

Skalierbare Videocodierung (SVC): Technologie und Anwendungen

T. Wiegand, H. Schwarz, D. Marpe, T. Schierl, K. Grüneberg,
Fraunhofer-Institut für Nachrichtentechnik – Heinrich-Hertz-Institut (HHI)
[\[wiegand | hschwarz | marpe | schierl | grueneberg\]@hhi.fhg.de](mailto:wiegand | hschwarz | marpe | schierl | grueneberg@hhi.fhg.de)

Einführung

Die Anzahl digitaler Übertragungswege nimmt ständig zu. Im Rahmen dieser Tagung wurden die Entwicklungen im Bereich Mobile TV wie DVB-H und MBMS (Multimedia Broadcast/Multicast Services der Mobilfunkbetreiber), sowie TV-Streaming über schnelle Internetzugänge diskutiert. Weiterhin sind auch drahtlose In-House-Netzwerke zunehmend in die Übertragungskette mit einbezogen.

Außerdem können audiovisuelle Medien auf ganz unterschiedliche Weise konsumiert werden: vom Handy-Display bis zur HD-Projektion im digitalen Heimkino. Sicherlich sind der Wiederverwertung desselben Inhalts hinsichtlich dieser Vielfalt auch praktische Grenzen gesetzt. Nicht jeder Inhalt lässt sich sowohl auf dem Großbildschirm als auch im Taschenformat verwenden.

Verschiedene Übertragungskanäle erfordern eine jeweils angepasste Codierung des Bildmaterials, die bei effizienter Ausnutzung der verfügbaren Datenrate eine möglichst gute Bildqualität sicherstellt, wobei Nutzer-Anforderungen, z.B. die verwendete Displaygröße oder die Betrachtungssituation, eine Rolle spielen.

Skalierbare Videocodierung stellt innerhalb desselben Bitstroms gleichzeitig mehrere Auflösungen, unterschiedliche Bildfrequenzen und variable Bildqualität zur Verfügung. Damit ist eine Anpassung an verschiedene Übertragungswege oder Endgeräte ohne erneute Codierung durch Auswahl von Teilen des Datenstroms möglich.

Die skalierbare Erweiterung des H.264/MPEG4-AVC-Standards

Skalierbare Videocodierung ist an sich nichts Neues. Frühere Ansätze und Standards für die skalierbare Codierung wurden jedoch nur selten angewendet, da für den zusätzlichen Nutzen stets ein erheblicher Verlust an Codiereffizienz in Kauf genommen werden musste und zudem die Komplexität sehr hoch war.

Im Rahmen der Standardisierung bei der ISO/IEC JTC 1/SC 29/WG 11 (Moving Picture Experts Group - MPEG) wurde die Suche nach möglichst effektiven Verfahren zur skalierbaren Videocodierung als eine von vielen Aktivitäten im Jahr 2001 aufgenommen. Für April 2004 wurde dann ein "Call for Proposals for Scalable Video Coding (SVC) Technology" ausgegeben bei dem es 14 Vorschläge gab: 12 basierten auf Wavelets und 2 auf einer Erweiterung von H.264/MPEG4-AVC. Bei allen Vergleichen der verschiedenen Verfahren stellte sich

schließlich im Oktober 2004 heraus, dass der Vorschlag des HHI auf Basis einer Erweiterung des H.264/MPEG4-AVC-Standards gegenüber den konkurrierenden Wavelet-Verfahren die weitaus besseren Ergebnisse liefert und weniger komplex ist.

Sowohl MPEG als auch die ITU-T SG16 Q.6 (Video Coding Experts Group – VCEG) beschlossen, das SVC-Projekt als eine Erweiterung des H.264/MPEG4-AVC-Standards [1] in ihrer gemeinsamen Standardisierungsorganisation dem Joint Video Team abzuschließen. Der Vorschlag des HHI wurde dafür als Arbeitsgrundlage gewählt und wird im folgenden mit SVC bezeichnet [2]. Eine wichtige Eigenschaft von SVC ist die, dass der größte Teil der H.264/MPEG4-AVC-Komponenten unverändert verwendet wird, insbesondere die bewegungskompensierte Prädiktion, die Intra-Prädiktion, Transformation, Entropiecodierung, das Deblocking-Filter sowie die Paketierung in NAL (Network Abstraction Layer) Units. Der "Base Layer" wird i.a. dem bisherigen H.264/MPEG4-AVC-Standard entsprechend codiert, so dass jeder standard-konforme H.264/MPEG4-AVC-Decoder diesen decodieren kann, wenn ihm ein SVC-Bitstrom angeboten wird. Hinsichtlich örtlicher und SNR-Skalierbarkeit führt SVC neue Tools ein.

Überblick

Das hier vorgestellte SVC-Verfahren ist ein sogenannter "Layered Video Codec", d.h. SVC operiert in verschiedenen Codierschichten und ermöglicht dabei Skalierbarkeit in drei Dimensionen:

- Zeitlich, d.h. variable Bildfrequenz
- Örtlich, d.h. variable Bildauflösung
- SNR, d.h. hinsichtlich der Bildqualität

Coder-Struktur und Effizienz der Codierung hängen von der Anwendung und ihren Anforderungen hinsichtlich der Skalierbarkeit ab. Bild 1 zeigt eine mögliche Coder-Struktur.

Innerhalb jeder örtlichen oder grob abgestuften („layered“) SNR Schicht werden die bewegungskompensierte Prädiktion und die Intra-Prädiktion wie in H.264/MPEG4-AVC angewendet. Die Redundanz zwischen den Schichten (Layers) wird durch zusätzliche Inter-Layer-Prädiktion ausgenutzt, wobei Bewegungs- und Texturinformation (Intra- und bewegungskompensiert) herangezogen werden. Als Basis-Repräsentation der Eingangsbilder jedes Layers wird durch Transformationscodierung eine NAL Unit mit Textur- und Bewegungsinformation erzeugt, im untersten Layer ist diese kompatibel mit der H.264/MPEG4-AVC-basierten Single-Layer-Codierung.

Die Wiedergabequalität dieser Basis-Repräsentation kann durch zusätzliche codierte sog. Progressive Refinement Slices verbessert werden. Im Gegensatz zu allen anderen NAL Units können diese an beliebiger Stelle abgeschnitten werden, wodurch eine fein abgestufte Skalierbarkeit in Bezug auf die Bildqualität gegeben ist, die auch zur flexiblen Datenratenanpassung geeignet ist.

