Automated video optimisation

Egg, 26 February 2024: For some months now, the ArchivistaBox desktop has offered the option of optimising videos so that they require as little storage space as possible in HD quality. The corresponding tools are now also available in the open source Linux distribution AVMultimedia. The new version 2024/II also includes the 'subtitleadjust' utility to efficiently edit or modify subtitle files.

https://archivista.ch/cms/wp-content/uploads/file/avmultimedia_vidopt_eng.mp4

Hurdles in the optimisation of videos

Let's start by asking why the solution presented here was created in the first place. Currently, video files can be obtained from many legal sources as part of private copying. Be it via MediathekView, via Internet television such as Yallo.tv or Teleboy.tv or even from YouTube (here the easiest way is via FreeTube). Unfortunately, these files are available in very different qualities.

This often results in considerable file sizes, which, for a private film archive for example, can put a lot of strain on the hard drive. A film downloaded from Teleboy in HD quality can easily take up 6 to 9 GB. Although MediathekView allows you to specify the quality when downloading, the attributes 'High', 'Medium' and 'Low' do not mean that the corresponding videos are only available in the same resolution.

Is Abos Schrift Ansicht	Hilfe								
Suche: Q+ spielfilm									
 ✓ Suche: 									
Thoma		• 🖻 🔻							
Thema	Titel	Datum	Zeit	Dauer	Größe [MB]	HQ	UT	Geo	URL
Fernsehfilm	🔒 Suche Mann für meine Frau - Spielfilm, 🕨 🛓 📕	20.10.2019	21:44:00	01:26:49	1259	2		<u> </u>	https://rodlzdf-a.akamaihd.net/dach/3sat/23/04/230416_suche_mann_f
Krause	Krauses Glück - Spielfilm, Deutschland 2016 🕨 🛓 📕	07.07.2019	22:25:00	01:28:12	1524	1		_	https://rodlzdf-a.akamaihd.net/none/3sat/19/07/190707_krauses_glueck
Kultur	Grimme-Preis für den Spielfilm "Hanne" 🛛 🕨 🛓 📕	21.08.2020	19:30:00	00:02:20	40	1			https://rodlzdf-a.akamaihd.net/none/3sat/20/08/200820_hanne_event/1
puzzle	"Running Against The Wind" - Ein äthiopis 🕨 🛓 📕	28.07.2023	12:14:00	00:29:08	423	~		_	https://rodlzdf-a.akamaihd.net/none/3sat/23/07/230728_1215_sendung
Spielfilm	🚔 🖾 Rivale - Spielfilm, Deutschland 2020 (🕨 🛓 📕	15.02.2024	21:02:00	01:31:34	1262	1			https://rodlzdf-a.akamaihd.net/dach/3sat/24/02/240215_rivale_spielfilm
Spielfilm	🖾 Dampfnudelblues. Ein Eberhoferkrimi (🕨 🛓 📕	17.07.2019	20:16:00	01:26:48	1263	1			https://rodlzdf-a.akamaihd.net/none/3sat/24/02/240212_dampfnudelblu
