem Dekodierer (engl. „Decoder“). Hieraus leitet sich der Begriff „Codec“ ab. Die rohen Videodaten werden zunächst auf Senderseite enkodiert. Als Sender kann man sich hierbei Film- und Fernsehproduktionen, Videokonferenzen, Sicherheitskameras oder Streamingplattformen vorstellen. Der Prozess des Enkodierens verwandelt das Video in einen Bitstrom, der dann über einen Übertragungskanal gesendet werden kann. Dabei dienen Satelliten, Mobilfunknetzwerke, Kabel, Datenträger oder das Internet als typische Übertragungskanäle. Die Kapazität dieser Übertragungskanäle variiert je nach Typ und Anwendung stark, insofern muss auch die Kompressionsleistung entsprechend angepasst werden. Wenn man jedoch beispielsweise ein HD Video von etwa 600 MB pro Sekunde enkodieren möchte und der Übertragungskanal nur eine Kapazität von etwa 10 MB pro Sekunde hat, entspricht das einer benötigten Datenreduktion auf etwa 1.7% der ursprünglichen Größe. Hierbei sind effiziente Kompressionsmethoden essentiell. Auf der Empfängerseite muss der empfangene Bitstrom dekodiert und das Video rekonstruiert werden, um es abspielen zu können. Als Empfänger dienen beispielsweise viele in Privathaushalten verwendete Geräte wie etwa Computer, Fernseher, Tablets oder Smartphones.

Klassische Videokompressionsmethoden basieren auf zwei Prinzipien der Informationstheorie: der sogenannten Redundanzreduktion und der Irrelevanzreduktion. Die Redundanzreduktion verwendet bereits vorhandene Daten zur Reduktion der zu übertragenden Datenmenge. Beispielsweise wird die zeitliche Korrelation von aufeinander folgenden Bildern genutzt um lediglich die Differenz kodieren zu müssen.
Moderne Videokompressionsmethoden verwenden eine Verfeinerung dieser Technik: Das Video wird in Bilder und diese in Blöcke unterteilt. Dann wird die zu übertragende Differenz pro Block minimiert indem der aktuellen Bildausschnitt aus den vorhandenen Daten so gut wie möglich geschätzt wird (im Fachjargon „prädizieren“ genannt). Dabei werden für die zeitliche Prädiktion („Inter-Prädiktion“) die vorhergehenden, d.h. bereits dekodierten Bilder als Referenz verwendet. Die entsprechende Auswahl wird dann im Bitstrom signalisiert, d.h. es werden genaue Angaben enkodiert, wo sich der als Schätzung gewählte Bildausschnitt befindet. Diese Technik erspart große

Mengen an Daten, induziert aber gleichzeitig eine zeitliche Abhängigkeit im Bitstrom: fehlen vorhergehende Bilder, kann das Video nicht mehr fehlerfrei rekonstruiert werden. In der Praxis, beispielsweise bei Live Übertragungen im Fernsehen, ist es jedoch von großer Bedeutung, dass das Video nach wenigen Sekunden verfügbar ist. Dafür muss gewährleistet werden, dass es innerhalb einer Übertragung in regelmäßigen zeitlichen Abständen Punkte gibt, ab denen man einschalten kann und ab denen das Video rekonstruierbar ist, auch wenn die vorhergehenden Daten dem Empfänger fehlen. Diese Eigenschaft wird in Fachkreisen auch „random access“ genannt, in diesem Zusammenhang lässt sich das als „wahlfreier Zugriff“ übersetzen. Daher werden in modernen Videokompressionsmethoden regelmäßig sogenannte „Intra-Bilder“ gesendet, die unabhängig von zeitlich vorhergehenden Daten sind. Zur Minimierung der zu übertragenden Differenz werden hier ausschließlich bereits dekodierte Bildausschnitte zur Schätzung des aktuellen Blocks verwendet, die sich innerhalb des aktuellen Bildes befinden.

