Capture-Cards and aspect-ratio for Dummies ;-)

V 1.00 29.08.02 by Der Karl; English translation by Arachnotron 25 Nov 2003

Foreword to the English version:

This guide, “Capture-Karten und aspect-ratio für Dummies” was originally written by Der Karl in German. I have tried to do an exact translation without making any changes to the contents. The German original can be found here:

http://www.videoxone.de/cgi-bin/load.pl?page=derkarl

Capture-cards and Aspect Ratio?

A discussion in a DVD-SVCD-Forum made clear that a lot is still unclear and confused about this subject. I therefore decided to clarify things and to hopefully make them more understandable. To keep things simple, I limit myself to the in Europe more common PAL-Norm (PAL B/G, H, I).

The complete definitions what a PAL signal should look like are published in ITU-R BT.470-3. Nowhere does it deal with resolutions, only with frequencies, levels, blanking.… In the publication ITU-R BT.601-4 (formerly "CCIR-601" or "Rec.601") is defined how to digitize these signals.

In the "EBU Technical Recommendation R92-1999" the ITU recommendations have been “translated” so a mere mortal can understand them. (Appendix).

What is full PAL?

THE resolution doesn’t exist. Since TV transmissions are analogue, there are no pixels. Pixels are needed for digital storage. To do this a procedure is needed how to convert the analogue signals in pixels a computer can understand. This procedure can be found in the publication ITU-R BT.601-4.

Common capture resolutions for PAL are 768*576, 720*576 and 704*576.

How much of those are correct?
ITU clearly states **720*576**!

It also clearly states, **how** these 720*576 have to be digitized. Unfortunately, nobody adheres to this. Actually, almost all cards I know of operate **incorrectly** at **720*576** resolution!

To understand what is being done incorrectly one has to know some backgrounds.

The TV picture

A PAL picture contains 625 lines. 50 half pictures or 25 full pictures are transmitted per second. Since we want to digitize full pictures we skip the half pictures (fields)- and interlace subject for the moment.

- *a full PAL picture = 625 Lines*
- *25 pictures / Second*

We have 25 full pictures per second. A second divided by 25 is 0.04 seconds or 40 milliseconds per full picture.

- a *full PAL picture = 40ms*

In 40 ms, 625 PAL-lines are transmitted. 40ms divided by 625 lines = 0.064 milliseconds or 64 microseconds per line.

- *A PAL line = 64µs*

so it takes 64µs for a TV to “write” one line.
Everything is still analogue at this point. No pixels!

What hat ITU-R BT.601-4 to say about this?

According to ITU-R BT.601-4 the video information (Luminance) is digitized at a frequency of 13.5 MHz. Luminance is the brightness information, as opposed to chrominance in which we are not interested at the moment.
The frequency is set to 13.5 as a compromise between PAL/SECAM and NTSC.

- *Sampling rate: 13.5 MHz*

what does this mean exactly?

The video information is sampled 13.5 million times per second. So every second, from 13.5 million analogue “moments in time” a digital representation is made.

