Image Coding II

Visual Communications

Arun N. Netravali

AT&T Bell Laboratories
Murray Hill, NJ 07974

Abstract. Commercial analog television was standardized over fifty years ago, mainly for entertainment, sports, and news, using over the air broadcast. It is only recently that technology of compression, integrated circuits, fiber optic transmission, switching, and storage has advanced to make digital video economically feasible. This is forcing the convergence of diverse industries such as communications, computing, entertainment, and information services. Digitizing video in a cost effective manner and integrating it with computers, telecommunication networks, and consumer products will produce a vast array of new products and services in the future. These will have a long lasting impact on entertainment, education, medicine, and will improve our productivity in daily life.

1 Introduction

Commercial analog television was invented and standardized over fifty years ago mainly for over the air broadcast of entertainment, news, and sports. While some upward compatible changes have been made, such as color, multichannel sound, close captioning, and ghost cancellation, the underlying analog system has survived several generations of technological evolution that has pervaded all other media. After 50 years of refinement, we are poised for a major disruption -- *digital video.* Digitizing video in a cost effective manner and integrating it with computers, telecommunication networks, and consumer products, promises to produce a large array of new products and services. Visual communications is poised to become inexpensive, ubiquitous, flexible, and a necessity for improving productivity in our daily life.

A number of technological advances over the past few decades will enable the growth of visual communications. They are:

1. High quality, standard compression algorithms to reduce the bit rate of the video signal.
2. Inexpensive integrated circuits for processing and storage of video signals.
3. High speed communication networks using fiber optic transmission and wide band switches.
4. High density and fast access storage media (magnetic, optical, or semiconductor).

5. Computer architectures that can handle video-rate pixels with ease.
6. Algorithms and implementations that can rapidly create photo-realistic images by computers.

This technological capability, along with a proven value of visual communications for group discussions, education, entertainment, and medical applications will make visual communications ubiquitous. Consumers will be able to access video programs from multimedia databases populated by a variety of content providers. Video conferencing and video telephones will improve our productivity in business transactions. Prepackaged interactive programs delivered on CD-roms will help medical and educational applications. Most importantly, computers will become friendly by the use of visual interfaces. This, together with falling prices, will lead to consumerization of personal computers resulting in an even richer set of services.

The current state of video processing can be best described as "pixel based." The common paradigm is to scan a video signal, convert it into an array of pixels and process pixels without regard to the semantic meaning of any groups of pixels (spatially and/or temporally). While some progress will continue to be made in pixel based processing, the big frontier to be conquered is: content-based processing. In this, the video signal is described as a set of objects in three dimensions undergoing transformations (rigid or non-rigid) over time. The goal will be to recognize object boundaries and process, store, and transmit objects rather than pixel at a time.

The two current dominant paradigms are:
a. point-to-point video (i.e., video telephone)
b. broadcast video (i.e., terrestrial, satellite, or cable TV.) As visual communications becomes flexible and pervasive, there will be many more. Multipoint video teleconference, as well as interactive television, where a consumer is interacting with the content rather than simply changing channels or connections, will become a common place. Computing, storage, communication networks that enable interactive, distributed and collaborative processing of video signals will give rise to exciting new services.

2 Enabling Technologies

In this section we present an overview of some of the technologies that make digital video economically viable.

1. *Compression* - Analog television signal when digitized produces far too many bits to economically process, store, transmit, or switch. For example, the CCIR-601 television standard results in 216 Mb/sec bitstream for a 6-MHz or 8-MHz analog signal. Digital compression reduces this data rate to manageable sizes consistent with the applications. Standard compression algorithms (JPEG for intra-frame compression; MPEG for inter-frame compression) have been devised. CCIR-601 standard provides "studio quality" at 216 Mb/sec; JPEG is

roughly equivalent at 35-45 Mb/sec; MPEG is equivalent at slightly less than 10 Mb/sec. Similarly, the "VHS quality" requires approximately 1.5 Mb/sec using MPEG and slightly over 4 Mb/sec with JPEG. For video telephone applications, where the motion in scenes is limited (and acceptable quality is lower than VHS), these standards have achieved impressive results with compression ratios of about a thousand. (See Figure 1.)

