The Optical Coherence Tomography Microsurgical Augmented Reality System (OCT-MARS): A Novel Device for Microsurgeries

Samantha Horvath

CMU-RI-TR-16-56

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Robotics.

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

September 2016

Thesis Committee:
Dr. George Stetten
Dr. John Galeotti
Dr. Roberta Klatzky
Dr. Thomas Furness

Sponsors:
NIH R01EY021641
NSF GRFP DGE-1252522
DoD CDMRP
Research to Prevent Blindness

Abstract

This thesis describes the development and testing of the Optical Coherence Tomography Microsurgical Augmented Reality System (OCT-MARS). This system allows surgeons to view real-time medical image data as an *in situ* overlay within the surgical field. There are a number of clinical applications for which real time, *in situ* visualization of otherwise transparent structures of the eye would be beneficial to surgeons. The primary motivating application for this project is the surgical treatment of glaucoma. We have built a projection system capable of producing flat and tilted images in the normal field of view of the microscope with sufficient brightness and resolution to be viewed under magnification. We have studied the perception of tilted surfaces under magnification and found that OCT images provide sufficient stereo information to be correctly perceived. Finally, we have tested stereo perception under magnification using surgically relevant tasks to evaluate the effectiveness of the system.

Acknowledgements

Advisors Dr. George Stetten and Dr. John Galeotti

Thesis committee members Dr. Roberta Klatzky and Dr. Thomas Furness

Fellow students, esp. Tejas Mathai, Jihang Wang, Kori MacDonald and Daniel Freer

The Drs. Horvath.

They told me I shouldn't do this.

They were probably right.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

This thesis concerns the design and testing of a novel Optical Coherence Tomography Microsurgical Augmented Reality System (OCT-MARS). The goal of this work is to allow surgeons to view real-time OCT data through a stereomicroscope during surgery, overlaid *in situ* into the existing microscope image. Our current clinical application focuses on surgery involving the anterior segment of the eye (cornea, lens, sclera, etc.).

In order to produce a system that works well with surgeons, the overall design must not only consider what information we can overlay but also what information the surgeons can perceive through the microscope, as well as what information they actually need to assist with their surgery. These considerations correspond to the three aims of the original proposal, covered in Chapters 3, 4, and 5, respectively. We review our assertion of them here as stated in the dissertation proposal, and describe the progress made towards each in the corresponding chapter of this dissertation.

What we can display: We are limited by the available technology as to what can be overlaid under the microscope. Resolution is of the utmost importance, since the overlaid images are viewed through the full optical magnification of the microscope (typically 10x to 30x). Image brightness is crucial to successful integration into the microscope optics. Flat (un-tilted) images at infinity are trivial to create within the microscope, but are of limited usefulness as they contain no depth information. Planar in-situ images are readily achievable, but are also limited in the information that they can contain (i.e. the image cannot contain depth information that does not conform to a planar surface.) Curved surfaces can be more informational and intuitive, but are difficult to create optically. Chapter 3 details our work to build a system that can provide the most information to the surgeon.

What users can perceive: Because the optics of the microscope distort essential perceptual cues, the capability to project an image containing certain information does not necessarily imply that the information is perceived correctly by the user. Why provide a complicated curved-surface image if the user cannot correctly perceive it? Chapter 4 covers our research to determine the limits of perception while viewing structures through a microscope.

What information surgeons need: If information can be displayed and correctly perceived, we must also ask, is it useful? In-situ information display has just as much capability to distract as it does to help. An integral part of the original proposal was collaboration with clinical researchers to determine what information they want to have overlaid, and how best to display it. This aim includes not only discussing these goals with clinicians, but also determining empirically how well they interact with the system. This preclinical validation relies on usability studies based on a pre-established set of key procedures. Chapter 5 describes the progress on this aim.

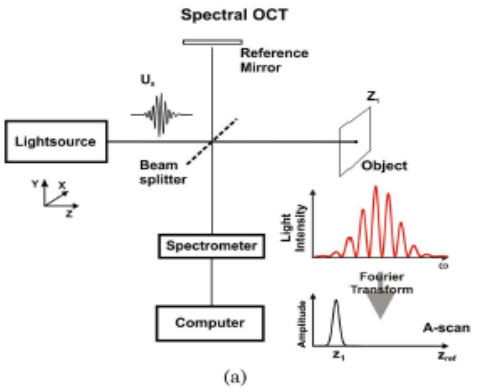
Chapter 2 describes the background and literature for the aims of this thesis. Chapters 3 through 5 describe the aims listed above. Concluding remarks are in Chapter 6.

2 BACKGROUND

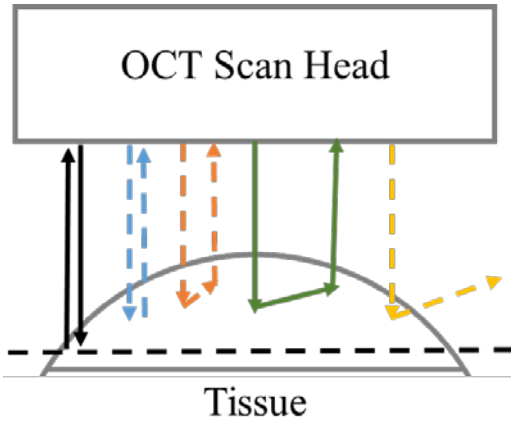
2.1 OCT DETAILS

Optical Coherence Tomography (OCT) is a promising technology for in-vivo imaging. OCT uses low coherence interferometry to create 3D reconstructions of shallow tissue (maximum 2 mm of penetration in translucent tissue) [1]. The specific type of OCT used in this project is spectral domain OCT (SD-OCT). Figure 2-1 shows the system diagram of a spectral domain OCT system. Reflected light from the sample is analyzed using an interferometer and a spectrometer to determine the reflectivity of the sample at different depths. Different depths exhibit constructive interference at different sets of optical wavelengths. Taking an inverse Fourier Transform of the spectrometer's line-scan pixels allows decoding of reflectivity measurements at each individual depth simultaneously.

The key function of OCT is to filter the returned light, in order to remove the scattered reflections while preserving the specular reflections from the target depth. This filtering is facilitated by the fact that most of the reflected scattered light will have a different optical path length than the specular reflections. OCT uses an interferometer to reject returned light with an optical path length that deviates from the preset reference distance (set by the reference mirror, see top of Figure 2-1), so that a large portion of the remaining light constitutes specular reflections from the desired depth.



2-1: Spectral (Fourier) domain OCT diagram from [1]



2-2: Light paths in OCT scanning. Dashed paths are lost or filtered out by the interferometer.

Figure 2-2 shows how light paths are filtered based on optical path length (OPL). Most of the light entering the tissue from the scanner is scattered away and does not return to the scan head (yellow path, far right). Specularly reflected light that does reenter the scan head must have an OPL that matches the reference path length (black path, far left) in order to contribute to the image signal. If the OPL does not match, the light is filtered out by the interferometer (blue path, center left). Light can also scatter multiple times within the tissue and then reenter the scan head. This light will usually be filtered out (orange path, center), but can remain if its OPL happens to be the same as the reference OPL (green path, center right). In that case, the light will be added to the signal and create noise in the image.

An inherent limitation of this filtering scheme is that an interferometer can only differentiate (and reject) light with an optical path length deviation greater than the coherence length of the light source. A short coherence length, accordingly, is advantageous for reducing the amount of scattered light that is accepted by the filter and accrues as either noise or spatial blur. Spatial blur, with reference to Figure 2-2, would be not disambiguating the blue path from the black path if the two paths are close to the same length, which reduces the axial resolution of the OCT

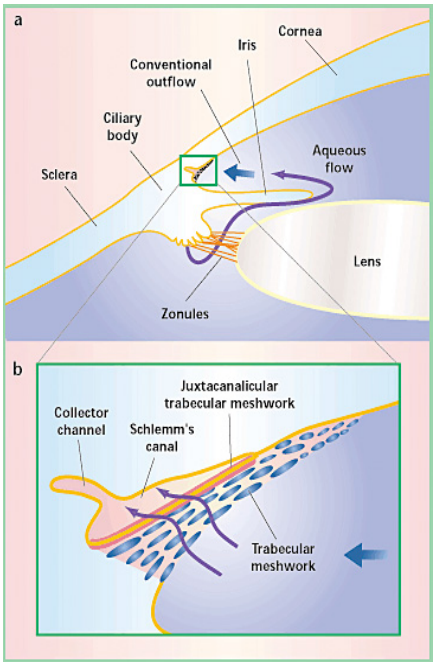
scanner. For better OCT axial resolution, a broadband light source is required, typically based on either a LED or swept source laser. (Since the light is usually channeled through single-mode fiber optic cable to a beam-forming lens, a Gaussian beam will emerge regardless of whether a laser or LED is used initially.)

A single reflected beam produces an A-scan, a column of image data representing the sample's structure through depth. 2D scanning mirror galvanometers are used to steer the OCT beam to form the images. Multiple co-planar A-scans form a 2D B-scan, in which the A-scans are typically nearly parallel. A 3D volume is formed from a stack of B-scans, or, equivalently, A-scans collected from a 2D grid of surface locations. Arbitrary planar cross-sections can be extracted from an acquired 3D volume, including C-scans (en-face planes, which lie at a constant depth) or I-scans (inclined planes, which are tilted to pass through different depths).

Our SDOCT system was jointly designed with Physical Sciences Inc. (PSI), who manufactured an initial system for us that we then further modified. This system is optimized for fast 3D volume scanning (maximum 12 bit A-scan rate of 127 kHz) at a larger-than-normal distance from the scan head. The extra space is essential for the geometry of the scan system to fit around the microscope and the working field of the surgeon. The system uses a graphical processing unit (GPU) for high-speed parallel processing of the spectrometer data and rendering the resulting image. Other systems for real time 3D scanning use dual GPU setups, with one GPU handling the processing of the A-scans, while the second GPU renders the volume [2]. As we do not render every pixel of our scanned volumes, a single GPU is sufficient. Essential to this work, the OCT system has been modified to improve the 3D scanning to allow for the extraction of 2D I-scan images to match our desired projection plane in the surgical field.

2.2 CLINICAL MOTIVATION

Our application for the device is in-situ intraoperative OCT during surgeries on the anterior segment of the eye. There are a number of clinical applications for which real time, in-situ visualization of otherwise transparent or occluded structures of the eye would be beneficial to surgeons. The primary motivating application for this project is the surgical treatment of glaucoma.



2-3: Critical structures for the drainage of the eye in the eye [3].

Glaucoma is an eye disorder that can cause loss of vision and damage to the optic nerve, and is the second leading cause of blindness worldwide [4]. Glaucoma is related (perhaps causally) to increased fluid pressure in the eye (referred to as intraocular pressure, or IOP). Reduction of IOP is a key goal in glaucoma therapy [5]. It has been shown that OCT can visualize important structures in the eye, not only macro structures such as the cornea, iris and sclera (shown in Figure 2-3, top), but also smaller ones including the trabecular meshwork and the canal of Schlemm [6] [7]. Schlemm's canal, which circles the eye at the limbus (where the cornea and sclera meet), is of particular interest to our work, as it is a main pathway for drainage for the eye. Procedures to treat elevated IOP include inserting tubes or cutting flaps in the sclera to drain the aqueous humor, inserting catheters into the canal of Schlemm, or inserting stents between the canal of Schlemm and the aqueous humor to create new drainage pathways [8]. Because the canal of Schlemm is small and runs through opaque tissue, it cannot be directly seen (even through a clinical microscope) when performing these procedures. OCT is the only noninvasive means available to see or otherwise image Schlemm's canal.

Preclinical validation of the system is included in Aim 3, but was significantly hampered by technical problems with the OCT system used to build the OCT-MARS.

2.3 AUGMENTED REALITY SYSTEMS

Augmented reality systems are a growing category of surgical guidance devices. Augmented reality refers to the integration of virtual objects with the real world, as opposed to virtual reality, which creates a fully synthetic environment [9]. In our case, the virtual object is an OCT image acquired during the surgical procedure. The implementation of an augmented reality system depends on how the surgeon is interacting with the surgical field. In open surgeries where there is no intervening imaging system, the source of the virtual object (some type of display, or an optical device such as a mirror allowing viewing of a display) is placed between the surgical field and the surgeon, and creates a translucent image. Head mounted displays create such images on a screen or through a semi-transparent mirror mounted in front of the operator's eyes, and are capable of producing stereoscopic images [9]. Projector based displays use a projector to create an image on the surface of the surgical field [10]. During laparoscopic surgeries, the field is viewed through a camera in the laparoscope, and so virtual objects may be introduced into the video stream that is viewed by the surgeon [11] [12]. The system in [11] uses a stereo laparoscope camera and registered laparoscopic ultrasound to produce a stereo augmented reality environment that is viewed through a 3D display. Augmented reality systems for microsurgery can introduce virtual objects through the microscope, most readily as a translucent overlay that is injected into the system just before the eyepieces, after the objective and primary magnification lenses [13] [14] [15].

There are currently some research prototype systems [16] and commercially available products from companies such as Bioptigen [17] that provide intra-operative OCT images. Most of these systems provide OCT images during surgery as an ex-situ view (i.e. on an external monitor). These systems are not augmented reality systems, as the additional information is not presented as a virtual object in the surgical field. However, they do represent some of the closest competitors to our system. The RESCAN system in [14] is one of only two augmented reality OCT system comparable to ours, but it suffers from a couple of key limitations. It provides OCT images through a head-up display integrated into the oculars of the microscope in such a way that the virtual overlays are displaced so as to not be coincident with the anatomy (not in-situ). The RESCAN system's images are each aligned along only one of the three Cartesian axes, and are rotated so as to be seen "from the side". Furthermore, these images are focused at infinity, and are therefore flat and always in focus, and cannot give a direct sense of depth or 3D shape. Shen et al. describes a system in [15] that uses a stereoscopic display (separate left eye / right eye images at infinity) to give the perception of depth. This method has the limitation of requiring (or assuming) knowledge of the user's pupil spacing to correctly render the image, and does not provide any accommodation cues as to depth (see below). Separate images would also have to be created for any assistant scopes or camera ports in use. Our system integrates

intraoperative OCT images in a way that differs from the state of the art systems in [14] and [15], both by our use of an in-situ *optical* virtual image, which exists at a defined depth and location [18], and by our insertion of the virtual image into the optical path of the microscope *before the objective lens*, allowing us to project that virtual image in-situ to the patient's ocular anatomy. Insertion before the objective lens is important because the microscope's optics distort the surgeons' view of 3D space, and so for the virtual object to appear in the right place it must be optically distorted in an identical fashion to the patient's anatomy, requiring that the virtual image pass through the entire optical path of the microscope.

2.4 3D PERCEPTION UNDER MAGNIFICATION

The OCT-MARS creates images that will be viewed under magnification, which inherently distorts the view of 3D space. Because these images will contain surgical targets and otherwise guide surgery, we need to examine the effects that magnification will have on the perception of our displayed images. We want to show OCT images in 3D space, along slanted B-mode and I-mode planes (and eventually also curved surfaces), and so we have developed experiments to measure how well users can perceive tilt and changes in depth.

When people view a surface that deviates from a perpendicular along the line of sight, they have available three classes of cues to judge its slant[1]: the so-called pictorial cues, which would be visible with monocular viewing; stereoscopic disparities, which are available only with two eyes; and oculomotor cues of convergence and accommodation. (Depending on viewing conditions, movement-based cues may also be available.) Considerable research on this topic has concentrated on the pictorial cues; more specifically, on how slant is signaled by gradient changes in spatial elements across a pictorial display. Gibson [19] famously used the term *texture gradient* to capture the progressive change in image statistics with distance from the viewer. Purdy, Gibson's student, attempted in his thesis work [20] to characterize gradient cues more systematically and found that several candidate properties of regular texture elements, such as principal axes and their ratios, were powerful cues to slant. In more recent research, Todd and associates [21] [22] have argued that local analysis is insufficient to compute slant from objectively 2D textures; region-based metrics are needed.

Stereoscopic cues to slant arise not from gradients in image statistics, but disparities between the retinal projections to the two eyes, which directly map to the distance of the local surface from the viewer. (The perception of metric depth, discussed here, should be distinguished from a qualitative sense of three-dimensional space, which can be achieved monocularly [23].) A fundamental problem in binocular depth perception is to establish correspondence between geometrically non-corresponding points on the two retinae that stem from a single point in space [24] [25] [26] [27]. To solve this problem from a computational standpoint, Marr and Poggio [25] made use of the constraints provided by the uniqueness of shape features and the continuity of edges. These constraints are not independent in either the 3D world or in images of it: Any image containing unique shape features such as objects, points, or corners, will necessarily contain edges: contiguous changes in an image parameter (e.g., brightness) within a region of space that are significant to human vision [28]. An image that is densely populated by edges that bound locally distinct shapes is well suited, when presented with appropriate disparities, to induce depth via stereo cues, although the stereo perception of slant in particular is further facilitated by local disparity discontinuities and relative gradients [29] [30].

Research on pictorial gradients seeks to understand how surface orientation is biologically computed in the absence of stereo disparities, which would directly convey that points on a slanted surface vary in depth from the viewer. In circumstances where judgments are made

[1] The term *slant* is used here for rotations of an image relative to the fronto-parallel plane, with the rotation axis centered on the image and oriented either horizontally (pitch rotation) or vertically (yaw rotation).

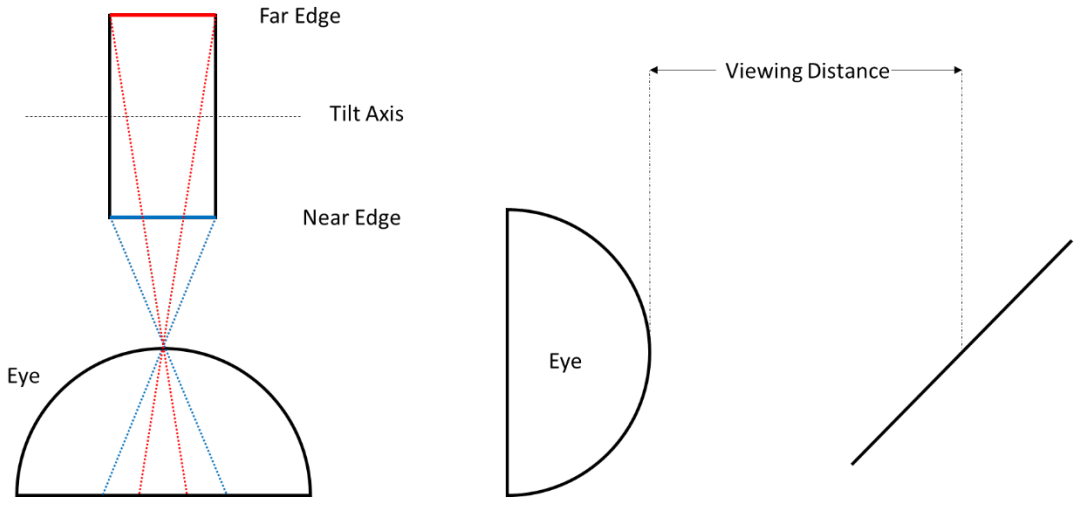
from a 2D display, this constitutes a cue-conflict situation, because pictorial cues in a 2D picture will necessarily be in conflict with stereoscopic and oculomotor cues, both of which would indicate the picture lies in the frontal plane. Even if one eye is closed, there is still a conflict between accommodation and pictorial cues in a 2D image.

Relatively little research has addressed the reverse problem, namely, how stereo disparities can cue slant in the absence of pictorial cues. Aim 2 includes the development of an algorithm for generating a novel image that would approximate statistical invariance over a large range of scales, and hence not give rise to gradient cues when slanted, while still affording stereo perception of local depth—and hence slant. In contrast to classical scale space [31] [32], in which statistics change across scale (typically monotonically due to Gaussian blurring), we are trying to maintain invariant statistics across scale. This generated image is used in the experiments in Aim 2 to investigate slant under magnification.

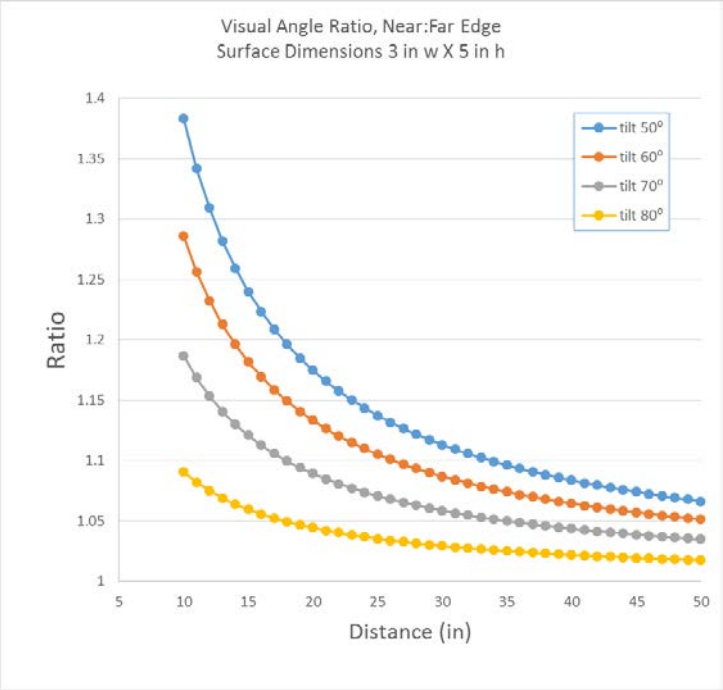
Zabulis and Backus [33] offered one approach to this problem, "Starry Night" textures, composed of points varying in flux. Pictorial cues of size and perspective were avoided by adjusting each point in the texture to a circular dot of constant radius. While this dynamic approach is possible when the displays are computer-generated, we want to produce stimuli that could be printed and presented under a microscope at different slants, in order to investigate how well slant can be perceived under conditions of magnified viewing from stereo cues alone.