Die Güte von Prädiktionen spielt bei der Effizienz von Video Codecs eine große Rolle. Störfaktoren, die generell bei diesen Prädiktionstechniken oft auftreten, sind beispielsweise Rauschen oder Artefakte, die aus anderen Teilen des Bildes stammen und sich in der Prädiktion wiederfinden. Jennifer Rasch arbeitete in ihrer Dissertation „Signal Adaptive Methods to Optimize Prediction Signals in Video Coding“ (dt. „Signaladaptive Methoden zur Optimierung von Prädiktionssignalen“) an sogenannten Prädiktionsfiltern, die solche Artefakte entfernen. Die untersuchten Filter basieren auf dem physikalischen Prinzip der Wärmeleitungsgleichung, das den Zusammenhang zwischen der zeitlichen und räumlichen Änderung der Temperatur in einem Körper beschreibt. Problem bei dieser Art von Filtern ist, dass im Bild vorhandene Kanten und Strukturen verschwinden, weil sie zu stark geglättet werden. Daher entwickelte Jennifer Rasch in ihrer Arbeit eine neuartige Filtermethode für Prädiktionen, die die unterliegende Bildstruktur mit einbezieht. Durch diese Signaladaptivität können Störungen herausgefiltert werden, während vorhandene Kanten erhalten bleiben.

Um die praktische Anwendbarkeit zu gewährleisten, ist ein weiterer essentieller Punkt bei neuen Kodierwerkzeugen deren Komplexität. Daher wird in der Dissertation von Jennifer Rasch ein starker Fokus auf die Komplexitätsreduktion der neuartigen Filtermethode gelegt. In der Arbeit wird gezeigt wie die initiale Komplexität um 70% verringert werden kann, während die Kodiereffizienz des signaladaptiven Prädiktionsfilters weitgehend erhalten bleibt.

Bei den heutigen hohen Anforderungen an moderne Kompressionsverfahren muss auf verlustbehaftete (engl. „lossy“) Kompressionstechniken zurückgegriffen werden. Hierbei kommt das Prinzip der Irrelevanzreduktion zum Einsatz, d.h. es werden Bildinformationen entfernt, die durch das menschliche Auge nicht oder nur wenig wahrnehmbar sind. Ein typisches Beispiel hierfür ist die Reduktion von Farbinformation im Bild: da die menschliche Wahrnehmung von Farbunterschieden sehr viel geringer ist als die von Helligkeitsunterschieden, wird das Bild in Helligkeits- und Farbkanäle unterteilt. Dann kann die Auflösung der Farbkanäle stark reduziert werden, ohne große, subjektiv wahrgenommene Qualitätseinbußen befürchten zu müssen.
Im Zuge der verlustbehafteten Kompression wird dies genutzt, um Details aus dem Bild zu entfernen, deren Verlust für das menschliche Auge im Idealfall nicht oder kaum wahrnehmbar ist, jedoch zu einem großen Maß an Dateneinsparung beiträgt.

Um garantieren zu können, dass Endgeräte (also die Empfängerseite) den Bitstrom lesen und das Video rekonstruieren können, werden internationale Standards für Video Codecs benötigt. Videokompressionsstandards sind so konstruiert, dass sie maximal viel Spielraum für eine individuelle Anpassung an Geräte und Anwendung bieten, während sie gleichzeitig Kompatibilität garantieren. Bei Videokompressionsstandards wird daher ausschließlich der Bitstrom spezifiziert (um dessen Lesbarkeit zu gewährleisten), die Implementierung des Enkodierers wird bewusst offengelassen. Die H.26x Serie von Videokompressionsstandards existiert seit über 25 Jahren und