In addition to producing the highest quality pictures at the lowest bit rate, compression algorithms must satisfy many other requirements. The compressed bits must be robust in the presence of transmission impairments. Abrupt changes, such as scene changes in the video material and channel changes introduced by the viewer, should not create transients that affect picture quality for a long time. If the compressed bit stream is stored in a digital VCR, many of the VCR functions that we are accustomed to TV. (e.g., fast forward - backward searches) should be easy to do. Editing of video and insertion of video within the compressed bitstream should be easy as well. Thus, as the use of the compressed video proliferates, many other requirements become important. While the cost of electronics continues to drop, compression algorithms must be manageably complex in order for them to get wide use. In applications such as digital broadcasting, digital cable, satellite, or interactive television, the cost of decompression done at the consumer end is the primary factor.

Figure 2 shows the approximate relationship between compression efficiency and the complexity of decompression. The compression efficiency is in terms of compressed bits per Nyquist sample, and therefore, pictures with different resolution and bandwidth can be scaled simply by proper multiplication to get the relevant bit rates. MPEG, which uses motion compensated transform coding, is about a thousand times more complex compared to PCM television. The relationship of cost to complexity depends upon the evolving technology. Very soon fast microprocessors will be able to decompress video signals entirely using only software.

Figure 3 shows the computational requirements for encoding and decoding of video at different resolutions. While encoding is still far too complex for even the most powerful processor of today, we are close to decoding done entirely in software.

As mentioned earlier, advances in computer vision and graphics are beginning to allow us to progress from pixel oriented coding to coding of objects in the scene. Techniques of computer vision can be used to identify a variety of objects from a complex scene. Each of these objects may contain hundreds of pixels and may be described by a modeling process available from computer graphics. In order to achieve a high degree of compression, proper models must be used. Since models that can handle arbitrary scenes are not well developed, a highly constrained application domain is usually considered. For example, in video phone applications, most scenes contain a head and shoulders view of a person in front of a static background. Head and shoulders views of people have been modeled as wire frames with some success. This type of work is still in its infancy and faces formidable challenges in segmentation of the scene into objects, building models for each segment, finding specific application domains, and dealing with the complexity of the encoder/decoder. A simpler approach is to segment an image into regions using techniques from computer vision and encode each region separately with coding strategy that is well matched to that region. Regions may not necessarily have semantic meaning and may not be rectangular blocks of pels as in block-transform coding. The principal difficulty of such an approach is the ability of the algorithm to segment images into areas similar statistics and perceptual characteristics. Extensive research is being carried out in model and segment based coding, but the promising results are yet to come.

2. *Multimedia Databases* - Once each of the mono media components of a multimedia signal (e.g., video, audio, data) are digitized, they need to be stored in a database server for a variety of services. Traditional databases store mostly alphanumeric characters as relations or tables. While much research and productization has been done to improve relational databases, they are not suited for storing multimedia signals. Most multimedia applications involving digital video require either simultaneous outputs to multiple subscribers or multiple outputs for a single subscriber. In either case very high bandwidth I/O is required. In addition, audio and video bits can be thought of as bitstreams, whereas computer data is usually bursty. Seamless integration of stream and bursty data presents a new challenge for multimedia databases. If digital video and its corresponding audio samples are stored separately within a database, then precise time synchronization of the video and audio outputs is required in order to create an effective presentation to the viewer. Additional functionality such as pause, restart, rewind, fast-forward, and fast-backward makes synchronization of audio and video even more difficult a challenge. The size of the video objects is usually several magnitudes higher than alphanumeric objects stored in the database. It appears that object oriented database technology in which objects of different types (e.g., video clips, audio clips, data,...) are stored directly is the potential solution. This is being developed in several laboratories and universities.

3. *Bridging* - In distributed and collaborative applications of multimedia services, it is often required to bridge multimedia signals. While bridging an analog audio signal is straight forward (simply add all the component audio signals) bridging video is not simple. In addition, bridging compressed bit streams without having to decompress, bridge, and then compress the bridged signal usually results in loss of quality and is expensive. In the case of video, a variety of options exist. Bridging multiple component video bit streams may be simply done by choosing one of the component video signals depending upon

voice activity. Another possibility is to display some customer-selected component video signals in different parts of the screen simultaneously. A more flexible way is to have different size windows displaying each of the component videos. In the case of data, bridging amounts to either multicasting to a select group of users or broadcasting to a large class of users. The trend in bridging is to allow users to control how each of the monomedia are bridged to create the most effective presentation to the viewer.