Magnification affects the perspective properties of images, and magnification >> 1 approaches an orthographic projection. Under orthographic projection, parallel lines do not converge and perspective is lost, removing pictorial cues such as scaling contrast, or how the scaling of individual elements varies over an image [21]. Magnification likewise has a profound effect on the perception of tilt through texture gradients, as it "flattens" the perceived gradient, which leads to the underestimation of tilt [20].

The degradation of perspective cues is created by the optics of the microscope, which enlarges the features of images and objects, but also increases their effective viewing distance. This increased viewing distance causes image elements at different depths to have less size disparity than expected, therefore lessening the perception of distance. Figures 2-4 and 2-5 show how the appearance of a tilted rectangle changes with distance. If the tilted surface is placed close to the viewer, there is a significant difference in the apparent size of the near and far edges. This visual angle ratio decreases as the viewing distance increases (Figure 2-5). Microscopes create larger optical viewing distances through magnification, and so the change of image element size over depth is less than is expected by the viewer.



2-4: Visual angle dependence on tilt and distance



2-5: Visual angle ratio comparing the near and far edges of a rectangular planar stimulus at various tilts. A stimulus tilted at 90 degrees is fronto-parallel with respect to the viewer. A larger ratio indicates and more obvious size difference between the sizes of the near and far angles as viewed by the eye. This ratio decreases as the viewing distance increases. The telephoto lenses in a microscope render images as if they are at a very far distance

Such changes to the monocular cues of images and objects viewed under magnification can cause conflicts between those cues and the depth inferred by stereopsis (e.g. the monocular cues suggest that the image is flat, while the stereo cues suggest tilt). Understanding how these conflicting cues interact is a key motivation the present studies of the perception of slant.

Depth perception is often enhanced by accommodation, which belongs to the oculomotor class of depth cues. Accommodation allows the viewer to infer depth by the change in their eye's focal distance and the related amount of blur for out-of-focus objects in their field of view. This is a very strong cue at higher magnifications [34]. Correct accommodation has been impossible to achieve for most augmented reality systems, but our optically in-situ virtual images naturally achieve "correct" accommodation, i.e. accommodation that is consistent between the virtual OCT image and the corresponding portion of the patient's anatomy. Unfortunately, the microscope will distort the accommodation cue for the entire field of view, including both the patient's anatomy and equivalently our virtual images (our images will still match the anatomy). This distortion arises from the significantly reduced depth of field of the microscope compared to the naked eye. Large amounts of accommodation indicate significantly more depth change using the naked eye versus the microscope. This unavoidable merging of conflicting cues could lead to an overestimation of differences in depth, which is not ideal for our purposes.

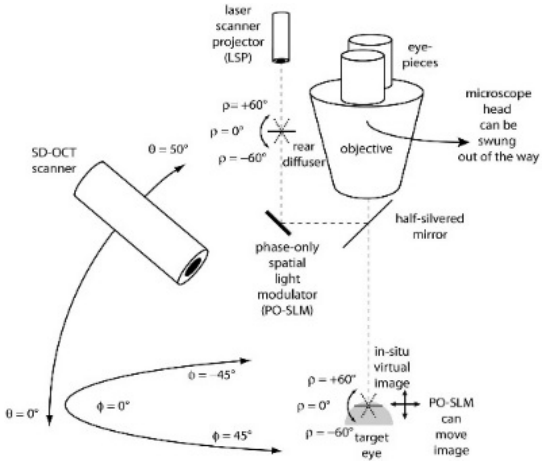
In order to provide the most useful information to surgeons through our device, we need to explore, quantify, and eventually understand how magnification affects the perception of depth. Aim 2 discusses experiments done to investigate the perception of slant under magnification.

3 AIM 1: THE OCT-MARS

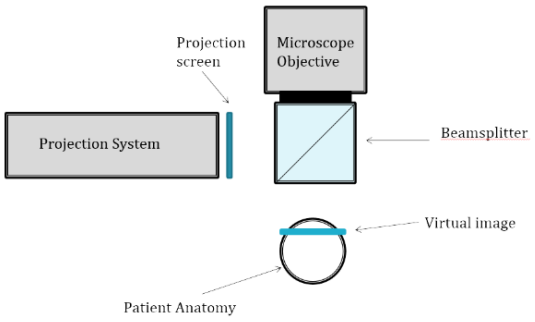
Figure 3-1 shows the layout of the OCT-MARS. A mirror is used to create a virtual image that appears to float in a defined 3D location (i.e., not at infinity). Images acquired from an OCT scanner are reflected by this mirror so that they appear in the correct anatomical location when looking through the microscope. This approach is novel compared to all other existing systems. The current version represents a significant step forward in design and function over the previous two versions developed by others in our lab and described in [35].

The OCT-MARS was implemented in two versions. The first version has a simplified image projection system (i.e. flat fronto-parallel images only). The second allows for optical repositioning of the in-situ images. As a base for our system we are using a Zeiss Lumera stereo surgical microscope, designed for eye surgery. This type of system is an ideal platform to build upon due to its popularity in surgical suites and its ability to integrate attachments (assistant scope, camera ports, etc.).

The OCT-MARS v1 (shown in figure 3-2) uses as its mirror a glass beamsplitter cube mounted under the objective lens of the microscope to insert the virtual image. The virtual image appears to float at a particular distance below the objective lens (the same as the total optical distance between the objective and the projection screen). This distance is set so that the virtual image is always at the focal plane of the microscope. The image is created using a novel projection system based on a laser scanning projector (LSP). The projector uses 3 collinear color laser beams to draw each pixel in the image by rastering the beam over the image space. A description of this system projections is published in [36].



3-1: OCT-MARS System Goals



3-2: OCT-MARS v1 Diagram

3.1 OCT SCANNER

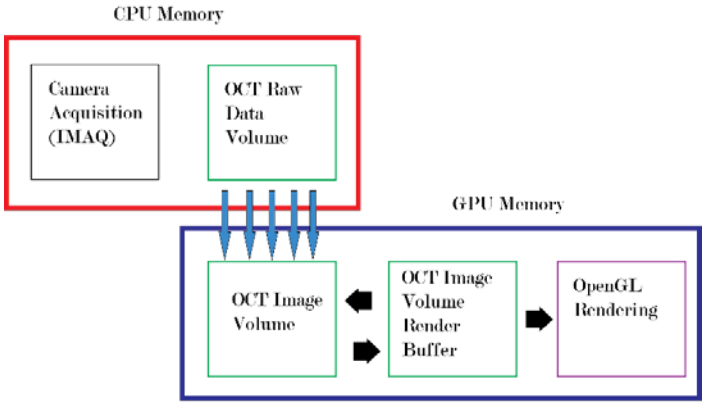
3.1.1 Hardware Basics

Our system is a spectral domain (SD-OCT) system (see Figure 2-1). Our broadband light source (as discussed in section 2.1) is a super luminescent diode (SLED, Exalos part # EXS8410-2413), which provides 50.8 nm of (3-dB) measured bandwidth, centered at 837.9 nm with a max power of 6.76 mW. The light source is divided between the reference and sample arms of the system using a fiber optic beamsplitter. Different splitting ratios were tested over the course of the project. The current setup sends 70% of the light to the reference arm and 30% to the sample. The reference arm consists of a free space (non-fiber) light path with the reference mirror set at the scan depth of the system. The sample arm sends the light though the scan head, which consists of two galvanometer mirrors (galvos) that steer the beam and a focusing lens. The sample arm light bounces off the tissue (the patient eye in the case of our

system), and returns to the optical system. The light beams returning from the tissue sample and the reference mirror are recombined to form and interference pattern (Michelson interferometer). This interference pattern is acquired by the LABView system using a diffraction grating and a line camera, and is analyzed to determine the reflectivity of the sample.

3.1.2 Real time 3D volume scanning

The projection system receives its source image from our custom OCT scanner (designed in partnership with PSI), also attached to the stereomicroscope. As described above, our scanner is optimized for high frame rates and a large working distance to accommodate the surgeon and tools. This, unfortunately, is at the expense of resolution and signal-to-noise ratio, which we have found are particularly problematic with our custom OCT scanner (discussed below in Section 3.6). The projection system must be calibrated so that the virtual image correctly overlaps the underlying anatomy.



3-3: OCT real time diagram

The original PSI OCT system was designed to provide high frame rates, but was not optimized/designed for continuous real time 3D volume acquisition and display. I took the lead role in supervising an undergraduate programmer to significantly modify PSI's system to provide these capabilities. Figure 3-3 shows our adapted 3D volume rendering scheme. Raw OCT image data are queued to allow processing without dropping frames from the current volume. The processed volume is then double buffered to allow for concurrent acquisition of processed frames and visualization of the previous volume.

3.1.3 Hardware Modifications

A number of modifications were made to the custom OCT scanner system to improve performance. These are in addition to hardware revisions and repairs discussed in section 3.6. A major change was the replacement of the reference mirror (see Figure 2-1). The PSI system was built with a protected gold first surface mirror, which provided adequate reflectance of the reference arm beam, but was difficult to align. Small deviations of the mirror angle caused the reference beam to not return to the spectrometer. This was replaced with a protected gold retroreflector. Unlike mirrors, retroreflectors bounce light beams directly back at the source making alignment of the system much easier. This change was a practical necessity for the constant adjustments and repairs required to the optical path of the scanner.

For continuous synchronization of volumes, we added a hardware trigger signal for the LabVIEW interface. This trigger is based on the voltage level of the slow galvo of the scan head, and it signals the beginning of each new volume. This trigger was custom built in our lab, and we modified PSI's LabVIEW software to synchronize based off of it, achieving correct, continuous volume acquisition.

3.1.4 Software modification – LabVIEW

As it was provided to us, the LabVIEW system could scan and save volumes, but not in real time. We added the capability to buffer full volumes to ensure that no frames are dropped, since image processing can take longer than image acquisition (eventual optimized software should be able to process images at the acquisition rate). Each volume acquisition begins on the signal of the hardware trigger. As raw frames are acquired, they are sent to the GPU for processing. Once a full volume of raw images is acquired, acquisition is paused until all frames have been processed and the resulting display image is returned. The system then waits for the next volume trigger and repeats.

3.1.5 Software modification – C++/CUDA

While LabVIEW handled the image acquisition and control of the scanner galvos, the bulk of the data processing and analysis was done with C++ and NVIDIA's Compute Unified Device Architecture (CUDA) to allow for GPU acceleration. The graphics card in the PSI system was upgraded with a state of the art NVIDIA Tesla workstation GPU. The CUDA programming for this project was done with significant collaboration with fellow graduate student Tejas Mathai.

The processing proceeds as follows: As images are received from the LabVIEW interface, the spectral data is processed (inverse Fourier transform) to obtain the reflectivity data that makes up the OCT image. Full volumes are acquired and can be "sliced" to extract I-mode images that are on tilted planes with respect to one axis (but presently not yet two axes) of the scanner's coordinate system. The slicing is done by computing the angle of the slice plane with respect to the coordinate system of the acquired volume, and accumulating the image rows in the plane to form a single image. Nearest neighbor interpolation is used to smoothly fill the sliced image. This ability is necessary if the projection plane of the OCT-MARS is not coplanar with the B-mode images.

In addition to volume acquisition and slice extraction, the CUDA framework is ideal for image analysis. Significant work on layer detection in human corneas has been done by fellow graduate student Tejas Mathai using the OCT-MARS scanner system [37].

3.2 LASER REFOCUSING SYSTEM

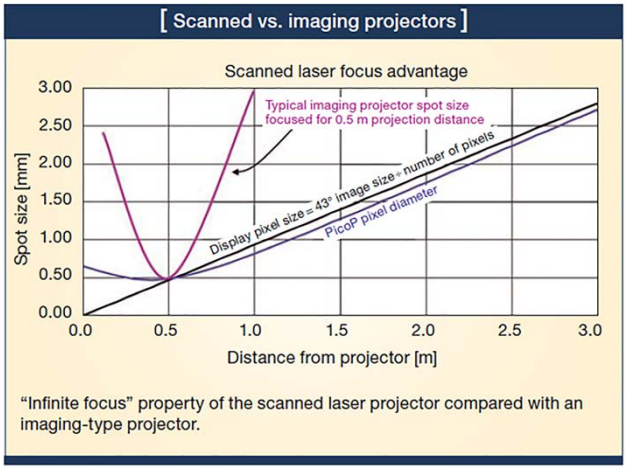
3.2.1 Overview

Although beyond the goals of the present dissertation, our laboratory aims to eventually incorporate a phase-only spatial light modulator for electronic focusing based on diffractive optics. To accommodate this long-term goal, we require a narrow bandwidth light source to allow the projection system to be used with such subsequent diffractive optics. We also require very small and very bright pixels. Small pixels are necessary to maintain image resolution after magnification. Bright pixels are necessary to withstand the light dilution caused by magnification. Laser based projection systems have the ability to fulfill all three requirements. Laser beams can be focused to very small, very bright spots, and in most cases they have relatively narrow bandwidth (often practically monochromatic). However, commercially available laser based projection systems (i.e. pocket projectors for giving presentations and watching videos) are typically designed for creating images viewable at much larger sizes than appropriate for magnification, and thus some modification is required for them to project tiny images. This section addresses the general issue of manipulating these types of laser projectors to control the width of both the individual laser beams (pixels) and the envelope that contains them (image). This problem is central to the design of a projection system suitable for our microsurgical augmented reality system.

3.2.2 Microvision Projector

For this project, we used Microvision's PicoP projection system, which combines red, green, and blue Gaussian-beam lasers into a single beam that can produce a full range of colors,

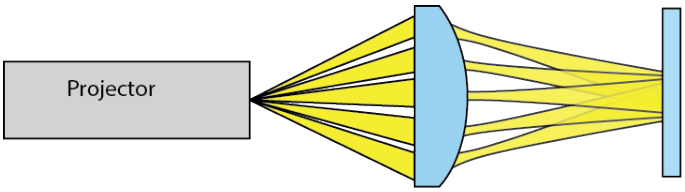
projecting an image on any surface (typically some sort of diffusing screen) by sweeping the laser beam in a sine wave raster pattern using a biaxial microelectromechanical systems (MEMS) mirror [38]. The combined laser beam is not fully collimated, as it must diverge slightly so that the pixels fill the image as the image expands (see Figure 3-4). In order to have the image and pixel sizes match at the manufacturer's target distances (0.5 m and further), the combined beam is internally focused to a waist at about 0.5 m in front of the projector [38]. After this point, the beam radius and desired pixel size are in good agreement, leading to the projector's "infinite focus" property, i.e., unlike traditional planar image source projectors it does not require focus adjustment. However, for very small image sizes (projected much closer to the projector than 0.5 m), the pixels are significantly oversized and overlapping, causing the image to blurred (Figure 3-4). Therefore, to get a small image from the projector, we cannot simply place a diffusing screen at such a close range.



3-4: Pixel size vs. distance from projector for the Microvision PicoP projector [24].

3.2.3 Single Lens Focusing

As we will show, for our desired range of pixel sizes (0.01 to 0.1 mm pixel radius) an "in focus" image cannot be produced by combining such a projector with a single lens. By "in focus" we mean an image in which the actual pixel size is not larger than the pixel spacing. A single lens will focus the laser beam to a waist near the back focal plane, but since the image cone is diverging entering the lens, it converges more slowly exiting the lens, forming a diffraction-limited spot past the back focal plane. (In comparison, a collimated image entering the lens will be condensed to a diffraction-limited spot at the back focal plane.) Increasing lens power to converge the image closer to the lens will, of course, also move the beam waist even closer to the lens, and so the image can never focus coincident with the beam waist (see Fig. 3-5). The light from the projector effectively behaves as two separate but simultaneous image systems, defined as follows: (1) The image cone is the envelope containing all of the light within the image. The width of the image cone at a given location along the optical axis determines the image size and the *desired* pixel size. The cone and the desired pixel size behave according to geometric (not Gaussian beam) optics. (2) The laser beam's profile behaves according to Gaussian beam optics, and the actual pixel size is determined by the radius of the laser beam at a given location along the optical axis.



3-5: Attempting to focus a laser-based projection system with a single lens. The projection screen and lens are positioned to project a small image, but the laser waist is unavoidably nearer to the lens, resulting in large and overlapping laser beams (i.e., pixels) on the projection screen.

3.2.4 Multiple Lens Focusing

In order to simultaneously control the location of the laser beam's waist and adjust the image cone to the desired size, we need more degrees of freedom to independently control the two systems. We have developed what we believe to be a novel application of a two-lens system to resize the image. We have published this work in [36]. As shown in Figure. 3-6, the first lens is selected and aligned to collimate the image cone. This also focuses the laser beam to a waist slightly past the back focal plane. The beam divergence is greatly increased, so while the image remains the same size along the collimated region, it becomes increasingly blurred as the pixels become larger. A second lens is then placed to refocus the laser beam and resize the image cone. The image cone is focused to a spot at the back focal plane of the second lens, while the laser beam is focused to a waist past the focal plane, near the desired image plane. As the distance between the two lenses increases, the waist approaches the back focal plane.

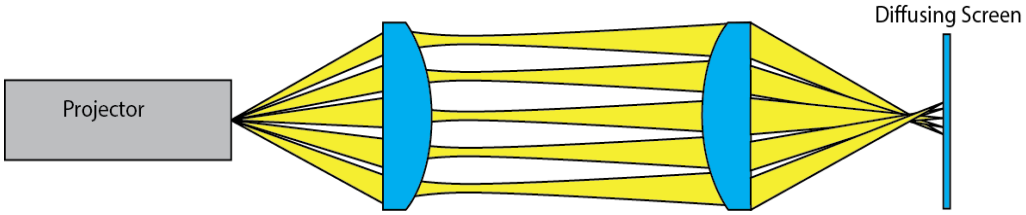
The third and final element of the system is the translucent diffusing screen, which is placed so as to create the in-focus image after the second lens. The distance between the second lens and the diffusing screen determines the intersection with the image cone, and thus the overall image size. Placing the diffusing screen closer to the back focal plane of the lens will, of course, produce a smaller image.

To quantify the behavior of the system with different lens arrangements, we modeled the behavior of the laser beam using Self's equations for beam propagation, which assume paraxial conditions [39]. Without loss of generality, the focal lengths of the first and second lenses are set at 30 mm and 40 mm, respectively, a physically realizable setup suitable for our purposes. The location of the new beam waist after each lens is given by Equation 1:

$$\frac{1}{s+\frac{Z_R^2}{s-f}}+\frac{1}{s'}=\frac{1}{f}$$

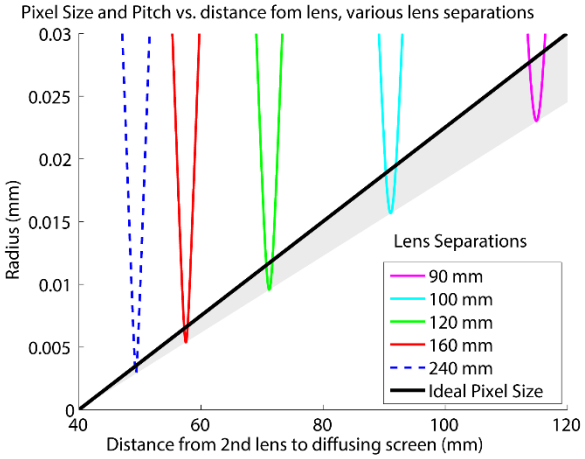
Equation 1

where s is the distance from the lens to the input waist, f is the focal length of the lens, s' is the distance from the lens to the output waist, and Z_R is the Rayleigh range of the input beam. Using Self's equations, it is straightforward to determine the waist size of the output beam and the output Rayleigh range [39]. The behavior of the image cone was modeled using geometric optical properties. The angle of the cone after the second lens is determined from the equivalent focal length (EFL) of the lens and the size of the cone entering the lens. This angle determines the size of the cone after the focal point. The ideal pixel size is the appropriate fraction of the cone size, as determined by the pixel-count of the projector. Figure 4 shows how the pixel size and in-focus image location vary with the distance between the two lenses (90-240 mm).