ist bis heute extrem erfolgreich. Ihre weltweite Verbreitung hat enorm zu der wachsenden Popularität von Videotechnologie beigetragen. Der im Jahre 2003 fertig gestellte Standard H.264/AVC („Advanced Video Coding“) ist bis heute einer der global am weitesten verbreiteten Standards: Jeder HDTV-Empfänger, jeder Blu-Ray-Player und die meisten Internetvideos verwenden heutzutage den H.264/AVC Standard. Schätzungen zufolge ist er in über einer Milliarde Geräten eingebaut. Seit 2013 ist sein direkter Nachfolger H.265/HEVC („High Efficiency Video Coding“) auf dem Markt, der im Vergleich bei gleichbleibender Bildqualität die benötigte Datenrate halbiert. Bis vor kurzem arbeiteten die beteiligten Firmen, Universitäten und Forschungsinstitute an dem Videokompressionsstandard der Zukunft H.266/VVC („Versatile Video Coding“), der im Juli 2020 fertig gestellt wurde. Auch hier lautete das selbstgesetzte Ziel, die benötigte Datenrate bei konstanter Bildqualität zu halbieren.

Der in Jennifer Raschs Dissertation neu entwickelte, signaladaptive Prädiktionsfilter wurde vom Fraunhofer Heinrich-Hertz-Institut im Zuge der Standardisierung des neuen Videokompressionsstandards H.266/VVC vorgestellt. Das patentierte Filtertool ist in der Lage, die Datenrate eines UHD-Videos bei gleichbleibender Bildqualität um bis zu 1 MB pro Sekunde zu verringern. Das entspricht pro Stunde einer Ersparnis von der Größe von etwa 1.5 DVDs. Angesichts des global zunehmenden, enormen Videodatenverkehrs wohnt dem Tool somit ein signifikantes Verbesserungspotenzial inne.

ARD/ZDF FÖRDERPREIS
FRAUEN + MEDIENTECHNOLOGIE

2. Preis 2021

Franziska Mertl

„Automatisierte Trainingsdatengenerierung zur Gesichtserkennung“

Masterarbeit

Hochschule für angewandte Wissenschaften München

Zusammenfassung

In einer digital vernetzten Welt gilt es für Medienhäuser, sich in einem immer breiter aufgestellten Medienangebot zu behaupten. Dabei ist es entscheidend, Inhalte auf schnellstem Wege und möglichst individuell, angepasst an die Bedürfnisse der Nutzer, zugänglich zu machen. Um allerdings das vielfältige Angebot für die unterschiedlichen Plattformen bestmöglich bedienen zu können, braucht es intelligente Werkzeuge, die durch die Übernahme automatisierbarer komplexer Aufgaben wertvollen kreativen Freiraum schaffen und das Serviceportfolio für Mitarbeiter und Medienendnutzer erweitern.

Für eine journalistisch qualitativ hochwertige Berichterstattung muss es Redakteuren möglich sein, neben externen Quellen auf gut gepflegtes internes Archivmaterial zugreifen zu können. Innerhalb der öffentlich-rechtlichen Rundfunkanstalten werden hierzu tagtäglich die hausintern oder im Verbund erzeugten Audio- und Videoinhalte durch Mitarbeiter des Archivs formal erfasst und inhaltlich erschlossen. Erst die genauen und umfassenden Metadaten ermöglichen den Redakteuren eine effiziente Recherche nach aktuell relevanten Inhalten. Folglich spielt das Archiv für einen effektiven und reibungslosen Produktionsablauf eine entscheidende Rolle und wirkt damit indirekt, aber maßgeblich an der Außenwirkung des Unternehmens mit.

Allerdings bringt die nahezu pausenlos und mengenmäßig ansteigende Generierung von Medieninhalten bisher bewährte interne Arbeitsprozesse an Grenzen. Wenn die Menge an Inhalten es nicht mehr zulässt, dass die formale Erfassung, inhaltliche Erschließung, Dokumentation und Bestandspflege zeitnah erfolgen kann, sind die Inhalte bis zu einer Bearbeitung nur unter erschwerten Bedingungen auffindbar und eventuell sogar veraltet. Das hätte zur Folge, dass ein Teil des wichtigsten Guts eines Medienunternehmens auf der Strecke bliebe und investierte Arbeiten sich nicht auszahlen würden. Daher gilt es, Archiv- und folglich auch Redaktionsmitarbeiter durch das Abnehmen automatisierbarer Arbeiten zu unterstützen und zu entlasten. Darüber hinaus treibt den Archivfachbereich des Bayerischen Rundfunks (BR) die Motivation an, das Dienstleistungsspektrum gegenüber den Redaktionen stetig zu verbessern und zu erweitern. Es ist angedacht, neben den bisher geführten Metadaten der manuellen Erschließung Zusatzinformationen wie die Filmmusikerkennung in Videos, Emotionen zu Personen, Sprach- zu Schriftgenerierung und viele weitere anbieten zu können. Das hebt die Wertigkeit des Archivmaterials und bietet überdies womöglich ungeahnte Möglichkeiten.

