HDTV Transmission Format Conversion and Migration Path

by

Lon E. Sunshine

S.B., Massachusetts Institute of Technology (1988)
S.M., Massachusetts Institute of Technology (1992)
E.E., Massachusetts Institute of Technology (1992)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1997

Author ______________________________
Department of Electrical Engineering and Computer Science
June 30, 1997

Certified by ______________________________
Jae S. Lim
Professor of Electrical Engineering
Thesis Supervisor

Accepted by ______________________________
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

HDTV Transmission Format Conversion and Migration Path

by

Lon E. Sunshine

Submitted to the Department of Electrical Engineering and Computer Science
on June 30, 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Abstract

The development of high-definition television (HDTV) in the United States has recently resulted in the establishment of a national standard for its coding and transmission. This standard dictates, among other specifications, certain limitations on the allowable resolution of transmitted HDTV signals. There are many of these allowable formats, permitting scene-dependent processing and encoding; however, this introduces the need for conversion from the production format to the transmission format at the transmitter, and from the transmission format to the display format at the receiver.

The adopted HDTV standard limits the resolution of television transmissions. In the future, having a path for migration to higher-resolution formats will keep television up-to-date with available technology. Requiring this migration path to be backward-compatible will be desirable so as to not render standard HDTV receivers obsolete. Since the original enhanced-resolution signal will often be available at the transmitter, it can be used to assist with the migration-path processing. Furthermore, performing computationally-intensive processing at the transmitter will reduce consumer costs and enable all receivers to benefit from the advanced signal processing.

This thesis develops a backward-compatible migration-path structure which is based upon hierarchical block processing and motion-compensated prediction. We demonstrate that having access to the original video provides a critical advantage to determining enhancement information. Furthermore, we show that even with only a small increase in digital data requirement, the resulting picture-quality improvement is notable.

Thesis Supervisor: Jae S. Lim
Title: Professor of Electrical Engineering

Dedication

To my mother and father
Rachelle and Herbert Sunshine

For showing me the way.

To my grandparents,
Mitzie and Sam Sunshine
Freda and Louis Topel

Who have always been and shall always be with me.

To Julia May and Lauren Elizabeth

The next generation.

Acknowledgements

I would like to take this opportunity to express my gratitude to all those who assisted—by way of technological, emotional, moral, or fiscal support—in this endeavor. I am fortunate to have had so many close friends, relatives, and colleagues who have provided assistance along the way, and I apologize for not being able to specify all of you by name; your involvement has not been overlooked and I take this opportunity to thank you first.

Additionally, there are a few individuals who have had a direct impact on the production of this thesis, and whose contributions I will not forget:

Thank you to Professor Jae S. Lim, my advisor and mentor, without whose support and guidance, none of this would have been possible. I'd also like to thank Professor David H. Staelin and Dr. Dan E. Dudgeon for their willingness to serve on my thesis committee.

Thank you to Professor Emeritus William F. Schreiber, for volunteering his time to provide me with technical assistance and guidance with respect to the HDTV process. Thank you to Dr. Andrew C. Singer and Dr. Shiufun Cheung, two of my closest friends and colleagues while at MIT, for their incessant pressuring of me to finish, and for their assistance with my thesis drafts, both of which are greatly appreciated. Thank you, also, to Dr. John Apostolopoulos, another close colleage whose technical input greatly aided my successful completion of this thesis.

A special thanks to Cindy LeBlanc, group secretary for (and den mother to) the Advanced Television Research Program, and to Denise "Cindy-on-Fridays" Rossetti for their day-to-day, behind-the-scenes actions which make everything run smoothly.

Finally, words cannot express the debt of gratitude I owe to my parents, Herbert and Rachelle Sunshine, for giving me the courage and strength needed to see this thesis through, for their daily words of support and reassurance, and for making me the person I am today.

L. Sunshine
Cambridge, MA
June 30, 1997

Contents

---

1 Introduction **10**

1.1 The HDTV Development Process . . . 10

1.2 HDTV Subsystems . . . 12

1.3 Thesis Outline . . . 14

2 HDTV Transmission Alternatives **15**

2.1 Transmission Formats . . . 15

2.2 Migration Path . . . 22

2.3 Format Conversion . . . 26

3 Deinterlacing **34**

3.1 Traditional Deinterlacing Methods . . . 34

3.1.1 Intraframe methods . . . 34

3.1.2 Interframe methods . . . 35

3.2 Migration Path Deinterlacing . . . 36

3.2.1 Block Processing . . . 38

3.2.2 Deinterlacing Modes . . . 39

4 Experimental Results **41**

4.1 Transmitter Implementation . . . 41

4.1.1 Deinterlacing Modes . . . 44

4.1.2 Block Size . . . 45

4.1.3 Motion Vectors . . . 46

4.2 Entropy Coding Results 48
    4.2.1 Deinterlacing Modes 48
    4.2.2 Frame Partitioning 48
    4.2.3 Motion Vectors 49
4.3 Objective Results 49
4.4 Subjective Results 60

5 Conclusions **67**

5.1 Summary 67
5.2 Future Directions 68

A Simulation Details **70**

B Experimental Data **72**

C Subjective Testing **74**

List of Figures

1.1 HDTV Transmitter 11
1.2 HDTV Receiver 13

2.1 Transmission Alternatives 18
2.2 Receiver Signal Processing 21
2.3 Standard HDTV System 24
2.4 Enhanced HDTV System 25
2.5 Future Advanced HDTV System 27
2.6 60-fps Progressive Scanning 28
2.7 30-fps Progressive Scanning 29
2.8 Interlaced Scanning 30
2.9 Anti-aliasing Regions for IS 33

4.1 Transmitter Deinterlacer 42
4.2 Adaptive Deinterlacer/Comparator 43
4.3 Block Subdivision Criteria 46
4.4 Entropy Scanning Order 47
4.5 Original Test Sequences 50
4.5 (continued) 51
4.6 MSE vs. Bit Rate 53
4.6 (continued) 54
4.6 (continued) 55
4.7 Access to Original Video 58

4.7 (continued) 59

4.7 (continued) 60

4.8 Deinterlacing Results – `mall` 61

4.9 Deinterlacing Results – `girl` 63

4.10 Deinterlacing Results – `tulipstext` 64

List of Tables

2.1 Grand Alliance HDTV Transmission Formats 16
2.2 Available Transmission Formats with Enhancements 23

3.1 Common Video Attributes and Proposed Deinterlacing Modes 39

4.1 MSE for Deinterlacing Experiments 53
4.1 (continued) 54
4.1 (continued) 55
4.2 MSE Improvement 56

B.1 Variable Block Size Data 72
B.2 Fixed Block Size Data 73
B.3 Intraframe Deinterlacing Error 73
B.4 No Progressive Original Data 73

C.1 Subjective Test Results 74
C.2 Pairwise Sequence Comparison 75

Chapter 1

Introduction

1.1 The HDTV Development Process

The development of high-definition television (HDTV) for terrestrial broadcasting in the United States has been ongoing for the past decade. Eventually, HDTV will replace the current color television standard (NTSC), which has been the national broadcasting standard since officially adopted by the FCC in 1953 [20]. Two of the most dramatic ways in which HDTV differs from NTSC is in picture resolution and aspect ratio: HDTV will have many more lines of resolution in the television picture than NTSC, and the screen aspect ratio—the ratio of the width to the height of the picture—of HDTV will more resemble that of a movie theatre.

The HDTV development process led to the formation, in 1993, of the Grand Alliance (GA), a consortium of industrial and educational institutions who worked together to define a national HDTV standard. Among the restrictions placed upon the GA in the development of such a system was that the HDTV signal be channel compatible with NTSC; that is, every HDTV transmission must fit within the 6 MHz channels currently allocated for NTSC broadcasts. Furthermore, there was a power constraint placed upon the HDTV broadcast to prevent interference with existing NTSC transmissions. Under these constraints, current compression and transmission technology can provide approximately 20 Mbps of data for each channel [23].

The GA developed an all-digital television system which was based upon the MPEG-2 video compression standard but had a specified set of allowable transmission formats[1] which could be used to encode high-definition video. This system was proposed to the Advisory Committee on Advanced Television Systems (ACATS), an advisory group to the Federal Communications Commission (FCC). ACATS approved the GA system and ACATS submitted this system to the FCC as its recommendation for adoption as a national standard [1].

---

[1] The term *transmission format* is used to describe the resolution (horizontal and vertical) and the scanning format (progressive or interlaced) of an image sequence immediately prior to encoding for its broadcast. The terms *production format* and *display format* will be used to describe the sequences produced at the output of the camera (with perhaps some post-processing) and at the display, respectively.

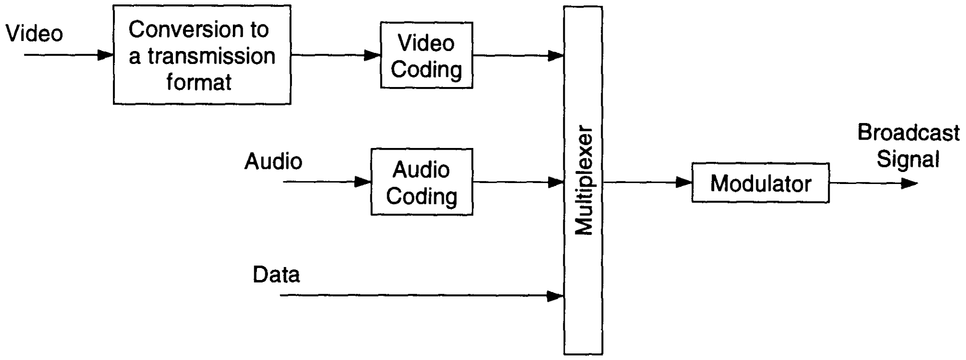


Figure 1.1 HDTV Transmitter Block Diagram. The five major subsystems of the HDTV Transmitter

In December, 1996, the FCC, after much deliberation, adopted the ACATS recommendation, with the exception of the proposed transmission formats [16]. The FCC decided that it is in the public's best interest to let market forces determine which transmission formats should be used. Their decision allows broadcasters to transmit video in any format, as long as it complies with the MPEG-2 constraints which include a maximum bit rate and a maximum resolution [17]. This system is currently being implemented and expected to be commercially available within the next 2 years.

Eventually, the evolution of relevant technologies may permit the broadcast of higher-resolution signals than are possible with MPEG-2 video coding. Over media such as cable, this situation may arise relatively soon. The concept of a *migration path* concerns the transition from standard HDTV to higher-resolution formats. Preferably, this migration will be done in backward-compatible manner in order to prevent the initial HDTV sets from becoming obsolete. Having a graceful migration path will allow higher-quality HDTV transmissions to be displayed by receivers which can handle them, but will still permit lower-caliber receivers to decode and display the standard HDTV picture.

This thesis addresses the migration path more closely. The role of format conversion will be investigated, and a backward-compatible implementation for yielding an increase in resolution over standard HDTV will be proposed. An assumption is made that an increase in the digital data rate will become available to the encoder, and this thesis will describe how this additional capacity can be used to effectively increase resolution.

1.2 HDTV Subsystems

A block diagram of the five major subsystems in the GA [34] HDTV transmitter is shown in Fig. 1.1. Each of these systems has its counterpart in the receiver which is shown in Fig. 1.2. When an input video stream enters the transmitter, it is first converted to an allowable transmission format. The video is then compressed and coded, and multiplexed with audio and other data before channel coding, modulation and broadcast.

Having several possible transmission formats adds flexibility to the system by permitting source-adaptive encoding. This can, in turn, increase picture quality. The format in which a particular program is encoded will depend on several factors including source material, scene content, and desired resolution. Some of the conversion options are outlined in more detail in chapter 2. The flexibility of multiple transmission formats, of course, adds some complexity to the system: since a transmitter can select any of several formats in which to encode the picture, each television receiver must be able to decode any of these formats or, otherwise, risk failing to receive a certain broadcasts. In the original GA proposal, there were six HDTV-quality formats and twelve standard-definition (SDTV) formats. The recent FCC decision now places less constraints on the transmission format used; however, there is still a resolution limitation to MPEG-2 video coding.

The video coder in the transmitter is based on the MPEG-2 standard, developed by the Moving Picture Experts Group. The GA system conforms to the MPEG-2 main profile implemented at high level (MP@HL), which contains the following attributes:

- Maximum bit rate of 80 Mbps
- Maximum sample rate of 62.6M samples/sec
- Upper bounds of 1920 samples/line, 1152 lines/frame, 60 frames/sec
- Progressive and interlaced scanning capability
- Motion-compensated prediction
- Bi-directionally predicted frames, in addition to intra-coded frames and forward-predicted frames
- Block-based, intra-frame and interframe compression

Compression is achieved by way of motion estimation/compensation and the discrete cosine transform (DCT) which serve to efficiently describe, respectively, the temporal and spatial

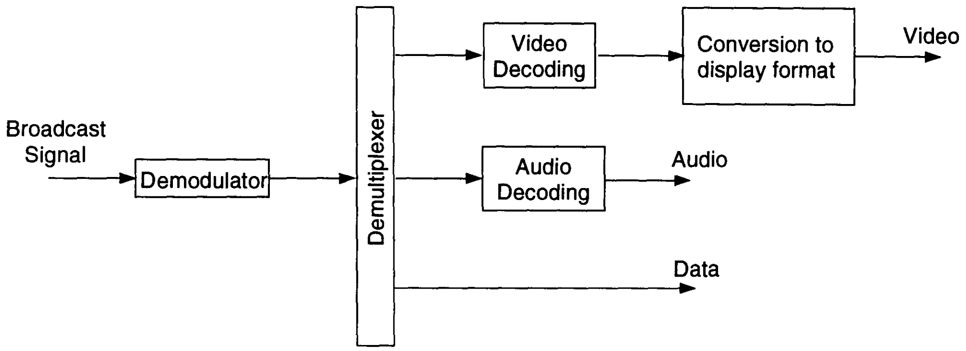


Figure 1.2 HDTV Receiver Block Diagram. The five major subsystems of the HDTV Receiver

redundancy in the video stream. Motion estimation is performed, on a block-by-block basis, in order to predict the current picture from the adjacent frames, and in many situations can give an efficient representation for many regions of the image. The DCT is widely known to have good energy compaction properties and fast algorithms with low-cost implementations make it attractive for use.

The audio subsystem works in parallel with the video coder. This system is based upon Dolby Labs' AC-3 digital audio compression system. The AC-3 coder uses a perceptual model of the human auditory system to encode, at CD-quality, 5.1 channels of audio (left, center, right, left surround, right surround, and low-frequency enhancement).

The coded video and audio streams, and possibly an auxilliary data stream, are passed to the multiplexer/transport system which packages the streams for transmission. Each 188-byte packet contains a 4-byte header and a payload consisting of one type of service (audio, video, text, etc.) The packet header contains synchronization information, a payload description, and a mechanism for encryption control to allow subscription services, such as pay-per-view, to be easily accommodated. The transport system ensures that audio and video signals are synchronized by multiplexing them together along with a control data stream and producing the transmitted bitstream.

The output of the multiplexer is channel-coded and then modulated for transmission by a vestigial sideband (VSB) technique. Specifically, 8-VSB is used, which consists of a four-level AM vestigial sideband signal with trellis coding. Each 188-byte packet from the transport system is stripped of its sync byte by the modulator, and the remaining 187 bytes are coded with a Reed-Solomon code, adding 20 parity bytes for error correction.

The HDTV receiver, shown in Fig. 1.2, contains blocks analogous to those in the transmitter, each performing the inverse function to receive, demodulate, decode and display the video program and sound. Typically, each receiver will have a display with fixed scanning (progressive or interlaced) and resolution, so the received signal will have to be converted to that display format. This format can, in general, not be compliant with the MP@HL sample rates, but the receiver will have to do the format conversion on its own.

1.3 Thesis Outline

The remainder of this thesis will be primarily concerned with the transmission format selection and conversion block in the transmitter. In particular, we will focus on converting a 1080-line, progressively scanned (PS) original at 60 frames per second to interlaced scanning (IS) by subsampling. We will then investigate a way to represent the missing lines such that an advanced receiver can display formats which have a higher resolution than what MPEG-2 allows.

The thesis is organized as follows: chapter 2 describes some possible transmission formats, their probable uses, and the requirements for conversion among them; the backward-compatible migration-path concept is also explained in greater detail. Chapter 3 discusses the deinterlacing problem. Traditional deinterlacing methods are addressed and deinterlacing is investigated when used within the framework of the migration path. Chapter 4 illustrates the performance of a basic migration-path system. Finally, chapter 5 summarizes the main contributions of the thesis and proposes direction for future work.

Chapter 2

HDTV Transmission Alternatives

The HDTV system proposed by the GA and endorsed by the ACATS incorporates several possible transmission formats. The availability of multiple formats differs from the existing NTSC standard, in which only one transmission format exists: a 525-line, 59.94 field-per-second, interlaced mode. Although the December, 1996 FCC decision did not include the adoption of the formats specified in the GA system, the decision made it possible to transmit in those specified formats plus many additional ones. In other words, the FCC decision has increased the number of possible HDTV transmission formats, not restricted them. One possibility, of course, is that broadcasters will decide to use only those formats which were included in the original GA proposal. Even if this is not the case, consideration of the GA proposal will provide insight into the advantages of a system with multiple formats without having to consider the myriad of formats possible as a result of the FCC decision.

This chapter describes the formats which would have been initially available with the GA system. The different GA formats cater to many different programming applications and thus are a fairly representative set of all possible HDTV formats. Some additional formats, which do not fall into the FCC-approved specifications, will also be proposed. The incorporation of such formats through the use of the migration path will be described.

2.1 Transmission Formats

The GA proposed that six transmission formats be available for immediate use for terrestrial broadcast HDTV. Each format is described by its spatial resolution, its temporal resolution (frame rate), and its scan format.[2] The six formats are shown in Table 2.1. Also shown is the number of samples per second (uncompressed) required by each.

---

[2] In this thesis, a production, transmission, or display format is denoted by X/Y/Z, where X is the number of horizontal lines, Y is the frames (or fields) per second and Z is the scanning mode: interlaced (IS) or progressive (PS). All HDTV formats which we will discuss, unless otherwise noted, have a 16:9 aspect ratio.

Spatial Resolution	Frame Rate	Scan Format	Samples per Second
720x1280	60 frames/sec	PS	55M
720x1280	30 frames/sec	PS	28M
720x1280	24 frames/sec	PS	22M
1080x1920	30 frames/sec	PS	62M
1080x1920	24 frames/sec	PS	50M
1080x1920	60 fields/sec	IS	62M

Table 2.1 Grand Alliance Transmission Formats. These formats will be part of the standard US HDTV system.

Five of these formats are based on progressively scanned video.[3] Progressive scanning (PS) provides better video quality than interlaced scanning (IS) because it avoids interlace artifacts such as interline flicker. Progressive scanning lends itself to transcoding application more readily than interlaced scanning and also provides an easier and more natural interface with computers and telecommunication networks [22]. Nevertheless, current production and display technologies are predominantly based on interlaced scanning; therefore, having an interlaced transmission format available may be initially advantageous from an economic viewpoint. On the other hand, many groups are opposed to the possibility of HDTV broadcasts in an interlaced format and hope to eventually phase it out completely if, contrary to their wishes, it should become available in the first place.

Each of the proposed transmission modes has higher vertical and horizontal resolution than NTSC video. Each also has a 16:9 aspect ratio, compared to the 4:3 aspect ratio of NTSC. Although it is desirable to have a 60 frame-per-second mode with over 1000 lines of resolution per frame (*e.g.*, 1080/60/PS), the number of samples per second (124M) is too high for satisfactory compression and transmission with current technology. In addition, such a mode would not comply with the MPEG-2 MP@HL standard. A 1080/60/PS mode may become available only as an extension to the current HDTV standard. For its realization, not only would there need to be an improvement in technology or an increase in the bandwidth or power permitted for transmission, but also there would have to be an additional protocol established for either coding this format directly or for coding the difference between an established HDTV format and the 1080/60/PS sequence. This, as we shall see, is the role of the HDTV migration path.

The transmission format used for a particular broadcast is essentially independent of the production and the display formats. Television cameras can scan scenes in their native format,

---

[3] In *progressive scanning*, every horizontal line in a frame is scanned, sequentially from top to bottom. In *interlaced scanning*, every other horizontal line in a frame is scanned, alternating between the even lines in one frame, the odd frames in the next frame, *etc.* This is discussed in more detail in section 2.3.

that is, cameras generate video in one particular format, and receivers can display video in a completely different format. Furthermore, the transmission format used during the broadcast might be different from both of these. Although it is possible to design and build a camera which can scan scenes in multiple formats, this is certainly not necessary—the same is true for displays—and adds an extraneous cost.[4]

With the NTSC system, the native format of virtually all cameras and displays is the same as the transmission format. With the GA system, since there is more than one possible transmission format, there is less incentive for cameras and displays to conform to any particular format; In fact, a camera can scan in *any* arbitrary format, as long as the format is converted to one of the allowable transmission formats prior to encoding, and a display can convert the decoded bit stream to its format (which is also arbitrary). Furthermore, the encoder is free to switch between any allowable formats on a frame-by-frame or scene-by-scene basis.

