INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313-761-4700  800-521-0600
Perception and filtering of interventional x-ray fluoroscopy image sequences

Aufrichtig, Richard, Ph.D.

Case Western Reserve University, 1994
PERCEPTION AND FILTERING
OF INTERVENTIONAL X-RAY
FLUOROSCOPY IMAGE SEQUENCES

by

RICHARD AUFRICHTIG

Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Thesis Advisor: David L. Wilson, Ph.D.

Department of Biomedical Engineering

CASE WESTERN RESERVE UNIVERSITY

May 1994
CASE WESTERN RESERVE UNIVERSITY

GRADUATE STUDIES

We hereby approve the thesis of

Richard Aufrichtig

candidate for the Ph.D.

degree.*

(signed) David L. Wilson
(chair)

Cecil L. Thorn

Sharon M. Simore

Mark D. Bagustanski

John P. Henry

date April 6, 1994

*We also certify that written approval has been obtained for any proprietary material contained therein.
PERCEPTION AND FILTERING OF INTERVENTIONAL X-ray FLUOROSCOPY IMAGE SEQUENCES

Abstract

by

RICHARD AUFRICHTIG

Due to the proliferation of interventional radiology the use of x-ray fluoroscopy is increasing. Examples of treatments include balloon angioplasty, atherectomy and neuroembolizations. X-ray fluoroscopic imaging at high frame rates and low x-ray dose is used to guide catheters, position interventional devices, and give feedback during intervention. We use image perception experiments, mathematical modeling and image processing to determine ways to reduce x-ray dose and optimize visualization.

The human visual system (HVS) temporally filters image sequences, and effectively improves the signal-to-noise ratio (SNR) of an image sequence. It is important to include this phenomenon in the analysis of fluoroscopy. Paired-comparison, min-contrast, and forced-choice perception experiments are used to compare pulsed fluoroscopy at reduced frame rates with conventional continuous fluoroscopy. Experimentally we find average dose savings of 22%, 38%, and 49%, for pulsed at 15, 10 and 7.5 acq/sec, respectively. Dose savings depend on object size, with less savings for smaller objects.

Theoretically we extend the framework on an ideal observer with three
models of the spatio-temporal response of the HVS: (M1) separable, (M2) non-separable, and (M3) non-separable with internal noise. With no free parameters, M1 predicts the average dose savings within a 3% difference, but does not describe the effect of object size. M2 and M3 explain the influence of size, and M3, with a single free parameter, fits the measurements best.

In similar experiments with single frames, we find that processing in the HVS effectively reduces noise variance by a factor of 3, corresponding to an effective averaging time of 100 msec. We suggest several variance reduction techniques to create a single last-image-hold frame having perception equal to the fluoroscopy sequence.

Simply reducing the x-ray dose decreases the SNR to unacceptable low levels. We describe an image processing technique to enhance fluoroscopy sequences. The method consists of object detection, followed by spatial and recursive, temporal filtering. Based upon the object detection we adjust filter parameters, such as to reduce noise without blurring the objects. A receiver operating characteristic (ROC) curve analysis is used to optimize matched filters for the object detection. Dependent upon the noise level, the method improves SNR from 6-10 dB, with dose savings approaching 95%.
To my Parents
I want to thank several people for their help, contributions and encour-
agements. First of all I thank my research and academic advisor, Dr. David
L. Wilson. Dave has been more than an advisor for me, — he has been
a true friend and a great mentor. I am very grateful for his strong sup-
port and wonderful guidance. He has in every way possible facilitated my
research and taught me the principles of medical imaging.

I want to thank the rest of my Ph.D. committee. Dr. Cecil W. Thomas
has shared many of his ideas and insights in signal and image processing.
as well as in visual perception. His office has always been open to me.
and he has provided me with not only technical advice, but also personal
encouragement. Dr. Grover C. Gilmore introduced me to the theories of
visual perception. and was a help in developing many of the perception
experiments. Dr. Mark S. Rzeszotarski taught me my first class in medical
imaging, and played a key role in making me decide for this research project.

I want to thank Dr. John R. Haaga for allowing me and fellow students to do experiments in the Department of Radiology at University Hospitals of Cleveland. I also thank Dr. Thomas R. Vrobel of MetroHealth Medical Center for providing the fluoroscopy image sequences used in this work.

I want to show my tremendous appreciation to Ping Xue. Over the last two years Ping and I have worked closely together. Ping has carried out many of the experiments reported in this work, and is continuing with his own research in this area. He has been a tremendous support, always willing to discuss experiments and theoretical derivations, any time of the day (or night).

I also want to thank all the graduate students that have passed through, or still work in the Biomedical Imaging Laboratory for their support and ideas, and for making the lab a pleasant research environment. In particular
I thank Mike Labbe, Paritosh Dhawale, and Ken Kump for offering help, suggestions, and good friendship.

I want to thank some special friends from CWRU: Ditte Wolin, Dave Igel, Wynee Tsao, Rich Diefes, and Allyson Rosen. They all contributed in making my stay in Cleveland particularly worthwhile.

I want to show my appreciation to Dr. Arne Talbot, who was the first to introduce me to Biomedical Engineering in general and CWRU in particular. I thank Dr. Michael R. Neuman and Dr. P. Hunter Peckham for inviting me to CWRU in the summer of 1989. It was this visit that convinced me to return to Cleveland and perform research in biomedical engineering.

I am very grateful for the financial support I received in my first years from the Danish Research Academy and the Knud Højgaard Foundation, and later from grants to David L. Wilson provided by the Whitaker Foun-
dation and the National Institutes of Health.

Finally, I want to thank my parents for their continuous encouragements.

support and believes in me.
Contents

1 Introduction  

2 Perceptual Comparison of Pulsed and Continuous Fluoroscopy  
  2.1 Introduction ........................................ 8  
  2.2 Materials and Methods ............................. 9  
     2.2.1 Equipment ................................... 9  
     2.2.2 Paired-Comparison and Min-Contrast Experiments . 11  
     2.2.3 Forced-Choice Experiments ................. 19  
  2.3 Results ........................................... 23  
     2.3.1 Paired-Comparison .......................... 23  
     2.3.2 Min-Contrast ................................. 28  
     2.3.3 Forced-Choice ............................... 33  
     2.3.4 Comparison of Methods and Dose Savings ...... 35  
  2.4 Discussion ....................................... 38  
     2.4.1 Comparison of Experiments ................. 38  
     2.4.2 Dose for Pulsed Fluoroscopy ............... 44  

3 Perception of Fluoroscopy Last-Image-Hold  
  3.1 Introduction .................................... 48  
  3.2 Material and Methods ............................ 49
4 A Model for Perception of Pulsed Fluoroscopy Image Sequences

4.1 Introduction ....................................... 71
4.2 Theory ............................................ 72
4.3 Simulation Models ................................. 81
  4.3.1 Model 1 ........................................ 82
  4.3.2 Model 2 ........................................ 83
  4.3.3 Model 3 ........................................ 87
4.4 Review Of Experimental Measurements .............. 88
4.5 Comparison Of Theory And Experiments ............. 91
4.6 Discussion ......................................... 95

5 Spatio-Temporal Filtering of X-ray Fluoroscopy using Object Detection .... 101

5.1 Introduction ...................................... 101
5.2 Algorithm ......................................... 105
  5.2.1 Object Enhancement ............................ 105
  5.2.2 Spatial Filtering ............................... 112
  5.2.3 Temporal Filtering ............................. 113
5.3 Methods ........................................... 115
  5.3.1 Test Images ................................... 115
  5.3.2 Object Enhancement ............................ 118
  5.3.3 Evaluation of Spatio-temporal Filtering .......... 124
5.4 Results ........................................... 127
6 Conclusion and Future Directions

6.1 Conclusion ........................................... 147
6.2 Future Directions ................................. 149
List of Figures

2.1 Paired-comparison phantom ............................................. 12
2.2 Plotted are the number of subjects that prefer either pulsed
    or continuous when comparing all disk sizes in the paired-
    comparison study ..................................................... 25
2.3 The fraction of subjects selecting pulsed-15 over continuous
    at Q/acq is plotted as a function of the pulsed acquisition dose  26
2.4 The paired-comparison equivalent-perception dose is plotted
    as a function of disk radius for pulsed-7.5, pulsed-10, and
    pulsed-15 ........................................................... 28
2.5 We compare the effect of interlaced and non-interlaced dis-
    play on the equivalent-perception dose for pulsed-7.5, pulsed-
    10, and pulsed-15 .................................................... 29
2.6 Average min-contrasts are plotted as a function of disk radius
    and dose/acq for interlaced display ................................ 31
2.7 Average min-contrasts are plotted as a function of disk radius
    and acquisition mode (pulsed-7.5, pulsed-10, and pulsed-15)
    for an interlaced display ......................................... 32
2.8 Average min-contrasts are plotted as a function of acquisition
    dose for pulsed-15, interlaced ..................................... 33
2.9 The fraction of correct responses from the forced-choice ex-
    periment are plotted as a function of disk contrast ........... 35
2.10 For a single subject, $u$ estimates are plotted as a function of dose for *pulsed-15* .................. 36  
2.11 We compare disk detection as measured with forced-choice and min-contrast measurements .................. 37  
2.12 We compare dose savings predicted by the paired-comparison, min-contrast and forced-choice measurements .................. 39  
2.13 Theoretical detection curves for an m-AFC experiment are plotted as a function of $m$ and disk contrast .................. 41  

3.1 The paired-comparison disk phantom .................. 51  
3.2 The paired-comparison cylinder phantom .................. 52  
3.3 Number of *single-frame* or *continuous* selections .................. 57  
3.4 The fraction of subjects selecting *single-frame* is plotted as a function of the *single-frame* acquisition dose .................. 59  
3.5 Equivalent-perception dose values are plotted as a function of diameter for the case of disks and cylinders .................. 61  
3.6 At equivalent-perception dose values, cylinder diameter is plotted as a function of disk diameter .................. 62  
3.7 Average min-contrasts are plotted as a function of disk and cylinder diameter and dose per acquisition .................. 69  
3.8 A method for obtaining a non-discrete equivalent-perception dose from min-contrast data for the case of a cylinder of diameter 5 pixel .................. 70  

4.1 The ideal observer model modified to characterize human detectability in image sequences .................. 75  
4.2 The temporal response of the human visual system as obtained from human image-perception flicker experiments .. 83
4.3 For each of four acquisition types, frequency power spectra are given for the input and output of the HVS temporal filter.

4.4 A perspective view of the surface and a contour map of the effect of spatio-temporal filtering by the HVS.

4.5 Equivalent-perception dose values predicted by Model 1 are compared with paired-comparison, min-contrast and forced-choice experiments for a disk of radius 8 pixels.

4.6 As a function of disk size and acquisition rate, the equivalent-perception dose for pulsed predicted by Model 2 is compared to paired-comparison experiments.

4.7 As a function of disk size and acquisition rate, the equivalent-perception dose for pulsed predicted by Model 3 is compared to paired-comparison experiments.

5.1 Flow diagram of the image sequence enhancement algorithm.

5.2 The oriented matched filter method to enhance line structures in a noisy fluoroscopy image.

5.3 The spatial filtering process with a $3 \times 3$ kernel.

5.4 The LUT used to control spatial and temporal filtering.

5.5 A set of hypothetical ROC curves.

5.6 Effect of object enhancement methods.

5.7 Computed ROC curves after enhancement.

5.8 An example of a manual segmentation, and of a thresholded matched filter output.

5.9 Three frames from a noisy fluoroscopy sequence.

5.10 Output from the object enhancement.

5.11 Output from the object guided spatio-temporal filter.
5.12 Improvement in SNR is plotted as a function of the frame number ................. 136

5.13 A guide catheter and guide wire with a corresponding cross-sectional profile .......... 140
List of Tables

2.1 Experimental conditions for the paired-comparison and min-contrast experiments .................................. 16

2.2 Experimental conditions for the forced-choice experiment... 21

2.3 For the paired-comparison experiments, we give the equivalent-perception dose and the 95% confidence interval .............. 47

3.1 For the paired-comparison experiments, we list equivalent-perception dose values .................................... 66

5.1 Fraction of standard dose for 6 test sequences ............... 118

5.2 Integrated area under the ROC curve obtained by matched filtering ....................................................... 129

5.3 Integrated area under the ROC curve obtained by unsharp mask processing ........................................... 131

5.4 The average improvement in SNR obtained by filtering ... 138

5.5 Estimated percent dose savings for the 6 image sequences obtained by the combined spatio-temporal filter ........... 139
Chapter 1

Introduction

Using x-ray fluoroscopy, cardiologists and radiologists perform interventional procedures such as balloon angioplasty, atherectomy, blood vessel embolization, transjugular intrahepatic portosystemic shunts (TIPS), biliary drainage, catheter ablation, etc. As the number, complexity, and duration of interventional procedures increases, x-ray fluoroscopy dose becomes an issue. Procedures lasting 4 hours with fluoroscopy times over 100 min are documented [1, 2], and occasionally neuro-embolization procedures last 6 to 8 hours. Earlier reports of radiation dose in angiography indicate that most dose is obtained from high-dose diagnostic angiography acquisitions rather than fluoroscopy [3, 4], but these studies do not include angioplasty or contrast power injectors. (Power injectors save x-ray dose to staff by allowing staff to exit the room during diagnostic acquisitions.) Staff dose in
angioplasty can far exceed that of a diagnostic exam [5], and, in coronary angioplasty studies, 74% of the dose to patients is obtained from fluoroscopy [6].

Fluoroscopy is also used for diagnosis in such studies as gastrointestinal exams, where a particularly large dose is used [7], and voiding cystograms of pediatric patients where dose to young patients is an issue.

Fluoroscopy images are quantum limited. Given the normal dose at the input to an image intensifier, the size of a pixel, and the acquisition rate, the average number of photons captured in a pixel is approximately 35 [8, 9]. Poisson statistics are obeyed, giving a standard deviation of 6, or 17% of the mean. Simply reducing dose increases the percent noise to unacceptable levels, particularly because interventional work creates a need for improved image quality.

There are several techniques for reducing patient and staff dose in interventional fluoroscopy imaging. One technique is the x-ray fovea where a semi-transparent collimator with a hole in the center attenuates the x-ray exposure in the periphery of the image [10, 11, 12, 13]. The center is imaged at full x-ray exposure, while the periphery is imaged with decreased x-ray exposure. An algorithm that accounts for beam hardening compen-
sates the image brightness in the periphery. The technique is targeted to interventional procedures where the center of the image is of principle interest and the image content in the periphery is either required or desirable. The exposure to patients is reduced by approximately 70%, while scattered exposure to operators is reduced by approximately 60%. A similar device is described by Rudin and Bednarek [14]. In this dissertation other dose saving methods are discussed.

One technique suggested for reducing x-ray dose is low-frame-rate pulsed fluoroscopy, hereafter called pulsed, where short x-ray pulses are used and images are acquired at reduced frame rates [15, 16, 17, 18, 19]. Typically, images are acquired at 15 acq/sec (image acquisitions per second) and displayed with gap-filling to provide a 30 frames/sec display; we call this pulsed-15. Systems with acquisitions rates of 10 and 7.5 acq/sec (pulsed-10 and pulsed-7.5) are also used. Pulsed fluoroscopy has other advantages such as reduced single-image motion blur and improved image quality of a last-image-hold image, but it has the disadvantage of a “choppy” display. Some complain about the choppy display being uncomfortable to watch, particularly when the image is panned. but hand-eye coordination is not greatly affected [20]. In at least one dose-sensitive institution, pulsed-15
is used almost exclusively in cardiac angiography applications. With the reduced motion in non-cardiac angiography, pulsed may have even greater applicability. In the case of pediatric radiology, pulsed fluoroscopy is being marketed as a dose savings feature.

Despite the growing availability of pulsed, a proper x-ray dose has not yet been established. In Chapter 2 we compare pulsed to conventional continuous fluoroscopy at 30 acq/sec, hereafter called continuous. (Note that for the purpose of imaging stationary objects, continuous and pulsed at 30 acq/sec are identical.) We maintain that one must consider the temporal response of the human visual system (HVS). For example, if the HVS averages over two frames, then a dose Q/acq at 30 acq/sec will give perception equivalent to a dose 2 Q/acq at 15 acq/sec. Note that if pulsed-15 is calibrated in this manner, there is no savings in the dose per minute. We establish a dose for pulsed called the equivalent-perception dose, that gives visualization equivalent to that of continuous. We use these measurements to calculate the dose savings from pulsed.

A useful feature on many modern x-ray fluoroscopy systems is last-image-hold (LIH) where a constant image is maintained on the monitor.

---

1Personal communication with Dr. Thomas R. Vrobel, Department of Medicine, MetroHealth Medical Center, Cleveland.
after x-ray exposure is terminated. LIH is a dose savings feature as it allows physicians to contemplate the last image and plan the next move in an interventional procedure. The use of LIH was recommended by the “Workshop on Fluoroscopy: Strategies for Improvements in Performance, Radiation Safety and Control” sponsored by the American College of Radiology (ACR) and the FDA [21].

One wants image quality in an LIH frame to be at least as good as in the preceding sequence of images. In the case of a sequence, temporal filtering in the human visual system (HVS) reduces perceived noise. This phenomenon is easily demonstrated by viewing a cine-loop sequence of noisy images containing low-contrast objects. When the sequence stops on a single LIH frame, there is no more temporal processing, and low-contrast objects disappear. In order to create an LIH frame with the same perceived image quality, noise reduction techniques must be applied. In Chapter 3 we determine the single-frame dose that gives perception equal to a continuous reference display, and we suggest some appropriate noise reduction techniques for the LIH frame.

In Chapter 4 we extend the ideal observer model to assess the performance of continuous and pulsed fluoroscopy. The ideal observer has its
roots in statistical decision theory and has been widely applied in radiology to evaluate imaging modalities [22, 23, 24, 25], image processing techniques [26, 27, 28, 29, 30], and image reconstruction methods [31, 32]. The model incorporates the spatio-temporal filter response of the human visual system (HVS). We study the visual perception of pulsed fluoroscopy image sequences, and the goal is to determine the effect of the HVS response on dose savings.

Another dose saving technique is to reduce dose and use image processing to compensate for the additional noise. To reduce noise, pixels nearby in space or time are typically combined in some fashion. Summing 10 pixels, each with 35 photons, gives 350 counts, and the standard deviation of the Poisson distribution is reduced from 17% to 5% of the mean. Simple averaging, however, usually results in unacceptable image blurring.

Image processing solutions on some existing medical products consist of time-domain filtering, time-domain filtering with motion detection to reduce motion blur, and spatial filtering for contrast enhancement of objects. “Motion” is detected whenever a pixel intensity changes more than a prescribed limit between the current frame and the last, and pixels are temporally filtered only where there is no motion. Unfortunately, image
noise often results in false "motion" detection. Typically, with an obese patient, the skin dose maximum is reached; the images are noisy because of the decreased number of x-ray quanta detected; "motion" is detected; and no smoothing is done.

In Chapter 5 we describe a method based upon spatio-temporal filtering with object detection. Particularly in non-cardiac angiography, most motion occurs in isolated, long, thin objects (catheters, guide wires, etc.). We demonstrate that such structures can be detected and roughly segmented using a matched filtering approach. We then locally adjust spatial and temporal filtering parameters to reduce object blurring. In contrast to motion-detection filtering using image differences, this scheme distinguishes between soft tissue and catheter motion.
Chapter 2

Perceptual Comparison of Pulsed and Continuous Fluoroscopy

2.1 Introduction

In this perception study, we determine the equivalent-perception dose for pulsed when viewing stationary disks. We use disk stimuli because they are frequently applied in single-frame perception studies [28, 33, 34, 35]. Because the disks are stationary, the experiments model clinical procedures with relatively little motion.

Using computer-generated phantoms, we investigate several independent variables, including interlaced and non-interlaced displays; acquisition rates of 7.5, 10, 15, and 30 acq/sec; and dose. Images are displayed sequen-
tially on a video screen in a continuous loop with the last frame followed by
the first (a cine loop). We examine low-contrast object detectability using
forced-choice experiments similar to many previous experiments on single
image frames [28, 36, 26, 37, 38] and to some recent experiments on cine
loops [39]. Such experiments are very time consuming when a cine display
is involved; hence, we also conduct more time-efficient paired-comparison
experiments where subjects compare the relative visibility of a contrast-
detail phantom in pulsed and continuous displays. Lastly, in min-contrast
experiments, subjects record the minimally detectable disk contrast.

2.2 Materials and Methods

2.2.1 Equipment

In this study, computer generated images are displayed on the video tachis-
toscope, a unique device developed by one of us (CWT) for image perception
studies. The video tachistoscope displays $512 \times 512 \times 8$ bit pixel
images. It offers variable frame times, starting as low as 1/417 sec. With
this flexibility, it is possible to operate in a non-interlaced 30 frames/sec
mode, or in a simulated 60 fields/sec interlaced mode. The tachistoscope is
controlled by a personal computer (Gateway 486 with 16 MBytes of RAM.)
Sioux City, South Dakota). On the video screen, a pixel is \( \approx 0.5 \) mm and the largest disk displayed is 1.6 cm in diameter. The viewing distance is 150 cm, approximately the minimal viewing distance in a cath lab. The average luminance of our display is 9.5 \( ft\cdot L \) at a gray-scale value of 134. Over the entire range from 0 to 255, the gray-scale to luminance curve for the video tachistoscope has a sigmoidal shape. However, a linear approximation is valid over the range of gray-scale values used. Over this range, one can convert pixel gray-scale values to luminance by multiplying by the factor 0.22\( ft\cdot L/gray-scale \) and subtracting an offset of 20\( ft\cdot L \).

