

Capture-Karten und aspect-ratio für Dummies ;-)

a) NT... systems



V 1.00
29.08.02
Der Karl

Capture-Karten und Aspect-Ratio?

In einer Diskussion im DVD-SVCD-Forum wurde klar, daß es zu diesem Thema noch einige Unklarheiten und Verwirrung gibt. Deshalb habe ich mich entschlossen, es mal etwas genauer – und hoffentlich verständlich - zu beleuchten.

Um es etwas übersichtlicher zu halten, beschränke ich mich auf die im europäischen Raum verwendete PAL-Norm (PAL B/G, H, I).

Sämtliche Definitionen, wie das PAL-Signal auszusehen hat, sind in der Rec. ITU-R BT.470-3 niedergelegt.

Da ist nirgendwo von Auflösungen die Rede, sondern nur von Frequenzen, Pegeln, Blanking..... In der Rec. ITU-R BT.601-4 (früher "CCIR-601" oder "Rec.601") ist definiert, wie diese Signale zu digitalisieren sind.

In der „EBU Technical Recommendation R92-1999“ sind die ITU-Empfehlungen „übersetzt“, so daß auch ein normal Sterblicher sie verstehen kann (Anhang).

Was ist eigentlich full PAL?

DIE Auflösung gibt es eigentlich nicht. Da die TV-Übertragung ja analog passiert, gibt es keine Pixel. Pixel benötigt man für die digitale Speicherung. Dafür braucht man eine „Handlungsanweisung“, wie die analogen Signale in computertaugliche Pixel zu übersetzen sind. Diese „Handlungsanweisung“ steht (als Empfehlung) in der Rec. ITU-R BT.601-4.

Gängige Capture-Auflösungen für PAL sind 768*576, 720*576 und 704*576.

Was davon ist nun richtig?
Die ITU sagt eindeutig **720*576** !

Die ITU sagt auch eindeutig, **wie** diese 720*576 zu digitalisieren sind. Leider hält sich kaum einer dran. Im Ergebnis arbeiten fast alle mir bekannten Karten in der Auflösung **720*576 falsch** !

Für das Verständnis, was da falsch gemacht wird, benötigt man etwas Grundlagenwissen.

Das Fernsehbild

Ein PAL-Bild besteht aus 625 Zeilen. Gesendet wird mit 50 Halbbildern bzw. 25 Vollbildern pro Sekunde. Da wir ja Vollbilder digitalisieren wollen, lassen wir die Halbbilder (fields) und das Interlace-Thema einmal beiseite.

- Ein Vollbild PAL = 625 Zeilen
- 25 Bilder / Sekunde

Wir haben also 25 Vollbilder pro Sekunde. Eine Sekunde geteilt durch 25 sind 0.04 Sekunden bzw. 40 Millisekunden pro Vollbild.

- Ein Vollbild PAL = 40ms

Es werden in 40ms 625 PAL-Zeilen übertragen. 40ms geteilt durch 625 Zeilen = 0.064 Millisekunden bzw. 64 Mikrosekunden pro Zeile

- Eine Zeile PAL = 64µs

Es vergehen also 64µs, während der Fernseher eine Zeile „schreibt“. Es ist immer noch alles analog. Keine Pixel!

Was sagt die ITU-R BT.601-4 dazu?

Nach ITU-R BT.601-4 wird die Videoinformation (Luminanz) mit einem Takt von 13.5 MHz digitalisiert. Luminanz ist die Helligkeitsinformation, im Unterschied zu Chrominanz, die uns im Moment nicht interessiert.

Die Festlegung des Taktes auf 13.5 MHz ist ein Kompromiß zwischen PAL/SECAM und NTSC.

- *Abtastfrequenz: 13.5 MHz*

Was heißt das nun genau?

Die Videoinformationen werden 13.5 Millionen mal pro Sekunde abgetastet.

Es werden also jede Sekunde aus 13.5 Millionen analogen „Momentanwerten“ digitale „Abbilder“ erstellt.

