Why video quality measurement is important

Video quality measurement (often called "video quality assessment" in scientific literature) presents an important interest for video services that target human observers, like television broadcast, VOD, DVD, Blu-Ray, Mobile TV, Web TV, because, in the end, **for human observers, only quality matters**.

Therefore, if you are in charge of a video service, measuring video quality must enable you :

- to choose the best equipment among the products available on the market,
- to adjust the parameters of the equipment you already have in order to maximize the video quality they provide
- to define the quality level you have and to measure the effort needed to reach the acceptability level you desire for your video service.
- to control the visual quality of the videos you product, process, encode, transcode, decode, buy or sell.

How to measure video quality

Video quality can be measured in two different ways:

- The first, called "subjective quality assessment", consists of the use of human observers who should score video quality during experiments called "quality assessment tests". Since video quality is a subjective notion, subjective quality assessment can be considered as the best method to measure video quality. However, many recommendations have to be followed when realizing subjective quality assessment tests, otherwise the results of these tests often become useless (because of the lack of precision on the quality scores obtained during these tests). These recommendations lead this method to be very complicated. More, realizing subjective quality assessments tests is a long and expensive process. To get more information about this subject, please read the AccepTV's white paper on the subjective quality assessment methods (available on www.acceptv.com).

- The second one is called "objective quality assessment". It consists of the use of a computational method called "metric" (or "quality metric" or "quality criterion") which produces values that score video quality. One of the properties required for an video quality metric is that it should produce objective scores well correlated with subjective quality scores produced by human observers during quality assessment tests. From a practical point of view, such a metric is an algorithm able to score (on a scale) the quality of a tested video which may have been distorted. While computing quality scores can be simple, producing meaningful visual quality scores, which means scores well correlated with subjective quality scores given by human observers, is much more complicated.

The different types of video quality metrics

Several video quality metrics have been proposed in the past and are described in literature. They can be roughly divided into three main categories:

- Full Reference (FR) metrics,
- Reduced Reference (RR) metrics,
- No Reference (NR) metrics.

To get more information about these categories, please read the AccepTV's white paper on FR, RR and NR metrics (available on www.acceptv.com).

Obviously, all these metrics are a function of the distorted video. FR metrics are also a function of the original video which is assumed to be free from distortions (called the "reference video"). RR metrics require a partial knowledge of the reference video (this knowledge is called the "reduced reference"). At last, NR metrics don't have any information about the reference video.

Some recent quality assessment techniques can hardly be classified in these three categories, for example when quality is computed from a watermark distortion.

For RR metrics, it is difficult to choose the types of features which are the most useful to describe visual quality and to efficiently exploit these features.

Finally, it is difficult for NR metrics without any information on the reference image to distinguish which part of the image signal is due to distortions and which part is due to the reference video.

Other differences come from the fact that existing metrics use various techniques. Some metrics first compute a map of distortions - this is only really possible for FR metrics - whereas the others quantify by a parameter the importance of each distortion type, depending on the distortion process (like the blocking effect in JPEG image coding for example).

Moreover, some metrics include a model of the Human Visual System (HVS), for example to locally weigh visual distortions, whereas others more directly use the video signal.

FR metrics are evidently considered as the best way to get good performances in quality assessment since they can use the maximum amount of data about the video whom quality has to be measured. But they cannot be used in applications where the original video is not available or, as in a transmission context with lossy compression, cannot be logically transmitted with the compressed video data. At this level, the problem with the metrics used in transmission contexts (RR or NR) is that most of them are designed for a limited number of predefined distortion types (lack of genericity). Moreover, most criteria are not used because they are not available as a software. Therefore, they need to be implemented. This is one of the reasons why the simple PSNR criterion (which is a FR criterion) is still too often used as its implementation is very easy.

High Definition (HD) video quality

The question of video quality measurement is even more important in HD than it was in SD. This is due to several reasons. Indeed, MPEG4/AVC (also known as H.264), which is the compression method used for HD, can provide a better compression than the well established MPEG2 standard. Most people consider that, compared to MPEG2, MPEG4/AVC can reduce the bitrate by 2 while preserving the same video quality. But there are a few limitations to this enhancement:

