

TECHNISCHE UNIVERSITÄT
CHEMNITZ

Fakultät für Informatik

CSR-15-01

ValidAX - Validierung der Frameworks AMOPA und XTRIEVAL

Arne Berger · Maximilian Eibl · Stephan Heinich · Robert Herms
Stefan Kahl · Jens Kürsten · Albrecht Kurze · Robert Manthey
Markus Rickert · Marc Ritter

Januar 2015

Chemnitzer Informatik-Berichte

ValidAX - Validierung der Frameworks AMOPA und XTRIEVAL

Vorhaben im Rahmen des Programms Validierung des Innovationspotenzials wissenschaftlicher Forschung - VIP

Schlussbericht



Autoren: Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Herms, Stefan Kahl, Jens Kürsten, Albrecht Kurze, Robert Manthey, Markus Rickert, Marc Ritter

Technische Universität Chemnitz

Fakultät Informatik

Professur Medieninformatik

Straße der Nationen 62

09107 Chemnitz

Zuwendungsempfänger

Technische Universität Chemnitz

Projektträger

VDI/VDE Innovation + Technik GmbH

Förderkennzeichen

03V0058 (ehemals 16V0058) – VIP0044

Vorhabenbezeichnung

ValidAX - Validierung der Frameworks AMOPA und XTRIEVAL

Laufzeit des Vorhabens

01.07.2011 bis 31.06.2014

Berichtszeitraum

01.07.2011 bis 31.06.2014

Kontakt

Prof. Dr. Maximilian Eibl

Professur Medieninformatik

Technische Universität Chemnitz

Straße der Nationen 62

09111 Chemnitz

Email: maximilian.eibl@informatik.tu-chemnitz.de

Web: <http://www.tu-chemnitz.de/cs/mi>

Haftungsausschluss

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Inhaltsverzeichnis

I. Kurzdarstellung.....	2
1. Aufgabenstellung.....	2
2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde	3
3. Planung und Ablauf des Vorhabens.....	4
4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde.....	6
5. Zusammenarbeit mit anderen Stellen	7
II. Eingehende Darstellung	7
1. Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen.....	7
AB 1: Flexible Mediatranskodierung für den Transport audiovisueller Medien	7
AB 2: Archivierungsstraße.....	13
AB 3: Workflowintegration	18
AB 4: Annotationsunterstützung	27
AB 5: Bilderkennung.....	33
AB 6: Web-Services	40
AB 7: Parallelverarbeitung	41
2. Wichtigste Positionen des zahlenmäßigen Nachweises.....	43
3. Notwendigkeit und Angemessenheit der geleisteten Arbeit	45
4. Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans	46
5. Während der Durchführung des Vorhabens dem ZE bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen	48
6. Erfolgte oder geplante Veröffentlichungen der Ergebnisse.....	48

I. Kurzdarstellung

1. Aufgabenstellung

Die Medienbranche ist seit Jahren ein Wachstumsmarkt in Bezug auf Gesamtvolumen und Vielfaltigkeit, und wird dies über die nächsten Jahre hinweg auch bleiben. Auslöser dafür ist die Digitalisierung, welche zunächst über das Internet, inzwischen aber auch über Digitalfernsehen und ab 2015 über Digitalradio den klassischen Umgang mit Medien revolutioniert. Dies führt zu folgenden Trends:

- Trend 1: Die Anzahl der Produzenten von Video- und Audiomaterial steigt mit der Digitalisierung: Die Anzahl der bundesweit gesendeten privaten TV-Kanäle hat sich beispielsweise in den Jahren 2004-2009 verdoppelt. Daneben entstehen im Internet verschiedenste zusätzliche Angebote. Gründe für diesen Zuwachs sind zum einen die gesunkenen Produktions- und Vertriebskosten beispielsweise für Webangebote im Vergleich zum klassischen Fernsehen. Zum anderen sind aber auch die herkömmlichen Vertriebswege wie Antennen durch die Digitalisierung nun in der Lage mehr Programme zu übermitteln, da die benötigte Bandbreite pro TV- oder Radiokanal deutlich kleiner ist als bei der analogen Übertragung (Stichwort: Digitale Dividende).
- Trend 2: Die Archivierung und Suche nach Video- und Audiomaterial wird für die Produzenten zunehmend zum Problem. Während beispielsweise die öffentlich-rechtlichen Fernsehsender dank verlässlicher Gebühreneinnahmen gut dokumentierende Archive unterhalten können, haben die privaten Anbieter hier große Schwierigkeiten. Ein Beispiel aus dem öffentlich-rechtlichen Bereich: Der in Hamburg ansässige NDR unterhält ein Archiv mit ca. 60 Mitarbeitern. Diese dokumentieren die in Hamburg produzierten Tagesschau-Sendungen und weitere Sendeformate. Dokumentieren bedeutet, dass zu den einzelnen Sendungen nicht nur das Videomaterial archiviert wird, sondern auch eine genaue Beschreibung der Inhalte. So können Sendehalte später gezielt recherchiert werden. Aber auch ein solches Archiv stößt in der Praxis an seine Grenzen: Zu Talkshows beispielsweise dokumentiert der NDR nur die Namen der Gäste sowie die Sendezeiten. Inhalte der Diskussion werden nicht einmal oberflächlich abgelegt. Eine Recherche nach bestimmten Aussagen und Meinungen der Gäste ist nicht möglich.

Nicht nur TV-Produzenten sind von der Flut ihrer eigenen Daten überfordert. Neue Internetplattformen wie zum Beispiel makeTV.de produzieren user generated content ohne ein Konzept für die Archivierung und Recherche zu haben. Die Strategie solcher Start-Ups ist klar auf die schnelle Marktdurchdringung hin ausgelegt. Der eigene Erfolg wird aber mit wachsender Datenmenge zum Fluch: nur noch wenige Produktionen erreichen die Aufmerksamkeit der Anwender, der größte Teil geht in der Masse der eigenen Daten verloren. Hier ist eine Strategie für Archivierung und Recherche dringend notwendig. Klassische Archive ziehen nach. Hier liegen Video- und Audiomaterial der letzten 100 Jahre vor, die durch Digitalisierung erfassbar und dauerhaft haltbar gemacht werden müssen.

- Trend 3: Die Anforderungen der Konsumenten und Produzenten an Video und Audiomaterial werden anspruchsvoller. Klassisches Lean-Back-TV tritt gegenüber den interaktiven Angeboten zurück. Stichwort hier ist Individualisierung: Anstatt sich im heimischen Wohnzimmer berieseln zu lassen möchte der Zuschauer genau den Film ansehen, der zu seinem Aufenthaltsort, seiner Stimmung und seinen Interessen passt.

Recommender Systeme im Web aber auch im Funknetz nutzen möglichst genaue Metadaten um Nutzern speziell auf sie zugeschnittenes Material anbieten zu können.

Die Aufgabe bestand darin, in diesem Kontext Anwendungsszenarien für die an der Professur Medieninformatik entwickelten Software-Frameworks AMOPA (Automated Moving Picture Annotator) zur Analyse audiovisueller Medien und Xtrieval (Extensible Retrieval and Evaluation Framework) zur Recherche in Dokumenten zu definieren und auf einen Einsatz in einer wirtschaftlichen Umgebung hin zu untersuchen.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Während der oben genannte Trend 1 das Marktpotential der hier vorgestellten Technik andeutete, bezogen sich Trend 2 und Trend 3 direkt auf die technischen Aspekte. Beide Trends beschreiben Probleme, die mit Hilfe der an der Professur Medieninformatik der TU Chemnitz seit 2004 entstandenen Vorarbeiten lösbar sind. So wurden Verfahren entwickelt, die audiovisuelle Medien recherchierbar machen. Dazu gehört zunächst eine möglichst weit automatisierte Inhaltsanalyse, die mit Hilfe von Bild- und Spracherkennungsalgorithmen Beschreibungsdaten (Metadaten) generieren kann. Diese Algorithmen lagen zum Projektstart als Sammlung im Framework AMOPA (Automated Moving Picture Annotator) vor. Daneben wurden Techniken entwickelt, welche die Recherche in den generierten Metadaten ermöglichen. Hierzu wurden innovative Information Retrievalansätze geschaffen und in aufwändigen Evaluationen getestet. Die Algorithmen lagen zum Projektstart als Sammlung im Framework Xtrieval (Extensible Retrieval and Evaluation Framework) vor.

Beide Frameworks haben ihre Vorzüge gegenüber den Ergebnissen anderer Forschergruppen in internationalen Kampagnen unter Beweis gestellt. Seit 2006 nimmt die Professur Medieninformatik am Cross Language Evaluation Forum CLEF teil. CLEF (www.clef-campaign.org) ist eine EU-finanzierte Evaluationskampagne, bei der Retrievalverfahren im internationalen Vergleich getestet werden. Hier hat sich die Gruppe in verschiedenen Wettbewerbskategorien zu Text-, Sprach-, Bild-, Video- und Multimediaretrieval sehr gut behaupten können und belegte seit der ersten Teilnahme 2006 regelmäßig den ersten Platz (z.B. in den Kategorien Bild-, Text- und Videoretrieval im Jahr 2009). Auch in komplexen Bereichen ist die Gruppe sehr erfolgreich: In der Kategorie Gridretrieval konnte sie 2011 weltweit als einzige neben der University of Berkeley teilnehmen (und lag in den Ergebnissen vorne).

Damit ist die prinzipielle Überlegenheit der an der Professur Medieninformatik geschaffenen Verfahren gegenüber anderen in der Forschung existierenden Ansätzen zwar belegt. Für eine wirtschaftliche Verwertung waren diese Verfahren allerdings noch nicht geeignet. Hierfür mussten zunächst zwei zentrale Aspekte bearbeitet werden:

- **Geschwindigkeit:** Die in Java implementierten Verfahren verarbeiteten Videoströme maximal in doppelter Echtzeit. D.h. der Bildanteil eines Videos von 90 Minuten Länge wird in 90 Minuten analysiert. Der Audioanteil benötigt ebenfalls 90 Minuten Länge. Diese Geschwindigkeit reicht aus, um bei den oben beschriebenen Evaluationskampagnen teilnehmen zu können. Für den Fall wirtschaftlicher Verwertung ist sie jedoch nicht akzeptabel. Hier mussten Verfahren zur Parallelisierung entwickelt werden, die es nun erlauben, die Analysen zu beschleunigen. In Hinblick auf Geschwindigkeitsoptimierung

wurden die Arbeitsbereiche AB 1 Medientranscoding, AB 5 Bilderkennung und AB 7 Parallelverarbeitung konzipiert.

- Systemanbindung: Die an der Professur Medieninformatik entwickelten Annotations- und Retrievalalgorithmen lagen als API-Sammlungen im AMOPA- und XTRIEVAL-Framework gebündelt vor. Um an den CLEF-Evaluationen teilnehmen zu können, wurden provisorisch Schnittstellen zur Einspeisung der dort verwendeten Text-, Bild-, und Videosammlungen sowie zur Eingabe von Suchanfragen und der Ausgabe der Rechercheergebnisse erstellt. Für einen belastbaren Gebrauch in einem kommerziellen Umfeld waren diese Schnittstellen nicht nutzbar. Hier waren zusätzliche Schritte notwendig, um vollständige Systeme zu erhalten. Diese Maßnahmen waren Bestandteil der Arbeitsbereiche AB 2 Archivierungsstraße, AB 3 Workflow, AB 4 Annotationsunterstützung und AB 7 Web-Services.

3. Planung und Ablauf des Vorhabens

Die in Abschnitt 2 angesprochenen technologischen Aufgaben wurden thematisch in sieben Arbeitsbereiche aufgliedert. Die zeitliche Zuordnung und die Abfolge der Arbeitspakete werden durch das beigefügte Diagramm (s. Abb. 1) ersichtlich. Vier Arbeitsbereiche starteten zu Projektbeginn:

- Arbeitsbereich 1: Mediatranscoding
- Arbeitsbereich 2: Archivierungsstraße
- Arbeitsbereich 4: Annotationsunterstützung
- Arbeitsbereich 5: Bilderkennung

Die übrigen Arbeitsbereiche starteten planmäßig sukzessive im Laufe des Projekts:

- Arbeitsbereich 6: Web-Services nach sechs Monaten
- Arbeitsbereich 3: Workflowintegration nach neun Monaten
- Arbeitsbereich 7: Parallelverarbeitung nach 15 Monaten

Der vollständige Arbeitsplan mit spezifischen Einzelzielen für die Arbeitspakete Arbeitsbereiche ist in Abb. 1 dargestellt. Die konkreten Beschreibungen der Ziele für die jeweiligen Arbeitspakete Arbeitsbereiche sowie die erreichten Ergebnisse können den Teilbereichen des Abschnitts II.1 dieses Berichts entnommen werden.

Die Meilensteinplanung kann ebenso dem Arbeitsplan (s. Abb. 1: Übersicht Arbeitsplanung) entnommen werden. Die Meilensteinplanung sieht zwei Meilensteine vor. Der erste Meilenstein lag nach dem ersten Quartal 2012. Dann konnten einige prinzipielle Fragestellungen aus den Arbeitsbereichen beantwortet sein und es herrschte grundsätzlich Klarheit über die Machbarkeit des weiteren Projektverlaufs. Die Konzeption der Mediatranscodierung lag vor. Die Realisierung des Annotationswerkzeugs für den Browser war realisiert. Der Testkorpus für die SW-Bilderkennung war vorhanden. Ebenso konzeptuelle Überlegungen für die Web-Services.

Der zweite Meilenstein lag Ende 2012. Ziele waren hier, dass die Archivierungsstraße an AMOPA gekoppelt und getestet war. Die grundsätzliche szenarienbasierte Vorgehensweise wurde modifiziert bestätigt. Die Annotationsunterstützung und die Bilderkennung auf SW-

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Für die Bearbeitung des Projekts wurden zunächst die in „2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde“ beschriebenen, an der Professur Medieninformatik geschaffenen Softwareframeworks verwendet und weiterentwickelt. Daneben wurde soweit wie möglich auf existierende Open-Source-Softwarelösungen zurückgegriffen mit dem Ziel, keine Komponenten zu programmieren, die bereits vorliegen. Diese sind im Einzelnen:

- OpenCV (<http://opencv.org>): Hierbei handelt es sich um eine in C bzw. C++ vorliegende Softwarebibliothek zur Bildverarbeitung. Grundlegende Algorithmen des State-of-the-Art sind hier bereits implementiert und frei zugänglich und unter DSB-Lizenz weiterverwendbar. Die Bibliothek umfasst eine Reihe von sehr schnell arbeitenden Filtern (z. B. Sobel, Canny, Gauß) sowie einige grundlegende Algorithmen zur Gesichtsdetektion, die als Grundlage für weitere Entwicklungen hergenommen wurden. Ebenfalls verwendet wurden die Boostingalgorithmen des Maschinellen Lernens.
- Apache Lucene (<http://lucene.apache.org>): Hierbei handelt es sich um eine Programmibibliothek, die die schnelle Suche in großen Textdatenbeständen ermöglicht. Sie wird eingesetzt um beispielsweise die Recherche in Annotationen zu ermöglichen.
- CMU Sphinx (<http://cmusphinx.sourceforge.net>): Das von der Carnegie Mellon University als Open Source zur Verfügung gestellte Toolkit Sphinx ermöglicht es, aus Audioströmen Sprache zu transkribieren. Es bildet auch den Kern der im Projekt erstellten Sprachanalyse, wobei Sphinx zunächst einmal für den englischen, insbesondere den US-englischen Sprachraum konzipiert ist. Für das Deutsche mussten entsprechende Adaptionen vorgenommen werden.
- Open Smile (<http://www.audeering.com/research/opensmile>): Der „Munich Versatile and Fast Open-Source Audio Feature Extractor“ erlaubt es, grundlegende Audio-Features zu extrahieren und zu filtern. Er wurde eingesetzt zur Extraktion solcher Features wie beispielsweise von MFCC.
- FFMPEG (<https://www.ffmpeg.org>): FFMPEG ist ein Kommandozeilenprogramm zur Bearbeitung, Transcodierung und der Wiedergabe von Video- und Audiodateien. Wir erstellen damit alle Zwischen- und Zielformate (MP4 für Analyse, MP4 für Vorschau, MPG für Produktion), da es zuverlässig und stabil läuft und sich gut mit SSH verteilt starten und steuern lässt.
- MindSqualls DLL (<http://www.mindsqualls.net>): MindSqualls ist eine .Net Programmibibliothek zur Ansteuerung von LEGO MINDSTORM NXT mittels einer Bluetooth- oder USB-Verbindung. Sie ist in der Programmiersprache C# geschrieben, kann allerdings durch jede andere .Net Sprache verwendet werden. Eingesetzt wurde sie, um dem entwickelten Kassetten-Roboter Steuerbefehle von einem Computer aus zu signalisieren.

Für die Bearbeitung der wissenschaftlichen Fragestellungen des Vorhabens validAX wurde im Wesentlichen auf Fachliteratur über die Universitätsbibliothek der TU Chemnitz zurückgegriffen. Diese verfügt über online-Zugänge zu den wichtigsten Ressourcen der Informatik wie zum Beispiel der ACM Digital Library.

5. Zusammenarbeit mit anderen Stellen

Innerhalb der TU Chemnitz wurden Fragen des Projekts sowie mögliche Folgeanträge oder weitere Vorhaben mit themenverwandte Professuren diskutiert:

- Professur für Nachrichtentechnik, TU Chemnitz
- Professur Prozessautomatisierung, TU Chemnitz
- Professur Digital- und Schaltungstechnik, TU Chemnitz
- Professur Technische Informatik, TU Chemnitz
- Professur Mediennutzung (Mediensoziologie / Medienpsychologie), TU Chemnitz
- Professur Medienkommunikation, TU Chemnitz
- Lehrstuhl für Verteilte Informationssysteme, Universität Passau
- Multi Sensor Based Image Processing Group (MSIP), Chemnitz

Besonders hervorzuheben ist die Zusammenarbeit mit dem Lehrstuhl für Verteilte Informationssysteme der Universität Passau, ebenfalls ein Drittmittelnehmer im Programm VIP, mit dem halbjährlich ein Doktorandenworkshop durchgeführt wird.

Gespräche über Fragen der aktuellen Forschungstätigkeit sowie zu weiteren potentiellen Forschungs- und Entwicklungsmöglichkeiten wurden mit zahlreichen Unternehmen (-verbänden) geführt. Die Gespräche reichten von einmaligen Kontakten im Rahmen der Präsentationen auf der CEBIT bis hin zu einem intensiven und kontinuierlichen Austausch über die gesamte Projektlaufzeit. Insgesamt bestand Kontakt zu über 120 Unternehmen (-verbänden). Die zehn intensivsten davon waren:

- Arbeitsgemeinschaft Regionalfernsehveranstalter in Sachsen (ARiS)
- Fink&Partner GmbH, Dresden bzw. München
- Intenta GmbH, Chemnitz
- Kabeljournal GmbH, Grünhain-Beierfeld
- Mugler AG, Oberlungwitz
- Novartis AG, Basel
- SACHSEN FERNSEHEN GmbH & Co. Fernseh- Betriebs KG, Chemnitz
- Sächsische Landesanstalt für privaten Rundfunk und neue Medien (SLM)
- Sächsisches Staatsarchiv, Dresden
- Telefonica O2 GmbH, München

Daneben wurde mit dem Mentor Dr. Stephan Roppel (Geschäftsführer Gutefrage.net, jetzt Director bei Tchibo) die Ausrichtung und Strategie der Projektgruppe besprochen. Insbesondere in der Anfangsphase war dieser Kontakt besonders wichtig, um einen Einblick in wirtschaftliches Denken zu erhalten.

II. Eingehende Darstellung

1. Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen

AB 1: Flexible Mediatranskodierung für den Transport audiovisueller Medien

Im Rahmen des Arbeitspakets Medientranskodierung mussten zunächst die Parameter und Besonderheiten untersucht und klassifiziert werden, die beim Umgang mit audiovisuellem Material auf analogen Videobändern von Bedeutung sind. Insbesondere wurden spezifische

Kalibrierungen und Konvertierungsoptimierungen erprobt und evaluiert, die erforderlich sind, um auch Material auf stark gealterten Videobändern in möglichst hoher Qualität weiterverarbeiten zu können. So sind bei der Umwandlung des analogen Videosignals am Ausgang des Videorecorders in ein digitales Signal für die Digitalisierung und Transkodierung bestimmte Normabweichungen und Verfälschungen zu beobachten, die aus verschiedensten Fehlerquellen stammen können.

Daher mussten die Eigenheiten der verwendeten Komponenten, wie Videorecorder, Signalkonverter, Encoder und Transcoding-Software katalogisiert und angepasst werden. Als Resultat konnten alle verwendeten Komponenten optimal kalibriert und parametrisiert werden. Insbesondere Geräte der Consumer-Klasse zeigten deutliche Qualitätsschwankungen, die nur in begrenztem Umfang an die Signaltreue hochwertiger Studio-Geräte angenähert werden konnten.

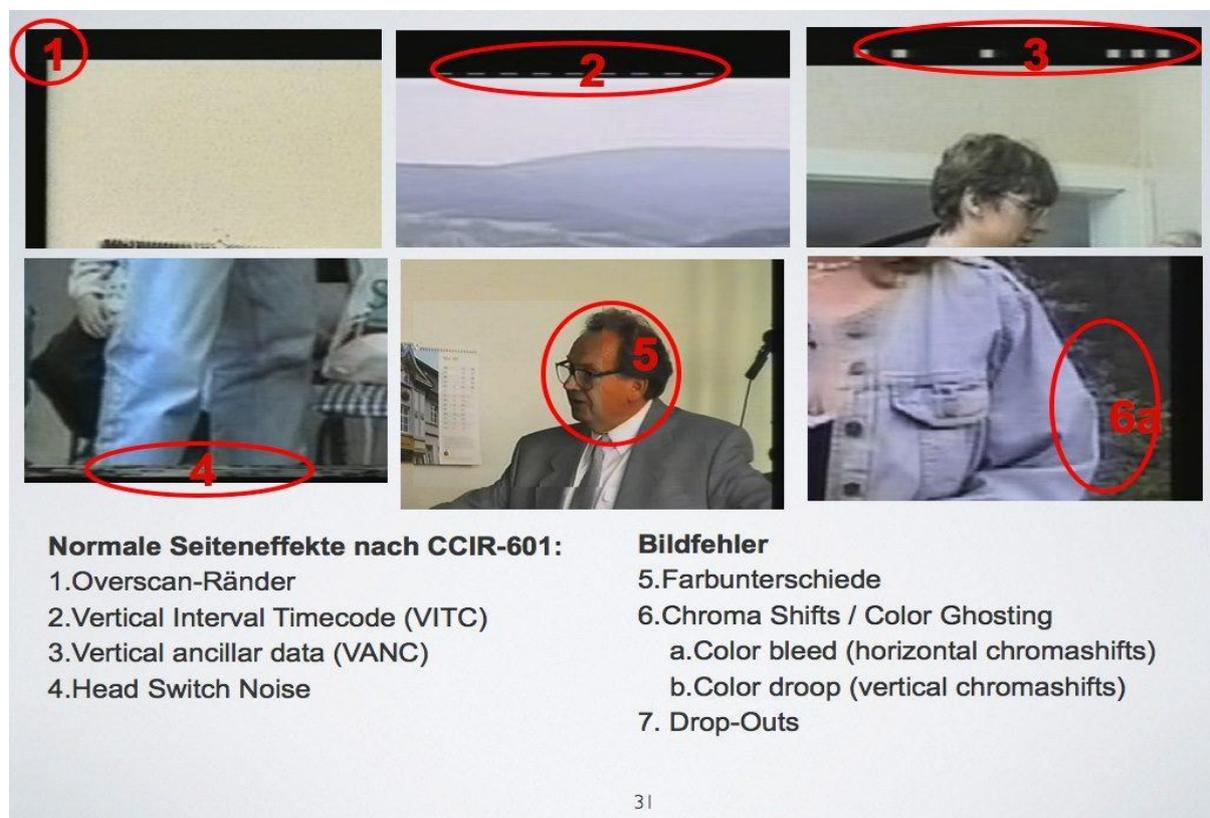


Abb. 2: Übersicht typischer Artefakte und Fehler, die bei der Digitalisierung von analogem (S)VHS Material entstehen und behandelt werden müssen.