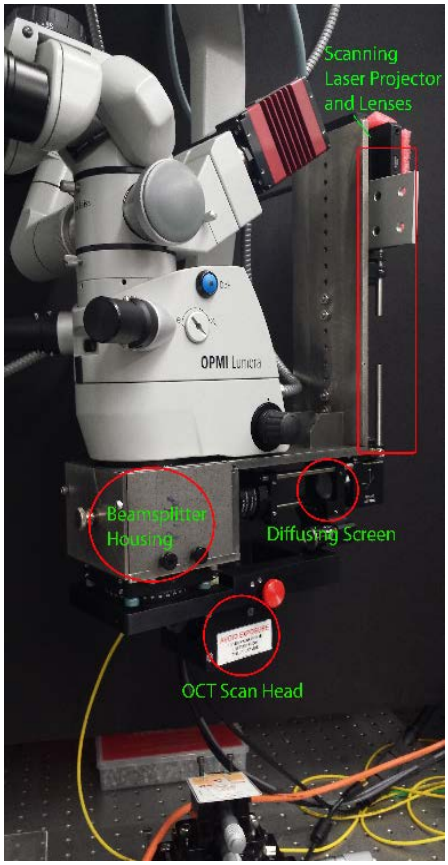
3-6: Projection System Lens Diagram

As the distance between the two lenses increases (increasing the beam size entering the second lens), the resulting beam-waist radius decreases, and the beam waist moves closer to the back focal point of the second lens. The solid black line in Figure 3-7 shows the "ideal" pixel size that corresponds with the image cone size at that location. The intersection between this line and a particular beam radius curve shows ranges at which the final image is in focus for the corresponding distance between the two lenses. The beam waist curve intersects the ideal pixel line in two places. The space between the intersection points of the beam radius curve and the pixel size line defines the depth of field over which the image will be in focus. Pixels that are somewhat smaller than the ideal pixel size would not typically cause a significant problem, since the blurring of subsequent optics, or even just the human eye, is likely to cause the spaces between pixels to be imperceptible. This depth of field decreases rapidly as the inter-lens distance increases and the beam waist radius decreases. Depth of field over all image sizes could be improved by increasing the focal length of the second lens, but doing so requires a larger diameter lens placed further away from the projector.



3-7 : Compare to figure 3-4 each inter-lens distance will result in a different beam profile. Only the portion of a profile that falls beneath the diagonal (indicated in gray) will result in non-overlapping "focused" pixels, determining the acceptable range of image distances (and thus image sizes) for that lens arrangement. Smaller separations between the lenses correspond to larger beam waists further from the back focal point, and larger depth of field.

3.3 OCT MARS VERSION 1



3-8: OCT-MARS v1 Assembly with Notes

Figure 3-8 shows the assembled OCT-MARS v1. The system is attached to the underside of the Zeiss Lumera microscope using an Invar steel backbone. Invar is a special alloy used to reduce the thermal expansion/contraction of the system, in order to maintain the alignment of the optics whether in a chilled or room-temperature operating room. The beamsplitter cube is housed in a closed box to reduce interference from ambient light. Additional neutral density filters are inserted after the diffusing screen to selectively reduce stray reflections from the illumination source within the microscope, which will pass twice through these filters. The OCT scan head is mounted below the beamsplitter box on a rotation stage that allows the scan head to be rotated in azimuth around the target.

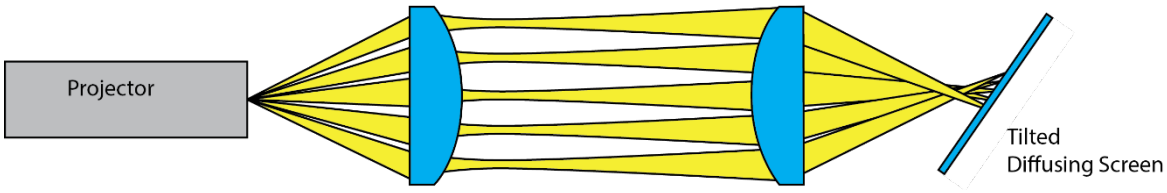
3.4 OCT MARS VERSION 2

Version 2 of the OCT-MARS (shown in Figure 3-9 and Figure 3-10) allows the manipulation and repositioning of the images projected through the microscope. In particular, it allows the tilt of the image displayed to be adjusted from orthogonal (to no longer appear fronto-parallel) using a swiveling projection screen. The version described here forms the basis for a system that would eventually be motorized in later iterations to permit real-time selection of any given plane through the data volume.

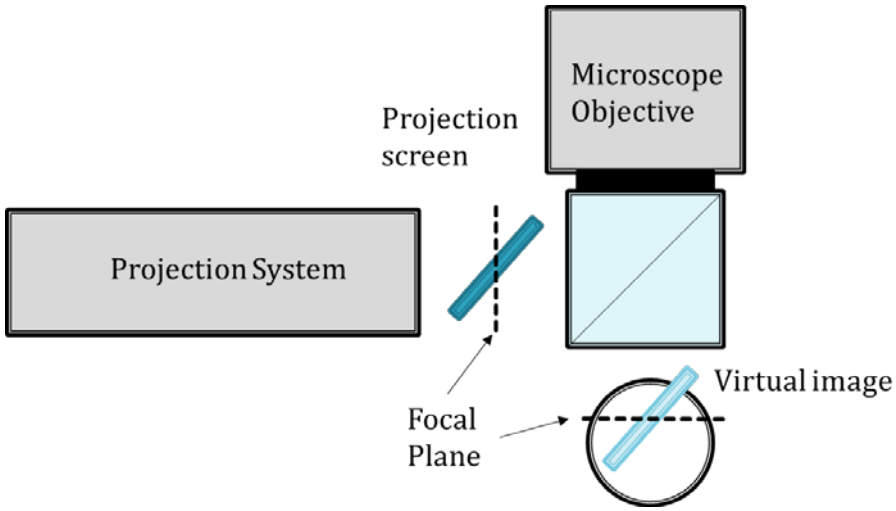
3.4.1 Design

Tilting the projection screen (as shown in Figure 3-9) creates an image that physically exists along a tilted plane with respect to the optical axis of the microscope. Portions of the image on this screen will be out of focus because the screen deviates from the image plane defined by the rest of the projection system. However, the image on the screen is designed to be viewed under relatively high magnification through the microscope, which has its own focal plane and depth. Structures and image content too far away from the microscope's focal plane (i.e., located

outside of the microscope's *depth of field*) will appear blurred regardless. We theorize that the blurring resulting from the microscope's shallow depth of field will be much larger than the blurring from the depth of field of the projection system. If the projector's depth of field is sufficiently larger than the depth of field of the microscope, then blurring due to tilting the projection screen will appear inconsequential.



3-9: Tilted projection system



3-10: OCT-MARS v2 diagram. Tilting the screen has the exact same departure from the focal plane of the projection system as from the focal plane of the microscope.

3.4.2 Depth of Field Analysis

The depth of field of the microscope is dependent on the numerical aperture of the system, the wavelength of light, and the resolution of the detector used to observe the system. For simplicity in our analysis, we use the resolution of the camera (AVT GT1920) that was used to acquire images for this thesis to determine resolution of interest. The physical resolution of the detector is important, as lower resolution systems can tolerate more blurring without losing information, leading to a larger depth of field. The numerical aperture of the system (or NA) is dependent on the magnification and aperture. The numbers used in our analysis are those for the f = 200mm Zeiss Lumera microscope maximum (24x) and minimum (4x) magnifications. Table 1 gives the specifications for the high and low magnifications.

Table 1: Specifications for the camera and microscope at high and low magnifications. Camera magnification our setup is 0.7x the microscope magnification.

Microscope Magnification	Camera Magnification	Camera Pixel Size (µm)
4	2.8	28.0
24	16.8	4.59

The numerical aperture of the camera mount is given by Equation 2, where N is the f-number of the optical system and m is the magnification. Camera images were acquired with an f/11

aperture, as this produced images with a qualitatively similar depth of field to viewing through the eyepieces.

$$NA = \frac{m}{2N(m-1)}$$

Equation 2

Equation 3 gives the formula for depth of field based on the magnification and numerical aperture:

$$DOF = \frac{\lambda \cdot n}{NA^2} + \frac{n}{m \cdot NA} \cdot e$$

Equation 3

where λ is the wavelength of the light source (537 nm, the wavelength of the green laser, is used for our calculations), n is the refractive index of the medium (air, n =1) and e is the camera resolution (dependent on magnification).

The depth of field is defined differently for our projection system. As we saw in Figure 3-7, the pixel size generated by our refocusing optics can be compared to the ideal pixel size (where pixels are non-overlapping and the full resolution of the laser projector is maintained). We consider the system to be "in focus" if the actual pixel size is less than or equal to the ideal pixel size. The depth of field is then the range of locations for the diffusing screen over which this is true. Table 2 shows the depth of field for the lens separations noted in Figure 3-7. The highlighted separation is the one used for the implementation of our system. The table also includes the depth of field calculated from the high and low magnification of the Zeiss microscope. We can see that the depth of field of the microscope is sufficiently smaller than that of the projector system for our range of magnifications.

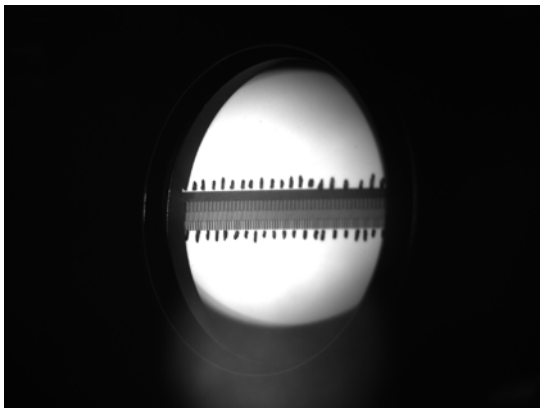
Table 2: Depth of field values for the stereomicroscope, and various projector designs. The highlighted design is used in the OCT-MARS

Optical System	Depth of Field (μm)
Microscope @ 4x	777.3
Microscope @24x	298.1
Projector (90mm)	2400
Projector (100mm)	1720
Projector (120 mm)	1000
Projector (160mm)	570
Projector (240 mm)	300

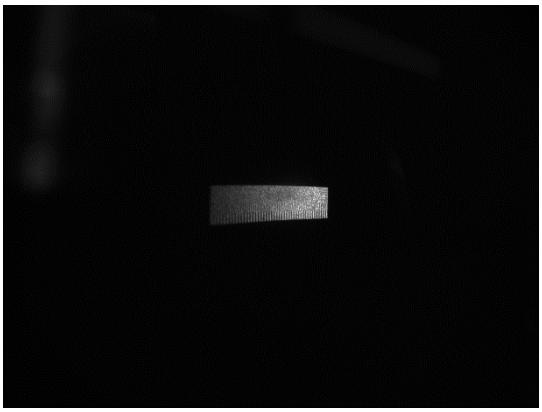
3.4.3 Verification

Qualitative verification of image focus was done using printed and projected versions of a resolution grating. The printed version represents an "ideal" projected image that has an infinite depth of field. Viewing the printed image at a tilted angle through the microscope allows us to qualitatively estimate the blurring caused by the microscope alone. If the hypothesis is correct that the scope blur overpowers the projector blur, the projected and printed gratings should have similar resolution limits when viewed. Figures 3-11 through 3-16 show the results of the qualitative verification at the high and low magnifications. At the high magnification, we can see that the printed and projected images have comparable blurring, indicating that most of the blur is the result of the microscope and not the tilt of the projection system. At the 4x magnification, we can see that the effect of the larger depth of field is mitigated by the reduced resolution of the system. However, the enlarged image of the projector pattern at 4x (Figure 3-16) does show some additional blurring over the printed version, indicating that the depth of

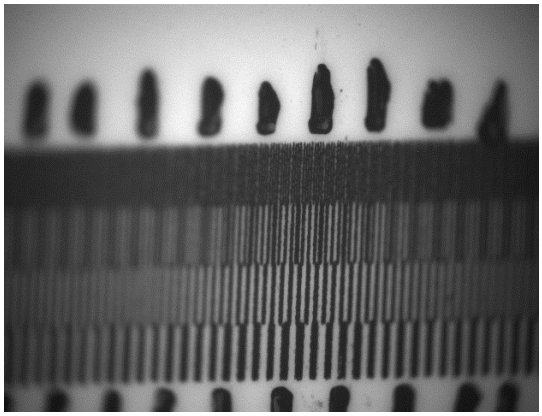
field of the projector is slightly too narrow at that magnification. However, this blurring does not noticeably effect the image as viewed through the eyepieces (Figure 3-12). Therefore it can be concluded that the tilting projection system gives a sufficiently in-focus image for the purposes of our implementation.



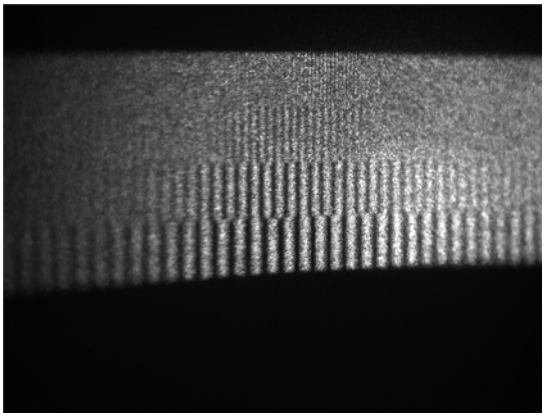
3-11: Printed image tilted to 45 degrees, viewed at 4x



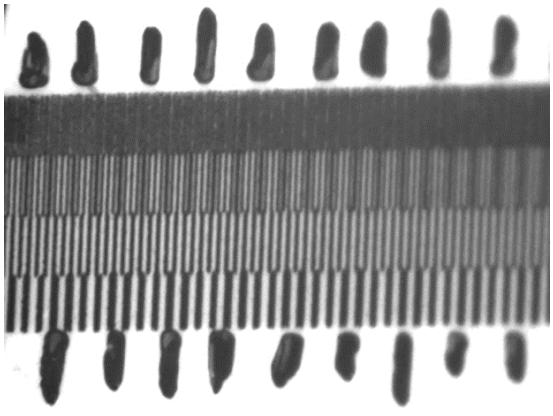
3-12: Projected image tilted to 45 degrees, viewed at 4x



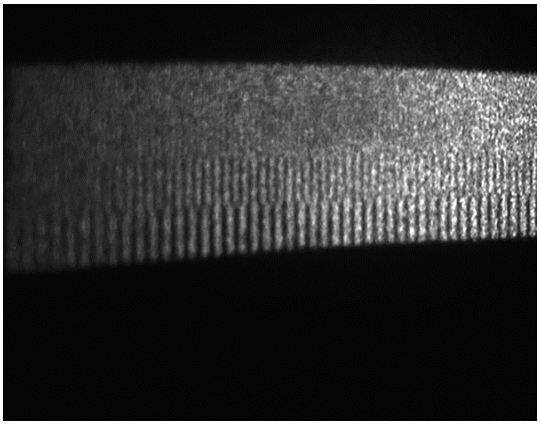
3-13: Printed image tilted to 45 degrees, viewed at 24x



3-14: Projected image tilted to 45 degrees, viewed at 24x



3-15: Printed image tilted to 45 degrees, viewed at 4x, enlarged to match 24x



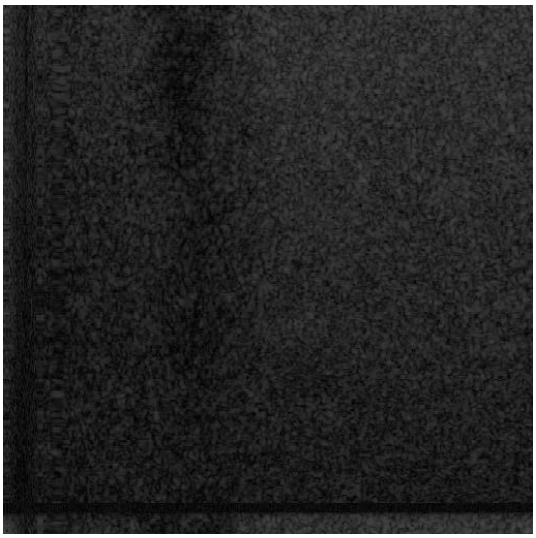
3-16: Projected image tilted to 45 degrees, viewed at 4x, enlarged to match 24x

3.5 CALIBRATION

Calibrations procedures were required for both version 1 and version 2 of the OCT MARS. As described below, the calibration consists of two steps: (1) the correct OCT image is selected from the data acquired by the scanner, and (2) the image is correctly aligned within the plane of the projection system.

3.5.1 Step 1

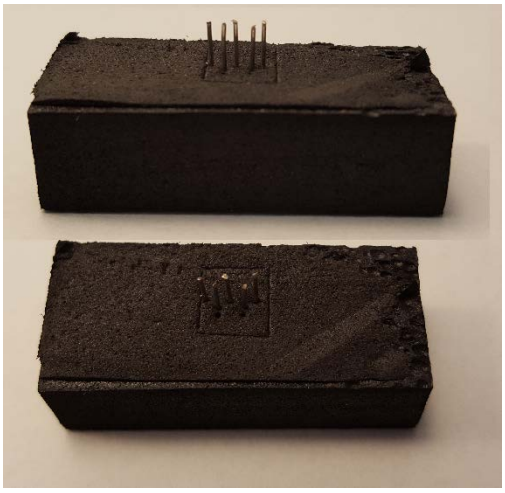
Step 1, which consists of extracting the correct OCT slice from a volume, is only required when the scan plane of the OCT scanner is not aligned with the projection plane. In that case, the desired slice for projection must be extracted from multiple slices in the volume. This extraction is done by pulling rows of OCT data from each image that are contained in the slice plane (Figure 3-17). These rows are identified in a calibration step where the scanner acquires a volume that contains a bright plane that lies within the desired projection slice. This highly reflective plane is easily identified by a simple maximum intensity search. The location of the bright calibration plane gives the optimal slice extraction for subsequent volumes. This step is not necessary for the OCT-MARS v2, as the projection and scan planes are parallel.



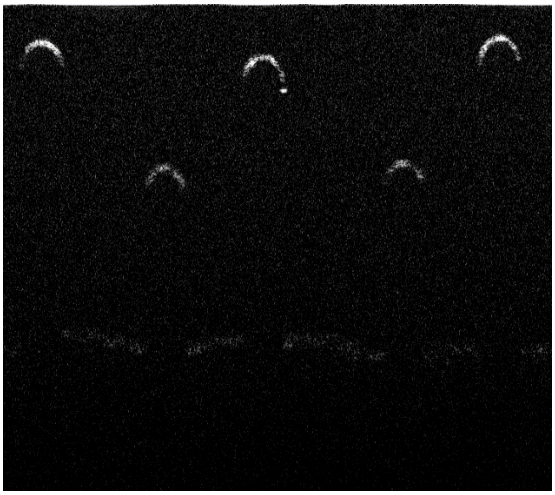
3-17: Example of an interpolated image slice. The image shows a flat plane imaged by the scanner at an angle.

3.5.2 Step 2

Step 2 takes either the B-scan or the extracted I-scan, and aligns it for projection in the microscope. This alignment accounts for any in-plane offsets or rotations between the scanner and the projection system. The alignment is done by performing an affine transform from the image space of the scanner to real world coordinates. This affine transform is computed by scanning a calibration phantom (Figure 3-18) that provides 5 distinct anchor points, visible in OCT (Figure 3-19). These anchor points are manually labeled in the projected images, and the transform between the two sets (real world points and projected points) is used to compute the affine transform. A least squares fit is used for increased reliability.



3-18: Five pin phantom used for calibration



3-19: OCT scan of the five pin phantom

3.6 LIMITATIONS OF THE EXPERIMENTAL OCT SYSTEM

The OCT scanner is a unique, custom system developed in partnership with PSI. As the first OCT scanner to our knowledge with such a long working distance, some aspects of device performance are themselves experimental. This research has uncovered some limits of the novel scanner, and for a few of these has also established a theoretical hypothesis. Our efforts going forward were also hampered in part by physical damage (see below) and by the departure from PSI of two key individuals involved in this project. A comparison between the performance of our machine and other scanners can be found in Figures 5-6 through 5-9 in section 5.3.

3.6.1 Initial System Damage

Regrettably, the device was damaged while in transit between PSI and CMU, requiring immediate replacement of the galvo driver boards. We later realized that additional optical components had been partially damaged, leading to eventual replacement of both the fiber optic beam splitter and the optical circulator. The prior damage repeatedly made it difficult to troubleshoot various aspects of system performance, and remaining problems are now believed to be a result of damage-induced internal misalignment within the spectrometer, leading to excessive signal fall off with depth. Each time a component began to fail, significant delays were introduced into the research by troubleshooting and lead times acquiring replacement parts. PSI provided partial support for the initial issues with the control board, but since then we have been left to take on responsibility. Thus far the system's SNR still falls off too sharply with depth, despite the best efforts of our lab.

3.6.2 Issues with PSI support

The software provided by PSI was not sufficient for the contracted purpose, was not easily modified by our lab, and had significant instabilities. Significant effort was required to make 3D volume acquisition function in real time. Little support was available for software modifications or hardware repairs. Lab members had difficulty diagnosing system problems due to scanty and poor documentation.