Wenn wir nun unsere PAL-Zeile von 64µs „Länge“ mit 13.5 MHz abtasten, bekommen wir

$13500000 \text{ Hz} * 0.000064 \text{ s} = 864 \text{ Abtastwerte (samples) pro Zeile}$

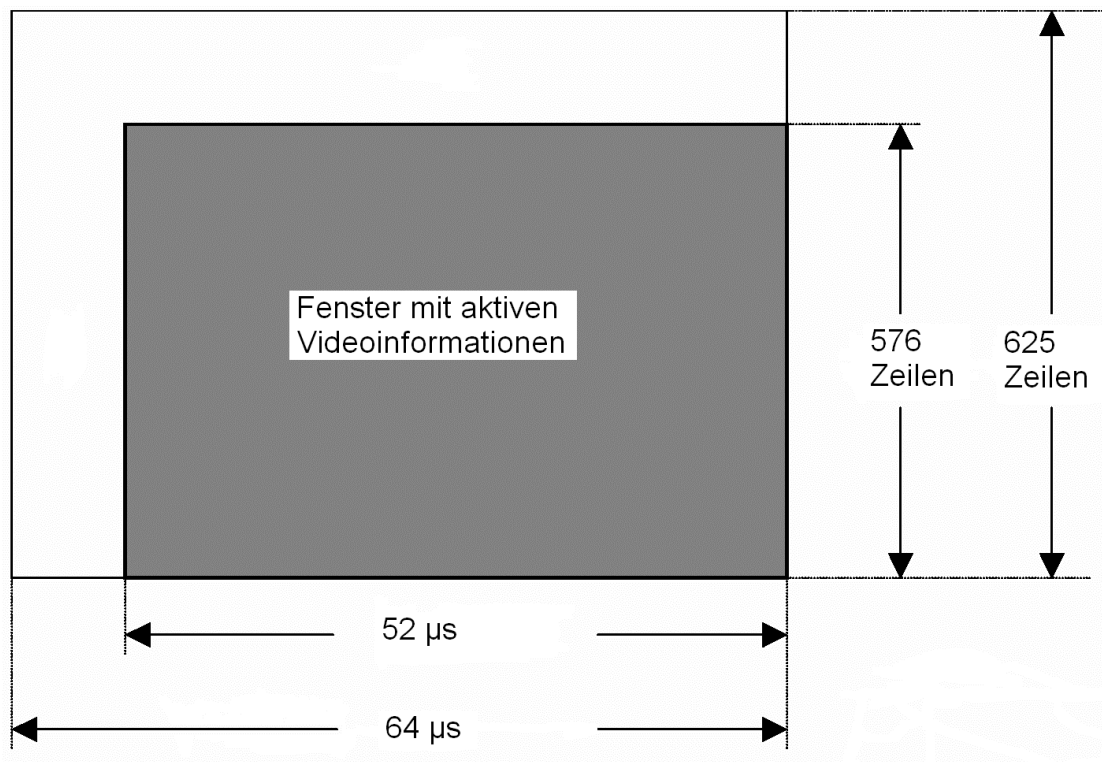
- *864 samples / Zeile*

Das aktive Videofenster.

Bisher sprachen wir über PAL mit 625 Zeilen und 64µs / Zeile.

Das ist allerdings ein Brutto-Wert, der noch eine Menge „Verwaltungsinformationen“ enthält.

Der tatsächliche Bereich, der „aktive“, also für das übertragene Video relevante Bildinformationen enthält, ist kleiner.



Zurück zur ITU

Wir haben jetzt also 576 aktive Zeilen a 52 μ s Länge, in denen die Bildinformationen enthalten sind, die wir für unser digitales Endprodukt benötigen.

- 576 Zeilen a 52 μ s

Die ITU sagt, von den 864 horizontalen samples sind **720** (line-)aktive, die in das Endprodukt eingehen und die restlichen 144 sind „Digital blanking“.

Das ist offensichtlich ein Widerspruch!

Wir haben eine Zeile von 64 μ s „Länge“, die nach dem (ITU-konformen) sampling aus 864 digitalen Werten besteht.

Wir rechnen wieder ein wenig:

$$64\mu\text{s} / 864 = 74.074\text{ns}$$

In einer Zeile von 64 μ s gibt es also 864 samples mit je 74.074 ns „Länge“.