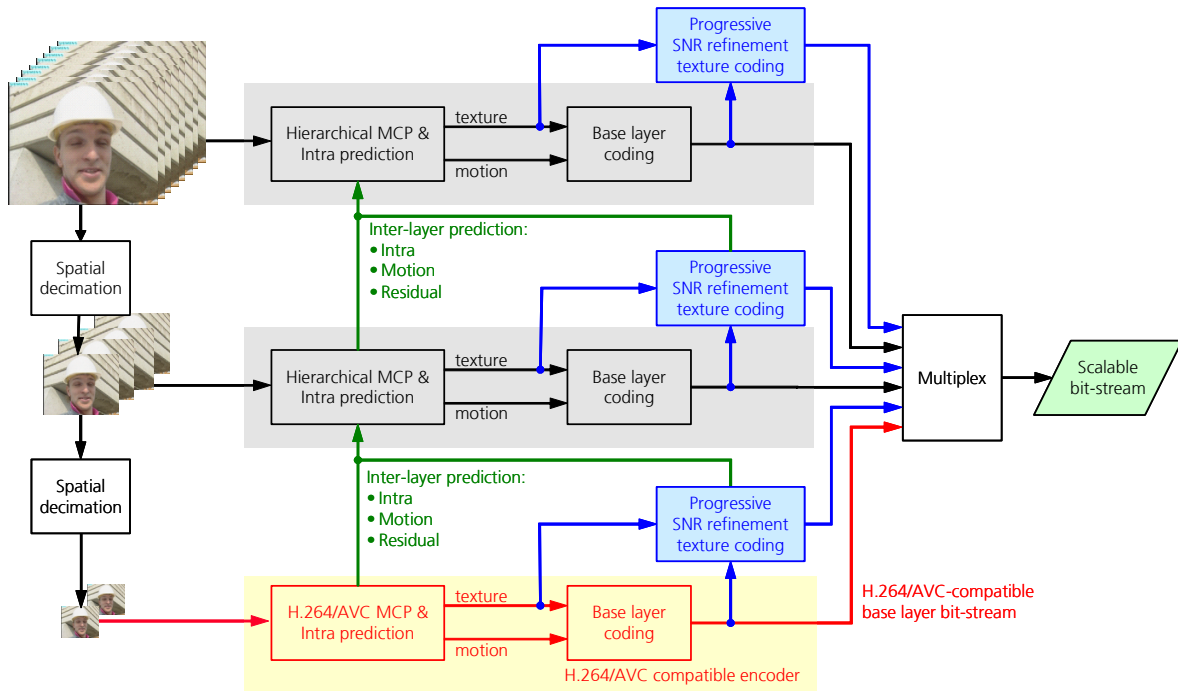


Bild 1 : Beispiel für eine SVC-Coder-Struktur mit drei örtlichen Layern

Zeitliche Skalierbarkeit und hierarchische Codierstrukturen

Im Gegensatz zu älteren Videocodierstandards wie MPEG-2 ist bei H.264/MPEG4-AVC die Reihenfolge, in der die Bilder codiert werden (Coding Order), vollständig von der Reihenfolge ihrer Präsentation (Display Order) entkoppelt. Jedes bereits codierte Bild kann als Referenzbild zur bewegungskompensierten Prädiktion nachfolgend codierter Bilder herangezogen werden. Dadurch ist die Codierung einer Bildsequenz mit beliebigen zeitlichen Abhängigkeiten möglich. Ohne jede Änderung des Standards lässt H.264/MPEG4-AVC die Erzeugung eines zeitlich skalierbaren Bitstroms durch hierarchische Prädiktionsstrukturen zu.

In regelmäßigen Intervallen werden sog. Key Pictures codiert, die ausschließlich vorangegangene Key Pictures als Referenz benutzen. Die Folge von Key Pictures stellen dabei die niedrigste zeitliche Auflösung einer Sequenz dar. Zwischen den Key Pictures werden in hierarchischer Weise prädizierte Bilder eingefügt. Durch das Hinzufügen jedes weiteren Layers wird die Bildfrequenz jeweils verdoppelt.

Bild 2 zeigt vier Varianten derselben Bildsequenz mit zunehmendem Abstand N von einem Key Picture zum nächsten. Um zwei Level zeitlicher Skalierbarkeit zu erreichen, wird nur ein bidirektional prädiziertes Zwischenbild (B-Bild) eingefügt, für drei Level wird zunächst das mittlere von drei Zwischenbildern von den nächstliegenden I- oder P-Bildern aus prädiziert (B_0), dann wird jeweils zwischen diesen B_0 -Bildern und dem benachbarten I- oder P-Bild ein weiteres Zwischenbild eingefügt (B_1). In Bild 2 sind die unterschiedlichen Level farblich hervorgehoben.



Bild 2 : Hierarchische B-Bilder

Durch eine geschickte Steuerung der Quantisierung kann mit hierarchischer Prädiktion eine besonders effiziente Codierung erreicht werden kann. Angefangen bei dem untersten (Basis-)Layer wird die Quantisierung des darauf aufbauenden (Enhancement-)Layers jeweils vergrößert. Das Ergebnis eines entsprechenden Experiments ist in Bild 3 dargestellt. In diesem Experiment wurde dieselbe Sequenz mit unterschiedlicher P-Bild-Distanz kodiert, wobei die Gesamtdatenrate konstant gehalten wurde.

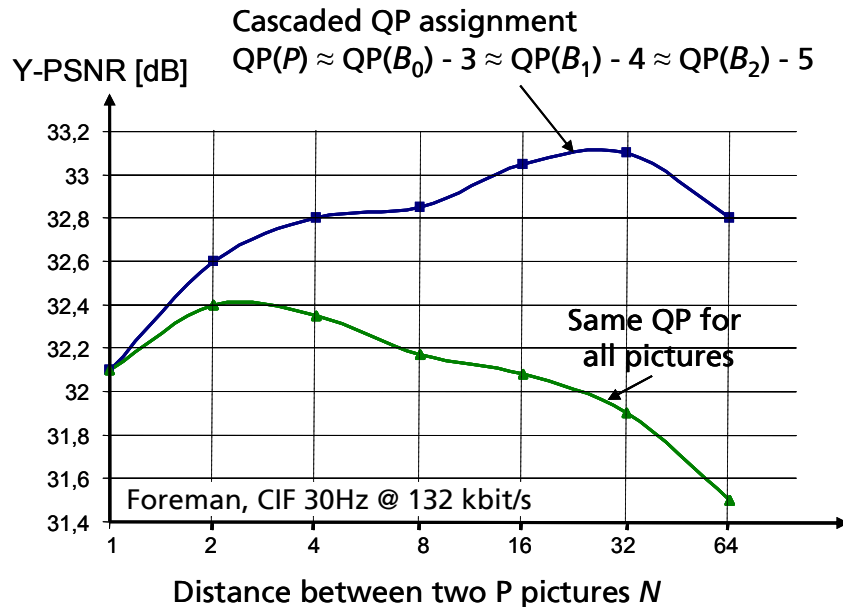


Bild 3 : Effizienzgewinn durch hierarchische B-Bilder

Die grüne Kurve zeigt das resultierende PSNR bei gleichem Quantisierungsparameter für alle Bilder. Wie zu erwarten, steigt die Effizienz bei Verwendung eines B-Bildes zunächst an, nimmt dann aber stetig ab. Die blaue Kurve zeigt, dass eine deutlich bessere Effizienz er-

reicht wird, wenn häufiger referenzierte Bilder (I , P , B_0) mit einem feineren Quantisierer bearbeitet werden als die weniger häufig (oder gar nicht) referenzierten (B_1 bzw. B_2). Erkauft wird dieser Effizienzgewinn durch eine erhöhte Verzögerung, wobei sich jedoch durch eine Beschränkung bei der Referenzierung zukünftiger Bilder die systematisch bedingte Verzögerung einer hierarchischen Prädiktionsstruktur steuern lässt [5].

Die hierarchische Prädiktionsstruktur verbessert außerdem die Effizienz der Tools für örtliche und SNR-Skalierbarkeit wie in den folgenden Abschnitten gezeigt wird.

Örtliche Skalierbarkeit

Zur möglichst effizienten Übertragung des Bildmaterials in mehreren örtlichen Auflösungen wird im Coder zunächst durch örtliche Filterung eine überbestimmte Bildpyramide erzeugt. Die Layer werden unabhängig voneinander mit eigenen Bewegungsparametern codiert. Um die Codiereffizienz gegenüber Simulcast zu erhöhen, wurden jedoch Inter-Layer Prädiktionsmechanismen eingeführt. Im Enhancement-Layer können Information der darunter liegenden Layer (z.B. Bewegungsvektoren) genutzt werden, soweit sich dadurch ein besseres Verhältnis von Bildqualität zu Datenrate ergibt. Hierzu sollten die zeitlichen Prädiktionsstrukturen der einzelnen Layer aufeinander abgestimmt sein. Anzumerken ist, dass alle zu einem bestimmten Zeitpunkt gehörigen NAL Units eine Access Unit bilden und daher im Bitstrom aufeinander folgen müssen.

Der SVC-Standard (im gegenwärtigen Entwurf) eröffnet dem Coder drei Möglichkeiten der Inter-Layer-Prädiktion. Über die im Folgenden beschriebenen Basis-Konzepte zur dyadischen örtlichen Skalierbarkeit hinaus kann die Auflösung auch in beliebigem Verhältnis abgestuft werden [3][4][6].

Inter-Layer Bewegungsprädiktion

Um die Bewegungsinformation des unteren Layers zu nutzen, wurden zusätzliche Makroblock-Modi für Enhancement-Layer eingeführt. Die Aufteilung eines Makroblocks ergibt sich aus der Hochtastung des entsprechenden 8×8 -Blocks im unteren Layer. Die Referenzbild-Indizes aus den entsprechenden Blöcken des Base-Layers werden kopiert, und die zugehörigen Bewegungsvektoren um den Faktor 2 skaliert. Diese Vektoren stellen Prädiktoren dar und können entweder direkt verwendet oder auf $\frac{1}{4}$ -Pixel-Genauigkeit verfeinert werden. Außerdem kann der skalierte Bewegungsvektor als Prädiktor für Bewegungsvektoren der konventionellen Makroblock-Modi im Enhancement-Layer verwendet werden.

