Motion-Induced Degradations of Temporally Sampled Images

by

Stephen Charles Hsu
B.S., California Institute of Technology
(1982)

Submitted in Partial Fulfillment
of the Requirements for the
Degree of

Master of Science
in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology
June 1985

Signature of Author ____________________
Department of Electrical Engineering and Computer Science
May 17, 1985

Certified by ____________________
William F. Schreiber
Thesis Supervisor

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

Motion-Induced Degradations of Temporally Sampled Images

by

Stephen Charles Hsu

Submitted to the
Department of Electrical Engineering and Computer Science
on May 17, 1985 in partial fulfillment of the requirements
for the Degree of Master of Science

Abstract

Current proposals for high-definition television (HDTV) standards do not plan to increase the effective temporal resolution, and as a result, motion-induced degradations will not improve. Since improvementscan come about only with a better understanding of the nature of these defects, the objectives of this research were (1) to categorize the subjective defects and (2) to explore their dependence on parameters of the image and of a model transmission system. The parameters include velocity, object size, presampling filter, transmission rate, and interpolation filter.

Properties of visual perception, along with signal and Fourier domain analysis of images, were used to explain the appearance, and predict the behavior, of defects in test patterns observed on a special-purpose 120 fps display system. The principal motion-induced defects, identified as blur, multiple images, and small-area flicker, were strongly dependent on observer eye movements. In the worst case, none of the transmission rates examined, regardless of the filtering scheme, would permit completely defect-free image reproduction at velocities typical of television. The nature of defects in future HDTV, and conditions for their elimination, are inferred from theory and observations.

Thesis Supervisor: Dr. William Schreiber

Title: Professor of Electrical Engineering

Contents

List of Figures

List of Symbols

Fd	δ-function impulse response
Fg	Gaussian impulse response
Fr	rectangular impulse response
Fr/2	rectangular impulse response (duration $T/2$)
Ft	triangular impulse response
Fu	offset triangular impulse response
s_0	original stationary image[1]
s_1	original moving image
s_2	camera image
s_3	transmitted image
s_4	displayed image
s_5	retinal image
s_6	perceived image
h_1	camera presampling filter
h_2	display interpolation filter
h_3	visual filter
h'_{1x}	spatial reflection of h_1
h'_{2x}	spatial reflection of h_2
$g'\|_t$	spatial reflection of h_3
v_1	image velocity due to object motion
v_2	image velocity due to eye motion
r	up-sampling rate
T	transmission frame period
$\mathcal{T}_v$	translation operator

[1] Corresponding Fourier transforms are capitalized, e.g., S_0

Acknowledgments

To my CIPG colleagues who put up with groping around the lab in the dark during my observations, I appreciated your cooperation. I would like to thank Steve Corteselli, Malik Khan, Jason Sara, Prof. Donald Troxel, and Russ Walker for invaluable assistance with the IPS and PCTV hardware, and Jean-Pierre Schott for advice on typesetting with LaTeX. I am grateful to my advisor, Prof. Schreiber, whose guidance was always insightful, if not frequent, and whose extraordinary patience granted me the freedom to work at my own deliberate pace. Finally, I dedicate this thesis to my parents for their enduring concern and sympathy.

Chapter 1

Introduction

Motion is the essence of television and the cinema. Ever more so than detail or color, motion catches the eye and engages the viewer. Motion communicates action; it tells the story.

On the other hand, motion is precisely the bane of these image transmission systems. There are two reasons for this. First, motion implies change, and reproduction of change requires transportation of information. Were it not for live-action motion, the information bandwidth could be vanishingly small. Second, every conventional scheme partitions time into intervals and for each transmits a frame to represent the evolving image during the whole interval. The dilemma is that, due to motion, a single instant is not representative of the whole, but an average over the whole does not describe any single instant very well. Thus, television and motion pictures do fairly well at rendering still images, but not so well with moving ones.

Because bandwidth must be finite (and small), the spatial and temporal resolution must be limited, of course. Severe limitations cause visible image degradations, that is, disagreeable discrepancies between the original and the observed reproduction. However, even if an image transmission system possessed ideal spatial characteristics, limited temporal resolution plus motion would still conspire to create apparent spatial image defects. Motion-induced defects are those which cannot appear on stationary image constituents — only on moving ones. As covered in this thesis, the subject of motion-induced defects is broader than that of motion rendition alone. Defective motion rendition means that the reproduced image betrays to the observer the underlying spatial and temporal quantization of movement. However, motion blur and geometrical distortions are motion-induced degradations as well.

The temporal fidelity of present-day systems leaves much to be desired. For example, flicker often appears on brighter parts of displays. Motion artifacts appear in both low-budget Saturday morning cartoons and serious cinematic productions (the author recalls Richard Attenborough's *Gandhi* in particular) because of excessive resolution in individual frames. More common is insufficient resolution of moving details, which, in the case of thin objects such as the legs of a flock of flamingos, can make them vanish into thin air.

Currently proposed standards for high-definition television (HDTV) such as the NHK 1125-line system [FUJI 82] and the Multiplexed Analog Components system [WIND 83] are designed primarily for the increase of spatial resolution. The frame rates of transmission and display are identical and differ little from those of NTSC or PAL; therefore, motion-induced defects cannot be fewer in the new standards. In fact, the defects may become more visible in the company of enhanced resolution in stationary parts of images and wider viewing angles. It is precisely this concern which motivates the present investigation.

1.1 Theories of Motion Rendition

There are two well formulated but incompatible theories of motion rendition. Common to both is the premise that the human visual system extracts certains signals out of the temporally discrete display, and that motion rendition is defective when the signals contain undesired spectral harmonics within the passband of the visual system. However, the theories differ in the nature of the extracted signals.

The older theory, which is closely related to apparent motion phenomena studied by psychologists, asserts that the relevant signal is that which represents object displacement as a function of time. The position of an object moving in discrete jumps is a staircase function, while that of an object moving smoothly is a ramp function

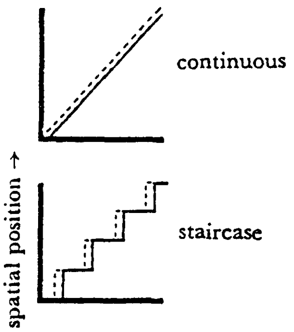


Figure 1.1: (a) Ramp; (b) staircase functions of object displacement vs. time (from [MORG 80A])

(Figure 1.1). If the spectral harmonics of the staircase are sufficiently attenuated by a temporal filter in the visual system, the motion rendition should be good and the perceived displacement function should be indistinguishable from a ramp [MORG 80A].

The second theory asserts that the relevant signal represents image brightness as a function of space and time. An image moving in discrete jumps can be shown to have a spatio-temporal spectrum with undesired harmonics in addition to the fundamental components possessed by a smoothly moving image. If the harmonics are sufficiently attenuated by the visual system's spatiotemporal filter, sometimes called the "window of visibility," the motion rendition should be good [WATS 83A]. Perhaps because the function of brightness vs. space and time is familiar to people in image processing, this theory has recently gained popularity.

Since the mapping between the spectra of the displacement signal and of the corresponding brightness signals is complicated, nonlinear, and not even one-to-one, the theories are quite different, not only qualitatively but also mathematically. The psychologist's question is, how can the theories be reconciled? The transmission system designer's question is, how can one or both of the theories be applied to improving motion rendition? (The material to be presented on visual perception will dissolve the incompatibility by disclosing independent pathways of vision. When experimental observations are discussed later, the second theory will be shown to be sufficient in practice for predicting unsatisfactory motion rendition.)

1.2 Objectives of this Research

Questions like the foregoing are part of the reason for studying motion-induced image degradations. We hope to gain a better understanding of visual perception. Eventually we would like to establish general bounds on accuracy of reproduction for subjectively defect-free images, for such data would be useful irrespective of the specific method of image coding and transmission. Finally, and more to the point, we need to evaluate existing and proposed image transmission systems.

As a modest step toward these goals, the specific objectives of this thesis are twofold. The first is to categorize the several kinds of subjective defects and to relate them to physical features of the displayed images. The second is to explore the dependence of these defects on variable factors in the original scene, the observer, and a canonical transmission system. The result will be a set of bounds on these variables within which individual defects are invisible.

Some studies concerning motion-induced defects already exist in the literature. Cunningham [CUNN 63] and Brainard [BRAI 67] reported problems with motion rendition in conjunction with conditional replenishment coding. Employing a frame buffer to convert a low-rate transmission to a 60 frames/sec (fps) display via frame repetition, Brainard decided that the greatest subjective change in rendition occurred between 12 and 15 transmitted fps. This figure is consistent with the long-established silent motion picture standard of 16 fps. Cunningham started with a 16 fps sequence, subsampled temporally by rate two or four, and then reconstructed the sequence by discrete temporal filters 0.5 sec in duration. These experiments noted tradeoffs between different defects.

Both of the preceding studies employed natural images, and thus the image conent and velocities were not controllable parameters. Moreover, the conclusions were primarily qualitative. Miyahara [MIYA 75] and Watson [WATS 83A] have performed psychophysical experiments to find the relation between frame rate and maximum velocity before defects were noticeable.

With existing studies such as these, what more is to be gained from the present investigation? For one thing, this thesis treats several variable factors jointly and consistently. Like the work of Cunningham, this study considers alternative temporal filters. While Miyahara's and Watson's observations were with pursuit eye movement and fixation, respectively, this study considers both modes of observation. Secondly, the categorization of defects in this study, believed to be novel, lessens ambiguity; for example, Miyahara gave thresholds for "jerkiness" but did not report his criterion for that defect.

1.3 Organization of the Thesis

The two chapters immediately following this introduction establish the background for the experimental work to be discussed in Chapters 4-7. The following is an annotated overview of the material to be presented.

As far as this research is concerned, the purpose of an image transmission system is to convey an image to a human observer. Consequently, the specifications of such a system should reflect the capabilities and limitations of the human visual system, abbreviated HVS. An ideal system should be able to reproduce the finest spatial and temporal features of images which the eye is capable of resolving and utilizing. On the other hand, an efficient system also takes advantage of the limitations of the HVS by not transmitting information which cannot be received. Accordingly, Chapter 2 reviews the relevant aspects of visual perception.

The canonical image transmission system adopted for this study consists of a camera with temporal filter and sampler, and a display with temporal interpolator. The function of the interpolator is to reconstruct a subjectively smooth time-varying image from discrete samples, which would show a sampling structure if viewed directly. Due to artifacts of interlaced display such as line crawl and interline flicker, only progressive scanning will be considered.

A mathematical model of the above system and the HVS is valuable for two reasons. One is that it provides a language for describing both physical and subjective characteristics of image defects. Another is that it assists in predicting the displayed image and its appearance, when give the original image and the parameters of the system and observer. Chapter 3 defines the model used throughout the thesis. In addition, in order to concentrate most of the mathematics into one chapter, it will also derive formulas for the reflected spatial filters referenced in the subsequent chapters.

Although many aspects of perception can be quantified, it is entirely too complicated to permit definitive predictions of image defects and image quality by modelling and analysis alone. First-hand observations of moving images are mandatory. A special-purpose display called the High Rate Television System — HRTV, as opposed to HDTV — was designed and assembled, largely from existing hardware and software subsystems, as part of this research project. The display rate is 120 fps, thought to be high enough to provide a "transparent display" [SCHR 83][LEE 83], that is, a display which is temporally continuous as far as the eye can tell, so that any image defects seen can be attributed to the signal processing rather than the display itself. Chapter 4 briefly discusses the features of the HRTV, which prescribe what kinds of moving images can be displayed and thus what subjective observations can be made.

Equipped with the HRTV, the first objective of this study was approached. It was verified that motion rendition is a multi-dimensional quality which can be degraded by multiple images, small-area flicker, and jerkiness. A classification of these and other defects based on subjective image features is presented in Chapter 5. Each type of defect is considered individually with a description of what it looks like, a specification of associated signal characteristics, and an explanation based on established principles of visual perception.

Once the classification of defects was set up, the second question of this research project could be studied. The amount of every defect present in an image as observed depends on many factors, including image characteristics, system parameters, and visual properties. For this reason, the major portion of the experimental work in this project consisted of viewing the large number of images which result from different combinations of the controlling factors. Chapter 6 examines the dependence of defects upon image velocity and basic temporal parameters of the transmission system. For each defect, the predicted behavior under joint parameter variation is compared with the results of observations.

The presentation of experimental results is continued in Chapter 7, which examines other image content factors. Though limited in number, these observations confirm that image characteristics, which are more or less beyond the control of the transmission system designer, play a significant role in the visibility of motion-induced defects.

Admittedly, the approach of this study is not without shortcomings. Besides questioning the approach of the experimental work, a critical review would be concerned with the generality of results based on specific test patterns. In the concluding chapter these issues accompany some extrapolation from the observations to high quality image transmission systems of the future.

Chapter 2

Some Aspects of Visual Perception

This chapter is a review of some properties of vision which are involved in the perception of time-varying images. The first section will deal with nonmoving spatiotemporal visual properties, because these fundamentals largely continue to apply to the perception of patterns in motion. The following section explains the role of eye movements in perception. Finally, the last section considers the perception of motion itself, as distinguished from the perception of objects in motion. The goal is to provide a framework of terminology, facts, and theories in which the subjective aspects of motion-induced image degradations may be investigated.

2.1 Spatiotemporal Properties of Vision

2.1.1 Spatial Vision

It is commonly said that the average human eye can resolve spatial detail on the order of one minute of arc, as measured by the ability to identify optotypes such as Snellen letters or Landolt C rings of calibrated sizes [LEGR 67]. However, this acuity figure is but a partial description of the response of the HVS to spatial variations in brightness.

A fuller description is the threshold contrast sensitivity to sinusoidal gratings. Among all the sensitivity data which can be found in the literature, no two curves

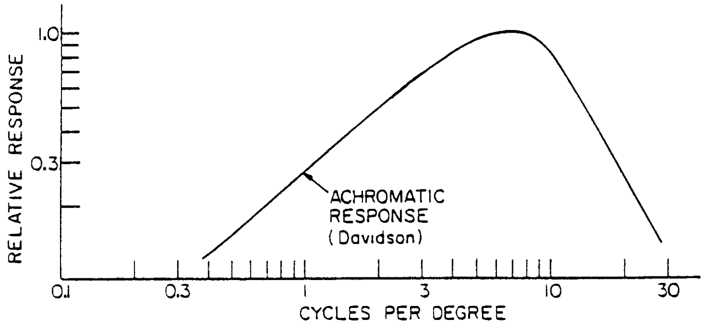


Figure 2.1: Threshold contrast sensitivity to spatial sinusoidal gratings (from [PRAT 78])

are identical, but they all show a maximum sensitivity around 6 cycles/degree (cpd) for foveal cone vision (Figure 2.1). 30-50 cpd is the figure usually quoted for the effective limit of perception. The hypothesized physiological reason for the midfrequency elevation is that a retinal neuron receives lateral inhibitory signals from surrounding receptors, in resemblance to the unsharp masking technique of image processing.

It is convenient to interpret these curves as linear modulation transfer functions (MTF) since sensitivity to square wave gratings [ISON 79] and visual effects like Mach bands [PRAT 78] can then be explained. However, at the same time it should be remembered that the HVS is anything but linear and that the relative sensitivity to different frequencies at threshold is not necessarily maintained for larger amplitude modulations.

2.1.2 Temporal Vision

One of the obvious temporal parameters of vision is the critical flicker frequency (CFF), to which the repetition rate of a light must be raised in order for visible flicker to cease. Classical experiments have extensively studied many stimulus factors which affect the CFF, but for this thesis the most relevant are mean intensity and area. All other things being equal, the CFF increases with mean intensity I according to the

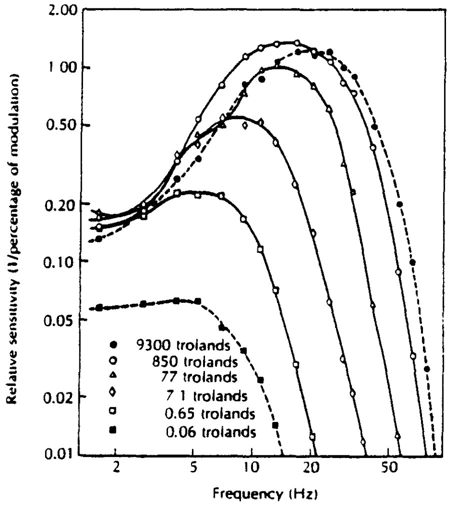


Figure 2.2: Threshold contrast sensitivity to temporal sinusoidal gratings (from [CORN 70])

Ferry-Porter law [FERR 92]:

$$\mathrm{CFF} = a \log I + b.$$

As a result, the brightest portions of a displayed image are most susceptible to flicker. Engstrom was one of the first to systematically investigate this relationship specifically for determining the display rate for television [ENGS 35], and it appears that elimination of flicker became the prime concern when the National Television Systems Committee (NTSC) designated a 60 Hz interlaced field rate in 1940.

A second property of flicker perception is that the CFF also has a logarithmic dependence on the area, A, of the region in question. This is given by the Granit-Harper law [BROW 65A]:

$$\mathrm{CFF} = c \log A + d.$$

Like spatial acuity, the CFF parameter is a limited description of the temporal response of the HVS. The threshold contrast sensitivity to a temporally modulated, but spatially uniform, field of illumination contains a midfrequency peak between approximately 10 and 20 Hz, increasing with mean intensity (Figure 2.2). In the language of linear systems the positive slope part of the curve is the characteristic of a differentiator; the negative slope part, an integrator. Finally, the physiological model which can explain this functional relation is recurrent inhibition, i.e., time-delayed negative feedback.

Perceived brightness is not linearly related to intensity, yet the brightness of a stimulus flickering faster than the CFF matches the brightness of a steady stimulus with equal average intensity. The temporal averaging takes place in the linear intensity domain, according to the Talbot-Plateau law [BROW 67]. Moreover, this law has been verified over a wide range of frequencies (up to 1500 Hz) and duty cycles (down to an 8 ns pulse) [LEGR 57].

In investigations with transient and step-function stimulation, the low-pass nature of the eye is called visual persistence [BROW 65A][COLT 80]. These studies generally find a subjective brightness decay lasting several hundred ms after a light source turns off. (The potentially longer-lasting phenomena of afterimages is an entirely separate issue [BROW 65B].) Though there might seem to be some connection between persistence and flicker fusion, there have not been any attempts to do the "obvious" thing, which is to relate the inverse Fourier transform of the threshold sensitivity curve to the subjective step response. This is because the curve is not really a MTF, and even if it were, large amplitude behavior is not expected to be described well by a linear model.

2.1.3 Spatiotemporal Vision

A natural generalization of spatial and temporal contrast sensitivity is the joint spatiotemporal contrast sensitivity, for which the stimulus intensity is

$$s(t,x) = a(1 + b\cos 2\pi f_t t \cos 2\pi f_x x).$$

This is a flickering grating with no motion. A perspective plot of the sensitivity surface found by Kelly clearly shows an interaction between the two frequency parameters where there is a fall-off in sensitivity for low f_t and low f_x (Figure 2.3). Vertical slices through the preceding solid show that the interaction is minimal at higher frequencies, where the sensitivity function becomes nearly separable (Figure 2.4).

Horizontal slices result in isosensitivity contours, and one of them can be selected to define a somewhat arbitrary spatiotemporal acuity limit. For example, the contour for

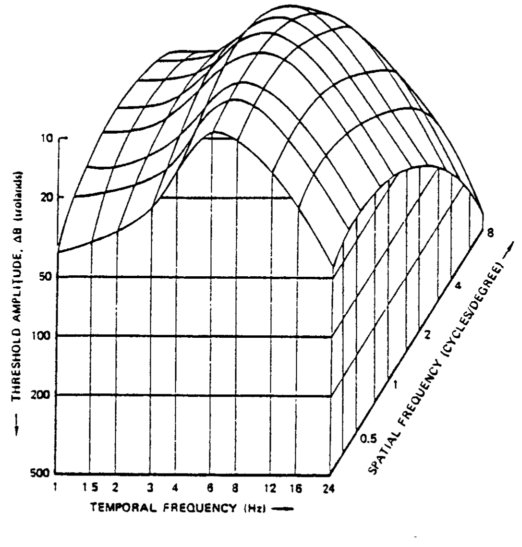


Figure 2.3: Perspective view of threshold contrast sensitivity to spatiotemporal sinusoidal gratings (from [KELL 72])

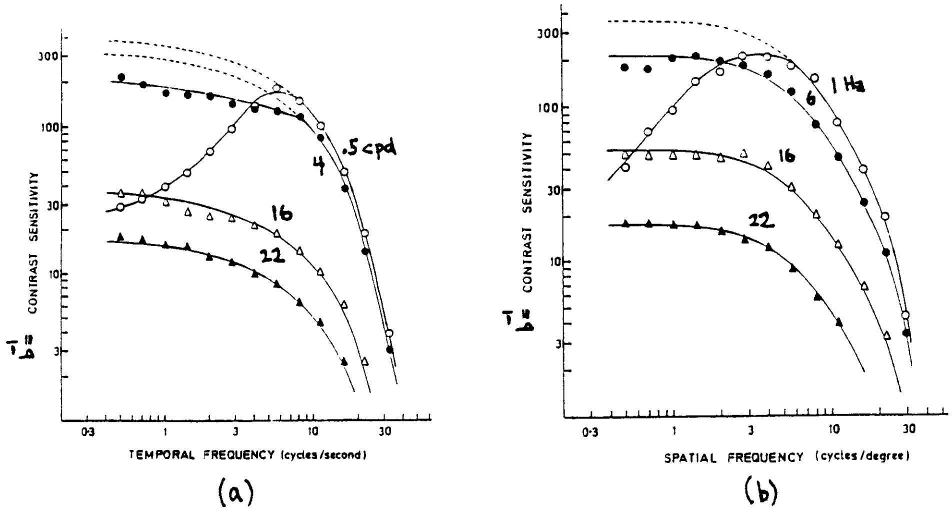


Figure 2.4: Cross-sectional view of threshold contrast sensitivity to spatiotemporal sinusoidal gratings: (a) spatial slices; (b) temporal slices (from [ROBS 66])

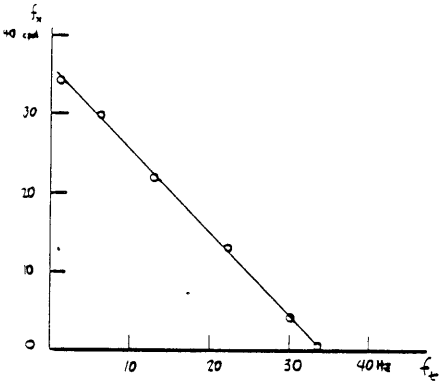


Figure 2.5: Spatiotemporal limit of resolution, taking $b^{-1} = 3$ contour in Robson's data

$b^{-1} = 3$, which can be deduced from Robson's graphs and is sketched in Figure 2.5, is curiously close to a straight line. High spatial frequencies can be detected only at low temporal frequencies, and similarly, high temporal frequencies can be detected only at low spatial frequencies.

Reflected into all four quadrants, this contour surrounds a diamond-shaped region of (f_t, f_x) space, a fact which points to possible image transmission bandwidth savings. Several investigators have proposed systems which exchange spatial and temporal bandwidth [TONG 83][LIMB 71], perhaps adaptively. Nevertheless, one should be wary of interpreting the spatiotemporal threshold function as a MTF when it comes to moving images. At first glance it seems that the minimum contrast for the moving, but apparently nonflickering, grating

$$s(t, x) = a(1 + b\cos 2\pi(f_t t - f_x x))$$

should be predictable from the previous measurements of contrast sensitivity for a grating of frequency (f_t, f_x). However, this reasoning is based on an assumption that moving gratings and flickering gratings receive identical processing in the HVS. Because of the manifestly different appearance of the two kinds of gratings and also because of some findings on dynamic acuity (Section 2.2.2), the general validity of that assumption is questionable. In fact, Chapter 3 will show why a system based on the spatiotemporal bandwidth tradeoff fails with moving images.

2.1.4 Masking

Masking refers to a reduction in the ability to detect a test stimulus due to the interfering effect of other stimuli in spatial or temporal proximity. Specifically, in forward temporal masking the test stimulus is presented after the masking stimulus.

In a series of experiments by Glenn, the test stimuli were gratings of spatial frequencies from 0.6 cpd to 12 cpd and the immediately preceding mask was a brief segment of a broadcast video signal. By measurement of the threshold contrast sensitivity to the gratings as a function of duration of stimulus presentation, Glenn was able to deduce that forward masking has a time constant of about 200 ms for all of the gratings [GLEN 83].

Another study, by Seyler, dealt with masking by the previous scene after a scene change. Immediately after a scene change, he cut the horizontal resolution to a fraction of normal and then restored the resolution exponentially. The longest tolerable restoration time constant was determined as a function of the resolution reduction fraction. A remarkable result is that a 780 ms recovery time causes a just-noticeable degradation when the fraction is 1/20 .

Whether temporal masking is merely a consequence of sensitivity to spatiotemporal frequencies, or an independent visual phenomenon, it is a limitation of the HVS which can be exploited by an image transmission system. In a region of an image which is newly revealed, as from change in object configuration or from a scene change, detailed spatial information cannot be utilized immediately by the observer. Therefore, such information, which lies at high spatial and temporal frequencies, might not be carried by the transmission system in order to reduce the channel bandwidth [GLEN 83].

2.2 Eye Movements

Much of the literature on moving-image transmission does not consider the observer's eye movements. In order to judge whether this oversight is justifiable, the properties of movements and of perception during movements must be examined. It will be argued that one kind of movement, ocular pursuit, cannot be neglected in the perception of moving images.

2.2.1 Types of Movements

Types of eye movements include tremor, drift, saccades, and pursuit. During fixation the eye is never perfectly stationary since it is subject to a random noise component of motion, called tremor, whose RMS amplitude is around 0.05°/sec [DITC 73]. It is known that tremor is essential for normal vision, for when an image is artificially projected into the eye so that it remains absolutely stationary on the retina despite ocular motion, the perception of the image fades soon after its onset [CORN 70]. On the other hand, because one is never consciously aware of ocular tremor and because it does not give rise to the perception of movement, this type of eye movement will be disregarded. Henceforth, the common-sense, but inaccurate, notion of a "stationary" retinal image will be adopted.

Drift is a slow, steady involuntary movement with a typical velocity of 0.1°/sec [YARB 67]. When drift has moved the gaze sufficiently far away from the desired point of fixation, a saccade quickly repositions the eye. Larger saccades are also used voluntarily to seek new fixation points, in which case the movement can have a peak velocity up to 800°/sec [LEGR 67].

Pursuit is a steady eye movement which tracks a moving target. Although pursuit can be started and stopped voluntarily, the observer has little conscious control over the actual velocity of movement. In particular, it is not possible to execute a smooth pursuit movement when the image contains no target to track; any attempt to do so invariably results in a series of saccades [YARB 67].

The initiation and control of pursuit involves two systems. The velocity system attempts to correct only velocity errors, regardless of the error in position. The saccadic system corrects the large initial position error and any subsequent displacements. Except for very predictable target motions such as sinusoidal oscillation, velocity changes always occur in discrete steps. Similarly, position corrections are nearly discrete, by virtue of the high speed of saccades [ALPE 62].

The time course of a pursuit movement begins with the observer's decision to track and/or with the onset of a moving target. Beginning from fixation, there is first a 150 – 250 ms latency period in which the eye remains at rest, a period which is shorter for faster targets. A saccade then brings the fovea onto the moving target, whereupon the eye immediately tracks at the target velocity. A gradual acceleration need not take place when the velocity is below 25 – 30°/sec. There is an additional 200 – 250 ms delay before perception becomes satisfactory, an effect comparable to forward temporal masking [ALPE 62].

The range of velocities over which smooth pursuit can be maintained is bounded below by about 0.2°/sec, the velocity at which pursuit first becomes dominant over drift [YARB 67]. The highest velocity at which some useful perception can take place is 100 – 200°/sec [YARB 67], but 30°/sec is about the limit for comfortable viewing [YUYA 82] and nearly undiminished acuity [LEGR 67].

2.2.2 Dynamic Acuity

The preceding point concerning acuity during pursuit, also known as dynamic visual acuity, demands further elaboration because of its possible implications for image transmission. To begin with, it might be asked why dynamic acuity should suffer when, after all, it appears to the observer that he is tracking the target successfully.

First, it has been proposed that the eye lacks the ability to turn fast enough during pursuit, but this is not likely since some studies have found pursuit velocities up to 500 – 600°/sec.

Another possibility is that positional errors during pursuit cause the target to fall outside of the fovea, where even static acuity is poor. However, Ludvigh found this to be too weak an effect to explain his experimental results which are detailed below.

A third explanation might be the activation of a central inhibitory process which reduces visual function during voluntary eye movements [RIGG 65]. For example, some studies maintain that saccadic suppression is responsible for preventing the conscious perception of widespread blur during saccades. However, a counterexample to this theory is that a stimulus briefly flashed in the middle of a saccade can be perceived in full resolution. Moreover, an analogous relationship with acuity during pursuit has not been established.

Residual retinal image motion due to velocity errors in pursuit may be the most obvious and understandable cause of lowered acuity [LUDV 58]. The limited temporal bandwidth of the HVS implies that such motion will cause blur. This fourth and final hypothesis might be tested by examining visual acuity for moving objects during *fixation*. Employing moving targets consisting of a pair of bright parallel lines separated by various values of D, Van den Brink demonstrated that it was not possible to resolve the lines when v/D was above 50-120 ms, a value which can be interpreted as a temporal time constant of vision [VAND 58]. From this it can be inferred that motion blur on the retina does cause acuity loss.

On the contrary, Westheimer found that resolution thresholds for randomly-oriented Landolt C and vernier targets did not change for linear translation up to 2.5°/sec. More interestingly, the gap size of the smallest resolvable Landolt C at that upper speed was only 2% of the distance moved during the 200 ms exposure time [WEST 75]. This experiment implies that, at least for small velocity errors, dynamic acuity is not diminished by retinal image motion.

The varied and sometimes contradictory explanations above show that it is better not to take dynamic acuity loss for granted. They also indicate that only direct measurements will resolve the problem. In order to interpret the following results in the context of image transmission systems, recall that the maximum comfortable viewing velocity was said to be 30°/sec. Another fact is that the typical speed of moving objects in broadcast television is $0.29 - 0.4W$/sec, where W is the screen width [MIYA 75]. Assuming 4:3 aspect ratio and a viewing distance of $4H$, it follows that the typical angular velocity is $5.5 - 7.6$°/sec.

An experiment to measure dynamic acuity with a Snellen letters identification task found acuity to drop linearly with velocity (Figure 2.6b). Although the deterioration is substantial over the range of velocities observed, it is minor up through 30°/sec, at which there is only a 15% drop [LUDV 47]. Similarly, another measurement using

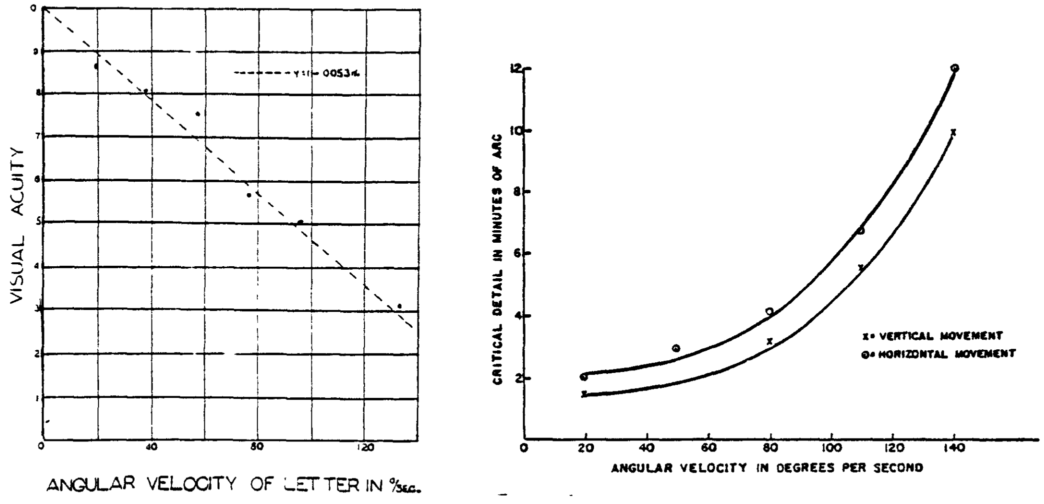


Figure 2.6: Dynamic acuity with (a) Snellen letters (from [LUDV 47]); (b) Landolt C test objects (from [LUDV 58])

Landolt C optotypes found insignificant acuity loss below 30°/sec (Figure 2.6b), even though the performance degraded as velocity to the third power [LUDV 58]. On the basis of these results, it appears that for the velocities of motion in television, dynamic acuity should be about the same as static acuity.

2.2.3 Conclusions

Measurements of dynamic acuity are performed under more or less steady-state conditions of ocular pursuit, well after the latencies of movement and masking have expired. Unless such conditions are attained when an observer looks at moving images in television, the measurements do not apply, and therefore, detailed spatial information on moving images cannot be utilized by the observer. It is most likely that "head and hand movements cannot be easily tracked" [DUBO 81] because they are short and unpredictable. Experimenting with images depicting mechanical toys, Miyahara found that the bandwidth of moving areas can be cut from 5 to 1.2 Mhz and still result in a picture which is rated 4, "perceptible but not annoying," on the five-grade impairment scale [MIYA 75].

However, these are examples of moving objects whose relatively small size permits rapid acceleration. Larger objects with greater inertia, and camera panning or zooming, can generate slower and/or more uniform motion. Furthermore, with the coming trend toward wide viewing angles in television displays [YUYA 82], movements not only can be faster but also can last longer before leaving the screen. Thus, the key premise of this thesis is that ***there will be moving image regions which can be successfully tracked by the observer and which should be reproduced at normal spatial resolution.*** In conclusion, ocular pursuit may not be safely neglected when evaluating the quality of a reproduced image.

2.3 Perception of Motion

A time-varying image can be represented mathematically by the spatial ensemble of functions $f_{xy}(t)$, each of which gives the intensity at one point (x, y) over time. However, this is not generally the way an image is represented at the conscious level of visual cognition. Rather, what is "seen" is the evolution of spatial forms, for this is closer to one's conceptual model of reality. For instance, the moving grating of Section 2.1.3 sets up periodic intensity modulations at any given point on the retina, but movement rather than flicker is perceived. Although all motion information is contained in the ensemble of functions, motion is psychologically different from mere spatiotemporal patterns. Accordingly, this section is concerned with the perception of motion itself, as distinguished from the spatiotemporal pattern of an object in motion, which was covered in the preceding section on dynamic acuity.

2.3.1 Processes of Motion Perception

When the configuration of a scene is different from one instant of time to another, not counting whole-field translation, it is possible to infer the presence of motion. Objects may occlude and uncover one another, or the distance between objects may change. However, the process of detecting motion by these visual clues is not considered to be motion perception per se.

There are three primary processes by which the HVS comes to perceive motion directly. The first process determines motion from eye movements, the second senses retinal images shifts, and the third analyzes contour displacements. These processes are logically independent from pathways for spatial detail perception and, except for the third process, independent from spatial form perception. For many stimuli more than one process can be active at once, and it may not be possible for the observer to isolate the contributions of each process; however, the existence of these processes can be demonstrated convincingly by suitable special cases. Historically, these processes were hypothesized in order to explain apparent movement, which is the perception of directed motion when the physical displacement of the stimulus is discrete. (As a matter of definition, apparent motion does not necessarily have to appear *smooth*). But there is reason to believe that these processes work the same way with either apparent or real, i.e., temporally continuous, motion [FRIS 72].

The pursuit eye movement is already familiar, so the perception of motion which arises from it will be explained first. Suppose a moving target is surrounded by a large featureless background; for instance, a small light bulb in a dark room. Since there are no changes in the image projected on the retina when the target is tracked successfully, the only explanation for the unmistakable perception of motion is that the eye movement itself is perceived. The autokinetic motion illusion can be attributed to this process [GREG 66]. Finally, motion perception from eye movement probably continues to work even in the presence of background features which can activate other processes.

The foregoing applies when the target is tracked but not when the image moves relative to the retina, as during fixation. It has been hypothesized that there exists a system of localized retinal receptors which are stimulated by image patterns shifting across them. An explicit model of a motion detector consists of two light sensors, connected to a comparator whose output is maximized when the two sensors' signals are offset by a fixed time delay in the proper order (Figure 2.7). The comparator performs some kind of subtraction or correlation to selectively detect patterns moving with the

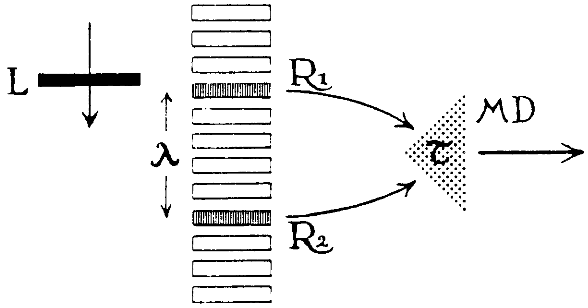


Figure 2.7: Paired-sensor model of movement detector with two sensors R_1 and R_2, light source L, and comparator MD (from [SCHO 67])

right speed and direction past the sensor pair. This model is but one implementation of a spatiotemporal filter which is tuned to the characteristic spectrum of a translating object [WATS 83B], which will be derived in the next chapter.

While there is some physiological evidence for the previous retinal system of receptors in some animals [LETT 59], its existence in the HVS is supported only by subjective effects. The phenomena of velocity adaptation and motion after-effects are strong evidence that motion detectors, of one kind or another, exist at or near the lowest levels of visual processing [SEKU 75][BECK 72]. Because the after-effects are diminished by large discontinuities during apparent motion, the range of an individual receptor is thought to be no more than about 0.25° [BANK 72]. Schouten predicted that a sensor pair with finite spacing would be susceptible to retrograde motion illusions, since any spatiotemporal pattern which illuminates the two sensors with the proper delay would activate the comparator. Indeed, he found that as a spoked disk observed under DC illumination rotated faster and faster, the apparent velocity would ultimately stop and reverse itself [SCHO 67].

The third process of motion perception also detects retinal image shifts, albeit indirectly through the displacement of contours and other spatial primitives which have already been extracted from the image by another part of the HVS. Such a process would have to reside at a higher cognitive level than the previous one, and consequently it lacks a quantitative model. In any event, only such a global process can account for certain apparent motion phenomena which cannot be explained by short-range retinal motion receptors [ANST 78][ANST 80][BRAD 78][ULLM 77].

2.3.2 More on Apparent Movement

Because apparent motion is the central feature when observing temporally sampled images, some further discussion is warranted on this subject. For image transmission, only apparent motion in a long sequence of images is of concern; however, the discussion here begins with apparent motion between a pair of images because this has received more attention in the psychological literature.

