Everything you wanted to know about Video codecs — but were too afraid to ask

David Wood
Head of New Technology, EBU

Digital video compression technology continues to evolve, and the choice of systems presents a difficult challenge for broadcasters and web content providers. In this article, the author explains some of the factors shaping the evolution of video compression technology, and offers some insights into the comparative performance of video compression systems. The article is based on a presentation given in spring 2003 to the EBU Technical Assembly in Moscow.

The "Lomo Compact" is more than just a humble camera made in St Petersburg. It is a camera artist's phenomenon. Advocates passionately claim that it takes the most beautiful pictures. And they do have a special and unique look. Taking pictures with it is now called "lomography", and there are clubs all around the world for people who are doing just that.

Figure 1
The author's "Lomo Compact" in action

The problem is that, when TV engineers look at the resulting pictures, what they see is almost horrifying – the colours are over-saturated, the gamma is wrong, there is vignetting, and more. In spite of this "distortion", many people love the camera's output. The point is that picture-quality evaluation is a very complex subject with many variables, and sometimes a lack of apparent logic. This is something we have to live with when we discuss quality in all areas, including picture coding and compression.

We cannot often discuss picture quality in simple terms of "good" or "bad". There are always caveats to add to each judgement, ranges of different types of picture content to consider, and other variables.

But, in spite of the complexity, we cannot avoid the generality of specifying picture quality. We have to grasp the nettle, and describe quality in a straightforward, intelligible and useful way. This is important in many areas, and particularly in understanding the way in which digital compression systems perform, and their effectiveness.

Figure 2
A typical "Lomo Compact" shot taken by the author: over-saturated colours, too much contrast, and vignetting – but the picture is vibrant

This is a critical time for digital video compression. The MPEG-2 system has served us well since the mid 1990s. The subsequent-generation system, MPEG-4 Part 4, included not just more advanced compression tools but also a potential new way of delivering multimedia with object and semantic coding. However, the part which has been most used is the one dealing with new compression tools. Another new generation is arriving with MPEG-4 Part 10 (H.264). Finally, Microsoft has entered the arena with a video codec in their Windows Media Series 9 offering, which they see as useful to both the broadcast and Internet worlds.

The objective of this article is to examine how we can evaluate these codecs, and how we can take strategic decisions about them. There is much more to say about this, than given here – but hopefully it will be food for thought.

The article does not explain the inner workings of these new codecs – there are excellent articles and papers from the codec developers that do that. Rather, the intention is to look down from 10,000 metres at the environment, to see where we are, where we are going and what can help us decide on a system.

Quality evaluation curves

When subjective quality evaluations are made following normal procedures, we arrive at curves which give the relationship between the average conception of what constitutes good or less-good picture quality – and a key variable, usually the bitrate. The type of curves that we use is illustrated in *Fig. 3*, taken from recent EBU studies (BPN 055). We might typically evaluate several codecs and arrive at a family of curves.

By comparing the qualities achieved at a given bitrate, or the bitrates needed to achieve a given quality, we can establish the difference in "quality delivery effectiveness" of the codecs, and this helps us to make a strategic judgement about if, and when, each type of compression might be used. Such curves form a basis for decision making in groups such as the EBU

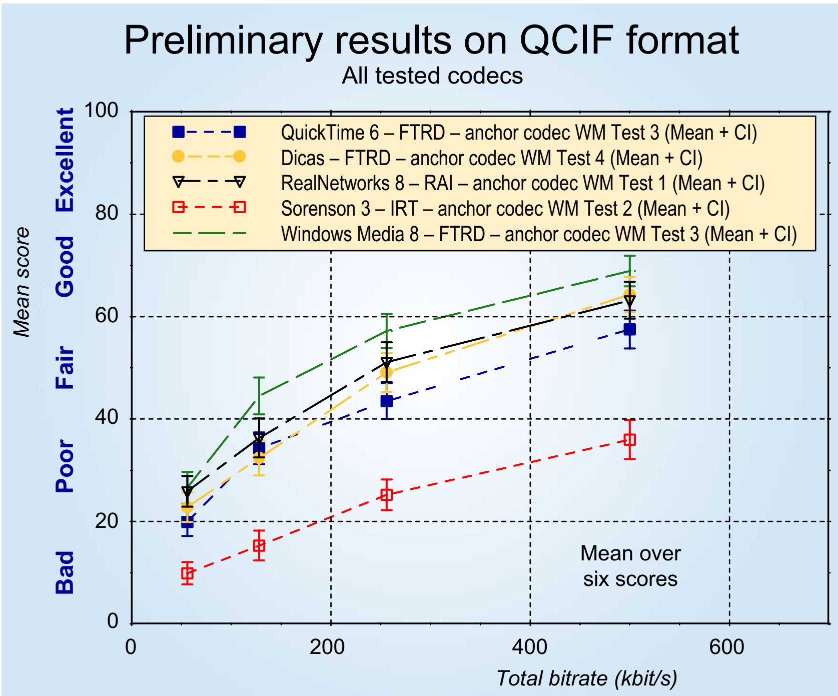

Figure 3
Example of the way subjective quality results are presented, taken from a recent EBU study of performance of codecs for web content