For the paired-comparison and min-contrast experiments, we use 10 subjects, all graduate students or faculty in the Department of Biomedical Engineering, CWRU. The average age is 26 \( \pm 4 \) years, and seven subjects are naive to the hypotheses. For the forced-choice experiment, three subjects (age 27 \( \pm 4 \) years) with normal visual acuity are used. Before the experiment, subjects are familiarized with the image sequences and the experimental procedure. Experiments are performed in a darkened room.
2.2.2 Paired-Comparison and Min-Contrast Experiments

In the paired-comparison experiments, the task is to compare the visibility of just-detectable disks in a contrast-detail phantom displayed in pulsed and continuous modes on either side of a single monitor. Using the same split-screen phantom, subjects also determine the disk which is just detectable or the min-contrast.

Design of Phantom

The contrast-detail phantom consists of disks of various sizes and contrasts on a noisy flat background (Fig. 2.1). Poisson noise approximating fluoroscopy is obtained using a random number generator [40].

The disk contrast, $C$, is defined below where $\mu_b$ and $\mu_d$ are the mean gray-scale values of the background and disk, respectively.

$$C = \frac{|\mu_b - \mu_d|}{\mu_b + \mu_d} \quad (2.1)$$

We assume the number of detected x-rays, $N$, to be Poisson distributed around a mean of $N_0$. $N \sim \text{Poisson}(N_0)$, and a linear conversion between $N$ and gray-scale ($N = \lambda \times \text{gray}\_scale$, where $\lambda$ is a constant). The noise
Figure 2.1: Paired-comparison phantom. The contrast of the disks is varied linearly along the y-axis. In 8 bit gray-scale, the mean values are 132, 131, 129, 127, 126, 124, 123 and 121 on a mean background of 134. The noise level of continuous is \( \sigma = 21.9 \) corresponding to a dose Q/acq. When the dose for pulsed varies from 1.0 to 2.4 Q/acq in steps of 0.2 Q/acq, the \( \sigma \)'s are 21.9, 20.0, 18.5, 17.3, 16.3, 15.5, 14.7, and 14.1, respectively. The contrast values specified on the figure are approximative. The exact disk contrasts can be calculated from the mean gray-scale values using Eq. 2.1. For clarity, noise is not added in this figure.

distribution of the gray-scale values is

\[
\text{gray-scale} \sim \frac{1}{\lambda} \text{Poisson}\{\lambda \mu\} \tag{2.2}
\]
where \( \mu \) is the mean gray-scale value and \( \lambda \mu = N_0 \). Finally, the gray-scale variance, \( \sigma^2 \), is
\[
\sigma^2 = \frac{N_0}{\lambda^2} = \frac{\mu}{\lambda}.
\]  

(2.3)

The noise in continuous mimics a dose \( Q/\text{acq} \) while the dose in pulsed varies from 1.0 to 2.4 \( Q/\text{acq} \). To maintain a constant image luminance while changing the noise level, we keep \( \mu \) constant and vary \( \lambda \). A noise level for continuous is established as described below: this gives a value \( \lambda_c \) corresponding to a dose \( Q/\text{acq} \). For a dose \( bQ/\text{acq} \), we use \( b\lambda_c \).

The baseline noise level of the contrast-detail phantom is established from measurements on a digital fluoroscopy x-ray unit (Siemens Polytron). Five images of a flat copper phantom (2.58 mm and 68 kVp) are obtained. The five are averaged to estimate the local mean, and the result is subtracted from single frames. No patterns are observed in these noisy images, and the variance is measured in several regions of interest (ROI’s). The value of \( \lambda \) is obtained from Eq. 2.3 where \( \mu = 140 \) is a mean taken from the five frames and \( \sigma^2 = 492 \) is \( 5/6 \) of the variance measured in a single subtracted noise image. The factor of \( 5/6 \) corrects the measured variance for the subtraction of a noisy estimate of the mean obtained from only 5 frames. This gives \( \lambda = 0.28 \pm 0.02 \), where the error estimate is calculated
from multiple ROI measurements. The mean number of counts per pixel 
\( N_0 = \lambda \mu \) is estimated to be 39. A similar number is obtained from a 
theoretical calculation using data such as pixel size and image intensifier 
dose [41].

For \textit{continuous}, we set \( \lambda_c = 0.28 \). To minimize the possibility of noise 
truncation, we set the mean value of disks in the center row of the contrast-
detail phantom to be 127, the mid-range gray-scale value of the 8 bit image. 
The mean background becomes \( \mu = 134 \). With these values, \( N_0 = 37 \) 
counts, and the standard deviation of the noise is \( \sigma = 22 \) gray-scales. For 
\textit{pulsed}, dose varies from 1.0 to 2.4 Q/acq giving \( \lambda \)'s from 0.28 to 0.67, 
respectively.

**Experimental Procedure**

The control acquisition mode (\textit{continuous}) and one of the three test modes 
(\textit{pulsed-15}, \textit{pulsed-10}, or \textit{pulsed-7.5}) are displayed simultaneously side-by-
side on a single screen. We use either non-interlaced 30 frames/sec or inter-
laced 30 frames/sec display. This requires gap-filling for the \textit{pulsed} acqui-
sitions. For example. a \textit{pulsed-15} acquisition with a non-interlaced display 
consists of the sequence: new noise frame, repeated frame, new noise frame,
repeated frame, etc., and each frame is on the screen for 1/30 sec. In the
interlaced case, the sequence is: new odd noise field, new even noise field.
repeated odd noise field, repeated even noise field, new odd noise field.
etc., where the odd and even fields make a frame and where each frame is
1/30 sec. Similarly, for pulsed-10 and pulsed-7.5, there are 2 and 3 gap-filling
frames, respectively. Image sequences are shown in continuous cine-loops.

Subjects are asked the following three questions in the order listed.

**Q1:** How many disks in each column can you detect?

**Q2:** For each disk size, which display (right or left) gives the better visi-

**Q3:** Considering all disk sizes, which display (right or left) gives the better visi-

For Q2 and Q3, the subjects are instructed to focus on disks having the
lowest visibility. The responses to Q2 and Q3 comprise what we call the
paired-comparison experiments. and the responses to Q1 comprise the min-

contrast experiments.

Presentations of different dose levels follow a Latin-square design [42],
and subjects have no time limit for their response. In Table 2.1, we specify
<table>
<thead>
<tr>
<th>Display Type</th>
<th>Acquisition Rate (acq/s)</th>
<th>Dose/acq (Q/acq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interlaced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 (C)</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>15 (P)</td>
<td>1.0, 1.2, 1.4, 1.6, 1.8, 2.0</td>
<td></td>
</tr>
<tr>
<td>10 (P)</td>
<td>1.6, 1.8, 2.0, 2.2, 2.4</td>
<td></td>
</tr>
<tr>
<td>7.5 (P)</td>
<td>1.6, 1.8, 2.0, 2.2, 2.4</td>
<td></td>
</tr>
<tr>
<td>Interlaced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 (C)</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>15 (P)</td>
<td>1.0, 1.2, 1.4, 1.6, 1.8, 2.0</td>
<td></td>
</tr>
<tr>
<td>10 (P)</td>
<td>1.6, 1.8, 2.0, 2.2, 2.4</td>
<td></td>
</tr>
<tr>
<td>7.5 (P)</td>
<td>1.6, 1.8, 2.0, 2.2, 2.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Experimental conditions for the paired-comparison and min-contrast experiments. In parentheses (C) and (P) refer to *continuous* and *pulsed* acquisition modes, respectively.

the experiments in terms of the independent variables.

In paired-comparison control experiments, we determine the number of frames required in the cine loop. For the case of *pulsed*-7.5, the slowest acquisition rate, we compare different numbers of new frames in a split-screen display. We compare 16 versus 4, 16 versus 8, 32 versus 4, 32 versus 8, and 32 versus 16. Three subjects view the images. In all cases, we find no obvious, consistent differences in perception. We also perform control
experiments for faster acquisition rates; i.e. pulsed-15 and continuous. In this case, 3 subjects view continuous and pulsed-15 at 2 Q/acq with cine loops consisting of 1, 2, 3, 5, 10 and 20 frames. With 2 and 3 frames, the image sequence gives a patterned jumpy appearance while no such pattern is evident with more frames. More disks are detected as the number of frames increases from 1 to 5. No significant difference is found between 10 and 20 frames. In paired-comparison and min-contrast experiments, we use between 8 and 32 frames.

When the bigger disks (radius 4, 8, and 16) are viewed in repeated experiments with the same display, observers tend to see the same number of disks, and each observer has a maximum deviation of ± 1 disk. However, for the smallest disk (radius 2) responses vary between 0 and 3 disks. Because of this high inter- and intra-observer variability, we exclude the small disk from subsequent analysis.

Data Analysis

Two analyses are applied to the paired-comparison data (responses to Q2 and Q3. above). First, we analyze the responses using hypothesis testing for a binomial distribution [19, 42]. We expect the visibility of pulsed at
higher dose levels to be better than continuous. Treating the decisions of the 10 subjects as realizations of a binomial distribution, $X \in B\{10, p\}$, where $p$ is the probability of preferring pulsed, we state the following null and alternative hypotheses:

$$H_0 : p \leq 1/2 \text{ vs. } H_1 : p > 1/2$$

Using a 5% significance level, we reject the null hypothesis if 8 or more subjects prefer pulsed [19, 42]. The same test method is used at low dose levels of pulsed where we expect a preference for continuous.

In the second analysis, we determine the real-valued equivalent-perception dose where 50% of the subjects prefer pulsed, and 50% prefer continuous. We fit the data using a PROBIT model which is often used to fit dose-response experiments in biology. In our case, the independent variable is pulsed dose, and the dependent variable is the fraction preferring pulsed. The PROBIT model consists of a modified integral of a Gaussian. The mean and variance of the Gaussian are estimated using a maximum likelihood method [43]. The Macintosh version of the SPSS program (SPSS 4.0, SPSS Inc., Chicago, Illinois) is used [44]. One output is a plot of the probability of subjects preferring pulsed with a best-fit PROBIT curve.
superimposed. Another output is a list of probabilities with confidence intervals in dose/acq. The equivalent-perception dose, and 95% confidence interval, is obtained from the line indicating a 50% probability for preferring pulsed. The program evaluates the fit of the model to the data using a Pearson Goodness of Fit test.

In the case of min-contrast, we record the number of disks seen (response to Q1) and convert this number to contrast. Contrast values from all 10 subjects are averaged and plotted.

To determine the equivalent-perception dose for pulsed, we fit a second-order polynomial to the average min-contrasts plotted as a function of dose. The equivalent-perception dose is where the curve equals the min-contrast from continuous.

2.2.3 Forced-Choice Experiments

Experimental Procedure

In the forced-choice experiments, a low-contrast disk is placed in one of 4 possible positions on the monitor, and the subject chooses its location with a verbal response. Using the notation of Green and Swets [45], this is a 4-alternative forced-choice experiment, or a 4-AFC experiment. Four rather
than more alternatives simplify the selection task. The disk is placed in each of the four positions an equal number of times in random order. To reduce experiment time, the screen is split into quarters, and 4 trials are presented with each screen presentation.

We use one disk size (radius = 8), and the independent variables are acquisition-type (*continuous and pulsed-15*), dose, and disk contrast (Table 2.2). In preliminary experiments, we use 4 additional contrasts to determine values for maximal (ceiling) and minimal (floor) responses. The display is non-interlaced. We use 48 trials for each combination of independent variables, giving a total of 768 trials ( = 4 disk contrasts × 4 acquisitions × 48 trials) or 192 screen presentations for each subject. The subject has unlimited time to examine the image and is not informed of the acquisition mode or disk contrast presented. Each subject usually has four 1.5 hour sessions.

**Data Analysis**

For each set of independent variables, we record the number of trials (48) and the number of correct choices. The fraction of correctly detected disks is computed and plotted versus disk contrast. Data are fitted to the detection
<table>
<thead>
<tr>
<th>Acquisition Mode</th>
<th>Acquisition Rate (acq/s)</th>
<th>Dose/acq (Q/acq)</th>
<th>Disk Radius (pixels)</th>
<th>Disk Contrast (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuous</td>
<td>30</td>
<td>1.0</td>
<td>8</td>
<td>0.3, 0.5, 0.75, 1.0</td>
</tr>
<tr>
<td>pulsed-15</td>
<td>15</td>
<td>1.2, 1.5, 1.8</td>
<td>8</td>
<td>0.3, 0.5, 0.75, 1.0</td>
</tr>
</tbody>
</table>

Table 2.2: Experimental conditions for the forced-choice experiment. The display type is non-interlaced. Disk contrasts are on a mean background of 134. The noise level of continuous is $\sigma = 21.9$, corresponding to a dose 1.0 Q/acq.

curve model proposed by Ohara et al. (Appendix A of [37]).

Briefly, the model hypothesizes a continuous decision variable internal to the observer with Gaussian probability density functions for the choices: disk present and no-disk present. The “distance” between the means of these two overlapping distributions is $d' = uC$, where $C$ is contrast and $u$ is a parameter to be determined. For a fixed disk contrast, it becomes easier to discriminate between disk and no-disk when $u$ increases. Ohara et al. develop an equation that relates $u$ to the probability of a correct choice in
an m-alternative forced-choice (m-AFC) experiment (Eq. A15 of [37])

$$prob(correct) = \int_{t=-\infty}^{\infty} [\Phi(t)]^{m-1} \frac{e^{-u(C-t)^2/2}}{\sqrt{2\pi}} dt$$

(2.5)

where $\Phi(t)$ is the cumulative Gaussian distribution. Estimates of $u$ are obtained using a maximum-likelihood method (Appendix B of [37]). We determine the effect of dose on $u$ for pulsed-15 displays.

**Number of Frames**

We perform a forced-choice control experiment to determine the number of different frames required in the cine loop. The number of frames examined is 4, 8, 16, and 32. We use 7.5 acq/sec and vary the contrast so as to create a detectability curve of fraction correct versus contrast. For each set of independent variables, 24 trials are given to two subjects.

There is a small tendency to increase the value of $u$ as the number of frames increases. However, a Mantel-Haenszel chi-square test [46] finds no statistical difference between cine loops consisting of 8, 16, and 32 frames (5% significance level). There is a statistical difference between 4 and 32 frames. This is consistent with psycho-physical experiments that determine a visual-sensory-memory “storage time” of 250 ms [47].

In paired-comparison and min-contrast experiments we use a minimum
of 8 frames (see Section 2.2.2). In the case of forced-choice, we use cine
loops consisting of 20 frames.

Because a single presentation consists of multiple image frames, there are
practical limitations on the space for storing images and the time for loading
images. Our engineering solution is to create images using precomputed
noise data. Twenty image frames of noise data are computed once using a
Poisson noise generator [40] and stored on disk. We load this background
noise data into an array consisting of \(20 \times 512 \times 512\) elements. Likewise,
we read and store 20,000 pixel values for each of the 4 disk contrasts into
a \(20,000 \times 4\) element array. These 2 large arrays are sources of input data
for creating the output image frames. To ensure random realizations of
image sequences, starting pointers into the noise arrays are picked randomly.
Additionally, we use random seed values each time the noise generation
program is run.

2.3 Results

2.3.1 Paired-Comparison

The bar graphs in Fig. 2.2 are results from comparing all disk sizes to-
gether (Q3) under conditions of different display modes (interlaced and
non-interlaced) and acquisition methods (pulsed-15, pulsed-10, and pulsed-7.5). Fig. 2.2A is the case for pulsed-15, non-interlaced. At 1.6, 1.8, and 2 Q/acq; 7, 9, and 9 subjects prefer pulsed-15 to continuous. No subjects prefer pulsed-15 at Q/acq, and only two prefer it at 1.2 and 1.4 Q/acq. The equivalent-perception dose, i.e. the dose where preference changes from continuous to pulsed-15, occurs somewhere between 1.4 and 1.6 Q/acq with almost perfect subject unanimity. Stars indicate significance at the 5% level. Very similar results are obtained in the case of pulsed-15, interlaced (Fig. 2.2B). With pulsed-10 and pulsed-7.5, non-interlaced, the equivalent-perception dose increases to the intervals 1.8–2.0 Q/acq and 1.8–2.2 Q/acq, respectively. Excellent subject unanimity is obtained, and once again, there is remarkable similarity between interlaced and non-interlaced results.

A short-coming of the preceding analysis is that the equivalent-perception dose is obtained in a discrete fashion over relatively large intervals. The PROBIT analysis gives a real-valued, equivalent-perception dose using all of the data in a curve fit. Figure 2.3 is an example fit to data obtained from the comparison of all disk sizes (Q3) for the case pulsed-15, interlaced. The x-axis is dose/acq and the y-axis is the fraction selecting pulsed-15. The solid curve is the model, and the dashed curves are the
Figure 2.2: Plotted are the number of subjects that prefer either pulsed or continuous when comparing all disk sizes (question Q3) in the paired-comparison study. All bars have heights of 10 corresponding to the 10 subjects. The dose/acq of continuous is 1.0 Q/acq, while pulsed varies between 1.0 and 2.4 Q/acq as indicated in the legend. Stars indicate significance at a 5% level (N=10). Experimental conditions are: (A) pulsed-15, non-interlaced; (B) pulsed-15, interlaced; (C) pulsed-10, non-interlaced; (D) pulsed-10, interlaced; (E) pulsed-7.5, non-interlaced; and (F) pulsed-7.5, interlaced.

95% confidence intervals. Note that all measurements fall within the confidence range. The equivalent-perception dose, where the fraction is 0.5, is
Figure 2.3: The fraction of subjects selecting pulsed-15 over continuous at Q/acq is plotted as a function of the pulsed acquisition dose. In this case, all disks are considered (question Q3) and the display is interlaced. The stars are measured data points, and the solid curve is the estimated PROBIT fit. Dashed lines indicate the 95% confidence limits along the x-axis. The equivalent-perception dose for pulsed-15 occurs at 1.53 ± 0.10 Q/acq, where the fraction selecting pulsed-15 is 0.50.

1.53 Q/acq.

Table 2.3 lists equivalent-perception dose values and confidence intervals. Measurements include responses to the comparison of all disk sizes
(Q3) as well as the comparison at each radius (Q2). With the exception of
one (see Table 2.3), all data fits are adequate as determined by the Pearson
Goodness of Fit test.

Effects of disk radius and acquisition rate on the equivalent-perception
dose values are shown in Fig. 2.4. For pulsed-15 with both interlaced and
non-interlaced displays, the equivalent-perception dose does not obviously
change as the disk radius varies from 4 to 16 pixels. However, when the
acquisition rate is reduced to 10 or 7.5 acq/sec. the equivalent-perception
dose increases as the disk radius decreases from 16 to 4 pixels. At a given
disk radius, the pulsed acquisition rate also affects the equivalent-perception
dose with an increase in rate causing a decrease in dose. As a rule, the
smaller the disk radius, the more acquisition rate affects the equivalent-
perception dose.

Note that the equivalent-perception dose values from interlaced and non-
interlaced experiments are nearly equal at all disk sizes (Fig. 2.4). When
disks are compared collectively (Q3), virtually identical interlaced and non-
interlaced results are obtained (Fig. 2.5). The average absolute difference
is only 1.8%.
Figure 2.4: The paired-comparison equivalent-perception dose is plotted as a function of disk radius for pulsed-7.5, pulsed-10, and pulsed-15. Both interlaced (A) and non-interlaced (B) displays are used. Values are estimated using the PROBIT analysis (see Figure 2.3 and Table 2.3).

2.3.2 Min-Contrast

In Fig. 2.6, we analyze min-contrast data (responses to Q1) for the case of pulsed-15, interlaced. At all x-ray dose values, min-contrast decreases
Figure 2.5: We compare the effect of interlaced and non-interlaced display on the equivalent-perception dose for *pulsed-7.5*, *pulsed-10*, and *pulsed-15*. All disks are considered in this paired-comparison (question Q3). Values are estimated using the PROBIT analysis, and the error bars indicate 95% confidence intervals.

as disk size increases. (This is not surprising – bigger disks are easier to see.) With regard to dose, results in Fig. 2.6 are consistent with previous results. As dose increases, min-contrast decreases. With the largest radius
of 16 pixels, nearly equal responses are obtained because the larger disks are
easy to visualize and saturate the detectability response. At all disk radii.
the continuous data lie between min-contrasts at 1.4 and 1.6 Q/acq. This
indicates an equivalent-perception dose within this interval, thus supporting
the paired-comparison results. Once again, there is good agreement between
interlaced and non-interlaced (not shown). Similar plots are obtained for
pulsed-10 and pulsed-7.5 (not shown).

In Fig. 2.7, we plot min-contrast data versus disk radius and acquisition
mode (pulsed-7.5, pulsed-10, and pulsed-15). A constant dose of 1.8 Q/acq
is used. As acquisition rate increases, the average min-contrast decreases
because there is effectively more temporal filtering by the HVS. Also, once
again, bigger disks are easier to see. Interlaced and non-interlaced give very
similar responses. and similar plots are obtained at other dose values (not
shown).

We calculate equivalent-perception dose values from the min-contrast
experiments. In Fig. 2.8, the curve is a second-order polynomial fit to the
min-contrast data as a function of x-ray dose. The horizontal line is the
min-contrast from the continuous measurement, and the vertical drop gives
the equivalent-perception dose. Estimating equivalent-perception dose in
Figure 2.6: Average min-contrasts are plotted as a function of disk radius and dose/acq for interlaced display. Contrasts are given on top of the bars. Min-contrasts for continuous at 1.0 Q/acq are plotted between the 1.4 and 1.6 Q/acq pulsed-15 data in order to demonstrate that the equivalent-perception dose lies somewhere between. We use 10 subjects and question Q1. A similar plot is achieved with a non-interlaced display (not shown).

