

# CHAPTER 7

## Summary, Extensions and Conclusions

---

### 7.1 Summary

#### 7.1.1 Network issues of packet video

#### 7.1.2 Subband coding as hierarchical source coding

### 7.2 Extensions

### 7.3 Conclusions

---

### 7.1 Summary

Future broadband integrated services digital networks will carry all types of information, such as broadcast television, video telephony, regular telephony and general sound services, data transfer or any combination thereof. By basing such a network on packet switching, these services can be dealt within a common format. Packet switching is more flexible than circuit-switching in that it can emulate the latter (while the opposite is not possible), and vastly different bit-rates can be multiplexed together. Since there is no fixed channel structure in such a system, it is well-equipped to handle a heterogeneous and changing mix of services. This thesis is the result of research into video coding and transmission over packet-switched networks. The emphasis of the research has been on video compression.

#### 7.1.1 Network issues of packet video

By allowing variable rate transmission, the paradigm imposed on circuit-switched systems of constant-rate / varying-quality may be changed into one of variable-rate / constant-quality. In addition, when variable rate signals are multiplexed statistically in the network, without reservation of capacity, a higher utilization of

the channel capacity is possible than with fixed capacity allocation. However, the disadvantage is that it creates strong dependencies between the entities of encoding, transmission and decoding. Therefore, design of packet video coders cannot be made in isolation; the global transmission system needs to be considered. A structured way of considering the entire system of signal processing and transmission is to place the required functions in their proper context, a network architecture model.

The main benefit of using a layered model is to incrementally limit the dependence on the video-format down through the layers. Owing to the format elimination, many of the layers may be shared between sessions. In the model, the network layers, which constitute a network node, should provide real-time service with a high throughput. We require that packets arrive in order although some of the delivered packets may contain bit-errors while others may be lost. The data-link control should be kept to a minimum, barring retransmission and allowing only error control for the framing information. The network layer must provide virtual-circuits and, for multi-party sessions, there has to be a multiple addressing capability together with packet duplication. The aim of the route selection should be to minimize the end-to-end delay. Furthermore, the network layer should provide for temporary as well as permanent virtual-circuits. If congestion control will be exercised within the network, it should be priority based. This will minimize the loss of perceptually important data and thereby simplify error recovery. Congestion control can be performed at the network interface to assure uninterrupted service of transit traffic. Priorities may not be necessary when sufficient capacity can be guaranteed for important information. This information will not be affected by congestion, as will the less important information, which has to vie for its share of the transmission resources. If dynamic guarantees are not provided, the capacity has to be allocated at call set-up and the call must then be monitored to assure

that it stays within its allocated quota.

The upper layers of the model reside at the customer's premises, and associated with them are the functions specific to packet video. Such functions include format conversion and the resynchronization at the application layer; source coding, ciphering and error recovery, which fall within the scope of the presentation layer, and error control coding at the transport layer. The session layer, as well as all the layers below, is completely transparent to the format of the video-signal. Thus, it is there that different information types (sound and data) are included to form a true integrated-services session. In addition, the session layer should provide for various session types, and negotiable quality and cost for each session. The session types are classified as point-to-point, multi-cast, multi-drop, and broadcast sessions. The former three types are negotiated end-to-end. Broadcast is considered to be based on permanent virtual circuits. Therefore the set-up would be negotiated locally with the receiving node. The session quality depends on two parts: source coding and transmission. The lossless session quality is determined by the coding, and the maximum permitted deterioration under network congestion is controlled by the use of the transmission priorities, or amount of guaranteed capacity, and error control coding. Packet video, compressed with hierarchical coding, would be advantageously transmitted with a hybrid of fixed and statistical capacity-sharing. This can protect visually important information but still yield utilization gain through statistical multiplexing of the non-fixed channels. The segmentation of a data stream into packets takes place at the transport layer. Lost packets must be detected to ensure that the error control encoder and decoder remain in phase. Transmission with fixed packet length appears favorable since it simplifies packetization, error control coding and routing.