In 1912 Wertheimer described two-image apparent motion effects which he named "optimal movement" and "phi movement." Supposedly, the appearance of optimal movement is equivalent to that of a real target in motion, while phi movement is an ill-defined pure motion sensation disembodied from the target [GRAH 65]. The requisite inter-stimulus interval (time between offset of the first presentation and onset of the second) for these phenomena to be seen depends on many factors, one of which is spatial configuration, but typical values are 60 – 200 ms. In addition, when the second stimulus is much brighter than the first, it is possible to produce delta movement, in which the perceived motion is directed from the second position to the first. Delta is entirely a perceptual effect and has nothing to do with retrograde motion caused by temporal undersampling, although these effects have on occasion been mistakenly linked [CHAT 54].

The phi phenomenon has often been cited to explain the smoothness of apparent motion which may be observed when fixating (as opposed to tracking) sequences of images longer than two time samples. But, as Le Grand has remarked,

> it is possible that for a greater number of successive images, the impression of saccadic or steady movement follows laws other than those applying to two images only [LEGR 67].

After all, when there are only two images, causality dictates that apparent motion cannot begin until after the second stimulus is shown, and nonuniformity of velocity probably cannot be detected in such a short interval. On the other hand, when there is a longer temporal context, apparent motion could stay in phase with the physical sequence of stimuli, and any nonuniformities might be more detectable. Thus, the emphasis is decidedly different in studies of two-image and of many-image apparent motion: in the former, the issue is the presence or absence of a motion percept, while in the latter the impression of motion is unquestionable and the issue is smoothness. These considerations suggest that the permissible spatiotemporal interval between successive target presentations for smooth apparent motion in a long sequence is not the same as for phi movement in an isolated pair of images.

Smoothness is difficult to measure psychophysically, so the approach taken by Morgan in his study of many-image apparent motion was to determine, by vernier alignment, the perceived location of a discrete stimulus at instants within the inter-stimulus interval. A motion continuity index was then computed to quantify how close the perceived location was to the theoretically interpolated location, and it was found that frequency components above 25 Hz in the displacement vs. time signal had little effect on the continuity (Figure 2.8). This finding implies that the applicable motion perception process has a temporal bandwidth of about 25 Hz [MORG 80A][MORG 79][MORG 80B].

2.4 Summary

In the introductory chapter, two differing theories of motion perception were presented, and a reconciliation was promised. In fact, there is no incompatibility between the theories because the HVS consists of independent processes of motion perception and spatiotemporal detail perception. It should be possible to perceive motion, even smooth motion, when a moving target occupies visibly discrete spatial locations. Conversely, in some cases it might be possible to detect nonuniform motion when discrete spatial positions are not discernible. Without any further information, one concludes that a reproduced image must be sufficiently smooth and continuous in both the displacement domain and spatiotemporal brightness domain to provide good rendition of moving objects.

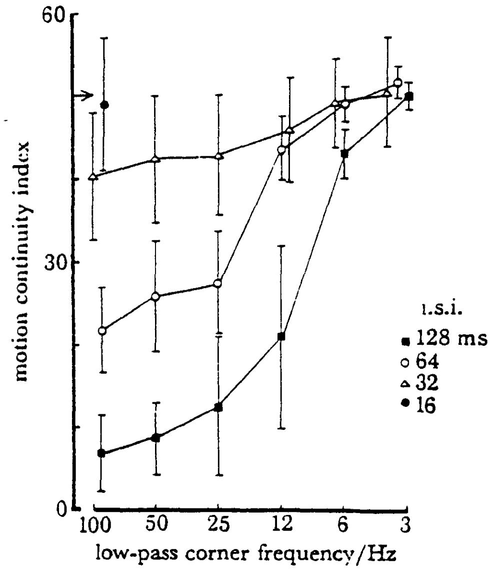


Figure 2.8: Motion continuity index (ideal = 50) as a function of the cutoff frequency of the electrical filter applied to the staircase displacement signal before it was used to deflect a CRT beam (i.s.i. = frame period); filters wider than 25 Hz have minor effect (from [MORG 80A])

This discourse on some aspects of visual perception has ranged from the psychophysical to the cognitive, from the easily quantifiable to the largely qualitative. Spatial and temporal contrast sensitivities are the foundations for resolution requirements of image transmission systems. Systems can also take advantage of bandwidth tradeoffs and masking to the extent that apparent resolution is unimpaired on moving objects trackable by ocular pursuit. Processes of motion perception are responsible for creating the impression of smooth motion from discrete stimuli. Given that the human visual system contains numerous pathways, each optimized to mediate a special class of visual information, it is remarkable that these pathways ultimately join at the conscious level of visual processing to form one unified perception of the observer's evolving environment.

Chapter 3

Analysis of Moving Images

This chapter will present a mathematical model of a basic image transmission system and the human visual system, which together transform the spatiotemporal distribution of light on a moving object in front of a camera into an image perceived by the observer. The linear model incorporates image and ocular motion, temporal characteristics of the transmission system, and spatiotemporal characteristics of vision. In conjunction with this model, the idea of reflecting a temporal filter into an equivalent spatial filter will be discussed at length. The signal-domain and frequency-domain notation presented here for the analysis of moving images will be used in subsequent chapters when predicting and interpreting the appearance of moving images in the observation experiments.

3.1 Notation

In the following treatment, signals will be represented by real functions of continuous time and space (t, x), admitting Dirac δ-functions for time-sampled or impulsive signals. The signals to be considered are constant in the y (spatial) direction and all motion is in the x direction, so the y direction is omitted.

First, the translation operator $\mathcal{T}_v$ will be defined in order to facilitate manipulations with signals representing moving images. $\mathcal{T}_v$ maps a function of time and space to another such function, and is defined as

$$\mathcal{T}_v\{u(t,x)\} \equiv u(t, x - vt).$$

If $w = \mathcal{T}_v\{u\}$, then an object moving with velocity v_u in image u will move at velocity $v_u + v$ in image w.

Some of the basic properties of $\mathcal{T}_v$, for arbitrary signals s and r, are

$$\begin{aligned} \mathcal{T}_v\{s+r\} &= \mathcal{T}_v\{s\} + \mathcal{T}_v\{r\} && \text{additive property} \\ \mathcal{T}_v\{sr\} &= \mathcal{T}_v\{s\}\mathcal{T}_v\{r\} && \text{multiplicative property} \\ \mathcal{T}_{v_1}\{\mathcal{T}_{v_2}\{s\}\} &= \mathcal{T}_{v_1+v_2}\{s\} && \text{composition property.} \end{aligned}$$

Furthermore, if $*$ denotes two-dimensional linear convolution of two signals, then

$$\mathcal{T}_v\{s * r\} \quad = \quad \mathcal{T}_v\{s\} * \mathcal{T}_v\{r\} \quad \text{convolution property.}$$

Finally, if $W(f_t, f_x)$ and $U(f_t, f_x)$ are the two-dimensional transforms of $w(t,x)$ and $u(t,x)$, then

$$w = \mathcal{T}_v\{u\} \longleftrightarrow W(f_t, f_x) = U(f_t + vf_x, f_x) \quad \text{transform property,}$$

where the Fourier transform F of a signal f is defined as

$$F(f_t, f_x) \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t,x)e^{-j2\pi(f_t t + f_x x)}dtdx.$$

Note. To be mathematically rigorous, the Fourier transform does not exist for infinite-energy signals such as δ-functions or periodic signals. An objection to Fourier transforms which is more serious than lack of rigor is that they do not adequately express the spatiotemporally varying local characteristics of an image. Furthermore, it can be argued that an observer looking at an image does not perceive simultaneously the entire signal from $-\infty$ to ∞ in either the spatial or temporal directions. In such a situation the appropriate frequency domain procedure is the short-term Fourier transform [RABI 78]. For instance, a certain flicker defect described in Chapter 5 is not obvious from an ordinary Fourier spectrum. However, in most of this thesis, as in some of the related literature [WATS 83A], ordinary Fourier transforms are used for simplicity; they often explain perceptual phenomena as well as short-term transforms do.

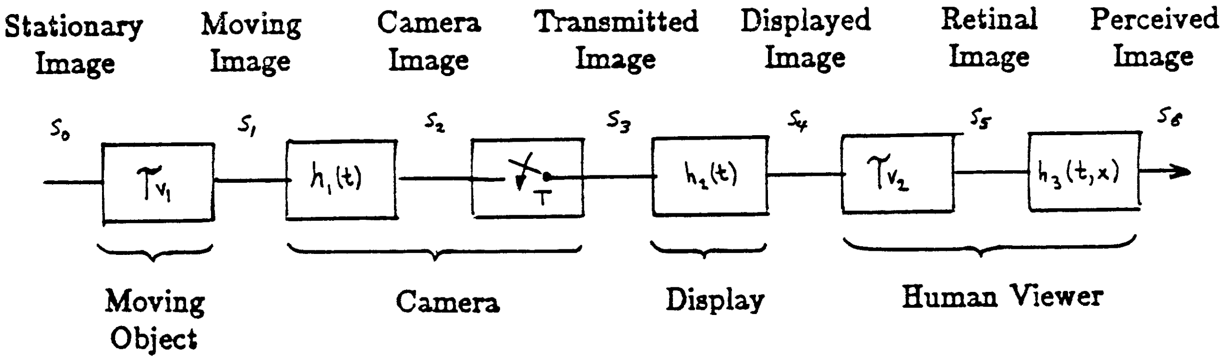


Figure 3.1: Block diagram of transmission system model

3.2 Canonical Image Transmission System

The linear model of the transmission system and human visual system consists of six stages, and is general enough to be applicable to either television or motion pictures. The block diagram of Figure 3.1 shows the model, which passes a stationary image through the translation operator imposing a constant velocity motion, the temporal presampling filter and sampler of the camera, the temporal interpolating filter of the display, and the translation operator and spatiotemporal filter of the HVS. The signals denoted s_0 through s_6 and their descriptive names, as indicated in the figure, will be used throughout this thesis. The components of the model will now be discussed individually in greater detail.

3.2.1 Moving Object

The input to the model is the stationary image $s_0(t,x) = s_{0x}(x)$, which is constant in t. The region of support of the Fourier transform of this image must be contained on the f_x axis, so $S_0(f_t,f_x) = S_{0x}(f_x)\delta(f_t)$,where $S_{0x}(f_x)$ is the transform of the 1-D signal $s_{0x}(x)$. The temporal bandwidth of a stationary image is zero. To illustrate the chain of signal transformations performed by this model, the example which will be continued through the rest of this section begins with the stationary image consisting of only a single narrow, vertical bar. The signal $s_0(t,x)$ and its transform $S_0(f_t,f_x)$

are shown in Figure 3.2, where points and lines represent impulses which project out from the page. All figures in this document are negatives: a dark point on the page represents a bright point in an actual image.

The translation operator $\mathcal{T}_{v_1}$ transforms s_0 into a moving image $s_1(t,x) = s_0(t, x - v_1 t) = s_{0x}(x - v_1 t)$. s_1 is the image of a narrow bar moving horizontally with velocity v_1. s_1 is an example of a purely translating image, which will be defined as any signal $u(t,x)$ satisfying

$$u(t,x) = u(0, x - vt) \quad \text{for all } t,x$$

for some constant v, the velocity of motion. An image consisting of an object moving across a constant flat background is purely translating, but if the background contains stationary image features it is not purely translating. However, as long as the object does not come too close to touching or occluding the stationary features as it moves, then the image would be considered to be purely translating in a spatiotemporal neighborhood of the object.

The region of support of the Fourier transform of the purely translating image s_1 must be contained within the line $f_t = -v_1 f_x$, since by the transform property,

$$S_1(f_t, f_x) = S_0(f_t + v_1 f_x, f_x) = S_{0x}(f_x)\delta(f_t + v_1 f_x).$$

The signal s_1 is shown in Figure 3.3, along with its transform.

3.2.2 Camera and Display

Since this research is concerned exclusively with temporal processing of images, it is assumed that the camera and display have ideal spatial characteristics — spatial filtering and sampling are of no concern.

The purpose of the presampling (antialiasing) filter $h_1(t)$ is to reduce the amplitude of temporal frequency components above the Nyquist frequency $1/2T$, where T is the period of the sampler. In television terminology, $1/T$ is the transmission frame rate. The output of the sampler, s_3, is a train of δ-functions in time, which are interpolated by $h_2(t)$ in the display, resulting in a signal s_4 which should be an approximate reproduction of the moving image s_1.

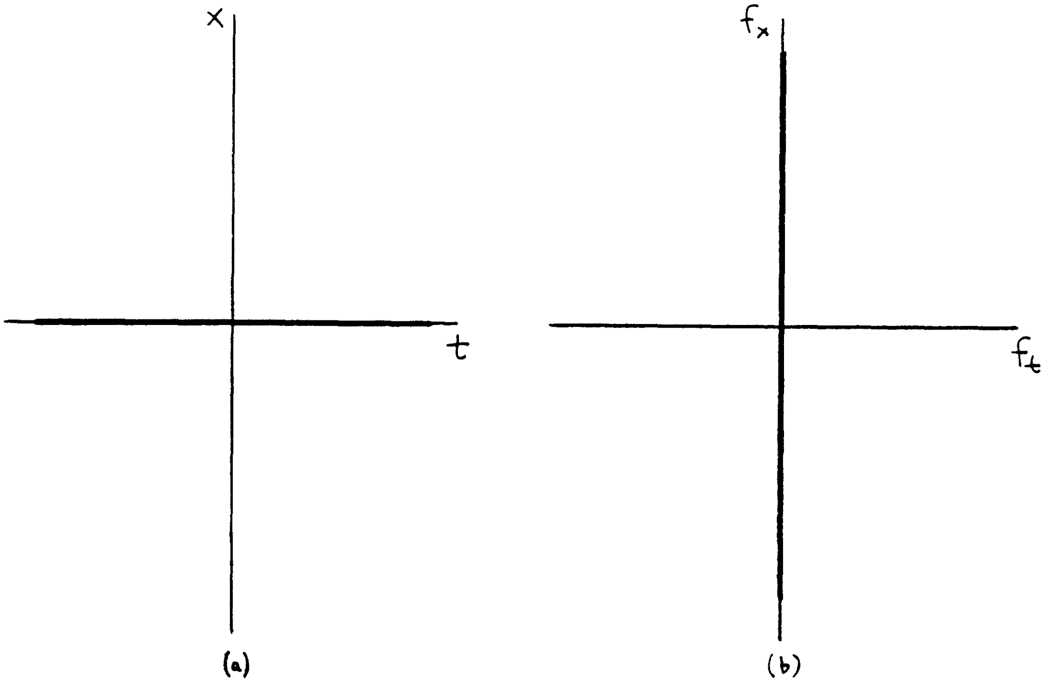


Figure 3.2: (a) Signal and (b) transform of a stationary vertical bar

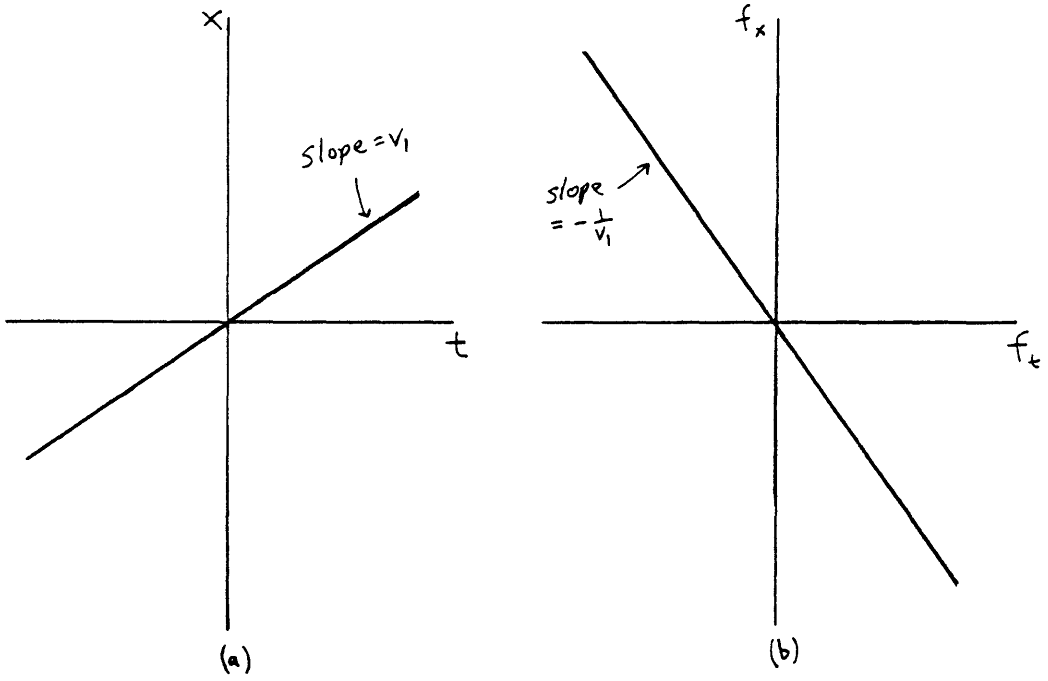


Figure 3.3: (a) Signal and (b) transform of a moving vertical bar

According to the sampling theorem, both the presampling filter and the interpolation filter should be ideal low-pass filters, with cutoff frequency $f_t = 1/2T$, to completely reject spurious harmonics. A sinc-type impulse response, however, is never used in common imaging systems, either as a spatial or a temporal filter, for it is not only impractical to implement but also, in the case of spatial filtering, it produces subjectively poor quality images [RATZ 80].

In a progressively scanned video camera tube, light is integrated on the photosensitive target for the entire frame period, and the target is almost completely discharged by the read beam. In a motion picture camera, the shutter is open during each frame for a period often much shorter than T, as dictated by mechanical requirements and for regulating the film exposure according to the brightness of the scene. Thus the presampling filter for television or movies is something like a rectangular pulse of duration T or less, and it is well known that such an impulse response cannot effectively remove aliasing components (at $1/2T$, the frequency response is down no more than 3.9 dB).

Practical display interpolation filters are hardly any closer to ideal, so that the additional filtering performed by the visual system will be necessary for apparent temporal smoothness. CRT phosphors used in standard black-and-white television work have an exponential decay time constant which is quite short relative to T, in which case $h_2(t)$ is made up of δ-functions. The frame rate might be up-sampled prior to display if frame storage were available at the receiver. For up-sampling by rate r, $h_2(t)$ would be a train of δ-functions spaced by T/r, not necessarily all equal in height or r in number. The impulse response corresponding to standard motion picture projection is a train of 2 or 3 rectangular pulses spanning T; again for mechanical reasons it is not possible to display a frame for 100% of the frame period, and the train of pulses is preferred over a single long pulse in order to raise the flicker frequency. (However, special-purpose non-intermittent projectors, typically employing rotating-prism optical systems, can achieve continuous temporal impulse responses of duration *longer* than T [WITT 55][ISHI 82].)

The input-output relationships of the camera and display stages of the model are

$$\begin{aligned} s_2(t,x) &= s_1(t,x) * h_1(t) & S_2(f_t,f_x) &= S_1(f_t,f_x)H_1(f_t) \\ s_3(t,x) &= T\textstyle\sum_{n=-\infty}^{\infty}\delta(t-nT)s_2(t,x) & S_3(f_t,f_x) &= \textstyle\sum_{k=-\infty}^{\infty}S_2(f_t-\frac{k}{T},f_x) \\ s_4(t,x) &= s_3(t,x) * h_2(t) & S_4(f_t,f_x) &= S_3(f_t,f_x)H_2(f_t). \end{aligned}$$

For proper normalization of s_4 with respect to s_1, H_1 and H_2 should have unity DC gain.

To continue the example of a moving narrow bar, let $h_1 = h_2 =$ ideal low-pass filters. The signals s_2, s_3, s_4, and their transforms are shown in Figure 3.4. All the signals have a cross-sectional waveform like $\sin x/x$, but only the region of the main lobe of this waveform is shown shaded. The sampling operation produces spectral replicas spaced by $\Delta f_t = 1/T$, and the output of the display interpolation filter, s_4, exactly reproduces the sampler input s_2. Of course, this is not a realistic example and it is given only to illustrate the camera and display transfer functions more clearly.

3.2.3 Human Viewer

To account for the possibility of eye movements, the displayed image s_4 is transformed by the translation operator $\mathcal{T}_{v_2}$. For observations while fixated on a stationary point on the display, v_2 is zero, while for observations with ocular pursuit, $v_2 = -v_1$ once accurate pursuit has been established. (There is a minus sign, for when the eye tracks a moving object, stationary points on the display move in the opposite direction with respect to retinal coordinates.) Except for pursuit, the model will not include properties of vision dealing explicitly with motion. The processes of motion perception are not easily incorporated into a model whose signals represent brightness vs. space and time. Accordingly, the final stage in the model is a spatiotemporal filter $h_3(t,x)$, whose characteristics were discussed in the previous chapter. The input-output relationship is

$$s_6(t,x) = s_5(t,x) * h_3(t,x) \qquad S_6(f_t,f_x) = S_5(f_t,f_x)H_3(f_t,f_x).$$

For the example of a moving narrow bar, suppose h_3 were the diamond-shaped window of visibility with maximum spatial frequency $f_{x0} > 1/2Tv_1$ and maximum

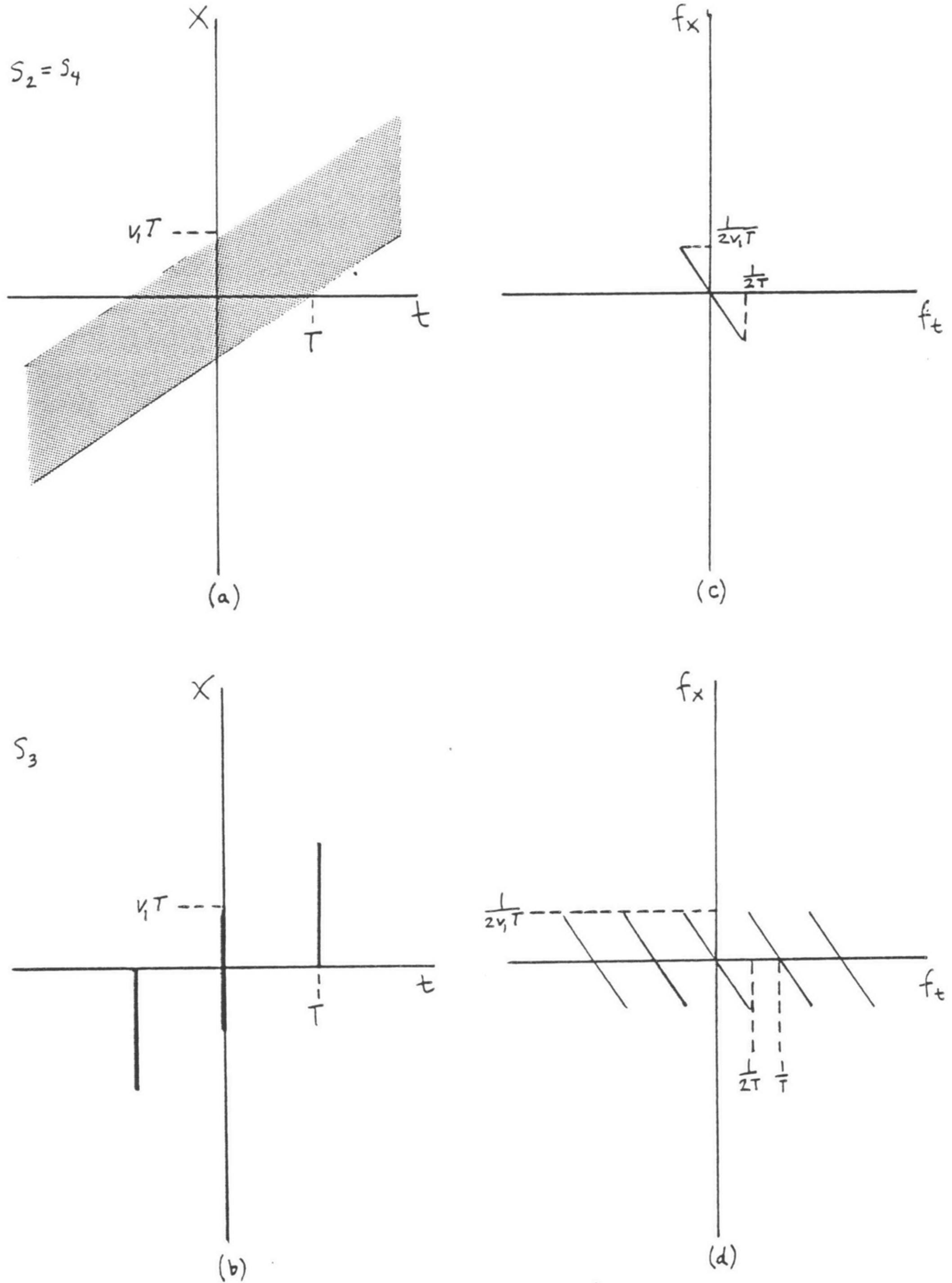


Figure 3.4: (a) Low-passed and (b) sampled moving vertical bar; (c,d) are the corresponding transforms

temporal frequency $f_{t0} < 1/2T$ (Figure 3.5a). With a fixated eye, the spectrum of the perceived image s_6 is that portion of $s_5 = s_4(t, x)$ which lies within the passband of the visual filter. The narrow temporal baseband of the eye has effectively removed spatial frequencies above $\left(\frac{v_1}{f_{t0}} + \frac{1}{f_{x0}}\right)^{-1}$.

On the other hand, if the same display were viewed with ocular pursuit, the translation operator $\mathcal{T}_{v_2} = \mathcal{T}_{-v_1}$ compensates for the motion of the object, leaving s_5 stationary with respect to the retina and also undoing the spectral tilt. Here, $s_5(f_t, f_x) = s_4(f_t + v_2 f_x, f_x)$. Although the visual filter h_3 remains the same as before, a different portion of the displayed image spectrum (in this case, none) is filtered out by h_3 (Figure 3.5b).

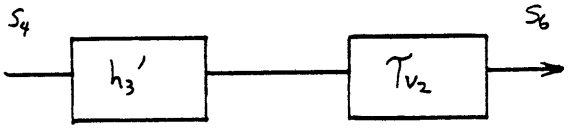
Since

$$h_3 * \mathcal{T}_{v_2}\{s_4\} = \mathcal{T}_{v_2}\{h_3' * s_4\}$$

where

$$h_3'(t, x) \equiv \mathcal{T}_{-v_2}\{h_3(t, x)\} = h_3(t, x + v_2 t),$$

an alternative way to model the human visual system response with eye movements is to transpose $\mathcal{T}_{v_2}$ and the filter to obtain the equivalent system:



By the transform property of the translation operator, the spectra of h_3 and h_3' are related by

$$H_3'(f_t, f_x) = H_3(f_t - v_2 f_x, f_x) = H_3(f_t + v_1 f_x, f_x),$$

so that if H_3 has the previous diamond-shaped passband, H_3' has a parallelogram-shaped passband, as shown in Figure 3.5c with the spectrum of the display image, S_4. The net effect of filtering the displayed image S_4 with H_3' before the translation operator is the same as filtering the retinal image S_5 with H_3.

The significance of this derivation is that H_3', the window of visibility in display coordinates, is parameterized by velocity. Thus, the union of the windows for $|v_2| <$

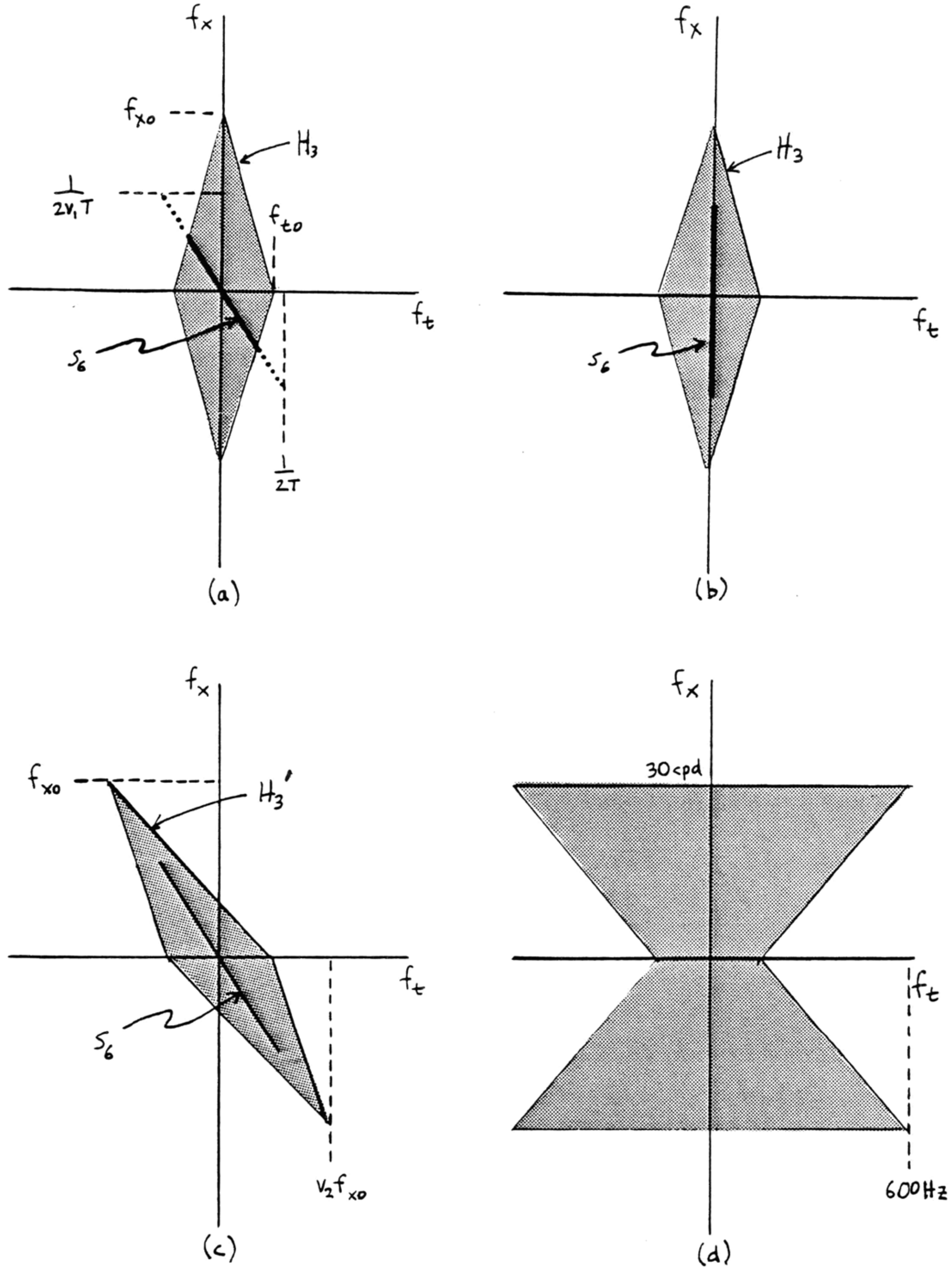


Figure 3.5: Window of visibility applied to moving images: (a) fixation and (b) pursuit in retinal coordinates; (c) pursuit in display coordinates; (d) union of windows for a range of pursuit velocities

$v_{\max}$, where $v_{\max}$ is some kind of bound on velocities of good ocular pursuit, is a very large window of spatiotemporal frequencies (Figure 3.5d), all of which need to be reproduced at the display to satisfy the HVS during arbitrary pursuit movements. If $f_{x0} = 30$ cpd and $v_{\max} = 20^\circ/\text{sec}$, then temporal frequencies go up to 600 Hz! In particular, images which have been band-limited to conform to the spatiotemporal limits of perception during fixation, as with image coding schemes which exchange spatial and temporal resolution, will not look satisfactory under pursuit.

This concludes the discussion of the chain of image processing steps which lead from a stationary object to the image as perceived by the observer. Before putting aside the formalism of the translation operator $\mathcal{T}_v$, however, it is appropriate to introduce and apply the notion of reflection.

3.3 Spatial Reflections of Temporal Filters

Because temporal patterns of illumination associated with moving objects are usually perceived as spatial patterns, the effects of temporal convolution usually appear to be the effects of a spatial convolution. In order to predict the appearance, the spatial reflection of a temporal (or spatiotemporal) impulse response will be derived. However, a complete discussion of the subjective appearance of these convolutions will be deferred to Chapter 5. In the model of transmission and visual systems described above, each of the three filters, h_1, h_2, and h_3, presents a slightly different case of reflection, so these will have to be examined individually.

3.3.1 Presampling Filter

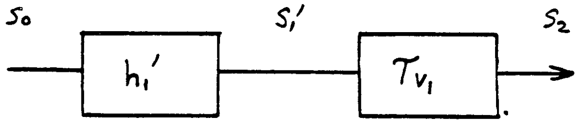
For the moment, let h_1 of the transmission system model be an arbitrary spatiotemporal filter $h_1(t,x)$ — not necessarily separable — instead of a strictly temporal filter. Transposing $\mathcal{T}_{v_1}$ and the filter h_1, we have

$$s_2 = h_1 * \mathcal{T}_{v_1}\{s_0\} = \mathcal{T}_{v_1}\{h_1' * s_0\} = \mathcal{T}_{v_1}\{s_1'\}$$

where

$$h_1'(t,x) \equiv h_1(t, x + v_1 t) \quad \text{and} \quad s_1' \equiv h_1' * s_0.$$

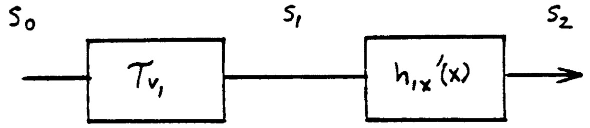
The transposed system is:



Since $s_0(t,x)$ is constant in time, the temporal filtering action of h_1' is irrelevant, so h_1' may be replaced by its time-average

$$h_{1x}'(x) \equiv \int h_1'(t,x)dt = \int h_1(t, x + v_1 t)dt.$$

Being a purely spatial filter, $h_{1x}'(x)$ can be placed after T_{v_1} with the same effect:



s_2 is not only a purely translating signal but also equal to s_1 *spatially* filtered by $h_{1x}'(x)$. Therefore, for a purely translating signal, every spatiotemporal filter is equivalent in effect to some 1-D spatial filter.

Furthermore, if h_1 is strictly temporal, then $h_1(t,x) = h_1(t)\delta(x)$ and

$$h_{1x}'(x) = \int h_1(t)\delta(x + v_1 t)dt = \tfrac{1}{v_1}h_1(-\tfrac{x}{v_1}),$$

which shows that every 1-D spatial filter is equivalent to exactly one 1-D temporal filter, and vice versa. We say that $h_1(t)$ is reflected into $h_{1x}'(x)$ by the constant velocity motion of the object since $h_{1x}'(x)$ and $h_1(t)$ are related by a simple geometric construction involving the line $t = -x/v_1$ (Figure 3.6).

This fact is important for three reasons. First, it justifies the simplifying assumption that the camera has perfect spatial characteristics, since any actual spatiotemporal filter can be represented in this model by a temporal filter. Second, it permits substitution

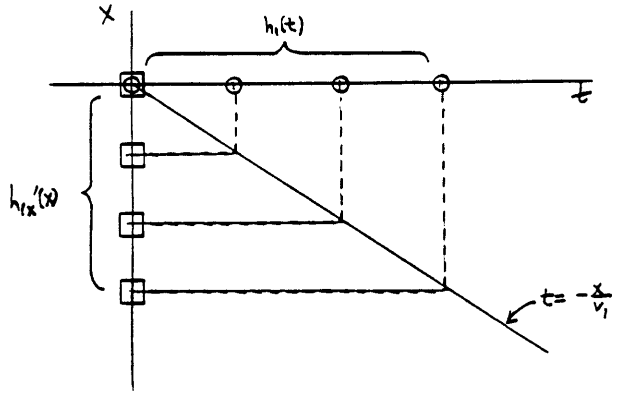


Figure 3.6: Geometric relation between reflected filters

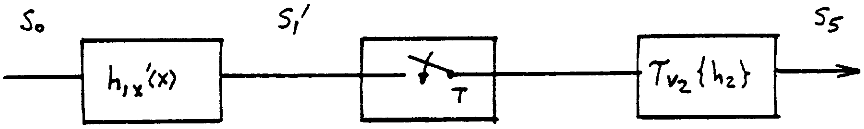
of costly temporal filtering by easier spatial filtering in a simulation program. Third, going back to the original motivation for this derivation, the effect of the presampling filter h_1 is seen by the HVS as a spatial filtering operation.

3.3.2 Interpolation Filter

The display interpolation filter h_2 is reflected into a spatial filter whenever the observer's eyes are in motion, but the particular situation of interest is ocular pursuit of a moving target, namely, when $v_2 = -v_1$. Since

$$\begin{aligned} s_5 &= \mathcal{T}_{v_2}\{s_4\} = \mathcal{T}_{v_2}\{h_2\} * \mathcal{T}_{v_2}\{s_3\} \\ \mathcal{T}_{v_2}\{s_3\} &= \mathcal{T}_{v_2}\{s_2\} \cdot \mathcal{T}_{v_2}\{T \textstyle\sum_n \delta(t-nT)\} = \mathcal{T}_{v_2}\{s_2\} \cdot T \textstyle\sum_n \delta(t-nT) \\ \mathcal{T}_{v_2}\{s_2\} &= \mathcal{T}_{v_2}\{\mathcal{T}_{v_1}\{s'_1\}\} = s'_1, \end{aligned}$$

the two translation operators cancel out, leaving the simpler equivalent model:



Because s'_1 is a stationary image, s_5 must be a temporally periodic image with period T. The spectrum of s_5 is concentrated at multiples of $1/T$, so it is meaningful to consider the image consisting of the DC temporal frequency component of s_5. This component can be isolated by replacing the filter $\mathcal{T}_{v_2}\{h_2\}$ with its time-average

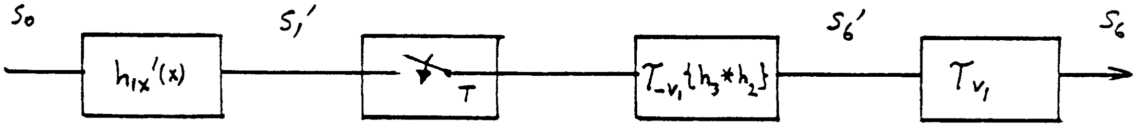
$$h'_{2x}(x) \equiv \int \mathcal{T}_{v_2}\{h_2\}dt = \int h_2(t, x - v_2 t)dt = \int h_2(t)\delta(x - v_2 t)dt = \tfrac{1}{v_2}h_2(\tfrac{x}{v_2}).$$

As far as the zero temporal frequency component of the retinal image s_5 is concerned, the temporal filter $h_2(t)$ is equivalent to the reflected spatial filter $h'_{2x}(x)$ when the moving image is observed in pursuit.

3.3.3 Visual Filter

The final case concerns the reflection of the temporal portion of the visual impulse response, $h_3(t)$, into a time-varying discrete spatial impulse response. This occurs only for fixated observation ($v_2 = 0$).