and ISO/IEC JTC1 MPEG. Although these simple curves of quality versus bitrate do provide useful information, behind them is a complex world. Each curve is actually just one line from a statistical distribution of lines which reflect codec performance with content of different "criticality", weighted by its "frequency of occurrence". In short, these curves are an approximation of reality.

The shape and location of these curves is nevertheless an important indicator of how the compression system performs. If the curves are parallel, this means the effectiveness difference is independent of the quality level. If they are not, it means the reverse is true. Usually, quality-versus-bitrate curves converge as the absolute quality rises: the differences between compression systems usually become less marked at higher qualities. In fact, when we do find a system whose curves remains parallel over the whole quality range, it will be time to open the champagne.

Using the curves of quality versus bitrate alone is by no means all we need to do when making strategic decisions about choosing a codec. What we really need to know is the relationship between "quality delivery effectiveness" and time, in the sense of months or years; i.e. we need to know how the performance of a particular codec might change and compare with other codecs – over the period of time when we are likely to be using that codec, and amortizing our investment in it. We need to estimate how long a particular codec is going to be the cheapest or best in its class. We have to ask ourselves: *when will something better arise, and better by how much?* We also need to consider the cost evolution of codecs.

These seem to be impossible questions to answer. They seem to require us to know beforehand when genius will strike the inventive minds of those working in the compression technology standardization groups. However, the situation is not quite so bleak. There are trends that we can spot. And there are predictions we can reasonably make.

Tendencies and predictions

To give us confidence, we may note how, in another area of technology, the so-called "Moore's law" has made it possible to plan strategically in many situations involving integrated circuits. Moore's law states that the complexity possible on a given IC size will double every 18 months. Gordon Moore's law has stood the test of time for several decades. It is based on a simple premise. Look at what has happened up until now and don't imagine that suddenly, and out of the blue, a trend will stop – it just doesn't happen like that.

So, is there something equivalent to Moore's law that is applicable to digital compression? Evidence suggests that there is.

Compression systems are collections of compression "tools" which are assembled together. Modern compression systems for mass media delivery are designed with asymmetric complexity. It is normal to load minimum complexity into the decoder and maximum complexity into the encoder. There are millions of decoders and only a few encoders, so this makes sense. In fact, normally the encoder doesn't need to be specified, just the decoder. In this case, all the encoder manufacturer has to do is to devise a box which will deliver a decodable signal. He can do whatever he wants inside the encoder box, as long as it gives a standard decoder-readable signal. So, over time, the overall performance can improve as manufacturers compete to make ever more "clever" encoders. This is quite possible. The characteristics of the content can be examined and the picture can be processed to give the most readily compressible signals.

But which collection of tools should make up a de-compression system? The choice is based firstly on practical considerations – what complexity does Moore's law currently allow in receivers? This is a moving target, but assumptions can be made about reasonable costs of receivers at the current time or in the near future. The choice is also based on what has been devised by the laboratories. If something hasn't been devised yet, you can't use it!

Mechanisms that influence codec quality

So, overall, we see two mechanisms influencing codec development and use. These are; firstly, the pattern of quality improvement which occurs after a set of compression tools has been agreed to be the best set for the moment. The second is the point later on, when it seems reasonable – because greater IC complexity is possible and knowledge has evolved – to create a new set of tools, usually adding to the last set, to create a new codec system. If you use all the old tools in the new set, you can arrange for pictures compressed under the old scheme to be decoded by the new decoder – and indeed this principle of backward compatibility is used in many of the MPEG systems.

One way to look at the evolution – both in terms of the quality effectiveness improvements that occur within a given set of tools, and with the assembling of new extended sets of tools – is that the process of compression becomes ever more "content adaptive". The compression system is able to adapt itself ever more intelligently to the type of content in the scene. We move from the "systematic" to the "adaptive".