Die maschinelle Bearbeitung von bestimmten Archivarbeiten wie der Verschlagwortung, Untertitelung und Metadatenpflege setzt allerdings eine gewisse künstliche Intelligenz voraus, da sich die vorliegenden komplexen Zusammenhänge nur sehr schwer über eine feste Struktur und durch direkte Programmierung erschließen lassen. In diesem Zusammenhang können unter dem Sammelbegriff Cognitive Services geführte Algorithmen der Schlüssel zum Erfolg sein, wenn es unter anderem darum geht, riesige Mengen an Medieninhalten möglichst rasch und detailliert metadatentechnisch aufzubereiten.

Um im großen Feld der Cognitive Services Fuß fassen zu können, bietet sich die Thematik der Gesichtserkennung in Videoinhalte als Einstieg an, da diese forschungstechnisch gut vorangetrieben wird und eine gute Basis für weitere Dienste zur Analyse von personenspezifischer Mimik oder Bewegungsabläufen geschaffen wird. Zudem können die den Gesichtern zugehörigen Namen Archivarbeiten wie der Verschlagwortung zugute-

kommen, da Persönlichkeiten in einem gewissen öffentlichen Interesse beliebte und entscheidende Schlagworte für die redaktionelle Recherche darstellen.

Mit den hierbei zum Einsatz kommenden Algorithmen des maschinellen Lernens beziehungsweise des *Deep Learnings* können Daten maschinell inhaltlich verstanden und interpretiert werden. Algorithmen dieser Art lernen über Trainingsvorlagen allgemeingültige Gesetzmäßigkeiten und Muster, die sich wiederum auf nicht angelernte Eingabedaten der gleichen Aufgabenstellung übertragen lassen. Für die Gesichtserkennung werden mittlerweile eine Vielzahl an sofort einsatzfähigen Softwarelösungen von einer mindestens genauso großen Anzahl an Entwicklungsfirmen angepriesen. Dazu ergaben vorab gesammelte Erfahrungen, dass solcherlei Fertigprodukte bei Anwendung auf BR-Videomaterial zu hohen Fehlerquoten führen, da das Gesichtserkennungssystem mit Trainingsmaterial, das nicht dem regionalen Anspruch des BRs hinsichtlich lokaler Prominenz, historischem Materialinhalt und deutscher Sprache genügt, befüttert wurde. Diese Tatsache bedingt, dass ein für den BR anwendungsspezifischer, aufbereiteter und voraussichtlich sehr umfangreicher Trainingsdatensatz bereitgestellt und zeitlicher Aufwand für die Trainingsphase eingeplant werden muss. Eine bequeme und zeiteinsparende Möglichkeit der Datensatzbeschaffung wäre auf extern zugeliefertes Trainingsmaterial zurückzugreifen, aber auch hier entspricht das Angebot überwiegend nicht den Anforderungen eines Medienhauses mit Fokus auf Regionalität. Um dennoch den notwendigen und aktuell zuhaltenden Datenpool bereitstellen zu können, gilt es, aus dem Videobestand des BRs geeignetes Trainingsmaterial herauszufiltern. Dabei wäre es wünschenswert, den Trainingsdatengenerierungsprozess weitestgehend zu automatisieren, um die damit verbundenen zeitlichen und personellen Ressourcen fortlaufend klein zu halten.

