



BEUTH HOCHSCHULE
FÜR TECHNIK
BERLIN

University of Applied Sciences

Scalable Video Coding

Streaming Media and Development, WS 2012/13

Vorgelegt von: Sharif Hanini, Denis Oczko, Martin Maurer

Betreuer: Professor Jürgen Lohr, Professor Robert Strzebkowski

18. Januar 2013

Inhaltsverzeichnis

1. Einleitung	1
2. Aktuelle Trends und Herausforderungen	2
3. H264-SVC in a Nutshell	4
3.1. Allgemein	4
3.2. Abgrenzung zu Adaptive Streaming	6
3.3. Videokodierung mit H264/AVC	6
3.4. Videokodierung mit H264-SVC	8
3.4.1. Baselayer und Enhancementlayer	8
3.4.2. Spatiale Skalierbarkeit	9
3.4.3. Temporale Skalierbarkeit	10
3.4.4. Qualitative Skalierbarkeit	10
3.5. H264-SVC mit MPEG DASH	12
4. H264-SVC Hands-on	14
4.1. SVC-Referenzcodec	14
4.1.1. Installation	14
4.1.2. Vorbereitung	15
4.1.3. Enkodieren	16
4.1.4. Dekodieren	17
4.1.5. Bitstreams extrahieren	17
4.1.6. Evaluation	17
4.2. Kommerzielle Anbieter/Verwendung	18
4.2.1. Google Video	18
4.2.2. Mainconcept SVC	18

5. Konklusion

19

A. Abbildungsverzeichnis

B. Literatur

1. Einleitung

In den letzten Jahren hat sich eine große Wandlung im Broadcastingumfeld ergeben. Während früher die Auslieferung von Videocontent in Standardauflösung über DVB-S und DVB-T in MPEG2 nahezu die einzige Art der Videodistribution war, ist der Distributionsprozess heute viel stärker diversifiziert und interaktiv.

Potentielle Kunden erwarten heutzutage die parallele Auslieferung des Videocontents in regulärer SD- als auch in FullHD-Auflösung. Weiterhin sind die Empfangsgeräte der Kunden sehr viel inhomogener in Bezug auf technische Anbindung als auch Dekodierfähigkeiten. Videoinhalte werden heutzutage sowohl stationär auf regulären Fernsehgeräten und modernen SmartTVs als auch mobil auf Tablettts und Smartphones konsumiert. Dabei gewinnt gleichzeitig auch die Interaktivität durch zusätzliche Medieninhalten im Rahmen des hybriden Fernsehen an Bedeutung. Dadurch stehen die traditionellen Anbieter dieser Dienste vor der großen Herausforderungen, wie sie ihre Inhalte an die Kunden über verschiedene Kanäle in verschiedenen Auflösungen gleichzeitig übertragen können. Mit Scalable Video Coding (SVC) steht eine Technologie zur Verfügung, die dieses zu Leisten im Stande ist.

In dieser Ausarbeitung wird Scalable Video Coding detailliert vorgestellt. Im ersten Teil der Arbeit wird vorgestellt welche aktuellen Herausforderungen sich im Broadcastinbereich stellen. Im zweiten Teil wird die Scalable Video Coding Extension des H264/AVC-Standards vorgestellt. In diesem Abschnitt wird zuerst ein allgemeiner Ausblick auf die Notwendigkeit und Vorteile von SVC gegeben. Danach wird die Videokodierung mit H264/AVC vorgestellt und anschließend die Neuerungen zwischen H264/AVC und H264-SVC gegenübergestellt. Im nächsten Abschnitt wird die Verwendung von Scalable Video Coding im Rahmen von MPEG DASH vorgestellt. Im nächsten XXXX

Im letzten Abschnitt wird abschließend eine Zusammenfassung über den Themenbereich Scalable Video Coding gegeben.

2. Aktuelle Trends und Herausforderungen

In den vergangenen Jahren dominierten verschiedene Themen das Broadcastingumfeld. Im Jahr 2009 befand sich die Broadcastingindustrie inmitten der Vorbereitung auf die Umstellung auf HD. Daher wurde dies als wichtigster Trend der nächsten Jahre gesehen[20, S. 1]. In Deutschland begann im Jahr 2010 beispielsweise die Ausstrahlung von ARD HD im regulärem Sendebetrieb.

In den Jahren 2010 und 2011 wurde die Distribution über mehrere Plattformen, das sogenannte „Multiplatform delivery“, als wichtigster Trend gesehen[20, S. 1]. Auch im Jahr 2012 liegt dies mit weitem Abstand vor anderen Trends weiterhin vorn. Einhergehend damit wurden als weitere wichtige Trends die Optimierung des Workflows und die Übertragung via IP genannt [19, S. 1]. Die Ausrichtung auf die Bedienung mehrere Distributionskanäle und Plattformen hat mehrere Gründe.

Einerseits haben sich die Sehgewohnheiten der Konsumenten geändert. Beispielsweise sank im Jahr 2011 die Zahl an amerikanischen Haushalten, welche keinen Fernseher besaßen, zum ersten Mal in 20 Jahren[17, S. 1]. Das bedeutet nicht automatisch, dass der Fernsehkonsum rückläufig ist und durch das Internet ersetzt wird, sondern dass es zu einer Medienkonvergenz kommt. Traditionelles Fernsehen und Internet verschmelzen. Fernsehen wird über das Internet geschaut und über Fernsehgeräte bzw. SmartTVs wird im Internet gesurft und Apps werden ausgeführt [4, S. 180].

Bereits im Jahr 2011 bestand über die Hälfte des ganzen Internetverkehrs der Konsumenten aus Videodaten. Bis zum Jahr 2016 wird sich der Wert auf ca. 56% steigern. Der Anteil an Videodaten für SmartTVs lag im Jahr 2011 bei 8% und wird bis zum Jahr 2016 auf 12% steigern[2]

Die Möglichkeit Webinhalte auf dem Fernsehgerät bzw. SmartTV zu empfangen, ermöglicht einerseits neuen Konkurrenten den Einstieg in den Markt. Sogenannter Over-the-top-Content, also im Internet frei verfügbare und durch Werbung finanzierte Inhalte, wie beispielsweise Youtube oder Hulu im amerikanischen Markt, können auf lange Sicht gegenüber etablierten Anbietern im Broadcastingumfeld Marktanteile gewinnen[10, S. 2].