As an illustration, consider the process of interlacing – which was the world's first video compression tool. When interlacing is applied, every other line is omitted in a two-frame cycle. This means – if you care to do the maths – that the high vertical temporal information content of the scene is dropped, and we benefit by halving the bandwidth of the signal. This is very effective, because the high vertical temporal information is the least noticeable part of the picture information if there is no vertically moving detail. However, the process is applied to every picture – whatever the scene is. It is thus a "systematic" compression tool.

As we move forward in knowledge about compression, we find ways of compressing information, not systematically, but based on what is contained in each scene. Interlacing is fine if the picture is static, and if there is no moving detail. If there is moving detail, it is blurred. Wouldn't it be good if we could drop the interlacing every time there is moving detail? Wouldn't it also be good if we could change the compression system depending on what content is in the scene? It is these aspects of compression systems that are getting better and better – the onward march of codec technology.

Incidentally, although interlacing was exactly right for the analogue age, it is a liability in the digital age, because we can do better with an adaptive digital compression system. But maybe in the next generation of broadcast systems we will indeed move to progressively-scanned production systems, which will be more "quality delivery effective" when compressed.

The 'Russian Steppes' of codec quality

If we put these elements together, and look at the history of codec development, we find a series of curves that are similar to those shown in *Fig. 4*. We have not included MPEG-1 in this figure, or even earlier codecs, for reasons of space. We know the pattern of quality development of the MPEG-2 codec. Substantial improvements in quality efficiency have been made since the tools were originally assembled. Indeed, improvements continue to be made even today.

A pattern of development cycles occurs, which result in longterm continuous gains in efficiency.

We have assumed here that we are only interested in one particular quality level. This might be, say, grade 4.5 pictures for professional-quality broadcasting. All the data of interest relates to the bitrate that you need in order to deliver this particular quality, using material of a given criticality (often taken to be scenes which are "critical but not unduly so").

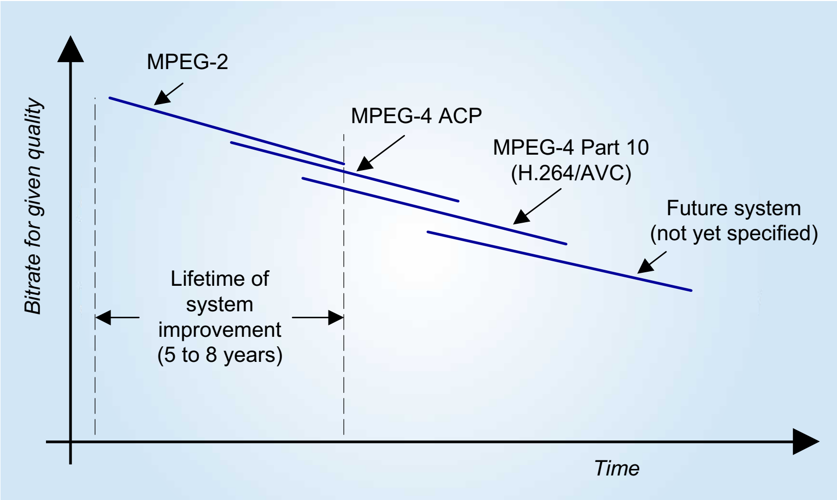

Figure 4
The evolution of open standard video compression systems

On this diagram are shown the curves for MPEG-2 and the MPEG-4 Part 4. We have also added a third curve for the more recent MPEG 4 Part 10 (or H.264) standard and a fourth curve for a future set of compression tools.

In practice these will be true curves and not just straight lines – and don't forget that there will be different curves for different quality levels. But to get to grips with the subject, we can start with a simplified picture. The quality effectiveness of a codec improves (less bitrate is needed for a given quality level) over a period of years after the tools have been standardized. There are then step changes when new codecs are developed. The quality effectiveness of one set probably overlaps with the next, but each new set of tools improves the quality effectiveness more than the last one can.

What we see is that a pattern of short-term development cycles occurs, which result in longterm continuous gains in quality efficiency, for the world to benefit from.

The author enjoys taking risks, and has examined the information available from quality evaluations of compression systems. Putting his neck on the block, his estimate is that the internal system improvement cycle is about 5 - 8 years. Furthermore, on average, the longterm gain in quality efficiency for SDTV (standard-definition TV) is about 5 - 10% per year. These are not laws of physics of course, and more research would yield more accurate values. The main thing is not the absolute values; it is an appreciation of what is happening and what should be factored into any thinking about codecs. For Windows Media Player 9 (WM9), we could suggest – based on the work done and reported in the EBU Information document I35 – that it may lie somewhere between the curves for MPEG-4 part 4 and MPEG-4 part 10, but its position would depend on the absolute quality level considered and how the internal elements were set up in the WM9 codec.

