



Universität des Saarlandes
Max-Planck-Institut für Informatik



Perceptually Driven Frame-Rate Manipulations

Masterarbeit im Fach Informatik
Master's Thesis in Visual Computing
von / by

Junaid Ali

angefertigt unter der Leitung von / supervised by

Piotr Didyk

betreut von / advised by

Piotr Didyk

begutachtet von / reviewers

Karol Myszkowski

Piotr Didyk

Saarbrücken, April 2017

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Statement in Lieu of an Oath

I hereby confirm that I have written this thesis on my own and that I have not used any other media or materials than the ones referred to in this thesis.

Einverständniserklärung

Ich bin damit einverstanden, dass meine (bestandene) Arbeit in beiden Versionen in die Bibliothek der Informatik aufgenommen und damit veröffentlicht wird.

Declaration of Consent

I agree to make both versions of my thesis (with a passing grade) accessible to the public by having them added to the library of the Computer Science Department.

Saarbrücken, April 2017

Junaid Ali

Abstract

Rising popularity of high frame-rates (HFR) in the films has sparked a lot of debate. Proponents of the technology claim it to be more realistic, smoother, and sharper. While opponents claim that it looks too realistic and smooth, rendering the films “cheap” and “soap-operatic”. Some critics also complain that HFR looks sped-up. Variable frame-rate technology gives filmmakers more control on manipulating frame-rates. However, there is no method or guide for the selection of frame-rate.

First part of this thesis investigates the urban legend that speed perception is dependent on frame-rate. To measure the dependence of frame-rate on speed perception, we performed three psychophysical experiments using basic stimuli (Gabor patches), animated content and real-world footage. The results show that there is no significant effect of frame-rate on speed perception. There was a lot of variation in participants’ responses. Inconsistencies in people’s opinions regarding HFR content is not new. There are studies which demonstrate the incongruity of participants’ opinions regarding HFR. This thesis concludes that frame-rate does not alter speed perception. Perhaps, lack of artifacts in HFR is some times being misconstrued by some imaginative viewers as difference in speed. Even if there is an effect, it is too small to be measured.

Flicker can be defined as jittery motion due to temporal variation in pixels. Since flicker is the most prominent difference between two frame-rates, we propose that it is also a major contributor to the “film-look”. This work explores the relationship between frame-rate and flicker. Given an input video, we present a model to calculate per-pixel flicker. We also present a flicker driven guide to use variable frame-rate technology. For this purpose, we take input video and goal flicker visibility maps, and solve an optimization problem for spatially and temporally variable, and spatially and temporally smooth frame-rate maps, such that perceived flicker matches with the goal flicker. We use these frame-rate maps as an input for the method proposed by Templin et al. to produce output frames, which have required flicker. By manipulating flicker with our technique, filmmakers can reap the benefits of HFR, without incurring the unwanted “soap-operatic” effects.

Acknowledgements

I would like to thank, Karol Myszkowski for his guidance and support for all these years. I would also like to thank my advisor, Piotr Didyk for immense help during my thesis. Furthermore, I am grateful to my parents and family for nagging me to finish my thesis quicker.

Contents

Abstract	v
Acknowledgements	vii
Contents	xi
List of Figures	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Problem	2
1.3 Proposed Solution	3
1.4 Contributions	3
1.5 Outline	4
2 Background and Related Work	5
2.1 Camera	5
2.2 Displays	6
2.3 Motion Artifacts	7
2.3.1 Motion Blur	8
2.3.2 Hold-type blur	8
2.3.3 Repeated Edges	8
2.3.4 Strobing	9
2.3.5 Flickering	9
3 Effects of Frame Rate on Speed Perception	11
3.1 Factors affecting speed perception	12
3.2 Experiment I: HFR Effects on Speed Perception using Gabor Patches . .	14
3.2.1 Setup	14
3.2.2 Results and Discussion	15

3.3	Experiment II: HFR Effects on Speed Perception using Animated Content	16
3.3.1	Setup	16
3.3.2	Results and Discussion	22
3.4	Experiment III: HFR Effects on Speed Perception Using Real-World Videos	23
3.4.1	Setup	23
3.4.2	Results and Discussion	25
3.5	Discussion	27
3.6	Implementation details	28
3.6.1	Optical Flow	28
3.6.2	Interpolation	29
4	Manipulation of Flicker using Variable Frame-rates	31
4.1	Previous Work and Background	32
4.1.1	Flicker Perception	32
4.1.1.1	Spatial Extent	32
4.1.1.2	Temporal and Spatial Contrast	33
4.1.2	Frame-rate	34
4.1.3	Flicker and Frame-Rate	36
4.2	Flicker Model	37
4.2.1	Retinal Images	38
4.2.2	Multi-scale Contrast Processing	40
4.3	Framerate Manipulations	41
4.3.1	Introducing Film-look	43
4.3.2	Framerate Adjustments using Target Flicker	44
4.3.3	Results	46
5	Conclusion and Future Work	51
6	Appendix	53
6.1	If Frame-Rate Affects Flicker non-Linearly	53
6.2	Results: HFR Effects of Speed Perception using Gabor Patches	55
	Bibliography	63

List of Figures

2.1	Amount of time sensor is exposed to light at different shutter angles. . . .	6
3.1	Examples of gabor patches used in experiment	14
3.2	The top and bottom surfaces are interpolated from the point obtained from experiment and control cases, respectively. The speed of the stimuli was fixed at 2° per second.	16
3.3	The top and bottom surfaces are interpolated from the point obtained from experiment and control cases, respectively. The speed of the stimuli was fixed at 8° per second.	17
3.4	The top and bottom surfaces are interpolated from the point obtained from experiment and control cases, respectively. The speed of the stimuli was fixed at 16° per second.	18
3.5	The results of difference between speed matches in experiment and control cases, for speed 2° per second.	18
3.6	The results of difference between speed matches in experiment and control cases, for speed 8° per second.	19
3.7	The results of difference between speed matches in experiment and control cases, for speed 16° per second.	19
3.8	Bunny scene.	20
3.9	Bird scene.	20
3.10	Butterfly scene	21
3.11	Skipping scene.	21
3.12	Squirrel scene.	22
3.13	X-axis shows names of the scenes, and Y-axis shows the matched speed of the test sequences. Each bar represents average of all the participants' responses. The results are also averaged over both the comparisons they were asked to make for every case.	22
3.14	Person scene. © <i>Junaid Ali</i>	24
3.15	Car scene. © <i>Junaid Ali</i>	24
3.16	X-axis shows type of the experiment and Y-axis shows number of times test sequence was perceived faster, compared to reference. Each bar represents sum of the times all participants perceived test sequence, at every speed-up and both videos, to be faster, for a particular type of experiment. Red lines mark 50 % and 100 % of the readings.	26

3.17	Each dot represents percentage of participants who perceived test to be faster, for that specific type of experiment. First four cases are for the “person” scene and the next four cases are for “car”	27
4.1	The spatial extent changes based on size and speed of the moving object.	32
4.2	Pipeline for calculating Flicker maps. We start with input frames, then calculate retinal images, build a Gaussian pyramid for every retinal image and calculate temporal contrast at every band of the Gaussian Pyramid. We take the maximum value of contrast across all bands of a pyramid, then we convert the values into flicker levels by multiplying contrasts with corresponding sensitivity values. We take maximum value of flicker across all pyramids to produce the final Flicker map.	37
4.3	Demonstration of scan-line of a bar moving from left to right. It shows discrete and continuous cases, as well as position of the scan-line as it moves and its projection on the retina. With the increase of frame-rate, the spatial extent of the flickering region shortens. The red lines show the spatial extent of the flickering region.	38
4.4	The readings from Makela et al. and its logarithmic fit.	42
4.5	Demonstration of variable frame-rate emulation method of Templin et al. Left figure shows uniform temporal sampling of video. Middle figure shows displacement of kernels. In the limit, kernels at t_1 and t_3 would merge into t_0 and t_2 , respectively. This would result in half of input frame-rate, as shown on the right of the figure.	42
4.6	Results from Daly et al. of effect of frame-rate on judder perception. The graph shows that as the frame-rat increases perceived judder decreases linearly.	43
4.7	Scenes used to demonstrate examples of enhancing film-look. In the bottom figure bars of different contrasts are moving from left to right with different speed. The figure shows that high frame-rate corresponds to high contrast and fast speeds.	47
4.8	For these results flicker values was fixed at 0 JND. The left column shows sample frames of scenes named alley, temple and mountain. The middle columns shows their corresponding optical flow. The column on the right shows frame-rate map.	48
4.9	Demonstration of flicker maps at different frame-rates. As the frame-rate gets lower flicker increase in certain regions.	48
4.10	Demonstration of difference of flicker maps between 60 FPS and other frame-rates. As the frame-rate decreases, flicker increases in regions with high contrast and speed.	48
4.11	Flicker maps overlaid on the frames. The results show that flicker is high around high contrast and fast moving parts such as, around the hand in alley scene, around the wings of the dragon in temple scene, and around the dark spots in the mountain scene.	49
4.12	Four bars, of different contrast moving from left to right with same the speed of 6 px per frame for 60 FPS. The results demonstrate that flicker is greater for the higher contrasts.	49
4.13	Four bars are moving from left to right with different speeds of 6,9,12 and 15 pixels per frame for 60 FPS. The results demonstrate that flicker is greater for the faster speed.	50
6.1	56

6.2	57
6.3	57
6.4	58
6.5	58
6.6	59
6.7	60
6.8	60
6.9	61

Chapter 1

Introduction

The recent trend of high frame-rates (HFR) in films has solved a lot of problems, such as strobing, ghosting or repeated edges and flickering. Psychophysical experiments performed by Wilcox et al. [45] showed that HFR fares better on realism, motion smoothness, blur/clarity, quality of depth, and overall preference parameters. However, HFR has also raised some questions, e.g. some people complained that it looks sped-up [30] and [28]. In the first part of the thesis, we present experiments performed to investigate the myth that speed perception is influenced by HFR.

Another critique on HFR is that it looks “soap-operatic” and “cheap”, and it does not have the “film-look” of traditional 24 frames per second (FPS) movies. Flickering is one of the prominent artifacts produced at such low sampling rates. When the temporal variations in a pixel are not fast enough, it appears jittery, this phenomenon is called flickering. Amount of perceived flicker is dependent on the frame-rate. Daly et al. [6] showed that flicker reduces linearly, as the frame-rate is increased. One of the prominent differences between two frame-rates is flicker [38]. We propose that it is also a major contributor to the “film-look”. In the second part of the thesis, we present a method to manipulate “film-look” by varying frame-rates, based on the visibility of desired flicker. To this end, we investigate the relationship of flicker and frame-rates.

1.1 Motivation

For almost a century film industry has been using 24 frames per second (FPS), but there has been a recent shift to employ HFR. “The Hobbit” trilogy was shot at 48 frames per second (FPS), “Billy Lynn’s Long Halftime Walk” was made at 120 FPS and James Cameron announced that HFR will be used to shoot sequel of the movie “Avatar”. The

traditional 24 FPS limits the filmmakers to a very narrow range of speeds that they can use without incurring unpleasant artifacts. Some of these artifacts exist in real-world and are enhanced, while others occur just due to discretization and under sampling. Due to these constraints filmmakers are somewhat restricted. The American cinematography manual suggests that, in order to avoid unwanted artifacts, an object should not cross the screen in less than 7 seconds, [31]. Due to high temporal sampling rate of HFR, it incurs less artifacts such as blur, judder etc. On the other hand, there has also been a lot of critiques on HFR. Some people say that it makes the movies look “soap-operatic” and “cheap-looking”, or content seems to be sped-up. Some viewers also complain about motion sickness. Most of these problems seem hard to quantify.

Even with employing HFR, filmmakers are restricted to use one of the integer divisors of screen refresh-rate. The established way of achieving unconventional frame-rates is by repeating certain frames, but this method leads to artifacts like strobing and judder. With the recently proposed method of Templin et al [38], it is now possible to emulate any frame-rate below screen refresh-rate. Apart from being able to set any unconventional frame-rate for a scene, Templin et al. also propose a method to set spatially and temporally variable frame-rate. This gives filmmakers even finer control of manipulating video content according to aesthetic requirements. However, they do not explain how to use variable frame-rate to get desired look. There is no guide of which frame-rate to set. It is also not clear, how the look of the video would change by setting a certain frame-rate.

1.2 Problem

First problem we address is the effect of frame-rate on speed perception. We investigate the myth that speed perception is dependent on frame-rate. There is anecdotal evidence that HFR looks sped-up ([30] and [28]). Perhaps these complaints are because people have been used to 24 FPS, which they associate to “flim-look” and they wrongly ascribe lack of certain artifacts in HFR as faster motion. We investigate if there is a quantifiable difference in speed perception between low frame-rate and high frame-rate content.

The second problem is to provide filmmakers with a content dependent guide of how to use variable frame-rate technique proposed by Templin et al. We investigate if the amount of “film-look” could be measured. We also investigate the relationship between “film-look” and frame-rate. Perhaps based on that, filmmakers can manipulate the “film-look”, by employing appropriate frame-rate.

1.3 Proposed Solution

To investigate the effects of frame-rate on speed perception this thesis shows results of three perceptual studies. We present the first study in Section 3.2. In this study, we used Gabor patches, of different spatial frequency and contrast. Second study is discussed in Section 3.3, in which we use animated content as stimuli. In the last study, presented in Section 3.4, real-world videos are used to investigate the effects of frame-rate on speed perception. We show that there is no effect of frame-rate on speed perception.

People are used to watching movies at 24 FPS. We propose that “flim-look” is a certain blend of artifacts, such as flicker, strobing, and repeated edges, which occur at such low frame-rates. Flicker or judder is the most prominent difference between two frame-rates [38]. Flicker can be defined as non-smoothness of motion, due to temporal variations in pixels. Other artifacts, such as strobing and repeated edges, which are associated with low frame-rates, are correlated with flicker. By manipulating flicker, we will also, proportionally, manipulate other artifacts. This allows us to manipulate “flim-look”. In order to do this, we measure the relationship of flicker and frame-rate. A per-pixel content dependent perceived flicker estimation model is proposed in Section 4.2. Using this model, filmmakers can know the exact amount of flicker in every pixel. Based on this model, we also provide a method to control “film-look”. We propose a novel method for filmmakers to make flicker driven frame-rate manipulation, in Section 4.3. Using this method, filmmakers can control “film-look” by choosing appropriate frame-rate.

1.4 Contributions

In the first half of this thesis, we investigate the myth that speed perception is dependent on frame-rates. In the second half, we study the relationship between frame-rate and flicker, and propose a method to make flicker driven frame-rate edits. Our contributions are:

- three psychophysical studies, statistical analysis of their results, and discussions, to investigate the urban legend that frame-rate affects speed perception. We debunk the myth that HFR looks sped-up.
- a metric to estimate content and frame-rate dependent, per-photoreceptor flicker, for complex motion, that accounts for luminance contrast and motion magnitude.
- a new technique to manipulate frame-rate, such that the visibility of perceived flicker could be controlled.

1.5 Outline

Chapter 2 explains working of cameras and displays, and the artifacts which arise from spatio-temporal discretization. Chapter 3 explains perceptual studies performed to measure effect of frame-rates on speed perception. Chapter 4 presents flicker strength estimation model and also proposes a novel flicker driven guide to manipulate frame-rates. Chapter 5 discusses the conclusion of the thesis and potential future work.

Chapter 2

Background and Related Work

Videos can be viewed as spatio-temporal sampling of the real-world. These Saptio-temporal slices of the real-world are captured one frame at a time with a camera. Then, these frames are presented, using displays, as videos. Both of these processes introduce certain artifacts in perception of moving images. This section talks about capture and presentation processes, and artifacts that occur as a result.

2.1 Camera

Videos are recorded frame by frame, using a camera. There are two basic types of cameras: mechanical and digital camera. Some details of working of these two major types of cameras are discussed below.

Images are captured with a camera by letting the light pass through a lens and then exposing it to the sensor. In case of mechanical shutter cameras, there is a mechanical rotatory disc behind the lens which allows some light to pass at every exposure time. Then, this procedure is repeated for every captured frame. The amount of light that is let through depends on the angle of the shutter. For example, if the angle is 180° , it means that the shutter has exposed the sensor to light for half the time between two consecutive captured frames. Figure 2.1 demonstrates different exposure times. The yellow bars indicate the amount of time film has been exposed to light, between two frames.

In digital cameras, sensor type shutters are used. In this case, there is no mechanical shutter; the sensor works as shutter itself. It captures the coming light row by row. Both mechanical and sensor shutter are rolling shutters. So, the whole film or sensor is not exposed to light at the same time. In case of mechanical shutter, it takes some time for

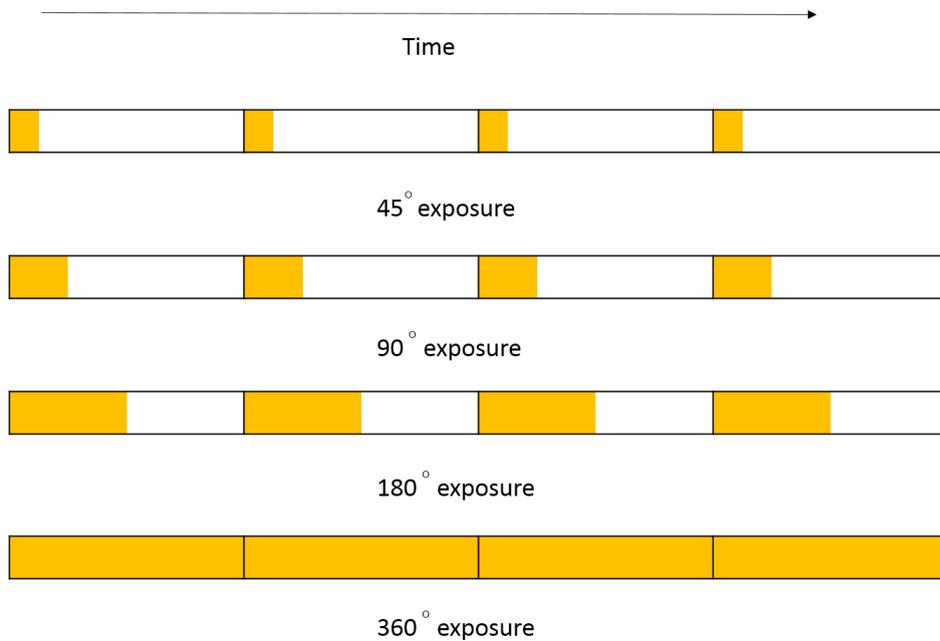


FIGURE 2.1: Amount of time sensor is exposed to light at different shutter angles.

the disc to rotate over the whole film. Similarly, in digital case pixels are exposed to light row by row. So, if a fast moving object is captured by digital camera, it could look slanted, since each position is captured at different times.

A hybrid shutter is an adapted mechanical shutter for digital cameras. It is used in order to compensate for slower sensor-type shutter. It could, potentially, help capture all the pixels at the same time instant.

2.2 Displays

Video are discrete samples (images) of space and time, shown using different types of displays. In this section, we explain two types of displays: cathode ray tube (CRT) display and liquid crystal displays (LCD).

CRT display is made of an electron gun in which electron beam is controlled by electro-magnetic deflection. The electron beam deflected by coils is thrown on a phosphor material, which, when hit by electrons, produces light. The luminance of phosphor material decays over time. The decay function is an intrinsic characteristic of the phosphor material used. Ideally, the luminance decay function of phosphor, after an exposure to the electron beam, should behave like an impulse function. However, it has a sharp peak near the beginning, followed by a shallow falloff. Phosphor persistence is the time for which phosphor stays illuminated, after being shot with electrons. The time of persistence is dependent on the type of phosphor and the presented stimulus. Shorter

persistence values of phosphors are desirable to avoid motion artifacts, such as motion blur. The electron gun shoots at the screen line by line, horizontally. The time between the end of one line and the start of the next line is called horizontal blank. Vertical blank is defined as the time between drawing the last line of a frame and the first line of the next frame. So, the time between vertical blanks is refresh rate of the screen. Color is created by using 3 types of phosphors, i.e. red, blue and green. Different electron guns are used to fire at these different colored phosphors. These electron guns are separated using a shadow map or aperture grill. The refresh rate of the screen depends on the speed of the electron gun, and the maximum resolution depends on the density of the phosphors. A linear voltage in CRT displays does not lead to a linear luminance function, hence a luminance scale has to be “gamma corrected” for correct luminance.

LCDs are made of cross light polarizers and liquid crystal sandwich. Back Light is constantly being shone on back polarizer. Light is then transmitted through the liquid crystals. Liquid crystals are padded with electrodes in the front and the back. By providing voltage through the electrodes, the orientation of the crystals can be changed. This allows to change the orientation of the light that reaches the second polarizer. After the second polarizer layer there are three color (red, blue and green) filters. All three of them combined, constitute a pixel. We can manipulate intensity and color of the pixel by manipulating the voltage provided to the liquid crystals. Each of the color filters, or sub-pixel, has its own electric field, which can be controlled separately. All the pixels in an LCD change their states simultaneously. Therefore, there is a concept of exposure time for a frame in LCD, which does not exist in CRT displays. Unlike CRT screens, LCDs also have a native display resolution, which is defined by the number of pixels.