Spielfilm	🚔 Im Weißen Rössl - Spielfilm, Österreich 🕨 🛓 📕	31.05.2018	15:33:00	01:39:03	1461	1		<u> </u>	https://rodlzdf-a.akamaihd.net/dach/3sat/24/02/240204_im_weissen_r
Spielfilm	Dampfnudelblues.Ein Eberhoferkrimi 📃 🕨 🛓 📕	17.07.2019	20:16:00	01:26:48	1263	1		_	https://rodlzdf-a.akamaihd.net/none/3sat/24/02/240212_dampfnudelblu
Spielfilm	Cop Secret - Spielfilm, Island 2021 🛛 🕨 🛓 📕	18.02.2024	23:20:00	01:35:53	1370	1			https://rodlzdf-a.akamaihd.net/none/3sat/24/02/240218_cop_secret_spi
Spielfilm	🚔 The Painted Bird - Spielfilm, Tschechien 🕨 🛓 📕	17.02.2024	23:17:00	02:42:55	2152	v			https://rodlzdf-a.akamaihd.net/dach/3sat/24/02/240217_the_painted_bi
Spielfilm	🚔 Jo Nasbø's Headhunters - Spielfilm, Nor 🕨 🛓 📕	19.05.2017	22:39:00	01:33:47	1351	1			https://rodlzdf-a.akamaihd.net/dach/3sat/23/01/230129_jo_nesboes_he
Spielfilm	Riva Jo Nesbø's Headhunters - Spielfilm, Norwegen	/Deutschland 2	011 :02:00	01:31:34	1262	1		<u> </u>	https://rodlzdf-a.akamaihd.net/dach/3sat/24/02/240215_rivale_spielfilm
Spielfilm	🚔 Große Freiheit - Spielfilm, Österreich/D 🕨 🛓 📕	15.02.2024	22:34:00	01:51:34	1545	v			https://rodlzdf-a.akamaihd.net/dach/3sat/24/02/240215_grosse_freiheit
Spielfilm	🚔 🖾 Große Freiheit - Spielfilm, Österreic 🕨 🛓 📕	15.02.2024	22:34:00	01:51:34	1545	1		<u> </u>	https://rodlzdf-a.akamaihd.net/dach/3sat/24/02/240215_grosse_freiheit
Spielfilm	lm Schwarzen Rössl - Heimatfilm, Österrei 🕨 🛓 📕	08.12.2017	15:19:00	01:31:10	1310	\checkmark		_	https://rodlzdf-a.akamaihd.net/none/3sat/24/02/240204_im_schwarzen
Spielfilm	🔒 Klondike - Spielfilm, Ukraine/Türkei 2022 🕨 🛓 📕	22.04.2023	22:48:00	01:36:10	1340	1			https://rodlzdf-a.akamaihd.net/dach/3sat/23/04/230422_klondike_spielfi
Spielfilm	🔒 🖾 Astrid - Spielfilm, Schweden/Dänem 🕨 🛓 📕	21.05.2020	20:15:00	01:57:57	1704	1		-	https://rodlzdf-a.akamaihd.net/dach/3sat/22/01/220123_astrid_spielfilm
Spielfilm	🚔 Astrid - Spielfilm, Schweden/Dänemark 🕨 🛓 📕	21.05.2020	20:15:00	01:57:57	1704	1	4		https://rodlzdf-a.akamaihd.net/dach/3sat/22/01/220123_astrid_spielfilm
Spielfilm	🔒 Tod im Strandhaus - Spielfilm, USA 2019 🛛 🕨 🛓 📕	10.05.2021	22:15:00	01:24:48	1220	1			https://rodlzdf-a.akamaihd.net/dach/3sat/24/02/240211_tod_im_strandh
Spielfilm	Weißer, weißer Tag - Spielfilm, Island/Sch 🕨 🛓 📕	14.02.2024	22:26:00	01:44:33	1492	\checkmark		_	https://rodlzdf-a.akamaihd.net/none/3sat/24/02/240214_weisser_weiss
Spielfilm	🚔 Postcard Killings - Spielfilm, USA/Deuts 🕨 🛓 📕	07.11.2022	22:15:00	01:40:11	1402	1			https://rodizdf-a.akamaihd.net/dach/3sat/24/01/240128_postcard_killing