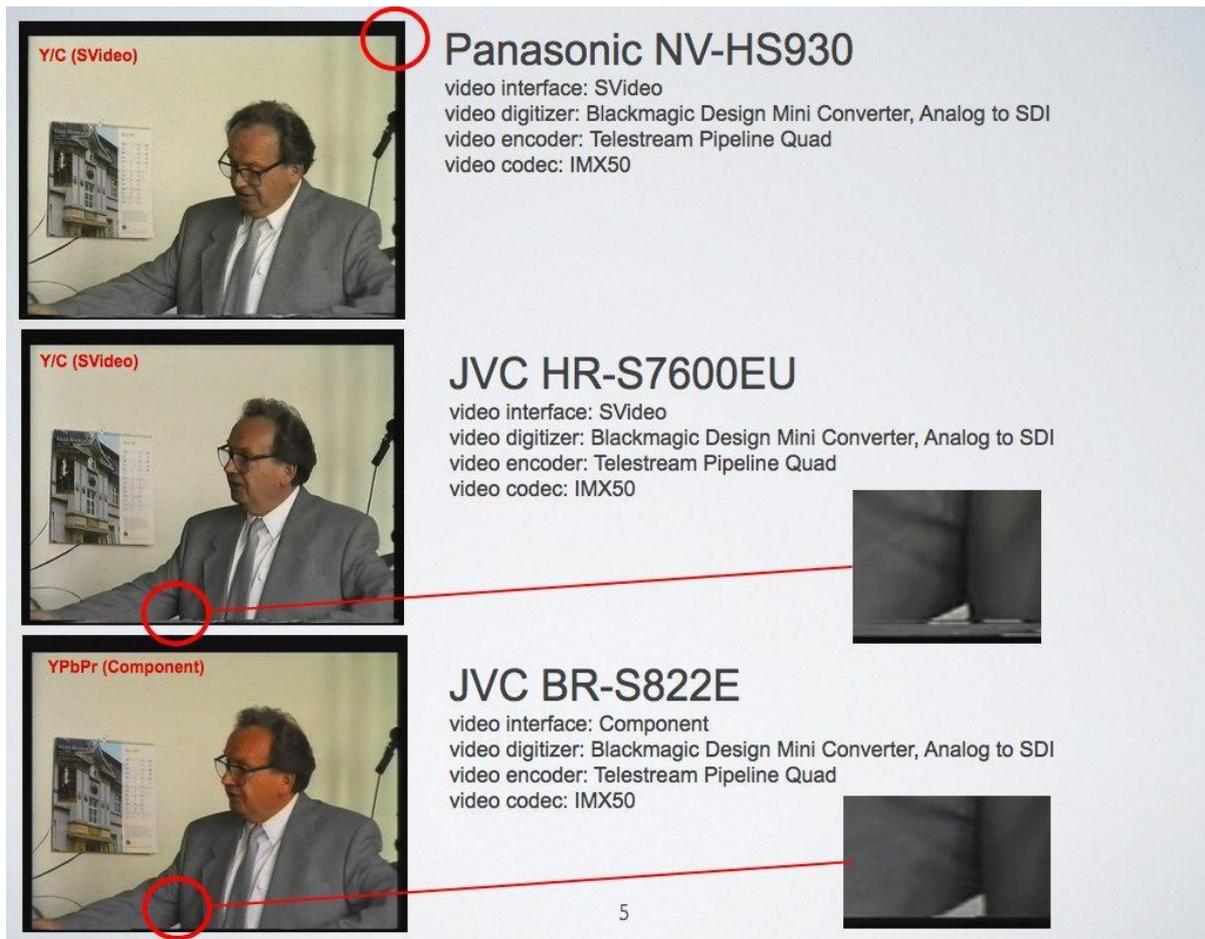


Abb. 3: Vergleich der Bildwiedergabe auf verschiedenen VHS-Playern bei gleicher Kalibrierung und Konfiguration. Insbesondere Randbildung und Stärke des "Head Switch Noise"

Der Einsatz eines speziellen Mess- und Kalibrierungsaufbaus ermöglichte anhand von RGB-, Waveform- und Vectorscope-Diagrammen die gerätespezifische Darstellungsqualität zu untersuchen. Die analoge Videotechnik der 1980/90 Jahre ist qualitativ bis zu einem Grad eingeschränkt, eine möglichst dichte Annäherung an das Optimum der Darstellung ist aber für eine weitere Verarbeitung von großer Bedeutung.

Das RGB-Überlagerungsdiagramm in Abb. 4 verdeutlicht die Kalibrierungseffekte. Es stellt auf der X-Achse die Breite des Bildes im Verhältnis zur Intensität der Luminanz auf der Y-Achse dar. Hierbei werden entlang des Diagramms drei Linienverläufe jeweils für die Chrominanz der drei Grundfarben Rot, Gelb und Blau eingezeichnet. Durch additives Mischen der Grundfarben entstehen so alle anderen Farben. Das EBU-Testbild macht so Schwächen und Farbabweichungen deutlich. Bei einer Idealen Darstellung müssten alle Linien im Bild nur aus fast exakten horizontalen oder vertikalen Linien bestehen und dort, wo sich Linien überlagern, müsste diese Überlagerung perfekt deckungsgleich sein.

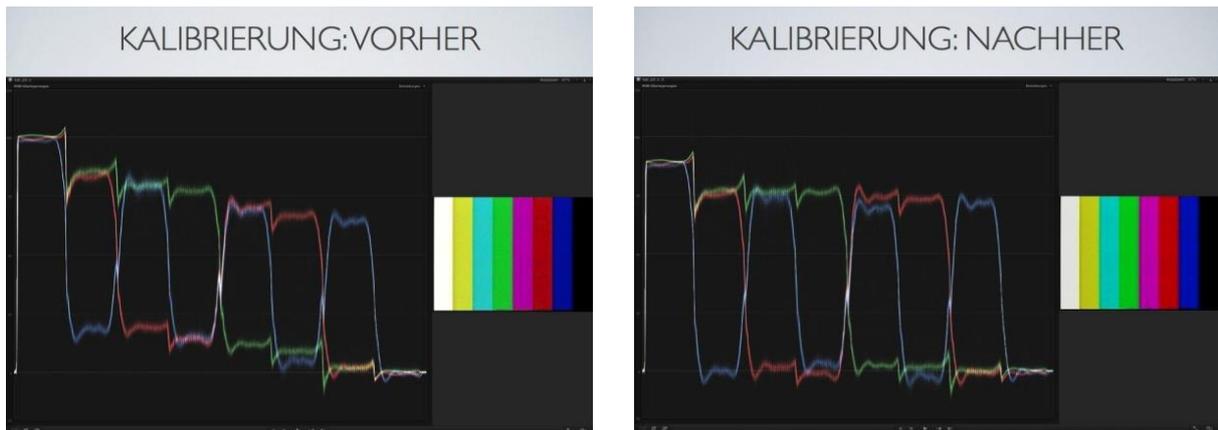


Abb. 4: Darstellung des EBU-Testbildes als RGB-Überlagerungsdiagramm, vor und nach der Kalibrierung

Das Vectorscope-Diagramm in Abb. 5 stellt die im Bild vorhandenen Farben (Chrominanz) mit ihrer jeweiligen Intensität (Luminanz) ins Verhältnis. Dabei wird jeder Farbton als Vector mit einem jeweils charakteristischen Winkel aufgetragen. Rottöne befinden sich im oberen Teil des Kreises, Grüntöne dagegen auf der gegenüberliegenden Seite unten. Bei Einsatz des EBU-Testbildes müssen sich die dargestellten Farbpunkte an bestimmten Stellen (eckige Klammern) zusammenballen. Treffen die Farbmaxima diese Markierungen nicht, liegt eine oder mehrere Farbverfälschungen vor, die ggf. durch Kalibrierung oder Transformation kompensiert werden können.

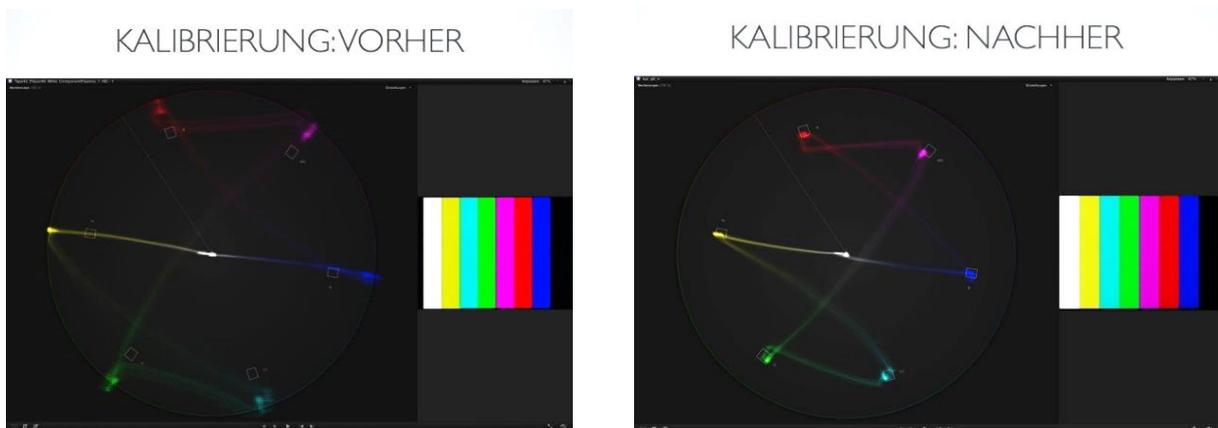


Abb. 5: Darstellung des EBU-Testbildes als Vectorscope-Diagramm vor und nach der Kalibrierung eines Players

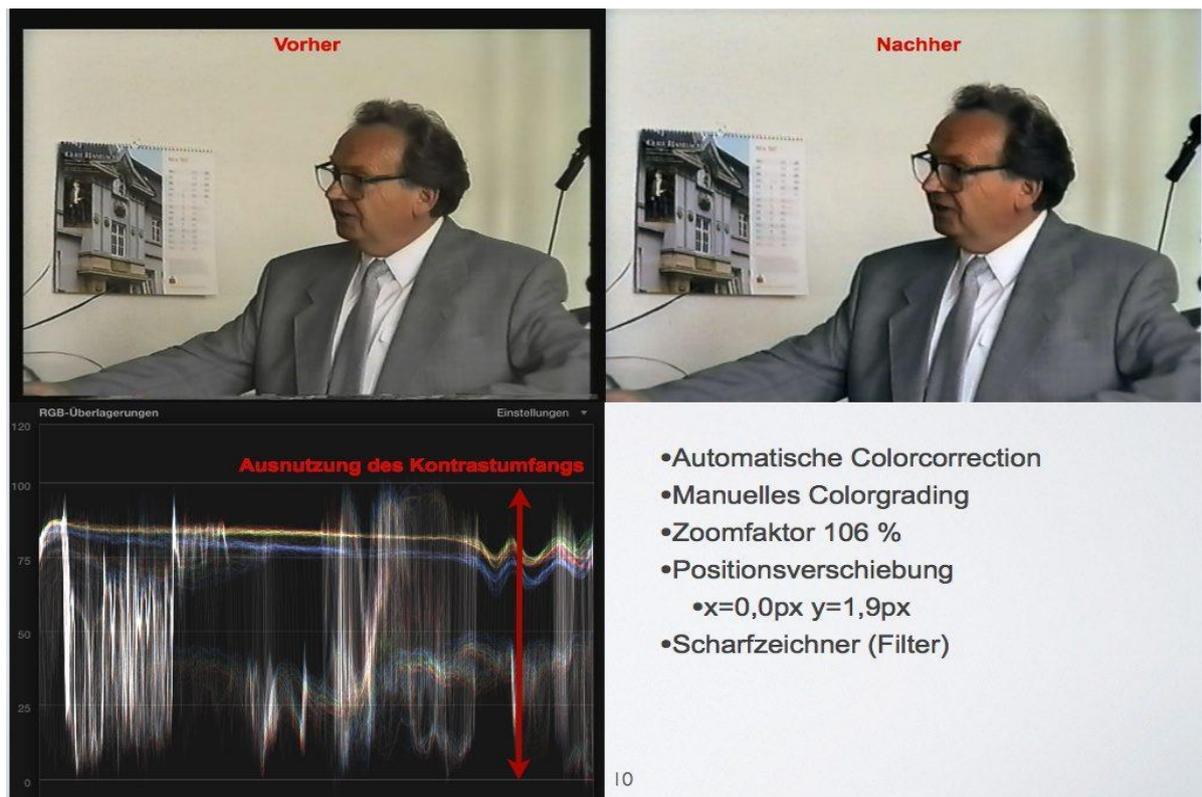


Abb. 6: Beispiel für die Leistungsfähigkeit der eingesetzten Kalibrierung und Optimierung als Vorher/Nachher-Vergleich für ein SVHS-Videoband

Ferner wurden Transkodierungsverfahren evaluiert und zusammen mit Kooperationspartnern bestimmte Archiv- und Nutzformate identifiziert. Durch den Einsatz einer integrierten Softwarelösung zur Verteilung sowie Konfiguration von Codecs ist es so möglich, digitalisiertes Videomaterial zeitnah in alle weiteren benötigten Formate umzuwandeln. Zu diesen Formaten gehört ein hochqualitativer Videocodec auf Basis von IMX50 im MFX-Container für die Archivierung, sowie einiger Proxy-Formate, die für andere Zwecke wie Videoanalyse, Webvorschau und Videoschnitt parallel erzeugt werden. Die Formate und Codecs können jederzeit an neue Erfordernisse angepasst werden und erlauben somit eine flexible Adaption neuer Szenarien oder geänderter Bedürfnisse.

Die im Rahmen der in den Untersuchungen zu den Arbeitspaketen 1.2, 1.3. und 1.4 erlangten Erkenntnisse wurden zusammengefasst und führten unter anderem zu einem Konferenzbeitrag¹ auf der LWA 2013. Die darin beschriebenen Formate und Vorgehensweisen, speziell die Entwicklung eines adaptiven und flexiblen Workflows für die Transkodierung von Material mit unterschiedlichen Farbräumen und Bildfrequenzen, stellen das Resultat der Untersuchungen bei den Kooperationspartnern, einer Analyse und Abstimmung von deren Bedürfnissen, als auch von Tests mit unterschiedlichem Beispielmateriale dar.

¹ Hems, Robert; Manthey, Robert; Ritter, Marc; Eibl, Maximilian (2013). Ein adaptiver Ansatz zum Ingest großer Bestände audiovisueller Medien unter heterogenen Anforderungen. In: Henrich, Andreas; Sperker, Hans-Christian (Hrsg.) LWA 2013 - Lernen, Wissen & Adaptivität, 7.-9.09.2013, Bamberg. - Bamberg: Proceedings LWA, 2013, S.268-273; <http://www.minf.uni-bamberg.de/lwa2013/proceedings/>

Aufgrund der Vielzahl der unterschiedlichen Anforderungen, der großen Varianz im Bereich der eingesetzten technischen Mittel, Systeme, Videoformate und Speicherkapazitäten, sowie der Unterschiede in Bezug auf die verfügbaren finanziellen Ressourcen, mussten die Methoden und Abläufe bei der Medientranskodierung anpassbar und in Abstimmung mit dem Nutzer erfolgen. Beispielhaft sind hier die in Tabelle 1 aufgeführten Formate F1, F2, F3 und F4 welche die Definition der Ergebnisdatenformate für die Transkodierungen des Kooperationspartners Kabeljournal GmbH darstellt.

Die im Berichtszeitraum durchgeführten Evaluationen des AP 1.5 dienten einerseits der Bestimmung der für die Transkodierungen notwendigen Parameter und andererseits als Entscheidungshilfe für die Abstimmung der Ergebnisformate mit den Kooperationspartnern. Hierzu wurde die Ähnlichkeit zwischen den Videobildern des Originalmaterials und denen des transkodierten Zielmaterials als objektives Vergleichsmaß verwendet. Da bei der Transkodierung verschiedene Parameter und Parameterkombinationen zu vergleichbaren Ergebnissen führen können, wurde die so entstehende Menge durch eine anschließende subjektive Bewertung, bei welcher auch die Kooperationspartner einbezogen wurden, zu einer einsatzgebietsspezifischen Ergebnismenge an Transkodierungsparametern reduziert. Eine anschließende Überprüfung der Echtzeitkriterien in der Verarbeitung und die Wahrung optimaler Bildqualität wurden durch die Anwendung dieser Parameter auf dem erstellten Testkorpus bestätigt.

Konfiguration \ Format	F1 (Analyse)	F2 (Preview)	F3 (Produktion)	F4 (Archiv)
Containerformat	MP4	MPEG	AVI	MXF (OP1a)
Videocodec	h.264	h.262	DV	IMX50
Auflösung	720 x 576	720 x 576	720 x 576	720 x 576
Chroma YUV	4:2:0	4:2:0	4:2:0	4:2:2
Bildrate	25 fps	25 fps	25 fps	25 fps
V-Bitrate	2 MBit/s	8,8 MBit/s	28,8 MBit/s	50 MBit/s
Audiocodec	AAC	MP3	PCM S16 LE	PCM S16 LE
Abtastrate	48 kHz	48 kHz	48 kHz	48 kHz
Kanäle	2	2	2	2
A-Bitrate	160 kBit/s	128 kBit/s	1,5 MBit/s	12.3 MBit/s

Tabelle 1: Verschiedene Zielformate für die Transkodierung als auch für den Testkorpus zur Evaluation der Anwendungsfälle Analyse, Preview, Produktion und Archivierung (aus: Herms et al. 2013)

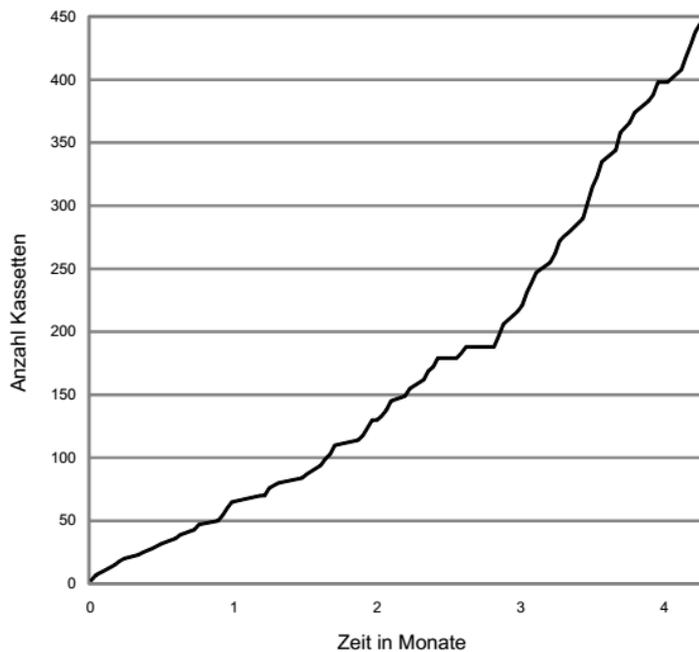


Abb. 7: Übersicht über den Zeitraum, in dem der Testkorpus (444 VHS-Kassetten) digitalisiert und transkodiert wurde

AB 2: Archivierungsstraße

AP 2.1: Aufbau der Hardware

Der Aufbau der Hardware wurde mit Beginn des Projektes durchgeführt. Aufgrund von Engpässen bei der Lieferung (siehe Abschnitt II.2) kam es 2011 zu Verzögerungen in der Beschaffung von Großgeräten. Dies betraf die DV-Player sowie den Digitalisierungscluster.

Für das automatische Be- und Entladen der einzelnen Player im Dauerbetrieb wurde ein Laderoboter konzipiert und realisiert. Gemäß Antrag wurde eine relevante Auswahl an Videoplayern vom Amateur- bis zum Profibereich abgedeckt. Hierzu zählen (S-)VHS, Betacam, und DV-Player. Die Player wurden durch eine fernsteuerbare Kamera ergänzt, die Aufnahmen der Playerzustände liefert und gleichzeitig die digitalisierten Kassetten dokumentiert. Für die Digitalisierung bzw. digitale Archivierung wurde ein entsprechender Rechencluster beschafft und in die Archivierungsstraße integriert.

Die DV-Player wurden in die Archivierungsstraße integriert, der Digitalisierungscluster konfiguriert und erfolgreich in Betrieb genommen. Weiterhin wurden die bereits entwickelten Arbeitsschritte des automatisierten Digitalisierungsprozesses erfolgreich miteinander verknüpft, so dass der Workflow zum Einspielen von verschiedenen Videobandformaten noch im ersten Quartal 2012 und damit bis zum ersten Meilenstein in Betrieb genommen werden konnte.

Hierfür wurde konkret eine Softwarelösung entwickelt, welche den Digitalisierungsworkflow steuert und gleichzeitig den Einspielprozess dokumentiert. Die Architektur sowie Realisierung und der Einsatz dieser Lösung sind in einer Publikation [Herms et al., 2012] wissenschaftlich festgehalten.

AP 2.2: Integration AMOPA

Die Integration des AMOPA-Systems stellt die Anbindung der Archivierungsstraße zum Analyse-Framework dar. Es handelt sich hierbei um zwei separat entwickelte Komponenten weswegen eine Integration in sechs Segmenten erforderlich war.

- Definition eines Analyse-Proxys für den Austausch von AV-Medien zwischen Archivierungsstraße und AMOPA (siehe AB 1).
- Entwicklung einer vereinfachten AMOPA-Integration mittels statischer Jobverarbeitung für den frühen und einfachen Einsatz als Analyse-Pipeline.
- Entwicklung einer erweiterbaren Analyse-Plattform auf Basis von AMOPA um flexibel auf neu auftretende Erfordernisse reagieren zu können.
- Eine Stand-Alone Version des AMOPA-Systems mit grafischer Benutzeroberfläche für Test- und Evaluationszwecke, sowie als Demo.
- Ein Identifikationssystem zur Organisation und Nachverfolgung des Gesamtprozesses über alles Systemgrenzen des Projekts hinweg.
- Eine zentrale Datenbank und Fileablage für die gebündelte Verarbeitung.

Der mit diesem Arbeitspaket assoziierte Anwendungsfall ist in Abb. 8 dargestellt und geht dabei auf die konkret ausgeführten Analyse-Module ein, abstrahiert gleichzeitig aber welche Art der AMOPA-Integration in diesem Fall Anwendung findet, da die im Kern verwendeten Module unabhängig davon arbeiten, ob sie als Stand-Alone Desktopanwendung, verteiltes Analysesystem oder skriptgesteuertes Stapelverarbeitungssystem verwendet werden.

Je nach den Bedürfnissen des Nutzers und den Erfordernissen seines Quellmaterials können verschiedene Analyse-Module aktiviert werden. Als Analyse-Module werden hierbei die verschiedenen Detektoren bezeichnet, die im AMOPA-Framework verfügbar sind und einzeln oder als Gesamtpaket auf ein konkretes AV-Medium angewendet werden können.

Die Ergebnisse des jeweiligen Moduls liegen zunächst in Textform vor. Durch die Integration in eine zentrale Datenbank und die Verwendung eines einheitlichen ID-Systems (AXID - AMOPA Xtrieval Identifier) ist die durchgängige Zuordnung und Wiederauffindbarkeit, sowie eine spätere Weiterverarbeitung gewährleistet.