Andererseits bieten Fernsehgerät mit Internetzugang auch für etablierte Anbieter im Broadcastingumfeld eine neue Einnahmequelle im Rahmen von HybridTV. Zusätzlich bietet es dem Zuschauer durch Zusatzinformationen einen Mehrwert und kann sich so positiv auf die Zuschauerzufriedenheit und Marktanteile auswirken[21]

Weiterhin nimmt die Anzahl unterschiedlichen Endgeräte mit heterogenen Dekodierfähigkeiten stark zu. Während Broadcastinganbieter früher im Rahmen von Ausstrahlungen in DVB-S oder DVB-T davon ausgehen konnte, dass nahezu alle Geräte die in MPEG2 ausgestrahlten Inhalte bis zu einer bestimmten Bitrate dekodieren können, gibt es heute einerseits Geräte am unteren Ende der Skala, die nur SD-Content im H264/AVC-Baseline Profile abspielen können, wie beispielsweise ältere Android Smartphones. Andererseits gibt es wiederum Geräte, die FullHD in H264/AVC-High Profile dekodieren können, z. B. reguläre Settopboxen und SmartTVs[10, S. 3].

Ein weiterer Grund für die Forcierung auf mehrere Distributionskanäle ist der heterogene Übertragungskanal zwischen Sender und Empfänger[10, S. 4]. Die Empfänger sind sowohl mit sehr hoher Bandbreite von teilweise 100Mbit/s angebunden, als auch mit einer relativ geringen Bandbreite von unter 1Mbit/s. Durchschnittlich beträgt die Bandbreite in Deutschland derzeit beispielsweise nur 5,8 Mbit/s und wird in den nächsten Jahren nur langsam steigen[11, S. 1]. Erschwerend kommt hinzu, dass im mobilen Bereich durch die Shared-Media-Charakteristik die Bandbreite stark schwanken kann.

Um diesen sich in Zukunft weiter wachsenden Herausforderungen zu begegnen, bieten sich skalierbare Videocodecs an. Diese sollen im folgenden näher vorgestellt werden.

3. H264-SVC in a Nutshell

3.1. Allgemein

Im Bezug auf einen Videostrom versteht man unter dem Begriff „Skalierbarkeit“ die Eigenschaft, dass Teile des Videostroms entfernt werden können und der Videostrom nach Entfernen weiterhin valide und abspielbar bleibt. Dieser Teilvideostrom bildet den Originalvideostrom mit verringerter Qualität ab. Videoströme, die diese Fähigkeit nicht besitzen, nennt man Singlelayer- Videoströme [12, S. 2].

Skalierbare Videocodecs befinden sich seit nunmehr 20 Jahren in der Entwicklung. Bereits ältere Videocodecs wie das angesprochene MPEG2 boten bereits durch standardisierte Erweiterungen (z. B. MPEG2 Video) die Möglichkeiten zur Übertragung von skalierbaren Videoströmen [12, S. 1].

Allerdings wurden diese Codec-erweiterungen nur selten benutzt, da der Bedarf nach Auslieferung von skalierbaren Videoinhalten noch vergleichsweise gering war. Dies ist durch die relativ homogene technische Ausstattung der Empfänger und deren Anbindung an das Internet zu erklären. Zusätzlich verringerte sich durch den Einsatz der SVC-Erweiterungen in älteren Codecs die Effizienz der Videokodierung stark, während gleichzeitigem Rechenaufwand anstieg [13, 1f].

In den vergangenen Jahren hat in Bezug auf die technisch Anbindung und Ausstattung der Empfänger eine starke Heterogenisierung eingesetzt, sodass skalierbare Videokodierung und Distribution immer mehr in den Fokus rückt. Weiterhin hat sich mit Verabschiedung von H264/AVC im Jahr 2007 [13, S. 2] ein Codec etabliert, der sehr viel effizienter kodiert als ältere Videocodecs [14, S. 2].

Entwickelt vom JVT¹ und auf dem regulären H264/AVC-Standard aufbauend, ist H264-SVC eine Erweiterung, die als Amendment 3 in Annex G in dieses Standard aufgenommen wurde[14, S. 10]. Damit benutzt H264-SVC die effizienten Enkodierfunktionen und technischen Features des H264/AVC-Videostandards und erweitert diesen, um den zu übertragenden Videostrom skalierbar zu machen[1, S. 59]. Zusätzlich bleibt dadurch die Rückwärtskompatibilität zu herkömmlichen H264/AVC-Dekodern gewahrt. [10, S. 601]

H264-SVC bietet im Broadcastingumfeld einige Vorteile gegenüber anderen derzeit eingesetzten Verfahren, wie beispielsweise Simulcast, bei welchem jeder Videostrom unabhängig voneinander kodiert und übertragen wird. [13, S. 1] Auf H264/AVC aufbauend, ist H264-SVC vollständig abwärtskompatibel zu herkömmlichen H264-Dekodern.

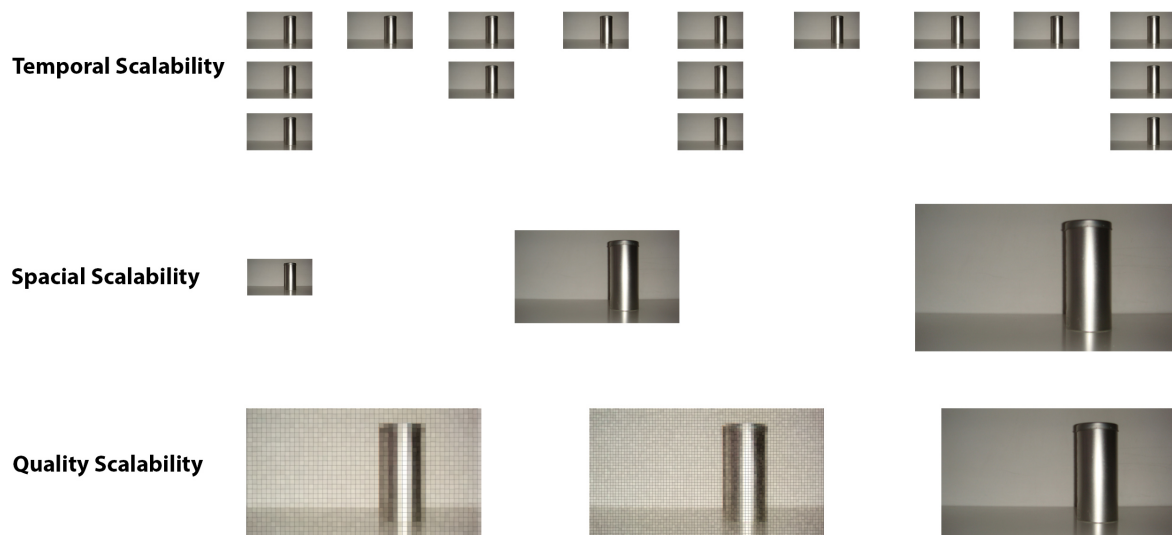


Abbildung 3.1.: Arten der Skalierbarkeit (Eigene Abbildung nach [7, S. 143])

Diese ignorieren die SVC-Erweiterung und dekodieren nur die regulären H264/AVC-Inhalt. Dies ist möglich indem neben einem regulärem H264/AVC-Videostrom, dem sogenannten Baselayer, weitere Videoströme, den sogenannten Enhancementlayers, eingebettet werden. Diese Enhancementlayers können sich gegenüber dem Baselayer in Bezug auf Framerate (Temporal), Auflösung (Spatial) und Bildqualität (Quality) unterscheiden (Siehe Abbildung 3.1) [10, S. 601]. In skalierbaren Videoströmen können diese drei Arten von Enhance-

¹JVT - Joint Video Team

mentlayern auch beliebig kombiniert werden, z. B. Spatiotemporale Enhancementlayer [12, S. 2].