The timing problems of packet video are twofold. First, there are variations in transmission delays and, second, there is no common time reference for the transmitter and the receiver so their clocks may differ. Packet delay jitter can be removed by buffering the packets at the receiver. Transmission delays are irrelevant for one-way sessions, such as broadcast television. Therefore all delay-jitter may be eliminated by a large buffer. For video telephony, we propose that video, as an information carrier, is subordinate to sound, and the video-delay has to be controlled to provide lip-synchronization. Thus, since buffering has to be limited, increased packet loss must be accepted. This occurs when the jitter exceeds than which the buffer can smooth. We also discuss the basic strategies for video resynchronization. Although clock recovery may be achieved from the received data stream, such techniques may not serve all types of video sessions. The frame rates of received video signals would therefore have to be altered to match the play-out rate. Temporal filtering can alleviate temporal discontinuities which may otherwise disturb the viewer.

### **7.1.2 Subband coding as hierarchical source coding**

Hierarchical source coding is believed to be a natural way of performing video coding since it categorizes video information into more important and less important sub-signals. Compression, transmission, and error recovery can therefore be adapted to the sub-signal's particular importance, which may be more efficient than treating the signal as a single bit-stream. We have investigated a hierarchical source coding scheme based on the technique of subband coding. Our method is based on three-dimensional analysis of the video which yields subbands with various combinations of high and low temporal and spatial frequencies. The development arises from our new results on the theory of general two-dimensional multirate filter banks.

The multirate theory is general, with no restrictions on the geometries of the periodic input and subsampling lattices. The subsampling is described by linear transformation, *i.e.*, scaling and rotation. The transformation is restricted to keep one axis of the subsampling lattice collinear with an axis of the input lattice (which does not restrict the lattice). The modulation caused by sub- and upsampling can then be derived. Given the subsampling, we can state the polyphase decomposition of two-dimensional filter banks leading to a formal matrix-relation between the time-domain and the polyphase domain. By decomposing the analysis and the synthesis filter banks into polyphase components, the input/output relation of a subband system is determined. This leads to necessary and sufficient conditions for alias-free and perfect signal reconstruction. (Curiously enough, it was found that the pseudo-circular form of the transfer matrix, which is sufficient for alias-cancellation in the one-dimensional case, is generally not sufficient for two-dimensional systems.) With these conditions in hand, design structures were derived for polyphase matrices. These structures enable the design of two-dimensional perfect reconstruction filter banks. These pertain to both paraunitary as well as non-paraunitary systems. The number of free variables are tabulated for these types of systems. We explained that a polyphase matrix of a FIR filter bank, for which the determinant is a monomial, has an inverse which is FIR as well. In this case, the two-dimensional polyphase matrices can be factored. This fact was used to justify a cascade structure for design of paraunitary systems. Two other design structures originate from the two-dimensional state-space equations. In addition, we discussed design based on the Smith form decomposition of a polynomial matrix. Finally, we derived conditions to test these filter banks for linear phase directly in the polyphase domain, which enabled us to develop a restricted design structure for linear phase systems.

The implementation uses separable filter banks which are non-paraunitary, but

yields perfect reconstruction when there is neither coding nor transmission loss. The aforementioned three-dimensional subband analysis yields eleven bands which are subsampled to their new Nyquist frequency. Of these eleven subbands one is DPCM encoded and ten are encoded by PCM. Their quantizers have a wide zero-level, which removes low-amplitude picture noise, and uniform symmetric levels for the non-zero parts. The quantization levels are represented by variable length codewords, with the most likely being the shortest. Only the non-zero values of the quantized prediction error and the PCM-encoding are transmitted and their locations are given by a run-length code. Our coding scheme has an architectural simplicity suitable for parallel implementation and provides a tractable integration in the network architecture. Furthermore, its implementation is multiplication-free; the only operations needed are integer additions and shifts. The simulation results indicate that a good video quality can be obtained at compression ratios of around 20 times with this low complexity coder. The simulations also verify that the quality level remains virtually constant when the transmission rate is allowed to vary. The extensions to the coding scheme that were outlined included non-separable subsampling and filter banks, more complex subband encoding, incorporation of motion compensation into the subband analysis, and the coding of color sequences.