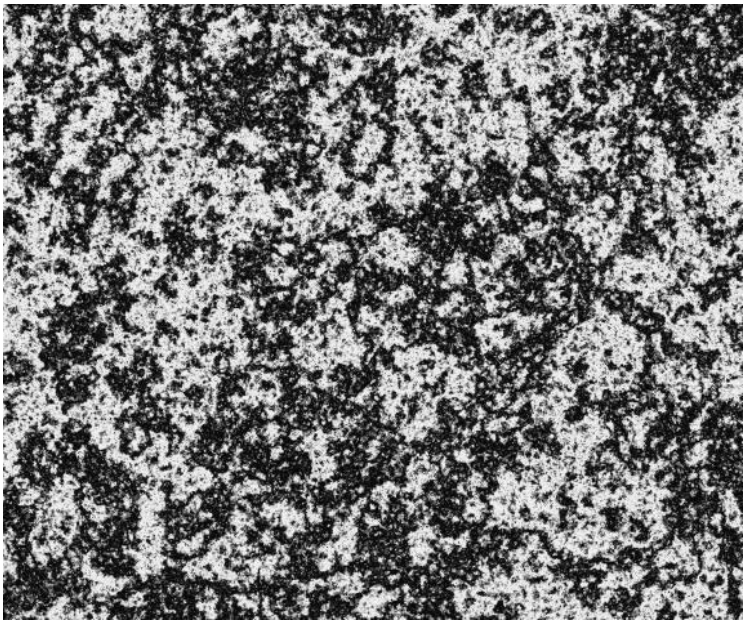
3.6.3 Theoretical Concerns

We suspect that the scan head's long working distance negatively affects resolution more than expected. The longer working distance (100 mm) and the deeper imaging range (6 mm) require a longer focal length and a wider beam waist. This combination would capture a larger percentage of scattered light. Since OCT cannot filter out 100% of scattered light, the SNR would be reduced. We hypothesize that the extra received scattered light causes a substantial decrease in SNR. We replaced the scan head optics with otherwise equivalent lenses to achieve shorter working distances. Empirically, the shorter the working distance we used, the better

our images looked, with SNR qualitatively but clearly increasing as working distance decreased. To achieve clinically relevant images with an acceptable form factor, we plan to eventually create a new scan head with an optical relay tube that can position a small lens closer to the patient while keeping as much of the scan head bulk as possible further away from the operating area. But for the present, reduction in SNR combined with the persistent hardware and software stability issues have precluded clinically relevant operation of the present device.

4 AIM 2: SLANT PERCEPTION IN MICROSCOPY

In addition to building the OCT-MARS, we have worked with collaborators to study how users perceive 3D shape and visual texture under magnified conditions. We are investigating stereo perception under magnification by attempting to quantify how fine a tilt angle can be resolved by subjects looking through a stereo microscope at a planar image (analogous to an in-situ tilted OCT slice viewed under magnification). One type of image we show comes from a class of special textures we created (shown in Figure 4-1) that provide stereo correspondences, but no pictorial cues, to aid in our studies by disambiguating the effects of different perceptual cues. OCT images displayed through our system will likely lack structure that gives strong perspective cues (parallel lines, right-angle corners, etc.), and so there is practical significance and scientific necessity to quantify how well users perceive tilt angle based on stereo only.



4-1: Our generated scale-invariant stimulus without perspective cues

The near-orthographic projection produced by the microscope has a significant effect on the perception of structure under the microscope. Parallel lines converge much more gradually than the user expects, causing a possibly systematic underestimation of the tilt angle for such structures. Scale-invariant textures (such as the one shown in Figure 4-1) allow us to investigate the performance of stereo tilt perception without pictorial cues. Our studies show that depth from stereo, and thus determination of tilt angle, is adequately perceived under stereo microscopic conditions to justify our in-situ display.

Aim 2 consists of four distinct parts: the design of our scale-invariant texture and three human subject experiments to analyze performance with that and control surfaces and the general ability of participants to perceive tilt. Experiment 1 analyzes our custom texture to determine if it achieves the goal of providing stereo cues without pictorial cues. Experiment 2 investigates the ability of humans to perceive the slant of a surface viewed under magnification both with and without stereo and pictorial cues. Experiment 3 tests the ability of subjects to match a tool to a slanted surface in a variety of conditions, including surfaces textured with real OCT data.

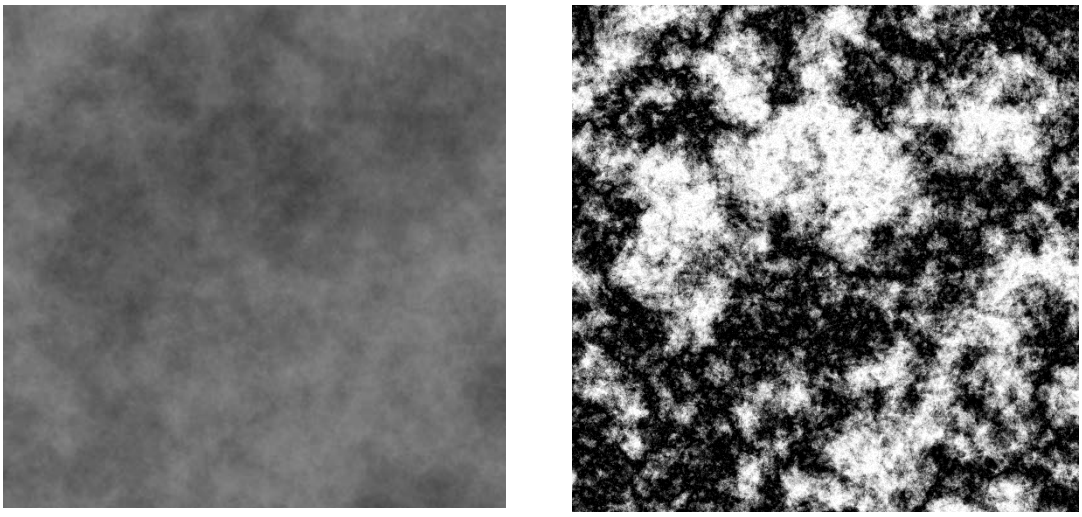
4.1 SCALE-INVARIANT TEXTURE GENERATION

We have developed an algorithm for generating images with three critical properties: First, as the detection of stereo correspondence is facilitated by the uniqueness of shape features and the continuity of edges (e.g., Marr & Poggio [25]), there should be dense edge content. Second, in

order to isolate the contribution of stereo cues, perspective changes under slant should be avoided (although foreshortening alone may not be sufficient to induce slant perception [40]). Third, given the known importance of gradient changes for pictorial slant perception, perceivable edge density should remain constant as the image is rotated relative to the viewer's line of sight. At a local level, rotation can be treated as a change in scale. Thus a critical feature of the generation algorithm, especially given our interest in slant under magnified viewing, is that the statistics of the resulting images are not indicative of viewing scale. In particular, as will be discussed later, the algorithm is constructed with the goal that the edge statistics (i.e., statistics of high-frequency content) are scale-invariant, while texture gradient information is suppressed by perturbation of low frequency content.

The goal of scale invariance suggests the use of a fractal image, which is scale-invariant by definition. However, fractals are constrained only to have equivalent non-regularity across scales, and as is exhibited by well-known fractal sets [41] [42], sub-regions can have relatively sparse, object-like shapes, the density of which changes with rescaling. The approach taken here was to determine a means of generating images with the desired properties *de novo* and then to test the extent to which they cued slant. Under the goal of scale invariance, the generated content consisted of overlapping edges with sufficient density that when zoomed in, edges that nearly spanned the original scale disappeared, but were replaced with newly emergent edge content that left the statistics unchanged. (These newly emergent edges are always in the generated image, but given the resolution limits of either a human eye or a computer monitor, they are too small to see when zoomed out.)

The scale-invariant texture shown in Figures 4-1 and 4-2 was designed by our lab in order to differentiate between pictorial cues and stereo matching, as both can affect 3d perception. The design and testing of this texture is detailed in [43]. This work is central to the following experiments, and was primarily done by an undergrad under my supervision, with assistance from other graduate students.



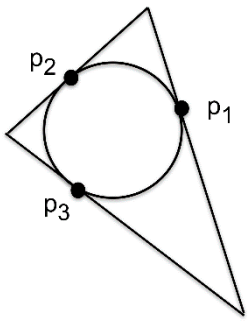
4-2: Scale-invariant image, before and after wrapping

4.1.1 Implementation

In general, if a source image is viewed within a rectangular aperture, zooming in by a factor of k will reduce the viewed source area to $1/k^2$ of its previous value, while the source data that are currently in view will now be assigned k^2 its previous number of pixels. Maintaining invariance under zoom thus requires k^2 times as many emergent edges in the entire image as there are vanishing edges.

To achieve this goal, the algorithm additively superimposes a large number of randomly shaped and oriented, overlapping triangles, where the scales of the triangles are distributed so as to maintain the inverse-square law. Shapes are used as the route to edges, because a directly generated edge is a 2D shape (albeit very small on one of the dimensions) that will change in size as it is rescaled. Triangles, in particular, were chosen as edge generators because the literature on slant from texture indicates certain shapes that are precluded *a priori*, particularly circles and rectangles, which will create undesirable perspective cues in the image, as well as shapes with constant parameters such as ellipses with a common axis ratio, and even families of blobs with underlying statistical regularities.

The process of generating a single triangle is illustrated by Figure 4-3: A circle (not rendered) of radius r was placed at a randomly selected point, three points were randomly selected around the circle, and tangent lines were drawn to these points, with the intersections constituting the three vertices of the triangle. Additional constraints were placed on the configuration so that the tangents intersected within a reasonable distance, avoiding near-parallel edges: Relative to the first point of tangency (p_1), p_2 was restricted to angular distances in the range <(15°-165°), (195°-345°)>, and p_3 must fall on the larger arc formed between p_1 and p_2 such that (1) the resulting triangle will contain the inscribed circle and (2) p_3 will not fall within 20° of p_1 or p_2.



4-3: Triangle generated from three tangent lines. The circle is not rendered; see text for description.

The scale of a triangle is defined as the radius of the largest circle that can be inscribed within it. To produce a distribution of scales, radii of circles used to generate triangles were drawn from a distribution of sizes according to

$$r = \frac{a}{\left(\frac{a}{r_{min}} - \gamma\right)}$$

Equation 4

where *a* is defined as:

$$a = \frac{1}{\frac{1}{r_{min}} - \frac{1}{r_{max}}}$$

Equation 5

In Equation 4, r represents the radius of the circle to be plotted, and r_{min} and r_{max} are the limits of the radii allowed. The variable γ is drawn from a uniform distribution between 0 and 1. Equation 4 achieves the goal of biasing the triangle distribution toward small scales.

To localize the triangles within the image, the centers of generating circles were placed randomly over its area. An extension on the boundaries was included in the computation as needed to generate triangles centered beyond them, so as to create uniform content, even at the extremities. The brightness values were then determined by additively superimposing the triangles, according to the rule that each pixel's final intensity equals the number of triangles covering it. Fifteen million triangles were sufficient to generate a "source image" at a resolution of 45,600 pixels wide by 31,200 pixels high with a broad range of intensities.

The source size was selected to allow printing at a large (12+ inch) size. The large source image also allows us to cut smaller patches that can be used to create experimental stimuli, either by printing or presenting on a monitor. In order for the displayed stimulus to exhibit the properties intended by the generation algorithm, its pixilation must be imperceptible to the observer under all viewing conditions across our range of slant. For our experiments, the monitor pixel size and viewing distance must be implemented so as to ensure that this goal is met. Another important constraint that applies to printed and computer-generated stimuli is that the pixel size of the source image must be as small as or smaller than the display pixel size. Higher-resolution source images are also advantageous to reduce quantization error when down-sampling. For the present studies, the goal of an ample margin of safety for pixel resolution and quantization, along with the requirement that the dimensions of the displayed stimulus must be a power of 2 for rendering purposes, led us to a resolution of 1 screen pixel to 3.1 source pixels for a stimulus presented en face. A tilted stimulus has a slightly lower/higher pixel ratio for the portions of the stimulus that are tilted toward/away from the viewer, respectively.

As will be shown in the analyses below, the result of the algorithm up to this point has a frequency spectrum in which spectral power decreases with frequency, as is to be expected from fractal content. The high power in the low-frequency components of the spectrum reflects random clustering of overlapping triangles in space (Poisson clumping), which is observable in initially generated images as visible patches of predominantly bright or dark regions. The contrast between these clumps was reduced under zooming, providing a potential gradient cue.

To eliminate this cue, a second step in the algorithm enforced dynamic range across all portions of the image. Specifically, the pixel values resulting from the addition of the triangles were remapped by a quantized repeating sinusoid with a phase shift of π and a frequency of 4 (i.e., the remapped pixel values wrap around between white and black 4 times across the range of original values). To accommodate to the dynamic range of a 3D monitor and avoid washout or blackness, the maximum and minimum values were curtailed, rather than reaching black (digital 0) and white (digital 255). The curtailed range was an experimental parameter here: In the generated version called *wrap 1*, the brightness values varied from 8 to 248, whereas in generated version *wrap 2*, they varied from 16 to 240, in both cases in increments of 8. Equation 6gives the formulation for the remapping step:

$$I_W = 8 * \mathrm{round}(16 + \beta \sin(2\pi f I_s - \pi))$$

$$\beta = 16 - \frac{p_{min}}{8}$$

I_W: wrapped image

I_s: input image scaled to $0 - 1$ range

p_{min}: desired minimal pixel value, must be a multiple of 8

β: constant to rescale to desired region

f: frequency in intensity space

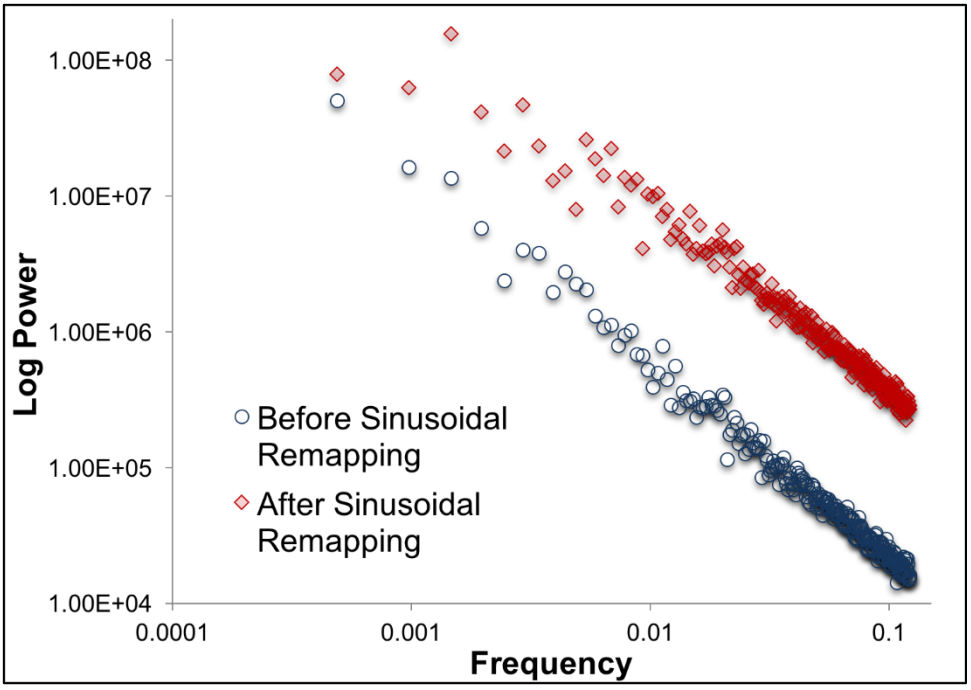
Equation 6

The parameter controlling the wrap value, p_{min}, is set to 8 for wrap 1 and 16 for wrap 2. The standard gamma correction (2.2) for Windows sRGB was used for both wraps. Sinusoidal wrapping and clamping the extremes of the intensity served to move the behavior of the image away from the edge cases of gamma correction and make it more robust to limitations of a non-ideal monitor. In order to print the texture, a professional grade painter (Canon PRO-1 or similar), which manages its own gamma, would be required to ensure optimal intensity matching with the source data.

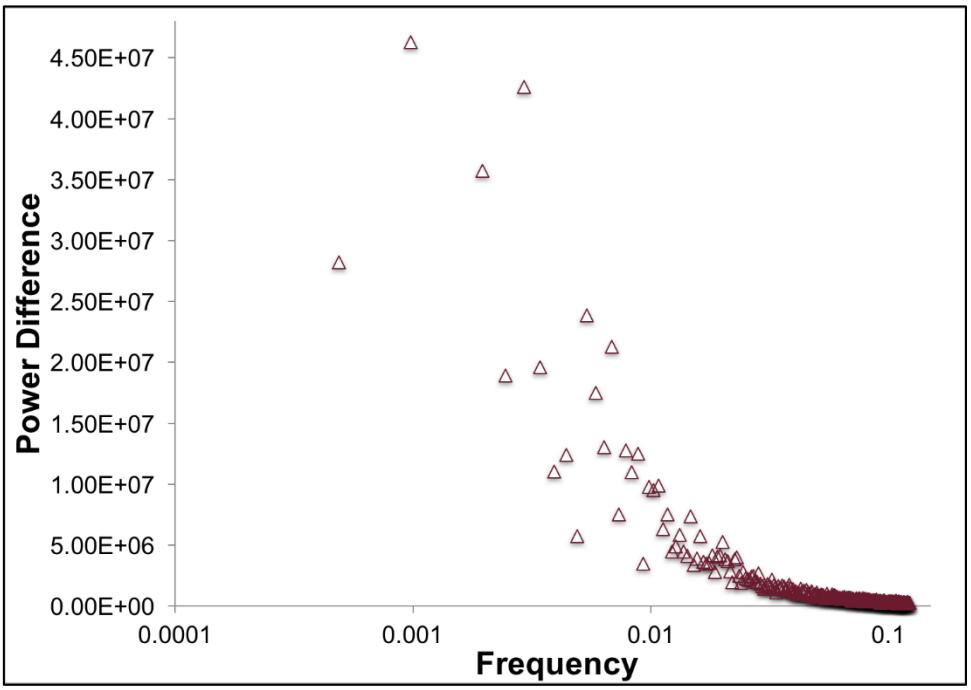
Figure 4-2 shows an example of a patch from the source image after step 1, and then in the second step after remapping (wrap 1, which ultimately proved more effective for our purposes).

4.1.2 Spectral analysis

The visual system is known to be sensitive to power distributions with a frequency fall-off of the form $f^{-\beta}$ [44] [45] which also characterizes the power spectra of many scenes in everyday visual experience (see [44]for review) as well as fractals [45]. The present images demonstrate such a power distribution before sinusoidal remapping, consistent with the fractal property of scale invariance. The value of the spectral fall-off was measured for patches drawn randomly from a source image by computing the 2D Fourier series and then converting it to a 1D (radial) spectrum. To eliminate quantization noise, this was done by extracting a neighborhood of 1D spectra around a diagonal from the center (the DC point) and averaging the power spectra [46]. The neighborhoods comprised angles of 45° to 50° degrees in .25° increments. The value of β was then extracted from a linear fit to the average spectrum in log/log space, as shown in Figure 4-4 (the "Before Sinusoidal Remapping" points). When this process was applied to six patches, all such fits yielded $\beta = 2.8$, with $r^2 > .98$.



4-4: Power spectra of patches from a source image generated by the algorithm.



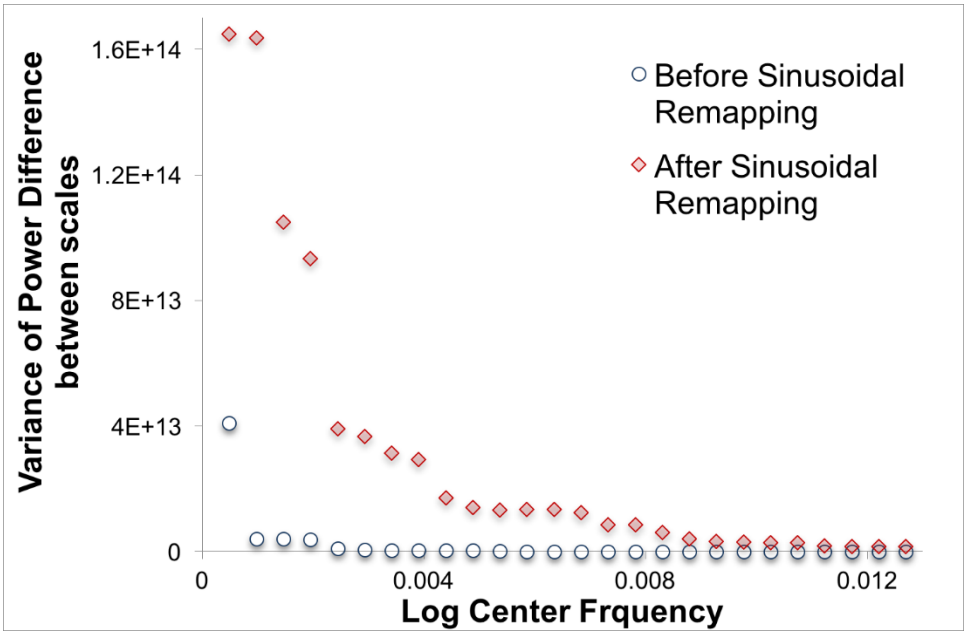
4-5: Difference in power of unmapped and remapped versions of a sampled patch, by spectral frequency. One outlier at the third frequency value, visible in Figure 4-4, is not shown.

Also shown in Figure 4-4 is the change in the spectrum after the second stage of the generation algorithm using sinusoidal remapping. This can be seen to amplify power overall, but more importantly, in comparison to higher frequencies, there is reduced relative power at the low frequencies associated with slow gradient change in the image.

Figure 4-5 shows in detail how the remapping affects the relative power spectra. Shown is the difference in spectral power between the unmapped and remapped version of the same image patch as a function of the frequency. The relative power loss due to remapping drops off sharply with frequency (essentially following a power function with exponent = -1.24, r2 = .95). This trend should promote the effects desired for stimuli that convey slant through stereo alone: Preservation of the high-frequency content should retain the edge information needed for stereo matching, while disruption of the low frequencies should reduce the pictorial information from the slowly changing gradient.

