Motion Compensated Interpolation For Television Standards Conversion

Eric Coll

May, 1986

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree
Master of Engineering

Department of Electrical Engineering
McGill University, Montréal, Canada

Abstract

A description of the structure of the ideal motion compensated television standards converter, consisting of a segmentation process, a motion estimation process, and an interpolation process is presented. The theory of linear processing techniques for television standards conversion, practical implementation of linear processing techniques, and the designs of existing standards converters are reviewed. An experimental motion compensated standards converter which allows the examination of the interpolation problem apart from the segmentation and efficient motion estimation problems is implemented. Interpolation apertures for various input conditions are described, and a rule for adapting the vertical aperture to the amount of aliasing in the input signal is given. Results arising from the development of the experimental converter, and from subjective testing of various converters are presented.

Sommaire

Une description de la structure d'un convertisseur idéal à compensation de mouvement pour les standards de signaux télévisés, consistant en des processus de segmentation, d'estimation du mouvement et d'interpolation, est présentée. La théorie des techniques linéaires de conversion de ces standards, l'exécution pratique de ces techniques, ainsi que la description des convertisseurs commerciaux existant sont passés en revue. Un convertisseur expérimental à compensation de mouvement qui permet l'étude des problèmes d'interpolation sans tenir compte des problèmes de la segmentation et de l'estimation efficace du mouvement est réalisé. Quelques ouvertures pour l'interpolation sont décrites et une règle pour l'adaptation des ouvertures verticales à la grandeur de l'aliasing des signaux d'entrées est définie. Les résultats provenant du développement du convertisseur expérimental ainsi que de tests subjectifs de différents autres convertisseurs sont présentés.

Acknowledgements

I would like to thank INRS-Télécommunications for supplying the facilities for the research and production of this thesis, and my fellow students for animating them.

Table of Contents

List of Figures

Chapter 1 Introduction

The goal of television standards conversion is to change the field rate (the number of fields per second) and the line density (the number of lines per field) of one television standard to those of a different standard without introducing distortion or artifacts to the output sequence.

Two distinct reasons motivate the development of the standards converter. One reason is for the exchange of information in the form of television broadcasts between areas which use different broadcast, transmission and display standards. In North America, a 525-line, 60 Hz standard is used, while in Europe and many other parts of the world, various 625-line, 50 Hz standards are used. Apart from information interchange between areas which use different television standards, a second reason for the development of the standards converter is for conversion between new and old standards within a particular area. If a new higher definition production standard were to be introduced, material produced using the new standard could be converted to old transmission and display standards before broadcast. If a new transmission and display standard were to be accepted, material produced using obsolete equipment could be converted to the new transmission standard. The high-quality standards converter is an essential ingredient in the acceptance of a new 1125-line, 60 Hz High Definition Television (HDTV) standard in Europe, where 50 Hz systems are currently used [1].

The state of the art in television standards conversion in terms of installed operating equipment is the Advanced Conversion Equipment (ACE) converter developed by the British Broadcasting Corporation (BBC) [2]. This converter is essentially a two dimensional vertical-temporal digital low pass filter, changing the field rate and line density through interpolation. The interpolation consists of taking a weighted linear combination of a number of lines in the temporal and spatial neighborhood of each desired output line to produce the output line. It requires the digital storage of four fields, and is referred to as a four-line, four-field standards converter. It is currently in general use in England, for live broadcasts as well as the exchange of programs on video tape. The Japanese national broadcasting corporation, NHK, has announced an experimental motion-compensated standards converter [3, 4] which incorporates principles similar to those proposed in this thesis. Little technical documentation on this NHK converter has been released. Investigation of the interpolation aspects of motion-compensated television standards conversion remains a strong area of research.

Two problems are apparent when viewing sequences converted with the ACE converter. The first problem is a noticeable loss of resolution when viewing 'normal' scenes. In this converter, the two-dimensional low pass filter has been specified to trade off motion judder (jerkiness) associated with large area motion for a very distinct loss of resolution. The second problem occurs when viewing sequences that contain a rapid pan of a high-contrast scene such as advertisements on hockey boards. In these sequences, the judder becomes quite evident, and moving objects are seen to jerk across the screen at the beat frequency between the two field rates.

The technique proposed in this thesis is to estimate the velocity of moving areas in order to allow the interpolation of output fields based on motion-compensated input fields. Estimates of the motion between a given output temporal position and temporally neighboring input fields are generated; the interpolated field is then generated by projecting input fields onto the temporal position of the interpolated field in the direction of the motion estimates. This procedure places objects in the correct spatial position in the interpolated field, completely eliminating judder, while maintaining a much higher degree of resolution than that afforded by the ACE converter. This technique is applicable to sequences containing general motion if the input fields are segmented into still and moving areas, and a separate motion estimate is generated for each moving area.

This thesis examines the details of motion-compensated interpolation apart from the segmentation and efficient motion estimation problems, by restricting the input sequences to be sequences containing uniform translational motion over the entire image, where a single motion estimate for the entire field can be reliably obtained. Eliminating the segmentation problem entirely and simplifying the motion estimation problem allowed the implementation of an experimental motion compensated standards converter. Using this converter, and the digital video capture/display facilities at the INRS/BNR Nun's Island Research facility, the structure of the ideal motion compensated standards converter was defined. Fundamental aspects of the interpolation process were investigated, and a scheme for adapting the vertical interpolation aperture to the amount of vertical aliasing in the input signal was developed. The experimental converter, along with existing converters, was used to process several critical test sequences. The results of subjective tests comparing the performance of the converters indicate superior treatment of sequences containing high speed motion and sequences containing high detail by the motion compensated standards converter.

The rest of the thesis is organized into 4 chapters. Chapter 2 describes the theory of standards conversion. Various techniques used for the conversion of television standards are described. A frequency sampling method for specification of interpolation apertures is described, and a brief survey of available types of motion estimation is given The chapter is concluded with a review of the design of several existing standards converters. Chapter 3 describes the proposed system, giving a complete description of the structure of the ideal motion compensated standards converter, as well as the experimental motion compensated converter constructed for this thesis. Chapter 4 gives the results of the development of the converter and the results of subjective tests comparing the experimental motion compensated converter to existing converters. Chapter 5 concludes the thesis.

Chapter 2 Theoretical Background

There are several different approaches to conversion of television standards, each trading off complexity, loss of resolution, and the addition of artifacts to varying degrees. This chapter provides the theoretical background for television standards conversion, both traditional linear techniques and the motion compensated technique proposed in this thesis. The structure of television signals is described, and the concept of a television standard is introduced. The theoretical basis for television standards conversion using linear processing is examined in detail, followed by a description of the options available in the practical implementation of a standards converter. A frequency sampling method of filter specification is discussed, and a brief survey of motion estimation algorithms is given. The chapter is concluded with an overview of the specific techniques used in several current standards converters.

2.1 The Television Signal

Television signals are three dimensional signals, having components in the vertical and horizontal dimensions as well as a temporal component. To produce a television signal, the image in a camera is scanned line by line, from top to bottom. One complete line scan of the image in the camera comprises a *frame*. In 2:1 *interlaced* systems, a complete scan of the image is formed in two passes. Each pass forms a *field*, which contains half of the number of lines in a frame. On the first

pass, even numbered lines are scanned, forming an 'even' field; on the second pass, odd numbered lines are scanned, forming an 'odd' field. These fields are transmitted as the output sequence at twice the frame rate. Sequential fields, when interlaced, form a frame.

A colour television signal is comprised of three independent colour component signals. These signals represent the intensities of three primary colours which are simultaneously scanned in the camera. For transmission, a luminance signal and two colour-difference signals are formed by weighted sums of the intensities of the three colour components. These signals are frequency multiplexed to obtain a composite signal, with the colour difference signals quadrature modulated on a subcarrier frequency.

The television signal can be thought of as a sequence of fields which are temporal samples of the spatially and temporally continuous image in the camera. Each field is composed of a number of lines, which are vertical samples of the spatially continuous image in the camera at that temporal sampling point.[1] Digital storage and processing of the television signal requires the transformation of the signal which is discrete in the vertical and temporal dimensions but continuous in the horizontal dimension to a signal which is discrete in all three dimensions. This transformation is achieved by *sampling* and *quantizing* the continuous signal to produce a digital sequence. The resulting digital representation of each field is an array of digital values called picture elements or *pels*.

Composite signals are sampled at various rates; typically either three or four times the frequency of the colour subcarrier. It is possible to reduce the amount of chrominance information in a sequence of images with little or no subjective effect, especially if the sequence contains motion. Taking advantage of this fact, the

[1] Since the line scanning process takes a finite amount of time, each line is in fact a vertical sample of the image at a different point in time. However, since finite-duration sampling results in only a small, constant temporal offset from one line to the next, each field is usually considered without loss of generality to be a sample of the image at a single point in time

colour difference signals in most systems are chosen to occupy approximately one half the bandwidth of the luminance signal [5]. The two colour difference signals may then be subsampled at a rate one half of that of the luminance signal since they occupy less bandwidth than the luminance signal. Component signals, when digitally processed or stored without being first formed into a composite signal, are sampled at the studio standard sampling frequency of 13.5 MHz

A typical structure for the digital storage of each field is a three dimensional array: the horizontal dimension indexes samples of each line (pels), the vertical dimension indexes each line; and the depth dimension indexes samples of the luminance and colour difference signals. Since the colour difference signals are subsampled 2:1, all of the samples of both colour difference signals may be contained in one array the same size as the luminance array; each line in the colour difference array would be composed of alternating pels from each of the colour difference signals.

2.2 Television Standards

The *line density* (the number of lines per frame), the *field rate* (the number of fields per second), the spectra of the three primary colours, and the weighting coefficients used to derive the composite signal are defined in a *standard*. Families of standards with similar field rates, line densities, and transmission schemes comprise *systems*. In North America, the standard used is referred to as NTSC, after the National Television System Committee [5]. The NTSC standard defines a 525 lines per frame, 60 fields per second, 2:1 interlaced system. A field rate of 59.94 Hz was chosen to correspond with the 60 Hz alternating current power supply in North America. In Europe, two systems are prevalent: the PAL system, and the SECAM system [5]. Both are 625 line, 50 fields per second, 2:1 interlace systems; the 50 Hz field rate reflects the power supply frequency in Europe. The *line rate* (the number of lines per second), of all three of these systems is very close; all require roughly the same channel bandwidth and transmit information at similar rates.

A system which has a substantially higher line rate than the 525/60 and 625/50 systems is the proposed 1125 line, 60 Hz, 2:1 interlace High Definition Television (HDTV) system. This system is intended to provide resolution approaching that of 35 mm film for production and display. NHK, the Japanese national broadcasting corporation, is currently broadcasting an 1125/60 system called MUSE [6]. It is expected that HDTV will eventually be accepted in North America for high definition production and specialized applications at first, and perhaps for general transmission and display uses in the future.

2.3 Linear Television Standards Conversion Theory

The fundamental problem in television standards conversion is to change the field rate and the line density of the input standard to those of the output standard without introducing *artifacts* or losing *resolution*. Various analog methods, such as optical conversion, where a camera using the output standard is pointed at a monitor displaying the input standard, and converters that use quartz delay lines to implement line and field stores were used before the advent of digital hardware. In this thesis, techniques applicable primarily to digital processing are examined.

The conversion of a television signal from one standard to another may be viewed as conversion from one *sampling structure* to another. Figure 2.1 shows the vertical-temporal time domain sampling structure of two different systems, the 525 line/60 fields per second NTSC system, and the 625/50 PAL system. For all practical purposes, (neglecting the fact that the NTSC temporal sampling rate is actually 59.94 Hz), this block illustrates the "lowest common denominator" between the two systems; one cycle in the vertical and temporal dimensions between the coincidence of the two sampling structures. To convert from one standard to another, it is necessary to find the values of the luminance and chrominance signals at the vertical-temporal points which comprise the desired sampling structure.

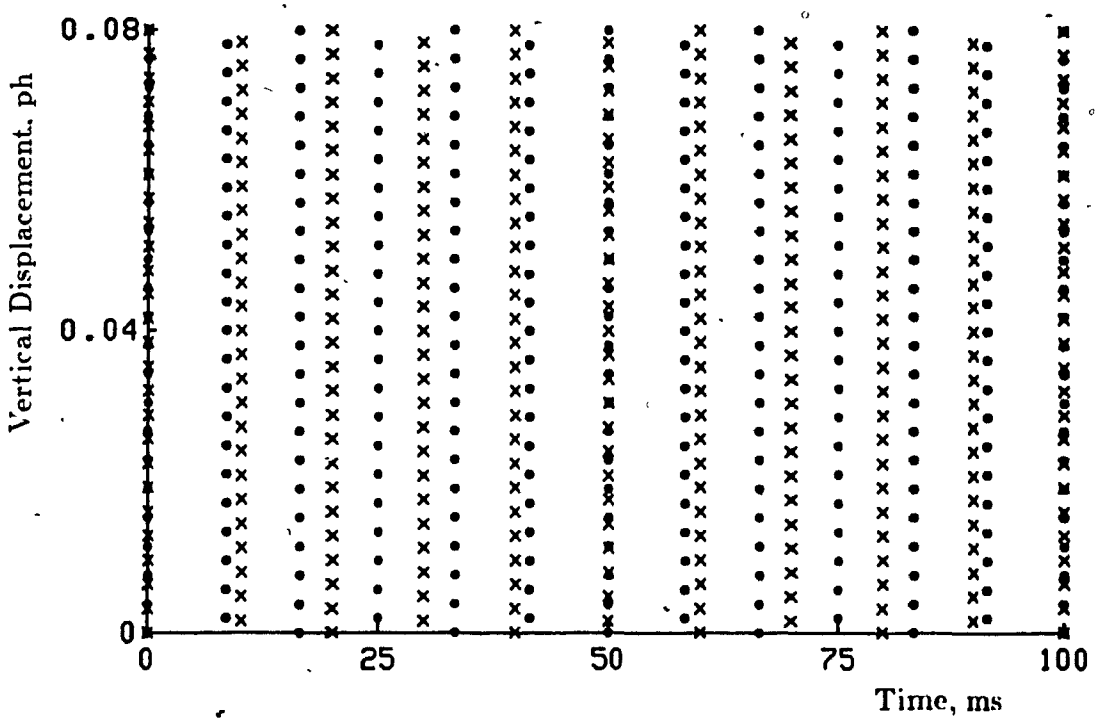

Figure 2.1 Time Domain Vertical-Temporal Sampling Structures
o - NTSC sampling structure, x - PAL sampling structure

The traditional method of sampling structure conversion is to use linear processing. Linear processing techniques use linear filters to convert from one vertical-temporal sampling structure to another. If the original signal has been bandlimited and sampled at a frequency sufficient to ensure that the sampled signal is not aliased, then perfect sampling structure conversion could theoretically be obtained. However, practical limitations such as the fact that the signal from the vertical-temporal sampling process (the camera) is aliased, and the difficulty in finding perfect linear filters to perform the sampling structure conversion, mean that traditional standards conversion techniques produce imperfect output. Non-linear methods of standards conversion such as the motion-compensated technique described in section 2.4.1.2 replace linear temporal filtering techniques with a complex non-linear process that produces superior results. This section derives the theory of traditional linear sampling structure conversion.

Conversion from one vertical-temporal sampling structure to another vertical-temporal sampling structure may be performed as a two step procedure. In the first step, the three dimensional continuous signal that was imaged in the camera is reconstructed from the sampled signal with a reconstruction filter. In the second step, the continuous signal is resampled at the desired vertical and temporal frequencies, producing a discrete signal sampled at the rates defined in the output standard. In practice, a one-step operation is performed on the input samples to produce the output samples directly. It is instructive to examine the effects of the two step reconstruction-resampling procedure in the frequency domain to gain an insight into the requirements for the one step procedure. The theory will be introduced in one dimension, and then extended to two dimensions.

2.3.1 One Dimensional Reconstruction and Resampling

The process of *sampling* a continuous signal generates a sequence of samples of the signal which are values of the continuous function at discrete points. If we wish to change the sampling structure, that is, generate samples of the signal at different discrete points, the original signal may be reconstructed from the sequence of samples, and then resampled at the desired points to form a new sequence.

To see how the continuous signal may be reconstructed, the effects of the sampling process are examined. In the one dimensional case, a discrete sequence $x(n)$ is formed by sampling a continuous signal $x_a(t)$ with period T. The values $x(n)$ are considered to be instantaneous values of $x_a(t)$ at $\{t = nT, n = ..., -1, 0, 1, ...\}$. A frequency domain representation $X(e^{j\omega})$ of the sequence $x(n)$ is the Fourier Transform, given in [7] as

$$X(e^{j\omega}) = \frac{1}{T} \sum_{r=-\infty}^{\infty} X_a(\frac{j\omega}{T} + j\frac{2\pi r}{T}) \qquad (2.1)$$

where $X_a(j\omega)$ is the frequency domain representation of $x_a(t)$. The sampling process causes the frequency spectrum of the continuous signal to be *replicated* at integer multiples of the sampling frequency (Figure 2.2(b) and (c)).

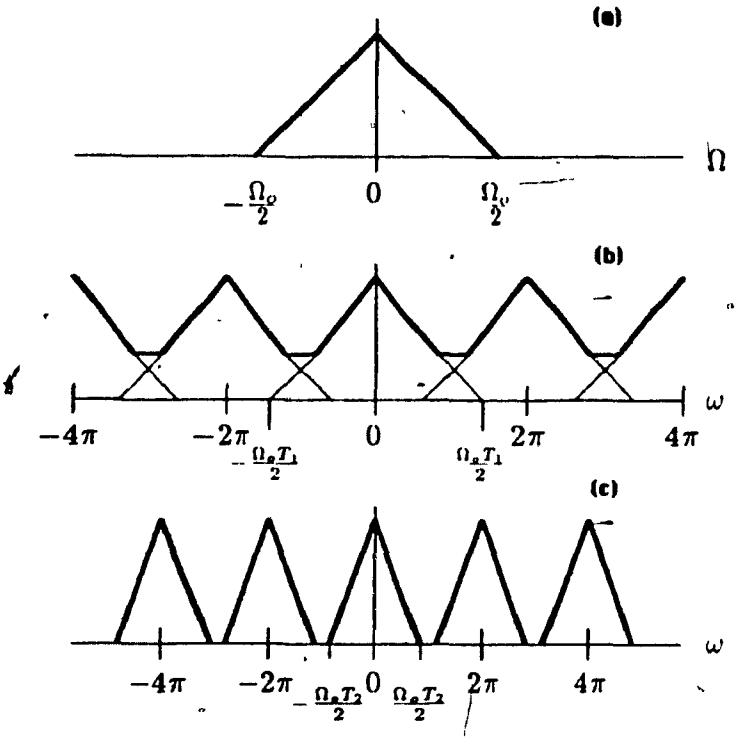

Figure 2.2 Aliasing in a Sampled Signal. (a) Frequency spectrum of continuous signal (b) Frequency Spectrum of signal aliased due to low sampling rate (c) Frequency spectrum of signal sampled at a rate sufficiently high to prevent aliasing (after [7]).