2.3 Motion Artifacts

Presence of motion can degrade the image quality with artifacts. However, it can also enhance the quality of images in some cases, as shown by Didyk et al. [8]. Since the human visual system (HVS) is more attuned to lower temporal frequencies and higher spatial frequencies, it tries to stabilize the moving image by tracking through the process called Smooth Pursuit Eye Motion (SPEM). Through SPEM eye keeps the moving image on fovea, which is exclusively made of cones and is sensitive to high spatial frequencies. Thus, the moving image appears sharp. Laird et. al [22] found that SPEM is perfect between the speeds of 0.625-2.5 *deg/sec* and the mechanism works up to 7 *deg/sec*. SPEM also works for complex motion. It requires 100-120 ms [19] to start tracking a random target moving in a random direction. Limitations of SPEM occur at very low speeds, such as below 0.15 *deg/sec* when other eye motions, e.g. drift eye movements,

are dominant, and at very high motion, such as above 80 *deg/sec*, as found out by Daly et al. [7].

2.3.1 Motion Blur

If the motion is too fast or there are multiple objects with different motion profiles SPEM fails. This leads to motion blur. Motion blur also occurs in real-world, if the object is moving at a high speed. This artifact occurs at the capture time due to inability of the camera to capture a sharp image.

Current cinema standard is to use 180° shutter angle at 24 FPS, which means an exposure time of $\frac{1}{48}$ secs is used. Motion blur occurs due to longer exposure time while capturing a moving object. Each frame is individually blurred. Moving images with sharp edges look blurred, but it has been shown by Hammett et al. [15] that blurred images look sharper when they are moving. This phenomenon is called “motion sharpening”. Motion blur is also used as an artistic tool to create softer and more fluid feeling of a moving object. Some times longer exposure times might be necessary in low light conditions to avoid image noise, and that could lead to unwanted motion blur while capturing a fast moving object.

2.3.2 Hold-type blur

Hold-type blur could be considered as opposite to motion blur. When the object is moving too fast to be tracked properly, the HVS incurs motion blur. Whereas, hold-type blur is a purely perceptual artifact, which occurs due to displays. When an object is moving the HVS tries to track the object through SPEM. In hold-type displays, e.g CRTs, same frame is shown until display is updated with a new frame. However, eye compensates for the motion that it expects to see. This causes a mis-match in what the HVS expects and the information it receives, resulting in hold-type blur. Unlike motion blur, in hold-type blur individual frames could be sharp, but still produce this artifact. Hold-type blur can be reduced by showing the frame for a shorter time, or by inserting a black frame, but the later solution would result in strobing artifacts.

2.3.3 Repeated Edges

If fast moving objects are displayed at low refresh-rates, there could be artifacts like doubling of edges or ghosting. When a video is displayed with higher frame-rate, the sampling rate is higher. In such case differences between consecutive frames are smaller,

so such artifacts are reduced. However, HFR has its own drawbacks, such as making the movies feel cheap-looking and “soap-operatic” ([28] and [9]). Another way to tackle ghosting is to introduce motion blur, but that could also alter the artistic feel of the video. Filmmakers could also choose not to use the speeds that are too fast, and which might produce such artifacts. However, both of these later solution are more intrusive, and it would be ideal that such artifacts could be reduced without compromising the “film-look”.

2.3.4 Strobing

Strobing is a camera based artifact and it occurs if short exposure times are used while capturing a moving object. The images seem to be frozen in time and give a feeling of stuttered motion. If exposure time is too long then motion blur is incurred, and if too short exposure times are used video will have strobing artifacts.

2.3.5 Flickering

The human eye works as a time averaging sensor. If a light is flashed twice, for it to be distinguished as two flashes there has to be an appropriate time between the flashes. Bloch’s Law [12] states that total luminance energy is equal to the product of luminance and duration of the signal. However, temporal integration can only be modeled up to a critical point using Bloch’s law, beyond which luminance energy is only dependent on input luminance. According to Bloch’s law, the time required for temporal integration is shorter in photopic conditions compared to scopic conditions. The information gathered by the HVS is integrated into a coherent smooth motion. If the presented information is sparse then flickering is observed. Critical flicker frequency (CFF) could be seen as the frame rate which should be set such that temporally varying light pattern seems to be coherent [16]. CFF is the value beyond which the HVS ceases to see flickering.

Flicker sensitivity and consequently CFF is dependent on 2 factor: Temporal contrast and spatial extent. CFF is increased with an increase of temporal contrast. Previous work shows that flicker sensitivity is higher at sharp edges. Kelly [17] found that flicker sensitivity at sharp edges was 10 times higher than at edge-less field, at spatial frequency 2-5 cpd. Spatial extent is another factor that affects CFF, Makela et al [24] found that CFF dropped to 40 Hz when spatial extent of the flickering stimulus was smaller than 0.3° . CFF is also dependent on the luminance of the display. According to Ferry Porter law CFF is directly proportional to logarithm of luminance of the stimulus.

People associate 24 FPS with “film-look”. Perhaps a certain mixture of the aforementioned artifacts, which are produced at low frame-rates such as 24 FPS, creates the illusive “film-look”. Most critics of HFR say that HFR movies look cheap, soap-operatic or perhaps too real. From the artifacts discussed above, flicker is the most distinguishing artifacts among different frame-rates [38]. By manipulating flicker we will proportionally change other artifacts. Hence, by manipulating flicker we would be able to manipulate “film-look”.

Chapter 3

Effects of Frame Rate on Speed Perception

There is a rising trend of HFR technology in films. Some studies, such as the one performed by Wilcox et al. [45], show that HFR is preferred. However, not everyone has a positive opinion about HFR. There is anecdotal evidence that HFR looks sped-up ([30] and [28]). In this section of the thesis, we investigate the myth that perceived speed is dependent on frame-rate.

There are several proposed motion perception models. Three famous motion models were proposed in 1980s: motion sensor model by Watson et al. [43], motion energy model by Adelson et al. [1], and elaborate Reichardt detector [41]. Although all of these models might be differently motivated, they are mathematically equivalent. All of these models follow the so called Motion From Fourier Components (MFFC) principle. This principle is based on converting the motion signal into sine wave patterns and specifying the location in frequency space, we are able to find out the parameters of motion [29].

Apart from the above mentioned motion models, there are also two recent speed perception models: a ratio model by Hammett et al. [13] and Bayesian Model by Stocker et al. [35]. Hammett et al. conducted perceptual experiment on adaptation effects of motion. They found that although the more established effect of an exponential decrease of perceived speed, after prolonged viewing of stimuli, existed under certain circumstances, perceived speed can also increase after adaptation. Based on this observation, they proposed a model which comprises of two temporally tuned mechanisms whose sensitivities reduce over time. Perceived speed is taken as ratio of these filters' output. So, the perceived speed is computed with comparison of few temporal channels. Stocker et al. [34] proposed Bayesian model of perceived speed. They modeled biases in speed perception through a Bayesian estimator. They proposed that estimation bias is determined by the

likelihood function, which fits the given data, and shape of the prior. They found that the prior function of the HVS assigns higher probability to lower speeds than higher speeds. Stimuli with high contrast yield a likelihood function with a sharp peak. To get the perceived speed, i.e. the posterior probability, the likelihood function should be multiplied with the prior, which they found to be biased towards the slow speed. For high contrast stimuli the likelihood functions have a sharp peak, so multiplication with the prior makes the peaks shift a little bit. Due to this reason high contrast stimuli yield a relatively better estimation of the speed. On the other hand, the likelihood functions for low contrast stimuli are more noisy and broader. So, the prior affects the posterior probabilities more, which causes an underestimation of speed perception.

For investigating the effects of frame-rate on speed perception, we looked into different attributes that affect speed perception. An overview of these attributes is given in Section 3.1. In Sections 3.2, 3.3 and 3.4, we present three psychophysical experiments and their results. In Section 3.5, we discuss the results of all three experiments in broader perspective. In the last Section 3.6 we discuss some relevant implementational details.

3.1 Factors affecting speed perception

Speed perception depends on several factors. In this section we discuss the effect of spatial frequency, contrast, luminance and adaption on speed perception.

One of the factors that affects speed perception is spatial frequency of the stimulus. Spatial frequency is defined as number of periods of a pattern contained in a visual angle and is measured in cycles per degree (cpd). Smith et al. [33] found that a stimulus with higher spatial frequency seems to move faster. In their experiments, they compared a test sinusoid grating of 1 cpd to reference gratings of 0.5, 1, 2 and 4 cpd at various speeds, in a two alternating forced choice (2AFC) setting. They found that for lower speeds, around 3 *deg/sec* or less, the matches for all spatial frequencies were correct. However, for higher spatial frequencies and higher speeds of the reference, perceived speeds of the test were underestimated. The greatest underestimation was observed at 4 cpd. It should be noted that the authors only tested till 4 cpd. Perhaps, for higher frequencies the effect would be greater. The authors also found that perceived speed is not dependent on direction of the motion.

Contrast is another factor that affects speed perception. Stone et al. [36] reported that lower contrast gratings were perceived to be moving slower in a forced choice experiment. They tested effects of speed perception for multiple contrast values in the

range from 2.5% to 50 %. They also performed the experiment for 70 % contrast value, and surprisingly this effect did not saturate. They found that stimuli with 70% contrast had to be slowed down 35% to match the stimuli with 10% contrast, which were both moving at 2 *deg/sec*. They also found that the absolute contrast had no effect on speed perception and the effect is a quasi-linear function of log contrast ratio. It was also found that effect of contrast on speed perception was sensitive to orientation. When orthogonal test and reference gratings were used, the errors in speed matches almost disappeared. In case of sequential presentation, as opposed to simultaneous presentation, the effect of contrast on speed perception was found to be weaker. Thompson et al [40], in a relatively recent work, confirmed the previous findings of underestimation of perceived speed. Additionally, they found that speed of test stimuli was overestimated up to 35%, when the reference and test stimuli with contrasts, 70% and 10%, respectively, and spatial frequency of 2 cpd, were tested at speeds 6 and 8 *deg/sec*. This shows that the effects of speed-up and slow down are symmetric and are dependent on the velocities. These later results are well explained by the ratio model, discussed before, but not very well explained with the Bayesian model of speed.

Hammet et al. [14] showed that low luminance at high speed leads to an overestimation of perceived speed. So, at low luminance objects appear to move faster. They used test luminances of 1.5 *cd/m²* and 0.13 *cd/m²* relative to reference luminances of 30 *cd/m²* and 2.54 *cd/m²*, respectively, at speeds of 1,2,3,4,8 and 16 *deg/sec*. 2/3 of the subjects mismatched the speed for 8 and 16 *deg/sec* and highest effect was noted at 16 *deg/sec*.

George et al. [11], in their work on orientation dependence of apparent speed perception, showed that if the visual elements are aligned with motion path then they appear to move faster. This effect was dominant at high speeds such as 40, 64 and 96 *deg/sec*. Maximum over-estimation was noted at 64 *deg/sec*. There was no effect for low speeds such as 4 *deg/sec*. Perceived speed decreased as the angle between motion axis and Gabor patches increased. When compared with non-oriented elements, speed of collinear sequences was overestimated and parallel sequences was underestimated.

Many perceptual studies, such as [26], [32] and research on speed and direction dissociation in hitting action [4], indicate that speed and direction are estimated independently. Hence, in this thesis we only handle speed instead of velocity. Similarly, acceleration perception of the HVS, both in terms of acuity [44] and accuracy of eye movement [42] is very poor.

Adaptation to the the factors discussed above can change our speed perception. Blackmore et al. [2] found that adaptation to high contrast reduces the perceived contrast of the subsequent stimuli. So, in our experiments we present the high contrast

stimuli for very short time. Similarly, Thompson et al. [39], in their experiments on adaptation effects on velocity perception, showed that adaptation to high velocity makes the low test velocities seem slower. The mismatch of the perceived velocity was found to be as high as 50%. Interestingly, they did not find any reliable results for overestimation of speeds. If the test velocity is greater than the adaptation velocity the effects of adaptation are diminished. In a relatively new findings by Hammett et al. [13], it was discovered that for low adaptation speeds (2 and 4 *deg/sec*) and a higher test speed ($> 8\text{deg/sec}$), perceived speed increases exponentially. This is similar to the decrease of perceived speed when adaptation speeds are high. To cater to all the adaption effects, we present the stimuli in random order such that the results are not affected by adaptation.

3.2 Experiment I: HFR Effects on Speed Perception using Gabor Patches

This section explains a perceptual study designed to determine the effect of frame-rate on speed perception. We used psychophysical test stimuli, Gabor patches, shown in figures 3.1a and 3.1b . We used such stimuli because we can manipulate other factor affecting speed perception in Gabor patches. This allowed us to isolate the effect of frame-rate in speed perception. We wanted to, later, be able to decompose complex videos in terms of spatial frequency, contrast, and speed, and thus, predict the perception of speed biases due to frame-rates.



FIGURE 3.1: Examples of gabor patches used in experiment

3.2.1 Setup

As discussed in section 3.1, there are several variables that affect speed perception. We consider three of the most important variables: contrast, spatial frequency and speed.

Three sample points from each of these attributes were considered. For contrast 25.5%, 53% and 75.5% Michaelson's contrast was considered. Spatial frequency was sampled at 0.5, 2 and 4 cpd. Speeds of 2, 8 and 16 *deg/sec* were considered. This makes a total of 27 points.

The luminance of the screen was measured with Minolta LS-100 luminance-meter. Due to low variability of luminance of the screen at the top, only the top portion was used. The experiment comprised of Gabor patches moving horizontally, from left to right. Reference and test sequences were temporally separated, inter stimulus interval (ISI) of 0.5 seconds was used. The stimuli were temporally separated because, as mentioned in section 3.1, speed perception is affected by adaptation effect of contrast, if stimuli are simultaneously presented. The mean luminance of the screen was 60 cd/m^2 . Reference was played for 0.75 ± 0.1 seconds, while the test sequence was played for 1.5 seconds. The difference and variability of presentation time of reference was motivated by the intention of discouraging participants to use length as a measurement of speed. Every sequence was shown for at least one visual degree, so that participants have a chance to gauge the speed.

The resolution of the screen was 1920 x 1200. The distance between the screen and participant was roughly 65 cm. No chin-rest was used, so the participants could move their heads freely. The experiment took place in a controlled office environment.

Ten naive subjects took part in the experiment. Participants were first shown the reference sequence and then the test sequence. They were asked to adjust the speed of the test sequence until it matched the reference. In 27 experiment cases frame-rates of test and reference were 60 FPS and 30 FPS, respectively. There was no limit on the number of replays. The refresh-rate of the screen was set at 120 Hz. There were also 27 control cases in which the frame-rate of both reference and test cases was 30 FPS. The control cases were put so that effect of other factors could be ruled out.

3.2.2 Results and Discussion

The results shown in the figures 3.2, 3.3 and 3.4, show graphs among contrast, spatial frequency and perceived speed. The blue surface shown in all the figures is the result of control sequences, i.e. where the frame-rates of reference and test were same. The red surface in the figures depicts the results of interpolated surface from the points where the frame-rate of reference and test were different. The obvious trend in the results is that the experiment cases are on top of the control cases. This means that a higher speed of higher frame-rates is matched to a lower speed of lower frame-rate; so speed appears to be perceived slower at higher frame-rates.

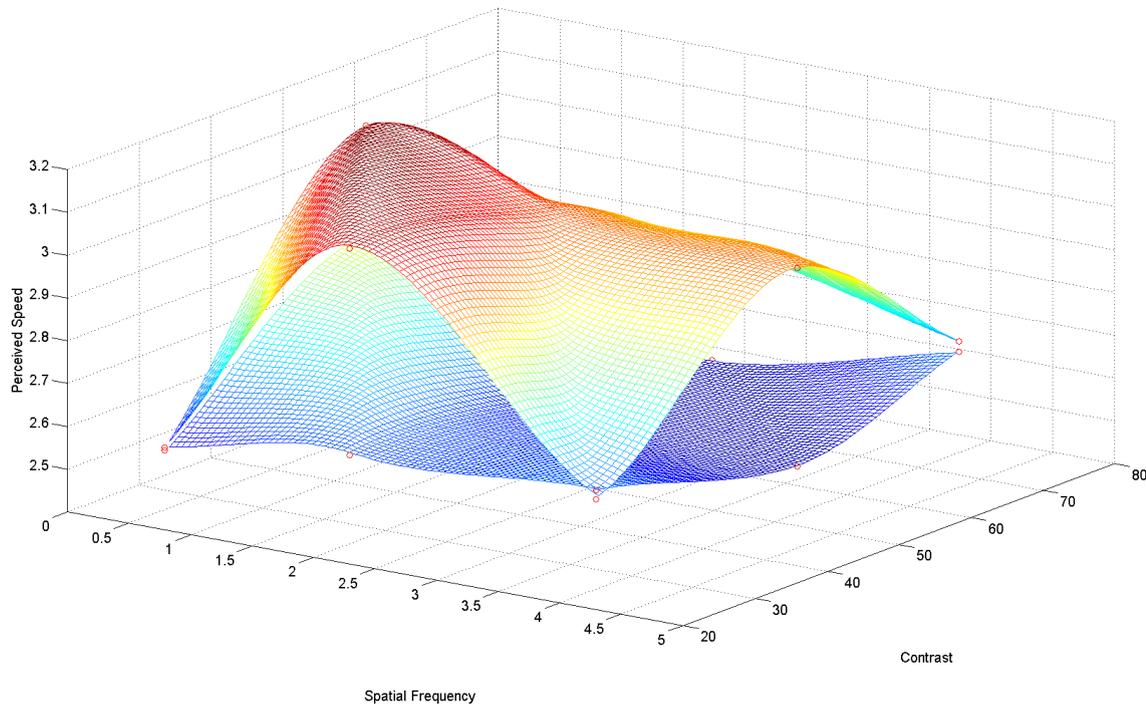


FIGURE 3.2: The top and bottom surfaces are interpolated from the point obtained from experiment and control cases, respectively. The speed of the stimuli was fixed at 2° per second.

The figures 3.5, 3.6 and 3.7 show interpolated surface from point differences between speed matches of test and control cases. The results show that at speeds 2° and 16° , the interpolated surface is hyperbola, and for 8° it is hyperbolic paraboloid shape.

More results for this experiment can be found in the appendix.

3.3 Experiment II: HFR Effects on Speed Perception using Animated Content

In previous experiment, we used simple stimuli to check if the frame-rate affects speed perception. In this section we investigate the effects of frame-rate on complex videos. For this purpose we designed another psychophysical experiment, whose details are given in the following section.

3.3.1 Setup

This experiment consisted of 5 sequence from big buck bunny video [20]. Screen-shots of the sequences used are shown in figures 3.8,3.9,3.10,3.11 and 3.12.

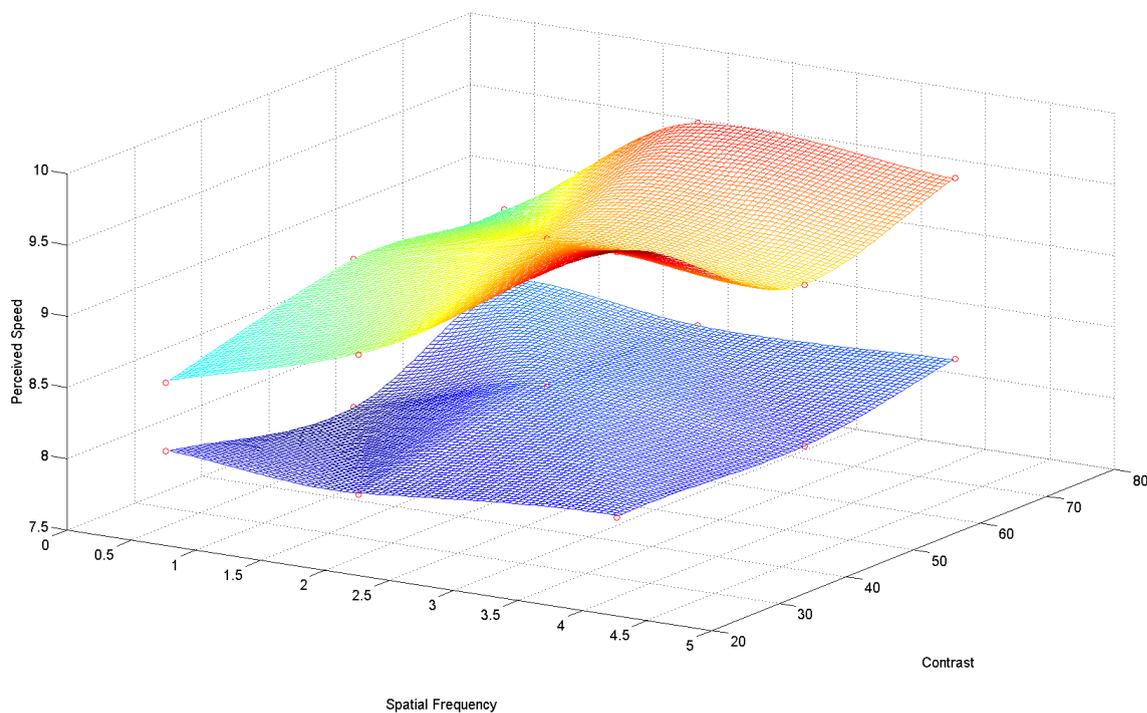


FIGURE 3.3: The top and bottom surfaces are interpolated from the point obtained from experiment and control cases, respectively. The speed of the stimuli was fixed at 8° per second.