Inter-Layer Residuumsprädiktion

Zu allen inter-codierten Makroblöcken in einem Enhancement-Layer wird ein Flag übertragen, ob dieser Modus verwendet wird. Ist dies der Fall, wird das Residuums-Signal des unteren Layers blockweise hochgetastet und als Prädiktion für das Residuumsignal des aktuellen Makroblocks verwendet, so dass nur noch die Differenz dazu codiert werden muss.

Inter-Layer Intra-Prädiktion

Darüber hinaus wird ein Makroblock-Modus eingeführt, in welchem die Prädiktion auf Basis des hochgetasteten rekonstruierten Signals für den entsprechenden Block des unteren Layers erfolgt. Da hierzu der Block vollständig dekodiert werden muss, was komplexe Algorith-

men hinsichtlich Bewegungskompensation und Deblocking einschließt, kann dieser Modus auf solche Bildbereiche beschränkt werden, die im unteren Layer intra-codiert wurden. Auf diese Weise wird bei der Decodierung jedes Layers maximal ein Durchlauf durch die Bewegungskompensation erforderlich [7].

Durch Auswahl des jeweils günstigsten Prädiktionsmodus lässt sich die Codiereffizienz optimieren. Beim Coder-Entwurf muss i.a. ein Kompromiss zwischen dem vertretbaren Rechenaufwand und der damit erreichbaren Performance gefunden werden.

Untersuchung der Leistungsfähigkeit

Die Leistungsfähigkeit der Prädiktionsmechanismen zur örtlichen Skalierbarkeit wurden mit Simulcast einerseits und Single Layer-Codierung andererseits verglichen. Der Base Layer wurde hierbei mit einer festen Datenrate codiert, hinsichtlich der örtlichen Enhancement-Layer wurden Datenrate und die verwendeten Intra-Layer Prädiktionsmechanismen variiert. Alle Coder wurden entsprechend [8] Rate-Distortion-optimiert. Die folgenden Bilder zeigen Ergebnisse für die Sequenz "Soccer" mit einem CIF- und einem 4CIF-Layer. Der Abstand zwischen aufeinander folgenden intra-codierten Bildern betrug 32 Bilder; simuliert wurde sowohl eine GOP-Länge von 16 Bildern Bild 4 (N=16) als auch eine reine IPPP-Sequenz Bild 5 (N=1).

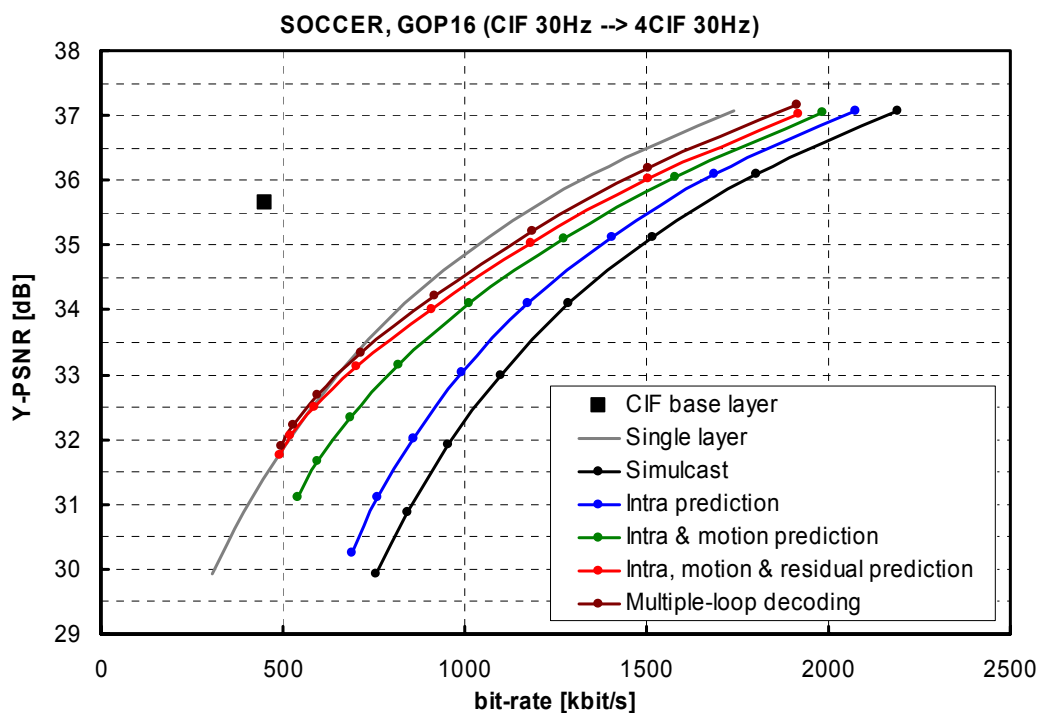


Bild 4 : Leistungsfähigkeit der Interlayer-Prädiktionsmechanismen bei hierarchischer GOP-Struktur

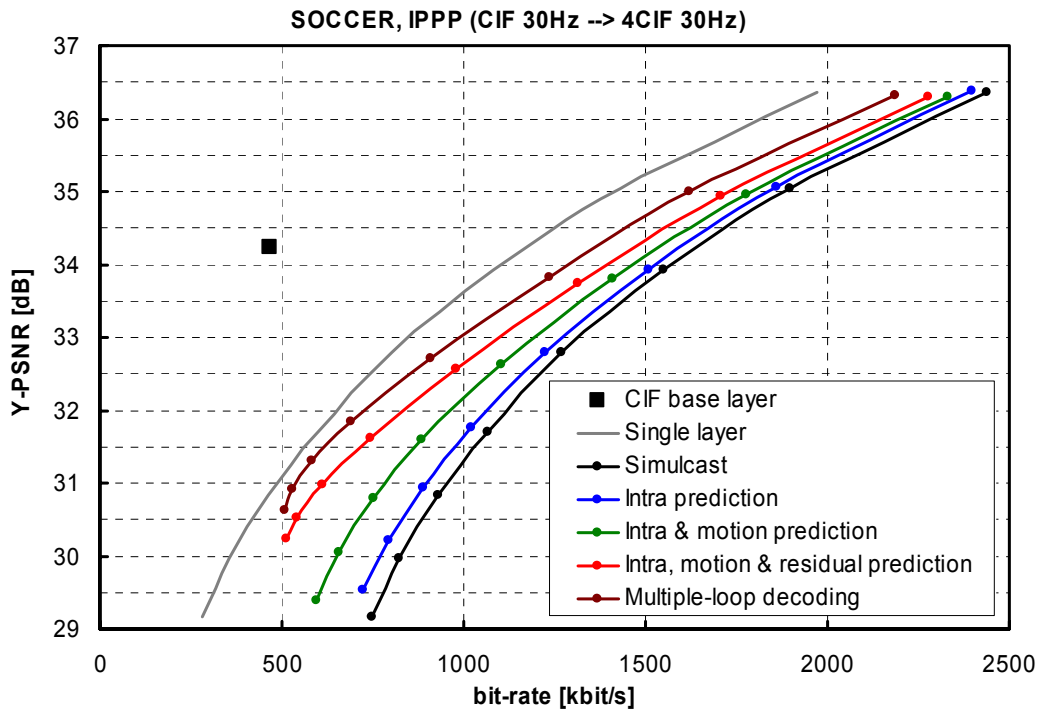


Bild 5 : Leistungsfähigkeit der Interlayer-Prädiktionsmechanismen bei IPPP-Codierung

Die schwarze Kurve repräsentiert eine Single-Layer-Codierung, die graue entspricht dem Simulcast-Fall. Bei der blauen, grünen und roten Kurve wurden nacheinander Inter-Layer-Intra-, Bewegungs- und Residuums-Prädiktion aktiviert. Bei diesen drei Kurven wurde die Inter-Layer-Prädiktion derart eingeschränkt, dass der Decoder die Bewegungskompensation nur einmal durchläuft. Die braune Kurve stellt die Effizienz örtlich skalierbarer Codierung ohne diese Restriktion dar, d.h. mit mehrfachem Durchlauf der Bewegungskompensation im Decoder. Durch einen Vergleich von Bild 4 mit Bild 5 zeigt sich die Effizienzerhöhung der Inter-Layer-Prädiktion durch die hierarchische Prädiktionsstruktur.