Because of the replication of frequency spectra, the sampling frequency must be at least twice the frequency of the highest frequency component in the continuous signal. If the signal is not *bandlimited*, that is, it has components at all frequencies, or if the sampling frequency is less than twice the highest frequency, component in the continuous signal, then the replicated spectra will overlap (Figure 2.2(b)) and the original signal cannot be recovered. This condition is known as *aliasing*.

To reconstruct the continuous signal $x_a(t)$ from the discrete sequence $x(n)$, it is necessary to restore the frequency spectrum of Fig. 2.2(c) to its original state (Fig. 2.2(a)) by passing the sequence through a reconstruction filter. The perfect

reconstruction filter must pass the baseband spectrum without attenuation, and reject completely all replicated spectra. The perfect reconstruction filter is the ideal low pass filter, which has unity gain in the passband, zero gain in the stopband, and a transition band of zero width (Figure 2.3).

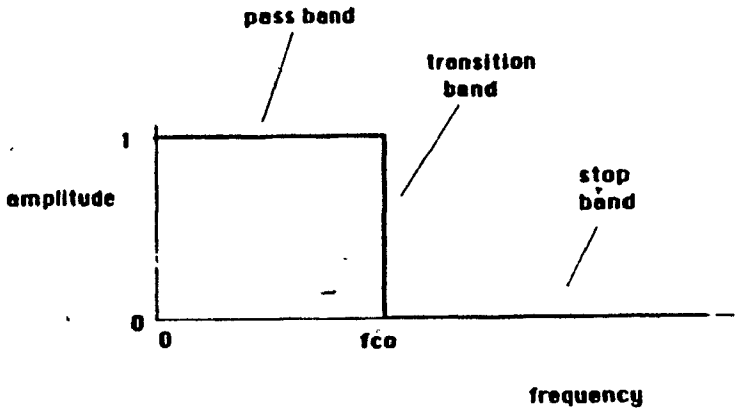

Figure 2.3 Ideal Low Pass Filter. The ideal low pass filter has unity gain in the passband, a transition band of zero width at the cutoff frequency f_{co}, and zero gain in the stopband.

If the reconstruction filter is imperfect, then the continuous signal is not reconstructed without distortion. If the reconstruction filter attenuates the baseband signal, then information from the signal is lost; if the reconstruction filter does not attenuate the stop band completely, then aliasing occurs because low frequency components from the replicated spectra are included in the high frequency areas of the reconstructed signal.

Once the continuous signal $x_a(t)$ is reconstructed from the sampled signal $x(n)$, the continuous signal may be resampled at the desired sampling frequency. This resampling process generates a new sequence $y(n)$ of samples of the continuous function $x_a(t)$, at points different than those of $x(n)$. In the frequency domain, the

baseband signal is again replicated, but at multiples of the new sampling frequency (Figure 2.4). If the resampling rate is reduced, aliasing as illustrated in Figure 2.2(b) may occur, necessitating prefiltering of the baseband signal to limit its bandwidth and avoid aliasing. This reconstruction and resampling process is a "hybrid" process involving mixed analog and digital processing. Sampling structure conversion may be performed much more easily using completely digital processing.

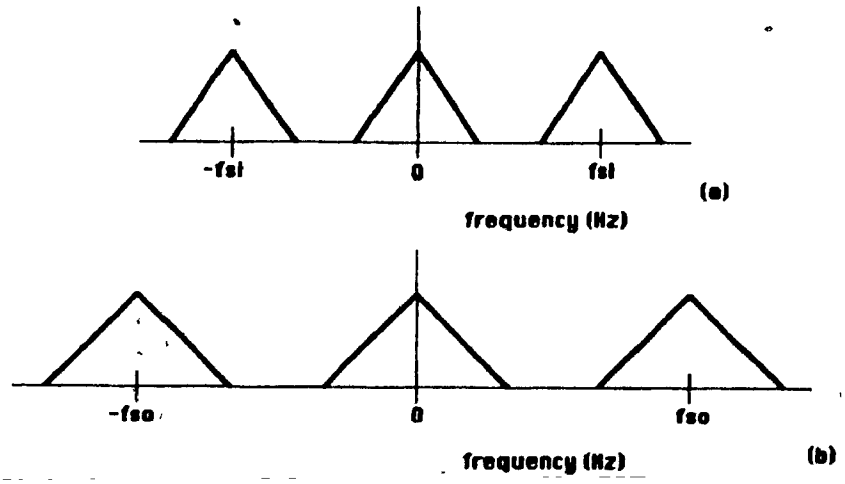

Figure 2.4 Frequency Domain Representation of Resampling The sampling process creates replications of the baseband frequency spectrum at multiples of the input sampling frequency f_{si}. When the sampling rate is changed to f_{so}, the entire frequency axis is scaled.

2.3.2 One Dimensional Digital Sampling Rate Conversion

Clearly, reconstruction of the continuous signal and then resampling it as a means of changing the sampling rate is inefficient. The alternative is to digitally process the input sequence of samples $x(n)$ to *interpolate* the output sequence $y(n)$, without reconstructing the continuous waveform $x_a(t)$ in an intermediate stage.

In the two step process, the continuous signal $x_a(t)$ was reconstructed from the sampled signal $x(n)$ (in the frequency domain) by multiplying the spectrum $X(e^{j\omega})$ by that of the reconstruction filter $A(e^{j\omega})$, which extracts the baseband spectrum. In the time domain, this corresponds to convolving $x(n)$ with the impulse response of the (continuous) reconstruction filter $a(t)$. If the values of the discrete sequence $x(n)$ are considered to be instantaneous values of the continuous sequence $x_a(t)$ at integer multiples of the sampling period T_1, *i.e.*

$$n = pT_1 \tag{2.2}$$

then the reconstructed signal is

$$x_a(t) = \sum_{p=-\infty}^{\infty} x(p)a(t - pT_1). \tag{2.3}$$

The resampling of the continuous function $x_a(t)$ at a new set of discrete points spaced at multiples r of the new sampling period T_2, can be expressed as choosing values of $x_a(t)$ at a new set of discrete points $t = rT_2$, $-\infty < r < \infty$:

$$x_a(rT_2) = \sum_{p=-\infty}^{\infty} x(p)\ a(rT_2 - p\bar{T}_1) \tag{2.4}$$

and so the new discrete sequence $y(r)$ can be described as

$$y(r) = \sum_{p=-\infty}^{\infty} x(p)\ a(rT_2 - pT_1). \tag{2.5}$$

When the ratio T_2/T_1 is a rational number, then the values of the function $a(t)$ are required at an infinite number of *discrete* points. The continuous function $a(t)$ may then be replaced with a discrete function $a(s)$, and the "hybrid" digital-analog-digital processing system of section 2.3.1 becomes a true digital processing system, interpolating $y(n)$ directly from $x(n)$.

Since (in the absence of an aliased input signal) the perfect reconstruction filter $a(t)$ is the ideal low pass filter, the sampling structure conversion problem resolves

into finding realizable approximations to the ideal low pass filter which trade off loss of information from the baseband for the inclusion of aliased spectra. In standards conversion, a finite-duration, continuous time domain function quantized in position and amplitude, whose frequency response approximates the ideal low pass filter is used to perform sampling structure conversion. These functions are known as *aperture functions*, *interpolation apertures*, or simply *apertures*. The discrete function $a(s)$ as derived on the previous page is an infinite-length function, usually required on a very dense sequence of discrete points for television standards conversion. In practice, the function $a(s)$ is truncated to a finite length and defined on a relatively coarse spacing. When a value $a(rT_2 - pT_1)$ is required, the argument $(rT_2 - pT_1)$ is quantized to the nearest point upon which $a(s)$ is defined, and the value of $a(s)$ at the quantized position is taken as the value of the aperture function. If the value of the argument is outside the domain of the finite-length aperture, the value of the aperture is defined as zero.

2.3.3 Two Dimensional Reconstruction and Resampling

Conversion of television standards implies conversion from one vertical-temporal sampling structure to another vertical-temporal sampling structure. For standards conversion, the television signal is processed in the vertical and temporal dimensions and reconstruction and resampling involves the use of two dimensional vertical-temporal filters. Two dimensional filters and signals are analogous in every way to their familiar one dimensional counterparts, except that the two dimensional filters and signals are defined over grids or *lattices* of discrete points.

A 2:1 interlaced 60 Hz television signal, viewed in the vertical-temporal dimensions has a time domain sampling structure as shown in Figure 2.5(a). The sampling process causes the replication of the baseband frequency spectrum centered on the points shown in Figure 2.5(b) for an NTSC signal [8].

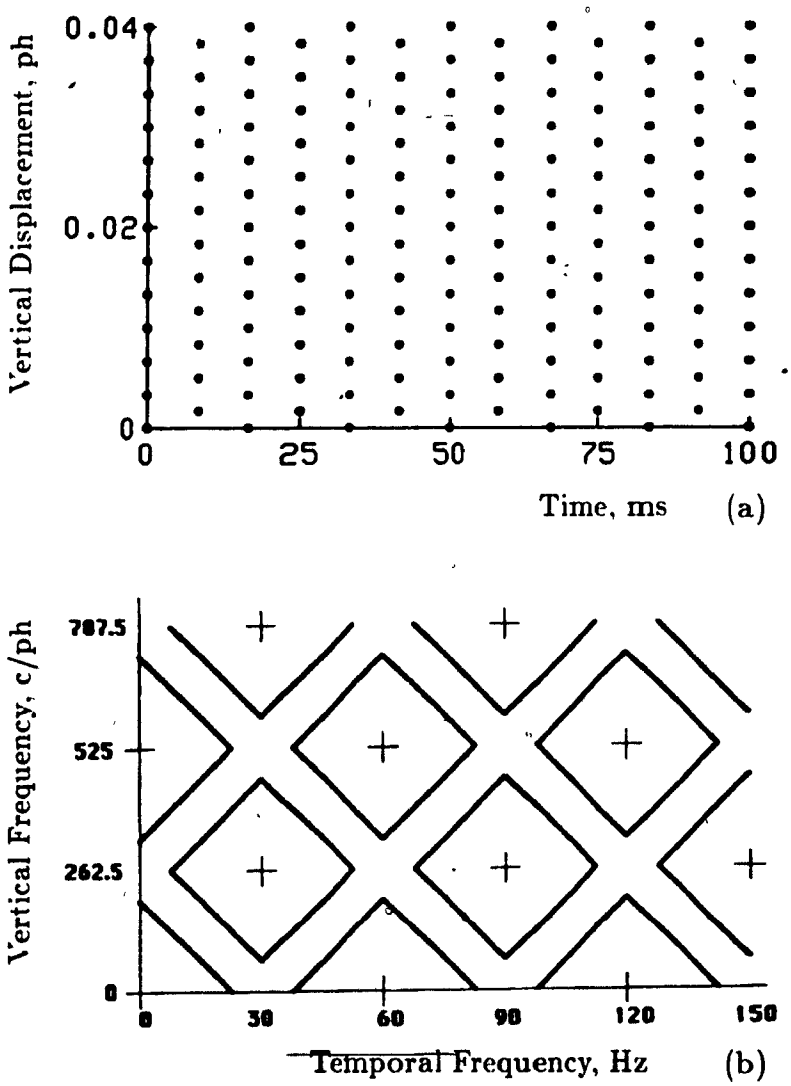

Figure 2.5 A Two Dimensional Vertical-Temporal Signal
(a) Time domain sampling structure
(b) First quadrant of frequency domain lattice

As in the one dimensional case, it is necessary to extract the baseband frequency spectrum using a reconstruction filter in order to reproduce the two dimensional continuous signal. The filtering may be performed as a cascade of two separable one dimensional filters, one in the vertical dimension and one in the temporal dimension, or a with a non-separable two dimensional filter. The use of separable one dimensional filters in cascade results in vertical and temporal aliasing, regardless of the order of the cascade because of the positioning of the replicated spectra. Figure 2.6 illustrates the extraction of the baseband signal in a two-step separable process. The spectrum of the input signal (Figure 2.6(a)) is filtered using a one dimensional vertical low pass filter with a cutoff frequency at about ± 262.5 c/ph, or half of the vertical sampling frequency. In the first quadrant, low vertical frequency components of the replicated spectra centered at 262.5 c/ph are passed along with the vertical baseband spectra. The resulting spectrum is shown in (b), with the aliased areas cross-hatched. Subsequent one dimensional filtering in the temporal dimension with a filter of cutoff frequency at about ± 30 Hz, or half of the field rate, produces the baseband spectrum as shown in (c). In the first quadrant, low frequency temporal components from the replicated spectrum centered at 30 Hz are passed along with the temporal baseband spectrum. Similar aliasing occurs in the other three quadrants of the frequency spectrum. The aliasing in the extracted baseband signal is unavoidable because of the positioning of the replicated spectra; cascading the two one dimensional filters in the reverse order results in a baseband signal which is aliased for the same reasons. The use of a non-separable two dimensional filter allows the extraction of the baseband spectrum without aliasing (Figure 2.7).

Once the signal is restored to being continuous in the vertical and temporal dimensions, it may then be resampled at the desired vertical and temporal frequencies to produce an output sequence in the desired standard.

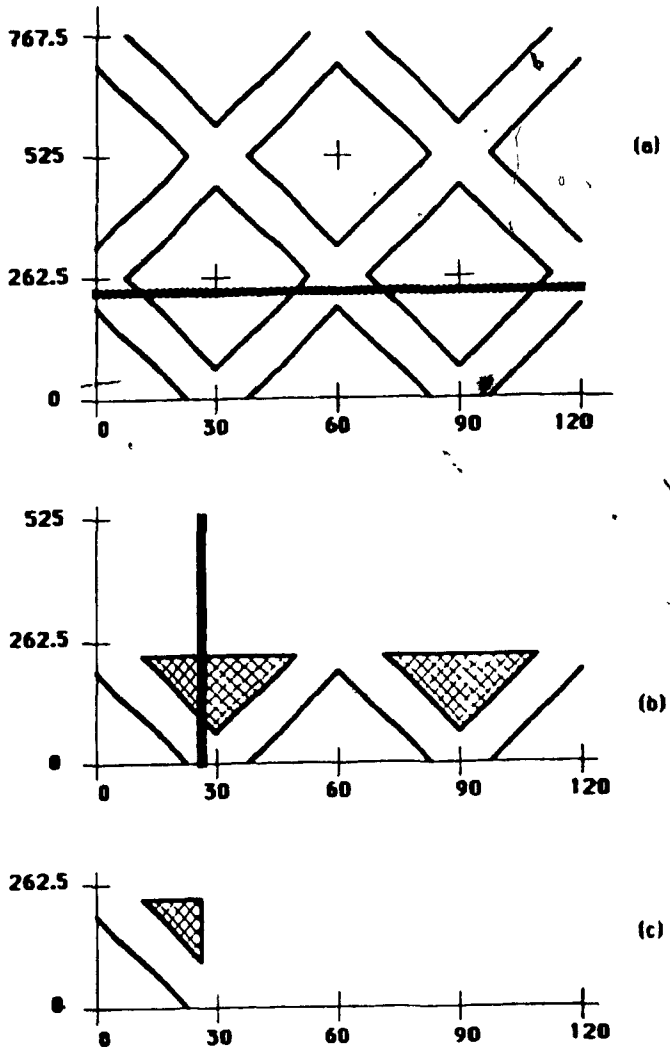

Figure 2.6 Two Dimensional Reconstruction Implemented with Vertical then Temporal Filtering The use of two filters in a separable process leads to aliasing due to the position of the replicated spectra in an interlaced system (a) In the first quadrant, initial vertical filtering passes low vertical frequency components of replicated spectra along with the vertical baseband spectra (b) Subsequent temporal filtering passes low temporal frequency components of replicated spectra along with the baseband spectra (c) The net result is an aliasing of vertical and temporal freqencies in all quadrants of the reconstructed frequency spectrum.

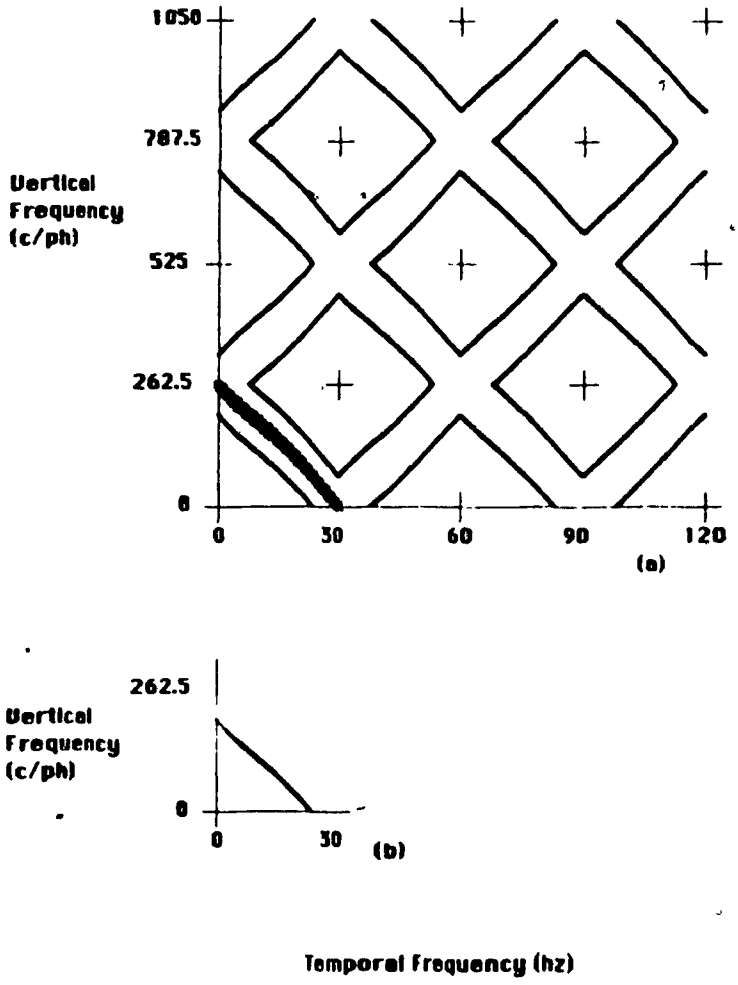

Figure 2.7 Non Separable Two Dimensional Reconstruction Filter. A filter with the diagonal shape as shown extracts the baseband signal without including frequency components from any of the replicated spectra

2.3.4 Two Dimensional Digital Sampling Rate Conversion

As in the one dimensional case, it is neither efficient nor necessary to actually reconstruct the continuous waveform in order to change the sampling structure. The sampling rate conversions in both the vertical and temporal dimensions may be simultaneously performed using a two dimensional aperture. For an input sequence $x(m,n)$, the output sequence $y(r,s)$ is given by

$$y(r,s) = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} x(p,q)\, a((rT_3 - pT_1),(sT_4 - qT_2)), \qquad (2.6)$$

where $a(t,v)$ is the continuous two dimensional aperture function, T_1 and T_2 are the existing sampling periods in each dimension, and T_3 and T_4 are the new sampling periods. As in the one dimensional case, the desired frequency response of the two dimensional aperture function is (in the absence of an aliased input signal) the frequency response of the perfect two dimensional reconstruction filter. In practice, the perfect reconstruction filter is not the ideal two dimensional low pass filter, because of the need to suppress aliased frequency components resulting from a insufficiently bandlimited input to the sampling process, at a cost of a loss of some baseband information.