Zusätzlich können mit einer in H264-SVC enkodierten Videoquelldatei gleichzeitig unterschiedliche Geräte z. B. Smartphones und FullHD-Geräte angesprochen werden. Die sonst übliche Transkodierung und das Vorhalten vorab enkodierter Teilvideodateien entfällt. Das erleichtert die Langzeitarchivierung und verringert gleichzeitig den benötigten Speicherplatz.

Bei Verwendung dieser H264-SVC erhöht sich der Overhead nur um 10% bis 20% verglichen mit regulärem H264/AVC. Im Falle eines stark fehleranfälligen Übertragungsmediums kann der Overhead sogar geringer ausfallen, als bei Simulcastübertragung. Zusätzlich bleibt der zur Enkodierung und Dekodierung benötigte Rechenaufwand nahezu gleich [14, S. 4].

3.2. Abgrenzung zu Adaptive Streaming

3.3. Videokodierung mit H264/AVC

Als Erweiterung des H264/AVC-Standards benutzt H264-SVC den gleichen Enkodierablauf und erweitert diesen in wenigen Bereichen, um Skalierbarkeit zu ermöglichen. Die Videokodierung mit H264/AVC erfolgt in vier Teilschritten. Zuerst wird der aktuelle Videoframe (F) in mehrere Slices unterteilt, die wiederum Makroblöcke enthalten [5, S. 146]. In jedem Makroblock sind die Luminanz und Chrominanzwerte gespeichert. Aufgrund des Chromasubsamplings liegen die Luminanzwerte in den meisten Fällen nur mit halbiert Auflösung vor.

Die Größe der Makroblöcke ist auf 16x 16 Pixel festgelegt. Makroblöcke können wiederum in verschiedene Makroblockpartitionen unterteilt werden z.B 8 x 8 Pixel Blöcke. Diese Makroblockpartitionen können selbst wieder in Sub-Makroblockpartitionen von beispielsweise 4 x 4 Pixel unterteilt werden. Diese so entstehende Unterteilung wird als *Tree Structured Motion Compensation* bezeichnet [7, S. 160].

Neben der Makroblockgröße unterscheidet man auch den Typ eines Makroblocks und damit die Art der Slices. I-Makroblöcke werden von vorherigen Makroblöcken des selben

Slices kodiert und dienen anderen Makroblöcke als Referenz. P-Makroblöcke referenzieren auf vorhergegangene I-Makroblöcke von vorherigen Frames und ermöglichen dadurch eine Steigerung der Kompressionsrate. B-Makroblöcke referenzieren sowohl auf vorhergegangene als auch auf nachfolgende I-Makroblöcke dieser Frames und ermöglichen so eine weitere Steigerung der Kompression [7, 164f].

Analog zu den Makroblöcken werden auch die Slices nach I, P und B-Slice unterschieden. I-Slices enthalten nur I-Makroblöcke, während P-Slices sowohl I-Makroblöcke als auch P-Makroblöcke enthalten können. B-Slices können I-Makroblöcke und B-Makroblöcke enthalten [7, 159f].

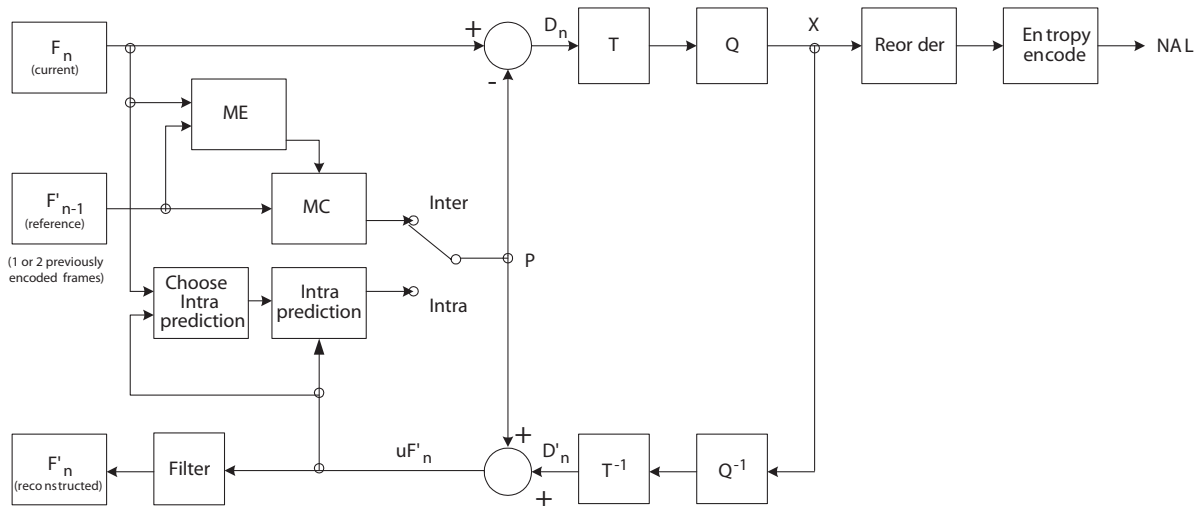


Abbildung 3.2.: Funktionsweise eines H264/AVC-Encoders [7]

Bei der Kodierung eines Frames (F) wird für jeden Makroblock eine Voraussage getroffen. Diese kann eine intra-basierte Voraussage sein, die auf der Kodierung von vorangegangenen Makroblöcken des gleichen Slice beruht. Alternativ kann es auch eine intra-basierte Voraussage sein, die auf der Referenzierung und Kodierung von Makroblöcken aus vorausgegangenen- und nachfolgenden Frames beruht. In beiden Fällen wird die Differenz zwischen dem aktuellen Makroblock und der Voraussage gebildet und die Abweichungen zwischen beiden Makroblöcken gespeichert [7, 161f].

