Codecs

Adrian Mackenzie

Codecs (coder-decoders) perform encoding and decoding on a data stream or signal, usually in the interest of compressing video, speech, or music. They scale, reorder, decompose, and reconstitute perceptible images and sounds so that they can get through information networks and electronic media. Codecs are intimately associated with changes in the "spectral density," the distribution of energy, radiated by sound and image in electronic media.

Software such as codecs poses several analytical problems. Firstly, they are monstrously complicated. Methodologically speaking, coming to grips with them as technical processes may entail long excursions into labryinths of mathematical formalism and machine architecture, and then finding ways of backing out of them bringing the most relevant features. In relation to video codecs, this probably means making sense of how transform compression and motion estimation work together. Second, at a phenomenological level, they deeply influence the very texture, flow, and materiality of sounds and images. Yet the processes and parameters at work in codecs are quite counterintuitive. Originating in problems of audiovisual perception, codecs actually lie quite a long way away from commonsense understandings of perception. Third, from the perspective of political economy, codecs structure contemporary media economies and cultures in important ways. This may come to light occasionally, usually in the form of an error message saying that something is missing: the right codec has not been installed and the file cannot be played. Despite or perhaps because of their convoluted obscurity, codecs catalyze new relations between people, things, spaces, and times in events and forms.

Patent Pools and Codec Floods

Video codecs such as MPEG-1, MPEG-2, MPEG-4, H.261, H.263, the important H.264, theora, dirac, DivX, XviD, MJPEG, WMV, RealVideo, etc., are strewn across networked electronic media. Roughly a hundred different audio and video codecs are currently in use, some in multiple implementations. Because codecs often borrow techniques and strategies of processing sound and image, they have tangled geneologies.

Leaving aside the snarled relations between different codecs and video technologies, even one codec, the well-established and uncontentious MPEG-2 coding standard, is extraordinarily complex in its treatment of images. MPEG-2 (a.k.a. H.262) designates a well-established set of encoding and decoding procedures for digital video formalized as a standard.[1] The standards for MPEG-2 are widely described. Many diagrams, definitions, and explanations of coding and decoding the bitstream are available in print and online.[2] Open source software implementations of the MPEG-2 standard offer a concrete path into its implementation. For instance, ffmpeg, "is a complete solution to record, convert and stream audio and video."[3] It handles many different video and audio codecs, and is widely used by many other video and audio projects (VLC, mplayer, etc.).

Economically, MPEG-2 is a mosaic of intellectual property claims (640 patents held by entertainment, telecommunications, government, academic, and military owners according to Wikipedia.[4] The large patent pool attests to the economic significance of MPEG-2 codecs. As the basis of commercial DVDs, the transmission format for satellite and cable digital television (DVB and ATSC), as the platform for HDTV as well as the foundation for many internet streaming formats such as RealMedia and Windows Media, MPEG-2 forms a primary technical component of contemporary audiovisual culture. It participates in geopolitical codec wars (e.g., China's AVC codec, versus the increasingly popular H.264, versus other versions such as Microsoft Windows VC-1—Windows Media 9).

Many salient events in the development of information and digital cultures (for instance, MP3-based file-swapping, or JPEG-based photography) derive from the same technological lineage as MPEG-2 (lossy compression using transforms). At a perceptual level, what appears on screen is colored by the techniques of "lossy compression" that MPEG-2 epitomizes. Codecs affect at a deep level contemporary sensations of movement, color, light, and time.

Trading Space and Time in Transforms

The MPEG standard is complex. Digital signal processing textbooks caution against trying to program it at home (which immediately suggests the desirability of doing so). They suggest buying someone else's implementation of the standard.[5] Where does this complexity come from? The purpose of the MPEG-2 standard developed in the early 1990s is generic:

This part of this specification was developed in response to the growing need for a generic coding method of moving pictures and of associated sound for various applications such as digital storage media, television broadcasting and communication. The use of this specification means that motion video can be manipulated as a form of computer data.[6]

How does a "generic coding method" end up being so complex that "it is one of the most complicated algorithms in DSP [digital signal processing]"?[7] MPEG-2 defines a bitstream that tries to reconcile the complicated psychophysical, technocultural, and political-economic processes of seeing. MPEG-2 puts more pictures, more often, in more places. It moves images further and faster in media networks than they would otherwise.

To do that, the code in MPEG-2 codecs reorganizes images at many scales. The code works to reorganize relations within and between images. Algorithmically, MPEG-2 combines several distinct compression techniques (converting signals from time domain to frequency domain using discrete cosine transforms, quantization, Huffman and Run Length Encoding, block motion compensation), timing and multiplexing mechanisms, retrieval and sequencing techniques, many of which are borrowed from the earlier, low-bitrate standard, MPEG-1.[8]

From the standpoint of software studies, how can these different algorithms be discussed without assuming a technical background knowledge? The technical intricacies of these compression techniques are rarely discussed outside signal processing textbooks and research literature. Yet these techniques deeply affect the life of images and media today. One strategy is to begin by describing the most distinctive algorithmic processes present, and then ask to what constraints or problems these processes respond. From there we can start to explore how software transforms relations.