Für eine derartige Automatisierung wäre ein möglicher Ansatz, Zusatzinformationen, mit denen das Videomaterial durch vorangegangene Produktionsschritte angereichert wurde, zu nutzen, um geeignete Referenzgesichter zu lokalisieren und den Trainingsvorgang zu erleichtern. Hierfür bieten sich Bauchbinden in bereits publiziertem Material an. Diese werden im Allgemeinen nur in Videosequenzen eingeblendet, welche eine sprechende, der Kamera überwiegend zugewandte und gut im Bilde stehende Person zeigen und zu dem dargestellten Gesicht auch den Namen mitliefern.

Im Rahmen der Masterarbeit „Automatisierte Trainingsdatengenerierung zur Gesichtserkennung“ von Franziska Mertl wird geklärt, ob über die Nutzung der in Bauchbinden geführten Namen verknüpft mit dem zeitgleich eingeblendeten Gesicht eine automatisierte Trainingsdatengenerierung gelingen kann. Dies wurde unter Verwendung von Videomaterial des Bayerischen Rundfunks (BR) geprüft, wobei eine Materialeingrenzung vorzunehmen war, um nicht den Rahmen zu sprengen. Hierzu gehörte auch die Wahl eines geeigneten Sendeformats, das im Durchschnitt ausreichend viele Baubinden einsetzt und dessen geführte Personengruppen überwiegend eine gewisse Relevanz in der Öffentlichkeit darstellen. Um die essentielle Frage zu klären, ob die über den Bauchbinden-Ansatz gewonnenen Gesichtsbilder sich für das Training eines Gesichtserkennungssystems eignen, wurde innerhalb einer Teststellung, in Zusammenarbeit mit einer ausgewählten Softwarefirma, geeignete Messungen festgelegt und durchgeführt. Zuvor galt es allerdings Möglichkeiten zur automatisierten Materialbereitstellung für den Analysedienst zu evaluieren und die Durchführbarkeit in einem Testintegrationsszenario in die BR-Infrastruktur verifizieren. Der Teststellung in der Form einer Machbarkeitsstudie wurde ein Theoriekapitel, das sich mit den Grundlagen der Mustererkennung, genauer gesagt mit den Algorithmen des maschinellen Lernens und *Deep Learnings* näher beschäftigt, vorangestellt. In diesem Teil der Arbeit wird

auch die Thematik des Trainingsprozesses genauer beleuchtet und mit Beispielen veranschaulicht. Nach der Klärung der grundlegenden Begrifflichkeiten ist der Übergang zur Gesichtserkennung als Beispiel eines Mustererkennungssystems gegeben. Der Leser erlangt dabei einen Eindruck über die Vielzahl an mathematischen Verfahren, die in der Regel nur Unteraufgaben innerhalb des Gesamtsystems dienen und in ihrer Raffinesse und deren Anwendung den Konkurrenzkampf zwischen Entwicklern befeuern, um die Analysegeschwindigkeit und -genauigkeit gleichermaßen zu verbessern.

Die automatisierte Videomaterialbereitstellung von Seiten des BRs gegenüber dem Analysedienst gelang mittels Programmier- und Datenbankskripten, durch welche die relevanten Sende- und Metadaten aus Sendeplanungs- und Archivsystem zusammengeführt und eine Materialbestellung ausgelöst wurde. Anschließend wurden mittels Texterkennung Personennamen aus Bauchbinden ausgelesen und zusammen mit den zeitgleich detektierten Gesichtern als Trainingsbilder abgelegt. Das dabei entstandene Set an Bildern ergab taugliche und zu sehr guter Klassifizierungsleistung führende Trainingsdaten. So konnte über das Antrainieren des Gesichtsklassifizierers auf Basis des generierten Trainingsmaterials eine Genauigkeit von $99{,}32\%$ erzielt werden oder mit anderen Worten, in 127 unbekannten Testvideos wurden von 1387 erkannten Personen nur 8 Gesichter einem falschen Personennamen zugeordnet. Dabei wurde sowohl das Quellmaterial für die Trainingsdatengenerierung als auch das Testmaterial für die finale Analyse auf Videomaterial des Sendeformats „Rundschau“ begrenzt. Allerdings sind Nacharbeiten hinsichtlich der Text- und Bauchbindenerkennung notwendig, um die sich ergebene Fehlerrate zu senken und somit weniger potentielles Trainingsmaterial zu verwerfen.