- Periodendauer für ein Pixel = 74.074 ns

Diesen Wert multiplizieren wir mit den 720 Pixeln, die ITU empfiehlt.

$$74.074\text{ns} * 720 = 53.3333\mu\text{s}$$

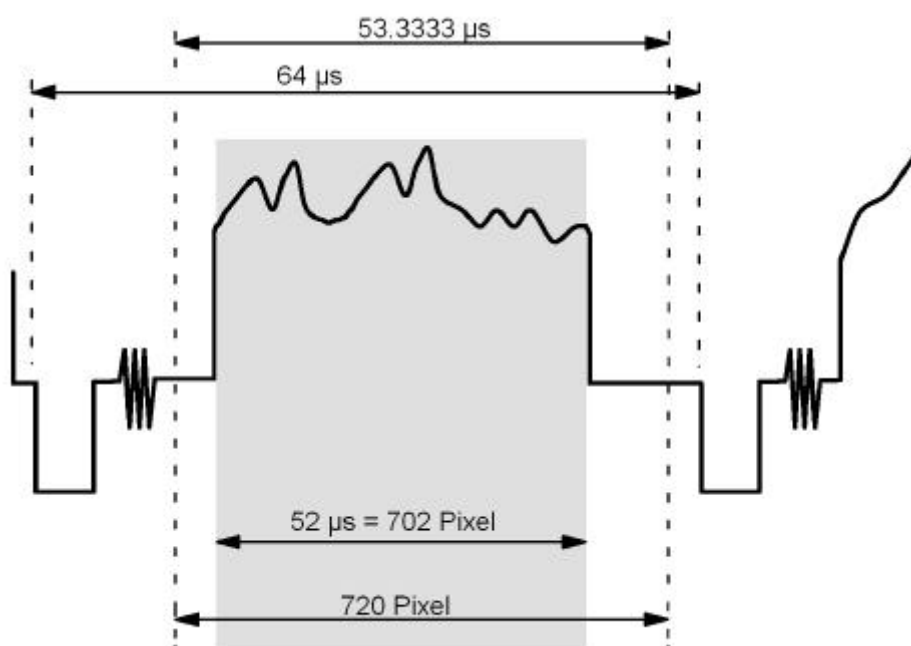
Das von ITU vorgeschlagene „Fenster“ ist also größer, als die 52 μ s vorhandenen Videoinformationen.

- 720 Pixel = 53.3333 μ s

Dividieren wir jetzt die „echten“ aktiven 52 μ s durch die errechneten Periodendauer / Pixel:

$$52\mu\text{s} / 74.074\text{ns} = 702.0007 = \sim 702 \text{ Pixel}$$

Es gibt also nur 702 aktive Pixel !



Das digitale Ergebnis nach ITU beinhaltet also 702*576 Pixel motion-area (der Bereich, wo etwas passiert) und rechts und links je 9 (in Summe 18) Pixel overscan, also schwarze Ränder. Diese 18 Pixel stimmen natürlich nie genau. Die Quelle ist immerhin analog und auch die Sendeanstalten arbeiten nicht sooo ungeheuer genau.

Spätestens jetzt sollte klar sein, wozu die 704*576er Auflösung gut ist. Sie entspricht präzise den 702*576 plus 2 mal 1 Pixel overscan, beinhaltet also die gleiche motion-area wie die 720*576 (wenn diese denn richtig nach ITU digitalisiert wurde).

Wenn eine Capture-Karte bzw. deren Treiber bei **720*576** in der Horizontalen den ganzen Bereich mit aktiven Videoinformationen füllt bzw. nur wenige Pixel Rand läßt, so verhält sie sich **falsch** bzw. **nicht ITU-konform**.

Was ist nun aspect-ratio?

Bisher war alles nur Vorspiel. Das soll ja angeblich der schönere Teil sein ;-).

Wir lassen PAL-plus und „echtes“ 16:9 mal beiseite, da es auch meines Wissens keine Capture-Karte gibt, die das unterstützt.

Es bleibt also das 4:3-Format.

Erst mal zur DAR (display aspect ratio). Eine DAR von 4:3 bedeutet, wenn man das Bild auf eine Leinwand von 4m Breite projizieren würde, die Höhe 3m wäre. Also 4m breit/3m hoch, oder 2m breit und 1.5m hoch, oder, oder..... kurz: ein Faktor von 1.3333333... (4/3)

Zurück zum Fernsehen. Wir haben als einzigen Fixwert die vertikalen 576 Zeilen. Diese 576 Zeilen multipliziert mit dem DAR-Faktor 1.33333.. ergibt 768 horizontale Punkte.

Damit haben wir schon mal unsere „Industrienorm“ 768*576.

Das scheint schon etwas seltsam, denn die 768*576 enthalten exakt die gleichen Bildinformationen wie die 702*576 nach ITU.