For instance, we could concentrate on what happens at the lowest levels of the picture, the "block" (8×8 pixels). Digital video typically arrives at the codec as a series of frames (from a camera, from a film or television source). Each frame or static digital image comprises arrays of pixels defined by color (chrominance) and brightness (luminance) values. Each frame then undergoes several phases of cutting and reassembling. These phases probe and re-structure the image quite deeply, almost to the pixel level. Digital video pictures are composed of arrays of pixels that have much spatial redundancy. Many adjacent pixels in an image of a landscape will be very similar, and it wastes stor-

age space (on a DVD) or bandwidth (on satellite transmitters or internet) to repeat the same pixel over and over. A sky could be mostly blue. Rather than transmit an exact replica of the sky, why not use an algorithmic process that transforms the blue sky into a quasi-statistical summary of the spatial distribution of blueness?

The so-called I-Picture or Intra-Picture is the product of one phase of encoding, *transform compression*. It is applied to selected frames. The I-Pictures effectively become key-frames in the MPEG videostream. This phase relies on spectral analysis carried out using Fourier transforms. What does spectral analysis do? Broadly speaking, it breaks a complex waveform into a set of component waveforms of different amplitude or energy. Many computational processes today rely on Fourier Transforms or on a particular variant of the Fourier Transform, the Discrete Cosine Transform (DCT). The DCT, implemented in silicon or C code, encodes complex signals that vary over time or space into a series of discrete component frequencies. They can be added together to reconstitute the original signal during decoding. Nearly all video codecs transform spatially extended images into sets of simple frequencies. This allows them to isolate those components of an image that are most perceptually salient to human eyes. These would include the brightest or most colorful components.

There is something quite counter-intuitive in transform compression applied to images. In what way can a videoframe be seen as a waveform? The notion of the transform is mathematical: It is a function that takes an arbitrary waveform and expresses it as a series of simple sine waves of different frequencies and amplitudes. Added together, these sine or cosine waves reconstitute the original signal. Practically, in encoding a given frame of video, the MPEG-2 code divides the 720×576 pixel DVD image into 8×8 pixel blocks. So application of the transform compression is not general or global. The image has been turned into in an array of small blocks that can be quickly transformed separately. This can be seen by freeze-framing a complex visual scene on a DVD. It will appear "blocky." The DCT sees each of these blocks as spatial distribution of brightness and color. It delivers a series of coefficients (or multiplicative factors) of different frequency cosine waves.

The decomposition of a spatial or temporal signal into a series of different frequency components allows correlation with the neurophysiological measurements of human hearing and sight. For instance, because the transform treats blocks as spectra of values, some of which are more significant to human eyes than others, it converts the spectrum values into a sequence in which the most

important come first. Components of the series that have small coefficients can be discarded because they will not be visually salient. In this way, a block can be compressed, transmitted or stored, and decompressed without ever sending any information about individual pixels. The cosine wave coefficients represent amplitudes of different frequency cosine waves. When the block is decoded (for instance, during display of a video frame on screen), the coefficients are reattached to corresponding cosine waves, and these are summed together to reconstitute arrays of color and brightness values comprising the block.

What stands out in transform compression is decomposition of the framed images through densely complex matrix manipulations occurring on the thousands of blocks. In contrast to film's use of linear sequences of whole frames, or television and video's interlacing of scan-lines to compose images, transforms such as DCT deal with grids of blocks in highly counterintuitive spectral analysis that has little to do with space. Blocks themselves are not fragments of pictures, but rather distributions of luminosity and chrominance that are packed into the bit stream.

Motion Prediction—Forward and Backward in Time

What does it mean to say that codecs catalyze new relations between people, things, spaces, and times in events and forms? Software has long been understood as closely linked to ideation or thought, particularly mathematical thought. Despite the mathematical character of the DCT compression just discussed, the thinking present in software cannot be reduced to mathematical thought, or not to mathematical thought as it is usually conceived. Codecs perhaps challenge cinematic and televisual perception even as they participate in making the world more cinematic or televisual. They deviate radically from the normal cinematic or televisual production of frames in a linear sequence. Video codecs are very preoccupied with reordering relations between frames rather than just keeping a series of frames in order. Indeed just as frames themselves are individually reconfigured as blocks of luminance and chrominance, the relation between frames is subject to calculated reordering in the interests of accelerated or compressed transport.

In order to gain purchase on the relation between frames, the MPEG codec again breaks the frame into an array of discrete "macroblocks" (usually four blocks put together). It compares successive frames to see how a specific macroblock shifts between frames. The working assumption behind the

motion-predicted encoding of video in MPEG-2 is that nothing much happens between successive frames that can't be understood as macroblocks undergoing geometric manipulations (translation, rotation, skewing, etc.). The fact that nothing much happens between frames apart from spatial transformation is the basis of the interframe compression and the generation of P and B pictures (forward and backward motion prediction, respectively). P (Predicted) and B (Backward) pictures, the pictures that accompany the I-Picture in a MPEG-2 bitstream are, therefore, really nothing like film frames. There will never be a flicker in an MPEG video because the boundaries between pictures are not constructed in the same way they are in film or even in television with its interlaced scanned images.