Starting from Figure 3.1 and using the techniques which by now are familiar, the translation operator $\mathcal{T}_{v_1}$ is successively transposed with h_1, the sampler, h_2, and h_3, until the following transposed model results:



Because observation is fixated, s_6 is an image containing motion with average velocity v_1. The image at the input to $\mathcal{T}_{v_1}$ in the transposed model, which is denoted by s_6', does not physically exist anywhere; however, it may be interpreted as the stationary mental image which results after the mental equivalent of pursuit has compensated for the motion.

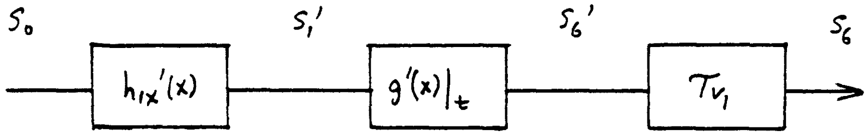
Letting $g \equiv \mathcal{T}_{-v_1}\{h_3 * h_2\}$, we write that

$$s_6'(t, x) = \int d\mu \int d\xi \; s_1'(\xi, \mu) T \sum_n \delta(\xi - nT) g(t - \xi, x - \mu).$$

Because s_1' is stationary, $s_1'(\xi, \mu) = s_1'(0, \mu)$, which allows simplification to

$$s_6'(t, x) = \int s_1'(0, \mu) T \sum_n g(t - nT, x - \mu) d\mu.$$

This expression shows that $s_6'(t,x)$ is the spatial convolution of $s_1'(0,x)$ with a time-varying spatial filter $g'(x)|_t = T\sum_n g(t-nT,x)$, which may be modeled as:



This result would not be useful except that $g'(x)|_t$ has a particularly simple form once we ignore the spatial filtering portion of visual response, writing $h_3(t,x) = h_3(t)\delta(x)$, and once we ignore $h_2(t)$ since it is typically of much shorter duration than $h_3(t)$. Then,

$$
\begin{aligned}
g(t,x) &= \mathcal{T}_{-v_1}\{h_3(t,x)\} = h_3(t)\delta(x+v_1 t) \\
g'(x)|_t &= T\textstyle\sum_n h_3(t-nT)\delta(x+v_1(t-nT)) \\
&= h_3(-\tfrac{x}{v_1})T\textstyle\sum_{n=-\infty}^{\infty}\delta(x+v_1(t-nT))
\end{aligned}
$$

This last expression states that the time-varying spatial filter to which s_1' is subjected is equal to the reflected visual temporal impulse response, multiplied by a sampling impulse train which shifts with time. For example, if $h_3(t)$ is a decaying exponential, then the corresponding $g'(x)|_t$ is a sampled decaying exponential (Figure 3.7).

3.4 Summary

The six-stage model of an image transmission system and visual response was manipulated using properties of the translation operator in order to compute the effective spatial convolutions which will be perceived by the observer. These reflected impulse responses will become more meaningful as soon as the subjective appearance of these convolutions are discussed in Chapter 5.

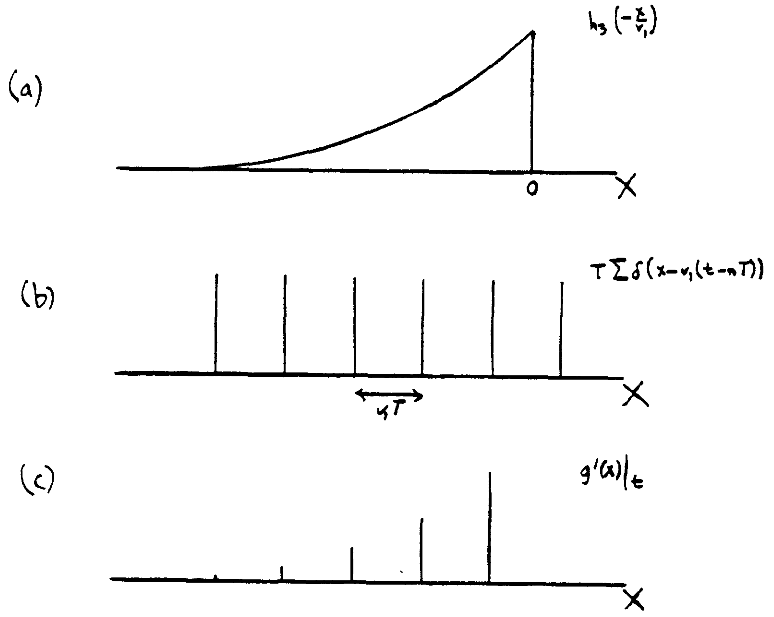


Figure 3.7: (a) Reflection of exponential visual impulse response; (b) sampling train; (c) sampled reflected impulse response

Chapter 4

Experimental Facilities

This study of motion defects required an accessible and convenient facility to generate and display high frame rate sequences at moderate spatial resolution. Motion picture film could have been created a frame at a time and then displayed by a high speed projector, but this method of experimentation is extremely inconvenient. Commercial and custom video processing systems which have been used in television research [PEL 83][JOHN 78][STAE 83] fail in accessibility and frame rate. The most accessible system is the Image Processing System within CIPG, which is equipped with a frame buffer using nominally NTSC scanning standards. Although the buffer cannot be loaded in real time, a trick for creating a moving display is dynamic modification of the tone scale mapping from buffer contents to displayed pel intensity [SCHR 84]. But this technique is not sufficiently convenient and also does not overcome the field rate limitation of NTSC.

In order to meet the requirements stated above, the High Rate Television System was configured. Due to the exploratory nature of this investigation, a larger investment in equipment was unwarranted. In the block diagram of Figure 4.1, all but the HRTV interface unit were existing components; therefore, only a minimum amount of new hardware had to be constructed.

This chapter will summarize the external user-visible aspects of the HRTV system, viewing conditions for the observations, and the software which simulates the image transmission system modeled in the preceding chapter. Implementation details have

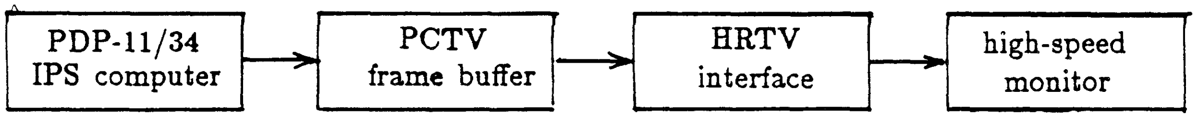


Figure 4.1: Block diagram of HRTV facility

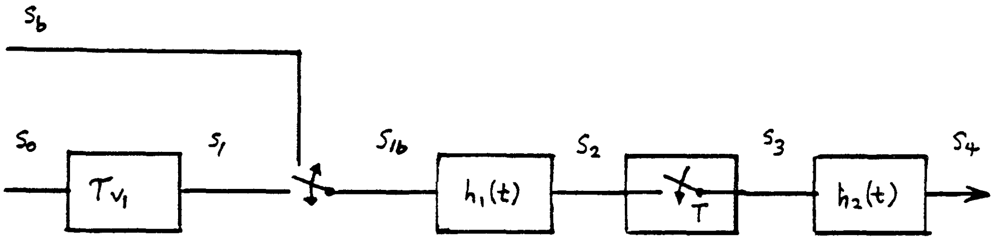


Figure 4.2: Block diagram of Transmission System Simulation

been relegated to Appendix A.

4.1 High Frame Rate Display

The image output device consists of a 120 fps raster on a CRT with white P4 phosphor, which was measured to decay to 50% intensity in 180 μs. The 120 active video lines are progressively scanned (horizontally) at a 16 KHz line rate and form a raster 4.5 cm high by 22 cm wide.

Each line displays 512 pels whose time-averaged intensities are adjustable in 256 equal-luminance steps from black to 16 ft-L. However, within a frame the pels in a vertical column are required to be the same intensity; i.e., images must be constant in the y direction, just as in the model. Storage for 509 individual frames permits the display of one 4.2 sec sequence or two 2.1 sec sequences. Under these mild restrictions, a variety of time-varying visual stimuli can be generated by means of the PDP-11/34 computer, loaded into image memory, and displayed by the HRTV system. Finally, a black or maximum-white stationary vertical line can be superposed on the display at any horizontal position to serve as a visual fixation point marker.

To operate the display, the observer selects and holds down a button, which displays the first frame of the chosen image sequence. This hold is essential to permit initial fixation on the target or on the fixation marker. When the observer releases the button, frames are displayed in succession. The last frame is frozen at the end instead of blanking the screen, in order to avoid an uncomfortable change in average display brightness.

4.2 Observation Conditions

A standard set of viewing conditions was established for formal observation sessions. At the standard viewing distance of 180 cm (8 times the raster width), the image subtends 1.5° × 7.2°. Neither the individual scan lines (80 cpd) nor the fundamental component of a maximum frequency, maximum contrast grating (36 cpd) is visually resolvable. Images always moved from left to right even though the HRTV could display right to left motion just as easily.

The room where observations were performed was made as dark as possible so that the luminance of the white cardboard surrounding the HRTV display was under 0.1 ft-L. Yuyama recommends that a HDTV display should have a peak luminance of over 30 ft-L for optimum picture quality [YUYA 82]. However, in low ambient lighting, the 16 ft-L display is more than adequate, and in fact a 16 ft-L test object of appreciable size, against a black background, would be glaringly bright.

All observations were made by the author, whose myopia was corrected to 20/15 as measured by the Snellen acuity test chart. The author's defective binocular fusion dictated monocular observations.

4.3 Simulation of Transmission System

In this research the images of interest consisted of one visual target moving in front of a stationary background. A single computer program was written to simulate the image transmission system of Figure 4.2, which is similar to the model of the preceding chapter. There is provision for an arbitrary stationary background s_b instead of requiring a constant background. Input parameters to this program permit selection of background s_b, moving target s_0, velocity v_1, presampling filter $h_1(t)$, the transmission sampling period T, and interpolation filter $h_2(t)$.

At the beginning of an image sequence the target is accelerated from standstill at 1/9 pels/frame^2 until the desired velocity v_1 is attained. Known as an "ease" by animators, this acceleration helps prevent the perception of defects normally associated with fixation during the latency period of ocular pursuit.

There is no inherent limit to velocities displayable on the HRTV, but in practice velocities above 10°/sec (at the standard viewing distance) cannot be measured easily because the presentation duration is limited by the field of view.

Since the filters and images are input as files rather than being built into the simulation program itself, these parameters are almost completely arbitrary. However, in the observations only the following sets of filters were used.

Four temporal presampling filters were tried as $h_1(t)$: a δ-function, a Gaussian pulse with standard deviation $\sigma = 0.3T$, and rectangular pulses of durations $T/2$ and T. The value of σ was chosen because it supposedly yields the optimum spatial presampling filter as determined by subjective evaluation of nonmoving images [RATZ 80][SCHR 85]. These four filters, denoted by Fd, Fg, Fr/2, and Fr, respectively, are illustrated in Figure 4.3, along with their Fourier transforms and with an ideal low-pass filter for comparison.

The HRTV's CRT phosphor decay waveform, if assumed to be exponential, would be the impulse response of a single-pole low-pass filter with cutoff at 4 KHz. Considering the time scale of human vision and the 120 Hz display frame rate, that impulse response is a δ-function for all practical purposes. It follows that the HRTV display's

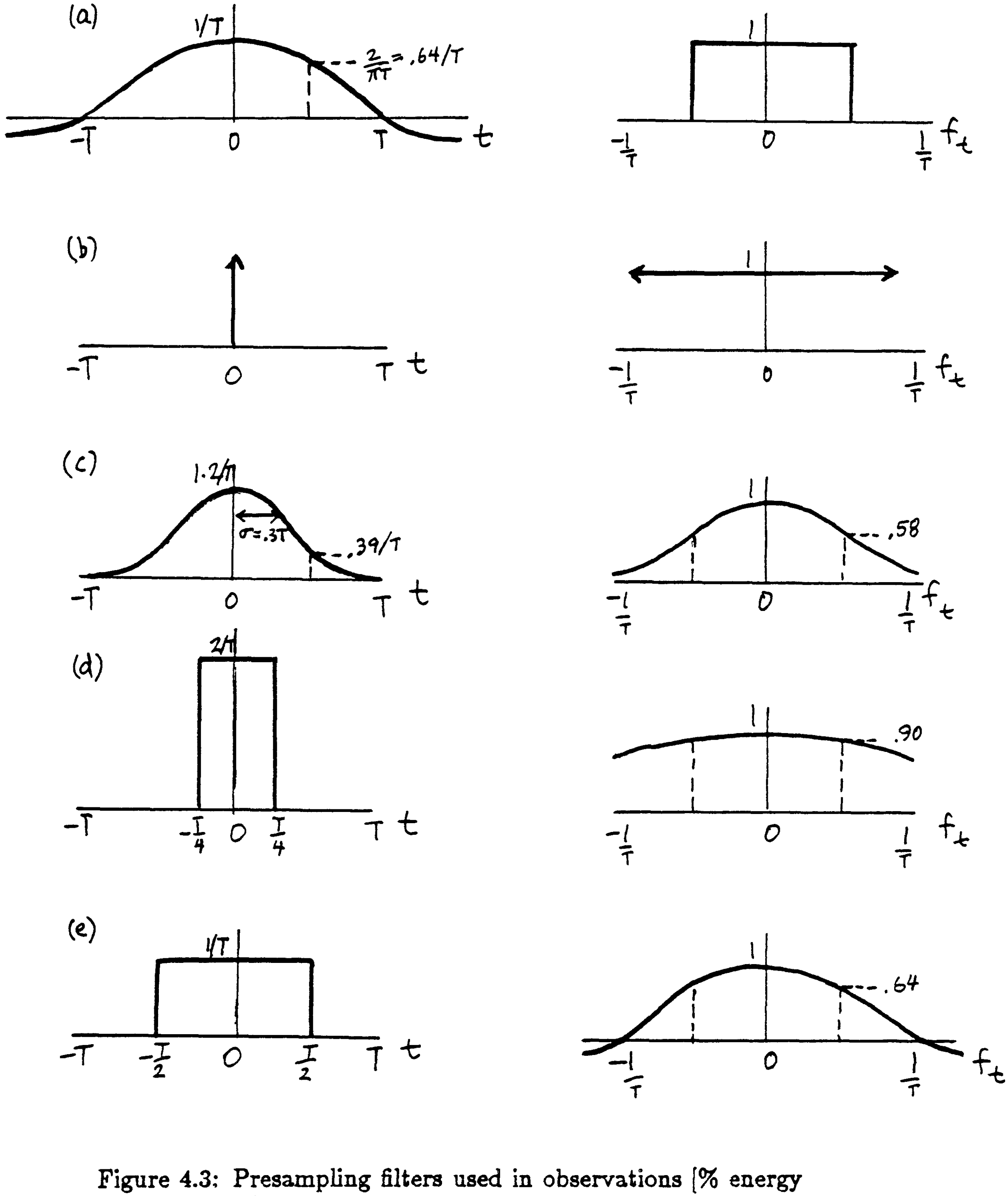


Figure 4.3: Presampling filters used in observations [% energy within $|f_t| < \frac{1}{2T}$]:(a) ideal sin x/x for comparison [100%]; (b) Fd = δ-function [0%]; (c) Fg = Gaussian, $\sigma = 0.3T$ [86%]; (d) Fr/2 = rectangular pulse, duration $T/2$ [47%]; (e) Fr = rectangular pulse, duration T [77%]

interpolation filter $h_2(t)$ is necessarily a train of impulses, spaced by T/r.

The four interpolation filters employed are sampled versions of the following continuous waveforms: a δ-function, a rectangular pulse of duration T (sample-and-hold), a triangular pulse of duration $2T$ that is sampled at the peak, and the same triangular pulse with sample points shifted by 1/2 of the sampling period. These filters are denoted by Fd, Fr,[1] Ft, and Fu, respectively, and are illustrated for rate $r = 3$ up-sampling in Figure 4.4.

The δ-function represents absence of frame rate up-conversion at the display, so that the display frame rate is effectively $1/T$, which must be a integral submultiple of 120 fps. The rest of the filters cause up-conversion to the display frame rate $r/T = 120$ fps. The rectangular pulse represents frame repetition, and the two triangular pulses represent linear interpolation. The filter Fu has the property that every displayed frame is a linear combination of two different transmitted frames; the filter Ft lacks this property since the display frame which is concurrent with a transmitted frame is derived from only that transmitted frame.

When $r = 1$, the filters Fr and Ft become identical to Fd since no interpolation is necessary. However, when $r = 1$, the filter Fu still consists of two equal impulses separated by 1 frame. Since no rate up-conversion is achieved in this case, it is not clear what, if anything, is to be gained by use of this filter.

Caveat. While the discrete filters which are actually used to simulate the presampling filter $h_1(t)$ are fair approximations to continuous-time filters, the discrete interpolation filter $h_2(t)$ cannot simply be thought of as an approximation to the continuous-time filter from which it is derived by sampling, because even the 120 fps display rate is sometimes not high enough to conceal the temporal structure. Indeed, $h_2(t)$ may be reflected into a spatial filter, which causes a characteristic image defect soon to be described.

[1] It will be clear from context when Fr indicates a continuous presampling filter or a discrete interpolation filter.

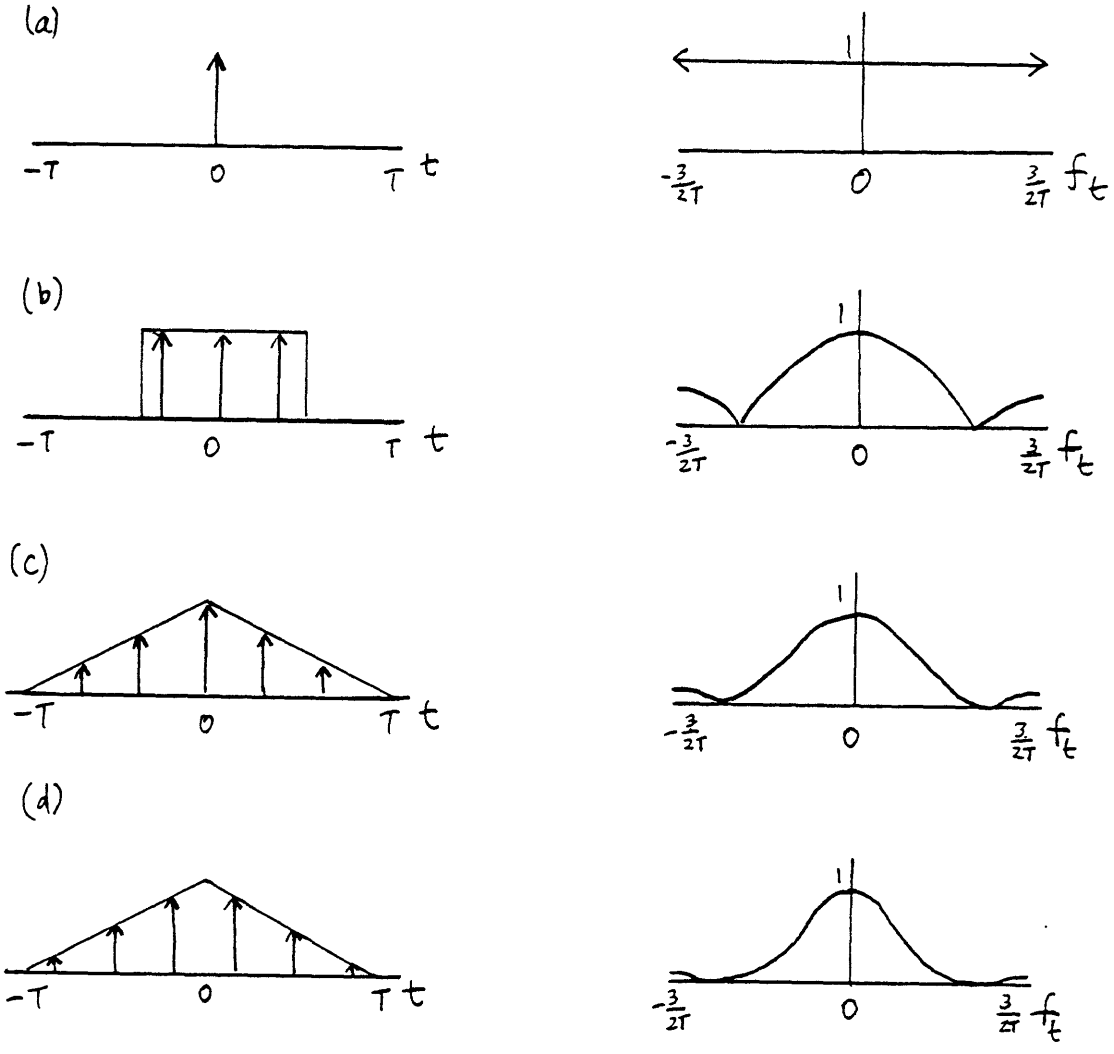


Figure 4.4: Interpolation filters used in observations are sampled from (a) Fd = δ-function; (b) Fr = rectangular pulse; (c) Ft = triangular pulse; (d) Fu = offset triangular pulse (The number of side lobes in the spectra will differ when the up-sampling rate is changed from $r = 3$.)

Chapter 5

Classification of Defects

Among all the defects caused by temporally sampling an image containing motion, perhaps the most striking and well known kind is the "stroboscopic" contrary motion which sometimes appears on moving gratings, such as wagon wheels and airplane propellers, in a television or motion picture display. However, because it occurs relatively infrequently and also because its appearance is more often amusing than annoying, this contrary motion defect is not of unique importance. Initial observations with the HRTV system have indicated that most image defects can be classified into one of four major classes. Multiple images and blur are two types of spatial defects, while large-area flicker and small-area flicker constitute the class of temporal defects. The tilt artifact is a type of geometrical defect, and finally, contrary motion and jerkiness are classified as velocity defects.

The objectives of this chapter are to describe, in turn, the subjective appearance of the various types of image defects, to define the defects with sufficient accuracy for unambiguous classification of observed images, and to relate appearance to physical characteristics of the displayed images. More emphasis will be placed on the spatial and temporal classes of defects since these are the subject of the experimental observations to be described in the next chapter.

With the image transmission system model given in Chapter 3, it is impossible to overlook the fact that eye movements determine the appearance of the perceived image s_6. The importance of ocular pursuit goes beyond enabling the perception of spatial details on moving objects, for the subjective appearance of almost every kind of defect is dependent on the velocity of the eye with respect to the display and moving objects.

When an observer views a purely translating image, it is fairly certain that he will track the motion, unless instructed to do otherwise. In this situation v_2 is predictable and equal to $-v_1$. Natural images may contain several moving objects so that the observer has a choice between pursuit at one of several velocities as well as fixation or a saccade. Fixation can never be ruled out because images usually contain stationary regions which the observer may wish to examine; in addition, even if one decides to track, the eye is stationary during the latency period.

As a result, an object with velocity v_1 (possibly 0) can be observed even when the eye moves at an unrelated velocity $v_2 \neq -v_1$. Since v_2 is fairly arbitrary, the most general study of image defects would have to consider a wide range of v_2's for any given v_1. The two fundamental cases studied in this thesis, however, are $v_1 \neq 0, v_2 = 0$ and $v_1 \neq 0, v_2 = -v_1$. In other words, only the defects of a moving object seen during fixation or during pursuit of *that* object are of primary concern.

5.1 Spatial Defects

5.1.1 Multiple Images

The first spatial image degradation is called the multiple images defect. It is the concurrent perception of two or more spatially displaced images of one object or object feature in the image s_6. To qualify as a multiple images defect it is not required that the onset and disappearance of the individual images be perceived as simultaneous, but only that at some instant two or more images appear to coexist. It is also not necessary that every aspect of the original object be visible in each of the multiple images or that they appear with equal brightness; it suffices that some salient feature appears two or more times at regularly spaced intervals.

By the preceding definition, an object which is perceived to successively occupy a discrete set of positions, but only *one* position at a time, is not an example of a multiple images degradation. Apparent concurrency is essential.

In terms of linear filtering, a moving image with this defect appears to be the result of spatial convolution with a multimodal impulse response — an impulse response with several peaks. As will be explained in detail, such filters can arise in practice by reflecting the interpolation filter of the display and the temporal filter of the visual system into the effective spatial filters $h'_{2x}(x)$ and $g'(x)|_t$, which were previously derived. Because in practice $h_1(t)$ is not multimodal, its spatial reflection does not induce multiple-image defects.

Pursuit

The appearance of multiple images during ocular pursuit was not always understood. Braunstein reported that a single small moving dot in a motion picture projected by three-fold frame repetition looked like a group of three moving dots when observed with ocular pursuit, and he advanced two alternative explanations. The first assumed the existence of a cognitive conversion process from "discrete position changes to continuous apparent movement," and the second considered movement of the eye with respect to each fixed position where a dot is repeated [BRAU 66A][BRAU 66B]. Kintz correctly validated the latter explanation [KINT 72], which would follow directly from the theory presented in Chapter 3.

A moving dot which is sampled at rate $1/T$ without a presampling filter and then interpolated by three-fold frame repetition would result in the displayed image s_4 shown in Figure 5.1. Since ocular pursuit subtracts out the mean velocity, the retinal image s_5 consists of three stationary dots which are illuminated in succession. If $1/T$ is above the CFF then no flicker is seen and the perceived image consists of the $f_t = 0$ component of s_5, which could have been predicted by reflecting the temporal filter $h_2(t)$ into $h'_{2x}(x)$ as shown in Figure 5.1e. Note that the distance between the multiple images is equal to the product of velocity and display frame period. This image is the prototypical example of the multiple images defect with pursuit, in which one object is perceived as three spatially displaced images.

A second basic example is with $s'_1(x)$ a rectangular pulse, which contains two abrupt

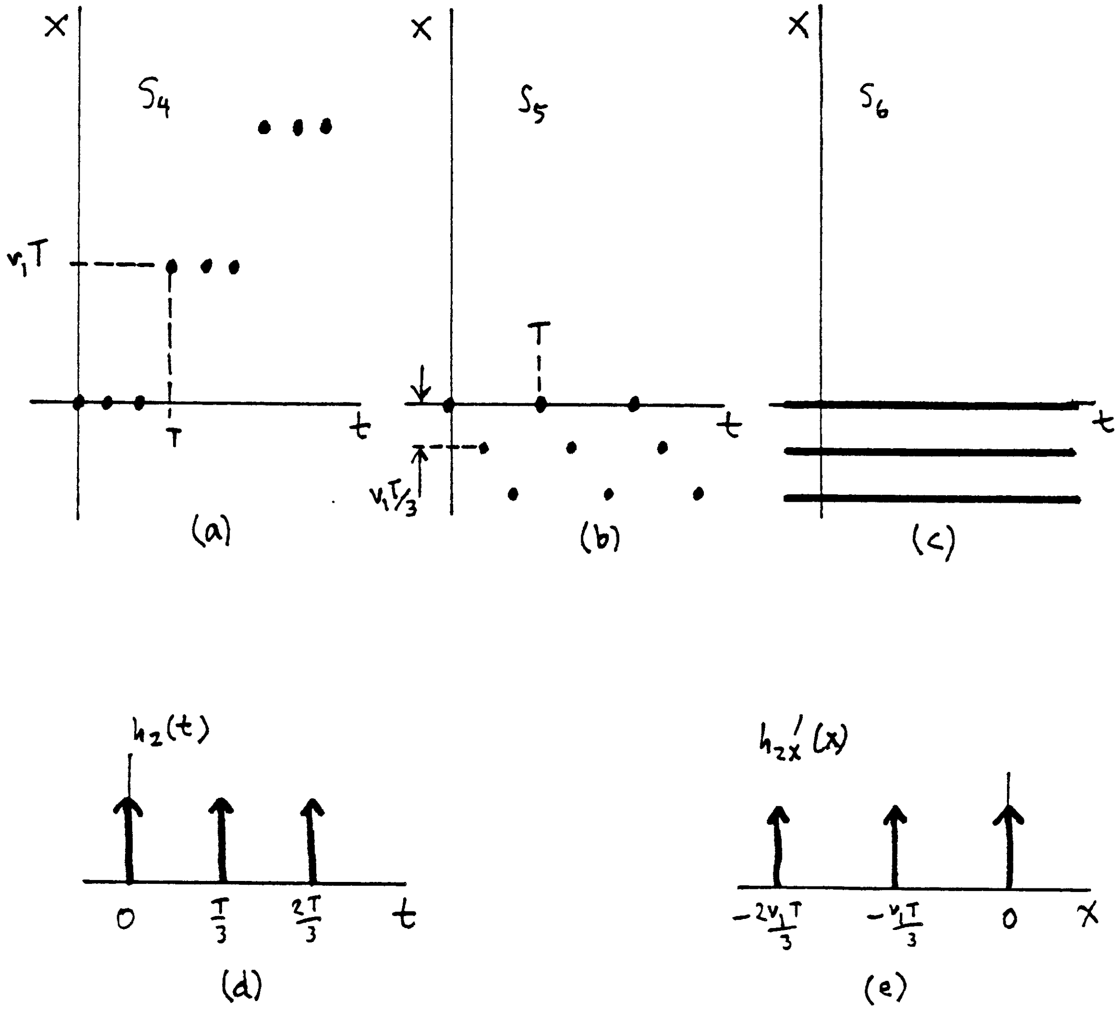


Figure 5.1: Multiple images during pursuit of a small object when $r = 3$: (a) displayed image; (b) retinal image; (c) perceived image; (d) interpolation impulse response h_2; (e) reflection of h_2

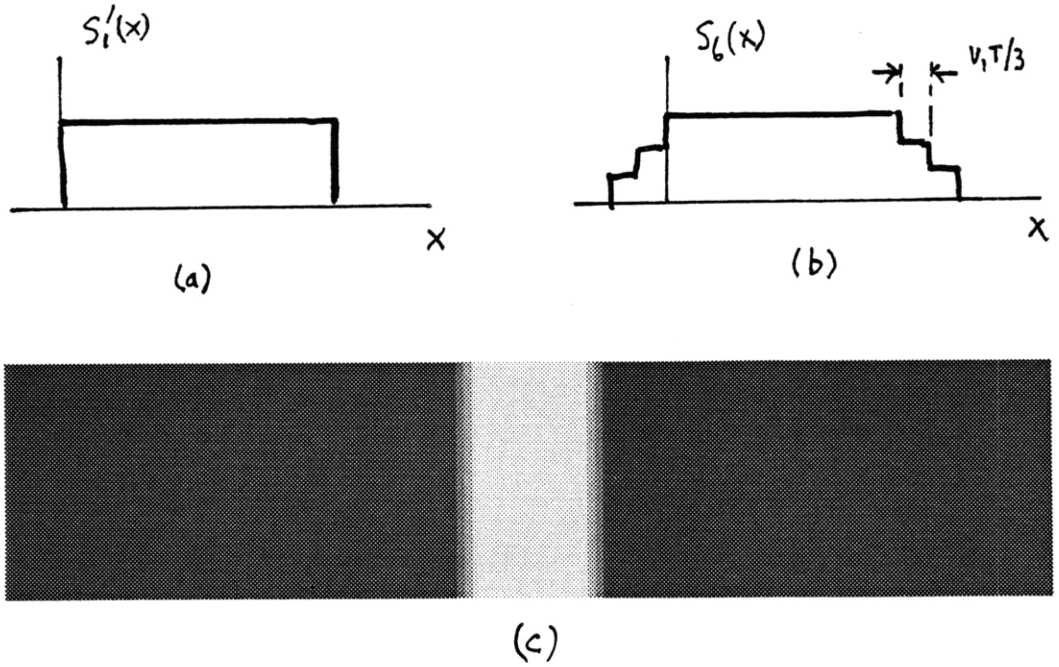


Figure 5.2: Multiple images during pursuit of a step when $r = 3$: graph of (a) moving image and (b) perceived image; (c) simulation of the image seen by viewer

dark to light steps. When this image is moving and viewed with pursuit, convolution of $s_1'(x)$ with the reflected impulse response $h'_{2x}(x)$ causes the image s_6 to make the transitions in three steps (Figure 5.2). The edges would appear as two stripes of intermediate brightness.

Figure 5.2c shows one typical image as seen by the viewer. Owing to the diverse appearance of defects which are all classified as multiple images under ocular pursuit, it is impractical to illustrate them all.

Fixation

The multiple images defect also arises under fixated observation of certain moving images, but the subjective appearance and the physical nature of such images differ greatly from those of multiple images under ocular pursuit. Consider a moving dot, sampled at rate $1/T$ without a presampling filter and viewed while fixating a stationary point (Figure 5.3). For a range of values of T, the perceived image s_6 consists of a

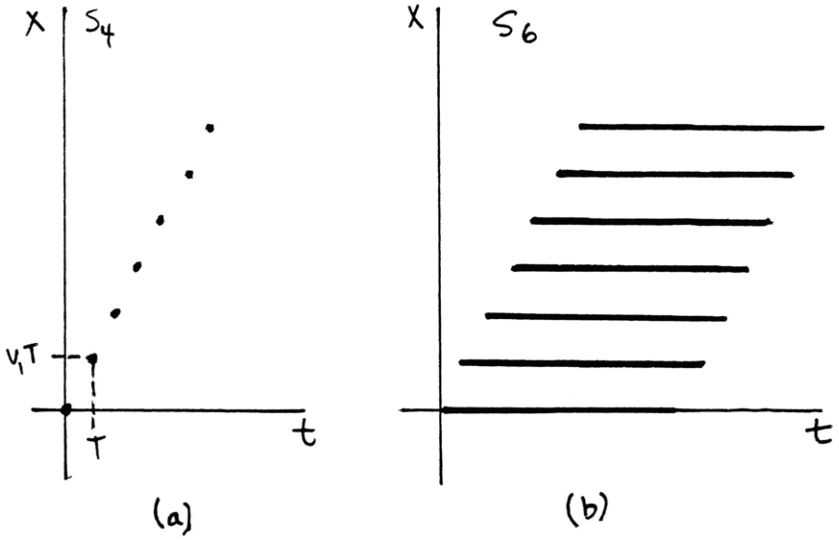


Figure 5.3: Multiple images during fixation: (a) displayed image; (b) perceived image

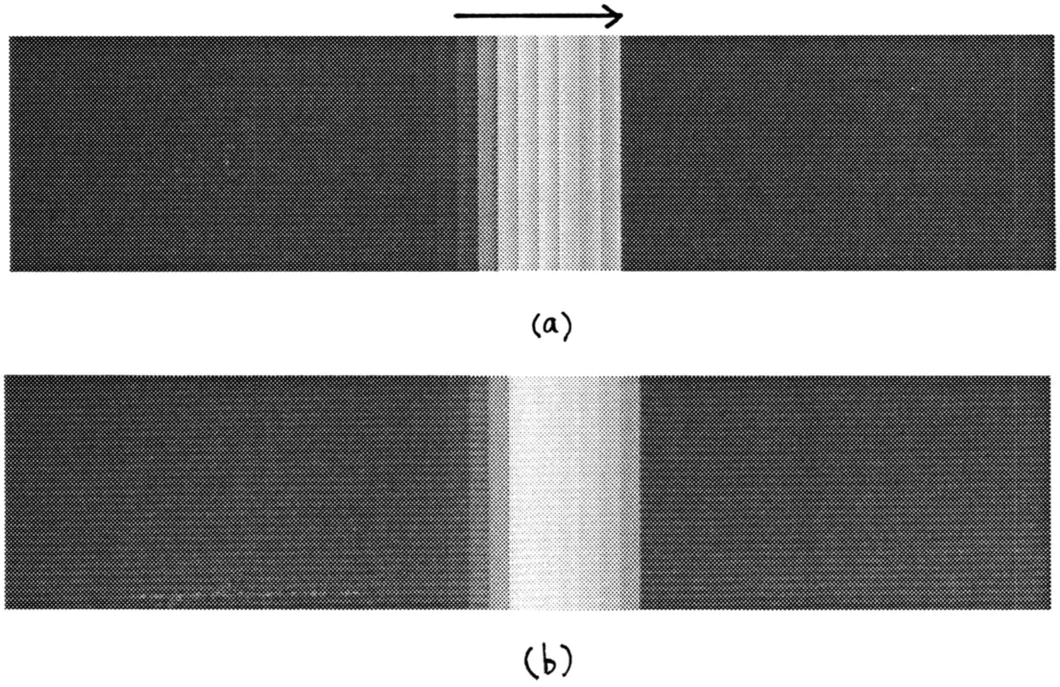


Figure 5.4: (a) Simulation of multiple images seen by the viewer during fixation with a rectangular pulse; (b) defect seen if $h_3(t)$ were exponential

set of dots in which the individual dots remain stationary but the group appears to move as new dots appear on the leading edge of the group and old dots disappear at the trailing edge. The dots may not be equally bright — the trailing dots tend to have less contrast than the rest.

These observations can be readily explained. Simply speaking, the persistence of vision embodied in $h_3(t)$ lasts much longer than T so that many dots appear to be illuminated simultaneously. The train of impulses in the effective spatial impulse response $g'(x)|_t$ corresponds to the existence of regularly space multiple images. In contrast to the situation of ocular pursuit, the distance between multiple images is equal to the product of velocity and *transmission* frame period. Finally, the width and shape of the envelope of $g'(x)|_t$ control the number of multiple images and their relative intensities. The foregoing explanation is reasonable even if the linearity of large-amplitude transient response is questionable, as long as $h_3(t)$ is taken to be the response to a specific input rather than a general convolution kernel.

Figure 5.4a is a "snapshot" of the multiple images perceived when the original s_1 is a rectangular pulse, whose width is comparable to the extent of $g'(x)|_t$. The trailing steps on the left lag $\gtrsim$ 100 ms behind the physical edge of the object. $g'(x)|_t$ is not subjectively close to an exponential decay, which is synthesized in Figure 5.4b. The Mach effect is probably responsible for the nonuniform brightness of individual steps and for accentuating the relatively small difference between steps. By necessity, Figure 5.4a is "simulated television reception" because no photographic camera responds temporally in the same way as the eye.

From the preceding sections it should be clear that multiple image defects under ocular pursuit and fixation are very different in nature. Both defects arise from multimodal reflected spatial impulse responses, but this is a superficial similarity. Underlying the differences is the fact that the relevant impulse response is $h'_{2x}(x)$ for pursuit but $g'(x)|_t$ for fixation. For pursuit the multiplicity of images is predetermined by the number of peaks in $h_2(t)$, which is a system parameter, whereas for fixation it is governed by temporal properties of vision. The inter-image separation depends on T for fixation but on T/r for pursuit. As a result of such dissimilarities, the existence of the defect under fixation in a given situation does not imply its existence under pursuit, nor vice versa. Finally, while multiple images under fixation indicate an limitation of low frame rates, the defect under pursuit is just an artifact of interpolation.

It might be noted that multiple images under fixation can be seen outside of television and motion pictures. Objects moving continuously under intermittent sources such as some street lights may exhibit such defects. Nevertheless, multiple images are not often experienced in normal observation of the real world, unlike blurred images, which are part of everyday experience.

5.1.2 Blur

Blur is the observed reduction of spatial resolution of a moving target, relative to the resolution of that target when at rest. The observable symptoms of blur include reduction of contrast, loss of texture and fine details, and less abrupt transitions at edges — the same degradations caused by improper focusing. Since this research assumes that all elements of the system have ideal spatial characteristics, all blur must be exclusively the result of temporal filtering reflected into spatial filtering by motion. There are three temporal filters, so there are potentially only three contributions to image blur.

