

CHAPTER 1

Introduction

1.1 Background and overview of research

1.1.1 Video communication

1.1.2 Transmission mode

1.1.3 Research description

1.2 Related work

1.3 Assumptions and scope of the research

1.4 Outline of the thesis

1.1 Background and overview of research

Today's laser and fiber technology offers transmission capacity that can easily handle high bit rates like those required for video transmission. This has led to an ongoing development of broad-band integrated services digital networks (B-ISDN). The concept of an integrated network is highly attractive since it can satisfy all of a user's needs for information transfer, regardless of its form.

1.1.1 Video communication

The first stage towards the goal of universal integrated services networks is currently being deployed world-wide and it will provide the customer with 144 kilo-bits per second (kbps) capacity, partitioned into three channels, which can be used for voice, data and low bit-rate video communication. However, a network that can carry all present types of services, as well as imminent ones, will require capacity in the order of 140 Mbps per customer. This dramatic increase in capacity can be attributed entirely to video communication. Present and anticipated video services may be classified into two categories: broadcast television and video telephony.

Broadcast television is a one-way transmission used mainly for entertainment and news services. It is currently supplied by aerial broadcast and terrestrial cable television (CATV). Anticipated services, beyond the currently available ones, will rely on the switching capabilities of integrated networks (lacking in CATV systems). Since the customers can be in direct connection with the program providers they can be given a direct influence on the selection of program-material. In its purest form, such a service is referred to as *video-on-demand* [JUD86]. Out of a given library of programs, the customers, rather than the program supplier, choose what should be transmitted to them. The preferred program is forwarded to the requesting customer, and to others whose choice coincide at that point in time. Hence, different selections may have to be provided to each and every customer, and a single program may be sent independently to several customers because they requested the program at slightly different points in time. All this constitutes a daunting communications-task even for the high-capacity networks that are envisaged. However, the program delivery may be restricted to start at specific hours so that at least multiple transmissions of single programs are avoided. A less demanding form is referred to as *democratic television*. Here the supplier offers a selection of programs and the interested customers cast votes; the most requested program will be provided. In both cases of user controlled programming, the service is based on the possibility of establishing a connection between customer and program supplier. Hence, the service may only be feasible when each supplier serves a small number of viewers, who may, however, be geographically dispersed. As an illustration consider a program supplier who specializes in Bergman films. With a *video-on-demand* service you can request at any time to have, say, "The Seventh Seal" sent to your home equipment (for example, monitor or video recorder). With *democratic television* the supplier may take requests for any of the available films for an hour or so before

the program starts. The provided program will be the most requested film which, by luck, may coincide with your choice, "The Seventh Seal".

Digital distribution of television, with quality commensurate with the existing analog formats (*e.g.* NTSC, PAL, SECAM) is expected to be possible at bit rates of 30–45 Mbps per channel [GAG88]. The expected introduction of high-definition television (HDTV) will bring viewers programs with a visual quality which is far beyond that of conventional television. HDTV attempts to overcome both the spatial and temporal distortion of regular TV. It will provide a high resolution image with a wide aspect ratio (possibly 16:9) and high-fidelity stereophonic sound. In short, it will bring a viewing experience equal to the best cinematic film. HDTV distribution, it is anticipated, can be accomplished at bit rates of 120–140 Mbps [GAG88].

Video telephony will probably lead to the most novel uses of the network services that will be offered with broadband ISDN. It is a connection oriented service and mainly used for person-to-person communication. It will include all the types of services which we associate with current telephone systems and for which video adds value. (The term telephony is not meant to imply a voice quality at the present level; it may be both high fidelity and stereophonic.) In video telephony we also include unidirectional information flows for query sessions. Among possible uses of video telephony, one will find teleconferences, remote-site education, database browsing, conversations, and surveillance. The limits of its usage will be determined by availability, cost and the users' ingenuity. Video telephony does not have the stringent quality requirements of broadcast television. The quality level may differ from one session to the next depending on the purpose of the session and the communicating parties' willingness to spend more for the higher transmission rate

associated with a high video quality. Consequently, the bit rates could be in the wide range of 56 kbps up to several mega-bits per second. Low bit rates are obtained through high compression and a reduction in spatial and temporal resolution. Transmission around 1 Mbps and up may be achieved by source coding solely, so that the resolution of regular TV can be retained.

The transmission rates for the various categories are summarized in Table 1.1. The uncompressed rates correspond to pulse-code modulated (PCM) video. Even though bit rates of tera-bits per second have been achieved with lasers and optical fiber links, it has not yet translated into usable capacity in a network. The throughput is presently limited by the electronic circuitry in network switches and interfaces. Consequently, it is still important to conserve capacity by seeking high channel utilization and by lowering the transmission rates of information sources by using compression (source coding).