SNR Skalierbarkeit

Neben einer Skalierbarkeit in zeitlicher und örtlicher Richtung bietet SVC als dritte Option die Skalierung hinsichtlich der Bildqualität. Diese Option lässt sich sowohl in groben Stufen (Layered SNR Scalability) als auch in feinen Abstufungen (Progressive Refinement) unterteilen.

Zur Erzeugung grober Qualitätsabstufungen wird die Quantisierung der Transformationskoeffizienten in mehreren Schritten verfeinert. Von der Hochtastung abgesehen, kommen dieselben Mechanismen zur Anwendung wie bei der örtlichen Skalierbarkeit. Ein Blockschaltbild des Coders ist in Bild 6 dargestellt.

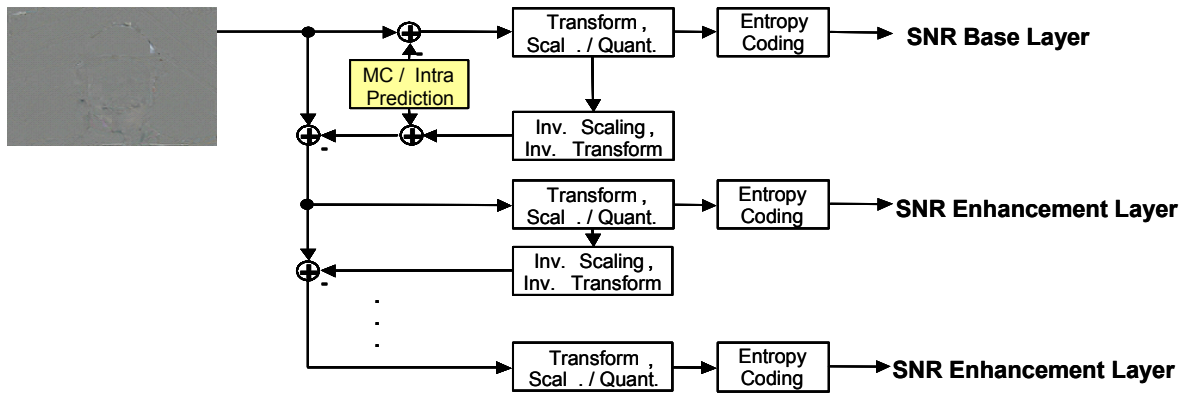


Bild 6 : Coder-Architektur für SNR Skalierbarkeit

Um zu feineren Abstufungen hinsichtlich Qualität und damit verbundener Datenrate zu kommen, wurden sog. Progressive Refinement (PR) Slices eingeführt. Von Layer zu Layer (in Bild 6 von oben nach unten) wird die Quantisiererstufe in etwa halbiert, was einer Erniedrigung des Quantisierungsparameters (QP) um den Wert 6 entspricht, und gleichzeitig wird eine Signaldarstellung gewählt, die decoderseitig eine Rücktransformation aller Koeffizienten in einem einzigen Schritt ermöglicht. Im Bitstrom werden zunächst die grob quantisierten Koeffizienten für alle Bildblöcke (SNR Base Layer) übertragen, die zugehörigen Verfeinerungen werden näherungsweise nach abnehmender Bedeutung für die Bildqualität, umsortiert und angefügt (s. Bild 7).

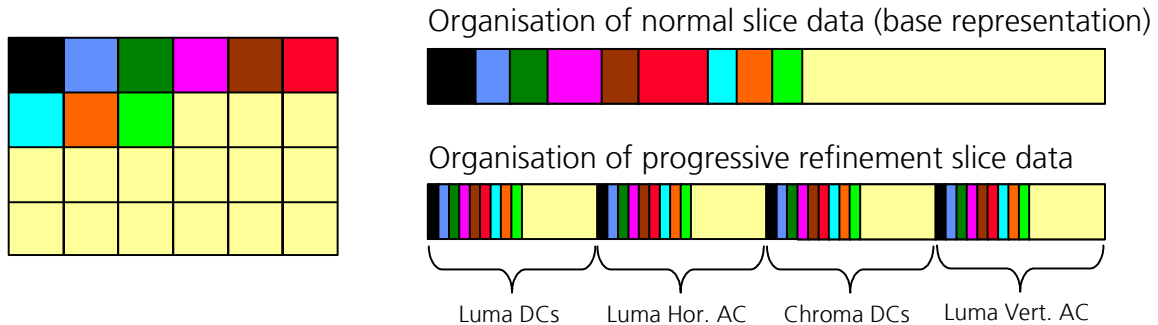


Bild 7 : Anordnung der Daten im FGS-Layer

Je mehr dieser Verfeinerungen zum Decoder übertragen werden, desto höher wird die Bildqualität. Dieser Teil des Datenstroms kann quasi an einer beliebigen Stelle abgebrochen werden. Dadurch lässt sich u.a. eine Anpassung an die gewünschte Datenrate auch nachträglich außerhalb des Coders erreichen.

Zur bewegungskompensierten Prädiktion von Bildern, die keine Key Pictures sind, werden die Referenzbilder in der höchsten verfügbaren Qualität herangezogen. Im Zusammenhang mit der hierarchischen Prädiktionsstruktur erhöht dies bei gleicher Komplexität die Codiereffizienz erheblich, im Vergleich zum früheren MPEG-4 FGS-Algorithmus, bei welchem stets nur die Base-Layer-Repräsentation verwendet wurde. Bei Key Pictures wird zur Bewegungskompensation nur die Base-Layer-Repräsentation verwendet um Fehlerfortpflanzung zu vermeiden (Bild 8).

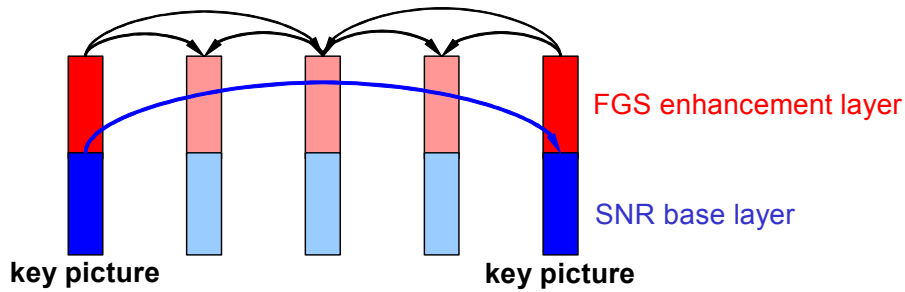


Bild 8 : Bewegungskompensierte Prädiktion mit FGS-Layer

Um die Codiereffizienz von PR-Slices insbesondere bei Anwendungen mit geringer Verzögerungszeit zu erhöhen, wurden sog. Leaky-Prediction-Konzepte für die Bewegungskompensation von Key Pictures in SVC integriert [3][4][9]. Weiterhin wurde die wahlweise Übertragung von Bewegungsparameterverfeinerungen als Teil der PR-Slices eingebaut [10].

Kombinierte Skalierbarkeit

SVC lässt eine beliebige Kombination in den drei Arten der Skalierbarkeit zu, d.h. ein Bitstrom kann dieselbe Bildsequenz mit verschiedener Bildfrequenz, in mehreren Auflösungen und mit variabler Bildqualität enthalten (s. Bild 9).