4.1.3 Invariance under scale

A further analysis directly assessed the effects of scale change on the power spectra. The spectral content was analyzed at two levels of scale, corresponding to a 2x magnified zoom-in. A square image patch measuring 8,192 pixels on each side was extracted from the source image and used to create two sub-patches measuring 4,096 pixels on each side. One was created by down sampling the patch to measure 4,096 pixels on each side, and the other by cropping it to extract the center square containing 4,096 pixels on each side. (The two sub-patches, which are equated in numbers of pixels, represent two levels of scale.) This was done both for the unmapped and remapped stages of generation. Of primary interest is whether or not differences in the spectral content at the two scales are systematic, particularly at the low frequencies where gradient cues would be represented. Figure 4-6 shows the running variance (across 25 successive frequency levels) in the difference between power at the two scales, for both the unmapped and remapped source images. After remapping, power differences due to scale change are highly unsystematic at lower frequencies, suggesting that the local scale changes produced by slant would not be cued by graded changes in the projected stimulus, but high frequency edge content is still preserved.



4-6: Running variance of power difference between scales by center frequency, for unmapped and remapped stages of generation. Scale changes in images with remapped pixel values result in highly unsystematic power differences at low frequencies, suggesting that scale changes are not inferable from image content. Since slant represents local changes in scale, it should likewise not be cued by the displayed stimulus.

4.2 EXPERIMENT 1

These first experiments assessed the performance of slant perception under conditions of stereo and non-stereo viewing, using our algorithm-generated scale-invariant stimuli and, for comparison, pictures of tiles that had clear pictorial perspective cues. "Generated" will be used to refer to our scale-invariant stimuli. The task was a forced-choice judgment, in which participants viewed a stimulus display that was either flat (i.e. frontal, perpendicular to the sagittal axis) or slanted 15° or 30°, and indicated its orientation direction by a 5-alternative response.

A large number of stimuli were created by cutting patches from a source image generated by our algorithm. Experiment 1A used a randomized assignment of stimuli to viewing condition, so that the measured level of slant perception would not reflect the contribution of a single stimulus. The following measures were obtained: (i) proportion of responses that were accurate as to slant direction, over all stimuli (5 possible directions) and when confining responses to stimuli reported as slanted rather than flat (4 possible slants); (ii) the ability to make a binary discrimination of stimuli into slanted vs. flat, using the d' statistic from signal detection theory (hit = slanted stimulus correctly identified, false alarm = flat stimulus called slanted) and (iii) a measure of information transmitted by the stimulus set in bits.

Experiment 1B tested whether individual stimuli might differ in the textural cues that they offer (i.e. some stimuli are more informative in mono than others), by obtaining responses from multiple participants for the same stimulus. Monte Carlo simulations were used to statistically test whether the response choices to algorithm-generated stimuli in the *Mono* condition were randomly distributed across the five slant alternatives.

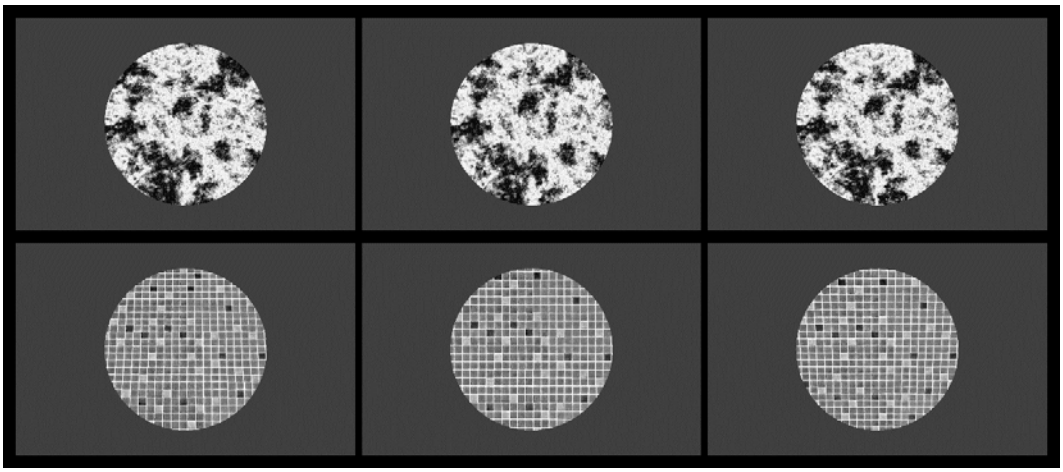
4.2.1 Experiment 1A

4.2.1.1 Participants

The participants were 12 young adults recruited from the Carnegie Mellon University community who were paid $10. All gave informed consent, and all experiments were conducted under IRB approval. All participants had stereo acuity of 32 seconds of arc or better, as assessed by a test (Vision Assessment Corporation).

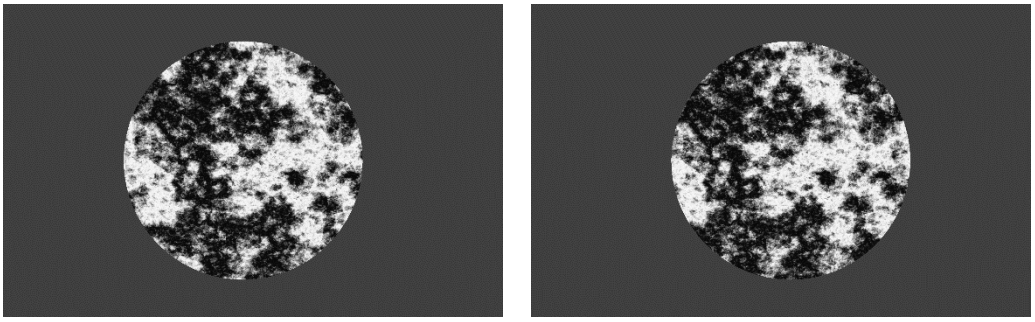
4.2.1.2 Stimuli

There were two classes of stimuli, algorithm-generated and pictures of tiles. Generated stimuli (after remapping) as rendered in 3D are shown in Figure 4-7, top. The control stimuli were images of ceramic tiles that offered clear pictorial perspective cues, as shown rendered in 3D in Figure 4-7, bottom. The on-screen size of a single tile was approximately 1.1 cm (38.7 pixels), and the grayscale values were broadly distributed with a modal value of approximately 130 (where "black"=0 and "white"=255).



4-7: Upper and lower rows show generated and control stimuli, respectively. Each column shows a different slant value: from left to right, they are top slanted inward (away from the viewer) 30°, flat, and bottom inward 30°

The generated stimulus set was constructed by extracting square patches from a large source image generated by the algorithm. The set comprised 6 sub-sets, as defined by two variables. The first was the *wrapping parameter* used in generation (2 possible parameter values), which affected the dynamic range of the source image as described previously. The second variable was the *region* of the patch in the generated source image. Specifically, slanted patches were rendered in 3D by rotating the generated source image around a line of bisection and dividing each half into thirds. For a given patch, its value on the region variable refers to whether it was drawn from the third that was nearest, in the middle, or furthest from the rotation axis. This variable was examined because any gradient cues in the source image should be strongest in the region nearest to the rotation axis.



4-8: Separate left- and right-eye views of a generated stimulus, shown rendered in stereoscopic 3D with the bottom slanted inward 30º.

For purposes of presentation, the patch from the source image was 3D rendered at the required slant using OpenGL, and then presented through a 3D-rendered circular aperture with radius 11.25 cm on the screen (397 pixels), surrounded by a dark-gray area that filled the screen. Under our image-generation approach, a key property of the image is that it must always contain higher frequency content (smaller edge fragments) than can be visually perceived at the maximal zoom level. During the OpenGL rendering process, each of the computer monitor's physical pixels within the viewing aperture must correspond to one or more generated pixels in the underlying image patch, with a many-to-one correspondence being preferable to avoid

potentially substantial down-sampling artifacts. This was accomplished by using patches that were 4,096 pixels on a side in the source image. When projected by Open GL rendering to the screen, they were locally sub-sampled by a factor ranging from .35 to .43, depending on slant. The radius of the circular aperture was sufficiently small that the edges of the square patch behind it were not exposed even at the most extreme slant, 30°. The subject was placed sufficiently far from the screen (50 cm) so that the individual pixels on the computer monitor corresponded to < 2 minutes of arc per pixel and could not be resolved under the experimental conditions.

There were two viewing conditions. In the *Stereo* condition, the participant viewed the stimulus (slanted or flat) with stereoscopic cues controlled by an Open-GL program using parallel axis asymmetric frustum perspective projection. The left and right "virtual camera" images for each eye were simulated using an eye separation with an inter-axial separation of 60 mm, and the two eyes were stimulated in alternating frames at 120 Hz (60 Hz per eye). Figure 4-8 shows a stereo pair.

In the condition contrasted with *Stereo*, the participant viewed an invariant frontal projection of the corresponding stimulus with both eyes, in essence showing both eyes the exact same perspective and pixels that would have been seen by a single eye in the *Stereo* condition. We label this binocular condition *Mono*, with the understanding that the same monocular cues are seen by both eyes. Although this condition constituted a cue-conflict situation (as described in the introduction), incorporation of the "tiles" images provided direct comparisons between a well-studied grid stimulus (tiles) vs. our present stimulus for both pictorial and pictorial + stereo viewing conditions.

4.2.1.3 Experiment design

Stimuli within each viewing condition were presented at one of 5 slant directions: flat, pitch rotated with the top in or bottom-in, and yaw rotated with the right in or left in (where "in" refers to into the screen, further from the viewer). The 4 non-flat slants could take on magnitudes of 15° or 30°.

The trials were divided into 4 blocks that alternated between *Mono* and *Stereo* viewing; the two possible viewing orders were counterbalanced across participants. Within each block there were 70 experimental trials: These included 60 trials with the generated set, 12 at each slant direction. Within each slant direction other than flat, the two magnitudes (15° or 30°) occurred equally often, and the values of region and wrapping parameter were represented equally within each condition. The remaining 10 trials in a block presented tiles, two at each slant direction (one for each magnitude) and two flat. The trials were randomly ordered within each block. Before the first block, there was a 20-trial practice block with 5 trials in each viewing condition, in the same order as would follow in the experiment proper, and each successive block was preceded by 5 practice trials, 1 at each slant.

A pool of 280 unique stimuli was created for the experiment, 240 drawn from the generated set and 40 from a source image of tiles, representing the constraints of the stimulus variables. Once the sequence of trials was determined by the control program for a given subject, a stimulus of the appropriate type was drawn from the pool and presented with the designated viewing condition and orientation. Thus each subject saw a unique order of stimuli, and the same source stimulus occurred randomly across experimental conditions.

4.2.1.4 Procedure

After informed consent, the participant was administered the stereo acuity test. He or she was then seated at a Dell computer with a 47.5 cm x 29.8 cm Samsung Syncmaster 2233RZ LCD screen, so that the head was at an optimal viewing distance for the stereo display (50 cm), as aligned with a mark on the adjacent wall. The participant was told that he or she could move the head laterally but not forward or backward, and the alignment was checked periodically by the experimenter. Each trial began with a signal cross that was viewed for 500 ms, followed by exposure of the stimulus. When ready to respond, the participant pressed a key on a number

pad to indicate its slant direction, at which point the selected key input and its corresponding slant were displayed on the screen for confirmation or correction if mistyped (responses were typically made without looking at the keypad, and occasionally the participant wished to correct a known erroneous response). The response keys were arranged in a cross for intuitive mapping to the slant options (e.g., the left arm of the cross corresponded to left edge inward, etc.) Participants had up to 7 sec to make the response.

4.2.1.5 Results

Initial analyses were performed to test whether the region variable (distance of patch from slant axis in source image) had a significant effect on accuracy (proportion correct on orientation out of 5 choices). Separate analyses of variance for the *Mono* and *Stereo* data showed no effect of region, ps > .10. Therefore data were pooled over region for all subsequent analyses.

Table 3 shows the proportion correct for each slant magnitude by viewing condition and stimulus category. As expected, there was a strong tendency to report generated stimuli viewed in the *Mono* condition as flat (overall, 46% of responses to generated stimuli, relative to 20% chance with 5 response alternatives), resulting in the higher accuracy for the 0° slant in the *Mono* condition. Considering the 54% of total responses that were other than "flat," 16% indicated the stimulus was slanted right edge in, 13% left in, 8% top in, and 17% bottom in. This indicates no particular bias to report pitch (top or bottom) vs. yaw (left or right) slants.

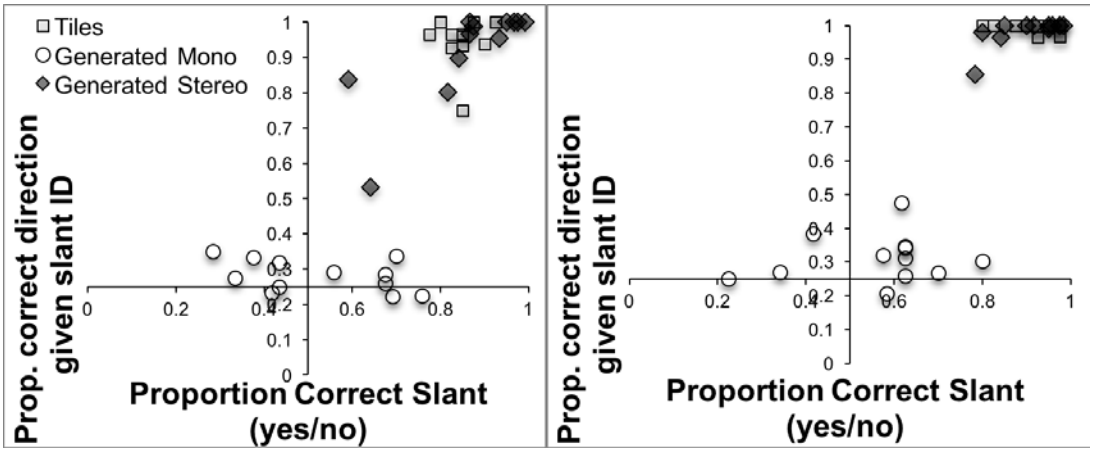
Table 3: Proportion correct identification of slant direction by viewing condition (Mono, Stereo) and stimulus (tiles, generated wrap 1, and generated wrap 2) in Experiment 1A.

	MONO			STEREO		
Slant	Tiles	Generated Wrap 1	Generated Wrap 2	Tiles	Generated Wrap 1	Generated Wrap 2
0°	0.65	0.46	0.48	0.73	0.78	0.76
15°	0.73	0.1	0.15	0.78	0.73	0.74
30°	0.94	0.13	0.22	0.95	0.9	0.9

Initial ANOVAs were conducted with factors of stimulus category (tiles, generated wrap 1 and wrap 2), experimental block (2 levels) and order of conditions (*Mono* first vs. *Stereo* first). These analyses were done separately for *Mono* and *Stereo* viewing conditions, because *Stereo* accuracy was generally high (tiles 84%, generated stimuli 81% and 80% for wrap 1 and 2). The ANOVA on *Stereo* showed no significant effects, indicating that the generated stimuli yielded accuracy under stereo viewing equivalent to the tiles, where perspective cues were also present[2]. The ANOVA on *Mono* showed only a significant effect of stimulus, $F(2,20) = 330.30$, $p < .001$, with the tiles (80% correct) superior to both of the generated stimulus categories (18% and 24%, respectively). Accuracy with tiles did not differ in the *Mono* and *Stereo* condition (by t test, $p = .30$).

Figure 4-9 shows the accuracy for each participant in Experiment 1A (left) and Experiment 1B (right, discussed in the next section). Clearly, the generated planar stimuli convey slant under *Stereo* viewing, whereas *Mono* accuracy clusters near chance.

[2] Because performance was essentially ceiling-level in these 5-alternative judgments, the data did not manifest a commonly found anisotropy in slant perception with metric responses, i.e., superior performance in matching pitch relative to yaw (e.g., [46]).



4-9: Left: accuracy in Experiment 1A for individual subjects, as assessed by two measures: the discrimination of slanted from non-slanted stimuli (x axis) and, conditional on accurate identification of slant, the identification of stimulus orientation direction (y axis). Both measures pool data from 15º and 30º slants. The axes are adjusted to cross at chance for both measures. Right: an equivalent graph for Experiment 1B.

To assess the effect of slant magnitude, the conditions with slanted stimuli were subjected to 3 X 2 (stimulus: tile vs. generated wrap 1 vs. generated wrap 2 X angle: 15° vs. 30°) ANOVAs, separately for *Mono* and *Stereo* viewing conditions. In the *Mono* condition, there were main effects of stimulus, $F(22) = 302.04$, $p < .001$ reflecting the obvious superiority of the tiles, which provided perspective cues. There were also effects of angle, $F(1,11) = 56.04$, $p < .001$, and an interaction, $F(2, 22) = 6.01$, $p = .008$. The latter effects arise from an advantage for the 30° angle for tiles (by t test, $p < .001$), but not generated stimuli. In the *Stereo* condition, the sole effect was of angle, $F(1, 11) = 28.90$, $p < .001$. This indicates that under stereo viewing, the generated stimuli and tiles were equally well identified as to orientation and showed an equivalent advantage for the 30° angle over the 15°.

The remaining analyses focused on the generated stimuli in the *Mono* conditions, where unlike the *Stereo* conditions or tiles viewed as *Mono*, substantial errors were made. To assess whether flat and slanted stimuli (in any direction) were discriminated, the d' statistic for flat/slant choices was computed for each subject, and t-tests were used to compare the values to zero. (This measure is independent of differential response frequencies.) Mean d' values at 15° were .02 and .01 for stimuli generated with algorithms 1 and 2, respectively, and at 30° the corresponding values were .02 and .21. None significantly exceeded zero, although the value for wrap 2 at 30° approached significance, $t(11) = 1.72$, $p= .06$ with a 1-tailed test. Thus, as shown by the clustering of data near chance on the x-axis of Figure 4-9, generated stimuli rendered as slanted vs. flat were essentially indiscriminable under *Mono* viewing.

The next analysis focused on the data represented by the y-axis of Figure 4-9, namely, accuracy on identifying the orientation of slant, given that a stimulus was identified as slanted at all. Mean accuracy on slant direction was tested against chance (25%) by t-test. Only the accuracy for generated wrap 2 at 30° (mean 38%) exceeded chance, $t(11) = 3.45$, one-tailed $p = .003$. Thus, for the generated stimuli using wrap 2 and subjected to 30° rotation, once identified as slanted they could be judged with respect to the direction at a level above chance.

Finally, the slant information available from the *Mono* rendered generated stimuli was further quantified in bits by an information-transmission statistic, which takes into account the response tendencies [47]. The value was .02 bits for wrap 1 at both orientations and also .02 bits for wrap 2 at 15°, but reached .04 bits for wrap 2 at 30°, again pointing to some orientation information available from this wrapping algorithm under the higher slant.

Although instructions did not indicate responses should be speeded, it is also of interest as to whether the time to respond varied with the stimulus and experimental variables, particularly for the conditions where stereopsis must be induced. An ANOVA with stimulus category (tiles,

generated wrap 1 and 2), experimental block (2 levels), and order of conditions (*Mono* first vs. *Stereo* first) was conducted on response latency. The *Mono* condition showed only an effect of stimulus, $F(2,20) = 5.52$, $p = .012$: Responses to tiles (mean latency = 2.5 sec) were faster than to generated stimuli (mean latency = 3.3 sec for each wrapping). The response latencies in the *Stereo* condition showed a stimulus X block interaction, $F(2,20) = 5.60$, $p = .012$, reflecting a drop in response latency for the generated stimuli, but not the tiles (tiles 1.9 sec in both blocks, generated stimuli decreased from 2.0 sec to 1.6 sec from Block 1 to Block 2). Importantly, the time to elicit a response under stereo viewing was no greater for generated stimuli than for tiles.

In short, the analyses indicate first that under stereo viewing, the generation algorithm produced planar stimuli that were quickly and accurately identified as to slant direction, equally so to stimuli that offered perspective cues. Second, multiple measures indicate that when viewed without stereo cues, stimuli generated with wrap 1 provided essentially no information about slant, even at 30°. However, stimuli generated with wrap 2 and presented at 30° did provide non-stereo slant cues with measurable impact, albeit small. Specifically, when using wrap 2 there was (a) marginally significant slant/non-slant discrimination, and (b) significantly above-chance identification of direction of orientation for stimuli identified as slanted.

4.2.2 Experiment 1B

The randomized assignment in Experiment 1A of source stimuli to viewing condition and slant means that the obtained level of slant perception does not reflect the contribution of a single stimulus. However, it remains possible that under non-stereo viewing, some individual stimuli generated by our process might have reliable slant cues or that the process might bias an illusory slant. To assess these possibilities, Experiment 2 obtained multiple responses to the same stimulus within a condition.

4.2.2.1 Method

The participants were 12 young adult volunteers who gave informed consent. The design of Experiment 1 was repeated, except that rather than assigning a stimulus patch randomly to a viewing condition and slant, a given patch was assessed at the same slant direction and magnitude across all subjects within a viewing condition. This resulted in six participants providing responses to the same stimulus at a particular slant direction and magnitude, with 48 generated stimuli being assessed in the flat orientation and 96 in the slanted.