This fashion gives 1.60, 1.78, and 1.87 Q/acq for pulsed-15, pulsed-10, and pulsed-7.5, respectively, for the case of the interlaced display. Similarly, values are 1.57, 1.95, and 2.08 Q/acq, respectively, when the display is
Figure 2.7: Average min-contrasts are plotted as a function of disk radius and acquisition mode (pulsed-7.5, pulsed-10, and pulsed-15) for an interlaced display. Disk contrasts are given on top of the bars. We use a constant dose of 1.8 Q/acq, 10 subjects, and question Q1. As acquisition rate increases, min-contrast decreases. In all cases, min-contrast decreases with increasing disk size. A similar plot is achieved with a non-interlaced display (not shown).
Figure 2.8: Average min-contrasts are plotted as a function of acquisition dose for pulsed-15, interlaced (squares). The solid curve is a second order polynomial fit \( y = -0.34x^2 + 0.68x + 1.27 \). The min-contrast for continuous (1.0 Q/acq) is marked on the y-axis (circle) and projected horizontally to the pulsed-15 curve. The dose where the two curves intercept \( (\approx 1.6 \text{ Q/acq}) \) is the equivalent-perception dose.

2.3.3 Forced-Choice

In Fig. 2.9, we show detection data from the 4-AFC experiment. Data points are averages over all subjects without any data normalization. The four panels are different display modes and dose values. For each sub-
ject, maximum likelihood estimates of $u$ are obtained. The smooth curves are calculated using $u$ values averaged over the subjects. (Virtually identical results are achieved by first averaging the detection data, and then estimating $u$ from the average data.) Note that the $y$-intercept of $1/4$ corresponds to the probability of making a correct choice purely by chance in a 4-alternative forced-choice experiment. The error bars are standard deviations obtained by quadratic averaging of the standard deviations from single subjects, $[p(1 - p)/n]^{1/2}$, where a binomial distribution is assumed ($p$ is the probability of a correct choice, and $n$ is the number of trials) [37].

Since the detection curves for continuous and pulsed-15 at 1.5 Q/acq are nearly identical, this further confirms that the equivalent-perception dose for pulsed-15 is very near 1.5 Q/acq.

As described in Section 2.2.3, the parameter $u$ quantitates the detectability of a data set. In Fig. 2.10, we show $u$ estimates (squares) from a single subject for pulsed-15. The smooth curve is a second-order polynomial fit. Also plotted is the value of $u$ for continuous (circle). By projecting this value horizontally to the smooth curve and dropping a vertical line we obtain the equivalent-perception dose of 1.57 Q/acq. For pulsed-15, the mean for three subjects is $1.54 \pm 0.17$ Q/acq.
Figure 2.9: The fraction of correct responses from the forced-choice experiment are plotted as a function of disk contrast. Data points are averaged over 3 subjects, and the bars indicate standard deviations (see text). Also shown are theoretical detection curves (Eq. 2.5). Values for $u$ are 254 for continuous and 201, 252, and 288 for pulsed-15 at 1.2, 1.5 and 1.8 Q/acq, respectively. Values of $u$ are averages over the 3 subjects. Note that a shift to the left indicates better detectability. Note also that the y-intercept of 1/4 is the probability of making a correct choice purely by chance in this 4-AFC experiment.

2.3.4 Comparison of Methods and Dose Savings

Figure 2.11 compares pulsed-15, forced-choice and min-contrast measurements. On the left, six curves from the forced-choice experiment are plotted
Figure 2.10: For a single subject, $u$ estimates (Eq. 2.5) are plotted as a function of dose for pulsed-15 (squares). The solid curve is a second-order polynomial fit ($y = -57.7x^2 + 289x - 29.9$). The $u$ estimate for continuous (1.0 Q/acq) is plotted on the y-axis (circle) and is projected to the curve. The dose at the intersection (1.57 Q/acq) is the equivalent-perception dose marked by the vertical line.

Corresponding to detection probabilities (fraction correct) of 50% through 95%. Data are at 1.2, 1.5 and 1.8 Q/acq. As expected, detection probability increases with increasing disk contrast. The bold curve on the right represents the min-contrast data. Note that the slopes of all curves are sim-
Figure 2.11: We compare disk detection as measured with forced-choice and min-contrast measurements. The 6 curves on the left, labeled forced-choice, are predicted dose and contrast values when the probability of detection is varied from 50% to 95%. We use Eq. 2.5 with $u = 252$, as obtained from Figure 2.9. The thick curve on the right is min-contrast data averaged over 10 subjects. Note that the min-contrast data is shifted to the right of the forced-choice curves towards higher contrasts. The disk has a radius of 8 pixels.

Interestingly, the min-contrast data are at much higher contrasts than the forced-choice curves. One interpretation is that subjects in the min-contrast experiment maintain a relatively high threshold for disk detection (see Section 2.4.1).
With regard to the equivalent-perception dose, we find excellent agreement between all measurements. From the equivalent-perception dose, we calculate a dose savings per unit time. The fraction of the continuous dose is

\[ f = \frac{(\text{equivalent-perception dose/\(acq\times\text{acq/sec}\)_pulsed})}{(\text{dose/\(acq\times\text{acq/sec}\)_continuous})} \]  

(2.6)

and the percent savings is \((1 - f)\)100%. For example, from the forced-choice, pulsed-15 experiments. \(f = (1.54 \ Q/\text{acq} \times 15 \ \text{acq/sec})(Q/\text{acq} \times 30 \ \text{acq/sec}) = 0.77\). and the percent dose savings is 23%.

In Fig. 2.12A for the case of an interlaced display, we give dose savings predicted by paired-comparison and min-contrast. In Fig. 2.12B, we show dose savings from the non-interlaced experiments and include the forced-choice, pulsed-15 data value. Excellent agreement between measurements is obtained. and the average absolute difference in dose savings is 2.7%.

2.4 Discussion

2.4.1 Comparison of Experiments

We present data from three measurements: (1) paired-comparison. (2) min-contrast, and (3) forced-choice. As described below. these task-oriented experiments complement one another.
Figure 2.12: We compare dose savings predicted by the paired-comparison, min-contrast and forced-choice measurements. The forced-choice experiment is done only for pulsed-15, non-interlaced. Interlaced (A) and non-interlaced (B) results are remarkably similar, and the average absolute difference is 6.6%.

The 4-AFC experiment removes the subjectivity of an observer threshold and measures true object detectability on a statistical basis. Much lower
contrasts are detected than in the min-contrast measurement (Fig. 2.11). That is, the forced-choice data, even at a 95% correct detection level, lie at much lower contrasts than the min-contrast data.

Others [28, 36, 26, 37, 38] use an 18-AFC design and define threshold for detection to be where the fraction correct is 50%. In Fig. 2.13, we plot detection curves (Eq. 2.5) for several m-AFC (m-alternative forced-choice) experimental designs. We vary m from 4 to 18 and set \( u = 252.4 \), the value estimated for pulsed-15 at 1.5 Q/acq. The y-intercept is the probability of selecting the disk when one cannot see it (zero contrast), and 4-AFC and 18-AFC give y-intercepts of 1/4 and 1/18, respectively. In general, forced-choice detection curves shift toward higher contrasts as \( m \) increases. This behavior of detection curves has also been confirmed experimentally [48].

The curves predict quantitatively the effect of an 18-AFC versus a 4-AFC experiment. At a fraction correct of 0.5, the threshold contrast for 1.5 Q/acq increases from 0.0032 (4-AFC) to 0.0069 (18-AFC). Referring to Fig. 2.11, if we plot this predicted 18-AFC data point, it approaches the min-contrast data but is still less than 1/2 the corresponding value from the min-contrast curve. In conclusion, min-contrast and paired-comparison measurements result from visual stimuli which are detected quite reliably:
Figure 2.13: Theoretical detection curves for an m-AFC experiment are plotted as a function of m and disk contrast. The value of $u$ is 252 and $m$ is varied between 4 and 18.

i.e. supra-threshold stimuli.

One can argue that supra-threshold stimuli are appropriate since objects in x-ray fluoroscopy imaging are above threshold. That is, the task in fluoroscopy is not simply to detect a catheter: rather, one wants to visualize it and manipulate it. Thus, although forced-choice detectability
is a well-defined objective quantity, it may not be the most appropriate measurement.

As compared to min-contrast, paired-comparison is more discriminating. That is, many times one can “see” the same numbers of disks on the right and left sides of the screen, but one side is decidedly “easier” to see. Also, min-contrast requires an observer threshold criterion that varies from one time to the next. These subjective observations result in a relatively high inter- and intra-observer variability in the min-contrast experiments. Paired-comparison is much more reproducible.

Despite measuring threshold or supra-threshold responses, all three experimental paradigms give surprisingly consistent results. Equivalent perception dose and corresponding dose savings are similar in all three experiments (Fig. 2.12). With respect to disk radius, paired-comparison and min-contrast give similar results (Figs. 2.4, 2.6, and 2.7). Both methods show no effect of interlaced versus non-interlaced display (Figs. 2.2, 2.4, 2.6, 2.7, and 2.12).

The cine loop experiments in this study are more time consuming than single frame presentations. Control experiments with pulsed-15 indicate that the loop should consist of several image frames. Because of the time
required to generate the image data, load it into display memory, and run
the experiments. We opt for more time-efficient paired-comparison rather
than forced-choice experiments. Hence, with forced-choice, we do not test
all the independent variables such as disk size and acquisition rate. Because
of the remarkably consistent experimental results, we think that additional
forced-choice experiments are unnecessary.

Equation 2.5 provides a convenient way to evaluate perception results.

From our forced-choice experiments, we find that $u$ depends on the type of
acquisition (continuous versus pulsed) and the dose per acquisition. Others
find $u$ to depend upon unsharp mask processing [49], gray-scale windowing
[38], and background noise structure [37].

We find no difference between interlaced and non-interlaced with regard
to the equivalent perception dose and future dose-savings experiments can
be done with non-interlaced displays only. We have not, however, compared
absolute detectability of interlaced versus non-interlaced with regard to
noise processing in the HVS or to image resolution (see [50] for a discussion
of resolution).

We can compare our experimental results with those of Whiting et al.
[39]. Although they did not experiment at different dose levels or discuss
dose savings of pulsed, they report data which support our conclusions. At a fixed dose/acq, as frame rate increases from 15 to 30 acq/sec, they find a relatively small decrease in the 50% threshold contrast as measured with a 6-AFC experiment. Assuming a continuous decision variable internal to the subject which is proportional to the change in gray-scale of an object over the square root of the dose per acquisition ($\Delta_{\text{gray-scale}}/\sqrt{\text{dose}}$), ¹ we estimate a dose savings of 35% for the case of pulsed-15 (Fig. 4 of [39]). In addition, Whiting et al. study the number of image frames required in a cine loop to saturate the detectability response [39]. As compared to these measurements, our control experiments indicate a less pronounced effect (Section 2.2.2).

### 2.4.2 Dose for Pulsed Fluoroscopy

Computer simulations are useful for these perception experiments. It is difficult to obtain all necessary images on an x-ray system because most systems do not allow easy adjustment of fluoroscopy dose. Also, digital storage of multiple, high-resolution fluoroscopy frames is impossible on most.

¹This assumption is consistent with two, but not a third, extension to the ideal observer model that we propose for describing perception of low-contrast disks in noisy sequences [51]. It is also valid in the Rose model which describes perception of low-contrast disks in single stationary images [35, 24].
if not all. conventional systems. An additional problem is that video camera lag creates unknown image filtering. Thus, it appears that simulation experiments are a good way to evaluate dose requirements for pulsed.

With the same ability to visualize low-contrast objects, our experiments predict average dose savings of approximately 22% (15 acq/sec), 38% (10 acq/sec), and 49% (7.5 acq/sec). Because our experiments are done using stationary objects, the results may better apply to non-cardiac applications with relatively little motion. Examples include general angiography (neuro, thoracic, peripheral limb, etc.), gastrointestinal studies, pediatric radiology, etc. As compared to cardiac applications, velocities are reduced and motion tends to be localized to moving catheters, etc. Future perception experiments should include effects of motion.

In general, the equivalent-perception dose measured with min-contrast and paired-comparison experiments depends upon disk radius for pulsed-7.5 and pulsed-10 (Figs. 2.4, 2.6, and 2.7). As the size of the object decreases, the equivalent-perception dose increases giving a decrease in dose savings. However, in the case of pulsed-15 there is no effect of disk radius. Dose savings are predicted at all disk sizes. We choose to report dose savings from the larger disks because they more closely match the "size" of objects
in fluoroscopy imaging. [52] (Although a catheter, or artery, is thin, it is very long.) Note also, that a paired comparison of all disks together (Q3) gives results very similar to the larger disks.

We have reported an extension to the ideal observer model which includes the temporal filtering characteristics of the HVS [53]. This model describes the average response, but not the effect of radius, on equivalent-perception dose. In a more recent model [51] we incorporate the spatio-temporal filter response of the human visual system (HVS), as measured in human image perception flicker experiments. As in previous ideal observer models with single stationary images [24], the output from our modified ideal observer is a signal-to-noise ratio (SNR), which relates observer performance to the detectability of noisy low-contrast objects. Equating the SNR of pulsed with the SNR of continuous, allows one to compute the equivalent-perception dose. The model quite accurately predicts effects of both acquisition rate and disk radius on the equivalent perception dose. Thus, it is the interaction between the acquisition rate noise characteristic and the spatio-temporal response of the HVS which gives dose savings.
<table>
<thead>
<tr>
<th>Display Type</th>
<th>Disk Radius (pixels)</th>
<th>Pulsed-15</th>
<th>Pulsed-10</th>
<th>Pulsed-7.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.53</td>
<td>1.89</td>
<td>2.07</td>
<td>(1.43, 1.65)</td>
</tr>
<tr>
<td>Interlaced</td>
<td>1.54</td>
<td>1.75</td>
<td>1.89</td>
<td>(1.45, 1.64)</td>
</tr>
<tr>
<td></td>
<td>1.54</td>
<td>1.89</td>
<td>2.14</td>
<td>(1.45, 1.62)</td>
</tr>
<tr>
<td></td>
<td>1.49</td>
<td>2.20</td>
<td>2.45</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.52</td>
<td>1.87</td>
<td>2.10</td>
<td>(1.40, 1.65)</td>
</tr>
<tr>
<td>Non-interlaced</td>
<td>1.56</td>
<td>1.65</td>
<td>1.91</td>
<td>(1.43, 1.71)</td>
</tr>
<tr>
<td></td>
<td>1.55</td>
<td>1.95</td>
<td>2.06</td>
<td>(1.41, 1.68)</td>
</tr>
<tr>
<td></td>
<td>1.52</td>
<td>2.31</td>
<td>2.37</td>
<td>(1.40, 1.68)</td>
</tr>
</tbody>
</table>

Table 2.3: For the paired-comparison experiments, we give the equivalent-perception dose and the 95% confidence interval (in parentheses). Values are estimated by the PROBIT analysis, and data include responses to the comparison of all disk sizes (Q3) (indicated by All) as well as the comparison at each disk radius (Q2) (indicated by 16, 8, or 4). With one exception (marked with **) the PROBIT model adequately describes all sets of experimental data, as determined by the Pearson Goodness of Fit test.
Chapter 3

Perception of Fluoroscopy
Last-Image-Hold

3.1 Introduction

In previous reports, we compare perception of low-frame-rate pulsed fluoroscopy \textit{(pulsed)} with that of conventional 30 acq/sec. continuous fluoroscopy (hereafter called \textit{continuous}) [9, 19, 53]. We compare the ability to see stationary, low-contrast objects and determine the dose value of \textit{pulsed} that gives perception equal to \textit{continuous}. We obtain surprisingly similar results using three experimental protocols: paired-comparison, min-contrast, and forced-choice. With equivalent perception, we measure average dose savings of 22\%, 38\%, and 49\% at 15, 10, and 7.5 acq/sec. respectively. We find a dependence upon object size with larger disks resulting in more dose.
savings. We have also calculated a savings from an other detectability study in noisy images sequences [39].

Presently, we extend these measurements and compare conventional, continuous fluoroscopy, acquired and displayed at 30 frames/sec, with a single-frame presentation that mimics an LIH frame. We use computer-generated, noisy image frames containing either low-contrast disks, or projected cylinders that simulate guide wires, catheters and arteries in interventional angiography procedures. We determine the single-frame dose that gives perception equal to a continuous reference display. We also suggest some appropriate noise reduction techniques for an LIH frame. Further, we compare the perception of disks and projected cylinders.

3.2 Material and Methods

3.2.1 Equipment and Subjects

Computer generated, $512 \times 512 \times 8$ bit pixel images are displayed on a video tachistoscope. On the video screen, a pixel is approximately 0.5 mm. and the largest object is 1.6 cm in diameter. The viewing distance is 150 cm. approximately the minimal viewing distance in a cath lab. Experiments are conducted in a darkened room. We measure and linearize the luminance
response of the display. Following linearization, 0 and 255 gray-scale units correspond to 0 and 18\textit{ft}\cdot\textit{L}, respectively. The \textit{continuous} presentation is a 30 frames/sec, non-interlaced display. Since we use stationary objects, \textit{continuous} mimics both conventional fluoroscopy acquisition, where x-rays are continuously produced, and pulsed fluoroscopy acquisitions at 30 acq/sec. A more detailed description of the experimental setup, the display equipment, and the method for generating noisy image sequences is described in a previous paper [9].

Three subjects (age 29 ± 4 years) with normal visual acuity are used. Two are naive to the hypotheses, and all subjects are familiarized with the image displays and experimental procedures.

### 3.2.2 Paired-Comparison and Min-Contrast Experiments

In paired-comparison experiments, the task is to compare the visibility of low-contrast objects in a contrast-detail phantom displayed in \textit{continuous} and \textit{single-frame} modes on separate parts of a single monitor. Objects are either disks (Fig. 3.1) or projected cylinders (Fig. 3.2). Using the same phantoms, subjects also determine the objects which are just detectable:
Figure 3.1: The paired-comparison disk phantom. A *continuous* reference at dose Q/acq (left) is compared to a *single-frame* display at various dose levels. The contrast of the disks varies approximately linearly along the y-axis, and exact contrasts can be calculated using gray-scale values in Eq. 3.1. In 8 bit gray-scale, mean values of the disks are 104, 107, 110, 113, 116, 119, 122, and 125 on a mean background of 128. The noise level of *continuous* is $\sigma = 29.2$, corresponding to a dose Q/acq, and we vary the dose for *single-frame* from 1.0 to 5.0 Q/acq. For clarity, noise is not added in this figure.

we call this the min-contrast measurement.
Figure 3.2: The cylinder phantom for paired-comparisons of continuous and single-frame. Sizes of the projected cylinders mimic guide wires, catheters, and arteries in interventional angiography applications. Four cylinder diameters (21, 11, 5 and 1 pixels) are shown in the 4 quadrants of the image. The cylinder projections are calculated using Eq. 3.2, and a 5 element convolution kernel, \( h = [.006, .210, .569, .210, .006] \), simulates x-ray system blur. The peak contrast (specified at the cylinder center-axis) varies linearly between the 6 cylinders, and is given in the figure. The mean background of the phantom is 128 gray-levels. The noise level of continuous is \( \sigma = 29.2 \), corresponding to a dose Q/acq, and the dose for single-frame varies from 1.0 to 5.0 Q/acq. For clarity, noise is not added in this figure.

Design of Phantom

The contrast-detail phantoms contain either disks or projected cylinders on a flat noisy background. The Poisson noise approximates fluoroscopy [9]
and is obtained from a random number generator [40]. In the reference presentation, continuous, the noise corresponds to a dose $Q$ per acquisition. while in the single-frame test presentation, dose varies from 1.0 to 5.0 $Q$ in steps of 0.8 $Q$. For a dose $bQ/acq$, the mean background gray level remains fixed and the standard deviation in gray levels is $\sigma/\sqrt{b}$ ([9], and Figs. 3.1 and 3.2), where $\sigma$ is the standard deviation corresponding to a dose $Q/acq$.

The object contrast, $C$, is

$$C = \frac{|m_{bg} - m_{obj}|}{|m_{bg} + m_{obj}|}$$  \hspace{1cm} (3.1)

where, $m_{bg}$ is the mean gray-scale value of the background, and $m_{obj}$ is either the mean gray-scale value of a disk or the mean gray-scale value along the center axis of a cylinder.

Assuming logarithmic conversion\(^1\) to remove exponential attenuation, the x-ray path-length, $\mu l$, for parallel rays through a cylinder is

$$\mu l(x) = 2\mu \sqrt{\frac{d^2}{4} - x^2} \quad \text{for} \quad -\frac{d}{2} \leq x \leq \frac{d}{2}.$$  \hspace{1cm} (3.2)

where $x$ is the distance from the center axis, $d$ is the cylinder diameter, $l$ is the geometric distance, and $\mu$ is an effective linear attenuation coefficient for

\(^1\)Although logarithmic, linear, or gamma conversions are used in fluoroscopy, we use the simplifying assumption of a logarithmic conversion.
a cylinder containing contrast material. For cylinder phantoms, we introduce x-ray system blur by convolving the projections with an approximate line spread function, a 5 element kernel. \( h = [.006, .210, .569, .210, .006] \), which is calculated from Fujita et al. [54]. Other details of the phantoms are described in the legends of Figs. 3.1 and 3.2.

**Experimental Procedure**

The reference display, *continuous*, consists of 20 different noise frames displayed in a cine-loop. The test display, *single-frame*, consists of a single noise frame which is continuously repeated. Test displays are constructed at 6 different dose levels. In order to ensure randomization, we use 8 realizations at each dose. Dose levels are presented counter-balanced in ascending followed by descending order [55]. A total of 144 trials (8 repetitions at 6 dose values for each of 3 subjects) is given for each cylinder and disk.

In the case of disks, subjects are asked the following three questions in the order listed.

**Q1:** How many disks in each column can you detect?

**Q2:** For each disk size, which display (right or left) results in better visibility?
Q3: Considering all disk sizes, which display (right or left) results in better visibility?

For Q2 and Q3, the subjects are instructed to focus on disks having the lowest visibility. The paired-comparison experiment uses Q2 and Q3, and the min-contrast experiment uses Q1. Similar questions are used for projected cylinders. The subjects have no time limit for their response.