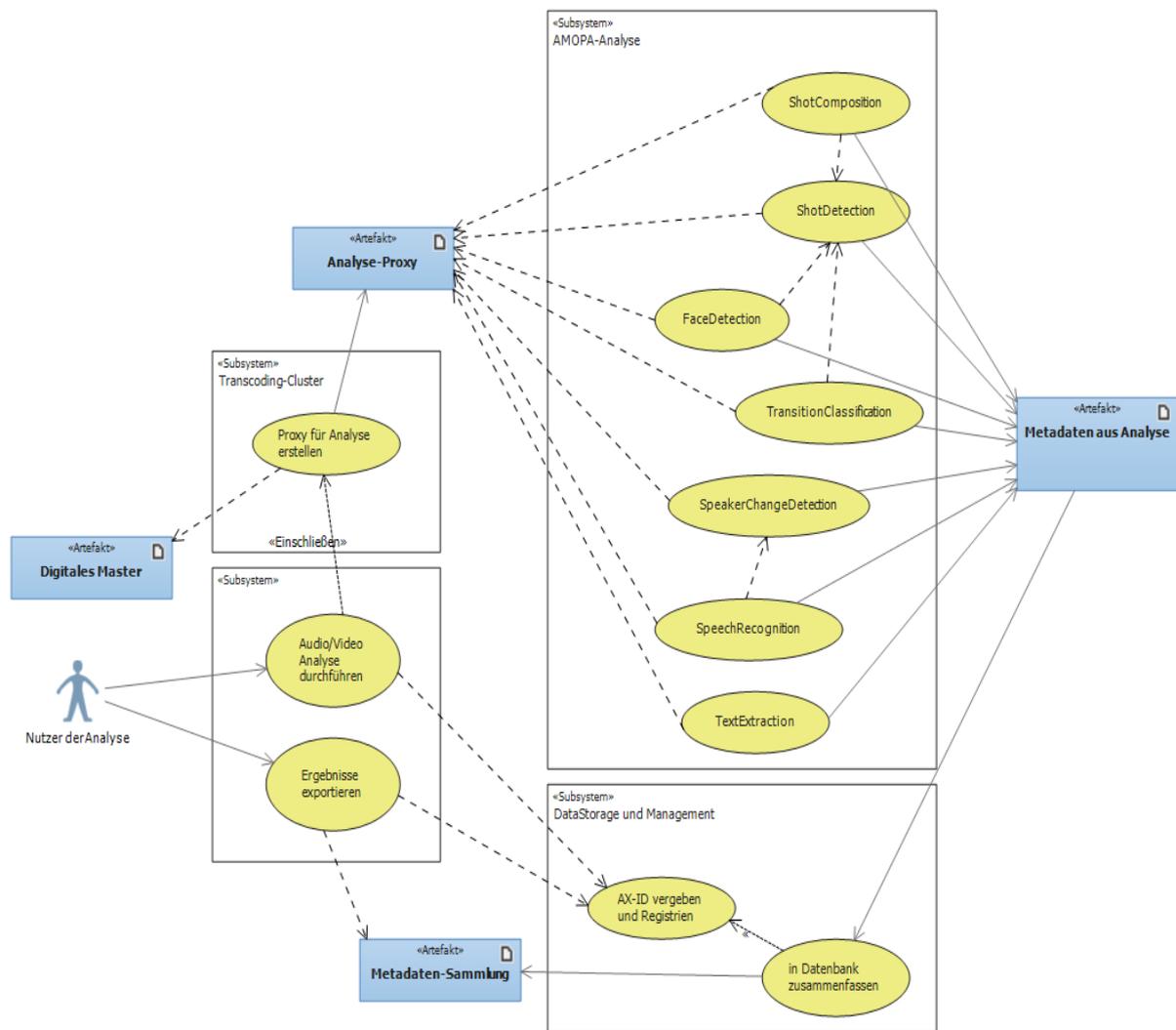


Abb. 8: Schematische Anbindung an AMOPA: Artefakte wie physische Datenträger oder Dateien sind blau, Anwendungsfälle gelb dargestellt. Subsysteme werden nur vereinfacht dargestellt und können weitere Elemente enthalten. Der Nutzer von AMOPA kann aktiv Anwendungsfälle, wie das Exportieren von Ergebnissen und die Audio/Video Analyse auslösen. Hierdurch werden im Hintergrund weitere Anwendungsfälle der Subsysteme Transcoding-Cluster, DataStorage und Management, und AMOPA-Analyse ausgelöst. Anwendungsfälle können als Voraussetzung oder Ressource auch von anderen Anwendungsfällen oder Artefakten abhängig sein (symbolisiert durch die unterbrochenen Linien), sie können auch neue Artefakte erzeugen (durchgezogene Linien).

Zur Analyse der durch die Archivierungsstraße erzeugten Videodateien kommt im ersten Schritt AMOPA als statische Jobverarbeitung zum Einsatz. Hierbei wird aus dem breitbandigen digitalen Master eine spezielle Videoversion als Analyse-Proxy für AMOPA erstellt und mittels FTP-Schnittstelle an die Analyse übergeben.

Da diese Form der Job-Verarbeitung zwar schnell realisiert werden konnte und so anderen Arbeiten zur Verfügung stand aber dennoch wenig flexibel ist und sich nicht an die Erfordernisse der erst später zu entwickelnden Workflow-Szenarien anpassen lässt, war eine parallele Entwicklung erforderlich. Zusätzlich wurde eine erweiterbare AMOPA-Plattform entwickelt, die auf dieselben AMOPA-Module wie die Jobverarbeitung aufbaut, aber in ihrer Softwarearchitektur besser an neue Erfordernisse angepasst werden kann. Als erste Variante der Plattform entstand eine Desktop-Applikation für Microsoft Windows Einzelplatzsysteme. Sie kann von Nutzern über eine grafische Benutzeroberfläche manuell gesteuert werden und dient zur Demonstration und Evaluation neuer Konfigurationen und Optimierungen.

Im Rahmen der in AB 3 entwickelten Szenarien ist es auf Basis dieser Plattform möglich, die Ausführung und Steuerung des AMOPA-Frameworks zu trennen und sowohl webbasierte Benutzeroberflächen anzubinden, wie auch die parallele Verarbeitung in verteilten Systemen bis hin zum Cloud Computing oder einer Service-orientierten-Architektur zu implementieren.

Da die genauen Anforderungen zukünftiger Szenarien im Verlauf des Projekts erst bestimmbar wurden und zum Beginn dieses AP noch nicht gänzlich bekannt sein konnten, war dieser flexible Ansatz wichtig. Gleichzeitig musste möglichst früh eine funktionierende Variante für die Archivierungsstraße bereitgestellt werden, weswegen sich die Zweigleisigkeit einer simplen und einer komplexeren aber dafür erweiterbaren Lösung anbot.

AP 2.3: Evaluation I: Pretest

Nach dem erfolgreichen Aufbau der Archivierungsstraße wurde von April bis Juli 2012 mit dem ersten Testlauf begonnen. Das Ziel war es, die Probleme bei der Digitalisierung zu identifizieren und eine hohe Qualität der Digitalisierung bei einer akzeptablen Größenordnung der aufkommenden Daten zu erreichen. Dazu wurden unterschiedliche Videobandformate wie VHS, DV und Betacam eingespielt und als dateibasierte Formate im System abgelegt.

Es wurde bereits in der frühen Testphase erkannt, dass insbesondere die Digitalisierung von VHS-Bändern eine Herausforderung darstellt, da die meisten Player dieses Formates zwar ein für den Heimgebrauch nutzbares Ausgangssignal bereitstellen, für die Digitalisierung im Profibereich dagegen musste auf Signalkonverter zurückgegriffen werden, um eine konstante Übertragung und die Ausfallsicherheit von Informationen zu gewährleisten. Da die Menge an Playern zum Teil unterschiedliche Qualitätsmerkmale aufweisen, musste das jeweilige Ausgangssignal kalibriert werden, um ein hochwertig qualitatives und einheitliches Niveau aller Player zu erreichen. Eine weitere Herausforderung stellte das Einspielen unterschiedlicher Videobandformate dar. Die Formate VHS und Betacam werden durch eine Signalkonvertierung und anschließende Enkodierung mittels Digitalisierungscluster als ein dateibasiertes Format im System gespeichert. Für das DV-Format war eine andere Strategie erforderlich, da auf DV-Bändern bereits digitales Material enthalten ist. Um das originale Material direkt in ein IT-basiertes System zu überführen, lag der Fokus zur Übertragung der Inhalte auf dem Schnittstellen-Standard IEEE 1394 (FireWire), welcher in den beschafften DV-Playern und im Digitalisierungscluster bereits verbaut wurde. Unter Berücksichtigung dieser Anforderung wurde das Einspielen durch eine eigenentwickelte Softwarelösung basierend auf Open-Source-Werkzeugen realisiert.

Der zweite Teil dieser Evaluation umfasste die Untersuchung von Datenmengen bezüglich optimaler Bedingung für unterschiedliche Anwendungsfälle. Hierzu wurden die drei Anwendungsfälle Archiv, Analyse und Online-Distribution betrachtet. Das Aufkommen der Datenmenge für den Fall Archiv ist aufgrund der Originaltreue am größten. Durch unterschiedliche Aufnahmebedingungen der Videobänder unterscheidet sich auch hierbei das Archivformat. Das DV-Material kann direkt von den Bändern abgegriffen und unverändert im Original mit einer Datenrate von 25 Mbit/s als Archivformat gespeichert werden. Bei den zu digitalisierenden Formaten steht die Enkodierung im Mittelpunkt, die nach Möglichkeit verlustfrei sein sollte. Eine hierbei aufgestellte Untersuchung ergab, dass sich eine Speicherung mit 50 Mbit/s als optimal erweist. Dies bestätigt auch das Regelwerk der ProSiebenSat.1 Media AG sowie Rücksprachen mit dem Lokalfernseher Kabeljournal. Für

den Anwendungsfall Analyse wurden verschiedene Datenmengen empirisch überprüft, um optimale Ergebnisse aus den Verfahren der Technologien zu erhalten. Die minimale Datenrate bei noch optimalen Ergebnissen variiert je nach Auflösung und Inhalt zwischen 1 und 2 Mbit/s. Bei der Online-Distribution geht es um das Betrachten von Videos über den Web-Browser. Die Datenmenge für eine akzeptable Bild- und Audiowiedergabe wird durch die Auflösung sowie dem Komprimierungsverfahren bestimmt und liegt zwischen 512 Kbit/s und 2 Mbit/s.

AP 2.4: Evaluation II: Massentest

Nach der Untersuchung unterschiedlicher Ein- und Ausgangsformaten sowie der grundlegenden Vorgehensweise, knüpfte im August 2012 die zweite Evaluation an. Der Fokus liegt dabei auf dem Einsatz der Technologien im Dauerbetrieb sowie der Verarbeitung großer Datenmengen.

Hierzu wurde ein umfangreicher Bestand heterogenen Materials zusammengestellt, der repräsentativ für den Großteil der Fernseharchive ist und eine Herausforderung für die entwickelten Technologien darstellt. Das in diesem Berichtszeitraum eingesetzte Material stammt u.a. vom Lokalfernseher Kabeljournal, der medienpädagogischen Einrichtung Filmwerkstatt e.V. und der Fakultät Medien der Hochschule Mittweida.

Zunächst wurde mit dem Digitalisieren bzw. Einspielen von VHS-Material begonnen. Dabei wurde im Laufe der Zeit die Speicherstrategie angepasst. Statt eines gesamten Speicherbereichs wurde nun für jeden Einspielkanal ein entsprechender temporärer Speicher zugewiesen und somit eine flexible Entlastung des Systems garantiert.

Aufgrund der Konfiguration zur Lastverteilung sowie Wartungen und entsprechenden Auswertungen von Ergebnissen konnten bis Ende 2012 für VHS lediglich 500 Stunden Videomaterial erfasst und gespeichert werden. Alle Störfälle wurden beseitigt, so dass die Evaluation im aktuellen Berichtszeitraum erfolgreich verlief.

Mit Beginn 2013 wurde ein Teilbestand des lokalen Fernsehsenders Kabeljournal GmbH im Umfang von 444 VHS-Kassetten mit einer Gesamtdauer von 1.450 Stunden innerhalb von vier Monaten im Dauerbetrieb digitalisiert und für die Archivierung aufbereitet. Hinzu kam ein Teil des Bestandes der Technischen Universität Chemnitz mit 100 VHS-Kassetten, die insgesamt ca. 200 Stunden umfassen. Eine Evaluation der Technologien bzgl. der Heterogenität von Eingangsformaten erfordert den Einsatz unterschiedlicher Standards. Daher erfolgte die Integration des Betacam-Formates. Hierzu wurde ein repräsentativer Bestand von 50 Videobändern mit einer Gesamtdauer von 30 Stunden von der medienpädagogischen Einrichtung Filmwerkstatt e.V. sowie der Fakultät Medien der Hochschule Mittweida akquiriert. Aufgrund der gleichen Bauweise aller im Vorfeld beschafften Betacam-Player und der Möglichkeit des direkten Abgreifens des Videosignals, ließ sich Betacam als Format problemlos und automatisiert mittels der entwickelten Einspieltechnologien verarbeiten. Sowohl digitale als auch analoge Beta-Formate konnten ohne Zwischenfälle eingespielt werden, entsprechend fielen die Ergebnisse der Evaluation positiv aus.

AP 2.5: Integration Xtrieval

Parallel zu der Umsetzung der im AB 3 entwickelten Workflow-Szenarien wurde eine Integration des Xtrieval-Systems in den Demonstrationsprototypen (siehe auch AP 4.3, AP 4.5) vorgenommen. Dabei wird Xtrieval als eigenständiges System für Suchanfragen verwendet. Es

erhält neu eintreffende Daten aus der Datenbank MetaBase in Form von speziellen Tabellen-Views und erzeugt hieraus einen Index. Über eine Schnittstelle lassen sich die Searchengine und der erzeugte Index nun von Nutzern wie Redakteuren oder Archivaren über das User-Interface des Prototypen "Thundercloud" ansprechen und nutzen. Die Möglichkeiten für Suchanfragen gehen dabei über die klassische einzeilige Suchanfrage, wie sie bei Suchmaschinen wie "Google" üblich sind, hinaus. So ist neben einer erweiterten Suchanfrage auf Feldebene auch eine Facettierung möglich. Im weiteren Projektverlauf könnten die Möglichkeiten von Xtrieval weiter ausgebaut werden und zum Beispiel Hervorhebungen erzeugt und dargestellt oder Feedback des Nutzers zur Verbesserung der Suchergebnisse herangezogen werden.

AP 2.5: Evaluation Xtrieval

In Kombination mit der Thundercloud wurden umfangreiche Tests der Integration von Xtrieval durchgeführt. Diese beinhalteten zum einen technische Tests. Hier wurden beispielsweise Unit-Tests eingeführt und durchgeführt. Zum anderen wurden klassische Retrievaltests durchgeführt, um die Retrievalgüte optimal auf den vorhandenen Videokorpus einzustellen.

AB 3: Workflowintegration

Im Arbeitsbereich Workflowintegration wurden Szenarien und Techniken entwickelt, um sowohl AMOPA und XTRIEVAL, als auch im Projekt entwickelte zusätzliche Komponenten in Arbeitsabläufe bei potentiellen und zukünftigen Verwendern integrieren zu können. Ziel des Arbeitsbereichs war die Realisierung von drei durchgängigen Anwendungsszenarien.

Im Mittelpunkt des Szenarios 1 "Kleine und mittlere Fernseharchive" standen dabei Digitalisierung und Archivierung. Szenario 2 "Sendung 2.0" baute auf digitalisiert vorliegendem Material auf und hatte besonders die automatisierte und personalisierte Weiterverbreitung von Inhalten zum Kern. Das Szenario "Sendung 2.0" hat dabei zwei Kernkomponenten: erstens die Automatisierung und zweitens die Personalisierung. Für beide Szenarien wurde 2013 ein Demonstrator erstellt und anschließend auf der CeBIT 2014 vorgestellt. Für das Szenario 3 wurden im Laufe des Jahres 2012 Kontakte zum Unternehmen Novartis aufgebaut. Das Szenario rückte die medizinische Medienverarbeitung in den Fokus.

Weil im Projekt keine eigene Stelle für Innovationsmanagement zur Verfügung stand, konnten die Arbeiten im AB nicht auf ein entsprechend vorbereitetes strategisches Konzept aufbauen. Stattdessen mussten zunächst nicht nur die entsprechenden Szenarien entwickelt werden, sondern auch zugrundeliegendes Innovationspotential evaluiert werden. Folgerichtig wurde der Arbeitsablauf für die Szenarien nicht wie ursprüngliche geplant in drei Phasen strukturiert, sondern in fünf Schritten, die zyklisch ineinandergreifen. Im Folgenden werden zwei erfolgversprechende Szenarien aus Schritt 1 kurz vorgestellt. Anhand der vier folgenden Schritte Analyse & Konzeption, Realisierung, Evaluation und Innovationspotential wird das Szenario 1 näher dargestellt.

- Schritt 1: Innovationspotential: Grundlegende Anknüpfungspunkte mit einer Vielzahl an Kooperationspartnern strategisch diskutieren.
- Schritt 2: Analyse & Konzeption: Workflows, Interaktionsschritte, Technologien untersuchen und Integrationsstellen für AMOPA / XTRIEVAL identifizieren.
- Schritt 3: Realisierung: Die definierten Konzepte zur Integration umsetzen.

- Schritt 4: Evaluation: Anwender in den neuen Workflows und ihrer Wartung schulen.
- Schritt 5: Innovationspotential: Tragfähigkeit und Relevanz der Szenarien unter wirtschaftlichen, wissenschaftlichen und praktikablen Gesichtspunkten evaluieren. Dieser Schritt setzt dabei an allen Referenz- und Differenzpunkte der Schritte 2 bis 4 an.

Schritt 1. Innovationspotential

Um Anknüpfungspunkte für erfolgversprechende Szenarien zu identifizieren und deren Potential abzuschätzen, wurden zunächst Gespräche zur strategischen Ausrichtung mit Kooperationspartnern in der Wirtschaft, in Hinblick auf Digitalisierung, Archivierung und Distribution geführt. Dazu gehören:

- Deutsches Rundfunkarchiv
- Behörde des Bundesbeauftragten für die Stasi-Unterlagen
- Hochschule Mittweida
- Lokalsender KabelJournal
- Lokalsender salve.tv
- Sächsische Landesmedienanstalt
- Sächsisches Staatsarchiv
- Mitglieder des Bundesverbandes Lokal-TV
- Senderverbund Regional Fernsehen (www.srf-media.de)
- Arbeitsgemeinschaft Regionalfernsehveranstalter in Sachsen (www.lokalfernsehen.de)

Anschließend wurden zwei Szenarien ausgewählt, die in Übereinstimmung mit dem Technologiepotential am vielversprechendsten erscheinen. Um diese sowohl im Hinblick auf Praxisnähe und Verwertbarkeit zu validieren, erfolgte die Definition der Use Cases in den zwei Szenarien in enger Zusammenarbeit mit den Partnern aus der Wirtschaft. Im Mittelpunkt des Szenarios 1 "Kleine und mittlere Fernseharchive" stehen Digitalisierung und Archivierung, während Szenario 2 "Sendung 2.0" auf digitalisiert vorliegendem Material aufbauend, besonders die automatisierte und personalisierte Weiterverbreitung von Inhalten zum Kern hat. Die Umsetzung von Szenario 2 ist dabei nicht Inhalt des Berichtszeitraums und wird daher im Folgenden nicht näher erklärt.

Schritte 2./3./4./5.: Szenario 1 "Kleine und mittlere Fernseharchive"

Dieses Szenario entstand in seiner Grundkonzeption in enger Zusammenarbeit mit dem Kooperationspartner Kabeljournal aus Baierfeld, einem der größten Lokalfernsehanbieter Sachsens. Es stellt die konkreten Bedürfnisse kleiner und mittlerer Programmveranstalter in den Mittelpunkt und bewegt sich damit sehr nah an einem realen Einsatzgebiet für die entwickelten Lösungen. Dies drückt sich auch in der Bereitschaft des Kooperationspartners aus, über den exemplarischen Probebetrieb hinausgehend sein gesamtes Programmarchiv für den Massentest und -betrieb zur Verfügung zu stellen.

Im Fokus des Szenarios steht dabei die Digitalisierung und Analyse vorhandener Videoband-Archive, wie sie in den allermeisten Regional- und Lokalfernsehsendern Sachsens seit den 1980er Jahren geführt werden. Diese Archive sind aus einigen Gründen von besonderer Bedeutung:

Zeitgeschichtliche Relevanz: Insbesondere die Lokalfernsehsender im Bereich der ehemaligen DDR, verfügen mit ihren Aufzeichnungen der Wendezeit mit regionalen Bezügen über zeitgeschichtliche Dokumente von herausragender historischer Bedeutung, deren Erhalt erstrebenswert ist.

Verfall des Archivmaterials: Die meisten Programmveranstalter, auch über die Grenzen der ehemaligen DDR hinaus, haben in der Vergangenheit nicht auf optimale Technologien und archivsichere Lagerungsmöglichkeiten zurückgegriffen. Daher befinden sich die vorhandenen Archive häufig bereits in prekärem Zustand und sind in den kommenden Jahren von massiven Materialverlusten bedroht.

Drohender Sendeschluss: Zudem ist die Zukunft einiger Lokalsender auch, zumeist aus personellen Gründen, aber auch aufgrund der sich wandelnden Medienlandschaft, unsicher. Der Erhalt der entsprechenden Archive über das Ende der Sendetätigkeit hinaus ist wünschenswert.

Mit der Digitalisierung dieser Videoband-Archive und der Erstellung digitaler Masterkopien wird der Grundstein für ein Fortbestehen all dieser Dokumente gelegt. Darüber hinaus werden diese Dokumente durch den Einsatz der automatischen Analyse und eine Suchindizierung erstmalig überhaupt für die Recherche zugänglich. Dies ist, nicht nur aus zeitgeschichtlichen Gründen für Historiker, sondern aus materiellen Gründen vor allem auch für die Zweitverwertung, relevant.

Das Szenario stellt einen idealen Testfall für die Archivierungsstraße des AB2 dar und bildet gleichzeitig dessen primären Use Case. Für die Integration der Archivierungsstraße, der Audio/Video-Analyse AMOPA und dem Suchsystem Xtrieval in den Workflow des Szenarios waren aber verschiedene Anpassungen und Neuentwicklungen erforderlich. Konzeptionell umfasst der Workflow zunächst die folgenden neun Schritte:

- 1. Schritt: Die Anlieferung von archivierten Videobändern lokaler TV-Sender (zumeist im Format SVHS, miniDV und Betacam).
- 2. Schritt: Die Registrierung und Etikettierung dieser Bänder für die spätere Nachverfolgung und Wiederauffindbarkeit mittels eines einheitlichen ID-Systems und QR-Codes.
- 3. Schritt: Automatisierter Ingest und Digitalisierung mit der Archivierungsstraße gemäß AB 2.
- 4. Schritt: Transkodierung der neu entstandenen Master-Kopie als Proxy-Dateien für die Weiterverarbeitung.
- 5. Schritt: Analyse der Proxy Version mit AMOPA zur Generierung von inhaltlichen Metadaten.
- 6. Schritt: Zusammenfassung der gewonnenen Metadaten, der vom Archiv-Besitzer mitgelieferten Metadaten, des Digitalen Masters und verschiedener Proxy-Versionen zu einem Archivierungspaket.
- 7. Schritt: Persistente Speicherung des Archiv-Pakets auf LTO-Magnetbändern und Übergabe der Magnetbänder als Austausch-Medium an den Archivbetreiber.
- 8. Schritt: Indexierung der Metadaten durch das Xtrieval-System.
- 9. Schritt: Bereitstellung des Indexes als Suchsystem für Archivbetreiber.