If we sample our 64µs "long" PAL line at 13.5 MHz, we get

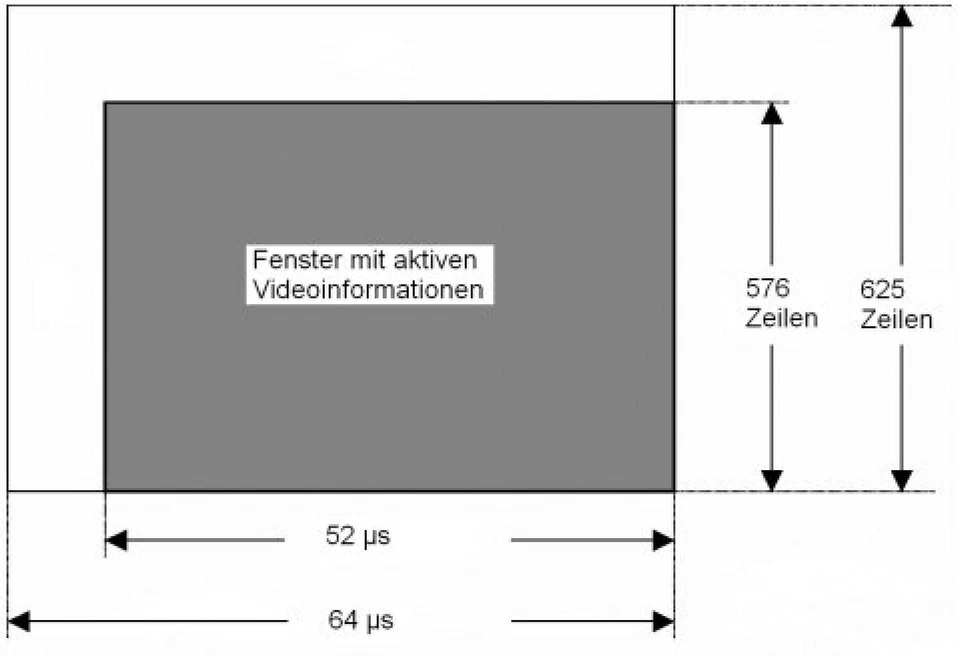
13500000 Hz * 0.000064 s = 864 samples per line

- *- 864 samples /line*

The active part of the video picture.

Up until now we talked about PAL with 625 lines and 64µs / line.
That is just a gross value, and contains a lot of " Administrative overhead "

The actual range that contains “active”, also for the transmitted video relevant, picture information is smaller.



Back to ITU

We have now have the 576 active lines at 52 µs per line length, in which the picture information is contained, which we need for our digital final product.
● 576 *lines a 52µs*

The International Telecommunication Union says that of the 864 horizontal samples, 720 are (line-) active. Those go into the final product, remaining 144 are "digital blanking".
That is obviously a contradiction!
We have a line of 64 µs "length", which after (ITU-conform) sampling consists of 864 digital values.
So again we calculate:

64µs / 864 = 74.074ns

So, in a line of 64 µs there are 864 samples with a “length” of 74.074 ns each.
● *Period duration for a single pixel = 74.074 ns*

we multiply this value by the 720 pixels, which ITU recommends.

74.074ns * 720 = 53.3333µs

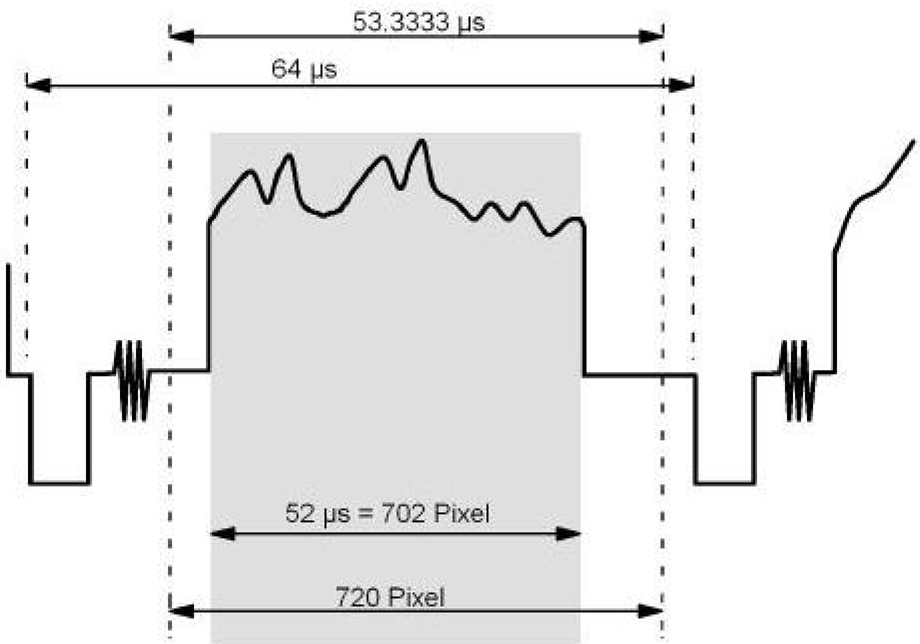
The "window" suggested by the ITU is thus larger, than the 52 µs existing video information.