Data Analysis

Two analyses are applied to the paired-comparison data (responses to Q2 and Q3) as described elsewhere [9]. Briefly, we first use hypothesis testing of a binomial distribution. We expect single-frame at high dose levels to be preferred over continuous. Using a 5% significance level, we can reject the null hypothesis that the probability of selecting single frame is less than 1/2. when single-frame is preferred in \( \geq 17 \) of 24 trials [42]. In order to obtain a non-discrete equivalent-perception dose, we fit a PROBIT curve to the fraction selecting single-frame [9, 43]. The equivalent-perception dose is obtained when the PROBIT curve gives a 50% probability for preferring single-frame. Fits are evaluated using a Pearson Goodness of Fit test.

Paired-comparison results are used to determine the relationship be-
between object size and equivalent-perception dose. For both cylinders and disks, we fit second-order polynomials to the equivalent-perception dose plotted as a function of object diameter. We further use the two polynomials to determine the relationship between disk and cylinder size.

In the case of min-contrast, we record the number of disks or cylinders seen (response to Q1) and convert this number to contrast. Min-contrasts are averaged over all trials and plotted. To determine an equivalent-perception dose for single-frame, we fit a line to the average min-contrasts plotted as a function of dose. The equivalent-perception dose is where the curve equals the min-contrast from continuous.

3.3 Results

The bar graph in Fig. 3.3A gives the comparison of continuous and single-frame when all disk sizes are considered (Q3). For the disk phantom, continuous is preferred to single-frame in 24 and 20 cases at 1.0 and 1.8 Q/acq., respectively. At 4.2 and 5.0 Q, single-frame is preferred in 17 and 22 trials, respectively. These results are significant at the 5% level and preference changes from continuous to single-frame somewhere between 2.6 and 3.4 Q/acq. In the case of projected cylinders, similar data are obtained.
Figure 3.3: In a paired-comparison, subjects select either *single-frame* or *continuous*. The bar plots show the number of *single-frame* (above the horizontal axis) and *continuous* (below the horizontal axis) selections. Phantoms are either disks (A) or projected cylinders (B), and all phantom sizes are compared (question Q3). All bars in this plot have a total height of 24 corresponding to the 24 trials. The dose/acq of *continuous* is 1.0 Q/acq, while *single-frame* varies between 1.0 and 5.0 Q/acq as indicated in the legend. Stars indicate significance at a 5% level (N=24).

and again the equivalent-perception dose lies somewhere between 2.6 and 3.4 Q/acq (Fig. 3.3B).

The PROBIT analysis allows us to predict a non-discrete, equivalent-perception dose using all of the data in a curve fit. In Fig. 3.4, we show
the PROBIT analysis for each of the four projected cylinders. The solid curve is the PROBIT curve, and the dashed lines are the 95% confidence intervals. The equivalent-perception dose is obtained when the fraction selecting single-frame is 0.5. As the diameter of the cylinder increases, the curves and the equivalent-perception dose shift to lower dose values. The shift is significant, as it is larger than the confidence intervals. Similar results are obtained for disks.

The fraction of subjects selecting single-frame is plotted as a function of the single-frame acquisition dose. Projected cylinders are used and diameter varies from 1 to 21 pixels in the four panels, as indicated. The stars are measured data points, and the solid curves are the estimated PROBIT fits. Dashed lines are the 95% confidence limits along the x-axis. Note that the data shift to smaller dose values as the cylinder diameter increases. The equivalent-perception dose for single-frame occurs when the fraction is 0.5; e.g., it is 2.6 ± 0.5 Q/acq, for the 11 pixel diameter cylinders. All data sets are adequately described by the PROBIT model, as determined by a Pearson Goodness of Fit test.

Table 3.1 lists the equivalent-perception dose and confidence intervals from the PROBIT analysis. Included are comparisons of all object sizes
Figure 3.4: The fraction of subjects selecting *single-frame* is plotted as a function of the *single-frame* acquisition dose. Projected cylinders are used and diameter varies from 1 to 21 pixels in the four panels, as indicated. The stars are measured data points, and the solid curves are the estimated PROBIT fits. Dashed lines are the 95% confidence limits along the x-axis. Note that the data shift to smaller dose values as the cylinder diameter increases. The equivalent-perception dose for *single-frame* occurs when the fraction is 0.5: e.g., it is 2.6 ± 0.5 Q/acq, for the 11 pixel diameter cylinders. All data sets are adequately described by the PROBIT model, as determined by a Pearson Goodness of Fit test.

...
We explore the dependence of equivalent-perception dose on object type and size. In Fig. 3.5, equivalent-perception dose values are plotted as a function of object size for both disks (Fig. 3.5A) and projected cylinders (Fig. 3.5B). The data are well fit to second-order polynomials which are plotted as solid curves. Note that at any given equivalent-perception dose, disk data are shifted to larger diameters than cylinder data.

We wish to determine equivalent sizes for disks and projected cylinders. Using equations from Fig. 3.5, we plot cylinder diameter as a function of disk diameter in Fig. 3.6. For equivalent-perception dose values, disk diameter is considerably larger than cylinder diameter. The curve is well-approximated by a straight line going through the origin with a slope of 0.32. Hence, a disk diameter of 16 pixels gives an equivalent-perception dose equal to that of a cylinder 5.1 pixels in diameter.

We next analyze the min-contrast responses (question Q1). In Fig. 3.7, min-contrasts from the 24 trials are averaged and plotted as a function of disk (Fig. 3.7A) and cylinder (Fig. 3.7B) diameter and x-ray dose. In general, the objects are easier to see as diameter and/or dose increase. To demonstrate the effect of object diameter upon equivalent-perception dose, we plot continuous min-contrasts between bracketing single-frame
Figure 3.5: Equivalent-perception dose values are plotted as a function of diameter for the case of disks (A) and cylinders (B). The solid curves are second-order polynomial fits to the data, and parameters are shown in the figure. Note that disks are shifted to larger diameters than cylinders.

min-contrasts. For example, in Fig. 3.7A, at a disk diameter of 4 pixels, the
Figure 3.6: At equivalent-perception dose values, cylinder diameter is plotted as a function of disk diameter. Values are obtained from the curve fits in Fig. 3.5. The relationship is well described by a line going through the origin with a slope of 0.32 (regression coefficient, $r = 0.999$).

Continuous min-contrast is plotted between single-frame min-contrasts at 4.2 and 5.0 Q/acq, which indicates that equivalent perception is obtained somewhere between these values. The min-contrast data predict that the
dose for equivalent perception depends upon object size with larger objects having a smaller equivalent-perception dose.

We can estimate non-discrete, equivalent-perception dose values from min-contrast experiments. An example is shown in Fig. 3.8, where average single-frame min-contrasts for a cylinder are plotted as a function of acquisition dose. The horizontal line marks the min-contrast for continuous. At the intersection, min-contrasts are equal, and the equivalent-perception dose is 4.0 Q/acq. When this analysis is applied to cylinder data, we obtain 4.3, 4.0, 3.1, and 2.9 Q/acq for cylinders 1, 5, 11, and 21 pixels in diameter, respectively. Similarly, we obtain 4.1, 3.8, 3.8, and 3.6 Q/acq for disks 2, 4, 8, and 16 pixels in diameter, respectively.

3.4 Discussion

The equivalent-perception dose measured with paired-comparison and min-contrast experiments depends both upon the type of object (disk or projected cylinder) and object size. In general, smaller objects require a higher equivalent-perception dose (Figs. 3.4, 3.5, and 3.7). This is consistent with earlier experiments comparing conventional, continuous fluoroscopy at 30 acq/sec with low-acquisition-rate pulsed fluoroscopy. Here too, the
equivalent-perception dose increased as disk size decreased [9].

As in a previous study [9], we find paired-comparison more reliable than min-contrast. Subjects often indicate that there is virtually no difference in the number of disks seen, but yet they have no trouble identifying which disks are seen better. Nevertheless, paired-comparison and min-contrast equivalent-perception dose values compare favorably. The trend of decreasing equivalent-perception dose with increasing size is preserved. On average, the absolute difference between the two measurements is 16% and 12% for cylinders and disks, respectively. The biggest discrepancy occurs for the largest diameters which are relatively easy to visualize. When the largest diameters are excluded, the mean absolute difference is reduced to less than 10% for both objects. Previously, we found that forced-choice experiments give results identical to those from paired-comparison and min-contrast measurements [9]. Also, forced-choice takes a long time to perform in experiments requiring image sequences. To economize experiment time and allow investigation of object size, we did not do forced-choice studies here.

It is interesting to compare results with disks and projected cylinders. The plot in Fig. 3.6 is very linear. With regard to equivalent-perception dose, a cylinder of diameter 10 pixels is equivalent to a disk of diameter
30 pixels. This is reasonable since the cylinder has virtually infinite extent along the long axis and, hence, covers a large area in that direction.

For the case of projected cylinders, the equivalent-perception dose ranges from \( \approx 2 \) to \( 4 \) Q/acq (Table 3.1 and Fig. 3.5B), and when all cylinders are considered, the equivalent-perception dose is \( \approx 2.9 \) Q/acq. Since many catheters and guide wires fall in the range from 1 to 10 pixels in diameter, we use a value of 3.0 Q/acq in subsequent analyses.

We infer that viewing a single frame at 3.0 Q/acq gives perception equivalent to continuous at 1.0 Q/acq. Thus, the perception of objects in a noisy sequence is similar to that of an LIH display averaged over 3 frames. To put it another way, the HVS effectively averages over approximately 3 frames, or 100 ms. We recognize that this analysis is a simplification since the temporal characteristic of the HVS is not simply an averaging filter [56, 57, 51]. In general, temporal visual integration depends on several factors, including contrast, stimulus duration, and luminance [58]. Remarkably, our estimated duration compares well with other numbers for persistence in the visual system. These are obtained from much different human perception experiments, and reported in the range of 50 to 250 msec [58, 47]. Also, Schade in his photoelectric analog of the visual system [59] estimates an
<table>
<thead>
<tr>
<th>Diameter (pixels)</th>
<th>Disks</th>
<th>95% confidence interval (Q/acq)</th>
<th>Cylinders</th>
<th>95% confidence interval (Q/acq)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Equiv convenience dose (Q/acq)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>(3.86, 4.83)</td>
<td>1</td>
<td>4.11</td>
</tr>
<tr>
<td>8</td>
<td>3.99</td>
<td>(3.17, 5.72)</td>
<td>5</td>
<td>3.63</td>
</tr>
<tr>
<td>16</td>
<td>3.26</td>
<td>(2.92, 3.62)</td>
<td>11</td>
<td>2.61</td>
</tr>
<tr>
<td>32</td>
<td>2.78</td>
<td>(2.41, 3.11)</td>
<td>21</td>
<td>2.02</td>
</tr>
<tr>
<td>Overall</td>
<td>3.29</td>
<td>(2.96, 3.64)</td>
<td>Overall</td>
<td>2.88</td>
</tr>
</tbody>
</table>

Table 3.1: For the paired-comparison experiments, we list equivalent-perception dose values and 95% confidence intervals (in parentheses). Results are presented for both disks and projected cylinders. Values are estimated by the PROBIT analysis, and data include responses to the comparison of all phantom sizes (Q3), as indicated by “overall”, as well as each phantom size (Q2). In all cases, the PROBIT model adequately describes all data, as determined by the Pearson Goodness of Fit test.

An effective visual storage time of 100 msec, which he relates to a 3 frame integration during television viewing [60].

In the event of a lower acquisition rate such as 15 acq/sec (pulsed-15), we can predict a new equivalent-perception dose for the single-frame display. The equivalent-perception dose can be interpreted as the ratio

\[
\frac{single\_dose}{continuous\_dose} = 3.0.
\]

A previous report finds equivalent
perception at 1.5 Q when \textit{pulsed-15} is compared to \textit{continuous} [9]; that is, \textit{pulsed-15.dose/continuous.dose} = 1.5. Because the reference display is \textit{continuous} in both cases we can use a transitive relationship to obtain \textit{single.dose/pulsed-15.dose} = 3.0/1.5 = 2.0.

Because of the relationship between dose and noise variance, our experiments predict that the variance of the LIH frame should be reduced by a factor of approximately 3 to achieve equivalent perception with the \textit{continuous} acquisition. There are several possibilities for reducing variance. First, one might simply increase the dose of the last frame by a factor of 3. This necessitates lengthening the x-ray pulse because changing the tube current requires time to change the cathode temperature and changing kVp affects image quality. Second, since the Poisson noise level in fluoroscopy is well approximated with a Gaussian amplitude distribution [61], linear spatial or temporal filtering results in a Gaussian with reduced variance [62]. As an example, one can temporally average the last 3 frames. Spatially, one can use a center-weighted 3 \times 3 kernel, like [1 1 1: 1 10 1: 1 1 1], which also reduces the variance by three, while minimizing spatial blur. Third, one can apply a temporal recursive filter [41], or temporal recursive filters modified to reduce motion blur [63, 64]. Finally, non-linear filters such as median [65],
or morphological filters [66, 67] can be applied. Although a Gaussian noise distribution is no longer assured in the case of non-linear filtering, as a first approximation, one can simply analyze the variance reduction in order to predict effects on image perception.
Figure 3.7: Average min-contrasts are plotted as a function of disk (A) and cylinder (B) diameter and dose per acquisition. Data from both continuous and single-frame displays are plotted. In order to demonstrate the effect of object size on equivalent-perception dose, we plot continuous min-contrasts between bracketing single-frame min-contrasts. The dose for continuous is 1.0 Q/acq. and dose values for single-frame are shown. We use question Q1. and 24 trials for each data point.
Figure 3.8: This demonstrates a method for obtaining a non-discrete equivalent-perception dose from min-contrast data for the case of a cylinder of diameter 5 pixels. Average min-contrasts are plotted as a function of acquisition dose for single-frame (circles). The solid curve is a line fit ($y = -0.0022x + 0.0322$). The min-contrast for continuous is marked on the y-axis (star) and projected horizontally to the curve. The dose at the intersection ($\approx 4 \, Q/\text{acq}$) is the equivalent-perception dose.
Chapter 4

A Model for Perception of Pulsed Fluoroscopy Image Sequences

4.1 Introduction

In this chapter, we extend the ideal observer model to assess the performance of continuous and pulsed fluoroscopy. The ideal observer has its roots in statistical decision theory and has been widely applied in radiology to evaluate imaging modalities [22, 23, 24, 25], image processing techniques [26, 27, 28, 29, 30], and image reconstruction methods [31, 32]. In most cases, single images are considered, and much less detection modeling has been done with regard to noisy image sequences [33, 39, 68].

In our model, we incorporate the spatio-temporal filter response of the
human visual system (HVS), as measured in human, image perception flicker experiments [69, 56, 57, 70, 71, 72, 73]. As in previous models, the final output is a signal-to-noise ratio ($SNR$). Equating the $SNR$ of pulsed with the $SNR$ of continuous, allows one to compute an equivalent-perception dose. Within this framework, we propose three models that include a separable spatio-temporal response of the HVS (Model 1), a non-separable, spatio-temporal interaction in the HVS (Model 2), and a non-separable, spatio-temporal interaction as well as an internal noise source in the HVS (Model 3). We compare all three models to experimentally determined equivalent-perception dose data.

4.2 Theory

The task presented to our human observers is to detect the presence of a disk of known size and location in a noisy background. This is called the signal-known-exactly problem, and it is well described by the “ideal observer” model which has roots in statistical decision theory [22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 39, 68, 35, 49]. The ideal observer consists of a matched filter, or a modification thereof, which creates an optimized signal-to-noise ratio ($SNR$). Object detection occurs when the $SNR$ exceeds a
threshold.

For a single image frame containing an object within a uniform noise background, the $SNR$ at the filter output is

$$SNR = \frac{S}{N} = \frac{\iint |\Delta S(u, v)|^2 \, du \, dv}{\sqrt{\iint |\Delta S(u, v)|^2 \, P_n(u, v) \, du \, dv}} \quad (4.1)$$

where $\Delta S(u, v)$ is the Fourier spectrum of the object, $P_n(u, v)$ is the noise power spectrum, and $(u, v)$ are the spatial frequency coordinates [24, 25, 26, 27, 28]. Referring to matched filter theory, $\Delta S(u, v)$ is the object template in the frequency domain, and the numerator $(S)$ is the square root of the peak power resulting from a correct match. The denominator $(N)$ is the amplitude of the filter output due to noise only.

Equation 4.1 is sometimes modified to include a filter, $V_s(u, v)$, representing the spatial frequency response of the $HVS$ [25, 28, 74].

$$SNR_s = \frac{S_s}{N_s} = \frac{\iint |\Delta S(u, v)V_s(u, v)|^2 \, du \, dv}{\sqrt{\iint |\Delta S(u, v)V_s(u, v)|^2 \, |V_s^2(u, v)| \, P_n(u, v) \, du \, dv}} \quad (4.2)$$

where the subscript, $s$, indicates the additional spatial filtering. Note that the new object template is $\Delta S(u, v)V_s(u, v)$.

Another modification includes the internal noise of the human observer [25, 26, 28, 49, 33]. Below, $N_s$ and $N_i$ are the input and internal noise sources following processing. Terms are expanded on the second line to
show the filtering processes operating on the input noise power, $P_n(u, v)$, and the internal noise power, $P_i$.

$$SNR_s = \frac{S_s}{\sqrt{N^2_s + N^2_i}}$$

$$= \frac{\iint |\Delta S(u, v)V_s(u, v)|^2 \, du \, dv}{\sqrt{\iint \left( |\Delta S(u, v)V_s(u, v)|^2 |V^2_s(u, v)| P_n(u, v) + |\Delta S(u, v)V_s(u, v)|^2 P_i \right) \, du \, dv}}.$$

In this formulation, $P_i$ is a constant which depends upon the observer and is independent of the input noise. $P_n(u, v)$ [25, 26, 28, 49, 33], although other possibilities exist [75, 76].

For the case of an x-ray fluoroscopy image sequence, we derive an extension to the above model that includes the temporal response of the HVS. We also develop techniques to quantify the temporal frequency characteristics of the input noise at various acquisition rates.

A block diagram of the model is given in Fig. 4.1. We assume that the object does not move. Hence, the object template is constant over time, and it is modified by the 3-dimensional spatio-temporal filter, $V_{st}(u, v, f)$, evaluated at the temporal frequency $f = 0$. The signal-to-noise ratio for the image sequence following spatio-temporal processing, $SNR_{st}$, includes the new object template $\Delta S(u, v)V_{st}(u, v, 0)$ and $V_{st}(u, v, f)$ at appropriate
Figure 4.1: This diagram outlines the ideal observer model modified to characterize human detectability in image sequences. The input image sequence is initially filtered by the spatio-temporal response, \( V_{st}(u, v, f) \), of the HVS. Next, the ideal observer detector which consists of a spatial matched filter, \( \Delta S(u, v) \), modified by the spatio-temporal response of the HVS at \( f = 0 \), \( V_{st}(u, v, 0) \), maximizes the signal-to-noise ratio of the test object. We include internal observer noise, \( P_i \), which is independent of the input sequence. Human detection occurs when the output signal-to-noise ratio exceeds a threshold.

\[
SNR_{st} = \frac{S_{st}}{\sqrt{N^2_{st} + N^2_i}} \quad (4.4)
\]

\[
= \frac{\iint |\Delta S(u, v)V_{st}(u, v, 0)|^2 \, du \, dv}{\sqrt{\iint |\Delta S(u, v)V_{st}(u, v, 0)|^2 |V^2_{st}(u, v, f)| \, P_n(u, v, f) \, du \, dv \, df + \iint |\Delta S(u, v)V_{st}(u, v, 0)|^2 \, P_i \, du \, dv}}.
\]

We next compute the frequency components of the noisy continuous and pulsed input image sequences. We assume that each image frame has the same spatially-uniform noise distribution, \( P_n(u, v) = \sigma^2 \), where \( \sigma^2 \) is the
noise variance of the frame. For a non-interlaced sequence, we consider the signal at each pixel as a temporal signal, $q_k(n)$, with a sampling interval $\Delta t$.

$$q_k(n) = \{\underbrace{a, a, \ldots, a}_{k}, \underbrace{b, b, \ldots, b}_{k}, c, \ldots\} \text{ for } k = 1, 2, \ldots \text{ and } n = 1, 2, \ldots, N.$$  

(4.5)

In the case of continuous, there are no repeated samples ($k = 1$), and we have the signal $q_1(n) = \{a, b, \ldots\}$. The time sequence $\{a, b, \ldots\}$ is white noise with variance $\sigma^2$. For $k > 1$, we have pulsed with $k - 1$ repetitions of each noise sample; e.g., when $k = 2$, $q_2(n) = \{a, a, b, \ldots\}$. The signal $q_k(n)$ can be expressed as

$$q_k(n) = h_k(n) \ast z_k(n)$$  

(4.6)

where $\ast$ denotes convolution. and the two sequences, sampled at an interval $\Delta t$, are given below.

$$h_k(n) = \{\underbrace{1, 1, \ldots, 1}_{k}\}$$  

(4.7)

$$z_k(n) = \begin{cases} 
q_1(m) & \text{for } m = 1, 2, \ldots, N/k \text{ and } n = k(m - 1) + 1 \\
0 & \text{otherwise (} n \leq N \text{)}
\end{cases}$$  

(4.8)

For the case $k = 1$, $z_1(n) = q_1(n)$, and for $k = 2$, $z_2(n) = \{a, 0, b, 0, \ldots\}$.