Error recovery was considered, since packet loss will occur in a network relying on statistical multiplexing. Error control coding is a good but not sufficient remedy; more packets may be lost than the code can correct. However, combined with error concealment by using visual redundancy, error control coding may solve the problem of packet loss. The general requirement for concealment is that the number of lost values and their locations can be deduced. The implemented coding scheme is augmented by an error concealment method. Synchronization flags are inserted into the coded bit stream, where each individual flag contains the address of the

starting location of the data that follows. Secondly, all subbands are packetized independently of one another. For ten of the eleven subbands, missing values are replaced by mean-valued samples. For the first DPCM-encoded subband the the corresponding area of the previous video frame was used to replace a missing area. Error recovery by concealment was simulated together with the coding scheme and was shown to be a good remedy for packet loss. However, for subband 1 the concealment should be based on motion compensated replacement from the previous frame; without it the replacement performs insufficiently in areas of motion. By using motion estimation, a more appropriate area could be found for replacement.

The transmission rates of subband coded video have been investigated. Ten sequences, each two minutes in duration, have been acquired and used to produce bit-rate statistics. The analysis of the output rates shows that the subbands have vastly different characteristics, ranging from nearly constant rate to highly varying (with a peak rate of approximately five times the time-average). The basic statistical properties have been tabulated and the gain achievable with statistical multiplexing has been calculated. This calculation shows that statistical multiplexing may yield a channel utilization of up to 7 times that possible with fixed multiplexing for highly varying subbands. Transmissions of subsets of all eleven subbands were also simulated. These cases show that a lowered rate can be obtained simply by excluding subbands from transmission which also reduces the visual quality of the decoded video signal. In light of all these results, the quality of transmission service was discussed. It is shown that subband 1, which exhibits a near-constant output rate, can be assigned capacity resources according to its peak rate without leading to any substantial loss of channel utilization. This is of importance since this subband is the most vital for the quality of the reconstructed video. Any other subband should have partial or no guaranteed capacity, depending on its importance and

statistical behavior.

## 7.2 Extensions

As indicated in Section 1.2 of the introduction, packet video has a short history. After this study, we believe that there are still unresolved issues. In this section we will discuss areas that could benefit from in-depth investigations. Subband coding of video will not be discussed, since the topic already appears in Sections 4.9 and 5.5.

We have not ventured to recommend a specific network architecture for packet video. The reason is that very different design philosophies suffice for video transmission [LAZ85, DEV88, CID88, TUR88, LAZ88]. Therefore, the solutions to issues like media access control, congestion control, capacity guarantees, use of priorities, and routing are unique to the underlying implementations. This results from the linkage between video encoder, channel and decoder, as discussed in Section 2.1.2. Among problems that do apply to packet video as a whole are format conversion, resynchronization, error recovery by concealment, and provisions for broadcast services.

The most difficult dimension for format conversion is the temporal one. Spatial conversion is performed in a straightforward manner, by filtering and resampling. In the temporal direction filter and resampling is insufficient and motion compensation has to be incorporated to avoid blurred motion. This is a general compatibility issue for television formats [REU89]. It may be interesting to investigate if temporal interpolation can be the single solution to frame rate conversion for cross-compatibility, resynchronization (see Section 2.4.2), and error concealment through the use of visual redundancy. The last, error concealment, becomes such a problem



in the following manner: if a network has a low probability of packet loss—repeating a frame to conceal a faulty one may be an adequate remedy if augmented by motion compensation. Alternatively, error concealment methods may be derived from ample work done in image restoration. However, these methods tend to be computationally complex which may make them unsuitable for real-time use.