The amount of blur introduced in the camera depends on the temporal impulse response of the camera and the velocity of motion, since the effective spatial filter is $h'_{1x} = \frac{1}{v_1} h_1(-\frac{x}{v_1})$. In television and photographic cameras each frame is formed by integrating light from the moving scene s_1 over some period of time $P \leq T$ so that $h_1(t)$ is approximately a rectangular pulse of duration P. This is reflected into a spatial pulse of width Pv_1, i.e., the distance moved during the exposure time. When this distance is comparable to or greater than the dimensions of the finest resolvable details in an image, then the blur degradation may be observed.

In the limit of infinite sampling rate ($1/T \longrightarrow \infty$) a fixating observer's temporal visual filter would introduce motion blur just as the camera does, because again there is a moving image and a stationary receptor. However, the effective spatial impulse response is $g'(x)|_t \longrightarrow h_3(-\frac{x}{v_1})$. In practice, the approximation $g'(x)|_t \approx h_3(-\frac{x}{v_1})$ is valid whenever conditions are such that the multiple images, spaced by $v_1 T$, cannot be

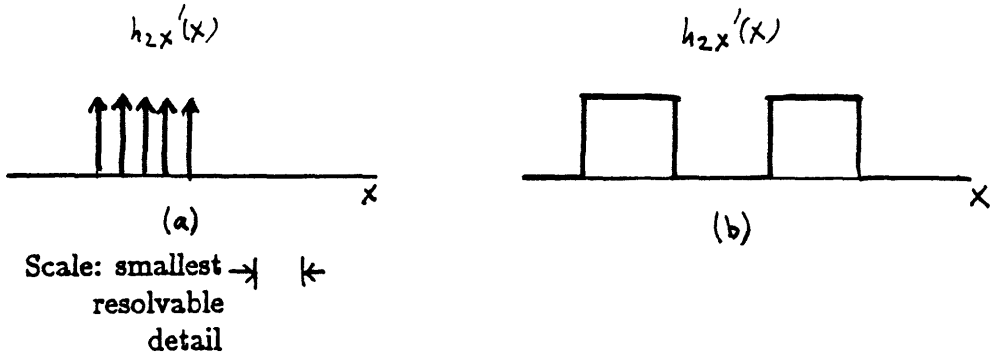


Figure 5.5: Relations between blur and multiple images: (a) blur as a degenerate case of multiple images; (b) blur concurrent with multiple images

resolved. Then blur would be perceived whenever $h_3(-\frac{x}{v_1})$ is of sufficiently wide extent. On the other hand, this blur is not considered to be an *image* defect since it would also occur under direct viewing of the original scene.

Spatial convolution with $h'_{2x}(x)$ during ocular pursuit of a moving displayed image is the third and final possible contribution to blur. If the reflected display interpolation filter were unimodal, then this source of blur would differ from the previous two sources in only one respect: here, the image s_6 is stationary — it pauses for duration T between steps — and the receptor (the eye) is in motion, rather than the other way around. For example, a CRT phosphor with an exponential decay curve would theoretically reflect into an exponential spatial impulse response.

On the other hand, $h'_{2x}(x)$ is often multimodal, as in frame rate up-conversion. Even if this is the case, $h'_{2x}(x)$ can still contribute to blur under one of the following two situations. First, the individual peaks which constitute $h'_{2x}(x)$ are too closely spaced for multiple images to be resolved under the circumstances, and at the same time the pulses collectively span a sufficient width (Figure 5.5a). It is said that a potential multiple images defect has degenerated into blur. Second, $h'_{2x}(x)$ does cause multiple images, but each of the peaks is sufficiently broad to individually create noticeable blur (Figure 5.5b).

The latter situation illustrates that multiple images and blur do not have to be mutually exclusive spatial degradations, since the individual replications of the image can themselves appear to be blurred. Besides an interpolation filter like the one in Figure 5.5b, another way for the defects to coexist is for $h'_{1x}(x)$ to be causing blur while $h'_{2x}(x)$ or $g'(x)|_t$ is inducing multiple images.

5.2 Temporal Defects

The discussion now turns to the class of temporal defects in moving images, which includes large-area and small-area flicker. Although both of these defects arise from incomplete concealment of the sampling structure, there are important differences in their subjective appearance. Moreover, stimuli which elicit one or the other defect are physically distinguishable. Hence it is vital to define the defects subjectively and to contrast their objective characteristics.

Large-area flicker will be presented first. While this defect cannot be said to be motion-*induced*, at least velocity has a mediating role. A better justification for discussing it is that large-area flicker is a reference against which small-area flicker, a true motion-induced defect, may be compared.

5.2.1 Large-area Flicker

Large-area flicker is a perceived temporally periodic brightness modulation which is global in scope; that is, it appears indiscriminately in all regions of the perceived image s_6 which are of sufficient brightness and extent. As long as the extent requirement is satisfied, the perceived appearance of this type of flicker is independent of motion. Furthermore, all affected regions appear to flicker synchronously, as far as the observer can tell. This definition is constructed in order to exclude small-area flicker, which is not a global defect.

Extent is significant when a small bright region of potential flicker is contained in a large, darker surround. For one thing, flicker perception is governed by the Granit-Harper relation, which lowers the CFF for stimuli of smaller area. Secondly, in the event that the bright region has a nonzero velocity in the retinal image s_5, its dimension in the direction of motion needs to be much greater than the distance moved between two consecutively displayed frames. In other words, the extent of the region's overlap between frames cannot be too small. The reason is that inside any fixed retinal field, a number of cycles of modulation need to be received before flicker is perceived. If the velocity is too high, the individual pulses of light from the moving region will fall on widely separated retinal fields and cannot induce the perception of flicker.

An example will clarify the preceding point. Suppose that there are small and large targets, both of which flicker when their motion is stopped or when they are observed under pursuit. If these targets are then set in motion and observed while fixating as shown in Figure 5.6, the small target would not appear to flicker but the large one would.

The frequency of the modulation which is perceived in large-area flicker is nearly always the display frame rate, although this is not necessarily true for arbitrary $\tilde{h}_2(t)$. In order to avoid dealing with original images that contain intentional flicker in the following analysis, suppose that the intensity is slowly-varying in the spatiotemporal region surrounding the point (t_0, x_0) in the camera image s_2. Then the corresponding region in the displayed image is described by

$$s_4(t,x) = h_2(t) * \left[T \sum_{n=-\infty}^{\infty} \delta(t-nT)s_2(nT,x)\right] \approx s_2(t_0,x_0)\tilde{h}_2(t),$$

where

$$\tilde{h}_2(t) = \sum_{n=-\infty}^{\infty} h_2(t-nT)$$

is the periodic extension of the display interpolation filter $h_2(t)$. One interpretation of this calculation says that the display image is equal to $\tilde{h}_2(t)$ modulated by a spatiotemporally varying envelope.

When no frame rate up-conversion is performed, the Fourier series for $\tilde{h}_2(t)$ will usually have a large component at the frequency $1/T$, which would be the frequency of any visible large-area flicker. When rate r up-conversion is performed, $h_2(t)$ is usually designed so that the $1/T$ component becomes too small to be visible; indeed, $h_2(t)$ typically has zeros at $f_t = k/T, \quad 0 \leq k < r$. The lowest significant Fourier component, and hence flicker rate, would have to be at the display frame rate r/T. Up-conversion by

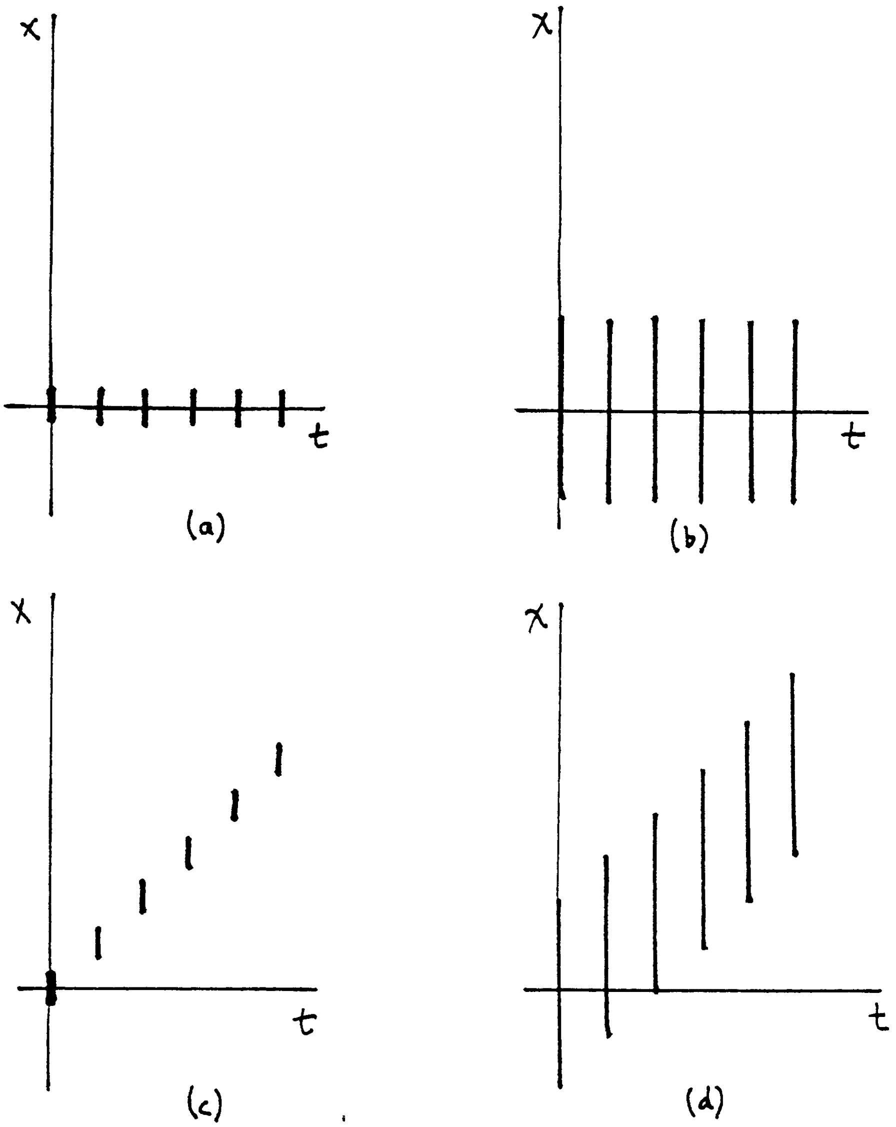


Figure 5.6: Effect of velocity on large-area flicker: (a,b) both objects flicker when stationary; (c,d) a speed at which only large object flickers

a filter $h_2(t)$ which does not adequately suppress the $1/T$ component is the conceivable, but unlikely, situation in which the large-area flicker rate is not equal to the display rate.

Even if this unlikely situation is conveniently ignored, large-area flicker should not be defined as "perceived temporal modulations occurring at the display frame rate." Such a definition violates the spirit of this chapter because there is no way for an observer to ascertain visually (i.e., without a measuring instrument or reference image of known flicker rate) whether the flicker he sees is actually occurring at the display frame rate.

5.2.2 Small-area Flicker

Small-area flicker shall be defined as a perceived temporally-recurring brightness variation which is localized to the vicinity of a moving, high-contrast edge. As will be shown, this defect occurs at the frequency of frame transmission, $1/T$, so that if large-area flicker of the same frequency exists in the perceived image, small-area flicker would be completely obscured. Since large-area flicker at a frequency higher than $1/T$ was not examined in the observations, it is not known whether it would mask the slower small-area flicker; however, it is likely that it would if the large-area flicker were sufficiently strong. Except for frequency, small-area flicker under ocular pursuit and under fixation possess few common properties. Accordingly, these two modes of observation will be examined individually.

Pursuit

During ocular pursuit, the perceived variation is caused by a physical alternation between lighter and darker levels of brightness. To begin by way of example, consider the moving step which was used before to illustrate multiple images (Figures 5.7a,b). It is assumed that the three-fold frame repetition has eliminated flicker, as intended, from the part of the image where $x > 0$. However, under certain conditions flicker can still be perceived on the intermediate steps between $x = -2v_1T/3$ and $x = 0$. The reason is that these two steps are refreshed only once or twice out of every three frames

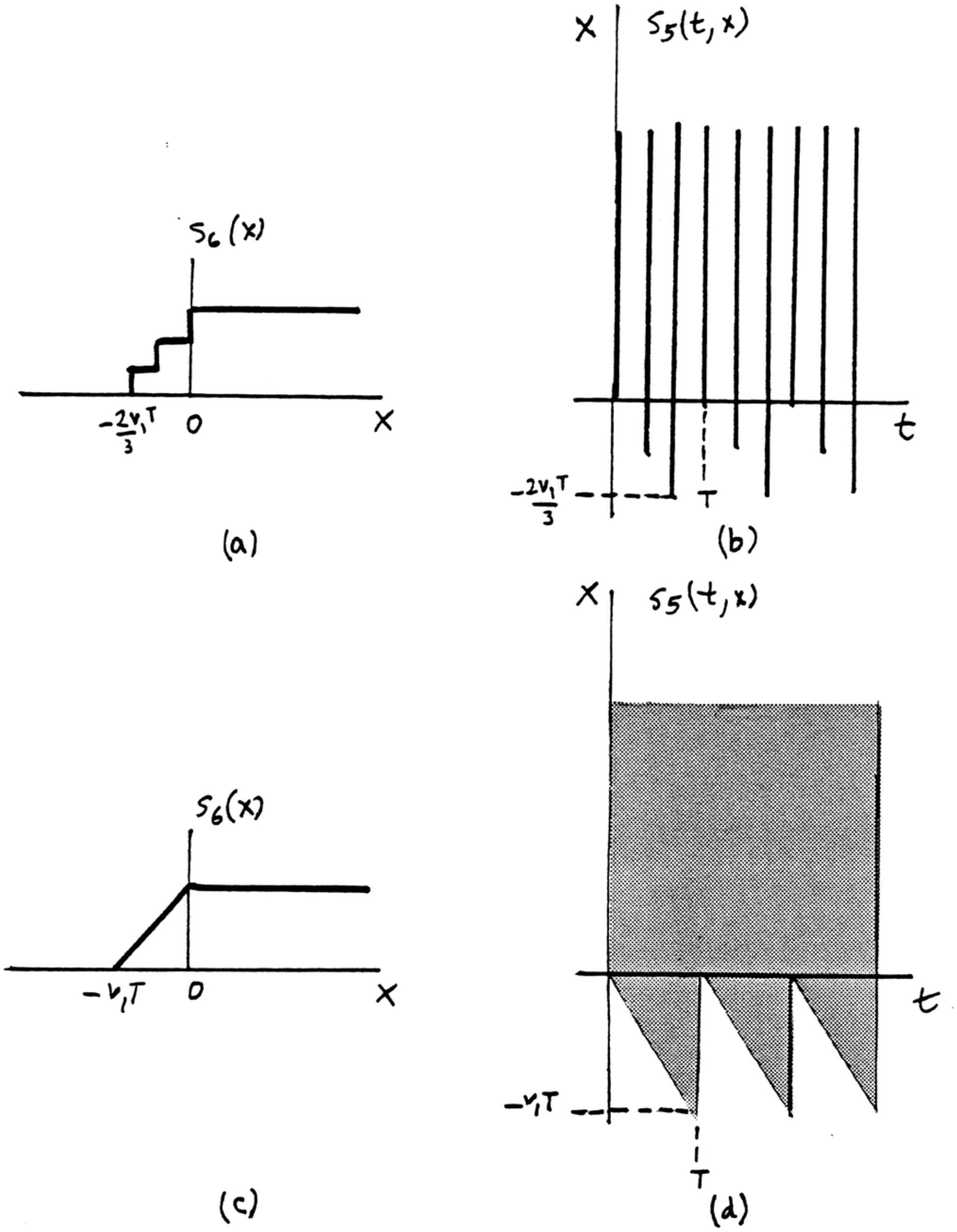


Figure 5.7: Small-area flicker during pursuit of a step: (a) perceived image and (b) retinal image for frame repetition when $r = 3$; (c) perceived image and (d) retinal image for continuous-time sample-and-hold

displayed during one transmission frame period T. Thus, this region contains energy at the fundamental temporal frequency $1/T$, which is seen as small-area flicker when $1/T$ is below the CFF.

Actually, neither a multimodal impulse response $h_2(t)$ nor a visible multiple images defect is necessary for there to be small-area flicker. A case in point is a moving step interpolated by a continuous-time rectangular pulse of full duration T (Figures 5.7c,d). During pursuit, the edge transition is a linear ramp, which prevents multiple images, yet small-area flicker is possible in the ramp portion of the image. In general, whenever an edge is convolved with a reflected spatial filter $h'_{2x}(x) \neq \delta(x)$, the region near the edge contains some energy at $f_t = 1/T$ which could result in small-area flicker.

Fixation

During fixation the localized region of small-area flicker necessarily moves with the target, and the subjective appearance of the defect is best described as a moving progression of brightness pulses. It can be distinguished from large-area flicker because the appearance of small-area flicker changes markedly when pursuit is initiated.

Again consider the moving step sampled without a presampling filter and interpolated with a rectangular pulse of duration T. Assume that $1/T$ is low enough that large-area flicker would be present if it were not for interpolation. At a point x_1 in the retinal image s_5, the brightness vs. time function $f(t) = s_5(t, x_1)$ is just a step function (Figures 5.8a,b). Likewise, there is no spatial position in s_5 at which a temporally periodic brightness modulation exists. For this reason, it is not immediately obvious how to explain the small-area flicker which is perceived in s_5 under the right conditions.

A reasonable explanation can be given if two propositions are accepted. First, the receptive field of a "flicker detector" in the HVS is sufficiently broad to span at least a few multiples of $v_1 T$, i.e., wide enough to integrate over a number of multiple image steps in the example of the figure. Second, a series of brightness increments in the same direction can stimulate a flicker detector as well as alternations between bright and dark periods can. This makes sense because of the low-frequency differentiator-like response of $h_3(t)$. Together, these statements suggest that a sequence of brightness level changes

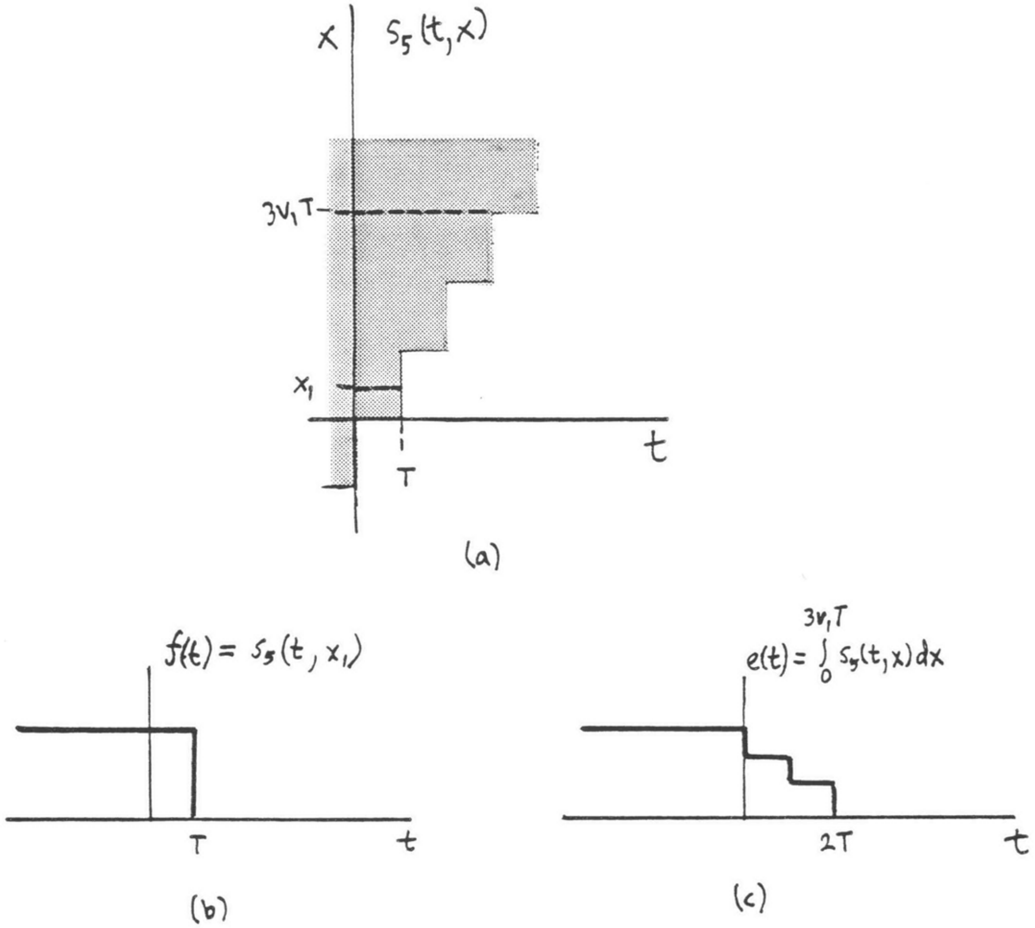


Figure 5.8: Small-area flicker on a moving step during fixation for continuous-time sample-and-hold: (a) perceived image; brightness vs. time (b) at a point x_1, and (c) over a receptive field

in distinct, but adjacent, parts of the retinal image s_5 induces a visual sensation which is indistinguishable from that induced by periodic modulation of a region stationary in s_5.

For an explicit but simplistic model of this effect, suppose that one of the receptive fields covers the region of s_5 in Figure 5.8a from $x = 0$ to $x = 3v_1T$. Uniformly weighted integration over this region results in a temporal signal $e(t)$ containing a number of steps (Figure 5.8c). Now it is evident that $e(t)$ contains some energy at frequency $f_t = 1/T$, which appears as small-area flicker.

It should be noted that, in general, images need not contain sharp edge transitions in the spatial direction or visible multiple image defects for small-area flicker to exist.

5.3 Defects in the Frequency Domain

Having related spatial and temporal image defects to properties of images primarily in the signal domain, it is interesting to review these defects from a frequency-domain standpoint. The image consisting of a rectangular pulse of length $L \gg v_1T$ will be processed with and without presampling and interpolation filters. The filter $h_1(t)$ will be Fd or Fr, and the filter $h_2(t)$ will be Fd or Fr for rate $r = 2$ frame repetition. Then, the resulting spectra for the retinal image $s_5(t, x)$, with and without ocular pursuit, will be examined for features which correspond to subjective image defects. Refer to Figure 5.9 for an outline of the development to follow.

Figure 5.10 shows the camera image $s_2(t, x)$ without and with the presampling filter $h_1(t)$, and their Fourier transforms. The spectrum of the unfiltered image (Figure 5.10c) has a sinc-like cross-section, and the region of support is indicated by a solid line. Since spectral components far beyond the main lobe are necessary to represent a sharp edge, this line extends beyond the pictured region of the (f_t, f_x) plane. The spectrum of the filtered image (Figure 5.10d) is attenuated above $f_t = 1/2T$, as indicated with a dotted line. (This is only an approximation, but it will suffice for explanatory purposes.)

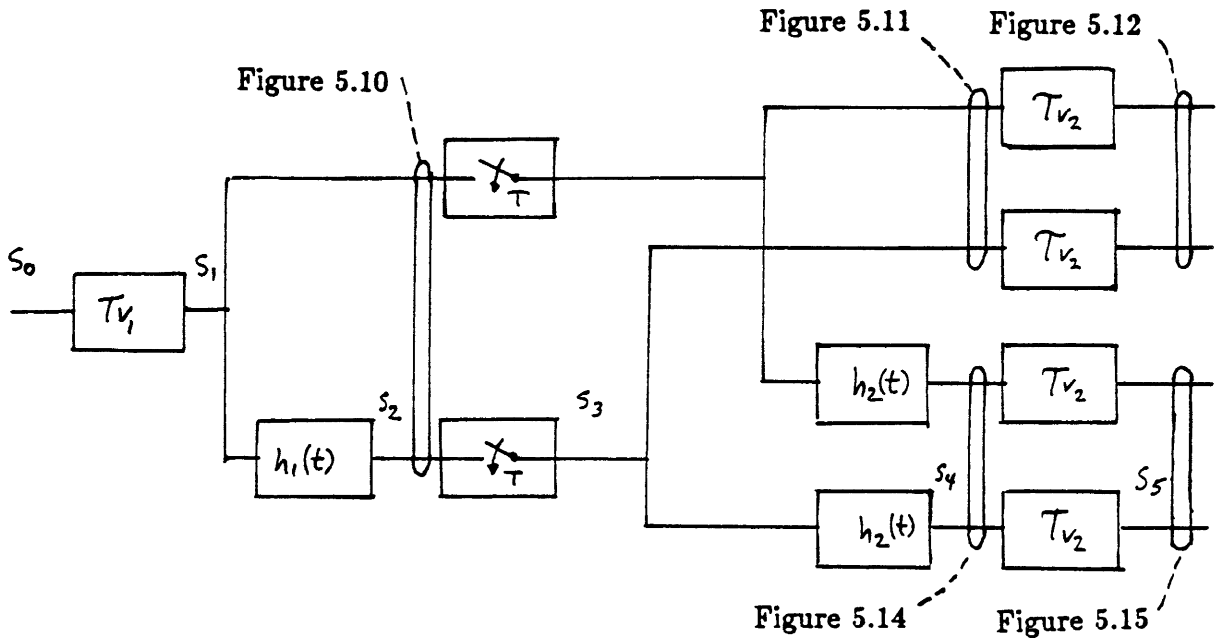


Figure 5.9: Flow diagram for development of image defects; points in this chart are keyed to succeeding figures which show the image at those points

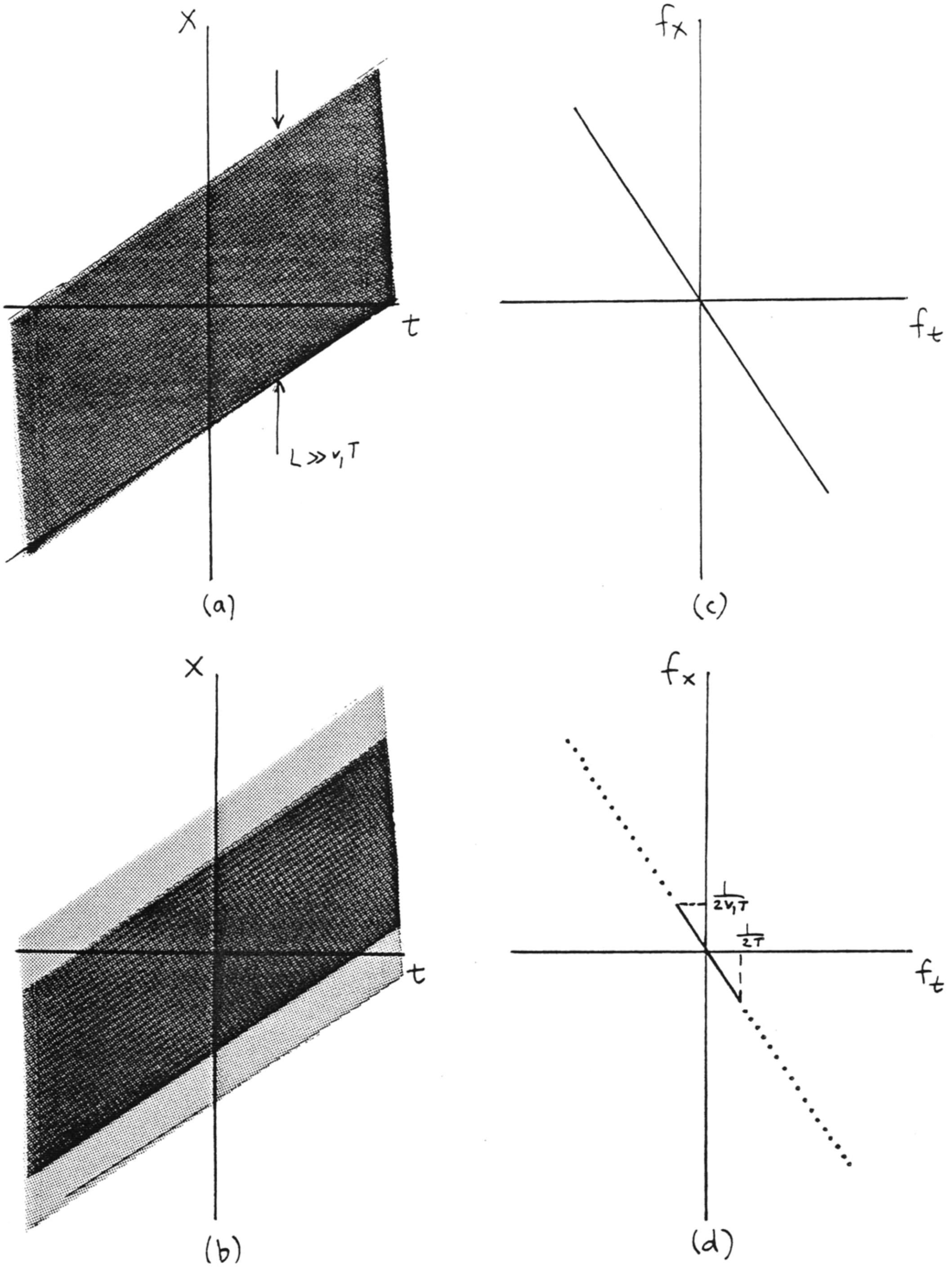


Figure 5.10: Camera image (a) without and (b) with presampling filter $h_1 = \mathrm{Fr}$; (c,d) are the corresponding spectra, where dotted lines indicate attenuation

5.3.1 Fixation and No Up-conversion

When the preceding images are sampled with period T and displayed without frame rate up-conversion, the displayed images $s_3 = s_4$ shown in Figure 5.11 are obtained. In fixated observation, which will be considered first, these are also the same as the retinal images, s_5. Spectral components *in the vicinity* of spatiotemporal frequencies $(f_t, f_x) = (0, k/v_1T)$ for $k =$ integer correspond to a spatially periodic structure with period v_1T, in other words, the multiple image defect. It is not necessary for the Fourier transform to have much amplitude at exactly k/v_1T — indeed, the transform of the original image s_0 could have *zeros* at k/v_1T — because a short-term Fourier transform would smear out any fine structure in the Fourier transform and thus remove the zeros. In Figure 5.11c the spectral amplitude is significant at these frequencies, and it is quite obvious from the signal domain graph that multiple images will occur. In Figure 5.11d the pre-sampling filter has reduced the amplitude at those frequencies so that the defect may be suppressed. Of course, since a rectangular impulse response is not an ideal low-pass filter, it is still possible that enough amplitude may be left to cause multiple images, depending on the original source image s_0.

Supposing that $1/T$ is below the CFF, the presence of spectral components in the vicinity of $f_t = 1/T$, the display and transmission frame rate, indicates that another potential defect will be large-area flicker. The presampling filter cannot affect flicker significantly.

5.3.2 Pursuit and No Up-conversion

The images displayed without frame rate up-conversion can also be observed under ocular pursuit. In this event, the signals of Figure 5.11 are skewed using properties of $\mathcal{T}_{v_2}$ to obtain the retinal images s_5 shown in Figure 5.12. The first thing to notice is that there are no longer any periodic groups of spectral components which lead to multiple images. Second, when a presampling filter is used, the tracked image appears blurred because spatial frequencies above $1/2v_1T$ have been attenuated, but without this filter the full spatial resolution of the original is reproduced. Finally, if $1/T$ is

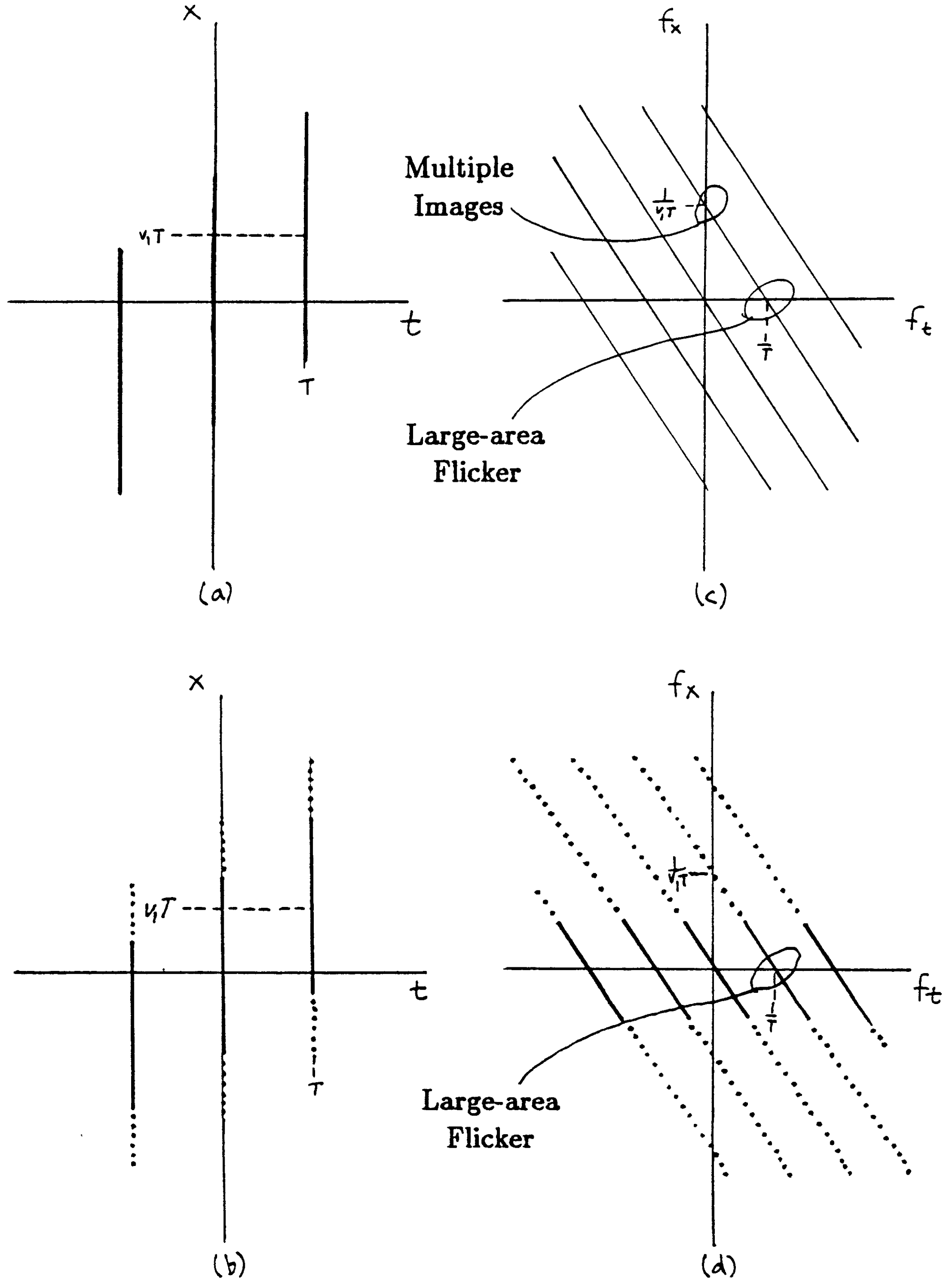


Figure 5.11: Display image, no up-conversion, (a) without and (b) with presampling filter h_1 = Fr; (c,d) are the corresponding spectra

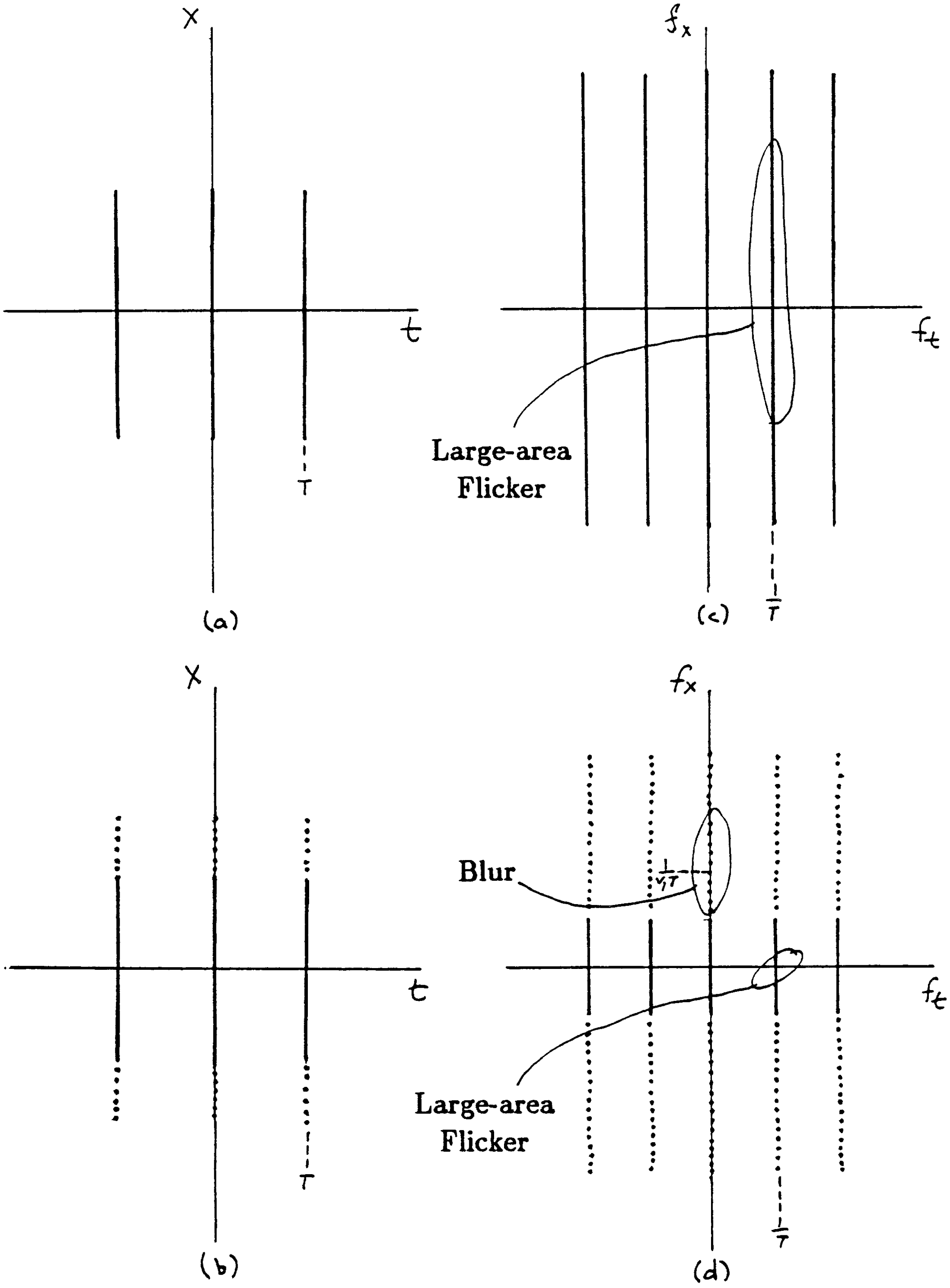


Figure 5.12: Retinal image during pursuit, no up-conversion, (a) without and (b) with presampling filter h_1 = Fr; (c,d) are the corresponding spectra

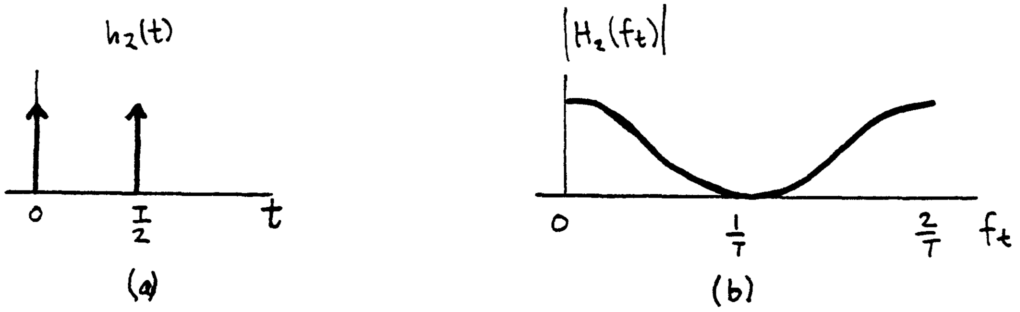


Figure 5.13: (a) Display interpolation filter for $r = 2$ frame repetition; (b) spectrum

below the CFF, the components at $f_t = 1/T$ indicate flicker.