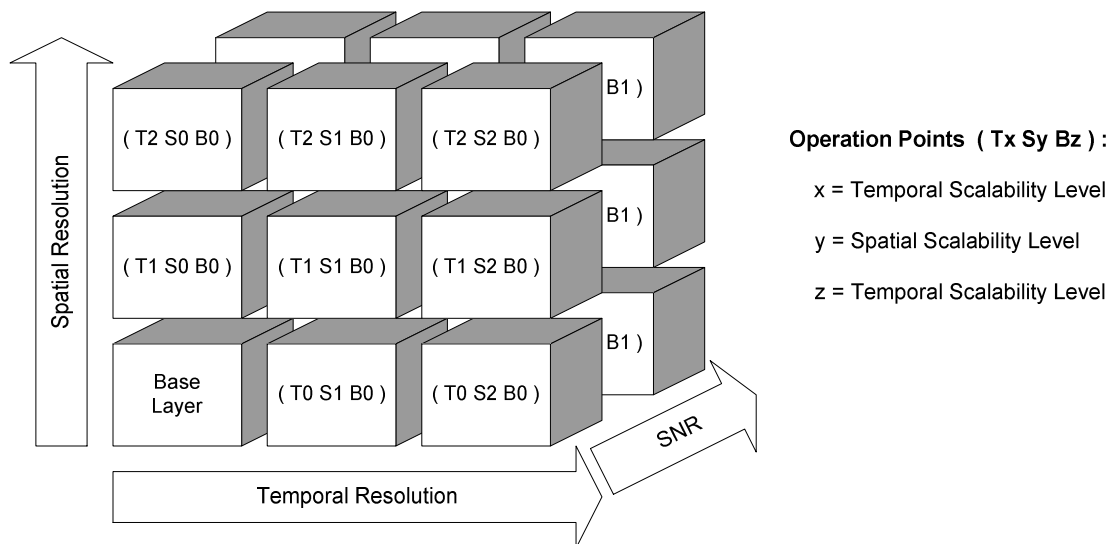


Bild 9 : Skalierbarkeitsdimensionen. zeitlich, örtlich und SNR (Qualität)

Es müssen jedoch nicht Daten für jede Kombination aus Bildfrequenz, Auflösung und Qualitätslevel in den Bitstrom eingebettet werden. In Bild 10 wird ein Beispiel gezeigt, das bei der geringsten örtlichen Auflösung auch nur eine reduzierte Bildfrequenz zur Verfügung stellt. Ein Decoder kann einen bestimmten Arbeitspunkt erreichen, wenn ihm auch die Daten der darunter liegenden referenzierten Level zur Verfügung stehen.

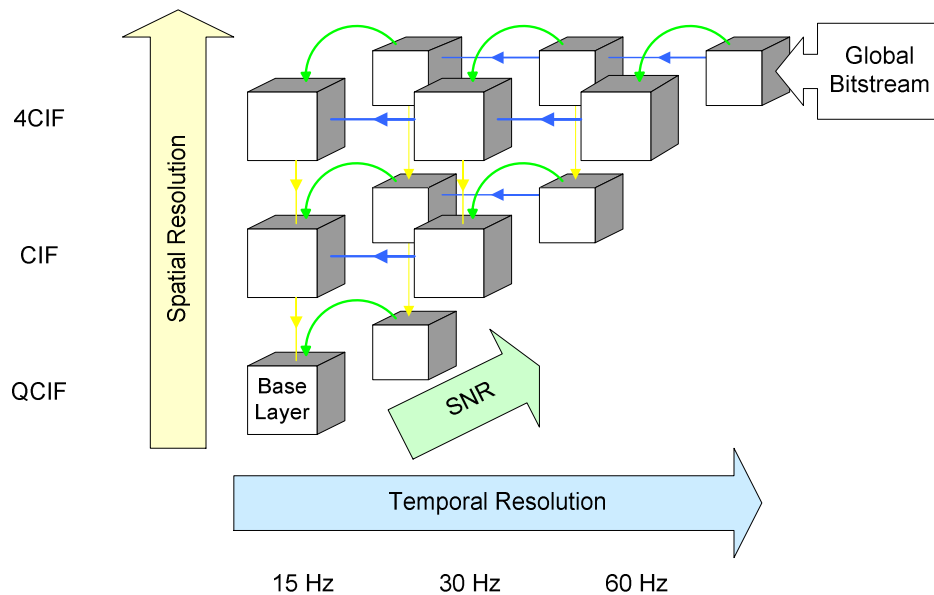


Bild 10 : Beispiel für mögliche Arbeitspunkte in den drei Skalierbarkeitsdimensionen zeitlich, örtlich und SNR (Qualität)

In Bild 11 wird die Codiereffizienz der kombinierten Skalierbarkeit anhand der Sequenz "Soccer" mit der Effizienz einer reinen Single-Layer-Codierung, einer rein örtlichen sowie einer reinen SNR-Skalierbarkeit verglichen. Der Abstand aufeinander folgender Intra-Bilder betrug 1,07 s (64 Bilder bei 60Hz); und es wurde stets eine dyadische hierarchische Prädiktionsstruktur mit einer GOP-Länge von 32 Bilder bei 60 Hz verwendet. Da die zeitliche Skalierbarkeit bei allen Bitströmen Bildfrequenzen von 1,875Hz bis 60Hz (bei 4CIF) unterstützt, wurde sie nicht separat untersucht.

Die schwarzen Kurven zeigen die Codiereffizienz einer Single-Layer-Codierung, wobei jeder Punkt einen separaten Bitstrom repräsentiert. Bei SNR-Skalierbarkeit wurde für jede örtliche Auflösung ein Bitstrom generiert, der durch die rote Kurve dargestellt wird. Die blauen Kurven zeigen die Effizienz der örtlich skalierbaren Codierung. Hierzu wurden je drei Bitströme zu jeder örtlichen Auflösung erzeugt, die jeweils entweder die niedrigste, die mittlere oder die höchste dargestellte Datenrate enthalten. Die Codiereffizienz der kombinierten skalierbaren Codierung, die alle dargestellten örtlich-zeitlichen Arbeitspunkte in demselben Bitstrom unterstützt, wird durch die grünen Kurven dargestellt. Wie aus dem Beispiel in Bild 11 deutlich wird, nimmt die Codiereffizienz der SVC-Bitströme mit zunehmender Anzahl unterstützter Arbeitspunkte ab.

Standardisierungsaktivitäten

Die SVC-Standardisierung als Erweiterung des H.264/MPEG4-AVC-Standards ist bereits weit fortgeschritten. Ihr Abschluss ist für Januar 2007 geplant. Derzeit wird u.a. noch an Verfahren für die Codierung von zeilenverschränkten Videosignalen gearbeitet.

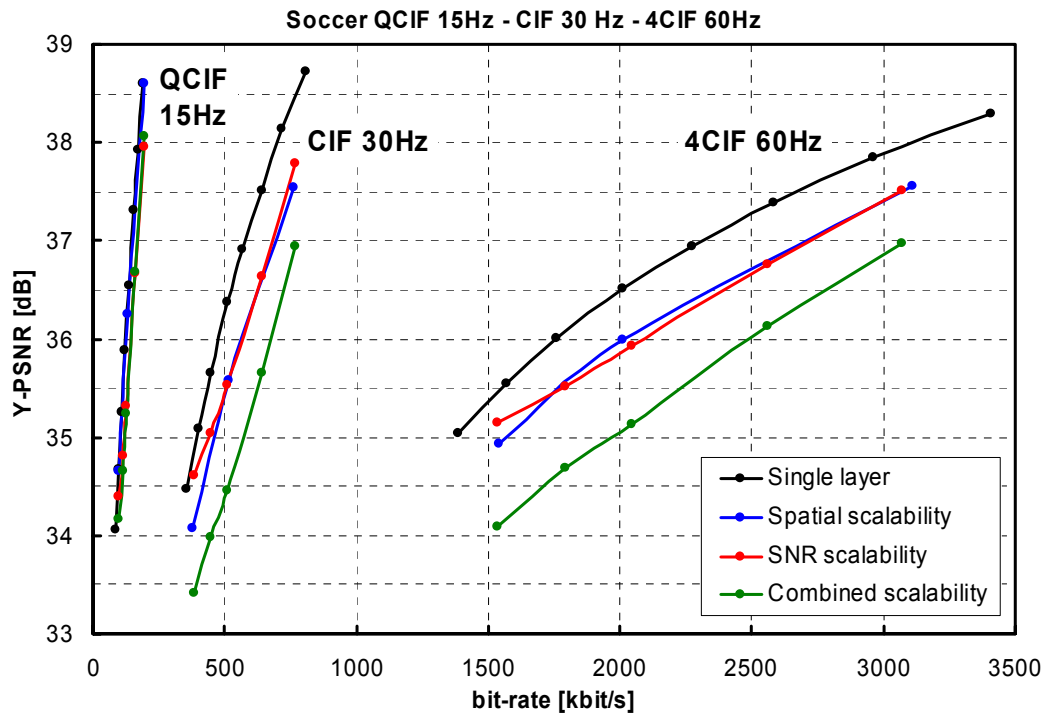


Bild 11 : Leistungsfähigkeit der kombinierten Skalierbarkeit in den Dimensionen zeitlich, örtlich und SNR (Qualität)