The problem of sampling structure conversion again resolves into choosing a truncated, quantized aperture function with frequency response which best approximates the perfect reconstruction filter, transforming (for example) the spectrum of Figure 2.8(a) to that of Figure 2.8(b) without lowering the subjective quality of the output sequence. The tradeoffs involved include on one hand, specification of the aperture so that the baseband spectrum is passed without attenuation while rejecting aliased frequency components from replicated spectra; and on the other hand, trading off the length of the aperture for simplicity of implementation.

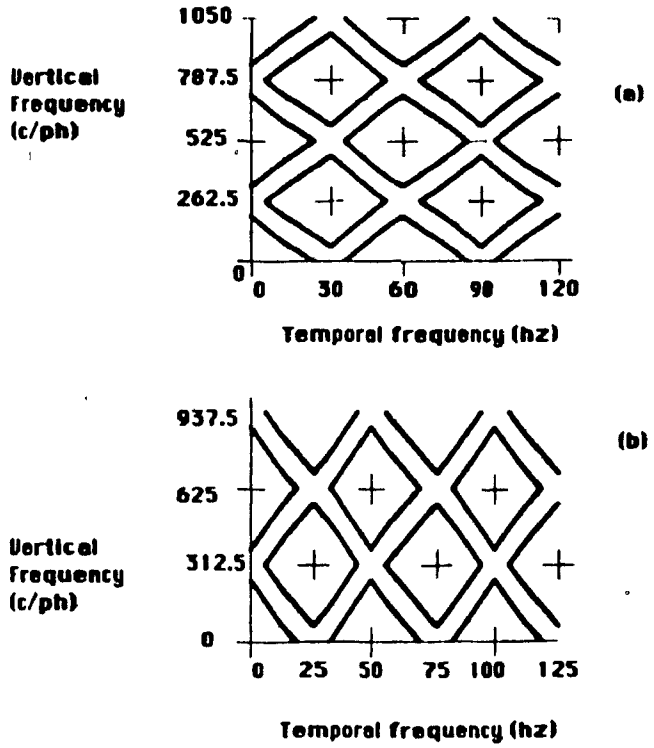

Figure 2.8 Two Frequency Domain Lattice Structures. The frequency domain representation of an NTSC signal, with replicated spectra at multiples of 30 Hz and 525/2 c/ph is shown in (a). Below is the frequency domain representation of a PAL signal, with replications at 25 Hz and 625/2 c/ph. (First quadrant)

2.4 Television Standards Conversion

The sampling structure conversion theory presented in the previous section assumes that the signal was sampled at at least twice the frequency of the highest frequency component in the continuous signal. Otherwise, the sampled signal is aliased, and the continuous signal can never be recovered. Since the input signal to a television camera is not sufficiently vertically and temporally prefiltered to limit its bandwidth, the signal is aliased in both the vertical and temporal dimensions. Low speed movement of high-detail images can result in the existence of high frequency temporal components in a signal, and significant aliasing occurs when the

signal is sampled at 50 or 60 Hz. For this reason, traditional methods of sampling structure conversion using linear processing can never achieve perfect results. Non-linear techniques, such as motion estimation, may be used to produce better results than linear techniques, notwithstanding the temporal aliasing in the sampled signal.

2.4.1 Choices for Interpolation in the Temporal Dimension

The goal of temporal interpolation is to interpolate fields at the temporal locations dictated by the output field rate, based upon information contained in temporally neighboring input fields. Output fields may be interpolated from one, two, or more input fields in the temporal neighborhood of the desired position. The interpolation may be performed using linear processing and an aperture function as an approximation to the perfect reconstruction filter, or the interpolation may be performed using non-linear techniques in an attempt to avoid the problems associated with temporal aliasing in the sampled signal.

2.4.1.1 Linear Approximations to the Perfect Reconstruction Filter

The least complex choice for a temporal aperture is a function which uses a single neighboring field to generate the output field. This aperture has the time-domain form shown in Figure 2.9(a), and the frequency response shown in 2.9(b). The effect of this "zero-order" function is to repeat input fields when the sampling rate is being increased, or to drop input fields when the sampling rate is being decreased.

Dropping or repeating entire input fields is the least complex method of changing the field rate, but seriously degrades the quality of the output sequence. In order to increase the field rate, for example from 50 Hz to 60 Hz, it is necessary to repeat every fifth field in the input sequence. To decrease the field rate from 60 Hz to 50 Hz, it is necessary to drop every sixth field in the input sequence.

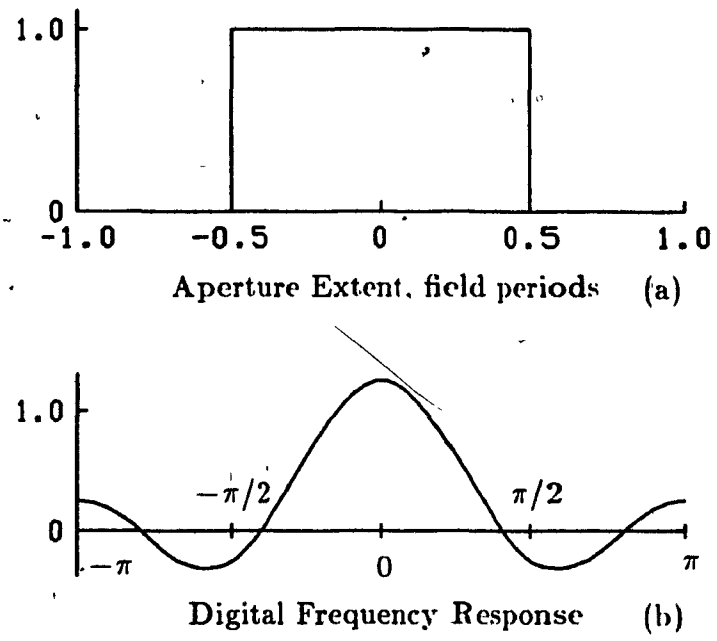

Figure 2.9 Interpolation Aperture Spanning One Input Field
(a) Time domain (b) Frequency domain

When viewing sequences with uniform motion, the repetition or dropping of fields causes a discontinuity in the motion. This discontinuity is referred to as *judder*, and is seen as a noticeable jerk of the image every time a field is repeated or dropped. The frequency of the discontinuity is the beat frequency between the input field rate and the output field rate. For 60 Hz to 50 Hz conversions, or the reverse, the beat frequency is 10 Hz. The human visual system has peak in temporal response at approximately 7 Hz [9], so the 10 Hz motion discontinuity is found to be highly objectionable by human viewers. Judder is the primary motion *artifact* introduced into an output sequence by standards conversion, and is a result of using an imperfect aperture function such as that of Figure 2.9 which attenuates the baseband and passes frequency components from replicated spectra.

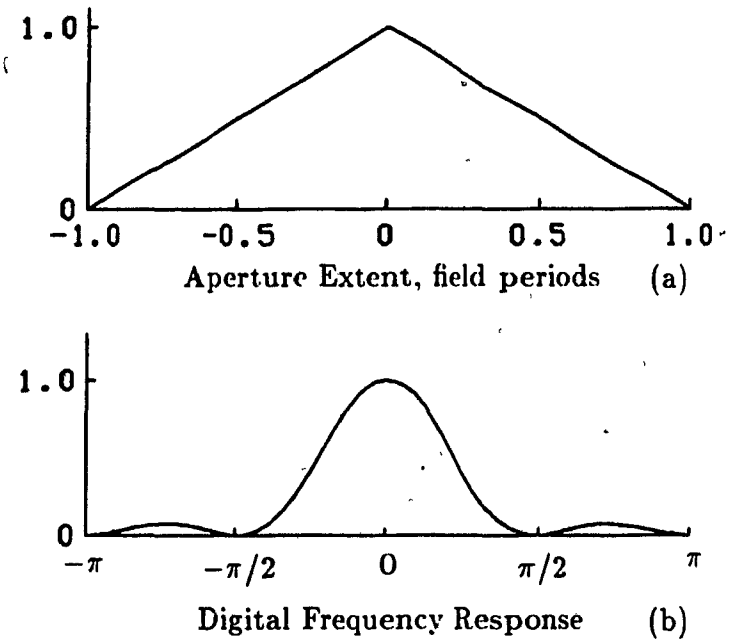

Figure 2.10 Interpolation Aperture Spanning Two Input Fields
(a) Time domain (b) Frequency domain

An aperture which uses two neighboring fields to interpolate each output field usually implements linear interpolation. "Linear interpolation" is not to be confused with "linear processing"; linear interpolation is a type of linear processing. Linear interpolation is the term used to describe an interpolation scheme which takes a weighted sum of the two closest fields to produce the output field at each desired temporal position. The weighting given to each input field is inversely proportional to the temporal distance from that field to the desired temporal position. Using a temporal dimension unit of one input field period, the output field F_o would be calculated as

$$F_o = (1.0 - \alpha)\ F_{i_p} + \alpha\ F_{i_f}, \qquad (2.7)$$

where F_{i_p} is the input field temporally preceding the output field position by α units, and F_{i_f} is the input field following, spaced $(1 - \alpha)$ from the output field position.

Linear temporal interpolation alleviates the judder problem to some extent by forming output fields which contain an average of the images in neighboring fields. Moving objects are *spatially* displaced in sequential fields of a sequence. When two such sequential fields are averaged together to form an output field, the effect is to *smear* the moving object. When viewed in motion, the smearing of moving objects results in a perception of loss of resolution, but an averaging of the motion discontinuities.

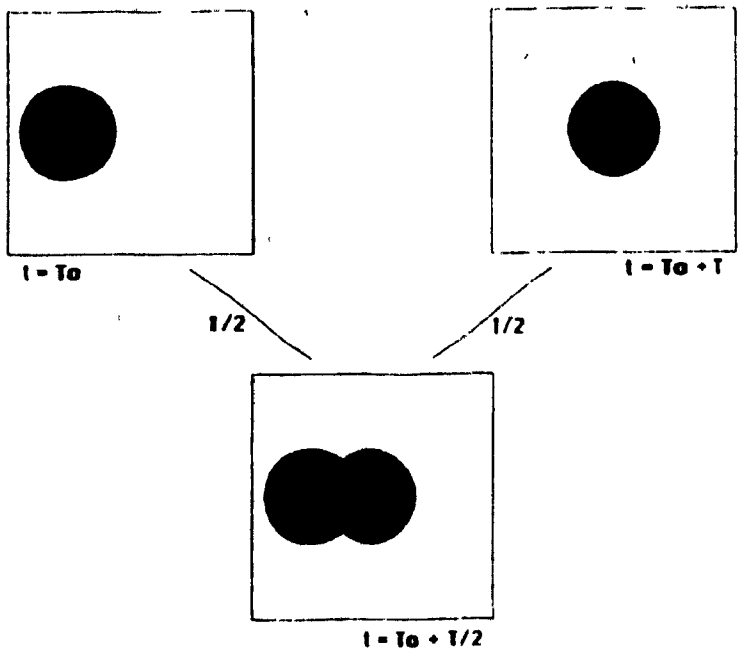

Figure 2.11 Averaging Two Input Fields To Form An Output Field. When interpolating a field (in this case half way between the input fields) by taking an average of two successive fields, the image of a moving object in the interpolated field becomes blurred.

By extending the time-domain extent of the aperture, more than two neighboring input fields may be used to interpolate each output field. The range of choices for the form of the aperture function is infinite. The function may be linear,

quadratic, polynomial, or defined empirically. Since (in the absence of aliasing) the ideal interpolation aperture has the frequency response of the ideal low pass filter, a logical choice for the aperture function is the truncated impulse response of a low pass filter, whose frequency response better approximates the ideal than the zero-order function, linear interpolation, or polynomial approximations.

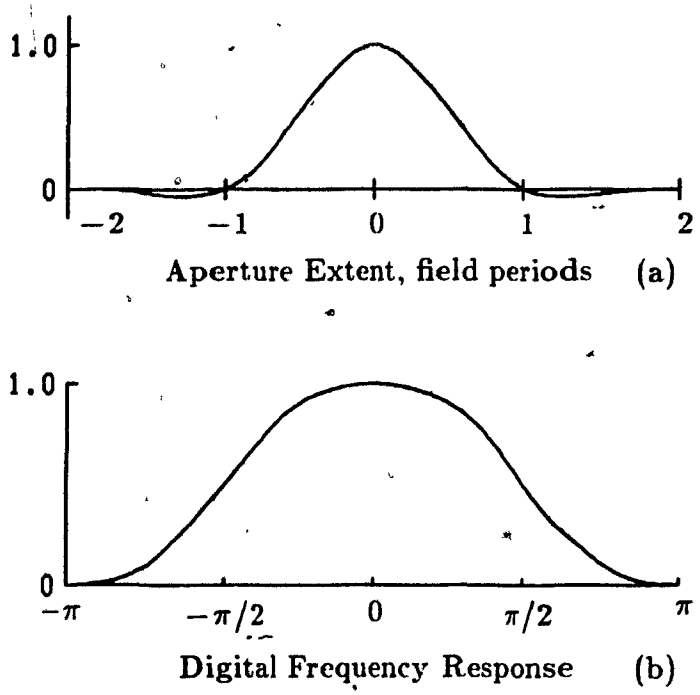

Figure 2.12 Interpolation Aperture Spanning Several Input Fields. (a) Time domain (b) Frequency domain. Note that this aperture is specified in the frequency domain in an attempt to better approximate the frequency response of the ideal low pass filter. The apertures of figures 2.9 and 2.10 were specified in the time domain.

The use of an aperture spanning more than two fields allows the specification of a filter having better passband performance and allowing less aliasing than lower-order methods previously described. Better approximation of the ideal leads to a perception of less judder; however, the judder is lessened at a cost of lower resolution because the output is a weighted average of a number of input fields.

2.4.1.2 Non-linear Temporal Interpolation

A highly complex, non-linear method of interpolating fields at the desired temporal positions is motion compensated interpolation of fields. The fundamental aim of motion compensated interpolation is to estimate the velocity of moving objects in a sequence, and to use the motion estimate to place moving objects in correct spatial positions at the desired temporal position. By placing objects in correct spatial positions, the 10 Hz motion artifacts introduced by the linear processing techniques of section 2.4.1.1 are eliminated entirely.

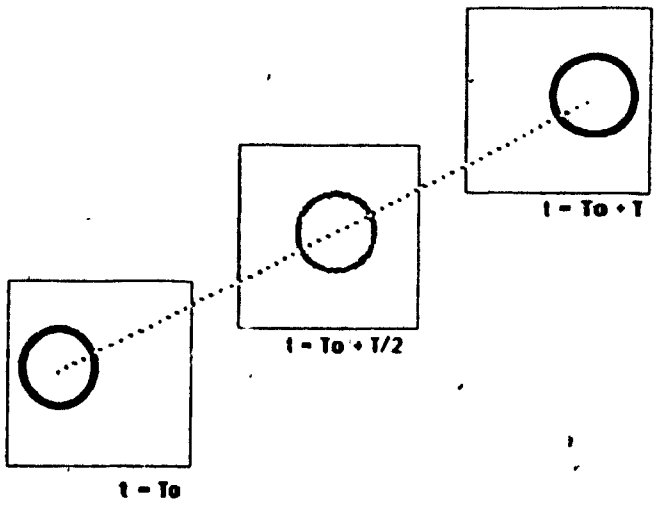

Figure 2.13 Projecting Moving Objects to the Correct Position The essence of motion compensated interpolation is to replace temporal filtering with a system that projects moving images to the correct spatial position in an interpolated field

An important factor in the design and implementation of a motion compensated interpolation system for television signals is the 2:1 field interlace scheme. The effect of an interlaced scanning system is that each field contains one half of the available vertical samples of the image in the camera: a frame formed by interlacing two

successive fields contains a maximal set of vertical samples of a stationary image. A field taken by itself is vertically aliased, so a field interpolated using only a single input field must suffer a loss of vertical resolution to suppress the vertical aliasing. To maximize the resolution of an interpolated field, information is taken from two sequential fields in the temporal neighborhood of the desired position.

In the case of a sequence of fields which contain the image of a stationary scene, two successive fields are combined into a single frame which contains vertical samples at twice the field rate. An interpolated field with a different line density may then be constructed using intra-frame interpolation with no loss of information. A motion compensated system allows the same principle to be used in the presence of motion in the scene by conceptually constructing a frame which is composed of two motion compensated fields. When a scene contains a moving object, sequential fields contain *spatially* displaced samples of the image of the moving object. If two such fields were to be combined to form a frame without any modification, the envelopes of the samples of the moving object from each field would not coincide, and the blurring effect illustrated in Figure 2.11 would occur. If, however, it was possible to adjust or *compensate* both fields so that the moving object occupies the same spatial position in each of the fields, then a frame formed by interlacing the two compensated fields would afford the same resolution as a frame constructed from two fields containing a stationary image.[†]

This *motion compensation* may be implemented by simply taking information from each input field with a displacement equal to a motion estimate from the interpolated field to the respective input field. In a scene with several moving objects, the interpolated field is segmented into different regions, and a separate motion estimate is found for each contiguous region. In the limit, a separate motion estimate may be generated at each pel in the interpolated field.

[†] If a sequence contains vertical motion at a rate equal to an odd number of frame lines per field period, then most of the samples of the moving object(s) in sequential fields will in fact be *identical* samples. In this case, the available resolution of the moving object(s) is reduced.

The compensation process may also be thought of as collapsing two fields onto the desired temporal position *in the direction of the motion estimates* in order to form a motion compensated frame to be used for intra-frame interpolation of a field with a new line density. In this interpretation, the only difference between forming a motion compensated frame when there are horizontally moving objects in the image, and forming a frame from a stationary image is that the motion estimate in the latter case is identically zero over the entire field.

The use of motion compensated temporal interpolation instead of traditional linear processing techniques can produce interpolated sequences free of motion artifacts notwithstanding the existence of temporal aliasing in the input signal. Motion compensated interpolation, forming the basis of the system proposed in this thesis, is treated more fully in Chapter 3.

2.4.2 Choices for Interpolation in the Vertical Dimension

The major problem in television standards conversion is the introduction of motion artifacts due to temporal interpolation. Spatial interpolation introduces artifacts which are subjectively less disturbing, mainly loss of vertical resolution in the interpolated fields. Consequently, highly complex non-linear processing techniques are not required to produce subjectively good quality vertically interpolated output. Vertical interpolation is customarily implemented with linear processing, using a finite length, quantized aperture. In the same way that the extent of the interpolation aperture may be such that one, two, or several input fields are used to interpolate each output field, the extent of a vertical aperture may be such that one, two or several input lines are used to interpolate each output line using linear processing techniques.