Service	Uncompressed	Compressed	Packet loss rate	Video Quality
HDTV	~ 1000	120-140	2.0 - 2.3	No degradation
Regular TV	216	30-45	6.2-9.3	Slight impairment
Video	135	1-5	55.6-278	Easily noticeable
telephony	30	56 - 384 kbps	723-4960	impairments

Table 1.1 *Transmission rates in Mbps (except where stated) for broadcast television and video telephony [GAG88]. The uncompressed rate refers to a PCM coded signal: for regular TV the rate is that of CCIR 601 with 4:2:2 sampling, the high video-telephony rate corresponds to CCIR 601 with 4:1:0 sampling (see [CCI85]), and the low rate is for the, so called, Common Intermediate Format (CIF) (see [NET88]). The given packet loss rate corresponds to one lost 1000-bit packet per hour of transmission at the compressed rate. The values should be taken $\times 10^{-9}$.*

1.1.2 Transmission mode

Given a channel, the fundamental issue is how to best divide its capacity between

users. Each user can be allotted a fixed amount of capacity which cannot be exceeded and which is wasted during periods when the sender is idle. We shall refer to this as the circuit-switched case. The capacity division may, in this case, be based on time-division or frequency-division multiplexing [SCH87]. The destination is uniquely determined by the locations of the sender's periodically available time-slots, or dedicated frequency band. Alternatively, the user can be allowed to establish a connection which does not guarantee any part of the channel-resource. When needed, the user will transmit as soon as sufficient capacity becomes available. When inactive, no capacity is used. This is usually achieved by time-division, and it can be accomplished by packet-switching [KUL84, TUR86]. Since the sender does not have a circuit to the destination (unlike the periodically available time-slots in the circuit switched case) for each block of information to be transmitted, the destination has to be specified. An information block with address and possibly other control information is referred to as a packet or a cell (packet-length may vary from packet to packet while cell length is always fixed). In the context of broadband networks, packet-switching is at times also called asynchronous transfer mode (ATM), for which the fixed length packet is referred to as a cell.

Traditionally, real-time services have been provided by circuit-switched technology, while packet-switched networks have been primarily used for data transmission. An example is the current [narrow-band] ISDN which offers two circuit switched channels for voice and data, each with 64 kbps capacity, and one packet switched channel of 16 kbps for control and data transfer. However, advances in switching technology have led to the development of fast packet switches which can handle the transmission rates necessary for broadband ISDN. The traditional way of handling real-time transmission should therefore no longer be assumed; the prospect of information transfer, real-time or not, over packet switched networks has to be

given equal consideration. Indeed, there exist strong arguments to justify such use.

- Integration of services in a network will be greatly facilitated if all of them are dealt with in a common format. That is, all information is transmitted as packets identical in appearance along a single route. Therefore, there is no difference between transmission of a single medium service, say voice, and one composed of all possible types (voice, video and data).
- Signals with varying rates are hard to accommodate in circuit-switched channels, resulting in wasted capacity, or in distortion when the bit-rate of the signal exceeds the assigned channel capacity. Packet networks do not impose a fixed capacity limit, whereby the bursty nature of a coded signal can be better accommodated without further degradation in quality or waste of capacity.
- Since a source only uses the network when it has information to transmit, but does not occupy capacity when idle, it is expected that packet-switching may yield a higher utilization of the channel capacity than what is obtainable with the fixed capacity allocation of circuit-switching.
- Packet switching is more flexible than circuit switching in that it can emulate the latter while the opposite is not possible. On a slotted network, it may be accomplished by enforcing a fixed and deterministic resource allocation which means that the packet network operates using time-division multiplexing.
- Packet-switched systems are well-equipped to handle a heterogeneous and changing traffic mix of services [TUR88]. This is a most important point for network providers since the network does not impose a channel structure which may bar introduction of new services. For example, new compression methods are easily taken advantage of since the transmission can readily be carried out at the new lower rate.

Of course, a price has to be paid for these advantages. The problems are related to a fundamental characteristic of packet networks, namely that the guaranteed and timely delivery (within a bounded delay, that is) of a packet is difficult to obtain. For the remainder of this thesis, we shall deal only with packet-switching and its impact on video compression and transmission.

1.1.3 Research description

With the background given above, we can now pose the following general problem of video transmission over packet-switched networks.

A synchronous signal with a high bit rate has to be compressed and delivered under distortion and time constraints by an asynchronous network, where the probability of information loss is nonnegligible.