Damit kommen wir zu einem weiteren Faktor: Der PAR (pixel aspect ratio).

Die PAR muß man dann berücksichtigen, wenn man zwischen den Formaten für verschiedene Geräte umrechnet.

Auf dem PC-Monitor gibt es nur eine PAR, nämlich 1:1. Die Pixel sind also quadratisch (square). Ein Pixel auf dem Monitor, wäre es 1mm hoch, wäre dann auch 1mm breit. Das erklärt sich aus den verwendeten Auflösungen (640*480, 800*600, 1024*768..) – immer DAR 4:3 und PAR 1:1. Ausnahme ist z.B. 1280*1024 und schon sind Kreise nicht mehr rund.

Geräte, die auf TV ausgeben (z.B. DVD-Player) arbeiten mit einer anderen PAR, nämlich 54/59 bzw. einem Faktor von ~1:1.0926. Diese „Pixel“ sind also „nonsquare“ (nicht quadratisch) und um den Faktor 1.0926 breiter als hoch.

Man könnte also die PAR als verzerrungs-/ entzerrungs- Faktor bezeichnen.

Es gibt noch mehr / andere verschiedene PARs. Kennen muß man eigentlich nur noch die generic-PAR (45/48 bzw. 1:1.066666).

Man kommt bei PAL also mit drei verschiedenen PARs (1:1, 54/59 und 45/48) aus.

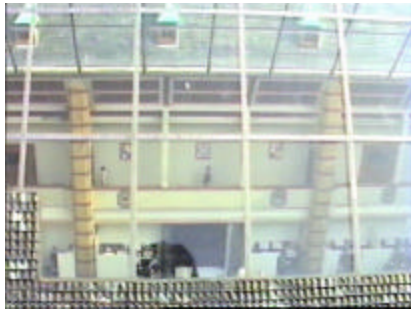
Wir belassen es mal dabei, damit es nicht zu kompliziert wird. Man kann damit schon auf 1 Pixel genau rechnen. Das reicht völlig ;-)

Schlußendlich: 702 (ITU) vs. 704 (MPEG2). Das sind natürlich alles Näherungswerte, die einen Kompromiß zwischen verschiedenen Videonormen und mathematischen Notwendigkeiten (Teilbarkeit) darstellen. Folgende Fixwerte haben wir: 576 Zeilen, 4:3 Bildformat und PAR 54/59. Noch einmal unsere „Industrienorm“: $576 * (4/3) = 768$

Diese 768 mit 54/59 „verzerrt“, also $768 \cdot (54/59) = 702.915$ ergibt rund 703 horizontale Pixel.
 Rechnerisch wäre die „richtige“ Auflösung bei PAR 54/59: also $703 \cdot 576$. Damit liegt sowohl die ITU,
 als auch die MPEG-Norm jeweils einen Pixel „falsch“.
 Diese Abweichungen kann man absolut vernachlässigen.

Overscan

Ich hatte es schon mal kurz erwähnt. Da gibt es noch was neben DAR und PAR, was man wissen muß. Je nach gewählter Capture-Auflösung hat man einen / keinen overscan-Bereich.
 Dieser overscan-Bereich muß bei der Umrechnung der Auflösungen natürlich berücksichtigt werden.
 Was zählt, ist die tatsächliche motion-area ohne overscan-Bereich, also ohne die schwarzen Balken.
 Alle Bilder enthalten **exakt** die gleichen Informationen!



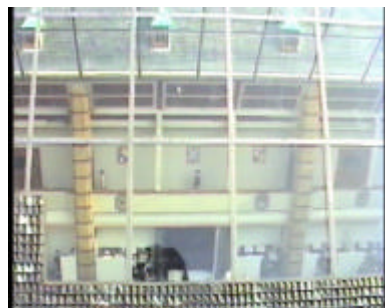
768*576
 kein overscan
 PAR 1:1



720*576
 kein overscan
nicht ITU-konform !
 Generic PAR 45/48



480*576
 SVCD 2/3 D1
 PAR 54/59 und
 auf 2/3 gestaucht
 + ca. 2*5 bis 2*6
 Pixel overscan



720*576
 D1 und DV
 2*8 bis 2*9
 Pixel
 overscan
 PAR 54/59



352*576
 half D1
 kein overscan
 PAR 54/59 und
 auf die Hälfte
 gestaucht



704*576
 kein, bzw.
 max 2*1 Pixel
 overscan
 PAR 54/59

Auflösungen umrechnen, aber richtig!