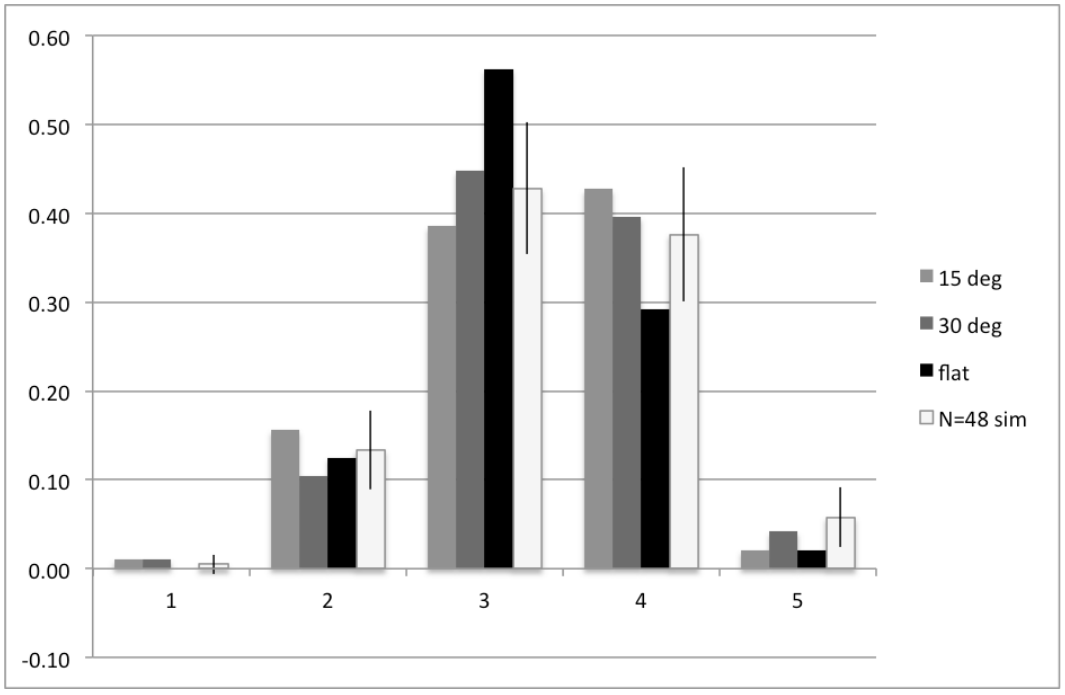
4.2.2.2 Results

Individual subjects' proportions correct in slant presence and direction for all conditions for this experiment are shown in Figure 4-9 (right), for purposes of comparison to Experiment 1A. The same pattern can be seen as previously, such that *Stereo* slant perception is generally excellent, as is perception of tiles without stereo cues, but the generated stimuli produce near-chance performance under *Mono* viewing conditions.

Because the main purpose of Experiment 1B was to test whether individual generated stimuli might offer reliable or biased non-stereo slant cues, the results of interest are the distribution of responses to a generated stimulus across the five response alternatives in the *Mono* condition. The analysis assesses whether the observed use of the response options is random or whether, alternatively, orientations of some stimuli are judged easily or consistently at some slant. Either of these latter alternatives would produce convergence on a single response.

In fact, only two of the generated stimuli rendered as *Mono* produced a common response by all subjects viewing them, and that response was incorrect and not biased toward a slant (i.e., "flat") in both cases. To statistically test whether the response choices to generated stimuli in the *Mono* condition were randomly distributed across the five slant alternatives, the number of unique slant responses used by the six subjects viewing a single stimulus was determined (regardless of which particular responses were chosen). For example, if three subjects chose top-in and three chose flat, that would constitute two unique responses out of the five possible for that stimulus. The empirical distribution of the number of unique responses was computed

for generated stimuli rendered as *Mono*, separately for sets with objective slants of 0° (flat), 15° and 30°. To assess whether these observed distributions of unique responses reflect random responding, Monte Carlo simulations were used to generate a random distribution for comparison. In each simulation, six random choices were made out of five alternatives for a given stimulus (representing the six participants and the five slant alternatives), where the probabilities of the slant responses were set to the marginal response tendencies that had been exhibited empirically. This was repeated for the total number of stimuli of the given slant magnitude, thus generating a distribution of number of unique slant responses based on random responding. One hundred such simulations were run for each slant magnitude, and the 95% confidence interval around the mean value of each number of unique responses was calculated. All 15 comparisons between observed and simulated values (5 numbers of unique slant responses X 3 slant magnitudes) were consistent with random responding (i.e., the observed proportion of that number of unique choices fell within the 95% confidence interval around the simulation mean). Figure 4-10 shows, for each slant magnitude, how the response options were distributed. Each bar is the proportion of stimuli with a particular slant that elicited a given number of distinct response options from the six subjects (regardless of which particular responses were chosen). The same results held when the two generated subsets defined by wrap algorithm were assessed separately. Thus the tests clearly indicate that for the population tested, Experiment 1A's aggregate data indicating ineffectiveness of non-stereo cues are not masking generated stimuli that are highly informative or biased on this basis.



4-10: In Experiment 1B, proportion of stimuli with the given slant that elicited the number of response options shown on the abscissa, in relation to a chance proportion generated by Monte Carlo simulation with 48 stimuli (the smaller set size).

4.2.3 Discussion

The image generation algorithm shown in section 4.1 largely met our goal of generating planar images that eliminate gradient cues when slanted (either in a 3D rendering or in a physical instantiation) but allow for stereoscopic depth processing. Under one parameterization, the algorithm yielded no evidence of *Mono* slant detection, let alone identification. The alternative parameterization appeared to allow a small amount of information to "leak" through under *Mono* viewing at the maximum slant; however, the ability to discriminate slant from flat still

did not statistically exceed chance. Experiment 1B demonstrated that these results were not the result of aggregating data across stimuli affording good and bad levels of slant perception.

The data showed a small but reliably detected level of pictorial gradient cues for wrap 2, for which the dynamic range of the images was the lesser. Further work is needed to establish why the lesser dynamic range (a reduction of information) somewhat counter-intuitively led to better (i.e., barely measurable) perception of slant. One possibility is that the higher contrast of wrap 1 allowed higher-frequency content to be perceived, at least at extreme levels, producing a masking effect [48]. It also remains to be determined whether the subtle gradient cues that emerged with wrap 2 are from edge statistics (which in theory, at least for the underlying triangular edges, should have been fully controlled) or some other monocular source (e.g., distributions of locally average intensities in the mid to low frequencies). These issues notwithstanding, we have demonstrated one algorithm that was completely successful by the present criteria, and another that offered very little monocular information regarding slant. We next describe the use of our generated images to measure the slant threshold under magnification from microscopic viewing.

4.3 EXPERIMENT 2

It is critical to assess the threshold level at which slant can be determined under magnification (The threshold is the minimum slant necessary to detect departure from a comparison plane.) Experiment 2 assessed absolute slant thresholds using the stimuli of Experiment 1A and 1B along with a control stimulus, diffuse glass. The experiment was done at two magnifications, 5x and 7.5x. They are referred to as Experiment 2A and 2B, respectively.

A substantial literature has assessed limits of the ability to perceive surface slant under various cue conditions. Knill [49] examined pitch-slant perception from textural cues alone and reported that the threshold was particularly high when discriminating low slant values; the 85% threshold for slant vs. fronto-parallel (i.e., the absolute threshold) was on the order of 30-40°. Comparing the effects of stereo versus textural cues, Knill and Saunders [50] reported 75% thresholds for pitch slant of approximately 15° for stereo only, 30° for textural only, and < 10° for the cue combination. Those values were obtained with rendered stimuli, however, not magnified, physically slanted surfaces.

Purdy [20] noted that the effect of magnification on a graded surface is to reduce the gradient magnitude by the inverse of the scaling factor. He confirmed that monocularly perceived slant is reduced as predicted by rescaling. Effects of the microscope go beyond magnification, however, because by virtue of its optics, each eye's viewpoint is projected toward infinity, effectively removing perspective. These conditions undermine the texture scaling cues proposed by Todd and associates [51] [21] [22] [52] to support planar slant perception from monocular gradients. Thus even if an image demonstrated perspective-induced gradients with slant, the resulting slant cues would be minimized under microscopic viewing conditions. In addition, magnification under the microscope reduces the range of disparities and hence the stereo slant cues. (As magnification increases, the diminishing viewing angle approaches orthographic projection, in the limit eliminating variation of disparities within the image and thus removing the ability of stereo vision to discriminate between different depths.)

4.3.1 Experiment 2A

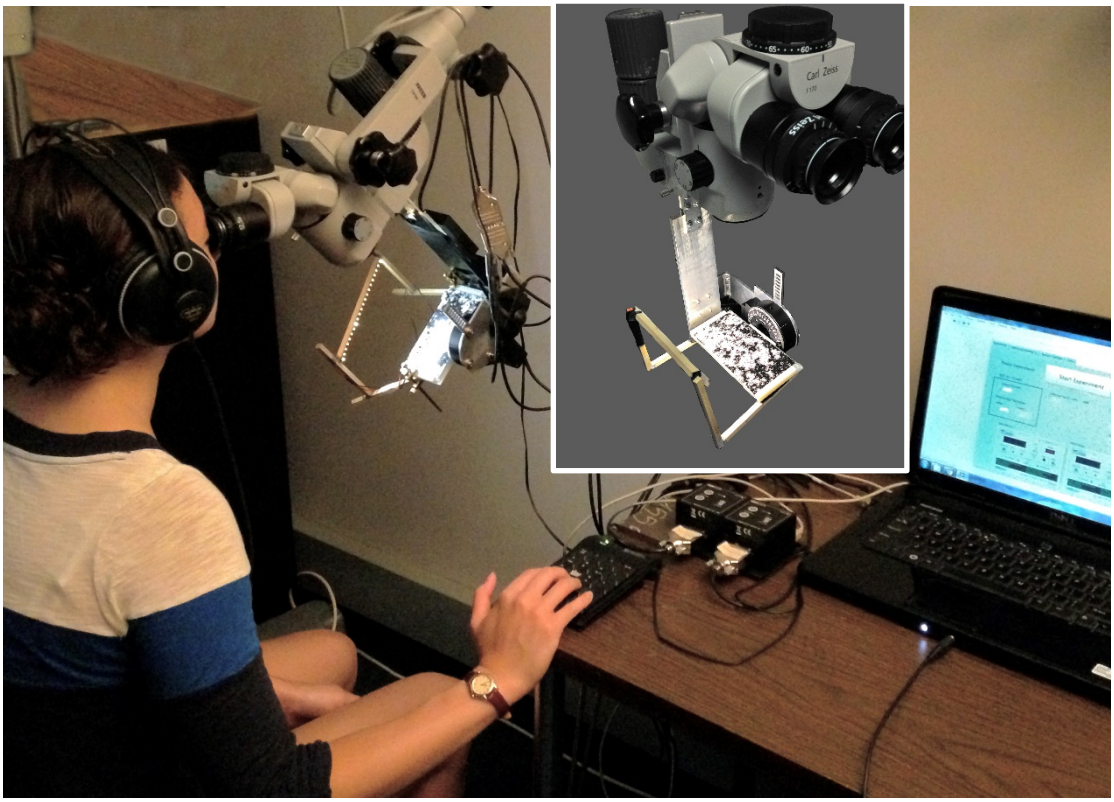
4.3.1.1 Participants

The participants were 8 young adult volunteers, 2 of whom were collaborators. All gave informed consent, and all experiments were conducted under IRB approval, including oversight of our protocols for maintaining subject privacy and handling human-subject data. All subjects had stereo vision of 32 seconds of arc or better, as assessed by a stereo acuity test (Vision Assessment Corporation).

4.3.1.2 Apparatus and Procedure

4.3.1.2.1 Microscope

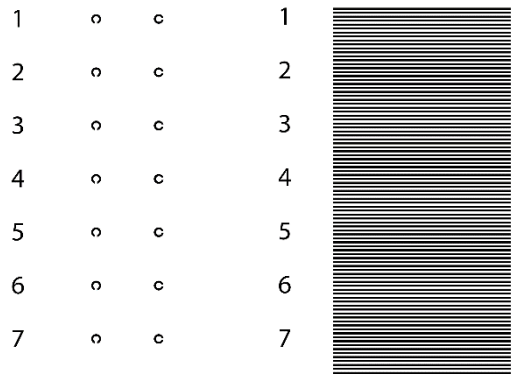
The experiment was run on a Zeiss OPMI-1 stereo microscope upgraded with Zeiss' most recent viewing binoculars and eye pieces. Attached to the microscope under its objective lens was a custom-made slide mount that allowed for computer-controlled rotation and focusing, as shown in Figure 4-11. A diffuse, dimmable lighting strip was attached to the rotating stage with a rigid mechanical bracket so as to maintain constant illumination regardless of slide slant. The lighting (FlexFire 24v Architectural Ultra Bright Natural White LED light strip) ran the length of the stage and was positioned to prevent reflected glare from being visible through the microscope. A wooden L-shaped structure affixed to the stage served as the mounting bracket for the LED strip (see Figure 4-11). A ThorLabs Motorized Rotation Stage (PRM1Z8) was used to rotate the viewed image, while a ThorLabs Motorized Translation Stage (MTS50-Z8) was used to move the stage vertically, allowing the subject and experimenter to control focus on the image being viewed. Both stages used a computer-controlled ThorLabs T-Cube DC Servo Motor Controller (TDC001) to drive the movement. A LabVIEW program, using ActiveX controls, controlled both stages, automated most of the experiment, prompted the experimenter to perform all the manual tasks, and recorded responses.



4-11: Experimental set-up with computer-controlled rotation and focusing stages attached to stereo microscope and response keypad.

The slant thresholds were measured with the following stimuli: two generated stimuli, chosen from the data of Experiment 1B because they produced broadly distributed slant responses when viewed as flat (rendering version 1), two versions of the tile stimulus from Experiments 1A and 1B and a featureless diffuse glass plate. Two additional stimuli were constructed for the adjustment of focal distance: one with two columns of Landolt C's (Figure 4-12, left), and the other with closely printed horizontal lines (Figure 4-12, right). The elements of both stimuli were sized as small as possible while still being clearly printed on our 4800 dpi printer (A Canon Pixma Pro-1 printer, with Canon Pro Platinum PT-101 Photo Paper). The first stimulus had one column of numbers, allowing each row to be identified, one column of C's opening downward, and another column of C's opening to the right. The vertical row spacing (center-

to-center) was 1.42mm. The closely printed horizontal lines were also assigned reference numbers to allow identification (also in 1.42mm vertical increments). Each black line was .127 mm wide, with a .127 mm white gap between each line, forming a grating pattern.



4-12: Focal distance of the microscope was adjusted for each subject using a semi-automated procedure that made use of two stimuli generated specifically for these experiments, a set of Landolt C's (left) and a grid of closely spaced horizontal lines (right). A sub-section of the panel used is shown.

4.3.1.2.2 Focusing Procedure

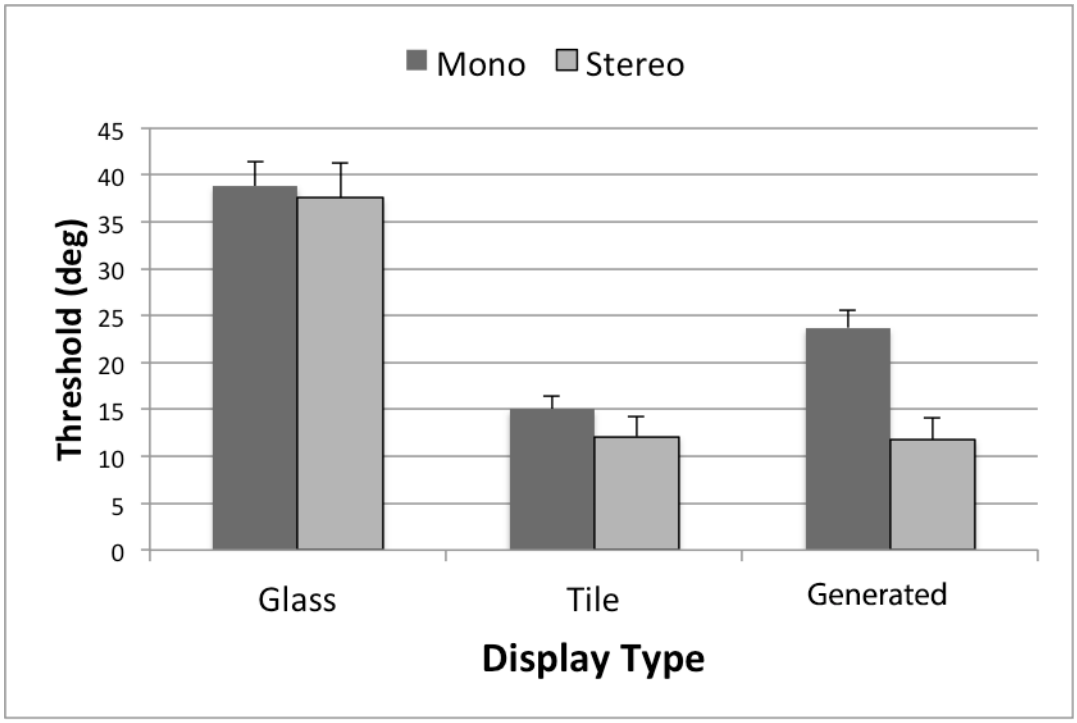
Once stereo acuity was confirmed, the dominant eye was identified by having the participant move the index finger quickly toward the eyes and determining the target eye. Then the microscope was set for the participant under binocular viewing as follows. The microscope and/or the participant's chair were adjusted to allow the particular individual to gaze straight forward through the stereo eyepieces. With the tile stimulus set at 45°, the participant iteratively adjusted both the inter-pupil eye-cup distance, until the circular field of view could be seen with each eye separately, and the eye relief, so that the field of view was not vulnerable to small head movements. Focal distance was then set by having the participant view the focusing stimuli without slant. While viewing the rows of Landolt Cs, the participant adjusted the slide height to find the upward and lower limits where any gap could be detected, and then further fine-tuned the height around the midpoint of that range so as to see as many gaps in Cs at the center as possible. The participant then reported the upper and lower row numbers of Cs where gaps were visible, with eye movements allowed. Finally, the participant viewed the grating stimulus at its center and, while keeping his or her eyes focused on the center, reported the top and bottom line numbers where the separate lines were visible. All focal distance measures were checked for any obvious asymmetries in upward and lower fields of view, and all such individual-specific adjustments were recorded for future use, simplifying non-continuous data collection. Adjustment of the microscope lasted approximately 20 minutes per subject.

4.3.1.2.3 Thresholds

The experiment then proceeded to the threshold measure. The microscope was set to a magnification of 5x. Two interleaved staircases, one ascending and one descending, were implemented with the method of Kaernbach [53] targeting 75% correct detections. The subject wore headphones with pink noise to mask the motor sounds. On each trial, one flat surface and one slanted surface were presented in random order and the subject responded with a keypress on a button panel to indicate which was more slanted (first or second) or "don't know" (DK) if uncertain. To avoid undesired motion cues from rotation, a computer-initiated vocal command through the headphones instructed the subject to close his or her eyes during the stimulus change. The initial slant of the test slide was ±15°. It was changed by a quantity D towards 0° after a correct response, and changed by 3*D away from 0° after an error, or by D away from 0° after a DK response. Initially D was 2°. After the 2nd and 4th reversal (correct followed by either incorrect or DK or either incorrect or DK followed by correct), the value of D was halved. If the test slant fell within 1°, D became .1°. The threshold measurement was ended after the

longer staircase reached the 8th reversal or ±45°. The value of the threshold was determined as the average of all slant values following the 4th through the 8th reversal.

All participants produced threshold measurements for each of the 3 displays (tiles, generated stimuli, and glass) both binocularly and monocularly, the latter using the dominant eye with the other eyepiece occluded. The monocular and binocular thresholds used different stimuli within each participant, and a Latin Square design was used to counterbalance for stimulus assignment and threshold order. The entire threshold procedure lasted approximately 3 hours. Participants took breaks as desired and could collect different thresholds in different sessions. Figure 4-11 shows the experimental set-up, except that room lights were turned off for actual experimentation.



4-13: Thresholds in Experiment 3 by display type and viewing condition. Error bars are one standard error of the mean.

4.3.1.3 Experiment 2A Results and Discussion

Figure 4-13 shows the thresholds by display type and viewing condition averaged over the two staircases. These results indicate the following: The lowest thresholds (12°) were achieved with stereo vision of the tiles and generated stimuli. These were equivalent despite the stimulus differences: Both offer edge features, but only tiles have a clear gradient when slanted. These threshold values are also comparable to those obtained for computer-generated images with texture and stereo cues [50]. Thresholds for rendered stimuli, but not tiles, were lowered by stereo viewing relative to monocular. Finally, diffuse glass produced thresholds essentially at the ceiling of our measurement, well above those of any other condition.

The statistical analysis supports these inferences. An initial ANOVA on display type, staircase direction, and viewing condition (monocular vs. stereo) found no effects of staircase direction or any interactions with that variable (all ps > .25). The only significant effects were of display type, $F(2,14) = 70.95$, $p < .001$, viewing condition, $F(1,7) = 10.58$, $p=.014$, and the interaction between them, $F(2,14) = 6.25$, $p = .011$. We used 2-tailed t-tests with alpha set at .05 to compare specific conditions of interest *a priori*; note that all significant effects are also significant by post hoc Bon Ferroni correction. When monocular and stereo conditions were compared within each stimulus, only the generated images showed a significant advantage for stereo. When tiles were compared to generated stimuli monocularly and in stereo, only the monocular conditions

showed an advantage for tiles. The featureless glass differed from both tiles and generated stimuli in both viewing conditions, and the mean threshold for glass approached the maximum value of 45° (which was in fact the modal value).