Die gewonnenen Erkenntnisse hinsichtlich des zeitlichen und qualitativen Mehrwerts und der Machbarkeit einer automatisierten Trainingsdatengenerierung erwiesen sich als so vielversprechend, dass der Weg der Automatisierung mittels maschineller Lernalgorithmen beim BR intensiv weiterverfolgt wird. Ein erster Schritt wird dabei sein das Gesichts-erkennungssystem soweit produktiv zu nehmen, dass die Analyseergebnisse im Media Asset Management (MAM) für die Recherche sichtbar gemacht und von dort auf die Produktionssysteme ausgeweitet werden.

ARD/ZDF FÖRDERPREIS
FRAUEN + MEDIENTECHNOLOGIE

3. Preis 2021

Maike Richter

„Lautheitsmessung von objektbasierten Audioszenen“

Bachelorarbeit

Hochschule der Medien Stuttgart

Titel der Bachelorarbeit: "Lautheitsmessung von objektbasierten Audioszenen"

Autorin: Maike Richter

Zusammenfassung

Seit einigen Jahren werden Forschungen im Bereich *Next Generation Audio (NGA)* angestellt. *NGA* soll unter anderem immersives Audio in hoher Qualität bieten und Interaktion mit dem Hörer ermöglichen. Hierzu werden neue Ansätze zur Produktion und Wiedergabe von Audioinhalten untersucht. Ein wichtiger Teil dieser Forschungen ist die *objektbasierte* Audiorepräsentation. Diese ermöglicht unter anderem die Wiedergabe von nur einer Anfertigung einer Tonmischung auf verschiedenen Lautsprecheranordnungen.
2018 entwickelte die European Broadcasting Union einen offenen Standard zum Rendering von objektbasierten Audioszenen, den sogenannten *EBU ADM Renderer (EAR)*. Ziel war es dabei, einen einheitlichen Renderer für die Produktion, Qualitätssicherug und die Evaluation bereit zu stellen.

Allerdings ergeben sich aus dem objektbasierten Ansatz auch weitere Herausforderungen, unter anderem die Frage nach der Lautheitsmessung von objektbasierten Audioszenen. *Lautheit* spielt vor allem in Broadcastingumgebungen ein wichtige Rolle, da wahrnehmbare Sprünge der Lautstärke zwischen zwei Sendeinhalten zu Zuschauerbeschwerden führen. Richtlinien zur Lautheitsmessung bestehen bislang aber nur für kanalbasiertes Audio.
Das Ziel dieser Thesis ist daher die Entwicklung eines Lautheitsmessverfahrens für objektbasierte Audioszenen. Das Verfahren soll einen Lautheitswert pro Audioszene generieren, der für alle Wiedergabesysteme gilt. Außerdem soll es eine möglichst hohe Kompatibilität zum kanalbasierten Lautheitsmessverfahren nach Rec. ITU BS.1770-4 aufweisen.
Im Rahmen dieser Arbeit werden bestehende Ansätze zur objektbasierten Lautheitsmessung aufgezeigt und im Kontext der Entwicklung des *EBU ADM Renderers* weitere Untersuchungen angestellt. Ein Schwerpunkt liegt hier auf der Untersuchung der Objektparameter in Bezug auf deren eventuellen Einfluss auf die Lautheit einer Audioszene. Auf Basis der Untersuchungsergebnisse wird daraufhin ein Lautheitsmessverfahren für objektbasierte Audioszenen entwickelt und implementiert. Zuletzt erfolgt eine Evaluation des entwickelten Verfahrens, welche messtechnische Untersuchungen und einen informellen, subjektiven Hörtest beinhaltet.