The length of the sequences at 60 FPS was 56, 196, 113, 178 and 172 frames. Full HD (1920 x 1080) at 60 FPS version of the movie was used to calculate optical flow, using the method proposed by Brox et al. [5]. Resolution of the display was 1920 x 1200. The stimuli were presented at the center of the screen with grey background. A GPU implementation of interpolation function from OpenCV library, [3], was used to compute real-time interpolation at 60 FPS. Since the interpolation was not fast enough for such high frame-rate, grayscale version of the video was used at half of full HD resolution. The GUI for the experiment was made using the library freeglut.

14 naive subjects took part in the experiment. They were given following instructions to read before starting the experiment.

“The experiment consists of a series of 20 trials. In each trial your task is to match the speed of two videos. The video on the left-hand side is the reference video and its speed does not change throughout the trial. The video on the right-hand side is the test video, the speed of which you are supposed to adjust so that it is the same as the speed of the reference video.

Initially, both videos are stopped. Pressing the left or right arrow keys will play back the corresponding video. To increase or decrease the speed of the

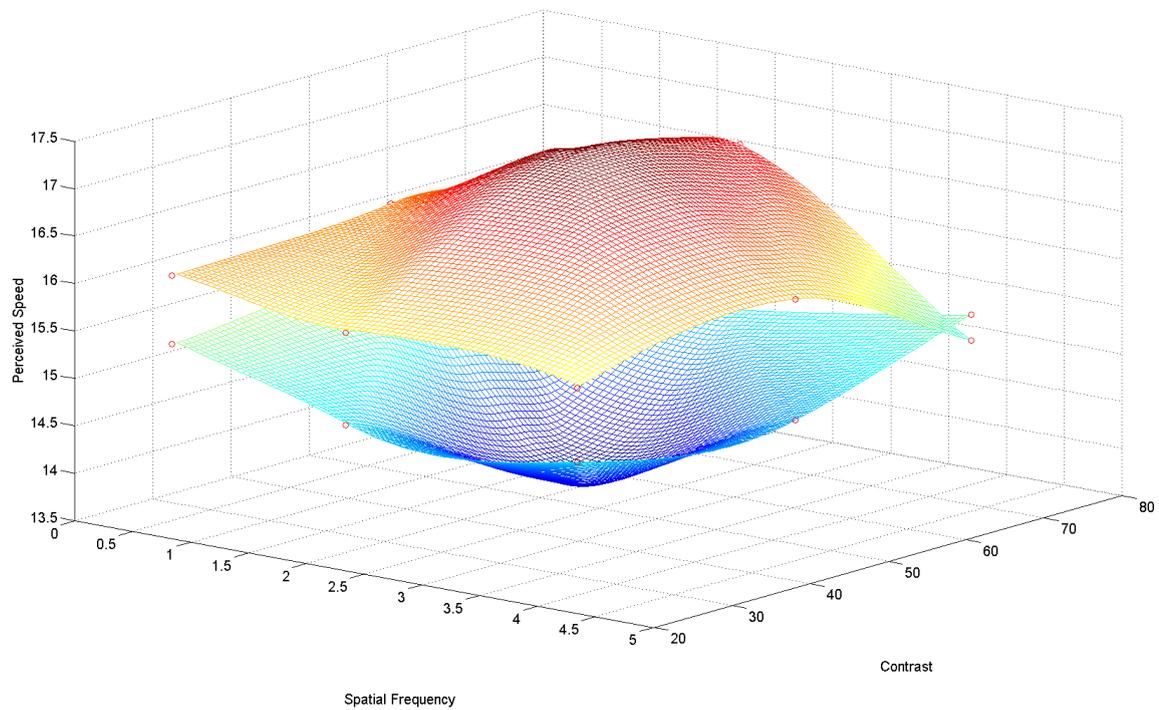


FIGURE 3.4: The top and bottom surfaces are interpolated from the point obtained from experiment and control cases, respectively. The speed of the stimuli was fixed at 16° per second.

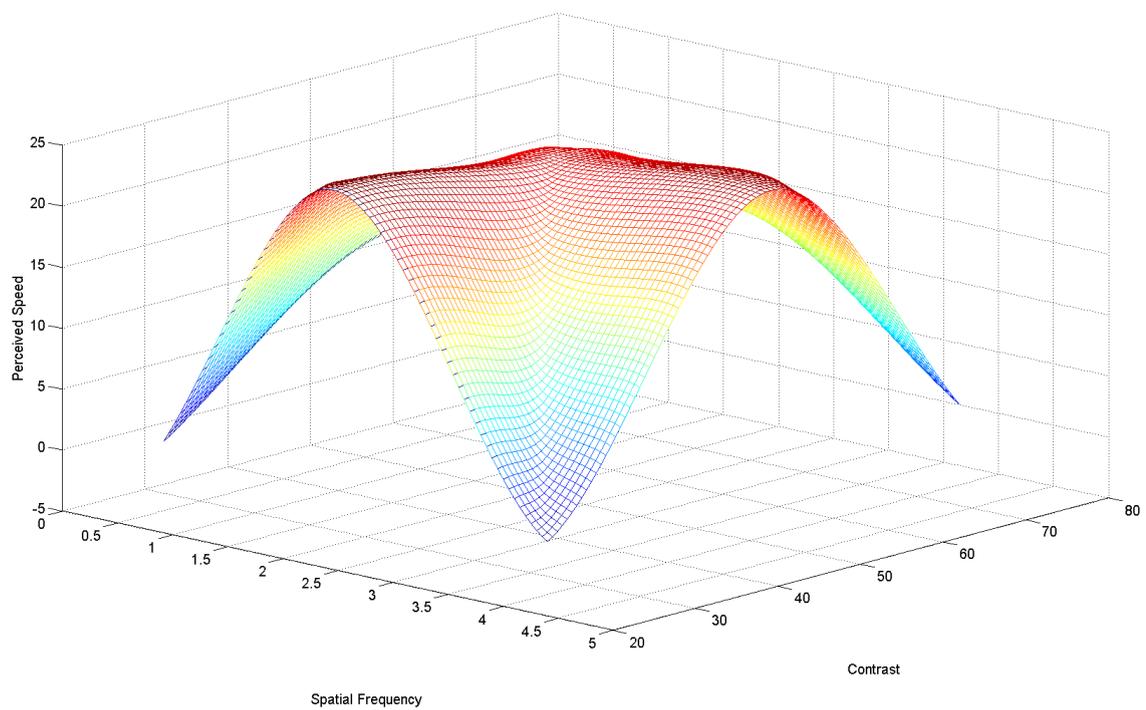


FIGURE 3.5: The results of difference between speed matches in experiment and control cases, for speed 2° per second.

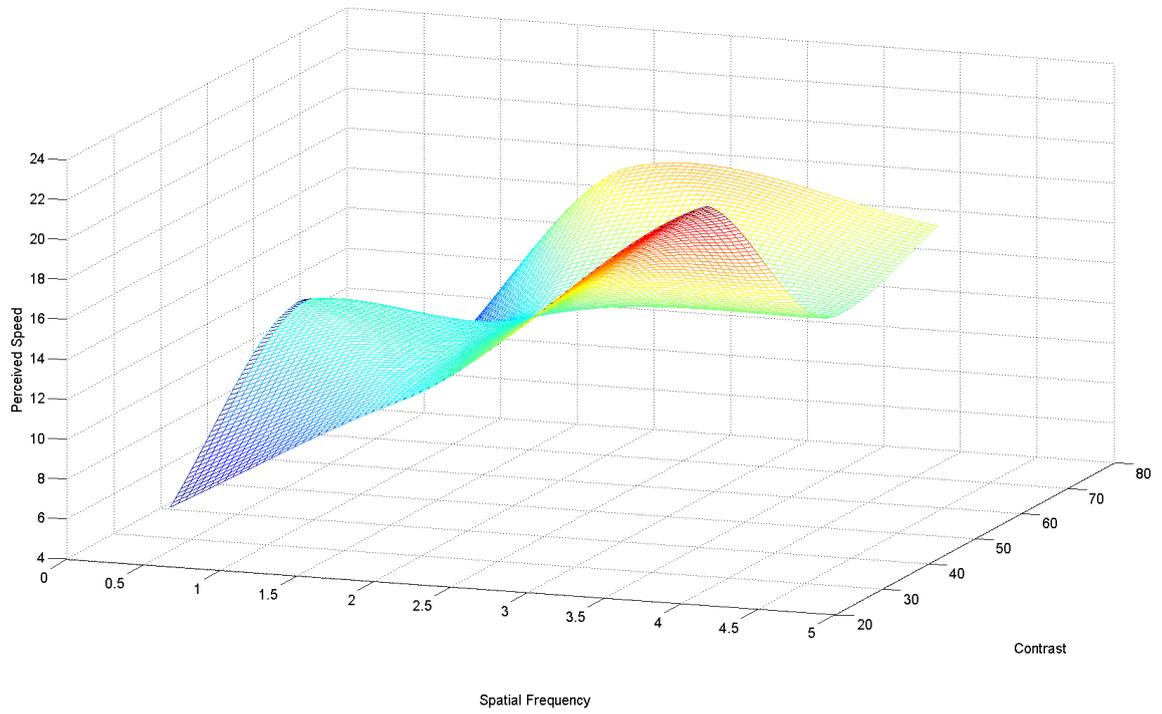


FIGURE 3.6: The results of difference between speed matches in experiment and control cases, for speed 8° per second.

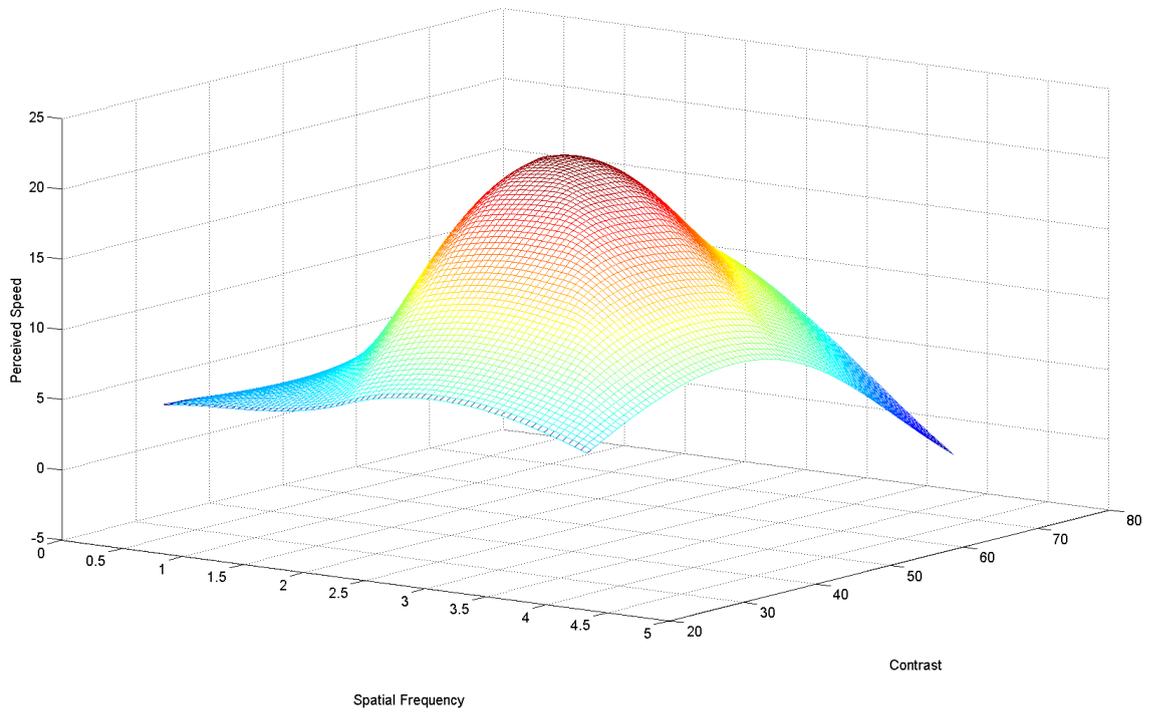


FIGURE 3.7: The results of difference between speed matches in experiment and control cases, for speed 16° per second.



FIGURE 3.8: Bunny scene.



FIGURE 3.9: Bird scene.

test video you press the plus or the minus key, correspondingly. There is no limit on the number of playbacks nor on the number of adjustments you can make, however, when one of the videos is currently playing you have to wait until it stops to play the other one or to make further adjustments to the speed. You can take as much time as you need to finish each trial. When you have matched the speed of the test video to that of the reference video, press the spacebar to proceed to the next trial.”

Each comparison was done twice by the participants. 10 trials in the experiment were control, where the frame-rate of both reference and test sequence were 30 fps. In other ten trials, frame-rate of reference was 30 while test was running at 60 FPS. For both



FIGURE 3.10: Butterfly scene



FIGURE 3.11: Skipping scene.

cases, the refresh-rate of the screen was 60 Hz. Both reference and test sequence were interpolated. So, if there were any artifacts caused by interpolation they were present in both reference and test. This was done so that participants were not biased because of artifacts. All the sequences were randomized to exclude any perceptual adaptation biases, such as discussed in section 3.1. The test videos were initialized randomly in the range of (0.8 – 1.2) times the regular speed. The new speed on every press of plus or minus buttons was calculated as follows:

$$New-Speed = Old-Speed \pm \exp(0.02 * Old-Speed) \quad (3.1)$$

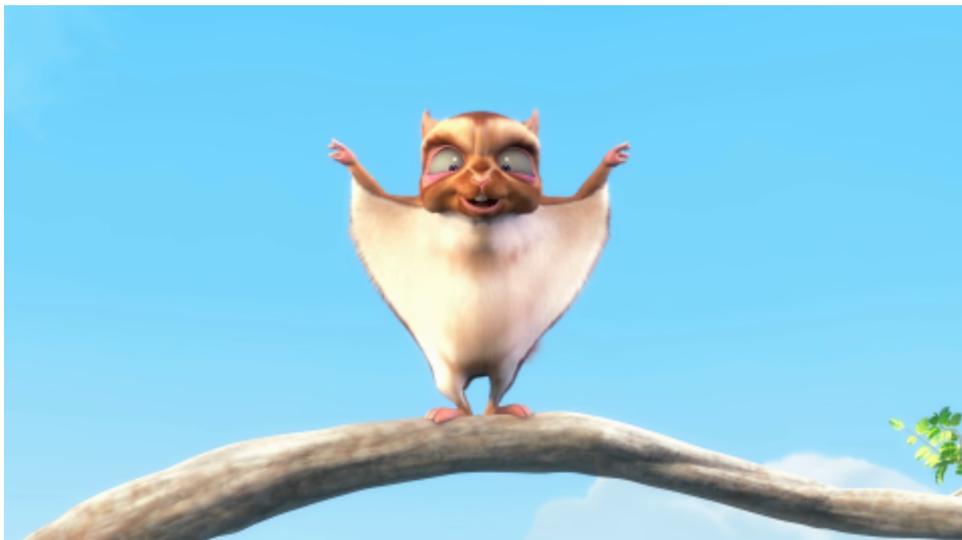


FIGURE 3.12: Squirrel scene.

3.3.2 Results and Discussion

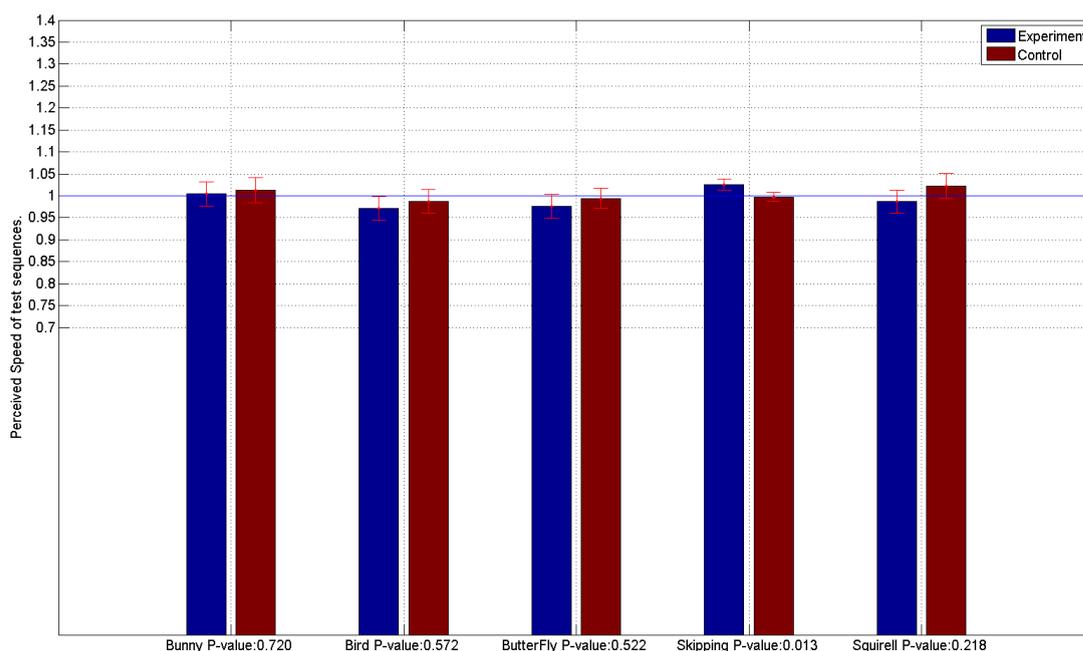


FIGURE 3.13: X-axis shows names of the scenes, and Y-axis shows the matched speed of the test sequences. Each bar represents average of all the participants' responses. The results are also averaged over both the comparisons they were asked to make for every case.

The bar graphs in figure 3.13 show blue bars for the trials where the frame-rates of the test and reference were 60 FPS and 30 FPS, respectively. The red bars are for control cases in which frame-rate of both the videos was 30 FPS. The figure shows

results for 5 scenes averaged over 14 participants. The participants performed every comparison twice. The results are also average of both the adjustments. The plots show the perceived speeds of test videos compared to the reference. The reference was always playing at the regular speed. The red error bars are standard error of mean (SEM). At the bottom of each scene p-value is mentioned, which is calculated using paired t-test.

There are a couple of interesting points to notice in this figure. For most of the scenes subjects perceived HFR faster than low frame-rates, however the p-values are not very low. This result is in agreement to the urban myth that HFR looks sped-up. In skipping scene, perceived speed, for HFR, was seen slower. One possible explanation could be that subjects associated temporal aliasing to faster motions. The speed of the object for skipping scene was fastest, which caused a lot visible temporal aliasing in the scene. Due to denser sampling in HFR, the motion became crisper, which was perceived as slower speed. All in all, the effect of frame-rate does not seem to be big or significant in most cases.

3.4 Experiment III: HFR Effects on Speed Perception Using Real-World Videos

Experiment 3.3 was performed on animated content, and in both the previous experiments subjects were asked to adjust the speed of test sequences until it matched the reference sequence. For this experiment, we recorded real-world sequences. In previous experiments a short interview was conducted from the subjects about the level of difficulty and they were asked if they found the experiments too straining. Most participants said that adjusting the speed was difficult and decisions for closer to veridical speeds were harder. Hence, in this experiment we used 2AFC setting and asked the participants to choose the sequence which looked faster.

3.4.1 Setup

The experiment comprised of two color sequences. Screen-shots of both the scenes are shown in figures 3.14 and 3.15. In the scene 3.14 a person, who was close to the camera, was moving across the screen. In the second scene, 3.15, a car, which was farther away from the camera, was moving across the screen. Both the scenes contained high angular velocity. In case of the “person” scene, the walking speed was low. However, since the person was closer to the camera, the angular velocity on the retina was still high. One of the reasons to select the car scene was that it contained high spatial frequencies and fast motion. This combination gave favorable results in experiment 3.2. The motivation



FIGURE 3.14: Person scene. © Junaid Ali



FIGURE 3.15: Car scene. © Junaid Ali

of choosing scenes with fast moving objects was, that it would create more artifacts at low frame-rate, such as those explained in section 2. These artifacts would be alleviated through HFR. As we mentioned earlier, perhaps people associate “film-look” to a certain blend of artifacts created by low frame-rate, which they also associate to fast motion. The myth that HFR affects speed perception could just be because of non-existence of those artifacts. By selecting such scenes, we wanted to test this hypothesis. Especially, in the “person” scene temporal aliasing was quite visible at lower frame-rate, and the

differences between 30 FPS and 60 FPS were more pronounced.

