

Verfahren zur Bewertung des Statistischen Multiplex bei der Video-codierung

Dipl.-Ing. Dagmar Driesnack, Peter Ostertag, Dr.-Ing. Rainer Schäfer, IRT Munich, Germany (driesnack@irt.de, ostertag@irt.de, schaefer@irt.de)

Kurzfassung

Dieser Beitrag beschreibt ein Verfahren zur objektiven Bewertung von Systemen mit statistischem Multiplex, das für den Vergleich verschiedener Systeme und zur Abschätzung des Gewinns an Datenrate eingesetzt werden kann. Für die Auswertung wurde zunächst eine Mittelwertbildung für objektive PSNR-Messungen begutachtet und zusätzlich mit den Verfahren DMOS und JND, als zwei Repräsentanten mit Modellierung subjektiver Eigenschaften, verglichen. In der Konsequenz wurde ein Verfahren entwickelt, das auf einer einfachen PSNR-Berechnung für jedes Einzelbild mit einer anschließenden Auswertung eines PSNR-Histogramms beruht. Dabei wird letztlich die Auftrittshäufigkeit niedriger PSNR-Werte als Maß für die Bildqualität eines Kanals im Statistischen Multiplex genutzt.

1. Statistisches Multiplexing

Für die gleichzeitige Verbreitung mehrerer Videoprogramme über einen einzigen Datenkanal mit meist konstanter Gesamtkapazität wird seit Ende der 90-er Jahre in nahezu allen Verbreitungswegen über DVB-T/T2, DVB-C/C2 und DVB-S/S2 häufig ein „Statistisches Multiplexing“ (im weiteren oft kürzer als StatMux bezeichnet) eingesetzt. Dabei wird die Datenrate für jeden einzelnen Videokanal dynamisch so geregelt, dass Kanäle mit besonders schwer zu kodierendem Videomaterial eine höhere Datenrate erhalten als Kanäle mit einem vergleichsweise einfach zu komprimierenden Inhalt.

Statistisches Multiplexing erlaubt es, die subjektive Qualität der ausgestrahlten Programme gegenüber der Verteilung mit konstanter Datenrate zu verbessern, oder die Anzahl der über den Gesamtkanal übertragenen Programme bei in etwa gleicher Qualität zu erhöhen. Für den in Ansatz gebrachten „Gewinn“ an Datenrate werden durch Hersteller – je nach Anzahl der im Multiplex gebündelten Videokanäle - oft Werte von 20-40% genannt.

2. Fragestellung und Testumgebung

Konkret wurde der Fragestellung „Wie stark verbessert sich die Bildqualität in einem Programm, wenn die Datenkapazität mehrerer Programme in einem gemeinsamen Pool gebündelt wird?“ nachgegangen.

Dazu wurde am IRT in Kooperation mit einem Hersteller (Ericsson) ein Payout-System mit Statistischem Multiplex für die Ausstrahlung von bis zu 7 HDTV-Programmen (720p/50) über H.264 installiert und konfiguriert, sowie eine Reihe (ca. 25) von Testsequenzen gesammelt, die eine möglichst typische Zusammenstellung von Programmmaterial repräsentieren. Alle Testsequenzen mit einer Gesamtlänge von je 10 min. wurden in jeweils gleicher Reihenfolge in einer Schleife zusammengesetzt und diese den einzelnen Videokanälen jeweils um 1:20 min versetzt aus 7 miteinander synchronisierten und unkomprimiert arbeitenden Videoservern (DVS Clipster und DVS Venice) zugeführt. Die Länge der Einzelsequenzen war mit 10 bis 30 s gegenüber diesem Versatz sehr klein, sodass eine annähernd zeitgleiche Ausstrahlung derselben Inhalte auf mehreren Kanälen vermieden wurde.

Der generierte Transportstrom wurde für die Analysen aufgezeichnet. Die Videosignale der einzelnen Kanäle wurden anschließend dekodiert, wiederum auf die Server aufgezeichnet, und standen für eine Auswertung per Software oder über externe Messgeräte mit objektiven Messverfahren [1] zur Verfügung (Bild 1).