5.3.3 Fixation and Frame Repetition

Now the sampled images of Figure 5.11 will be displayed by frame repetition. The display interpolation filter $h_2(t)$ and its transform appear in Figure 5.13, and the resulting displayed images s_4 are shown in Figure 5.14. These are equal to retinal images, s_5, under fixated observation. Notice that the zero at $f_t = 1/T$ removes signal components at the transmission frame rate. (More generally, if each transmitted frame is displayed r times, zeros will occur at $f_t = \frac{1}{T}, \frac{2}{T}, \ldots, \frac{r-1}{T}$.) If the CFF is just above $1/T$ then improvement over the images displayed without interpolation is expected since large-area flicker should disappear.

On the other hand, there is a possibility of small-area flicker at $f_t = 1/T$, although this is not obvious from the Fourier transform because there is in fact a zero at that frequency. The proper resolution of this discrepancy essentially requires a short-term Fourier transform using a window whose temporal extent is at least several multiples of T and whose spatial extent is the size of a flicker detector's receptive field. Such an analysis would find not a zero but actually a small peak at $(f_t, f_x) = (1/T, 0)$ corresponding to small-area flicker.

Spatial defects are not reduced by the interpolation since the low temporal frequency spatial components are unaffected. In particular, multiple images will be present under the same conditions which cause this defect in the absence of display frame rate up-

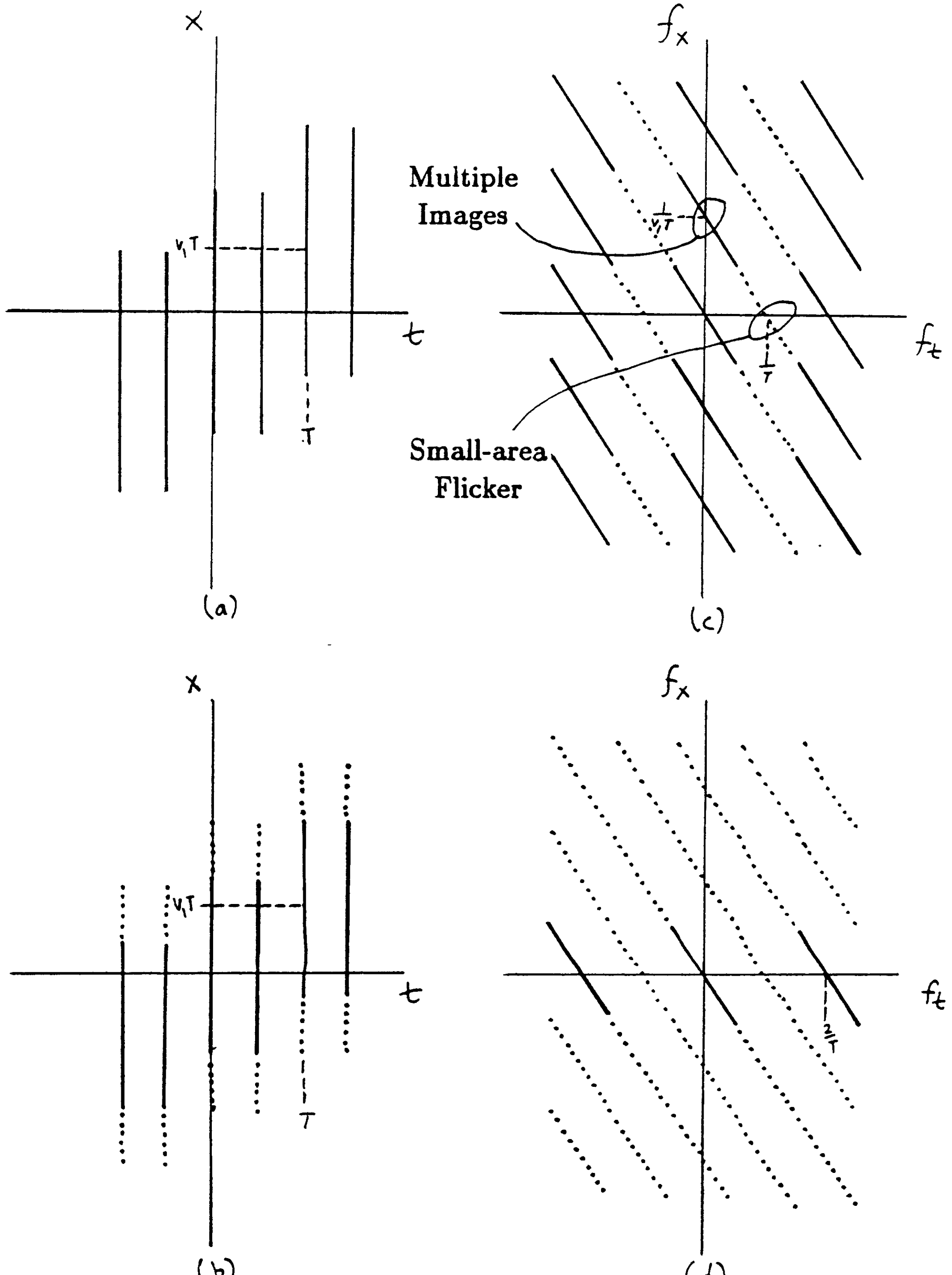


Figure 5.14: Display image, frame repetition, (a) without and (b) with presampling filter h_1 = Fr; (c,d) are the corresponding spectra

conversion.

5.3.4 Pursuit and Frame Repetition

Finally, if the images displayed by frame repetition are observed with ocular pursuit, the retinal images shown in Figure 5.15 are obtained. The image with a presampling filter benefits from frame repetition only by flicker suppression. The image without a presampling filter now exhibits groups of spectral components in the vicinity of $(f_t, f_x) = (0, 2k/v_1T)$ for k = integer, corresponding to a spatially periodic structure with period $v_1T/2$; i.e., the multiple images defect. (More generally, if each transmitted frame is displayed r times, spectral components are grouped around $(f_t, f_x) = (0, rk/v_1T)$.) Furthermore, spectral components around $f_t = 1/T$ correspond to small-area flicker seen near the edges of moving objects.

5.4 Geometrical Defects

All along, this thesis has been ignoring image variations in the y vertical direction and has also been modeling images as if the entire spatial field were sampled concurrently. It is not true that the entire field is either sampled or displayed all at once in a television camera or CRT; e.g., in the HRTV the top line of the display is scanned 1/120 sec earlier than the bottom line. (However, it is closer to being true in motion pictures.) For ordinary broadcast television this deviation from the ideal model has no visible effect, but when computer-generated images are examined on the HRTV, at least two defects are apparent which belie the assumption. To begin with, when large-area flicker is visible a faint downward motion, corresponding to the downward progression of scan, is perceptible.

A more disconcerting artifact, however, is the tilt effect on rapidly moving objects observed under ocular pursuit. In the absence of frame interpolation, a rectangular box translating rightward with sufficient velocity v_1 actually appears skewed in such a way that the top is v_1T to the right of the bottom, as depicted in Figure 5.16. (Under fixation, the box would retain the correct shape, but such a large v_1 would typically

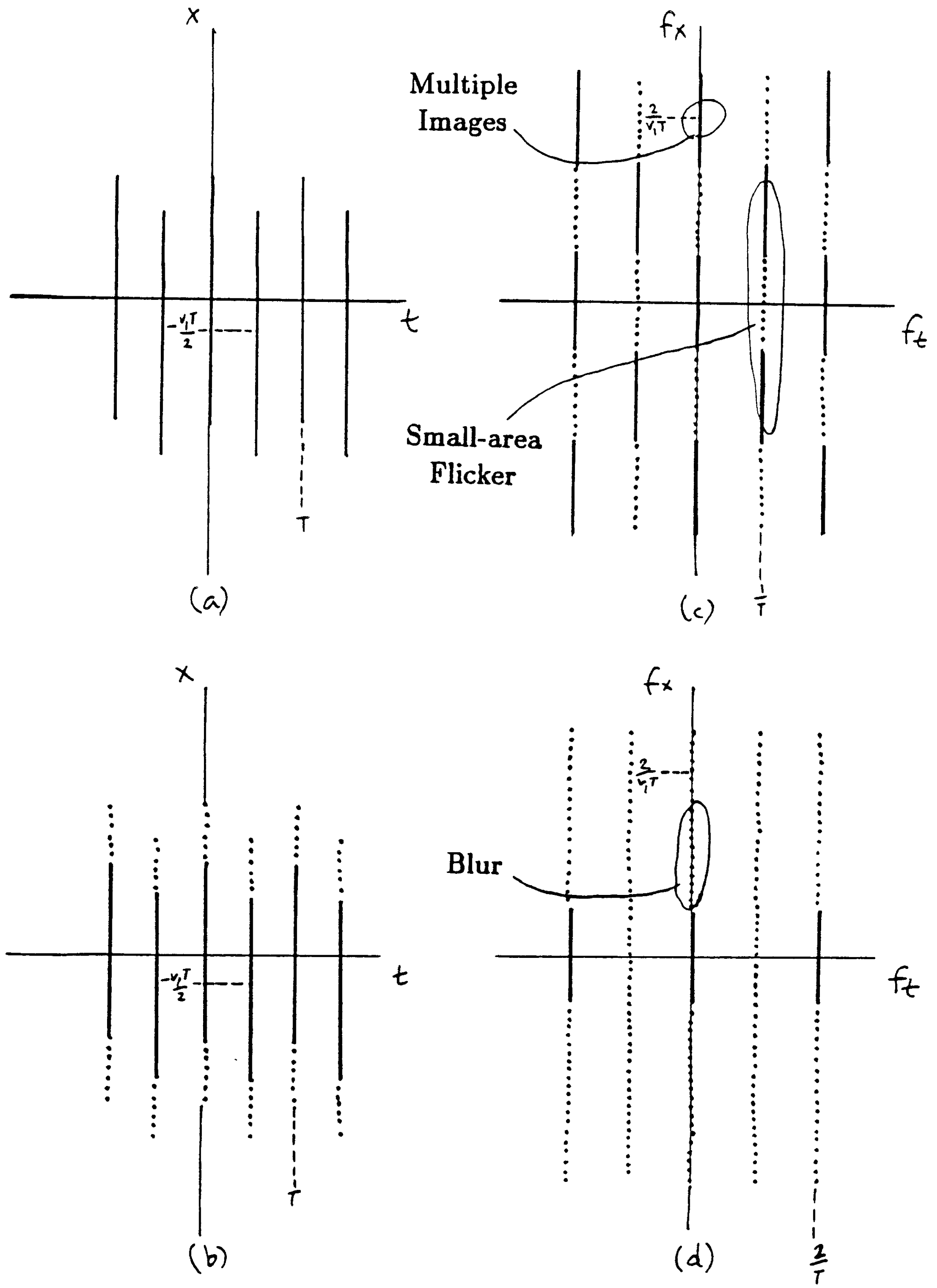


Figure 5.15: Retinal image during pursuit, frame repetition, (a) without and (b) with presampling filter $h_1 = \mathrm{Fr}$; (c,d) are the corresponding spectra

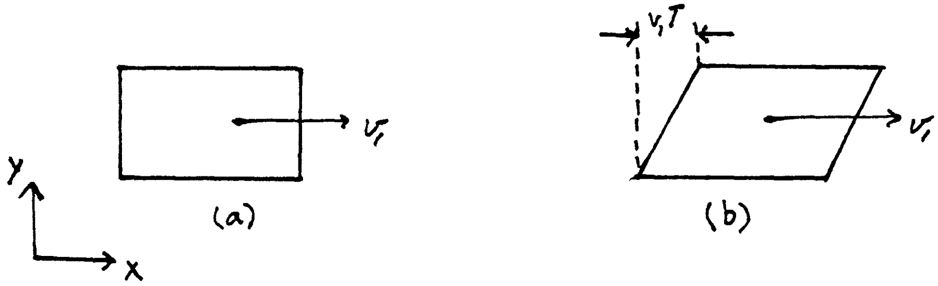


Figure 5.16: Tilt effect during pursuit: (a) original image; (b) perceived image with tilt

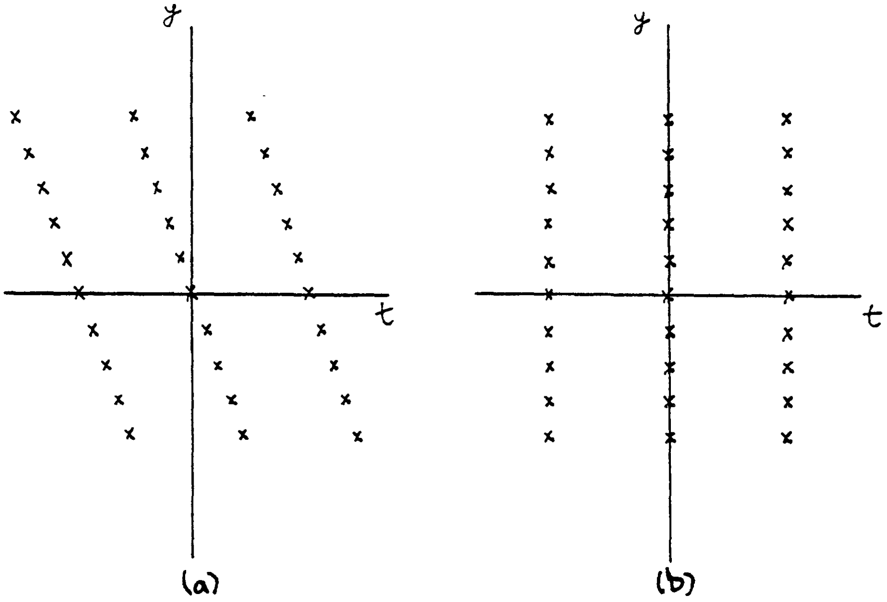


Figure 5.17: (a) Actual and (b) ideal sample points in temporal-vertical plane

induce multiple images, if not blur.)

The reason why tilt is seen is that, although the top and bottom edges are vertically aligned in the displayed image s_4, the eye moves a distance v_1T during the interval between the top and bottom scan lines. Ignoring the time necessary to scan horizontally, the sample points in the temporal-vertical plane for actual and idealized scanning are shown in Figure 5.17. In general, when samples are taken on one grid but displayed on a different one, geometrical distortion will result. This statement is obviously true when a xy sampling grid is skewed spatially. Here, the temporal-vertical skew imparted by the discrepancy between sampling grids is reflected by the translation operator $\mathcal{T}_{v_2}$ into an apparent horizontal-vertical skew.

The tilt effect does not exist under ocular pursuit when a CRT is fed directly from a television camera because then the sampling grids of image source and display would be identical. Under fixation, tilt with the opposite direction of slant would exist but is not likely to be noticeable using ordinary cameras due to motion blur from h_1. Furthermore, seldom in natural images are there rapidly moving, right-angled objects which would obviously look wrong to the observer when skewed away from the vertical.

5.5 Velocity Defects

An image is said to possess a velocity defect when the observed velocity departs from the correct or expected velocity. This class of defects includes contrary motion and jerkiness. Contrary motion will not receive attention in the experimental observations of the next chapter since it is already a well understood form of temporal aliasing, but it will be presented here because the frequency-domain analysis is rewarding. A discussion of jerkiness, on the other hand, is vitally important to this thesis because this defect is often considered to be the crux of defective motion rendition.

5.5.1 Contrary Motion

The "stroboscopic" contrary motion defect exists when the observed velocity of a translating, spatially periodic image is ambiguous and may depend on the manner of observation, or when this observed velocity is contrary to that determined from other visual cues.

The prototypical situation exhibiting this defect is the image $s_1(t,x) = 1+\cos 2\pi(x-v_1t)/x_0$, a grating translating at velocity v_1, temporally sampled at rate $1/T$ with no presampling filter and then displayed with no interpolation. The samples $s_3(nT,x) = 1+\cos 2\pi(x-v_1nT)/x_0$ do not uniquely determine v_1 since the same samples could be produced with any velocity $v_1' = v_1 + kx_0/T$ for $k =$ integer.

The first case to consider is when the grating moves within a stationary window: either the entire display, as in interlace line crawl, or a small region, as on a rotating spoked wheel. Given the appropriate initial stimulus for initiating ocular pursuit, it is often possible to track at several different alias velocities, each resulting in a stationary grating on the retina. While fixating, only the alias velocity with minimum $|v_1'|$ (not necessarily for $k=0$) is usually visible; e.g., when $x_0/2T < v_1 < x_0/T$ the lowest alias velocity is $v_1 - x_0/T$, which is in the opposite direction of motion.

A simple interpretation in terms of aliased spectral components is shown in Figure 5.18. The transform of $s_3(t,x)$ contains a pair of δ-functions at $f_x = \pm 1/x_0$ and one δ-function at $f_x = 0$. The dashed lines of Figure 5.18a group together the baseband components, first order harmonics, etc.

It is evident that when $x_0/2T < v_1 < x_0/T$, the temporal frequency of the order ± 1 harmonics is lower than that of the "baseband" (Figure 5.18b). Under fixation, the HVS tends to associate the three circled frequencies as a group; hence, the velocity $v_1 - x_0/T$ is perceived. Moreover, with various choices of pursuit velocity v_2, any one of various different groups of frequencies can be shifted to zero temporal frequency, as in Figures 5.18c,d.

The second case of contrary motion is where the spatially periodic pattern is a static feature of an object translating at velocity v_1. If pursuit is attempted by the observer, only one pursuit velocity, $v_2 = -v_1$, can be accomplished and there is no velocity ambiguity. For fixation, an alias velocity $v_1' = v_1 + kx_0/T$ might be perceived if v_1' is close to 0, and a visual paradox occurs: the object is seen to move at v_1 but the pattern on the object looks (nearly) stationary. Other alias velocities cannot be

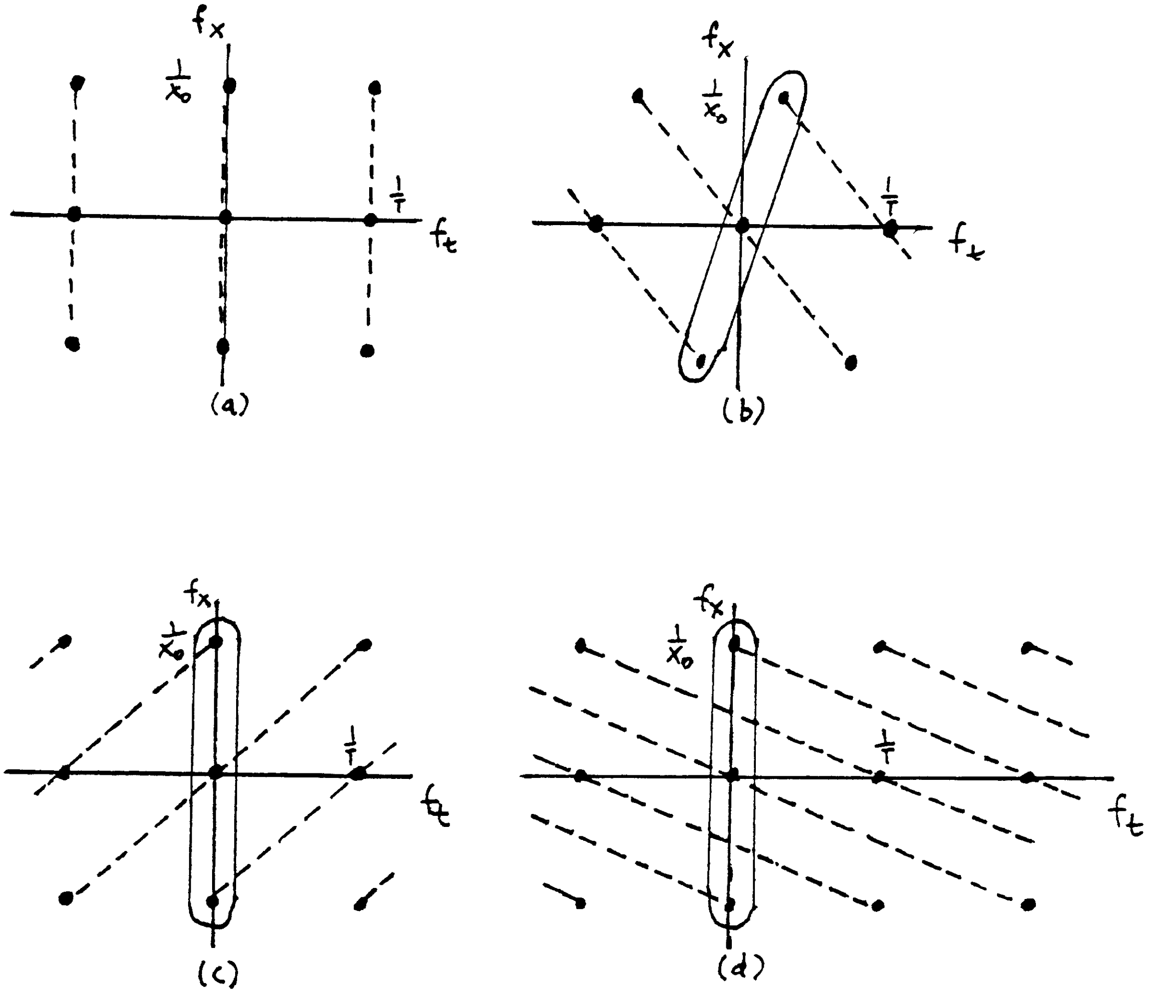


Figure 5.18: Spectra of contrary motion defects: (a) fixation when $v_1 = 0$; (b) fixation when $v_1 = \frac{3}{4}x_0/T$; (c) pursuit with $v_2 = -v_1 - x_0/T$; (d) pursuit with $v_2 = -v_1 + 2x_0/T$

detected easily because the object velocity dominates the perceived velocity field.

5.5.2 Jerkiness

Whereas contrary motion causes a gross error in the observed mean velocity, the jerkiness defect involves only deviations about the mean. Jerkiness is the perception of temporal fluctuations in velocity.

At first glance, the preceding definition merely restates one's common-sense notion of "jerkiness." However, not every subjective defect which is commonly considered to be "jerkiness" can be ascribed to a direct sensation of velocity fluctuations. More often than not, "jerkiness" is inferred from the presence of other defects which are associated with temporally sampled motion. On closer inspection, the motion itself typically appears smooth even when spatial or temporal image defects exist.

This specialized definition of jerkiness is necessary in order to eliminate overlap with the definitions for multiple images and small-area flicker. A multiple images defect alone is clearly not a case of jerkiness because it contains no temporal fluctuations. Small-area flicker and jerkiness are distinct defects because they concern temporal fluctuation in brightness and velocity, respectively. During an actual observation, the two kinds of fluctuation are sometimes hard to tell apart; nevertheless, they are at least conceptually different defects.

For an example of jerkiness under ocular pursuit, consider again the retinal image s_5 of Figure 5.7. When T and v_1T are sufficiently large the edges no longer exhibit flicker but instead appear to oscillate back and forth with temporal period T. With a more realistic object which contains details on its surface, not just the edges but also the entire object might appear to be oscillating.

For jerkiness under fixation, the observed velocity fluctuations correspond to discrete jumps of distance v_1T occurring with period T. Presumably the reason jerkiness is perceived is that the low-frequency components of the displacement vs. time function are large enough in magnitude to pass through the temporal filter in the velocity pathway of the visual system, as hypothesized in Section 2.3.2.

So far it has been customary to relate defects to features in the image spectrum;

however, it does not make sense to do this for jerkiness since the defect does not directly involve spatiotemporal patterns of brightness.

Jerkiness is often singled out as the primary motion-induced defect, but the observations of images using the HRTV have suggested otherwise. They have found that before v_1 and T become large enough to induce the direct perception of temporal velocity fluctuations, small-area flicker and multiple images have already become very prominent defects. This finding is significant for the following reason. Chapter 2 and this section have tried to emphasize that there are distinct pathways in the visual system for spatiotemporal brightness information and velocity, and that a transmission system should provide a reproduced image which is adequate for all pathways. Since jerkiness seems to appear only after spatial and temporal classes of defects have appeared, we may focus our attention on the spatiotemporal pathway. Properties of apparent motion need not be of concern. In particular, we are more or less justified in following the brightness vs. space-time theory of motion rendition which is espoused by Watson [WATS 83A] and others, and which has been the continuing theme throughout most of this thesis.

To sum up, directly perceived temporal fluctuations in velocity are not the principal degradations in temporally-sampled images. Consequently, only in the colloquial sense of the word can "jerkiness" be equated with defective motion rendition.

5.6 Conclusions

Image defects arising under two modes of eye movements, pursuit and fixation, have been analyzed in this chapter. A third common situation should be mentioned, however. During pursuit of a moving target the eye might be swept past a stationary object on the display. When the display is not temporally continuous, ocular-motion-induced defects can appear on the stationary object. With a few assumptions, though, the retinal image of the stationary object ($v_1 = 0, v_2 \neq 0$) is identical to the one obtained by fixating a certain moving image with related parameters. Therefore, this case does not have to be examined separately in experimental observations. A sufficient condition for equivalence is that the interpolation filter $h_2(t)$ performs rate r up-conversion, is made up of only δ-functions, and has a periodic extension $\tilde{h}_2(t)$ (Section 5.2.1) that consists of δ-function pulses all of equal height. For example, the four filters Fd, Fr, Ft, and Fu satisfy this condition. Then, to produce the corresponding moving image to be viewed under fixation, the original stationary image s_0 should be translated at velocity v_2, sampled at rate r/T, and displayed without interpolation. If r/T is above the CFF, which is highly likely, the only possible defect will be multiple images, spaced v_2T/r apart.

In this chapter a consistent classification of motion-induced image defects was developed based on observations of images on the HRTV system. Qualitative descriptions of multiple images, blur, large-area flicker, small-area flicker, tilt, contrary motion, and jerkiness were provided. Then, for most of the defects, the signal domain and frequency domain descriptions were correlated with subjective appearances. The classification is especially convenient for the next chapter, where the dependence of defects on variation of image parameters will be investigated.

Chapter 6

Dependence of Defects on Major Parameters

The preceding chapter defined, and described the subjective appearance of, a number of motion-induced defects which can be observed under the appropriate conditions. The logical sequel, of course, is to empirically determine what those conditions are by means of systematic observation of test images, guided by theoretical predictions.

This chapter explores the effects and interactions of velocity, transmission frame rate, presampling filter, and interpolation filter on the spatial and temporal classes of image defects. Except for velocity, these are the key temporal parameters of an image transmission system over which the designer has direct control. To begin with, the procedures for observing and recording data will be explained. Then, for each type of image degradation, a hypothesis about the effects of the four parameters will be stated and subsequently compared with the results of the observations. Interpretations and further analysis of the results will follow when appropriate.

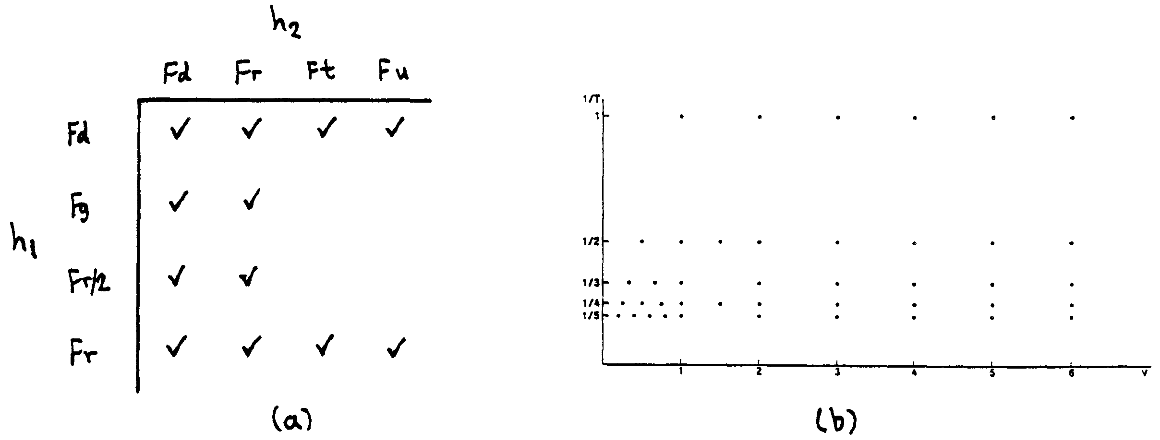
6.1 Observation Procedures

There is certainly more to an image defect than merely absence or presence, and no doubt it would be better to measure the subjective magnitude of a perceived degradation. For simplicity, though, this thesis disregards the problem of scaling suprathreshold visual responses. Verbal comments regarding the subjective appearance and strength of defects will be included from time to time, but the systematically recorded data indicates only whether the degradation is "detected," i.e., above an arbitrary threshold. In the experiment on blur, a specific reference image was chosen to establish a physical threshold, but for the other defects the threshold was set by the observer on an internal psychological scale, based upon prior experience with the range of image variation.

The physical dimensions of the HRTV and standard viewing conditions have been specified. Thus, throughout most of this chapter and the next, the quantities of luminance, position, and time will be reckoned in more convenient, normalized units rather than physical units. The relationships with physical units are 1 pel = 0.014°, 1 frame = 1/120 Hz = 8.3 ms, 1 pef (pels/frame) = 1.7°/sec, and 255 luminance units = 16 ft-L. The display frame period in these units is defined to be 1 so that the transmission frame period T equals the up-sampling factor r.

Of the four variables of interest in this chapter, only two are scalar quantities, v and $1/T$. This suggests that a convenient way to diagram the dependence of an image defect on these two parameters is to draw the boundary between the below- and above-threshold regions of the velocity/frame-rate plane. These boundaries will be referred to as threshold boundary curves. Since defects generally increase with greater v and smaller $1/T$, it is usually clear which side of the boundary is below threshold and which is above; therefore, no explicit indication is necessary. Boundaries will be drawn with vertical and horizontal segments halfway between experimental sample points to emphasize that the "true" threshold location is unknown. Finally, velocity/frame-rate maps for different defects, combinations of filters, or eye movements can be superimposed for comparison. Theoretical predictions and observation data alike can be presented in the same fashion.

The two nonvarying parameters of this series of observations are the time-averaged uniform background luminance of 5 units, and the moving target which is a rectangular box 64 pels wide and 64 units in average luminance. The twelve pairs of filters h_1, h_2 tabulated in Figure 6.1a are selected from the set of filters introduced in Chapter 4. The remaining two parameters, v and $1/T$, are jointly chosen from the sample points



		h_2			
		Fd	Fr	Ft	Fu
h_1	Fd	✓	✓	✓	✓
	Fg	✓	✓		
	Fr/2	✓	✓		
	Fr	✓	✓	✓	✓

Figure 6.1: Levels of the variable factors in the observations: (a) pairs of filters h_1, h_2; (b) $v, 1/T$ sample points in velocity/frame-rate plane

plotted in the map of Figure 6.1b. At these points T is an integral number of frames and vT is an integral number of pels, as required by the HRTV system. Exception: when h_2 = Fd, the frame period $T = 5$ cannot be examined because the necessary peak luminance level of the object, $T \cdot 64 = 320$, is outside the range 0-255 and cannot be displayed.

Preliminary surveys were conducted first, in order to identify ranges of combinations of h_1, h_2, v, and T which would be of interest for each individual type of defect. For all twelve combinations of h_1, h_2, the image corresponding to each sample point in Figure 6.1b was examined and every defect perceived was recorded. After this was completed, observations focused on individual defects. Two procedures were employed to collect the data reported in this chapter, randomized presentation and free observation.

Randomized presentation, used only to study blur, resembles the "haphazardly-ordered method of limits" [GUIL 54]. The reference image, produced with $v = 1.5$, $T = 2$, $h_1 = h_2 =$ Fr for a barely perceptible amount of blur, was compared to test images in a randomized sequence. Blur was recorded whenever the blur of the test image was equal to or greater than the blur of the reference. In the first and second viewing sessions, every test image of interest was displayed with the reference and was examined as long as necessary to arrive at a decision. After this was done, the test images which were judged differently from one session to another were examined again in a tie-breaking session. During this third session, each image was displayed two times at random. If an image was judged differently in the two presentations, the tie was retained.

For defects other than blur it was felt that standard psychophysical methods were not appropriate, mainly because a suitable threshold reference was not definable. Free observation is essentially the same as the survey procedure except that only the interesting ranges were examined and attention was directed to one defect at a time. The number of interesting images for any one defect was small enough to allow viewing all of them during one session. The time spent per image was not limited, and pairwise comparisons between images of different parameters were unrestricted in order to promote a more uniform psychological threshold.

6.2 Blur

Blur perceived under fixated observation is not considered in this study. It is not of practical importance because the component of blur caused by motion in the retinal image exists even when viewing the original scene directly and is independent of the image transmission itself. Moreover, the component of blur caused by the camera can be studied better under ocular pursuit with h_2 set to Fd.

Predictions. In the velocity/frame-rate plane, the threshold boundary curve for blur under ocular pursuit is predicted to be a straight line through the origin (Figure 6.2). For a fixed choice of presampling and interpolation filter, the spatial extents of both $h'_{1x}(x)$ and $h'_{2x}(x)$ are roughly proportional to vT, so that blur should be perceived when vT is greater than some constant β. Of course, $h'_{2x}(x)$ sometimes causes multiple images rather than blur at higher velocities. In this event, the region of the map where blur exists is additionally bounded on the right by the multiple images threshold, which is not shown here.

The position of the threshold, i.e., the value of β, should depend on the types of

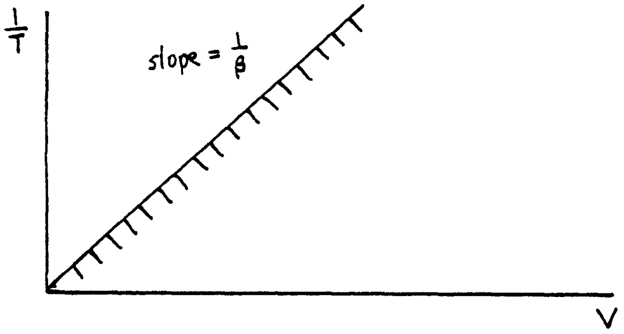


Figure 6.2: Predicted threshold for blur

filters. If $h_1 = h_2 =$ Fd then there is no reason to have any blur. As h_1 and/or h_2 are changed to successively wider types of filters, the boundary should swing towards the left.

Results. In informal observations, the perceived amount of blur was, without exception, a non-decreasing function of both velocity v and frame period T when other parameters were held constant.

The blur thresholds obtained experimentally by comparison with a reference image are found to follow straight-line trends, as far as one can tell from the sparse grid of sample points, with β values ranging from 2 to 4 (Figure 6.3). Notable deviations from this trend are the curves for $h_2 =$ Fr or Ft, whose slopes flatten out between $T = 2$ and $T = 1$. The explanation for this behavior is that when $T = 1$, no display up-conversion takes place and these interpolation filters degenerate into Fd. Thus the spatial extent of $h'_{2x}(x)$ is 0, which is much narrower than would be the case if its width were truly proportional to vT.