This thesis is the result of research dedicated to the above problem. It covers both video compression and video transmission, which we henceforth shall refer to collectively as “packet video”.

The starting point of the research was video compression. However, we soon came to realize that it is an issue intimately linked to issues of the signal’s transmission. That leads us to take a more general approach to the original problem, and one that goes as far as to consider architectural issues of a transmission system. The general thought behind our broad investigation is to identify the issues upon which video imposes specific requirements. Equally important, we try to point out the functions which can be designed without particular regard to video because general real-time constraints are sufficient.

Within the wider scope of our investigation, we still devote the larger part of the research to the video coding problem. The source coding method is based

on the technique of subband coding. That choice is well-founded: subband coding appears to compare favorably to other methods such as transform coding, vector quantization, and predictive coding with regard to compression and quality, computational complexity, and possibility for parallel implementation. In addition, we have found that multi-dimensional subband coding provides a simple framework on which to implement packet video.

We develop theoretic results for non-separable subband coding of two-dimensional signals pertaining to sub- and upsampling, alias-free and perfect signal reconstruction, and filter bank design. Based on these results, a coding scheme for packet video is derived and tested in simulations. Appropriate error recovery, based on concealment of errors, is devised for the coding scheme and is evaluated in simulations as well. The results are encouraging with good quality, medium compression, and high resilience to packet loss.

1.2 Related work

Packet video has a short history. A part of the concept came out of the picture telephone effort at Bell Laboratories, namely the advantage of statistical multiplexing of video sources [HAS72]. That issue was again investigated as part of a wider trial by Koga *et alii* [KOG81]. However, most of the other issues that we associate with packet video did not surface until the research in fast packet switching had reached a convincing status. Experiences from integrated network projects have been reported during the last five years and they include MAGNET at Columbia University [LAZ85, PAT85, LAZ86, LAZ87, LAZ88], Prelude at CNET* [THO84, LEB86, GON86a, GON86b, DEV88], and PARIS at IBM [CID88]. MAGNET,

* *Centre National d'Etudes des Télécommunications, Lannion, France.*

which started as a project on integrated local area networks, is based on a slotted-ring topology and is now evolving to become a metropolitan area network based on interconnected rings. Considerable attention in the project is given to intelligent network control. Prelude, a wide area network prototype, centers around the development and evaluation of a fast packet switch. The PARIS project aims at the development of private integrated networks. A project that has its main thrust in video communications is carried out by Verbiest and his colleagues at BTM.* Their test system consists of a variable rate video coder and an ATM switch [VER88a].

Discussions on general aspects of broadband transmission that include video can be found in [KUL84, CHI84a, CHI84b, TUR86, TUR88, DEC88, COU88, TER88, KAR89a]. Video coding for packet transmission has been considered by a number of authors [VER86, KAR87, KAR88a, OHT88a, KAR88b, OHT88b, VER88a, VER88b, VER88d, KAR89a, NOM89]. Other issues of importance for packet video include error recovery [LEE87, KAR87, AAR88, KAR88b, OHT88a, NOM88, OHT88b, VER88b, VER88d, KAR89a, NOM89], video resynchronization [COC85, VER86, VER88b], and statistical analysis and modeling of the variable bit rate of coded video [HAS72, KOG81, SEN87, MAG87, HUA88, MAG88, VER88a, DOU88, VER88c, NOM89]. The abstracts of the recent *Second International Workshop on Packet Video* give a good overview of new results in the field [PVW88].

In recent years, subband coding has gained attention as a powerful method for compressing still images and video. Theory and coding results can be found in [VET84, VBR86a-b, WOO86, GHA86, GHA87, SMI87b, WES87, KAR87, GHA88, LEG88a-b, WES88a, KAR88a-b, ANS88]. Moreover, the filter banks, which are integral parts of subband coding, have been thoroughly researched for one-

* *Bell Telephone Manufacturing Co., Antwerp, Belgium.*

dimensional systems. For an overview of results, the reader is referred to [VET86, SMI87a, VET87, VAI87a, VET89] and references therein. It is only proper that a two-dimensional signal, such as an image, should be processed in a two-dimensional system. Little attention has been given to such filter banks. Initial discussions and results, which would serve this application, have appeared in [VAI87b, ADE87, VIS88, KAR88c, ANS89, KAR89c].