Im nächsten Schritt werden diese Abweichungen quantisiert und die Koeffizienten in einem Zickzack-Verfahren angeordnet [7, 198f]. Dieses Ausleseverfahren ist für die anschließende Entropiekodierung besser geeignet, da so ähnliche Koeffizienten nacheinander stehen und

diese somit einfacher zusammenfassbar sind [5, S. 30]. Bei H264/AVC kommen mehrere Verfahren zur Entropiekodierung zum Einsatz. Neben Variable-Length Codes (VLC) auch Context-Adaptive Arithmetic Coding (CABAC), Context-Adaptive Variable Coding (CAVLC) und Exp-Golomb [7, 198f].

Im letzten Schritt werden die entropiekodierten Koeffizienten zu einem Videostrom zusammengefasst und durch den Network Abstraction Layer (NAL) übertragen[7, 161f]

3.4. Videokodierung mit H264-SVC

3.4.1. Baselayer und Enhancementlayer

Zur Erzeugung von skalierbaren Videostreamen werden H264-SVC Videostreame in zwei Arten von Layer unterteilt. Einerseits gibt es den sogenannten Baselayer, welcher in jedem H264-SVC Videostrom vorhanden sein muß. Zusätzlich gibt es auf diesen aufbauend mehrere Enhancementlayer, die Videostreame mit unterschiedlichen Qualitätsstufen (Auflösung, Framerate und Qualität) enthalten.[6, S. 1]

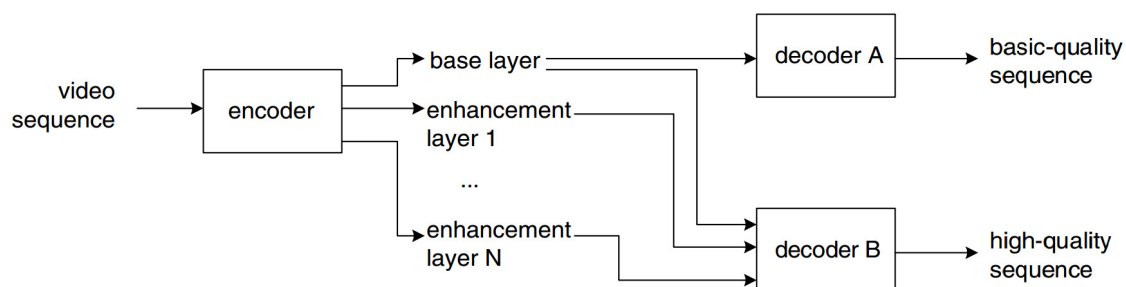


Abbildung 3.3.: Base- und Enhancementlayer [7, S. 142]

Der Baselayer wird im regulärem H264/AVC-Modus enkodiert. Die Kodierung erfolgt mit festgelegter Group of Pictures bzw. GOP und freier Refrenzierung der einzelnen Frames innerhalb jeder GOP. Dadurch kann der Baselayer auch von regulären, nicht H264-SVC-fähigen H264/AVC Videodecodern dekodiert- und angezeigt werden. Bei diesen Geräten werden die Enhancementlayer ignoriert. [7, S. 144]

H264-SVC-fähige Geräte dekodieren sowohl den Baselayer als auch den Enhancementlayer. Die einzelnen Enhancementlayer erhöhen mit jeder Iteration des Enhancementlayers die

Qualität in Bezug auf Auflösung, Bitrate oder Framerate. Bei der Dekodierung wird aus dem im Enhancementlayer gespeicherten Deltawerten bzw Differenzen zum Baselayer und dem Baselayer selber das Videobild rekonstruiert [7, S. 144].

3.4.2. Spatiale Skalierbarkeit

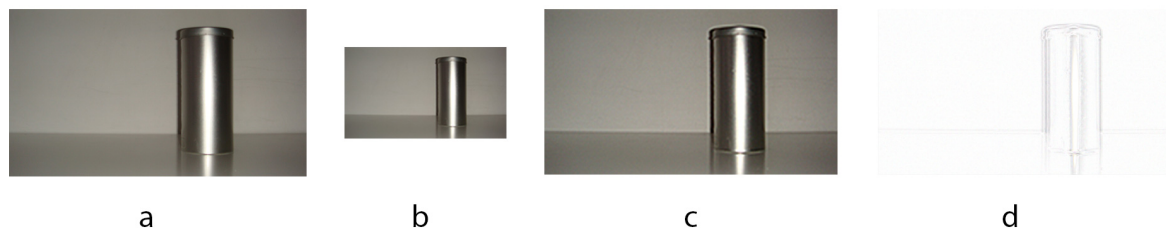


Abbildung 3.4.: Enkodierung von Baselayer und Enhancementlayer
(Eigene Abbildung nach [7, S. 143])

Die Videoenkodierung der Spatialen- bzw. Auflösungs-Enhancementlayer erfolgt, indem zuerst für jedes Frame des Videoeingangssignal (Siehe Abbildung 3.4 a) der Baselayer (Siehe Abbildung 3.4 b) mit verringerter spatialer Qualität in H264/AVC enkodiert wird.

Um die Differenzen des Baselayers zum Originalvideobild zu erhalten, wird der zuvor erzeugte Baselayer auf die ursprüngliche Größe skaliert (Siehe Abbildung 3.4 c) und die Differenz beider Bilder extrahiert (Siehe Abbildung 3.4 d). Aus diesem Differenzbild wird der erste Enhancementlayer enkodiert. Jeder weitere Enhancementlayer enkodiert die Differenz zum vorherigen, hochskalierten Enhancementlayer. [7, 142ff]

Die Dekodierung der mit H264-SVC enkodierten Videoströmen erfolgt, indem zuerst der Baselayer dekodiert wird (Siehe Abbildung 3.5 a) und auf die ursprüngliche Auflösung hochskaliert wird (Siehe Abbildung 3.5 b). Anschließend werden die Enhancementlayer dekodiert (Siehe Abbildung 3.5 c) und deren Bildinformationen zum bereits dekodierten Baselayer hinzugefügt (Siehe Abbildung 3.5 d), um das vollständigen Videobild wiederherzustellen.[7, S. 144]



Abbildung 3.5.: Dekodierung von spatialen Enhancementlayern
(Eigene Abbildung nach [7, S. 143])

3.4.3. Temporale Skalierbarkeit

Zur Erzeugung verschiedenerer temporaler- bzw. zeitlicher- Enhancementlayer verwendet H264-SVC, die regulären in H264/AVC angewandten hierarchischen B- und P Frames [18, S. 3]. Zur Erzeugung der temporalen Enhancementlayer werden die einzelnen Frames der in höchster Framerate vorliegende Videodatei auf die einzelnen Layer aufgeteilt. Die Länge der GOP ist dabei relativ frei wählbar. Der Baselayer enthält die Qualitätsstufe mit der geringsten Framerate. In diesem wird nur ein geringer Teil der GOP gespeichert, beispielsweise I-Frame des ersten und letzten Frames der GOP und ein weiterer B-Frame. [7, S. 144]

Die Enhancementlayer kodieren sukzessive jeweils höhere Frameraten, indem sie auf die Frames der vorhergehenden Layer referenzieren. Im ersten Enhancementlayer werden weitere B-Frames des Originalvideodatei gespeichert. Diese B-Frames referenzieren die I-Frames des Baselayers. Im nachfolgenden Enhancementlayer werden wiederum weitere B-Frames des Originalvideodatei gespeichert. Damit referenziert dieser nun den Baselayer und den ersten Enhancementlayer. [8, S. 292].