A vertical aperture with an extent of one field line is the least complex type of vertical interpolation aperture. The form of this function is identical to that of Figure 2.9. The effect of this "zero-order" aperture is to choose the nearest line

to the desired vertical position as the output line, in effect, dropping lines from the input when the line density is being decreased, and repeating lines when the line density is being increased. Repeating or dropping lines from an input field to form an output field is the least complex method of changing the line density, but introduces *spatial* artifacts to the interpolated field. Diagonal lines will have a discontinuity added, and fine horizontal structures may be eliminated entirely. These spatial artifacts, which are due to the poor frequency response of the aperture of Figure 2.9, may be reduced by finding a better approximation of the ideal low pass filter.

An interpolation aperture which uses two neighboring input lines to interpolate each output line usually implements linear interpolation, in the same manner as temporal linear interpolation described in section 2.4.1.1. Linear interpolation, being a better approximation to the ideal low pass filter, produces better results than the zero-order system described above. The spatial discontinuities are replaced with a less severe aliasing of the old sampling structure onto the new sampling structure. This aliasing produces a sequence of lines which vary in resolution. The lines where the sampling structures coincide are perceived and being *sharp*, because they are copied directly from the input, while the remainder of the lines are perceived as being *blurred* because they are an average of two input lines. Spatial artifacts occur at the beat frequency between the frequency corresponding to the input line density and that corresponding to the output line density.

A vertical interpolation aperture may have an extent which covers more than two input lines, so that each output line is a weighted linear combination of a number of input lines. As in field rate conversion, the range of choices for the form of the aperture function is infinite. The truncated impulse response of a low pass filter specified in the frequency domain, whose frequency response better approximates the frequency response of the ideal low pass filter is a logical choice for the interpolation aperture.

2.5 Frequency Sampling Specification of Interpolation Apertures

Interpolation apertures used in standards converters are often specified by their frequency response in an attempt to obtain a finite length time domain function with a frequency response approaching that of the perfect reconstruction filter. To generate the aperture, the frequency response of a digital low pass filter is specified at a set of discrete points. The impulse response corresponding to the frequency specifications of the filter, when truncated and quantized, forms the interpolation aperture. The impulse response of the filter is obtained by transforming the frequency specifications using the Inverse Discrete Fourier Transform [7]. This method is useful since it provides samples of the impulse response at a set of discrete points, generating the discrete aperture function $a(s)$ of section 2.3.2 directly. First, the one dimensional case will be considered and then the two dimensional case.

2.5.1 One Dimensional Filter Specification

The one dimensional Discrete Fourier Transform (DFT) of the periodic sequence $\tilde{x}(n)$ of length N is defined [7] as one period of the periodic sequence $\widetilde{X}(k)$:

$$X(k) = \widetilde{X}(k)\ R_N(k) \tag{2.8a}$$

$$= [\sum_{n=0}^{N-1} \tilde{x}(n)\ W_N^{kn}]\ R_N(k) \tag{2.8b}$$

where R_N is a rectangular function which extracts one period of length N, and

$$W_N = e^{-j\frac{2\pi}{N}}. \tag{2.9}$$

The Inverse Discrete Fourier Transform of $\widetilde{X}(k)$ is defined as

$$x(n) = \tilde{x}(n)\ R_N(n) \tag{2.10a}$$

$$= [\sum_{k=0}^{N-1} \widetilde{X}(k)\ W_N^{-kn}]\ R_N(n) \tag{2.10b}$$

The frequency response of the filter is considered to be specified at N discrete points equally spaced along the frequency axis at intervals of $\frac{2\pi}{N}$, spanning the frequency range from 0 (dc) to 2π (the sampling frequency). This sequence is designated as $X(k)$, and $\widetilde{X}(k)$ is defined as $X(k)$ repeated with period N. $\widetilde{X}(k)$ is the Discrete Fourier Series (DFS) associated with the frequency response of the filter.

A purely real DFS has the property that its transform is conjugate symmetric; the real part of the transform is an even function, and the imaginary part odd. For the impulse response of the filter to be real, the frequency response of the filter is specified such that $\mathrm{Re}[\widetilde{X}(k)] = \mathrm{Re}[\widetilde{X}(-k)]$, and $Im[\widetilde{X}(k)] = 0$ for all k, *i.e.* $\widetilde{X}(k)$ is specified to be symmetrical about 0 and purely real. The latter condition also implies that the filter has zero phase response. Note that because $\widetilde{X}(k)$ is periodic with period N, $X(k)$ is symmetrical about 0 and about π.

Evaluating 2.10 for all n generates the DFS $\tilde{x}(n)$, which is periodic in N. Any one period of $\tilde{x}(n)$ may be chosen as the IDFT of $X(k)$. If

$$x(n) = \left\{ \tilde{x}(n),\ n = -\frac{N-1}{2},\ \ldots,\ -1,\ 0,\ 1, \ldots,\ \frac{N-1}{2} \right\} \qquad (2.11)$$

is chosen, $x(n)$ is a familiar representation of the impulse response of the filter specified by $X(k)$. The impulse response obtained using 2.10 may be thought of as samples of a continuous time domain function. By calculating the impulse response using 2.10, N equally spaced samples of the entire length of the impulse response are obtained. Samples of the impulse response are spaced at the inverse of the value of the frequency at which the signal is sampled. For example, a signal which is temporally sampled at 60 Hz will have samples of the impulse response spaced at 1/60 second using the above method. Figure 2.14 illustrates a filter specified with eight frequency samples. The resulting aperture is then specified at eight points spanning four field periods.

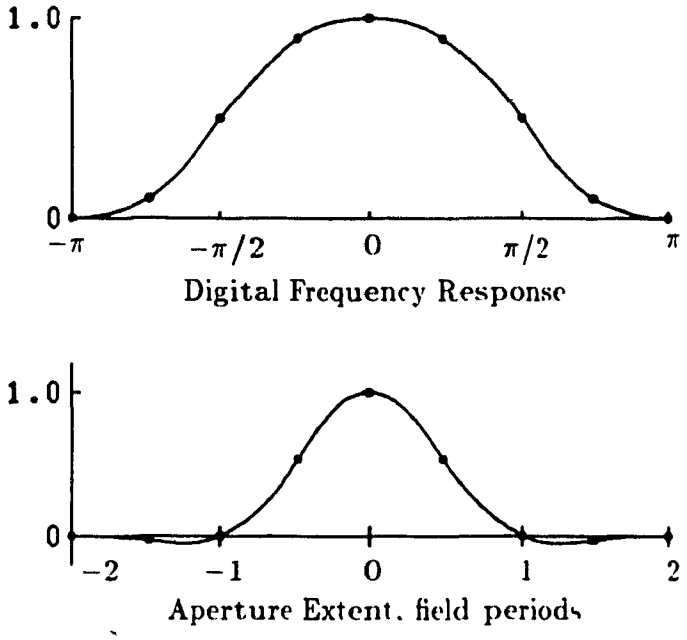

Figure 2.14 A Frequency Specified Interpolation Aperture. Note that both the frequency specification and impulse response are defined only at the discrete points shown

Increasing N by specifying more closely spaced points in the *frequency* domain increases the length of the impulse response obtained with the IDFT. If more closely spaced samples of the *time* domain function are required to avoid positional quantization noise, it is necessary to increase the length of the frequency domain function $\tilde{X}(k)$ by padding it with zeroes:

Define, for some whole number P,

$$\tilde{Y}(k) = \begin{cases} \tilde{X}(k) & \text{for } k = 0, 1, \ldots, N-1 \\ 0 & \text{for } k = N, N+1, \ldots, PN-1 \end{cases} \tag{2.12}$$

Note that $\tilde{Y}(k)$ is periodic with length $N_y = (P+1)N$.

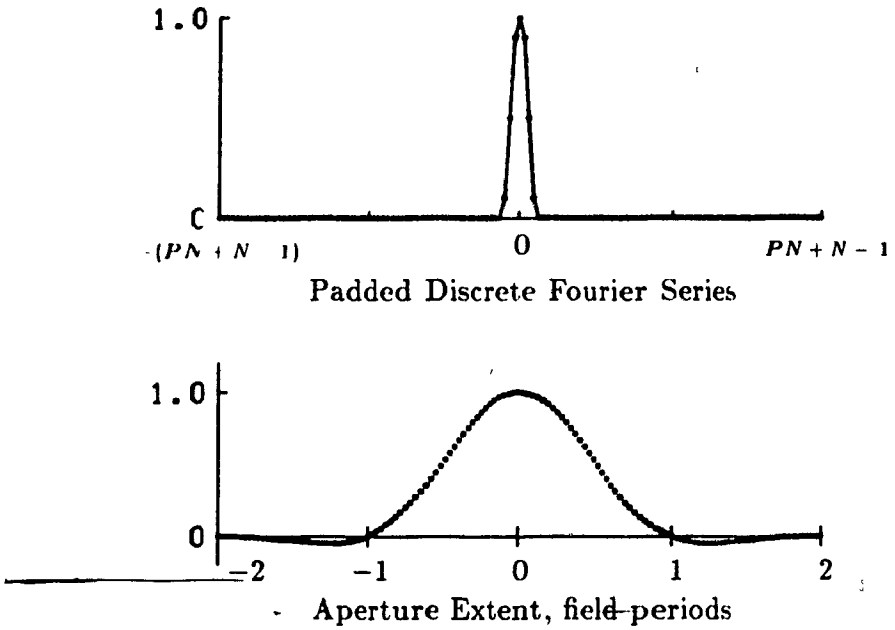

Figure 2.15 A Padded Frequency Specified Interpolation Aperture. Note that both the frequency specification and impulse response are defined only at the discrete points shown, but the density of "samples" of the impulse response obtained is much higher than in figure 2.14

From 2.10 we have

$$\tilde{y}(n) = \frac{1}{N_y} \sum_{k=0}^{N_y - 1} \tilde{Y}(k)\, W_{N_y}^{-kn} \tag{2.13a}$$

$$= \frac{1}{(P+1)N} \sum_{k=0}^{N-1} \tilde{X}(k)\, W_{(P+1)N}^{-kn} \tag{2.13b}$$

$$= \frac{1}{P+1} \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}(k)\, W_N^{-k\left(\frac{n}{P+1}\right)} \tag{2.13c}$$

$$\tilde{y}(n) = \frac{1}{P+1}\, \tilde{x}\left(\frac{n}{P+1}\right). \tag{2.14}$$

Thus, $\tilde{y}(n)$ is periodic with period $(P+1)N$, and consists of $\tilde{x}(n)$ plus samples interpolated between those of $\tilde{x}(n)$ with a density factor of $P+1$.

By selecting appropriate values for P and N, an aperture function of the correct length, and sampled at a suitable density may be generated. Figure 2.15 illustrates how the frequency samples of figure 2.14 is padded with zeroes with a factor P = 15 to form a longer DFS, and generate a more densely sampled impulse response.

2.5.2 Two Dimensional Filter Specification

An interpolation aperture which can be used for combined vertical-temporal sampling structure conversion is often specified by its frequency response in an attempt to find a finite length time domain function whose frequency response is a good approximation to the ideal. The frequency specification is on a two dimensional grid of points which ranges in both dimensions from 0 to 2π. If M is defined as the length of the sequence in the vertical dimension, and N as the length of the sequence in the temporal dimension, the two dimensional DFT is defined as

$$X(k,l) = \widetilde{X}(k,l)\ R_{MN}(k,l) \tag{2.15a}$$

$$= [\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(m,n)\ W_M^{km}\ W_N^{kn}]\ R_{MN}(k,l) \tag{2.15b}$$

where R_{MN} is a two dimensional rectangular function which extracts one period of $\widetilde{X}(k,l)$, and

$$W_M = e^{-j\frac{2\pi}{M}}, \tag{2.16}$$

$$W_N = e^{-j\frac{2\pi}{N}}. \tag{2.17}$$

The two dimensional Inverse Discrete Fourier Transform is defined as

$$x(m,n) = \tilde{x}(m,n)\ R_{MN}(m,n) \tag{2.18a}$$

$$= [\frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} X(k,l)\ W_M^{-km}\ W_N^{-kn}]\ R_{MN}(m,n) \tag{2.18b}$$

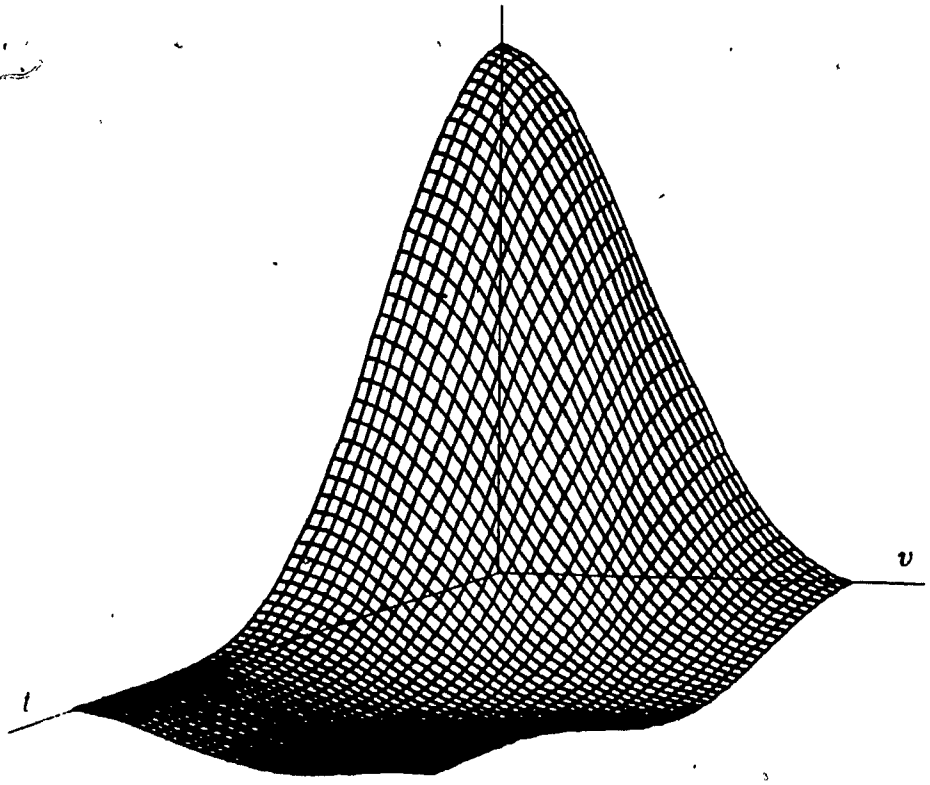

Figure 2.16 A Two Dimensional Aperture One quadrant of a two dimensional aperture generated with the frequency sampling method is shown. This quadrant spans two field lines in the vertical dimension and two field periods in the temporal dimension.

The sequence $X(k,l)$ is specified over an M by N grid of points with the same symmetry constraints as the one dimensional case in both dimensions. $\tilde{X}(k,l)$ is formed by repeating $X(k,l)$ in the vertical dimension with period M and in the temporal dimension with period N. Equation 2.18 is evaluated for all m and n to obtain $\tilde{x}(m,n)$. The sequence $x(m,n)$ is defined as one period (in each dimension) of $\tilde{x}(m,n)$. If we choose

$$x(m,n) = \tilde{x}(m,n), \quad m = -\frac{M-1}{2}, \ldots, -1, 0, 1, \ldots, \frac{M-1}{2}$$

$$n = -\frac{N-1}{2}, \ldots, -1, 0, 1, \ldots, \frac{N-1}{2}$$

a familiar representation of the two dimensional aperture results.

As in the one dimensional case, the spacing of samples of the impulse response is in each dimension a function of the sampling frequency. If more closely spaced samples of the impulse response are required, the length of the frequency domain sequence $\widetilde{X}(k,l)$ must be increased by padding with zeroes in a manner analagous to the one dimensional case:

Define for some whole numbers P and Q.

$$Y(k,l) = \begin{cases} X(k,l) & \text{for } k = 0,1,\ldots,M-1 \\ & \text{and } l = 0,1,\ldots,N-1 \\ 0 & \text{for } k = M, M+1,\ldots,PM-1 \\ & \text{and } l = N, N+1,\ldots,QN-1 \end{cases} \tag{2.19}$$

Note that $\tilde{Y}(k,l)$ is periodic with lengths $M_y = (P+1)M$ and $N_y = (Q+1)N$. From 2.18 we have

$$\tilde{y}(m,n) = \frac{1}{M_y N_y} \sum_{k=0}^{M_y-1} \sum_{l=0}^{N_y-1} \tilde{Y}(k,l)\, W_{M_y}^{km}\, W_{N_y}^{ln} \tag{2.20a}$$

$$= \frac{1}{(P+1)M\ (Q+1)N} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \widetilde{X}(k,l)\, W_{(P+1)M}^{-km}\, W_{(Q+1)N}^{kl} \tag{2.20b}$$

$$= \frac{1}{(P+1)(Q+1)}\ \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \widetilde{X}(k,l)\, W_M^{k(\frac{m}{P+1})}\, W_N^{l(\frac{n}{Q+1})} \tag{2.20c}$$

$$\tilde{y}(m,n) = \frac{1}{(P+1)(Q+1)}\ \tilde{x}(\frac{m}{P+1}, \frac{n}{Q+1}). \tag{2.21}$$

The sequence $\tilde{y}(m,n)$ is periodic with period $(P+1)M$ in the vertical dimension, and period $(Q+1)N$ in the temporal dimension, and consists of $\tilde{x}(m,n)$ plus samples interpolated between those of $\tilde{x}(m,n)$ with a density factor of $P+1$ in the vertical dimension and $Q+1$ in the temporal dimension. As in the one dimensional case, an aperture function of the correct length in each dimension, and sampled at a suitable density in each dimension may be generated by selecting appropriate values for P, Q, M and N. Figure 2.16 illustrates one quadrant of a vertical-temporal aperture generated using this method.

2.6 Motion Estimation

Estimating the velocity of moving objects in a television sequence is a current area of research. There are several methods available for estimating motion, all of which attempt to find, for a pel or a block of pels in a particular field, areas in neighboring fields which *correspond* to that pel or block of pels. Thus, motion estimation is often referred to as the *correspondence problem*, and a field of motion vectors, one vector for each pel in the field where motion is being estimated, is often called the *correspondence field*. For the greater part, algorithms have been developed for video *coders* rather than video standards converters; and produce motion estimates that are suitable for differential motion compensated coders, but not necessarily suitable for standards conversion. In a coder, a small error in the motion estimation, especially in a flat or uniform portion of the image, results only in a small increase in the error signal. In a standards converter, a small error in the motion estimation places a moving object, or more frequently, a *portion* of a moving object in the wrong place in an interpolated field. Placement of a portion of an object apart from the rest of the object in a field is subjectively displeasing, since the human visual system is very sensitive to spatial discontinuities.

Two primary methods exist for solving the correspondence problem: spatio-temporal gradient techniques, and block-matching techniques. Gradient techniques generate a motion estimate for each pel using an algorithm that follows the gradient of intensity differences between fields. Block-matching techniques generate a motion estimate for a region by minimizing a distortion measure between a region in one field and displaced regions in the following field. Gradient algorithms, while useful for coding because of their relative simplicity of calculation, may not be useful for standards conversion because of a long convergence time and lack of robustness in the presence of noise and sharp gradients. Block matching algorithms that average their estimate over a large area, and lend themselves to matching objects instead of arbitrary regions may be more useful.