The freedom of the transmitter to choose the format in which to broadcast does not imply that every HDTV encoder must be able to convert to more than one of the allowable formats, but having multiple formats allows the transmitter some flexibility to decide in which format to broadcast the video. The main reason for allowing multiple transmission formats is to allow for scene- or source-dependent encoding. For example, fast moving sequences such as the video of sporting events may require encoding at a high temporal rate, typically 60 frames per second (fps), to realize smooth motion. Alternatively, video which has been generated from film can be encoded at a lower temporal rate (24 fps) without loss of information, since the source itself is at this rate. Because there is essentially a constant bit-rate capacity of the 6-MHz broadcast channel, there is a tradeoff between temporal resolution and spatial resolution. That is, video at a lower frame rate can be encoded at a higher spatial resolution (number of lines) than video at a higher frame rate. Furthermore, by decreasing the temporal and spatial resolution of a signal so that only half the channel capacity is used, two pictures can be transmitted over the same 6-MHz channel.

In some situations, a particular transmission format seems ideally matched to a particular production format. For example, when an event is progressively scanned at 720 lines and 60 fps, it seems logical to use the 720/60/PS transmission format, or when the source is a 1080-line interlaced sequence, the logical transmission choice is the 1080/60/IS format; however, there is often no transmission format which is matched to the source format. For example, 1080-line PS video at 60 fps may be produced by a camera. In this case, there is no 1080/60/PS allowable transmission format and an HDTV transmitter can convert the sequence to 720/60/PS by spatial downsampling, to 1080/30/PS by temporal downsampling, or to 1080/60/IS by interlacing. Also, since the format

---

[4] In all further discussion we shall assume that cameras and displays only produce video in one format.

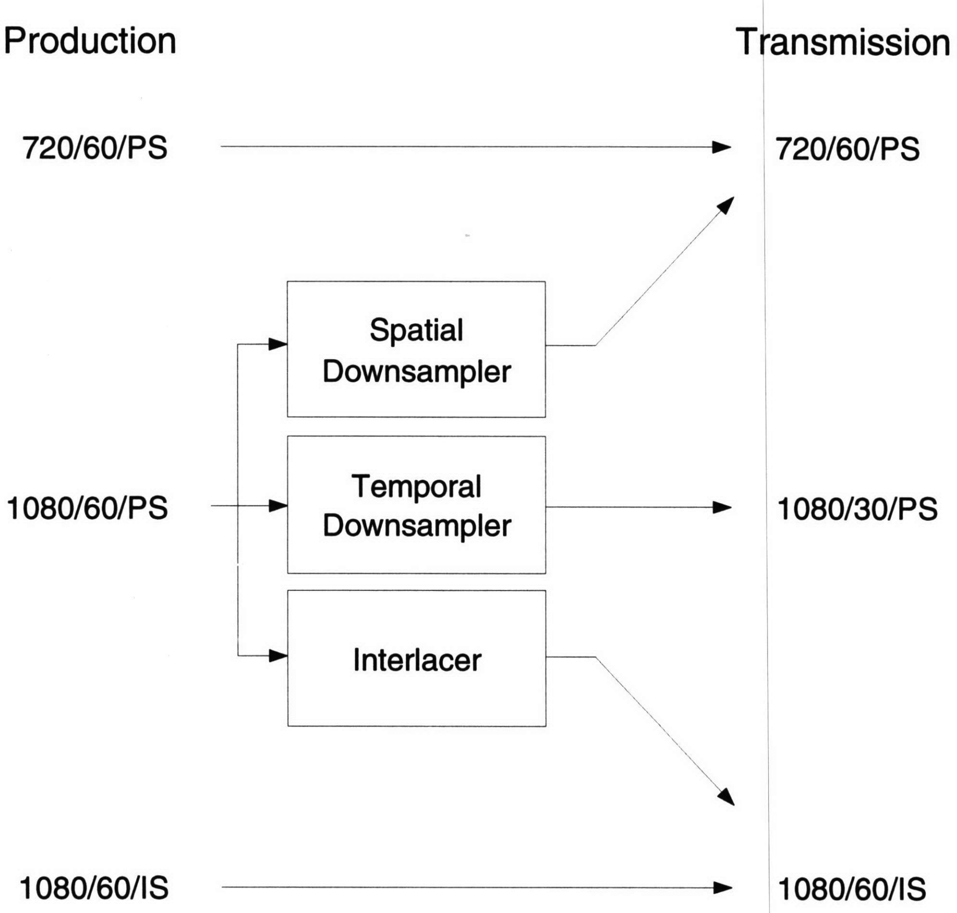


Figure 2.1 Transmission Alternatives. The production format can be converted to one of the allowable transmission formats. Three typical production formats are shown here with the signal processing required to convert to three of the six GA transmission modes. The 720/60/PS and the 1080/60/IS production formats are trivially converted to their corresponding transmission formats; however, it is possible and, in certain situations, may be desirable to convert these formats to some other allowable transmission modes.

chosen can be scene-dependent, the transmitter may switch between allowable formats during the course of a particular transmission. Some of the possible transmission alternatives just described are shown in Fig. 2.1.

Even in the first two situations above, where the transmission format choice seems clear, this may not actually be the case. Again, consider an interlaced source and a situation when progressive displays have become more widespread and essentially universal. In this case, if the IS format is transmitted, all receivers will have to convert from IS to PS in order to display the signal. The cost of deinterlacing will lie in the receivers, and better deinterlacers will cost more for the consumer. Alternatively, if the conversion from IS to PS is performed at the transmitter, then a sophisticated deinterlacer can be used without affecting the price of consumer television receivers. Also, as deinterlacing technology improves with time, *all* televisions would be capable of displaying pictures of equivalent resolution, since the deinterlacing is performed before transmission. Otherwise, televisions would display pictures at different qualities, depending on when they were built with respect to the course of deinterlacing evolution. A second situation where it may be desirable to convert from an allowable format to a less natural choice is when a 1080/30/PS source is converted to 60 fps for transmission. This may seem unlikely since any increase in the frame rate could be done at the receiver and because an increase in the temporal rate would come at a cost of decreasing the spatial resolution to 720 lines; however, it might be desirable in the case where a 30-Hz frame rate is not sufficient to prevent motion jitter but sophisticated temporal interpolation can be performed successfully at the transmitter. It is also possible to interpolate prior to transmission (*i.e.*, not in real time). Either way, the cost of interpolation is incurred only once, at the transmitter, instead of requiring complicated, real-time interpolators in all receivers. A third possibility for conversion to a less natural format arises when it is desirable to multiplex two signals on the same channel. For example, a 1080/30/PS source format can be transmitted, or it may be converted to the 720/30/PS format and combined with another 720/30/PS picture. These two pictures could then be transmitted in the same bandwidth required to transmit the original 1080/30/PS video.

If our source is not one of the available transmission formats—1080/60/PS material, for example—it must be converted to one. Each of the six proposed GA formats has its advantages which makes it particularly useful. The 720/60/PS mode is important because it maximizes the spatial resolution given the 60 Hz frame rate. This mode is likely to be used when there is fast motion in the picture sequence and it is necessary to transmit at the highest possible frame rate to yield smooth motion and avoid jerkiness. The 1080/30/PS mode is also important because it maximizes the frame rate given the 1080-line spatial resolution requirement. This mode is likely to be used when a 30-Hz frame rate is adequate for smooth motion. In this case, it is desirable to broadcast as many lines of video as possible. The 1080/60/IS format will be important initially, when interlaced scanning plays a significant role in production and display formats. The movement of camera and display technology toward progressive scanning is likely to drastically reduce the popularity of interlaced scanning and the 1080/60/IS mode. The 720/30/PS mode will be useful in situations where the spatial resolution of the 1080/30/PS can be sacrificed in return for the ability to transmit multiple sequences over a single channel as previously mentioned. Since the 720/30/PS mode has less than half the samples per frame of the 1080/30/PS mode, it is possible to multiplex two signals at the lower resolution into the bandwidth which would otherwise be used for one signal at the higher resolution. This can effectively double the number of programs available to viewers. The 720/30/PS mode will also be useful for scenes which are difficult to code at the higher resolution of 1080/30/PS, since twice as many bits per pixel will be available for coding.

A 1080/60/PS source will likely be converted into one of the four transmission modes discussed above. The other two transmission modes are more apt to be used when the source is 24 fps film. The 1080/24/PS mode is the probable alternative for transmission in this case, since conversion to a higher frame rate for transmission would increase the data requirement but not increase the resolution. The 720/24/PS mode complements the 1080/24/PS mode in the same manner that the 720/30/PS mode complements the 1080/30/PS mode: it allows for two sequences to be transmitted over the same 6-MHz channel or provides additional bits for scenes which are difficult to code.

At the receiver end of the process, there is less flexibility than on the production and encoding side. Whereas the transmitter is free to decide which format to use for encoding, the receiver has no such freedom and must be able to handle any of the transmission formats or fail to receive certain broadcasts. Assuming that each receiver will only be capable of displaying one format, a decoder must be able to convert from any of the possible transmission formats to its display format. This is required so that every receiver is able to display any broadcast. An example of the processing that needs to be done for three different types of displays and three of the transmission formats is shown in Fig. 2.2. Each decoder should have the capability of converting from any of the possible transmission formats to its display format, which need not be one of the transmission formats. The cost of a display will be affected by the complexity of the conversion between these formats. Receiver costs can be kept down by using simple conversion algorithms for transmission formats which are not frequently used (*e.g.*, the interlaced format when it is being phased out.)

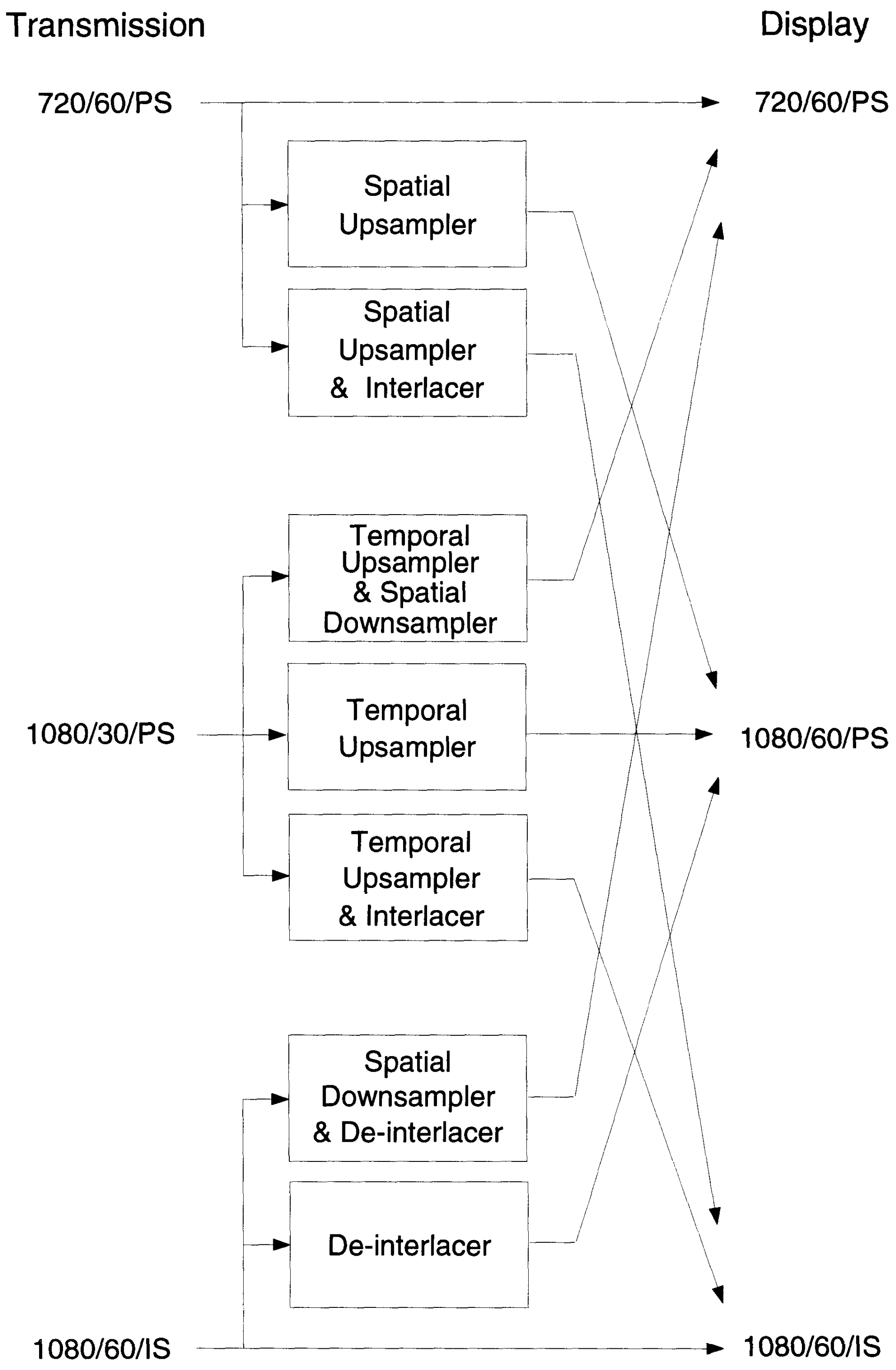


Figure 2.2 Receiver Signal Processing. Each receiver must be able to convert any of the six possible transmission formats (three are shown here) to its display format. Note that if the 1080/60/IS mode is phased out, receivers will not need to have deinterlacers, removing a potentially computation-intensive task and, thereby, reducing the cost.

2.2 Migration Path

As HDTV evolves, it can be expected that additional bits will become available for picture coding. These additional bits may come about in several ways. The improvement of image compression techniques may reduce the number of bits necessary to encode an image. Similarly, advances in modulation techniques may increase the number of bits capable of being reliably transmitted over a 6-MHz channel, or in areas closer to the transmitter the higher signal-to-noise ratio may be exploited [28]. Alternatively, the FCC may allocate additional bandwidth to the broadcasters.[5] For example, after the NTSC system is phased out, the FCC may allow broadcasters to transmit over portions of the channels previously used for NTSC. The FCC may also allow transmitters to broadcast at a higher power after the NTSC stations are gone. The increase in signal-to-noise ratio can be traded off for a higher bit capacity. These last two situations are speculative and rely on FCC actions, but if they do arise, they will yield a significant increase in the available bit rate. Finally, over media such as wire or optical fiber, it is likely that a 6-MHz channel can support a higher bit rate than a terrestrial channel. Furthermore, cable (and perhaps satellite) service providers may not be limited to a 6-MHz channel.

Additional bits can be readily incorporated into the compression algorithms for any of the valid formats. By arbitrarily reducing the quantization step size, for example, the available bits can be used to minimize the difference between the coded picture and the original. At some point in this process, the two pictures might become perceptually indistinguishable. From this point on, reducing the quantization error does not improve picture quality. If all the additional bits have not been used, then some will have been wasted.

Another way to use additional bits is to add to the list of available transmission formats. Specifically, supplementary bits may support additional formats with higher resolution, such as those shown in bold in Table 2.2. As production technology advances, it may be possible to transmit higher resolution pictures at 60 fps. Of course, transmission in these enhanced formats must be compatible with the original formats so all receivers can decode them: since these enhanced formats are not MPEG-2 compliant, they cannot be simply encoded using the FCC-approved HDTV standard. Instead, a method for migrating to the new formats in a backward-compatible way must be implemented. This migration path from standard HDTV to enhanced HDTV is illustrated in Figs. 2.3 and 2.4.

The standard HDTV system will operate as shown in Fig. 2.3. In this example input video is generated in the 1080/60/PS format. Since this is not an allowable transmission format as

---

[5] Because of the high demand for spectrum allocation, this might be somewhat of a pipe dream.

Spatial Resolution	Frame Rate	Scan Format
720×1280	60 frames/sec	PS
720×1280	30 frames/sec	PS
720×1280	24 frames/sec	PS
1080×1920	30 frames/sec	PS
1080×1920	24 frames/sec	PS
1080×1920	60 fields/sec	IS
1080×1920	**60 frames/sec**	**PS**
1440×2560	**60 frames/sec**	**PS**

Table 2.2 Available Transmission Formats with Enhancements. These are some of the transmission formats which may be available with advanced HDTV systems. The two formats in bold are examples of the many possible enhanced-resolution options which may be available.

shown in Table 2.1, it must be converted. In this case the 1080/60/IS mode is selected and the "standard" video bits are broadcast over the channel. A receiver will then decode the 1080/60/IS signal and convert it to the appropriate display format. As stated earlier, the display format is not necessarily linked to the transmission format. Three possible display formats are identified in Fig. 2.3. Although the 1080/60/PS display format is possible, it will not contain more information than the transmitted interlaced signal.

An enhanced HDTV system may operate as shown in Fig. 2.4. If the elements in the dashed boxes are ignored, what remains is the standard HDTV system discussed above. A standard receiver can decode the standard video bits and use them to display a picture. The transmitter, in addition to sending the standard bits will also broadcast some video enhancement information. These bits may be generated from two mechanisms: an adaptive filter and a residual encoder. First, a copy of the standard video decoder in the transmitter decodes the "standard" bit stream. The decoded sequence is then converted back to the original "enhanced" transmission format (in this case, 1080/60/PS) by an appropriate signal processing module. As an example, the system in Fig. 2.4 requires a deinterlacer to convert from the transmitted (and decoded) 1080/60/IS format to the original 1080/60/PS enhanced transmission format. The deinterlacer need not be a fixed processor: the parameters of the deinterlacer can be determined adaptively in order to minimize the difference between the original and the reconstructed pictures, since the transmitter has access to both. The deinterlacing parameters can then be transmitted. This deinterlacing module is the first source of enhancement bits. Even in the absence of other enhancement information, the deinterlacing parameters may be used by an advanced receiver to maximize the quality of an interpolated image.

The second source of enhancement information is a residual encoder which encodes the

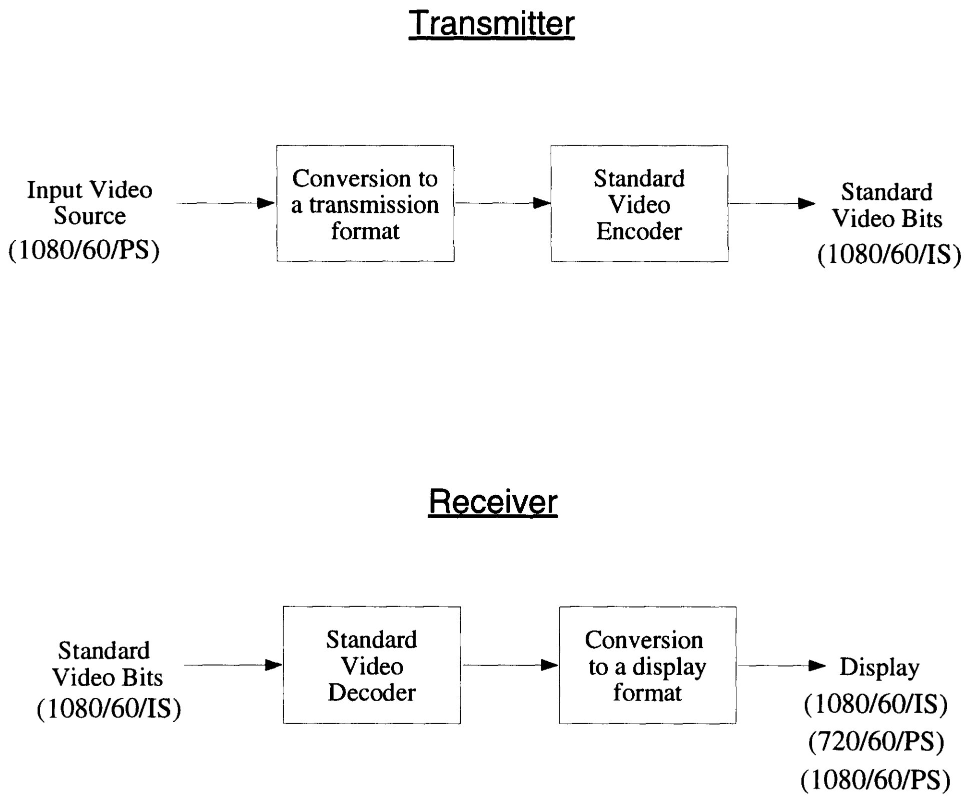


Figure 2.3 Standard HDTV System. The input video source is converted to an allowable transmission format by the video encoder. The standard video receiver, which can interpret any of the allowable transmission formats, decodes the received bit stream and converts it to an appropriate display format. This display format is receiver-dependent.

difference between the original 1080/60/PS sequence and the interpolated sequence. All the enhancement bits are then transmitted as side information. Receivers have access to these enhancement bits as well as to the standard video. Because the enhancement bits are separate, a standard video decoder can ignore the enhancement information and just produce an image equivalent in resolution to the 1080/60/IS transmission standard. Advanced HDTV receivers can use the enhancement information to improve upon the resolution of the signal. For example, a simple advanced receiver can use the information about the deinterlacer to upsample the received signal in the best possible way (as determined by the encoder). A more complex receiver could use the same information in addition to the encoded residual information to produce an even higher fidelity reconstruction.