Of additional interest is the problem of including the location information into a bit stream in order to achieve the erasure property. One method is suggested in Section 3.2 but it is one of many possible methods. With the erasure property, the information in a packet is decodable independently of the information in other packets. It may therefore be possible to remove delay-jitter after decoding, at the application layer, where resynchronization by frame rate conversion would be performed (see Section 2.4.2). Thus, a single storage of video-frames may suffice for signal recombination, resynchronization, error concealment and jitter removal. This would reduce memory requirements in consumer equipment.

In Section 3.3.1 we sketched out the procedure for broadcast transmissions. This is highly important since entertainment appears to be one of the strongest driving forces of B-ISDN. Given our broadcast scenario of transmission over permanent virtual-circuits, we should investigate how the routes should be selected to offer the best balanced load over the broadcast region. The actual set-up procedure should be determined, for which it is to be able to “break in” on a signal at any time and still obtain synchronization. Note that standards for broadcast format must include, not only the video signal and its compression method, but also any error correcting code and synchronization (location) information stuffed into the compressed video signal.

Finally we wish to discuss an issue related to network performance evaluation.

Performance analyses usually result in measures of queue sizes, delay characteristics and probabilities of packet loss. It is our belief that these measures of network performance do not reflect perceptual quality, which is essential for real-time video and sound services. Let us take as a starting point the maximum transmission-delay that the service can tolerate. The margin with which that delay limit is met is irrelevant since it only corresponds to the waiting time in the jitter-removal buffer at the destination. If, on the other hand, the limit is exceeded the packet is considered lost. Consequently, both delay characteristics and effects of buffer over-flow result in packet loss. However, you cannot use packet loss probabilities to evaluate quality degradation of sound and video sources. Instead, such an evaluation should be made according to some distortion measure, such as mean-square error. Thus, a combined utilization of both information theory and queueing theory is necessary. From an information theoretic point of view, the source sees a channel with capacity, which varies with the network-load along the path of the circuit. This way, a queueing analysis, as that of Yin [YIN88], could be extended to include a mean-square error analysis of the received signal. The concept is in part akin to information theory with a delay constrained signals (see for example [ENG87]). We shall illustrate the importance of the above. Assume that a priority scheme is being used. If a priority assignment undervalued the importance of the associated information, then the priority scheme may perform worse, in a perceptual sense, than transmission without priorities. Also, let this underscore the importance of finding realistic source models for video signals and video bit-rate sources.

### **7.3 Conclusions**

The purposes of the research reported in this thesis were twofold. The first was to develop subband coding as a video coding method that could be comparable to

the best existing video coding methods. This led us to apply the theory of two-dimensional linear systems to the problem of filter bank design. Some elementary applications of lattice theory led to the derivation of conditions for alias cancellation and perfect reconstruction. The second purpose was to study the use of subband coding for video-telephony services of future packet-switched broadband integrated services digital networks. To this end, a hypothetical network architectural model for fast packet switched systems was proposed. In this model we placed the coding function and other functions necessary for reliable video transmission.

The simulation-results prove subband coding to be on par with established methods like discrete transform coding and vector quantization. Our implementation of subband coding has a low computational complexity and a structure well-suited for parallel implementation. It shows good visual quality at about 20 times compression. Moreover, it is a likely candidate for applications as high-definition television coding. Its use as hierarchical source coding was tested through simulated packet video transmission. These simulations showed that packet loss can be remedied. Also, subband coding allow the users to negotiate quality of service. This permits the users to set-up a session with a visual quality and transmission cost appropriate to the session's purpose. The general outlook on the entire topic of packet video is positive. We believe that its advantages are indisputable and the solutions to its associated problems appear to be tractable.