The object of this thesis is to study the interpolation aspects of motion compensated standards conversion; no conclusions as to the most suitable method of motion estimation for standards conversion will be made. Nonetheless, some method of motion estimation must be used to perform motion compensated interpolation, so a brief survey of motion estimation algorithms is presented.

2.6.1 Gradient Techniques

Gradient techniques for motion estimation use the spatio-temporal gradient constraint equation [10] to estimate the motion between frames based on the gradient of the intensity of the image. The gradient constraint equation is given as

$$\mathbf{v} \cdot \nabla_{\mathbf{x}} u + \frac{\partial u}{\partial t} = 0, \tag{2.22}$$

where $\mathbf{v}$ is the two dimensional spatial velocity estimate between the two frames, $\nabla_{\mathbf{x}} u$ is the spatial gradient of the intensity, and $\frac{\partial u}{\partial t}$ is the temporal derivative of the intensity at the point $\mathbf{x}$.

The first algorithm, and the basis for subsequent improvements, was proposed by Limb and Murphy [11], and Cafforio and Rocca [12]. In this algorithm, the frame is divided into a number of rectangular blocks, and 2.22 is evaluated over all of the points $\mathbf{x}$ within a block by approximating the spatial gradient of the intensity and the temporal derivative with finite differences which may be measured from the image. Linear regression is used to find a minimum mean square error estimate for $\mathbf{v}$, which is then assigned to the entire block.

An improvement in the form of a temporal recursion was added to improve the convergence rate, by replacing the frame difference with a displaced frame difference. In general, recursive gradient techniques attempt to form an estimate of the motion at each pel by updating the estimate at a spatially or temporally neighboring pel.

An algorithm by Netravali and Robbins [13] based on the previous algorithms gives an individual estimate at each pel. This algorithm uses a spatial recursion, where the estimate from a spatially neighboring pel is updated using a steepest descent algorithm on the intensity to form the estimate for the given pel. The magnitude of the correction is in proportion to the magnitude of the gradient, and a constant ϵ which controls the rate of convergence. Improvements to this algorithm consist of improvements in the factors which control the convergence rate, and the use of gradient information at several spatially neighboring points to improve the robustness of the algorithm.

Another type of improvement on the original gradient algorithms is a method which uses a sliding block of pels to give a block-averaged estimate at each pel [10]. This estimate is obtained by updating the estimate for the same pel in the previous frame by finding an estimate which best satisfies a more general form of 2.22 over a small three dimensional block of pels centered on the pel in question. This algorithm evidently maintains the robustness of the block approach while giving an estimate at each pel.

2.6.2 Block Matching Techniques

A different technique for estimating motion is the block matching technique. This technique essentially minimizes some form of distortion measure between the block in question and displaced blocks in the following field. The fields may be divided into rectangular blocks of equal size, or segmented into contiguous regions using some qualitative criterion. The basis of the block matching technique is the definition of a mean distortion function $D(i,j)$ [14] between an M by N block in the reference field U_R and a displaced block in the following field U.

The distortion function is given as

$$D(i,j) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} g(u_r(m,n) - u(m+i, n+j)), \quad -p \le i,j \le p, \qquad (2.23)$$

where $u(m,n)$ is the intensity at spatial position (m,n), $g(x)$ is a monotonically increasing distortion measure, and the search is limited to a range of $\pm p$ in each spatial dimension. The value of $D(i,j)$ is calculated for each displacement in the specified range, and the coordinates (i,j) corresponding to the minimum of $D(i,j)$ are said to be the *direction of minimum distortion*. This definition of the distortion function leads to an estimate of displacement quantized to integer values. If fractional pel estimates are required, the integer direction of minimum distortion is found, then a search of positions displaced $\pm q$ about the last direction of minimum distortion is performed for $q = 2^{-1}, 2^{-2}, \ldots$, until the desired accuracy is obtained. The intensity values $u(m+i, n+j)$ for fractional (i,j) must be spatially interpolated from surrounding pels.

Refinements of the basic technique consist mainly of varying the search pattern to avoid having to search every (i,j) displacement in the range $\pm p$, and selecting the function $g(x)$. The most complex search pattern is an iterative search of four points and the center of a square of logarithmically decreasing size [14], with a minimum mean square error distortion function, *i.e.* $g(x) = x^2$. Other types of search patterns use specific sets of trial vectors; a selected menu of possible displacements [15, 16]. The vectors in the sets are decreased in magnitude at each step until the desired accuracy is reached. A third type of search pattern is the "conjugate direction" search [17], which finds a minimum independently in each direction, then searches along the direction of the two independent minima. [15] and [17] use a mean absolute error function, *i.e.* $g(x) = |x|$, while [16] uses a log absolute error function. These refinements decrease the number of computations required by a significant factor by eliminating many candidate search positions based on a knowledge of the statistics of the video signal, at a cost of being more prone to divergence than the basic algorithm which searches all displacements.

For standards conversion, arbitrary division of the field into rectangular blocks leads to the introduction of discontinuities in the interpolated field at block boundaries. The human visual system is highly sensitive to such spatial discontinuities. A better approach is to segment the field into regions of constant motion such as the images of moving objects, and search for corresponding regions in the following field. This strategy eliminates the block-boundary discontinuity problem, but places more significance on the segmentation problem. Implementation of a region-matching search simply requires the replacement of the summation over a rectangular area with a summation over a general region, perhaps indicated by a template passed from a segmentation process. Block matching techniques may be more robust than gradient techniques, and do lend themselves to object or region matching, but at a cost of significantly higher computational requirements.

2.7 Existing Standards Converters

The first standards converters used were optical converters, that essentially recorded the image of a monitor displaying the input standard with a camera using the output standard. The interpolation apertures were determined by the impulse response of the phosphor used on the display tube. Analog electronic standards converters became possible with the advent of the quartz delay line, allowing the implementation of a line store. The converters produced using this technology were of the repeat/drop variety, since interpolation requires high speed multipliers and adders as well as a minimum of two field stores. Digital processing allows the implementation of highly complex systems that were not easily implemented with analog quartz delay lines, and produce superior subjective results while having the added advantages of needing no periodic calibration or adjustment due to 'drifting', burning in, or similar analog problems.

Digital standards conversion systems which have been produced include systems that implement line density conversion using frequency-specified apertures and temporal linear interpolation, and systems that use two dimensional vertical-temporal apertures. All existing converters trade off algorithmic complexity for temporal and spatial artifacts to varying degrees.

2.7.1 Separate Vertical-Temporal Interpolation

The most primitive digital standards conversion systems implement field rate conversion using an aperture which spans two input fields (linear temporal interpolation), and line density conversion using an aperture which spans several input lines, in a separable process. An example of this type of converter is DICE, or Digital Intercontinental Conversion Equipment, developed by the Independent Broadcast Authority (IBA) in 1976 [18].

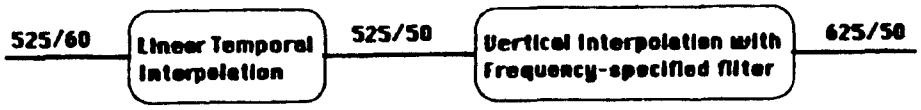

Figure 2.17 Block Diagram of the DICE Converter The DICE Converter developed by IBA implements linear temporal interpolation

The IBA DICE converter converts 525/60 signals to 625/50 signals or vice-versa using two 525 line field stores. In the 525/60 to 625/50 direction, the field rate of the 525/60 input field sequence is first converted to 50 Hz using two-field linear temporal interpolation, yielding a 525-line, 50 Hz sequence. The 525 line fields are then line converted to 625 line fields with intra-field interpolation and a frequency-specified vertical aperture. In the reverse direction, the line density of the 625/50

fields are first line converted to 525 lines per frame so that the same field stores may be used for both directions. The resulting 525 line fields are then used to interpolate 525 line output fields at the 60 Hz output rate using linear temporal interpolation.

The algorithm used in this converter is virtually the minimum possible improvement over the algorithm used in analog quartz delay line converters; but since digital processing in any form offers a large advantage in maintenance, reliabilility and noise immunity over analog equipment, and since real-time implementation of the converter with the technology that was available at the time of the design required a certain simplicity of design, this algorithm was well received and remains the basis of many of the converters available commercially. Subjective viewing of sequences converted with the linear temporal interpolation scheme reveals the introduction of significant motion artifacts at the beat frequency between the two field rates. A motion discontinuity is particularly noticeable at the point in time where a field is repeated or dropped, because of the linear weighting scheme.

2.7.2 Combined Vertical-Temporal Interpolation

The state of the art in terms of installed commercially available standards conversion equipment is the Advanced Conversion Equipment (ACE) converter designed by the BBC research department [2]. It is essentially a two dimensional low pass filter. This converter is a four-field, four line standards converter, for 525/60 to 625/50 conversion and vice-versa. Sixteen lines in the temporal and spatial neighborhood of each output line are taken in a weighted linear combination to form the output line. The weights are defined by a two dimensional aperture, which is the truncated and quantized impulse response of a two dimensional low pass filter specified by its frequency response.

This two dimensional aperture algorithm is the logical improvement on the linear temporal interpolation, vertical aperture scheme. The frequency response of the two dimensional aperture is specified to virtually eliminate the motion artifacts

introduced when using two field linear temporal interpolation. The judder is eliminated for most cases by averaging the motion discontinuities over four fields, at the expense of a noticeable loss of resolution. Under extreme conditions, such as a rapid pan of a high-contrast scene, severe motion artifacts become visible.

2.7.3 Motion Compensated Interpolation

A motion compensated television standards converter for conversion from the 1125 line 60 Hz HDTV system to 50 Hz systems is the experimental NHK converter [4]. This converter uses separate vertical and temporal sampling structure conversion, first converting two fields in the temporal neighborhood of the output field to the output line structure, and then using motion compensated interpolation to interpolate the output field.

The 1125/60 to 625/50 conversion process begins with the interpolation of 625-line non-interlaced frames from the input fields using intra-field interpolation in moving areas and inter-field interpolation in stationary areas. These 625/60 frames are passed to a section which performs linear temporal interpolation, and also to a section which performs motion compensated interpolation. The line-converted frames are divided into four quadrants, and a motion estimate is independently generated for each quadrant. Each motion estimate is then used to interpolate an entire frame. Application of a single motion estimate to the entire frame avoids the introduction of discontinuities at quadrant boundaries but leads to obvious problems when there are several large moving objects in a scene. An edge detector is included to allow the special treatment of object boundaries. The five output frames, one generated by linear interpolation, and four generated with the motion estimates from one of the quadrants applied to the entire field, are compared to find the candidate with the minimum error using some measure of quality.

No information on the type of aperture used for line conversion, the type of motion estimation used in the converter, or the basis for selecting the best 'can-

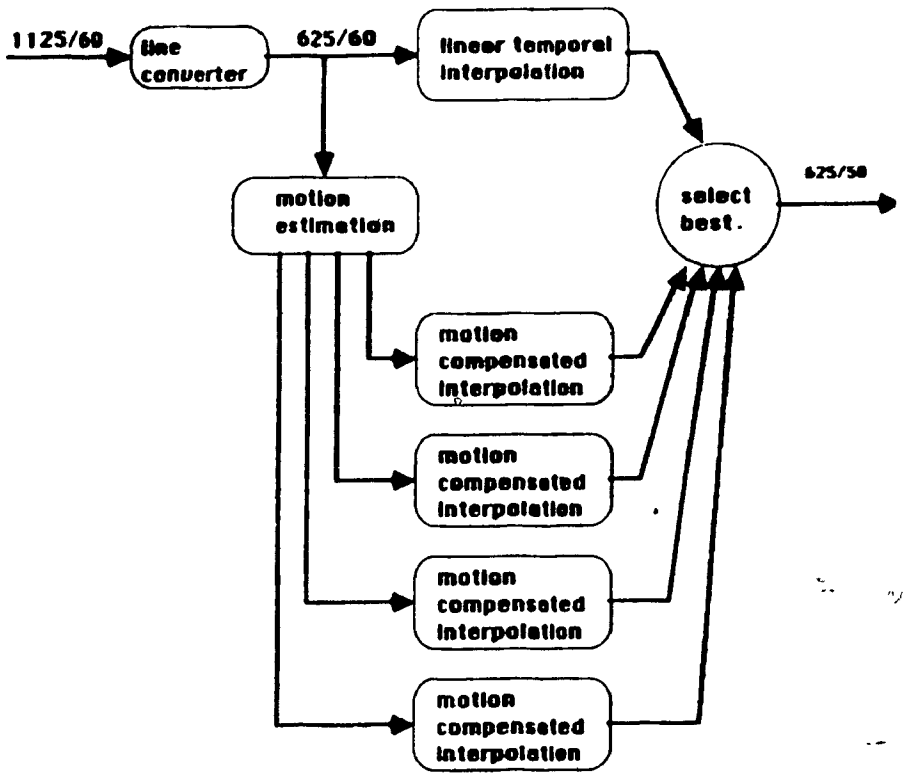

Figure 2.18 Block Diagram of the NHK Converter. The experimental motion compensated converter developed by NHK, the Japanese national broadcasting corporation, uses a single motion estimate applied to the entire field to interpolate a field. Four different motion estimates are tried (one from each quadrant of the image). The "best" interpolated field is selected from the four motion-compensated interpolated fields, and a field interpolated with linear temporal interpolation

didate' interpolated field has been given. The converter was demonstrated at the IWP 11/6 (HDTV) meeting in Tokyo, and is said to produce significantly better results than the most sophisticated fixed aperture converter [4]. It should be noted that this system changes the line density from 1125 lines per frame to 625 lines per frame. Since the input sequence is vertically sampled at about twice the rate of NTSC or PAL sequences, problems with vertical aliasing are not as severe as those occurring in 525/60 to 625/50 and 625/50 to 525/60 converters.

Chapter 3

Motion Compensated Television Standards Conversion

The television standards conversion system proposed in this thesis is a two-field motion-compensated interpolation system. The object of motion compensation is to place moving objects at the correct spatial position in output fields, eliminating the motion artifacts associated with linear processing, while maintaining a high degree of resolution. First, a description of the structure of an ideal motion compensated interpolation scheme is presented. In the second section, the experimental system implemented for this thesis is described.

3.1 The Ideal System

An ideal motion compensated television standards converter consists of three distinct processes: a *segmentation* process to segment the interpolated field into contiguous regions, a *motion estimation* process which uses luminance information from the input field sequence as well as higher-level information from the segmentation process to provide motion estimates at each pel in the interpolated field; and an *interpolation* process which uses information from the input field sequence, the segmentation process and the motion estimation process to interpolate fields at the desired temporal positions.

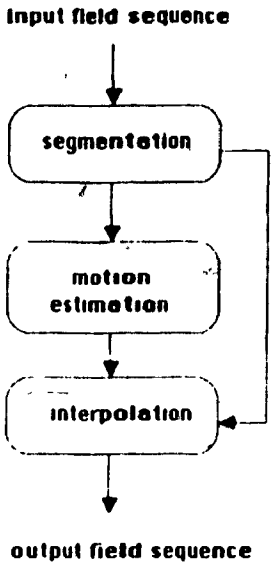

Figure 3.1 The Ideal Motion Compensated Interpolation System. The ideal motion compensated interpolation system consists of a segmentation process, a motion estimation process, and an interpolation process

3.1.1 Segmentation

The purpose of the segmentation process is to classify two dimensional spatial regions in the interpolated field into one of four categories: moving regions, stationary regions, newly covered background, and newly exposed background. In the simplest case, an object moves with uniform translational motion in front of a stationary background. The segmentation process should identify four regions in the interpolated field. One region is the leading edge of the moving object which obscures background that was visible in the previous field. A second region is a region of background trailing the moving object, which was not visible in the previous field. A third region is the remainder of the moving object, and the last region is the remainder of the field, the stationary background (Figure 3.2).

A more complex scene with several moving objects, would be segmented into a number of regions classified as one of the four categories as appropriate. Segmentation provides the motion estimation and interpolation processes with information necessary to interpolate the best possible field while at the same time lowering the computational requirements of the system.

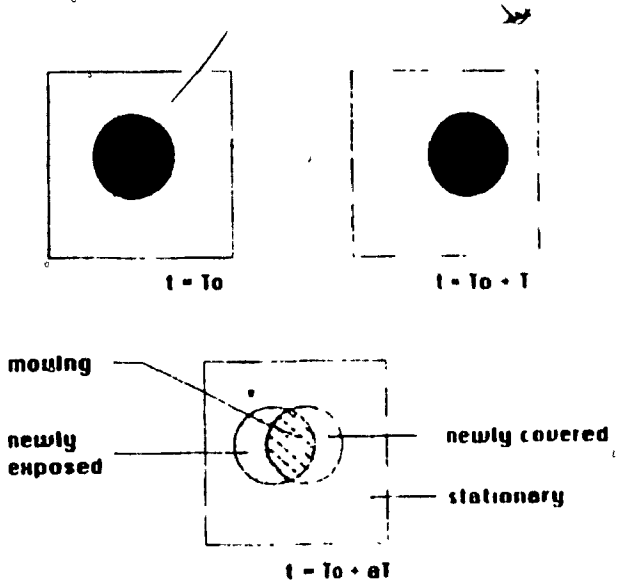

Figure 3.2 Segmentation of a Simple Scene. The segmentation process should identify regions of constant motion, newly exposed background, newly covered background and stationary regions.

The classification of a region as stationary eliminates the need to produce a motion estimate for that region. Identification of regions as newly exposed or newly covered background allows the interpolation process to be signaled that information regarding those particular regions exists only in one of the two input fields temporally neighboring the interpolated field. By making use of this identification, the interpolation process can avoid the use of incorrect information, preserving the integrity of the edges of a moving object.

The identification of contiguous regions of constant motion allows the motion estimation process to generate useful motion estimates for a scene more complex than one with a single moving object. The identification of an entire moving object as being a region allows the use of a motion estimation algorithm which generates an estimate for the whole object, instead of for individual pels or arbitrary rectangular blocks. Generating motion estimates by matching such regions has several advantages. The primary advantage of region matching is the avoidance of possible spatial discontinuities within an object because the region boundaries have been chosen to be the object's boundaries. If motion estimates are calculated for arbitrary rectangular blocks, subjectively displeasing spatial discontinuities may occur at block boundaries in the interpolated field. Another advantage is a decrease in the computational requirements of the motion estimation process. The area covered by all of the arbitrary rectangular blocks which would contain a portion of the object is likely to be much higher than the area covered by the object itself, implying the calculation of a distortion measure over many more pels when using rectangular block searches. As well, a region matching algorithm is likely to be more robust than an algorithm which calculates a motion estimate separately at every pel. A final advantage of region matching is the possibility of an improvement in performance of the search algorithm since searching for a particular shape of object implies the inclusion of information of a higher level than simple luminance values.