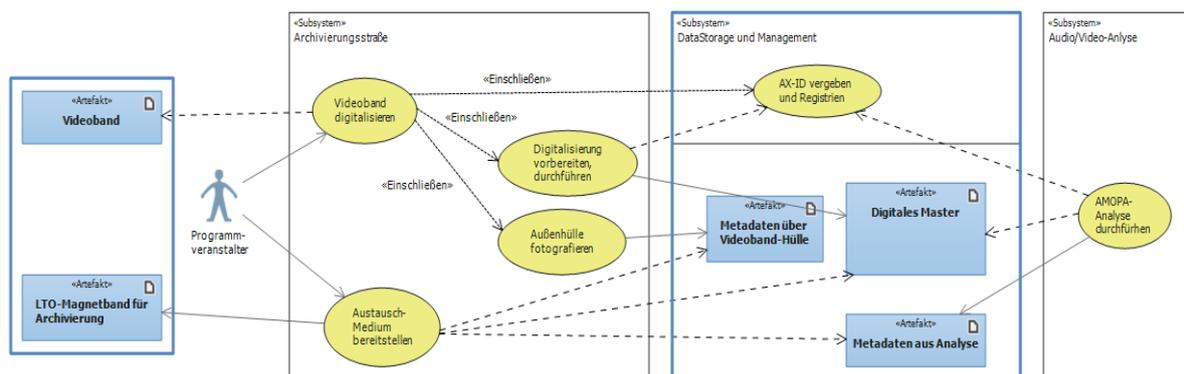


Abb. 9: Use Case für Szenario 1 - Digitalisierung

In Abb. 9 sind die Punkte 1. bis 7. als Use Case-Diagramm zusammengefasst, wie sie sich für den Nutzer von außen darstellen. Die internen Abläufe sind deutlich komplexer und bleiben dem Nutzer dabei verborgen. Er übergibt das Original-Videoband und erhält es später zusammen mit einem neuen LTO-Tape zurück auf welchem die Digitale Masterkopie, sowie die Analyse-Ergebnisse in Dateiform gespeichert sind. Dieses Paket kann dann vom Nutzer gemäß seiner eigenen Archivierungsstrategie genutzt werden. Die Abb. 9 stellt Teile des Szenarios als UML-Use Case Diagramm dar. Blaue Rechtecke stellen Artefakte wie physische Datenträger oder Dateien dar. Subsysteme können weitere Elemente enthalten, die zur Vereinfachung hier nicht dargestellt sind. Gelbe Ellipsen stellen Anwendungsfälle dar. Der Programmveranstalter (Nutzer) löst diese Anwendungsfälle aktiv aus (durchgezogener Pfeil auf den ausgelösten Anwendungsfall). Hierdurch können im Hintergrund verborgen weitere, untergeordnete Anwendungsfälle ausgelöst werden, wie Digitalisierung vorbereiten, AX-ID vergeben, Außenhülle fotografieren. Anwendungsfälle können als Voraussetzung oder Ressource von anderen Anwendungsfällen oder Artefakten als Vorbedingung abhängig sein, wie Metadaten über Videoband-Hülle, Digitales Master, Metadaten zur Analyse. Anwendungen können außerdem neue Artefakte erzeugen, wie zum Beispiel das Photographieren der Außenhülle zusätzliche Metadaten über die Videoband-Hülle mit sich bringt.

In Abb. 10 werden die Punkte 8 und 9 dargestellt. Nach der Zusammenfassung der Analyse-Ergebnisse und aller sonstigen Metadaten kann ein Suchindex erzeugt werden. Mittels Xtrival wird damit die Schlagwortsuche und Darstellung in einem Suchinterface möglich. Der Nutzer kann dadurch die Metadaten-Datenbank nicht mehr nur anhand der AXID durchsuchen, sondern Volltextsuchen über alle Analyse-Daten durchführen, wie den gefundenen Texteinblendungen im Bild und der mittels Spracherkennung übersetzen Stimmen. Darüber hinaus lässt sich auch die Struktur eines gefundenen Videos darstellen, da Schnitte und Szenenübergänge im Videobild, sowie erkannte Gesichter visualisiert werden.

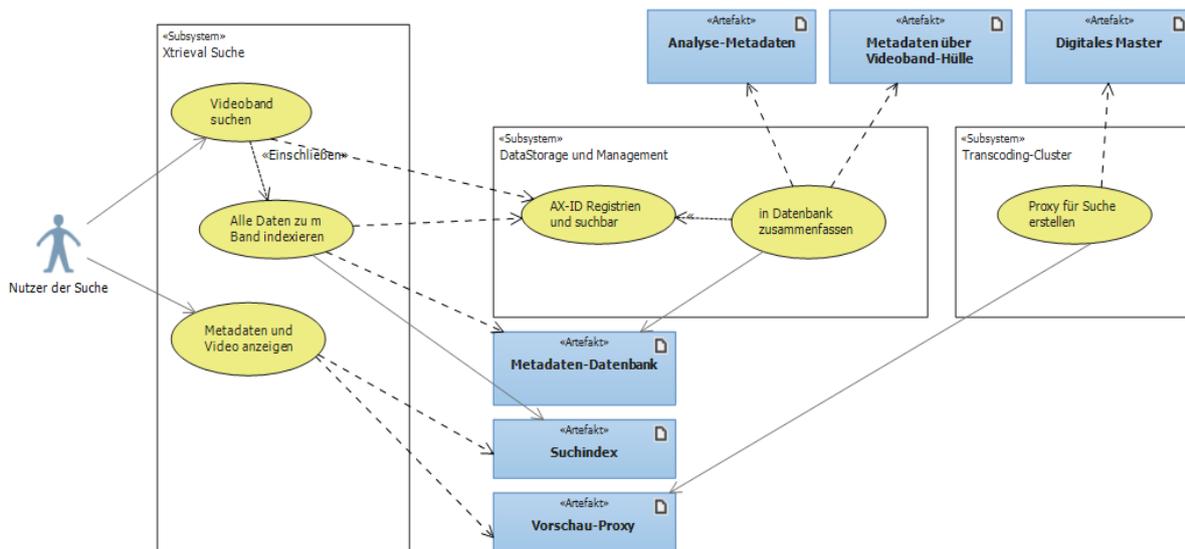


Abb. 10: Use Case für Szenario 1 - Suche und Erschließung mit Xtrieval

Das hier dargestellte Szenario bildete die Grundlage für die Evaluation und den Massentest der Archivierungsstraße aus AB 2. Während zu Beginn des Tests, der Fragilität des Materials geschuldet, außerordentlich behutsame Tests durchgeführt wurden, konnte in weiteren Iterationen eine Steigerung in Stabilität und Durchsatz des Systems erreicht werden. Seitens des Kooperationspartners wurde inzwischen das gesamte Archiv für die Digitalisierung nach Muster dieses Szenarios bereitgestellt.

Diese Ergebnisse ermutigte unseren Kooperationspartner dazu, die Digitalisierung den weiteren Mitgliedern der ARiS (Arbeitsgemeinschaft Regionalfernsehveranstalter in Sachsen) zu empfehlen. Diese führten im Anschluss an validAX 2014 zu einem Pilotprojekt mit der Sächsischen Landesmedienanstalt (SLM) zur Ausweitung der Digitalisierung auf alle 58 Lokalfernsehsender in Sachsen.

Szenario 2a: Personalisierte Zugänge zum digitalen Medienarchiv

Für Szenario 1 wurde eine Vielzahl flexibler Lösungen zu einem leistungsstarken Framework zusammengefasst. Hier standen die Bedürfnisse kleiner und mittlerer Lokalfernsehsender im Vordergrund. Auf dieses Framework aufbauend kann aber eine Vielzahl an Anwendungen entwickelt werden, die an unterschiedliche Nutzungskontexte anschließen. Zur praxisnahen Demonstration dieser Möglichkeiten wurde ein verteiltes System entwickelt, das an einem konkreten Anwendungsfall exemplarisch demonstriert, wie flexibel und anwendungstauglich das Framework ist. Im Folgenden werden zunächst das Framework und seine Komponenten vorgestellt und danach der Anwendungsfall beschrieben. Abschließend werden weitere Anwendungsfälle, die auf diesem Framework aufbauen können, skizziert.

Skalierbares Framework

Das Framework besteht aus den drei Komponenten MetaBase, Xtrieval-Integrator und Thundercloud. MetaBase ist ein Data-Warehouse System zur zentralen Speicherung und Aggregation der Analysedaten, Xtrieval-Integrator ist eine spezifizierte Schnittstelle zur redaktionellen Suche in den Datenbeständen (siehe AP 2.5). Auf technischer Ebene ist Thundercloud ein integratives System für Retrieval und Ergebnis-Visualisierung. Der unmittelbare Vorteil dieser Demonstrationsplattform liegt in ihrem modularen Aufbau und

der konsequenten Anwendung von State-of-the-Art-Schnittstellen und Techniken. Dies ermöglicht es im weiteren Verlauf des Projekt und darüber hinaus neue Funktionen und Module hinzuzufügen. Das System bietet die Möglichkeit der einfachen Skalierung und Personalisierung. So können Daten neuer Analyse-Module mit entsprechenden Interface-Elementen für Retrieval und Visualisierung integriert werden. Das System lässt sich auch mit geringem Aufwand für die Verarbeitung größeren Datenmengen skalieren oder beispielsweise in eine Cloud-basierte Lösung transferieren.

Personalisierbare Benutzer-Schnittstelle

Thundercloud ist auf Nutzerseite - trivial gesagt - eine webbasierte graphische Oberfläche für die Recherche in den audio-visuellen Datenbeständen, die im Rahmen des Projekts digitalisiert und analysiert wurden. Nutzer können je nach Anwendungszweck textuelle oder multimediale Retrievalanfragen stellen; diese werden vom System entsprechend aufbereitet textuell und multimedial visualisiert.

Annotationsunterstützung und Workflow digitaler und analoger Daten

Im Rahmen von Szenario 1 wurden eine Datenbank und ein entsprechendes Datenbank-Schema entwickelt. Damit können Bänder und digitalen Master mit einheitlichen Video-IDs über den gesamten Workflow hinweg verknüpft werden. Ein Anwender, der die Retrodigitalisierung überwacht, kann einerseits analoge Bänder, LTO-Master und digitale Master via QR-Scan zuordnen (s. Abb. 11). Andererseits können Metadaten hier bequem editiert werden.

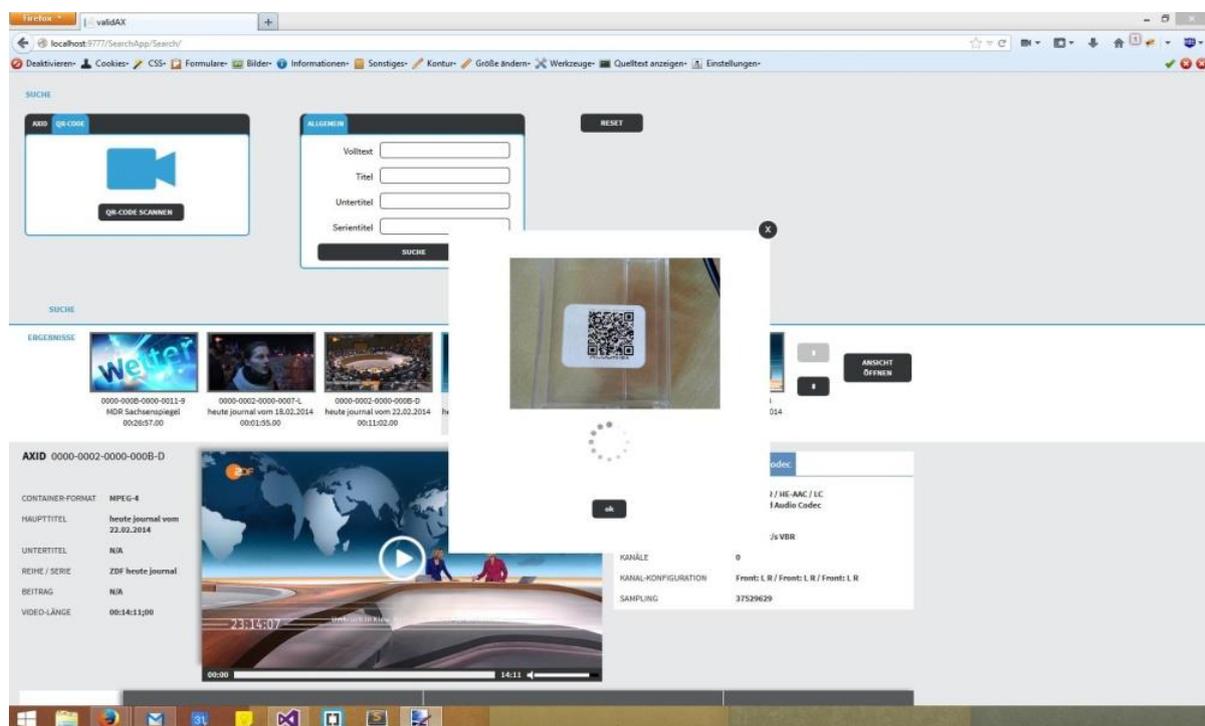


Abb. 11: Einlesen des QR_Code

Videosuche

Abb. 12 stellt die zwei wichtigsten Elemente der Suche vor. Im oberen Teil sind exemplarisch zwei Such-Widgets dargestellt, mit denen Archivare sowohl Known-Item Suchen mithilfe der AXID durchführen können, als auch komplexere Suchanfragen mithilfe verknüpfter Suchterme

stellen können. Im unteren Teil ist das dynamische Ergebnis-Gitter dargestellt. Dieses wird entsprechend der Suchanfrage automatisch aktualisiert und bietet einen grundlegenden Überblick über die Suchtreffer.

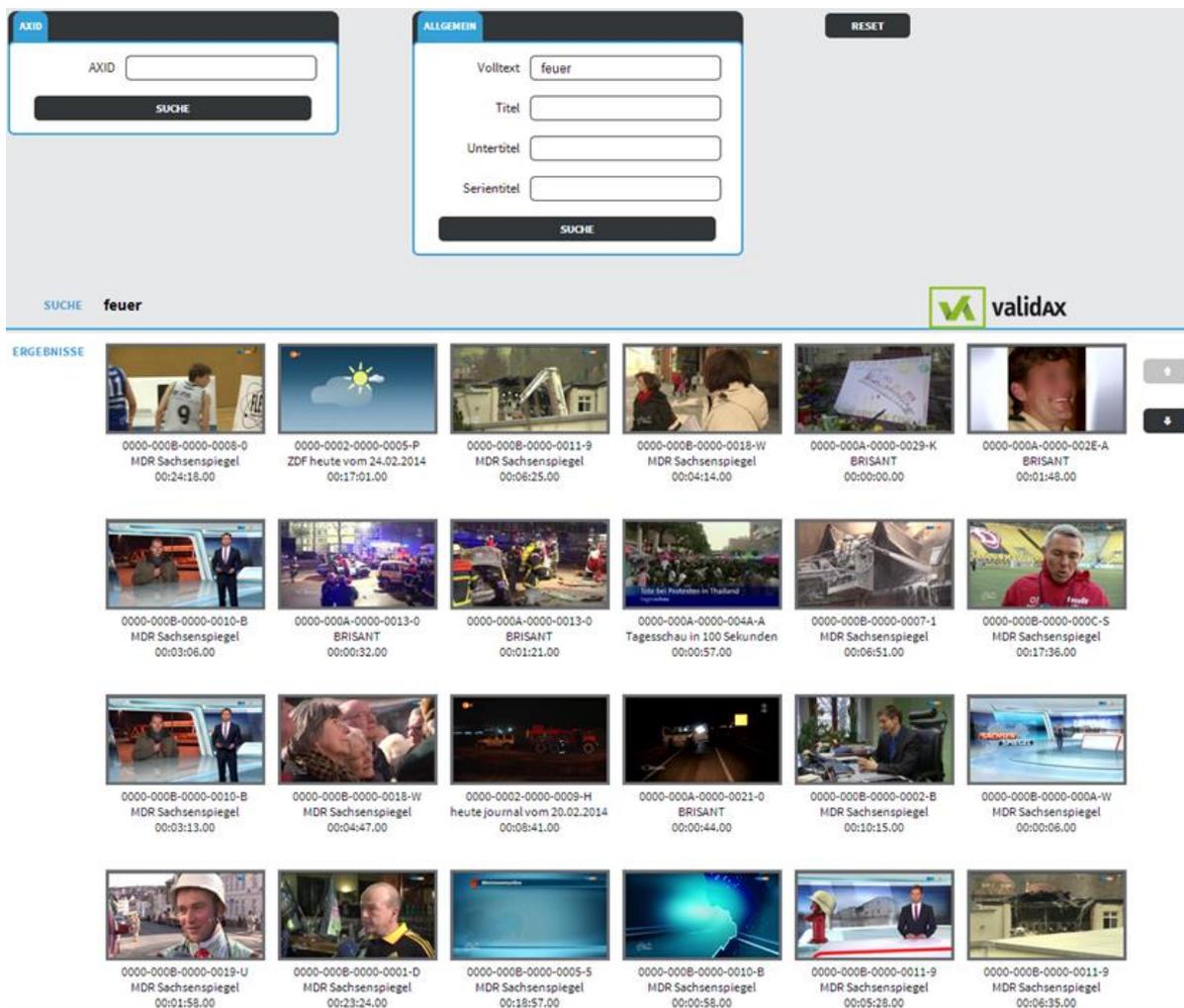


Abb. 12: Screenshot der Suche

Visualisierung zeitbasierter Metadaten

Ein grundlegender Überblick über alle Suchergebnisse ist in den seltensten Fällen ausreichend. Eine detaillierte Visualisierung der Suchtreffer in einem einzelnen Videobeitrag wird nach der Auswahl eines Videobeitrages aus dem Ergebnisgitter zugänglich. Diese ist in Abb. 13 dargestellt. Im oberen Teil befinden sich von links nach rechts drei Elemente. Ganz links die wichtigsten, das gesamte Videoband betreffenden Metadaten, mittig der Videobeitrag, rechts detaillierte Metadaten, die das gesamte Videoband betreffen. Im unteren Teil nimmt die Visualisierung der zeitbasierten Metadaten den größten Bildschirm-Bereich ein. Hier werden, parallel zur Zeitleiste unter dem Video die Ergebnisse der Videoanalyse in anwenderfreundlicher Art dargestellt. In Abb. 13 sind Schnittgrenzen und die Ergebnisse der Spracherkennung visualisiert.

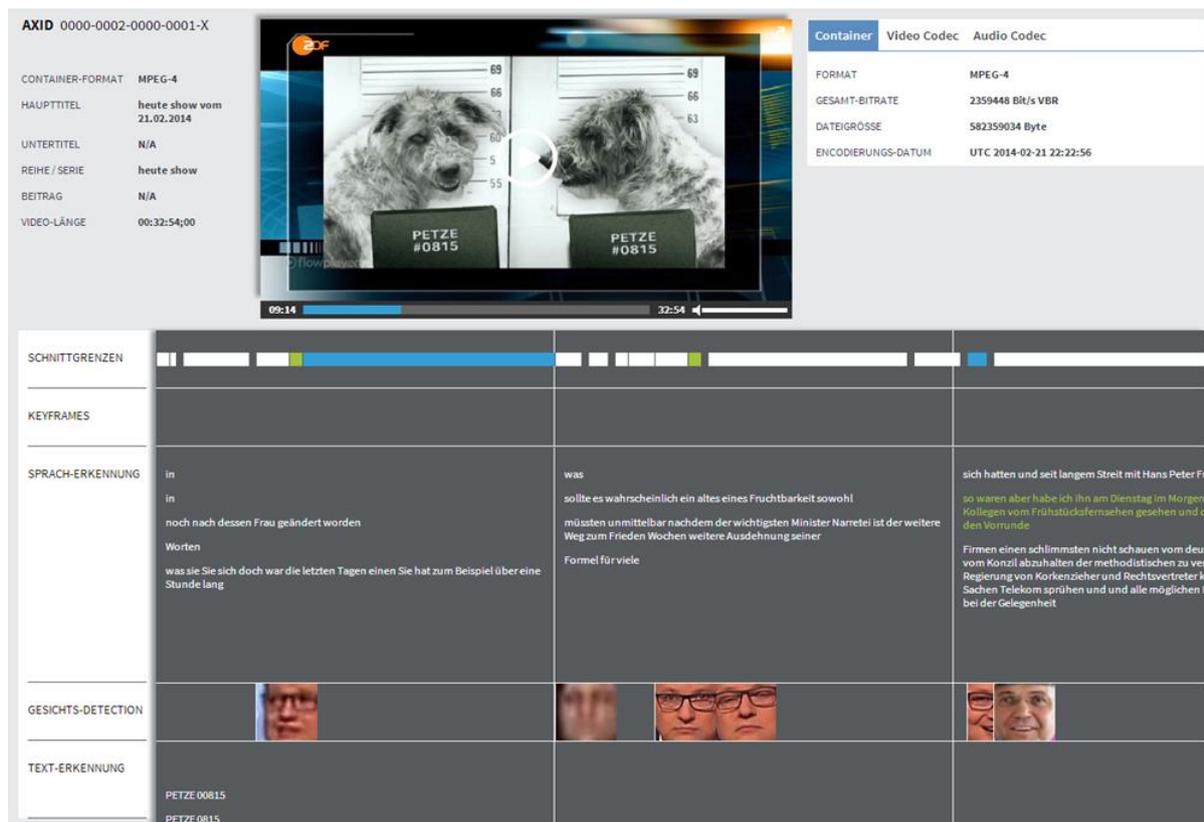


Abb. 13: Screenshot Visualisierung zeitbasierte Metadaten (Schnittgrenzen und Ergebnisse der Spracherkennung)

Diese Darstellung zeitbasierter Metadaten hat sich in qualitativen Nutzertests mit Videoredakteuren als geeignet erwiesen. So ist es möglich, sowohl personalisierte Suchanfrage-Widgets als auch entsprechende zeitbasierte Ergebniszeilen je nach Bedarf der Verwender zu implementieren.

Personalisierung und Adaption

Thundercloud wird in seiner ersten Ausbaustufe im Rahmen der Fachmesse CeBIT 2014 präsentiert um so Feedback und Anregungen durch weiter mögliche Nutzergruppen zu erhalten. So ist beispielsweise eine Umsetzung der Anwendung für einen interaktiven Multitouch Tisch prototypisch implementiert und evaluiert worden.

Szenario 2b: automatisierte Zugänge

Es wurde ferner eine Variante des Szenario 2 „Sendung 2.0“ in Richtung Automatisierung weiterentwickelt. Es profitiert direkt von den Ergebnissen der Demo-Plattform „Thundercloud“, der Xtrieval-Integration aus AP 2.5 sowie der Data-Warehousing Lösung „MetaBase“. Das Szenario wurde dabei in Form eines Plug-ins an die bestehende Plattform „Thundercloud“ angedockt.

Es ermöglicht alternativ zu der redaktionellen Recherche in den Analyse-Daten auch die automatische oder semi-automatische Zusammenstellung neuer Fernsehsendungen basieren auf den Ergebnissen der Suche. Ein hierfür entwickelter Algorithmus berechnet die passendste Kombination geeigneter Videos und erlaubt es diese neue Fernsehsendung zu einem Videoclip zu rendern, der wiederum der Datenbasis MetaBase hinzugefügt werden kann.

Szenario 3: Medizinische Medienverarbeitung

Im Laufe des Jahres 2013 verdichtete sich ein zusätzlicher Aspekt der Nutzung von AMOPA/XTRIEVAL: die Medizinische Medienverarbeitung. Hierzu liefen erste Tests für ein Szenario. Bereits im Vorjahr wurden im Rahmen einer studentischen Abschlussarbeit Überlegungen zur Spracherkennung in der Chirurgischen Praxis angestellt. Frau Christina Lohr (Studiengang Bachelor Angewandte Informatik mit Schwerpunkt Medieninformatik) hat dazu die Sprachmodelladaptation von CMU Sphinx, das auch für die Audioanalyse innerhalb von AMOPA genutzt wird, für den Einsatz in der Medizin untersucht. Ihre Untersuchung entstand im Rahmen einer Kooperation mit dem Klinikum Chemnitz. Sie erhielt für ihre Ergebnisse den Best Paper Award auf dem Studentensymposium Informatik der TU Chemnitz.