The finding that the threshold for generated stimuli under monocular viewing was substantially below that of the glass indicates the contribution of non-stereo cues. Given the optics of the microscope, which preclude a strong cue from convergence, the implicated cue is accommodation. In other research, we have shown that accommodation (i.e., the blur gradient) is a strong cue to surface depth under conditions of magnification [34].

4.3.2 Experiment 2B

We repeated the procedure of Experiment 2A at 7.5x to investigate differences in thresholds based on magnification. We hoped that there would be no significant difference in performance between these two magnifications, and so a short study was done to confirm this.

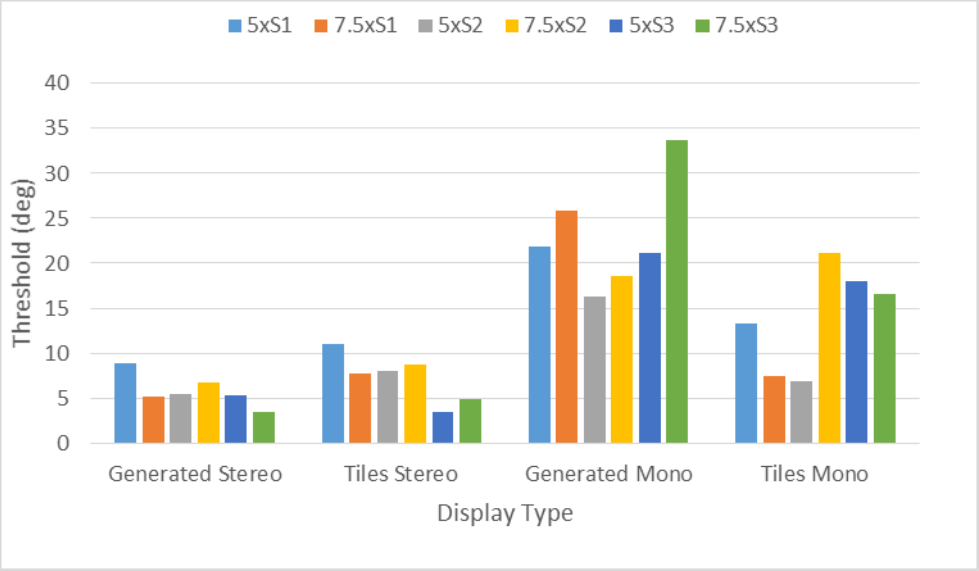
4.3.2.1 Participants

The participants were 3 young adult volunteers, 2 of whom were collaborators. All gave informed consent, and all experiments were conducted under IRB approval, including oversight of our protocols for maintaining subject privacy and handling human-subject data. All subjects had stereo vision of 32 seconds of arc or better, as assessed by a stereo acuity test (Vision Assessment Corporation). All three subjects also participated in Experiment 2A to allow direct comparison of performance.

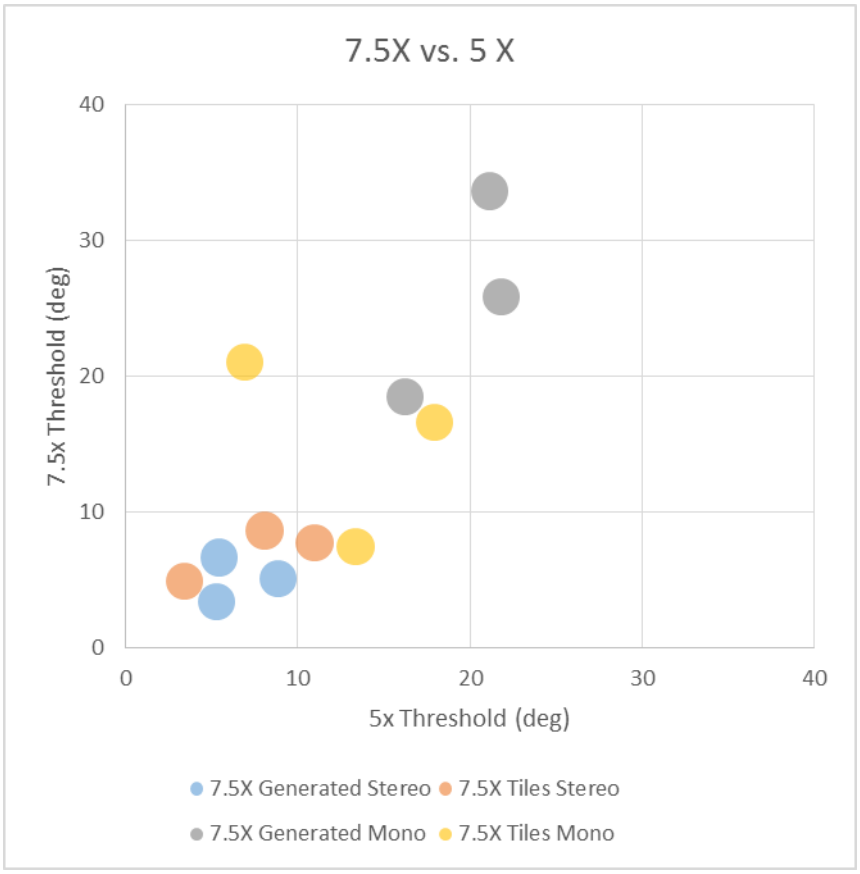
4.3.2.2 Methods

The same setup and procedure from Experiment 2A was used, with the microscope set at 7.5x for the course of the experiment rather than 5x. The frosted glass stimulus was excluded, as it was highly unlikely that there would be any appreciable difference in perception of the blank stimulus.

4.3.2.3 Results and discussion



4-14: Experiment 2B results. Thresholds of the three subjects is compared directly to their 2A 5x performance



4-15: Subject performance on 7.5x plotted against their 5x performance

With the small number of subjects (N=3), we are presenting their data directly in Figures 4-14 and 4-15. Clearly there is little to no difference in the thresholds acquired at 5x and 7.5x. In the Generated *Mono* condition, however, all three subjects performed worse at 7.5X than 5X (Figure 4-15, grey bubbles). We hypothesize this may be due to barely-visible printer dots creating misleading pictorial cues. We therefore conclude that moderate changes in magnification have little effect on the perception of slant when stereo cues are available, but magnification may produce a small degradation effect under monocular vision when gradient cues are suppressed.

4.4 EXPERIMENT 3

4.4.1 Overview

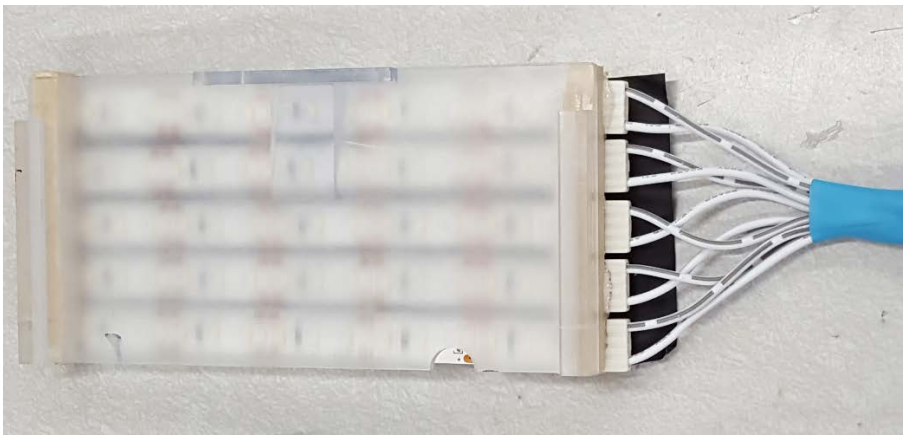
The third experiment looked at the performance of stereo perception under magnification in a more task-relevant framework. Instead of asking subjects to compare two surfaces with non-biologically relevant data, this experiment required that they perform an actual task, under conditions using actual OCT data. Experiment 3 used the same base apparatus as Experiment 2, and the same focusing procedure for each subject. The subject interacted with the microscope setup using a specially designed trackable tool, and their task was to align the tool with the tilted surface presented to them. In addition to the textures used in experiments 1 and 2, an actual OCT image was also used. We measured how well the user aligned the tool with the surface. Correct perception of the tilt of the surface is required for this task, although additional error and/or variability may be introduced by subjects' ability to control the alignment of the hand-held tool. The tool is designed to provide minimal pictorial cues itself, and it mimics the construction of a fixation ring, a surgical tool commonly used in ophthalmology. Our tool was tracked with the LEAP Motion Tracker, an infrared stereo tracker designed for tracking fingers

and styli above keyboards. The tracker was mounted on our tilting microscope apparatus so that is has a clear view of the entire stage (Figure 4-16).

4-16: LEAP Motion tracker mounted above the rotating stage.

4.4.2 Lighting Changes

The lighting for the microscope system was changed from the wooden L-bracket used in Experiment 2 to an array of backlights mounted under the rotating stage (Figure 4-17). The lighting consisted of a grid of LEDS (FlexFire 12v Ultra Bright Natural White LED light strip) affixed to an aluminum backing sheet. This sheet was attached to the rotating stage with enough space between to ensure that the individual LEDs were sufficiently diffused when viewed through the glass plate with the printed stimuli attached. This backlighting change was required to allow free movement of the alignment tool, and to prevent the tool from casting shadows during the experiment.



4-17: Experiment 3 backlight array, attached to the underside of the slide stage

4.4.3 Alignment Tool

The design of the tool for Experiment 3 was inspired by actual surgical tools used in ophthalmology. There are numerous tools with ring and loop like structures that are used in glaucoma treatment and in more general procedures. A selection of tools that influenced our

ring tool is shown in Figures 4-18 through 4-21. Of these tools, the spatulas (Figure 4-19) require the most precise alignment between their orientation and that of the patient's eye.

4-18: Fixation rings used in many procedures to stabilize the eye

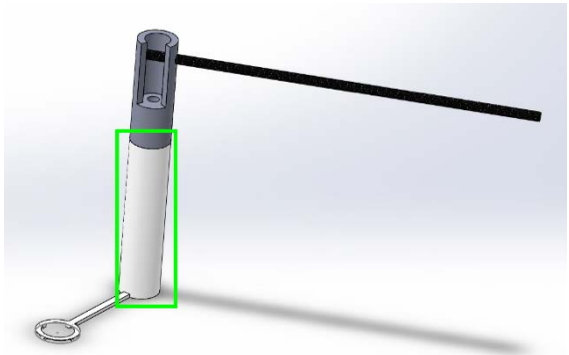
4-19: Cyclodialysis Spatula

4-20: Wilder serrated lens scoop

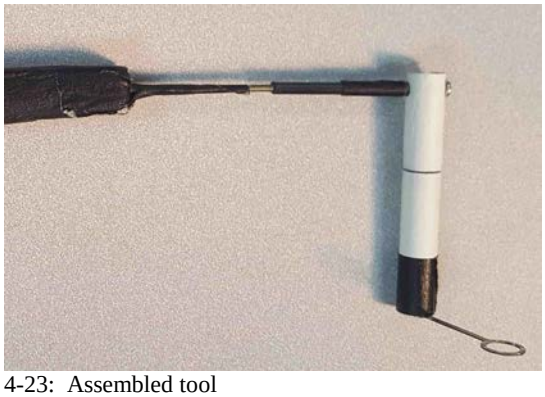
4-21: Arlt lens loop

A CAD drawing of our ring tool is shown in Figure 4-22. The LEAP Motion Tracker tracks the location of the shaft of the tool (highlighted in green). When the ring is held parallel to the surface, the white shaft will be normal to the surface. The size of the ring is designed to fit within the central field of view of the microscope at our primary magnification (5x). The ring is marked with iridescent paint so as to aid stereo viewing. The finished tool is shown in Figure 4-23. The ring is made of laser cut steel. The trackable shaft is 3D printed, and allows for the adjustment of the angle of the handle. The handle is made from metal and carbon fiber rods, with a foam grip for comfort and usability.

In order to prevent subjects from actually resting the tool on the surface, and therefore immediately perceiving the correct angle, a small guide hair is glued to the underside of the ring tool (Figure 4-24). This hair is angled to come into contact with the surface 2-3 mm before the tool itself, and bends visibly when contact is made. Subjects were instructed to monitor the status of the hair and use it to avoid making contact with the surface.



4-22: Solidworks CAD of ring tool. Trackable "finger" portion is highlighted in green.

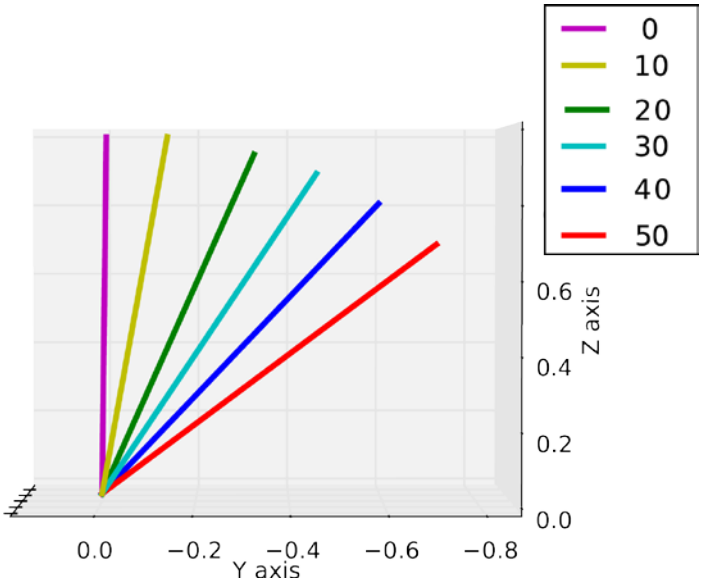


4-23: Assembled tool

4-24: Ring as viewed through the stereo microscope, with generated stimulus in background. The guide hair can be seen.

4.4.4 LEAP Motion Tracker

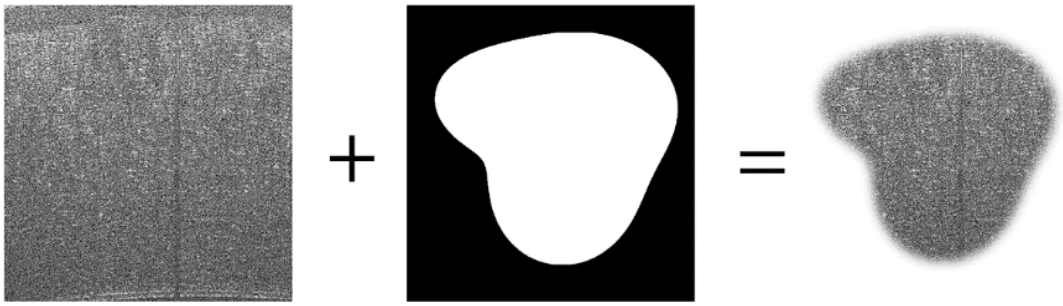
The LEAP Motion Tracker uses stereo infrared cameras to track fingers and tools. The tracker creates a depth map that is used to generate a 3D location for the tool with respect to the tracker's coordinates system. The tracker's API provides us with the tool/finger's XYZ location and direction vector. This direction vector can be used to determine the angle the tool is being held at (we constrain our analysis by projecting to angles to the rotation plane of the tilt surface). For our experiment, the surface presented to the subject could be one of six possible angles (0 to 50 in 10 degree steps), where 0 means that the surface is flat with respect to the objective of the microscope. For each angle, we calibrated our tracking system by measuring the tool's direction vector when perfectly placed on the surface. An affine mapping between these ideal measurements and the angles we wanted to represent was computed. The mapping was used to transform the vectors into our desired coordinate system, and to correct any systemic bias in the tracker. Figure 4-25 shows the calibration vectors before the affine remapping. The vectors were already close to ideal, showing little bias in the tracking system.



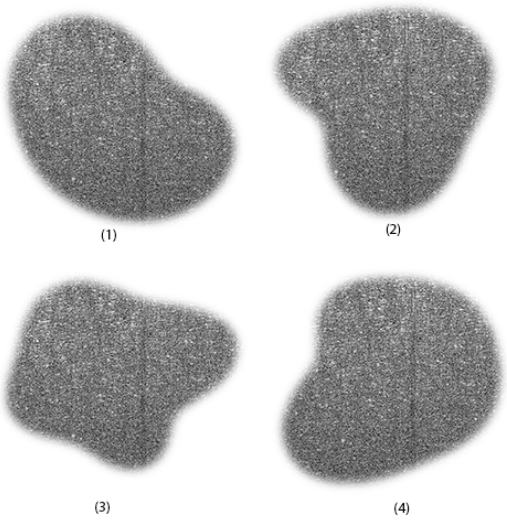
4-25: Calibration vectors before remapping to target coordinate system

4.4.5 Masked OCT Stimulus

In addition to the scale-invariant and control stimuli used in the earlier experiments, the tilt matching experiment used a stimulus derived from actual OCT data. Figure 4-26 shows the construction of the stimulus. An OCT scan of an ex-vivo pig cornea was used as the base of the image. The rectangular boundary of the image is masked by a randomly generated curve to remove perspective cues. This mask is generated by randomly perturbing and filling a spherical curve. A circular curve with a fixed radius is generated in polar coordinates, and a number of lobes (3 or 4 lobes were used for our image masks) were then added to the curve. The lobes are created at evenly spaced locations with randomly generated radii, and smoothly joined into the base circle. Once the random curve is generated, it is filled to create a binary mask (see Figure 4-26). The masked OCT image was printed on a white background. Four of these stimuli were prepared for use in the experiment, and are shown in Fig. 4-27. A number of candidate stimuli (same underlying image with different boundary masks) were tested to determine which would be used in our experiment. A short pilot study was done with a single subject to eliminate candidate boundaries that provided obvious perspective cues in the *Mono* condition.



4-26: Construction of masked OCT stimuli



4-27: OCT images, masked with randomly generated boundaries

4.4.6 Experiment 3A

4.4.6.1 Participants

The participants were 12 young adults recruited from the Carnegie Mellon University community who were paid $15. All gave informed consent, and all experiments were conducted under IRB approval. All participants had stereo acuity of 32 seconds of arc or better, as assessed by a test (Vision Assessment Corporation).

4.4.6.2 Apparatus and Procedure

Experiment 3A used the same Zeiss OPMI-1 stereo microscope and motorized slide mount design as used for Experiment 2. The lighting bar from Experiment 2 was replaced with the backlight array described in section 4.4.2. The focusing procedure from the previous experiments was also used. The microscope was set to a magnification of 5x.

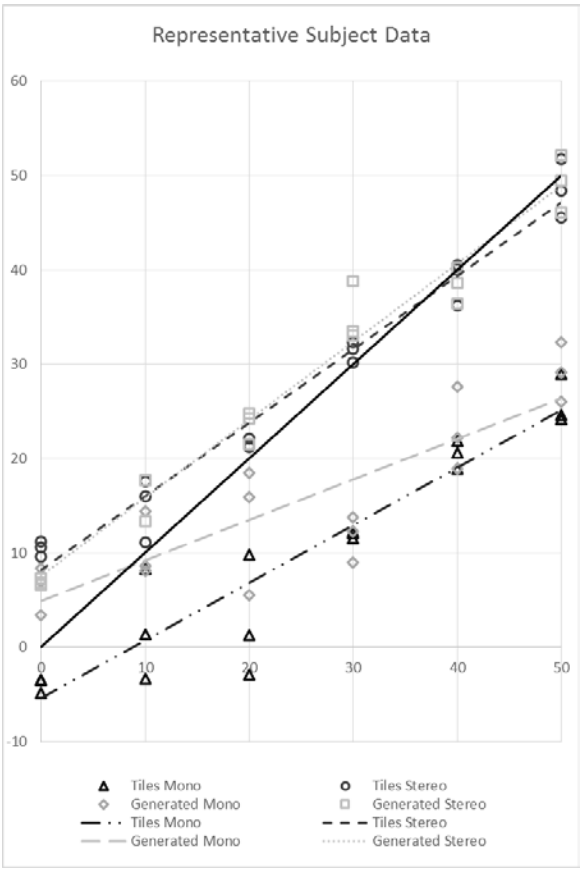
After the microscope was set up for the participant's focus depth, they moved on to the slant matching task. The experiment consisted of four blocks of 18 trials (3 repetitions of 6 discrete angles, randomized) each. Each block used a different stimulus/viewing condition pairing. Two stimuli were used, the generated scale-invariant image and the tiles. Both *Mono* and *Stereo* conditions were tested. As before, the non-dominant eye was blocked for the *Mono* condition. The order of pairings was varied between subjects using a Latin Square. Participants wore headphones which relayed verbal instructions and played white noise to cover the sound of stage movements. The tool was held in the subject's right hand, which was allowed to rest on a "floating" arm support. Each trial began with the subject's eyes closed as the stage adjusts to the randomized angle for that trial. When instructed, the subject opened their eyes, and moved the tool into the view of the microscope to begin aligning with the surface. Once they were satisfied with the alignment, the experimenter recorded the 3D position of the tool. Once recording was completed, the subject was instructed to close their eyes and pull the tool back from the stage as the next trial began. The stage movement was automated with LabVIEW and data recording was done automatically with a Python script.