3.1.2 Motion Estimation

The second process in the ideal motion compensated television standards converter is the motion estimation process. The purpose of the motion estimation process is to generate motion estimates at each pel in the interpolated field. The required motion estimates are two vectors $\mathbf{v_p}(\mathbf{x})$ and $\mathbf{v_f}(\mathbf{x})$, which are the spatial displacements $\mathbf{v_p}$ and $\mathbf{v_f}$ from the pel at $\mathbf{x}$ in the interpolated field to the corresponding points in the input field preceding the desired temporal position and

in the input field following the desired temporal position respectively; the motion estimation process must generate two correspondence fields for each interpolated field. Since the object of television standards conversion is to generate an output sequence of subjective quality equaling that of the input, the correspondence field generated for motion compensated interpolation must be accurate enough to eliminate motion artifacts, and be free of error to avoid introducing spatial artifacts into the interpolated field.

The correspondence fields for each interpolated field may be generated using one of two approaches. The first approach is to generate the correspondence fields at the desired temporal position by generating motion estimates between the preceding input field and the following input field, and then deriving the correspondence fields at the desired temporal position. This approach is the least complicated approach, but is subject to several key weaknesses. If a correspondence field is generated between input fields, from which the correspondence fields at the desired temporal position are to be derived, it is necessary to interpolate the *motion estimates* themselves. One way of doing this is to project the motion estimates between input fields onto the plane of the desired temporal position, finding their spatial intersection with the plane, and assigning geometrically scaled values to the pel in the interpolated field closest to the intersection as the motion estimate for that pel (Figure 3.3).

Clearly, not all pels in the interpolated correspondence fields will have a value assigned to them; missing values would have to be filled in through intra-field interpolation. A more serious problem is the emphasis of errors. If the correspondence field between two input fields contains an error, then projecting the vector in error onto the plane of the desired temporal position places an entry in the interpolated correspondence fields which is not only incorrect, but also in the wrong spatial position. To correct errors, median filtering or averaging of the interpolated correspondence fields must be performed before they are used to interpolate the output to avoid the introduction of spatial discontinuities.

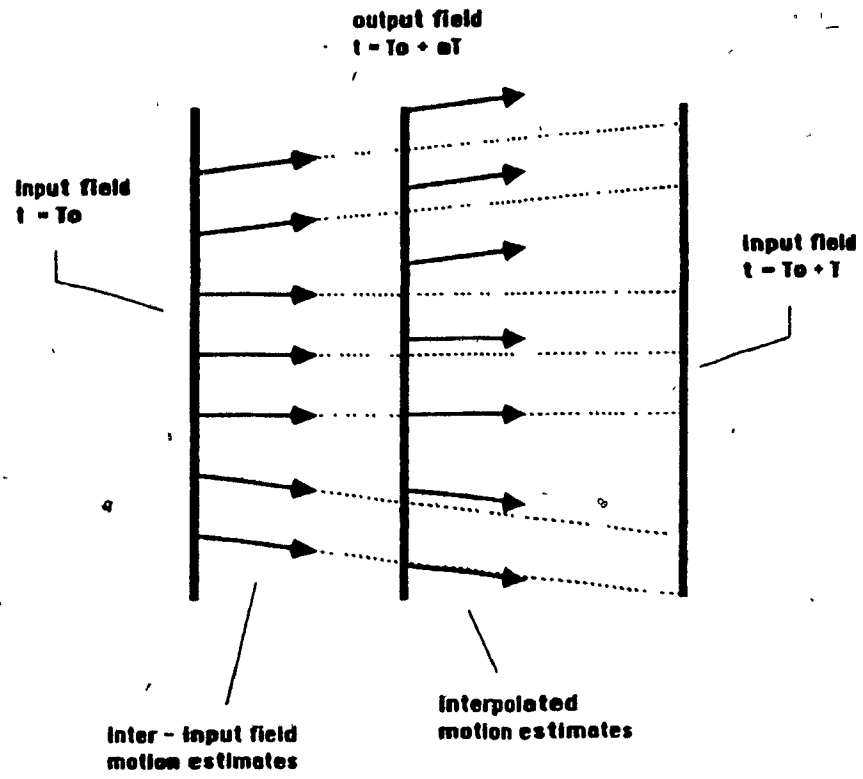

Figure 3.3 Interpolating Motion Estimates. Motion estimates required at the temporal position of the interpolated field may be derived from motion estimates between neighboring input fields. Only the interpolated motion estimates from the output field to the following input field are shown.

The correspondence fields for the interpolated field may also be derived by simply using the motion estimates that exist at the same spatial positions in the correspondence field between input fields. Aside from the obvious error introduced by this approximation, *spatial* interpolation of the motion estimates is required due to the difference in line structure between the input and output fields.

The second approach to generating the correspondence fields at the desired temporal position avoids the problems associated with interpolating the correspondence fields by generating them directly. One method of generating the correspondence

fields directly requires the assumption of constant motion between the two neighboring input fields. In this scheme, the three dimensional horizontal-vertical-temporal displacement vectors from each output pel to the corresponding points in the input field preceding and the input field following are collinear. The motion estimate is calculated in an iterative fashion. Initially, the spatial components of the motion estimate are set to zero. A distortion measure is calculated between the luminance at the spatial position in the input field preceding and following the desired temporal position. If the distortion is below a certain threshold, the process stops; if not, the spatial displacements are adjusted, and the distortion measure is recalculated.

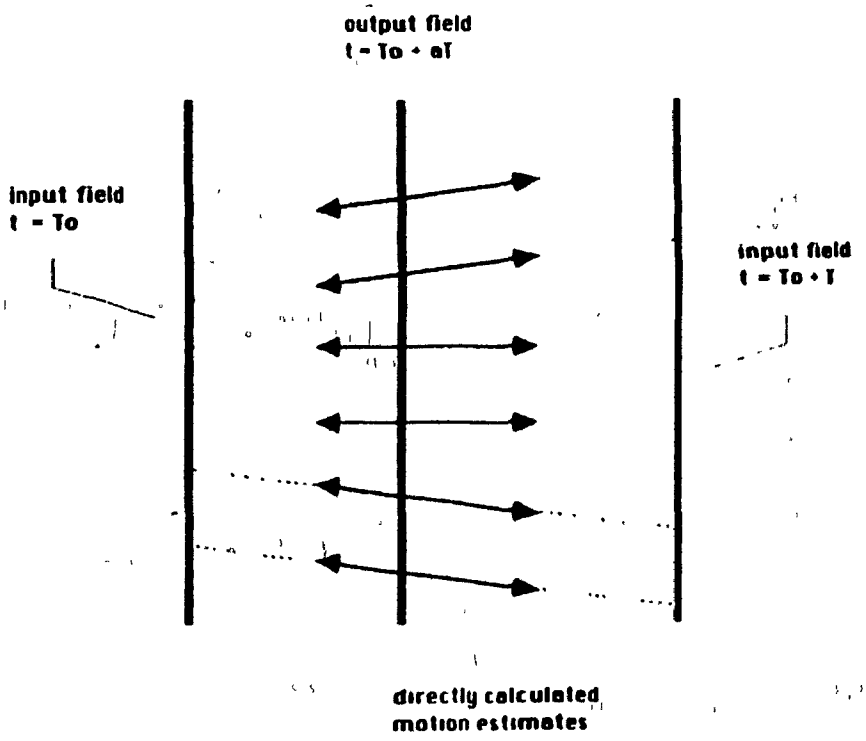

Figure 3.4 Direct Calculation of Motion Estimates. Calculating motion estimates directly is preferable to interpolating the estimates. The dashed lines connect corresponding areas in the input fields, and pass through spatial points upon which the interpolated field is defined.

This iterative procedure continues until an acceptable level of distortion is reached, or until the spatial displacement reaches a predetermined limit. Choice of the distortion measure and the method by which the spatial displacements could be adjusted is wide. The distortion measure could be calculated from luminance values in each input field over a small region. A good choice for the 'region' would be a contiguous moving region identified by the segmentation process. Possibilities for the displacement adjustment include methods based on the gradient constraint equation, and simple grid searches within defined limits.

3.1.3 Spatial Interpolation

The final process in the motion compensated television standards converter is the interpolation process, which generates interpolated fields using information from the segmentation process, the motion estimation process, and the input field sequence. The essence of the interpolation process is to calculate the value of each pel in the interpolated field based on motion compensated input fields. Recall from Chapter 2 that each input field contains one half of the available vertical samples of the image in the camera; a frame formed by interlacing two successive fields contains a maximal set of vertical samples of a stationary image. A field taken by itself is vertically aliased, so a field interpolated using only a single input field must suffer a loss of vertical resolution to suppress the vertical aliasing. To maximize the resolution of an interpolated field, information for each pel is taken from the input fields preceding and following the interpolated field with a spatial displacement equal to the motion estimate supplied by the motion estimation process.

Due to the use of fractional motion estimates, and the fact that line density of the interpolated field sequence is different than that of the input field sequence, it is necessary to *spatially* interpolate information required from the input fields using intra-field interpolation. The spatial interpolation is performed using an aperture function as described in section 2.3, using either separable horizontal and vertical

apertures, or a two dimensional horizontal-vertical aperture. Information is interpolated from each of the two neighboring input fields at a displacement equal to the appropriate motion estimate, then averaged to produce each pel in the interpolated field. In effect, regions of the image in each input field are projected onto the plane of the interpolated field in the direction of the motion estimates (Figure 3.5).

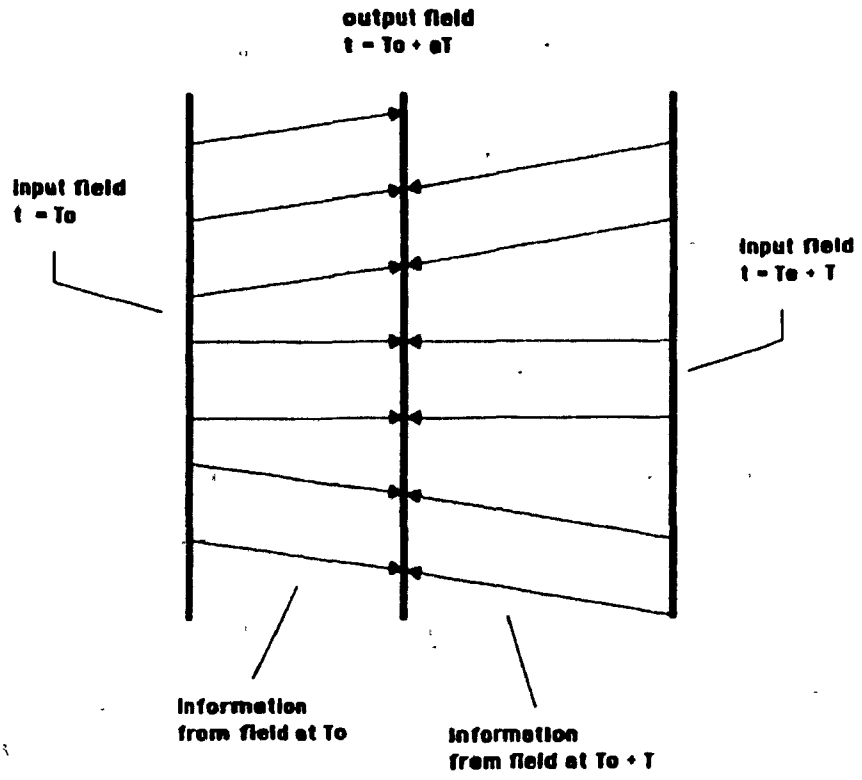

Figure 3.5 Motion Compensated Interpolation. Motion compensated interpolation is implemented by taking information from neighboring fields, in the direction of the motion estimate from the neighboring field to the interpolated field.

To maintain the integrity of object edges in the interpolated field, the interpolation process makes use of information from the segmentation process. In the case of newly exposed background, information about that portion of the image exists only

in the field following the interpolated field: in the case of newly covered background, only in the field preceding. In both of these cases, the interpolation process must use only information from one field. If the interpolation process attempts to extract information from both fields, the interpolated edge of the object will be an average of the edge and the background behind it, resulting in a blurring of the edge.

3.1.3.1 Adaptive Vertical Interpolation

To maximize the resolution in an output sequence while avoiding the introduction of spatial artifacts, it is necessary to *adapt* the vertical aperture to the amount of vertical aliasing in each particular region. The need for adapting the vertical aperture arises from the field-interlace structure of the television signal. When there is no motion in the scene being imaged in a television camera, pairs of sequential fields are independent sets of samples of the image; the image is sampled at a rate equal to the number of lines per frame in the vertical dimension, and the number of fields per second in the horizontal dimension. However, when there is vertical motion in the scene, the sets of vertical samples which comprise sequential fields may no longer be independent; in fact, when the entire scene moves with vertical motion equal to an odd number of frame lines per field period, sequential fields are *identical* except for a few lines at the top and bottom of the fields. In this case, the vertical sampling rate drops to one half of that for scenes with no vertical motion, causing severe vertical aliasing, and necessitating the adjustment of the frequency response of the vertical aperture.

When there is no vertical motion in the scene, the vertical aperture is specified to have a very high cutoff frequency since the vertical aliasing in the sequence is minimal. The high cutoff filter allows the retention of a maximum amount of information from the baseband spectrum, generating an output sequence with resolution approaching that of the input sequence. When there is vertical motion in the input scene, the effective vertical sampling rate changes, necessitating the adjustment of the frequency response of the aperture.

For vertical motion equal to an odd number of frame lines per field period, the vertical sampling rate is effectively decreased by a factor of two, causing the sampled signal to be significantly aliased in the vertical dimension. To interpolate an output field of acceptable quality, it is necessary to scale back the cutoff frequency of the aperture function to a value approximately one half that of the no-motion case. By reducing the cutoff frequency of the vertical aperture, the range of the vertical frequency spectrum which contains significant aliasing is suppressed, but at a cost of losing information from what was the baseband frequency spectrum; the net effect is a distinct loss of resolution.

With vertical motion that is not exactly an odd number of frame lines per field period, that is, vertical motion which does not cause lines in sequential fields to be identical samples of the image, the realm of non periodic sampling is entered. Sampling theory, and conventional concepts of frequency domain representations are no longer applicable. Empirically, it may be considered that the effective sampling rate has decreased from the original sampling rate to a value equal to or greater than one half of the original sampling rate, as these are the boundary conditions from the two cases above. Following this argument, the cutoff frequency of the aperture should be reduced to a value inbetween that of the no motion case and the vertical motion of an odd number of frame lines per field case to reject aliased vertical components.

The implementation of an adaptive vertical interpolation scheme to maximize the resolution of interpolated fields requires a rule for switching between apertures. Since the aperture must be adapted to the amount of vertical aliasing in each region, the vertical motion estimate for that region may be used as the basis for the switching rule. Ideally, a continuously-variable aperture would be used; in practice, one of several fixed apertures could be chosen, based on the magnitude of the vertical motion estimate. A switching rule for selecting one of three apertures, one with a low cutoff frequency to suppress aliased vertical frequency components

when the vertical motion is an odd number of frame lines per field period, a second aperture with a high cutoff frequency to retain maximum baseband information when the signal is not aliased due to vertical motion, and a third aperture with an intermediate cutoff frequency for vertical motion of a magnitude between these two boundary conditions, would have the form

$$\begin{aligned}
REM_2[\mathbf{v}_{\mathbf{f}_{x_2}}(\mathbf{x}) - \mathbf{v}_{\mathbf{p}_{x_2}}(\mathbf{x})] &= 0: \text{ choose high cutoff frequency} \\
0 < REM_2[\mathbf{v}_{\mathbf{f}_{x_2}}(\mathbf{x}) - \mathbf{v}_{\mathbf{p}_{x_2}}(\mathbf{x})] &< 1: \text{ choose intermediate cutoff frequency} \\
REM_2[\mathbf{v}_{\mathbf{f}_{x_2}}(\mathbf{x}) - \mathbf{v}_{\mathbf{p}_{x_2}}(\mathbf{x})] &= 1: \text{ choose low cutoff frequency} \\
1 < REM_2[\mathbf{v}_{\mathbf{f}_{x_2}}(\mathbf{x}) - \mathbf{v}_{\mathbf{p}_{x_2}}(\mathbf{x})] &< 2: \text{ choose intermediate cutoff frequency,}
\end{aligned}$$

where $\mathbf{v}_{\mathbf{p}_{x_2}}(\mathbf{x})$ is the vertical motion estimate at spatial position $\mathbf{x}$ from the interpolated field to the preceding field, measured in frame lines per field period, $\mathbf{v}_{\mathbf{f}_{x_2}}(\mathbf{x})$ is the vertical motion estimate to the following input field, and the function $REM_2(\bullet)$ takes the remainder of the argument after dividing by 2. Since the motion estimates are customarily quantized, the argument of the switching rule frequently takes on values of exactly 0 and 1.

3.1.3.2 Horizontal Interpolation

In order to avoid spatial discontinuities due to coarsely quantized motion estimates, it is necessary to use fractional pel motion estimates. This implies that the luminance and the chrominance values from the input fields must be *horizontally* as well as vertically interpolated at fractional points between the discrete lattice upon which the input fields are defined. The television signal does not suffer from the effects of an interlace scanning scheme in the horizontal dimension; with accurate motion estimates, the effective horizontal sampling rate is constant and independent of any motion in the scene. This allows a one-time specification of the horizontal aperture used for spatially interpolating information when a fractional motion estimate is encountered.

3.2 The Experimental System

The purpose of constructing the experimental motion compensated standards converter is the examination of interpolation aspects of the motion compensated standards conversion problem, apart from the segmentation problem and the efficient motion estimation problem. Separate examination of the interpolation problem requires the implementation of several key simplifications to the ideal system of section 3.1.

The major simplification is the restriction of the input sequences to be sequences containing only uniform translational motion over the entire image. If the translation is strictly horizontal, this type of sequence is known as a *pan*. By restricting the input to have identical motion throughout the field, the segmentation problem is entirely eliminated, and the motion estimation problem is significantly simplified since only a single estimate for the entire field need be generated.

A second simplification is the use of a simple and effective although inefficient motion estimation scheme. The scheme used is a block matching technique which produces motion estimates from one *input* field to the next. To generate motion estimates from the temporal position of the output field to the two neighboring input fields, the motion estimate vector from the input field preceding the output position to the input field following is scaled in proportion to the distance between the output position and the respective input. This technique, while the most reasonable approach for sequences with constant translation over the entire field, is less useful for general scenes which have several moving objects due to the difficulty in interpolating the correspondence fields at the output temporal position which are accurate enough to avoid the introduction of spatial discontinuities in the output field. With these two key simplifications, an experimental motion compensated television standards converter may be constructed to investigate the interpolation aspects of the problem.

3.2.1 Motion Estimation Algorithm

The motion estimation algorithm used in the experimental motion compensated standards converter implemented for this thesis generates a single motion estimate for the entire field using a block matching technique as described in section 2.6.2. The distortion measure used is $g(x) = x^2$, giving a minimum mean square error measure as in [14]. The search pattern used is a simple grid search of all possible integer displacements within specified limits followed by a fractional pel search. The input field is divided into four quadrants, and a motion estimate is generated for each quadrant using the luminance information. The average of the motion estimates from each quadrant is used as the motion estimate for the entire field. Averaging the motion estimates from four quadrants tends to average erroneous results from areas where the block matching technique diverges, such as flat areas, or areas which are shift-invariant in one dimension, such as vertical bars.