In the situation when the 1080/60/PS original is converted to 1080/30/PS for transmission, the interlacer and deinterlacer in Fig. 2.4 would be replaced by a temporal downsampler and temporal upsampler. When the 720/60/PS format is used, a spatial decimator and spatial interpolator would be substituted.

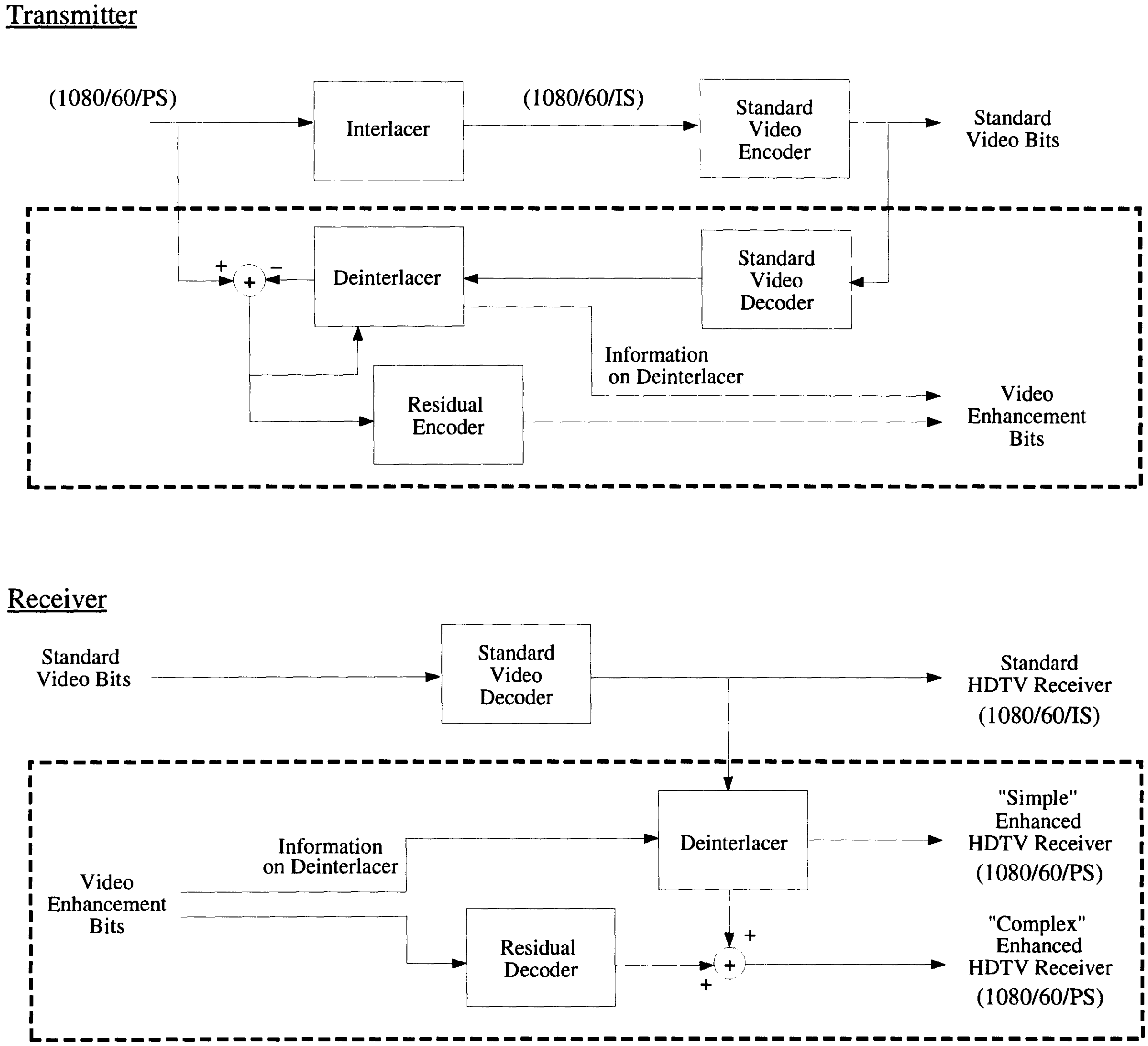


Figure 2.4 Enhanced HDTV System. The input video is converted to one of the allowable transmission formats and encoded with the standard HDTV encoder. The transmitter then mimics the decoder and can interpolate the decoded signal in some intelligent way. Finally, the difference between the original and the interpolated video can be encoded with a residual encoder. Both the filter information and the encoded residual can be transmitted along with the standard video bits. Standard receivers will decode the standard bits, but advanced receivers can use the filter information and the encoded residual to produce enhanced resolution video.

A major advantage of the migration-path system is its flexibility. When relatively few bits are available, only the information about the deinterlacer or upsampler might be transmitted, whereas if many more additional bits are available, the encoded residual can be sent. Inexpensive receivers could be built which only use the interpolation information. More sophisticated (and, therefore, more expensive) receivers could also be built which use the filter information as well as the coded residual. Furthermore, this idea of flexible migration could be extended to future generations of HDTV systems, as shown in Fig. 2.5. In this example, the input video is at some "super-enhanced" resolution, which is converted into the enhanced resolution image and then into the lower-resolution standard video bit stream. As with the enhanced encoder in Fig. 2.4, the standard decoded picture is compared to the enhanced resolution original and video enhancement bits are generated. This image now corresponds to the enhanced HDTV resolution. To yield a higher resolution, this enhanced HDTV image is interpolated and subtracted from the super-enhanced resolution original, and a second set of video enhancement bits is generated. At the receiver, the decoder can use as many levels of the enhancement bits as it is capable of translating in order to display standard, enhanced, or super-enhanced resolution video. This backward-compatible migration can be implemented repeatedly, as future generations of HDTV evolve.

2.3 Format Conversion

As mentioned in the previous section, conversion between transmission formats and production or display formats is an unavoidable aspect of the GA HDTV system. In the situation where the source is of the 1080/60/PS format, conversion to the 720/60/PS, 1080/30/PS, or 1080/60/IS format for standard encoding is likely. Conversion to the 720-line format is generally performed using spatial anti-aliasing filtering and is straightforward. Conversion to the other two formats is the focus of this section.

To compare the conversion of 1080/60/PS video to each of these two formats, consider Fig. 2.6, which represents the 1080/60/PS format. Figure 2.6(a) depicts two consecutive frames of progressively scanned video: horizontal lines are scanned at the same vertical location at each time sample, and the vertical sampling period is denoted by T_V. The frames are located $T_t = \frac{1}{60}$ seconds apart temporally. Figure 2.6(b) shows the sampling locations in the vertical/temporal plane, and Fig. 2.6(c) shows the associated frequency domain aliasing grid with a hypothetical frequency spectrum superimposed.

Figure 2.7 illustrates the 1080/30/PS format which is similar to the 1080/60/PS format; the difference is the slower temporal sampling rate and, therefore, lower temporal resolution. As

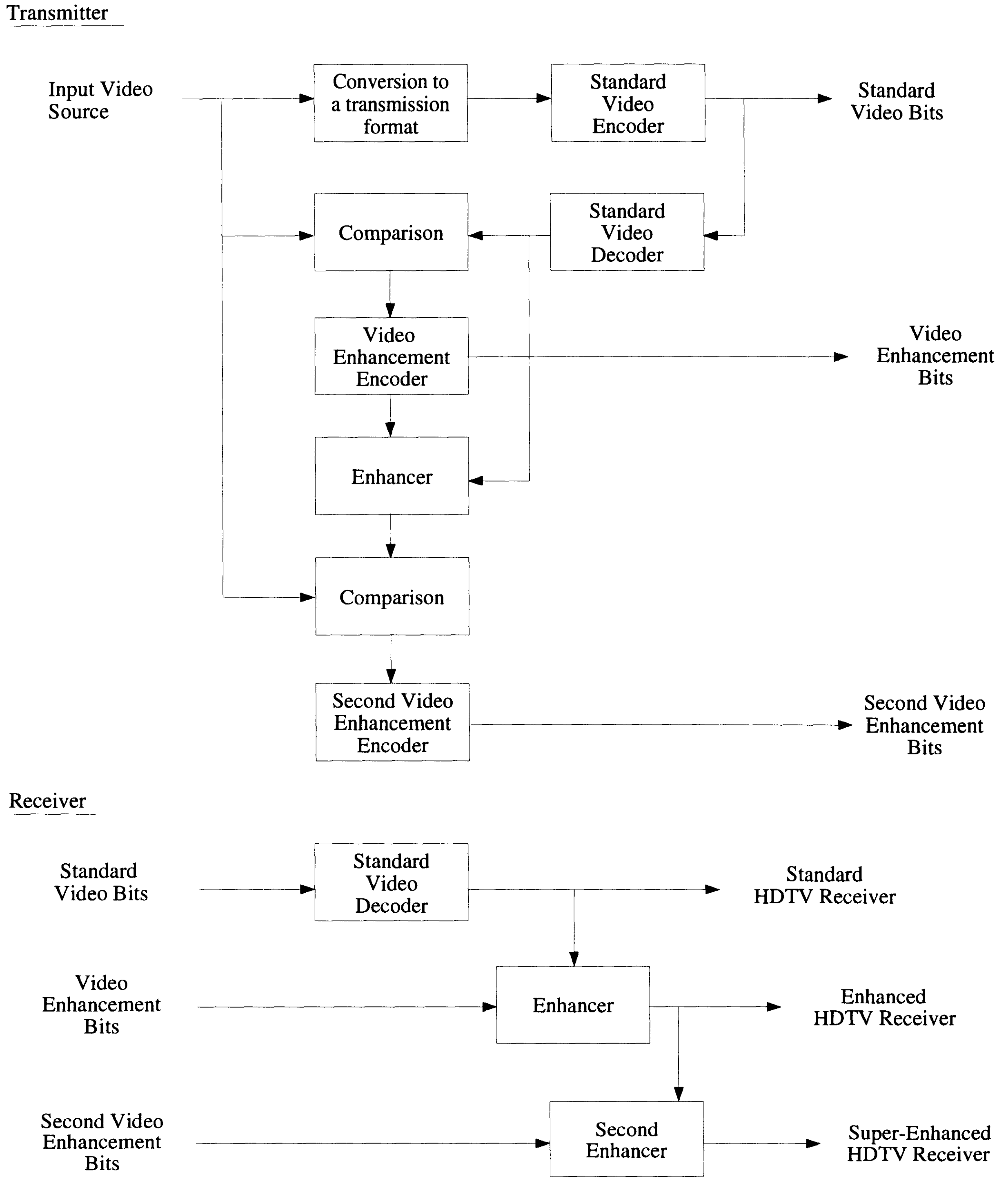
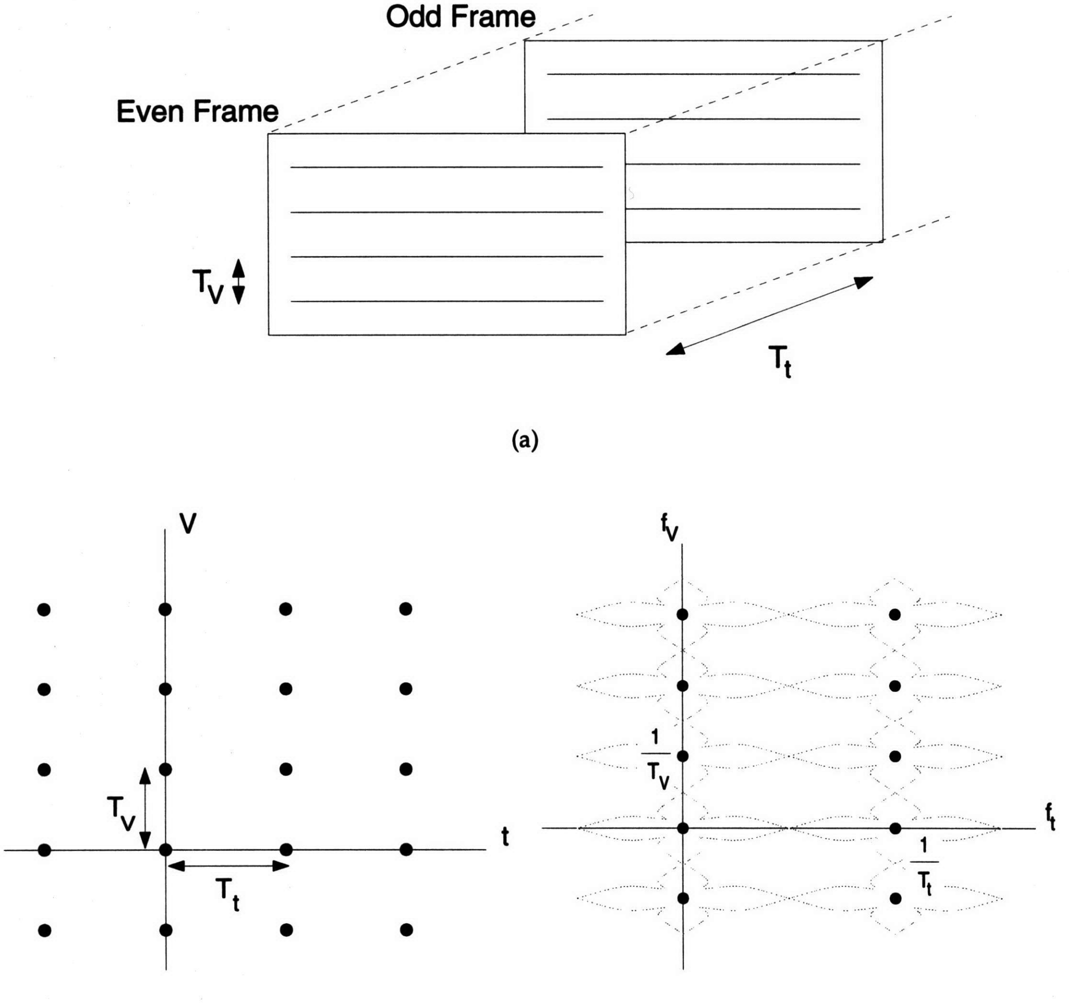


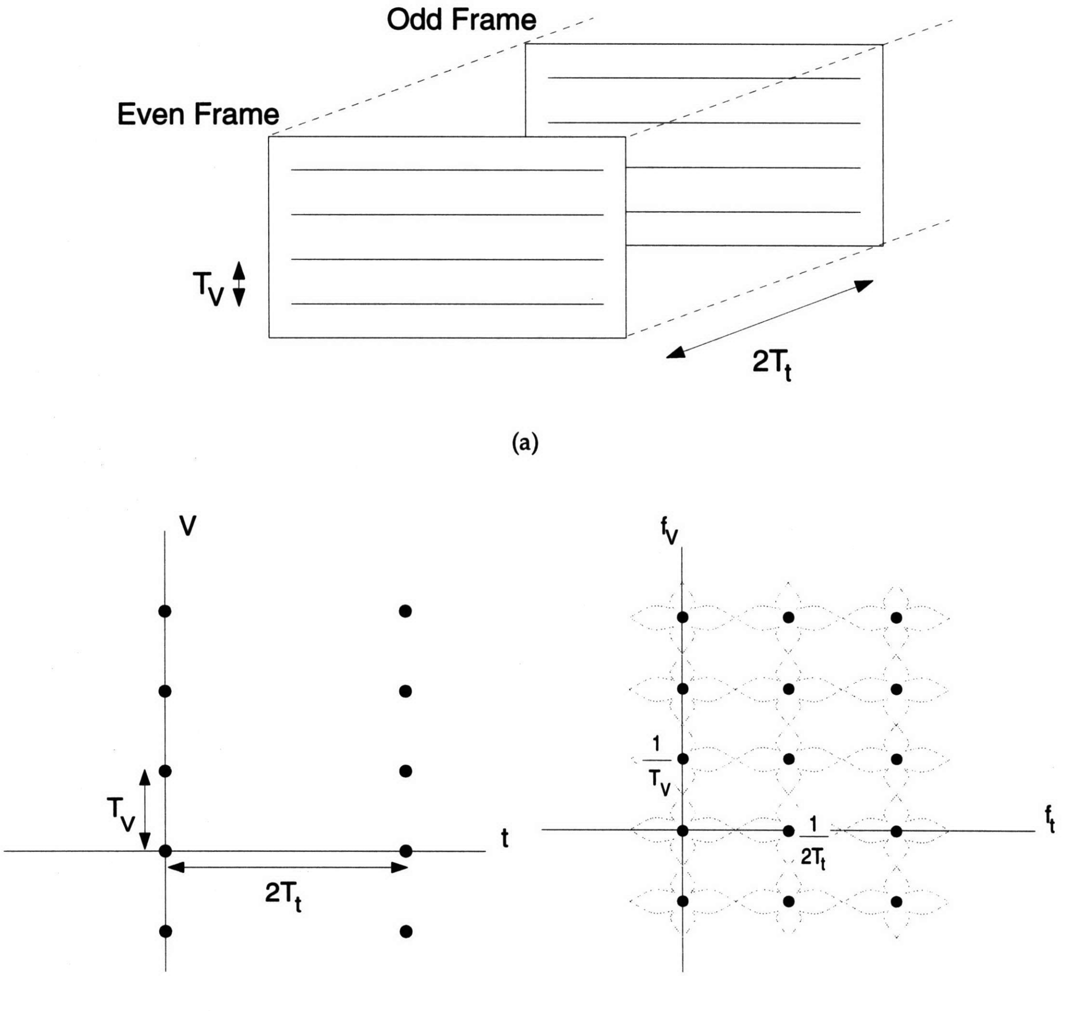
Figure 2.5 Future Advanced HDTV System. The migration of HDTV will incorporate multiple levels of enhancement, each increasing the resolution of the transmitted video.



(a)

(b) (c)

Figure 2.6 60-fps Progressive Scanning. Horizontal lines in each frame are scanned (a) at the same vertical position. The vertical/temporal sampling grid (b) is rectangular, as is the associated frequency domain aliasing grid (c). Adjacent frames are $\frac{1}{60}$ seconds apart.

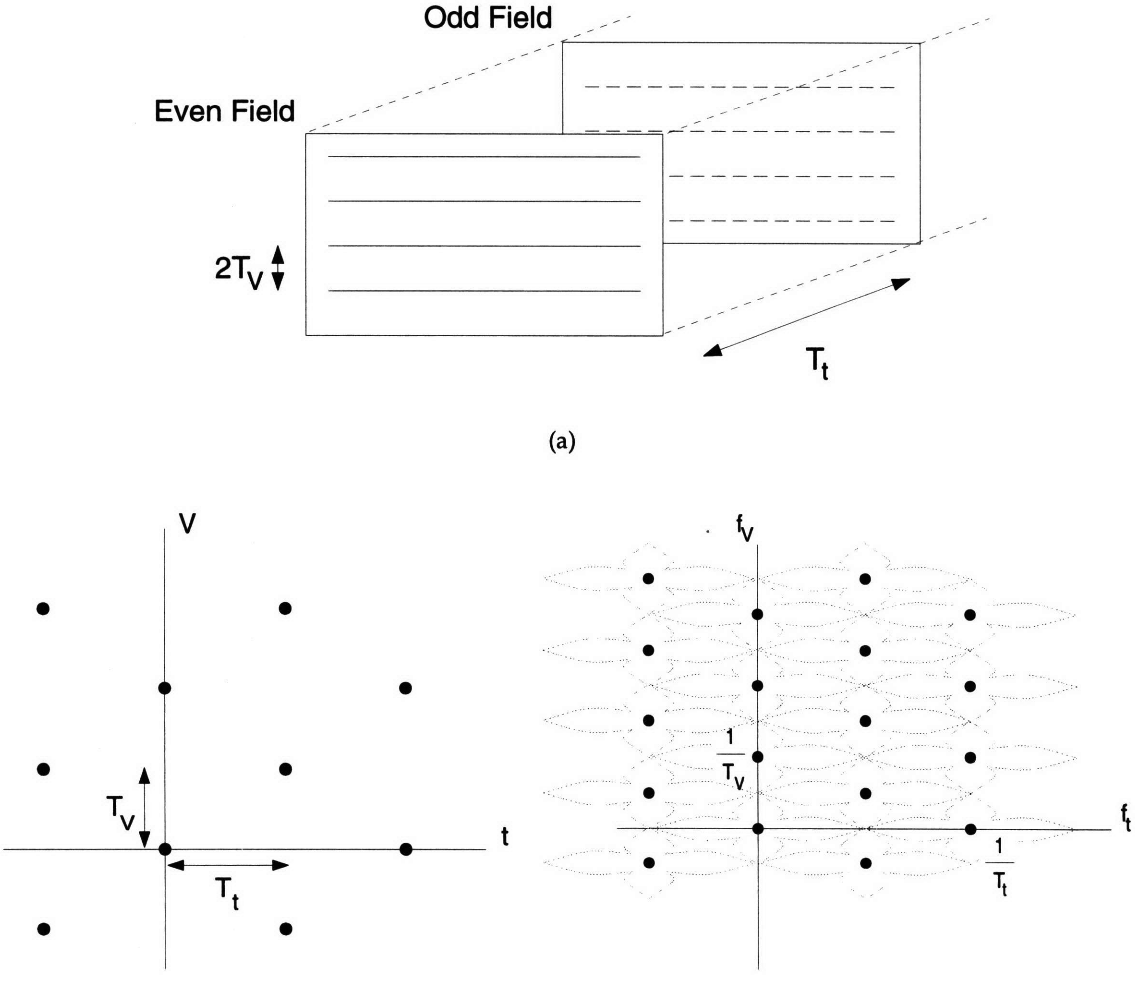


(a)

(b)

(c)

Figure 2.7 30-fps Progressive Scanning. Horizontal lines in each frame are scanned (a) at the same vertical position. The vertical/temporal sampling grid (b) is rectangular, as is the associated frequency domain aliasing grid (c). Adjacent frames are $\frac{1}{30}$ seconds apart.