Speicherung

Parallel zum eigentlichen Codierungsstandard wird in MPEG auch an einer Erweiterung des MP4-Fileformats zur Speicherung von SVC-Bitströmen gearbeitet. Dabei werden mehrere Ziele verfolgt. Zum einem soll es einfach möglich sein, den kompatiblen Base Layer mit einem herkömmlichen H.264/MPEG4-AVC-Decoder aus dem File zu lesen. Ein weiteres wichtiges Ziel bei der Speicherung ist, den notwendigen Overhead so gering wie möglich zu halten. Dennoch soll sowohl die Vielfalt der möglichen Arbeitspunkte adressierbar bleiben, als auch ein einfacher Zugriff auf eine kleine Anzahl unterschiedlicher Stufen ermöglicht werden. Um diese Stufen nicht mit den bisher diskutierten Begriffen Layer bzw. Level zu verwechseln, wurde der Begriff "Tier" eingeführt (s. Bild 12).

Bei bestimmten Anwendungen ist es vorteilhaft, wenn aus einem gespeicherten skalierbaren Bitstrom die oberen Layer entfernt werden können, ohne die gesamte Datei bearbeiten zu müssen. Dieses sog. Erosion Storage Feature kann z.B. bei der Aufzeichnung von Überwachungsvideos eingesetzt werden, die über einen bestimmten Zeitraum in hoher Auflösung und Qualität gespeichert werden sollen, die aber aus Platzgründen nach einer bestimmten Zeit überschrieben werden. Auch eine Anwendung für Videorecorder mit Festplatte ist möglich. Mit Hilfe von SVC kann anstelle des vollständigen Löschens eine stufenweise Erosion stattfinden: Nach einer definierten Zeit kann die Qualität, Auflösung oder Bildfrequenz durch Löschen einzelner Layer (bzw. Tier) reduziert werden, so dass bei gleichem Speicherbedarf über einen längeren Zeitraum auf das aufgezeichnete Material zugegriffen werden kann. Für

das Fileformat bedeutet dies, dass die einzelnen Layer / Tier möglichst separierbar anstatt in einem Track gemultiplext abgelegt werden sollten.

Transport

Teil des H.264/MPEG4-AVC Standards ist der sog. Network Abstraction Layer (NAL), der die Übertragung von Bitströmen z.B. mit Hilfe des Internet-Protokolls unterstützt. Bei herkömmlicher Single Layer-Codierung beginnt jede NAL Unit mit einem Header von einem Byte, der Informationen über die Art der Payload enthält. Für SVC wurde dieser Header um ein weiteres Byte erweitert, welches bei SVC jede NAL Unit einem Arbeitspunkt in den drei Skalierbarkeitsrichtungen örtlich-zeitlich-SNR zuordnet.

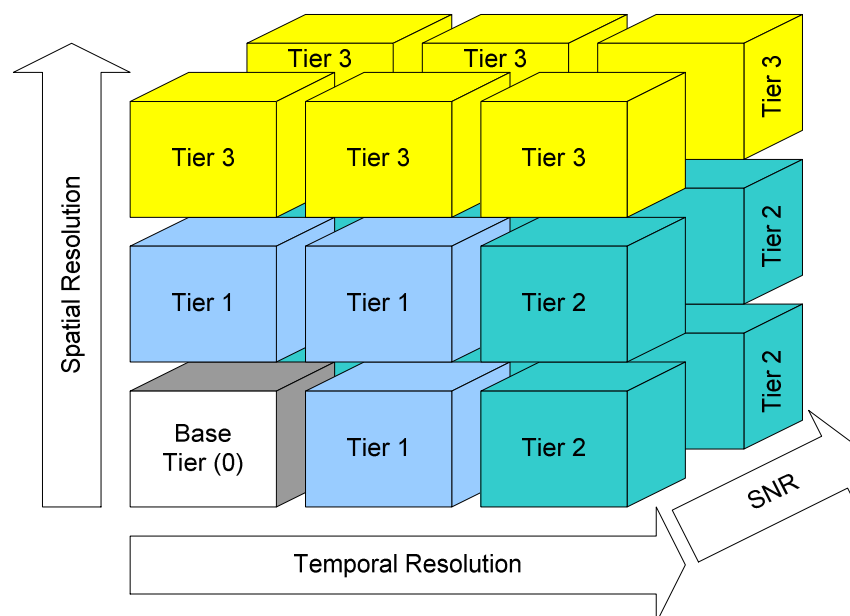


Bild 12 : Zusammenfassung mehrerer Level im MP4 Fileformat

Innerhalb der IETF wird derzeit ein entsprechendes Payloadformat für RTP (Real-Time Transfer Protocol) entwickelt. Die Wunschvorstellung bei der Übertragung eines skalierbaren Videostromes durch ein reales Netzwerk ist in Bild 13 dargestellt. Ein Server sendet einen Datenstrom per Broadcast oder Multicast in das Netzwerk, und dort, wo die verfügbare Datenrate nicht ausreicht, werden automatisch nur die oberen Layer entfernt. Allein auf der Ebene der Internet-Übertragungsprotokolle ist ein solcher Automatismus jedoch nicht zu erreichen, insbesondere wenn eine Kompatibilität zu den derzeitig verwendeten Streaming-Verfahren gefordert ist.

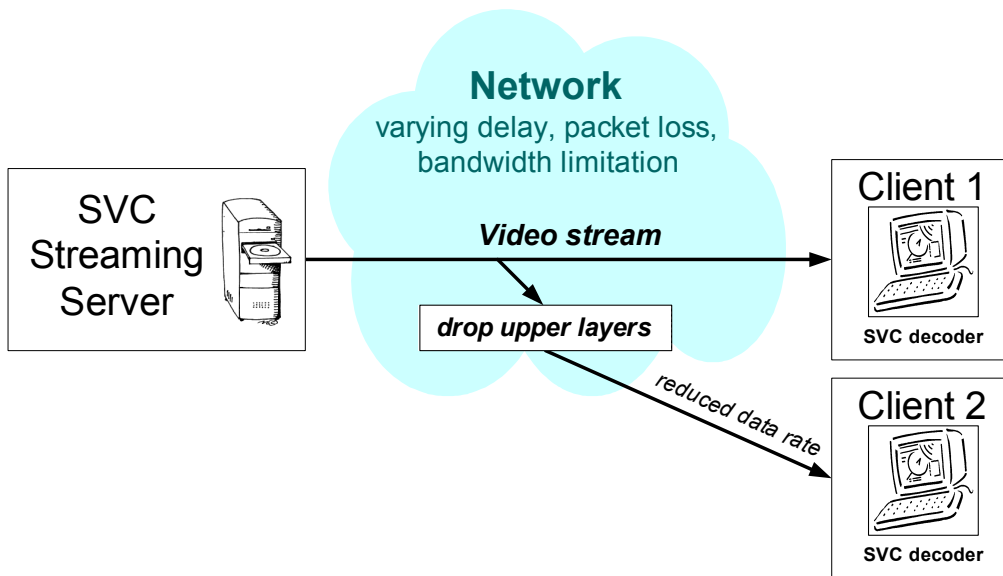


Bild 13 : Gewünschtes Netzwerkverhalten bei der Versorgung von Endgeräten mit unterschiedlicher Datenrate

Ein mögliches Konzept ist in Bild 14 dargestellt: Vorausgesetzt wird, dass bestimmte Elemente im Netzwerk existieren (z.B. Gateways mit erweiterter Funktionalität), die nicht nur Pakete weiterleiten, sondern abhängig vom Inhalt eines Pakets Entscheidungen über Weiterleitung oder Entfernung bestimmter Pakete fällen können. Ein solches MANE (Media Aware Network Element), das im Bereich des Zugangsnetzwerks angesiedelt sein sollte, könnte z.B. für jeden Layer eine einzelne RTP-Verbindung zu einem Server aufbauen und abhängig von der verfügbaren Datenrate mehr oder weniger Layer an den Client weiterleiten. Zur Weiterleitung kann der gesamte Teilstrom über eine einzelne RTP-Verbindung gesendet werden. Dies ist z.B. beim Passieren einer Firewall und bei einer Net Address Translation von Vorteil.