The experiment consisted of two videos shown side-by-side, horizontally. One of the videos was reference and other was test. Each of the them could be played by pressing left or right arrow keys. All the controls were disabled, when one of the videos were being played. The person scene originally had 137 frames and the car scene had 129 frames. 0-3 frames were dropped randomly at every play from the beginning and the end, so that participants did not match the length of the sequence. 7 speed-ups of test sequence were picked i.e $\pm 1\%$, $\pm 2\%$, $\pm 4\%$ and 0% . 0% was added to check if there is an effect of frame-rate on speed perception. Other speed-ups were considered to measure the amount of any speed perception difference. The difference in speed perception for different frame-rates did not seem very apparent. Our hypothesis was that the speed-up effect that people claimed to see in HFR was due to lack of artifacts. Since, the effect was not very apparent, we hypothesized that it might be below JND. So, we wanted to operate below the discrimination threshold to capture an effect which might be subliminally there. McKee, [27], found that JND for speed discrimination is 5% . We planned to perform experiments with higher speed-ups if needed. We considered the following pairings of reference and test: (30 FPS, 60 FPS), (60 FPS, 60 FPS), (30 FPS, 30 FPS), and (60 FPS, 30 FPS). Both (30 FPS, 60 FPS) and (60 FPS, 30 FPS), reference and test pairs, were considered because, then, we would not change the speed of only one frame-rate compared to the other frame-rate. This was done to get rid of any potential perceptual biases, and to make the measurements more robust. Cases of (30 FPS, 30 FPS) and (60 FPS, 60 FPS) were added as a control. In total there were $4 * 7 * 2 = 56$ trials.

14 videos, 7 for for each scene, were pre-generated at 60 FPS and 30 FPS. 30 FPS videos were generated by dropping the alternate frames. The experiment was two alternating forced choice (2AFC). 23 subjects were asked to choose the faster of the two presented videos. One of the videos was reference video and the other one was test video. Reference video always had 0% speed-up. The test video could have one of the speed-ups stated above. Positions of reference and test were also randomly chosen, from left or right. This was done to avoid biased results, in-case participants preferred left or right.

3.4.2 Results and Discussion

Figure 3.16 shows results aggregated over 28 participants, for 2 videos over 7 speeds. We asked the participants to choose the faster of two presented sequences. If test sequence was chosen to be faster we assigned it value 1 and if reference was chosen to be faster

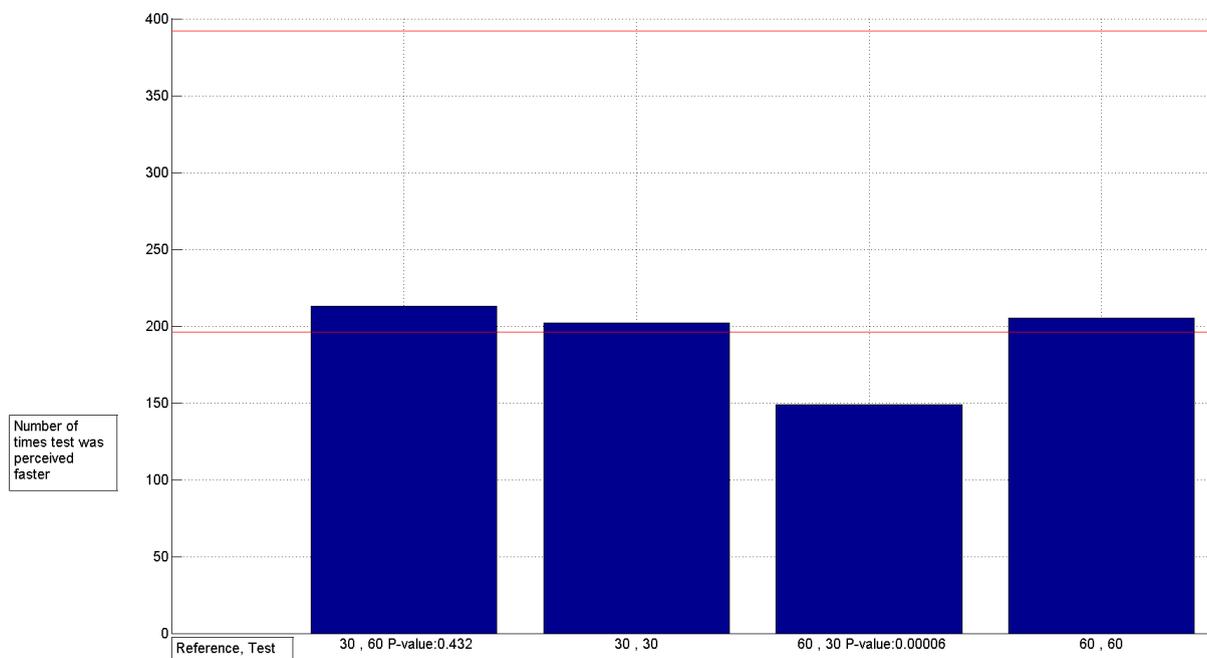


FIGURE 3.16: X-axis shows type of the experiment and Y-axis shows number of times test sequence was perceived faster, compared to reference. Each bar represents sum of the times all participants perceived test sequence, at every speed-up and both videos, to be faster, for a particular type of experiment. Red lines mark 50 % and 100 % of the readings.

we assigned it value 0. The figure shows summed-up results of these experiments. For example, all the results of reference and test pair, which were of type of (30 FPS , 60 FPS), are stacked in one vector and then summed up. Total length of the this vector is (number of participants * number of video * number of speed-ups) $28 * 2 * 7 = 392$. This is done for all four cases. If there were an effect of speed-up, we would see significantly lower or higher values than 50%. The difference between the first bar and third bar, in the figure, is only that test and reference positions are switched. X-axis shows the type of the experiment. Y-axis shows the number of times people chose test to be faster. The experiment was 2-AFC and subjects were asked to pick faster of the two presented videos. This produces a binary and ordinal data; ordinality 0 means test cases were slow and 1 means that test cases were fast. If we consider speed of test cases, as the variable, then these binary numbers give the data ordered categories of slow or fast. Hence, p-value were computed using Mann-Whitney U test.

As seen in the figure the results are consistent with the results shown in experiment 3.3 i.e. 60 FPS is perceived faster than 30 FPS. This effect is more prominent in the third bar graph. The trend, of seeing 60 FPS faster, also exists in the first bar graph. However, the effect is not as strong as it is in the third graph. So, the results do not

seem symmetric. This leads one to believe that perhaps the effect of frame-rate, on speed perception, is not very robust.

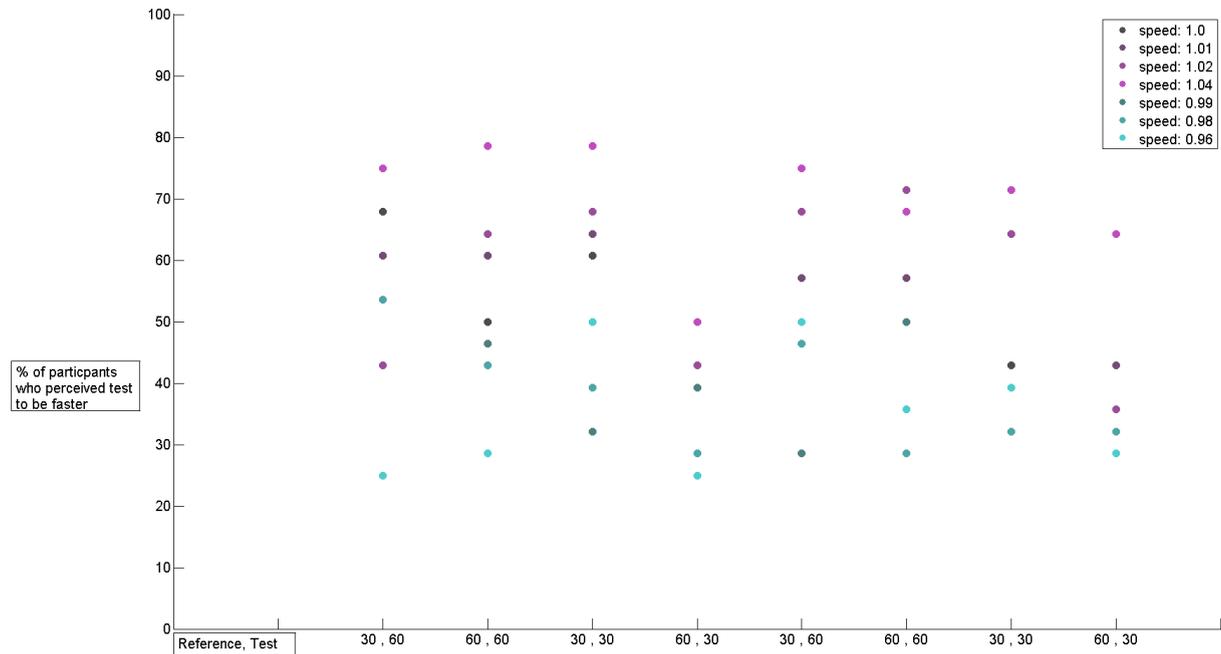


FIGURE 3.17: Each dot represents percentage of participants who perceived test to be faster, for that specific type of experiment. First four cases are for the “person” scene and the next four cases are for “car”

Figure 3.17 shows results of different types of experiments, which are given in the x-axis. The percentage of the test cases chosen as faster, out of all experiments of that type, are given on y-axis. The first four results are for the “person” video and the next four are for the “car” video. The figure shows that at least in one of the cases 60 FPS appears faster than 30 FPS. Surprisingly, the results are not repeated in other cases. Hence, it could not be definitively concluded that the frame-rate affects speed perception.

3.5 Discussion

We investigated the urban legend, [30], that HFR looks sped-up, but we found mixed results. Some people saw HFR as faster and some people saw it slower. One of the reasons could be that maybe the HVS is not very precise at speed discriminations and that most of the results are accidental. One conclusion that could be drawn for certain is that there was not enough over-whelming evidence of the effect of frame-rate. Even if there is any effect, the size is quite small. By looking at a complex scene, in two different

frame-rates, the perceived speed difference was not very obvious. People are used to 24 FPS videos and HFR is a new technology, in which motion seems smoother as shown by Wilcox et al. [45]. We also confirmed this finding from the accounts of participants of our studies. HFR also alleviates some of the artifacts discussed in Section 2. Perhaps, the lack of artifacts, which takes away from the “film-look”, is hard to express, and some people attribute it to difference in speed. However, when explicitly asked to adjust the speeds of videos, in two different frame-rates, they fail to see any difference.

Another interesting point to notice is the difference between the direction of the effect in simple stimuli, such as Gabor patches, and complex videos. In experiment 3.2, people perceived low frame-rates as faster. On the other hand, in complex videos, in most cases, people saw HFR a bit faster. Perhaps, in videos the HVS has more high level content dependent mechanism involved, which are not very easy to break down into basic stimuli, such as used in 3.2. Maybe a more detailed study of biological procedures, involved in speed perception estimation, is required to be able to break down perception of complex scenes into building blocks or basis function of speed estimation.

3.6 Implementation details

To change the speed of the videos, at the same frame rate, we need to calculate interpolated frames. If we want to adjust the speed during the experiments, as we did in experiment described in Section 3.3, we need to interpolate at the frame-rate of the input video. Since we were using content at 60 FPS, we needed to compute new frames in less than 16.67 *milliseconds*. For getting smooth interpolation, we, also, needed to calculate accurate and coherent optical flow. Some of the details of optical flow and interpolation used in this thesis are given in the following sections.

3.6.1 Optical Flow

To perform similar experiments as described in section 3.2, on complex videos, we need to be able to alter the speed of the sequence, while keeping the frame-rate fixed. In order to do that we need optical flow between consecutive images in the sequence. State of the art optical flows use variational energy functionals. To calculate optical flow of the videos in our experiment, an off-the-shelf method of Brox et al. [5] was used. Some of the details of their methods are relevant to set the variables to get good results. Hence, in this section we explain some of the details of their method.

There are two components of an energy functional for calculating optical flow: fidelity term or data term and smoothness term. Data term matches the features between

images. Brox et al. use grey value and gradient constancy assumption i.e. the grey value and the gradient of a tracked pixel remains constant in the consecutive images.

The data term looks as follows:

$$E_{Data}(u, v) = \int_{\omega} |I(X) - I(X + U)|^2 + \gamma |\nabla I(X) - \nabla I(X + U)|^2, \quad (3.2)$$

Where I is the image, $X = (x, y, t)$ i.e. pixel (x, y) at time t , ω is whole image domain and U is flow vector at a pixel i.e $U = (u, v, 1)$.

Data term only does pixel wise search. If the gradient of the functional vanishes, then only normal flow is available. This is called aperture problem. In this case, calculation of optical flow at those locations fails. To cater to this problem, Brox et al. employ a spatio-temporal smoothness term, which works as hole-filling. The smoothness term is given as follows:

$$E_{Smoothness}(u, v) = \int_{\omega} \Psi(|\nabla_3 u|^2 + |\nabla_3 v|^2), \quad (3.3)$$

The smoothness term helps optimize for a piecewise smooth optical flow function. The final energy is sum of equations 3.2 and 3.3. To find the optical flow we have to minimize the energy for unknowns u and v .

A GPU implementation of this method can be found in the library OpenCV, [3]. Optical flow calculated for all the sequences used in the previous experiments is calculated using that implementation. Smoothness term is set to 14.0, weight of gradient importance or gamma is set to 75, lagged non-linearity is set to 30, warping iterations are set to 200 and the linear solver around it is set to 100.

3.6.2 Interpolation

Interpolation is performed using “interpolateFrames” method of the library OpenCv [3]. The interpolation method is based on pixel reprojection technique introduced by Mark et al. [25]. The method takes two frames, forward and backwards optical flow, *step-size* of interpolation and optionally occlusion maps. Forward optical flow has the displacement vectors for each pixel from frame one to frame two, and vice versa in case of backward optical flow. The interpolation is performed by warping first frame with amount of the *step-size* using forward optical flow, and by warping the second frame with $1 - (\textit{step-size})$. Both warped images are blended together for robust results and the final interpolated results are returned. In this way, a lot of occlusions are avoided. However, there might still be some occlusions. In order to deal with them the interpolation function also takes occlusion maps. Occlusion maps are the marked pixels which are occluded from frame one to frame two. Based on the *step-size* and the

optical flow between two frames, the occluded region is identified, and it is ignored in the computation. Due to motion between two frames, there also might be dis-occlusion. This area is filled up using in-painting.

Chapter 4

Manipulation of Flicker using Variable Frame-rates

Recent developments in cinematography led to several feature films being shot and presented using high frame-rates which provide smoother motion and less temporal artifacts such as strobing and jittering. Despite many benefits, the productions were widely criticized for their “cheap” and “soap-operatic” look. To address this problem, it is possible to use a so-called varying frame-rate which, in contrast to standard 24 and 48 frames-per-second choices, equips an artist with a much finer control over frame-rate and the resulting appearance. However, the specific choice regarding the frame-rate is made using trial and errors. In this work, we propose a semi-automatic technique that enables frame-rate edits which account for perceptual implications of the changes. Since flicker is the most prominent difference between two frame-rates, we propose that it is also a major contributor to the “film-look”. We explore the relationship between frame-rate and flicker, and manipulate the frame-rate according to the desired flicker visibility. To this end, we present a computational model that calculates per-pixel strength of visible temporal variation in the input video. Perceived flicker, in a video sequence, is dependent on the amount of temporal contrast, and speed and size of the moving objects. Due to this reason, we need content dependent flicker model. We present a model which takes video sequence as an input and generates content dependent per-pixel flicker map. Using our flicker estimation model, we propose a guide to use variable frame-rate technology. We propose an optimization procedure that computes spatio-temporal varying frame-rate map resulting in the desired flicker, which we use to set local frame-rates, using technique introduced by Templin et al. [38]. Flicker is an intuitive attribute and our tool helps filmmakers make flicker driven local frame-rate manipulations.

The rest of the chapter is arranged as follows: in Section 4.1, we discuss background and previous related work. In Section 4.2, we present flicker model. In Section 4.3, we present to methods to use our flicker model in order to make frame-rate manipulations. Lastly in Section 4.3.3, we show the results.

4.1 Previous Work and Background

Low frame-rates produce several artifacts, such as motion blur, repeated edges, strobing and flickering. These have been discussed in Section 2. For our work in this chapter, most important of these artifacts is flicker. In Section 4.1.1, we present background and related work on flicker perception. In Section 4.1.2, we present relevant work regarding frame-rate. Then, we discuss influence of frame-rate on flicker perception, in Section 4.1.3.

4.1.1 Flicker Perception

Flickering occurs due to temporal variation in a pixel. Flicker is perceived if the frequency of these temporal variations is not high enough. Beyond a certain frequency, flicker is not visible. This frequency is called critical flicker frequency (CFF).

CFF is dependent on two factors, spatial extent and contrast.

4.1.1.1 Spatial Extent

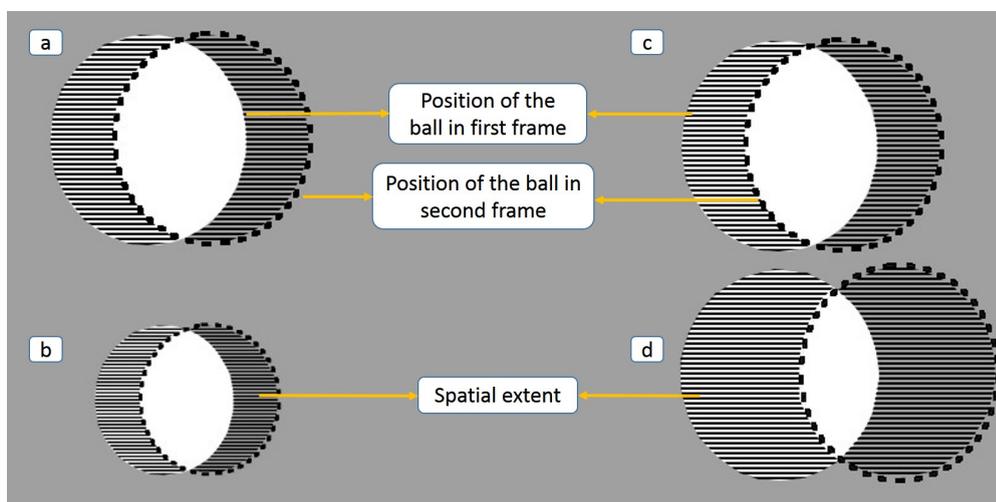


FIGURE 4.1: The spatial extent changes based on size and speed of the moving object.

Spatial extent is the flickering region, which is shown in line-patterned region in Figure 4.1. Spatial extent is dependent on two factors: the absolute size of the moving object and its speed. Figure 4.1 shows different scenarios involving both the attributes. It shows a white ball moving on a gray background. Since the color of the ball and background is same in all cases, the only difference between flicker perception is based on spatial extent. The spatial extent of the ball in Fig. 4.1a is bigger compared to the ball in Fig. 4.1b because of the bigger radius. The balls in cases “c” and “d” are of same the radii, but since the ball in Fig. 4.1d is moving faster it has a bigger spatial extent. Increasing the absolute size or speed of an object can potentially increase the number of pixels with non-zero temporal contrast, which increases perceived flicker. Spatial extent is capped by the size of the object.

Flicker sensitivity at a pixel is directly proportional to the amount of the flicker in the neighborhood. Makela et al. [24] performed experiments to check the relationship of flicker sensitivity to spatial extent at different frequencies and different eccentricities. They used sinusoidally flickering uniform circular spots. They tested 1, 3, 10 and 30 Hz temporal frequencies and eccentricities of 0, 5, 10 and 20 degrees. They used 2AFC experimental setting, and participants were asked to choose one out of two exposures which contained the stimulus. For every wrong choice the contrast of the stimulus was increased by $0.1 \log_{10}$. The results showed that as the spatial extent increases, flickering sensitivity also increases. At lower temporal frequencies, sensitivity is higher in the fovea, however at higher frequencies and higher spatial extents sensitivities are higher at higher eccentricities. After a certain increase in spatial extent, the sensitivities start to plateau. As the temporal frequencies increase, the point where sensitivities start to plateau comes at larger spatial extents.

4.1.1.2 Temporal and Spatial Contrast

CFF increases with increase of temporal contrast. Previous work shows that flicker sensitivity is higher at sharp edges. Kelly [17] found that flicker sensitivity at sharp edges was 10 times higher than at edge less field at spatial frequency 2-5 cpd.

To calculate visibility of flicker of a pixel, we need to perform Fourier analysis, considering the varying values of the pixel as 1D signal, over time. Kelly [18] presented a simple equation for flicker visibility,

$$E(f) = ae^{bf}, \quad (4.1)$$

Where f is the fundamental frequency of flickering pixel, while a and b are determined by the amplitude of the fundamental frequency. If the observed energy at a pixel, according to equation 4.1 is more than predetermined threshold, flicker would be visible at that pixel. Since we are working with fixed frame rate the fundamental frequency of all the pixels would be the frame rate. So, flicker is only determined by the absolute amplitude of the fundamental frequency. Farrell et al. [10] used this equation to predict the thresholds for video display terminals.