(a)

(b)

(c)

Figure 2.8 Interlaced Scanning. Horizontal lines in adjacent frames are scanned (a) at the alternating vertical positions. The vertical/temporal sampling grid (b) is quincunx, as is the associated frequency domain aliasing grid (c).

shown in Fig. 2.7(a) the progressively scanned video frames are now located $2T_t = \frac{1}{30}$ seconds apart. Figure 2.7(b) shows the sampling locations in the vertical/temporal plane, and Fig. 2.7(c) shows the associated frequency domain aliasing grid. Comparing the aliasing grid in Fig. 2.7(c) with that in Fig. 2.6(c), we see that, as expected, the spectrum of the 30-fps video is half as wide temporally; therefore, temporal aliasing occurs if the signal is not bandlimited properly.

Conversion from 1080/60/PS to 1080/30/PS can be performed using linear filtering and 2:1 subsampling. The anti-aliasing filter required[6] is a lowpass filter in the temporal direction. This is straightforward to implement, since the filter is one dimensional. Conversion back to the 60-fps format is also easily implemented using zero-insertion and lowpass filtering. The problem with this approach toward format conversion is that the filtering causes too much blurring. If one visualizes what happens at a scene-change, one realizes that the blending of images occurs. Also, moving objects leave behind residuals as they travel from one region of the screen to another, creating "streakiness" or image echoes in the video. Therefore, one concludes that the format conversion should be done in a different way; preferably a manner which will retain the high-frequency temporal characteristics of the video stream. One way of accomplishing this is to do little or no temporal filtering prior to subsampling and to tolerate temporal aliasing. If it becomes necessary or desirable to temporally upsample the sequence, motion-compensated estimation and prediction is one type of processing which can potentially accomplish the goal of reconstructing the high temporal frequencies.

Figure 2.8 shows the 1080/60/IS format. Due to interlacing, the time, space, and frequency characteristics of this format are more complicated than those of the two progressively scanned formats. As shown in Fig. 2.8(a), the vertical locations of the scanned lines alternate between fields. The vertical/temporal sampling grid for the interlaced signal is shown in Fig. 2.8(b). Two attributes of this quincunx lattice are noteworthy. First, with adjacent frames of the 1080/60/PS video occuring at $T_t = \frac{1}{60}$ seconds apart, each horizontal line is scanned every $\frac{1}{30}$ seconds. Second, due to interlacing, the vertical distance between horizontal lines in the same field is $2T_V$, since adjacent lines are not scanned in the same field. A two-dimensional frequency lattice (with frequency aliasing for a sample spectrum shown) due to interlaced scanning is provided in Fig. 2.8(c) [29] [7].[7]

These differences have some important ramifications. First, pictures with low vertical detail can have approximately twice the temporal resolution with interlaced scanning. Second, the 30-fps format can support higher vertical frequency components in conjuction with its higher temporal

---

[6] We assume that aliasing is an undesirable product of format conversion, however, it should be noted that preventing aliasing comes at the price of blurring the video; thus, some aliasing might be tolerable if it improves sharpness in the downsampled picture.

[7] There are other shapes (especially rectangular) which can be used to demonstrate this frequency aliasing pattern. The rhombus has the desirable property that for low temporal frequencies, the vertical resolution can be maximized.

frequencies. Third, with IS there is an ambiguity between frequency components containing high vertical frequencies with low temporal frequencies and components containing low vertical frequencies with high temporal frequencies since these frequency combinations alias to the same point in the two-dimensional spectrum; *i.e.*, $H(\frac{1}{2T_V}, 0) = H(0, \frac{1}{2T_t})$ after sampling. This can be seen by comparing the sampled (interlaced) version of a sequence which alternates between black and white frames at the field rate with the sampled version of a still image whose horizontal lines alternate between black and white. The two interlaced signals are identical.

Conversion from 1080/60/PS to 1080/60/IS can also be performed using linear filtering and subsampling. As stated above and shown in Fig. 2.9(a), one lowpass filter which prevents frequency aliasing is shaped like a rhombus in the vertical/temporal plane. This filter is more complex than the lowpass filter needed for conversion to 1080/30/PS because it is two-dimensional and is inherently non-separable. Alternatives to using this ideal lowpass filter include separable approximations of this filter. These filters tend to attenuate the high vertical and temporal frequencies, softening (blurring) the picture in both dimensions. Other alternatives to this filter are the rectangular anti-aliasing filters shown in Fig. 2.9(b) and Fig. 2.9(c). These filters preserve high frequencies in one dimension at the expense of high frequencies in the other dimension. When filtering is done to convert the interlaced picture back to 1080/60/PS, some aspect of image sharpness is gone. In order to preserve detail in space and time, format conversion should not be performed using linear lowpass filters. For example, aliasing can be allowed to occur during the interlacing process if motion-compensated prediction and estimation have the potential to recapture the sharpness and high detail when used in format conversion back to the higher sampling rate. In this thesis, conversion to 1080/60/IS will be performed without the associated lowpass filtering. Chapter 3 will further address the conversion from 1080/60/IS back to 1080/60/PS.

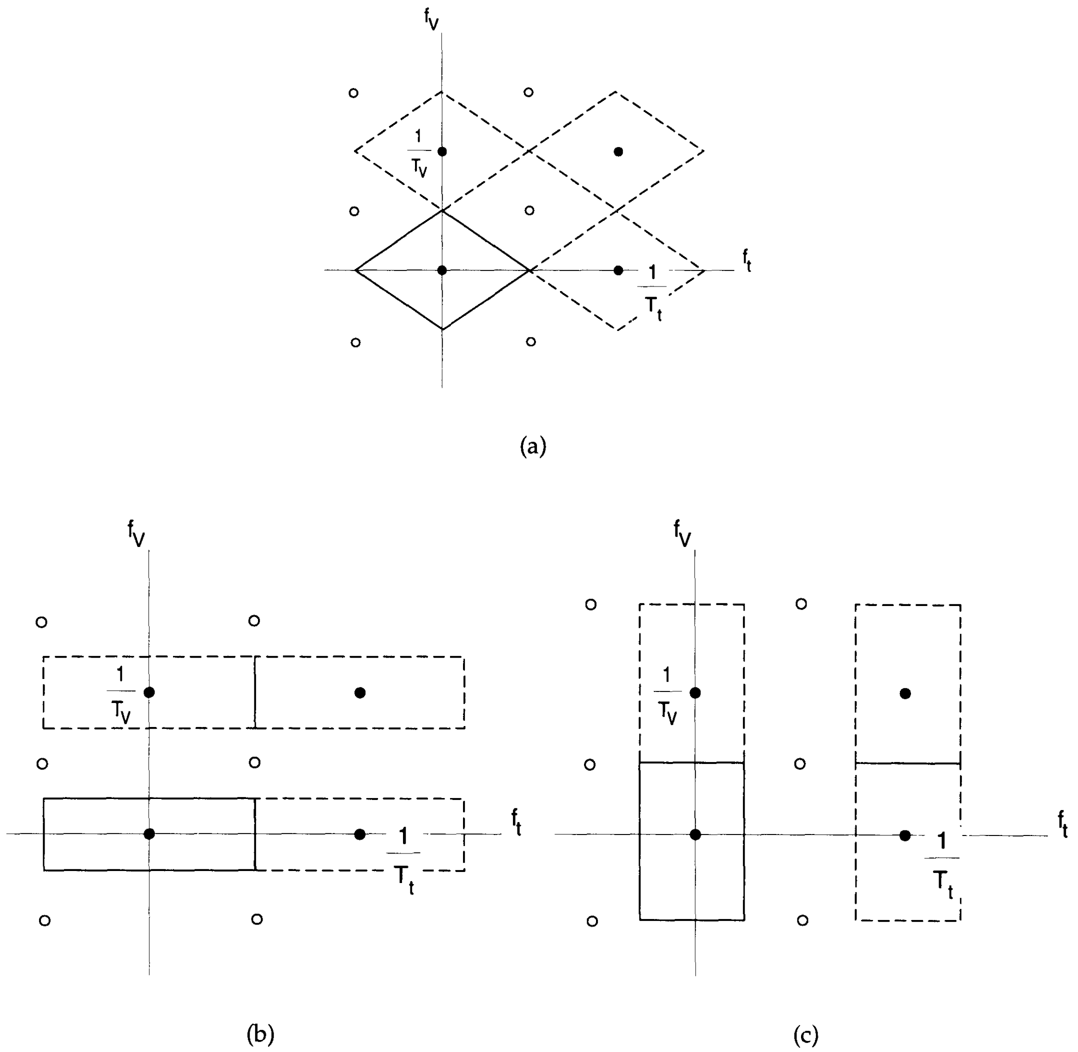


Figure 2.9 Anti-aliasing Regions for Interlaced Scanning. Three choices for the passband of ideal anti-aliasing filters for use prior to interlacing. The regions bounded by solid lines correspond to the baseband region with its periodic replication in the discrete two-dimensional frequency plane denoted by dashed lines. The open circles correspond to the locations of the replicated spectrum after interlacing.

Chapter 3

Deinterlacing

The well-studied problem of deinterlacing dates back as long as the current NTSC system has been in use. Interlaced scanning was introduced as part of the NTSC standard to reduce the signal bandwidth requirement while maintaining the vertical sampling interval, allowing smooth motion rendition while eliminating the flicker associated with a slower frame rate. Because cameras and televisions were both built using interlaced scanning, deinterlacing was not needed. The realization that progressively scanned video avoids some of the artifacts of interlaced-scanned pictures, coupled with the advancing technology of cameras and displays, has heightened recent interest in deinterlacing.

Deinterlacing methods can vary in complexity. One common aspect to all traditional methods is that they do not make use of a progressively scanned original. In many cases, this is not possible because no progressive sequence exists; that is, the source material is interlaced. Another common feature is that deinterlacing is performed using only the information from the video signal, with no assistance from additional side information. As shall be shown, deinterlacing under the migration path concept deviates from traditional methods in these two respects.

3.1 Traditional Deinterlacing Methods

Traditional deinterlacing methods make no use of side information or of a progressive original; the only information they use is the interlaced video itself. These deinterlacers are inherently of two classes: intraframe and interframe. Intraframe deinterlacers use only information in the current field for interpolation. Interframe deinterlacers are characterized by the ability to use previous and, perhaps, subsequent fields to interpolate the current field.

3.1.1 Intraframe methods

Intraframe deinterlacers are essentially vertical upsamplers (or, as is the case with 2:1 interlacing, line doublers). Several simple methods exist for performing line doubling, including line repetition[8] and linear interpolation from adjacent lines. The drawback with these and other linear filtering approaches is the limited possible resolution. In order to prevent aliasing when sampling, the original (progressive) sequence must be bandlimited to $\frac{1}{4T_V}$ in the vertical direction. The resulting lowpass sequence represents the best resolution possible with linear filtering, which is unable to differentiate between aliased frequencies. Furthermore, if the line doubling is performed using a crude filter (for example, line repetition), there will be frequency components in the upsampled images which are improperly introduced, producing jagged edges. Because vertical filtering to prevent aliasing essentially yields half the number of lines of resolution, HDTV-quality video cannot be preserved. A conclusion we may draw is that intraframe deinterlacing must use nonlinear processing to maintain HDTV quality.

Nonlinear methods of line doubling, while more complicated to implement than linear filtering, can often yield better perceptual results. Because these methods are not tied to the frequency spectrum in a purely multiplicative way, some frequency components which are apparently aliased by the downsampling (interlacing) process may be reconstructed. This can allow the proper rendition of sharp diagonal edges, for example. Nonlinear methods hold the promise that interlacing followed by deinterlacing can be done without vertical bandlimiting, yielding a higher-resolution result. In [26], a low-complexity line doubling algorithm is described which performs adequately in many applications.

3.1.2 Interframe methods

A deinterlacing system based solely on intraframe methods has the desirable feature that no buffering of previous or subsequent frames is needed. Therefore, the cost of frame memory is reduced. However, due to the high temporal redundancy in most image sequences, interframe deinterlacers provide a significantly better opportunity to deinterlace HDTV-quality images. As an example, consider a finely detailed, stationary scene: it is likely that intraframe methods, even fairly complex, nonlinear ones, will not be able to properly provide an accurate, sharp, interpolated image. However, the simple interframe method consisting of field repetition will do a *perfect* job of reconstruction. It is for reasons such as this that interframe processing is highly desirable as part of a deinterlacing system designed to produce truly high-definition pictures.

Three-dimensional linear filtering is one fairly straightforward interframe method for deinterlacing, but it has its deficiencies. As stated in section 2.3, temporal linear filtering causes blurring

---

[8] The term *line repetition* is used to describe the copying of an adjacent line from the same frame (or time sample). The terms *field repetition* and *frame repetition* are used interchangeably to describe the copying of a line from an adjacent field or frame at the same vertical location.

and streaking between frames. In addition, the storage of several frames is required to prevent temporal aliasing.

More effective approaches to deinterlacing often involve nonlinear processing. For example, methods such as three-dimensional median filtering are relatively simple to implement. Alternatively, many algorithms incorporate a variation of motion detection, estimation, and/or compensation. Due to the normally high correlation between consecutive frames, effective motion-compensated interpolation can yield highly accurate field reproduction. Consider again the highly detailed, stationary scene: if the motion is correctly estimated, perfect reconstruction can be performed.

The simplest of these motion-adaptive algorithms involves attempting to detect the presence of motion at a given spatial location. Areas which are stationary from frame to frame are deinterlaced using field repetition, while moving areas use spatial interpolation techniques. Slightly more complex estimation methods, which only use one or two adjacent frames often consist of matching pixels of the current frame with pixels from abutting frames while attempting to determine a motion vector. Elaborate motion estimation procedures using multiple preceding and succeeding fields are, of course, also possible. These block-matching algorithms can outperform linear interpolation methods, but have their limitations. One major shortcoming is the increased computational complexity required by motion-estimated processing. Another shortfall—one shared by all traditional deinterlacers—is that interpolation is performed using only the interlaced lines. That is, lines in the original sequence which are removed during the interlacing process are unused while deinterlacing. Thus, motion compensation is performed using estimates of lines that are to be reconstructed.

Other variations on deinterlacing can be found in the literature. For example, [36] proposes a recursive method for deinterlacing; that is, a method which uses the previously deinterlaced frame to predict the current frame. This method has certain theoretical advantages, but in practice only yields small benefits at the expense of requiring complete knowledge of the past. This might be impractical for bootstrapping in the middle of a telecast.

3.2 Migration Path Deinterlacing

As described earlier, the goal of the HDTV migration path is to allow for increased-resolution transmission formats while maintaining backward-compatibility with the original formats. There are several advantages to be gained by implementing deinterlacing as part of an HDTV migration path system. One benefit is that deinterlacing can be executed using knowledge of the source at full resolution. Because the progressive source may be available, motion compensation and other deinterlacing processes can be compared in the context of how well they predict the original sequence. Traditional deinterlacing only enables the comparison of different methods for reconstructing an estimate of the missing lines. Another advantage to migration path deinterlacing is that most of the computationally intensive processing, including motion estimation and the comparison of different deinterlacing methods (modes) can be done exclusively at the transmitter. This provides two major benefits. First, most of the expense of deinterlacing in advanced HDTV systems may be centralized at the transmitter rather than being distributed among the receivers. Therefore, the cost to consumers is reduced. Second, sophisticated processing can be performed at the transmitter, allowing all receivers to display the enhanced HDTV images equally well. The idea of supplying enhancement information may also be useful in other cases that are not for the migration path; for example, to help convert existing NTSC or SDTV-quality video for progressively scanned displays.

The expense incurred by the migration path is an increased digital data requirement. Extra bits of information are required to describe to receivers how to perform the deinterlacing. If more sophisticated processing is done at the transmitter, more bits may be required.

The remainder of this thesis investigates a practical method for adequately performing deinterlacing in accordance with the "migration path" concept. Desired aspects of the involved processing include:

- A low-cost receiver structure, allowing only for relatively simple processing and minimal frame storage requirement.
- A small to moderate augmentation bit stream, disallowing, fore example, the coding and transmission of DCT coefficients which would generally require many additional bits to accurately code.

If the cost of receivers can be minimized, consumers will be more inclined to purchase new television sets. As previously stated, shifting the processing to the transmitter will allow all receivers to benefit from the advanced signal processing. By restricting ourselves from transmitting DCT coefficients as enhancement data, we will see the improvements which can be gained when the allowable amount of side information is small. Furthermore, since the processing of the DCT coefficents would require more extensive decoding, restriction from using them may help to lower the receiver cost.

3.2.1 Block Processing

Processing the interlaced signal on a block-by-block basis is a characteristic common to both traditional and migration-path systems. While this is not absolutely necessary, it does provide a relatively simple, yet adequate way of partitioning an image into smaller regions. Partitioning images into rectangular regions is a common image processing method, exemplified by the MPEG and JPEG compression standards in which frequency domain analysis and motion estimation (for MPEG) are performed on individual blocks. Partitioning enables one to take advantage of the locality of image characteristics. For example, adjacent pixels often exhibit similar movement (displacement) from frame to frame. In the context of deinterlacing, there is reason to expect a great deal of correlation between interpolation modes of nearby pixels.

Partitioning is most simply implemented using fixed-size regions, always dividing each frame into 8 × 8 square blocks, for example. This approach has the advantage that no information is required to specify the block size. As a result, all allocated bits can be used for coding the individual blocks' contents. If the information to be coded for each block determines which deinterlacing mode to use, then on average, there will be a constant bit requirement for each frame; the additional bandwidth required for side information will be approximately constant. If the available bandwidth is greater than necessary, it will go unused; if it is less, then some information will have to be omitted from the transmission.

A more flexible method allows variable-size blocks in a hierarchical scheme. In such an approach, we start with large blocks and determine the best way of deinterlacing each. Then, we subdivide those blocks for which the deinterlacing does not provide adequate performance as determined by some error criterion such as mean-square error (MSE). We can continue this process as long as additional bits are available. The end result is a non-uniform partitioning of each frame into variable-size blocks. Of course, there is some overhead required in specifying the partitioning, but this scheme adds flexibility in two ways. First, larger regions which are highly predictable are not subdivided and are coded with few bits, allowing bits to be concentrated where they are needed most. Second, the total number of bits used per frame is no longer constant. Therefore, when a greater number of bits are available, a finer partitioning can be performed. When fewer bits are available, coarser partitioning may be done. Either way, bits are allocated by priority until they are expended.

Image Attribute	Deinterlacing Mode
Spatial Correlation	Intraframe
Stationary Region	Field Repetition
Translational Motion	Motion Compensation
Exposed Region	Backward Field Repetition
Panning	Motion Compensation
Rotational Motion	
Zooming In	
Zooming Out	

Table 3.1 Common Video Attributes and Proposed Deinterlacing Modes

3.2.2 Deinterlacing Modes

As previously stated, deinterlacing for HDTV requires both nonlinear and interframe processing. Furthermore, scene-adaptive processing can also be greatly advantageous. For a stationary scene, a region may be best reproduced by field repetition. For a moving scene, perhaps motion compensated interpolation is the most beneficial. At a scene change, intraframe processing may be needed, in the form of line repetition or another spatial interpolation method. Obviously, the possible variations or modes of deinterlacing are innumerable. By restricting ourselves to a few modes, we shall show that we can achieve substantial coding gain with a manageable overhead.

The deinterlacing modes which are most useful depend on the scene content. The goal of this research is to select a few modes which do well in general, representing all types of scene characteristics. Some of the more common scene characteristics and proposed deinterlacing modes are listed in Table 3.1.