A key point to grasp is that there is no reason why these cycles, or the internal improvements within a set of tools, should stop. We can continue to expect new sets of tools to be developed over time, with higher IC complexities. Indeed, a set of tools using *wavelets* instead of DCT (Discrete Cosine Transform) techniques will probably be the next set after H.264. Broadcasters need to consider how these improvement cycles might affect their choice of codecs.

Improvements in codec performance

New sets of improved tools are assembled when standards are agreed, but what is the magic that allows the improvements? Digital video compression systems all have a common structure. Basically, they extract redundant information from the picture – in the sense that the picture can be reconstructed in the receiver without this redundant information having to be explicitly sent. Next, the system makes approximations of the signal, where needed, to reduce the bitrate beyond that possible with the redundancy alone and, finally, the system finds the most efficient way to send this data.

The compression process occurs in three consecutive steps:

1) motion compensation;
2) transform coding;
3) statistical coding.

The first stage of compression is motion compensation. Firstly, the system finds out if any parts of the picture have occurred before, and if they have, we send information on where they occurred rather than the parts themselves. After this is done, we pass what is left to the next stage of compression – the transform coding. In this stage, we convert the signal from the time domain (the real world) to the frequency domain (we express the signal as a group of frequency components). Then we drop the frequency components of low value (because these are the least noticeable) and pass what is left to the next stage – the statistical coding. Here we examine the digital words arriving over a period of time, and recode the ones that appear most often as the shortest words. This being done, we pass the signal out to the world.

So, how can we make these three stages more effective – by using ever more processing power? For the first stage, motion compensation, we could use ever larger search areas to find where the same part of the picture has occurred before. For the second stage, we can divide the picture into blocks which are smaller, or adapted in size according to what is in them. For the third stage, we can increase the number of signals examined, and use more sophisticated ways to match word length to frequency of occurrence. It's not really that simple, but at least these are the basic ways of improving codec performance.

So for how long can improvements continue to be made? Clearly, the bitrate needed to convey a high-quality video signal cannot be reduced to zero. Furthermore, there is no such thing as an entirely "free lunch" in the world of compression. Although applying more compression can increase the average quality effectiveness, further compression can mean that if and when the codec fails (in the sense that the scene content is impaired), the failure can be more dramatic – it may be a case of fewer failures, but more dramatic ones.

Also, the more compression that is applied, the less "headroom" there is in the signal. If you use up all the redundancy, there is more risk that passing a second time through a codec or other picture-processing system will cause visible impairments. Compression systems used in an environment where there is going to be more processing have to be lighter than those where there will be none. Compression systems used in an environment where the audience is going to be annoyed by the slightest failure also have to be lighter than when there is only a normally-attentive audience. These considerations make choosing a system more difficult, but they are the facts of life.

Having said that, there are probably at least two cycles to come after MPEG-4 Part 10. We should be able to look forward to at least the next ten years bringing longterm improvements in quality delivery effectiveness.

Quantitative approximations in the comparisons of codec performance

It is always possible to find manifold reasons *not* to draw quantitative conclusions about codec performance – but this would be of no help in deciding the strategy. Instead, we can hope to draw reasonable conclusions if we accept and understand the hidden elements – in particular that the results are dependent on the absolute quality level and on the content. It matters what quality level you are talking about, and what is in the scene being viewed. These can change the results.

For the purpose of obtaining a general understanding of the differences between video compression systems, the author believes that existing information – though less than ideal for drawing these kinds of conclusions – can lead to reasonable rules of thumb for material that is not unduly critical.

The available results of quality evaluations in two areas have been examined by the author:

- The first is the relationship between MPEG-2 (the world's most successful codec) and the subsequent-generation system MPEG-4 Part 4 (also known as MPEG-4 Visual).
- The second is the relationship between the quality achievable with MPEG-2 and the new system H.264 (i.e. MPEG-4 Part 10).

Both MPEG-4 Part 4 and MPEG-4 Part 10 are considerably more complex than MPEG-2 in terms of the processing needed in the receiver, but this is to be expected and is "permissible" because of Moore's Law.

The relationships between these three systems, based on the author's observations, are shown in the accompanying text box – for several important quality levels.

MPEG-4 Part 4 is considered at three quality levels:

- SDTV – standard-definition television (i.e. PAL/SECAM quality);
- CIF – Common Interchange Format, which has the resolution of a quarter SDTV picture and is used for broadband Internet delivery in some cases;
- QCIF – which is a quarter CIF. This is also used for the delivery of web content video.