With $h_1 = h_2 =$ Fd no blur can be detected at all, as expected. As h_1 is varied through the progression of filters Fd, Fr/2, Fg, and Fr, the threshold curves shift leftward (Figures 6.3a,b), though more dramatically when $h_2 =$ Fd than when $h_2 =$ Fr since the latter additional source of blur partly obscures the effect of h_1. As far as blur thresholds are concerned, the presampling filters Fr and Fg appear to be essentially equivalent. The same can be said for Fr/2 and Fd when $h_2 =$ Fr. As h_2 is varied through Fd, Fr, Ft, and Fu, the thresholds also move leftward (Figures 6.3c,d). As

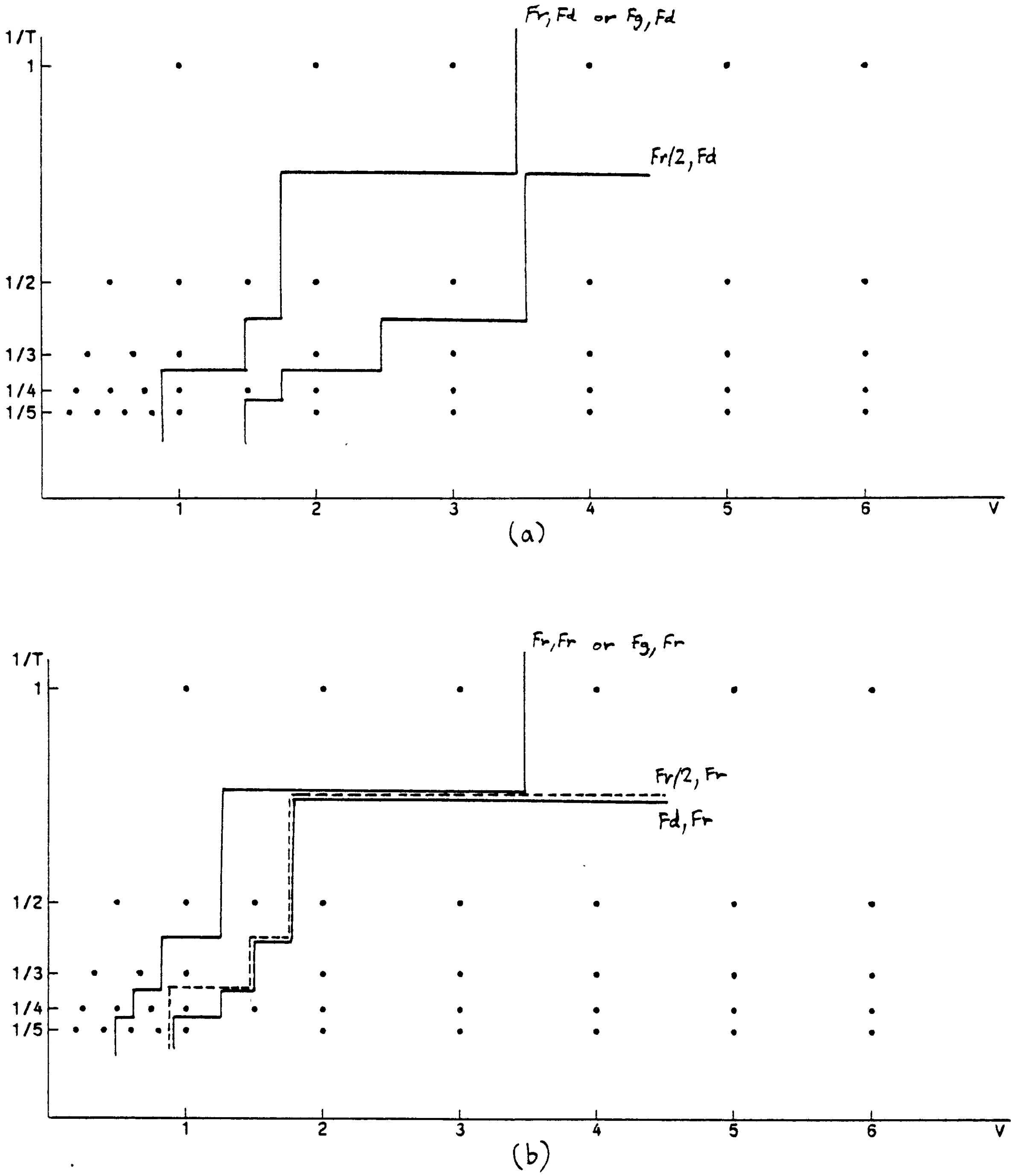


Figure 6.3: Measured thresholds for blur plotted two ways; filters are denoted by ordered pair h_1, h_2. Vary h_1, fixing (a) $h_2 =$ Fd, (b) $h_2 =$ Fr

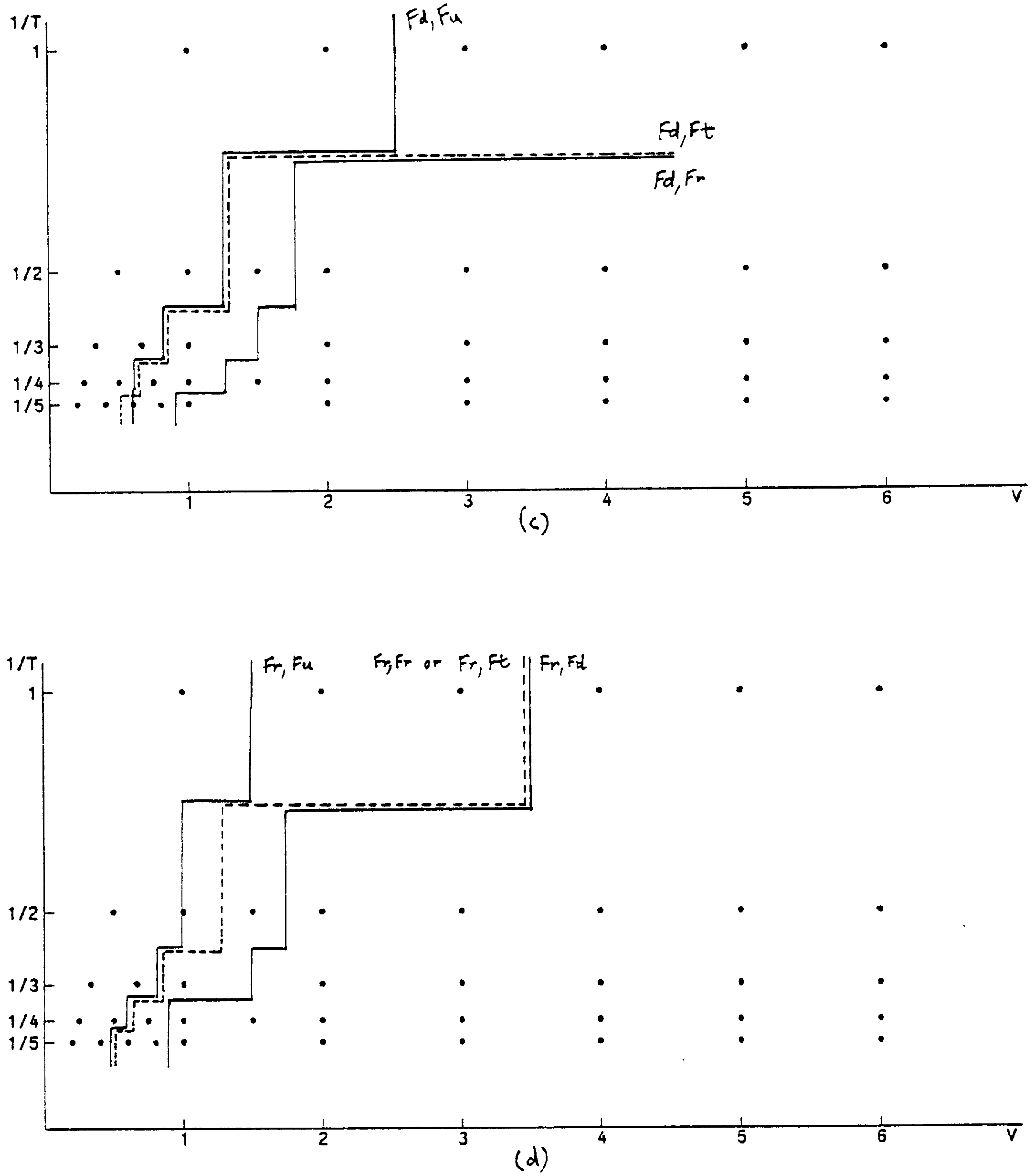


Figure 6.3, continued: Vary h_2, fixing (c) h_1 = Fd, (d) h_1 = Fr

required for self-consistency, the curves for h_2 = Fd, Fr, or Ft merge when $T = 1$, because these filters all become Fd. When $T > 1$, the display interpolation filters Ft and Fu are nearly equivalent for blur, and Fr is equivalent to both of these when $h_1 =$ Fr.

6.2.1 Further Analysis

Because the test image is devoid of detail, except for the leading and trailing edges of the object, it is reasonable that the observed amount of blur is related to the steepness of the edge transition in the retinal image s_5. Among transitions between two fixed brightness levels, the risetime of the edge should be an index for comparison.

The risetime can be defined as the distance for the signal to rise between 10% and 90% of the full amplitude swing. For example, the reference image moving at velocity 1.5, observed under ocular pursuit, is filtered by the effective impulse responses $h'_{1x}(x)$ and $h'_{2x}(x)$ of Figure 6.4. The discrete filter which the HRTV transmission system simulation program uses to approximate a continuous-time presampling filter is depicted. The sharp step in the original image s_1 becomes a tapered transition in s_5, with a risetime of 3.5 pels. (The small steps in s_5 are close enough together that only blur, not multiple images, is perceived.)

For every image observed in the blur experiment, the risetime was computed using the method above, and it was found that risetime values near the blur threshold are mostly in the range from 2 to 3.5. There was no noticeable change of threshold risetime with velocity, up to the highest velocity recorded for any blur threshold in this experiment. This indicates that visual acuity was not significantly impaired by ocular pursuit.

It thus appears that risetime can be a suitable index of blur for one particular test image; however, in order to compare the blurs of arbitrary-size steps, it would be necessary to account for the step contrast and mean brightness. Isono does this by defining an image sharpness metric based on the "perceived edge gradient" of a step after passing through a nonlinear excitatory-inhibitory model of the visual system [ISON 84]. The sharpness is calculated by $S = \Delta V / \Delta X$ where ΔV is the amplitude

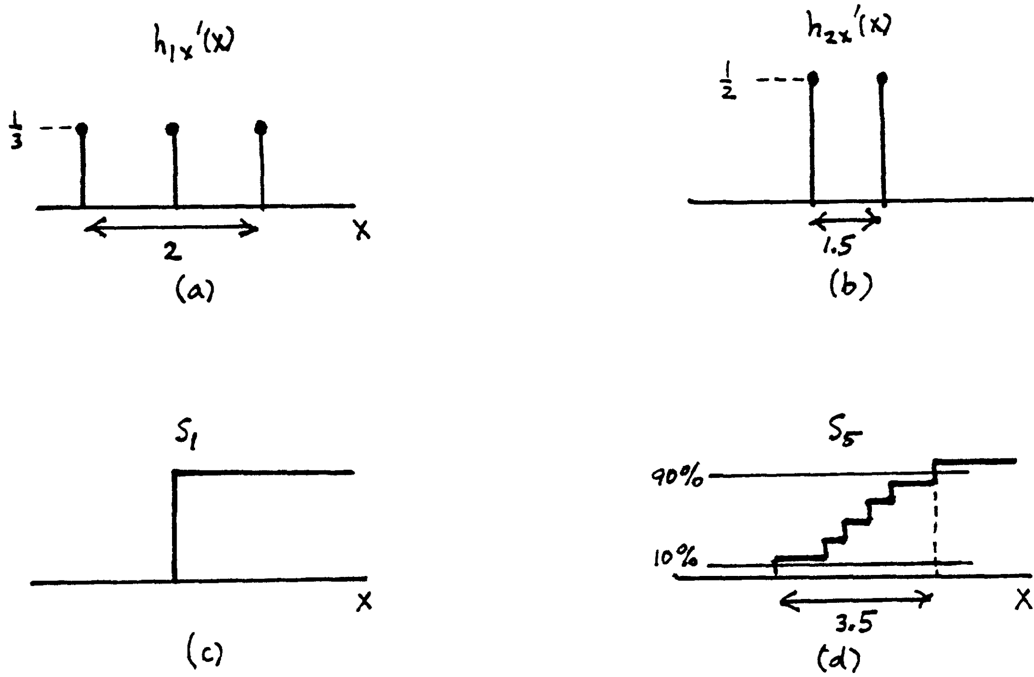


Figure 6.4: Risetime of a blurred edge: (a) reflected presampling filter; (b) reflected interpolation filter; (c) original step; (d) filtered step showing risetime of 3.5 pels

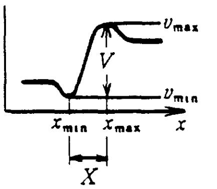


Figure 6.5: Sharpness measured by perceived edge gradient method (from [ISON 84])

difference between the Mach band overshoot and undershoot, and ΔX is the spatial distance between these extrema (Figure 6.5).

Unfortunately, it is not possible to apply this visual model and sharpness metric to the test images of the observations, for even though explicit formulas and parameter values for the visual model are provided by Isono, the published information contains many inconsistencies.

6.3 Multiple Images during Pursuit

Predictions. Since the perception of multiple images requires the visual system to detect a spatially periodic pattern, a first-order prediction can be made by assuming that the defect will be perceived whenever that period exceeds some threshold distance α. Under ocular pursuit, the distance between peaks in the reflected filter $h'_{2x}(x)$ must be v since $h_2(t)$ is a train of pulses spaced 1 time unit apart. By this reasoning, the threshold curve should be a vertical line at velocity α (Figure 6.6). But the region of defects stops short of $T = 1$, for at that transmission frame rate no display interpolation takes place ($h_2 =$ Fu is an exception, as always).

One boundary curve ought to hold for all filters h_1, h_2 which permit multiple images to exist. If $h_2 =$ Fd then no display up-conversion takes place and thus no defect is expected. All other choices of h_2 should behave identically and result in the curve shown as long as $h_1 =$ Fd. If $h_1 \neq$ Fd then there should be sufficient camera blur to conceal the multiple images, the reason being that among $h_1 =$ Fr/2, Fg, or Fr, the narrowest type of filter is Fr/2, whose reflection $h'_{1x}(x)$ has a spatial extent of $vT/2$. Since $T \geq 2$ is necessary for multiple images anyway, the spatial extent of $h'_{1x}(x)$ will be at least v. Thus, it should effectively interpolate between the peaks of $h'_{2x}(x)$ and suppress the defect.

Results. Once systematic observations of the multiple images defect under ocular pursuit began, it became obvious immediately that the visibility of the defect depends upon much more than merely the spacing between image repetitions. Not only does the visibility appear to increase with v, it also depends on T, and, moreover, the latter

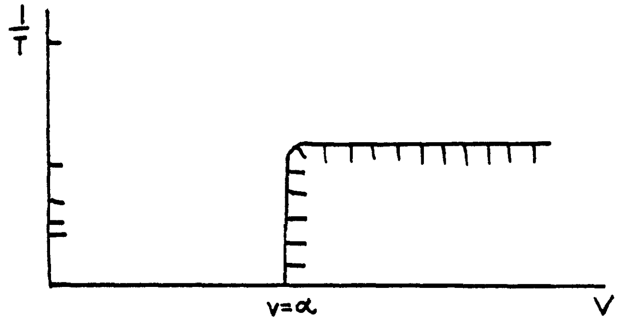


Figure 6.6: Predicted threshold for multiple images during pursuit

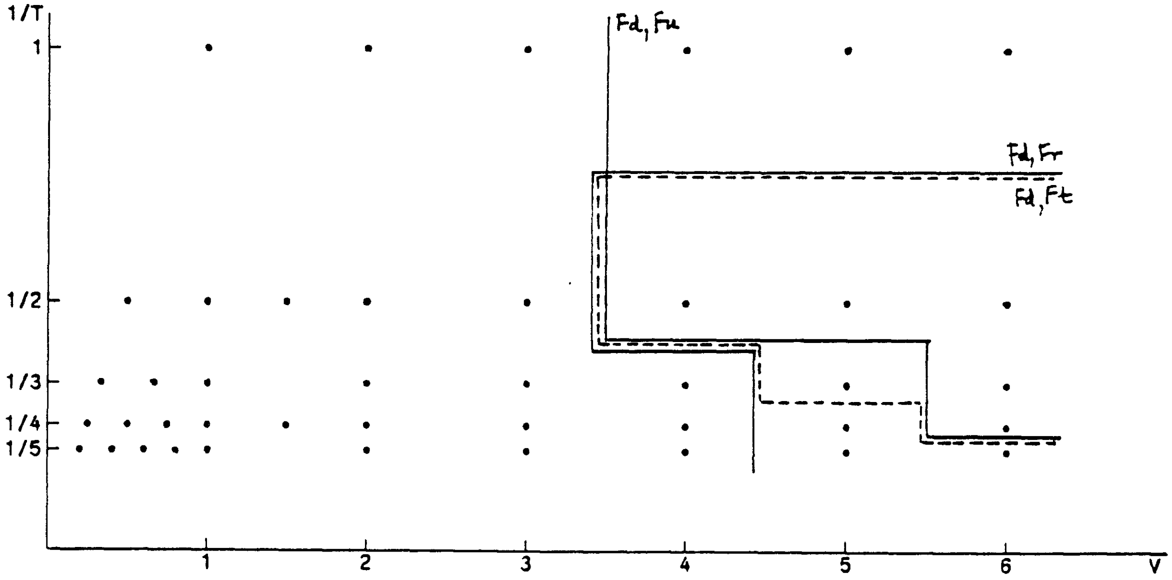


Figure 6.7: Measured thresholds for multiple images during pursuit

is an inverse dependence. This finding does not imply that lower transmission frame rates are more desirable; a more satisfactory interpretation is that higher up-sampling factors are preferable. Subjectively, as the filter $h'_{2x}(x)$ increases in width, holding v constant, the distinctness of the multiple images fades. Finally, among the individual steps of a multiple images defect when the number of steps is large, the steps nearer the dark background are more distinct, in accordance with Weber's law.

The foregoing comments on suprathreshold visibility are complemented by the experimentally derived threshold curves (Figure 6.7). As the transmission frame period T increases, a greater inter-image separation is required for the defect to remain distinguishable from blur. The threshold velocity "constant" α ranges from 4 to above 6 pef.

Even though the form of the boundary curve was predicted incorrectly, the hypothesized suppression of multiple images by $h_2 = \text{Fd}$ or by $h_1 \neq \text{Fd}$ was verified. With $h_1 = \text{Fd}$ fixed, when h_2 is varied through the progression Fr, Ft, and Fu, the lower part of the threshold curve shifts to the right, apparently because of decreasing contrast.

6.3.1 Further Analysis

As the number of multiple images increases, the contrast between adjacent steps must decrease since the total luminance difference between object and background is constant. On the one hand, more multiple images should increase visibility of the periodic pattern. On the other hand, the decreasing contrast per step is a simultaneously opposing effect which reduces visibility. The observations show that decreasing contrast is the dominant effect.

That being the case, the negative slope of the threshold boundary curve in the velocity/frame-rate map can be explained readily. Since the threshold contrast sensitivity of the HVS to spatial gratings increases with period (up to a point), decreasing contrast levels necessitate larger periods for detection; i.e., smaller frame rate $1/T$ demands greater v.

The three factors discussed so far which affect visibility of multiple images can now be summarized in Figure 6.8. A fourth factor which cannot be neglected is the apparent

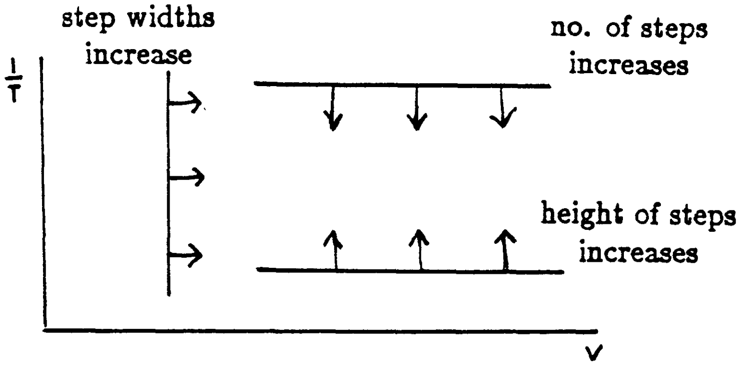


Figure 6.8: Factors affecting multiple images during pursuit are step width, number of steps, and contrast of steps

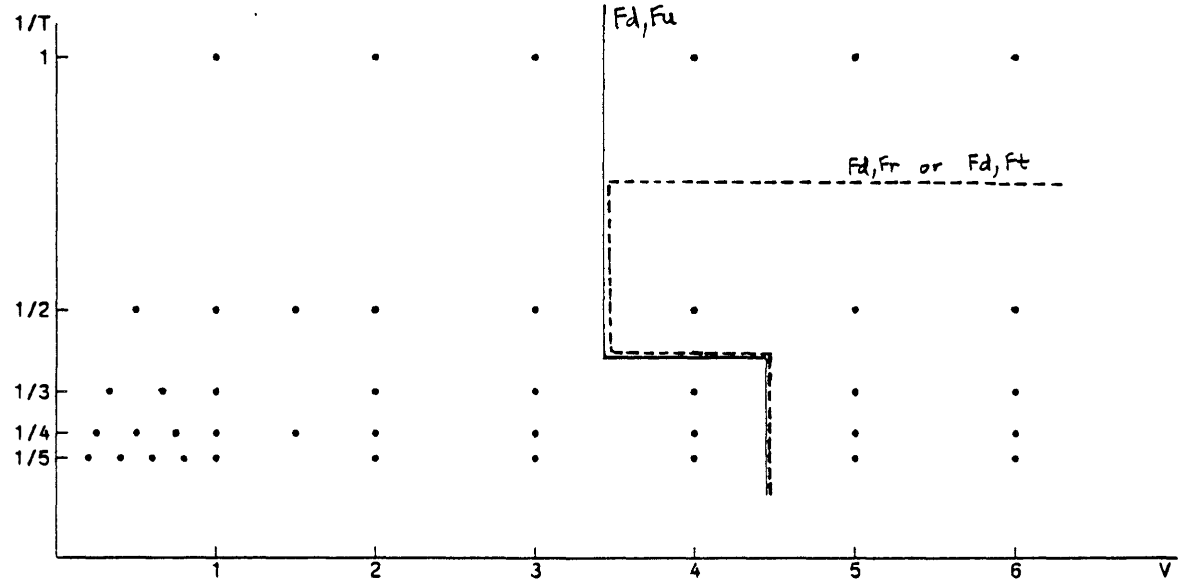


Figure 6.9: Measured thresholds for multiple images during simulated ideal pursuit

sharpness of the individual images or steps, as controlled by the characteristics of the original object, camera blur, and dynamic acuity. If the camera blur of $h_1 = \mathrm{Fr}/2$, Fg, or Fr is sufficient to eliminate the defect, then softening the abrupt edges of the original object should also have the same mitigating effect on multiple images.

While greater v does increase the inter-image separation, it also shortens the duration of presentation of the moving target by the HRTV display. Short presentation alone potentially reduces the ability of the HVS to detect spatial patterns; furthermore, short duration and greater velocity make accurate pursuit more difficult to achieve, and the resulting retinal blur would reduce the apparent visibility of the defect. Indeed, it was often the case that multiple images were recorded as being detected even though they could be seen only part of the time in a group of consecutive trials, especially when $v = 6$ and $T \geq 3$. Because practice and attentiveness seemed to improve detection, the random fluctuation in visibility is most likely attributable to chance errors in pursuit.

Difficulty in pursuit is an undesirable variable when examining the intrinsic visibility of the multiple images defect. It is possible to simulate long duration perfect pursuit by performing the translation operator $\mathcal{T}_{v_2}$ within the HRTV system, in effect presenting s_5 on the display instead of s_4. The average velocity in such a display is zero and can be observed under fixation for a simulation of ideal pursuit. Moreover, the image can be displayed for the entire duration of the HRTV memory.

This simulation was performed, yielding the multiple images defects thresholds of Figure 6.9. These curves fall to the left of those obtained in real ocular pursuit when $T \geq 4$ and $h_2 =$ Ft or Fu, confirming the suspicion that retinal blur or shortened presentation is reducing the apparent defect visibility.

The following hypothesis might explain why the velocity threshold α for $T \leq 3$ or $h_2 =$ Fr does not increase when real pursuit is substituted for ideal pursuit. It is likely that a blurred or less contrasty target is an inherently more difficult stimulus to track. Because the multiple images under either $T \leq 3$ or $h_2 =$ Fr are intrinsically more visible, they are easier to track accurately and thus do not become less visible in real pursuit.

To investigate whether lack of a distinct target is a hindrance to pursuit, a filtered moving object could be overlaid with a sharp moving marker to guide the pursuit system. Such an experiment is within the capabilities of the HRTV system but was not performed.

6.4 Multiple Images during Fixation

Predictions. Just as for multiple images observed under pursuit, the initial prediction is for this defect to be visible when the spacing of the images exceeds α. Now, since the spacing is vT, the threshold curve in the velocity/frame-rate plane should be a line through the origin of slope $1/\alpha$ (Figure 6.10). In this mode of observation there is no fundamental restriction that the region of defects must exclude $T = 1$, since these multiple images are produced by $g'(x)$, independent of the display up-conversion method or lack thereof.

The filtering action of h_1 may control the defect visibility by blurring the sharp edges of the individual images. As h_1 becomes a progressively wider type of filter, from Fd through Fr, the slope of the boundary should decrease; nevertheless, it should remain a straight line since all images along any such line are subject to the same $h'_{1x}(x)$. Finally, when h_1 = Fr the defect should disappear completely because the spatial extent of $h'_{1x}(x)$ will be vT.

The choice of interpolation filter h_2 is not expected to affect the threshold curves at all because this filter is not reflected into an equivalent spatial filter in the absence of ocular movement and because it is of much shorter duration than $h_3(t)$.

Results. It was found that the visibility of multiple images under fixation increases with both v and T. Subjectively, it appears that the increase with T is due to enhanced contrast between adjacent multiple image steps.

The threshold curves obtained experimentally are largely supportive of the predictions (Figure 6.11). Until they flatten out to the right, straight-line trends are followed by the curves for h_1 = Fd and h_1 = Fr/2, and, with a little imagination, even for h_1 = Fg. The values of α corresponding to these trends range roughly from 6 for Fd to 16 for Fg.

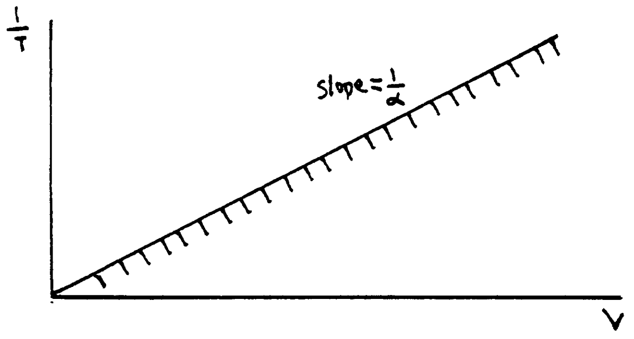


Figure 6.10: Predicted threshold for multiple images during fixation

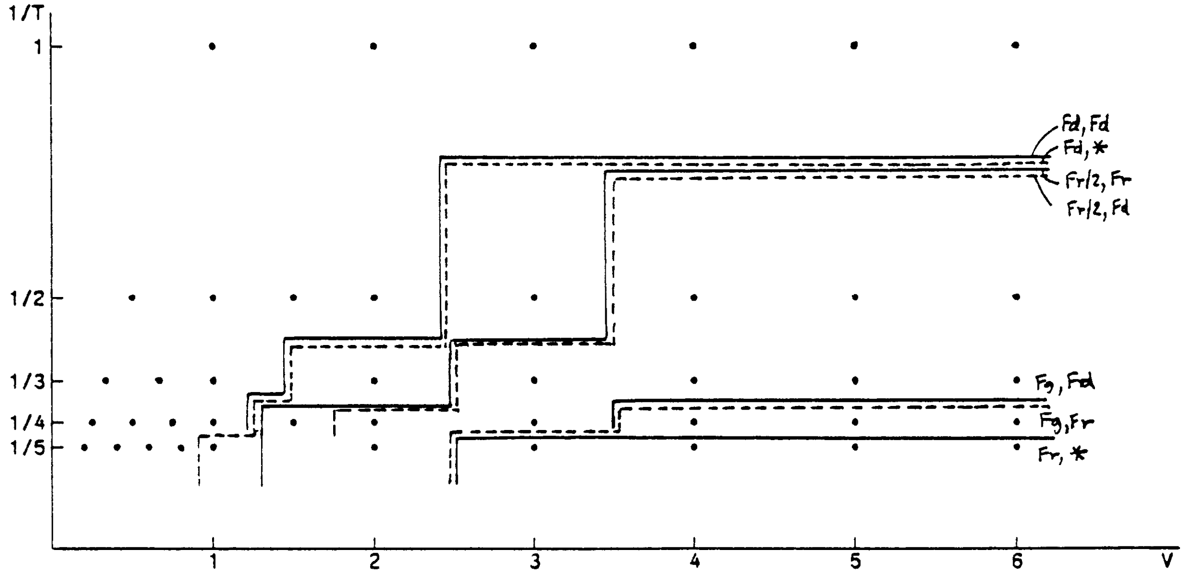


Figure 6.11: Measured thresholds for multiple images during fixation ;
* = Fr, Ft, or Fu

The subjective equivalence of images interpolated by different filters was noted during pairwise comparisons. In addition, there are only insignificant differences in threshold curves caused by change in h_2. These findings confirm the hypothesized irrelevance of h_2 for multiple images under fixation.

As h_1 is changed through the progression of filter types Fd, Fr/2, Fg, and Fr, the threshold curve moves toward the lower right as predicted; however, even the widest filter $h_1 = \text{Fr}$ was not always sufficient to eliminate the defect at $T = 5$. The difference between $h_1 = \text{Fr}/2$ and Fg is much more significant than differences between other pairs of adjacent curves on the map. The appearance of near-threshold multiple images under either $h_1 = \text{Fd}$ or Fr/2 is that of a periodic pattern of distinct edges, while the appearance under either $h_1 = \text{Fg}$ or Fr is that of a blurry low-amplitude sinusoidal grating.

6.4.1 Further Analysis

With images not examined in this study or with higher velocities, the visibility of multiple images will not necessarily increase monotonically with velocity as it does in the experimental observations. If the spacing between images increases too far, then the eventual decline in the HVS's sensitivity to low-frequency spatial gratings would begin to reduce the visibility. More directly, the greater velocity shortens not only the total duration of presentation but also the time the object remains near the retinal fovea, tending to reduce the ease of defect detection. However, even if a combination of the preceding factors does indeed cause the multiple images defect to decrease in visibility and eventually disappear with further velocity increases, that phenomenon is unlikely to be of practical value in an image transmission system.

More pertinent is the following explanation for the appearance of enhanced contrast with larger values of T. Consider two adjacent steps in the multiple image pattern, each vT pels pels wide. The perceived brightness versus time functions of the two regions differ only by a time shift; i.e., they are $f(t)$ and $f(t-T)$. Because the temporal filter $h_3(t)$ has a time constant much longer than T, the instantaneous contrast between the regions is related to $|f(t) - f(t-T)|$ and is monotonically increasing with T. By

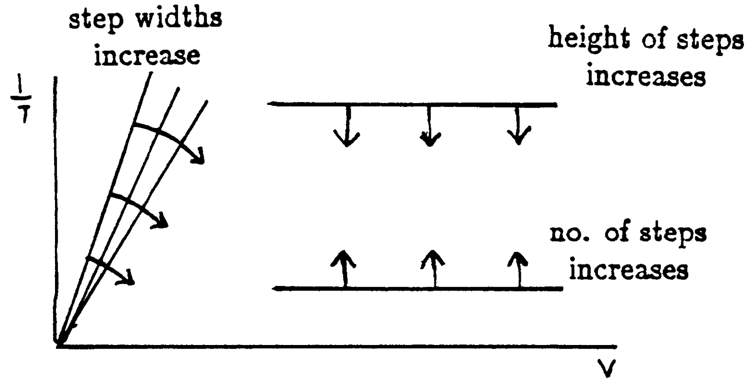


Figure 6.12: Factors affecting multiple images during fixation are step width, number of steps, and contrast of steps

this argument, the lower frame rates would be degraded by multiple images of greater contrast and, hence, greater visibility. Likewise, decreasing contrast at higher frame rates is a reasonable explanation for the flattened portion of the threshold curves.

As in the ocular pursuit situation, it is conceivable that a greater multiplicity of simultaneously visible images could assist in the detection of the periodic pattern. The number of images, roughly equal to the duration of visual persistence divided by the transmission frame period, bears an inverse relationship with T. The three factors, inter-image separation, contrast, and multiplicity, are compared in the velocity/frame-rate map of Figure 6.12. Contrast and multiplicity are again in opposition, with contrast the dominating effect; however, this time the directions of increasing effect are reversed from those under pursuit.

Incidentally, the threshold curve for $h_1 = h_2 =$ Fd should also serve as the threshold for multiple images observed as the eye sweeps past a stationary object presented at *display* frame rate $1/T$, due to the considerations of Section 5.6.

6.5 Large-area Flicker

Predictions. The behavior of flicker is presumably very simple, since it should depend mainly on the relation between the display frame rate and the CFF for the object being observed, taking into account its brightness and area. If the object is viewed under ocular pursuit, the velocity should be irrelevant; thus, the threshold boundary should be a horizontal line at the CFF, above which flicker is invisible. Under fixated observation, the preceding threshold should also apply, but an additional lower threshold may come into play since vT must be much less than the dimensions of the object in order for flicker to be observed on the object itself, as explained in Chapter 5. These two considerations lead to the predicted thresholds of Figure 6.13.

The presampling filter h_1 should be irrelevant, but among the standard choices for h_2, only h_2 = Fd will allow the display frame period to be longer than 1. With any other interpolation filter, absolutely no flicker will be detected.

Results. In the experimental observations, the flicker defect was recorded when it was visible on the moving object, in the background, or both. Although it would have been more informative if flicker on the object were recorded separately, in practice it was difficult to decide sincerely whether or not the object observed under fixation was flickering when it was surrounded by a clearly flickering background. On the other hand, the additional information is not of direct value to display design since obvious flicker is unacceptable whether it appears on the object or on the background.

On account of the above experimental difficulty, the vT threshold is not exhibited in the results shown in Figure 6.14. The data requires little discussion since all other predictions have been confirmed for both modes of observation. The flicker threshold lies between $T = 2$ and 3, which translates to 60 and 40 Hz, respectively. Incidentally, flicker in the background, which is darker than the object, was invisible in most of the trials of $T = 3$.

6.6 Small-area Flicker during Pursuit

Predictions. The perception of small-area flicker during pursuit requires detecting a temporal modulation in a small region in the vicinity of a moving edge. The area of this potential region of flicker is roughly proportional to vT, by the same reasoning as for blur. For any given frame rate $1/T$, there should be a "threshold area" A above which

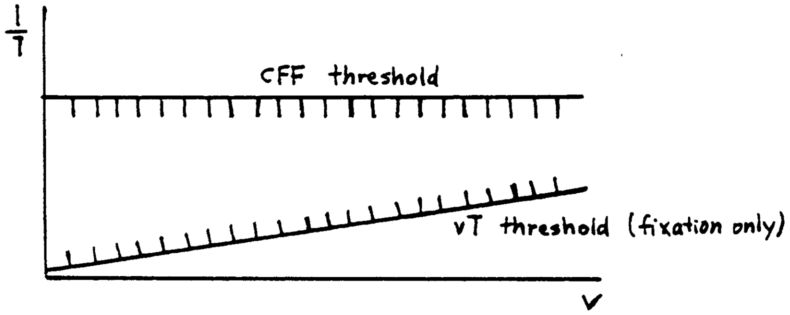


Figure 6.13: Predicted threshold for large-area flicker

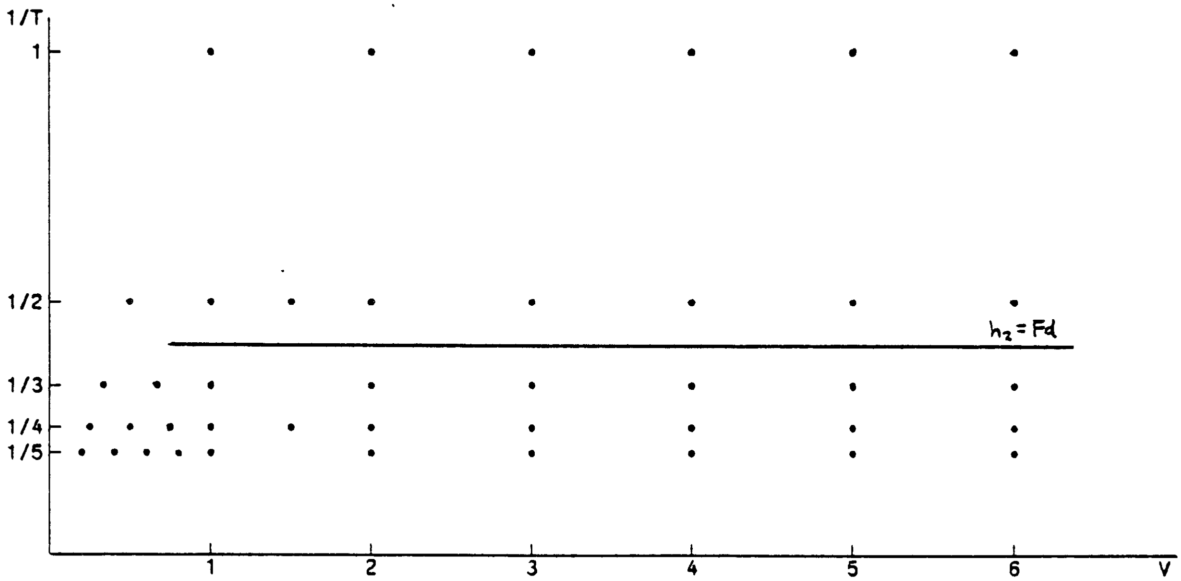


Figure 6.14: Measured thresholds for large-area flicker for h_2 = Fd with any h_1

flicker is visible, and as $1/T$ increases this threshold should rise since the CFF is an increasing function of stimulus area. Assuming only that the latter function is concave (specifying a logarithmic function would be spurious at this level of approximation), it can be deduced that the threshold boundary in the velocity/frame-rate plane is some concave curve of positive slope (Figure 6.15).

The region of defects is expected to grow as the filters h_1 and h_2 are changed to successively wider types of filters, since the area of the edge transition region will increase. The limiting case of $h_2 =$ Fd precludes small-area flicker, for not only does the reflected filter $h'_{2x}(x)$ vanish, the absence of interpolation would also cause large-area flicker to dominate over small-area flicker.

Results. The sparse set of samples in the velocity/frame-rate plane does not permit verification of the true shape of the threshold boundary curve, but the observation data does find the the threshold value of vT to be greater at higher frame rates (Figure 6.16).

Contrary to predictions, the effect of broadening h_1 or h_2 is to move the boundary curve to the right. This implies that the greater width of the transition region is more than offset by the reduction in amplitude of temporal modulation at points in the region. Nevertheless, none of the filter combinations is able to suppress completely the small-area flicker defect for the highest velocities when $T = 5$.

When $h_2 =$ Fr, changing h_1 from Fr/2 to Fg to Fr progressively removes and finally eliminates the defect at $T = 4$ but has no influence on the defect at $T = 5$. $h_1 =$ Fd and Fr/2 are equivalent in effect. The effect of h_1 is not as large when $h_2 =$ Ft instead of Fr. No difference between images interpolated by $h_2 =$ Ft or by $h_2 =$ Fu can be observed, doubtless because Fu is only negligibly wider than Ft at large values of T.

6.6.1 Further Analysis

By analogy to multiple images under ocular pursuit, the shortened duration of stimulus presentation necessitated by greater velocity potentially reduces the ability of the HVS to detect temporal patterns of illumination. On the other hand, experience finds that the poorer accuracy of pursuit due to high velocity or an indistinct target tends to increase the likelihood of observing small-area flicker. This happens because

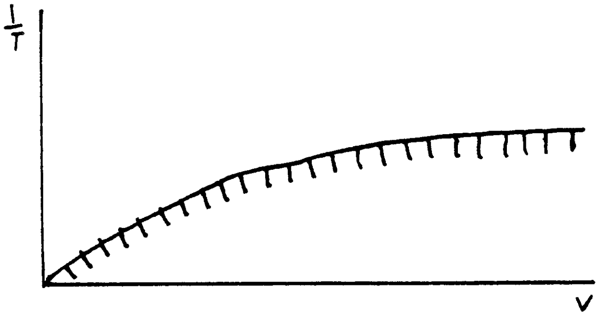


Figure 6.15: Predicted threshold for small-area flicker during pursuit

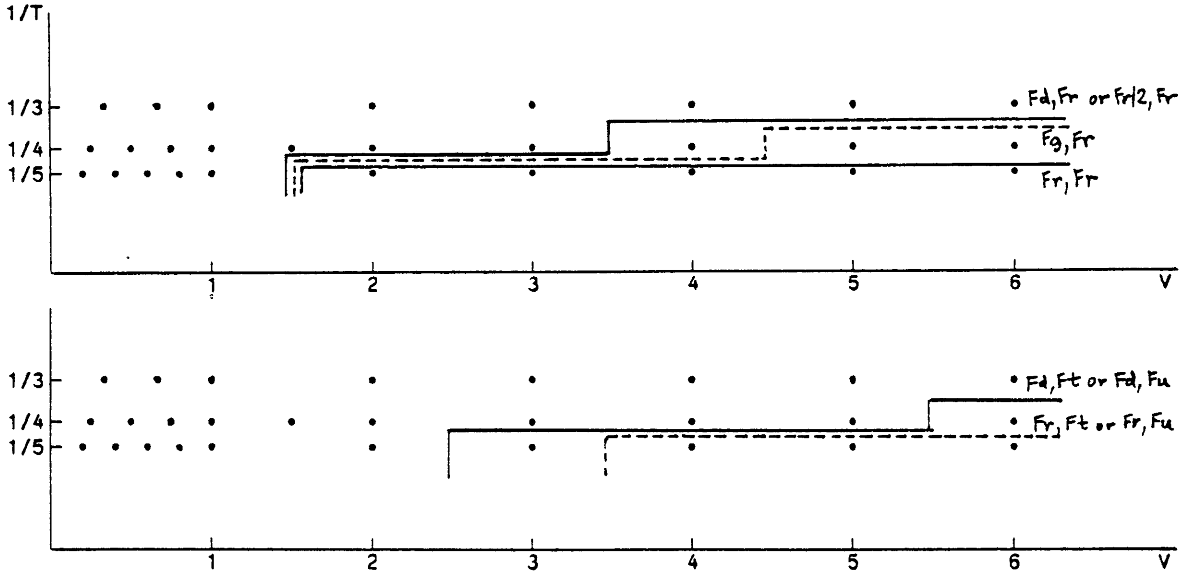


Figure 6.16: Measured thresholds for small-area flicker during pursuit

a small amount of the stronger small-area flicker of fixation is being detected.