The fundamental mode proposed is an intraframe mode. Due to the high spatial redundancy of most images, having a deinterlacing method which does not require temporal processing is attractive. Intraframe interpolation is useful in situations where the vertical motion of the sequence is an odd number of lines per frame; when there is little temporal correlation between adjacent frames; or where temporal correlation can not be well exploited by the other proposed deinterlacing modes. Intraframe interpolation is also useful as a backup when motion vectors (or other temporal deinterlacing information) become corrupted during transmission. For this research, we have decided to use the Martinez-Lim algorithm [26] to perform intraframe deinterlacing because of its adequate performance in many circumstances and its relative ease of implementation.

The incorporation of temporal processing to complement this intraframe mode is essential.

Table 3.1 indicates three interframe modes which were studied in this research. As asserted earlier, temporal processing for deinterlacing is necessary for the recovery of a high-quality sequence. Drawing experience from traditional deinterlacers which often try to detect and exploit regions of no motion, we conclude that the use of a field-repetition mode is particularly inviting due to the existence of stationary regions in many types of scenes. A more general motion-compensated mode can also be beneficial for moving regions. When true motion compensated interpolation is performed (as opposed to merely field repetition) it is necessary to include the relevant motion vectors as side information.

Certain scene characteristics are not well-estimated by either the intraframe mode or the motion-compensated modes described above. For example, when objects pass in front of one another, some regions are occluded and others are exposed. Regions which are uncovered have little relationship to areas in previous frames, however these regions tend to remain exposed for a period of time, and therefore, correlation often exists between these regions and regions in subsequent frames. For this reason, a backward field repetition mode is suggested. Inclusion of this mode, which requires non-causal processing, is of minimal cost from a frame storage point of view: because the Grand Alliance system uses B (bi-directional) frames, receivers are already required to store multiple future frames.

Once non-causal processing is seen to be beneficial at little additional cost, the use of backward motion compensation mode is easily accommodated. Backward motion compensation will be most advantageous in situations where forward motion compensation fails, such as at a scene change or during a scene containing panning, where new material enters the picture at each frame. Image content which appears due to moving objects, panning, or scene changes is often not adeqately predicted by forward motion estimation.

Other attributes such as rotational motion and zooming frequently occur in video. Special modes for each of these actions might be advantageous but, like the motion-compensated mode, would require parameterized information in addition to the mode specification. For example, in a rotation mode, the angle and axis of rotation would have to be specified; in a zoom mode, the rate of magnification and a stationary point would have to be described. In addition to this extra information which must be sent, the task of computing such parameters would be expensive (although transmitter-based); the job of using such information would also complicate receivers. Another difficulty of incorporating additional modes such as these is that the number of allowable modes increases, working against one of our fundamental goals of restricting the number of modes. Instead of encompassing these modes, our experiments will rely on the modes previously described. We shall show that these modes will do an adequate job of deinterlacing image sequences which contain zooming and rotational motion, obviating the need for such specialized modes.

Chapter 4

Experimental Results

Deinterlacing experiments were performed on a variety of image sequences. A migration path deinterlacing system was formulated and simulated in software. Several aspects of the system were varied to determine the potential of such a system. The mechanics of both the transmitter and the receiver of a migration path system were implemented, and outcomes were compared on the basis of MSE as well as picture quality.

This chapter describes the experimental setup and certain results.[9] First, the specifications of the implemented system are provided. Included are details regarding deinterlacing decisions, motion vector computation, and entropy coding requirements. Next an objective appraisal of the migration path system using the mean-square error criterion is provided. Finally, subjective results in terms of picture quality are stated.

4.1 Transmitter Implementation

Figure 4.1 shows a block diagram of the processing performed by the encoder in our experiments. This figure depicts the processing which occurs inside the dashed box of the transmitter portion of Fig. 2.4.

Each frame output from the standard video decoder is partitioned into blocks. The block size chosen is dependent on the total number of bits available for enhancement and on whether or not variable-size block partitioning is to be used. Regardless of the partitioning scheme, all blocks are initially of the same size. The adaptive deinterlacing module deinterlaces every block with each of the different deinterlacing modes; in a real system this could be done in parallel. Each of the resulting blocks is passed to a comparator that determines the mode which performs the best for that region. The choice of deinterlacing mode, along with any relevant parameters (*e.g.*, motion vectors) is stored in preparation for entropy coding and transmission. Once all the blocks are processed by the adaptive deinterlacing module, and the best deinterlacing mode has

---

[9] More details about the experimental implementation are provided in Appendix A.

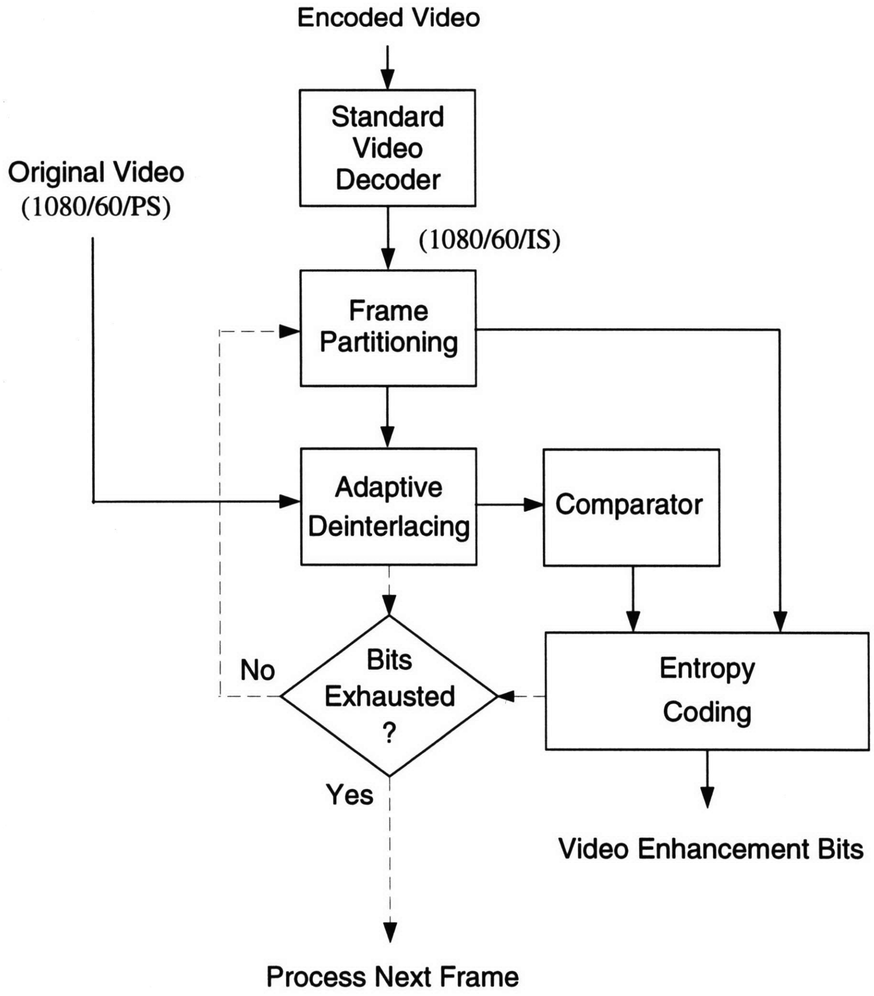


Figure 4.1 Block Diagram of Deinterlacer. This is a block diagram depicting the processing that occurs in the simulated transmitter of our migration path deinterlacing system. The functionality shown here corresponds to the processing performed in the dashed box of Fig. 2.4. Solid lines indicate the flow of image or enhancement data. Dashed lines indicate the flow of control information.

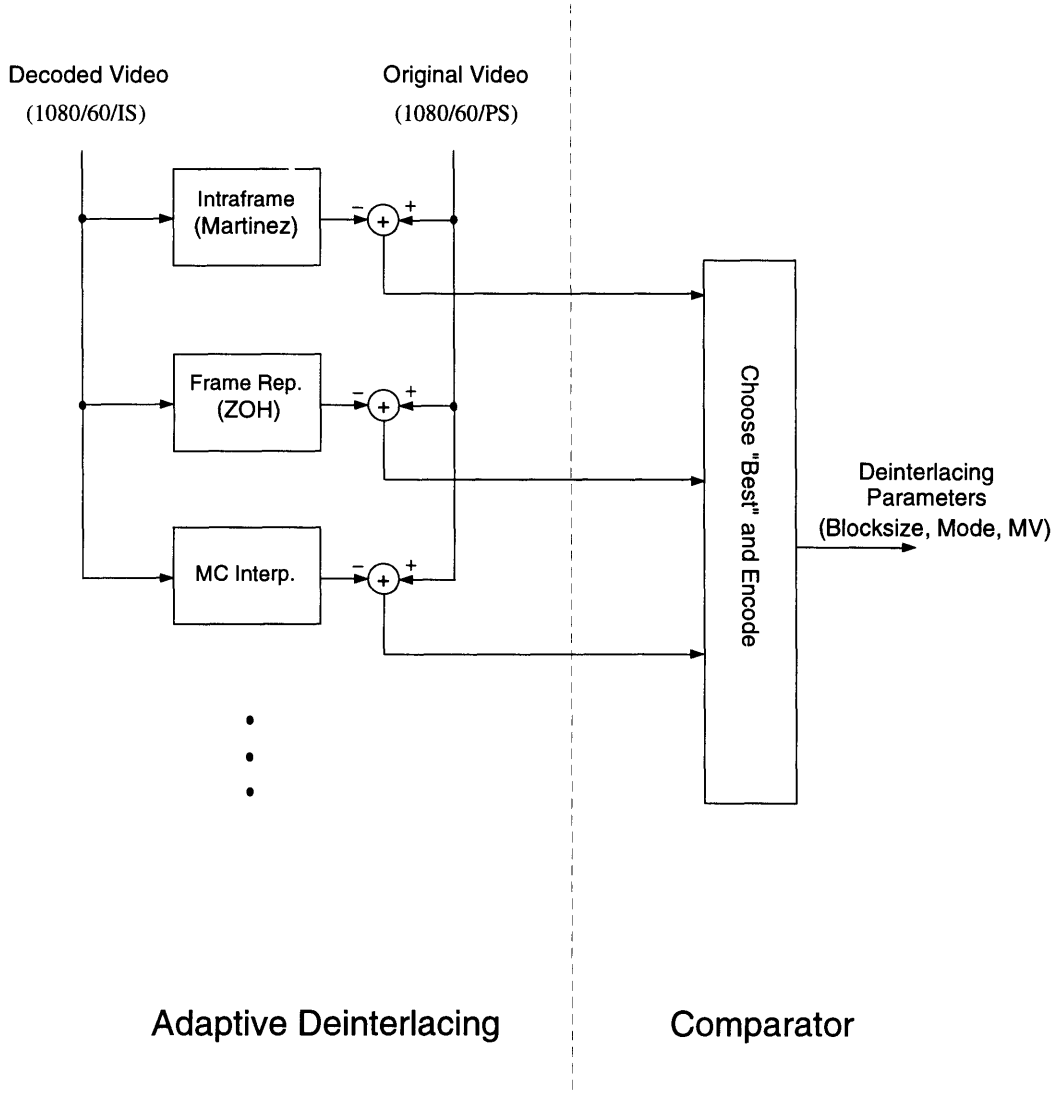


Figure 4.2 Adaptive Deinterlacing and Comparator The adaptive deinterlacing module and comparator process each block of the decoded, interlaced image. Different methods of deinterlacing are performed and the comparator selects the best.

been chosen for each, the entropy coder estimates the total bit requirement for the current frame, including the partition information. If the total number of bits allocated has not been exhausted, the original partition is refined by subdividing those blocks which need it most. The new blocks are then passed to the adaptive deinterlacer which iterates on the refined partition. This process continues until the number of bits required to encode the current frame meets the number allowed by the side-information channel. At that point the deinterlacing modes, parameters, and partition are entropy coded for transmission. The next frame is then loaded for processing.

Fig. 4.2 shows a more detailed view of the adaptive deinterlacing and comparator modules. Each block of the interlaced sequence enters the adaptive deinterlacer. This block is deinterlaced using multiple deinterlacing modes. The adaptive deinterlacer also receives as input the corresponding blocks from the original high resolution video, which in our simulations was the 1080/60/PS format. As mentioned in chapter 3, having access to the original simplifies the task of comparing the different deinterlacing algorithms.

It is in the adaptive deinterlacer that motion detection and estimation are performed.[10] The algorithm which we have used for motion estimation was block-matching with the mean-square error criterion. Again, having access to the 1080/60/PS original simplifies the choice of the best motion vector once an error measure is chosen; however, it is relevant to note that any motion-estimation algorithm would work and with any error criterion: since receivers do not need to do any motion estimation, they do not need to know the algorithm used for motion estimation at the transmitter, only the result, *i.e.*, the selected motion vector.

The comparator selects which of the deinterlacing modes to use for the current block. The selection criteria which we have used are described in the following section; however, it is important to note that from the receiver's point-of-view, how the comparator chooses the mode is unimportant. The receiver only needs to know which mode and what parameters to use for a given block.

4.1.1 Deinterlacing Modes

Simulations were performed using the deinterlacing modes described in Chapter 3. The intraframe mode (INTRA) and forward field repetition mode (FZOH) were always active in the adaptive deinterlacing module. Experiments were conducted with and without using motion vectors to determine whether the advantage gained by the motion-compensated deinterlacing mode (FMV) was worth its higher overhead. Additionally, experiments using and omitting the backward field repetition mode (BZOH) were conducted in order to determine its relative merits. In our experiments, a backward motion-compensated mode was not used. In regions where smooth motion occurs, it can be expected that forward and backward motion compensation perform similarly. The types of situations which suffer most from the omission of this mode—regions containing panning and scene changes—are where new objects are introduced into the scene.

---

[10] Since motion estimation is also performed by the standard HDTV encoder, these motion vectors are also available to the adaptive deinterlacer; however, not every block will have a motion vector in the standard video stream, and the motion vectors for those blocks that do are not necessarily the best to use for deinterlacing, thus, the adaptive deinterlaces computes new motion vectors for each block.

As mentioned in the previous section, the migration-path deinterlacer can use any arbitrary criterion to determine which mode to use for the current block. During most of our experments we made our deinterlacing decisions based solely upon the MSE improvement; however some of our simulations included weighing the selection of the modes based on their coding requirements—in particular, the FMV mode was not selected as often because of its inherently higher overhead. We also ran some tests where the FMV mode was not allowed for some of the smaller *i.e., 2* × 2 blocks. These adjustments to the mode selection scheme were seen to have some merit and warrant futher research.

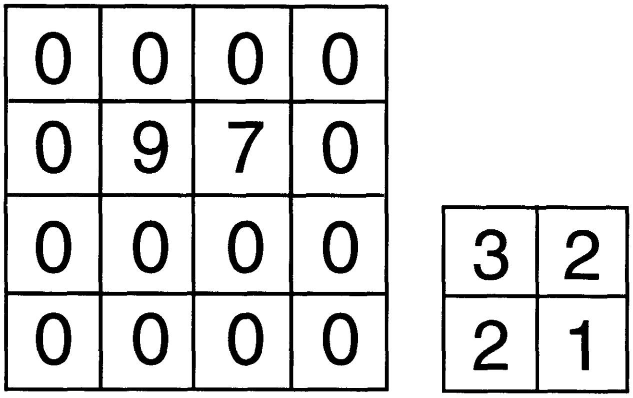
4.1.2 Block Size

Simulations were also run to assess the relative performance of processing with fixed-size blocks versus using variable-size blocks. When variable-size blocks are used, the partition information must be encoded and transmitted in addition to the mode information. We shall see that despite this extra overhead, the use of variable-size blocks is beneficial. Also, as stated earlier, fixed-size blocks constrain the side-channel-capacity requirement. Fixed block size experiments were conducted using 16 × 16 and 8 × 8 blocks. Experiments with variable-size blocks were conducted by initially partitioning the image into 16 × 16 blocks, every other line of which was missing due to the interlacing, and systematically subdividing into quadrants those blocks which required it most. The smallest that blocks were ever subdivided was into 2 × 2 pixels.

When a variable-size block-partitioning strategy is used, a prioritization must be established for selecting blocks to subdivide. Some possibilities include sorting blocks by total error, by error per pixel (MSE), or by reduction in error per bit of side information. Arguments exist in favor of each of these alternatives.

Subdivision of blocks by total error would tend to favor the larger blocks; thus, if two blocks had approximately the same MSE, the larger block would be subdivided first. Also, since the number of additional bits required for subdividing a block and coding the subblocks is approximately uncorrelated with block size, subdivision based on total error would tend to use the additional bits to reduce a larger error. Thus, if the assumption that picture quality and total error are correlated is accurate, this strategy is a logical choice.

Subdividing based on error per pixel also has certain merits. In this case the additional bits are used to improve pixels which are collectively inaccurate. Comparision of these two alternative methods is depicted in Fig. 4.3. In Fig. 4.3(a), a 4 × 4 block with a total error of 16 and, therefore, a mean error of 1 is shown. Fig. 4.3(b) depicts a 2 × 2 block with total error of 8 but a mean error of



0	0	0	0
0	9	7	0
0	0	0	0
0	0	0	0

(a)

3	2
2	1

(b)

Figure 4.3 Block Subdivision Criteria. The blocks which are subdivided first depend on the criterion used. In this example each square and its number represents a pixel and its estimation error. In our processing the 4 x 4 block would have a higher priority for subdivision if it were done based on total error, and the 2 x 2 block would be subdivided first if it were based upon error per pixel.

2. Which of the two subdivision criteria is used determines which block would be subpartitioned first.

An alternate option for the subdivision criterion is to divide based upon the improvement in error per additional bit of side information. This method, in some sense, gets the most out of the additional bit stream. These three options for subdivision were implemented and compared. One of the interesting results we found was that the interpolation error was approximately the same no matter which of these three options for subdivision was used. This result seems to convey the message that the worst blocks under one of these criteria tend to be the worst under the other criteria. One advantage of the migration-path system is that regardless of the subdivision prioritization scheme, receivers will be able to decode the video. The results shown later in the chapter were compiled using the total error method for subdivision decisions.

4.1.3 Motion Vectors

The computation of vectors for motion-compensated interpolation is an important task. A migration-path system performs this operation at the encoder. This has several advantages. First, the majority of the computational cost lies at the transmitter end. This is obviously desirable, since the receivers' cost can be kept down, making them more affordable for consumers. Second, the computation can be performed using the original image. If motion estimation were performed

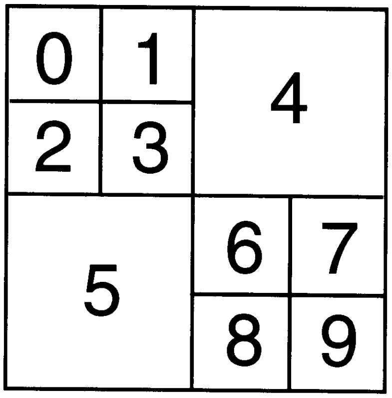


Figure 4.4 Entropy Scanning Order. Information from blocks to be coded is scanned from left to right and top to bottom in the frame. If a block is to be subdivided, its subblocks are scanned with a similar orientation. In this example each numbered square represents a block of pixel with the same deinterlacing parameters (mode or motion vector). This figure depicts the order by which the blocks would be scanned, starting at 0 and continuing to 9.

at the decoder, the vectors computed would be based on estimates of the missing lines. Since motion estimation is done at the encoder, the original lines can be used, necessarily yielding better estimates. Third, the error criterion and accuracy used can be easily varied at the encoder, allowing all receivers to benefit. If motion estimation were done at the decoder, the error measure (such as mean-square error), and the accuracy (for example, full- or half-pixel) would be fixed by the receiver. If a receiver could only do full-pixel motion estimation, then it might have difficulty deinterlacing a video stream which was encoded using half-pixel accuracy. Furthermore, if new innovations in motion estimation were to come about, existing receivers would not be up-to-date. When motion estimation is done at the encoder, advanced methods can be implemented, with all receivers benefitting.

When motion vectors were used in these experiments, they were computed using full-pixel accuracy. Comparisons were made, on a block-by-block basis, between the luminance components of the current and previous frames. Only the missing lines in the frame to be deinterlaced and corresponding lines of the previous frame were considered; for example, only the four missing lines of an 8 × 8 block in the current frame are used when determining the error with an 8 × 8 block in the previous frame (from which only four lines are used, also). When the vertical motion offset is even, the original lines in the previous frame can be used; for an odd offset, the previous frame's lines were estimated using linear interpolation; recursive motion processing—estimating the motion from previously deinterlaced frames—was not used. The best motion vectors, based on mean-square error, were selected.

4.2 Entropy Coding Results

After the partitioning and mode selection process is completed, the encoder has generated one or more streams of digital information. Because of the statistical redundancy in these streams, entropy coding techniques can be used to compress the data. In our experiments, the data in each stream was organized consistent with left-to-right scan order. When blocks were subdivided, all sub-regions were scanned before the next block. This is illustrated in Fig. 4.4. The computed entropy of each stream was used to approximate the total bit requirement of the side channel. Because the bit requirement is based on the entropy and not on predetermined codebooks, there is an underlying assumption that each stream's statistics can be well-predicted or that adaptive coding will converge to the true statistics relatively quickly.