Worauf kommt es beim Umrechnen der Auflösungen an?

Erst einmal auf die Quell- und Zielauflösung. Davon die tatsächliche Motion-Area bzw. coded Pixel. Dabei muß der overscan berücksichtigt werden. Schließlich die Quell- und Ziel-PAR.

Hat das Quell- bzw. Zielvideo keine square pixel (PAR 1:1), wird als Zwischenschritt immer mit PAR 1:1 gerechnet.

-Quellvideo: Tatsächliche motion-area (weglassen der schwarzen Ränder) ermitteln, „entzerren“ mit source-PAR, um auf PAR 1:1 (Zwischenschritt) zu kommen, „echte“ DAR errechnen.

-Zielvideo: Gewünschte Zielauflösung festlegen, mit Ziel-PAR „entzerren“, unter Berücksichtigung der Faktoren Overscan, Teilbarkeit, Block-/Macroblock-Optimierung ähnliche DAR (wie Quellvideo) erreichen und das Quellvideo durch cropping an die Ziel-DAR weiter anpassen.

Rechenbeispiele

Beispiel 1:

Fangen wir mal mit dem einfachsten Fall an. Wir haben ein VHS-Video mit TV-Karte (BT8x8) in der Auflösung von 768*576 digitalisiert. Zielauflösung soll 640*480 für MPEG4 bzw. Divx sein. Beide Auflösungen haben square-Pixel und eine DAR von 4:3 bzw. 1.3333. Theoretisch könnte man das 1:1 umrechnen, wenn die schmutzigen Ränder der Aufnahme nicht wären. Die Aufnahme muß also erst einmal gecroppt werden.

Angenommen, wir müßten um die schwarzen Balken und „ausgefranst“ Ränder loszuwerden, oben 8, unten 10, links und rechts je 4 Pixel abschneiden. Als tatsächliche motion-area bzw. coded pixel ergeben sich also 760 (768-(2*4)) horizontal und 558 (576-(8+10)) vertikal. Also ist unsere source-DAR $760/558=1.362$.

Wir wollen aber $640/480=1.3333$ erreichen. Da die DAR nach oben abweicht, muß die Horizontale angepaßt werden.

Wir nehmen also die Vertikale als Bezug: $558*1.333333=744$. Das Video muß also um weitere 16 Pixel (760-744) horizontal gecroppt werden. Also *zusätzlich* z.B. links und rechts je 8.

Wir croppen unser Video also jetzt: oben 8, unten 10, links und rechts je 12 Pixel.

Ergibt $744*558$ coded pixel und eine DAR von $744/558=1.333333$

Das können wir jetzt nach 640*480 „resizen“.

Beispiel 2:

Quelle ist, (wie schon bei Beispiel 1) ein BT-capture von VHS. Ziel ist diesmal eine SVCD.

Die coded pixel von 760*558, also auch die DAR von 1.362 übernehmen wir einfach mal aus Beispiel 1.

Erreichen wollen wir SVCD 2/3 D1 (480*576). Als Vorgabe sollen rundrum 2 Blocks (16 Pixel) TV-Overscan entstehen. Horizontal $480-(2*16)=448$, Vertikal $576-(2*16)=544$.

Unsere Zielauflösung (coded Pixel) ist also 448*544.

Das wird jetzt auf PAR=1:1 entzerrt (Zwischenschritt).

Wir erinnern uns: SVCD ist auf 2/3 „gestaucht“. Also $448*3/2=672$.

Zusätzlich muß neben der Stauchung noch die PAR berücksichtigt werden, also noch $672*1.0926\approx 734$

Die "echte" square-Zielauflösung mit PAR=1:1 wäre also 734*544. Jetzt wird die tatsächliche DAR ausgerechnet. $734/544=1.34926$.

Unsere "echte" DAR für das Zielformat ist also 1.34926.

Die DAR unseres Ausgangsformates ist (siehe oben): $760/558=1.362$.

Wir müssen also unser Ausgangsformat durch cropping weiter anpassen.
Da die DAR wiederum noch oben abweicht, wird ebenfalls die Horizontale angepaßt.
 $558 \cdot 1.34926 = 752.887$, also (wg. Teilbarkeit durch 2) ~ 752 .