- *- 720 pixel = 53.3333 µs*

We now divide the "genuine" active 52 µs by the calculated period duration/pixel:

52µs / 74.074ns = 702.0007 = ~702 pixel

There are only 702 active pixels!



The digital result after ITU contains a 702*576 pixel motion-area (the range, where something happens) and both on the left and on the right 9 (in total 18) pixels each overscan, thus black edges. These 18 pixels of course not completely exact. The source is analogue after all, and also the broadcasting stations do not works all that precise.
By now it should be clear what this 704*576 resolution is good for. It corresponds precisely to 702*576 plus 2 times 1 pixel overscan, and therefore contains the same motion area as the 720*576 (when digitized correctly according to ITU).
If a capture-card and/or it's driver at 720*576 resolution fills the whole horizontal range with active video information or leaves only a few pixels in the border, then it behaves incorrect and is not ITU compliant.

What is aspect ratio?

So far everything has been foreplay. That is supposed to be the best part anyway. ; -) .
We leave PAL plus and "genuine" 16:9 time aside, since to my knowledge there is currently no capture-card that supports it.

So, the 4:3-format remains.

Let's deal with DAR (display aspect ratio) first. A DAR of 4:3 means that if one would project he picture on a canvas of 4m wide, the height would be 3m. Or 4m wide and 3m high. Or 2m wide and 1.5m high, or, or..... Briefly: a factor of 1.3333333.... (4/3)

Back to television. Our only vertical constant is the number of lines, 576 These 576 lines multiplied by the DAR-factor 1.33333 gives 768 horizontal points.

Here we have our “industry norm” 768*576.

That is a bit odd, since these 768*576 contain exactly the same picture information as the 702*576 ITU.

Thus we come to another factor: The PAR (pixel aspect ratio). One must consider the PAR if one converts between the formats for different devices.
On the PC monitor there is only one PAR, i.e. 1:1. The pixels are square. A pixel on the monitor, if it were 1mm high, would then also be 1mm wide.
This explains the resolutions that are used (640*480, 800*600, 1024*768..).
Always, DAR 4:3 and PAR 1:1. An exception is e.g. 1280*1024; and already, circles are no longer round.
Devices that output signals for TV (e.g. DVD-Player) use another PAR, 54/59, or a factor of ca. ~1:1.0926. So these pixels are nonsquare and about a factor 1.0926 wider than high.

You could see the PAR as a distortion or stretch factor.

There are more / other PAR’s. You only need to know about one other, the generic- PAR (45/48 or 1:1.066666).

With PAL, you can get by using three different PAR’s (1:1, 54/59 and 45/48). We leave it at that, so it doesn’t get too complicated. With these you can calculate with up to one pixel precision. Close enough ;-)

Finally: 702 (ITU) vs. 704 (MPEG2). Those are of course approximations, representing a compromise between different video specifications and mathematical necessities (divisibility).
We have the following constants: 576 lines, 4:3 picture format and PAR 54/59. Once again our "Industry norm": 576*(4/3) = 768

This 768 at 54/59 "distorts", so 768*(54/59) = 702.915 results in ca. 703 horizontal pixels. Mathematically, the correct resolution at PAR 54/59 would therefore be 703*576. With this, both the ITU and the MPEG-norm are about one pixel “off”. These deviations can safely be ignored.

Overscan

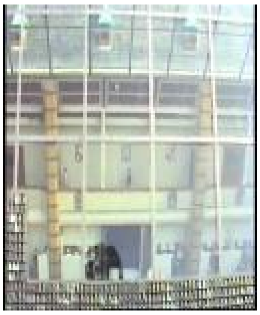
I already mentioned it briefly before. There is another thing apart from PAR, which one must know. Depending upon the selected capture resolution one either has one or has no overscan range. Of course, this overscan range must be considered when converting between resolutions. What counts, is the actual motion-area without overscan area or, without the black bars. All pictures contain the exact same information!