Spielfilm	Tagundnachtgleiche - Spielfilm, Deutschla 🕨 🛓 📕	07.01.2024	00:18:00	01:42:36	1465	v		_	https://rodlzdf-a.akamaihd.net/none/3sat/24/01/240106_tagundnachtgle
Spielfilm	Sehnsucht nach Rimini 🛛 🕨 🛓 📕	04.08.2018	17:30:00	01:28:37	1263	-		_	https://rodlzdf-a.akamaihd.net/none/3sat/24/01/240120_sehnsucht_nac
Spielfilm	Ein Mord mit Aussicht - Spielfilm, Deutschl 🕨 🛓 📕	07.05.2019	20:15:00	01:26:57	1240	\checkmark		.	https://rodlzdf-a.akamaihd.net/none/3sat/22/07/220710_mord_mit_auss
Spielfilm	🔒 🖾 Die Kinder von Windermere - Spielfil 🕨 🛓 📕	27.01.2020	22:15:00	01:27:56	1229	v			https://rodlzdf-a.akamaihd.net/dach/3sat/24/01/240126_die_kinder_von
Spielfilm	Rat mal, wer zur Hochzeit kommt 🛛 🕨 🛓 📕	12.08.2018	20:15:00	01:28:08	1265	~		1	https://rodlzdf-a.akamaihd.net/none/3sat/22/02/220219_rat_mal_wer_z
Spielfilm	Freies Land - Spielfilm, Deutschland 2019 🛛 🕨 🛓 📕	04.10.2021	22:15:00	02:04:06	1744	1	1	1	https://rodlzdf-a.akamaihd.net/ebu/3sat/24/01/240112_freies_land_spie
Spielfilm	🔤 In Berlin wächst kein Orangenbaum - S 🕨 🛓 📕	13.01.2024	23:39:00	01:28:18	1272	1			https://rodlzdf-a.akamaihd.net/none/3sat/23/05/230531_in_berlin_waec
Spielfilm	🖾 Freies Land - Spielfilm, Deutschland 20 🕨 🛓 📕	04.10.2021	22:15:00	02:04:06	1744	1		_	https://rodlzdf-a.akamaihd.net/ebu/3sat/24/01/240112_freies_land_spie
	Vause Vause Vause vause spelifim	Crause Krauses Guice - Spelfilm, Deutschind 2019 > Valur Grimme - Pre-Sir ür efts pelfilm, Thane" > vazze "Running Against The Wind" - Ein Attripse	Crause Krause Guide - Speltim, Detection and 2019 P = 0 > 0.7.7.2019 Subur Grimme-Free's for den Speltim Hanne P = 21.08.2020 system "Running Against The Wind" - Ein Attoignes P = 21.08.2020 system "Running Against The Wind" - Ein Attoignes P = 21.08.2020 system "Running Against The Wind" - Ein Attoignes P = 15.02.2024 spelfilm Campfruideblues. Ein Eberhoferkrimi P = 31.05.2018 spelfilm Campfruideblues. Ein Eberhoferkrimi P = 18.02.2024 spelfilm Campfruideblues. Spelfilm, Tochechen. P = 19.05.2017 spelfilm Campfruideblues. Spelfilm, Tochechen. P = 19.05.2017 spelfilm A forde Frieht-1 Spelfilm, Oxterreich. P = 19.05.2017 spelfilm A forde Frieht-1 Spelfilm, Oxterreich. P = 19.05.2017 spelfilm A forde Frieht-1 Spelfilm, Oxterreich. P = 10.05.2020 spelfilm A forde Frieht-1 Spelfilm, Oxterreich. P = 10.05.2020 spelfilm A forde Frieht-1 Spelfilm, NetworeIORen.	