To explore the extent of these influences on visibility of the defect in question, simulated ideal pursuit was observed for h_1 = Fd or Fr/2. The threshold curves obtained in this experiment compare very closely with corresponding ones for real pursuit, indicating that the effects of duration and pursuit error on small-area flicker are not substantial.

One interesting difference between real and simulated pursuit, however, is that the edges of the object observed in simulated pursuit, with $T = 5$ and $V \geq 4$ ($V \geq 3$ for h_2 = Fr), exhibit not only flicker but also a cyclic leftward motion. The apparent motion stimulates the visual system's velocity receptors intensely and unmistakably. The fact that this phenomenon is not perceived during real pursuit seems to suggest that either the channel associated with these receptors is inhibited during pursuit, or, more probably, that random offsets due to inaccurate pursuit prevent sufficient stimulation.

6.7 Small-area Flicker during Fixation

Predictions. In this type of flicker defect, no region of the image by itself can be responsible for a periodic temporal modulation, so it is not really possible to predict the visibility of small-area flicker under fixation as a function of v and T by defining and calculating the expected area of the defect. Nevertheless, since there is nothing to suggest a better alternative, the distance moved per transmitted frame, vT, will be used to predict the visibility. The threshold boundary will then be the same as for ocular pursuit (Figure 6.15).

It is possible to predict the general effect of changing among different types of filters, however. As the spread of $h'_{1x}(x)$ becomes sizable with respect to vT, the rate of brightness change at a fixed location as an edge passes it becomes more gradual. In addition, when h_2 = Ft or Fu rather than Fr, these brightness changes occur smoothly rather in abrupt steps. Therefore, as either h_1 or h_2 is changed to wider types of impulse responses, the region of defects should shrink. Finally, absence of display interpolation

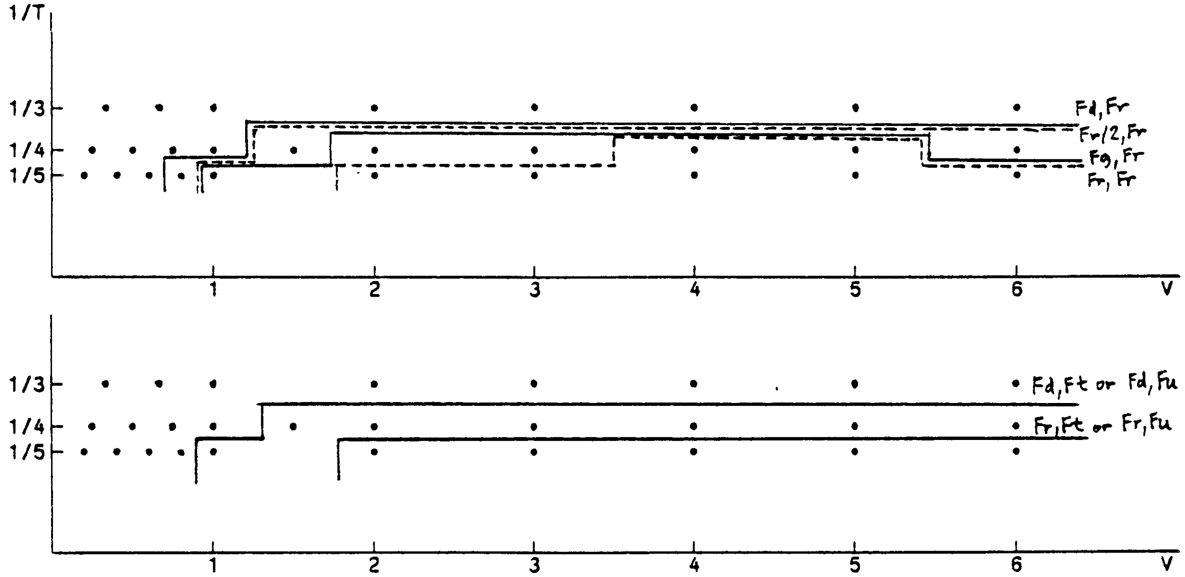


Figure 6.17: Measured thresholds for small-area flicker during fixation

(h_2 = Fd) precludes small-area flicker.

Results. At lower velocities ($v \leq 3$) the rising slope of the threshold curve is evident in the observation data for all combinations of filters (Figure 6.17). As h_1 is changed through the progression of filters from Fd to Fr, this portion of the boundary retreats toward the right. Whether h_2 is Fr, Ft, or Fu makes no difference. No filters are sufficient to eliminate the defect when $T = 5$. On the other hand, any filter is completely satisfactory when $T \leq 3$. The latter statement is also true for small-area flicker under pursuit, suggesting that the CFF in these instances of small-area flicker is just above 30 Hz.

At increasing velocities a sight distinction between images filtered by h_2 = Fr and by h_2 = Ft or Fu appears. Those filtered with Ft or Fu exhibit the small-area flicker defect in a smaller subset of the velocity/frame-rate map.

Not previously accounted for, however, is the downward turn of some of the boundary curves. An explanation might be that the larger spatial separations, vT, between successive brightness fluctuations exceed the span of lateral interaction in the eye. Although this upper velocity threshold appears only for two combinations of filters within

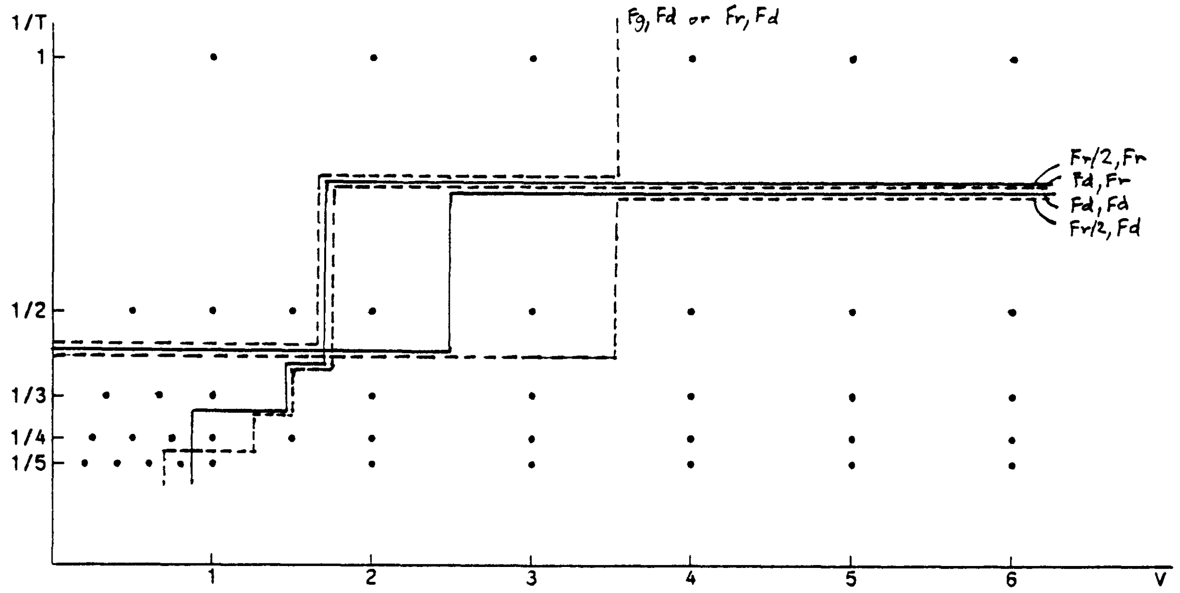


Figure 6.18: Composite defect thresholds

the range of this study, it would probably occur for all filters eventually.

6.8 Relationships Among the Defects

A substantial amount of data has been presented in this chapter, emphasizing the unique behavior of each defect as the velocity, frame rate, and filters are changed. While this approach increases our understanding of each defect in isolation, it does not point out the correlations among defects. Nor does it suggest how to choose the best parameters for an image transmission system. Therefore, composite defect thresholds and tradeoffs between defects need to be considered.

One quality metric of an image transmission system might be the maximum velocity completely free of any kind of defect. A composite defect threshold is the boundary of the union of the regions of blur, multiple images under pursuit and fixation, small-area flicker under pursuit and fixation, and large-area flicker. The composite thresholds for six combinations of h_1, h_2 are shown in Figure 6.18. The remaining combinations are not shown; because their composite thresholds are the same as their blur thresholds and because these thresholds are positioned entirely to the left of the curves shown, they cannot maximize the defect-free velocity.

Two facts may be noted from this data. First, the best pair of filters to use, i.e. the one whose composite threshold lies farthest to the right, depends on the transmission rate. When $T \leq 2$ then $h_1 = \mathrm{Fr}/2, h_2 = \mathrm{Fd}$ is no worse than any other choice. It represents a compromise between $h_1 = h_2 = \mathrm{Fd}$, which gives better sharpness, and all other filters, which give better smoothness. When $T = 3$ or 4 then $h_1 = \mathrm{Fd}, h_2 = \mathrm{Fr}$ provides the best sharpness while suppressing large-area flicker. When $T = 5$ then $h_1 = \mathrm{Fr}/2, h_2 = \mathrm{Fr}$ appears to be the best. While the boundary curve crossover shown between $T = 4$ and 5 has not been shown to be statistically significant, it could be explained by the presence of small-area flicker in the image filtered by $h_1 = \mathrm{Fd}, h_2 = \mathrm{Fr}$.

Second, very little of the velocity/frame-rate plane is entirely free of defects. Specifically, a 60 fps transmission is sufficient for subjectively perfect reproduction of a step moving up to $\approx 6°$/sec. 40 fps is enough for up to $\approx 2.5°$/sec, and 30 fps is enough for $\approx 2°$/sec, but these figures are considerably smaller than the typical velocities of motion in television, said to be 5.5 – 7.6°/sec in Chapter 2. A 120 fps transmission, with no display up-conversion, seems to be adequate for the highest velocities examined in this study (10°/sec); however, it is possible that a display and transmission rate between 60 and 120 fps would also be enough.

The conclusion that the best interpolation and presampling filters are the narrower ones is surprising in light of everything known about aliasing. This result was obtained because the quality metric penalizes all defects equally, and just-noticeable blur appears at lower velocities than other defects. Because no foreseeable system will transmit much more than 60 complete frames per second, and because motion cannot be limited to $\approx 6°$/sec, in practice the composite thresholds will be exceeded for much of the time. Therefore, what may be more relevant than the location of any threshold is how gracefully degradations develop beyond it and the relative contributions of the individual defects to the overall subjective quality. After these factors are considered, the optimum selection of filters may be quite different.

With regard to the effects of h_1 and h_2, the data of this chapter may be summarized as follows: as h_1 and h_2 are changed to wider types of filters, with all other factors held constant, the region of blur defects grows and the regions of motion quantization defects shrink or stay the same. This observation agrees with the well known inverse relationship between the amount of blur and the visibility of aliasing artifacts.

Now, since all defects cannot be completely avoided, the filters h_1 and h_2 should be chosen for the optimum subjective tradeoff among defects. Obtaining the optimum requires evaluation of suprathreshold defects. One approach would be subjective quality rating of images with several simultaneously visible defects. Another would be construction of separate psychophysical magnitude scales for each defect, followed by a rule for summation of impairments. Unfortunately, these techniques are beyond the scope of this research. The present set of data is insufficient because two systems with similar threshold boundary curves may behave very differently far away from the boundary. For example, the blur thresholds for h_1 = Fd and Fr, holding h_2 = Ft fixed, are identical below $T = 1$. However, farther to the right the blur appears to be much worse for h_1 = Fr than for h_1 = Fd, as one might expect.

Finally, a difficulty which is certain to arise in any attempt to balance blur against suprathreshold multiple image defects is that the former is most noticeable during pursuit but the latter is often more pronounced when $v_2 \neq -v_1$, as during fixation. When h_2 = Fd and the observer does not track the object, a nontrivial presampling filter h_1 is desired in order to reduce the visibility of multiple images, but if the observer does track, h_1 = Fd is best for the sharpest image. When h_2 = Fr the situation is slightly different since multiple images could also arise during pursuit, and thus some camera filtering might be desirable even for pursuit. Nevertheless, for any CRT display interpolation scheme, a wide h_1 is not better *a priori* than a narrow one, nor vice versa, because the transmitter cannot predict with certainty when viewers will be fixating or when and what they will be tracking.

6.9 Summary

Observations of test images with the HRTV display system have found combinations of three system parameters and one image parameter — transmission frame rate, camera presampling filter, display interpolation filter, and velocity — which allow reproduction of a moving object completely free of defects. The best region of the velocity/frame-rate plane is small v, large $1/T$. As v increases and $1/T$ decreases, moving objects suffer successively from blur, multiple images, and then small-area flicker. The experimental behavior of the threshold boundary curves under change of h_1 and h_2 was anticipated, and later rationalized, on the basis of properties of the HVS. Generally, the wider filters induce blur earlier in order to hold off multiple images until later. Just as multiple images and small-area flicker take on subjectively distinct forms during pursuit and during fixation, the threshold curves depend heavily on eye movements, too.

The experiments spanned the range from 24 to 120 fps in order to assess the potential improvements afforded by increasing the transmission frame rate over that used in present-day television, assuming a high-rate display. Recall that the widest filter h_1 = Fr, despite all the blur it causes, is not sufficient for good motion rendition at 24 fps, and sometimes not even at 30 fps. That being the case, it is unlikely that a satisfactory tradeoff between blur and other defects can be achieved at those frame rates. Tradeoffs are perhaps more promising above 30 fps, since at least small-area flicker defects disappear. For instance, some intermediate class of filter between h_1 = Fr/2 and h_1 = Fg might largely suppress multiple images during fixation without too much blur (Figure 6.11). However, it is probable that only above 60 fps can motion up to 10°/sec be reproduced with practically no visible degradations.

Chapter 7

Dependence of Defects on Image Content

In the previous chapter all observations were performed with a single test object and background. However, the characteristics of the source image are expected to affect the visibility of defects. Among others, some factors are the size and configuration of objects, contrast and brightness, the amount of high frequency detail on the object, and the texture of the background.

Here, a limited exploration of these factors is presented. The threshold boundary curves for multiple images are measured for three different combinations of object and background. The first is a narrow vertical bar against a plain background, which is used to examine the role of size. The next image uses the same moving object as in the last chapter, with a brighter constant background. The third image is used to investigate spatial masking by adding texture to the background of the preceding image.

7.1 Effect of Object Size

In this series of observations, the background is uniform with a luminance of 5 units, and the moving target is a vertical bar one pel wide and 64 units in luminance. In many ways, this target tests the worst-case performance of an image transmission system. Four pairs of filters h_1, h_2 are selected, each being either Fd or Fr. v and $1/T$

are chosen from the same sample points as before.

Blur

It was easy to detect blur on the moving bar because the perceived width and brightness of this target, unlike the previous wide target, is greatly affected by reflected spatial filtering. Indeed, when the amount of blur was very great (h_1 = Fr, v_1T large) the displayed target was scarcely brighter than the background and thus difficult to distinguish.

Multiple Images during Pursuit

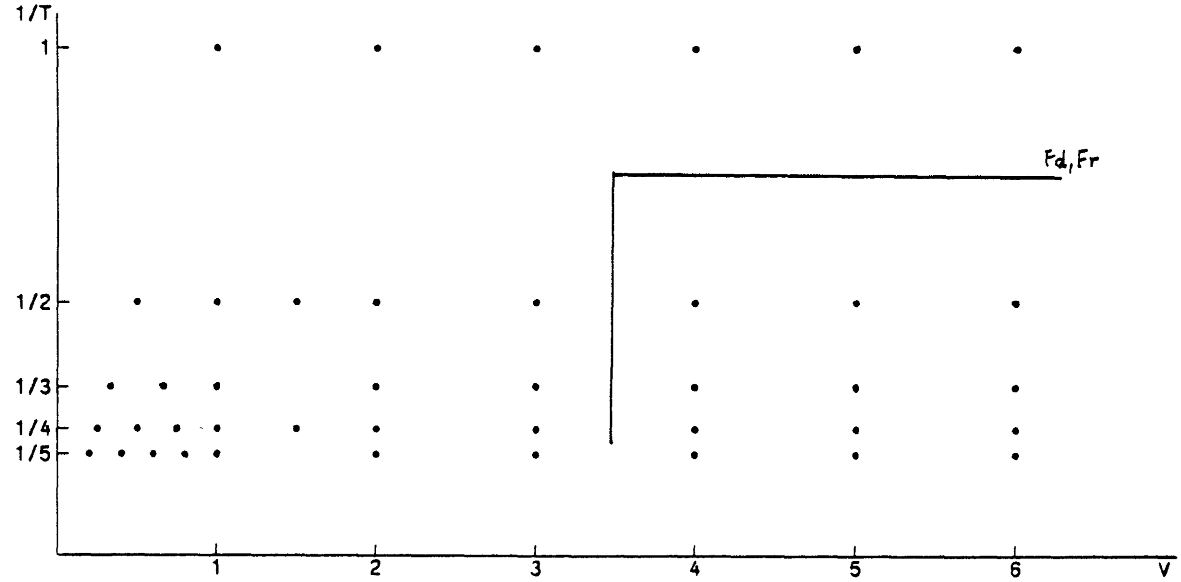
Among the filter pairs of this experiment, multiple images during pursuit should appear only when h_1 = Fd and h_2 = Fr. The predicted threshold curve for this defect (Figure 6.6) was not followed by the observations of a 64-pel-wide object in Chapter 6 because, for large T, the decreasing contrast between steps makes it more difficult to distinguish them.

On the other hand, for a narrow vertical bar the threshold inter-image separation α seems to be constant, with $\alpha = 4$, following the predicted behavior exactly (Figure 7.1). This α is the same as the worst-case minimum separation between multiple images with a wide target (Figure 6.7).

The sum of the intensities of the separate images in the defect must be independent of the number of images, T, so that as T increases, each image becomes less bright. With a narrow bar, each image is surrounded by a dark background; thus, even with decreasing brightness, the contrast ratio between the image and surround is still sufficient for discrimination. With the previous wide target, the relevant contrast is the one between adjacent steps, which falls rapidly with increasing T.

Multiple Images during Fixation

The defect threshold boundary curves with h_1 = Fd (Figure 7.2) are similar to the ones obtained with the wide target, except that they do not flatten out above



Figurc 7.1: Measured thresholds for multiple images of a narrow bar during pursuit

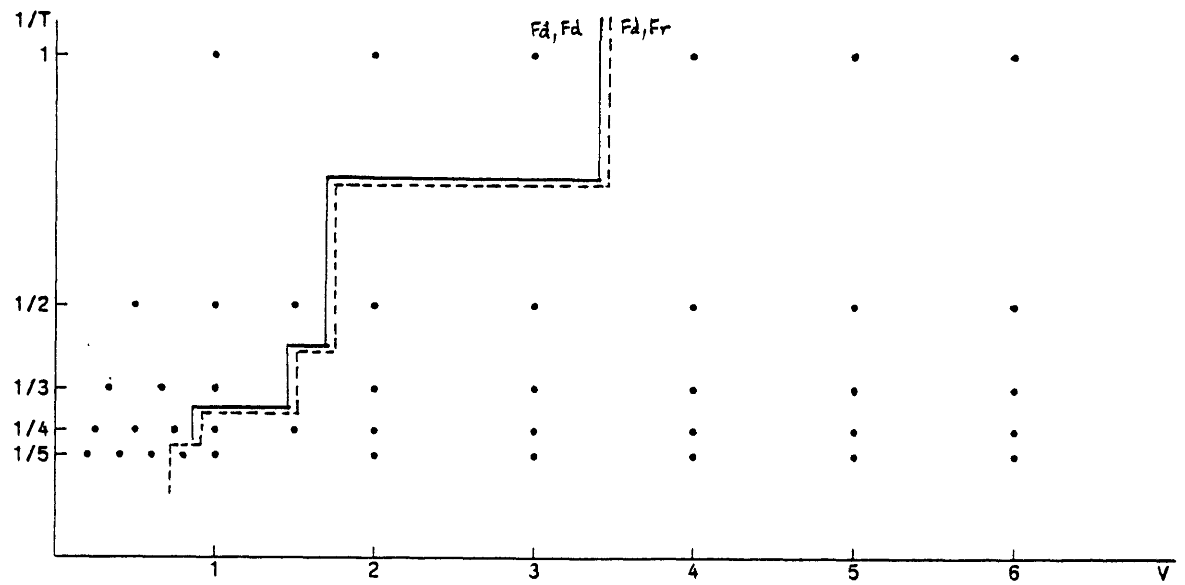


Figure 7.2: Measured thresholds for multiple images on a narrow bar during fixation

$1/T = 1/2$. In fact, defects can be observed even at the maximum frame rate, so that it appears that in the worst case a much higher frame rate is needed to avoid multiple images at these velocities. The curves follow a straight-line trend with parameter $\alpha = 4$, which is somewhat smaller than the value of α with the wide target. From this α it can be calculated that a multiple images defect whose spatial frequency is above 18 cpd is not visible.

The foregoing result agrees with one experiment by Watson [WATS 83A], which found that the required frame rate for a narrow bar to appear spatiotemporally continuous was linearly related to velocity, with a coefficient of 17 cpd. His data also shows that display rates in the range of a few hundred frames/sec would be needed for continuous motion up to 20°/sec.

Multiple images could not be detected when h_1 = Fr, even though $h_1 = h_2$ = Fr resulted in this defect for a wide moving object. The reason is that the narrow object is so thin that camera blur obliterates its contrast against the background.

Temporal Defects

Large-area flicker was visible in the background at 30 Hz or less, but the target itself was never observed to flicker. The CFF for small regions is lower than for large ones. Perhaps for the same reason, there was no small-area flicker, either.

7.2 Effects of Background Brightness and Texture

A brief examination of the effects of background brightness and texture completes this short chapter. The following experiment observes multiple images under fixation for the original 64-pel-wide target. The background is either uniform, with a luminance of 32 units, or a 0.125 cycles/pel square wave, alternating between 24 and 40 units. The presampling filter is Fd and the interpolation filter is Fr.

Figure 7.3 shows the measured threshold curves, both of which lie clearly to the right of the Fd,Fr curve for the original dark background (Figure 6.11). Therefore, the diminished contrast increases the minimum inter-image separation, $v_1 T$, for detection

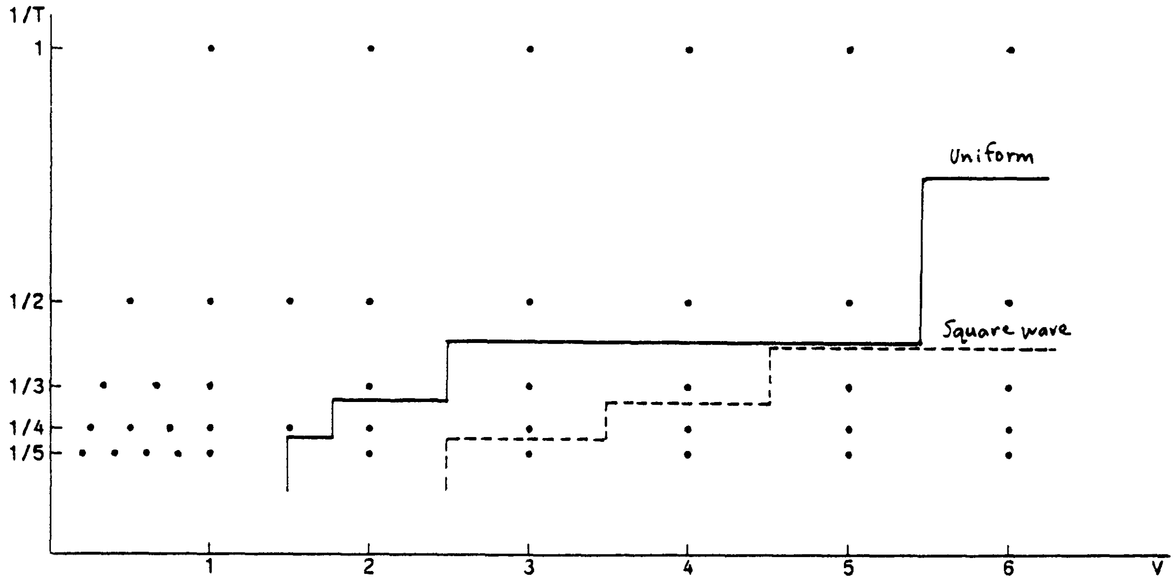


Figure 7.3: Measured thresholds for multiple images during fixation with uniform and grating backgrounds

of the defect. In addition, the presence of a background grating makes it more difficult to identify multiple images.

A narrow vertical bar is hardly any more representative of objects found in natural images than a rectangular box. Similarly, the scene behind the moving object is usually more complex than a grating. Nevertheless, these observations have indicated at least some of the possible variability in defect thresholds due to changing image characteristics.

Chapter 8

Discussion and Summary

Despite the large number of observations reported in Chapters 5-7, the results are for a very narrow subset of all image transmission systems representable by the canonical model, and thus they do not lend themselves directly to conclusions about imaging systems in general and HDTV in particular. One reason for caution is that there are shortcomings in the simulation using the HRTV system which restrict the generality of the results. Rather than being concealed, however, the known limitations of this experimental approach should be discussed as a guide to interpreting the results and also as a catalog of potential refinements.

Limitations aside, a number of somewhat speculative conclusions can be drawn from what has been said so far. First, an interesting way to think about some aspects of spatiotemporal images is to analyze analogous aspects of still images. Second, the original idea that the HRTV display rate is high enough for all image defects observed on it to be attributable to the signal processing was found to be incorrect. Thus, an unforeseen offshoot of this research is the ability to make some statements about the requirements for an ideal display. Third, similar statements can be made about the camera and transmission channel. Finally, possible differences between defects as observed here and as seen in HDTV systems are conjectured.

A list of suggested extensions to the work and a summary of the main points conclude this investigation of motion-induced image defects.

8.1 Display Limitations and Consequences

8.1.1 Problems with the Implementation

It is acknowledged that the experimental methodologies employed did not adhere in many respects to orthodox procedures of psychophysical measurement. In large part the ad hoc procedures were necessitated by limitations of the HRTV facility. (On the other hand, painstaking measurements would probably be considered inappropriate at this stage of investigation.) The limitations of the facility include spatiotemporal resolution, viewing angle, inability to present two simultaneous images, and environmental conditions. Future experimental studies should correct some of these deficiencies.

Because the HRTV is designed to display at 120 fps, the only transmission and display rates which could be examined are its integral submultiples. This unfortunately leaves a large region between 60 and 120 fps where observations would be valuable. Moreover, because velocities must be chosen so that v is an integer when $T = 1$, the selection of velocities is sparse. Not being able to vary the velocity in fine steps precludes procedures such as the method of limits or the method of adjustment [GUIL 54].

The viewing distance was chosen so that the selection of angular velocities would not be overly sparse. As a result, the field of view is so small (7.2°) that objects moving at what should be moderate velocities (say, 10°/sec) cannot remain in view for much time. Although this may be more representative of viewing situations in present-day broadcast television, it is a drawback when trying to study steady-state pursuit so that results may be extended to future wide-screen displays.

The inability to simultaneously show two or more images side-by-side hinders subjective testing techniques such as paired comparison, ranking, or matching with a scale of standard images. Although two images could be loaded in the HRTV for immediate access, a memory task still intervenes in the process of comparing two images presented sequentially. Alternating between images feels much more awkward than, for instance, comparing two photographs presented together.

Finally, the environmental conditions in the room containing the HRTV display are less than satisfactory. The location is not comfortable for undistracted, concentrated evaluation of images for extended periods — obviously a hindrance for audience testing. More seriously, the ambient illumination cannot be maintained at a suitable level for viewing a television screen. Overhead lighting is too bright, and subdued daylight alone is closer to optimum but variable. The observations were performed under the most repeatable level, which is achieved at night with lights off. The principal objection to this level is that the time constants of human vision are prolonged under dark adaptation, meaning that sensitivities to temporal image defects and jerkiness are lower than those under more realistic television viewing conditions. In fact, the CFF is maximized when the intensity of the surround is equal to the mean intensity of a flickering test target [BROW 65A].

8.1.2 Inherent Problems

Even supposing that all of the aforementioned deficiencies were corrected and that precise measurements of defect thresholds could be obtained, the HRTV would still have two intrinsic limitations. They are the finite display frame rate and the two-dimensionality (t, x) of displayable images.

When this research project was first conceived, it was thought that a 120 fps display would look temporally continuous. As a result, image defects that truly exhibit a fundamental property of some transmission frame rate are confounded with those that are the fault of the display technique (e.g., multiple images under ocular pursuit). In addition, the discrete interpolation filters, $h_2(t)$, are not adequate approximations to continuous-time filters; for instance, Ft does not really simulate continuous-time linear interpolation between frames. On the other hand, these findings are informative in themselves. It will be interesting to speculate on the frame rate necessary for a temporally continuous display and on the different results obtainable from continuous interpolation (Section 8.3).

The HRTV was designed to display images constant in the y spatial dimension, in order for the rate and volume of data feeding the display to be no more than for ordinary still pictures. Although all types of motion-induced defects which could exist in 3-D images in t, x, and y are accounted for by observing 2-D images in t and x, there are measurements which cannot be performed in 2-D. The problem is not that objects move in directions other than horizontal or in curved trajectories, for curves are essentially straight within small neighborhoods and oblique motions could have been investigated by physical rotation of the HRTV monitor. Rather, the chief shortcoming is that it is not possible to judge subjective image quality outside of the familiar context of natural images. Thus, tradeoffs between image defects cannot be evaluated, and quality rating scales such as those established by the CCIR [ALLN 83] are not usable.

Another consequence of the restricted class of displayable images is that blur can be measured only by its effect on a step transition. In natural images, the loss of texture and fine details due to blur may be more objectionable than blur at object boundaries, but these effects cannot be assessed adequately with the HRTV since vertical texture patterns do not look realistic.

8.2 Analogies between Space and Time

The image transmission model of this thesis is unusual because it ignores all spatial characteristics of the system, dealing with the spatial dimension only when motion causes a spatiotemporal interaction. The discipline of spatial image processing is considerably more advanced than its temporal counterpart. Therefore, perhaps the better-known spatial concepts could be brought to bear on temporal processing problems if the analogies between space and time were exhibited.

8.2.1 Similarities and Differences

In spatial image processing, the two dimensions of an image are x and y, while in the temporal image processing of this thesis, the two dimensions are t and x. Many concepts in one domain possess a corresponding concept in the other. Edges in an xy image correspond to trajectories of motion and scene changes in an tx image. High frequencies are required to preserve the sharpness of both spatial and temporal edges. In reconstructing an xy image from periodic samples, inadequate spatial interpolation leaves behind a visible sampling grid structure and jagged edges; similarly, in reconfixed product of resolutions, the relative resolutions could differ by as much as 2.5 to 1 with minimal noticeable change. Consequently, it has been asked whether spatial and temporal resolutions are likewise interchangeable [SCHR 84]. (Note: the term "exchange of resolution" was used in a previous context to refer to the nonrectangular shape of the window of visibility.)

If resolution is equated with the reciprocal of the amount of blur, then one answer is that in the presence of motion, temporal filtering is reflected into spatial filtering. Thus, the total amount of blur is a combination of motion-induced blur, which is proportional to $v_1 T$, and ordinary spatial blur. The combination rule is probably closer to a sum than a product, so that for constant bandwidth there must be an apportionment of resolutions which analytically minimizes the total blur. For example, if the total blur were given by $X + v_1 T$, where X = spatial blur, then the sum is minimized under the constraint $XT = C$ by $X = \sqrt{v_1 C}$, $T = \sqrt{C/v_1}$. There may in fact be a range of indifference to changes in relative resolutions about the optimum; however, this is not really analogous to the findings of Baldwin. For one thing, temporal resolution itself is not being perceived here. More importantly, the optimum ratio of orthogonal resolutions is so highly dependent on the velocity v_1 that it is unlikely for any fixed ratio to be acceptable across different images.

8.3 Requirements of the Ideal Display

The example of a narrow moving bar demonstrates that even 120 fps can be insufficient as a display frame rate (or transmission frame rate, for that matter), essentially because image features can move farther than their own widths between adjacent frames. In particular, the 72 fps display of the spatiotemporal interpolator currently under construction by Lee and Wang [LEE 83] will not always appear temporally continuous.

Accordingly, it is valid to ask what the display frame rate should be so that it does not contribute additional defects to those inherent in the transmitted signal. There are two viewpoints, depending on whether sophisticated source coding will be employed structing a time-varying image from temporal samples, inadequate interpolation leaves behind a temporal structure possibly seen as flicker and motion defects. Chapter 5's definition of multiple images — two or more images visible at one time instant – is paralleled by its definition of flicker – several cycles of modulation at one spatial location. The threshold contrast sensitivity functions for temporal and spatial gratings are very similar-looking curves. Lastly, spatial and temporal resolutions are equivalent in the sense that the uncoded transmission channel bandwidth is inversely proportional to their product.

With all these similarities between images in x and y, and images in t and x, it might be thought that good spatial filtering and sharpening techniques are also good for temporal processing. For example, the subjective optimality of the sharpened Gaussian spatial interpolation filter [SCHR 85] might carry over to the time domain. However, there are enough underlying differences between space and time that it shouldn't be taken for granted even it were true. First, space is basically isotropic, but time is irreversible. Second, there is no analog to ocular pursuit in an xy image. Finally, the ability of an imaging system to reproduce fine-grained details by transmitting high spatial frequency information is very much desired, but the analogous ability to reproduce rapid temporal oscillations in the original is usually not appreciated by the observer. In other words, although flicker can sometimes be detected at about 50 Hz or so, hardly any usable visual information is contained in the fine temporal detail. As pointed out earlier, the HVS is inclined to perceive evolving spatial forms over time functions. Hence, in the time dimension there is a wide discrepancy between the small bandwidth needed to carry usable information to the HVS and the large bandwidth required to preserve sharp temporal edges.

8.2.2 Exchanging Orthogonal Resolutions

A consideration in the correspondence between xy and tx concerns trading off orthogonal resolutions. In evaluating sharpness by judging entire images rather than focusing attention on specific features, Baldwin found that the optimum apportionment of spatial resolutions between x and y was 1 to 1 [BALD 40]. However, for a to enable reconstruction of a very high resolution signal at the receiver, or whether conventional frame transmission, as in the canonical model, will be used.

8.3.1 Transparent Display

A display will be called transparent if, given an ideal input signal of infinite spatiotemporal bandwidth, the reproduced image would satisfy every critical demand of the HVS in the categories of resolution and freedom from artifacts (colorimetric accuracy, dynamic range, and viewing angle are not of concern here). It follows that such a display would look temporally continuous.

If, in the spirit of this study, the difficulty of making a spatially transparent display is disregarded, what are the temporal requirements for transparency? Supposing, as in Chapter 3, that the maximum pursuit velocity is about $v_{\max} = 20°/\text{sec}$ and the spatial frequency limit is $f_{x0} = 30$ cpd. Preserving detailed spatial information during pursuit requires only that the passband of $h_2(t)$ extend out to at least $v_{\max} f_{x0} = 600$ Hz; with a short-persistence phosphor, that is easy to fulfill. Even with a low-frequency display, none of the spectral harmonics will enter the window of visibility corresponding to pursuit (Figure 8.1a).

The problem starts when the observer fixates, for then the first order harmonic will cross the window unless the display frame rate is on the order of $v_{\max} f_{x0}$ (Figure 8.1b). Things are not quite that bad because both the amplitude of image spectra and the sensitivity of the HVS begin declining much before f_{x0}. In Chapter 7 multiple images were visible only below 18 cpd. Accordingly, the harmonic could be allowed to cross the window of visibility at that frequency, even as the spatial resolution is maintained at 30 cpd. As a result, 360 fps would become sufficient.