Daneben wurden gemeinsam mit der Juniorprofessur Visual Computing erste Untersuchungen zur medizinischen Bilderkennung, insbesondere bei Endoskopaufnahmen in den Nasennebenhöhlen vorgenommen.

Als zweiter Anwendungsbereich wurde die Möglichkeit der automatisierten Segmentierung von OCT-Scans (OCT – engl. Optical Coherence Tomography), wie sie bei der Untersuchung von Patienten mit altersbedingter Makuladegeneration (AMD) entstehen, untersucht. Dabei stand zunächst die Detektion, Visualisierung und Bewertung des Verlaufes des Retinalen Pigmentepithels (RPE), welches bei AMD Patienten teils erheblich deformiert ist, im Vordergrund. Aus technischer Sicht lag dabei der Fokus auf der Kombination vorhandener Tools und Verfahren und deren Adaption auf die Domäne der OCT-Scans der Netzhaut. Grundlegende Verarbeitungsketten und Prozesse konnten aus dem AMOPA Framework extrahiert und um semantischen Informationen sowie Plausibilitätskriterien erweitert für die Detektion des RPE und der Bestimmung des Schädigungsgrades dieser Netzhautschicht genutzt werden.

Aus den Arbeiten entstand eine Publikation (Kahl et al. 2014) sowie eine weiterführende Kooperation mit der Novartis AG (siehe Ausführungen dazu in Abschnitt 4. Voraussichtlicher Nutzen), die ein Pilotprojekt finanziert, welches untersucht, inwiefern die Segmentierung der Scans der Netzhaut durch Verfahren der Bildverarbeitung auch auf andere Gewebeschichten der Netzhaut ausgeweitet werden kann. Ziel ist eine zuverlässige Detektion der einzelnen, für den Bereich der AMD relevanten Schichten der Photorezeptoren. Die Visualisierung geschädigter Bereiche soll Mediziner bei der Diagnose und Wahl einer geeigneten Therapie unterstützen. In Kooperation mit der Novartis AG und dem Klinikum Chemnitz sollen hier im Bereich der Datenverarbeitung, Datenvisualisierung und semantischer Bildanalyse Maßstäbe auf dem Gebiet der medizinischen Bildverarbeitung gesetzt werden.

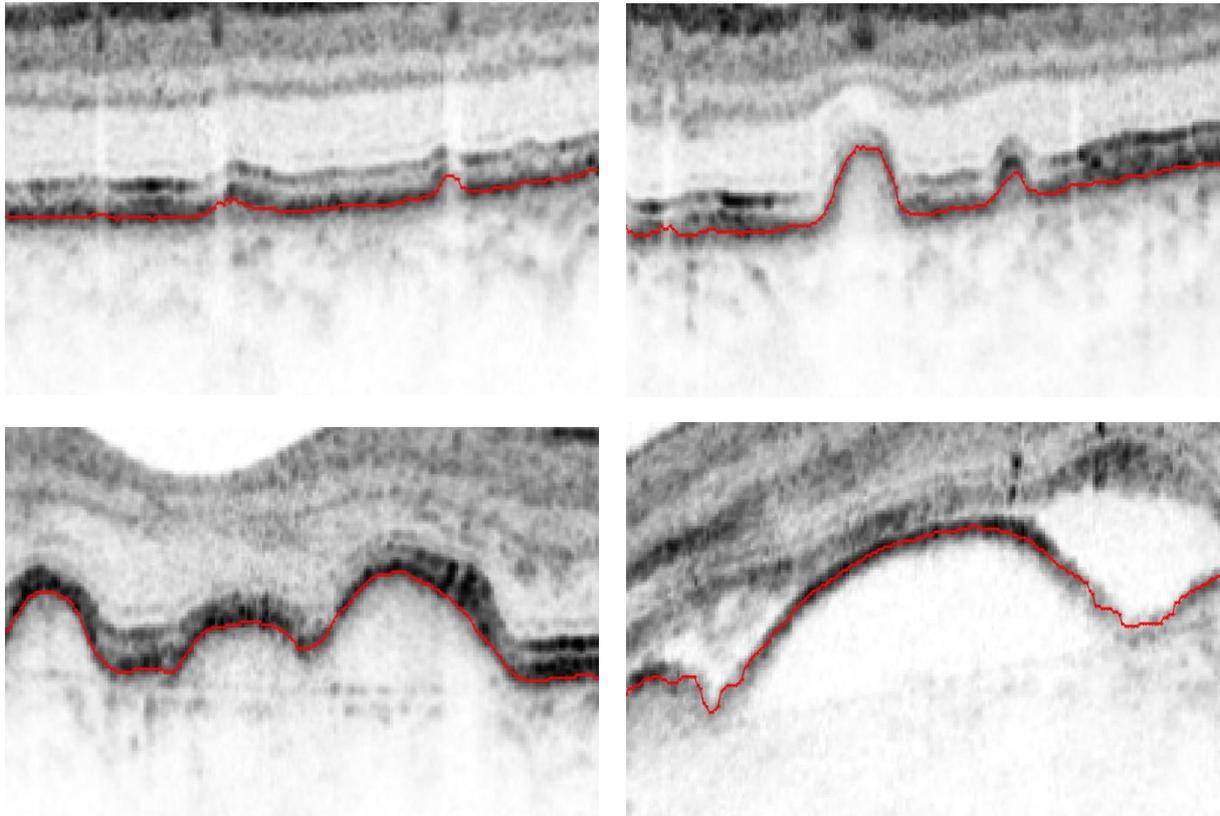


Abb. 14: Beispiele für die erfolgreiche Detektion des Verlaufs des RPE bei geringer Schädigung (obere Reihe) und massiver Deformation (untere Reihe). Auch die Abbildung des RPE-Verlaufs in kontrastarmen Bereichen ist möglich (Bildmaterial Universitätsklinikum Freiburg)

AB 4: Annotationsunterstützung

AP 4.1 Browser

Für das Annotationswerkzeug als browserbasierte Applikation wurden die Möglichkeiten zur Realisierung der notwendigen Annotationsschritte in Webbrowsern untersucht und eine darauf aufbauende, ergonomische Umsetzung konzipiert. Eine darauf aufbauende Realisierung wurde geplant und unter Einbeziehung ergonomischer Aspekte für die einfache und effiziente Benutzung durch Anwender optimiert. Die Umsetzung wurde im ersten Quartal 2012 und somit bis zum Erreichen von Meilenstein 1 abgeschlossen.

AP 4.2 Objektverfolgung

Eine bereits bestehende Szenenwechselerkennung zerlegt das Video in einzelne zeitlich- und räumlich kohärente Kameraeinstellungen, die als Schnitte bezeichnet werden. Aus diesen werden Schlüsselbilder ermittelt, die alle einem Schnitt zugehörigen Bilder geeignet repräsentieren. Der nachfolgende Prozess der Objektdetektion wird nur auf diesen repräsentativen Schlüsselbildern ausgeführt, was die Verarbeitungsgeschwindigkeit im Vergleich zur Bild-für-Bild-Analyse deutlich beschleunigt. Die detektierten Objekte werden nun in der direkten Nachbarschaft über ein Block-Matching-Verfahren vorwärts und rückwärts gerichtet im Videostrom verfolgt. Dazu werden nicht nur reine Helligkeitswerte, sondern auch Kanteninformationen und andere texturbasierte Merkmale genutzt, die beliebig erweiterbar sind.

AP 4.3 Integration Formalerfassung

Im Berichtszeitraum wurden die von Archivaren und Dokumentaren verwendeten Arten und Angaben zu Formaldaten der unterschiedlichen Kooperationspartner wie z.B. Kabeljournal und weiterer Quellen gesammelt. Anschließend erfolgte eine Sortierung, Gruppierung und Strukturierung der Daten unter Einbeziehung des Wissens und der Erfahrung der Kooperationspartner. Die Integration dieser Ergebnisse in das Annotationswerkzeug wurde unter Einbeziehung des Bereiches Workflowintegration begonnen und im zweiten Quartal 2013 abgeschlossen. Es steht nun ein Werkzeug zur Verfügung, dass die manuelle Eingabe sehr unterschiedlicher und auch komplexer (z.B. mehrfach hierarchisch strukturierter) Metadaten erlaubt und auf unterschiedliche Anwendungskorpora zugeschnitten werden kann.

AP 4.4 Text-Alignment

Text-Alignment stellte sich als ein wichtiges Mittel heraus, Videos sekundengenau recherchier- und navigierbar zu machen. Im Projekt wurde daher hier ein besonderer Fokus gelegt. Das Text-Alignment hilft bei der zeitlichen Zuordnung von Wörtern einer vorhandenen Transkription zu einem Video. Sind Transkriptionen vorhanden, wie zum Beispiel bei vorgelesenen Nachrichten, kann auf die automatische Spracherkennung verzichtet werden. Dies wurde im diesem Arbeitspaket auf zwei verschiedene Weisen gelöst. Zum ersten wird ein heuristischer Ansatz gewählt, der die Wörter in Silben unterteilt und anschließend zeitlich über das Video interpoliert. Der zweite Ansatz bezieht eine Phonemerkennung mit Hidden Markov Modellen ein. Somit ist eine Phonem-genaue zeitliche Zuordnung möglich. Die Anwendung beider Verfahren wird im folgenden Kapitel beschrieben.

Heuristische Methode zum Text-Alignment

Eine der einfachsten Methoden die Zeiten zu berechnen, wäre durch folgende Formel:

$$t_{\text{Wort}}(i) = \frac{T}{n} \cdot i$$

mit:

T die Gesamtzeit des Signals,

n die Anzahl der Wörter und

i der Index des Wortes, dessen Zeitmarke mit $t(i)$ berechnet werden soll.

Hier werden also die Zeitmarken gleichmäßig über die Gesamtdauer der Aufnahme verteilt. Als mögliche Erweiterung könnte man eine Start- und eine Endzeit festlegen, die angibt, ab wann und bis wann überhaupt gesprochen wird.

$$t_{\text{Wort}}(i) = \frac{T - (T_{\text{Start}} + T_{\text{Ende}})}{n} \cdot i + T_{\text{Start}}$$

T_{Start} und T_{Ende} werden durch Sprachdetektion berechnet.

Das Signal wird auf die Merkmale spektraler Flux, Rollover-Frequenz und Zero-Crossing-Rate hin untersucht. Eine Kombination dieser Merkmale liefert eine Aussage darüber, ob es sich in einem Signal um Sprache handelt oder nicht. So wird nun vom Anfang des zu analysierenden Videos das erste Vorkommen von Sprache gesucht und somit der Starzeitpunkt ermittelt.

Dasselbe geschieht mit dem Ende des Videos, um den Endzeitpunkt zu bestimmen. Der Bereich der Sprache wird somit eingeschränkt.

Obwohl das Ergebnis damit sicherlich genauer wird, bleiben unterschiedliche Sprechgeschwindigkeiten hier ebenfalls unberücksichtigt. Es wird davon ausgegangen, dass für jedes Wort die gleiche Zeit zur Aussprache benötigt wird und dass keine längeren Sprechpausen gemacht wurden.

Es ist ein weiterer Schritt implementiert worden, der Wörter und Zahlen in Silben zerlegt. Der Vorteil davon ist, dass lange Wörter mehr Zeit bekommen als kurze, so dass sich die Zeiten besser annähern.

Text-Alignment mit Hilfe von MAUS

Eine komplexere aber auch genauere Variante des Text-Alignment geschieht über die Integration des Münchener Automatischen Segmentationssystems (MAUS). Dieses dient unter anderem der Erkennung von Phonemen in einem Sprachsignal. Das Projekt wurde am Institut für Phonetik und Sprachverarbeitung der Ludwig-Maximilians-Universität München entwickelt und ist als Freeware-Paket für Linux erhältlich. Außerdem existiert ein MAUS-Webservice, der ähnliche Funktionen anbietet.

Grundsätzlich werden bei dieser Methode vier Phasen durchlaufen, wie in Abb. 15 skizziert ist. Im ersten Schritt wird eine Text-Normalisierung angewendet, bei der Satzzeichen aus der Texteingabe entfernt und Abkürzungen sowie Zahlen, wie Daten und Tageszeiten, in die ausgeschriebene Form umgewandelt werden. So wird '5' umgewandelt in 'fünf'. In der nächsten Phase wird der Text in Phoneme umgewandelt. Ein möglichst gutes Ergebnis dieser Umwandlung hängt von der Korrektheit bzw. Vollständigkeit des phonetischen Lexikons ab, das dabei genutzt wird. In der dritten Phase werden Aussprachevarianten erstellt. Dabei wird ein kreisfreier, gerichteter Graph erzeugt, der die Aussprachevarianten, nach ihrer Wahrscheinlichkeit gewichtet, enthält. In der letzten Phase wird, unter Anwendung des Viterbi-Algorithmus, das eingehende Sprachsignal zum am besten passenden Pfad des Graphen zeitlich ausgerichtet, wobei kontinuierliche Hidden-Markov-Modelle eingesetzt werden.

Implementierung des Frameworks

Vor dem Einsatz von MAUS müssen zunächst die Phasen eins und zwei als zusätzliche Vorarbeit erledigt werden, bevor das Programm genutzt werden kann.

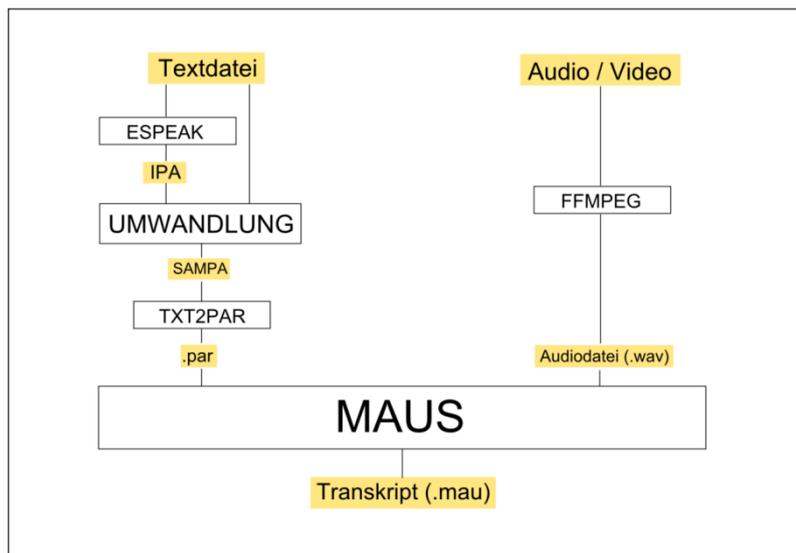


Abb. 15: Workflow bei Integration von MAUS

Für unser Ziel, den Worten einer manuell generierten Transkription Zeitmarken zuzuweisen, lässt sich demnach MAUS zwar nutzen, jedoch ist hier noch eine entsprechende Umwandlung der Eingabe erforderlich. MAUS kann verschiedene Formate verarbeiten, es benötigt jedoch immer sowohl die Wörter selbst als auch ihre phonetische Beschreibung in der SAMPA-Notation (Speech Assessment Methods Phonetic Alphabet). Zunächst muss die manuelle Transkription durch eine solche Beschreibung vervollständigt werden. Jedes Wort muss dazu entsprechend der deutschen Aussprache analysiert werden, was mithilfe des Programms „Espeak“ erledigt werden kann. Es entsteht ein Arbeitsablauf mit mehreren Zwischenschritten.

Mithilfe eines Python-Skriptes wurde dieser Ablauf automatisiert, so dass als Eingabe die Transkription sowie eine Audio- oder Videodatei mit dem gesprochenen Text genügen. Als Ausgabe entsteht eine Datei, die jedem Phonem eine Zeitmarke zuordnet. MAUS produziert also sogar etwas mehr als gebraucht wird, benötigt werden nur Zeitmarken für jedes Wort. Es reicht hier, nur die Zeitmarken des ersten Phonems eines jeden Wortes zu erfassen.

Zur Überprüfung der korrekten Arbeitsweise des Algorithmus gibt es die Möglichkeit, die Zeitmarken Eintrag für Eintrag manuell zu prüfen. Die Werte sind in Samples gegeben, sie sollten noch in Sekunden umgerechnet werden. Vor allem bei mehreren zu prüfenden Transkriptionen stellt die Überprüfung einen hohen Aufwand dar. Um diesen Vorgang zu vereinfachen und um wertvolle Zeit zu sparen, wurde eine einfache Webanwendung erstellt, in der die Video- bzw. Audiodatei abgespielt werden kann, wobei zusätzlich die gesprochenen Worte zur richtigen Zeit dargestellt werden können.

Arbeitsablauf

Grundvoraussetzung zur Benutzung des Pythonskriptes sind zwei Dateien: Eine Video- bzw. Audiodatei in einem Format, das von FFmpeg gelesen werden kann, und eine Textdatei, in der das gesprochene Wort, in der richtigen Reihenfolge gespeichert ist. Satzzeichen dürfen enthalten sein, die Wörter sollten durch Leerzeichen oder Zeilenumbrüche voneinander getrennt sein.

Aus der Textdatei werden nun die einzelnen Worte nacheinander, mithilfe von „Espeak“ in Lautschrift umgewandelt. Espeak unterstützt nur die Ausgabe in IPA, dem Internationalen Phonetischen Alphabet, da MAUS jedoch von einer Umschreibung im SAMPA-Format ausgeht, muss eine weitere Umwandlung stattfinden. Nun werden die Wörter aus der ursprünglichen Textdatei Zeilenweise getrennt und hinter jedem Wort die entsprechende Umschreibung in SAMPA notiert. Als Trennzeichen wird das Tabulatorzeichen genutzt. Mithilfe des Skriptes „txt2par“, das bereits in der MAUS-Sammlung enthalten ist, kann aus der entstandenen Textdatei nun die nötige Partiturddatei erzeugt werden. Aus der Video- oder Audiodatei wird durch Nutzung von Ffmpeg eine Wave-Datei mit nur einem Kanal erzeugt.

Kontrolle auf Korrektheit

Die erzeugte Partiturddatei enthält eine Liste von Phonemen in der Reihenfolge, in der sie im Audiosignal vorkommen. Außerdem wird zu jedem Phonem die Startzeit und die Länge in Samples gespeichert, ebenso der Index des Wortes zu dem das jeweilige Phonem gehört. Aus diesen Daten heraus ist es zunächst schwer zu erkennen, ob das Ergebnis zufriedenstellend ist, vor allem aufgrund der Zeitangabe in Samples, die zuallererst in Sekunden umgerechnet werden sollte. Aber selbst dann ist ein manueller Vergleich der Daten aufwändig.

MAUS-Ausgabe:
 Datei auswählen: ausgabe.mau
 MAU: 0 1919 -1
 <p:>
 MAU: 1920 1439 0
 n
 MAU: 3360 3359 0
 a:

Originaltranskript
 Datei auswählen: ausgabe_words.txt
 Nach
 den
 tagelangen
 massiven
 Protesten
 in

Audio/Videodatei
 Datei auswählen: input.webm
 LOAD

Proteste in Brasilien
Lebensbedingungen.

Nach den tagelangen massiven Protesten in Brasilien ist Staatspräsidentin Rousseff auf die Demonstranten zugegangen. Sie kündigte einen Dialog mit der Protestbewegung an Rousseff versprach schärfer gegen die Korruption vorzugehen und einen konkreten Plan für bessere **Lebensbedingungen**. Sie gehen auf die Straße, seit Wochen. Auch vergangene Nacht protestieren tausende Brasilianer gegen Korruption und soziale Missstände nun wendet sich die Präsidentin des Landes erstmals direkt an die Demonstranten im Fernsehen gibt sie ihnen Recht von ihrem Engagement lernen wir, dass wir

Abb. 16: Webanwendung stellt Beitrag mit Transkript dar (Beitrag aus den heute-Nachrichten ©ZDF): Das aktuell gesprochene Wort „Lebensbedingungen“ wird im Video angezeigt und im Transkript hervorgehoben. Die einzelnen Wörter des Transkripts sind anklickbar und erlauben eine wortgenaue Navigation.

Eine JavaScript-basierte Webanwendung erlaubt das Abspielen des Audio- oder Videosignals mit gleichzeitiger Darstellung des aktuell gesprochenen Wortes. Die Anwendung (s. Abb. 16) ist in jedem modernen Browser ausführbar. Es handelt sich um eine HTML-Datei, die mit JavaScript auskommt, es wird also kein Server benötigt. Als Eingabe der Anwendung kann direkt das von MAUS ausgegebene Format genutzt werden. Die komplette Transkription wird angezeigt, und durch einen Klick auf ein Wort der Transkription kann im Video zur

entsprechenden Stelle gesprungen werden. Somit muss nicht das gesamte Video zur Prüfung gesichtet werden. Diese Funktionen dienen einerseits zu Demonstrationszwecken, andererseits dazu, schnell und effektiv das Ergebnis von MAUS auf Korrektheit zu überprüfen.

AP 4.5: Videoplayer

Für die Demonstration einer integrierten Lösung, die alle Workflow-Komponenten verbindet, wurde der Prototyp "Thundercloud" entwickelt. Es handelt sich hierbei um eine Web-Applikation, in der die erreichten Ergebnisse und Lösungen des Projekts dargestellt werden

The screenshot displays the 'Thundercloud' demonstrator interface. At the top, there is a search bar with 'mdr' entered and the 'validax' logo. Below this, a row of video thumbnails is shown, each with a unique ID and 'MDR Sachsenpiegel' as the source. The main video player shows a scene with a crane and a building. To the right of the player, a table displays technical details: Container (MPEG-4), Video Codec (MPEG-4), Audio Codec (1730967 BR/s VBR), Dateigröße (355506013 Byte), and Encodierungs-Datum (UTC 2014-02-11 18:54:42). Below the player, there are sections for 'SCHNITTGRENZEN' (a timeline), 'KEYFRAMES' (a sequence of image thumbnails), and 'SPRACH-ERKENNUNG' (a text analysis grid). The text analysis grid contains several columns of text, including phrases like 'weiter ein tausend fünf hundert Arbeiter Teilabriß Ankara fünf a die', 'alle Zuschauer war der Porsche kann dem Jahr ein Einzelkind doch seit Deuter hat der Geländewagen einen kleinen Bruder denn es gibt Nachwuchs im Hause Porsche DMark Rahmen der Essener seit heute ebenfalls und rauschender den Leipzig zuhause und wie das bei Familien Zuwachs um mögliche ist wohl das natürlich Asphalt groß gefeiert und zwar auch weil mit der Erweiterung des Werks anderthalb tausend Soldaten stehen', 'knapp eine halbe Milliarde Euro a a', 'drei Jahren die Tausen so der Bundeswirtschaftsminister tausend fünf hundert neue industrielle Arbeitsplätze in Deutschland', 'für zahl achtzehn stehene achtzig Prozent dieser Arbeitsplätze seien an Menschen aus Mitteldeutschland vergeben worden wie eine große Suche nach abgewanderten Arbeitsfeld die vor zwei Jahren wurde der Bau der neuen Hallen für die Kassenrollen über Lackiererei begonnen hat', 'die Tausenden die Hessen wie soll im Jahr produziert werden', 'derzeit Entscheidungen treffen und der Zitaun mit die Unterstützung der Gestaltungen Freistaat Sachsen verstanden', and 'mindestens zwei'.

Abb. 17: Demonstrator "Thundercloud" in der Detailansicht eines Videos mit Videoplayer und Analyse-Ergebnissen

können und welche die prototypische Implementation der unter AB 3 entworfenen Anwendungsszenarien umsetzt.

Das User-Interface greift dabei auf die entwickelte Datenbank MetaBase (siehe AP 3, Szenario 2 zu Skalierbares Framework) zu, in welcher die Ergebnisse der audiovisuellen Analyse, als auch die Daten der Formalerfassung aggregiert sind. Damit bindet es mittelbar auch die Ingest-Straße an. Ferner bedient sich die Recherche- und Suchmaske des ebenfalls angebotenen Xtrieval-Framework für die Abwicklung der vom Nutzer gestellten Suchanfragen.