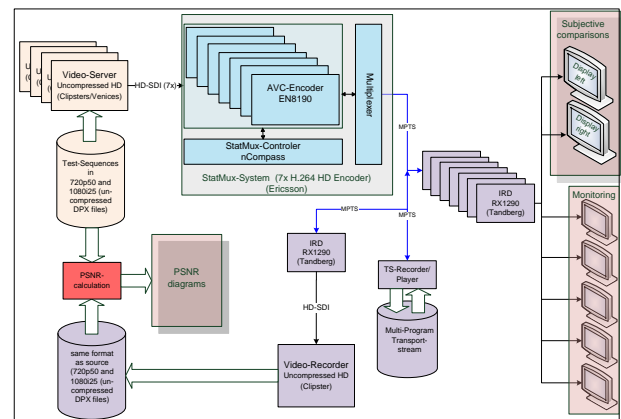


Bild 1 Blockschaubild des Testaufbaus

Jedem Kanal wurde eine für anspruchsvolles HD-Material als kritisch zu bewertende Datenrate von 6 Mbit/s zugeordnet und mehrere Kanäle schrittweise zu einem Datenraten-Pool zusammengefasst. Diese Konfigurationen wurden als Szenarios T0..T6 definiert, wobei T0 einer unabhängigen Kodierung der 7 Kanäle mit jeweils 6 Mbit/s constant bit-rate (CBR) und T6 einer gemeinsamen statistischen Kodierung aller Kanäle in einem Datenraten-Pool von $7 \cdot 6 \text{ Mbit/s} = 42 \text{ Mbit/s}$ entspricht.

Als Referenzen wurden zusätzlich CBR-Kodierungen der Testschleife mit Datenraten zwischen 5 und 12 Mbit/s vorgenommen (vgl. Abschnitt 5). Die Testschleife enthielt zusätzlich eine Stressesequenz mit künstlich generiertem Rauschen variabler Fläche, das den Encoder in jedem Fall überfordert und bei der Auswertung meist nicht herangezogen wurde. Das Vorhandensein dieser Sequenz in „Parallel-“ bzw. „Nachbarkanälen“ erlaubt jedoch die Beurteilung der Empfindlichkeit des Statistischen Multiplexes gegenüber Störeinflüssen durch andere Videokanäle, z.B. der Ausstrahlung stark verrauschten Archivmaterials oder dem Ausfall einer Signalzuführung.

3. Ergebnisse mit bekannten Verfahren

Tabelle 1 zeigt die Auswertung für die bekannten Verfahren PSNR, DMOS und PQR für die Szenarien T0..T6. Zur Berechnung von DMOS und PQR wurde das System PQA500 von Tektronix eingesetzt. Auffallend war, dass die für die Gesamtsequenz vom PQA500 ausgegebenen Werte für DMOS und PQR sehr hohe (schlechte) Werte zeigen. Dies wird auf das Vorhandensein der stark unterschiedlichen Sequenzen in der Testschleife zurückgeführt und mit dem Hersteller diskutiert. Daher wurden die hier angegebenen Messwerte für die Einzelbilder aus den Logfiles des PQA500 genutzt und daraus Mittelwerte errechnet (die getroffenen Aussagen gelten jedoch auch für die Gesamtbewertungen des PQA500).

Szenario	PSNR [dB]	DMOS	PQR [JND]
T0	45,786	12,158	2,695
T1	45,574	11,372	2,716
T2	45,583	10,321	2,643
T3	45,197	9,789	2,619
T4	45,408	9,381	2,599
T5	45,383	9,977	2,645
T6	45,090	9,768	2,639

Tabelle 1 Auswertung mit konventionellen Verfahren

Grundsätzlich ist für alle Verfahren festzustellen, dass die Mittelwertbildung über die gesamte Sequenz keine oder – bezüglich des möglichen Wertebereiches (DMOS 0..100, 1 JND „just noticeable difference“) – kaum signifikante Einflüsse des StatMux zeigt. Dies ist prinzipiell auch plausibel, da eine Verbesserung einiger kritischer Sequenzen auch Verschlechterungen anderer Sequenzen nach sich zieht. Auffallend ist aber, dass neben PSNR auch das rechenaufwändige und dem subjektiven Empfinden besser angepassten Verfahren PQR keine auswertbaren Tendenzen zeigt. Auch das DMOS-Verfahren liefert eine bezüglich des Wertebereiches nur marginale und wenig verlässlich erscheinende Tendenz. Alle Verfahren scheinen daher nur sehr begrenzt für die Bewertung des StatMux geeignet.

4. Verfahren mit Auswertung der PSNR-Histogramme

Dennoch ist aus einer Reihe von Videoqualitätsbewertungen am IRT bekannt [2,3], dass PSNR-Auswertungen durchaus eine sehr gute Korrelation mit subjektiven Ergebnissen besitzen, insbesondere dann, wenn damit ausschließlich Ergebnisse desselben Kompressionsverfahrens und desselben Encoders verglichen werden. Beide Voraussetzungen treffen bei der Analyse des Statistischen Multiplex zu.