The entropy of each stream was computed in two ways: the zero-state entropy, which depends on each symbol's relative frequency, and the one-state entropy, in which the frequencies of pairs of symbols are important. Coding of each of the different streams is described below.

4.2.1 Deinterlacing Modes

The digital data requirement for coding the deinterlacing mode stream was computed using the one-state entropy, which was typically 5-10% lower than the zero-state entropy. One might expect this, since spatial correlation of image sequences would tend to indicate spatial correlation among deinterlacing decisions. One interesting phenomenon which was apparent in analyzing the results was that the stream entropy increased with the partition fineness. That is, if more partitioning is done, the entropy (per symbol) of the resulting stream is greater. This seems to occur since finer partitioning decreases the correlation between adjacent blocks.

4.2.2 Frame Partitioning

When variable-size blocks are used, the partition information must be encoded. The partition can be described as a binary stream where a **1** indicates that the current block should be subdivided and a **0** dictates that it should not. This stream is effectively compressed using runlength encoding, describing the stream by integers which represent the number of **0**s between each pair of **1**s, followed by entropy coding of the integer stream. The one-state entropy of the integer stream was found to be significantly (on the order of 10-20%) lower than the zero-state entropy.

4.2.3 Motion Vectors

As stated earlier, motion vectors were computed to full-pixel accuracy. The maximum displacements used were 32 pixels in the horizontal direction and 16 pixels in the vertical direction. These correspond to an object moving across the width or height of the screen in approximately one second. Coding of motion vectors was performed on each component separately. The entropy of the horizontal and vertical vector streams were computed separately, with and without first differentially encoding them. Results consistently showed that the one-state entropy of these streams without differential encoding was the lowest. Typically, the average bit requirement for transmitting an individual motion vector was 3-4 times higher than the average bit requirement for coding the deinterlacing mode.

4.3 Objective Results

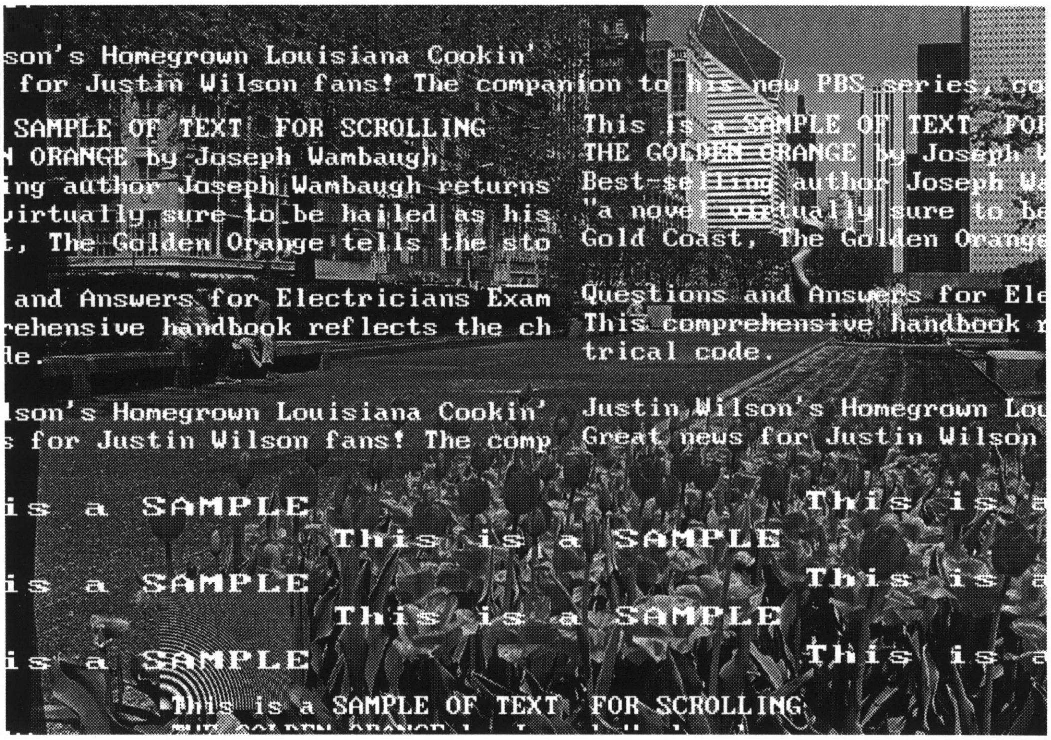
With a set-up as described in the previous sections, video sequences were processed. The number of enhancement bits was varied in order to evaluate deinterlacing-error reduction as a function of the side channel capacity. Figure 4.5 shows the three sequences which we shall discuss here. The first sequence, hereafter referred to as `mall`, is a scene containing people walking by a fountain in a shopping mall. Figure 4.5(a) displays one frame of this sequence. The other two sequences, `girl`, and `tulipstext` are synthetic sequences which are extremely detailed and are, in general, much more difficult to code than most real life sequences. These sequences will further serve to identify some of the attributes of the migration-path processing. The `girl` sequence contains three distinct regions. The majority of each frame shows a girl in front of various items of different colors and textures. This part of the sequence pans to the right and then zooms in and toward the upper left. Inset in the upper left corner of the sequence is a clip of video which contains a woman walking down a street. The third distinct region is the "Super 8" rotating test pattern and "high resolution" color bar. These last items contain high spatial frequency components from which aliasing would normally arise. Figure 4.5(b) shows a frame of this sequence. The third sequence, herein referred to as `tulipstext` and shown in Fig. 4.5(c), depicts scrolling text on a still, detailed city background.

Experiments were conducted to investigate the improvement in deinterlacing using the migration-path model. The numerical results which follow are in terms of the mean-square error between the original (1080-line progressive) video and the processed (deinterlaced) sequence. While the MSE is not necessarily the optimal criterion to use for comparing images, we shall show that it is, in general, well-correlated with picture quality for the purposes discussed here.

(a) `mall`

(b) `girl`

Figure 4.5 Original Test Sequences.



(c) `tulipstext`

Figure 4.5 (continued)

Also, it is simple to formulate and compare objectively. The baseline used for comparison is intraframe deinterlacing using the Martinez-Lim algorithm. The MSE is computed as the average over the deinterlaced pixels of the sum of the three colors' (RGB) squared errors.[11] The additional bits required for deinterlacing will also be compared. Frequently, the coding requirement for the deinterlacing information will be given as a percentage of the data for standard 1080/60/IS high-definition video, which is approximately 0.32 bits per pel.

Figure 4.6(a) plots the deinterlacing error as a function of bit rate for the `mall` sequence.[12] The horizontal axis indicates the number of bits per pixel required to code the side information in excess of the bit requirement for coding the standard HDTV video stream. The vertical axis shows the mean-square error. Each line shown in this figure represents a group of experiments conducted with the same set of deinterlacing modes as the legend indicates. For example, the top line shows data from deinterlacing using only the intraframe (INTRA) and forward field repetition (FZOH)

---

[11] Each color component takes on integer values from 0 to 255.

[12] The complete data for the three plots in Fig. 4.6 is provided in Appendix B.

modes, while the bottom line plots the data from deinterlacing using all four modes. Different points along the same line, indicated by an 'x' were computed by varying the number of excess bits available. These points are connected by straight lines in order to estimate performance in the in-between regions.

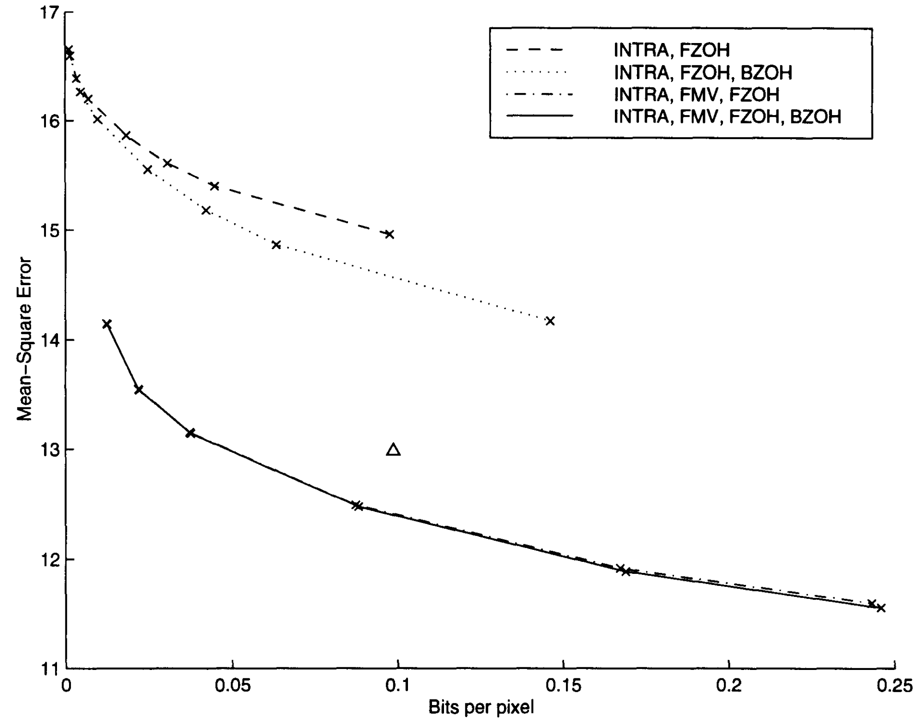
The point denoted with a triangle is the result of deinterlacing using all of the available modes but with a fixed (8 × 8) block size. All the points on the lines correspond to experiments in which the block size can vary from 16 × 16 to 2 × 2 pixels; the left-most point on each line corresponds to all 16 × 16 blocks and represents the fewest enhancement bits required for our experiments. When zero additional bits are used, the system, unable to encode either partition information or deinterlacing mode information, will always use the intraframe mode. The corresponding point is not shown in Fig. 4.6(a), but will be discussed in the comparison below.

Figure 4.6(b) and Fig. 4.6(c) show the corresponding results for the `girl` and `tulipstext` sequences, respectively. Collectively, these three graphs provide some useful intuition regarding the performance of our system.

Remark 1
Significant MSE improvement can be obtained from a small number of bits.

As mentioned above, the left-most point on each of the lines in Fig. 4.6 indicates the MSE using a fixed partition with 16 × 16 blocks. In our variable blocksize scheme, this was the largest block size allowed and therefore, represents, the coarsest frame partition. Table 4.2 shows the MSE improvement corresponding to these points along with the additional bit requirement for each. The column labelled Δ **MSE** indicates the percent reduction of the mean-square error from the error due to intraframe deinterlacing. The overhead is given as a percentage of 0.32 bits per pixel, the approximate coding requirement for standard HDTV.

As shown in this table, large reduction in the MSE can be realized with a small number of additional bits. Furthermore, if we look at the curves in Fig. 4.6, we consistently see the steepest slope occurs for a relatively small amount of side information with a diminishing, yet still significant, return as we increase the bit rate. Table 4.1(a) lists the improvement in MSE as a function of bit rate for the `mall` sequence. These values correspond to the bottom curve in Fig. 4.6(a) with the exception of the top row of the table, which contains the error due to intraframe deinterlacing. Corresponding results for the `girl` and `tulipstext` sequences are shown in Tab. 4.1(b) and Tab. 4.1(c), respectively.



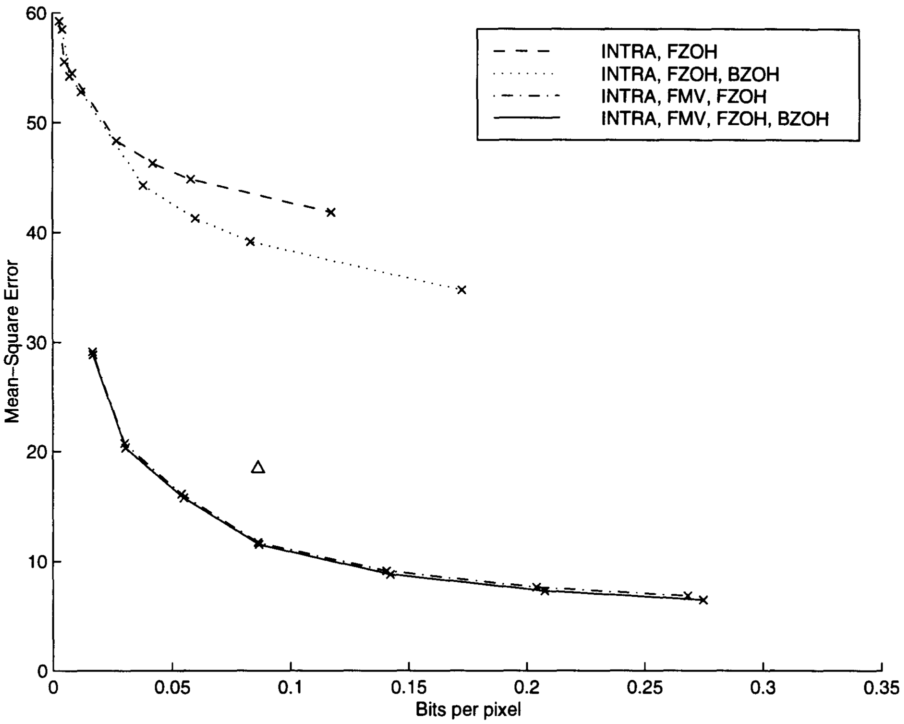
(a) `mall` Sequence Statistics

Figure 4.6 MSE vs. Bit Rate. Graph of mean-square error vs. bits per pixel for the migration path deinterlacing system. Points connected along the same line are generated using the same set of deinterlacing modes. The point plotted with a triangle corresponds to fixed 8 × 8 blocks with all modes in use (the same as the bottom line).

Blocks per 64 pixels	**MSE**	**Δ MSE**	**Bits/pel**	**Overhead**
N/A	16.9	N/A	0.0000	0.0%
0.25	14.1	16.4%	0.0124	3.9%
0.33	13.5	19.9%	0.0220	6.9%
0.5	13.1	22.3%	0.0376	11.7%
1.0	12.5	26.2%	0.0882	27.6%
1.5	11.9	29.7%	0.1691	52.8%
2.0	11.6	31.7%	0.2460	76.9%

(a) `mall` Sequence Statistice

Table 4.1 MSE for Deinterlacing Experiments.These values are from deinterlacing the sequences using all four modes (INTRA, FMV, FZOH, BZOH). The top line is the result when no side channel is available; only the INTRA mode is used.



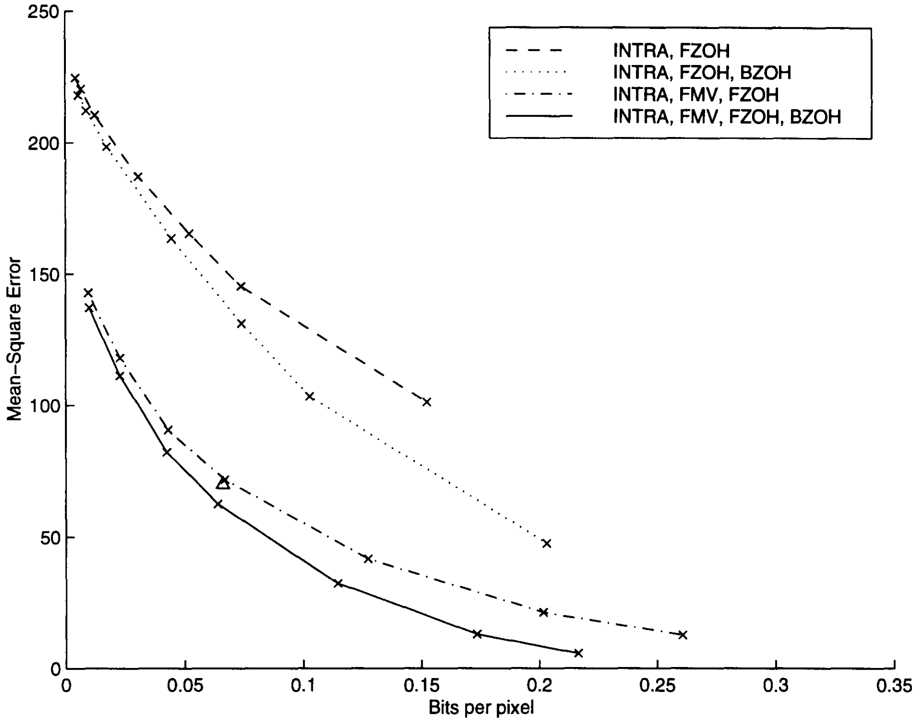
(b) `girl` Sequence Statistics

Figure 4.6 (continued)

Blocks per 64 pixels	MSE	Δ MSE	Bits/pel	Overhead
N/A	78.5	N/A	0.0000	0.0%
0.25	28.8	63.3%	0.0169	5.3%
0.33	20.3	74.1%	0.0305	9.5%
0.5	15.8	79.9%	0.0549	17.1%
0.64	11.5	85.4%	0.0867	27.1%
1.0	8.8	88.8%	0.1424	44.5%
1.5	7.3	90.7%	0.2077	65.0%
2.0	6.4	91.9%	0.2748	85.9%

(b) `girl` Sequence Statistics

Table 4.1 (continued)



(c) `tulipstext` Sequence Statistics

Figure 4.6 (continued)

Blocks per 64 pixels	MSE	Δ MSE	Bits/pel	Overhead
N/A	284.06	N/A	0.0000	0.0%
0.25	137.21	51.70%	0.0100	3.1%
0.33	111.34	60.80%	0.0228	7.1%
0.5	82.26	71.04%	0.0425	13.3%
0.65	62.48	78.00%	0.0639	20.0%
1.0	32.39	88.60%	0.1145	35.8%
1.5	12.99	95.43%	0.1737	54.3%
2.0	5.76	97.97%	0.2165	67.7%

(c) `tulipstext` Sequence Statistics

Table 4.1 (continued)

Sequence	Δ MSE	Bits/pel	Overhead
`mall`	1.6%	0.0010	0.3%
	1.9%	0.0014	0.4%
	16.3%	0.0123	3.8%
	16.4%	0.0124	3.9%
`girl`	24.6%	0.0029	0.9%
	25.6%	0.0041	1.3%
	63.0%	0.0167	5.2%
	63.3%	0.0169	5.3%
`tulipstext`	20.9%	0.0044	1.4%
	23.3%	0.0055	1.7%
	49.7%	0.0095	3.0%
	51.7%	0.0100	3.1%

Table 4.2 MSE Improvement for Deinterlacing Experiments.These values show the percent reduction of the MSE which can be achieved using a minimal amount of side information. The four entries for each sequence correspond to the leftmost points of the plots in Fig. 4.6.

Remark 2

MSE improvement due to the motion-compensated prediction mode easily compensates for this mode's higher overhead.

Because of the necessity to transmit a motion vector along with the mode information for motion-compensated interpolation, the overhead required for each instance of using this mode is several times higher than the overhead for each instance of the other modes. Therefore, we should require our deinterlacing system get enough improvement from this mode to justify its use. Fig. 4.6 suggests that this requirement is met. For example in Fig. 4.6(a), we see that, at the same bit rate, use of the FMV mode in conjunction with the other three modes reduces the MSE by approximately 13–15% from the level attainable using only the other three modes. For the `girl` and `tulipstext` the results are even more dramatic: a reduction in MSE of 45%–75% and 35%–85%, respectively. For all three sets of experiments, the amount of improvement increases with bit rate. This is understandable, for at higher bit rates, there is a finer division of the picture and thus more opportunity to take advantage of the motion-compensated mode.

On an interesting side note: motion-compensated estimation is a method which will be extensively used to encode standard HDTV video; however, we have not attempted to use these motion vectors to assist in deinterlacing. It is likely that the motion vector needed for deinterlacing will frequently be the same as the motion vector used for coding the corresponding lines of the standard video. In this case, it is not necessary to transmit the same motion vector twice, only to indicate to the deinterlacer that it should use the corresponding motion vector from the standard video stream. By applying this strategy, it may be possible to get the deinterlacing benefit of the motion-compensated mode while reducing the overhead required.

Remark 3

MSE improvement from the use of variable-size blocks more than offsets the need to code the partition information.

The points plotted with a triangle in each of the three figures correspond to the performance achieved using all four deinterlacing modes but fixing the block size at 8 × 8 pixels. When we compare this performance with that due to equivalent processing but with varying block sizes, we notice two related results. First, at the bit rate required by the fixed block size experiment, the MSE can be reduced by using variable-sized blocks. This reduction is approximately 4%, 36%, and 12% for the `mall`, `girl`, and `tulipstext` sequences, respectively (see Tables B.1 and B.2). Alternatively, the bits required to achieve the same error performance are fewer for the variable-sized block experiments. Our numerical results determined that reduction of the enhancement bit rate by 50%, 42%, and 15% for the three sequences is possible by using variable-sized blocks.