Larimer et al. [23] performed an experiment showing that even if spatial contrast is zero, flicker is visible based on temporal variations. The stimuli used in their experiment were bars moving from left to right. One set was white bars moving against black background. In second set 75% of pixels of the bar were black and 25% were white. In this set the ratio of black and white pixels in the background was reversed. In the last set all pixels on the screen were assigned black or white color randomly, which means that there was no distinguishing spatial contrast at the edges of the bars. They asked the participants to adjust the speed of the bars such that flicker just becomes noticeable. The thresholds for last category were highest. The results showed that flicker was still visible at zero spatial contrast, as long as there was temporal contrast.

When content with strong spatial contrast moves, it generates strong temporal contrast. The relationship between spatial and temporal frequency is given by following formula:

$$f_t = f_s v, \quad (4.2)$$

If a sinusoidal pattern with spatial frequency f_s is moving with the velocity v , it would produce the same flickering as stationary pattern flickering with frequency f_t . In a scene displayed with refresh-rate beyond CFF, if there is no motion, there will be no visible flicker because of the lack of temporal contrast. If an object with uniform color moves across the scene, it will only, potentially, generate flicker around the edges. This will only happen when it generates non-zero temporal contrast. So, flicker perception is based on temporal contrast.

4.1.2 Frame-rate

There has been a rising trend to use HFR technology in films. Films such as “The Hobbit” and “Billy Lynn’s Long Halftime Walk” were filmed and presented at HFR. The review of HFR are conflicting.

Kuroki [21] investigated the relationship between frame-rates and depth perception. He used moving random-dot stereograms at different speeds and at frame-rates

60, 120 and 240. Six subjects were asked if the test pattern appeared to be “nearer, in the same position, or farther compared with the reference pattern in depth”. He found that depth discrimination was better at 240 FPS compared to 120 and 60 FPS. Kuroki also reported that over-all perception of 240 FPS clips were “smooth motion and natural depth impression”.

Wilcox et al. [45] performed psychophysical experiments to investigate which frame-rate is preferred by the participants. They used 3 recorded S3D video clips. Comparison was done for different combinations of frame-rates (24, 48, 60), shutter angles (180°, 270°, 358°) and exposure time (24FPS-180°, 48FPS-180° and 48FPS-358°). In each trial participants were shown a pair of clips. They were asked to rate which clip they preferred based on the attributes: Realism, Motion smoothness, Blur/clarity, Quality of depth and over-all preference. For each of these attributes they could provide rating on a five-point scale from clip 1 to neutral to clip 2. The results showed that viewers preferred higher frame-rates compared to lower frame-rates in all measures. This preference was more pronounced when 24 FPS was compared with either 48 or 60 FPS. The preference differences between 48 FPS and 60 FPS were only statistically significant in the scene which was more prone to flicker.

The attributes used by Wilcox et al. [45] might have a bias among them. If participants rate one of the clips higher on the attribute “Realism”, it is likely that they might also be inclined to rate it high on over-all preference. These attributes do not seem to be independent, so drawing any conclusion from the results of individual attributes does not seem justified.

There have also been some negatives reviews about HFR. There is anecdotal evidence ([30] ; [28]) that HFR seems sped-up. However, there has been no published studies to this effect. Reviewing the movie “Billy Lynn’s Long Halftime Walk”, Engobar [9] concluded that even though the movie was crisp and very clear, it failed to portray the “film-look”. Some people complain that HFR looks like soap-opera, comparing it to TV which is generally shown at a higher frame-rate than films.

Michelle et al. [28] did two extensive studies on the viewers of the first two parts of “The Hobbit” trilogy. The authors identified two major viewing modes of the participants: transparent mode and mediated mode. In transparent mode viewers have more immersive experience and have an emotional affect or they identify with certain character. Meditated mode is more objective, in which viewers “focus on the constructed nature of the text as an aesthetic object and media production”. Viewers in transparent mode were more forgiving of artifacts that arose due to cinematic advancements. The mode of viewer-ship could change during the course of a film. The type of viewer-ship

varies from person to person, it is also content dependent and could sometimes be deliberately changed. Data also showed that if viewers had more experience with stereoscopic 3D (S3D), they were more likely to focus on drawbacks of HFR and 3D. They asked the participants about their preference of HFR. They found that there might not be one simple answer. Some people might prefer HFR, while others might dislike it.

The literature hints that preference of HFR comes down to personal preference, however, benefits of using HFR are clear.

Templin et al. [38] proposed a technique to emulate temporally and spatially variable frame-rates on a fixed refresh-rate screen. We can view a video as uniformly sampled signal in time. Templin et al. proposed that instead of using uniform kernels for sampling we displace alternating kernels in time. They presented psychophysical experiments to map the displacement of sampling kernels to perceived frame-rate. Their technique of frame-rate emulation (between 24-96 FPS) fared better, when tested on real-world content, than reference which was shown on 24, 48 or 96 FPS. The effect of shutter angle was not very significant on lower speed and higher frame-rates.

4.1.3 Flicker and Frame-Rate

It is not very well understood how variable frame-rate should be employed. Since the most apparent difference between two frame-rates is flicker, we investigate its behavior at different frame-rates. The relationship between flicker and frame-rate is not very well studied. Daly et al.[6] performed a psychophysical experiment, in which they investigated judder or flicker. They used gabor patches and complex images, as stimuli. The participants were presented with two options and were asked to pick the one with most judder, in 2IFC (interval forced choice) setting. They presented the normalized raw results. The results of their experiments relating frame-rate and judder are shown in Figure 4.6. The figure shows that as frame-rate increases, judder reduces linearly. Xu et al. [47] filed a patent which explains a method to control judder visibility. However, they do not measure judder as a function of frame-rate, they, rather implicitly, assume this connection based on findings of Daly et al.[6].

These results, however useful, do not tell us how much flicker is produced on the retina, given a scene.

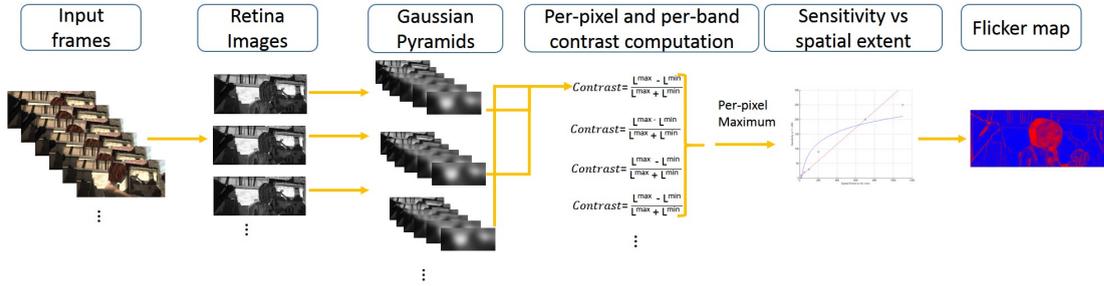


FIGURE 4.2: Pipeline for calculating Flicker maps. We start with input frames, then calculate retinal images, build a Gaussian pyramid for every retinal image and calculate temporal contrast at every band of the Gaussian Pyramid. We take the maximum value of contrast across all bands of a pyramid, then we convert the values into flicker levels by multiplying contrasts with corresponding sensitivity values. We take maximum value of flicker across all pyramids to produce the final Flicker map.

4.2 Flicker Model

Didyk et al. [8], presented a flicker visibility model. However, they did not measure the amount of flicker. They, also, only consider simple motions. We present a model to calculate content dependent per-pixel flicker on the retina for complex motion, given an input video. Flicker is produced at a pixel due to temporal variations. These variations occur due to motion in the scene. We assume that all the motion in the scene is being tracked, through a process called SPEM. If every moving pixel is being tracked, we have to compensate for the motion to calculate flicker. Due to this reason, we generate projection of presented frames on retina. Then, we compute temporal variations or temporal contrast on every photo-receptor. Flicker sensitivity is dependent on the area or spatial extent of the flickering region. To account for the dependence of flicker sensitivity on spatial extent, we use Gaussian pyramids. We generate a Gaussian pyramid for every retinal image. For every band of generated Gaussian pyramids, we calculate temporal contrast and convert it into multiple of 1 JND, using results from Makela et al. [24]. At every band, the corresponding kernel size models spatial extent. For a given photo-receptor, we take maximum flicker produced at all the bands.

We follow the pipeline explained in the figure 4.2. We assume that we are given input video and goal flicker maps. In Section 4.2.1 we discuss how we calculate retinal images. Lastly, in Section 4.2.2 we explain how we model spatial extent in flicker calculations. As a result, we get a map of perceived flicker.

4.2.1 Retinal Images

We assume that all the motion in the given video is being tracked through a process called smooth pursuit eye motion (SPEM). Through this process, eye keeps the tracked pixel in the fovea. Fovea consist of only cones cells and it is most sensitive region of the retina, to high spatial frequencies. To keep tracked pixel in the fovea, eye has to move with the same speed as the moving object. We call the projection of the presented frames on the retina, retinal images.

We are working with discrete images instead of continuous signal as in real-world. Our eye keeps integrating information even between consecutive presented video frames. The eye takes samples between consecutive frames, expecting the tracked pixel to move, but finds old information available. If, at that location, the pixel value differs from our tracked pixel, it generates a non-zero temporal contrast. This causes flickering sensation. The perceived flicker depends on the size of the flickering region, value of the temporal contrast and eccentricity of the projected position on the retina.

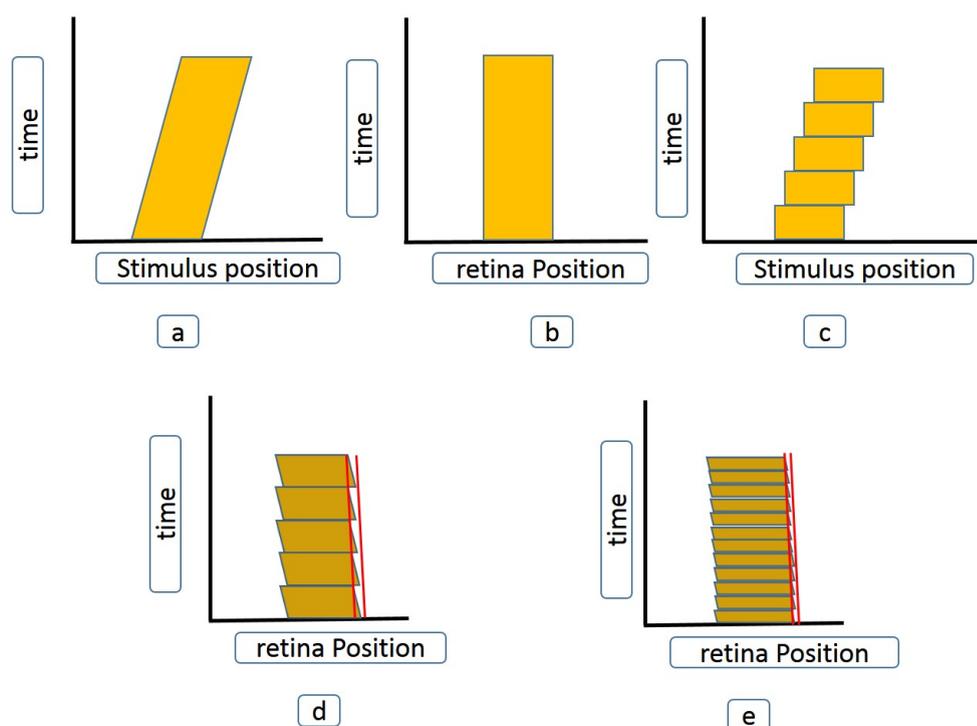


FIGURE 4.3: Demonstration of scan-line of a bar moving from left to right. It shows discrete and continuous cases, as well as position of the scan-line as it moves and its projection on the retina. With the increase of frame-rate, the spatial extent of the flickering region shortens. The red lines show the spatial extent of the flickering region.

Figure 4.3 shows scan-line of a bar moving from left to right. Fig. 4.3.a shows the position of the stimulus, if the samples were infinite. Fig. 4.3.c demonstrates position of

the stimulus on the screen when samples are discrete. If our eyes are tracking the bar, we would like to keep the position of the bar fixed in the fovea and keep the speed of the tracked object on the retina, zero. This is demonstrated in Fig. 4.3.b for continuous case and 4.3.d and 4.3.e for discrete cases.

In discrete cases, scan-line of the moving bar would be jagged. Each block represents a scan-line and the time duration for which the frame stays on the screen. Due to SPEM, our eyes try to compensate for the motion of the bar. However, since we use discrete samples and hold-type displays, the position of the bar does not stabilize on the retina, as shown in Fig. 4.3.d. It shows that due to SPEM, eye keeps tracking even between two frames. However, the information available is from the previous frame. Therefore, position of the bar does not stay fixed on the retina, as it does in the real-world. Due to this phenomenon flicker is perceived around the edges. One way to lower the flicker is to increase the number of samples as shown in Fig. 4.3.e. As the number of the samples increase the spatial extent of flickering region decreases and flicker frequency increases, making the flicker less visible.

Given an input video, we consider a temporal window of W frames starting from the reference frame. In this window, we generate retinal images. The number of retinal images, generated between two frames, are determined by the maximum optical flow in this window, let's call it M . Every frame in the window is first mapped to the reference frame. This is achieved through warping the frames with accumulated optical flow. Then we generate M retinal images, for every frame in this window, through warping with optical flows given as follows:

$$flow_j^1 = \frac{-j}{M} * opticalflow_k : j \in \{1...M\}, k \in \{1...W\} , \quad (4.3)$$

$opticalflow_k$ is the optical flow of k^{th} frame in the window, which gives correspondences between frame k and $k + 1$. Total number of retinal images produced in a window are $W \times M$. This way motion between two consecutive retinal images is not more than 1 pixel. Eye takes continuous samples of presented content. We generate such high number of retinal images in order to emulate working of the eye as closely as possible. Dis-occlusion in the retinal images are filled using inpainting. We use the inpainting method proposed by Telea et al. [37]. We take the implementation provided in the library OpenCv [3]. The algorithm is based on fast marching method. This ensures that pixels are filled in order of their closeness to known pixels. In this way information is propagated from known regions to the unknown region.

To get symmetric flickering region around edges we also generate retinal images, by first mapping every frame to the first frame in the window and then warping with optical flows given as follows:

$$flow_j^2 = \frac{j}{F} * opticalflow_k : j \in \{1...M\}, k \in \{1...W\} , \quad (4.4)$$

By warping the presented frames, in the way explained above, we align the images on the retina. This enables us to calculate temporal contrast on every photo-receptor.

Motion blur is a purely perceptual phenomenon. If the speed of a moving object is too fast to be tracked properly, they seem blurry. To account for motion blur, we perform temporal smoothing. In our experiments we mostly used 60 FPS content. The kernel size for 60 FPS videos was set to 1.5 frames and it was appropriately adjusted for producing retinal images at lower frame-rates. This is further discussed in Section 4.3.

4.2.2 Multi-scale Contrast Processing

As discussed in the section 4.1.1.1, the size of the flickering region influences amount of flicker perceived at a pixel. To incorporate spatial extent, we generate Gaussian pyramid for every retinal image produced through the process explained in previous section. First, we calculate temporal contrast on every photo-receptor, at each level of the Gaussian pyramid, in the temporal window of W frames. We use Michaelson contrast, which is given by,

$$C_{i,j}^g = \frac{(L_{i,j}^{max})^g - (L_{i,j}^{min})^g}{(L_{i,j}^{max})^g + (L_{i,j}^{min})^g} , \quad (4.5)$$

$C_{i,j}^g$ represents contrast at location (i, j) , and at the pyramid level g . $(L_{i,j}^{max})^g$ and $(L_{i,j}^{min})^g$ are maximum and minimum luminance values, respectively, at every photo-receptor at pyramid level g . $(L_{i,j}^{max})^g$ and $(L_{i,j}^{min})^g$ are calculated using following formulae.

$$(L_{i,j}^{max})^g = \max_{k=1}^n (L_{i,j}^g)^k , \quad (4.6)$$

$$(L_{i,j}^{min})^g = \min_{k=1}^n (L_{i,j}^g)^k , \quad (4.7)$$

In equations 4.6 and 4.7 we take maximum and minimum luminance, respectively, at location (i, j) of all the retinal images produced in window W at a fixed pyramid g , which are given by:

$$n = W \times M , , \quad (4.8)$$

We calculate temporal contrast for retinal images produced by equations 4.3 and 4.4, separately. Then, we take the maximum value at every location.

Kernel size of the Gaussian at every pyramid level corresponds to spatial extent. We convert temporal contrast values, calculated on all the levels, into multiple of flicker JND (just noticeable difference) by multiplying it with corresponding sensitivities given by Equation 4.11. Then, we take the maximum flicker value for every retina location, on all the pyramid levels.

$$flicker_{i,j} = \max_{g=1}^q (C_{i,j}^g * sensitivity^g) \quad (4.9)$$

q is the number of pyramid levels, which are determined by the size of the frames. Maximum function is non-differentiable and non-smooth. Since we are using warping to generate retinal images, they have some anomalies. In order to avoid getting outlier maximum values, we use soft maximum. Soft maximum is an approximation of maximum function, it gives smoother and differentiable maximum. We use the following approximation of *max* function, used in equation 4.9.

$$soft-maximum = \frac{1}{N} \ln \left(\sum_{x \in R} \exp(Nx) : x = C_{i,j}^g * sensitivity^g \right) \quad (4.10)$$

R is set of flicker maps on the all Gaussian pyramid levels. If N is large, equation 4.10 approximates maximum function. We used $N = 4$ in our implementation.

In this work we use results from Maekla et al. [24]. Since we want the most conservative estimate, we take the concave envelope of all the reading, at different eccentricities. The fitted function is shown in Figure 4.4.

The blue function shows fitting using log function and is given by,

$$Sensitivity = 52 * \log(spatial\ extent) - 163, \quad (4.11)$$

4.3 Framerate Manipulations

Usually displays allow fixed refresh-rates. This limits us to either use the same or any lower divisor of that refresh-rate, as frame-rate. Using method proposed by Templin et. al [38], one can set any frame-rate, lower than the refresh-rate of the screen.

A video sequence is uniform spatio-temporal sampling of real-world. The temporal sampling-rate is the frame-rate of the video. Templin et. al proposed that by sampling non-uniformly, in time, we can emulate appearances of different frame-rate. Figure 4.5 shows uniform temporal sampling kernels on the left. Templin et al. proposed that by displacing the sampling kernel, such as shown in the Figure 4.5, one can emulate

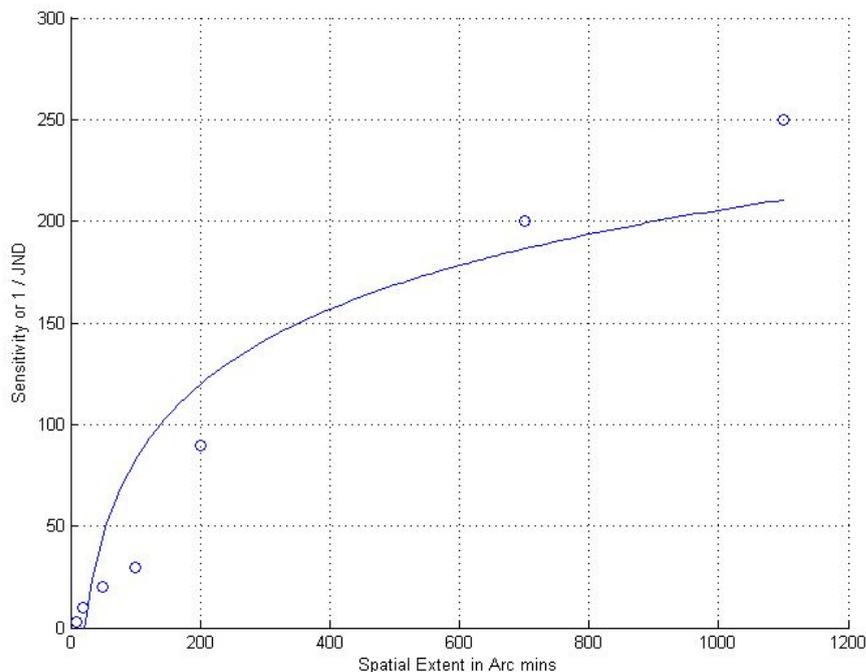


FIGURE 4.4: The readings from Makela et al. and its logarithmic fit.

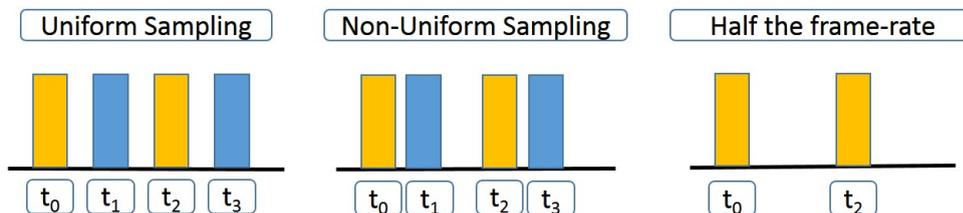


FIGURE 4.5: Demonstration of variable frame-rate emulation method of Templin et al. Left figure shows uniform temporal sampling of video. Middle figure shows displacement of kernels. In the limit, kernels at t_1 and t_3 would merge into t_0 and t_2 , respectively. This would result in half of input frame-rate, as shown on the right of the figure.

frame-rates between input frame-rate and its half. Their experiments show that the displacement of the kernels is proportional to the inverse of perceived frame-rate. This process can be repeated again to emulate arbitrarily low frame-rate. Therefore, we can set different displacements for any pixel, if we know its optical flow, and emulate spatially and temporally variable frame-rate.