Graue Kraues Gluck - Spelitim, Deutschland 2010 ▶ ▲ 0.70.2019 222500 Subur Grimme-Freis für der Spelitim Hanne" ▶ ▲ 2.08.2020 1214.600 syzzie "Running Against The Wind" - Ein striogia. ▶ ▲ 2.08.2020 1214.000 systelfitim 🖵 Dampfrudeblues. Ein Betrioferkrimin, Nature Schlein, Nature Sc	Grane Kraues Gluck - specifium, Deutschand 2018 ▶ ▲ 0.07.019 222.500 01.07.019 Subur Grimme-Freiß fürde specifium Hanne" ▶ ▲ 1.08.202 1.03.000 00.02.20 syzzie "Running Against The Wind" - Ein Athiogia. ▶ ▲ 1.08.202 1.02.000 00.02.20 syzzie "Running Against The Wind" - Ein Athiogia. ▶ ▲ 1.02.0224 1.02.000 0.13.14 specifilm □Dampfnudeblues. Ein Eberhoferkrimi. ▶ ▲ 1.07.2019 2016.000 0.12.85.43 specifilm □ampfnudeblues. Ein Eberhoferkrimi. ▶ ▲ 1.07.2019 2016.000 0.12.85.43 specifilm □ampfnudeblues. Ein Eberhoferkrimi. ▶ ▲ 1.07.2019 2016.000 0.12.85.43 specifilm △ nofback- specifilm, Nature Assistilm, Nature Nature Nature Nature Nature Nature Nature Nature Nature Natu	Value Knames Guick - Speeling, Deutschand 2010 P ≥ 10,07,0710 22,2500 012,812 1544 Subur Grimme, Presid in den Speling, Manne Manne > 2,08,200 10,8202 30,000 00,22,020 40 Suzzie "Running Against The Wind" - Ein Athiogiss. 2,80,7020 11,80200 00,29,08 423 Spelfiffin CD Dampfrudeblues. Ein Eberhoferkrimi (L) 2,80,70201 21,6000 01,26,48 1263 Spelfiffin CD Spelfing, Deutscherkrimi (L) 2,80,70201 21,600 01,26,48 1263 Spelfiffin Campfrudeblues. Ein Eberhoferkrimi (L) 2,80,70201 21,600 01,26,48 1263 Spelfiffin Cam Secret - Spelfiffin, Notrow-Secret - Spelfifin, Notrow-Secret - Spelfiffin, Notrow-Secret - Spelfiffin, No	Grane Kraues Guick - Spellinin, Deutschand 2010 ▶ ▲ 0.77.2013 22.25:00 012.81:12 12.14 2 Subur Grimme, Preis Tranches, Dirith Hanner, Marken ▲ 1.08.2020 101.83:12 12.44 2 Suzzie "Running Against The Wind" - Ein Stringes, - ▲ 1.08.2020 102.80:10 002.90:8 43.2 2 Spelfillin Clin Scripter Mind Clin Scripter Mind 1.50.22042 1.20.200 012.84:8 126.3 2 Spelfillin Clin Scripter Mind Market Mind 1.50.2018 1.50.20	Grane Kraues Guick - spetifilin, Deutschand 2019 ▶ ▲ 0.07.019 222500 012812 124 2 Subur Grimme-Freiß für des Spetifilin, Hanne" ▶ ▲ 2.08.2020 1214.00 0.02.20 40 2 Spetifilin Common Against The Wind" - Ein Attroigen. ▶ ▲ 2.08.2020 1214.00 0.02.20 40 2 Spetifilin Common Against The Wind" - Ein Attroigen. ▶ ▲ 1.50.2024 120.2000 0.12.6.48 1263 2 Spetifilin Common Against The Bancherkrimin ▶ ▲ 1.70.70219 2016.00 0.12.6.48 1263 2 1263	Kraue Kraues Suitx- Specific III, Deutschand 2010 P = 10 0.7.07.2019 22.25:00 012.8121 1.9.4 P = 1 Suitur Grimme-Preis Tird ers Specific III Name > = 1 21.08.2020 1.9.00 00.22.90 4.0 2