3 Applications

A large number of applications of visual communications technology already exist. For example, the concept of video phones has been around for several decades. However, the newer implementations of video phones are quite different from the earlier ones. Person to person video telephony can now be implemented by incorporating a special card in the back plane of a personal computer which may be attached either to a local area network or to a telephone line which can handle different bit rates. On the entertainment front, digital video is making hundreds of channels available on the current cable television systems as well as in the newer generation of high power satellites.. In this section we will describe two emerging applications enabled by digital video.

Interactive Television enables the consumer to exert both coarse and very fine grain control over the contents of the programming being viewed. By contrast, conventional television only allows a viewer to select among a number of programs being broadcast to a large audience. For the most part, each member of the audience for each of these standard programs sees identical contents. With an interactive television program, each viewer can directly control what he is watching. Interactive TV programs can be thought of as a collection of media elements together with software that controls the flow of the program in response to consumer inputs and directs how the media elements are used to create the aural or visual presentation. Media elements, such as audio and video clips, still images, graphics, text, etc., are the primitives of the presentation and are created as part of the process of producing the ITV application. The process and technology for creating ITV applications borrows heavily from computer graphics. Each simultaneous viewer of an interactive TV program can be thought of as totally independent. This requires a separate instance of an application executed by the ITV system for each viewer (although the media elements are shared for economy of storage). In addition, communication bandwidth must be allocated for each viewer to connect computing and storage facilities in the network to customer premises equipment. Overall, delivery system resources must be allocated for each active viewer -- a situation that more closely resembles the telephone network than a conventional broadcast or cable TV system. The development and evolution of interactive television on a large scale presents numerous technical challenges in different areas. However, most of the technical issues are sufficiently developed to allow trials to ascertain consumer interest in this new form of communication.

Distributed, Collaborative Virtual Space - The virtual workspace can be created by connecting several users with a flexible multimedia collaboration between them. Figure 4 shows a configuration for virtual meeting rooms. Virtual Meeting Room 1 involves user A and B as well as shared programs and data from server 1, whereas Virtual Meeting Room 2 involves a multimedia connection between user A and C. Control of these rooms allows users to participate from their own offices. Users may be people or programs. In addition, any of the users may share data that is generated by other users or that can be obtained from a database server. One user may participate in multiple rooms; each room conducting a conversation with a specific context. Interaction with the screen by one user may be seen by designated others. Thus, the attempt is to create a common visual space resembling communication that takes place when all the participants are in the same room.

Summary

In this short paper, we described some key video technology necessary for further evolution of visual communications. Falling prices of hardware and bandwidth, coupled with computer and telecommunication infrastructure will result in faster growth of visual communication than ever before.

References

1. A. N. Netravali, B. G. Haskell, *Digital Pictures: Representation and Compression, and Standards.* Plenum (Second Edition), 1994
2. M. Kunt (Editor), *Digital Television.* Special Issue Proceedings of IEEE, July, 1995