However, it is unclear which frame-rate to use to get the illusive “film-look”. People associate “film-look” to 24 FPS. We propose that “film-look” is a blend of artifacts, such as those discussed in Section 2, which occur at low frame-rates. Templin et al. [38] reported that perceived intensity of flicker was the most important criterion while matching stimuli shown in veridical frame-rates to the ones modified by their algorithm.

Flickering is an artifact produced due to low temporal sampling rate. Flicker is the most prominent difference between two frame-rates. By manipulating frame-rate in order to get desired flicker in the scene, we would also increase other artifacts proportionally. In this way, we manipulate “film-look” by making flicker driven frame-rate edits.

In the following sections we present two ways to locally manipulate frame-rate using our flicker model. First method describes how we can introduce “film-look” by increasing flicker in the scene. Second method takes goal flicker maps, and produces frame-rate maps, which are used to set appropriate frame-rate to achieve goal flicker.

4.3.1 Introducing Film-look

Filmmakers are forced to pick higher frame-rate, due to fast motion in the film. However, movies usually have a wide range of motion. As discussed earlier, using higher frame-rates could give rise to unwanted “soap-operatic” look. This could be due to lack of artifacts, in slow moving objects or objects with low temporal contrast variations. Dally et al. [6] performed a study which collected subjective scores of flicker visibility at different frame-rates, using complex videos and Gabor patches as stimuli. The results showed that flicker increases linearly as the frame-rate is lowered, as shown in figure 4.6. Using these results, we can introduce “film-look” in the scene. To this end, first

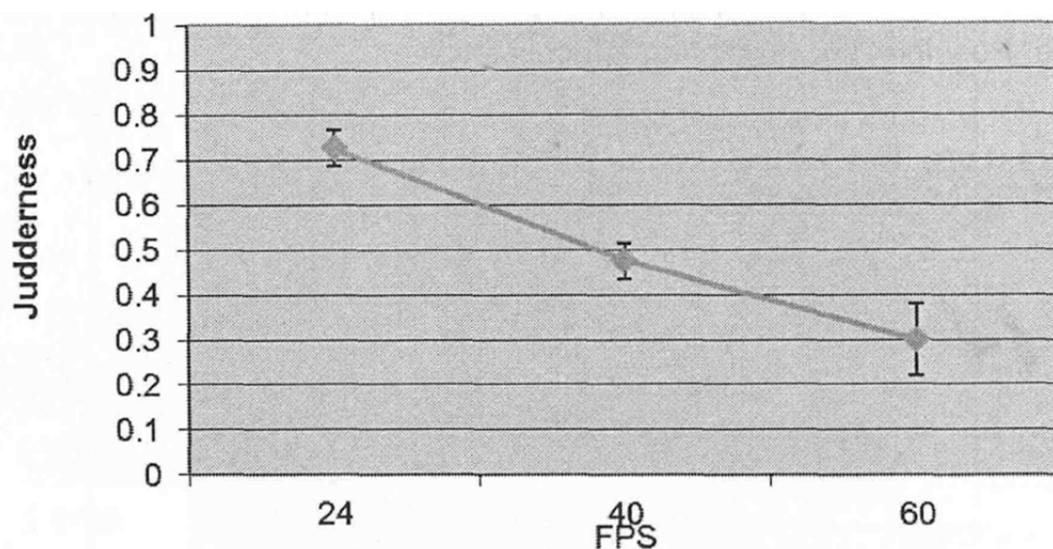


FIGURE 4.6: Results from Daly et al. of effect of frame-rate on judder perception. The graph shows that as the frame-rat increases perceived judder decreases linearly.

we measure flicker strength with the model explained in Section 4.2. The model gives us flicker-maps, with amount of per-pixel flicker. Then, assuming that input frame-rate was used to get rid of annoying flicker, we can introduce “film-look”, by setting highest frame-rate at the pixels with highest flicker, and half of the input frame-rate at the

pixels which have the lowest flicker. The flicker values in between are linearly mapped to the frame-rate values. In this way, areas with most flicker are shown at input frame-rate, and the areas with least flicker are shown with half of the input frame-rate. This introduces more flicker in the scene.

4.3.2 Framerate Adjustments using Target Flicker

The method proposed in the previous section is not very flexible. It does not give the filmmakers the ability to make local changes. In this section, we propose a method through which we can optimize for per-pixel frame-rate which produces required per-pixel flicker. This technique helps filmmakers make local adjustments to frame-rates which produce desired flicker.

Given a video sequence at input frame-rate, f_{inp} , and goal flicker maps, we generate video frames which produce desired amount of flicker. We define a content dependent flicker functional, F , which models the amount of flicker in a given frame of the video. We can evaluate F , for a given frame, using the model described in Section 4.2. In the Section 4.2, we have already modeled dependence of flicker on temporal contrast and spatial extent. Therefore, for a specific frame of the given video, F depends on the location, (x, y) , and the frame-rate function, $f(x, y)$.

In our experiments, we used input videos at 60 FPS. We take four constant samples of the frame-rate function, given by:

$$f(x, y) = fr : fr \in \{60, 30, 20, 15\}, \quad (4.12)$$

For these frame-rate values we use corresponding window sizes, W , of 5, 2.5, 1.67 and 1.25. We evaluate the flicker functional, $F(x, y, f(x, y))$, at these frame-rates. Then, we interpolate the functional between these points. Since we are considering frame-rates between $\frac{f_{inp}}{4}$ and f_{inp} , we also assume that goal flicker maps are given at $\frac{f_{inp}}{4}$ FPS. In this way, there is not going to be any conflict. After interpolating flicker functional for every discrete frame-rate between $\frac{f_{inp}}{4}$ to f_{inp} , we proceed to create frame-rate maps. For every pixel of the frame-rate map, first, we take the frame-rate value which produces flicker that is closest to the goal flicker in terms of squared difference. The corresponding equation is given as follows:

$$f_0(x, y) = \underset{p}{\operatorname{argmin}} (F_{goal}(x, y, f_0(x, y)) - F(x, y, p))^2 : p \in \{\frac{f_{inp}}{4} \dots f_{inp}\}, \quad (4.13)$$

$F_{goal}(x, y, f_0(x, y))$, is a painted goal flicker mask, whose values we can look up. Then, we perform spatio-temporal Gaussian smoothing of the frame-rate maps. Using the

method explained by Templin et al. [38], we have the capability to emulate spatially and temporally variable frame-rate. Their method takes frame-rate maps to emulate required frame-rates.

The steps mentioned, above, to produce frame-rate maps can be theoretically justified by solving the the following optimization problem for smooth frame-rate maps which produce goal flicker:

$$E_{flicker} = \int_{\Omega} \frac{1}{2} (F_{goal}(x, y, f_0(x, y)) - F_{out}(x, y, f(x, y)))^2 + \alpha |\nabla_3 f^2| d\Omega, \quad (4.14)$$

Ω is image domain, $F_{goal}(x, y, f_0(x, y))$ and $F_{out}(x, y, f(x, y))$ are realization of flicker function F . $F_{goal}(x, y, f_0(x, y))$ is input flicker map, $f_0(x, y)$ is frame-rate map which produces the goal flicker, $F_{out}(x, y, f(x, y))$ is output flicker map, which is dependent smooth frame-rate i.e. $f(x, y)$. The first term in the equation is data or fidelity term, which makes sure that the output flicker map is as close to goal flicker map as possible. $|\nabla_3 f^2|$ term penalizes 3D spatio-temporal gradient of frame-rate map $f(x, y)$, in other words the frame-rate map is spatially and temporally smooth. α controls the effect of smoothness term.

As demonstrated in the figure 4.6, the results from Daly et al. [6] show that as the frame-rate increases perceived flicker decreases linearly. Our own experimentation also shows similar results. Therefore, we assume that perceived flicker and frame-rate are related linearly. Using linearity we can expand the flicker functions as follows:

$$F_{goal}(x, y, f_0(x, y)) = \Lambda(x, y)f_0(x, y) + C(x, y), \quad (4.15)$$

$$F_{out}(x, y, f(x, y)) = \Lambda(x, y)f(x, y) + C(x, y), \quad (4.16)$$

Where $\Lambda(x, y)$ and $C(x, y)$ are some unknown functions, which are constant w.r.t frame-rate. Now, we can write equation 4.14 as follows:

$$E_{flicker} = \int_{\Omega} \frac{1}{2} (\Lambda(x, y)f_0(x, y) + C(x, y) - \Lambda(x, y)f(x, y) - C(x, y))^2 + \alpha |\nabla_3 f^2| d\Omega, \quad (4.17)$$

Equation 4.17 can be written as:

$$E_{flicker} = \int_{\Omega} \frac{1}{2} (f_0(x, y) - f(x, y))^2 + \frac{\alpha}{\Lambda(x, y)} |\nabla_3 f^2| d\Omega, \quad (4.18)$$

Assuming that $\Lambda(x, y)$ is not zero, we can treat $\frac{\alpha}{\Lambda(x, y)}$ as a parameter which controls smoothness.

Above equation just explains homogeneous diffusion, and there exists a closed form solution. Homogeneous diffusion is equivalent to Gaussian smoothing. The standard deviation of the Gaussian is related to the diffusion time as follow:

$$\sigma = \frac{1}{2} \sqrt{t}, \quad (4.19)$$

Therefore, if the diffusion time is larger we would have to choose bigger standard deviation of the Gaussian kernel. In our experiments we used Gaussian kernel of 11×11 pixels. For temporal smoothing, we smooth a frame with two of its neighboring frames. Since we were using video sequences containing fast motion, naively blending the adjacent frames produced artifacts. Therefore, for temporal smoothing we used motion compensation.

4.3.3 Results

The Results are shown using Dell screen, with resolution 1920×1200 . We made a mapping of gray values and their corresponding luminance values using Minolta LS-100 luminance meter. Average distance between the screen and the viewer is ca. 60 cm. Therefore, one visual degree computes to ca. 40 pixels.

We present the results of the method introduced in Section 4.3.1, by using a video sequence from Sintel dataset [46]. Screen-shot of the scene that was used is given in the figure 4.7. The results demonstrate that the flicker was increased in the areas which were smooth, while the areas which already had flicker stayed the same. We also performed experiments on a synthesized example, the screen-shot is shown in 4.7. In this example, we used bars, of different contrasts, moving at different speeds from left to right, on a grey background. We calculated the flicker using temporal contrast and optical flow. Then we mapped the highest flicker values to the input frame-rate, and lowest flicker values to the half of the input frame-rate. This introduced more flicker in the bars with low contrast, and low speed. For both of these examples, spatial extent was taken as optical flow vector, both magnitude and direction.

To demonstrate the results for Section 4.3.2 we used 3 scenes from Sintel dataset [46] and a synthetic bars sequence. Screen shots of the scenes, from sintel dataset, along

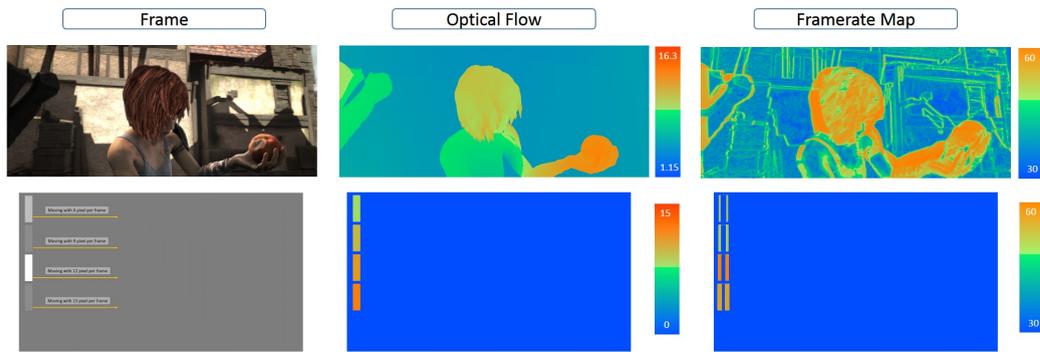


FIGURE 4.7: Scenes used to demonstrate examples of enhancing film-look. In the bottom figure bars of different contrasts are moving from left to right with different speed. The figure shows that high frame-rate corresponds to high contrast and fast speeds.

with their corresponding optical flows, flicker maps and frame-rate maps are shown in figure 4.8. These results show that our model predicts higher amount of flicker around the objects with high optical flow. Due to faster motion the spatial extent of the moving objects is increased which produces flickering. The last column of the figure shows frame-rate map, to produce constant flicker value of 0 JNDs. The objects moving with higher speeds are assigned higher frame-rates.

Figure 4.9 shows the dependence of flicker on frame-rate. It shows that as the frame-rate decreases perceived flicker around the objects with high speed and high contrast, increases. The difference maps shown in the figure 4.10 demonstrate that flicker increases as the frame-rate decreases. The Figure 4.11 shows flicker maps overlaid over frames. It demonstrates that our metric correctly predicts flicker around fast moving objects.

As we discussed in the Section 4.2, perceived flicker depends on contrast and spatial extent. The spatial extent can either increase due to faster speed or large size of the object. To demonstrate that our flicker metric is sensitive to the these attributes we also performed test on bars, with different contrast and speed values. Figures 4.12 and 4.13 demonstrate that higher flicker is predicted for fast speed, high contrast objects and at low frame-rates.

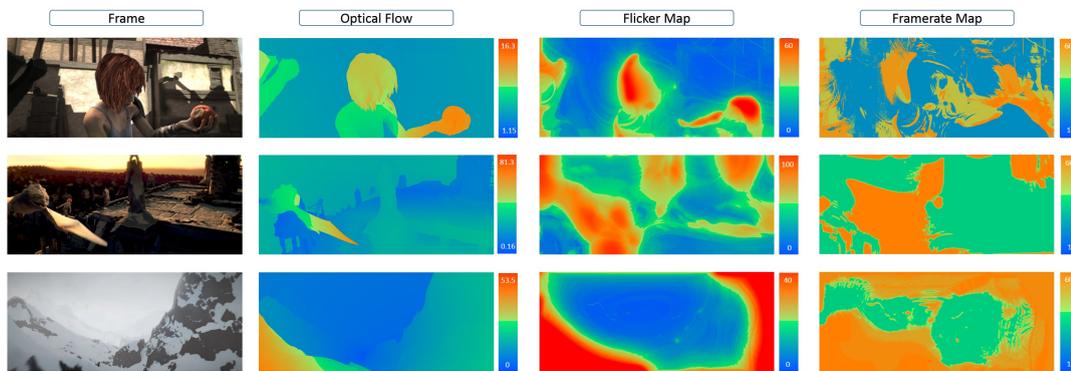


FIGURE 4.8: For these results flicker values was fixed at 0 JND. The left column shows sample frames of scenes named alley, temple and mountain. The middle columns shows their corresponding optical flow. The column on the right shows frame-rate map.

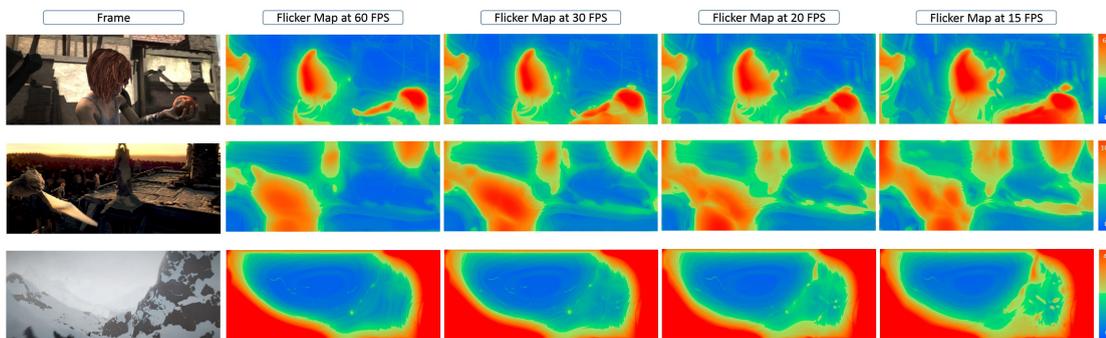


FIGURE 4.9: Demonstration of flicker maps at different frame-rates. As the frame-rate gets lower flicker increase in certain regions.

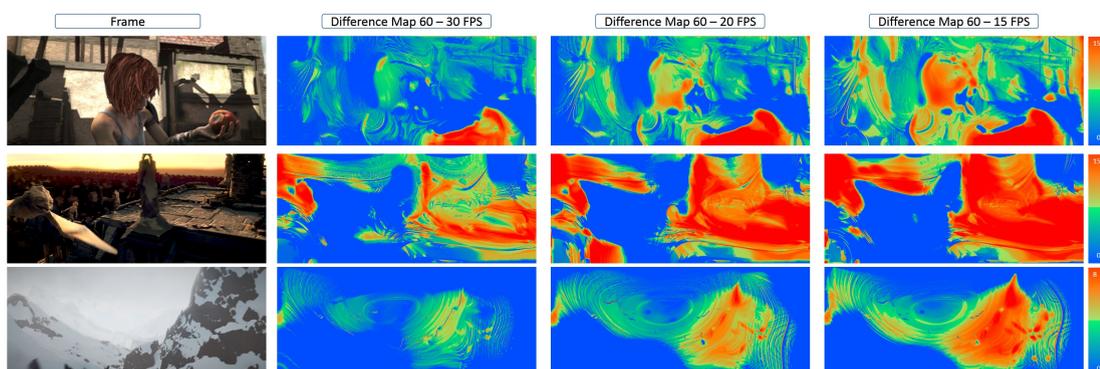


FIGURE 4.10: Demonstration of difference of flicker maps between 60 FPS and other frame-rates. As the frame-rate decreases, flicker increases in regions with high contrast and speed.

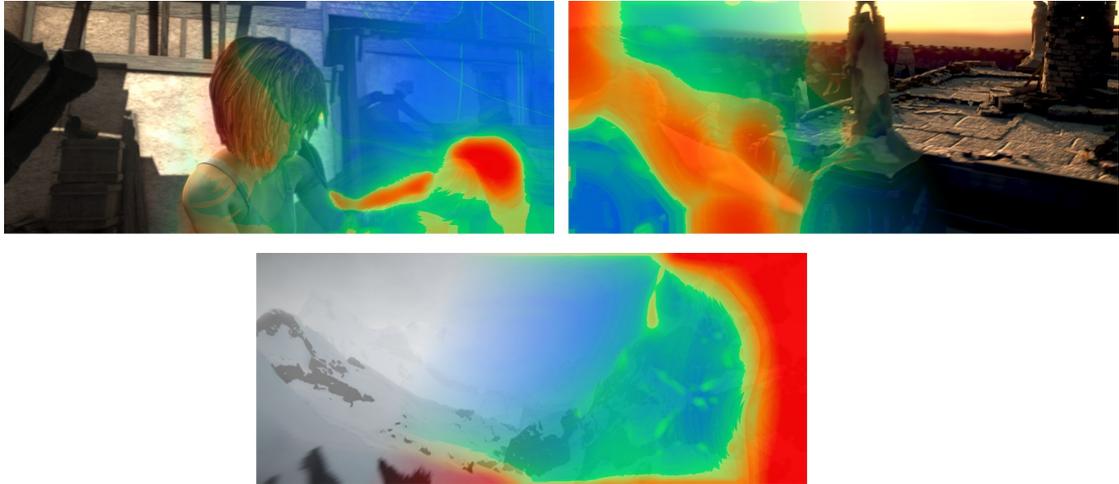


FIGURE 4.11: Flicker maps overlaid on the frames. The results show that flicker is high around high contrast and fast moving parts such as, around the hand in alley scene, around the wings of the dragon in temple scene, and around the dark spots in the mountain scene.

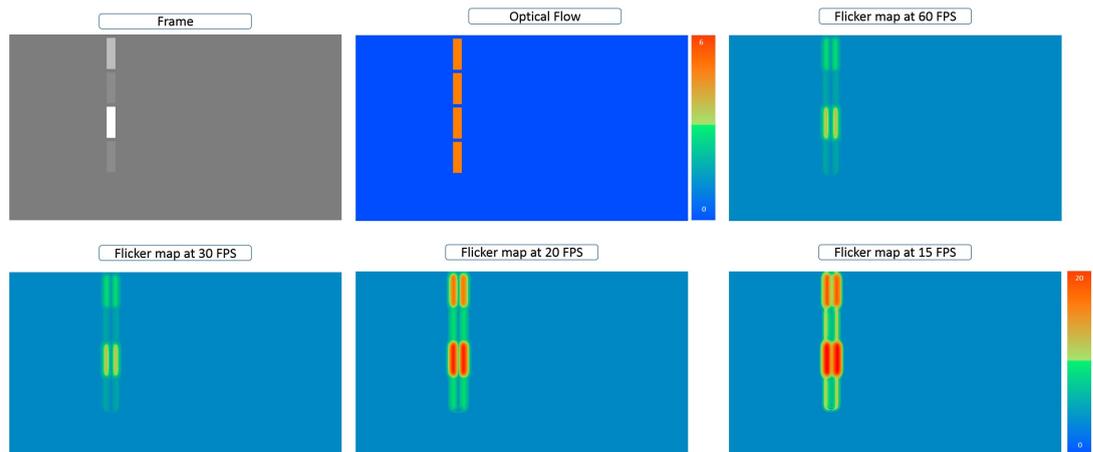


FIGURE 4.12: Four bars, of different contrast moving from left to right with same the speed of 6 px per frame for 60 FPS. The results demonstrate that flicker is greater for the higher contrasts.