And for those who would like to add at this point that a private film archive can always be stored on a hard drive, it should be noted that a feature film such as 'The Good Shepherd' (obtained via MediathekView) requires approx. 6.5 GB of space. Extrapolated to one TB, this is 130 films. After optimisation with the tools presented here, the same film still requires approx. 1.2 GB. Of course, video files can also be converted using standard programmes such as FFMPeg or VLC. However, it is not always easy to "get" the right parameters. This is because different parameters are required depending on the quality of the original files.



Automated optimisation with vidopt

The solution presented here was developed for precisely this reason. "Guessing" the optimum values for ffmpeg often worked, but not always. If a film is available in 50 frames/second, it can be compressed quite easily with 25 frames/second. However, 25 frames/second should not be used for files with 29.97 or 30 frames/second. With 60 frames/second, it makes more sense to carry out the optimisation at 30 frames/second.

A film with 1920×1080 pixels can be scaled to HD quality with 1280×720 pixels relatively "safely". The situation is not quite so simple if the film has 1440×720 pixels. Should it be optimised to 960×480 pixels?

Video files can also contain many audio tracks. Of course, it can make sense to preserve these for films that contain several languages. However, this is not the case if there are two or more audio tracks (e.g. eac3 and aac). In this case, one audio track is sufficient.

The videos are optimised in three steps. Before this can be started, the files to be optimised must be placed in a folder (e.g. /home/archivista/data/in for AVMultimedia). The following three programmes must then be started one after the other:

```
vidopt1 /home/archivista/data/in
vidopt2 /home/archivista/data/in
vidopt3 /home/archivista/data/in
```

It is important to note that vidopt2 can only be started once all files have been

processed with vidopt1. The same applies to vidopt3. The console program 'ffmpeg' is used for optimisation. If you want to know how these programmes work, you can take a look at the log files.

- tail -f /home/data/archivista/av.log
- tail -f /home/data/archivista/vidopt1.log
- tail -f /home/data/archivista/vidopt2.log

After running the three programs, the optimised videos can be found in the directory. The original film files can be found in the following directory:

/home/archivista/data/vidbackup

Finally, three comments: Firstly, the optimisation requires considerable computing power. A minimum of 8 CPU cores or 16 threads should be available. Secondly, the more CPU cores, the better, as all CPU cores plus/minus are used for the optimisation. Thirdly, the optimisation only takes place if ffmpeg is not already running. This ensures that the corresponding computer is not overloaded with jobs.



Optimisation of subtitle files

Newer streams sometimes also have one (or more) "subtitle tracks". Older films, on the other hand, are often only available in English. German subtitles can often be found for more well-known works (e.g. via opensubtitles.org).

However, these subtitles somehow don't fit. Be it because, for example, the subtitles are available at 24.97 frames/second, but the film runs a little "faster" at 25 frames/second. Even such a small difference results in a difference of 162 frames or 6.48 seconds over a 90-minute feature film (3600*1.5*25.0-3600*1.5*24.97). This is of course enough to "destroy" any cinematic enjoyment.

Subtitle files can also contain the same text passages several times. Here is an example from the Swiss television news programme of 21 February 2021:

00:03.600 --> 00:03.620Mit Live-Untertiteln von SWISS TXT 00:05.200 - -> 00:05.220Mit Live-Untertiteln von SWISS TXT 00:07.380 --> 00:07.400Guten Abend, willkommen zur "Tagesschau". 00:08.400 --> 00:08.420Guten Abend, willkommen zur "Tagesschau". 00:10.000 --> 00:10.020Guten Abend, willkommen zur "Tagesschau". 00:10.640 --> 00:10.660Das sind unsere Schlagzeilen von heute Mittwoch:

With video players (e.g. mpv), this currently leads to the same subtitles being displayed several times, although one display would be sufficient. When optimising video files, any subtitles are automatically optimised. The example above becomes:

00:00:03.520 --> 00:00:05.409 Mit Live-Untertiteln von SWISS TXT

00:00:07.380 --> 00:00:10.020 Guten Abend,^M willkommen zur "Tagesschau".