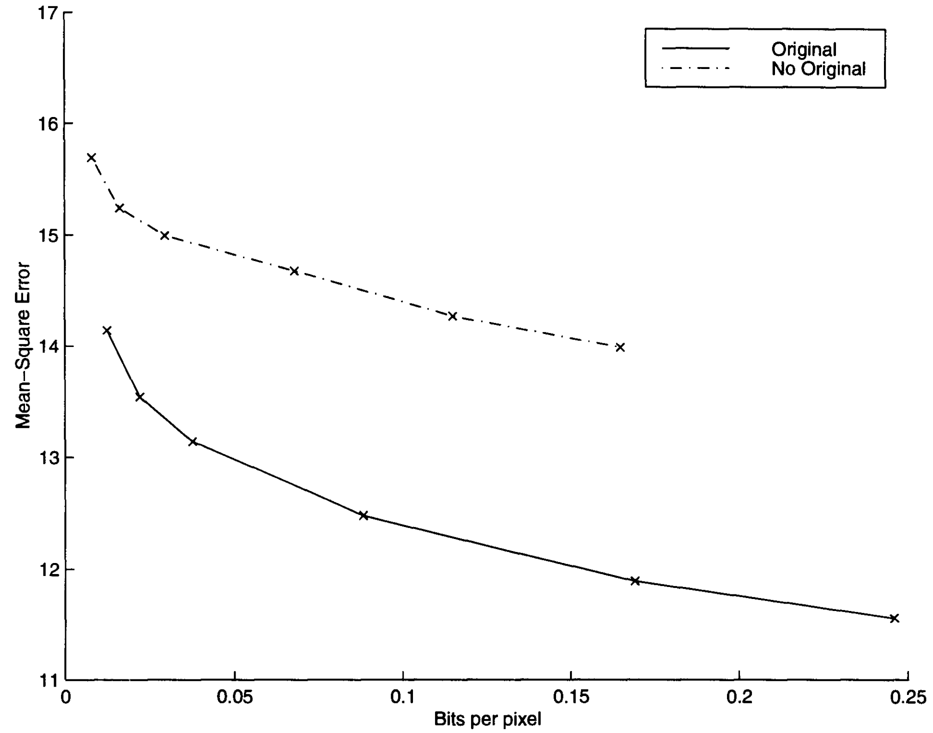
As stated earlier, when fixed-size blocks are used the number of blocks is predetermined and therefore the overall bit rate is somewhat constrained. For the `mall`, `girl`, and `tulipstext` sequences, the overhead required using the four deinterlacing modes is 30.9%, 26.9%, and 20.1% respectively. It should be noted that it is possible to reduce this overhead by, for example, limiting the number of blocks which can use the FMV mode; however, this reduction in overhead comes at the price of not being able to use the best deinterlacing mode for some blocks.

There is another interesting result in comparing variable-sized and fixed-size block processing for the `mall` sequence: while only a small difference in MSE is possible at the same bit rate, a large difference in data rate is possible at the same MSE. This suggests that perhaps the MSE improvement between the two corresponding points on the lower curve in Fig. 4.6(a) may not warrant the additional bits required.

Remark 4

Having use of the backward field repetition mode in addition to the forward field repetition mode can provide substantial MSE gain.

Sequences such as `mall` and `girl` are dominated by panning and zooming. In scenes such as these, addition of the backward field repetition mode does not provide substantial improvement beyond what can be achieved using the forward field repetition and motion-compensated modes. This is confirmed by Fig. 4.6(a) and Fig. 4.6(b), where the two lower curves are virtually indistingishable. In the case when the FMV mode is not used, the BZOH mode can provide some benefit. Conversely, in scenes such as `tulipstext` where the background is stationary and fore-



(a) `mall` Sequence Statistics

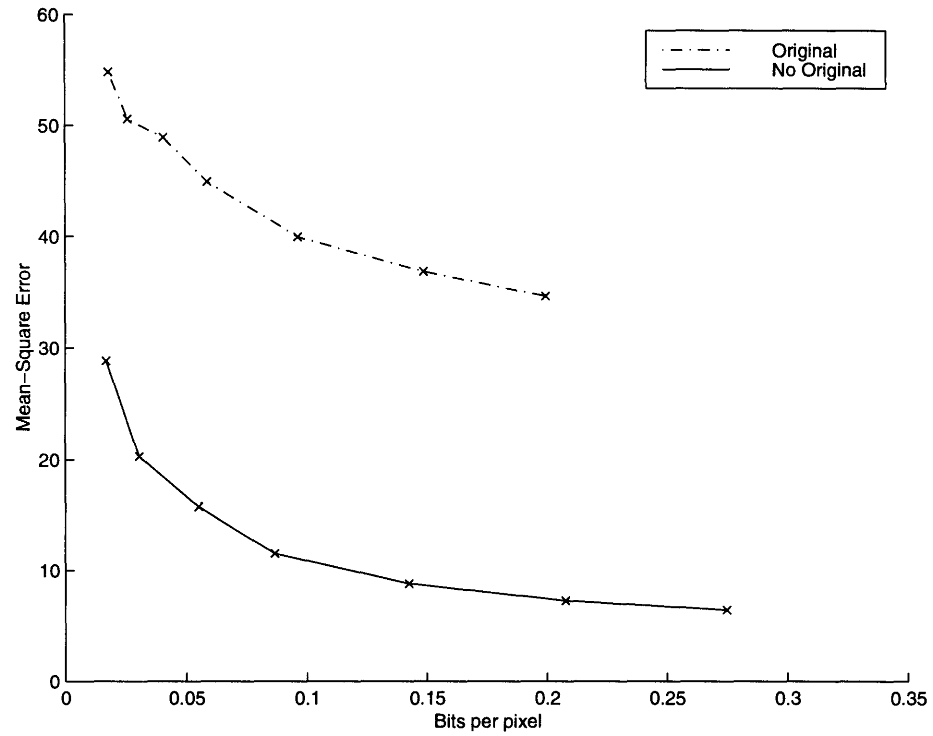
Figure 4.7 Access to Original Video. Graph of mean-square error vs. bits per pixel to compare importance of having access to original video while computing motion vectors.

ground images are moving, there is the potential for great benefit from the BZOH mode. Regions of the background which are exposed due to the passing of foreground objects will generally be identical in subsequent frames. Therefore, during the frame in which they are exposed, use of the BZOH will be advantageous. This is verified in Fig 4.6(c) where the bottom curve depicts an improvement of 5–65% as compared to the curve above it. Again, the relative benefit increases as the total enhancement bit rate increases.

A final set of experiments was conducted to ascertain the relative importance of having access to the original video. The bit rate vs. MSE results are shown in Fig. 4.7.[13] In this experiment, the `mall`, `girl` and `tulipstext` sequences were deinterlaced using all four modes; however, the motion vectors used were estimated based on the assumption that the original video was not available. That is, the missing lines were estimated by linear interpolation and then motion estimation was based on these missing lines. Deinterlacing was then performed using these motion

---

[13] The complete data for the three plots in Fig. 4.7 is provided in Appendix B.



(b) `girl` Sequence Statistics

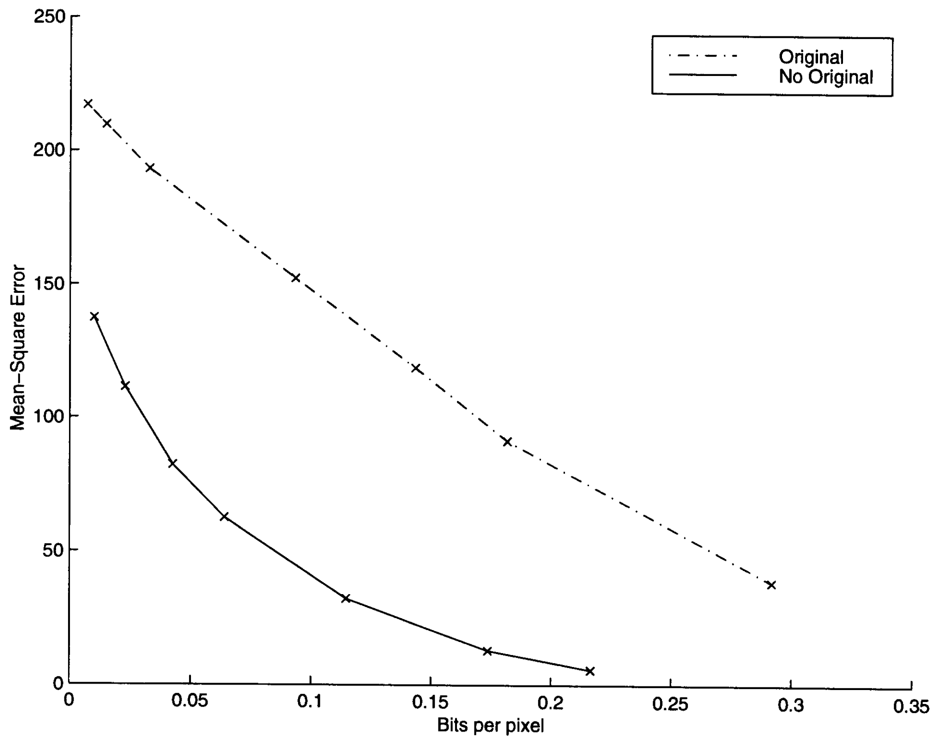
Figure 4.7 (continued)

vectors—the deinterlacing mode was chosen by matching blocks to the original video. The results are shown by the top lines in Figs. 4.7(a), (b) and (c). The bottom lines in these figures, shown for reference, are the same as the bottom lines in Fig. 4.6(a), (b), and (c), where access to the original video is assumed.

Remark 5

Having access to the original video greatly improves deinterlacing performance.

Figure 4.7 indicates the benefit possible from having access to the original video when computing the motion vectors. When executing these experiments we used the original video for selection of the deinterlacing mode. This assumption can only reduce the deinterlacing error, yielding better results than would realistically be possible when no original was present. Therefore, we can infer that, in general, the system performance without the original video available would be worse than indicated by the top lines in Figs. 4.7(a), (b), and (c).



(c) `tulipstext` Sequence Statistics

Figure 4.7 (continued)

4.4 Subjective Results

The numerical results in the previous section suggest that there is a undeniable advantage to deinterlacing using a migration-path system. However, our underlying assumption that mean-square error is correlated with picture quality still needs to be verified. The deinterlacing improvement to the `mall`, `girl`, and `tulipstext` sequences is not only evident by looking at the MSE comparison but also by viewing on a high resolution display.[14]

When intraframe deinterlacing is performed on the `mall` sequence, artifacts are most noticeable in the regions containing diagonal detail; for example, along the railing and brick floor on the right side of the scene and in the bench on the left side of the scene. An excerpt of this area is displayed in Fig. 4.8. The flicker seen in these areas is quite disturbing. When the migration path experiments were conducted, dramatic improvement could be seen. For example, even without

---

[14] More information regarding the conduct and results of subjective tests is included in Appendix C.

(a) Original (b) Intraframe

(c) 0.098 bits/pel (fixed block size) (d) 0.038 bits/pel (variable block size)

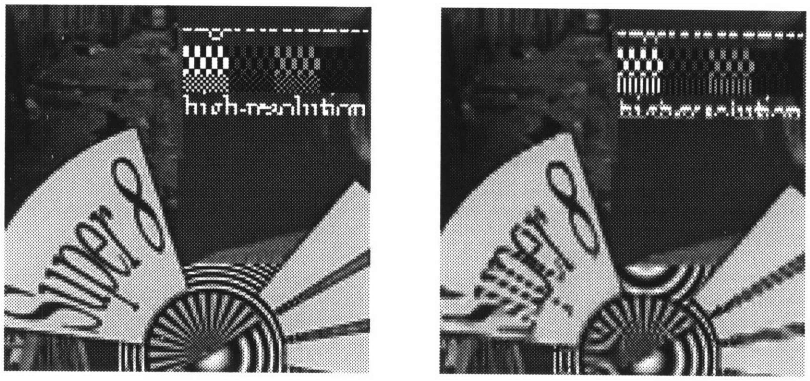
Figure 4.8 Deinterlacing Results for the `mall` sequence. A section of the `mall` sequence which shows the original as well as different versions of the deinterlaced sequence.

the motion vector mode at 0.021 bits/pel (6.25% overhead), the flicker is clearly reduced. This confirms the validity of **Remark 1** in the context of picture quality. Experiments with the FMV mode showed further improvement. For example, using a fixed block size at 0.098 bits/pel (Fig. 4.8(c)), nearly all the deinterlacing artifacts are removed. The variable block size at a similar rate (0.088 bits/pel) showed only minor improvement. These latter two experiments, however, correspond to approximately a 30% coding requirement. Reducing the side information to 0.038 bits/pel (11.6%) is possible with variable block size coding (Fig. 4.8(d)), yielding the same quality as the fixed block size coding. This performance is quite acceptable and is indicative of the advantage one can gain from the migration path. These picture-quality results support the MSE results in the previous section, from which we also concluded that the gain from fixed-sized blocks to variable-sized blocks would be minimal for the `mall` sequence.

Experiments with the `girl` sequence showed similar trends, but as this sequence is much more difficult to code—probably more difficult than typical HDTV-quality pictures—it would require a large side channel to completely restore the quality of the original. Nevertheless, with the migration-path implementation, we can still see substantial improvement over the intraframe deinterlaced sequence as depicted in Fig. 4.9. Intraframe deinterlacing produces many objectionable artifacts: aliasing in most areas of the picture, striations in the letters in the test pattern, and large amount of flickering throughout the scene. Remarkably, the letters underneath the color bar, although blurred, are not too distracting; however, this is mainly due to their vertical movement by an odd integer number of pixels per frame.

Migration-path coding of the picture with fixed blocks at 0.086 bits/pel (4.9(c)) dramatically reduces the flicker and much of the aliasing throughout the frame. The only region which looks worse is the color bar which, although sharper, has some granular artifacts. Using a variable block size at approximately the same rate continues to make noticeable improvements (Fig. 4.9(d)), especially in the rotating test pattern. The variable block size method (Fig. 4.9(e)) also allows us to deinterlace with about the same quality as the fixed block size strategy, but at 0.031 bits/pel (9.7% overhead).

The `tulipstext` sequence, shown in Fig. 4.10 is another example where the migration path approach gives a large improvement to picture quality. Because the background region is of very high resolution, intraframe deinterlacing has difficulty reproducing the original detail (Fig. 4.10(b)); however, the migration path system has a great deal of success (Fig. 4.10(c)). This picture shows the benefit which can be gained from the backward field repetition mode. The four modes combine to reproduce a sharp, clear image of much higher fidelity than is possible with intraframe deinterlacing alone.



(a) Original

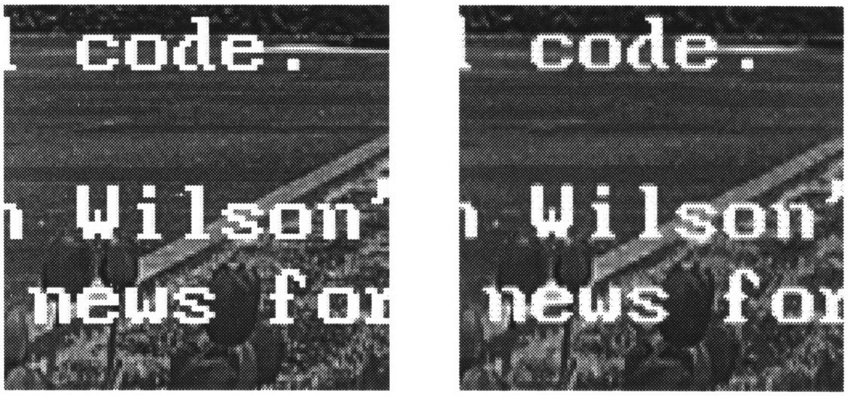
(b) Intraframe

(c) 0.086 bits/pel (fixed block size)

(d) 0.085 bits/pel (variable block size)

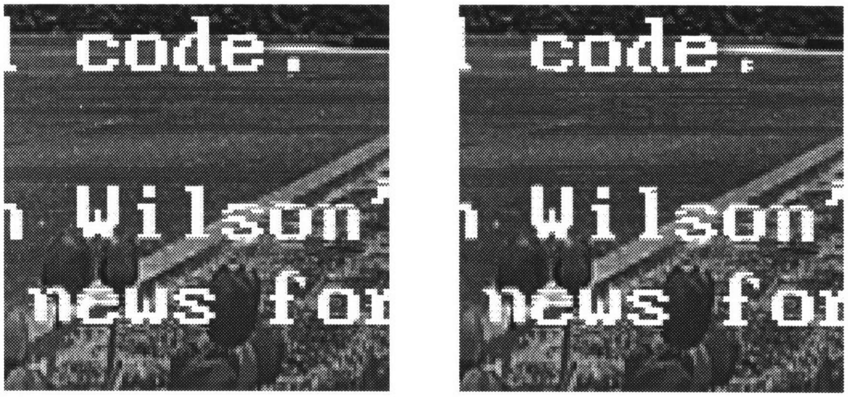
(e) 0.031 bits/pel (variable block size)

Figure 4.9 Deinterlacing Results for the `girl` sequence. A section of the `girl` sequence which shows the original as well as different versions of the deinterlaced sequence.



(a) Original

(b) Intraframe



(c) 0.066 bits/pel (fixed block size)

(d) 0.064 bits/pel (variable block size)

Figure 4.10 Deinterlacing Results for the `tulipstext` sequence. A section of the `tulipstext` sequence which shows the original as well as different versions of the deinterlaced sequence.

One interesting result which we found with the `tulipstext` sequence is that in some regions, the use of variable-sized blocks was detrimental. Because the background region was not moving, some blocking effects could be seen in some of the larger blocks (16 × 16) which overlapped foreground and background regions. For example, Fig. 4.10(d) was produced using variable block sizes and approximately the same number of bits as Fig. 4.10(c) which was generated using a fixed block size. In the former, we notice a horizontal blocking artifact immediately above the letters "Wil". In the moving video, this disturbance is quite visible. Using fixed blocks may have produced similar problems, but because the block size is smaller (8 × 8), the perceptual effect was not noticeable.

This unexpected result reminds us that MSE is not the only criterion which should be considered in deinterlacing. It is important to weigh perceptual effects as well when making decisions regarding block partitioning and deinterlacing mode selection.

In general, the picture-quality results were consistent with what was predicted from the objective results. In particular, we found that deinterlacing with the migration path system gave noticeable improvements in picture quality, even with only a small amount of side information. Also, the following conclusions were drawn:

Remark 6

Using only deinterlacing modes (no coded residual) can capture and restore most of the high-definition detail lost in the interlacing process.

All of the processing that we performed was done without the coding of a residual image. As seen in Fig. 4.8(d), Fig 4.9(d), and Fig. 4.10(c), it is possible to recover much of the high definition detail without resorting to using the residual.[15]

Remark 7

A small number of simple modes is probably sufficient to implement the migration-path system.

The four modes which we have chosen seem to do an adequate job of representing information useful in deinterlacing. Although the sequences which we studied contained regions with panning, zooming, and rotating features, the four modes were able to accommodate all these attributes. Therefore, it is probably not necessary to require highly specialized modes for deinterlacing—those outlined in Table 3.1 are, for the most part, sufficient.

---

[15] This is not to imply that coding a residual is not useful. It is certainly true that, given enough of a side channel, coding the residual can give you perfect reconstruction of any original sequence. Even with fewer enhancement bits, a good reconstruction might be possible; however this was not investigated.

Remark 8
MSE seems to be well-correlated with picture quality in the context of our processing.

This conclusion is perhaps the most important result in order to practically implement the migration-path system. Focusing on the MSE is a good first order approach to determining how to partition blocks and to choose deinterlacing modes and parameters. It is important, though, to keep in mind that perceptual effects are important as well and must remain a key consideration in the processing.

Subjective results also confirmed what the objective results suggested regarding access to the original video: such access can greatly improve the performance of the migration path deinterlacing system. At bit rates such as 0.09 bits/pel (28% overhead) for the `mall`, 0.12 bits/pel (38% overhead) for the `girl`, and 0.093 bits/pel (29% overhead) for the `tulipstext` sequences, the pictures are clearly inferior to those shown in Fig. 4.8, Fig. 4.9, and Fig. 4.10 which are at much lower rates. For the `tulipstext` sequence, because inaccurate prediction of the motion vectors reduces the selection frequency of the FMV mode, we found that the blocking artifacts, although still present, were somewhat reduced; however, the reconstruction of the lettering had clearly deteriorated. As mentioned in the previous section, these experiments were conducted in a way which produced a best-case picture for the situation where no original is available. Thus, in a more realistic setting, the advantage to having the original would be even more pronounced.

