



Post-modern video

by A. Lippman

Digital video is no longer a special-purpose curiosity, but is an integral component of all emerging computer-mediated communications systems. Moving pictures, conferencing, entertainment systems, movie production, and video sequencing are some of the areas that have been influenced. Advances in digital representations of moving pictures are already responsible for important changes in our communications systems. Novel conferencing and entertainment systems made possible by compression are infecting home and work environments. Now, inroads are being carved into theatrical movie production, and new analysis techniques are making digital picture archives realistic and useful. Video sequence understanding has suddenly become a vital research domain, rivaling restoration and compression as a locus of interest. Applications in digital video range from digital libraries to surveillance and home television.

The emergence of digital video

Digital video and its associated delivery channels have always been envisioned by many as a supporting leg for any national or global information infrastructure (NII or GII), but its actual role has been either obscure or in doubt. The first compression standards (H.261 and MPEG, Moving Pictures Experts Group) predate any focused effort on an NII or GII and were motivated by specific applications. H.261 (1988–1990) was intended for teleconferencing with limited motion, and MPEG-1 (1990–1991) was intended for low-resolution video sequence. But these first standards have smoothly migrated from segregated applications to systematic components of universal digital communications systems. While not nearly perfect, compressed digital video works both for computer network communications and for consumer appliances. A single, extensible family of standards serves these disparate needs.

At the high end of the video quality spectrum, high-definition digital terrestrial broadcasting was first suggested in June, 1990, by General Instrument Corporation as a combination of compression with digital modulation. The initial picture representation was a hybrid coder, much like its predecessors, H.261 and MPEG, but the significant addition was the combination with a general-purpose, transmission-based digital delivery channel. The picture standard ultimately accepted by all North American proponents of High Definition Television (HDTV) systems became MPEG-2 (4–9 megabits per second, 3–6 times more than MPEG-1, and supporting the required interlaced video). When the United States government articulated a need for a national information infrastructure (Vice President Gore in 1992), the indigenous television industry advanced its image format as a basis for that infrastructure. Today, it is entirely possible that the imaging aspects of HDTV development may be eclipsed by the utility of the digital broadcast channel. There is more interest in bit radiation than in consumer picture resolution—thus any actual broadcast is likely to be a broad mix of sound, picture, and raw data.

In 1994, the focus of research and development changed from HDTV to video-on-demand (VoD). VoD may not prove to be cost-effective or even interesting for many years, but the development effort is significant because it requires a mature picture representation and a commitment to high bandwidth digital delivery channels. The coincidence of these elements implies two things: (1) that moving picture information can be easily integrated into computer interactions (multimedia), and (2) whatever infrastructure is built will have sufficient bandwidth for general picture and data use. The research and development emphasis is also broadened to include issues of archiving, picture understanding, interactive manipulation, and real-time retrieval.

©Copyright 1996 by International Business Machines Corporation.



By 1996, the “video-centric” view had developed the following main characteristics:

- Digital video is a consumer item, with inexpensive decoding hardware that will soon become implicit firmware.
- Development of HDTV has spawned both new picture formats and generalized terrestrial delivery systems.
- Video-on-demand has resulted in storage and networking solutions for high-speed, synchronous data.

Taken together, the foregoing characteristics suggest that now is the time to rethink the evolution of entertainment networks and the broader impact of imaging research in society. Television is slowly becoming an interactive, networked data type, and networks are increasingly becoming video-ready. Mass media broadcasting may be “running out of breath”—for the first time in history, the channels we use for the delivery of entertainment may not be technologically restricted to broadcast; and the data carried in those channels may not appear as the same program in every household, or even as a program at all.