768*576
no overscan
PAR 1:1

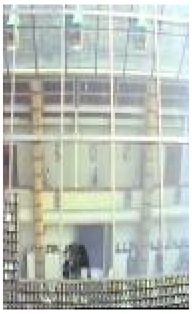
720*576
no overscan
not ITU-compliant!
Generic PAR 45/48

720*576
D1 and DV
2*8 to 2*9
pixel
overscan
PAR 54/59

704*576
no or rather
max 2*1 pixel
overscan
PAR 54/59



480*576
SVCD 2/3 D1
PAR 54/59 and
squished to 2/3
+ ca. 2*5 to 2*6
pixel overscan



352*576
half D1
no overscan
PAR 54/59 and
compressed to half

Calculating resolution conversions, but doing it correctly!

What is the most important when converting resolutions?
First, the source and target resolution. From these, the actual motion area or coded pixels. You must take the overscan into account. And finally, source and target PAR.
If the source- or the target-video is not square pixel (PAR 1:1), always use PAR 1:1 an intermediate step.

- -Source video: determine actual motion area (get rid of the black borders), resize with source PAR to get to PAR 1:1 (intermediate step), calculate “real” DAR.

- -Target video: determine the desired target resolution, resize with target PAR, taking into account the factors overscan, dividibility, block/macro block optimization, comparable DAR (to source video) and adjusting the source video to the target DAR by cropping.

Example calculations

<u>Example 1:</u>
Let’s start with the simplest case. We have digitized a VHS video using a TV-card (BT8x8) at resolution 768*576. The target resolution is 640*480 for MPEG4 or Divx. Both resolutions are square-pixel and have a DAR of 4:3 or 1.3333. Theoretically one could convert 1:1, if it weren’t for those dirty edges in the recording. So the recording first has to be cropped.

Assume that in order to get rid of the black bars and the tattered edges, we have to crop 8 pixels on top, 10 at the bottom and 4 pixels both left and right. The actual motion area turns out to be 760 (768-(2*4)) horizontal and 558 (576-(8+10)) vertical. So our source DAR is 760/558=1.362.

However, we want to get to 640/480=1.3333. Since the DAR is to high we have to adjust horizontally.

So we take the vertical as reference: 558*1.333333=744. We now have to crop the video a further 16 pixels (760-744) horizontally. That means an *additional* 8 pixels both left and right.

So we end up having to crop our video 8 pixels on top, 10 at the bottom but 12 Pixel both left and right.

This results in 744*558 coded pixels and a DAR von 744/558=1.333333.
This we can resize to 640*480.

Example 2:
The source is, (just like in example 1) a BT-capture from VHS. The target now is a SVCD.

The coded pixel, 760*558, and also the DAR of 1.362 we can copy from example 1.

Our target is SVCD 2/3 D1 (480*576). Additionally, there must be 2 blocks (16 pixels) TV-overscan all around. Horizontal 480-(2*16)=448, Vertical 576-(2*16)=544.
So our target resolution (coded Pixel) is 448*544.
This we now convert to PAR=1:1 (intermediate step).
We remember: SVCD is "squished" to 2/3. So 448*3/2=672.
Additionally, next to the squishing we have to take the PAR into account, so also 672*1.0926=~734

The "true" square target resolution with PAR=1:1 should therefore be 734*544. Now we calculate the actual DAR. 734/544=1.34926.
So the "true" DAR for the target is 1.34926.

The DAR of our source is (see above): 760/558=1.362.

We have to adjust our source format further by cropping. Since the DAR is again to high, we again adjust the horizontal
558*1.34926=752.887, so (because must be dividable by 2) ca. 752.

Our source material must be cropped to 752*558, and afterwards resized to 448*544 and letterboxed to 480*576.

Example 3:
This time the source is a TV-Capture of a 16:9-movie with generic-PAR (720*576). The generic-PAR can be recognized by the fact that the entire 720 Pixel contain horizontal picture information. Target is again a SVCD.

To get rid of the black bars we have to crop 58 Lines both at the top and at the bottom. So now we have 720*460 coded Pixel.

Since I have a fetish for cleanliness and hate dirty edges, I also crop 2 pixels left and right, resulting in 716*460 coded Pixels.

Now convert horizontally: (with generic-PAR) 716*1.06667=~764.