Chapter 5

Conclusions

5.1 Summary

The deinterlacing experiments discussed in the previous chapter illustrate not only the feasibility of using a side channel to aid in deinterlacing but also the benefit in doing so. Furthermore, these experiments show that even with only a few simple deinterlacing modes available—and no residual encoding—a great improvement can be realized over intraframe deinterlacing. Each of the deinterlacing modes describes a deinterlacing algorithm which is applied to a block of pixels. By comparing the deinterlacing performance of each mode upon a particular block, we can determine and use the best mode for that block. Two of the improvements to deinterlacing implied by these experimental results are 1) the gain from using variable-size deinterlacing blocks, and 2) the gain from having a progressive original at the transmitter.

Variable-size blocks add efficiency to deinterlacing by allowing larger blocks in easier-to-code regions and smaller blocks in more difficult areas. The end result is that the bits of side information are spatially concentrated where they are most needed. Also, if there were sufficient frame buffering done, the side information could be unequally distributed temporally as well. By efficiently using these bits, it is possible to improve picture quality while reducing the side channel requirement. The areas of video sequences where this is most apparent are around the edges of objects, where larger blocks overlap regions of distinctly different characteristics, and in finely-detailed areas which are only well-predicted by small regions. Using fixed-size blocks of a sufficiently small area would also enable one to deinterlace well around the edges of objects and in detailed regions; however, such a strategy would incur a high bit requirement in those larger regions which do not need to be so finely divided in order to be well-predicted and interpolated.

The second way in which variable-size blocks add efficiency is they provide a straightforward method to fully use all the side information allowable. For a given number of supplementary bits, a frame can be subdivided up to the point when those bits are exhausted. If more bits are available, a finer subdivision can be performed. When less bits are usable, a coarser partitioning is done. This scheme allows for full consumption of the additional bits; this would not necessarily be true in the situation where fixed-size blocks were used. In the latter case, the number of sub-blocks is constant, therefore the number of bits available may be more than needed to code the deinterlacing information in which case some bits may go unused. Alternatively, the number of available bits may be less than needed to code the deinterlacing information, in which case only partial information can be transmitted. Either way, the deinterlacing efficiency will be reduced.

Having access to the progressively-scanned original at the encoder is of great benefit for a migration path system. While it is certainly possible to encode deinterlacing information without having a progressive original, the task is much simplified by its presence. A key advantage to having the original is the improved accuracy in computation of motion vectors. Because the lines to be estimated are known, it is easy to find the block which best matches the lines of interest. When the original is not available, the lines being predicted must be approximated, yielding substandard estimates. In the absence of the original, a more sophisticated attempt at determining scene motion may be desirable. It should be noted that the migration-path structure is suitable for such processing, since complicated calculations and high frame storage can be accomplished at the encoder only, keeping costs of decoders down; however, the encoder might have to be extremely complicated in order to accurately estimate motion. Similarly, having the original simplifies the task of determining which interlacing mode to choose. Using the original to predict the missing lines is a simple process which may obviate the need for complicated deinterlacing modes.

Finally, we can conclude from our experiments that a noticeable improvement in deinterlacing can be realized with a small amount of side information and that a simple error criterion such as the mean-square error is an effective first-order approach to determining deinterlacing parameters.

5.2 Future Directions

As terrestrial HDTV begins to be broadcast over the next few years, interest in the migration path will be at a peak. Although it may take some time for additional bandwidth to be available for the broadcasters, cable and satellite companies may be able to implement systems sooner. Regardless, there is a substantial amount of development which must be done before an efficient migration-path system can be effective. First, our results concerning the advantages of such a system must be confirmed by incorporating the migration-path concept into a real MPEG coder, determining deinterlacing information based upon coded data rather than the uncoded sequences in our experiments. Once an MPEG coder is in use, experiments should be performed to decide the tradeoff between using digital data to code the MPEG-based standard video or to code the migration-path stream; that is, if we are given 20 Mbps to encode video, should all 20 Mbps be used to encode the standard video, or should a lesser amount (*e.g.*, 18 Mbps) be used to code the standard video and the remaining data (2 Mbps) be used to represent migration-path information. Second, efficiency of the system can be studied in various ways, to include weighing selection of FMV mode by its overhead requirement in addition to just its reduction in error and using the standard video motion vectors. Third, incorporation of perceptual modeling for mode selection and partitioning may help to improve picture quality while reducing the side channel. Finally, investigation of the migration path concept for use with other transmission formats should be performed. If the 1080/30/PS transmission format is used, it may be desirable to temporally upsample the received signal to 1080/60/PS. The determination of parameters for efficient processing will be quite different than for the deinterlacing problem. For example, when deinterlacing, intraframe information is present for use; however, this is not the case for temporal upsampling. Other format conversions may require their own distinctive processing. Investigation in this area is warranted.

The development of the US high-definition television standard has opened the door for a wide variety of technological development. Although the new standard has been developed and is now being implemented, the HDTV migration path will receive considerable research attention over the next several years to come.

Appendix A

Simulation Details

This appendix summarizes the details regarding the simulations described in chapter 4 in order that the results may be reproduced, if desired.

Motion vectors were pre-computed and stored for each sequence to save computation time. These vectors were computed at three resolutions: 16 x 16, 8 x 8, and 4 x 4. To compute the vectors, each sequence was converted from the RGB format to its luminance (Y) component. Each frame of the sequence was then divided into the desired block size. The previous frame was interlaced and then each line was interpolated using the Martinez-Lim algorithm. For each block in the current frame, we selected the motion vector which minimized the MSE between the same size block in the previous frame. As stated in chapter 4, motion vectors were computed to full-pixel accuracy with maximum displacements of 32 horizontally and 16 vertically. The MSE was computed using the lines of the block in the current frame which would be missing due to the interlacing process and the corresponding lines from the block of the previous frame. For the experiments where access to the original video was tested, the lines of the current frame used in the computation of motion vectors were not the original video lines but instead were constructed using linear interpolation from the spatially adjacent lines.

In order to conduct the bit-rate vs. MSE experiments, we began by deciding into how many total blocks we would divide each frame and by selecting which subset of the deinterlacing modes we would use. Each frame was then interlaced and divided into 16 x 16 blocks.[16,17] For each block, the mode and parameters which minimized the deinterlacing error as compared to the original video were selected and stored. The blocks were then sorted by deinterlacing error, and, the 250 blocks with the highest error were subdivided (if the total number of blocks did not exceed the number originally chosen). Those 250 blocks were divided into quarters (yielding 1000 new blocks) and deinterlaced similarly. This process continued until the frame was divided into the chosen number of blocks; however, the smallest that blocks were ever divided was into 2 x 2 pixels, and when the motion-compensated mode was implemented for these blocks, the motion

---

[16] For the fixed block size experiments (corresponding to those points denoted by a triangle in Fig. 4.6), each frame was divided into 8 x 8 blocks and, of course, no subdivision occured.

[17] Half of the lines in each block were missing due to the interlacing.

vector for the next larger 4 × 4 block was used.

Once the frame had been divided into the pre-determined number of blocks, trivial cases of block subdivision were removed. For example, suppose there were a 4 × 4 block which would have been deinterlaced using the INTRA mode; however, because of the large deinterlacing error from this block, it was subdivided into four 2 × 2 blocks, and the deinterlacing mode for each was the INTRA mode. Then, because the subdivision didn't improve the deinterlacing at all, the 4 × 4 block was restored and represented as one INTRA block.

While each frame was processed as described above, three sequences were generated: a deinterlacing-mode sequence, a frame-partition sequence, and a motion-vector sequence (only for those cases when the motion-compensated mode was selected). The sample entropy of these streams was computed as described in chapter 4. From this sample entropy, we estimated the number of bits per pixel required to code the interlaced lines. Finally, the deinterlacing information was used to produce a recontructed video sequence, from which the MSE with the original sequence was determined. This gave us the necessary data to plot the graphs in Fig. 4.6 and Fig. 4.7.

Appendix B

Experimental Data

This appendix provides the data from which the graphs and numerical results in chapter 4 were determined.

Sequence	**Deinterlacing Modes**							
	INTRA, FZOH		INTRA, FZOH BZOH		INTRA, FMV FZOH		INTRA, FMV FZOH, BZOH	
	MSE	**Bits/pel**	**MSE**	**Bits/pel**	**MSE**	**Bits/pel**	**MSE**	**Bits/pel**
`mall`	16.65	0.0010	16.59	0.0014	14.15	0.0123	14.14	0.0124
	16.39	0.0034	16.27	0.0046	13.55	0.0219	13.54	0.0220
	16.20	0.0070	16.01	0.0096	13.15	0.0372	13.14	0.0376
	15.86	0.0183	15.55	0.0247	12.49	0.0873	12.47	0.0882
	15.61	0.0307	15.18	0.0423	11.92	0.1674	11.89	0.1691
	15.40	0.0449	14.87	0.0634	11.60	0.2432	11.55	0.2460
	14.96	0.0977	14.17	0.1463				
`gts`	59.22	0.0029	58.45	0.0041	29.12	0.0167	28.82	0.0169
	55.50	0.0050	54.22	0.0072	20.71	0.0301	20.30	0.0305
	54.46	0.0083	52.77	0.0121	16.10	0.0540	15.75	0.0549
	48.34	0.0268	44.28	0.0381	11.70	0.0862	11.51	0.0867
	46.30	0.0422	41.28	0.0601	9.11	0.1407	8.79	0.1424
	44.84	0.0581	39.15	0.0833	7.60	0.2041	7.26	0.2077
	41.84	0.1175	34.74	0.1727	6.80	0.2683	6.44	0.2748
`tulipstext`	224.70	0.0044	217.98	0.0055	142.82	0.0095	137.21	0.0100
	220.21	0.0067	212.06	0.0089	118.24	0.0228	111.34	0.0228
	210.49	0.0124	198.42	0.0175	90.74	0.0430	82.26	0.0425
	187.09	0.0305	163.46	0.0444	71.70	0.0668	62.48	0.0639
	165.35	0.0520	131.34	0.0740	41.65	0.1275	32.39	0.1145
	145.39	0.0738	103.35	0.1028	21.18	0.2017	12.99	0.1737
	101.19	0.1525	47.53	0.2031	12.82	0.2608	5.76	0.2165

Table B.1 Variable Block Size Data This is the data which generated the lines in Figs. 4.6(a), (b), and (c), where variable-size blocks were used.

Sequence	MSE	Bits/pel
`mall`	12.98	0.0987
`girl`	18.45	0.0862
`tulipstext`	69.87	0.0660

Table B.2 Fixed Block Size Data This is the data for the experiments in Figs. 4.6(a), (b), and (c) which used fixed (8 × 8) blocks and all four deinterlacing modes (the points denoted by triangles).

Sequence	MSE
`mall`	16.91
`girl`	78.53
`tulipstext`	284.06

Table B.3 Intraframe Deinterlacing Error This is the mean-square error from intraframe deinterlacing using the Martinez-Lim algorithm.

Sequence	MSE	Bits/pel
`mall`	15.70	0.0080
	15.24	0.0162
	14.99	0.0296
	14.68	0.0678
	14.27	0.1150
	13.99	0.1648
`gts`	54.81	0.0180
	50.56	0.0258
	48.92	0.0407
	44.91	0.0587
	39.95	0.0963
	36.82	0.1485
	34.60	0.1995
`tulipstext`	217.04	0.0071
	209.62	0.0149
	193.26	0.0328
	152.29	0.0932
	118.71	0.1433
	91.16	0.1817
	38.63	0.2918

Table B.4 No Progressive Original Data This is the data which generated the upper lines in Figs. 4.7(a), (b), and (c), where the motion vectors were computed without access to the 1080/60/PS original.

Appendix C

Subjective Testing

Subjective tests were conducted with a panel of four expert viewers. Two series of tests were conducted. In the first series, the viewers were asked to rank, on a scale of 1 (lowest) to 10 (highest), the different pictures in terms of picture quality. In the second series, the panelists were shown different pairs of sequences and asked to determine which was better, or if the picture quality was the same.

Sequence	**A**	**B**	**C**	**D**	**E**	**F**
	Original	Intraframe	Fixed Block Size	Variable Block Size		No Original
`mall`	8.25	4.0	.099 Bits/pel 6.75	.088 Bits/pel 6.5	.038 Bits/pel 6.5	.090 Bits/pel 5.25
`girl`	9.5	2.75	.086 Bits/pel 5.5	.087 Bits/pel 7.25	.031 Bits/pel 5.25	.122 Bits/pel 4.75
`tulipstext`	9.25	3.25	.066 Bits/pel 8.25	.064 Bits/pel 6.75	.043 Bits/pel 5.75	.093 Bits/pel 7.25

Table C.1 Subjective Test Results This table shows the average score given to each of the sequences studied in the first series of subjective tests.

The sequences used in the subjective tests are shown in Tab. C.1. The first sequence (column A) was the enhanced-resolution original and the second sequence (column B) was the result from intraframe deinterlacing. The next three sequences were outputs of the migration-path system using all four deinterlacing modes with access to the original video. The last sequence (column F) was obtained by using the migration-path system but having no access to the enhanced-resolution original. Because the different viewers had different scales on which they based their decisions, it is difficult to attach too much significance to the scores in Tab. C.1; however, some general conclusions can be drawn. One such conclusion is that the migration-path results clearly improve upon the intraframe results, even at low digital data rates (column E). Also, the migration path results where the transmitter has access to the original video (specifically, column E for `mall` and `girl`, column C for `tulipstext`) outperform the results where no original is used to compute the motion vectors (column F) even though the latter uses much higher bit rates. The result in column E for `tulipstext` performs worse than in column F due to the blocking artifacts discussed in chapter 4.

Sequence		
`mall`	**C**	**D**
	1.5	2.5
	C	**E**
	1.5	2.5
	D	**E**
	2.5	1.5
	E	**F**
	4	0
`girl`	**C**	**D**
	0	4
	C	**E**
	2	2
	D	**E**
	4	0
	E	**F**
	4	0
`tulipstext`	**C**	**D**
	4	0
	C	**E**
	4	0
	D	**E**
	2	2
	C	**F**
	4	0

Table C.2 Pairwise Sequence Comparison This table shows the score given to each of the sequences studied in the second (pairwise) series of subjective tests.

Due to minor variations in each person's relative evaluation scale, the numerical results from the first series of experiments only give us a rough feeling for which sequences were better (*e.g.*, for `mall`, does a score of 6.75 for sequence C really indicate that it has better picture quality than D or E, which score 6.5). To get a more accurate judgement for which sequences look the best, we conducted pairwise comparisons. In each of these comparisons, one point was given to the picture deemed to be better, and one half point was given to each sequence if they were judged to be of equal quality. The results of these comparisons are shown in Tab. C.2.

The results in Tab. C.2 are consistent with what we expected from the objective tests. In particular, `mall` and `girl` show that the MSE is reasonably well-correlated with visual quality. In fact, three out of four panelists determined that the quality of C and E were the same for `mall`, and the one who differed noted only the slightest difference. For `girl`, all the viewers thought C and E were of the same quality. For `tulipstext`, blocking artifacts in the variable block size experiments caused these sequences to be judged inferior to the fixed block size; however, the viewers unanimously agreed that processing with access to the original was clearly superior to processing without it for all sequences. It should be noted, however, that for `tulipstext`, the rotating color bar in the sequence corresponding to column F actually looked better than that for column C due do the degenerate vertical motion of an odd number of pixels per field.

References

---

[1] Advanced Television Systems Committee. *ATSC Digital Television Standard,* September 16, 1995.

[2] Advanced Television Test Center, Inc. *digital* HDTV Grand Alliance System Record of Test Results. Alexandria, VA, October 1995.

[3] Peter J. Burt and Edward H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications,* COM-31(4):532–540, April 1983.

[4] Shih Chang Hsia, Bin Da Liu, and Chien-Hsiu Huang. An NTSC to HDTV video conversion system by using the block processing concept. *IEEE Transactions on Consumer Electronics,* 40(3):216–224, August 1994.

[5] L. Chiariglione, editor. *Signal Processing of HDTV,* P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands, 1988. International Workshop on Signal Processing of HDTV, Elsevier Science Publishers B.V.

[6] L. Chiariglione, editor. *Signal Processing of HDTV, II,* P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands, 1990. International Workshop on Signal Processing of HDTV, Elsevier Science Publishers B.V.

[7] C. K. P. Clarke. Interpolation. In C. P. Sandbank, editor, *Digital Television,* pages 317–374. John WIley & Sons, Ltd., 1990.

[8] Gerard de Haan and Henk Huijgen. New algorithm for motion estimation. In Chiariglione [6], pages 109–116.

[9] Danny de Vleeschauwer and Ignace Bruyland. Non-linear interpolators in compatible HDTV image coding. In Chiariglione [5], pages 431–438.

[10] T. Doyle. Interlaced to sequential conversion for EDTV applications. In Chiariglione [5], pages 421–430.

[11] E. Dubois and L. Chiariglione, editors. *Signal Processing of HDTV, IV,* P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands, 1993. International Workshop on Signal Processing of HDTV, Elsevier Science Publishers B.V.

[12] Manfred Ernst and Thomas Reuter. A universal standards converter concept for HDTV applications. In Chiariglione [6], pages 51–59.

[13] L. Capodiferro *et. al.* A screen format converter for HDTV. In Yasuda and Chiariglione [38], pages 599–606.

[14] M. A. Isnardi *et. al.* A single channel, NTSC compatible widescreen EDTV system. Technical report, David Sarnoff Research Center, 1987.

[15] Ti-Hao Chiang *et. al.* Compatible coding of digital interlace HDTV using prediction of the even fields from the odd fields. In Yasuda and Chiariglione [38], pages 523–530.

[16] Federal Communications Commission. Fourth report and order: FCC 96-493, December 27, 1996.

[17] International Organisation for Standarisation. *Generic Coding of Moving Pictures and Associated Audio Information: Video*, October 4, 1996. Recommendation H.262, Draft Interational Standard.

[18] Daryl K. Jensen and Dimitris Anastassiou. Digitally assisted deinterlacing for EDTV. *IEEE Transactions on Circuits and Systems for Video Technology*, 3(2):99–106, April 1993.

[19] Michel Kerdranvat. Hierarchical motion estimation and motion information encoding. In Chiariglione [6], pages 101–108.

[20] Milton S. Kiver and Milton Kaufman. *Television Electronics*. Van Nostrand Reinhold Company Inc., 1983.

[21] Tero Koivunen. Motion detection of an interlaced video signal. *IEEE Transactions on Consumer Electronics*, 40(3):753–760, August 1994.

[22] Jae S. Lim. A proposal for an HDTV/ATV standard with multiple transmission formats. *SMPTE Journal*, 102(8), August 1993.

[23] Jae S. Lim. A migration path to a better digital television system. *SMPTE Journal*, 103(1), January 1994.

[24] Jae S. Lim and Lon E. Sunshine. HDTV transmission formats and migration path. *International Journal of Imaging Systems and Technology*, 5(4):286–291, Winter 1994.

[25] Bede Liu and André Zaccarin. A comparative study of coding of interlace sequences with motion compensation. In Yasuda and Chiariglione [38], pages 423–430.

[26] D. M. Martinez and J. S. Lim. Spatial interpolation of interlaced television pictures. In *Proceedings of ICASSP*, 1989.

[27] Michael T. Orchard and Aggelos K. Katsaggelos. Backward motion compensation for interlaced HDTV. *Signal Processing:* Image Communications, 5(5-6):487–501, December 1993.

[28] Michael O. Polley. *Spectrum-Efficient Modulation and Channel Coding for Terrestrial Broadcasting*. PhD thesis, Massachusetts Institute of Technology, February 1996.

[29] R. S. Prodan. Multidimensional digital signal processing for television scan conversion. *Philips Journal of Research*, 41(6), 1986.

[30] Jouni Salonen. Edge and motion controlled spatial upconversion. *IEEE Transactions on Consumer Electronics*, 40(3):225–233, August 1994.

[31] Ralf Schäfer, Peter Kauff, Shing-Chi Chen, and Matthias Leptin. A comparison of different HDTV standards using subjective and objective criteria. In Chiariglione [6], pages 61–71.

[32] William F. Schreiber. The FCC digital television standards decision. Letter to the FCC, January 28, 1997.

[33] L. Stenger, L. Chiariglione, and M. Akgun, editors. *Signal Processing of HDTV, V*, P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands, 1994. International Workshop on Signal Processing of HDTV, Elsevier Science Publishers B.V.

[34] The Grand Alliance. The U.S. HDTV standard. *IEEE Spectrum*, pages 36–45, April 1995.

[35] Lawrence J. Thorpe. HDTV production – the technical dilemma within the progressive versus interlace debate. In Yasuda and Chiariglione [38], pages 501–512.

[36] Feng-Ming Wang, Dimitris Anastassiou, and Arun N. Netravali. Time-recursive motion compensated deinterlacing. In Chiariglione [6], pages 635–642.

[37] Dietrich Westerkamp and Hartmut Peters. Comparison between progressive and interlaced scanning. In Chiariglione [5], pages 15–23.

[38] H. Yasuda and L. Chiariglione, editors. *Signal Processing of HDTV, III*, P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands, 1992. International Workshop on Signal Processing of HDTV, Elsevier Science Publishers B.V.