- First, a progressive HD picture (1920x1080 pixels) contains about 2 million pixels whereas a standard definition (SD) picture (720x576 pixels) contains about 400 000 pixels. Therefore, a progressive HD picture contains 5 times the number of pixels of a SD picture (and remember we just said that MPEG4/AVC can compress twice better, but not five times better, than MPEG2).
- Even with interlaced contents, the problem is still the same (if you compare the number of pixels by second for interlaced HD and interlaced SD, you will have the same ratio of 5).
- More, even in a 720p picture (1280x720 pixels, also known as "HD Ready" by consumers), the problem remains. Indeed, a 720p picture contains about 1 million pixels. But 720p HD is replacing interlaced SD (576i) and therefore, there is still 5 times more pixels in 720p than in interlaced SD.
- From another point of view, since you compress much more the information, you are much more sensitive to transmission errors. One simple way to express that would be : you compress twice more, then you have to transmit twice better. This is even more important with HD IPTV. Since IPTV is based on the Internet, when a Internet packet is lost, up to 7 video packets (TS packets) may be lost. And 7 video packets may correspond to the beginning of a 300-frame GOP (in long GOP encoding). Even if this is very rare in reality, it means that loosing one packet of an I-frame in long GOP could correspond to the lost of up to 12 seconds of video.
- At last, consumers have very high expectations about HD. Most consumers expect to find in the SD to HD transition, the same quality improvement they found when DVB arrived. But DVB arrival was a transition from analog to digital television, which provided a quality improvement on video contents (except contents with fast motion and rich details, because of blocking effect visibility, on sports content for example, like football). The transition from SD to HD is different because it is mainly a transition from digital SD to digital HD. Therefore it is harder to provide the same quality improvement than the one that occurred when DVB arrived.

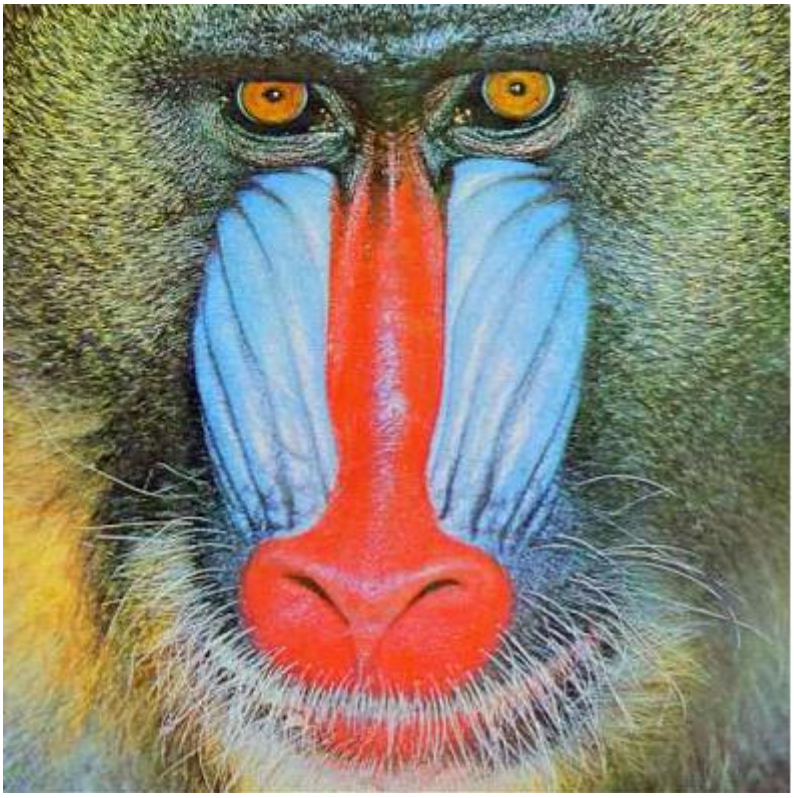
Conclusion

If you are in charge of video or television services deployment, you should now be convinced that you need to measure and optimize the video quality you produce.

AccepTV can help you in this mission and provide software and hardware solutions for video quality measurement.

You have to be aware that there are many simple algorithms or mathematical equations which claim to be able to compute video quality scores (like PSNR, MSE and even quite a lot of simple algorithms developed in academic research labs). Creating a program that computes scores for videos is really simple: any programmer with video programming skills can create programs like that every minute. But **the difficulty is to create an algorithm that computes scores that are highly correlated with subjective quality scores (which means creating a program that output scores which are very close to the quality scores given by human observers during subjective video quality assessment tests in normalized conditions)**.

Therefore you should avoid using these simple methods (like PSNR) because, even if they give you quantitative scores, these scores have no relation (or very few) with subjective judgements of video quality. For example, consider the two following images:



For the first image (fruits), human observers give a mean opinion score (MOS) of 1.69 (out of 5). This score corresponds to a "bad quality".

For the second image (mandrill), human observers give a mean opinion score (MOS) of 4.62 (out of 5). This score corresponds to a "very good quality".

For the first image (fruits), the PSNR gives a score of 29.40 dB.

For the second image (mandrill), the PSNR gives a score of 26.45 dB.

Therefore the PSNR considers that the first image has a better quality than the second image, whereas the observers say the opposite! This difference can be explained in good part by the masking effect. The masking effect is a visual perception (in fact it also exists in audio perception) that indicates how a signal (distortions in our case) can be invisible when they are surrounded by a lot of activity. On the left image, distortions are visible because they are mainly located on homogenous areas (on the grey background for example) whereas on the right image the distortions are mostly invisible because they are located on very contrasted areas (due to the animal hair). That's why the human visual system doesn't see the distortions whereas the PSNR takes them into account.

For More Information: http://www.acceptv.com
info@acceptv.com
Tel : +33 (0)6 10 71 11 64