Der für AP 4.5 entwickelte Videoplayer kann auf diese Weise nicht nur das zu jedem analysierten Video vorliegende Vorschau-Proxy anzeigen, sondern auch die Ergebnisse der

verschiedenen Analyse-Komponenten zeitabhängig darstellen. So werden auf einer Zeitachse alle relevanten Daten, wie gefundene Gesichter oder gesprochene Sprache genau dort dargestellt, wo sie sich innerhalb des Videostroms ereignen.

Da diese Komponente des Demonstrators von zentraler Bedeutung für das User-Interface und alle damit verbunden Anwendungsfälle ist, wurde das Arbeitspaket 4.5 zeitlich vorgezogen und bereits vor Fertigstellung des AP 4.3 bearbeitet. Dies bietet zudem den Vorteil, dass der Prototyp schon im Frühjahr 2014 auf der CeBIT vorgestellt werden konnte und das dabei entstehende Feedback bis zum Ende des Projekts in der noch fehlenden User-Interfaces für die Formalerfassung einfließen konnte.

AB 5: Bilderkennung

AP 5.1 Aufbau Testkorpus SW, AP 5.2 SW-Bildverarbeitung und AP 5.3 Evaluation

Im Bereich Bilderkennung wurden zunächst Verfahren zur Bild- und Videoanalyse sowie die Ergebnisse der Analysen des AMOPA-Frameworks untersucht um geeignete Bestandteile für den Testkorpus zur Analyse von Schwarzweißmaterial bestimmen zu können. Dabei konnten Hinweise auf charakteristische Merkmale bestimmter Objektklassen, wie z.B. Gesichter, bestimmt werden, welche für die Umsetzung der nachfolgenden Arbeitsbereiche von Bedeutung sind.

Der Aufbau des Testkorpus selbst schritt etwas langsamer voran als geplant, da parallel die Umplanungen der Archivierungsstraße unterstützt wurde. Dennoch konnte der Testkorpus bis zum ersten Meilenstein wie geplant aufgebaut werden und stand dieser für die Extraktion von Merkmalen zur Objektidentifikation und entsprechender Deskriptoren zur Verfügung. Neben eigenen Entwicklungen wurden hierfür die in industriellen Systemen eingesetzten bzw. im wissenschaftlichen Forschungsumfeld gut untersuchten und beschriebenen Merkmale wie Textur, Geometrie, Struktur und Form auf ihre Eignung für dieses Projekt analysiert. Um die vielfältigen Einsatzmöglichkeiten der so ermittelten Verfahren und damit eine große Robustheit bezüglich des Videodateninhaltes zu gewährleisten, wurden diese ebenfalls auf weiteren frei verfügbaren Testkorpora wie z.B. dem INRIA-Datensatz, dem DaimlerChrysler Pedestrian Classification Benchmark Dataset, und dem TUD-Brussels-Datensatz am Beispiel der Fußgängerdetektion getestet. Zur Förderung der Nachhaltigkeit und der Kooperation im Bereich der Forschung an der TU Chemnitz wurden einige der verwendeten Verfahren im Rahmen einer gemeinschaftlichen Forschung mit der Professur Nachrichtentechnik für die Erkennung von Fußgängern angewandt. Erste daraus resultierende Ergebnisse wurden in (Ritter et al., 2012) als Konferenzbeitrag auf dem Forum Bildverarbeitung veröffentlicht, das als Schnittstelle der angewandten Bildverarbeitung zwischen Wirtschaft und Wissenschaft agiert. Die Verfahren zur Anwendung der Bilderkennung auf SW-Material wurden somit erfolgreich bis zum Meilenstein 2 implementiert und evaluiert.

Darüber hinaus wurde die Gesichtsdetektion auf den zweimal täglich ausgestrahlten 100 Sekunden langen Videosequenzen der zusammengefassten Tagesschau der letzten eineinhalb Jahre evaluiert. Diese wurden mit der Szenenwechselerkennung vorverarbeitet und die Detektion auf die so extrahierten 16 000 repräsentativen Schlüsselbilder beschränkt. Der hier konzipierte Gesichtsdetektor schneidet im Vergleich zu State-of-the-Art-Verfahren wie von (Viola & Jones, 2001) mindestens gleichwertig ab. Während der Viola & Jones basierte kaskadierte Haar-Merkmal-Detektor ca. 5.000 Gesichter korrekt auffindet und eine Falsch-

Alarm-Rate von 2.000 Patches aufweist, findet der angepasste Gesichtsdetektor 4.300 Gesichter bei einer deutlich geringeren Falsch-Alarm-Rate von nur 360 Bildausschnitten. Der Verlust der Trefferquote wird durch die drastische Reduktion der Falsch-Alarm-Rate mehr als egalisiert, da eine große Anzahl falscher Metadaten den Index schnell nachhaltig unbrauchbar macht.

AP 5.4: Aufbau Testkorpus

Um die in diesem AP angestrebte Reduzierung der Trainingsmenge zu erreichen wurde zunächst eine Test- und Beispieldatenmenge aufgebaut, anhand derer sich charakteristische Hauptmerkmale, z.B. von Objekten, in verschiedenen Arten von Bild- und Videomaterialien unterschiedlicher Herkunft, unterschiedlicher Aufnahmetechnologien und individueller Weiterverarbeitungsschritte weitestgehend extrahieren lassen. Zu diesem Zweck wurden Bild- und Videodaten sowohl aus unterschiedlichsten öffentlich zugänglichen Quellen (Tagesschau, Brisant, ...) bezogen als auch die beim Einspielen der Videos der Kooperationspartner anfallenden Daten herangezogen. Zusätzlich wurden hochqualitative und georeferenzierte Kamera- und Videoaufnahmen von ausgewählten Objekten und Gebäuden in Chemnitz sowie von ausgewählten Strecken im Chemnitzer Umland unter unterschiedlichen Bedingungen, wie z.B. Tageszeit, Wetter und Jahreszeit erstellt, um daraus allgemeine Merkmale für spezifische Objekte und Gebäude generieren zu können, die inhärent eine Reduzierung der Trainingsdatenmenge ermöglichen sollen.

Die notwendige Annotation der äußerst heterogenen Daten gestaltet sich als schwierig und äußerst zeitaufwendig. Diese Herausforderungen seien stellvertretend auf dem Datensatz der zusammengefassten Tagesschau in 100 Sekunden am Beispiel zur Gesichtsdetektion erläutert.

Datensatz	TS100
Ausstrahlungszeitraum	27.03.2011 bis 23.10.2012
Anzahl Videos	1.011
Anzahl Einzelbilder	2.581.100
Episodenlänge	Σ 28:40:44h; $\mu=102,12s$; $\sigma=13,31s$
Videocodec	H.264
Videoauflösung	512x288
Bildrate	25 fps
Datenvolumen	8.548 MB
Extrahierte Schnitte	16.165
Schnittlänge	$\mu=6,39s$; $\sigma=8,21s$
Datenvolumen Schlüsselbilder	306 MB (JPEG)

Tabelle 2:Übersicht zur Datenkollektion Tagesschau in 100 Sekunden, die von der Schnitterkennung aufbereitet wurde. (aus: Ritter 2014, S. 215)



Abb. 18: Auszug aus dem TS100 Datensatz mit menschengefüllten Szenen und Bauchbinden (aus: Ritter 2014, S. 216)

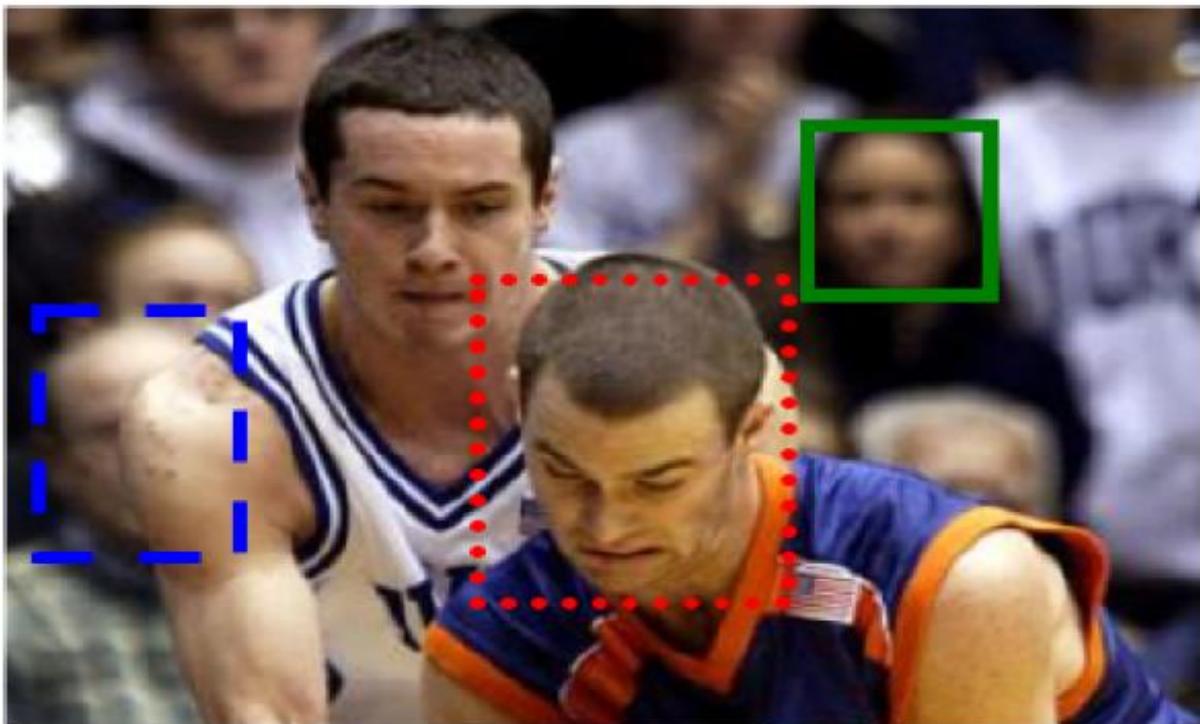


Abb. 19: "Challenges in face labeling. For some image regions, deciding whether or not it presents a 'face' can be challenging. Several factors such as low resolution (green, solid), occlusion (blue, dashed), and pose of the head (red, dotted) may make this determination ambiguous." (aus: Jain et al., 2010²)

Die Dissertation von Marc Ritter beschreibt den Korpus: "Die deutschlandweit bekannte Nachrichtensendung Tagesschau wird mehrmals am Tag mit aktuellen Meldungen auf den öffentlich-rechtlichen Fernsehsendern ausgestrahlt und von einem Millionenpublikum konsumiert. Da die einzelnen Episoden stark in ihrer Länge variieren, wird sich in der nachfolgenden Untersuchung auf inhaltlich zusammengefasste Versionen in Form der Tagesschau in 100 Sekunden (TS100) fokussiert, die stark heterogene aber brandaktuelle Themen enthalten, mehrmals am Tag aktualisiert werden und über Webcast frei verfügbar sind. Um aussagekräftige Analysen über eine große Vielfalt an unterschiedlichsten Themen zu gewährleisten, wurden diese Episoden lückenlos über einen Zeitraum von fast eineinhalb Jahren aufgenommen. Um wesentliche Aussagen zur Leistungsfähigkeit verschiedener Gesichtsklassifikatoren zu erhalten, wurde als Vorverarbeitungsschritt eine Schnitterkennung angewendet. Aus den 16:156 gefundenen Schnitten wurde jeweils ein repräsentatives Schlüsselbild ermittelt. Da diese möglichst scharf erscheinen sollen, wurden in einem naiven Ansatz die Bilder mit den geringsten Bewegungsänderungen aus jeder Bildsequenz ausgewählt. Eine detaillierte Übersicht über diesen Datensatz inklusive statistischer Informationen über die durchschnittliche Länge der Episoden und Schnitte beinhaltet (s. **Fehler! Verweisquelle konnte nicht gefunden werden.**). Ein optischer Eindruck über die starke inhaltliche Heterogenität der Themengebiete und die Bildqualität lässt sich (s. Abb. 18) entnehmen." (Ritter 2014, S.215ff4)

So unterliegt bereits die Annotation von Gesichtern in heterogenen und inhaltlich unbeschränkten Anwendungsdomänen keiner trivialen Vorgehensweise. Nach Jain et al. (2010, §5) (s. Abb. 19) ist sich beim Labeling von Gesichtern mit zahlreichen Mehrdeutigkeiten auseinanderzusetzen. Eine Methode dieser Problematik zu begegnen besteht in der Erstellung eines quantitativen Qualitätsmaßes für Gesichtsregionen, bei dem Regionen unterhalb eines Schwellwerts einfach für den Annotationsprozess abgelehnt werden. Trotzdem verbleibt die Konstruktion eines "satisfactory set of objective criteria" (ebd., S.4) ungelöst. Eine weitere Alternative stellt Crowd-Sourcing dar, bei der die gleichen Daten von mehreren Personen annotiert werden. Eine zu berücksichtigende Bedingung beim Annotationsprozess für Gesichter könnte beispielsweise lauten, dass beide Augen stets sichtbar im Bild auftreten müssen.

Der Datensatz TS100 enthält eine große Anzahl an Crowded Scenes (siehe Mahadevan et al. 2010³). Menschenansammlungen (s. Abb. 18 - Thematik: "Proteste im Jemen" und "Spitzenkandidatin") sowie größere Menschengruppen (ebd., Thematik "Lage in Libyen" und "Rösler kandidiert") stellen für allgemeingültige Beschränkungen oder Annotationsanweisungen in der Form "Markiere alle Gesichter, die mindestens 32 x 32 Pixel groß sind!" eine große Herausforderung dar. Hier tauchen oftmals viele Grenzfälle auf, bei denen die Entscheidung über die Erkennbarkeit eines Gesichts trotz unzureichender Auflösung, Verdeckung oder abgewandtem Blick wesentlich von der individuellen Auffassung des Betrachters abhängt. Die Identifizierung, Normung und allgemeingültige Disambiguierung solcher Grenzfälle kann hier nur an konkreten Beispielen erfolgen. Folgerichtig führen solche

36 ² Jain, Vidit ; Learned-Miller, Erik: Fddb: A Benchmark for Face Detection in Unconstrained Settings / University of Massachusetts, Amherst. 2010 (UM-CS-2010-009). Forschungsbericht, 11 S.

³ Mahadevan, Vijay ; Li, Weixin ; Bhalodia, Viral ; Vasconcelos, Nuno: Anomaly detection in crowded scenes. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, S. 1975-1981

selektiven Constraints zu einer thematischen Einschränkung, was der automatisierten Analyse und Annotation von beliebigen heterogenen Beiträgen im Projektrahmen entgegensteht. Bei der Applikation auf geboosteten Gesichtsklassifikatoren spielen beispielsweise auch Kompressionsartefakte keine untergeordnete Rolle mehr. (siehe Ritter 2014, S.217f4)

Die enorme Größe der Datenbestände, die beschriebenen Problemstellungen und auftretende Mehrdeutigkeiten sowie eingeschränkte Ressourcen verhindern eine auch nur annähernde vollständige Annotation des Ground Truth. Daher wird im nachfolgenden Arbeitspaket vordergründig eine Optimierung der Precision für ausgewählte Beispielklassen durchgeführt.

AP 5.5: Reduktion & Evaluation

Dieses Arbeitspaket teilt sich auf diesen und den nachfolgenden Berichtszeitraum auf. Gesichter treten in den vorliegenden Videos am häufigsten auf. Daher beschäftigt sich das Arbeitspaket in vorliegenden Berichtszeitraum damit, die bereits in AMOPA integrierte Technologie der Gesichtsdetektoren zu nutzen und den langwierigen Trainingsprozess durch eine entsprechende Reduktion der Trainingsdaten weiter zu verkürzen. Ein Einblick in die Charakteristik des trainierten Klassifikators TUC_FD (rechte Spalte) ist in Tabelle 3 gegeben. Gegenüber konventioneller Fachliteratur (linke Spalte) benutzt dieser eine geringere Auflösung (1. Zeile) und einen um Faktor 8 reduzierten Trainingsdatensatz mit Erweiterten Haar-Merkmalen in einer 29-stufigen Kaskade mit einer kumulierten Anzahl von 3.922 aggregierten schwachen Klassifikatoren. Eine wesentliche Verfahrensmodifikation besteht in der Einführung von harten Lernbedingungen, die dem trainierten Detektor in jeder Kaskadenstufe auferlegen, alle Gesichter des Trainingsdatensatzes korrekt zu erlernen, womit im Umkehrschluss keine Fehler gemacht werden, was wiederum der eigentlichen Lernidee augenscheinlich konträr gegenüberzustehen scheint.

Bezeichnung	Viola&Jones	TUC_FD
Auflösung	24 x 24 Pixel	20 x 20 Pixel
Trainingsdaten (Gesicht : Nicht-Gesicht)	9.832 : 10.000	1.200 : 1.200
Testdaten (Gesicht : Nicht-Gesicht)	k.A.	1.200 : 1.200
Merkmalstyp	Haar-Merkmale	Erweiterte Haar-Merkmale
# Merkmale	180.000+	122.364
Kaskadentiefe	38	29
# Merkmale des Detektors	6.061	3.922
Besonderheiten	Spezieller Detektor mit zwei Merkmalen für Stufe 1	Beschneidung ab Stufe 27 bei globalem Minimum des Trainingsfehlers

Tabelle 3: Gegenüberstellung wesentlicher Eigenschaften des trainierten und datenreduzierten Detektors der Technischen Universität Chemnitz (TUC_FD) mit dem klassischen Gesichtsdetektor von Viola & Jones (links). (aus: Ritter 2014, S.204)

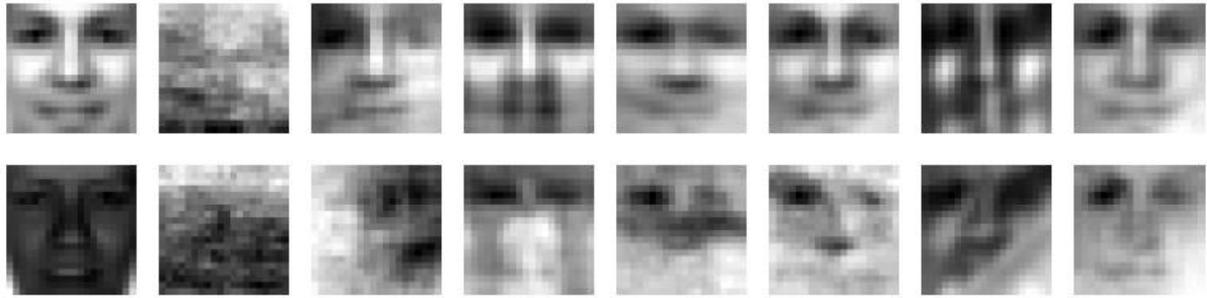


Abb. 20: Entwicklung von Mittelwert (1. Reihe) und Standardabweichung (2. Reihe), gemittelt über 1.200 Bilder der positiven Klasse (1.v.l.) sowie der negativen Klasse mit zunehmender Kaskadenlänge, wobei die Stufen 1,5,10,15,20,25 und 29 (v.l.n.r.) abgebildet

Den Lernfortschritt des datenreduzierten und unter harten Lernbedingungen trainierten Detektors sowie die sinkende Diskriminierbarkeit zwischen den beiden Klassen der Gesichter und Nicht-Gesichter illustriert Abbildung 21. Die jeweiligen Samples werden durch den sukzessiven Verlauf des Bootstrap-Prozesses vor jeder Kaskadenstufe gewonnen und zeigen, dass sich die beiden Klassen zunehmend annähern. Durch diese zunehmende Ähnlichkeit verkleinert sich der Lernfortschritt des maschinellen Lernverfahrens stetig und stagniert schließlich, womit das Risiko eines Übertrainierens im Gegenzug steigt. Dabei werden positive Detektionen in der späteren Anwendung fälschlicherweise aussortiert, weshalb innerhalb einer ersten Testreihe Detektoren mit unterschiedlichen Anzahl an Kaskadenstufen erzeugt und auf Testkollektionen vorab getestet wurden. Gemäß dieser empirischen Voruntersuchungen auf Datenmaterialien der lokalen und öffentlich-rechtlichen Fernsehsender sowie verschiedener datenreduzierter Klassifikatoren erwies sich der dargestellte Detektor als leistungsstärkster.

Um den Eindruck der Leistungsstärke des datenreduzierten Klassifikators im realen Testumfeld zu validieren, wurde der TS100-Datensatz aus dem vorherigen Arbeitspaket herangezogen und auf den durch die Schnitterkennung extrahierten 16.195 Schlüsselbildern gegenüber fünf frei verfügbaren State-of-the-Art-Gesichtsdetektoren (siehe Tabelle 4) unterschiedlicher Technologiestände verglichen (siehe Tabelle 5). Tabelle 5 verdeutlicht, dass der datenreduzierte Klassifikator weitläufig besser abschneidet und seine Mitkonkurrenten dominiert und nur dem in der Literatur verbesserten und aufwendigeren Technologie des kaskadierten Entscheidungsbaum mit Stufenklassifikatoren (Haar_tree) unterliegt, was darauf hinweist, dass die Datenreduktion in Kombination mit den harten Lernbedingungen erfolgreich durchgeführt wurde.

Detektor ID	Methode	Detektorgröße	Stufen	Schwache Klassifikatoren
Haar_default	Kaskadierter Entscheidungsstumpf mit diskretem AdaBoost	24 x 24	25	2.913
Haar_alt	Kaskadierter Entscheidungsstumpf mit Gentle Boost	20 x 20	22	2.135
Haar_alt2	Entscheidungsbaum mit zwei Knoten und Gentle AdaBoost	20 x 20	20	2.094
Haar_tree	Entscheidungsstumpf mit Gentle AdaBoost mit einem Baum von Stufenklassifikatoren	20 x 20	47	8.468
LBP	Kaskadierter Regressionsbaum mit Gentle AdaBoost und LBP	24 x 24	20	139
TUC_FD	Kaskadierter Entscheidungsstumpf mit Gentle Adaboost	24 x 24	29	3.922

Tabella 4: Überblick unterschiedlicher State-of-the-Art-Detektoren zur frontalen Gesichtsdetektion. Die oberen fünf Detektoren sind in OpenCV enthalten und nutzen Haar-Merkmale oder Lokale Binäre Muster (LBP). TUC_FD bezeichnet den datenreduzierten Detektor der Technischen Universität Chemnitz, der in dieser Arbeit entwickelt wurde. (aus: Ritter 2014, S.204)

Detektor ID	Skalierung	Detektionen	Korrekte Detektionen	Falsche Detektionen	Präzision (%)	Detektionsgröße in Pixel (μ ; σ)	\emptyset Laufzeit pro Bild (ms)
Haar_default	1,1	8.608	5.387	3.221	62,6	82,4; 35,0	117,5
	1,3	5.717	4.836	881	84,6	87,4; 34,7	59,1
Haar_alt	1,1	5.711	5.006	705	87,7	84,0; 31,9	119,6
	1,3	4.804	4.677	127	97,4	88,3; 31,7	61,1
Haar_alt2	1,1	6.556	5.287	1.269	80,6	80,1; 34,0	111,9
	1,3	5.147	4.990	247	95,2	90,2; 32,9	57,3
Haar_tree	1,1	4.317	3.991	326	92,4	82,7; 30,9	149,1
	1,3	3.557	3.485	72	98,0	87,4; 30,6	68,1
LBP	1,1	6.856	4.554	2.302	66,4	81,4; 29,8	55,3
	1,3	3.940	3.610	310	91,6	78,5; 33,8	36,4
TUC_FD	1,1	4.173	3.780	393	90,5	71,9; 36,8	2.686,9
	1,3	2.682	2.560	122	95,5	71,5; 33,1	1.130,0

Tabella 5: Vergleich der Ergebnisse verschiedener State-of-the-Art-Klassifikatoren zur frontalen Gesichtsdetektion auf den 16:156 repräsentativen Schlüsselbildern des TS100 Datensatzes. (aus: Ritter 2014, S.221)

AB 6: Web-Services

AP 6.1: Konzeption

Es wurde ein Konzept entwickelt, welches flexibel erweiterbar ist. Im Rahmen des Konzeptes gibt es verschiedene Bereiche, die sich an die Arbeitspakete anlehnen. So ist vorgesehen das die AV-Medien über verschiedene Kanäle zugeführt werden können. Die Analyse kann für jeden Mandanten getrennt konfiguriert und zusammengestellt werden. Der Status der Zuführung und der Fortschritt der Analyse sollen über den Web-Service kommuniziert werden.