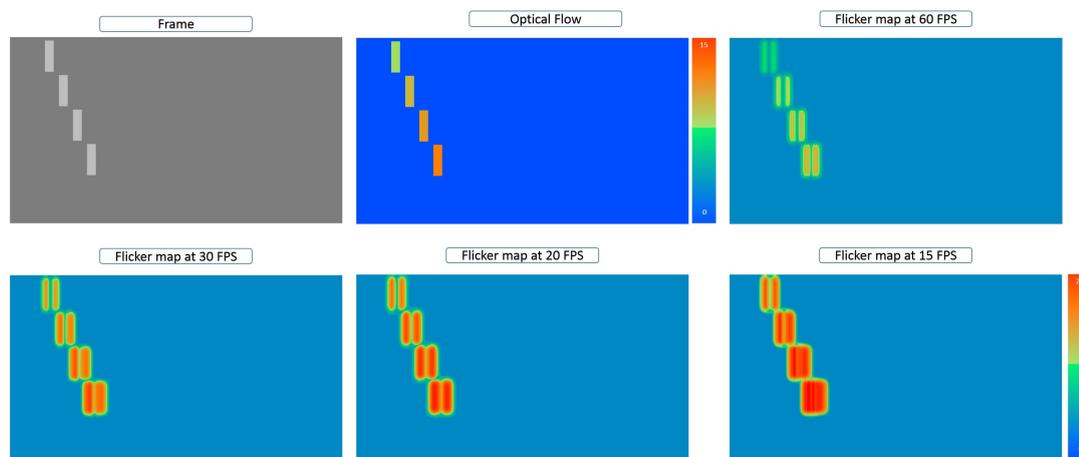


FIGURE 4.13: Four bars are moving from left to right with different speeds of 6,9,12 and 15 pixels per frame for 60 FPS. The results demonstrate that flicker is greater for the faster speed.

Chapter 5

Conclusion and Future Work

This thesis explored perceptual implications of HFR. Specifically, we tested the myth that speed perception is dependent on frame-rate. We came to the conclusion that there is no significant effect of frame-rate on speed perception. Even if there is an effect, it is very small and probably negligible. A literature survey of factors that effect speed perception was done. We identified three most important factors affecting perceived speed: contrast, spatial frequency and speed. We took different sample points of these attributes. Then, we designed and performed three psychophysical experiments, using different kinds of stimuli, such as Gabor patches, complex animated sequences, and real-world sequences. In some experiments, users were asked to adjust speed, while in other experiments they were asked to choose one of the presented videos that appeared faster. Control sequences, where both the presented input videos were of same frame-rate, were used to separate the effect of frame-rate from other attributes. Then, we performed statistical analysis and calculated p-values to show confidence in the readings. In all these experiments, we did not get any conclusive or overwhelming evidence regarding quantifiable effect of frame-rate on speed perception. Some cases showed some differences in perception for different frame-rates, but the effects were either too small or not consistent.

In the second part of the thesis, we researched flicker perception and its relationship with frame-rate. We presented a model to evaluate content dependent per-photoreceptor flicker for complex motion, in a given sequence of images. We proposed that “flim-look” is a blend of artifacts that occur due to low frame-rates. Among these artifacts, flicker is the most prominent difference between two frame-rates. By manipulating flicker, we also manipulate other artifacts, proportionately. Hence, we can manipulate “flim-look” by changing flicker. Based on the presented model, we also presented methods to use variable frame-rate technology. Firstly, we presented a method to increase the

“film-look”, in a given high frame-rate video, by increasing flicker. Then, we presented a method to make flicker driven frame-rate manipulations. To this end, we took a video and flicker masks as input, and generated spatially and temporally varying frame-rate maps to match input flicker masks. We, then, used these frame-rate maps to set appropriate per-pixel frame-rate, using variable frame-rate technology.

A future direction could be to make a perceptual metric for “film-look”. We can investigate and quantify the artifacts that contribute to the “film-look”. In our work, we assumed that every pixel is being tracked through SPEM, so all the pixels stays in the fovea. With the increase in popularity of wide field of view displays, perhaps a general purpose flicker metric could be developed. In this metric, we could model perceived flicker for both fovea and periphery, taking their perceptual differences into consideration.

Chapter 6

Appendix

6.1 If Frame-Rate Affects Flicker non-Linearly

$$E_{x,y,f} = \int_{\Omega} \frac{1}{2} (F_{goal}(x, y, f_0) - F_{out}(x, y, f(x, y)))^2 + \alpha |\nabla f|^2 d\Omega, \quad (6.1)$$

Necessary conditions for the extremum are given by Euler-Lagrange equation which is as follows:

$$\frac{\partial E_{flicker}}{\partial f} = (F_{goal}(x, y, f_0) - F_{out}(x, y, f(x, y))) \frac{\partial F_{out}(x, y, f(x, y))}{\partial f} - \alpha \Delta f = 0, \quad (6.2)$$

The goal is to find $f(x, y)$ and to do so we can use warping. First step would be to introduce fixed point iteration,

$$F_f^{out}(x, y, f^k) (F_{goal}(x, y, f_0) - F_{out}(x, y, f^{k+1}(x, y))) - \alpha \Delta f^{k+1} = 0, \quad (6.3)$$

$F_f^{out}(x, y, f^k)$ is evaluated at old time stamp, while the rest is computed at current iteration. Fully implicit computations might be very complex and slow.

Since, we cannot say anything about convexity relationship of F^{out} and $f(x, y)$, we would have to break the computations in small increments,

$$f^{k+1} = f^k + df^k, \quad (6.4)$$

We are interested in finding $f(x, y)$ and it appears in the argument of F^{out} , as we broke down the computations in smaller increments we can linearize 6.3 around (x, y, f^k) which gives us

$$\begin{aligned} F^{out}(x, y, f^{k+1}(x, y)) &= F^{out}(x, y, f^k(x, y) + df^k(x, y)) \\ &\approx F^{out}(x, y, f^k(x, y)) + F_f^{out}(x, y, f^k)df^k(x, y), \end{aligned} \quad (6.5)$$

If we put above results in 6.3 we get following equation as follows:

$$F_f^{out^2}(x, y, f^k)df^k + F_f^{out}(x, y, f^k)(F_{goal}(x, y, f_0) - F_{out}(x, y, f^k(x, y))) - \alpha \Delta(f^k + df^k) = 0, \quad (6.6)$$

$F_f^{out^2}$ is square of the first derivative of F_{out} with respect to f , in above equation we have to solve for df^k .

Since we are working with images, we have to discretize our equations. $F_{out}(x, y, f)$, $F_{goal}(x, y, f_0)$, $f(x, y)$ and $df(x, y)$ can be represented in discrete form as $F_{i,j,f_{i,j}}^{out}$, F_{i,j,f_0}^{goal} , $f_{i,j}$ and $df_{i,j}$. As Laplace is linear operator we can separate it,

$$\alpha \Delta(f^k + df^k) = \alpha \Delta f^k + \alpha \Delta df^k, \quad (6.7)$$

We can use central differences for approximating double derivatives, which is give as follows:

$$\Delta f^k = \sum_{l \in x, y} \sum_{\tilde{i}, \tilde{j} \in N(i, j)} \frac{f_{\tilde{i}, \tilde{j}}^k - f_{i, j}^k}{h_l^2} \quad (6.8)$$

In above equation $N(i, j)$ are neighbors of pixel (i, j) , whereas l represents the dimension and h_l the grid size, we take the grid size to be 1. Putting all the discretizations explained above in equation 6.6 and solving for $df_{i,j}$ we get,

$$df_{i,j}^k = \left(-F_f^{out2}(i, j, f_{i,j}^k) - \frac{1}{h_l^2}\right)^{-1} + F_f^{out}(i, j, f_{i,j}^k)(F_{i,j,f_0}^{goal} - F_{i,j,f_{i,j}^k}^{out}) - \alpha \sum_{l \in x,y} \sum_{\tilde{i}, \tilde{j} \in N_l(i,j)} \frac{f_{\tilde{i}, \tilde{j}}^k - f_{i,j}^k}{h_l^2} - \alpha \sum_{l \in x,y} \sum_{\tilde{i}, \tilde{j} \in N_l(i,j)} \frac{df_{\tilde{i}, \tilde{j}}^k}{h_l^2},$$

We have the discretized version of almost all the terms, and we can see that the equation that we have to solve is linear. So we solve the discretized version of equation 6.6, using Gauss Seidel approach. We do so by introducing fixed point iterator g .

$$df_{i,j}^{k,g+1} = \left(-F_f^{out2}(i, j, f_{i,j}^k) - \frac{1}{h_l^2}\right)^{-1} + F_f^{out}(i, j, f_{i,j}^k)(F_{i,j,f_0}^{goal} - F_{i,j,f_{i,j}^k}^{out}) - \alpha \sum_{l \in x,y} \sum_{\tilde{i}, \tilde{j} \in N_l(i,j)} \frac{f_{\tilde{i}, \tilde{j}}^k - f_{i,j}^k}{h_l^2} - \alpha \sum_{l \in x,y} \sum_{\tilde{i}, \tilde{j} \in N_l(i,j)} \frac{df_{\tilde{i}, \tilde{j}}^k - df_{i,j}^k}{h_l^2} - \alpha \sum_{l \in x,y} \sum_{\tilde{i}, \tilde{j} \in N_l^-(i,j)} \frac{df_{\tilde{i}, \tilde{j}}^{k,g+1}}{h_l^2} - \alpha \sum_{l \in x,y} \sum_{\tilde{i}, \tilde{j} \in N_l^+(i,j)} \frac{df_{\tilde{i}, \tilde{j}}^{k,g}}{h_l^2},$$

N^- represent already visited neighbors of pixel (i, j) hence they are evaluated at new iteration level i.e. $g + 1$, N^+ represents neighbors which are yet to be visited and so they are evaluated at old time stamp.

We do not know the exact function through which frame-rate function, $f_{i,j}$, and the flicker functional, $F_{i,j,f_{i,j}}$, are related to each other. Therefore, we sample the input video at different frame rates and interpolate for the values in between.

6.2 Results: HFR Effects of Speed Perception using Gabor Patches

The graphs show veridical speed of the reference, which is always 30 FPS, on X-axis. The perceived speed is shown on y-axis. In experiment turns (green lines) test patches were 60 FPS. The refresh-rate of the screen was 120 FPS. For each of the given figures,

contrast and spatial frequencies are fixed; their corresponding values are mentioned on top of each figure. The error-bars shown in the figures are standard errors of mean (SEM). P-values are resultant of paired t-tests. They are also mentioned in the graphs.

Another interesting, yet not surprising, thing to observe here is that speed perception works much better for lower speeds. People tend to underestimate higher speeds, which is explained by Bayesian theory. So, the underestimation of the speed that we see here, could just be as a result of well-established results discussed in Section 3. The difference between different frame-rates could just be accidental.

For lowest spatial frequency (0.5 cpd) speed perception, for both frame-rates, looks almost similar. The highest effect, $\approx 1^\circ$, for 0.5 cpd was at medium speed (8° per second) and 53% contrast. For convenience, the speeds, in rest of the section, are just written as number of degrees.

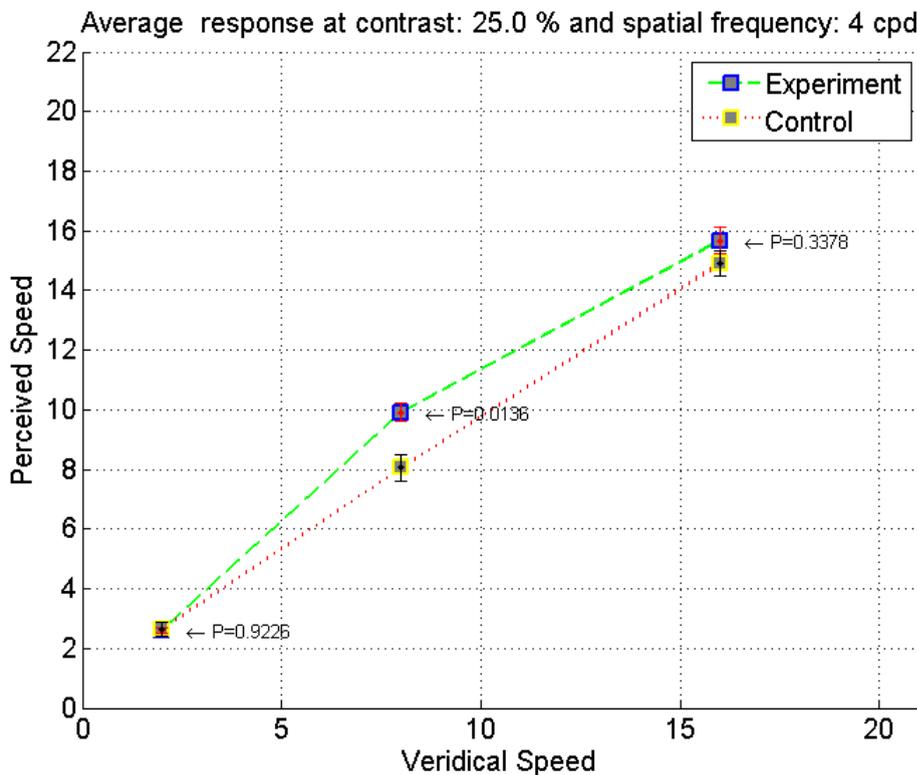


FIGURE 6.1

At 2 cpd, a higher effect is observed at higher speed, albeit the p-values are not very low. Perhaps, more experimentation would be beneficial here. The effect is around 0.5° for 2° , 1° for 8° and $\approx 1 - 3^\circ$ for 16° . In most cases, at the speed 8° , p-values are lower and SEM for test and control never cross each other. The highest response for 2 cpd is at 16° and 53 % contrast, perhaps partly due to speed underestimation.