While there is little doubt that more bits will enter most living rooms than will leave, the notion that those bits will simply be an ever-expanding set of fully packaged programs is outdated (or at least open to question). A more convenient view of the home communications interface is the sum of two networks, one fully symmetric and the other predominantly one way. This network will carry a combination of long lead time, scheduled broadcasts, mutually agreed upon multicasts, and demand-based data. Video, therefore, is more than television programs or conference images; it is a maturing part of consumer and professional communications. The papers in this section of the *IBM Systems Journal* explore moving picture processing with the previously described evolving scenario in mind.

The papers in this section

The Media Bank is an example of a digital library where the content is represented as a set of objects, stored in multiple formats, perhaps redundantly, and distributed among a diverse set of storage and processing servers. Both delivery and assembly are on demand, under the control of the recipient program, and are guided by objects that contain the assembly rules. This specific research, discussed by Lippman and Kermode, is targeted at building support for an entertainment infrastructure that is personalized and responsive to local community needs. Related work is occurring elsewhere and commercial systems for wide-area distribution of multimedia information are being proposed.

Video understanding is crucial to support an archive and perusal primitives. It can often be done asymmetrically, with far more energy applied to the analysis than is expended in retrieval. For example, a recognition system may function by reducing the data in a video sequence to a small number of significant parameters. Recognition can then be performed rapidly on the diminished data set. In movies, we may wish to create annotations that denote changes in the scene, that expose telling events in the story, or that pick out the presence of various actors. Such an annotated movie will be viewed more often than it is recorded, so analysis cost is spread over each screening. Likewise for libraries, the value of an archive is in its index. While original content can often be obtained via delayed electronic or physical delivery, simplifying the search is the role of the archivist. Picard explores this topic in the paper on video and image libraries.

Any infrastructure needs to address issues of security and intellectual property. With respect to images, this is an especially interesting task because images can be cut and pasted, torn and sheared, and segmented and reassembled into entirely new works where the train of ownership is lost. At what point is a picture of actor Gene Kelly, where the dancer has been changed into an animal and the street into a forest, no longer *Singing in the Rain*? Similarly, a recent magazine cover showed a composite picture comprised of over 1400 individual blocks, each of which was a minified image. Is this a new work or 1400 copyright lawsuits? Walter Bender et al. describe attempts to indelibly watermark and add data to such images, both as carriers of annotation and as ownership stamps.

An implicit picture model dictates the vocabulary of picture understanding and the repertoire of actions one may take to compress that sequence. Two-dimensional models incorporate various textural and motion assumptions and thereby simplify processing pictures that obey the model. Some models jointly serve the goal of efficient representation (compression) and understanding. Pyramids, for example, were first published as an efficient picture repre-



sentation and then investigated as a way to identify content at diverse scales. Hierarchical, block-based motion analysis is a direct by-product of this work. In his article in this section, Bove describes object-oriented references to multimedia explored in an environment where flexible hardware allows test and exportation of new representations and primitives.

A final view of video as an exploratory space is presented by Lucente: three-dimensional television. This is an ongoing topic of research in the laboratory and the exposition here addresses one critical component of that research: how it is possible to express an image that simultaneously supports multiple perspectives and viewpoints. Unlike stereoscopic imaging, where each viewer sees the same perspective, a holographic image re-creates the actual picture space. Lucente shows how a digitized version of that space can be efficiently maintained.

Taken as a package, the papers in this section all support video as an element of personalized, creative interactions. The vision is an infrastructure where image representations and understanding is sufficient to support practical, humanly stated operations and interactions. We think this is what the word “multimedia” should have meant.

Andrew Lippman *MIT Media Laboratory, 20 Ames Street, Cambridge, Massachusetts 02139-4307 (electronic mail: lip@media.mit.edu).* Dr. Lippman received both his B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology. In 1995 he completed his Ph.D. studies at the École Polytechnique Fédérale de Lausanne, Switzerland. He is currently associate director of the MIT Media Laboratory and a lecturer at MIT. He holds seven patents in television and digital image processing. His current research interests are in the design of flexible, interactive digital television infrastructure.

Reprint Order No. G321-5605.