1.3 Assumptions and scope of research

Although we have tried to keep our packet video investigation general, some assumptions and restrictions have been made. First, visual communication with images, graphics and facsimile is not explored since such information can be treated as data for which the delay constraints are not so stringent. Second, no information is exempt from transmission error. Therefore, there are no perfect channels for transmission of side information and we will not consider control messages as an aid to synchronization, error recovery, and congestion control during the data-transfer phase. Third, we assume that transmission rates are restricted more by processing speed (for functions such as routing) in the network nodes than by link capacity. In addition, the links are assumed to be “fairly reliable,” *i.e.*, loss due to congestion dominates over bit errors due to noisy links. This leads to a design philosophy which places minimal functionality within the network nodes.

For the theoretic development of the subband coding scheme there are few restrictions: the theory is valid for signals defined on any sampling lattice. However, the sampling structures of the input and the subsampling must be representable as lattices and sampling structures which can only be represented as a union of shifted lattices are not considered. For the design of the video coding system we limit the theory by considering only separable subsampling patterns (horizontal and vertical

subsampling) and separable filters.

The services we have in mind for the coding scheme belong to those of video telephony. Thus, a lowered visual quality is accepted so that the transmission rate and cost can be substantially reduced. A proper trade-off between rate and quality has to be evaluated according to human perception of video (not solely mean-square error), as given by [ROG83, GIR88]. Since we assume that packet loss is unavoidable in the packet-switched environment, the coding scheme has to be designed with robustness in mind so that it lends itself to integration with practical error control methods. It is also beneficial if the compression ratio is controllable so that the trade-off between transmission cost and video quality can be decided at the start of each new session.

1.4 Outline of the thesis

In Chapter 2 the design considerations of packet video coding and transmission are analyzed in more detail. These pertain to variable rate and transmission, source coding, error recovery, transmission delays and resynchronization. The first discussion examines the pros and cons of variable rate coding and transmission. Second, we propose a generic structure for a packet video coder, referred to as hierarchical source coding, which offers a framework for resolving the complications resulting from packet video transmission. Error recovery is proposed in two forms: the traditional, by error control coding, and a technique based on concealment of error through the use of visual redundancy. The timing issues of packet video involve end-to-end delays and its statistical variations, as well as discrepancies in the clock frequencies between the sender and the receiver. We discuss these issues and propose a resynchronization method based on temporal interpolation.

The integration of video coding and transmission functions into a network architecture is presented in Chapter 3. Video transmission considerations pertaining to the network layers of the model are covered. The aim is to make these layers transparent to the underlying format of the video-signal (*e.g.* PAL, NTSC). These layers need only to provide real-time transmission with high throughput. Video-specific functions, such as compression and error recovery, are fitted into the higher layers of the model. The application layer is presented as the holder of format-conversion functions such as analog to digital conversion (and vice versa), and video resynchronization. Video compression and error recovery by concealment are associated with the presentation layer. The session layer has to allow session types which can meet the needs of any service and provide for user selected quality of these services. In conjunction with the transport layer, we discuss packetization, error recovery by channel coding and the trade-offs encountered when choosing packet length. The chapter is concluded by a discussion of means to achieve congestion control.

Beginning with Chapter 4, we address the source coding problem and investigate subband coding as a method for video compression. Chapter 4 develops the general theory of two-dimensional subband coding. Given a general subsampling structure, the related polyphase decomposition of a filter bank is derived and is used to find the input/output relationship of the subband analysis and synthesis system. We pose the conditions for alias-free and perfect reconstruction filter banks and give structures for the synthesis of perfect reconstruction filter banks through paraunitary and non-paraunitary systems. Conditions are given in the polyphase domain for the linear phase response of two-dimensional systems.

By restricting the theory to separable sampling lattices and filter banks, we can

easily extend the subband coding to encompass all three dimensions of a video signal. This is done in Chapter 5. The three-dimensional coding scheme gives eleven subbands out of which ten are encoded by PCM and one is DPCM encoded, followed by run length encoding of the PCM values and the prediction error. The scheme is evaluated with regard to computational complexity. Error recovery based on concealment through the use of visual redundancy is proposed. The chapter ends with a discussion of possible extensions to the proposed coding scheme.

The coding scheme derived in Chapter 5 has been simulated and the results are presented in Chapter 6. The simulations cover both lossless and lossy transmission. The coding scheme and its error recovery mechanism were found to perform well. The coding scheme was also used in a simplified version in order to make feasible the simulation of several video sequences, each 2 minutes in duration, which generated statistics of the output rate. All the results are then used and tied together in a discussion of quality of service pertaining to coding quality and transmission quality.

Chapter 7 concludes the thesis. Two appendices are included of which one covers modeling attempts for video sources, and the other is a study of image extension for subband coding. All references are given at the end of the thesis and they are indexed by the three first letters of the first authors last name together with the publication year.