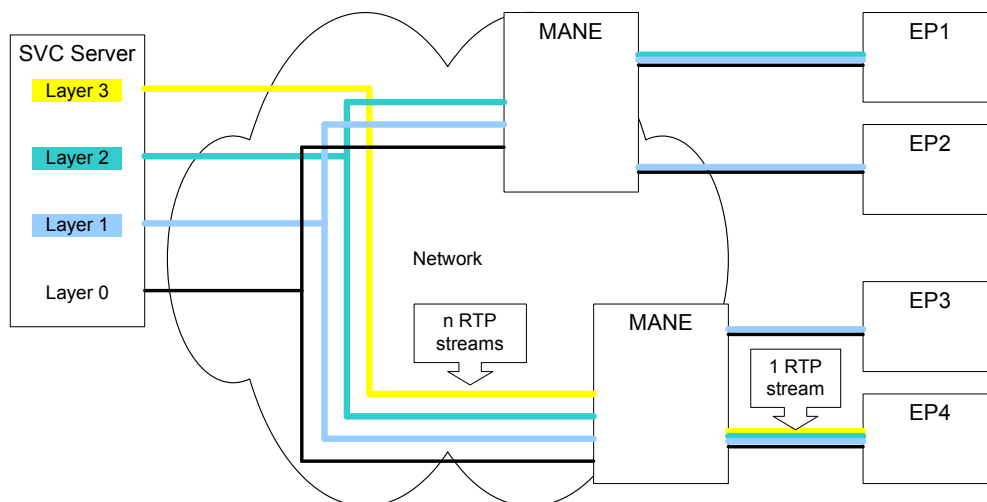


Bild 14 : Datenraten-Anpassung durch Media Aware Network Elements (MANEs)

Anwendungen

Uneven Level Protection

Die Kombination aus skalierbarer Videokodierung und einem Fehlerschutz, der die einzelnen Layer mit unterschiedlicher Redundanz versieht, eignet sich insbesondere für den Einsatz im mobilen Bereich, wo es auf größere Fehlerrobustheit ankommt, da sich die verfügbare Datenrate häufig ändert. Bei der IETF wird unter der Bezeichnung Uneven Level Protection derzeit ein entsprechendes Übertragungsprotokoll (nicht nur für SVC) standardisiert. Durch den stärkeren Schutz unterer Layer fällt die Verbindung bei Übertragungsfehlern nicht vollständig aus, sondern es können zunächst nur die zu den oberen Layern gehörigen Pakete nicht mehr korrigiert werden, eine Basisqualität wird jedoch nach wie vor empfangen.

Mobiles Ad-Hoc-Netzwerk

Die zunehmende Verbreitung drahtloser Netzwerke, insbesondere auch mobiler ad-hoc Netzwerke (MANETs) bringt neue Möglichkeiten für Media Streaming. Mehrere Server können den gleichen Inhalt auf verschiedenen Wegen zum Endgerät senden (s. Bild 15), wobei eine spezielle Fehlerschutzkodierung (Raptor-Codes) dafür sorgt, dass sich unabhängig voneinander generierte Codesymbole verschiedener Quellen ergänzen [11]. Bei optimaler Übertragung können auch die schwächer geschützten oberen Layer komplett decodiert werden. Bei ungünstigen Empfangsverhältnissen z.B. auf Grund eigener Bewegung oder durch den Ausfall einiger Übertragungswege reicht die Anzahl der empfangenen Symbole nur für die Decodierung unterer Level bzw. des Base Layers.

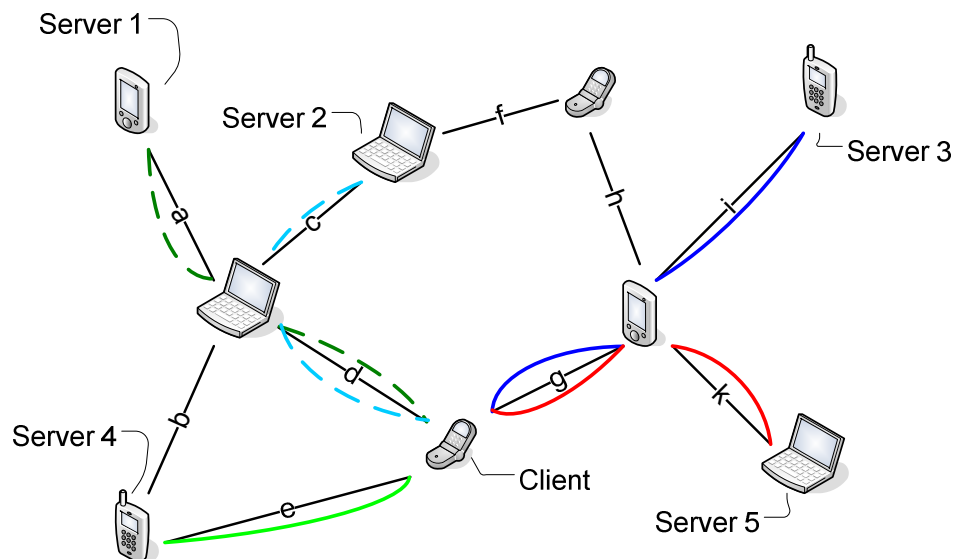


Bild 15 : Mobiles Ad-Hoc-Netzwerk (MANET)

Bild 16 zeigt, wie die Übertragung auf Verbindungsverluste innerhalb des MANET reagiert. Bis zum Zeitpunkt A wird der Client von den Servern 3, 4 und 5 mit dem SVC-Strom versorgt, die in ihrer Gesamtheit zur kompletten Dekodierung aller Layer ausreichen. Zum Zeitpunkt A

geht die Verbindung zu Server 4 verloren, die Redundanz der verbleibenden zwei Bitströme reicht nur zur Dekodierung der unteren beiden Layer. Zum Zeitpunkt B empfängt der Client einen Strom von Server 1 und die volle Qualität kann wiederhergestellt werden, bis zum Zeitpunkt C die Verbindungen zu den Servern 1 und 3 gleichzeitig ausfallen. Daraufhin kann die Fehlerkorrektur nur noch der Base-Layer wiederherstellen.

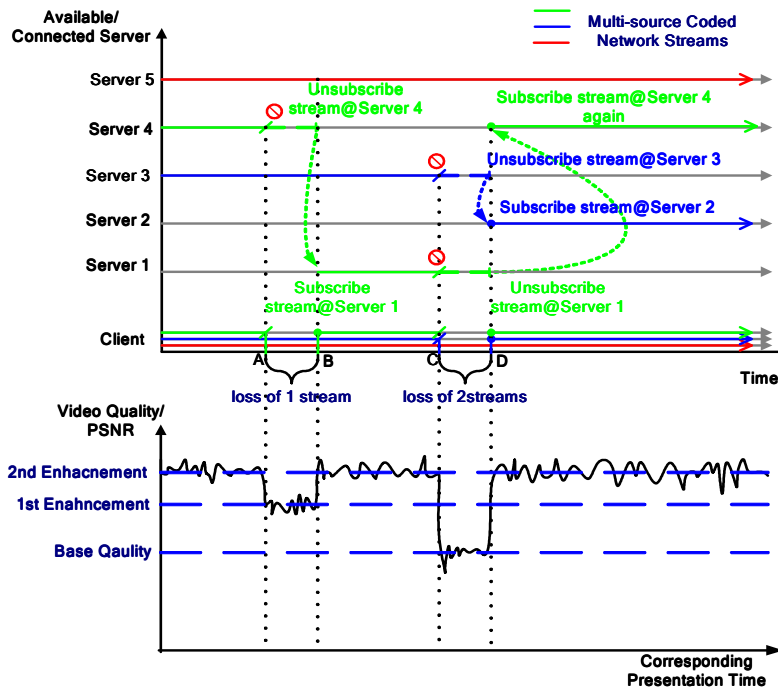


Bild 16 : Simulation einer SVC-Übertragung mit speziellem Fehlerschutz über ein MANET

Videoüberwachung und "Erosion Storage" im Heimnetzwerk

Das europäische Forschungsprojekt ASTRALS [12] hat sich die Entwicklung einer neuartigen Architektur für das überwiegend drahtlose Heimnetzwerk zum Ziel gesetzt (s. Bild 17). Mit einem leistungsfähigen Residential Gateway und optimierten drahtlosen Streaming-Verbindungen werden u.a. neue Features der Überwachungstechnik implementiert, z.B. die automatische Erkennung einer stürzenden Person. Datenübertragung und Speicherung erfolgt zum Teil mit SVC. Zeitliche Skalierbarkeit und Erosion Storage Features sind in diesem Zusammenhang von besonderem Interesse.