Dieses Prinzip setzt sich bis zum letzten Baselayer fort. Bei der Dekodierung muß man daher alle Base- und Enhancementlayer dekodieren, die dem gewünschten Enhancementlayer vorangehen. [8, S. 292].

3.4.4. Qualitative Skalierbarkeit

H264-SVC enkodierte Videoströme mit qualitativer Skalierbarkeit besitzen mehrere Layer mit unterschiedlicher Bildqualitätsstufen. Diese Art der Skalierbarkeit wird auch als SNR-

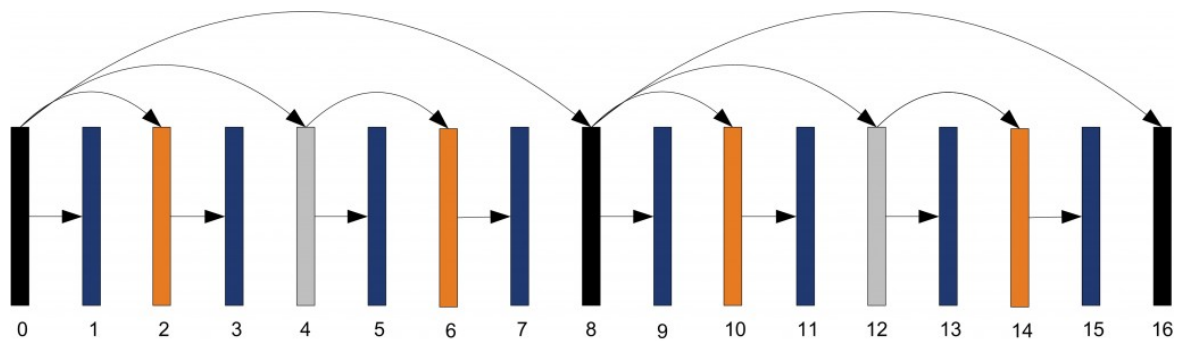


Abbildung 3.6.: Temporale Skalierbarkeit [15]

Skalierbarkeit bezeichnet. Im Detail unterstützt H264-SVC drei unterschiedliche Implementierung für qualitative Skalierbarkeit. [6, S. 10]

Die Basisimplementation ist das sogenannte Coarse-grain Quality Scalable Coding (CGS). Dies kann als Sonderfall der spatialen Skalierbarkeit angesehen werden. Zuerst wird aus dem Videoeingangssignal der Baselayer mit geringer Qualität (z. B. Bitrate) generiert. Anschließend werden in nachfolgenden Enhancementlayers die Differenz zwischen dem I-Frame des Originalvideos und dem I-Frame des vorherigem Layers mit festen Bitraten kodiert [18, S. 4]. Innhalb der GOP werden nur die Referenzbilder im eigenen Layer kodiert. Referenzierungen zwischen den einzelnen Layern findet nichts statt.[12, 8f]. Skalierbarkeit wird in diesem Fall durch das komplette Entfernen einzelner Enhancementlayer erreicht, wodurch allerdings die Flexibilität und Effizienz eingeschränkt ist. [9, S. 2] Die Anzahl der Layer ist auf bestimmte Bitraten beschränkt und der Wechslen zwischen verschiedenen Layern ist nur an an den GOP-Grenzen bzw. I-Frames möglich [13, S. 7].

Eine verbesserte Implementation ist das sogenannte Medium Grain Scalability (MGS), bei welchem innerhalb der GOP des Baselayers auch auf die Referenzframes des nachfolgendem Enhancementlayers zugegriffen werden kann. Dies erlaubt eine flexiblere Skalierbarkeit, erhöht aber die Fehleranfälligkeit. [6, S. 8]

Die dritte Implementation ist die sogenannte Fine Grain Scalability (FGS). Diese Art der Skalierung baut auf CGS auf und verfeinert diese Methode. Die Bitrate der einzelnen Layer wird dynamisch an die aktuell verfügbare Bandbreite angepasst. Referenzierung einzelner Frames innerhalb der GOP wird nur im Baselayer durchgeführt. [6, S. 8]

3.5. H264-SVC mit MPEG DASH

MPEG DASH (Dynamic Adaptive Streaming over HTTP) definiert ein HTTP-basierendes Streamingverfahren. [3, S. 1]. Durch die Verwendung von HTTP gegenüber anderen Protokollen wie RTP ergeben sich mehrere Vorteile. In vielen Einsatzbereichen (z. B. Unternehmen) sind aus Sicherheitsgründen nur absolut notwendige Ports geöffnet, unter anderem HTTP auf Port 80. Zusätzlich entfällt die Auslieferung der Streaminginhalte über einen dedizierten Streamingserver. Stattdessen können reguläre Webserver zur Auslieferung benutzt werden. Dadurch sinken für Streaminganbieter die Kosten für die Auslieferung des Contents [16, S. 3].

Zur Auslieferung des Videocontents spezifiziert MPEG DASH ein XML-Dokument, das sogenannte Media Presentation Document (MPD), welches die einzelnen Segmente der Videodatei definiert und referenziert. Auf diese einzelnen Segmente kann per HTTP-GET zugegriffen werden [3, S. 1].