So our "true" DAR is 764/460=1.6609.

Again I define 2 blocks TV overscan. This time only horizontally. That's only logical ;-)

Target format is 448 (480-(2*16)) converted to PAR 1:1 - like in example 2 - = 734.

Try to reach closest possible matched DAR: so 734/1.6609 = ~442, but also Block- / Macroblock-optimized (Dividable by 32) = 448.
The actual target-DAR will be 734/448=1.6384.

Taking again the vertical as reference: 460*1.6384 = 753.664 (convert with Generic-PAR) 753.664/1.06666 = 706.56 or taking in account dividibillity by 2: 706.

Result: crop the source to 706*460, resize to 448*448 and letterbox to 480*576

Example 4:
Target is again (like 3.) a TV-Capture of a 16:9 movie with generic PAR (720*576).
Target is Half-D1 (352*576).
Cropping and DAR like in 3.

Because of the strong squishing for Half-D1 we now can only use one block TV overscan.

Target format 336 (352-(2*8)) at PAR 1:1 converted (336*2*1.0926 =734).
So we have the same target DAR as in 3.

We again crop the source to 706*460.
But now we resize to 336*448 and letterbox to 352*576.

And once again to Overscan ;-)
We have to distinguish between the overscan in broadcasted/encoded material (see the topic "overscan") and the TV overscan we define ourselves during encoding.
Here the purpose is to leave parts black that the TV won't depict anyhow. This saves bit rate, which can be invested in the actually visible picture and improves quality.

So for Example **2** and **3**: Since we letterbox ourselves, we don't have to pay attention to the SVCD-overscan. Only the "coded pixel" and the correct PAR count.

For whomever finds all these calculations to troublesome (and error-prone), we recommend the excellent FitCD: http://shh.dvdboard.de/

Though the program is freeware, the author would appreciate every € as a "Thank you" for all the time he put into it.

Workings of capture-chips:

BT848/878 (almost all TV-cards):

PAL is digitized at 17.735 MHz. Scaling produces the needed horizontal resolutions. This results for PAL in the 64 µs window 1135.04 horizontal pixels. Those are sent through the "scaler" and "Ultralock". This results in 944 pixels square format or 864 CCIR- compatible (720 horizontal with correct 54/59 PAR). On this, a window is imposed which is 768 resp. 720 pixels wide.
So the resulting window is 52.0678 µs (for 768) or 53.33 µs (for 720).
The BT-Chips digitize too much (768) or exactly correct (720).

So much for theory. The recommendations by Brooktree/Conextant are o.k. The reality looks different. The digitized range, which the capture driver passes on, is always substantially smaller (about 200-300 ns less). Besides this, at 720 the scaling is completely wrong.

That leaves only one conclusion: the drivers are wrong.

Philips SAA7108/7113 (Asus/Nvidia cards):

These digitize at 13.5 MHz (ITU-conform 720 Pixel in a 53.33 µs-window) The chip (or driver) passes on a 704 pixel horizontal part of this.
The result is o.k. (Measured on my Asus V3800 ultra deluxe).
It is assumed that the PAR with the Phillips chip is always correct, since this chip has no scaler, so wrongly programmed drivers cannot distort anything.

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

Though this thing is probably no longer in use anywhere, nevertheless out of interest: It digitizes at 13.5 MHz
The 720 or 704-resolution should be correct.
Scaling produces the square resolutions. So dependent on the driver.
Sadly I cannot confirm this (any more)

ATI with integrated Video-Chip:

About this one cannot find any info.
The 720 resolution is definitely not ITU compliant. So you have to use the generic-PAR.
The 704 resolution should be more or less correct. This could depend on the driver however, since apparently a scaler is used.

EBU Technical Recommendation R92-1999
Active picture area and picture centring in analogue and digital 625/50 television systems

EBU Committee	*First Issued*	*Revised*	*Re-issued*
PMC	1998	1999	

Keywords: 16:9, Cameras television, Interfaces video (digital), telecines

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.60l specifies a line length of 720 luminance pixels (13.5 MHz sampling). ITU-R Recommendation BT.470[2], specifies a line length of 52 µs 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**