The simplified motion estimator generates motion estimates of quarter-pel accuracy. Fractional pel accuracy is achieved by performing the integer-grid search, and then searching eight points displaced $\pm q$ in each dimension about the most recent direction of minimum distortion, for $q = \frac{1}{2}$ and then again for $q = \frac{1}{4}$. Luminance values at fractional positions between pels for the motion estimator are interpolated using bilinear interpolation. Bilinear interpolation is a separable horizontal-vertical linear interpolation which uses the four pels spatially neighboring the fractional position. First, interpolation is carried out in the vertical dimension, generating two intermediate values at the same vertical position as the fractional position. Then, these two intermediate values are interpolated horizontally to give the luminance value at the fractional position (Figure 3.6).

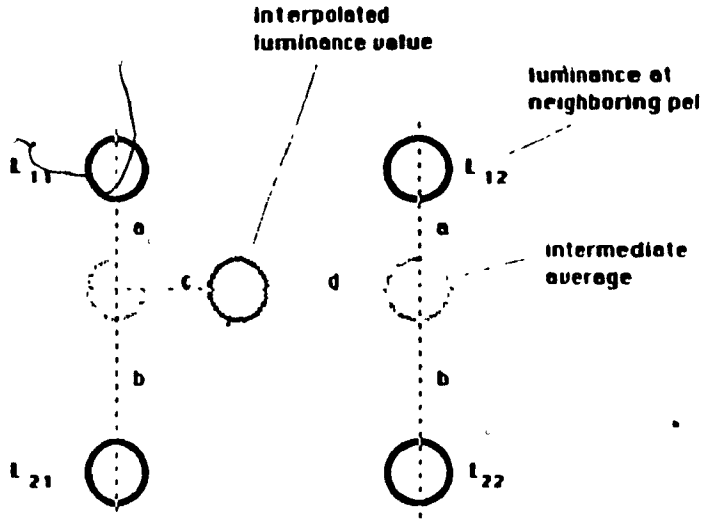

Figure 3.6 Bilinear Interpolation Luminance values at fractional displacements for the motion estimation algorithm are spatially interpolated using bilinear interpolation The luminance l, interpolated at the fractional position is
$l_i = (1-c)\,((1\quad a)l_{11} + (1\quad b)l_{21}) + (1\quad d)\,((1\quad a)l_{12} + (1\quad b)l_{22})$

This motion estimation algorithm generates a motion estimate between input fields. Motion estimates at each desired temporal position are interpolated from motion estimates generated between the input field preceding the desired position and the input field following the desired position. The magnitude of each interpolated estimate is simply the estimate between input fields scaled by the temporal distance between the desired temporal position and the respective input field; *i.e.* the motion estimate at the interpolated field is linearly interpolated from the motion estimate between input fields. This simplification is allowed by the fact that the single estimate applies to the entire field.

3.2.2 Spatial Interpolation

The proposed system implements vertical interpolation for line structure conversion in a similar manner to other converters that have separable vertical-temporal interpolation processes. Due to the use of fractional pel motion estimates, horizontal interpolation of the luminance and chrominance values from spatial positions between the points of the discrete lattice in the input fields is required. Vertical and horizontal interpolation is implemented with two separate apertures; the vertical aperture spans four field lines, and the horizontal aperture spans four pels. Both apertures are generated using the method of section 2.5 with N=8 and P=15, yielding a discrete vertical aperture defined at intervals of 1/32 of a field line, and a discrete horizontal aperture defined at intervals of 1/32 of a pel. Using these separate apertures, the sixteen spatially closest pels are combined to interpolate the value of the luminance and the chrominance at fractional pel positions specified by the motion estimate. While the interpolation could be carried out with a non separable two dimensional horizontal-vertical aperture with greater efficiency, and perhaps better subjective results, separate apertures are used in the proposed system to allow the independent specification of the vertical and horizontal frequency response of the apertures, which simplifies the subjective optimization of the apertures.

The proposed system uses two vertical apertures with different frequency responses depending on the vertical component of the motion estimate. One vertical aperture has a high cutoff frequency for regions with little vertical aliasing, while the other has a low cutoff frequency to suppress severe aliasing. An aperture of intermediate cutoff frequency was not implemented in the experimental converter. Since the switching rule and the frequency responses of the apertures were determined subjectively by the author during the development of the experimental system, they are included with the results in the following chapter.

Chapter 4 Results

The results obtained from this research may be divided into two categories: results arising from the development and implementation of the experimental motion compensated standards converter, and results arising from subsequent subjective testing of the converter. The system of section 3.2 was implemented using FORTRAN software running on a Digital Equipment Corporation VAX 11/780 computer, and the High resolution Digital Video sequence Store (HDVS) system [19] at the BNR/INRS-Télécommunications research lab at Nun's Island for capture and display of video sequences. Results arising from the development and implementation of the experimental converter include the determination of the ideal system configuration which is presented in Chapter 3 and the subjective determination of the frequency response of the vertical and horizontal interpolation apertures, and a switching rule for adapting the vertical aperture frequency response to the amount of vertical aliasing in the input signal. In subjective tests, the experimental motion compensated standards converter was compared against several types of existing standards converters. The converters were compared using several test sequences including a slow pan of a 'typical' high detail, high frequency content scene, a sequence containing high speed translation of a high contrast scene, and a sequence containing vertical motion of one frame line per field period in a scene with high frequency content, which causes severe vertical aliasing.

4.1 Development of the Experimental System

4.1.1 Interpolation Apertures

The interpolation apertures chosen for the experimental motion compensated standards converter were interpolation apertures which produced the best subjective output quality for the various test sequences, given the system defined in section 3.2. The system was developed in three stages: first, a system which produced subjectively good quality output with stationary images was developed. Then, the system was modified to generate motion estimates, and to produce subjectively good quality output for panning sequences. Finally, the system was modified to accept general translational motion. For each stage, the subjective effects of many different apertures were examined. Using the frequency sampling method of aperture specification, apertures with differing cutoff frequencies and transition band characteristics were generated in an iterative fashion. Once an aperture was generated, it was incorporated in the experimental motion compensated standards converter and was used to convert several test sequences. By viewing the results of these tests, the merits and weaknesses of each aperture could be evaluated. In all cases of interpolation aperture specification, there exists a tradeoff between the suppression of aliased frequency components to eliminate artifacts and the retention of baseband signal information to maximize resolution. Spatial artifacts and arithmetic overflow identified apertures which did not sufficiently attenuate aliased frequency components, while loss of resolution identified apertures which attenuated the baseband frequency spectrum. Over a period of several months, the tradeoffs involved in aperture specification were defined, apertures were chosen for each characteristic input condition, and a switching rule for adapting the aperture to the input conditions was defined. In each of the three stages of development, interpolation apertures were chosen to generate the best *subjective* output for the particular input sequences.

The first stage of development of the experimental converter was a converter whose input was restricted to still scenes. In the case of an input sequence containing the image of a still scene, sequential fields are independent sets of vertical samples of the image spaced at intervals of exactly two frame lines. This is the maximum vertical sampling rate attained in a given system under any conditions. Because of the high vertical sampling rate, the sampled signal resulting from a still scene is relatively free of vertical aliasing. This allows the tradeoff between suppression of aliased frequency components and the retention of baseband information to be heavily weighted in favour of retaining the maximum possible amount of information from the baseband.

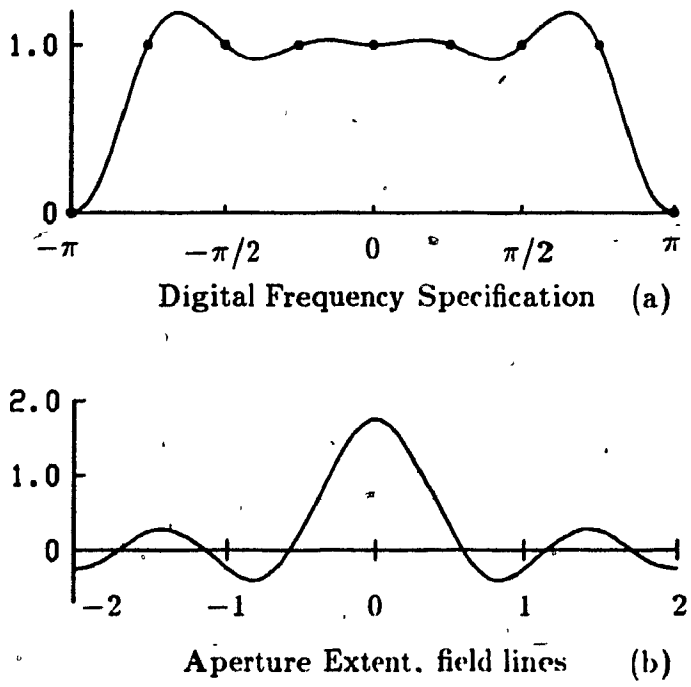

Figure 4.1 Vertical Aperture for Stationary Scenes. (a) Frequency domain (b) Time domain This high cutoff frequency aperture, specified at eight points in the frequency domain using the frequency sampling method, is used for stationary scenes since the amount of vertical aliasing in the input signal is minimal.

It was found that a vertical interpolation aperture specified to have a very high cutoff frequency and sharp transition band (Figure 4.1) produced interpolated fields with maximum resolution and contrast. Apertures with lower cutoff frequencies and smoother transition bands produced interpolated fields with significantly lower resolution, due to attenuation of the baseband frequency spectrum.

Refining the system to accept horizontal motion necessitated the introduction of a motion estimation process and the capability to use the motion estimates in the interpolation process. It was found that motion estimates of 1/4 field line accuracy were sufficient to eliminate artifacts that would be caused by an overly coarse motion estimator. Lower accuracies produced movement discontinuities, while higher accuracy produced no subjective improvement. To implement fractional pel motion estimates in the horizontal dimension, it is necessary to horizontally interpolate information at fractional positions from neighboring pels. A horizontal aperture spanning two pels, implementing linear interpolation (an inverse-distance weighting scheme identical to that described in section 2.4.1.1) was implemented, but it was found that the degradation in horizontal resolution was unacceptable. A four-pel aperture with an empirically specified frequency response was chosen to give the best possible horizontal resolution. The form of this subjectively chosen aperture is shown in Figure 4.2. Since the field-interlace structure of the television signal does not cause the effective horizontal sampling frequency to vary with the magnitude of horizontal motion in a sequence, a single fixed aperture was used for all horizontal interpolation. For vertical interpolation, the aperture used for stationary input (Figure 4.1) again afforded the best resolution and contrast since there was no vertical motion in the input sequence.

The final stage in the development of the motion compensated standards conversion system was the modification of the system to accept general translational motion. The vertical aperture selected for still scenes caused spatial artifacts in

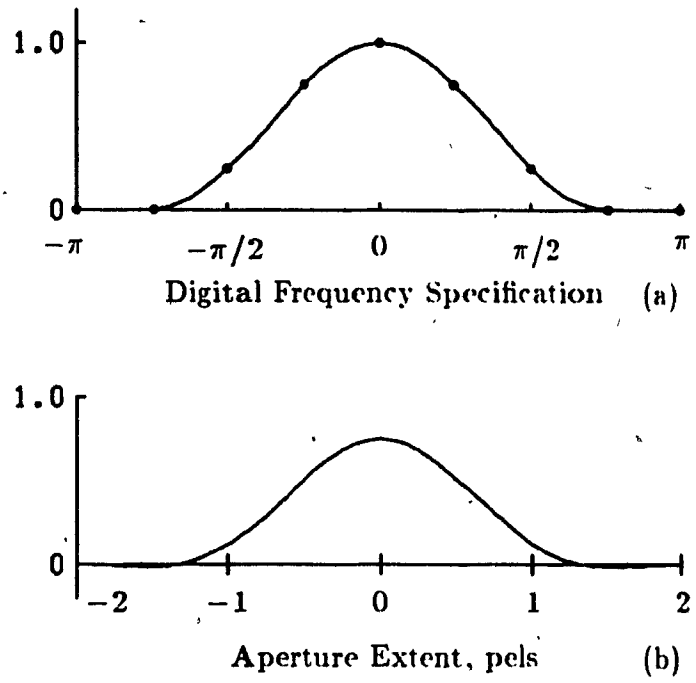

Figure 4.2 Horizontal Aperture. (a) Frequency domain (b) Time domain This aperture, specified at eight points in the frequency domain, was subjectively chosen to produce the best horizontal resolution in interpolated fields

the form of an aliasing of the input vertical sampling structure onto the interpolated fields, and motion artifacts in the form of vertical motion discontinuities to occur when there was vertical motion in a sequence. To eliminate these artifacts, it was necessary to suppress vertically aliased components of the input frequency spectrum by reducing the cutoff frequency of the vertical aperture. A test sequence consisting of a scene with high frequency content being moved at a constant rate of one frame line per field period (causing worst-case vertical aliasing) was recorded, and a vertical aperture which suppressed visible artifacts yet retained a maximum degree of resolution was chosen (Figure 4.3).

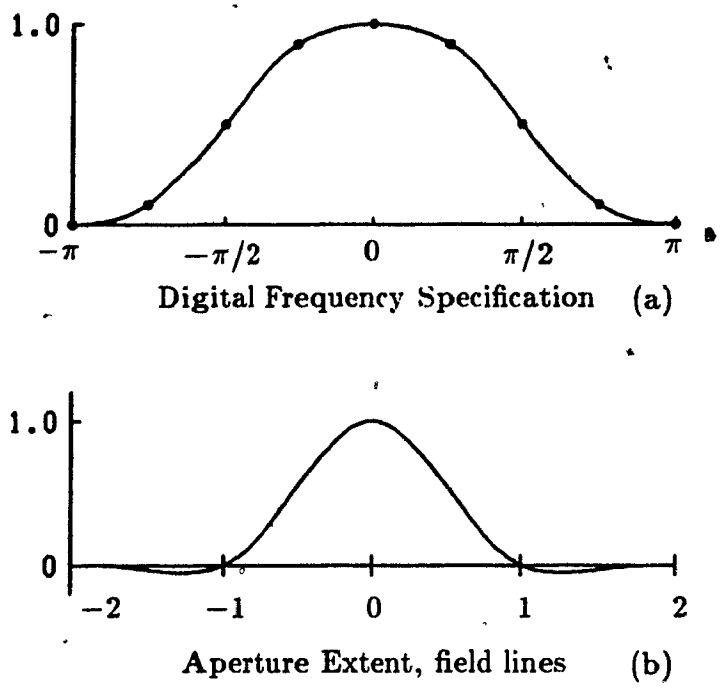

Figure 4.3 Vertical Aperture for Scenes with Vertical Motion. (a) Frequency domain (b) Time domain. This low cutoff frequency aperture, specified at eight points in the frequency domain, is used for scenes with vertical motion to supress aliased frequency components existing in the input signal.

4.1.2 Adaptive Vertical Interpolation

To maximize the subjective quality of sequences interpolated from a sequence containing varying amounts of vertical motion, it was necessary to adapt the frequency response of the vertical aperture to the amount of aliasing in the input sequence. Due to the difficulty in recording test sequences with constant motion, and practical time constraints on the research, a separate aperture for vertical interpolation when the vertical motion was of a magnitude between integer numbers of frame lines per field was not chosen. The aperture used for worst-case vertical motion was used when the vertical motion was not equal to an even number of frame lines, *i.e.* when the input signal had any noticeable degree of vertical aliasing.

Since the motion estimation process provided vertical motion estimates quantized to 1/4 field line per field period, or equivalently, 1/2 frame line per field period, the switching rule used in the experimental converter was

$$\begin{aligned} REM_2[\mathbf{v}_{\mathbf{f}x_2} - \mathbf{v}_{\mathbf{p}x_2}] &= 0: \text{ choose high cutoff frequency} \\ REM_2[\mathbf{v}_{\mathbf{f}x_2} - \mathbf{v}_{\mathbf{p}x_2}] &= \frac{1}{2}: \text{ choose low cutoff frequency} \\ REM_2[\mathbf{v}_{\mathbf{f}x_2} - \mathbf{v}_{\mathbf{p}x_2}] &= 1: \text{ choose low cutoff frequency} \\ REM_2[\mathbf{v}_{\mathbf{f}x_2} - \mathbf{v}_{\mathbf{p}x_2}] &= 1\frac{1}{2}: \text{ choose low cutoff frequency} \end{aligned}$$

where $\mathbf{v}_{\mathbf{p}x_2}$ is the vertical motion estimate from the interpolated field to the preceding field, measured in frame lines per field period, and $\mathbf{v}_{\mathbf{f}x_2}$ is the vertical motion estimate to the following input field, and the function $REM(\bullet)$ is as defined in Chapter 3.

The effect of this simplified switching rule was to provide maximum resolution when these was no vertical motion, maximum possible resolution while suppressing the severe vertical aliasing that occurs when the vertical motion is an odd number of frame lines per field period, and to provide sub-optimal resolution while suppressing vertical aliasing when the vertical motion had a magnitude between the two boundary conditions.

4.2 Subjective Testing

Once the development of the experimental system was complete, subjective tests were carried out. The intent of the subjective testing was not to produce a definitive study on the subjective quality of motion compensated standards conversion, but to confirm the conclusions of the author, and to characterize the motion compensated standards converter in relation to other popular conversion schemes.

Subjective testing was conducted based on the method for full-range quality grading of broadcast television pictures outlined in Chapter 7 of [20], and standard procedures established at the Nuns' Island research facility. The experiments were carried out in the subjective testing facility at Nuns' Island, which consists of a high-quality Tektronix monitor, space for three viewers seated four picture heights from the monitor, and a computer terminal, in an atmosphere of subdued lighting and noise. This subjective testing facility approaches the CCIR recommendations for subjective experiments concerned with domestic television [20]. Since the HDVS system currently records and displays stored sequences at 60 fields per second, and 525 lines per field, converted sequences cannot be displayed at the intended field rate or line density. A sequence which has been converted from a 525/60 standard to a 625/50 standard is displayed at 60 fields per second, 525 lines per field regardless of the intended field rate and line density. Such a sequence appears to be magnified vertically and speeded up compared to the original when viewed on the 525/60 display. This effect is not considered to disturb the subjective evaluation of the converters, since the subjects are not shown the 'original', and are evaluating primarily the magnitude of motion artifacts and relative resolution.

A small group of mixed expert and non-expert viewers rated presentations of various converters acting on several test sequences. For consistency, all converters were operating in the 525/60 to 625/50 mode. The presentations were shown in the randomized order suggested in [20] to average the effects of subject adaptation over the course of the experiment. The subjects were divided into four groups of three, and rated the presentations in two separate sessions one day apart. Each subject was shown each presentation a total of four times over the two days, twice the first day and twice the second day: separating the experiment to take part over two days averages day-to-day variations in the subjects' evaluations.

Four converters including the experimental motion compensated standards converter were used in the subjective test. These converters were

- A "zero-order" converter which drops fields for frame rate conversion and uses a four-line aperture for line density conversion;
- The DICE converter which uses two-field linear temporal interpolation and a four-line aperture for line density conversion;
- The ACE converter which uses a two dimensional four-line, four-field aperture for frame rate and line density conversion; and
- The experimental motion compensated converter.