Unser Ausgangsmaterial muß also auf $752 \cdot 558$ gecroppt werden, dann nach $448 \cdot 544$ "resized" werden und auf $480 \cdot 576$ "letterboxed" werden.

Beispiel 3:

Quelle ist diesmal ein TV-Capture eines 16:9-Kinofilms mit generic-PAR ($720 \cdot 576$). Die generic-PAR erkennen wir daran, daß die gesamten 720 Pixel horizontal Bildinformationen enthalten.
Ziel ist wiederum eine SVCD.

Um die schwarzen Balken zu entfernen, müssen oben und unten je 58 Zeilen gecroppt werden.
Wir haben jetzt $720 \cdot 460$ coded Pixel.

Weil ich ein Sauberkeitsfetischist bin und keine schmutzigen Ränder mag, werden links und rechts auch 2 Pixel gecroppt, also $716 \cdot 460$ coded Pixel.

Nun horizontal entzerren (mit generic-PAR) $716 \cdot 1.06667 = \sim 764$.

Unsere "echte" DAR ist also $764/460 = 1.6609$.

Ich gebe mal wieder je 2 Blocks TV overscan vor. Diesmal nur horizontal - ist ja logisch ;-)

Zielformat 448 ($480 - (2 \cdot 16)$) auf PAR 1:1 entzerrt - wie in Beispiel 2 = 734.

Möglichst ähnliche DAR erreichen: also $734/1.6609 = \sim 442$ also Block- / Macroblock-optimiert (Teilbarkeit durch 32) 448.
Die tatsächliche Ziel-DAR ist also $734/448 = 1.6384$.

Nehmen wir (wieder) die Vertikale als Bezug: $460 \cdot 1.6384 = 753.664$ (mit Generic-PAR verzerren)
 $753.664/1.06666 = 706.56$ bzw. wg. Teilbarkeit durch 2 ~ 706 .

Man muß also die source auf $706 \cdot 460$ croppen, nach $448 \cdot 448$ resizen und auf $480 \cdot 576$ "letterboxen".

Beispiel 4:

Quelle ist wieder (wie 3.) ein TV-Capture eines 16:9-Kinofilms mit generic-PAR ($720 \cdot 576$).
Ziel ist Half-D1 ($352 \cdot 576$).
Cropping und DAR wie in 3.

Für Half-D1 dürfen wir (wegen starker Stauchung) nur mit einem Block TV-overscan arbeiten.

Zielformat 336 ($352 - (2 \cdot 8)$) auf PAR 1:1 entzerrt ($336 \cdot 2 \cdot 1.0926 = 734$).
Wir haben also die gleiche Ziel-DAR wie in 3.

Wir croppen die source wiederum auf $706 \cdot 460$.
Hier allerdings auf $336 \cdot 448$ resizen und auf $352 \cdot 576$ "letterboxen".

Und nochmal zum Overscan ;-)

Wir müssen unterscheiden zwischen dem overscan im gesendeten/encodeten Material (siehe Topic Overscan) und dem TV-overscan, den wir beim encoden selbst festlegen.
Hier geht es darum, Bereiche schwarz zu lassen, die der Fernseher ohnehin nicht darstellt. Das spart Bitrate, die dem wirklich sichtbaren Bild zugute kommt – dient also besserer Qualität.

Für Beispiel 2 und 3 gilt: Da wir selbst "letterboxen", müssen wir auf den SVCD-overscan keine Rücksicht nehmen. Es gelten nur die "coded pixel" und die richtige PAR.

Wem diese ganze Rechnerei zu aufwändig (und fehlerträchtig) ist, dem sei shhs hervorragendes FitCD empfohlen: <http://shh.dvdboard.de/>

Das Programm ist zwar Freeware, der Autor freut sich aber sicher über jeden € als „Dankeschön“ für die viele investierte Zeit.

Arbeitsweise der Capture-Chips:

BT848/878 (fast alle TV-Karten):

PAL wird mit 17.735 MHz digitalisiert. Die benötigten Horizontalaufösungen werden durch "scaling" erzeugt.

Das ergibt für PAL im 64 µs-Fenster 1135.04 horizontale Pixel. Die werden durch den "scaler" und "Ultralock" geschickt. Daraus entstehen 944 Pixel für das square-Format bzw. 864 für das "CCIR-kompatible" (720 horizontal mit korrekter 54/59 PAR). Darüber wird ein "Fenster" gelegt, was 768 respektive 720 Pixel breit ist.