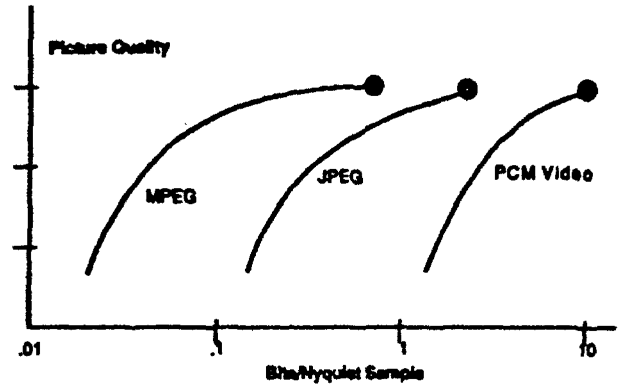

Fig. 1. A rough measure of the rate/quality of picture coders

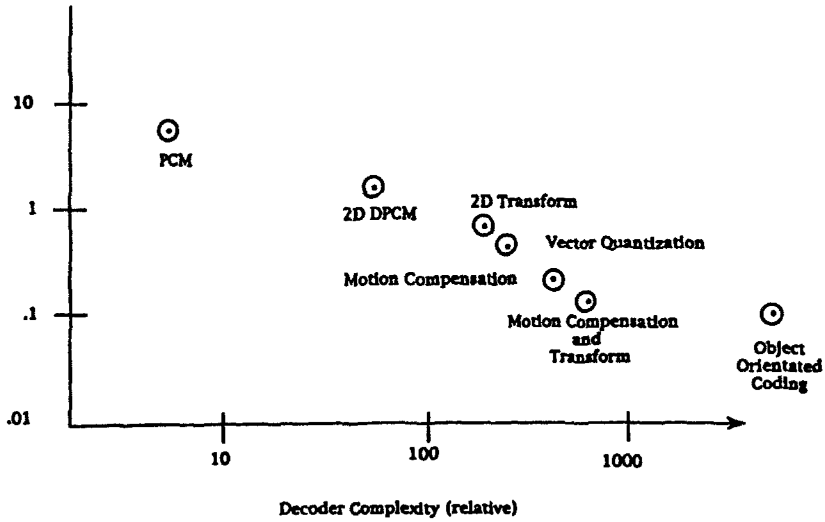

Fig. 2. Compression versus complexity of video decoding

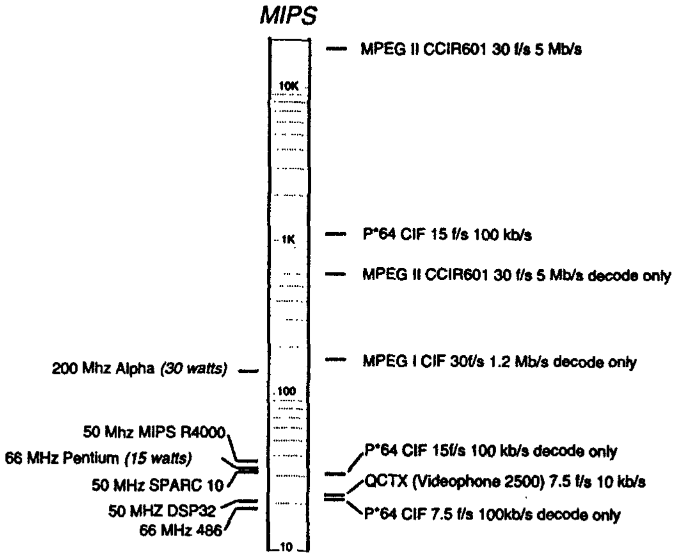

Fig. 3. Computational requirements (in millions of instructions per second, MIPS) for video encoding and decoding at different resolutions

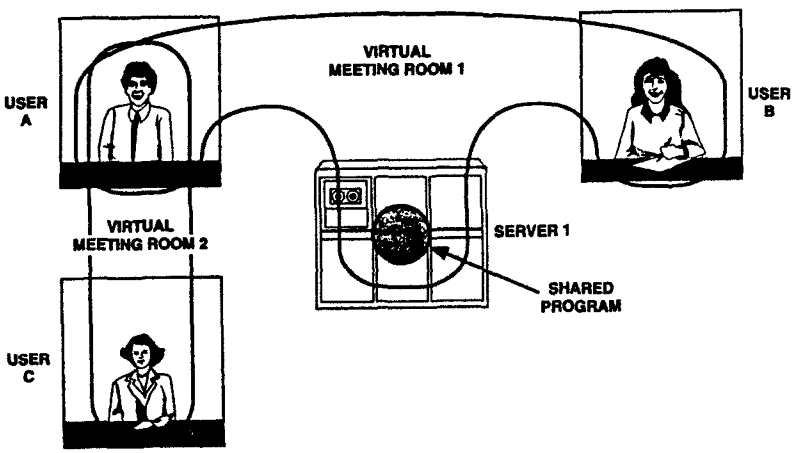

Fig. 4. Virtual Meeting Rooms