



Best way to measure performance

From: <https://superuser.com/questions/338725/compare-two-video-files-to-find-out-which-has-best-quality>

I work in video quality research, and it's hard to give a simple answer to your question. What you want is a program that gives you a [Mean Opinion Score](#) (MOS) of a video, i.e. a number between 1 and 5, or between 0 and 100, which corresponds to the quality as perceived by a human being.

Why you cannot simply compare bitrate/resolution/etc.

Just comparing video resolution won't tell anything about the quality. In fact, it may be completely misleading. A 1080p movie rip at 700MB size might look worse than a 720p rip at 700MB, because for the former, the bitrate is too low, which introduces all kinds of compression artifacts.

The same goes for comparing bitrate at similar frame sizes, as different encoders can actually deliver better quality at less bitrate, or vice-versa. For example, a 720p 700MB rip produced with XviD will look worse than a 700MB rip produced with x264, because the latter is much more efficient.

You would also have to define how a final "integral score" (the MOS) is composed of the individual quality factors. This *heavily* depends on several things, including but not limited to:

- the type of videos you are comparing (cartoons, movies, news, etc.)
- their length
- their viewing audience
- their original frame size
- their original "quality" before they were encoded

We're not even talking about how humans would perceive the videos. Let's assume you have a friend who is watching movies because he or she enjoys crisp details and high motion resolution. They would be much more critical when seeing a low quality rip than a friend who is just watching movies for *their content*. They

probably would not care about the quality so much, as long as the movie is funny or entertaining.

There are different types of video quality metrics!

Let me give you a list of what I think of is most commonly used for basic evaluation of video quality today. There exist several video quality metrics, which can be classified according to which kind of information is used to determine the quality. In principle and very simply speaking, you distinguish between the following:

- **No-reference metrics** – They just have one video as input and output a quality score. In your case you are looking for a *no-reference metric*, because you often do not even have the original video. Such a metric will take one video and output one quality score. Here are [some examples](#) of problems a NR metric will detect (e.g. blurring).
- **Full-reference metrics** – They have two inputs, one being the original input video and the other being the encoded video. For example, you could take a DVD movie, then create two rips from it, and use a full-reference metric to estimate the quality loss between the original DVD movie (i.e. the MPEG-2 video on the disc) and your rips. This will take a long time to compute, but it's more accurate.

The above metrics look at video coding quality, but there are also metrics that incorporate problems like initial loading times and stalling events when streaming video (e.g. [ITU-T P.1203](#)).

What software can I use?

Here is a list of ready-to-use tools that you can use to test some metrics (some are for Windows only):

- [VMAF – Video Multi-Method Assessment Fusion](#) by Netflix (more info [here](#))
- [VQMT – Video Quality Measurement Tool](#) by the EPFL in Lausanne, Switzerland
- [MSU Video Quality Tool](#), a commercial software
- [ITU-T P.1203 Implementation](#) for analysis of HTTP streaming quality
- [ITU-T P.1204.3 Implementation](#) for analysis of H.264, H.265 and VP9-encoded bitstreams

- [FFMetrics](#), a Windows GUI for several video quality metrics available in FFmpeg

Now what metrics are there?

PSNR, PSNR-HVS and PSNR-HVS-M

For starters, [PSNR](#) (Peak Signal-to-Noise Ratio) is a very simple-to-use but somewhat poor method of assessing video quality. It works relatively well though for most applications, but it does not give a good estimation of how humans would perceive the quality.

PSNR can be calculated frame-by-frame, and then you would for example average the PSNR of a whole video sequence to get the final score. Higher PSNR is better.

PSNR-HVS and PSNR-HVS-M are extensions of PSNR that try to emulate human visual perception, so they should be more accurate. [VQMT](#) and [MSU](#) can calculate PSNR, PSNR-HVS and PSNR-HVS-M between two videos.

SSIM, MS-SSIM

[Structural Similarity](#) (SSIM) is as easy to calculate as PSNR, and it delivers more accurate results, but still on a frame-by-frame basis. You will find some implementations under the Wikipedia link, or you can use [VQMT](#) or [MSU](#). These tools also include MS-SSIM, which gives better (i.e., more representative) results than SSIM, as well as a few other derivatives.

The results should be similar to PSNR. Again, you need to compare a reference to a processed video for this to work, and both videos should be of the same size.

VMAF

[Video Multi-Method Assessment Fusion](#) by Netflix is a set of tools to calculate video quality based on some existing metrics, which are then fused by machine learning methods into a final score between 0 and 100. Netflix have [explained the whole thing here](#):

[VMAF] predicts subjective quality by combining multiple elementary quality metrics. The basic rationale is that each

elementary metric may have its own strengths and weaknesses with respect to the source content characteristics, type of artifacts, and degree of distortion. By 'fusing' elementary metrics into a final metric using a machine-learning algorithm - in our case, a Support Vector Machine (SVM) regressor - which assigns weights to each elementary metric, the final metric could preserve all the strengths of the individual metrics, and deliver a more accurate final score.

You can also [use ffmpeg to calculate VMAF scores](#).

VQM

The [Video Quality Metric](#) was validated in the [Video Quality Experts Group](#) (VQEG) and is a very good full-reference algorithm. You can download VQM for free or use the implementation from MSU.

When you register and download, you want to use the *NTIA General Model* or the *Video Quality Model with Variable Frame Delay*.

ITU P.1204.3

This is an ITU-T standard for bitstream-based evaluation of video quality. It is a short term video quality prediction model that uses full bitstream data to estimate video quality scores on a segment level (for segments of ~10 seconds length).

A reference implementation can be [found on GitHub](#).

Other Metrics

- [PEVQ](#) is a standardized full-reference metric under ITU-T J.246. It aims at multimedia signals, but not HD video.
- [VQuad-HD](#) is another full-reference metric standardized as ITU-T J.341. Since it's newer, it's better suited for HD video.

Both of them are commercial solutions and you'll not find a software to download for them.

There are also some ITU standards on no-reference metrics, such as [ITU-T P.1201](#) and [ITU-T P.1202](#), which work with parameters from the bitstream for IPTV streaming. [ITU-T P.1203](#) can be used for adaptive streaming cases.

Summary

If you just seek to compare simple objectively measurable criteria like:

- Frame size
- Bit rate
- Frames per second
- Video resolution

... a simple call to `ffmpeg -i` should give you all the details you need at the beginning. Also have a look at the `-vstats` option. You could then summarize this in a spreadsheet. Note that when you encode videos, x264 for example will log stuff like PSNR straight to a file if you need to, so you can use these values later.

As for how to weigh these criteria, you should probably emphasize the bit rate – but only if you know that the codec is the same. You could generally say that when both videos use x264, the one with higher bitrate is better. Even more generally, you should choose a lower resolution when you have two videos with the same bitrate, since the degradation due to upscaling is not as bad as the degradation due to low bitrate.

Comparing different codecs according to their bit rate is *not possible* unless you know more about the content and the individual encoding settings. Frame rate is a very subjective thing too and should be counted into your measurements if it is well below 25 Hz.

To summarize, heavily emphasize the bitrate if it's the only thing you have. Don't forget to use your eyes, too :)