Each of the converters was used to process four different test sequences. These sequences were selected as sequences which comprise typical as well as the worst-case input conditions for the converters. The sequences were

- "QUILT", a slow pan of a high detail scene with high frequency content;
- "FAST PAN", a high speed strictly horizontal pan of a high contrast scene consisting of words and phrases such as those found on large advertisements;
- "DIAGONAL", the same scene as used in 'Quilt', being translated at one pel per field in the horizontal direction and one frame line per field in the vertical direction (causing worst-case vertical aliasing); and
- "VERTICAL", high speed, strictly vertical translation of the same scene used for 'Fast Pan'.

The presentations were stored digital sequences of 48 fields, displayed in colour for approximately 10 seconds in palindromic fashion (played start to finish, then finish to start, then repeat). A period of approximately 10 seconds between presentations was allowed for the subjects to record their evaluations on a scale of A (highest rating) to E (lowest rating). The subjects were encouraged to use the whole range of choices, and no suggestions were given for "appropriate" responses.

Once the subjective tests were complete, the subjects' responses were tabulated and averaged by assigning a numerical value to each of the letter scores A to E. 'A', the highest rating, was assigned a value of 1.0; 'B', 0.75; 'C', 0.5; 'D', 0.25; and 'E', the lowest rating, was assigned a value of 0. The overall results of each converter acting on the four sequences are shown in graphical form Figure 4.4.

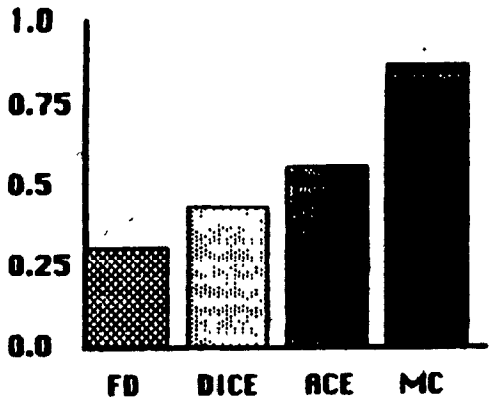

Figure 4.4 Overall Results of Subjective Testing. Averaged subjective evaluations of the four converters, the zero-order field-dropping converter (FD), the DICE converter, the ACE converter and the experimental motion compensated converter (MC) are presented.

These results show a steadily increasing perception of quality in relation to the complexity of the converter. It is more useful to break down these results into subjective evaluation of each sequence processed by the converters, and to analyze the statistical characteristics of the measurements. Figure 4.5 shows the averaged quality assessment for each converter for each of the four test sequences, and Table 4.1 presents the mean and standard deviation for each of the sixteen presentations.

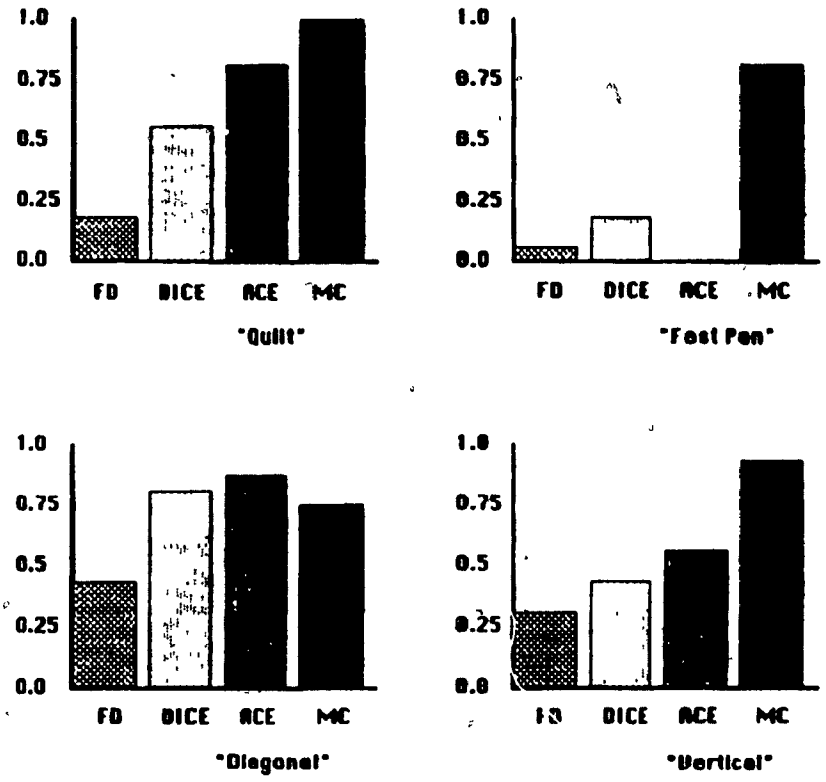

Figure 4.5 Results of Subjective Testing For Each Test Sequence. Averaged subjective evaluations of the four converters acting on each of the four test sequences are presented. Note the progressive increase in quality rating for the 'typical' sequence 'Quilt', the equal ratings for the vertically-aliased sequence 'Diagonal', and the superior performance of the motion compensated converter for the 'Fast Pan'.

For the sequence 'Quilt', a slow pan of a scene with high detail and high frequency content intended to test the converters under 'typical' conditions, the results reflect the overall averages of Figure 4.4. The experimental motion compensated converter exhibited markedly better vertical resolution than either the DICE or the ACE converters, presumably due to the non-adaptive nature of the vertical interpolation apertures in the latter two converters, which requires those converters to use a worst-case vertical motion aperture for all sequences. By using a low

cutoff-frequency vertical aperture for sequences containing no vertical motion, the ACE and DICE converters attenuate some of the baseband frequency spectrum and thus lose vertical resolution. To see if the distribution of mean scores for the sequences presented in Figure 4.5 are meaningful, the standard deviations of the measurements are examined. The standard deviation is calculated as

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x - x_i)^2} \qquad (4.1)$$

where $\bar{x}$ is the sample mean. The standard deviation gives an indication of the spread of the distribution of evaluations for each sequence, and in particular, the root-mean-square value of the variations about the mean of possible evaluations [21]. Table 4.1 presents the mean scores and standard errors for each of the sixteen presentations.

Converter	Sequence							
	QUILT		DIAGONAL		FAST PAN		VERTICAL	
	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
FD	0.20	.17	0.47	.21	0.14	.15	0.30	.17
DICE	0.63	.17	0.79	.17	0.22	.15	0.48	.17
ACE	0.84	.18	0.83	.18	0.05	.09	0.57	.21
MC	0.96	.09	0.73	.22	0.80	.15	0.94	.11

Table 4.1 Analysis of Subjective Evaluations of Various Converters. This table gives the mean and standard deviation of the subjective evaluations of each of the four converters, the field-dropping zero-order converter (FD), the DICE converter, the ACE converter, and the experimental motion compensated converter (MC), acting on the four test sequences.

Examining the standard deviations for the evaluation of the sequence Quilt, it is seen that the standard deviations are significantly large compared to the difference between the sample means, implying that a larger group of subjects is needed before statistically proven conclusions may be drawn for this sequence.

Averaged results for the sequence 'Diagonal', a sequence containing the same scene as in 'Quilt', but being translated with vertical motion at one frame line per field as well as slow horizontal motion, indicate that all three of the multi-field converters scored virtually identical subjective quality ratings. The standard deviations of the evaluations are much greater than the differences between the means for this sequence, allowing the conclusion that there is no real difference between these three conversion schemes. In order to suppress the aliased vertical frequency components in the input signal, all conversion schemes must use vertical interpolation apertures which attenuate some of the baseband frequency spectrum, losing vertical resolution.

The results for the two sequences containing high speed translation of a high contrast scene, 'Fast Pan', and 'Vertical', indicate clearly the superiority of motion compensated interpolation over linear techniques for these types of sequences. The mean of the evaluations for the experimental motion compensated converter for both of the rapid translation sequences are several standard deviations away from the next best rated converter. For the sequence 'Fast Pan', the motion artifacts introduced by all three linear converters caused very low subjective quality ratings. The experimental motion compensated converter was able to produce an interpolated sequence completely free of motion artifacts, with good resolution.

It was noted during the preparation of the experiment that the motion discontinuities created by the linear converters for the sequence 'Vertical' were not as visible as those for the sequence 'Fast Pan', even though the two sequences contained the same scene being translated at the same speed. It was further noted that if a subject's head was tilted (or the monitor placed on its side) so that the vertical translation of the sequence 'Vertical' became horizontal translation, the motion artifacts were equally as visible as those in the sequence 'Fast Pan'. Presumably, the human visual system is much better at tracking horizontal motion than vertical motion; this phenomenon would suggest that the results for the sequence 'Vertical'

would be the same as those for 'Fast Pan' if the vertical translation were viewed as horizontal translation.

These subjective tests, while by no means comprehensive, allow several fundamental conclusions to be drawn concerning the advantages of motion compensated television standards conversion over traditional linear processing techniques. These conclusions can be summarized as: better conversion of high-detail, high frequency content scenes, equally good conversion of slow-moving scenes with severe vertical aliasing, and clearly superior conversion of scenes with high speed, large area motion.

Chapter 5 Conclusion

This chapter concludes the thesis, outlining the results obtained from the research, and the conclusions that may be drawn from them, as well as outlining the work that remains to be done in the field of motion compensated television standards conversion.

This thesis has defined a structure for the ideal motion compensated standards converter, and examined the interpolation aspects of the motion compensated television standards conversion problem. An experimental motion compensated converter was developed and implemented using FORTRAN software running on a VAX 11/780 computer and the HDVS digital video capture and display system at the BNR/INRS Nuns' Island research facility. Interpolation apertures for various input conditions were chosen, and a switching rule for adapting the vertical interpolation aperture frequency response to the amount of aliasing in the input signal was defined. Three main conclusions may be drawn from the results of subjective testing of the experimental motion compensated standards converter and three popular standards converters using linear processing techniques.

The first conclusion is that motion compensated interpolation can interpolate sequences from a typical high detail, high frequency content input sequence with better resolution and contrast than converters using linear processing techniques. This increase in performance is attributable to the fact that linear processing tech-

niques rely on spatial averaging of moving objects to reduce motion artifacts, reducing perceived resolution. Further, converters using linear techniques have fixed vertical interpolation apertures for all input conditions, implying that the fixed vertical interpolation aperture must be specified for worst case conditions, reducing vertical resolution needlessly in normal conditions. Motion compensated standards converters eliminate motion artifacts without loss of resolution by placing moving objects in the correct spatial positions in interpolated fields, and can adapt their vertical interpolation apertures to the amount of vertical aliasing in a sequence, allowing maximal vertical resolution at all times.

The second conclusion that may be drawn from the results is that motion compensated standards converters cannot produce better interpolated fields than linear processing techniques when a sequence contains low speed vertical motion that causes severe vertical aliasing. In fact, all of the multi-field standards converters tested produced virtually identical output under these conditions. However, when the vertical motion becomes large enough to cause the linear converters to introduce motion artifacts to the output sequence, the motion compensated converter can produce superior output even when there is severe vertical aliasing.

The third, and perhaps most important conclusion to be drawn from the results obtained from this research is that motion compensated standards converters can interpolate superior quality, artifact-free sequences from input sequences such as high speed pans of a high contrast scenes that cause converters using linear processing techniques to fail. This advantage can be attributed to the use of nonlinear processing, which produces interpolated sequences free of motion artifacts notwithstanding temporal aliasing in the input signal when high speed translation occurs.

It must be noted that these conclusions are based on sequences which require no segmentation. A system which can reliably segment the image into moving and stationary areas, and object edges, and then generate efficient, reliable motion estimates for each moving area remains to be developed. The development of a system which can interpolate fields containing images of independently moving objects, without introducing spatial discontinuities to the interpolated field remains as the next step in the development of the ideal system.

For motion compensated television standards conversion, it must be considered that the output field, and the errors in interpolation due to poor segmentation or poor motion estimation are viewed directly by the subject, as opposed to motion-compensated *coding* schemes, where the only result in the event of segmentation or motion estimation errors is an increase in the transmitted bit rate. It is possible that it will be extremely difficult to segment complex input fields well enough to interpolate a subjectively better sequence than the two dimensional vertical-temporal interpolation process used in the ACE converter. If this is the case, motion compensation could be used when large area motion, such as high speed panning occurs, but for complex scenes, traditional temporal interpolation would be employed. This suggestion introduces a fourth problem to the motion compensated television standards converter: the need to make a real time *decision* as to which technique would produce a better output sequence for a given input.

Whether a motion compensated system which handles general input, or a system which combines the best of traditional linear processing techniques and motion compensated interpolation to form a converter which handles general input is developed, it may be concluded that motion compensation can eliminate the motion artifacts associated with high speed large area motion often seen in sequences interpolated using existing converters.

References

[1] R. Zavada, Moderator, "Workshop 2: Systems and Standards," presented at the High Definition Television Colloquium, Ottawa, May, 1985

[2] C. K. P. Clarke & N. E. Tanton, "Digital Standards Conversion: Interpolation Theory and Aperture Synthesis," BBC Research Department Report 1984/20, December 1984.

[3] J. Kumada, "Developmental State of Various HDTV Equipment including the MUSE System," presented at the High Definition Television Colloquium, Ottawa, May, 1985.

[4] —, "1125/60 - 625/50 Motion Compensated Standards Converter," NHK Science and Technical Research Laboratories, May, 1985.

[5] W. K Pratt, *Digital Image Processing*. John Wiley and Sons, New York, 1978.

[6] T Fujio, "The Present state of HDTV: What it takes and what should be done," presented at the High Definition Television Colloquium, Ottawa, May, 1985.

[7] A. V. Oppenheim and R. W Schafer, *Digital Signal Processing* Prentice-Hall, New Jersey, 1975.

[8] E. Dubois, "The Sampling and Reconstruction of Time-varying Imagery," *Proceedings of the IEEE*, vol. 73, pp. 502-522, April 1985.

[9] D. H. Kelly, "Flickering Patterns and Lateral Inhibition," *J Opt. Soc. Am.*, vol. 59, pp. 1361-1370, 1969.

[10] R. Paquin & E. Dubois, "A Spatio-Temporal Gradient Method for Estimating the Displacement Field in Time-Varying Imagery," *Computer Vision, Graphics and Image Processing*, vol. 21, pp. 205-221, 1983

[11] J. O. Limb & J A. Murphy, "Measuring the Speed of Moving Objects from Television Signals," *IEEE Trans. Commun.*, vol. COM-23, pp. 474-478, Apr. 1975

[12] C. Cafforio & F. Rocca, "Methods for Measuring Small Displacements of Television Images," *IEEE Trans. Inform. Theory*, vol. IT-22, No. 5, pp. 573-579, Sept. 1976.

[13] A. N. Netravali & J D. Robbins, "Motion Compensated Television Coding: Part I," *Bell System Tech. Journal*, vol. 58, No. 3, pp 631-670, Mar. 1979

[14] J. R. Jain & A K Jain, "Displacement Measurement and its Application in Interframe Image Coding," *IEEE Trans. Commun*, vol. COM-29, No. 12, pp. 1799-1808, Dec. 1981.

[15] T. Koga et. al., "Motion-Compensated Interframe Coding for Video Conferencing," in *Proc. National Television Conference*, NTC '81, pp. G5.3.1-G5.3 5

[16] Y. Ninomiya & Y Ohtsuka, "A Motion-Compensated Interframe Coding Scheme for Television Pictures," *IEEE Trans Commun*, vol. COM-30, No 1, pp. 201-211, Jan. 1982.

[17] R. Srinivasan & K R. Rao, "Predictive Coding Based on Motion Estimation," in *Proc Int. Conf. on Comm.*, ICC '84, Vol. 2, pp. 521-526

[18] J. Baldwin, "Digital Standards Conversion," *IBA Technical Review*, vol. 8, pp. 84-93, September 1976.

[19] R. Paquin, "Laboratoire de Traitement des Signaux Video de l'INRS-Télécommunications et les Recherches Bell-Northern," Volumes 1-4, 1983-1985

[20] J. Allnatt, *Transmitted-picture Assessment.* John Wiley and Sons, Chichester, England, 1983

[21] W. B. Davenport, Jr., *Random Processes.* McGraw-Hill, 1970

Bibliography

General References:

J. Allnatt, *Transmitted-picture Assessment*, John Wiley and Sons, Chichester, England, 1983.

E. Dubois, "The Sampling and Reconstruction of Time-varying Imagery", *Proceedings of the IEEE*, vol. 73, pp. 502-522, April 1985.

T. S. Huang, ed., *Image Sequence Processing and Dynamic Scene Analysis*, Springer Verlag, New York, 1983

H G. Musmann, P. Pirsch, and H. J Grallert, "Advances in Picture Coding", *Proceedings of the IEEE*, vol. 73, pp 523-548, April 1985.

A V. Oppenheim and R W. Schafer, *Digital Signal Processing*, Prentice-Hall, New Jersey, 1975.

W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, New York, 1978.

Television Standards Conversion:

J. Baldwin, "Digital Standards Conversion", *IBA Technical Review*, vol. 8, pp. 84-93, September 1976.

C K. P. Clarke & N. E Tanton, "Digital Standards Conversion: Interpolation Theory and Aperture Synthesis", BBC Research Department Report 1984/20, December 1984.

, "1125/60 - 625/50 Motion Compensated Standards Converter", NHK Science and Technical Research Laboratories, May, 1985.

Motion Estimation:

C. Cafforio & F. Rocca, "Methods for Measuring Small Displacements of Television Images", *IEEE Trans. Inform. Theory*, vol. IT-22, No. 5, pp. 573-579, Sept. 1976.

J. R. Jain & A. K. Jain, "Displacement Measurement and its Application in Interframe Image Coding", *IEEE Trans. Commun*, vol COM-29, No 12, pp 1799-1808, Dec. 1981

T. Koga et. al., "Motion-Compensated Interframe Coding for Video Conferencing", in *Proc. National Television Conference*, NTC '81, pp. G5.3.1-G5.3.5

J O. Limb & J. A Murphy, "Measuring the Speed of Moving Objects from Television Signals", *IEEE Trans Commun*, vol COM-23, pp 474-478, Apr. 1975

A. N. Netravali & J. D Robbins, "Motion Compensated Television Coding: Part I", *Bell System Tech Journal*, vol 58, No. 3, pp 631-670, Mar 1979

A N. Netravali & J D. Robbins, "Motion Compensated Television Coding: Some New Results", *Bell System Tech. Journal*, vol. 59, No 9, pp. 1735-1745, Nov. 1980.

Netravali et al, "Video Signal Interpolation Using Motion Estimation", United States Patent 4,383,272, May 10, 1983

Y. Ninomiya & Y. Ohtsuka, "A Motion-Compensated Interframe Coding Scheme for Television Pictures", *IEEE Trans. Commun* , vol. COM-30, No 1, pp 201-211, Jan. 1982

R. Paquin & E. Dubois, "A Spatio-Temporal Gradient Method for Estimating the Displacement Field in Time-Varying Imagery", *Computer Vision, Graphics and Image Processing*, vol. 21, pp. 205-221, 1983

R. Srinivasan & K. R. Rao. "Predictive Coding Based on Motion Estimation", in *Proc. Int. Conf. on Comm.*, ICC '84, Vol. 2, pp. 521-526.