Both $z_k(n)$ and $q_k(n)$ have a time duration of $N \Delta t$, but for $k > 1$, $z_k(n)$ con-
tains zeros which reduce its variance. The variance and power spectrum \([77]\) of \(z_k(n)\) are respectively \(\sigma^2/k\) and

\[
P_{z_k}(f) = \frac{\sigma^2}{k} \Delta t. \tag{4.9}
\]

The power spectrum of \(q_k(n)\) is obtained from \(H_k(f)\), the Fourier transform of \(h_k(n)\), and \(P_{z_k}(f)\) as follows.

\[
P_{q_k}(f) = |H_k(f)|^2 P_{z_k}(f) = \left| \frac{\sin(k\pi f \Delta t)}{\sin(\pi f \Delta t)} \right|^2 \frac{\sigma^2}{k} \Delta t. \tag{4.10}
\]

Note that we include \(\Delta t\) so that the integral of \(P_{q_k}(f)\) over \(-1/(2\Delta t) \leq f \leq 1/(2\Delta t)\), yields the true power of an analog signal \([78]\).

For continuous at 30 acq/sec \((k = 1\) and \(\Delta t = 1/30\) sec), the power spectrum of the noisy input becomes

\[
P_{q_1}(f) = \frac{\sigma^2}{30} \quad \text{for} \quad -15 \leq f \leq 15. \tag{4.11}
\]

Similarly, for pulsed-15 \((k = 2\) and \(\Delta t = 1/30\) sec), the spectrum becomes

\[
P_{q_2}(f) = \left| \frac{\sin(2\pi f/30)}{\sin(\pi f/30)} \right|^2 \frac{\sigma^2}{2 \cdot 30} \quad \text{for} \quad -15 \leq f \leq 15. \tag{4.12}
\]

The 3-dimensional input noise power spectrum required in Eq. 4.4 becomes

\[
P_n(u, v, f) = P_{q_k}(f) = \left| \frac{\sin(k\pi f \Delta t)}{\sin(\pi f \Delta t)} \right|^2 \frac{\sigma^2}{k} \Delta t. \tag{4.13}
\]
If we display a single frame, then $\Delta t \to \infty$, and we find that $\int_{-1/2\Delta t}^{1/2\Delta t} P_{V_1}(f) V_{st}^2(u, v, f) df \to V_{st}^2(u, v, 0) \sigma^2$ and Eq. 4.4 reduces to the previous result for a single image frame (Eq. 4.3).

$$SNR_s = \frac{\iint |\Delta S(u, v)V_{st}(u, v, 0)|^2 \, du \, dv}{\sqrt{\iint \left( |\Delta S(u, v)V_{st}(u, v, 0)|^2 |V_{st}^2(u, v, 0)| \sigma^2 + |\Delta S(u, v)V_{st}(u, v, 0)|^2 P_i \right) \, du \, dv}} \quad (4.14)$$

We apply the modified ideal observer model to fluoroscopic imaging of low contrast objects. We assume that the x-ray counts are Poisson distributed about a mean of $N_0$ and that the object has a small number of counts, $\Delta N$, above the background [24], thus approximating additive noise [68, 79]. Substituting $N_0$ for $\sigma^2$ and $\Delta N S(u, v)$ for $\Delta S(u, v)$ in Eq. 4.4, we get

$$SNR_{st, k} = \frac{\iint |\Delta N S(u, v)V_{st}(u, v, 0)|^2 \, du \, dv}{\sqrt{\iint \left( |\Delta N S(u, v)V_{st}(u, v, 0)|^2 |V_{st}^2(u, v, f)| \right) \, du \, dv \, df}}$$

$$= \frac{\sqrt{\frac{\sin(k\pi/\Delta t)}{\sin(\pi/\Delta t)}}^2 \frac{N_0}{k} \Delta t \, du \, df \, + \iint |\Delta N S(u, v)V_{st}(u, v, 0)|^2 P_i \, du \, dv}{\sqrt{\frac{\sqrt{N_0^2}}{\Gamma_k \cdot N_0 + \gamma}}} \quad (4.15)$$
\[
\Gamma_k = \int \int |S(u, v)V_{st}(u, v, 0)|^2 \left| V_{st}^2(u, v, f) \right| \left| \frac{\sin(k \pi f \Delta t)}{\sin(\pi f \Delta t)} \right|^2 \Delta t \frac{du \, dv \, df}{k}
\]

(4.16)

and

\[
\gamma = P_i \int \int |S(u, v)V_{st}(u, v, 0)|^2 \, du \, dv.
\]

(4.17)

In order to compare with previously published results, we now evaluate the equations for the special case of detecting a disk in a single noisy frame.

The signal, \( s \), is a disk of radius \( r_d \), given by the circle function.

\[
s(r) = \text{circle} \left( \frac{r}{r_d} \right) = \begin{cases} 
1 & \text{for } r \leq r_d \\
0 & \text{otherwise}
\end{cases}
\]

(4.18)

With no internal noise \((P_i = 0)\), \( \gamma \) becomes zero. For a single frame display, \( \Delta t \rightarrow \infty \) and \( \Gamma_k \rightarrow \int \int |S(u, v)V_{st}^2(u, v, 0)|^2 \, du \, dv \). If we assume no spatial filtering by the HVS, \( V_{st}(u, v, f) \overset{\text{def}}{=} V_i(f) \). Converting to polar coordinates we obtain

\[
\int \int |S(u, v)V_{st}(u, v, 0)|^2 \, du \, dv = \int \int |S(u, v)V_i(0)|^2 \, du \, dv \\
= 2\pi V_i^2(0) \int_0^\infty |S(q)|^2 \, dq \\
= 2\pi r_d^2 V_i^2(0) \int_0^\infty [J_1(2\pi r_d q)]^2 \frac{dq}{q} \\
= \pi r_d^2 V_i^2(0) = a_d V_i^2(0)
\]

(4.19)
where \( J_1 \) is the Bessel function and \( a_d = \pi r_d^2 \) the disk area. Substituting into Eq. 4.15, the \( SNR \) reduces to

\[
SNR = \sqrt{a_d \frac{\Delta N}{N_0}} \sqrt{N_0},
\]  

which is the Rose model giving the \( SNR \) for a low-contrast disk in a single image frame [24, 35, 80].

In our experiments, we compare the detectability of \textit{continuous} and \textit{pulsed} with objects of equal shape, \( S(u, v) \), and contrast, \( \Delta N/N_0 \). We are interested in the x-ray dose where \textit{continuous} and \textit{pulsed} have equal detectability; i.e. \( SNR_{st,1} \overset{\text{def}}{=} SNR_{st,k} \) for \( k > 1 \). We equate \( SNR \)'s obtained from Eq. 4.15. substitute \( N_{cont} \) for the mean number of counts in the \textit{continuous} case \( (k = 1) \) and \( N_{pulsed,k} \) for the case of \textit{pulsed} with \( k - 1 \) repetitions, and rearrange to obtain

\[
(\Gamma_1 N_{cont} + \gamma) N_{pulsed,k}^2 - \left( \Gamma_k N_{cont}^2 \right) N_{pulsed,k} - \gamma N_{cont}^2 = 0.
\]  

Solving for \( N_{pulsed,k} \) we get

\[
N_{pulsed,k} = \frac{\Gamma_k N_{cont}^2 (-)}{2 (\Gamma_1 N_{cont} + \gamma)} \sqrt{\frac{\Gamma_k^2 N_{cont}^4 + 4 \gamma N_{cont}^2 (\Gamma_1 N_{cont} + \gamma)}{\Gamma_1 N_{cont} + \gamma}}.
\]  

Assuming proportionality between the acquisition dose, \( Q \), and the number of counts, \( N \) \( (Q = \eta N) \), where \( \eta \) is the counts-to-dose proportionality
constant) and setting $\gamma^\prime = \eta \gamma$ we get

$$Q_{\text{pulsed}, k} = \frac{\Gamma_k Q_{\text{cont}} + \sqrt{\Gamma_k^2 Q_{\text{cont}}^2 + 4 \Gamma_k^2 Q_{\text{cont}} + \gamma^\prime (\Gamma_1 Q_{\text{cont}} + \gamma^\prime)}}{2 (\Gamma_1 Q_{\text{cont}} + \gamma^\prime)} Q_{\text{cont}}. \quad (4.23)$$

With no internal noise, $\gamma^\prime = 0$. Eq. 4.23 reduces to

$$Q_{\text{pulsed}, k} = \frac{\Gamma_k}{\Gamma_1} Q_{\text{cont}}. \quad (4.24)$$

An interesting scenario exists when we further require the spatio-temporal response to be separable, i.e. $V_{st}(u, v, f) \overset{\text{def}}{=} V_s(u, v)V_t(f)$. In this case, $\Gamma_k$ becomes

$$\Gamma_k^{\text{sep}} = \int \left| S(u, v) V_s^2(u, v) \right|^2 du dv \int |V_t(f)|^2 \left| \frac{\sin(k\pi f \Delta t)}{\sin(\pi f \Delta t)} \right|^2 \frac{\Delta t}{k} df. \quad (4.25)$$

When identical objects are compared, the ratio $\Gamma_k^{\text{sep}}/\Gamma_1^{\text{sep}}$ in Eq. 4.24 is independent of both $S(u, v)$ and $V_s(u, v)$. and we get

$$Q_{\text{pulsed}, k} = \frac{\int |V_t(f)|^2 \left| \frac{\sin(k\pi f \Delta t)}{\sin(\pi f \Delta t)} \right|^2 \frac{\Delta t}{k} df}{\int |V_t(f)|^2 \Delta t df} Q_{\text{cont}}. \quad (4.26)$$

### 4.3 Simulation Models

We calculate the equivalent-perception dose using 3 models.
4.3.1 Model 1

This model assumes that the observer has no internal noise \( P_i = 0 \) and that the spatio-temporal response of the HVS is separable. In this case the equivalent-perception dose is calculated using Eq. 4.26.

We model the temporal response, \( V_t(f) \) with the second order low-pass filter

\[
V_t(f) = \frac{f_c^2}{\sqrt{f^4 + 2f_c^2 (2\beta^2 - 1) f^2 + f_c^4}} \tag{4.27}
\]

where \( f_c \) is the cut-off frequency and \( \beta \) is a damping factor [81]. This model gives a good fit to curves from human image perception flicker experiments performed by Kelly [56]. A least-squares fit of flicker data to Eq. 4.27 gives parameter estimates \( f_c = 10.3 \) and \( \beta = 1.44 \) (see Fig. 4.2).

Filtering the noisy input signal (Eq. 4.10) with the low-pass filter (Eq. 4.27), the power spectrum of the output signal becomes, \( V_t^2(f) P_{\text{in}}(f) \).

Note that \( Q_{\text{pulsed},k}/Q_{\text{cont}} \) from Eq. 4.26 can be interpreted as the ratio between the integrated output power of \textit{pulsed} and \textit{continuous}. Input and output spectra in Fig. 4.3 show the effects of acquisition rate and temporal filtering.
Figure 4.2: This is the temporal response of the human visual system as obtained from human image-perception flicker experiments. Normalized data (stars) are obtained from Fig. 7 by Kelly [56]. The smooth curve is a low-pass filter model fit to the data (Eq. 4.27). Parameters are $f_c = 10.3$ Hz and $\beta = 1.44$.

4.3.2 Model 2

This model assumes a spatio-temporal interaction in the visual system, but no internal observer noise. Many studies show that there is a dependency between flicker sensitivity and spatial frequency [69, 57, 70, 71, 82, 83].
Figure 4.3: For each of four acquisition types identified in the legend, frequency power spectra are given for the input (A) and output of the HVS temporal filter (B). The input curves are given by Eq. 4.10 with $\sigma^2 = 1$ and $\Delta t = 1/30$ sec. With decreasing acquisition rates, the input power spectra have progressively more energy at lower frequencies. The low-pass filter of the HVS, specified by Eq. 4.27 and shown in Fig. 4.2 attenuates the frequency content above $\approx 2$ Hz.

In Model 2, we use an analytical surface description of the HVS's spatio-temporal response and apply Eq. 4.24 for the calculation of the equivalent-perception dose.

Kelly describes a model of the 2-dimensional contrast sensitivity surface.
\[ V_\varepsilon'(q, f) = 4\pi^2q \left[ 6.1 + 7.3 \log \left( \frac{f}{3q} \right) \right] \exp \left[ -\frac{4\pi q}{45.9} \left( \frac{2\pi f}{q} + 2 \right) \right] \] (4.28)

where the spatial frequency in the radial direction, \( q \), is given in cycles per degree (cpd), and the temporal frequency, \( f \), is in Hz (Eq. 8 by Kelly) [69]. The model has been fit to data measured using flickering sine wave stimuli [69]. The lower limit for a correct model fit is reported to be 0.1 cpd. Similar, we adopt 0.1 Hz as the lowest valid temporal frequency for which the model can be applied. Further, Kelly shows that all higher spatial and temporal frequencies of interest are characterized well by the model. Equation 4.28 specifies the spatial response in one direction. We further assume it to be rotationally symmetric in the spatial frequency domain, and get

\[ V_\varepsilon(u, v, f) = V_\varepsilon' \left( \sqrt{u^2 + v^2} \cdot f \right) . \] (4.29)

Figure 4.4 shows the effect of the spatio-temporal figure on the integrand of Eq. 4.16. The curves are shown as a function of temporal frequency, \( f \), and horizontal spatial frequency, \( u \), with the vertical spatial frequency, \( v = 0 \). Note that as compared to continuous, pulsed-7.5 is attenuated at higher temporal frequencies (compare Figs. 4.4A with 4.4B and 4.4C with 4.4D). In addition with a smaller disk relatively more energy is present at higher
Figure 4.4: The integrand of Eq. 4.16, \(G(u, v, f)\), is evaluated as a function of temporal frequency, \(f\), and horizontal spatial frequency, \(u\), with vertical spatial frequency, \(v = 0\). In each plot, both a perspective view of the surface and a contour map are shown. The signal spectrum, \(\Delta S(u, v)\), is that of a disk. (A) is continuous \((k = 1)\) and disk radius, \(r_d = 16\) pixels; (B) is pulsed-7.5 \((k = 4)\) and \(r_d = 16\) pixels; (C) is \(k = 1\) and \(r_d = 1\) pixels; and (D) is \(k = 4\) and \(r_d = 4\) pixels.

spatial frequencies (note the shift on the contour map when comparing Figs. 4.4A with 4.4C and 4.4B with 4.4D).

The triple integral of Eq. 4.16 is estimated by partitioning the volume into rectangular cells and calculating the sum using trapezoidal approximation. The cells have dimensions \(\Delta f = 0.1\) Hz by \(\Delta u = 0.2\) cpd by
\( \Delta v = 0.2 \) cpd. The temporal frequency is integrated from 0.1 Hz to 15 Hz while the two spatial frequencies limits are 0.1 cpd and 25 cpd. With the specified integration limits, the integrals reach an asymptotic value as the cell size decreases to the dimensions given.

4.3.3 Model 3

Model 3 includes the spatio-temporal dependency of Model 2 (Eqs. 4.28 and 4.29) as well as an internal observer noise source, \( P_i \). The equivalent-perception dose is calculated from Eq. 4.23 with the normalization, \( Q_{cont} = 1 \). Using a least-squares algorithm [84] Eq. 4.23 is fitted to experimental equivalent-perception dose data. The single free parameter in the fit is the product, \( \eta P_i \), the internal noise scaled by the counts-to-dose proportionality constant. We report the influence of internal noise in Model 3 using the ratio of the internal observer noise and the external noise, \( N_i/N_{st} \). It follows from Eqs. 4.15-4.17 that this ratio can be written as \( (\Gamma_k/\gamma)^{1/2} \) where \( k = 1 \) for continuous. Note, that only one single value of \( \eta P_i \) exists for all acquisition rates and disk sizes.
4.4 Review Of Experimental Measurements

We briefly review measurements reported elsewhere [9, 19, 53]. In all cases, our objective is to study the perception of low-contrast, stationary disks displayed in a sequence of noisy images. We use three experimental paradigms:

(1) paired-comparison, (2) min-contrast, and (3) 4-alternative forced-choice.

In all three experiments, it is our goal to find the equivalent-perception dose where the test display (pulsed-1.5, pulsed-1.0, or pulsed-0.5) gives a result equivalent to the control display (continuous).

For the paired-comparison and min-contrast study, we use two computer-generated, contrast-detail disk phantoms displayed side-by-side on the monitor. Poisson distributed noise is added to simulate fluoroscopy. The left side is the control display (continuous) at a fixed dose, Q/acq. The right side is the test display (pulsed). In different trials the dose of pulsed is varied from 1.0Q to 2.6Q to determine the equivalent-perception dose.

In paired-comparison subjects identify which display (left or right) allows one to visualize the low-contrast disks better. Responses are collected for each disk size as well as for all disk sizes together (the "global" response).
Equivalent-perception occurs at the dose level of *pulsed* where 50% of the subjects prefer the test method and 50% prefer the control method. In the min-contrast part we determine the contrast of the disk which is just detectable in the phantom. The equivalent-perception dose is obtained when an average min-contrast of *pulsed* equals *continuous*.

The 4-alternative forced-choice (4-AFC) test use a different display. Here a single disk is placed on a noisy background in one of four positions. and the subject determines the position of the disk. Forced-choice experiments are done with *continuous* at Q/acq. and *pulsed* at 1.2, 1.5 and 1.8 Q/acq. Data are analyzed according to a maximum-likelihood method [37], and equivalent-perception is specified by equal threshold detectability.

From the equivalent-perception dose, we calculate a dose savings per unit time. The fraction of the *continuous* dose is

\[ F = \frac{(\text{equivalent-perception}\_\text{dose}/\text{acq} \times \text{acq}/\text{sec})_{\text{pulsed}}}{(\text{dose}/\text{acq} \times \text{acq}/\text{sec})_{\text{continuous}}} \]  

(4.30)

and the percent savings is \((1 - F)\) 100%.

Some experimental results are shown later in comparisons with theoretical predictions. Important experimental results from Aufrichtig et al. [9] are:
• Both min-contrast and paired-comparison show a dependency of the equivalent-perception dose upon disk size. Forced-choice is not done as a function of disk size.

• For the larger disks, we predict dose savings of 22%, 38%, and 49% for \textit{pulsed-15}, \textit{pulsed-10}, and \textit{pulsed-7.5}, respectively.

• With regard to \textit{pulsed} and \textit{continuous} comparisons, we find remarkable similarity between the supra-threshold experiments (min-contrast and paired-comparison) and the detectability experiment (forced-choice). That is, the average absolute differences in the equivalent-perception dose as determined by the three measures is approximately 3%.

• The experimental methods are complementary – the forced-choice experiment properly eliminates the subjectivity of the observer threshold while the paired-comparison study is much more time efficient.

• There is no difference between interlaced and non-interlaced display on equivalent-perception dose.
4.5 Comparison Of Theory And Experiments

Figure 4.5 compares the equivalent-perception dose predicted by Model 1 to experimental data. The bars show the theoretical predictions, as well as the measurements from paired comparison, min-contrast and forced-choice experiments for a disk of radius 8 pixels. For paired-comparison, similar measurements are obtained from “global” responses to all disk sizes [9]. We find excellent agreement between theory and measurements, and the average absolute difference in equivalent-perception dose is 3%. Converting the numbers into dose savings, Model 1 predicts 22%, 37%, and 47% for pulsed-15, pulsed-10, and pulsed-7.5, respectively. As described previously, Model 1 predicts no effect on disk size with respect to the equivalent-perception dose.

In Fig. 4.6, we compare equivalent-perception dose values predicted by Model 2 with data from paired-comparison experiments [9]. Model 2 qualitatively gives the correct effects of disk size and acquisition rate. In particular, unlike Model 1, it demonstrates that smaller disks require a larger equivalent-perception dose. However, Model 2 systematically overestimates
Figure 4.5: Equivalent-perception dose values predicted by Model 1 are compared with paired-comparison, min-contrast and forced-choice experiments for a disk of radius 8 pixels. The different experimental measures, which are taken from Fig. 12 by Aufrichtig et al. [9], agree well. The forced-choice experiment is done only for pulsed-15. Model and measurements also match very well, and the average absolute difference in equivalent-perception dose is 3%.

In the case of Model 3 we use paired-comparison data and a least-squares fit (see section 4.3.3) to estimate the single internal noise parameter, and the comparison to data is improved (Fig. 4.7). With Model 3 we retain the
Figure 4.6: As a function of disk size and acquisition rate, the equivalent-perception dose for pulsed predicted by Model 2 is compared to paired-comparison experiments [9]. Although absolute agreement is imperfect, equivalent-perception dose increases with smaller disk size in both the data and the model. The numbers on top of the bars are the experimental and predicted values for the equivalent-perception dose.

effect of disk size, and we have a much better quantitative agreement with experimental data. The average absolute difference between the model fit and the experimental data is 7%.
Figure 4.7: As a function of disk size and acquisition rate, the equivalent-perception dose for *pulsed* predicted by Model 3 is compared to paired-comparison experiments [9]. The numbers on top of the bars are the experimental and predicted values for the equivalent-perception dose. Agreement is quite good, and the average absolute difference in equivalent-perception dose is 7%.

Again, there is only **one single value** of the internal noise parameter for all acquisition rates and disk sizes. The single estimated value determines the ratio, $N_i/V_{st}$ which depends upon acquisition rate and disk size. For
continuous. \( N_i/N_t \) is 84\%, 82\%, and 67\% for disks with radius 16, 8, and 4 pixels, respectively. These numbers are reduced to 48\%, 45\%, and 36\% for pulsed-7.5, where the external noise component is increased due to less temporal filtering.

4.6 Discussion

By modifying the ideal observer model to include the spatio-temporal characteristics of the HVS, we describe our previous experimental results in perception of low-acquisition-rate pulsed fluoroscopy [9]. The mechanism for dose savings is complex. In the formulation of SNR, as acquisition rate changes, the denominator (noise) changes while the numerator remains constant. In Fig. 4.3A, we find that at a dose Q/acq, noise energy moves to lower frequencies as acquisition rate decreases. Surprisingly, the effect is to increase the noise integral because lower temporal frequencies are less attenuated (Figs. 4.2 and 4.3B)! The explanation for the paradox is simple. Data in Fig. 4.3A are simulated for a dose Q/acq. Indeed, detection at low acquisition rates is inferior to that of continuous when both are acquired at a dose Q/acq. It is only when the dose/acq is increased beyond the equivalent perception dose that improved detection occurs. In terms of Fig. 4.3
for the case 7.5 acq/sec, this means that when dose is increased beyond
2.1 acq/sec (the equivalent-perception dose from Fig. 4.5), the noise power
curve shifts downward. the integral of the curve in Fig. 4.3B is decreased.
and the SNR increases as compared to continuous. Thus, it is the inter-
action between the acquisition rate noise characteristic and the low-pass
characteristic of the HVS which gives dose savings.