MPEG-4 Part 10 is considered at the HDTV level (roughly four times SDTV) and at the SDTV and CIF levels.

How does MPEG-4 Part 4 compare with MPEG-2?

- 15 - 20% better at SDTV
- 20 - 30% better at CIF
- 30 - 50% better at QCIF

How will MPEG-4 Part 10 compare with MPEG-2?

- 20 - 40% better at HDTV
- 40 - 50% better at SDTV
- 50 - 60% better at CIF

The percentages shown in the text box are the reductions in bitrate that would provide the same picture quality.

These observations suggest that, at higher picture-quality levels (HDTV and SDTV), the percentage gains over MPEG-2 are less than at the CIF and QCIF levels. The gains when using MPEG-4 Part 10 are substantial at the lower picture-quality levels (50 - 60% at the CIF level) and diminish at HDTV levels (20 - 40%). But would such gains be sufficient to justify a change of compression system? And would such gains justify using one of the post MPEG-2 systems if starting a new service from scratch?

Licensing cost

A further factor to consider carefully is the cost of compression systems. New systems often have high initial equipment costs, as the research and development spend is amortized. But there is more to choosing a compression systems than looking purely at the hardware costs, and the manufacturers mark-up. Licensing costs must also be considered. All standardization bodies offer specifications which are licensed on "fair, reasonable and non-discriminatory" terms. It sounds excellent, but no one is sure exactly what it means – especially the word "reasonable".

MPEG-2 licensing is based on a ***charge per receiver*** – actually about 2.5 USD. This is a system which is easy for everyone to work with – they know up-front what the costs are likely to be. But, the new licensing regime being planned for MPEG-4 Part 4 is different. The proposal here is that, for services which involve the user paying something – or for services to mobiles – the licensing is ***charged per hour of use***. It is no secret that the world's broadcasters are anxious about charges per use, and they see this as a deterrent to using such codecs. The regime to be adopted for MPEG-4 Part 10 is not yet decided, but Microsoft has announced that they will not charge per use with Windows Media Player 9.

One of the reasons for the wish to change licensing agreements to one based on use – rather than on a receiver levy – may, ironically, be linked to the very success of worldwide standards bodies such as ISO/IEC JTC1 MPEG. In times past, when a particular company held the patents on a system, they could charge royalties on receivers from other manufacturers, and this would provide an ongoing income. In times present, when multiple manufacturers hold the patents on an open-standard system, and these same companies are making receivers, then paying licence fees on a receiver may amount to paying themselves for the right to use their own system. Their preference would naturally be to get someone else to pay royalties, and this has to be the broadcaster – with money from those paying to view the service.

The question of licence fees, based on either usage or on the receiver, may be linked to who holds the patents and what core business they are in. This may be an important issue in the years ahead – only time will tell.

Conclusions

What initial conclusions on codec development can we draw?

- Choose the compression system as close as you can to the date of service. This is not the first thing to do – it is the last thing to do. That way you will get the highest quality effectiveness, and the lowest overall costs. Make the business decisions first, before the final technical decisions.
- Note that there is no law of physics that says improvements in codec effectiveness will stop – they will not.
- Looked at globally, there are some signs that the MPEG-4 Part 4 system may have been, or is being, overtaken by the technology of MPEG-4 Part 10 (H.264). There may be lessons here about when to adopt new technology.
- Costs matter as well as technical quality. MPEG-4 Part 4 may be hampered by the current proposals for charge-per-use licensing – and MPEG-4 Part 10 similarly, if the same philosophy is applied.

David Wood is Head of New Technology at the EBU Headquarters in Geneva. He is a graduate of the Electronics Department at Southampton University in the UK and the UNIIRT in Ukraine. He worked for the BBC and the former IBA in the UK, before joining the EBU.

Within the EBU, David Wood is currently Secretary of the Digital Strategy group, the On-line Services group and the Television Quality Evolution group.

Acknowledgements

Discussions and papers from many sources were needed to prepare this article. Particular thanks are due to Ken McCann, who leads the DVB-AVC group, for his insight.

Incidentally, other fans of the Lomo Compact camera – apart from the author – are said to include President Putin of Russia.

References

[1] EBU I34-2002: **The potential impact of Flat Panel displays on broadcast delivery of television**.

[2] EBU I35-2003: **Further considerations on the impact of Flat Panel home displays on the broadcasting chain**.

[3] BPN 055: **Subjective viewing evaluations of some commercial internet video codecs – Phase 1, May 2003**.

These documents can be obtained from **Mrs. Lina Vanberghem** at EBU Headquarters.

21 July 2003