Zur Untersuchung des Einflusses einzelner Objektparameter auf die Lautheit einer Audioszene erfolgt zunächst die Erstellung objektbasierter Audioszenen mit Testsignalen. Für diese Audioszenen wurden Kanalrepräsentationen für alle unterstützten Lautsprecheranordnungen gerendert und deren Lautheit nach Rec. ITU 1770-4 ermittelt.
Anschließend werden den Testsignalen unterschiedliche Parameterwerte zugeordnet und die jeweils kalkulierten Lautheitswerte miteinander verglichen.
Die Untersuchungen ergeben, dass lediglich die Objektparameter zu Position und Gain eines Audioobjekts relevante Einflüsse auf die Lautheit einer objektbasierten Audioszene haben und damit bei der Entwicklung eines Lautheitsmessverfahrens für objektbasierte Audioszenen berücksichtigt werden müssen.

Die Implementierung des Lautheitsmessverfahrens erfolgt in Form eines Python Skriptes.
Zur Lautheitsberechnung einer objektbasierten Audioszene werden die Audiosamples der einzelnen Audioobjekte, sowie die beschreibenden Parameter aus den Metadaten eingelesen. Die Audiosamples werden in Abhängigkeit der zugehörigen *gain*-Werte bearbeitet. Dann durchlaufen die Audiosignale die einzelnen Module zur Lautheitsberechnung analog zur kanalbasierten Messung: *Filterung*, *Mittelwertbildung*, *positionsabhängige Gewichtung*, *Summation* und *Gating*.
Im Gewichtungsmodul werden die Positionsparameter der entsprechenden Audioobjekte aus den Metadaten herangezogen.
Zur messtechnischen Evaluierung wird das entwickelte Messverfahren zur Berechnung der Lautheitswerte unterschiedlicher Beispielszenen verwendet. Außerdem werden die Szenen auf alle Lautsprecheranordnungen gerendert und die Lautheitswerte der gerenderten Szenen berechnet. Anschließend werden Verlgeiche zwischen dem Lautheitswert der ADM-Szene und den Lautheitswerten der zugehörigen gerenderten Szenen angestellt.
Ergänzend zur messtechnischen Evaluation wird ein informeller Hörtest durchgeführt. Die Prüfung des entwickelten Lautheitsmessverfahrens ergibt schließlich, dass das Verfahren insgesamt zufriedenstellende Ergebnisse liefert.
Allerdings wird eine Problematik bezüglich des Renderings auf 2.0-Lautsprecheranordnungen aufgedeckt. Der EBU ADM Renderer arbeitet hierbei im Gegensatz zum Rendering auf alle anderen Lautsprecheranordnungen nicht lautheitserhaltend. Der generische Lautheitswert einer objektbasierten Audioszene kann deshalb nicht mit der Lautheit beim Rendering auf eine 2.0-Lautsprecheranordnung übereinstimmen. Zur Lösung dieses Problems empfiehlt die Autorin eine entsprechende Anpassung im Stereo-Panner des EBU ADM Renderers. Sobald diese Anpassung vorgenommen ist, gelten die mit dem generischen Lautheitsmessverfahren berechneten Lautheitswerte einer objektbasierten Audioszene auch bei der Wiedergabe auf einem 2.0 Stereo-Lautsprechersystem.

Das im Rahmen dieser Bachelorarbeit entwickelten Lautheitsmessverfahren ermöglicht die Generierung zuverlässiger Richtwerte zur Lautheitskontrolle von objektbasierten Audioszenen.
Zudem können auf Grundlage des entwickelten Verfahrens Programmerweiterungen für Digital Work Stations (DAWs) zur Lautheitsmessung objektbasierter Audioszenen erstellt werden. Des Weiteren besteht die Möglichkeit, den Lautheitswert einer Audioszene in den Metadaten der ADM-Szene einzuspeichern. Damit ist die Lautheitsberechnung nur einmal im Produktionsablauf notwendig.
Bei der Einbettung einer objektbasierten Audioszene in ein Sendeprogramm, kann der Lautheitswert der Szene direkt aus den Metadaten gelesen werden und bei Bedarf können entsprechende Anpassungen der Lautheit vorgenommen werden.