With no free parameters. Model 1 gives a good quantitative agreement
with the average equivalent-perception dose data. With the same ability to
visualize low-contrast objects. the model theoretically predicts dose savings
of 22%, 37%, and 47% for pulsed-15, pulsed-10, and pulsed-7.5. which is in
excellent agreement with the measured average dose savings of 22%, 38%.
and 49%, respectively. However, because this model incorporates a separa-
bale spatio-temporal response. all spatial filtering characteristics divide out
from the expression for the equivalent-perception dose (Eq. 4.26). Hence,
this model does not predict the dependence of equivalent-perception dose
upon disk size.

Model 1 incorporates a temporal response curve. using data from human
image perception flicker experiments performed by Kelly [56]. In these ex-
periments, a circular cosinusoidal stimulus pattern is used. Temporal vari-
ation is introduced by flickering the pattern sinusoidally in counterphase. Contrast sensitivity is measured with the method of adjustment, in which the subject adjusts the amplitude of the stimulus to be just detectable. Kelly’s flicker curves, measured at the peak spatial response, provides a single low-pass curve that is used in our simulation experiments. Similar curves are observed in other studies [57, 70].

It is remarkable that Model 1 fits the average responses so well. Not only is the experimental measurement of flicker very different from our experiments with disks in noise, the model predictions are done with no free parameters. Perhaps this match is not fortuitous. Note that the equivalent-perception dose tends to measure temporal characteristics of the HVS. Although we know there is variation in spatial contrast sensitivity across the human population [81, 85], temporal dissimilarities are less well investigated. Also, studies of the temporal contrast sensitivity are usually performed with alternating spatial stimuli [69, 57, 70, 71, 72, 73], hereby creating a spatio-temporal linkage. Because our experiments and model primarily respond to temporal characteristics, these results may imply relatively little variation in the temporal response.

Models 2 and 3 incorporate a model of the spatio-temporal charac-
teristic of the HVS proposed by Kelly [69]. This model describes many features of the HVS response. As the spatial frequency decreases, the temporal response asymptotically approaches the fixed Ganzfeld flicker response [69, 58]. Also, the spatial response asymptotes with decreasing temporal frequency [69]. Similarly shaped experimental spatio-temporal response curves are reported in other studies [57, 70, 71, 72, 73]. However, Kelly’s model has not been rigorously fit to responses from many subjects. In particular, it is not well defined at low spatial and temporal frequencies. Since the spatio-temporal response at \( f = 0 \), \( V_{st}(u, v, 0) \), is necessary to predict the equivalent-perception dose (see Eqs. 4.15-4.17), an ill-defined DC response may produce errors. Another potential inaccuracy results from the assumption of circular symmetry in the spatial frequency domain. Measurements of spatial contrast sensitivity along different meridians have found the acuity to be highest in the vertical and horizontal direction and in general, not circularly symmetric [56, 86, 55].

Despite these shortcomings, adding the non-separable spatio-temporal filter characteristic in Model 2 qualitatively gives the correct response for the effect of disk size. That is, it demonstrates that smaller disks require a larger equivalent-perception dose (Fig. 4.6). Once again, it should be
remembered that this is obtained with no free parameters using a spatio-temporal response model derived from entirely different experiments. It is not surprising that the quantitative absolute agreement is imperfect.

In Model 3, we introduce one free parameter describing the internal noise of the subject. An internal noise source has previously been incorporated in order to describe single-frame experiments [25, 26, 28, 49, 33]. It has been shown in psychophysical studies that the detectability of square objects [25, 26, 28, 49], and disks [33] superimposed in radiographic noise is predicted from this modified model. The internal noise has been attributed to inappropriate tuning of the observers perceptual filter, to variability in the detection procedure and decision threshold level, and to neurophysiological fluctuations [25, 26, 28, 68, 49, 33, 75, 76, 87, 88]. We assume the internal noise to be independent of the external noise, and model it by a single parameter. in a manner similar to several other studies [25, 26, 28, 49, 33]. However, Burgess et al. [75, 76] split the internal noise into two components: a constant factor, and an induced internal noise which is proportional to the external. Pelli [68, 88] has measured the noise level at the retina as a function of spatio-temporal frequency and has found the noise spectrum to be flat, except at very low spatio-temporal frequencies. We opt for the
simplest model, thereby limiting Model 3 to a single free parameter.

With Model 3, we retain the effect of disk size, and we have much better absolute quantitative agreement with the data (Fig. 4.7). As a result, the model can be used to predict the equivalent-perception dose for other object sizes. When small disks (radius 1.2 pixels) are used as input to the model, results identical to disks of radius 4 are predicted. Similarly, the model predicts no change in equivalent-perception between disks of radius 16 and 32. Thus, the model predicts that the range of disk sizes used in our experiments [9] cover the range of varying responses.

There are several potential refinements. Most importantly, more complete spatio-temporal data should be obtained. In particular, care should be taken at low spatial and temporal frequencies. It might be interesting to measure both the spatio-temporal characteristic and the equivalent-perception dose characteristic of a single subject.

In summary, Model 3 gives the best description at all disk sizes, but because of its simplicity and good absolute accuracy. Model 1 may be most useful for many future purposes. Examples include display assessment of temporal and spatial filtering as well as temporal interpolation of reduced frame-rate pulsed fluoroscopy.
Chapter 5

Spatio-Temporal Filtering of X-ray Fluoroscopy using Object Detection

5.1 Introduction

A different approach from pulsed fluoroscopy is to reduce dose and use image processing to compensate for the additional noise. To reduce noise, pixels nearby in space or time are typically combined in some fashion. Summing 10 pixels, each with 35 photons, gives 350 counts, and the standard deviation of the Poisson distribution is reduced from 17% to 5% of the mean. Simple averaging, however, usually results in unacceptable image blurring.

Image processing solutions on some existing medical products consist
of time-domain filtering, time-domain filtering with motion detection to reduce motion blur, and spatial filtering for contrast enhancement of objects. “Motion” is detected whenever a pixel intensity changes more than a prescribed limit between the current frame and the previous, and pixels are temporally filtered only where there is no motion. Unfortunately, image noise often results in false “motion” detection. Typically, with an obese patient, the skin dose maximum is reached: the images are noisy because of the decreased number of x-ray quanta detected; “motion” is detected: and no smoothing is performed.

Relatively little is published on image enhancement hardware for x-ray image sequences, particularly low-dose fluoroscopy images. Edmonds et al. have developed a real-time system for fluoroscopic filtering and applied it to gastrointestinal studies [7]. They create an output image from a weighted combination of spatially and temporally filtered input images. Temporal filtering is favored where no motion is detected and spatial filtering is favored where there is motion. Few system details are published. A product from Digivision includes a Wallis-type spatial filter for contrast enhancement as well as a recursive temporal filter [89]. Riederer et al. have developed a matched filter-type algorithm for optimizing the signal-to-noise ratio of the
contrast signal in a sequence of angiographic images [90].

An advanced method for noise reduction in image-sequences is motion-compensated temporal smoothing [91, 92, 93, 94]. Motion compensation requires that the instantaneous optic-flow field be explicitly computed prior to the filtering step. In the simplest case where two frames are averaged, the previous frame is warped such that object movement is reduced or eliminated. Averaging of the current frame with the warped version of the previous frame gives motion-compensated filtering. Temporal smoothing by a low-pass filter (such as a mean or median filter) reduces the noise variance, and motion compensation prevents blurring of moving regions. Motion estimation in sequences of coronary angiography images have been reported for image interpolation between image frames and for image coding [95, 96]. However, due to the noise level and the large interframe motion, optical flow methods may not provide accurate results. Further, the calculation is computationally demanding, and may not be suitable for real-time implementation.

Statistically based, spatial averaging filters have been developed for signal-dependent, Poisson noise [97, 98], which is the case for quantum-limited fluoroscopy images. This work has recently been extended to time
domain filtering by Kalivas and Sawchuck [94] and has been applied to x-ray images by Chan, Katsaggelos and others [99, 100, 101]. For the case of temporal filtering motion is estimated between two consecutive frames with a block-matching algorithm [101].

We propose spatio-temporal filtering with object detection. Particularly in non-cardiac angiography, most motion occurs in isolated, long, thin objects (catheters, guide wires, etc.). We demonstrate that such structures can be detected and roughly segmented using a matched filtering approach. We then locally adjust spatial and temporal filtering parameters such as to reduce object blurring. In contrast to motion-detection filtering using image differences, this scheme distinguishes between soft tissue and catheter motion. Note that blurring of soft tissues due to movement is generally not a problem.

In this paper, we describe the algorithm and optimize filter parameters using a receiver operating characteristic (ROC) curve analysis, which is computed without using human observers. We report numerical results from the processing method and describe several qualitative observations.
5.2 Algorithm

A flow diagram is shown in Figure 5.1. We spatially process each input image to create an object-enhanced image consisting of large pixel values where there is a high likelihood of an object being present. We consider various methods for accomplishing this task. The object-enhanced output images are used as control images. They feed look-up-tables (LUT's) to adapt spatial and temporal filter weights. In the case of temporal filtering, we apply a recursive temporal filter with spatially-dependent filter parameters. The method reduces filtering if an "object" has moved to or from a pixel. Similarly, we apply a spatially-varying noise reduction filter that again uses the object-enhanced image as a control image. The noise reduction filtering processes are implemented in a conservative manner such as to produce an enhanced sequence with reduced noise without blurring objects of interest.

5.2.1 Object Enhancement

Features of interest such as catheters, guide wires, arteries, and interventional devices, are all line-like structures. We wish to create an object-enhanced, background-removed image consisting of large pixel values where
Figure 5.1: This is a flow diagram of the algorithm. The input consists of a noisy fluoroscopy image sequence. Each image in the sequence is processed to create an object enhanced image consisting of large pixel values where there is a high likelihood of an object being present. This "object likelihood image" is used to control both spatial and temporal filter weights. That is guided by the object enhancement so that less spatial filtering is done across image edges, and less temporal filtering is done when object movement is detected. The result of the algorithm is a noise suppressed image sequence.

there is a high likelihood of such an object. Currently, we investigate two
spatial processing methods; matched filtering and unsharp mask processing with various blurring kernels. The methods rely on a finite base of support and are amenable to hardware implementation.

**Matched Filtering**

We wish to maximize the output due to the object (signal) relative to the background noise. For the case of white, additive noise, a matched filter is the linear filter that maximizes the ratio of the peak signal to the average noise power [102, 103]. Although x-ray fluoroscopy noise follows a signal-dependent Poisson distribution, one can assume additive Gaussian noise for small contrasts which result from a small change in the number of counts above background levels [24, 68, 79]. X-ray system fluoroscopy noise is not exactly white [104]. Also, tissue structure is even further from a white noise source and presents significant possibilities for false detection. Nevertheless, the matched filter is attractive, and we develop an enhancement scheme based upon the concept.

A matched filter consists of a convolution with a kernel which is a transposed version of the 2D object (signal) of interest [105, 106]. We often know the shape and size of an object but not the orientation. A solution is to use
oriented matched filters and take a maximum of several filter outputs. We utilize this concept and implement a method very similar to that described by Chaudhuri et al. for arteries in retinal images [106].

Our purpose is to detect objects like arteries and catheters. A mathematical model consists of the x-ray projection of a cylinder. The x-ray path-length, \( \mu l \), for parallel rays through a cylinder is

\[
\mu l(x) = 2\mu \sqrt{\frac{d^2}{4} - x^2} \quad \text{for} \quad |x| \leq d/2 .
\]  

(5.1)

where \( x \) is the distance from the center axis, \( d \) is the cylinder diameter, \( l \) is the geometric distance, and \( \mu \) is an effective linear attenuation coefficient for a cylinder containing contrast material [107].

Using Eq. 5.1, we now create a 2D matched filter kernel

\[
k(x, y) = k_o(x, y) - m .
\]  

(5.2)

where

\[
k_o(x, y) = \begin{cases} 
-\sqrt{\frac{d^2}{4} - x^2} & \text{for} \quad |x| \leq d/2 \text{ and } |y| \leq L/2 \\
0 & \text{for} \quad d/2 \leq |x| \leq W/2 \text{ and } |y| \leq L/2 
\end{cases}
\]  

(5.3)

\[
m = \frac{1}{LW} \sum_{x=-W/2}^{W/2} \sum_{y=-L/2}^{L/2} k_o(x, y) .
\]  

(5.4)
and $L$ and $W$ are the length and width of the cylinder, respectively. The term $k_0$ corresponds to a cylinder placed on a background with zero mean noise.

Several issues arise. First, the minus sign occurs since catheters and arteries are negative excursions. (In general catheters and blood vessels have higher absorbance to x-ray as compared to surrounding soft tissue, making the arteries appear darker than the background. An exception exists when CO$_2$ is used as a contrast agent. The gas displaces the blood and reduces absorbance [108]). Second, a matched filter is indifferent to a constant multiplicative term [102]. and we ignore the term $2\mu$ in Eq. 5.1. Third, the general definition of the matched filter contains a flat base region on either side of the projected cylinder for $d/2 \leq |x| \leq W/2$. We investigate the noise reduction effect of the base. Fourth, since we desire a background subtracted result, we create a high-pass kernel by subtracting the mean value. $m$ [106].

The convolution kernel specified by Eqs. 5.2-5.4 detects projected cylinders oriented vertically. We create kernels oriented at different angles using the transformation

$$k_\theta(x, y) = k(x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta).$$ (5.5)
where $\theta$ specifies the rotation angle. By sufficiently zero-padding $k(x,y)$ prior to rotation we insure that the correct size of $k_\theta(x,y)$ is achieved. Also, bilinear interpolation is used to obtain $k(x,y)$ at real-valued arguments. An input image is filtered with each oriented matched filters, and a set of intermediate output images is created. The maximum at each pixel position $(x,y)$ creates the final matched-filter enhanced image, $E_{\text{match}}(x,y)$,

$$E_{\text{match}}(x,y) = \max_{n=0}^{N-1} \{ I(x,y) \ast k_{2\pi n \theta}(x,y) \}.$$  \hspace{1cm} (5.6)

where $I(x,y)$ is the input image, $\ast$ symbolizes convolution, and $N$ is the number of oriented kernels. In our experiments, we most often use $N = 4$. The filtering scheme is illustrated in Fig. 5.2.

**Unsharp Mask Filtering**

For comparison purposes, objects are also enhanced with an unsharp mask filter. An advantage of this method is its computational simplicity. In unsharp mask processing, a smoothed version of the input image is subtracted from the original in order to create a high-pass filtered image [105, 49]. The equation is

$$E_{\text{unsharp}}(x,y) = I(x,y) - I(x,y) \ast h(x,y)$$  \hspace{1cm} (5.7)

Figure 5.2: The figure illustrates the oriented matched filter method to enhance line structures in a noisy fluoroscopy image. Initially the image is processed with four 2D matched filters consisting of projections of a cylinder. The 2D filter kernels with cylinders oriented at 0°, 45°, 90°, 135° can be calculated from Eq. 5.5. After matched filtering, a maximum operation is applied at each pixel position to give the final object enhanced image.

where \( h(x, y) \) is the blurring kernel. We investigate three blurring kernel structures: (1) a flat kernel, \( h_{flat}(x, y) \), (2) a Gaussian kernel, \( h_{gau}(x, y) \), and (3) a cone-shaped kernel, \( h_{cone}(x, y) \). The three kernels are:

\[
h_{flat}(x, y) = \frac{1}{\Sigma_{NUM}}
\]  
(5.8)
\[ h_{\text{gauss}}(x, y) = \frac{\exp\left(-\frac{1}{\pi \sigma^2}(x^2 + y^2)\right)}{\Sigma_{\text{NUM}}} \]  
\[ h_{\text{cone}}(x, y) = \frac{1 + \sqrt{x^2 + y^2}}{\Sigma_{\text{NUM}}} \]  
\[ \text{(5.9)} \]
\[ \text{(5.10)} \]

where \(|x| \leq p/2\) and \(|y| \leq p/2\) and \(\Sigma_{\text{NUM}}\) is a normalization factor consisting of the sum over the \(p \times p\) numerator values.

### 5.2.2 Spatial Filtering

Adaptive spatial processing is done with a \((2n + 1) \times (2m + 1)\) kernel. The kernel elements are defined by

\[ M(x, y) = \sum_{i,j} \frac{1}{\alpha_{i,j}} \left[ \begin{array}{cccccc}
\alpha_{-n,-m} & \alpha_{-n,-m+1} & \cdots & \alpha_{-n,0} & \cdots & \alpha_{-n,m} \\
\alpha_{-n+1,-m} & \alpha_{-n+1,-m+1} & \cdots & \alpha_{-n+1,0} & \cdots & \alpha_{-n+1,m} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\alpha_{0,-m} & \alpha_{0,-m+1} & \cdots & \alpha_{0,0} & \cdots & \alpha_{0,m} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\alpha_{n,-m} & \alpha_{n,-m+1} & \cdots & \alpha_{n,0} & \cdots & \alpha_{n,m}
\end{array} \right] \]  
\[ \text{(5.11)} \]

For each pixel in the object enhanced image, \(E(x, y)\), the \(\alpha_{i,j}\)'s are obtained from a look-up-table (LUT).

\[ \alpha_{i,j} = \text{LUT}(|E(x_0, y_0) - E(x_i, y_j)| \cdot \gamma_{sp}, \ z_{sp}) . \]

for \(i = -n, \cdots, n\) and \(j = -m, \cdots, m\).  
\[ \text{(5.12)} \]
Figure 5.3: The spatial filtering process is illustrated in this figure for the case of a $3 \times 3$ kernel. For each $3 \times 3$ region in the input image we calculate the argument to the LUT as the absolute difference between the center pixel and the surround. The filter weights are derived from the LUT specified by Eq. 5.13, and shown in Fig. 5.4. The output center pixel becomes a linear sum of the input weighted by the corresponding LUT values, represented by $\alpha_{i,j}$.

where $E(x_i, y_j)$ are the pixels values of $E$ corresponding to the kernel positions, and $\gamma_{sp}$ and $\beta_{sp}$ are LUT parameters used for the spatial filter. The method is illustrated in Fig. 5.3. The LUT function is a variation of a logistic sigmoid function defined by

$$LUT(x, \gamma, \beta) = \frac{1}{1 + \exp(-\gamma(x + \beta))}$$  \hspace{1cm} (5.13)

where $\gamma$ is the gradient of the function, and $\beta$ is the bias that indicates the amount of horizontal shift. The LUT function is plotted in Fig. 5.4 with $\gamma_{sp} = -0.07$ and $\beta_{sp} = -100$

### 5.2.3 Temporal Filtering

A spatially-dependent, recursive temporal filter is applied. In the temporal filter, the estimate of a pixel intensity $\hat{I}_t(x, y)$, at coordinates $(x,y)$ and time
t. is given by

\[ \hat{I}_t(x, y) = \hat{I}_{t-1}(x, y) + \kappa[I_t(x, y) - \hat{I}_{t-1}(x, y)] \quad (5.14) \]

where \( I_t \) is the measured intensity at the current time instant, and \( \kappa \) is a spatially-varying gain factor. The filter is applied independently to each pixel. \( I_t(x, y) \) in the image. When \( \kappa = 1 \), there is no filtering \( (\hat{I}_t = I_t) \); when \( \kappa = 0 \), the current frame is ignored \( (\hat{I}_t = \hat{I}_{t-1}) \); and when \( 0 < \kappa < 1 \), the filtered noise variance is \( \kappa/(2 - \kappa) \) of the input noise variance [109].

We control the value of \( \kappa \) with outputs from the object enhancement algorithm. A control image is obtained by subtracting the present image, \( E_t \), from the last one, \( E_{t-1} \). The value for \( \kappa \) comes from a LUT operation applied at each pixel location \( (x, y) \).

\[ \kappa = \text{LUT}(|E_t - E_{t-1}| \cdot \gamma_{tmp} \cdot \beta_{tmp} \cdot (\kappa_{max} - \kappa_{min}) + \kappa_{min}. \quad (5.15) \]

where \( \gamma_{tmp} \) and \( \beta_{tmp} \) are LUT parameters for the temporal filter, and \( \kappa_{min} \) and \( \kappa_{max} \) are limits of \( \kappa \). Since the absolute value is taken, a large value of the argument indicates the presence of an object in the current or previous frames, but not both. In this case, reduced filtering is desired, and we desire an upper limit, \( \kappa_{max} \) (\( \kappa_{max} \leq 1 \)). When the LUT argument is small, either there is no object present, or an object is present in both frames, and a
small value $\kappa_{\text{min}}$ ($\kappa_{\text{min}} \geq 0$), is desired. The LUT function is plotted in Figure 5.4 for $\gamma_{\text{tmp}} = 0.02$ and $\beta_{\text{tmp}} = -200$.

### 5.3 Methods

#### 5.3.1 Test Images

We perform experiments on real x-ray fluoroscopy images in order to optimize the object enhancement filters and to evaluate the filtering methods.

We acquire $512 \times 512$ images from an 8 bit digital fluoroscopy x-ray unit. A very low-dose acquisition is simulated by adding noise to the images. The added noise is signal-dependent, Poisson distributed and truncated to the digitization range. More specifically, we assume the number of detected x-rays, $N$, to be Poisson distributed around a mean of $N_0$. $N \sim \text{Poisson}\{N_0\}$, and a linear conversion between $N$ and gray-scale ($N = \lambda \times \text{gray\_scale}$, where $\lambda$ is a constant). The noise distribution of the gray-scale values is then

$$g(x, y) \sim \frac{1}{\lambda} \text{Poisson}\{\lambda f(x, y)\}$$

(5.16)

where $f(x, y)$ and $g(x, y)$, are the gray-scale values in the original and noise-added sequence, respectively.