Das resultierende Fenster ist also 52.0678 µs (für 768) bzw. 53.33 µs (für 720).

Die BT-Chips digitalisieren etwas zuviel (768) bzw. präzise richtig (720).

Soviel zur Theorie. Die Vorgaben (Empfehlungen) von Brooktree/Conextant sind also o.K. In der Praxis sieht das völlig anders aus. Der digitalisierte Bereich, den der Capture-Treiber übergibt, ist immer deutlich kleiner (schätzungsweise 200-300 ns weniger). Außerdem wird bei 720 völlig falsch skaliert.

Das läßt nur einen Schluß zu: Die Treiber machen MIST!

Philips SAA7108/7113 (Asus/Nvidia-Karten):

Es wird mit 13.5 MHz (ITU-konform 720 Pixel im 53.33 µs-Fenster) digitalisiert. Der Chip (oder Treiber) übergibt davon einen Ausschnitt von 704 Pixeln horizontal. Neuere WDM-Treiber übergeben auch 720.

Das Ergebnis ist o.K. (nachgemessen an meiner Asus V3800 ultra deluxe).

Es ist davon auszugehen, daß die PAR bei den Philips-Chips immer stimmt, da der Chip keinen scaler hat, unsauber programmierte Treiber also nichts verbiegen können.

Micronas/Intermetall VPX3224/25 (z.B. Elsa Erazor III ViVo, Guillemot Cougar VE):

Obwohl das Ding wohl kaum noch irgendwo funktioniert, trotzdem Interessehalber.

Es wird mit 13.5 Mhz digitalisiert. Die 720 bzw. 704-Auflösung sollte also korrekt sein.

Die square-Auflösungen werden über "scaling" erzeugt. Hängt also am Treiber.

Ich kann das leider nicht (mehr) prüfen.

ATI mit dem "integrierten" Video-Chip:

Dazu findet man leider überhaupt keine Infos.

Die 720er Auflösung ist definitiv nicht ITU-konform. Es muß also mit generic-PAR gerechnet werden.

Die 704er sollte ungefähr stimmen. Das kann aber vom verwendeten Treiber abhängen, da hier scheinbar auch ein scaler verwendet wird.

EBU Technical Recommendation R92-1999

Active picture area and picture centring in analogue and digital 625/50 television systems

<i>EBU Committee</i>	<i>First Issued</i>	<i>Revised</i>	<i>Re-issued</i>
PMC	1998	1999	

The EBU is aware of a certain amount of confusion about the active picture area in the implementation and use of digital signals conforming to the ITU-R Recommendation BT.601 [1].

Recommendation BT.601 specifies a line length of 720 luminance pixels (13.5 MHz sampling). ITU-R Recommendation BT.470[2], specifies a line length of 52 μ s for 625 line analogue signals. This corresponds to 702 luminance pixels.

This apparent difference can lead to difficulties in conversion, especially if users wish to maintain the correct aspect ratio of the pictures.

Recommendation BT.601 accommodates modest variations in the position and length of analogue blanking which arise before a signal is digitised or when digital signals pass through any subsequent analogue process.

The EBU *recommends* that:

- In 625-line television systems sampled to ITU-R Rec. BT.601 part A, only the central 702 luminance samples of the digital active line (samples 9-710 inclusive) and their associated chrominance samples should be used to carry the active picture. The remaining 18 luminance samples and their associated chrominance samples may be used to carry picture information only but for no other purpose. It cannot be guaranteed that picture information in these samples will be displayed in either 4:3 or 16:9 aspect ratio images.
- The centre of the picture should retain its position throughout all production processes unless there are creative reasons to deliberately do otherwise.

Notes:

- *The centre of the image is located between pixels 359 and 360 in the horizontal direction.*
- *The centre of the image is located and midway between line 479 (field II) and line 167 (field I) in a 625-line interlaced raster in the vertical direction. (The vertical centre would be located between lines 332 and 333 in a notional progressively scanned raster.)*[3]

Bibliography

- [1] ITU-R Recommendation BT.601-5: **Studio encoding parameters of digital television for standard 4:3 and widescreen 16:9 aspect ratios**
- [2] ITU-R Recommendation BT.470-3: **Conventional television systems**
- [3] ITU-R Recommendation BT.1358: **Studio parameters of 625 and 525 line progressive scan television systems**