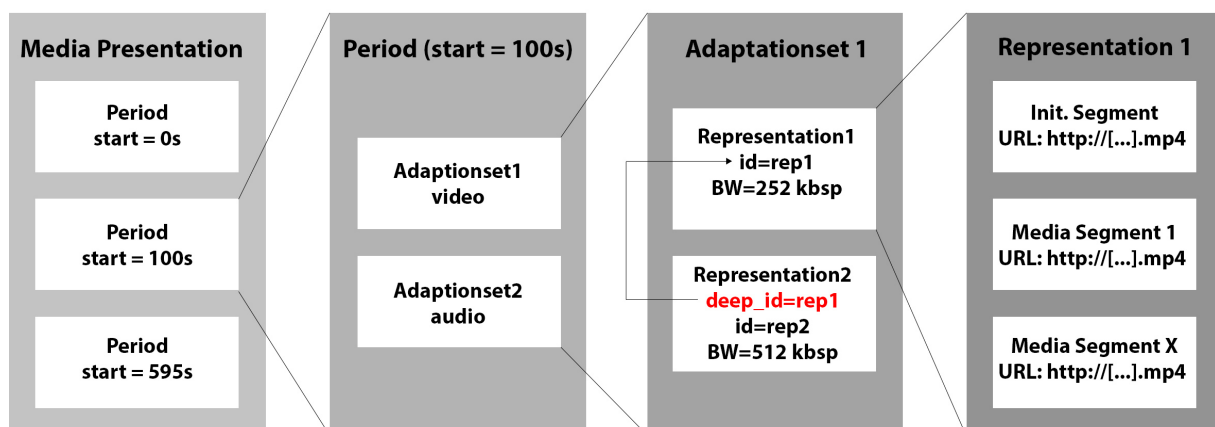


Abbildung 3.7.: Media Presentation Document und H264-SVC
(Eigene Abbildung nach [3])

Jedes in der MPD definierte Segment besteht aus einer Abfolge von Zeiteinheiten bzw. Periods, deren Intervall beliebig gewählt werden kann. In meisten Fällen beträgt dieses Intervall 2 bis 10 Sekunden. Jede Zeiteinheit besteht aus einem oder mehreren Adaptionssets z. B. jeweils für die jeweiligen Audiodaten, Videodaten oder Untertitel. Jedes Adaptionset besteht wiederum aus verschiedenen Representationen für unterschiedliche Endgeräte (z. B.

Unterscheidung via Bitrate oder Auflösung). In den einzelnen Representationen werden jeweils wiederum einzelne Segmente gespeichert, die durch eine eigene URL per HTTP-GET aufrufbar sind [3, S. 1].

Durch diese modulare Struktur eignet sich MPEG DASH sehr gut zur Auslieferung von H264-SVC enkodierte Videoströmen. Die einzelnen Baselayer und Enhancementlayer können jeweils in eigene Representationen gespeichert werden und innerhalb der Adaptationsets referenziert werden. [16, S. 14].

Die Zuordnung der einzelnen Layer erfolgt dabei über zusätzliche Attribute in der MPD-Datei. Jede Representation, die ein Layer definiert, erhält ein ID-Attribut und ein DependencyId -Attribut. Dieses gibt an auf welche vorherige Layer, sich dieser Layer bezieht und welche zuvor dekodiert werden müssen[16, S. 16].

4. H264-SVC Hands-on

4.1. SVC-Referenzcodec

Zur Evaluation der H264-SVC-Erweiterung wird der *JSVM-COdec* (Joint Scalable Video Model) verwendet. Dieser ist der Referenzcodec des Joint Video Team (JVT), der ISO/IEC Moving Pictures Experts Group (MPEG) und der ITU-T Video Coding Experts Group (VCEG). Neben dem Encoder und Decoder sind im JMVC-Softwarepakets auch Multiplexer und Testprogramme enthalten.

Das JMVC-Softwarepaket ist im Quelltext frei verfügbar. Der Quelltext für das Softwarepaket ist in C++ geschrieben und auf dem Webservice der Rheinisch-Westfälische Technische Hochschule Aachen (RWTH-Aachen) in einem CVS-Repository gehostet. Um dieses Abzurufen ist eine geeigneten CVS-Software wie Tortoise notwendig

Die hier verwendete Version des Encoder ist die Version XXXX.

Authentifizierungstyp	pserver
Hostadresse	garcon.ient.rwth-aachen.de
CVS-Pfad	/cvs/jvt
Benutzername	jvtuser
Passwort	jvt.Amd.2
Modulname	jsvm oder jsvm _{red}

Tabelle 4.1.: Zugangsdaten für das CVS-Repository der RWTH Aachen

4.1.1. Installation

Nach dem lokalen Download aus dem CVS der RWTH-Aachen befindet sich der Quellcode mit allen benötigten Klassen und Bibliotheken im Unterordner *jsvm/h264Extension/build/windows*.

Zur Vereinfachung der Installation werden bereits vorkonfigurierte Projektmappen für Microsoft Visual Studio mitgeliefert. Nach Öffnen der Projektmappe in Visual Studio wird mit der Menüoption *Create Batch/ Projektmappe erstellen* der Quellcode kompiliert. Nach erfolgreichem Kompilieren sind alle ausführbaren Programme im Unterordner *jmvc/bin* zu finden.

Die im Test verwendeten Beispielvideodateien können als YUV-Dateien unter folgender URL downgeloadet werden: <ftp://ftp.tnt.uni-hannover.de/pub/svc/testsequences/>

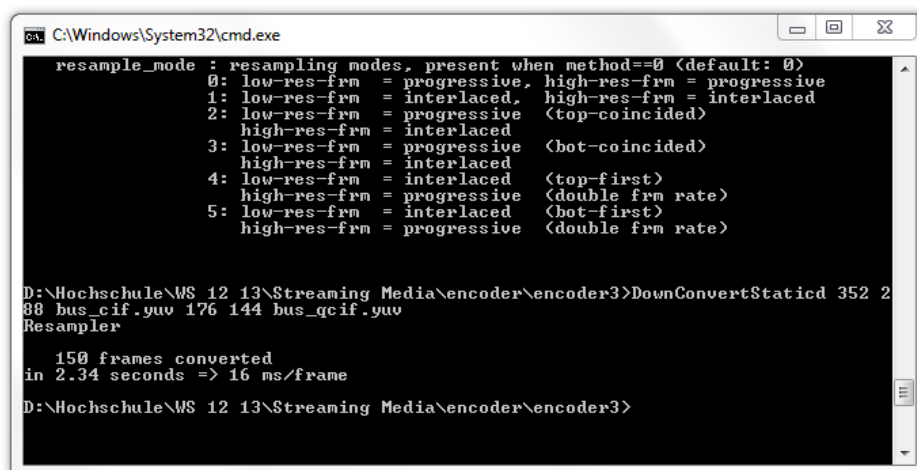
4.1.2. Vorbereitung

Der Referenzeencoder benötigt für die Erstellung eines beliebigen skalierbaren Videostroms mehrere Versionen eines Quellvideos, das sich in Bezug auf Bildgröße, Framerate oder Bildqualität unterscheidet. Da diese unterschiedlichen Versionen meist nicht vorliegen, besitzt der Referenzeencoder bereits ein Unterprogramm zur Erstellung dieser Videodateien. Das Programm heißt „DownConvertStatic“ und wird mit folgender Syntax aufgerufen:

```
DownConvertStatic 352 2 88 bus_cif.yuv 176 144 bus_qcif.yuv
```