In einem weiteren Schritt wurde daher ein Bewertungsverfahren gesucht, das die Intention des StatMux-Verfahrens näher berücksichtigt. Dessen Aufgabe besteht darin, das Auftreten von besonders vielen - subjektiv störenden - Codierartefakten in kritischen Sequenzen mit bereits niedriger Qualität zu reduzieren und dabei durchaus mehr Codierartefakte in unkritischen Sequenzen mit ohnehin besserer Bildqualität zuzulassen. Daher wurde in einem ersten Schritt aus dem zeitlichen Verlauf des PSNR (Bild 2) das Histogramm (Bild 3), d.h. die Auftrittshäufigkeit $h(x)$ bestimmter PSNR-Werte ermittelt. Bereits hier ist die geringere Häufigkeit schlechter PSNR-Werte von T6 gegenüber T0 zu erkennen.

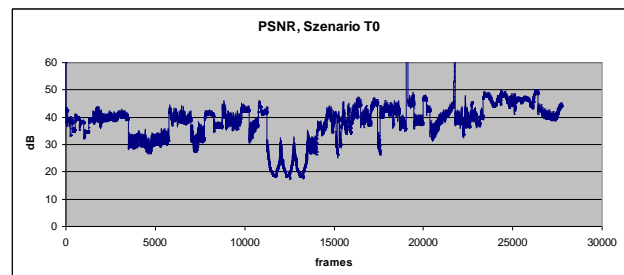


Bild 2 Zeitlicher Verlauf des PSNR für die Testschleife (Szenario T0)

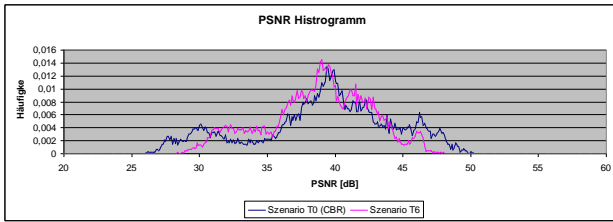


Bild 3 Histogramm des PSNR für die Testschleife (Szenarien T0 und T6)

In einem zweiten Schritt wurde als ein Maß für den Anteil des subjektiv gestörten Materials die Häufigkeit H des Auftretens von PSNR-Werten unterhalb einer bestimmten Grenze $PSNR_k$ bestimmt:

$$H(PSNR_k) = \int_0^{PSNR_k} h(x)$$

Die Auftrittshäufigkeit für verschiedene $PSNR_k$ ist eine Schwellenfunktion (Bild 4) und zeigt deutlich, wie mit einer zunehmenden Anzahl von Kanälen im StatMux die Häufigkeit niedriger PSNR-Werte abnimmt – auf Kosten der Häufigkeit besonders guter PSNR-Werte (denn die Sättigung, dass – nahezu - 100% der Testschleife einen PSNR-Wert von $PSNR_k$ oder schlechter aufweist, ist schon bei niedrigeren Werten gegeben).

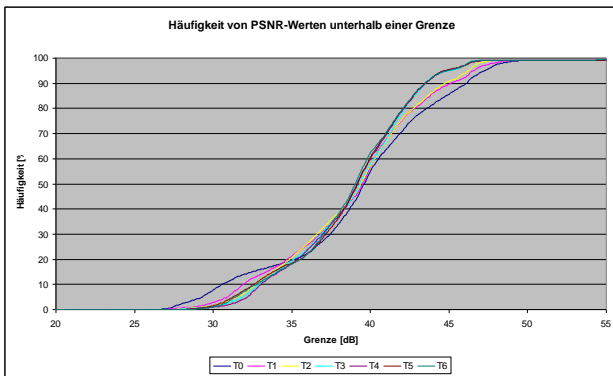


Bild 4 Auftrittshäufigkeit von PSNR-Werten unterhalb einer Grenze (Szenarien T0..T6)