Video-Broadcasting

Die Vorteile skalierbarer Videokodierung für den Fernsehrundfunk liegen auf der Hand: Eine rückwärts-kompatible Einführung erweiterter Services ist ebenso möglich wie die Möglichkeit, einen Layer hoher Qualität verschlüsselt zu senden, während der Base Layer frei verfügbar ausgestrahlt wird. Die teure Ressource der Übertragungskapazität wird höchst effizient genutzt, bei der Übertragung von HD ist ein (inhaltsgleicher) SD-Kanal bereits mit inbegriffen. Höchst interessante Anwendungen sind die gemeinsame Ausstrahlung von SD + 1080i oder SD + 720p, aber in nicht allzu ferner Zukunft auch 1080i + 1080p. Bei DVB-H käme die Ska-

lierbarkeit zwischen QCIF und CIF (bzw. QVGA und VGA) in Betracht, aber auch CIF + SD zur gleichzeitigen Ausstrahlung von DVB-H und DVB-T.

Aber auch die Hemmnisse liegen auf der Hand: Mit der Einführung neuer Codierverfahren tut man sich hierzulande besonders schwer, und die Kompatibilität nutzt dort nichts, wo die bereits eingeführten Systeme noch mit MPEG-2 arbeiten.

Durch Medienkonvergenz und Triple Play kommt jedenfalls einiges in Bewegung, und auf dem PC ist H.264/MPEG4-AVC bereits Stand der Technik.

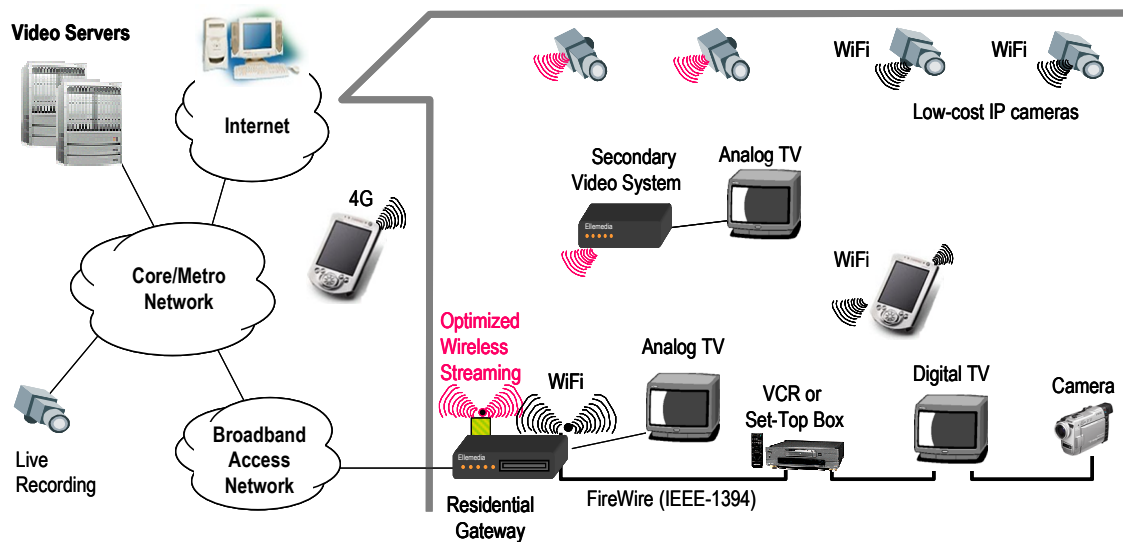


Bild 17 : Architektur des ASTRALS Heimnetzwerks [12]

Zusammenfassung

Die Skalierbare Videocodierung in Erweiterung des H.264/MPEG4-AVC-Standards deckt auf Grund ihrer Flexibilität bei hoher Effizienz viele Anwendungsfälle ab. Dabei wird der nur einmal erforderliche zusätzliche Codieraufwand durch neue Features aufgewogen:

- Bei der Bereitstellung bzw. Archivierung verringert sich der Speicherbedarf.
- Endgeräte mit begrenzten Ressourcen (z.B. Handhelds) decodieren nur den von ihnen benötigten Teil des Datenstroms.
- Die nachträgliche Entfernung von Teilen des Bitstroms (Erosion Storage), z.B. bei der Langzeitaufzeichnung von Überwachungskameras, ist möglich.
- Eine Datenraten-Anpassung während der Übertragung ist mit minimalem Aufwand möglich.

Die hier dargestellten Erweiterungen des H.264/MPEG4-AVC Standards um Skalierbarkeitstechniken wurden hinsichtlich ihrer Codiereffizienz untersucht. Entsprechend den Anforderungen der jeweiligen Anwendung kann ein Kompromiss zwischen der Codiereffizienz und der Anzahl verfügbarer örtlich-zeitlicher Arbeitspunkte gefunden werden.

Danksagung

Teile der hier vorgestellten Arbeiten wurden im Rahmen des IST-Projekts M-Pipe gefördert [13].

Abkürzungen

EP	End Point
HD	High Definition (TV)
MANE	Media Aware Network Element
MANET	Mobile Ad-Hoc Network
MCP	Motion Compensated Prediction
NAL	Network Abstraction Layer
JSVM	Scalable Video Model
JVT	Joint Video Team
PR	Progressive Refinement
PSNR	Peak Signal to Noise Ratio
QP	Quantization Parameter
SD	Standard Definition (TV)
SVC	Scalable Video Coding: Scalable Extension of H.264/MPEG4-AVC
ULP	Uneven Level Protection

Referenzen

- [1] ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audiovisual Services," Version 3, 2005
- [2] H. Schwarz, et al., "Technical Description of the HHI proposal for SVC CE1," ISO/IEC JTC1/WG11, Doc. m11244, Palma de Mallorca, Spanien, Okt. 2004
- [3] J. Reichel, H. Schwarz, M. Wien (Editors), "Scalable Video Coding - Joint Draft," JVT Dokument Q201, Nizza, Frankreich, Okt. 2005
- [4] J. Reichel, H. Schwarz, M. Wien (Editors), "Joint Scalable Video Model JSVM-4," JVT Dokument Q202, Nizza, Frankreich, Okt. 2005
- [5] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B pictures," JVT Dokument JVT-P014, Poznan, Polen, July 2005.
- [6] E. François and J. Vieron, "Extended spatial scalability: a generalization of spatial scalability for non-dyadic configurations," ICIP 2006 (eingereicht)
- [7] H. Schwarz, T. Hinz, D. Marpe, and T. Wiegand, "Constrained interlayer prediction for single-loop decoding in spatial scalability," Proc. of ICIP 2005, Genua, Italien, Sep. 2005.
- [8] T. Wiegand, et. al, "Rate-constrained coder control and comparison of video coding standards," IEEE Trans. CSVT, Vol. 13, pp. 688-703, Juli 2003.
- [9] J. Ridge, X. Wang, Y. Bao, M. Karczewicz, "Low-delay, low-complexity scalable bit-rate video coding," ICIP 2006 (eingereicht)
- [10] M. Winken, H. Schwarz, D. Marpe, T. Wiegand, "Adaptive motion refinement for FGS slices," JVT Dokument JVT-Q031, Nizza, Frankreich, Okt. 2005.
- [11] Thomas Schierl et al: "*SVC-Based multi-Source streaming for robust video transmission in Mobile Ad-Hoc Networks*", IEEE Wireless Communications Magazine (eingereicht)
- [12] Forschungsprojekt ASTRALS, Förderkennzeichen FP6-IST-0028097.
- [13] Forschungsprojekt M-Pipe, Förderkennzeichen FP6-IST-004790.