If intrapicture compression is the first major component of MPEG-2, motion prediction between frames is the second. Interpicture motion prediction compression relies on forward and backward correlations, and in particular on the calculation of motion vectors for blocks. In the process of encoding a video sequence, the MPEG-2 codec analyzes for each picture how blocks have moved, and only transmits lists of motion vectors describing the movement of blocks in relation to a reference picture or keyframe, itself coded using DCT transform compression. This fundamentally alters the framing of images. We have already seen that rather than the raw pixel being the elementary material of the image, the block becomes the elementary component. Here the picture itself is no longer the elementary component of the sequence, but an object to be analyzed in terms of sets of motion vectors describing relative movements of blocks and then discarded. The P and I pictures, after encoding, are nothing but a series of vectors describing how and where macroblocks move. Decoding the MPEG stream means turning these vectors back into arrangements of blocks animated across frames.

Motion prediction takes time to work out, but heavily compresses the videostream. Transform compression is fast to calculate, but yields quite a large amount of data. Hence, the actual ratio of intraframe and interframe pictures in a given bitstream is heavily weighted toward motion prediction. In an MPEG datastream, the precise mixture of different frame-types (I, P-forward, and B-backward) is defined at encoding time in the Group of Pictures (GOP) structure. It is usually 12 or 15 frames in a sequence such as I_BB_P_BB_P_BB_P_BB_P_BB_. One intracoded frame is followed by a dozen or so block motion-compensation frames. The combination of forward-prediction and backward-prediction found in the GOP means that the MPEG bitstream

effectively treats the video stream as a massive doubly linked list.[9] Each item in the list is itself a list describing where and how (rotated, translated, skewed) each block should be placed on screen.

The ratio of different frame types to each other affects the encoding time because motion compensation is much slower to encode than the highly optimized block transforms. Codecs must make direct tradeoffs between computational time and space. The tradeoffs sometimes result in artifacts visible on screen as, for example, blocking and mosaic effects. At times, motion prediction does not work. A change in camera shot, the effect of an edit, might mean that no blocks are shared between adjacent frames. In that case, a well-designed codec falls back on intraframe encoding.

From Complicated to Composite

Many of the complications and counterintuitive orderings of the MPEG-2 codecs arise because they try to negotiate a fit between network bandwidth constraints (a commercially marketed service), viewing conventions (the rectangular frame of cinema and television), embodied perception (sensations of motion, light, and color), and cultural forms (fast-moving images or action). They respond to the economic and technical need to reduce the bandwidth required to circulate high-resolution digital pictures and sounds. As a convention, the MPEG-2 standard refers implicitly to a great number of material entities ranging from screen dimensions through network and transmission infrastructures to semiconductor and data storage technologies. The generic method of encoding and decoding images for transmission relates very closely to the constraints and conditions of telecommunications and media networks. And the codec more or less performs the function of displaying light, color, and sound on screen within calibrated psycho-perceptual parameters.

However, the way the MPEG-2 codec pulls apart and reorganizes moving images goes further than simply transporting images. Transform compression and motion estimation profoundly alter the materiality of images, all the while preserving much of their familiar cinematic or televisual appearance. Like so much software it institutes a relational ordering that articulates realities together that previously lay further apart.

Notes

1. ISO/IEC 13818-1, I. I. (1995). "Information technology—Generic coding of moving pictures and associated audio information: Systems." ISO/IEC 13818-2 (1995). "Information technology—Generic coding of moving pictures and associated audio information: Video."

2. S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, p. 225; Wikipedia, 2006, MPEG-2, available at http://en.wikipedia.org/wiki/MPEG-2/ (accessed Jan. 12, 2006).

3. ffmpeg, *FFMPEG Multimedia System, 2006.*

4. Wikipedia, "MPEG-2," 2006.

5. S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing,* p. 225.

6. ISO/IEC 13818-2 (1995) (E), vi.

7. S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing,* 225.

8. ISO/IEC 11172-1 (1993).

9. Donald Knuth, *The Art of Computer Programming*, p. 280.

Computing Power

Ron Eglash

Computational power plays an accelerating role in many powerful social locations. Simulation models, for example, sneak into our medical decisions, speak loudly in the global warming debate, invisibly determine the rates we pay for insurance, locate the position of a new bridge in our city, plot the course of our nation's wars, and testify in the courtroom both for and against the defense. Other applications in which computing power matters are molecular biology, communication surveillance, and nanotechnology. Social scientists concerned with the relations of power and society commonly examine who has money, who owns property, and who owns the means of production. But the ownership