Listing 4.1: Aufruf DownConvertStatic

Anschließend befinden neben der ursprünglichen YUV-Datei weitere Videodateien mit unterschiedlicher spatialer Qualität im Verzeichniss.



```

C:\Windows\System32\cmd.exe
resample_mode : resampling modes, present when method==0 (default: 0)
0: low-res-frm = progressive, high-res-frm = progressive
1: low-res-frm = interlaced, high-res-frm = interlaced
2: low-res-frm = progressive <top-coincided>
   high-res-frm = interlaced
3: low-res-frm = progressive <bot-coincided>
   high-res-frm = interlaced
4: low-res-frm = interlaced <top-first>
   high-res-frm = progressive <double frm rate>
5: low-res-frm = interlaced <bot-first>
   high-res-frm = progressive <double frm rate>

D:\Hochschule\WS 12 13\Streaming Media\encoder\encoder3>DownConvertStaticd 352 2
88 bus_cif.yuv 176 144 bus_qcif.yuv
Resampler

150 frames converted
in 2.34 seconds => 16 ms/frame
D:\Hochschule\WS 12 13\Streaming Media\encoder\encoder3>

```

Abbildung 4.1.: Referenzeencoder

4.1.3. Enkodieren

Zur Erzeugung eines skalierbaren Videostroms werden mehrere Konfigurationsdateien benötigt. Einerseits wird eine allgemeine Konfigurationsdatei benötigt, in welcher die Eigenschaften der Quelldatei definiert werden. Dazu zählen u.a. der Dateiname, die Auflösung, die zu enkodierenden Frames und die Größe der Group of Pictures. Zusätzlich werden die layerspezifischen Konfigurationsdateien in dieser Konfigurationsdatei verlinkt.

```
# JSVM Main Configuration File
```

```
OutputFile          test.264    # Bitstream file
FrameRate           30.0        # Maximum frame rate [Hz]
FramesToBeEncoded   150         # Number of frames
GOPSize             16          # GOP Size (at maximum frame rate)
BaseLayerMode       2           # Base layer mode
SearchMode          4           # Search mode
SearchRange         32          # Search range (Full Pel)
NumLayers           1           # Number of layers
LayerCfg            layer0.cfg  # Layer configuration file
```

Listing 4.2: Konfigurationsdatei des Encoders

Weiterhin wird für jeden Layer eine zusätzliche Konfigurationsdatei benötigt. In dieser spezifischen Datei werden u.a. der Dateiname der herunterskalierten Quelldatei, die Auflösung, Bitrate und die Framerate für diesen Layer definiert.

```
# JSVM Layer Configuration File
```

```
InputFile           BUS_CIF30.yuv # Input file
SourceWidth         352           # Input frame width
SourceHeight        288           # Input frame height
FrameRateIn         30           # Input frame rate [Hz]
FrameRateOut        30           # Output frame rate [Hz]
InterLayerPred      2            # Inter-layer Pred.
```

Listing 4.3: Konfigurationsdatei für einen Layer

Der Encoder **H264AVCEncoderLibTestStatic** wird über folgenden Befehl aufgerufen:

H264AVCEncoderLibTestStatic pf Konfigurationsdatei.cfg

Listing 4.4: Aufruf H264AVCEncoderLibTestStatic

Wenn die Enkodierung erfolgreich abgeschlossen ist, gibt er Encoder anschließend eine Meldung über die enkodierten und eingebetteten Layer aus. Die entsprechende enkodierte Videodatei befindet sich Im Verzeichniss des Enkoders unde trägt die Dateendung .h264.

```

CA: C:\Windows\system32\cmd.exe
AU 248: P T2 L0 Q0 QP 35 Y 31.4242 U 39.4237 U 41.5078 15984 bit
AU 248: P T2 L1 Q0 QP 34 Y 28.8106 U 37.7434 U 39.9488 89696 bit
AU 244: B T3 L0 Q0 QP 36 Y 31.5760 U 39.4932 U 41.7155 8296 bit
AU 244: B T3 L1 Q0 QP 35 Y 28.1347 U 38.1003 U 40.1135 59832 bit
AU 242: B T4 L1 Q0 QP 37 Y 29.6180 U 38.5340 U 40.5434 7368 bit
AU 246: B T4 L1 Q0 QP 37 Y 29.2819 U 38.3331 U 40.5182 7128 bit

SUMMARY:
      bitrate  Min-bitr  Y-PSNR  U-PSNR  U-PSNR
352x288 @ 0.9375  73.6378  73.6378  34.2445  40.0239  41.6075
352x288 @ 1.8750  95.2941  95.2941  33.5317  39.9046  41.5085
352x288 @ 3.7500 119.6775 119.6775 32.9233 39.8249 41.4088
352x288 @ 7.5000 150.0600 150.0600 32.5389 39.7384 41.3042
352x288 @ 0.9375 239.6316 239.6316 34.1508 39.5330 41.0978
352x288 @ 1.8750 369.6759 369.6759 31.9007 38.6521 40.2152
352x288 @ 3.7500 565.6416 565.6416 30.3757 38.1891 39.7394
352x288 @ 7.5000 848.1448 848.1448 29.1988 38.0126 39.5286
352x288 @ 15.0000 902.9626 902.9626 29.5107 38.1703 39.7392

Encoding speed: 5865.360 ms/frame, Time:1466340.000 ms, Frames: 250
D:\Hochschule\WS 12 13\Streaming Media\encoder>

```

Abbildung 4.2.: Referenzencoder

4.1.4. Dekodieren

4.1.5. Bitstreams extrahieren

4.1.6. Evaluation

Inputfile muss unkomprimiertes YUV-Video sein -> Großes Files, Umkodieren notwendig
Nicht Geschwindigkeitsoptimiert/langsam/nur Singlecore (10s pro Frames)

4.2. Kommerzielle Anbieter/Verwendung

4.2.1. Google Talk

4.2.2. Mainconcept SVC

5. Konklusion

Die Verwendung von H264-SVC zur Distribution von Videoformaten bringt entscheidende Vorteile gegenüber herkömmlichen Verfahren, wie das Simulcastverfahren, also der Übertragung mehrerer unabhängiger Videoströme in unterschiedlicher Qualität.