00:00:10.640 --> 00:00:13.307 Das sind unsere Schlagzeilen^M von heute Mittwoch:

This means that the subtitles are not displayed multiple times (sometimes "flickering") during playback. The separate 'subtitleadjust' tool is called up automatically during video optimization. If a separate subtitle file is available, 'subtitleadjust' can also be called up individually. To eliminate any discrepancies in speed or start, there are

corresponding options in 'subtitleadjust':

subtitleadjust in.srt out.srt [change=*x.xxx|+hh:mm:ss|hh:mm:ss]

After the name of the output file, the speed and the start position can be adjusted. Here are two examples:

subtitleadjust in.srt out.srt *0.999

This results in the time data in out.srt being entered correspondingly slower (correction from 25 to 24.97 frames/second). If the start position is to be shifted forwards by 20 seconds, this can be solved in this way:

subtitleadjust in.srt out.srt -00:00:20

Sometimes subtitle files also have character sets that are not consistent. The programme subtitlefix8859-1 is available for this purpose. The following command can be used to add a subtitle file to a video:

subtitle vidwithout.mp4 subtitle.vtt videowith.mp4



Final remarks

The programmes vidopt1, vidopt2, vidopt3 and subtitleadjust presented here are available in source code on every AVMultimedia distribution under /home/cvs/arcihvista/jobs. The wrappers vidopt1, vidopt2, vidopt3 and subtitleadjust are available under /usr/bin so that they can be called easily.

There are also wrappers to adjust video and subtitle files accordingly. The most important tools are

```
videoonly video.mp4 videoonly.mp4 [tracknr]
videocombine videoonly.mp4 audioonly.mp4 video.mp4
videocombine2 videoonly.mp4 audio1.mp4 audio2.mp4 video.mp4
audioonly video.mp4 audioonly.mp4 [tracknr]
subtitle video.mp4 subtitle.srt|subtitle.vtt
videosubtitle.mp4 [deu|eng|fra]
subtitle2 vidonly.mp4 audio1.mp4 audio2.mp4 sub.srt|sub.vtt
vidout.mp4 [lang]
subtitleonly video.mp4 subtitle.srt [tracknr]
subtitlefix8859-1 subtitle8859-1.srt subtitleutf8.srt
```

All these little helpers are small wrappers that make working with the 'ffmpeg' console programme a little easier. An example of this is videocombine2:

```
#!/bin/bash
if [ ! "$1" ];then
    echo "$0 videoonly.mp4 audio1.mp4 audio2.mp4 video.mp4"
    exit 1
```

```
else
in="$1"
a1="$2"
a2="$3"
out="$4"
ffmpeg -i "$in" -i "$a1" -i "$a2" \
        -map 0:v:0 -map 1:a:0 -map 2:a:0 -c copy "$out"
fi
```

If you often work with ffmpeg, you may not need the utilities. However, for those who do not work with ffmpeg on a daily basis, these tools may be helpful. It should also be noted that there are currently no graphical tools for any of the helpers presented here. However, calling them up via the console is deliberately kept simple so that this should not be a problem in day-to-day work.



Finally, I would like to make the following comment. The programmes presented here comprise approx. 1500 lines of code. The programmes were applied to several thousand video files. The error rate is 1 to 2 per thousand, which were not satisfactorily optimised with the test files. It had to be established that a file was optimised too much or too "weakly".

Several hundred hours of work were necessary before the tools ran as they currently do. This work was motivated by the fact that it is not practical for a private film archive to have to purchase several hard drives with many dozens of TB.

Compressing videos always leads to certain losses in quality. Measured against the price saved on hardware, this is probably the lesser evil. In the author's opinion, a home cinema experience can certainly be achieved with a standard projector, even if "only" HD quality (1280×720 pixels) is available. With this in mind, have fun with AVMultimedia 2024/II.

Note: Customers of the ArchivistaBox should know that these tools are of course also available in the current version of the ArchivistaBox. For the optional applications (e.g. LosslessCut), the flag 'optos' can be set for customers. If this is the case, the opt.os file is also installed during the update. And because the flag is available for the first time with version 2024/II, the update to version 2024/II must be carried (this time only) out twice.