For our evaluation we use 6 image sequences, acquired from a cardiac
Figure 5.4: The LUT used to control spatial and temporal filtering are plotted are shown. The LUT function, \( LUT(x, \gamma, \beta) \), is a variation of a logistic sigmoid function and is defined in Eq. 5.13. The function has two parameters which controls the gradient, \( \gamma \), and the amount of horizontal shift, \( \beta \). The slope at the center of the curve is \( \gamma/4 \), as indicated by the thin lines. The half point of the LUT, when \( LUT(x, \gamma, \beta) = 0.50 \), occurs for \( x = -\beta \), as shown by the dotted lines. The spatial filtering weights, \( \alpha_{i,j} \)'s, are calculated from the dashed curve, which has \( \gamma_{sp} = -0.2 \) and \( \beta_{sp} = -25 \). With this parameter setting the LUT falls off very quickly as the argument \( x \) increases. The temporal filter parameter \( \kappa \) is calculated from the solid curve, which is derived with \( \gamma_{tmp} = 0.01 \) and \( \beta_{tmp} = -500 \). This curve is further scaled from \( \kappa_{min} = 0.50 \) to \( \kappa_{max} = 0.90 \). With this setting, \( \kappa_{max} \) is archived when the argument is large, hereby minimizing temporal filtering, while more filtering is done when the argument is small.

cath lab (Siemens Hicor at MethoHealth Medical Center, Cleveland, Ohio).
consisting of between 62 and 124 frames. The cardiac fluoroscopy images are from a variety of views, and includes objects such as interventional guide catheters (7–11 F), diagnostic pig-tail catheter (7 F), guide wire for guide catheter (0.38"), interventional guide wires (0.14"), and inflated balloons.

We add noise using $\lambda = 2.50$, 0.50 and 0.75, and want to investigate the effect this noise addition has on the fluoroscopy dose level. Assuming inverse proportionality between dose, $Q$, and noise variance, $\sigma^2$, we can calculate the fraction between the simulated dose and the standard dose in the original sequence as:

$$F_{Q, \text{sim}} = \frac{Q_g}{Q_f} = \frac{\sigma^2_g}{\sigma^2_f}$$

where the subscripts $f$ and $g$ refer to original and noisy sequence, respectively. The variances are calculated from a preselected $100 \times 100$ region of interest (ROI), that only contains background tissue (i.e. no objects are present). Forty images are averaged to estimate the local mean, $\mu_f$, and the result is subtracted from each single frame. The local variance, $\sigma^2_f$ is then estimated from the background subtracted frames (The factor of 40/41 corrects the measured variance for the subtraction of a noisy estimate of the mean obtained from only 40 frames). A similar calculation is done for
<table>
<thead>
<tr>
<th>sequence</th>
<th>Fraction of standard dose, $F_{std}$</th>
<th>$\lambda$</th>
<th>2.50</th>
<th>0.75</th>
<th>0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.17 ± 0.05</td>
<td>0.04 ± 0.02</td>
<td>0.04 ± 0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.14 ± 0.03</td>
<td>0.05 ± 0.01</td>
<td>0.03 ± 0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.15 ± 0.03</td>
<td>0.05 ± 0.01</td>
<td>0.03 ± 0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.16 ± 0.04</td>
<td>0.05 ± 0.02</td>
<td>0.04 ± 0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.13 ± 0.02</td>
<td>0.04 ± 0.01</td>
<td>0.03 ± 0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.13 ± 0.03</td>
<td>0.04 ± 0.01</td>
<td>0.03 ± 0.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: This table lists the fraction of the standard dose, $F_{Q, \text{sim}}$, for the 6 test sequences used in the study. The table gives means and standard deviations calculated from 40 frames in each sequence. Three noise levels, $\lambda = 2.50, 0.75,$ and $0.50$, are used for all sequences. $F_{Q, \text{sim}}$ is calculated using Eq. 5.17, and the local variances are estimated from preselected background subtracted 100 x 100 regions of interests (ROI's), which have no objects present.

the noise added sequences to obtain $\sigma_g^2$.

Table 5.1 lists the fraction of the standard dose for the 6 test sequences used in the study. The Siemens Hicor is calibrated to 40 $\mu$R/sec at the input to the image intensifier. Using Eq. 5.17, and the values from Table 5.1 it follows that with $\lambda = 2.50, 0.75,$ and $0.50$, we simulate dose levels corresponding to approximately 6, 3, and 1 $\mu$R/sec.

### 5.3.2 Object Enhancement

For the case of the matched filter, several filter parameters need to be optimized. They are: (1) the length of the projected cylinder, $L$. (2) the
diameter, $d$, and (3) the width of the kernel, $W$, or the size of the base regions, $(W - d)/2$. For the case of the unsharp mask filters, we wish to determine the best shape (flat, Gaussian, or cone) as well as the optimal kernel size, $p$. Finally, the two methods are compared.

A visual inspection of the enhancement result is not satisfying because it is insensitive to small variations in the output. Instead, we use a quantitative measure that borrows from receiver operating characteristic (ROC) curve analysis.

**ROC Analysis**

The enhancement process should enhance objects of interest while not enhancing other structures. If we simply threshold an object enhanced image, we create a segmentation where pixels above and below threshold are marked as object and background, respectively. In general, pixels may be correctly identified as object (true-positive), incorrectly identified as object (false-positive), correctly identified as background (true-negative), or incorrectly identified as background (false-negative). We call these TP, TN, FP, and FN, respectively. A true-positive fraction (TPF) and a false-positive
fraction (FPF) are computed.

\[
TPF = \frac{\text{number of pixels correctly determined as object}}{\text{total number of object pixels}} = \frac{TP}{TP + FN}
\]  
(5.18)

\[
FPF = \frac{\text{number of pixels incorrectly determined as object}}{\text{total number of background pixels}} = \frac{FP}{TN + FP}
\]  
(5.19)

By varying the threshold level from low to high, we decrease TPF and increase FPF.

The analysis of such data arises in psychophysical signal-detection theory. A plot of TPF versus FPF for various threshold levels is called a receiver operating characteristic (ROC) curve [110, 111, 112]. ROC curve analysis is a very well accepted method for evaluating observer performance, and the technique has been widely applied in radiology to evaluate imaging modalities [112, 113, 114] and image processing techniques [115, 116]. We apply this methodology to analyze results from the object enhancement process.

There are some more definitions. The false-negative fraction (FNF) is the frequency with which object pixels are determined as background, and FNF = 1-TPF. Similarly, the true-negative fraction (TNF), also called specificity, is TNF = 1-FPF. Because of the complimentary relationship
between these measures, only two need to be reported, and generally TPF and FPF are used.

Figure 5.5 shows some hypothetical ROC curves. The more closely the curve approaches the upper left corner ([FPF, TPF] = [0, 1]), the better is the discrimination between object and background. Higher ROC curves are better since a greater TPF is recorded for each value of FPF. The straight line, where TPF = FPF, indicates pure chance. Thus, we desire the object enhancement method to create an ROC curve well above the straight line. Different methods exist for evaluating ROC curves [112, 113, 114]. We adopt one of the most frequently used, the area under the ROC curve, \( A_z \). A chance performance gives \( A_z = 0.5 \), while a higher value indicates an improved performance.

In human observer studies, an ROC curve often consists of \( \approx 5 \) data points [111], obtained by asking the observer to use five confidence threshold levels. To calculate the area, \( A_z \), a continuous curve is fitted through the data points. and often, a binormal form is assumed [111]. In the case of computerized segmentation, we compute FPF and TPF from a large number of samples and with small increments in threshold. For example, a typical object enhanced image consists of \( 512 \times 512 \) or 262,144 pixels.
Figure 5.5: A set of hypothetical ROC curves indicating the general shape of the curve at different enhancement performances. Curve 1 is the diagonal, which is the worst performance, indicating that the method selects object and background randomly. The area under Curve 1 is $A_1 = 0.50$. As the area under the curves increases the performance is improved. Thus Curve 5 gives the best performance and $A_5 = 0.92$.

Gray-values are scaled between 0 and 255, giving 256 possible threshold values. In this case, we get relatively smooth curves, and we compute the integral numerically using trapezoidal integration.
Calculation of ROC Curves

In order to calculate the ROC curve one must first obtain an image which contains the true positions of the objects. This is done by manually segmenting the images by drawing regions of interests (ROI's) around catheters in the image set. The software package DIP Station™ (Hayden Image Processing Group; Boulder, CO) is used to create the binary overlay image of the objects of interest.

Following manual segmentation, we apply the object enhancement method. The enhanced image is thresholded at 256 levels, and at each threshold we count the number of object and background pixels, and calculate the FPF and TPF. In this calculation we exclude areas of the image which are either outside of the image intensifier or have been collimated out. The ROC curve is plotted and $A_2$ is estimated for a variety of experimental conditions as outlined below.

Experiments

We compute ROC curves for 15 different parameter combinations of the matched filter kernel. We choose $(d, L, W) = [(1.7, 7); (1.9, 15); (1.11, 15); (3.7, 7); (3.9, 9); (3.9, 15); (3.11, 15); (3.13, 15); (3.15, 15);...
we compute ROC curves for various unsharp mask blurring kernels. We examine all 3 shapes (flat, cone, and Gaussian), each at 3 sizes \( p = [7: 11: 15] \). We compute \( A_z \)'s for all the ROC curves, and for each image at a given noise level, we obtain 24 values of \( A_z \). We chose optimal parameters and method based upon a maximum \( A_z \) value.

### 5.3.3 Evaluation of Spatio-temporal Filtering

All image sequences are filtered with the LUT parameters fixed to the levels specified in Fig. 5.4. In addition, the temporal filter gain, \( \kappa \) is constrained between \( \kappa_{\text{min}} = 0.5 \) and \( \kappa_{\text{max}} = 0.9 \).

The performance of the filter is measured by the signal-to-noise ratio improvement calculated in dB for each frame \([99. 101]\).

\[
I_{SNR}(t) = 10 \log_{10} \left( \frac{\sum_{x,y} [g_t(x,y) - f_t(x,y)]^2}{\sum_{x,y} [\hat{f}_t(x,y) - g_t(x,y)]^2} \right).
\] (5.20)

Above, \( f \) is the original sequence, \( g \) is the sequence being filtered after noise is added, and \( \hat{f} \) is the filter output. An average improvement, \( I_{SNR} \) is obtained by averaging the values of \( I_{SNR}(t) \) after the recursive filter has reached a steady state, which typically occurs within 10 frame times.

A second measure is the amount of dose reduction that can be obtained
from filtering. This is defined by

$$R_{Q, \text{filt}} = \frac{Q_f - Q_s}{Q_f} = \frac{\sigma_g^2 - \sigma_f^2}{\sigma_f^2}.$$  \hspace{1cm} (5.21)

where $\sigma_g^2$ and $\sigma_f^2$ are the noise variances of noisy and filtered image frames, respectively. Like in Section 5.3.1 proportionality between dose and variance is assumed, and the variance is calculated in a preselected background subtracted ROI. In order to relate this measure to the previously specified fraction between the simulated dose and the standard dose of the original sequence, $F_{Q, \text{sim}}$, we use the same $100 \times 100$ ROI’s to estimate the variance in the filtered sequences.

The above measure allows us to compute the fraction of the standard dose obtained after filtering, $F_{Q, \text{filt}}$, which can be calculated from Eqs. 5.17 and 5.21 as

$$F_{Q, \text{filt}} = \frac{Q_f}{Q_f} = \frac{\sigma_f^2}{\sigma_f^2} = \frac{F_{Q, \text{sim}}}{1 - R_{Q, \text{filt}}}.$$  \hspace{1cm} (5.22)

For spatial filtering only, the dose reduction can vary from 0 to $\frac{8}{9}$. Zero reduction occurs for the case where an object pixel is surrounded by eight background pixels, or vice versa. In this case, the arguments to the LUT are high, and no filtering is done. Maximum dose reduction occurs when all nine pixels belong to the same state (background or object), in which case
the arguments to the LUT are small resulting in maximum filtering. For the case of temporal filtering, the dose reduction can vary from $1 - \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} - 2$ to $1 - \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}}$, dependent upon the arguments to the LUT.

We calculate performance measures for all six image sequences. In order to investigate the effect of spatial and temporal filtering separately, we evaluate the SNR improvement for each of the two steps individually, and then combined. Finally, the filtering scheme is compared to the 2D Kuan spatial filter [97]. The Kuan filter is a noise smoothing filter which adapts to local changes in image statistics based on a non-stationary mean and non-stationary variance Poisson noise image model [97]. This filter is previously applied to fluoroscopy images [100, 101], and for these images it has been shown to perform better than the 2D Kak filter [117] which is optimized for Gaussian noise.

A critical issue with respect to spatial processing is the amount of image blur that is introduced as a result of filtering. Often a numerical value for noise reduction will be insensitive to edge blurring, and in general the best measure of performance is obtained from human subject evaluation [105]. We have selected the LUT parameters for both the spatial and temporal filters so as to minimize blur. The tradeoff is that less noise reduction is
obtained.

Using a typical frame from a noise added sequence (\( \lambda = 2.50 \)) which contains a guide wire and a guide catheter we compare the described LUT controlled spatial filter, with a regular averager and with the Kuan filter. We visually evaluate a portion of the frame before and after filtering, and investigate cross-sectional profiles through the guide-wire.

5.4 Results

5.4.1 Object Enhancement

Figure 5.6 shows an example of the output from matched filtering and unsharp masking. Figure 5.6A is the original image while 5.6B has added noise with \( \lambda = 0.5 \). Figure 5.6C-D and Fig. 5.6E-F is the match filter and unsharp mask output, respectively. The image contains a 10 F guide catheter with a 0.014" guide wire, which are typical image structures found in cardiac interventional imaging. The object enhancement is done with the \((5.15.15)\) cylinder kernel, and the unsharp mask is with the \(15 \times 15\) cone shaped kernel. Best performance is obtained with the matched filter especially for the noisy frame.

Figure 5.7 shows ROC curves obtained after applying matched filtering
Figure 5.6: This figure illustrates the effect of the object enhancement methods. The original $512 \times 512$ image (A) and the noise added ($\lambda = 0.50$) image (B) contains a 10 F guide catheter with a 0.014" guide wire. Both images are processed with the matched filter and with the unsharp mask filter. The size of the matched filter kernel is $15 \times 15$, and it contains a projected cylinder with a diameter of $d = 5$ pixels and a length of $L = 15$ pixels. For the unsharp mask processing a $15 \times 15$ cone shaped blurring kernel is used. The matched filter enhanced images are shown in (C) and (D), for the original and the noisy image, respectively. Similar, (E) and (F) are unsharp mask outputs. Notice that the matched filter results in a much better enhancement, especially for the noisy case.
<table>
<thead>
<tr>
<th>matched filter</th>
<th>Area under ROC curve, $A_z$</th>
<th>image 1 $\lambda$</th>
<th>image 2 $\lambda$</th>
<th>image 3 $\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d$ $L$ $W$</td>
<td>- 0.75 0.50</td>
<td>- 0.75 0.50</td>
<td>- 0.75 0.50</td>
</tr>
<tr>
<td>1</td>
<td>7  7  7</td>
<td>0.70 0.55 0.53</td>
<td>0.76 0.62 0.60</td>
<td>0.86 0.61 0.58</td>
</tr>
<tr>
<td>1</td>
<td>9  15 7</td>
<td>0.83 0.66 0.64</td>
<td>0.87 0.74 0.71</td>
<td>0.93 0.79 0.76</td>
</tr>
<tr>
<td>1</td>
<td>11 15 7</td>
<td>0.83 0.67 0.64</td>
<td>0.87 0.75 0.72</td>
<td>0.94 0.80 0.77</td>
</tr>
<tr>
<td>3</td>
<td>7  7  9</td>
<td>0.72 0.58 0.56</td>
<td>0.77 0.64 0.63</td>
<td>0.88 0.71 0.66</td>
</tr>
<tr>
<td>3</td>
<td>9  9  9</td>
<td>0.78 0.64 0.62</td>
<td>0.83 0.70 0.68</td>
<td>0.92 0.80 0.76</td>
</tr>
<tr>
<td>3</td>
<td>9  15 9</td>
<td>0.84 0.72 0.71</td>
<td>0.87 0.79 0.76</td>
<td>0.94 0.86 0.84</td>
</tr>
<tr>
<td>3</td>
<td>11 15 9</td>
<td>0.84 0.73 0.71</td>
<td>0.88 0.80 0.77</td>
<td>0.94 0.87 0.85</td>
</tr>
<tr>
<td>3</td>
<td>13 15 9</td>
<td>0.85 0.74 0.72</td>
<td>0.89 0.81 0.79</td>
<td>0.95 0.88 0.86</td>
</tr>
<tr>
<td>3</td>
<td>15 15 9</td>
<td>0.85 0.74 0.73</td>
<td>0.90 0.82 0.80</td>
<td>0.95 0.89 0.86</td>
</tr>
<tr>
<td>5</td>
<td>7  7  7</td>
<td>0.74 0.59 0.57</td>
<td>0.78 0.64 0.62</td>
<td>0.90 0.72 0.68</td>
</tr>
<tr>
<td>5</td>
<td>7  15 7</td>
<td>0.84 0.74 0.72</td>
<td>0.86 0.79 0.75</td>
<td>0.94 0.87 0.86</td>
</tr>
<tr>
<td>5</td>
<td>11 13 7</td>
<td>0.84 0.73 0.72</td>
<td>0.88 0.80 0.77</td>
<td>0.95 0.88 0.86</td>
</tr>
<tr>
<td>5</td>
<td>11 15 7</td>
<td>0.85 0.75 0.74</td>
<td>0.88 0.82 0.79</td>
<td>0.95 0.89 0.87</td>
</tr>
<tr>
<td>5</td>
<td>13 15 7</td>
<td>0.85 0.76 0.76</td>
<td>0.90 0.83 0.81</td>
<td>0.95 0.90 0.88</td>
</tr>
<tr>
<td>5</td>
<td>15 15 7</td>
<td>0.86 0.77 0.76</td>
<td>0.91 0.84 0.83</td>
<td>0.95 0.91 0.89</td>
</tr>
</tbody>
</table>

Table 5.2: This table gives values for the integrated area under the ROC curve, $A_z$, obtained by matched filtering. The filter parameters are cylinder diameter, $d$, cylinder length, $L$, and kernel width, $W$. Results from three different images are shown. ROC calculations are done for the original images (indicated by -) and at 2 noise levels: $\lambda = 0.75$ and 0.50. In all cases, the highest value of $A_z$ is obtained by the $(5, 15, 15)$ kernel.

and unsharp masking to the two images shown in Fig 5.6A-B. Figure 5.7A shows results from the original image, while Fig 5.7B is for $\lambda = 0.5$. The areas, $A_z$ under the ROC curves are shown in Tables 5.2-5.3, which also includes results at different noise levels and from two other image sets.

The results indicate that the matched filter in all cases is superior to the
Figure 5.7: Computed ROC curves after enhancement of Fig. 5.6A (A) and Fig. 5.6B (B) are shown. As indicated, the solid lines result from matched filtering, while the dashed lines are unsharp mask processing. The parameters for the two enhancement methods are given on the figure. These are (d, L, W)’s representing cylinder diameter, cylinder length, and kernel width for the case of the matched filter, and kernel size and type for the unsharp mask. Best performance is obtained with the (5, 15, 15) matched filter. For this filter, $A_z = 0.91$ and 0.83 for the original and noise added image, respectively. The largest area under the ROC curve for the unsharp mask is obtained with a $15 \times 15$ cone shaped kernel, which gives $A_z = 0.79$ and $A_z = 0.59$ for the two images.
Table 5.3: This table gives values for the integrated area under the ROC curve, $A_z$, obtained by unsharp mask processing. The three blurring kernels, $h_{uni}$, $h_{gau}$, and $h_{cone}$ of size $p \times p$ are specified in Eqs. 5.8-5.10. The ROC calculations are done for the same image sets that were used in Table 5.2. The $15 \times 15$ cone shaped kernel performs the best, but the $A_z$ values are significantly lower than those obtained by matched filtering.

Unsharp mask processing. Improved performance (larger $A_z$) is achieved as the width and length of the mask increases. Saturation of $A_z$ occurs approximately for $L = 15$ and $W = 15$. The best blurring kernel for the unsharp mask processing is the $15 \times 15$ cone shaped kernel. The worst kernel is the Gaussian, and virtually no enhancement is achieved with this filter. As the performance is close to chance ($A_z \approx 0.5$). As the noise increases ($\lambda$ decreases) the performance of the unsharp mask processing decreases drastically, while the matched filter is less sensitive to the noise. Based
upon these results, we chose the matched filter with \((d, L, W) = (5, 15, 15)\) for the purpose of object enhancement.

Figure 5.8 illustrates another property of object enhancement with the matched filter. With proper thresholding, the filter output can be used to generate a binary image containing the objects of interest. In Fig. 5.8A a manual segmentation of the catheter and guide wire from Fig. 5.6A is shown. Without processing it is not possible to get a binary representation of these objects (Fig. 5.8B). However, after matched filtering we get a very good representation of the catheter in the original image, as shown in Fig. 5.8C. Even when noise is added \((\lambda = 0.50)\) significant parts of the objects can be detected by a thresholding operation (Fig. 5.8D).