Im praktischen Anwendungsfall, dass gleichzeitig ein Videostrom an mobile Endgeräte, HD-Ready-Geräte und Geräte mit FullHD- Auflösung übertragen wird, braucht man durch die Nutzung von spatialen Enhancementlayers keine dedizierten Videodateien für bestimmte Endgeräteklassen zu enkodieren und archivieren. Stattdessen wird ein Baselayer für alle Endgeräte enkodiert, der nur den Videostrom in SD-Auflösung enthält. Mobile Endgeräte dekodieren ausschließlich diesen Baselayer. Geräte mit HD-Ready-Auflösung dekodieren zusätzlich den ersten Enhancementlayer in HD-Ready-Auflösung, der die Differenz zum Baselayer enthält. Für Geräte mit FullHD-Display werden alle Enhancementlayer dekodiert. Dadurch ist man sehr flexibel in Hinblick auf die entsprechenden Endgeräte und kann sehr einfach für weitere Geräte zusätzliche Enhancementlayer definieren. [10, S. 602]

Nachteilig an H264-SVC ist allerdings, dass es derzeit keine große Anzahl an kommerziellen Anbieter von SVC-basierten Applikationen gibt, die den kompletten Workflow von Enkodierung, Übertragung und Dekodierung abbilden. Dies ist daran zu begründen, dass in der Regel nicht SVC-basierende Lösungen einfacher zu implementieren sind, da sie auf etablierten Techniken wie reguläres H264/AVC basieren und dadurch mittlerweile einen gewissen Reifegrad erreicht haben.

Allerdings schreitet die Diversifikation der Endgeräte und Übertragungsmedien weiter voran, sodass auf lange Sicht durch Simulcast nicht alle Geräteklassen, Qualitätsstufen und Übertragungsmedien abbildbar sind und somit nicht für jedes Gerät die optimale Videoübertragung verfügbar ist.

Auf lange Sicht wird daher der Bedarf an skalierbaren Videocodecs weiter steigen und damit wird auch die Verbreitung von H264-SVC zunehmen [10, S. 603]

A. Abbildungsverzeichnis

3.1. Arten der Skalierbarkeit	5
3.2. Funktionsweise eines H264/AVC-Encoders	7
3.3. Base- und Enhancementlayer	8
3.4. Encodierung von Baslayer und Enhancementlayer	9
3.5. Dekodierung von spatialen Enhancementlayers	10
3.6. Temporale Skalierbarkeit	11
3.7. Media Presentation Document und H264-SVC	12
4.1. Referenzencoder	15
4.2. Referenzencoder	17

B. Literatur

- [1] Chuan-Yu Cho. “Overview on Scalable Video Coding - II”. In: (2010). URL: http://vc.cs.nthu.edu.tw/home/paper/codfiles/cycho/200507272019/Overview_on_Scalable_Video_Coding_-_II.ppt.
 - [2] Cisco Visual Networking Index: Forecast and Methodology. 2012. URL: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html.
 - [3] Heinrich Hertz Institut Fraunhofer. “Advanced MPEG Dash”. In: (2012). URL: <http://www.hhi.fraunhofer.de/de/kompetenzfelder/image-processing/research-groups/multimedia-communications/advanced-mpeg-dash.html>.
 - [4] Ralf Hohlfeld. *Crossmedia -s Wer bleibt auf der Strecke?: Beiträge aus Wissenschaft und Praxis*. 2010.
 - [5] Nikolaus Hottong. *Digitale AV-Technik WS 09/10*. 2009.
 - [6] Ronaldo Husemann Javier Del Ser Valter Roesler Aitor Rodríguez Pedro Sánchez Iraide Unanue Iñigo Urteaga. “A Tutorial on H.264/SVC Scalable Video Coding and its Tradeoff between Quality, Coding Efficiency and Performance”. In: (2011). URL: http://cdn.intechopen.com/pdfs/15138/InTech-A_tutorial_on_h_264_svc_scalable_video_coding_and_its_tradeoff_between_quality_coding_efficiency_and_performance.pdf.
 - [7] Iain E. Richardson. *H.264 and MPEG-4 Video Compression - Video Coding for Next-Generation Multimedia*. 2008.
 - [8] Iain E. Richardson. *The H.264 Advanced Video Compression Standard*. 2011.
-

- [9] Patrick Seeling Lina J. Karam Rohan Gupta Akshay Pulipaka und Martin Reisslein. “H.264 Coarse Grain Scalable (CGS) and Medium Grain Scalable (MGS) Encoded Video: A Trace Based Traffic and Quality Evaluation”. In: (2012). URL: http://mre.faculty.asu.edu/CGS_MGS_Traffic.pdf.
 - [10] Ellie Sader. “Scalable Video Coding (SVC) & Broadcast Delivery of 1080P”. In: (2008).
 - [11] Jörg Schieb. “Das Internet wird immer schneller - aber nur langsam”. In: (2012). URL: http://wdrblog.de/joergschieb/archives/2012/08/internet_tempo.html.
 - [12] Heiko Schwarz, Detlev Marpe und Thomas Wiegand. “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard”. In: *IEEE Transactions on Circuits and Systems for Video Technology* (2007). URL: http://ip.hhi.de/imagecom_G1/assets/pdfs/Overview_SVC_IEEE07.pdf.
 - [13] Puech Shahid Chaumont. “Scalable Video Coding”. In: (2010). URL: http://cdn.intechopen.com/pdfs/15732/InTech-Scalable_video_coding.pdf.
 - [14] Offer Shapiro. “H.264/SVC (Scalable Video Coding) - New Video Compression Standard”. In: (2009). URL: http://www.isccast.com/podcasts.isc365.com/powerpoint/09ISCWest_DI08.pdf.
 - [15] Stefan Slivinski. *SVC Demystified: What is Temporal Scalability?* 2011. URL: <http://www.lifesize.com/videoconferencingspot.com/?p=1406>.
 - [16] Yago Sánchez. “Title: Scalable Video Coding based DASH for efficient usage of network resources”. In: (). URL: http://www.w3.org/2011/09/webtv/slides/w3c_presentation_v4.pdf.
 - [17] Brian Stelter. “Ownership of TV Sets Falls in U.S.” In: (2011). URL: <http://www.nytimes.com/2011/05/03/business/media/03television.html>.
 - [18] Mathias Wien. “Performance Analysis of SVC”. In: (). URL: http://ip.hhi.de/imagecom_G1/assets/pdfs/PerformanceAnalysis_SVC_IEEE07.pdf.
 - [19] Joe Zaller. “Analyzing Where is Money Being Spent in the Broadcast Industry The 2012 BBS Broadcast Industry Global Project Index”. In: (2012). URL: <http://>
-

`blog.devoncroft.com/2012/08/08/analyzing-where-is-money-being-spent-in-the-broadcast-industry-the-2012-bbs-broadcast-industry-global-project-index/`.

- [20] Joe Zaller. “Tracking the Evolution of Broadcast Industry Trends 2009–2012”. In: (2012). URL: `http://blog.devoncroft.com/2012/04/09/tracking-the-evolution-of-broadcast-industry-trends-2009-2012/` (Stand 11.11.2012).
- [21] *Zukunft des Fernsehens: HbbTV für zusätzliche Informationen aus dem Internet*. 2011. URL: `http://www.techfacts.de/news/technik/zukunft-des-fernsehens-hbbtv-fuer-zusaetzliche-informationen-aus-dem-internet`.