AP 6.2: Zuführung AV-Medien

Als ein erster und einfacher Übertragungskanal wurde FTP gewählt. Mit FTP können Dateien unterschiedlichen Typs und unterschiedlicher Größe an einen Server übertragen werden. Der Server nimmt diese entgegen und löst automatisch die Analyse mittels AMOPA aus.

AP 6.3: Abrufen Metadaten

Der Web-Service umfasst eine Komponente, mit dessen Hilfe die Metadaten abgerufen werden können. Dabei hat jeder Mandant einen eigenen Bereich, der Mittels Web-Technologien wie HTTPS, Passwörtern im Rahmen der HTTP Authentifizierung und Client-seitigen Zertifikaten geschützt werden kann. Die Abgerufenen Metadaten liegen in einem XML-Format vor und entsprechenden Daten, die AMOPA für generiert hat. Mit diesen Daten könnten die Mandanten eigene Retrieval-Funktionen konstruieren.

AP 6.4: Retrieval

Die Daten aus AMOPA werden mit Hilfe des Xtrieval Frameworks nun in einen neuen Web-Service integriert. Dafür wurde neben den bereits bestehenden Verbindungen zwischen dem Xtrieval Framework und den Retrieval Frameworks Apache Lucene und Terrier nun eine neue Verknüpfung mit Apache Solr hinzugefügt. Apache Solr ist ein solcher Web-Service, der zwar auch auf Basis von Apache Lucene arbeitet, aber darüber hinaus verschiedene Web-Schnittstellen zur Verfügung stellt. Für Apache Solr gibt es bereits diverse Web-Oberfläche um den darin indizierten Datenbestand zu durchsuchen, jedoch sind für den Bereich AV-Medien weitere Anpassungen notwendig. Allgemein lässt sich aber sagen, dass dadurch neue Benutzeroberflächen einfacher auf die indizierten Daten zugreifen können.

AP 6.5: Abrufen AV-Medien

Für den Abruf von AV-Medien mit Hilfe des Web Service wurde eine Streaming-Lösung entwickelt. Der Web-Service besitzt nun die Fähigkeit, Daten mit Hilfe des Content-Range Headers aus dem HTTP-Standard als Teile einer Datei zu übertragen. Moderne Browser mit der Unterstützung für Codecs wie H.264 oder WebM können so weitgehend frei einzelne Teile eines Videos herunterladen und gezielt in einem Video hin und her springen.

Der Informationsaustausch über das World Wide Web (WWW) ermöglicht sowohl die Kommunikation mit anderen Webservices für den Datenaustausch als auch die Gewinnung von zusätzlichen Metadaten durch die Adressierung geeigneter Ressourcen. Wird nach der automatischen Spracherkennung bspw. die Entität „Brandenburger Tor“ detektiert, so können verwandte Begriffe wie „Deutschland“, „Berlin“, „Unter den Linden“ und „Mauer“ gewonnen werden, welche einen sehr positiven Einfluss auf die Recherchierbarkeit in Archiven haben. Darüber hinaus können analysierte Medien durch weitere Inhalte des Webs (bspw. Google

Maps, Wikipedia, Flickr, etc.) angereichert und in geeigneter Weise visuell repräsentiert werden.

Voraussetzung für die Extraktion einer Entität des entsprechenden Transkriptes der Audiospur, ist ein umfassender Wortschatz der Spracherkennung. Um dies zu garantieren, müssen Sprachmodelle auf situative Gegebenheiten adaptiert werden. Die Hypothesen einer Spracherkennung eignen sich weniger als Adaptionen, da die Transkripte Erkennungsfehler unterliegen können und Vokabular außerhalb der Wissensbasis der Spracherkennung nicht abgedeckt wird. Das Web als Informationsquelle bietet hierzu einen enormen Umfang an Adaptionen für die Schätzung von Sprachmodellen. Da die Aufbereitung spezifischer Adaptionen mit sehr viel Aufwand verbunden ist, wurde eine Methode zur unüberwachten Sprachmodelladaptation entwickelt: Die Spracherkennung erzeugt mittels eines generischen Sprachmodells in einem ersten Dekodierschritt ein Sprachtranskript der Audiospur. Aus diesem Transkript werden dann Stichwörter bzw. Entitäten extrahiert, um diese als Suchanfragen im Web gezielt zu verwenden, wobei die zurückgelieferten Dokumente (Artikel, Feeds, etc.) nach Relevanz und Zeit priorisiert werden. Alle Dokumente werden schließlich zu einem Adaptionenkorpus zusammengefasst und normalisiert, um den Anforderungen für die Weiterverarbeitung zu entsprechen. Der akkumulierte Korpus dient zum Training eines spezifischen Sprachmodells, welches verwendet wird, um das generische Basismodell mittels einer linearen Interpolation zu adaptieren. Ebenfalls erfolgt die Erweiterung des Aussprachewörterbuchs durch das zusätzlich aufkommende Vokabular, welches mittels eines Graphem-zu-Phonem Prozessors automatisiert phonetisch beschrieben wird. Nach der Adaptionenphase wird das adaptierte Wörterbuch und adaptierte Sprachmodell in einem zweiten Dekodierschritt der automatischen Spracherkennung angewendet, die somit bzgl. Inhalt und Vokabular justiert ist.

Diese Methode wurde durch zwei Experimente evaluiert. Zum einen kann der Erfolg im Spoken Document Retrieval nachgewiesen werden, bei der die Mean Average Precision um 11,7% reduziert werden konnte. Zum anderen wurde eine Verbesserung des Media Enrichment (Anreicherung mit zusätzlichen Inhalten) um 12%, basierend auf dem Fehlermaß der extrahierten Entitäten, ermittelt.

[AB 7: Parallelverarbeitung](#)

AP 7.1: Konzeption

Das AMOPA-Framework erlaubt durch seine flexible multi-threading Architektur die vollständige Auslastung darunter liegender Multikern-Rechensysteme. Eine weitere Möglichkeit, die Verarbeitungsgeschwindigkeit weiter zu optimieren, besteht darin, die Algorithmen bzw. die Ausführung der Berechnung von mehreren Videos gleichzeitig über mehrere Rechner innerhalb eines Rechenverbundes zu verteilen. Das kann im Analysecluster per schnellem Infiniband von effektiv 5 Gb/s oder in beliebigen Netzwerken über TCP/IP mit 1 Gb/s erfolgen.

Zudem bietet die Ausnutzung modernster Graphikkarten (GPU) einen immer weiter verbreiteten Ansatz, um Algorithmen durch massive Parallelisierung zu beschleunigen. Moderne Graphikkarten erreichen heutzutage gegenüber der aktuellen Prozessorgeneration

bei wissenschaftlichen Berechnungen mit großen Matrizen und Gleichungssystemen in Abhängigkeit vom zugrundeliegenden Algorithmus eine Beschleunigung um bis zu Faktor 10.

Für alle Komponenten des Xtrieval Frameworks wurde die Thread-Sicherheit hergestellt. Das heißt im speziellen Fall, dass mehrere Verbraucher gleichzeitig eine Komponente nutzen können, ohne dass es dabei zu Dopplungen oder anderen Inkonsistenzen kommt, die sonst mit einer parallelen Verarbeitung in Verbindung gebracht werden. Die Verarbeitungsgeschwindigkeit hängt dadurch hauptsächlich von der Anzahl der Verbraucher (z.B. dem Indizierer) ab und kann darüber skaliert werden.

AP 7.2: Realisierung AMOPA

Zuerst soll der Ansatz mit internen durch Graphikkarten gestützten Beschleunigung genauer eruiert werden. Dazu gehört die Evaluation unterschiedlicher Programmierarchitekturen für die Grafikprozessoren (GPGPU) wie CUDA, OpenCL, OpenGL oder DirectCompute hinsichtlich Geschwindigkeit, Konvertierungsfähigkeit und Portabilität. In Testszenarien befinden sich die Algorithmen zur Gesichtsdetektion und zur Szenenwechselerkennung in der Umsetzung. Genaue Ergebnisse werden mit Abschluss des Arbeitspaketes im nächsten Berichtszeitraum erwartet.

AP 7.2: Realisierung AMOPA

Aus den Vorbereitungen zum Arbeitspaket aus dem vorhergehenden Berichtszeitraum, wurde sich für eine Implementierung mittels OpenCL entschieden, die sich in der besonders universellen und flexiblen Einsatz- und Nutzbarkeit verschiedenster Prozessoren in unterschiedlichsten Hardware-Konfigurationen begründet. Um eine prinzipielle Anwendbarkeit zu gewährleisten, war es notwendig, das bestehende AMOPA-Framework um Transferroutinen zu erweitern, die einen Datenaustausch für unterschiedliche Objekt- und Datentypen erlauben.

Aufgrund der relativ kurzen Realisierungsphase wurde sich für eine Implementierung der Farbraumkonvertierungsroutinen und von Kantenhistogrammen entschieden, die sowohl für die Szenenwechselerkennung als auch bei der Objektdetektion in der Prozesskette von AP 5 relevant sind.

AP 7.3: Evaluation AMOPA

Bei Nutzung der Grafikprozessoren über OpenCL (CUDA etc.) ist besonders zu beachten, dass die beschleunigte Berechnung des eigentlichen Algorithmusteils die Zeitdauer des Datentransfers zwischen Arbeitsspeicher und Grafikprozessor mindestens aufwiegen muss, um einen tatsächlichen Vorteil zu erzielen. Es konnte nachgewiesen werden, dass die Berechnung von einzelnen Helligkeitswerten innerhalb der Farbraumkonvertierungsmethoden um bis zu Faktor 5 beschleunigt werden konnte, wobei die Beschleunigung des gesamten Verfahrens von der zugrundeliegenden Bildgröße und dem verfügbaren Grafikkartenspeicher abhängig ist. Experimentelle Versuche auf der zugrundeliegenden Hardware des Analyseclusters wiesen eine Beschleunigung inklusive aller Datentransferroutinen in PAL-Auflösung von etwa Faktor 2 auf.

Um zukünftig auch eine verteilte Analyse mit Hilfe des AMOPA-Frameworks auf mehreren Rechnern zu ermöglichen, wurden verschiedene Lösungen untersucht und miteinander verglichen. Eine Grundlage bilden moderne XML-Datenbanken. Ein Vergleich dieser wurde als

Beitrag bei der Konferenz LWA 2013 dem wissenschaftlichen Publikum präsentiert (siehe Neumerkel & Manthey (2013)).

AP 7.4: Realisierung, Konzept für Xtrieval und mehrere Verbraucher

Die Bearbeitung von AP 7.2 und AP 7.3 nahm mehr Zeit in Anspruch als ursprünglich vorgesehen. Gleichzeitig erwies sich die Parallelisierung von Xtrieval als weniger aufwändig als gedacht, da auf die Erfahrungen aus AP7.2 und AP 7.3 zurückgegriffen werden konnte. Für die Anforderung der Praxisanwendung erkannten wir die Notwendigkeit der Berücksichtigung mehrerer Verbraucher.

Die Möglichkeit zur parallelen Nutzung der Komponenten hängt maßgeblich von der Anzahl der gleichzeitig arbeitenden Verbraucher ab. Deshalb wurde ein Konzept für die Implementierung dieser Verbraucher erstellt und in ersten Tests überprüft. Dabei sind jedoch bei verschiedenen Komponenten, die auf externe Bibliotheken zugreifen, Probleme aufgetreten. Hier musste die parallele Verarbeitung noch einmal nachgebessert werden. Die Identifizierung der betroffenen Komponenten hat sich als komplex erwiesen, da manche Fehler bei der parallelen Verarbeitung nicht deterministisch provoziert werden können und eine Vielzahl an Versuchen notwendig war.

2. Wichtigste Positionen des zahlenmäßigen Nachweises

Auf Grundlage der relativen Anteile an der bewilligten Fördersumme für das Vorhaben validAX (s. Abb. 24) werden nachfolgend die wesentlichen Positionen aufgeführt. Die Illustration verdeutlicht, dass 84,6% der Gesamtsumme in den Positionen 0812 und 0850 verausgabt wurden. Mit knapp 72,4% hat die Position 0812 (wissenschaftliches Personal) den größten Anteil.

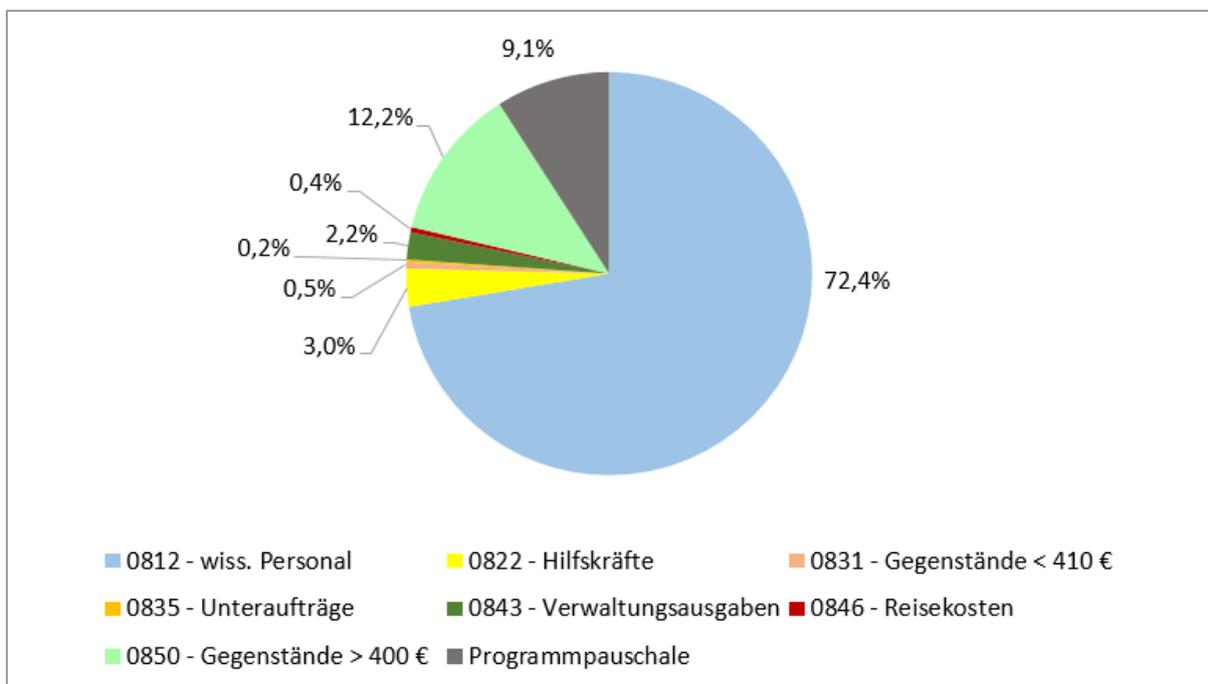


Abb. 21: Relative Verteilung der für das Vorhaben validAX bewilligten Positionen

Die Gegenüberstellung der tatsächlichen Ausgaben in den bewilligten Positionen (s. Abb. 22) zeigt die Abweichungen der tatsächlichen Ausgaben von den bewilligten:

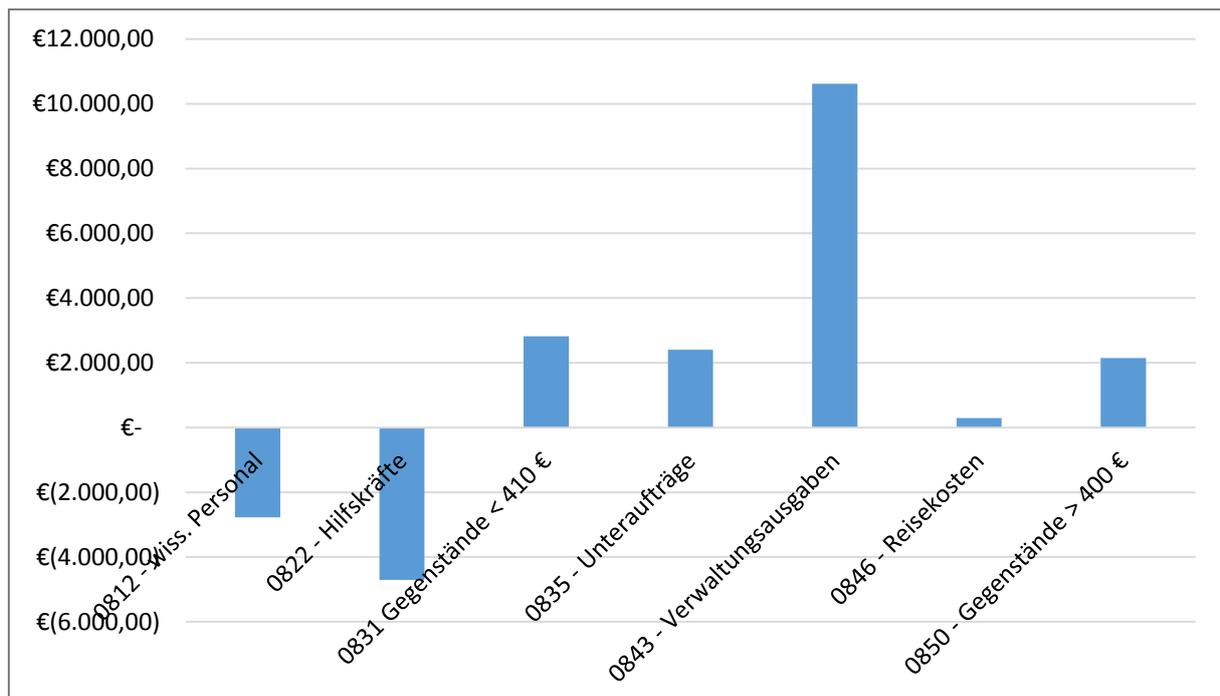


Abb. 22: Nicht verausgabte Mittel und zuviel verausgabte Mittel über die einzelnen Posten

In den Posten 0812 (Wissenschaftliches Personal) und 0822 (Hilfskräfte) wurden insgesamt 7.483,46 € mehr benötigt als ursprünglich veranschlagt. Dies liegt zum einen an Gehaltserhöhungen für Angestellte in den Jahren 2013 und 2014 sowie der Einstellung erfahrener Mitarbeiter mit einer entsprechend hohen Stufe. Zum anderen wurden mehr wissenschaftliche Hilfskräfte eingestellt, um die aufwändige Technik zu betreuen und die Mitarbeiter zu entlasten.

In den übrigen Posten wurde dafür insgesamt 18.280,77 € eingespart. Den größten Posten machen hier mit über 10T€ die Verwaltungsausgaben aus, die zum Großteil nicht benötigt wurden.

Die Position 0850 (Investitionen für Gegenstände >400 EUR) bildete mit einem Anteil von knapp 12,2% die zweitgrößte Ausgabeart. Aufgrund des Tsunamis in Japan im Frühjahr 2011, der Flutkatastrophe in Thailand im Herbst 2011 und dem begrenzten Umfang des Marktes für die vorgesehenen Geräte, ohne nennenswerte Lagermengen, entstanden erhebliche Lieferprobleme bei der Beschaffung, welche unter diesen Bedingungen nicht fristgerecht bzw. bei manchen Geräten gar nicht möglich war. Folglich mussten zeitintensive Umplanungen in der Bauweise der Archivierungsstraße durchgeführt und auf alternative Geräte ausgewichen werden. Dies verschob die finanziellen Planungen und führte dazu, dass die Beschaffung dieser Geräte z.T. erst Anfang 2012 vorgenommen wurde und sich der Aufbau der Archivierungsstraße verzögerte.

Bei der Konkretisierung des Bedarfs und der Preise im Rahmen der tatsächlichen Anschaffung der Geräte stellte sich heraus, dass eine Neukalkulation notwendig wurde. Nachfolgend wird kurz auf die wichtigsten Gegenstände und die veränderte Kalkulation eingegangen, die im Vorhaben validAX genutzt und unter dieser Position bewilligt wurden:

Nr.	Bezeichnung	Anzahl	Einzelpreis, Plan €	Gesamtpreis, Plan €	Gesamtpreis, tatsächlich €	Differenz €
1	Sony J-30 / SDI	5	15.000,00	75.000,00	40.995,50	-34.004,50
2	JVC BR-DV3000E	5	1.900,00	9.500,00	17.347,22	+7.847,22
3	Digitalisierungsrechencluster	1	50.000,00	50.000,00	73.997,58	+23.997,58
4	Videokamera	1	8.000,00	8.000,00	8.009,65	+9,65
Summe				142.500,00	140.349,95	-2.150,05

Tabelle 6: Übersicht über angeschaffte Geräte (>400 €)

Sony j-30/SD und HVR-M25AE

Insgesamt wurden drei verschiedene Gerättypen für das Einspielen von unterschiedlichem Videomaterial vom Consumer- bis zum High End-Bereich angeschafft. Die J-30-Geräte erlauben die Wiedergabe von Digital Betacam-, MPEG IMX-, Betacam SX-, Betacam SP- und Betacam-Kassetten, die HVR-Geräte die Wiedergabe von HDV, DVCAM und DV in unterschiedlichen Auflösungen. Daneben wurden noch Geräte für VHS und S-VHS genutzt. Die unterschiedlichen Gerätetypen erlauben die Verarbeitung qualitativ sehr unterschiedlichen Ausgangsmaterials.

Digitalisierungscluster

Der Digitalisierungscluster bestand aus folgenden Einzelkomponenten:

- 1x TELESTREAM Pipeline HD Netzwerk Encoder zum Digitalisieren von HD-Material
- 1x TELESTREAM Pipeline SD Netzwerk Encoder zum Digitalisieren von SD-Material
- 3x Precision T5500: N-Series Mid-Tower zum Einspielen, als temporärer Speicher und zur Transkodierung in verschiedene Medienformate
- FOCUS ProxSys PX-50 LTO Edition, für die Archivierung auf LTO-Bändern

Player und Cluster wurden technisch miteinander verbunden. Es wurde ferner ein Roboter entwickelt, der die Player automatisch befüllt.

3. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die erzielten Ergebnisse in den sieben inhaltlichen Arbeitsbereichen (siehe Abbildung 1) konnten nur durch die gezielte Förderung in Form einer personell stark aufgestellten Projektgruppe mit einer entsprechenden Ausstattung erzielt werden. Vorarbeiten lagen in Form der Frameworks AMOPA und Xtrieval vor. Ferner lag zu Projektbeginn eine grundlegende technische Ausstattung vor. Auf dieser Basis wurden notwendige Anschaffungen und Arbeitsaufgaben realisiert, die die Einzelkomponenten zum einen optimierten und ausbauten, zum anderen in einen durchgängigen Workflow setzten. Diese komplexe Tätigkeit wäre mit Bordmitteln einer Professur nicht zu stemmen gewesen. Für Unternehmen war die perspektivische Verwertbarkeit noch zu unklar, um das Risiko eines solchen Projekts zu tragen.

Die im Rahmen des Projekts angeschaffte Technik wurde so konzipiert, dass keine nicht unbedingt notwendigen Neuanschaffungen getätigt wurden. Vielmehr wurde darauf geachtet, möglichst auf die Ressourcen der Professur zurückzugreifen. Wo zusätzliche Anschaffungen notwendig wurden, wurde auf eine möglichst kostengünstige Lösung gesetzt.