The problem could become more intractable when the eye moves with a non-zero velocity unrelated to v_1. The result in the preceding paragraph holds for defects observed when the eye sweeps across a stationary object. However, the observer could be tracking an object which is moving at $-20°/\text{sec}$ at the time another object moving at 20°/sec passes near the center of the field of view. In that event, the first order harmonic of the object moving at 20°/sec will cross the window below 18 cpd unless

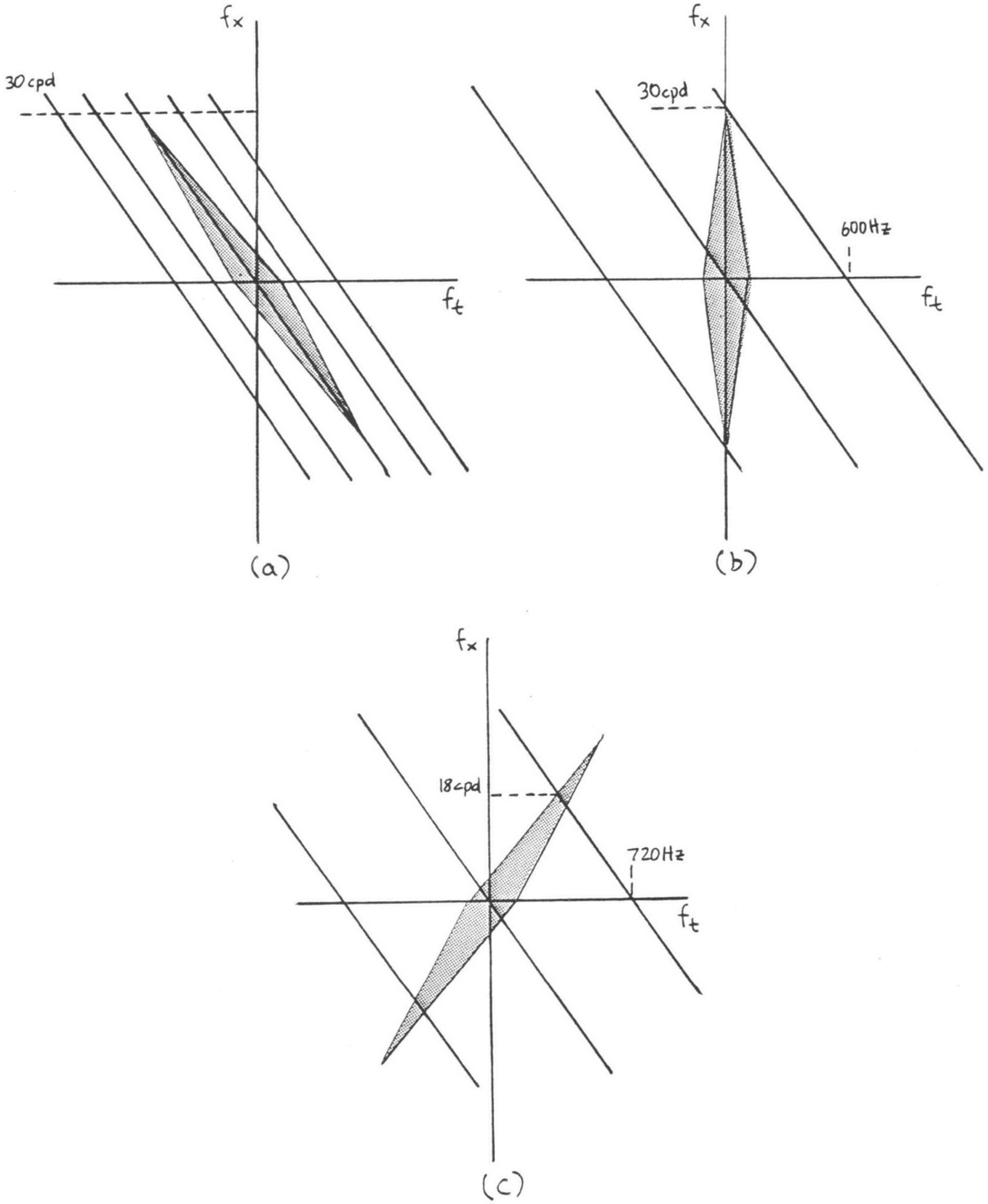


Figure 8.1: Spectrum of displayed image s_4 with passband of HVS superposed: (a) pursuit of a 20°/sec object; (b) fixation; (c) eye movement at velocity opposite that of the object

the display rate is doubled, to 720 fps (Figure 8.1c). h_2 cannot be used to filter out the resulting multiple images defect because it lies at $f_t = 360$ Hz; doing so would sacrifice moving details.

One other unrelated movement is a saccade, whose velocity can reach hundreds of degrees/sec. It is clearly absurd to contemplate a display that is transparent during saccades, but then where should the line be drawn on unrelated movements?

In view of the proposed frame rates, all of which are an order of magnitude higher than those in present-day systems, a temporally transparent display is not technologically feasible in the foreseeable future. Either some reduced resolution on moving objects observed under pursuit, or some multiple image defects, must be tolerated. The usual smoothness vs. aliasing subjective tradeoff thus exists at any realistic display rate even when the input signal is ideal. Of course, relaxing the resolution requirement also reduces the burden on the channel and camera. Alternatively, wide viewing angles and rapid motion could be forbidden by decree, but that would hardly be an acceptable solution!

8.3.2 Interpolation from Frames

The foregoing applies to image transmission systems using motion-compensated coding and reconstruction, for ultimately the receiver of such a system would be able to calculate an unblurred image for any time instant desired. The second viewpoint returns to the conventional system of the canonical model, where the problem is to determine the best way to interpolate a sampled image $s_4(t, x)$ with given parameters h_1 and T. A related question asks how much up-sampling should be performed so that it is equivalent to infinite up-sampling, as far as the HVS is concerned. Partial answers can be based on the results of Chapter 6.

First, suppose the transmission frame rate $1/T$ is greater than the CFF. Assuming that defects on a moving object are of interest only during pursuit of that object or fixation, then the best interpolation filter is Fd, with no tradeoff necessary. In other words, the display frame rate should also be $1/T$, and the phosphor should be as short persistence as possible.The reason is that Fd causes neither blur nor multiple images under pursuit. Other filters cannot remove defects inherent in s_4 and can only make things worse.

The preceding conclusion follows from the experimental observations because the images used there were purely translating. In more general images, however, defects can also be observed when the eye moves at a velocity unrelated to v_1, e.g., when it sweeps over a stationary object. Then, the selection of h_1 must compromise between blur during pursuit and multiple images during any eye movement.

Furthermore, the frame rate is more likely to be less than the CFF, in which case some interpolation is mandatory because large area flicker is overwhelmingly more objectionable than any other defect. The choice between sample-and-hold or linear interpolation between pairs of frames is a matter of balancing multiple images under pursuit and small-area flicker (more with sample-and-hold) against blur (more with interpolation), and the system parameters h_1 and T will influence the tradeoff.

When the display technology is a CRT, the continuous-time filters have to be approximated by Fr and Ft. Obviously, the higher the display rate the better, but only because a larger up-sampling rate r increases the spatial frequency r/v_2T of the multiple images defect during eye movements. Other defects are hardly affected, if at all, by raising r, since the amount of blur under pursuit due to h_2 is proportional to the constant v_2T, multiple images under fixation are independent of h_2, and small-area flicker should be independent of the fine structure of h_2. If the persistence of the phosphor is very short, as in the HRTV, a calculation similar to that in the previous section predicts that a 360 fps display is needed in the worst case for the multiple images defect to merge into blur. Then the display would appear as if infinite up-sampling were employed. A longer persistence phosphor could probably be used to compensate for a lower display frame rate without causing more blur than continuous interpolation, as long as the decay time constant is still substantially shorter than T.

An electronic display whose elements hold a constant intensity until changed at refresh time has long been sought [ENGS 35] but has yet to become practical. If such a display were available, the refresh rate would not need to be very high at all. Operated at $1/T$ fps, it would directly implement a sample-and-hold interpolation filter. The image defects would be identical to those of infinite up-sampling on a CRT.

8.4 Requirements of the Ideal Transmitter

The specifications of the displayed image having been set first, in accordance with the needs of the HVS, the transmitter end of the system (including the camera) can then be designed to supply the receiver with the information necessary for image reconstruction. The two viewpoints concerning the camera parallel those of the ideal display.

A necessary condition for a subjectively perfect reproduction on a hypothetical transparent display is that all of the spatiotemporal frequencies within the window of visibility, as pictured in Figure 3.5, have to be passed from the original scene to the display. Without intelligent signal processing the camera and channel bandwidth, and hence the sample rate, needed to preserve moving spatial resolution up to, say, 20°/sec would be ridiculous. Uncoded transmission is wasteful because in any spatiotemporal neighborhood of the image, only a small part of the spatiotemporal spectrum corresponding to the local velocity is utilized. In a motion-compensated system, frame rates can be much lower due to adaptation. The temporal bandwidth in the direction of motion is usually very narrow, so that the Nyquist criterion is effectively not violated even when the camera integration time is made very short ($h_1 \longrightarrow \mathrm{Fd}$) to preserve sharpness.

Getting back to the canonical model, the problem might be to determine the best way to filter before sampling for a given transmission frame rate $1/T$. Wider filters reduce the visibility of multiple images during fixation and small-area flicker (when $1/T \lesssim 30$ fps) at the expense of increased blur. Conventional practice has been to prefer blur over poor motion rendition, but HDTV may require a different tradeoff.

There is not really much freedom to select h_1 in ordinary cameras. It has to be a flat-topped pulse of duration no longer than T. However, this restriction is removed if the camera is operated at a rate faster than $1/T$ (oversampled), so that h_1 can be implemented as a digital filter.

8.5 Differences between HDTV and the Observations

Although no HDTV images have actually been observed during this research, it is speculated that defects in HDTV will differ only in degree, not in kind, from those observed on the HRTV. Conclusions based on the HRTV have been very conservative in that they either propose very high frame rates or else predict irreconcilable tradeoffs. When natural images are observed, as opposed to test patterns, several factors will tend to ameliorate motion-induced defects. First, many objects in the real world lack sharp edges. Second, spatial masking due to image busyness may help conceal defects when they do occur. Third, due to the limited depth of field in the optical system of cameras, objects in the background and foreground are more than likely to be defocused anyway. For these reasons, the findings based on HRTV observations may be too pessimistic.

However, one area in which advanced image transmission systems should improve over present-day television is viewing angle. Therefore, as previously hypothesized, blur under ocular pursuit may be more easily detected than in ordinary television viewing or in HRTV observations. Whether this is true hinges on dynamic acuity, which will be reexamined in the next section.

8.6 Problems to be Resolved

It has become painfully obvious that aspects of image perception involving the time domain are difficult to observe and quantify. As a result, there are still many gaps in the knowledge essential for a truly vision-based design of a transmission system for time-varying images.

An example of a problem which should be resolved using subjective considerations is frame rate conversion by fractional ratios, e.g., NTSC to PAL [BALD 76], or telecine. Interlace has been avoided completely in this study, but owing to its long history it deserves further scrutiny as a camera, channel, or display scanning method. In retrospect, the two-dimensional approach of the HRTV was found to preclude subjective assessment of overall image quality; however, the apparatus could still be useful for more fundamental psychophysical studies of contrast sensitivity to moving gratings, temporal masking, apparent motion, etc. One last problem which is raised by motion compensation is determining what novel defects may be generated and what the bounds on tolerable defects would be.

The following five issues are more closely related to the investigations of this thesis than the above. The first concerns presampling and interpolation filters. Because only four types of filters were used for h_1 and only four for h_2, the observed differences in defects between filters has not been shown to be caused by differences in the actual shape of the filter, for among filters for the same T, the width varies as well as the shape. More work is necessary to reach a conclusion about Gaussian vs. sample-and-hold, for instance. There may also exist better shapes.

Second, eye movements that are unrelated in velocity to the object on which defects are observed should be considered more carefully, for only scant attention has been given to this issue. Some of the conclusions drawn in this thesis may need to be modified. Following the development in Chapter 3, the spatial reflection $h'_{2x}(x)$ can be incorporated into $g'(x)|_t$, resulting in an effective impulse response for arbitrarily chosen v_1, v_2. The generalized impulse response when $v_1 + v_2$ is not too small should be approximately

$$g''(x)|_t \approx h_3(-\tfrac{x}{v_1+v_2})T \sum_{n=-\infty}^{\infty} h'_{2x}(x + (v_1 + v_2)(t - nT)),$$

which reduces to $g'(x)|_t$ if $v_2 = 0$.

Third, motion-induced defects in color images have to be studied because what applies to monochrome most likely does not apply to color. Miyahara found that during pursuit the velocity threshold for what he called jerkiness was six times higher for a magenta object than a white one [MIYA 75]. The difference might be due to properties of color vision, although the lower luminance of the colored object and possibly the additional blur imparted by slow red and blue phosphors may have been partially responsible for the reported effect.

Fourth, to better assess the frequency and amount of motion defects in HDTV, there is a need for some statistics on velocity and duration of motion in images. Existing measurements of source characteristics are more concerned with interframe correlations than with velocities [COLL 76][KRET 52]. Data is needed not only for broadcast television [TADO 68] but also for 70 mm motion pictures, since the characteristics of future HDTV programming are expected to differ from those of television today.

Finally, the main assumption of this thesis, as stated in Section 2.2.3, remains to be validated or qualified for anticipated HDTV viewing conditions through experiment. The required spatial resolution for moving objects should be measured using natural-looking synthetic scenes with controllable parameters such as v. It may be a simple matter to determine the threshold at which blur is detectable, but the point at which blur becomes objectionable is an entirely different, subjective, and perhaps more vital, question. If it turns out that images blurred by $v_1 T$ in the direction of motion are acceptable on the average, then there is not much point in further investigations of methods to increase the spatial resolution of moving objects without artifacts (although, recall that even $h_1 = \mathrm{Fr}$ is not always sufficient to assure good motion rendition).

8.7 Summary

The main points of the thesis were the following.

- When the observer tracks a moving object, nearly undiminished spatial acuity implies that the spatiotemporal window of visibility is extended far beyond its static limits. The conclusions and recommendations of this research are contingent on this assumption.

- The subjective appearance of a temporal convolution is that of the corresponding spatial convolution.

- The principal defects can be distinguished as blur, multiple images, large-area flicker, and small-area flicker. There are also jerkiness, tilt, and contrary motion. Depending on the observer's fixation or pursuit, these defects differ in appearance and occupy distinct portions of the image spectrum.

- Jerkiness is mediated by the motion pathway of vision, which is distinct from channels carrying spatiotemporal brightness information. However, jerkiness can generally be disregarded because it occurs at much lower transmission frame rates than other artifacts of motion quantization.

- Broadening the presampling filter increases blur, attenuates multiple images and small-area flicker, but does not affect large-area flicker.

- Broadening the interpolation filter attenuates multiple images under pursuit (except complete absence of interpolation precludes the defect), small-area flicker, and large-area flicker. It also increases blur, but has no effect on multiple images during fixation.

- When blur is considered on an equal basis with any other defect, the filters which maximize the region of defect-free reproduction of a moving step image in the velocity/frame-rate plane are the narrower ones. No transmission frame rate is sufficient at velocities typical of television, and tradeoffs do not seem favorable below 60 fps.

- A display frame rate of 120 fps is not enough for "transparency." Multiple images during pursuit are a display artifact.

- The nature of motion-induced defects in HDTV can be extrapolated from HRTV observations. A transparent display is not feasible, but a transparent camera and channel may be possible with adaptive signal processing.

Some of the statements above may now seem obvious without having to be based on observations of test images; however, the value of the experiments was that they enabled such facts to be recognized. Although the conclusions given here are far from establishing optimum tradeoffs for acceptable image reproduction, the two original objectives have been largely fulfilled. Defects have been separated by appearance and correlated with signal characteristics, and velocity/frame-rate thresholds for the defects have been measured for a specific subset of television systems.

Since the whole point of television, after all, is to reproduce moving images, it is hoped that temporal resolution will no longer be neglected in the evaluation of imaging systems and that properties of vision will be used to analyze and eventually conquer motion-induced degradations of temporally sampled images.

Appendix A

Implementation of HRTV Facility

A technical description of the HRTV system appears in this appendix. The hardware consists of the PDP-11/34 IPS computer, the PCTV frame buffer, the HRTV interface, and the high-speed monitor. The principal software components, running under Unix, are `vh.c`, which simulates the canonical image transmission system, and `hrtv.c`, a driver which facilitates interfacing to the HRTV hardware.

A.1 Hardware

A.1.1 General Description

The IPS system and PCTV were preexisting components and are documented elsewhere [TROX 81][ALSI 79]. This section provides a medium-level description of the PCTV output [HSU 84] and the HRTV interface.

If vertical blanking is ignored for ease of explanation, the format of the video data provided by the PCTV is simply a sequence of 512 pels per line, 512 lines per frame, $\approx$ 30 frames per second. All frames are identical. Basically, the HRTV interface samples a different PCTV line during each successive HRTV frame period (1/120 sec) and replicates that line 120 times on the screen of the high-speed monitor.

Since the line rate of the HRTV is identical to that of the PCTV (16 KHz, almost NTSC) and the frame rate is exactly four times faster than the PCTV, the HRTV is synchronous with the PCTV.

The user has a choice of display modes. In 512-mode an image sequence 512/120 = 4.3 sec long can be displayed. In 256-mode either the first half (the B channel) or the second half (the A channel) can be displayed alone. When start button B or A is depressed (A, in 512-mode), the first four frames are displayed cyclically; when released the rest of the frames are displayed. At the end the last frame is frozen, unless repeat mode is enabled to restart the sequence automatically.

A.1.2 High-speed Monitor

The CRT display is a Ball Miratel Model TE12 [BALL 74] monochrome monitor with two changes. The first was to increase the vertical sweep rate from the usual 60 Hz to 120 Hz, by adjusting an internal potentiometer. The second was to replace the original diode clamp DC restorer with an active black-level bias amplifier with feedback gated during the horizontal blanking period. Without this enhancement the cathode bias would vary enormously with change in image content. The black level was then adjusted to make the screen look completely dark under the observation conditions. The white level with DAC inputs set to 255 was measured to be about 16 ft-L with a Photo Research Litemate III/Spotmate telephotometer.

The nonlinear transfer characteristic of the CRT was compensated by a digital look-up table (tone scale memory) whose contents were derived from measurements of screen intensity with 17 equally-spaced DAC input codes. An Eastman Color Monitor Analyzer was used to measure intensity.

The half-life of the phosphor was estimated at 180 μs using a UDT PIN-10 photodiode. Because the diode was measured with a high-impedance load, its voltage was not linear with intensity, but the precise time constant was not of concern.

The look-up table and DC restoration may not have been completely successful in making the screen intensity proportional to DAC input code. A fine spatial or temporal grating alternating between levels X and Y did not necessarily appear equal in brightness to a constant patch of intensity $(X+Y)/2$, particularly if one of X and Y were 0. The main drawback was that constant subjective brightness could not be maintained as T was varied (with $h_2 = Fd$), in apparent contradiction to the Talbot-Plateau law. Moreover, when the CRT was switched from, say, intensity X at 120 fps to intensity TX at $120/T$ fps the display not only increased in brightness but also gave a momentary bright flash at the switch instant. This might be explained by a hardware problem or the transient response of the HVS.

A.1.3 Notes on the HRTV Interface

This section is to be read only in conjunction with the full circuit and timing diagrams [HSU 85]. The sole purpose is to comment on aspects which may not be evident from the drawings, not to give a complete explanation of operation.

Each HRTV frame is approximately synchronous with half of a PCTV field, and two fields make up an interlaced PCTV frame. During each half-field one of the two line stores refreshes the display while the other acquires a selected line from the PCTV data stream; the roles are switched in successive half-fields.

The temporal counter (CTZ<7:0>) starts at 128 and runs up to 255, incremented once every four half-fields by TCLK. At the start of each half-field ($\overline{\text{RSYNC}}$) the line delay counter is preset from the temporal counter and advances until 255, at which time a line is acquired (GRAB). Hence, numbering the lines of the first ("odd") field $0, 2, 4, \ldots, 510$ and those of the second ("even") field $1, 3, 5, \ldots, 511$, the lines are acquired in the interleaved sequence: $254, 510, 255, 511, 252, 508, 253, 509, \ldots, 0, 256, 1, 257$.

The previous paragraph is true only in 512-mode. For 256-mode-B (256-mode-A) the delay presets range from 128 to 191 (192 to 255), so that only the first (second) half of the interleaved sequence is acquired and displayed.

The horizontal counter is preset to CTX<9:0> = 512–109 upon detecting PCTV horizontal sync ($\overline{\text{HSYNC}}$). After 109 clocks, PCTV pel data becomes valid for 512 clocks; then CTX overflows. HHUNT low prevents preset by an equalization pulse in the middle of a line. The composite sync PROM synthesizes three sync waveforms for the high speed monitor and produces $\overline{\text{SAMPLE}}$ during CTX = 557.

If the PCTV sync is active during CTX = 557, a vertical serration is present in the left half of the line; that is how vertical sync is detected to produce $\overline{\text{VSYNC}}$, which clears the vertical counters (CTY<7:0>). VHUNT (= CTY8) low prevents clear during the second and third lines of serrations. The vertical format PROM selects normal sync + video output (SYNCSEL = 0), normal sync + blanking (1), equalization + blanking (2), or serration + blanking (3), in the correct sequence to display two HRTV frames during one field. It also generates $\overline{\text{RSYNC}}$ when CTY = 10 and 128, to signal the imminent start of a new half-field. TCLK is generated when VSYNC is discovered during an even value of CTY, that is, only after an even field. (In the event that the PCTV is running in a noninterlaced mode with only one field per frame, all fields are of the even type and there are only 256 distinct video lines. The temporal counter will then be advanced by TCLK every *two* half-fields.)

A.2 Software

A.2.1 Interface to the HRTV

Due to the interleaved access scheme of the HRTV, image data to be displayed sequentially must be written into the PCTV frame buffer in a special order. Also, since the first four frames are shown cyclically while a start button is depressed, those frames should be loaded with identical data. The routines in `hrtv.c` form a driver for the HRTV hardware which insulates the programmer from the interleaving and the special case of frames 0-3.

There are eight modes for loading the HRTV, and the HRTV interface must be set manually to 256-mode or 512-mode as specified. Mode 0 displays 509 frames with the HRTV in 512-mode, and mode 1 displays 254 frames palindromically, also in 512-mode (this was rarely used). Modes 2 and 3 display 253 frames in the A and B channels, respectively, with the HRTV in 256-mode. Adding 5 to the mode number reflects the image in the spatial dimension.

A.2.2 Transmission System Simulation

As outlined in Chapter 4, the program **vh** synthesizes a moving test pattern from user-supplied parameters. Instructions for using **vh** are found in the source code. Much of the complexity of the algorithm was for eliminating unneeded computation without restricting the stationary background to be constant intensity. The resulting program takes about 6 sec to generate and load a typical 256-mode image. The rapid response time is a great convenience when a group of images is to be compared, since at most two separate images can be loaded for immediate display on the HRTV.

The amount of computation would be very high if the convolutions with h_1 and h_2 had to be computed for every point in the synthesized display image s_4. However, when the moving image $s_1(t,x) = s_0(x - v_1 t)$ is purely translating, the camera image

$$s_2(t,x) = s_1(t,x) * h_1(t) = (s_0 * h'_{1x})(x - v_1 t)$$

is also purely translating so that only one time instant needs to be computed, by filtering s_0 with the reflected impulse response $h'_{1x}(x)$.

The displayed image can be computed by

$$s_4(t,x) = A(t,x) * h_2(t)$$

where

$$A(t,x) \equiv T \sum_{n=-\infty}^{\infty} \delta(t - nT)(s_0 * h'_{1x})(x - nv_1 T) = A(t - T, x - v_1 T)$$

is a periodic signal, and

$$h_2(t) = \sum_{m=0}^{N-1} \delta(t - \tfrac{mT}{r}) f_m, \quad r = \text{upsampling rate}, \quad \sum f_m = 1.$$

Since A is periodic, s_4 is periodic in the same way and needs to be computed only within the interval $0 \le t < T$. The discreteness of s_4 means that it has to be computed only for $t = 0, \frac{T}{r}, \ldots, \frac{(r-1)T}{r}$, and h_2 can be implemented as a digital polyphase filter. Pels of s_4 outside the interval are obtained by the appropriate spatial shift.

Complications arise because, in the extended model of Figure 4.2, the moving image s_{1b} contains both stationary and moving parts and is not purely translating. The object represented by s_0 occupies some finite interval of space, and $s_0(x)$ should be defined to be 0 outside that interval. A mask function $m(x)$ is defined to be 1 outside that interval and 0 inside. If the object is opaque,

$$s_{1b}(t,x) = s_0(x - v_1 t) + s_b(x)m(x - v_1 t).$$

(Fully transparent points in the object can be handled by setting $s_0(x) = 0$ and $m(x) = 1$ at that point.)

Now, the camera image

$$\begin{aligned} s_2(t,x) &= s_{1b}(t,x) * h_1(t) \\ &= (s_0 * h'_{1x})(x - v_1 t) + s_b(x)[(m * h'_{1x})(x - v_1 t)] \end{aligned}$$

is not purely translating but it is a linear combination of two signals that are. The displayed image is then

$$s_4(t,x) = \underbrace{A(t,x) * h_2(t)}_{A'(t,x)} + s_b(x)\underbrace{[B(t,x) * h_2(t)]}_{B'(t,x)}$$

where

$$B(t,x) \equiv T \sum_{n=-\infty}^{\infty} \delta(t - nT)(m * h'_{1x})(x - nv_1 T) = B(t - T, x - v_1 T).$$

It follows that A' and B' are periodic in the same way as A and B, and they need to be computed only for $t = 0, \frac{T}{r}, \ldots, \frac{(r-1)T}{r}$. When A' and B' are precomputed and saved, only one multiplication and one addition is needed to get each output pel in s_4.

For a large part of s_4, both operations can be avoided. Because $s_0 = 0$ outside some interval and h'_{1x} is of finite width, there must exist L and R such that for $x < L$ or $x > R$, $A'(t,x) = 0$ and $B'(t,x) =$ (a constant depending only on t) for $t = 0, \frac{T}{r}, \ldots, \frac{(r-1)T}{r}$. For such x a precomputed value of $s_b(x)\cdot$ (constant depending on t) is simply copied to the output. The result of this optimization is that the multiplication and addition are performed only for pels on or near the moving object, which is typically a small fraction of the total number of pels in s_4.

Bibliography

[ALLN 83] Allnatt, J., *Transmitted-Picture Assessment* (New York: Wiley, 1983).

[ALPE 62] Alpern, M., "Types of Movement," ch. 5, in *The Eye,* vol. 3, H. Davson, ed. (New York: Academic Press, 1962), pp. 63-142.

[ALSI 79] Alsip, D.A., "An Advanced Color Image Display Processor," S.M. Thesis (MIT Dept. of Electr. Eng., June 1979).

[ANST 78] Anstis, S.M., "Apparent Movement," ch. 21, in *Handbook of Sensory Physiology, Vol VIII: Perception,* R. Held, H.W. Leibowitz and H. Teuber, eds. (New York: Springer-Verlag, 1978), pp. 655-673.

[ANST 80] Anstis, S.M., "The Perception of Apparent Movement," *Phil. Trans. R. Soc. London* **B 290**, no. 1038 (July 1980), pp. 153-168.

[BALD 40] Baldwin, M.W., "The Subjective Sharpness of Simulated Television Images," *Proc. IRE* **28**, no. 10 (Oct. 1940), pp. 458-468.

[BALD 76] Baldwin, J.L.E., "Digital Standards Conversion," *IBA Tech. Rev.* **8**, Digital Video Processing – DICE (Sept. 1976).

[BALL 74] TE Video Monitor Instruction Manual, IM1006 Rev. H (Ball Brothers Miratel Div., Aug. 1974).

[BANK 72] Banks, W.P. and Kane, D.A., "Discontinuity of Seen Motion Reduces the Visual Motion Aftereffect," *Perception and Psychophysics* **12**, no. 1B (July 1972), pp. 69-72.

[BECK 72] Beck, J. and Stevens, A., "An Aftereffect to Discrete Stimuli Producing Apparent Movement and Succession," *Perception and Psychophysics* **12**, no. 6 (Dec. 1972), pp. 482-486.

[BRAD 78] Braddick, O. and Adlard, A., "Apparent Motion and the Motion Detector," in *Visual Psychophysics and Physiology,* J.C. Armington, ed. (Academic Press, 1978), pp. 417-426.

[BRAI 67] Brainard, R.C., Mounts, F.W., and Prasada, B., "Low-resolution TV: Subjective Effects of Frame Repetition and Picture Replenishment," *Bell Syst. Tech. J.* **46**, no. 1 (Jan. 1967), pp. 261-271.

[BRAU 66A] Braunstein, M.L., "Interaction of Flicker and Apparent Motion," *J. Opt. Soc. Am.* **56**, no. 6 (June 1966), pp. 835-836.

[BRAU 66B] Braunstein, M.L. and Coleman, O.F., "Perception of Temporal Patterns as Spatial Patterns During Apparent Movement," *Am. Psychologist* **21**, no. 7 (1966), p. 645.

[BROW 65A] Brown, J.L., "Flicker and Intermittent Stimulation," ch. 10, in *Vision and Visual Perception,* C.H. Graham, ed. (New York: Wiley, 1965), pp. 251-320.

[BROW 65B] Brown, J.L., "Afterimages," ch. 17, in *Vision and Visual Perception,* C.H. Graham, ed. (New York: Wiley, 1965), pp. 251-320.

[BROW 67] Brown, E.F., "Low-resolution TV: Subjective Comparison of Interlaced and Noninterlaced Pictures," *Bell Syst. Tech. J.* **46**, no. 1 (Jan. 1967), pp. 199-232.

[CHAT 54] Chatterjee, N.R., "Study of Delta Movement in Motion Pictures," *Indian J. Psych.* **29** (1954), pp. 155-159.

[COLL 76] Coll, D.C. and Choma, G.K., "Image Activity Characteristics of Broadcast TV," *IEEE Trans. Commun.* **24**, no. 10 (Oct. 1976), pp. 1201-1206.

[COLT 80] Coltheart, M., "The Persistence of Vision," *Phil. Trans. R. Soc. London* **B 290**, no. 1038 (July 1980), pp. 57-69.

[CORN 70] Cornsweet, T.N., *Visual Perception* (New York: Academic Press, 1970).

[CUNN 63] Cunningham, J.E., "Temporal Filtering of Motion Pictures," Sc.D. Thesis (MIT Dept. of Electr. Eng., May 1963).

[DITC 73] Ditchburn, R.W., *Eye-movements and Visual Phenomena* (London: Oxford Univ. Press, 1973).

[DUBO 81] Dubois, E., Prasada, B., and Sabri, M.S., "Image Sequence Coding," ch. 3, in *Image Sequence Analysis,* T.S. Huang, ed., (New York: Springer-Verlag, 1981), pp. 229-287.

[ENGS 35] Engstrom, E.W., "A Study of Television Image Characteristics, Part 2: Determination of Frame Frequency for Television in Terms of Flicker Characteristics," *Proc. IRE* **23**, no. 4 (Apr. 1935), pp. 295-310.

[FERR 92] Ferry, E.S., "Persistence of Vision," *Am. J. Sci.* **144**, no. 261 (Sept. 1892), pp. 192-207.

[FRIS 72] Frisby, J.P., "Real and Apparent Movement: the Same or Different Mechanisms?," *Vision Res.* **12**, (1972), pp. 1051-1055.

[FUJI 82] Fujio, T., "Future Broadcasting and High-definition Television," *NHK Tech. Monogr.,* no. 32 (Tokyo, June 1982), pp. 5-13.

[GLEN 83] Glenn, W.E. and Glenn, K.G., "Masking Effects in Imaging," SMPTE Psychophysical Committee "White Paper" presented at Optical Society of America meetings (New Orleans: Oct. 20, 1983), pp. 1-10.

[GRAH 65] Graham, C.H., "Perception of Movement," ch. 20, in *Vision and Visual Perception,* C.H. Graham, ed. (New York: Wiley, 1965), 575-588.

[GREG 66] Gregory, R.L., *Eye and Brain* (New York: McGraw-Hill, 1966).

[GUIL 54] Guilford, J.P., *Psychometric Methods,* 2nd Ed. (New York: McGraw-Hill, 1954).

[HSU 84] Hsu, S.C., "Interfacing External Hardware to the IPS PCTV," CIPG Memo PCTV-23 (Apr. 1, 1984).

[HSU 85] Hsu, S.C., CIPG Drawings File, no. MISC-18.

[ISHI 82] Ishida, T. and Masuko, H., "Development of High-definition Television Equipment, Part 6: 70-mm Film Laser Telecine," *NHK Tech. Monogr.,* no. 32 (Tokyo, June 1982), pp. 57-61.

[ISON 79] Isono, H., "A Re-examination of Contrast Threshold Differences Between Spatial Sine-wave and Square-wave Gratings," *Vision Res.* **19**, no. 5 (1979), pp. 603-607.

[ISON 84] Isono, H., "A New Objective Method of Evaluating Image Sharpness," *NHK Lab. Note,* no. 296 (Feb. 1984).

[JOHN 78] Johnston, R., Mastronardi, J., and Mony, G., "A Digital Television Sequence Store," *IEEE Trans. Commun.* **26**, no. 5 (May 1978), pp. 594-600.

[KELL 72] Kelly, D.H., "Adaptation Effects on Spatio-temporal Sine-wave Thresholds," *Vision Res.* **12**, no. 1 (Jan. 1972), pp. 89-101.

[KINT 72] Kintz, R.T. and Witzel, R.F., "Role of Eye Movements in the Perception of Apparent Motion," *J. Opt. Soc. Am.* **62**, no. 10 (Oct. 1972), pp. 1237-1238.

[KRET 52] Kretzmer, E.R., "Statistics of Television Signals," *Bell Syst. Tech. J.* **31**, no. 4 (July 1952), pp. 751-763.

[LEE 83] Lee, C. and Wang, J., "An Overview of the Temporal and Vertical Interpolators," CIPG Memo ATRP-T-17 (Dec. 15, 1983).

[LEGR 57] Le Grand, Y., *Light, Color, and Vision,* English Translation (New York: Dover, 1957).

[LEGR 67] Le Grand, Y., *Form and Space Vision,* English Translation (Bloomington: Indiana Univ. Press, 1967).

[LETT 59] Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., and Pitts, W.H., "What the Frog's Eye Tells the Frog's Brain," *Proc. IRE* **47**, no. 11 (Nov. 1959), pp. 1940-1951.

[LIMB 71] Limb, J.O. and Pease, R.F.W., "A Simple Interframe Coder for Video Telephony," *Bell Syst. Tech. J.* **50**, no. 6 (July 1971), pp. 1877-1888.

[LUDV 47] Ludvigh, E., "Visibility of the Deer Fly in Flight," *Science* **105** (1947), pp. 176-177.

[LUDV 58] Ludvigh, E. and Miller, J.W., "Study of Visual Acuity During the Ocular Pursuit of Moving Test Objects, I: Introduction," *J. Opt. Soc. Am.* **48**, no. 11 (Nov. 1958), pp. 799-802.

[MIYA 75] Miyahara, M., "Analysis of Perception of Motion in Television Signals and its Application to Bandwidth Compression," *IEEE Trans. Commun.* **23**, no. 7 (July 1975), pp. 761-768.

[MORG 79] Morgan, M.J., "Perception of Continuity in Stroboscopic Motion," *Vision Res.* **19**, no. 5 (1979), pp. 491-500.

[MORG 80A] Morgan, M.J., "Analogue Models of Motion Perception," *Phil. Trans. R. Soc. London* **B 290**, no. 1038 (July 1980), pp. 117-135.

[MORG 80B] Morgan, M.J., "Spatio-Temporal Filtering and the Interpolation Effect in Apparent Motion," *Perception* **9** (1980), pp. 161-174.

[PEL 83] Video Sequence Processor (VSP) Brief Technical Description (Palo Alto, CA: Picture Element Limited (PEL), 1983).

[PRAT 78] Pratt, W.K., *Digital Image Processing* (New York: Wiley-Interscience, 1978).

[RABI 78] Rabiner, L.R. and Schafer, R.W, *Digital Processing of Speech Signals* (Englewood Cliffs, NJ: Prentice-Hall, 1978).

[RATZ 80] Ratzel, J.N., "The Discrete Representation of Spatially Continuous Images," Ph.D. Thesis (MIT Dept. of Electr. Eng., Aug. 1980).

[RIGG 65] Riggs, L.A., "Visual Acuity," ch. 11, in *Vision and Visual Perception,* C.H. Graham, ed. (New York: Wiley, 1965), pp. 251-320.

[ROBS 66] Robson, J.G., "Spatial and Temporal Contrast-Sensitivity Functions of the Visual System," *J. Opt. Soc. Am.* **56**, no. 8 (Aug. 1966), pp. 1141-1142.

[SCHO 67] Schouten, J.F., "Subjective Stroboscopy and a Model of Visual Movement Detection," in *Models for the Perception of Speech and Visual Form,* W. Wathen-Dunn, ed. (MIT Press, 1967), pp. 44-55.

[SCHR 83] Schreiber, W.F., "TV Systems with Storage: Some Preliminary Thoughts," CIPG Memo ATRP-T-2 (May 12, 1983).

[SCHR 84] Schreiber, W.F., "CIPG First Year Progress Report," CIPG Memo ATRP-T-31 (June 13, 1984).

[SCHR 84] Schreiber, W.F., "Psychophysics and the Improvement of Image Quality," *SMPTE J.* **93**, no. 8 (Aug. 1984), pp. 717-725.

[SCHR 85] Schreiber, W.F. and Troxel, D.E., "Transformation Between Continuous and Discrete Representations of Images: A Perceptual Approach," *IEEE Trans. Pattern Anal. & Mach. Intell.* **7**, no. 2 (Mar. 1985), pp. 178-186.

[SEKU 75] Sekuler, R., "Visual Motion Perception," in *Handbook of Perception, Vol. 5: Seeing,* E.C. Carterette, ed. (New York: Academic Press, 1975).

[SEYL 65] Seyler, A.J. and Budrikis, Z.L., "Detail Perception after Scene Changes in Television Presentations," *IEEE Trans. Inf. Theory* **11**, no. 1 (Jan. 1965), pp. 31-43.

[STAE 83] Staelin, D.H., "Proposal to the Advanced Television Research Program for Initiation of Research Concerning Resolution-Preserving Interpolation of Video Frames," CIPG Internal Memorandum (July 27, 1983).

[TADO 68] Tadokoro, Y. *et. al.,* "Moving Velocity of Objects and its Visual Effects through Television" (in Japanese) NHK Tech. Rept. 11 (Sept. 1968), pp. 422-426.

[TONG 83] Tonge, G.J., "Signal Processing for Higher Definition Television," IBA Rept. E8D 7/83 USA (1983).

[TROX 81] Troxel, D.E., "An Interactive Image Processing System," *IEEE Trans. Pattern Anal. & Mach. Intell.* **3**, no. 1 (Jan. 1981), pp. 95-101.

[ULLM 77] Ullman, S., "The Interpretation of Visual Motion," Ph.D. Thesis (MIT Dept. of Electr. Eng., May 1977).

[VAND 58] Van den Brink, G., "The Visibility of Details of a Moving Object," *Optica Acta* **5**, Supplement: International Colloquium on Physical Problems of Colour Television (Jan. 1958), pp. 44-49.

[WATS 83A] Watson, A.B., Ahumada, A.A., and Farrell, J.E., "The Window of Visibility: A Psychophysical Theory of Fidelity in Time-Sampled Visual Motion Displays," NASA Tech. Paper 2211 (1983).

[WATS 83B] Watson, A.B. and Ahumada, A.J., "A Look at Motion in the Frequency Domain," in *Proc. ACM SIGGRAPH/SIGART Interdisciplinary Workshop on Motion: Representation and Perception* (Toronto, Apr. 1983), pp. 1-10.

[WEST 75] Westheimer, G. and McKee, S.P., "Visual Acuity in the Presence of Retinal-Image Motion," *J. Opt. Soc. Am.* **65**, no. 7 (July 1975), pp. 847-850.

[WIND 83] Windram, M.D., Morcom, R., and Hurley, T., "Extended-Definition MAC," *IBA Tech. Rev.* **21** (Nov. 1983), pp. 27-41.

[WITT 55] Wittel, O. and Haefele, D.G., "Continuous Projection Problems," *SMPTE J.* **64** (June 1955), pp. 321-323.

[YARB 67] Yarbus, A.L., *Eye Movements and Vision,* English Translation (New York: Plenum, 1967).

[YUYA 82] Yuyama, I., "Large-Screen Effects," *NHK Tech. Monogr.,* no. 32 (Tokyo, June 1982), pp. 14-20.