### 5.4.2 Spatio-temporal Filter

Figure 5.9(A-C) shows three frames from a fluoro sequence with added noise \((\lambda = 0.75)\). Figure 5.10(A-C) shows the output from the enhancement algorithm applied to the three frames. In Figure 5.11(A-C) we show the same three frames, after spatio-temporal filtering of the 60 frame sequence. Notice that the background noisy is greatly reduced, and that there is no blurring of the catheters, which have moved significantly from frame 4.
Figure 5.8: This figure shows an example of a manually segmented object (A), and a result of thresholding the original (B) and the matched filter output (C and D). The original image is shown in Fig 5.6A. This image was manually segmented to create the true positives and true negatives used in the ROC analysis. Figure (B) shows that without object enhancement it is not possible to threshold the original image to create a binary representation of the catheters. A much better result is obtained after image enhancement with the (5, 15, 15) matched filter kernels. In (C) the threshold level is chosen from the ROC analysis. This gives a FPF of $\approx 5\%$ and a TPF of $\approx 75\%$ for the original image in Fig 5.6A. Similar, the threshold level chosen in (D) gives a FPF of $\approx 5\%$ and a TPF of $\approx 60\%$ for the noise added image in Fig 5.6B.

through 30 and 60.
Figure 5.9: Three frames from a noisy fluoroscopy sequence are shown. The frame times are: (A) 5, (B) 30 and (C) 60. The noise is added using Eq. 5.16 with $\lambda = 0.75$.

Figure 5.10: The figure shows output from the object enhancement with the (5, 15, 15) matched filter, applied to the 3 frames shown in Fig. 5.9. Large pixel values in the images represents a high likelihood of an object being present. These enhanced images are used to control the amount of spatial and temporal filtering.

In Fig. 5.12 we plot the SNR improvement, $I_{SNR}$ for the sequence shown in Fig. 5.9 on a frame per frame basis. The figure shows the improvement for the case of object detection guided $3 \times 3$ spatial filtering, temporal filtering, and combined spatio-temporal filtering. Also, we show the performance
Figure 5.11: Output from the object guided spatio-temporal filter applied to the sequence shown in Fig. 5.9 after (A) 5, (B) 30 and (C) 60 frames. The background noise is significantly reduced as a result of the filtering process. There is no noticeable blur of the catheters, which are in motion during the cardiac cycle.

using the Kuan filter with two different window sizes: $3 \times 3$ and $5 \times 5$. For the 60 frame sequence, $I_{SNR}$ is lowest for the $3 \times 3$ Kuan filter (≈ 3.6 dB), while the $5 \times 5$ Kuan filter and the object guided $3 \times 3$ spatial filter give similar improvements of approximately 7.3 dB and 7.9 dB, respectively. Object guided temporal filtering alone is less effective with respect to $I_{SNR}$ than the spatial filter, giving an average improvement of 4.1 dB after 10 frames. However, the highest $I_{SNR}$ value is achieved by combining the two filters giving an SNR improvement of approximately 9.7 dB.

In Table 5.4 we list the mean and standard deviation of the SNR improvements calculated for the 6 image sequences at all three noise levels. Comparing spatial filtering alone, the object detection guided spatial filter
Figure 5.12: The improvement in SNR, $I_{SNR}$, for the sequence shown in Fig. 5.9 is plotted as a function of frame number. As indicated above, the figure shows the improvement for the case of object detection guided $3 \times 3$ spatial filtering, temporal filtering, and combined spatio-temporal filtering. Also, we show the performance using the Kuan filter with two different window sizes: $3 \times 3$ and $5 \times 5$. The highest SNR improvement is obtained by the spatio-temporal filter combination.

performs better than the $3 \times 3$ Kuan filter, and comparable to the $5 \times 5$ Kuan
filter. In general the SNR improvement increases with increasing noise levels (decreasing \( \lambda \)’s). This can be observed for all filter types, but is most prominent for the spatio-temporal combination. When temporal filtering is used, standard deviations of the \( I_{SNR} \) are higher than for spatial filtering alone. This is due to the variable degree of temporal filtering, which is dependent upon the amount of motion that occurs between successive frames.

Table 5.5 shows estimated percent dose savings, \( R_{Q, filt} \), obtained by the combined spatio-temporal filter. These are calculated using Eq. 5.21 for the same 100 \( \times \) 100 ROI’s that were used for the variance calculations in Table 5.1. At all three noise levels, the savings are very high, approaching 95%. The tendency is that somewhat higher savings are obtained from the most noisy sequences, where \( \lambda \) is 0.50. The fraction of the standard dose after filtering, \( F_{Q, filt} \) is the ratio between \( F_{Q, sim} \) which is given in Table 5.1, and \( (1 - R_{Q, sim}) \) which can be obtained from Table 5.5. Using data from the two tables and referring to 40 \( \mu R/sec \) as the standard dose level at the input of the image intensifier we estimate the dose level obtained by filtering to approximately 26, 34, and 78 \( \mu R/sec \) for \( \lambda = 0.50, 0.75 \) and 2.50, respectively.
<table>
<thead>
<tr>
<th>seq.</th>
<th>$\lambda$</th>
<th>spatial</th>
<th>temp.</th>
<th>spatio-temp.</th>
<th>$I_{SNR}$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>$5.8 \pm 0.1$</td>
<td>$3.0 \pm 0.5$</td>
<td>$6.1 \pm 0.6$</td>
<td>$3.1 \pm 0.1$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.75</td>
<td>$6.2 \pm 0.1$</td>
<td>$3.5 \pm 0.6$</td>
<td>$6.3 \pm 0.8$</td>
<td>$3.4 \pm 0.1$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.50</td>
<td>$7.5 \pm 0.1$</td>
<td>$3.9 \pm 0.1$</td>
<td>$9.9 \pm 0.4$</td>
<td>$3.2 \pm 0.3$</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>$6.2 \pm 0.1$</td>
<td>$3.5 \pm 0.6$</td>
<td>$6.3 \pm 0.8$</td>
<td>$3.4 \pm 0.1$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.75</td>
<td>$7.3 \pm 0.1$</td>
<td>$4.1 \pm 0.1$</td>
<td>$9.7 \pm 0.4$</td>
<td>$3.6 \pm 0.1$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.50</td>
<td>$7.5 \pm 0.1$</td>
<td>$4.1 \pm 0.1$</td>
<td>$10.4 \pm 0.6$</td>
<td>$3.5 \pm 0.2$</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>$6.0 \pm 0.1$</td>
<td>$4.1 \pm 0.1$</td>
<td>$8.5 \pm 0.5$</td>
<td>$3.2 \pm 0.2$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.75</td>
<td>$6.3 \pm 0.1$</td>
<td>$4.2 \pm 0.0$</td>
<td>$9.5 \pm 0.3$</td>
<td>$3.4 \pm 0.2$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.50</td>
<td>$6.3 \pm 0.1$</td>
<td>$4.2 \pm 0.0$</td>
<td>$9.5 \pm 0.3$</td>
<td>$3.4 \pm 0.2$</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>$7.6 \pm 0.2$</td>
<td>$3.7 \pm 0.2$</td>
<td>$9.3 \pm 0.3$</td>
<td>$3.6 \pm 0.2$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.75</td>
<td>$8.0 \pm 0.2$</td>
<td>$4.2 \pm 0.1$</td>
<td>$10.0 \pm 0.3$</td>
<td>$3.7 \pm 0.1$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.50</td>
<td>$8.0 \pm 0.2$</td>
<td>$4.2 \pm 0.1$</td>
<td>$10.0 \pm 0.3$</td>
<td>$3.7 \pm 0.1$</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>$8.1 \pm 0.3$</td>
<td>$4.0 \pm 0.1$</td>
<td>$9.7 \pm 0.5$</td>
<td>$3.7 \pm 0.0$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.75</td>
<td>$8.6 \pm 0.3$</td>
<td>$4.0 \pm 0.1$</td>
<td>$10.9 \pm 0.4$</td>
<td>$3.7 \pm 0.0$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.50</td>
<td>$8.6 \pm 0.3$</td>
<td>$4.0 \pm 0.1$</td>
<td>$10.9 \pm 0.4$</td>
<td>$3.7 \pm 0.0$</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>$7.9 \pm 0.2$</td>
<td>$3.9 \pm 0.1$</td>
<td>$9.4 \pm 0.4$</td>
<td>$3.6 \pm 0.1$</td>
</tr>
<tr>
<td>2.50</td>
<td>0.75</td>
<td>$8.3 \pm 0.2$</td>
<td>$4.1 \pm 0.1$</td>
<td>$10.6 \pm 0.3$</td>
<td>$3.7 \pm 0.1$</td>
</tr>
</tbody>
</table>

Table 5.4: This table lists the average improvement in SNR, $I_{SNR}$, as calculated from Eq. 5.20. The table gives means and standard deviations calculated for the 6 image sequences. The $I_{SNR}$ values are calculated for object detection guided 3 x 3 spatial filtering, temporal filtering, and combined spatio-temporal filtering. The $I_{SNR}$ is also calculated for two spatial Kuan filters: a 3 x 3 and a 5 x 5. For all 6 sequences, the highest SNR improvement is obtained by the spatio-temporal filter combination.

We have chosen LUT parameters that minimize spatial and temporal blur. Figure 5.13 illustrates how well the object guided spatial filter performs around catheter edges. Figure 5.13A shows part of an image from an original sequence and a corresponding cross-section through the image. The two valleys in the cross-section clearly indicate the presence of the guide
Table 5.5: This table shows estimated percent dose savings, $R_{Q, fit}$, for the 6 image sequences obtained by the combined spatio-temporal filter. The values are calculated using Eq. 5.21 for the same $100 \times 100$ ROI's that were used to estimate the fraction of the standard dose, $F_{Q, sim}$, for the noisy sequences given in Table 5.1.

catheter and guide wire. In Fig. 5.13B we show the image and profile after noise addition with $\lambda = 2.50$. The images and profiles in Fig. 5.13C-F are obtained after filtering with the $3 \times 3$ object guided spatial filter, a $3 \times 3$ averager, a $3 \times 3$ Kuan filter, and a $5 \times 5$ Kuan filter, respectively. From the figure it can be seen that the averager and the $5 \times 5$ Kuan result in significant spatial blur, as the profiles through the catheters are wider and have a smaller amplitude. Both the $3 \times 3$ object guided spatial filter and the $3 \times 3$ Kuan filter have minimal blur, but more noise reduction is obtained with the object guided filter.
Figure 5.13: The figure shows part of a guide catheter and guide wire with a corresponding cross-sectional profile along the dashed line. (A) is from an original image, while (B) is noise added at $\lambda = 2.50$. (C) is obtained after filtering the noise image with the $3 \times 3$ object guided spatial filter. (D) is processed with a $3 \times 3$ averager. A $3 \times 3$ Kuan filter is used in (E), and a $5 \times 5$ Kuan filter in (F). Most spatial blur occurs in (D) and (F). The $3 \times 3$ object guided spatial filter has minimal blur, and more noise reduction than the $3 \times 3$ Kuan filter.

5.5 Discussion

As compared to motion-detection filtering using image differences [41, 63],
the present method incorporates a priori knowledge about moving struc-
tures of interest. The object-enhanced image gives a likelihood for the presence of long, thin features such as catheters and guide wires, and it is used in control functions for both spatial and temporal filtering. The object-enhanced image seems quite appropriate for reducing spatial blurring of structures of interest. With respect to SNR improvement, the $3 \times 3$ spatial filtering performs better than the $3 \times 3$ Kuan filter, and similar to the $5 \times 5$ Kuan filter (Fig. 5.12 and Table 5.1). However, blurring of edges is significantly reduced with the $3 \times 3$ LUT controlled spatial filter as compared to the $5 \times 5$ Kuan filter (Fig. 5.13).

Object-enhanced image differences are used in the temporal control function (Eqs. 5.14-5.15), and temporal filtering is reduced whenever objects of interest move from one frame to the next. Thus, although the object enhancement may be imperfect and also enhance other stationary features in the image such as bone edges, etc., this does not create a problem. These features will simply be temporally filtered if they remain stationary between two consecutive frames.

The object enhancement step is critical to the success of the algorithm. Small kernel, edge-enhancing filters such as a Sobel filter are inappropriate because we wish to enhance the entire object, not just its edges. Further, the
performance of these operators is highly degraded in the presence of noise at fluoroscopic levels [106]. All methods that we currently evaluate take advantage of the spatial averaging characteristics of a large kernel convolution operation. We find these methods superior to a morphological approach investigated earlier [118]. This morphological filter, like others, tends to perform poorly in the presence of large, Gaussian noise variances [119, 120].

The advantage of using the unsharp mask filters for object enhancement is computational speed. The flat kernel is fastest, as it is separable and can be implemented with running averages. The Gaussian kernel is likewise separable, but virtually no enhancement is obtained with this shape. The best performance is obtained with the cone shaped blurring kernel. This kernel with a negative excursion better matches the intensity profile of the artery, and gives a better estimate of the background.

The ROC analysis indicates that significantly better enhancement is achieved with the oriented matched filters (Tables 5.2 and 5.3). Typically, we use cylinders in four orientations. More orientations (6, 8 and 12) show no increase in the area, \(A_z\), under the ROC curve. As the number of orientations increase, the sensitivity (TPF) increases, but simultaneously, the specificity (1 - FPF) decreases. However, with only two orientations smaller
A₁’s are obtained. Four orientations seems to be a reasonable compromise.

The shape of the filter output is important. The matched filter used here has a relatively broad output, and we use this feature to create a good control image.

We suggest the use of large filter kernels that necessarily require a large number of computations. However, there is a finite base of support and this allows parallel processing and computation with fast neighborhood processors. Current technology of 2D convolver chips approach the demands of the algorithm [121]. Computations resulting from multiple orientation used in the matched filtering can be performed in parallel. Also, the 0° and 90° kernels are separable.

There are other possibilities for improved object enhancement. Model based approaches similar to that used in edge detection [105] may also be possible. Dallas et al. presents an interesting approach based upon dynamic programming [122, 123]. Hoffman et al. have described a double-square box region-of-search algorithm for vessel tracking [115, 124], but this method requires user interaction to indicate starting points. Earlier, we investigated hysteresis thresholding [64]. This method results in a superior object enhanced image by greatly reducing false positives: however, it necessarily
requires serial processing. In summary, some of these methods may be promising, but they will add computational complexity. Since our targeted application is filtering of x-ray fluoroscopy, we are obligated to consider algorithms which can run in parallel.

The algorithm uses LUT's to control both spatial and temporal filtering. In earlier implementations [64], we used a binary decision to regulate spatial filtering. In general, a LUT approach allows one to blend the amount of spatial processing from one region to the next. Without such blending, one can see differences in texture. The input to the second LUT which controls the temporal filtering consists of object-enhanced image differences. This approach is more reliable than non-enhanced image differences which are very noisy. With the matched filter approach, we basically maximize the signal-to-noise ratio in the control image. Once again, the use of a LUT is much superior to a threshold approach. We adjust the LUT parameters to create conservative filtering. However, one can easily adjust the parameters such as to create more noise reduction at the expense of possible spatio-temporal blurring.

To evaluate the noise reduction characteristics on a sequence, we use average SNR improvements as defined by Eq. 5.20. This measure has also
been used by others [99, 101]. Although numerical evaluation is attractive, it does not account for small defects such as occasional edge blurring in a single image or ghosting that occurs in an image sequence when residual gray-values remain. Hence, in addition to numerical evaluation, one must perform visual evaluation to insure that the objects of interest remain intact after filtering (Fig. 5.13).

Despite the shortcomings of a numerical variance evaluation, it is of interest to predict a dose savings. An indication of the dose savings is obtained by assuming inverse proportionality between x-ray dose and noise variance. When the variance is calculated from ROI’s containing no moving objects, the estimated dose savings approximate 95% (Table 5.5). Using the same relationship between dose and noise variance, we are able to estimate the dose level that is achieved by spatio-temporal processing. A standard dose value at the input to the image intensifier is 40 \( \mu R/\text{sec} \). In our simulations, we use \( \lambda = 0.50, 0.75 \) and 2.50, corresponding to a dose of \( \approx 1, 3, \) and 6 \( \mu R/\text{sec} \). After spatio-temporal processing, the noise reduction of the image sequences results in an equivalent dose of 26, 34, and 78 \( \mu R/\text{sec} \) respectively. These dose savings are significantly larger than what can be achieved using low frame rate pulsed fluoroscopy [9, 19], and
also larger than savings from semi-transparent collimators, like the x-ray fovea [10, 11, 12, 13].

Currently, we opt to error on the side of conservative noise-reduction processing whereby our "gold standards" are high-dose input image sequences. We try to ensure that the filtered images resemble the input images as closely as possible. From the above calculations it follows that for $\lambda = 0.50$, 0.75, we do not recreate the $40 \mu R/sec$ of the original sequence. However, with $\lambda = 2.50$ we are able to create a sequence that contains even less noise than the original.

An interesting possibility is to contrast enhance the objects of interest. Currently, we reduce noise but do not change gray-scale values of the detected objects, so as to increase their detectability. Image perception experiments have been used to study detectability in pulsed fluoroscopy at reduced acquisition rates [9, 19, 107, 52]. These experiments can also be applied to evaluate image processing methods, and there exists a need to carry out some basic image perception studies to determine the appropriateness of image enhancement approaches.
Chapter 6

Conclusion and Future Directions

6.1 Conclusion

X-ray dose is a critical issue in fluoroscopy when long, complex studies are performed. The technology is at a point of saturation: images are x-ray quantum limited and any further reduction in dose may impair the physicians’ ability to perform interventional tasks. The focus of this work has been to describe methods to reduce x-ray dose without degrading image quality. The inspiration for the work was driven by two facts (1) little was known regarding proper dose levels in pulsed fluoroscopy and (2) current fluoroscopy systems are not incorporating advanced image processing solutions for noise reduction.
First, our work on establishing proper dose levels for pulsed fluoroscopy is unique. Prior to this work, industry had not provided the consumer with rational information, and x-ray vendors were giving conflicting information about dose levels in pulsed mode. We have chosen to investigate pulsed fluoroscopy perception at a fairly basic level. This includes perception experiments with stationary low-contrast objects in noisy image sequences as well as a model of the phenomena.

We describe several interesting aspects of fluoroscopy perception. Pulsed fluoroscopy of stationary objects gives average dose savings of 22%, 38%, and 49% at 15, 10 and 7.5 acq/sec, respectively. Our model that includes the spatio-temporal response of the human visual system (HVS) describes effects of acquisition rate and of object size with only one free parameter. Since the model uses an independently obtained spatio-temporal filter, a good fit indicates that the measured response is due to the interaction between the acquisition rate noise characteristic and the spatio-temporal characteristic of the HVS.

Second, we have developed an image enhancement algorithm to reduce noise in x-ray fluoroscopy without blurring the objects of interest (catheters, guide wires, etc.). The ultimate goal is to reduce radiation dose without sig-
nificantly degrading image quality of the quantum limited images. We have investigated a novel approach consisting of filtering with object-detection that allows us to achieve this.

6.2 Future Directions

While this work has initiated an important research area, much work still needs to be done. With respect to perception studies this work has focused on stationary objects. The obvious extension is to included motion so as to determine the contrast detectability properties of pulsed acquisitions for moving objects. The question to ask is: “Which allows easier detection of moving objects: pulsed at reduced acquisition rates with reduced single-image motion blur or continuous at standard acquisition rates?” As for the case of stationary objects, the size of the moving objects may become a factor. Preliminary results from experiments with motion indicate that pulsed-15 saves dose as compared to pulsed-30 [52]. However, the savings decrease for high velocities and small objects. These results may have direct implications for clinical imaging. In non-cardiac applications, where motion is reduced, pulsed-15 is probably adequate and can be used at reduced dose. With regard to cardiac applications, perhaps smaller objects, such as guide
wires, are better seen with \textit{pulsed-30}, while \textit{pulsed-15} is sufficient for imaging larger arteries. An interesting idea is to change the acquisition rate during the cardiac cycle as the velocity of objects change during contraction and filling.

The incorporation of motion also introduces a challenge in the ideal observer model. For the stationary case we describe a “signal-known-exactly” model. For the case of motion one may assume a “signal-and-velocity-known exactly” model. However, basic eye tracking perception experiments must be carried out to verify this model extension. An immediate problem may be that tracking does not occur prior to detection. However, it is possible that knowing the velocity assists the detection process. That is, the detection response may be delayed a short period of time in which the visual system is tracking the moving objects.

Another future aspect is to develop and evaluate techniques for reducing the “choppiness” of reduced acquisition-rate. \textit{pulsed} fluoroscopy sequences. One particular problem is the visual impression of two catheters when a catheter moves rapidly. Visualization can be improved by interpolating the motion using techniques similar to those required in motion-compensated filtering or in the object detection method. One problem with temporal
interpolation is that a frame delay is introduced. However, results from the current use of *pulsed* indicate that such a delay would have little influence on the hand-eye coordination. By extending the described ideal observer model to include an external temporal filter, it is possible to use the model to evaluate the effect of interpolation. In other words, one may use the model to investigate the effect of applying linear interpolation to *pulsed* instead of gap-filling or frame repetition.

This model extension can also be used to predict the minimal amount of temporal filtering of *pulsed* that will result in equivalent visualization with *continuous* at a given (e.g. 50%) dose savings. That is, by including a temporal filter one can use the model to optimize the filter gain to achieve specific dose savings. Based upon the predictions, paired-comparison, min-contrast or forced-choice experiments can be used to validate this extended model.

As in previous ideal observer models, the final output from our extended model is a signal-to-noise ratio (*S.N.R*). We have used this *S.N.R* output as a relative measure in which we equate the *S.N.R* of *pulsed* with the *S.N.R* of *continuous*, to compute an equivalent-perception dose. It would be very interesting to use the *S.N.R* as an absolute term, and relate this term directly
to the detectability of low contrast objects, hereby allowing one to predict
the threshold contrast of a specific object at a given acquisition rate.

In conclusion, further research is needed in all three investigated areas:
perception experiments, modeling, and filtering. Perception experiments
must be used to evaluate filtering results, and further modeling may help
to better understand the spatio-temporal processes occurring in the visual
system, when it is stimulated by low-contrast objects in a noisy environ-
ment.
Bibliography


153