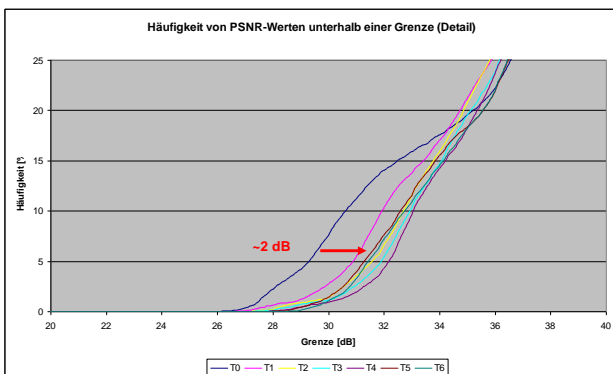


Bild 5 Auftrittshäufigkeit von PSNR-Werten unterhalb einer Grenze (Szenarien T0..T6), vergrößertes Detail

Die Schwellenfunktion „versteilert“ sich also mit zunehmender Anzahl der Kanäle im StatMux. Je nach dem vom Hersteller eingesetzten StatMux-Algorithmus wird sich die Funktion mit steigender Anzahl von Kanälen einer Sprungfunktion annähern, d.h. dass – bestimmte Randbedingungen u.a. im Quellmaterial und bei der Konfiguration vorausgesetzt - alle Kanäle immer und zu jeder Zeit denselben Ziel-Wert aufweisen.

Mit Hilfe dieser hier angewandten Methode kann also die Qualitätsoptimierung durch StatMux in Form einer Verkleinerung der Auftrittshäufigkeit niedriger PSNR-Werte eindeutig nachgewiesen und auch quantitativ erfasst werden. Das Verfahren zeigt ebenfalls, dass der „Gewinn“ in kritischen Sequenzen auch durch einen „Verlust“ bei weniger kritischen Sequenzen bezahlt wird. Für das getestete System ist ebenso festzustellen, dass der Qualitätsgewinn durch StatMux vor allem bei der Bündelung von bis zu vier Kanälen (T3) stark ansteigt und darüber nur noch vergleichsweise gering zunimmt.

5. Interpretation

In der Praxis wird meist die Frage nach dem „Gewinn an Datenrate“ durch einen StatMux relevant. Im Vergleich mit Referenzen aus einer CBR-Codierung kann hier zumindest grob qualitativ eine Aussage getroffen werden: Aus Bild 5 wird ersichtlich, dass durch Einsatz von StatMux der Bereich von „schlechten“ PSNR-Werten im Bereich von 30 dB um rund 2 dB nach rechts hin zu besseren PSNR-Werten verschoben wird. Daraus lässt sich die Frage ableiten, wie weit die Datenrate bei einer CBR-Codierung erhöht werden müsste, um dieselbe Verbesserung zu erzielen. Bild 6 zeigt den PSNR-Verlauf für die CBR-Codierung der Testschleife im Bereich von 5..12 Mbit/s. Daraus ist abzuschätzen, dass eine Verbesserung von 2 dB, ausgehend von der Datenrate von 6 Mbit/s, bei rund 11 Mbit/s erreicht wird. Dies entspricht in etwa einem Gewinn von 45 %. Diese einfache Abschätzung basiert auf PSNR-Mittelwerten über eine gesamte Testschleife und berücksichtigt daher nicht den steileren Verlauf der Kurve hin zu niedrigeren PSNR-Werten.

Daher wurden für eine weitere Abschätzung diejenigen Anteile der CBR-Testsequenzen genutzt, die tatsächlich den PSNR-Wertebereich um 30 dB aufweisen. Hier entsprach eine Veränderung von 2 dB einem Gewinn an Datenrate von ca. 38%. Bild 7 und Bild 8 zeigen den PSNR-Verlauf für 6 und 10 Mbit/s.

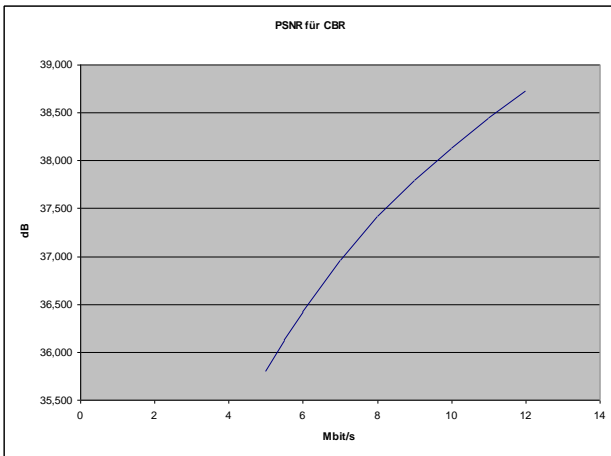


Bild 6 PSNR vs. Datenrate für CBR-Codierung (Mittelwert für Gesamtschleife)