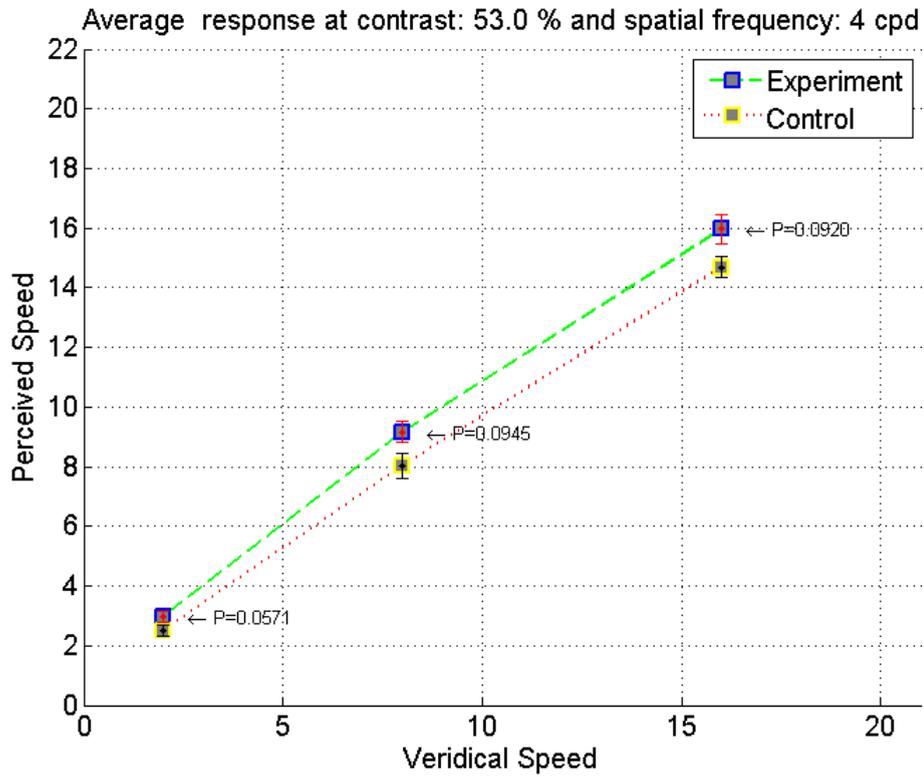


FIGURE 6.2

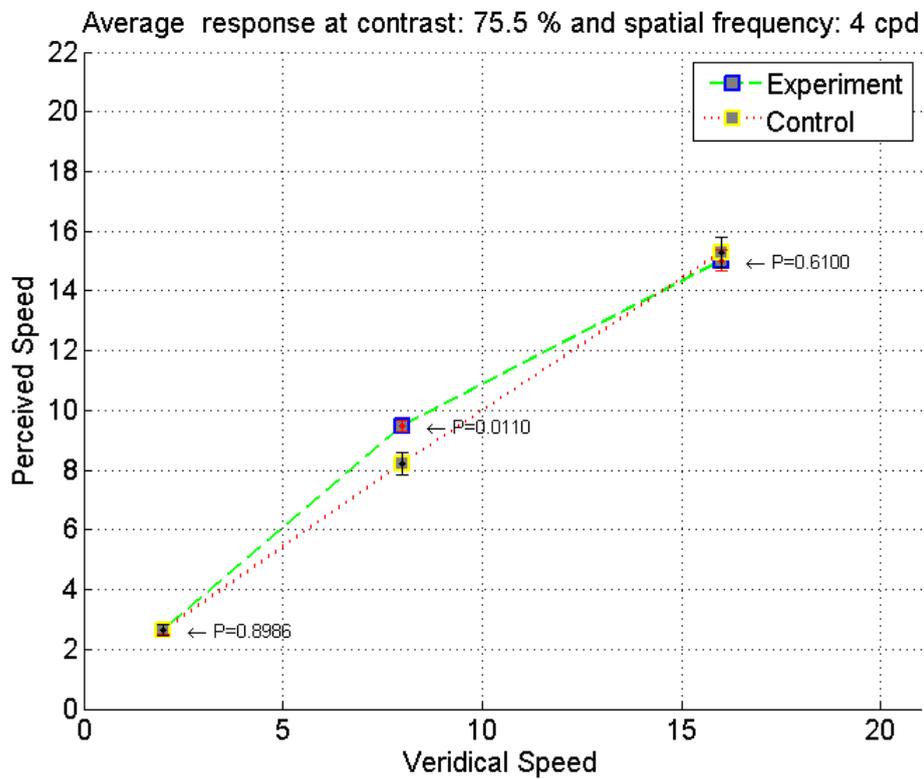


FIGURE 6.3

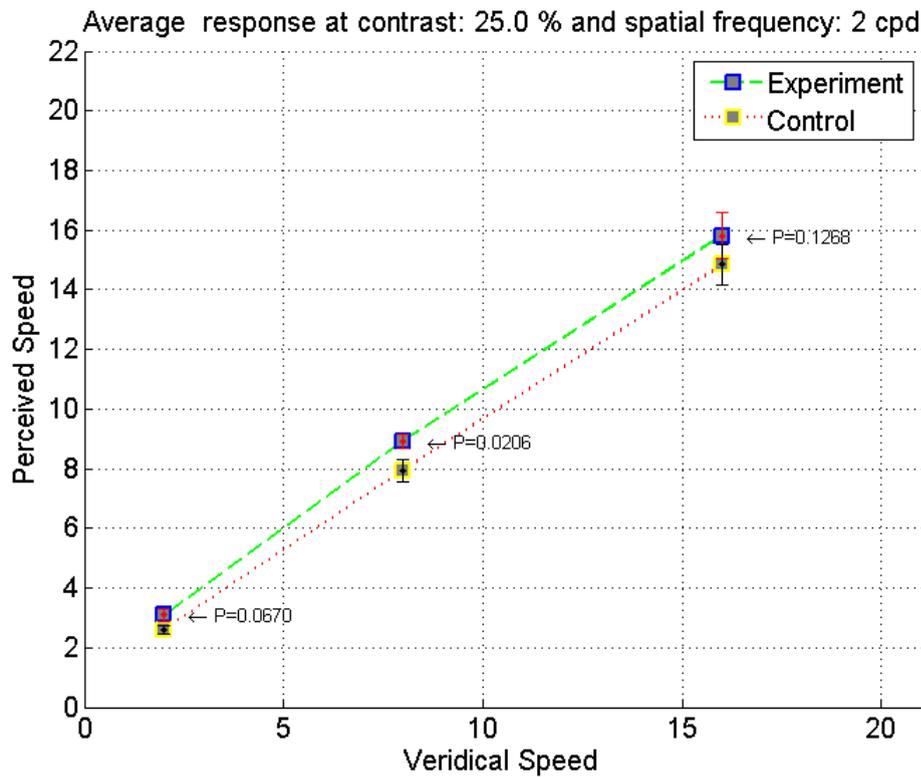


FIGURE 6.4

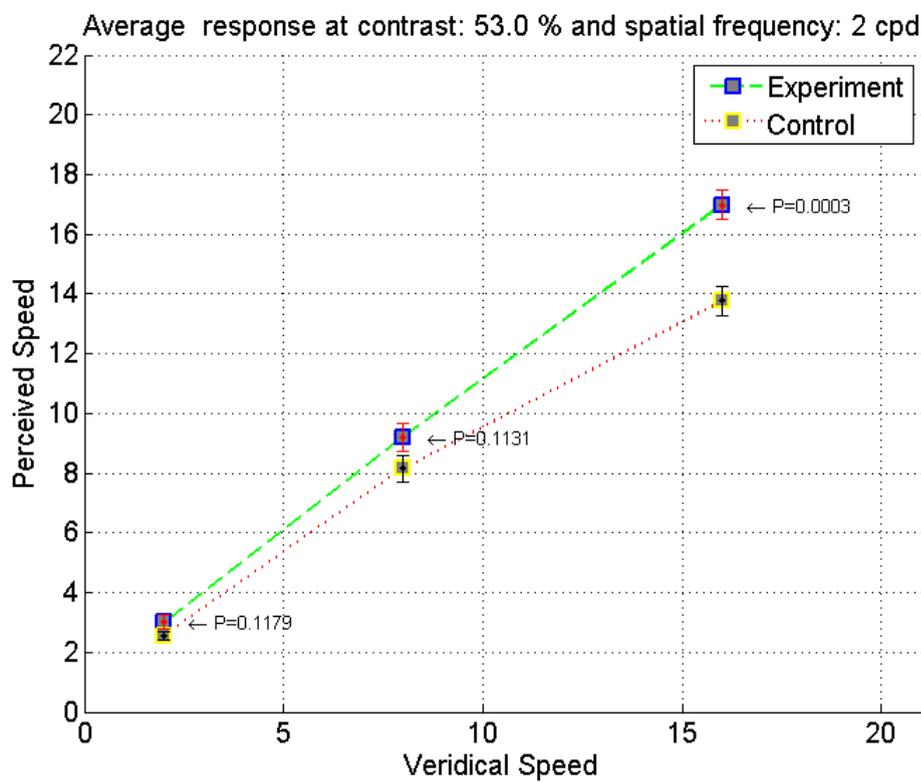


FIGURE 6.5

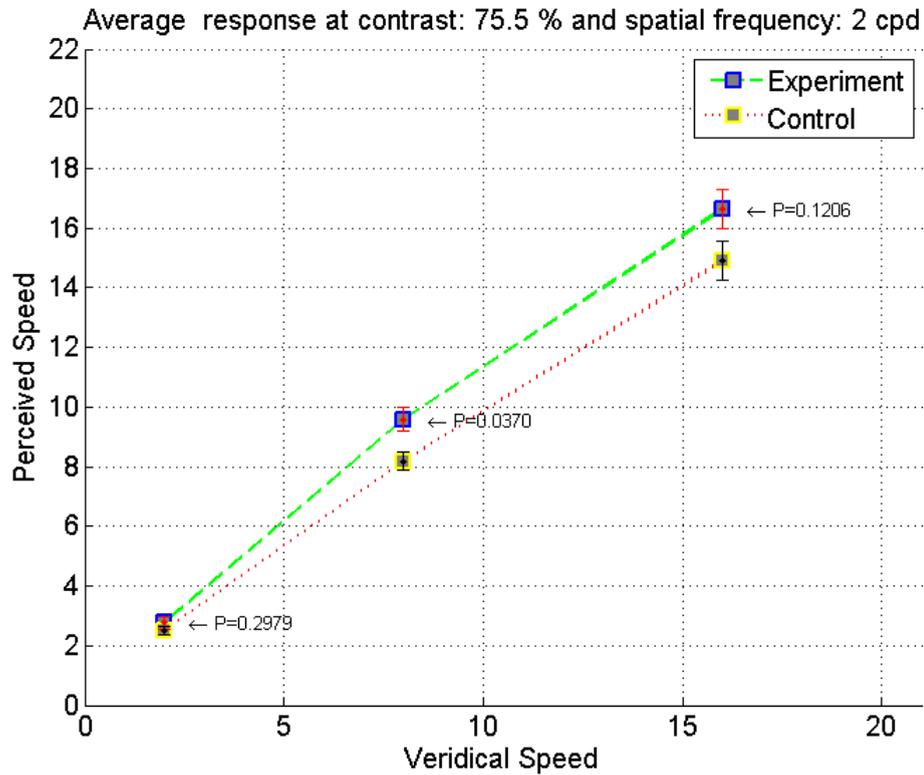


FIGURE 6.6

At 4 cpd, the speed estimation of low and mediums speeds is quite accurate for different contrast values. On the other hand, at higher spatial frequency and higher speed is underestimated. When the contrast is considerably higher, perhaps it provided more information resulting in an accurate estimation of even a higher speed as can be seen 6.9. Underestimation of speed is again, mostly, observable for 8°.

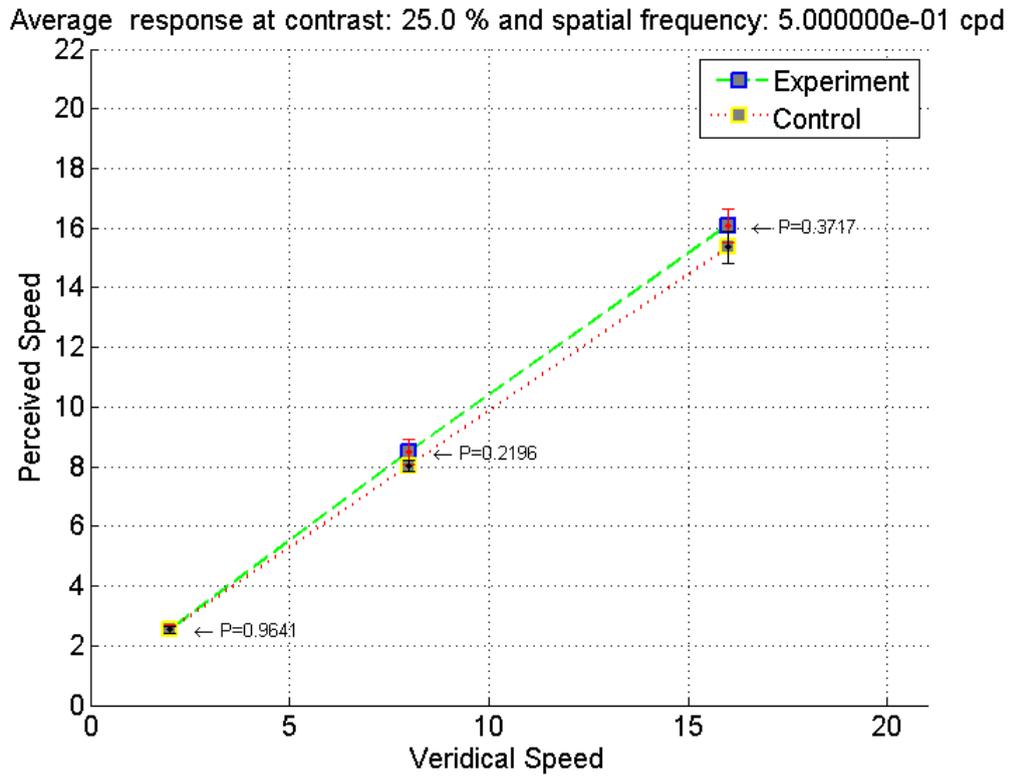


FIGURE 6.7

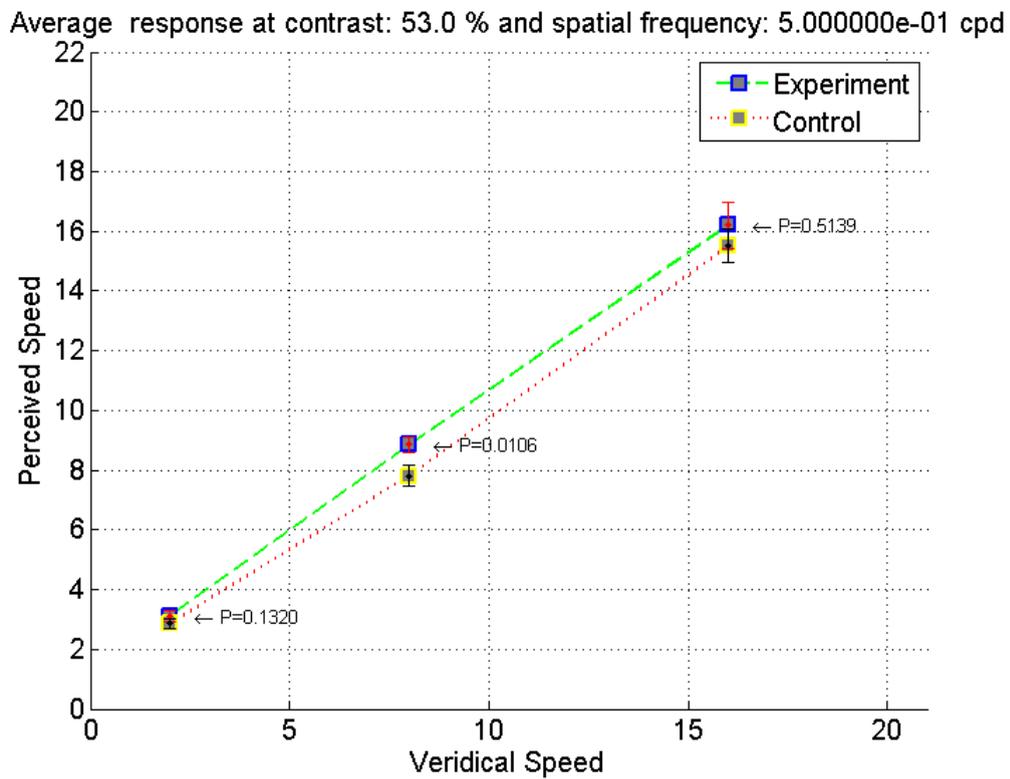


FIGURE 6.8

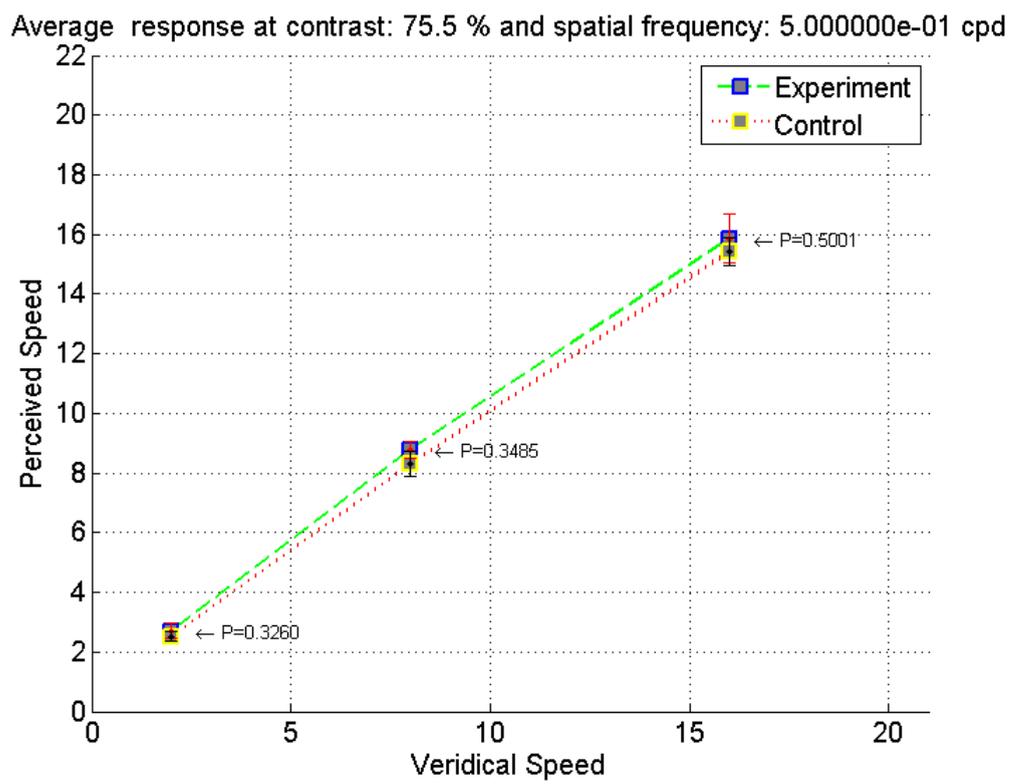


FIGURE 6.9

Bibliography

- [1] Edward H Adelson and James R Bergen. Spatiotemporal energy models for the perception of motion. *JOSA A*, 2(2):284–299, 1985.
- [2] Colin Blakemore, James PJ Muncey, and Rosalind M Ridley. Stimulus specificity in the human visual system. *Vision research*, 13(10):1915–1931, 1973.
- [3] G. Bradski. *Dr. Dobb's Journal of Software Tools*.
- [4] Anne-Marie Brouwer, Tom Middelburg, Jeroen B. J. Smeets, and Eli Brenner. Hitting moving targets. *Experimental Brain Research*, 152(3):368–375, 2003.
- [5] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *Computer Vision-ECCV 2004*, pages 25–36. Springer, 2004.
- [6] Scott Daly, Ning Xu, James Crenshaw, and Vikrant J Zunjarrao. A psychophysical study exploring judder using fundamental signals and complex imagery. *SMPTE Motion Imaging Journal*, 124(7):62–70, 2015.
- [7] Scott J. Daly. Engineering observations from spatiovelocity and spatiotemporal visual models, 1998.
- [8] Piotr Didyk, Elmar Eisemann, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Apparent display resolution enhancement for moving images. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2010, Los Angeles)*, 29(4), 2010.
- [9] Daniel Engber. It looked great. it was unwatchable., 2016. <https://goo.gl/26CNeS>.
- [10] JE Farrell, Brian L Benson, and Carl R Haynie. Predicting flicker thresholds for video display terminals. In *Proc. SID*, volume 28, pages 449–453. Citeseer, 1987.
- [11] Sbastien Georges, Peggy Seris, Yves Frgnac, and Jean Lorenceau. Orientation dependent modulation of apparent speed: psychophysical evidence. *Vision Research*, 42(25):2757 – 2772, 2002.
- [12] Andrei Gorea and Christopher W. Tyler. New look at bloch's law for contrast. *J. Opt. Soc. Am. A*, 3(1):52–61, Jan 1986.
- [13] Stephen T Hammett, Rebecca A Champion, Antony B Morland, and Peter G Thompson. A ratio model of perceived speed in the human visual system. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1579):2351–2356, 2005.

- [14] Stephen T. Hammett, Rebecca A. Champion, Peter G. Thompson, and Antony B. Morland. Perceptual distortions of speed at low luminance: Evidence inconsistent with a bayesian account of speed encoding. *Vision Research*, 47(4):564 – 568, 2007.
- [15] Stephen T Hammett, Mark A Georgeson, and Andrei Gorea. Motion blur and motion sharpening: temporal smear and local contrast non-linearity. *Vision Research*, 38(14):2099 – 2108, 1998.
- [16] Michael Kalloniatis and Charles Luu. Temporal resolution, 2009. <http://webvision.med.utah.edu/book/part-viii-gabac-receptors/temporal-resolution/>.
- [17] DH Kelly. Effects of sharp edges in a flickering field. *JOSA*, 49(7):730–732, 1959.
- [18] DH Kelly. Diffusion model of linear flicker responses. *JOSA*, 59(12):1665–1670, 1969.
- [19] R. J. Krauzlis and S. G. Lisberger. Temporal properties of visual motion signals for the initiation of smooth pursuit eye movements in monkeys. *Journal of Neurophysiology*, 72(1):150–162, 1994.
- [20] Janus B. Kristensen. Big buck bunny, 2008. <http://bbb3d.renderfarming.net/download.html>.
- [21] Yoshihiko Kuroki. Improvement of 3d visual image quality by using high frame rate. *Journal of the Society for Information Display*, 20(10):566–574, 2012.
- [22] Justin Laird, Mitchell Rosen, Jeff Pelz, Ethan Montag, and Scott Daly. Spatio-velocity csf as a function of retinal velocity using unstabilized stimuli, 2006.
- [23] James Larimer, Jennifer Gille, and James Wong. 41.2: Judder-induced edge flicker in moving objects. In *SID Symposium Digest of Technical Papers*, volume 32, pages 1094–1097. Wiley Online Library, 2001.
- [24] Pia Makela, Jyrki Rovamo, and David Whitaker. Effects of luminance and external temporal noise on flicker sensitivity as a function of stimulus size at various eccentricities. *Vision Research*, 34(15):1981 – 1991, 1994.
- [25] William R Mark, Leonard McMillan, and Gary Bishop. Post-rendering 3d warping. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 7–ff. ACM, 1997.
- [26] Nestor Matthews, Bruce Luber, Ning Qian, and Sarah H. Lisanby. Transcranial magnetic stimulation differentially affects speed and direction judgments. *Experimental Brain Research*, 140(4):397–406, 2001.
- [27] Suzanne P McKee. A local mechanism for differential velocity detection. *Vision research*, 21(4):491–500, 1981.
- [28] Carolyn Michelle, Charles H Davis, Craig Hight, and Ann L Hardy. The hobbit hyperreality paradox polarization among audiences for a 3d high frame rate film. *Convergence: The International Journal of Research into New Media Technologies*, page 1354856515584880, 2015.
- [29] Shin'ya Nishida. Advancement of motion psychophysics: review 2001–2010. *Journal of Vision*, 11(5):11–11, 2011.

- [30] rexcrk. Gamefaqs, 2014. <http://www.gamefaqs.com/boards/227-movies-at-the-theater/68078160?page=1>.
- [31] Rod Ryan. American cinematographer manual. 1993.
- [32] Tiffany Saffell and Nestor Matthews. Task-specific perceptual learning on speed and direction discrimination. *Vision Research*, 43(12):1365 – 1374, 2003.
- [33] A.T. Smith and G.K. Edgar. The influence of spatial frequency on perceived temporal frequency and perceived speed. *Vision Research*, 30(10):1467 – 1474, 1990.
- [34] Alan Stocker and Eero P Simoncelli. Constraining a bayesian model of human visual speed perception. In *Advances in Neural Information Processing Systems*, pages 1361–1368, 2004.
- [35] Alan A Stocker and Eero P Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, 9(4):578–585, 2006.
- [36] Leland S. Stone and Peter Thompson. Human speed perception is contrast dependent. *Vision Research*, 32(8):1535 – 1549, 1992.
- [37] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1):23–34, 2004.
- [38] Krzysztof Templin, Piotr Didyk, Karol Myszkowski, and Hans-Peter Seidel. Emulating displays with continuously varying frame rates. *ACM Transactions on Graphics (TOG)*, 35(4):67, 2016.
- [39] Peter Thompson. Velocity after-effects: the effects of adaptation to moving stimuli on the perception of subsequently seen moving stimuli. *Vision research*, 21(3):337–345, 1981.
- [40] Peter Thompson, Kevin Brooks, and Stephen T Hammett. Speed can go up as well as down at low contrast: Implications for models of motion perception. *Vision research*, 46(6):782–786, 2006.
- [41] Jan PH Van Santen and George Sperling. Elaborated reichardt detectors. *JOSA A*, 2(2):300–321, 1985.
- [42] Scott NJ Watamaniuk and Stephen J Heinen. Perceptual and oculomotor evidence of limitations on processing accelerating motion. *Journal of vision*, 3(11):5–5, 2003.
- [43] Andrew B Watson and Albert J Ahumada. Model of human visual-motion sensing. *JOSA A*, 2(2):322–342, 1985.
- [44] Peter Werkhoven, Herman P. Snippe, and Toet Alexander. Visual processing of optic acceleration. *Vision Research*, 32(12):2313 – 2329, 1992.
- [45] Laurie M Wilcox, Robert S Allison, John Helliker, Bert Dunk, and Roy C Anthony. Evidence that viewers prefer higher frame-rate film. *ACM Transactions on Applied Perception (TAP)*, 12(4):15, 2015.
- [46] J. Wulff, D. J. Butler, G. B. Stanley, and M. J. Black. Lessons and insights from creating a synthetic optical flow benchmark. In A. Fusiello et al. (Eds.), editor, *ECCV Workshop on Unsolved Problems in Optical Flow and Stereo Estimation*, Part II, LNCS 7584, pages 168–177. Springer-Verlag, October 2012.

- [47] N. Xu, I.I.I.K.D. THURSTON, S. Daly, J.E. Crenshaw, and A. SHE. Systems and methods to control judder visibility, October 22 2015. WO Patent App. PCT/US2015/017,110.