So wurde beispielsweise die automatische Ladestation der Archivierungsstraße nicht über industrielle Laderoboter, sondern über eine Lego-Mindstorms-Lösung realisiert.

Die Weiterentwicklungen wurden szenarienbasiert durchgeführt. Zu den einzelnen Szenarien wurden Gespräche mit Unternehmen und Verbänden geführt, bei denen die aktuelle Technik vorgestellt und das Entwicklungspotential besprochen wurde. In diesen Gesprächen ging es der Projektgruppe vor allem darum, für eine spätere Verwertbarkeit sinnvolle Weiterentwicklungen zu definieren.

Sowohl die vorgenommenen Investitionen als auch die erzielten Forschungsergebnisse wurden in die Infrastruktur am Standort Chemnitz integriert. Die gewonnenen Erkenntnisse wurden in ihren wesentlichen Teilen in die Lehre der Professur Medieninformatik eingearbeitet. Damit wird die Professur Medieninformatik bzw. die TU Chemnitz als kompetenter Ansprechpartner für privatwirtschaftliche Unternehmen der Medienbranche nachhaltig gestärkt. Die Beiträge zu internationalen Konferenzen sowie die hervorragenden Ergebnisse bei internationalen Vergleichskampagnen im Bereich Retrieval haben zur Stärkung und Schärfung dieses Profils beigetragen.

Letztendlich zeigen die in Richtung einer Verwertung ausgerichteten Folgeprojekte, dass die Maßnahmen der Projektgruppe zielgerichtet und erfolgreich waren.

4. Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

Ziel des Projektes war es die an der Professur Medieninformatik erstellten Softwareframeworks AMOPA und Xtrieval auf ihr Verwendbarkeit in wirtschaftlich geprägten Szenarien hin zu untersuchen. Dazu wurden drei exemplarische Szenarien realisiert, die unterschiedliche technische Einsatzmöglichkeiten beispielhaft darstellten und untersuchbar machten. Auf Messen wie der CEBiT wurde die Technik dem breiten Fachpublikum vorgestellt und stieß auf großes Interesse.

Auch die wissenschaftliche Community zeigte großes Interesse an der Technik. Wissenschaftlich dokumentiert sind die Projektergebnisse in 24 referierten Publikationen und zwei Dissertationen.

Daneben entsprangen auf der Basis des Projekts mehrere Folgeprojekte, teils öffentlich gefördert, teils privatwirtschaftlich finanziert. Die Folgeprojekte zielen darauf ab, die gewonnenen Erkenntnisse und Projektergebnisse im Sinne einer Wirtschaftlichen Orientierung weiterzutreiben. Die Zahl der Projekte und die unterschiedliche Finanzierung zeigen, dass ein breites Interesse an der entwickelten Technik besteht.

Während Projektlaufzeit ist ein Folgeprojekt mit Bezug zu validAX entstanden:

- „Freiluftlabor ‚Neue Mobilität‘ am Sachsenring“ im Rahmen des Schaufensters „Elektromobilität verbindet“: Hier untersucht die Professur Möglichkeiten der automatischen Straßenzustandsbestimmung auf Basis von onboard Kameras im Automobil. Basis für die Untersuchung bilden die Ergebnisse aus AMOPA.

Nach Projektende wurden folgende öffentlich geförderte Projekte gestartet, bzw. akquiriert:

- „H-RoC: Human Robot Cooperation - Kooperation von Menschen und mobilen Robotern in unstrukturierten Umgebungen“ ist ein durch das Sächsische Staatsministerium für Wissenschaft und Kultur gefördertes Projekt zur Kooperation mit Robotern in unstrukturierten Umgebungen. Die Professur Medieninformatik entwickelt hier die Ergebnisse der Audioanalyse weiter um eine audiobasierte Steuerung von Robotern zu ermöglichen.
- „StayCentered - Methodenbasis eines Assistenzsystems für Centerlotsen“: Im Rahmen des BMBF-Programms „Vom technischen Werkzeug zum interaktiven Begleiter – sozial- und emotionsensitive Systeme für eine optimierte Mensch-Technik-Interaktion“ (InterEmotio) wurde ein Grundlagenorientierter Vorschlag eingereicht, der zur Unterstützung von Centerlotsen im Flugverkehr dienen soll. Die Projektskizze wurde begutachtet und eine Aufforderung zum Vollartrag erteilt. Die Professur Medieninformatik ist mit Themen der Audioerkennung vertreten, die die Ergebnisse von ValidAX in den Bereich der Emotionserkennung weitertreibt.

In einer Mischung aus öffentlicher und privatwirtschaftlicher Finanzierung ist entstanden:

- Stiftungs juniorprofessur Media Computing mit Begleitprojekt „localizeIT – Lokalisierung visueller Medien“: Im Sommer 2014 startete die Stiftungs juniorprofessur Media Computing mit Dr. Marc Ritter, der im Rahmen von ValidAX promovierte. Die Professur wird von den Unternehmen Intenta GmbH, 3DInsight GmbH, 3D-Micromag AG, und der IBS Software & Research GmbH gestiftet. Das Begleitprojekt localizeIT wird über das Innoprofile-Transfer-Programm der Initiative Unternehmen Region durch das BMBF gefördert. Die Thematik in localizeIT ist eine Weiterführung der Ergebnisse zur Bilderkennung aus validAX und dem Vorgängerprojekt sachsMedia (2007-2012) und erweitert das Spektrum der Untersuchung auf Verfahren im dreidimensionalen Raum.

Rein privatwirtschaftliche aus ValidAX entstandene Vorhaben sind:

- „Pilotprojekt zum Erhalt von Medienmaterial sächsischer privater Fernsehveranstalter mittels Digitalisierung und Archivierung zwecks weiterer Nutzung für Medien, Wissenschaft und Kultur“, Forschungsvertrag mit der Sächsischen Landesanstalt für privaten Rundfunk und neue Medien (SLM): Ausgangspunkt des Pilotprojekts die Überzeugung, dass die in den letzten zwei Jahrzehnten von den Lokalfernsehsendern Sachsens produzierten Sendematerialien als Kulturgut bewahrungswürdig sind, da sie in besonderer Weise die Lebensverhältnisse der Menschen Ostdeutschlands seit der Wiedervereinigung dokumentieren. Ziel ist es im Rahmen dieses FuE-Projekts, anhand exemplarischer Videomaterialien Möglichkeiten der Digitalisierung und Annotation von TV-Sendungen zu explorieren. Die gewonnenen Erfahrungen und Ergebnisse zu Technik und Kosten sind Grundlage für eine zu definierende Archivierungsstrategie.
- „OphthalVis: Datenverarbeitung und Visualisierung - Neue Tools für die effiziente medizinische Versorgung degenerativer Netzhauterkrankungen“, Auftragsforschung für die Novartis AG. Das Projekt wird in erster Linie durch die oben beschriebene Stiftungs juniorprofessur Media Computing in Kooperation mit der Juniorprofessur Visual Computing durchgeführt. Verfahren der Bilderkennung werden hier auf ihre Eignung für den medizinischen Kontext Netzhauterkrankungen hin untersucht.

5. Während der Durchführung des Vorhabens dem ZE bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Projektlaufzeit begann die Initiative D-WERFT (<http://dwerft.de>), gefördert von 2014 bis 2017 im Rahmen des Wachstumskerne-Programms der Bundesregierung (<http://www.unternehmen-region.de/de/8386.php>, WK 1703). Diese Initiative verfolgt das Ziel, die Arbeitsprozesse der Produktion, Archivierung und Distribution von audiovisuellen Medieninhalten zu unterstützen. Der Wachstumskern ist dabei eng mit der lokalen Potsdamer Wirtschaft verbunden. Inhaltlich unterscheidet sich D-WERFT von validAX vor allem durch die Fokussierung auf die Filmwirtschaft, die andere Anforderungen stellt als die von validAX realisierten Einsatzszenarien.

6. Erfolgte oder geplante Veröffentlichungen der Ergebnisse

Im Verlauf des Vorhabens haben die Wissenschaftlichen Mitarbeiter erzielte Teilergebnisse auf verschiedenen nationalen und internationalen Fachkonferenzen und Workshops vorgestellt. Gesondert hervorzuheben sind dabei zwei Abgeschlossene Promotionen:

- Dr. Ing. Arne Berger promovierte zum Thema „Prototypen im Interaktionsdesign - Klassifizierung der Dimensionen von Entwurfsartefakten zur Optimierung der Kooperation von Design und Informatik“. Herr Berger forschte im Arbeitsbereich 3 – Workflowintegration. Seine Dissertation entstand in diesem Kontext.
- Dr. rer. nat. Marc Ritter promovierte zum Thema „A Generic Approach to Component-Level Evaluation in Information Retrieval“. Herr Ritter forschte im Arbeitsbereich 5 – Bilderkennung. Auch seine Forschungsergebnisse flossen in die Dissertation ein.

Daneben entstanden folgende referierte Publikationen:

- Kahl, Stefan; Ritter, Marc; Rosenthal, Paul (2014). Automatisierte Beurteilung der Schädigungssituation bei Patienten mit altersbedingter Makuladegeneration (AMD). In: Puente León, Fernando; Heizmann, Michael: Forum Bildverarbeitung, 27.11. - 28.11.2014, Regensburg, S. 179 - 190. - Karlsruhe : KIT Scientific Publishing, 2014.
- Herms, Robert; Ritter, Marc; Wilhelm-Stein, Thomas; Eibl, Maximilian (2014). Improving Spoken Document Retrieval by Unsupervised Language Model Adaptation Using Utterance-Based Web Search. In: INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association, Singapore, September 14-18, S. 1430-1433 [URL: http://www.isca-speech.org/archive/archive_papers/interspeech_2014/i14_1430.pdf, ISSN: 1990-9770]
- Ritter, Marc; Heinzig, Manuel; Herms, Robert; Kahl, Stefan; Richter, Daniel; Manthey, Robert; Eibl, Maximilian (2014). Technische Universität Chemnitz at TRECVID Instance Search 2014. In: Proceedings of TRECVID 2014, November 2014, Washington.
- Markus Rickert and Maximilian Eibl. 2014. A proposal for a taxonomy of semantic editing devices to support semantic classification. In Proceedings of the 2014 Conference on Research in Adaptive and Convergent Systems (RACS '14). ACM, New York, NY, USA, S. 34-39. DOI=10.1145/2663761.2664225 [<http://doi.acm.org/10.1145/2663761.2664225>]
- Wilhelm-Stein, Thomas; Herms, Robert; Ritter, Marc; Eibl, Maximilian (2014). Improving Transcript-Based Video Retrieval Using Unsupervised Language Model Adaptation, In: CLEF 2014, 15.-18. September 2014, Sheffield, UK. - Springer International Publishing, 2014. -

Lecture Notes in Computer Science No. 8685, S. 110-115. [ISBN/ISSN: 9783319113814, 9783319113821; DOI/URL: doi:10.1007/978-3-319-11382-1]

- Fritzsche, Thomas; Müller, Stefanie; Berger, Arne; Eibl, Maximilian (2014). Location Based Video Flipping: Interactive Prototype navigated by HbbTV remote control. In: ACM TVX2014, 25.-27. Juni 2014, Newcastle upon Tyne, GB, 2014. [URL: http://wsicc.net/2014/proceedings/wsicc2014_submission_5.pdf]
- Berger, Arne; Fritzsche, Thomas; Heidt, Michael; Eibl, Maximilian (2014). Location Based Video Flipping: Navigating Geospatial Videos in Lean Back Settings. In: ACM TVX 2014 Newcastle, UK, 2014
- Berger, Arne; Heidt, Michael; Eibl, Maximilian (2014). Towards a Vocabulary of Prototypes in Interaction Design - A Criticism of Current Practice. In: Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience. Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I. S.25-32. - Berlin, Heidelberg: Springer, 2014
- Wilhelm-Stein, Thomas; Eibl, Maximilian (2013). A Quantitative Look at the CLEF Working Notes. In: 4th International Conference of the CLEF Initiative, CLEF 2013, Valencia, Spain, September 23-26, 2013. - Berlin Heidelberg: Springer, 2013. - Lecture Notes in Computer Science Vol. 8138 [ISBN/ISSN: 978-3-642-40801-4 (Print) 978-3-642-40802-1 (Online) DOI: <http://dx.doi.org/10.1007/978-3-642-40802-1>]
- Manthey, Robert; Herms, Robert; Ritter, Marc; Storz, Michael; Eibl, Maximilian (2013). A Support Framework for Automated Video and Multimedia Workflows for Production and Archive. In: HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013. - Berlin Heidelberg: Springer. LNCS Vol. 8018, Part III, S. 336-341.
- Storz, Michael; Ritter, Marc; Manthey, Robert; Lietz, Holger; Eibl, Maximilian (2013). Annotate. Train. Evaluate. A Unified Tool for the Analysis and Visualization of Workflows in Machine Learning Applied to Object Detection. In: HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013. - Berlin Heidelberg: Springer. LNCS Vol. 8008, Part V, S. 196-205. [ISBN/ISSN: ISBN 978-3-642-39341-9; DOI: http://dx.doi.org/10.1007/978-3-642-39342-6_22]
- Herms, Robert; Manthey, Robert; Ritter, Marc; Eibl, Maximilian (2013). Ein adaptiver Ansatz zum Ingest großer Bestände audiovisueller Medien unter heterogenen Anforderungen. In: LWA 2013 - Lernen, Wissen & Adaptivität, 7.-9.09.2013, Bamberg. - Bamberg: Proceedings LWA, 2013, S. 268 – 273. [URL: <http://www.minf.uni-bamberg.de/lwa2013/proceedings/>]
- Neumerkel, Tom ; Manthey, Robert (2013). Funktionsumfang und Eignung von XML-Datenbanken für Multimedia- und Metadaten. In: LWA 2013 - Lernen, Wissen & Adaptivität, 7.-9.09.2013, Bamberg. - Bamberg : Proceedings LWA, 2013, S. 274 - 281
- Ritter, Marc; Herms, Robert; Manthey, Robert; Eibl, Maximilian (2013). Ein ganzheitlicher Ansatz zur Digitalisierung und Extraktion von Metadaten in Videoarchiven. In: Proceedings des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013), Potsdam, 19. bis 22. März 2013, S.362-371. - Glückstadt: Hülsbusch, 2013 [ISBN/ISSN: ISBN 978-3-86488-035-3; DOI/URL: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:kobv:525-4234>]
- Rickert, Markus; Eibl, Maximilian (2013). Evaluation of Media Analysis and Information Retrieval Solutions for Audio-Visual Content through their Integration in Realistic

Workflows of the Broadcast Industry. In: Research in Adaptive and Convergent Systems (ACM RACS, 2013), 01.-04.10.2013, Montreal, Quebec, Kanada, S. 118-121. - New York: ACM Press. [ISBN/ISSN: 978-1-4503-2348-2]

- Wilhelm-Stein, Thomas; Schürer, Benjamin; Eibl, Maximilian (2013). Identifying the Most Suitable Stemmer for the CHiC Multilingual Ad-hoc Task. In: CLEF 2013 Evaluation Labs and Workshop, 23 - 26 September, Valencia - Spain, 2013 [ISBN: 9788890481055, ISSN:2038-4963; URL: <http://www.clef-initiative.eu/documents/71612/b9672d04-c08d-4840-877f-8d664eec89cf>]
- Kürsten, Jens; Wilhelm, Thomas; Richter, Daniel; Eibl, Maximilian (2012). Chemnitz at the CHiC Evaluation Lab 2012: Creating an Xtrieval Module for Semantic Enrichment. In: CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012 [<http://www.clef-initiative.eu/documents/71612/fd9ccc73-42d7-48b2-9008-a4cfee57189f>]
- Wilhelm, Thomas; Kürsten, Jens; Eibl, Maximilian (2012). Chemnitz at CLEF IP 2012: Advancing Xtrieval or a baseline hard to crack. In: CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012 [<http://www.clef-initiative.eu/documents/71612/a77f637c-b3eb-4b57-9b14-ab7e1cb47ebc>]
- Kürsten, Jens; Eibl, Maximilian (2012). Comparing IR System Components Using Beanplots. In: Information Access Evaluation. Multilinguality, Multimodality, and Visual Analytics: Third International Conference of the CLEF Initiative, CLEF 2012, Rom, 17. - 20. September 2012, LNCS 7488, S.136-137. [ISSN 0302-9743, ISBN 978-3-642-33246-3, http://dx.doi.org/10.1007/978-3-642-33247-0_15]
- Herms, Robert; Manthey, Robert; Eibl, Maximilian (2012). Framework für Ingest mit Annotation technischer Randbedingungen. In: LWA 2012, 12.-14.09.2012, Dortmund. [<http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-103912>]
- Ritter, Marc; Manthey, Robert; Lietz, Holger; Thomanek, Jan; Wanielik, Gerd (2012). An empirical study on image features for pedestrian detection systems. In: Forum Bildverarbeitung 2012, 29.-30.11.2012, Regensburg, S.315-326.
- Lietz, Holger; Ritter, Marc; Manthey, Robert; Wanielik, Gerd (2011). Improving pedestrian detection using MPEG-7 descriptors. In: Advances in Radio Science. - 11. 2013, 8, S. 101 - 105.
- Wilhelm, Thomas; Kürsten, Jens; Eibl, Maximilian (2011). A Tool for Comparative IR Evaluation on Component Level. In: 34th International ACM SIGIR conference on research and development in Information Retrieval, 24. - 28. Juli 2011, Beijing, China, S. 1291-1292. [ISBN 978-1-4503-0757-4; <http://dx.doi.org/10.1145/2009916.2010165>]
- Becks, Daniela; Eibl, Maximilian; Jürgens, Julia; Kürsten, Jens; Wilhelm, Thomas; Womser-Hacker, Christa (2011). Does Patent IR Profit from Linguistics or Maximum Query Length? CLEF 2011 Labs and Workshop, Notebook Papers, 19-22 September 2011, Amsterdam, (PDF: 8 S.) [http://clef2011.org/resources/proceedings/Becks_Clef2011.pdf, ISBN 978-88-904810-1-7]

Chemnitzer Informatik-Berichte

In der Reihe der Chemnitzer Informatik-Berichte sind folgende Berichte erschienen:

- CSR-08-01** Johannes Steinmüller, Holger Langner, Marc Ritter, Jens Zeidler (Hrsg.), 15 Jahre Künstliche Intelligenz an der TU Chemnitz, April 2008, Chemnitz
- CSR-08-02** Petr Kroha, José Emilio Labra Gayo, Using Semantic Web Technology in Requirements Specifications, November 2008, Chemnitz
- CSR-09-01** Amin Coja-Oghlan, Andreas Goerdts, André Lanka, Spectral Partitioning of Random Graphs with Given Expected Degrees - Detailed Version, Januar 2009, Chemnitz
- CSR-09-02** Enrico Kienel, Guido Brunnett, GPU-Accelerated Contour Extraction on Large Images Using Snakes, Februar 2009, Chemnitz
- CSR-09-03** Peter Köchel, Simulation Optimisation: Approaches, Examples, and Experiences, März 2009, Chemnitz
- CSR-09-04** Maximilian Eibl, Jens Kürsten, Marc Ritter (Hrsg.), Workshop Audiovisuelle Medien: WAM 2009, Juni 2009, Chemnitz
- CSR-09-05** Christian Hörr, Elisabeth Lindinger, Guido Brunnett, Considerations on Technical Sketch Generation from 3D Scanned Cultural Heritage, September 2009, Chemnitz
- CSR-09-06** Christian Hörr, Elisabeth Lindinger, Guido Brunnett, New Paradigms for Automated Classification of Pottery, September 2009, Chemnitz
- CSR-10-01** Maximilian Eibl, Jens Kürsten, Robert Knauf, Marc Ritter, Workshop Audiovisuelle Medien, Mai 2010, Chemnitz
- CSR-10-02** Thomas Reichel, Gudula Rünger, Daniel Steger, Haibin Xu, IT-Unterstützung zur energiesensitiven Produktentwicklung, Juli 2010, Chemnitz
- CSR-10-03** Björn Krellner, Thomas Reichel, Gudula Rünger, Marvin Ferber, Sascha Hunold, Thomas Rauber, Jürgen Berndt, Ingo Nobbers, Transformation monolithischer Business-Softwaresysteme in verteilte, workflowbasierte Client-Server-Architekturen, Juli 2010, Chemnitz
- CSR-10-04** Björn Krellner, Gudula Rünger, Daniel Steger, Anforderungen an ein Datenmodell für energiesensitive Prozessketten von Powertrain-Komponenten, Juli 2010, Chemnitz
- CSR-11-01** David Brunner, Guido Brunnett, Closing feature regions, März 2011, Chemnitz

Chemnitzer Informatik-Berichte

- CSR-11-02** Tom Kühnert, David Brunner, Guido Brunnett, Betrachtungen zur Skelettextraktion umformtechnischer Bauteile, März 2011, Chemnitz
- CSR-11-03** Uranchimeg Tudevtagva, Wolfram Hardt, A new evaluation model for eLearning programs, Dezember 2011, Chemnitz
- CSR-12-01** Studentensymposium Informatik Chemnitz 2012, Tagungsband zum 1. Studentensymposium Chemnitz vom 4. Juli 2012, Juni 2012, Chemnitz
- CSR-12-02** Tom Kühnert, Stephan Rusdorf, Guido Brunnett, Technischer Bericht zum virtuellen 3D-Stiefeldesign, Juli 2012, Chemnitz
- CSR-12-03** René Bergelt, Matthias Vodel, Wolfram Hardt, Generische Datenerfassung und Aufbereitung im Kontext verteilter, heterogener Sensor-Aktor-Systeme, August 2012, Chemnitz
- CSR-12-04** Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Knauf, Jens Kürsten, Albrecht Kurze, Markus Rickert, Marc Ritter, Schlussbericht zum InnoProfile Forschungsvorhaben sachsMedia - Cooperative Producing, Storage, Retrieval and Distribution of Audiovisual Media (FKZ: 03IP608), September 2012, Chemnitz
- CSR-12-05** Anke Tallig, Grenzgänger - Roboter als Mittler zwischen der virtuellen und realen sozialen Welt, Oktober 2012, Chemnitz
- CSR-13-01** Navchaa Tserendorj, Uranchimeg Tudevtagva, Ariane Heller, Grenzgänger - Integration of Learning Management System into University-level Teaching and Learning, Januar 2013, Chemnitz
- CSR-13-02** Thomas Reichel, Gudula Rüniger, Multi-Criteria Decision Support for Manufacturing Process Chains, März 2013, Chemnitz
- CSR-13-03** Haibin Xu, Thomas Reichel, Gudula Rüniger, Michael Schwind, Softwaretechnische Verknüpfung der interaktiven Softwareplattform Energy Navigator und der Virtual Reality Control Platform, Juli 2013, Chemnitz
- CSR-13-04** International Summerworkshop Computer Science 2013, Proceedings of International Summerworkshop 17.7. - 19.7.2013, Juli 2013, Chemnitz
- CSR-13-05** Jens Lang, Gudula Rüniger, Paul Stöcker, Dynamische Simulationskopplung von Simulink-Modellen durch einen Functional-Mock-up-Interface- Exportfilter, August 2013, Chemnitz
- CSR-14-01** International Summerschool Computer Science 2014, Proceedings of Summerschool 7.7.-13.7.2014, Juni 2014, Chemnitz
- CSR-15-01** Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Herms, Stefan Kahl, Jens Kürsten, Albrecht Kurze, Robert Manthey, Markus Rickert, Marc Ritter, ValidAX - Validierung der Frameworks AMOPA und XTRIEVAL, Januar 2015, Chemnitz

Chemnitzer Informatik-Berichte

ISSN 0947-5125

Herausgeber: Fakultät für Informatik, TU Chemnitz
Straße der Nationen 62, D-09111 Chemnitz