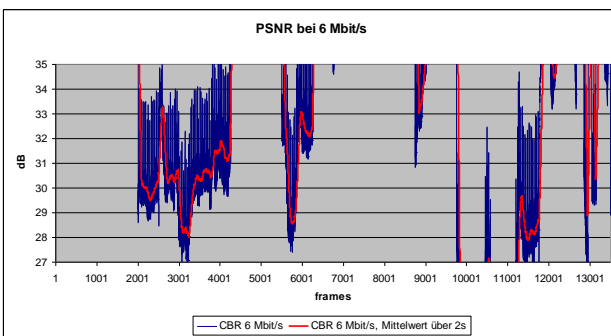


Bild 7 PSNR vs. Zeit für CBR-Codierung bei 6 Mbit/s

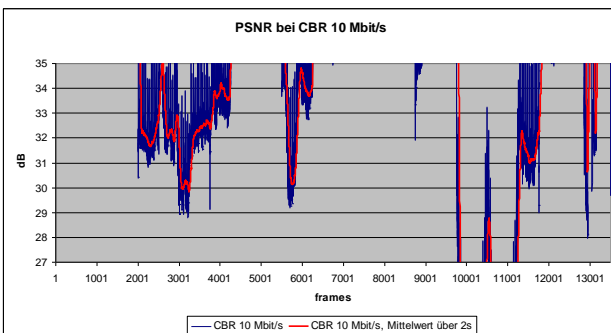


Bild 8 PSNR vs. Zeit für CBR-Codierung bei 10 Mbit/s

Die Abschätzungen geben Auskunft über die Ersparnis an Datenrate gegenüber einer CBR-Ausstrahlung, die zum Vermeiden der Zeiträume mit sehr schlechtem PSNR – und in der Regel auch störenden Kodierartefakten – erforderlich ist. Sie können jedoch noch nicht direkt als Gewinn an Datenrate in Ansatz gebracht werden, da sich bei einer Reduktion der mittleren Datenrate je Kanal ein gegenläufiger Trend in Form einer Verschiebung der Gesamtkurven in Bild 4 hin zu kleineren PSNR-Werten ergibt.

Hersteller nennen im allgemeinen Gewinne zwischen 20% und 40%, allerdings ohne allgemein zugängliche Hinweise auf die eingesetzten Beurteilungsverfahren oder genaue Kenntnis des Testmaterials.

6. Erfahrungen und Ausblick

Überprüfungen haben tendenziell gleiche Ergebnisse auch bei StatMux-Systemen unterschiedlicher Hersteller bestätigt. Als Bezug für die Bewertung eines StatMux wird dabei immer die CBR-Codierung desselben Encoder-Typs herangezogen. Daher lassen sich gut Aussagen im Vergleich zwischen einer CBR-Codierung und einem StatMux-System desselben Encoder-Typs treffen. Ein direkter Vergleich zwischen unterschiedlichen Encodern scheint aufgrund der dafür festgestellten Unterschiede zwischen subjektiver Wahrnehmung und absoluten PSNR-Werten [3] ohne weitere Maßnahmen noch nicht möglich.

Das vorgestellte Verfahren wurde mit PSNR-Messungen realisiert. Es ist ebenso anwendbar auf der Basis von DMOS- und PQR-Messungen und soll in einem weiteren Schritt mit diesen Skalen überprüft werden.

7. Danksagung

Diese systematischen Tests wurden am Rande einer mehr betrieblich orientierten Untersuchung zur Nutzung des Statistischen Multiplex für die HD-Programme von ARD und ZDF durchgeführt und im Nachgang detaillierter ausgewertet. Der Fa. Ericsson wird an dieser Stelle für die leihweise Bereitstellung eines StatMux-Systems mit ausreichender Kanalanzahl, sowie für die begleitende Unterstützung bei Inbetriebnahme und Konfiguration gedankt.

8. References

- [1] Tektronix: „Picture Quality Analyzers“, <http://www.tek.com/picture-quality-analyzer>
- [2] Schäfer, Rainer: „HD Ausstrahlungsformat“, FKT 10/2008, S.568 ff
- [3] Driesnack, Dagmar: „Untersuchung der Bildqualität von HD- und SD-Signalformaten auf modernen Flachdisplays in Abhängigkeit von der Datenrate und dem Codierverfahren“ Diplomarbeit an der FH Mittweida und dem IRT, Mittweida, Juni 2006