4.4.6.3 3A Results

Figure 4-28 shows representative data from Experiment 3A for a single subject. Linear fits were done for the data points in each condition. The slopes of these fit lines give a good measurement of perception of relative angle. An ideal subject would have a slope of 1, indicating that each N degrees of tilt change are perceived as N degrees, and an intercept of 0, indicating no systematic response bias. However, subjects with non-zero intercepts but reasonably high slopes and linear fits can still accurately perceive relative tilt. Figure 4-29

shows the average slopes for each of the four conditions. These slope values for the stereo conditions are somewhat lower than tilt matching slopes for tasks where surfaces are observed using direct vision or haptics. Shaffer et al show that tilts observed with direct vision tend to be overestimated with a gain of 1.5 [54] . Freyberger et al. describe haptic tilt rendering schemes that yield tilt estimation slopes of ~0.85 [55] . Neither of these studies, however, used matching implements and provided feedback comparable to the present testbed.

As expected, we found better performance in stereo viewing conditions compared to *Mono*. Stereo performance for the two stimuli, tiles and scale-invariant, is essentially equivalent. *Mono* performance, however, shows better slope matching (i.e. perception of relative slant) in the tile stimulus. This is due to the perspective cues (gradient from regular pattern and objectively parallel lines) present in the image. The generated image viewed in *Mono* shows the worst performance, as subjects have neither stereo nor perspective cues.

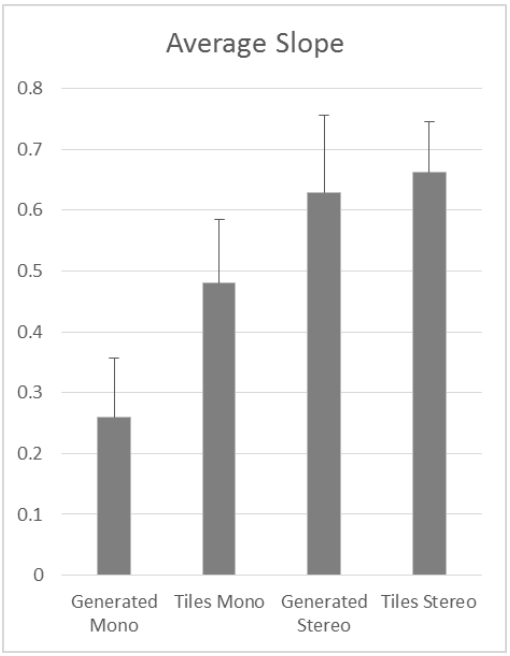


4-28: Representative Subject Data

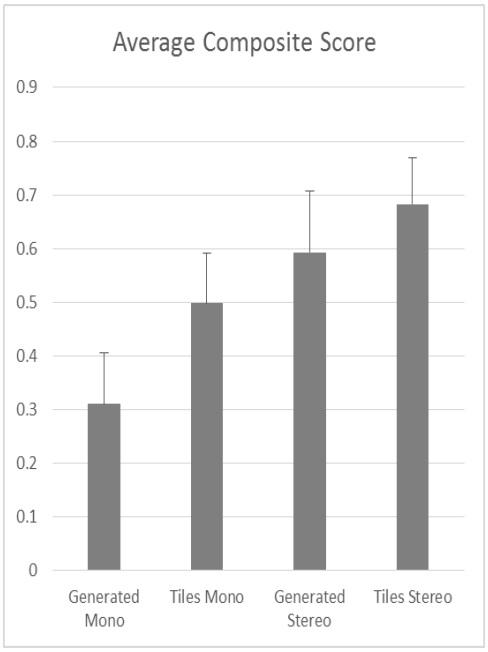
An initial ANOVA on slope found that effects of viewing condition (*Mono* vs. *Stereo*), $F(1,11) = 12.066$, $p = .005$, and stimulus type, $F(1,11) = 5.455$, $p = .039$, were significant, but their interaction was not. As performance also depends on the intercept (bias) and linearity of the fit function, we further analyzed a composite performance score (S) that ranges from zero to 1 and combines the relevant measures of slope (s), intercept (i) and R-squared value (see Equation 7). Average scores for the four conditions are shown in Figure 4-30. We found a significant effect for viewing condition only on this score, $F(1,11)=8.814$, $p=.013$ This indicates that when overall performance (slope, intercept and fit) is considered, there is no significant difference between the two stimuli, although a trend can be seen for superior performance with tiles.

$$S = \frac{2 - \text{abs}(1 - s) - \text{abs}\left(\frac{i}{50}\right) - (1 - R^2)}{2}$$

Equation 7



4-29: Average slopes for the 4 conditions in Experiment 3A. Error bars are 1 s.e.m.



4-30: Average composite scores for the 4 conditions in Experiment 3A. Error bars are 1 s.e.m.

4.4.7 Experiment 3B

This experiment was similar to the preceding, but used the masked OCT images described in section 4.4.5, to determine how effectively the slant of an OCT image from the cornea can be determined under the microscope.

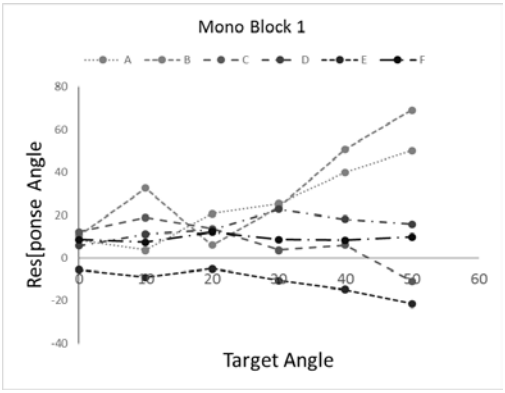
4.4.7.1 Participants

The participants were 6 young adults recruited from the Carnegie Mellon University community who were paid $15. All gave informed consent, and all experiments were conducted under IRB approval. All participants had stereo acuity of 32 seconds of arc or better, as assessed by a test (Vision Assessment Corporation).

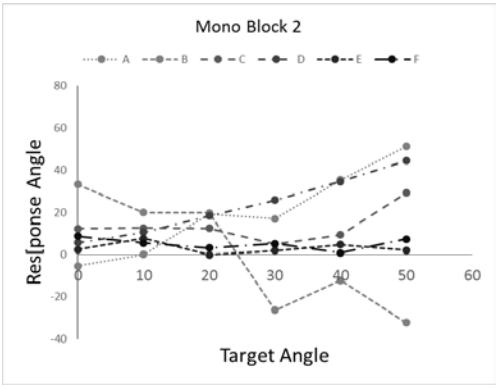
4.4.7.2 Method

The same procedure was used as in 3A, but with the masked OCT images instead of our previously used set of stimuli. The experiment consisted of four blocks of 12 trials (2 repetitions of 6 discrete angles, randomized) each. Each block used a different stimulus, to minimize the effects of learning, with two blocks performed in the *Mono* viewing condition and two performed in the stereo condition. The starting viewing condition and the stimulus/viewing condition pairings were varied between subjects using a Latin Square.

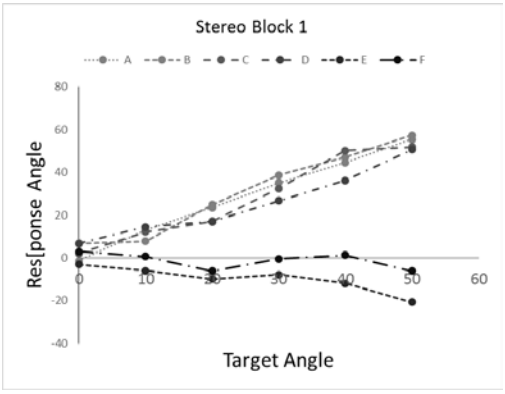
4.4.7.3 3B Results



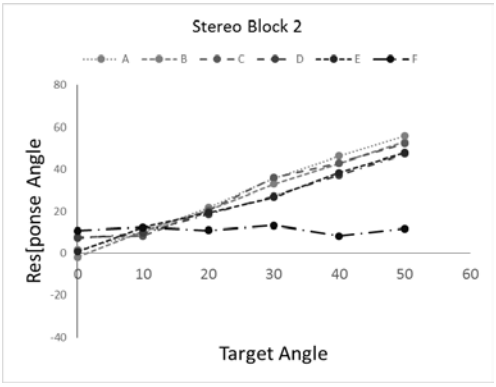
4-31: 3B Result, First Mono Condition Block



4-32: 3B Result, Second Mono Condition Block



4-33: 3B Result, First Stereo Condition Block

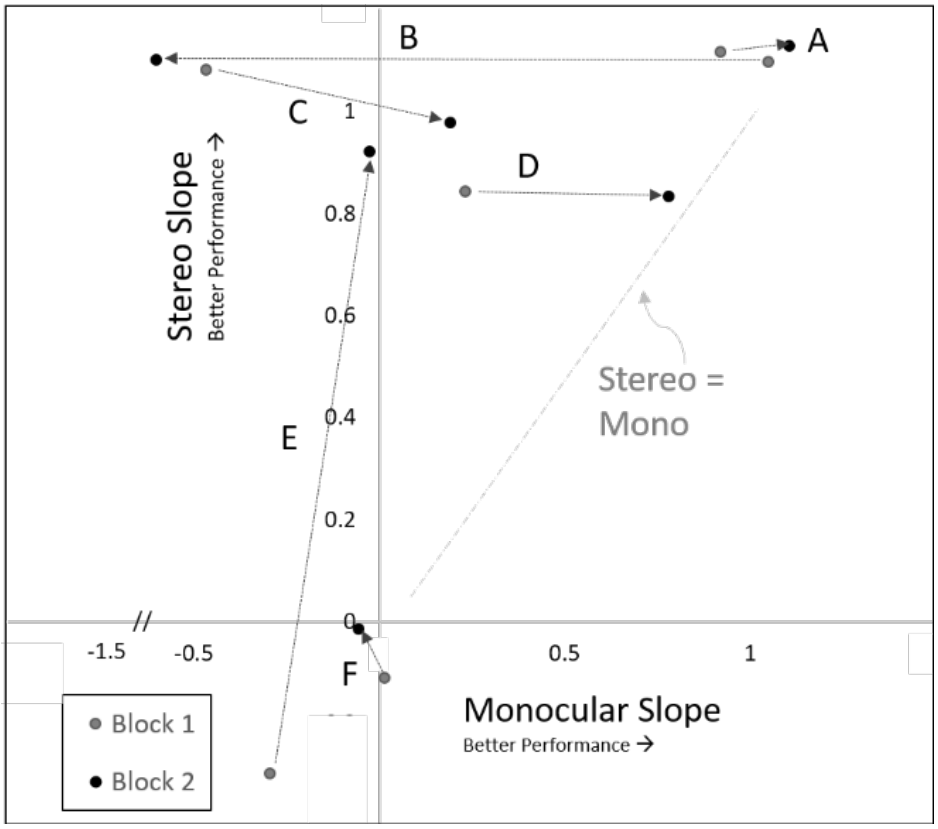


4-34: 3B Results, Second Stereo Condition Block

Figures 4-31 through 4-34 show the results for the first and second blocks of the two viewing conditions in Experiment 3B. The *Mono* condition blocks showed little consistency across subjects' performance, on average showing only a slight increase in response angle with target angle (this is likely due to the effect of accommodation at more extreme tilts). The stereo blocks showed well-defined slopes, with most subjects having clearly perceived the stereo information in the image. In particular, subject E showed a learning effect by the second stereo block, such that their performance greatly improved from the first stereo block.

The question might be raised as to whether the stereo performance with this type of image depends on monocular cues. If so, there should be a correlation between subjects' *Mono* and stereo performance. To test this, Figure 4-35 looks at slopes in the stereo condition in relation to the *Mono* performance for each of the two blocks, by subject. While the *Mono* performance was highly variable between subjects (large dispersion of points along the x axis), the stereo performance tended to be above a slope of 0.8, demonstrating that the stereo information in the masked OCT image is sufficient for determining slant, independent of the perception of monocular cues. An exception, Subject F, was notable for their complete failure at both the *Mono* and Stereo matching tasks in all blocks. We can also clearly see the learning effect for Subject E (large distance between the stereo performance of the first and second blocks).

Given the small N for this study and the variability across blocks and subjects, omnibus tests are unlikely to be significant. Indeed, an ANOVA on the slope data did not show any significant effects, though viewing condition was marginally significant ($p = .056$). However, an ANOVA on the composite score used in 3A showed significant viewing condition effect, $F(1,5) = 7.984$ $p=.037$.



4-35: Comparison of performance between stereo and monocular conditions in each block. Performance is measured by the slope (ideal slope of 1). Arrows connect the same subject across the two blocks. Points should lie on the diagonal if stereo performance relied on monocular cues. Only Subject F clusters near the diagonal.

5 AIM 3: PRECLINICAL VALIDATION

5.1 OVERVIEW

Planned preclinical validation of the OCT-MARS included a subjective component consisting of discussions and surveys with surgeons using the system, and an objective component consisting of usability experiments to quantify interaction with the system. The subjective studies were to be geared toward determining what features/information surgeons would like to have provided through the in-situ overlay system during procedures, based on surveys and interviews with surgeons after they use the OCT-MARS. The objective studies would look at how well surgeons can interact with and absorb information from the system.

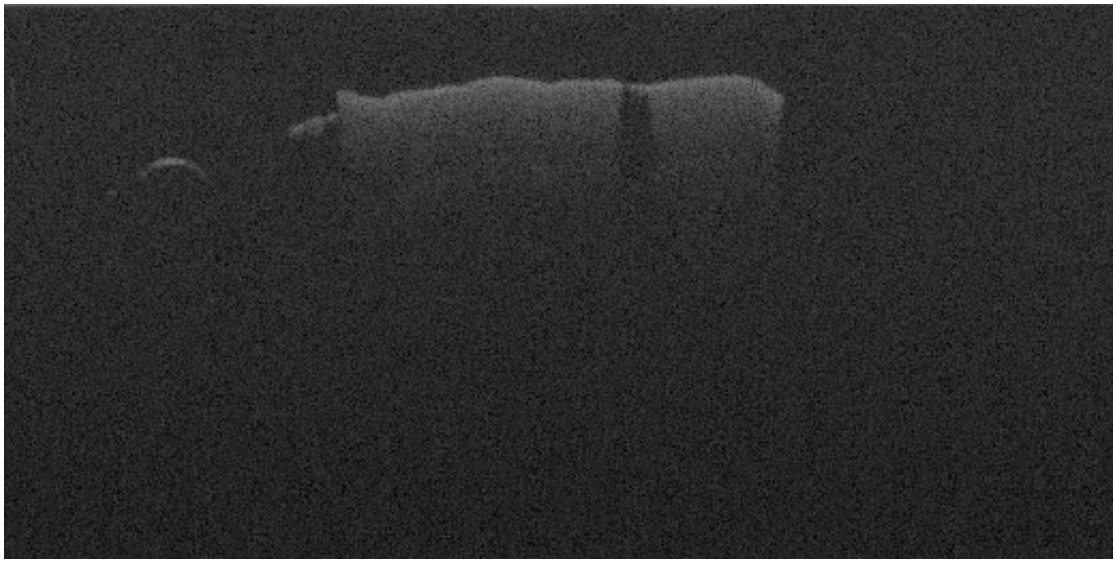
Unfortunately, the image quality limitations of our OCT scanner precluded preclinical experimentation, as already discussed in section 3.6. Substantial time was spent integrating the current scanner with our Zeiss microscope, and replacing the scanner at this time is prohibitively expensive. Described below is our initial work on phantom designs, with images from our current scanner, and with those images overlaid to demonstrate the basic performance of the OCT-MARS. Preclinical research has instead taken the form of psychophysics studies using our own OCT data (see Experiment 3 in section 4.4.)

5.2 CUSTOM PHANTOMS

Our initial test phantoms used a clear gel wax as the base for the structure (shown in Figure 5-1). This wax can be molded at relatively low temperatures and is optically clear. These phantoms were created in silicone molds to approximate eye-shaped masses. Early and late versions of the phantoms are shown in Figures 5-1 and Figure 5-2. Graphite particles and opaque paint were used to create structures within the gel (Figure 5-3). These phantoms were used for most of the initial testing and analysis of our scanner, including dealing with the aforementioned technical issues. Due to the limited imaging capabilities of our OCT system, developing more elaborate custom phantoms would have been non-advantageous.

5-1: Initial prototype of the anterior segment eye phantom

5-2: Blue phantom with paint

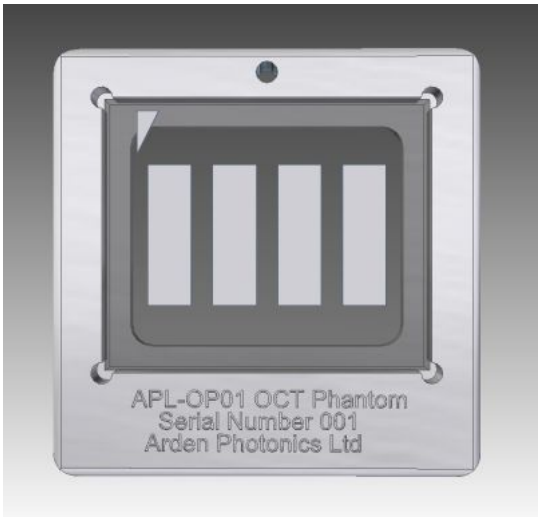


5-3: OCT scan of paint phantom. The surface of the paint stripe can be seen, as well as the top of a needle inserted into the gel

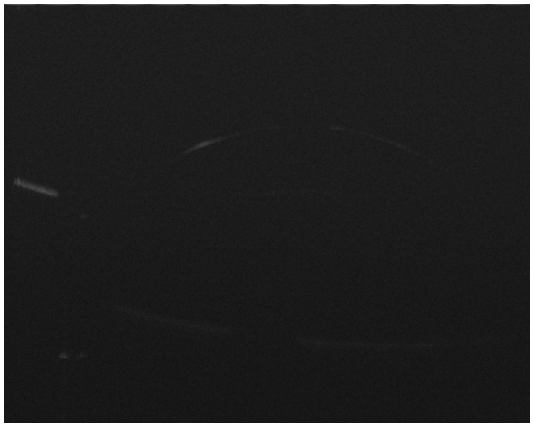
5.3 COMMERCIAL/INDUSTRY PHANTOMS

To assist in the analysis of the performance of the system, we acquired a number of commercially produced phantoms. The first was an anterior chamber phantom (Rowe Technical Design, AC-OCT) that modeled the cornea, iris, and lens (see Figure 5-4) [56]. The second was a high resolution laser etched phantom (Arden Photonics, APL-OP01) to be used for calibration and to quantify OCT image quality (see Figure 5-5). Unfortunately, our scanner was unable to adequately image most subsurface structures in the corneal phantom (Figure 5-6), and could not resolve the full layers of depth in the latter phantom (Figure 5-7). Figures 5-8 and 5-9 shows the quality of images that the phantoms produce with a functioning scanner.

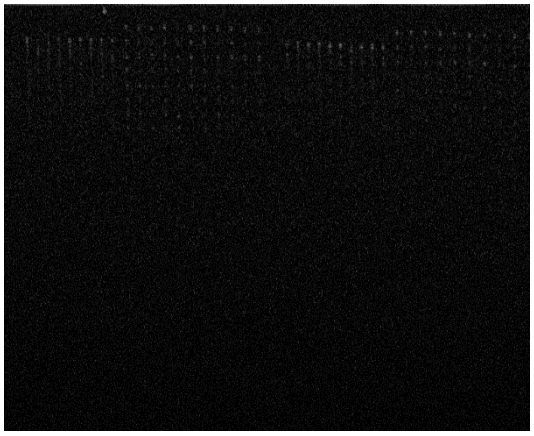
5-4: Anterior chamber phantom



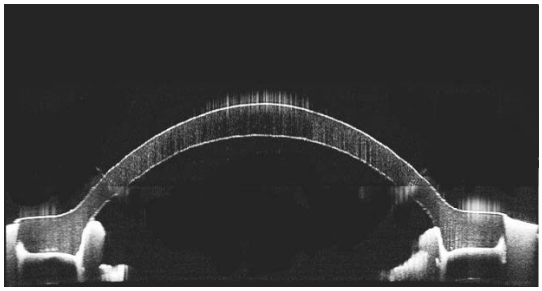
5-5: Precise calibration phantom



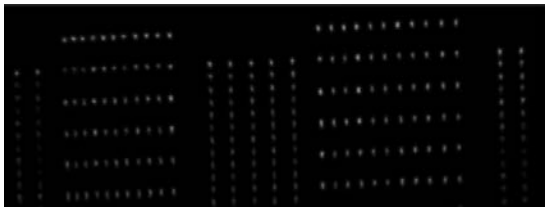
5-6: OCT scan of anterior chamber phantom, using our scanner. The surfaces of the cornea and iris are barely visible



5-7: OCT scan of precise calibration phantom, using our scanner. We can see the signal fall off with depth



5-8: OCT image of anterior chamber phantom taken with a different scanner, from [56].



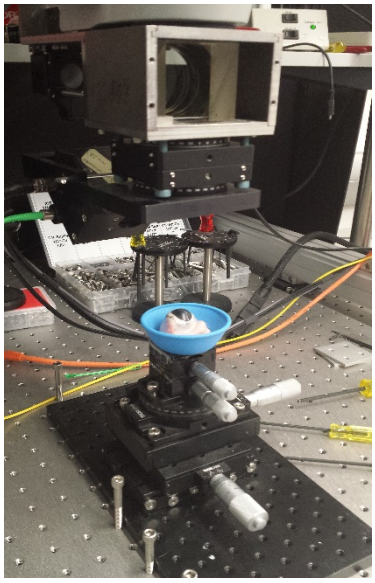
5-9: OCT image of calibration phantom taken with a different scanner, from [57].

5.4 SYSTEM TEST IMAGES

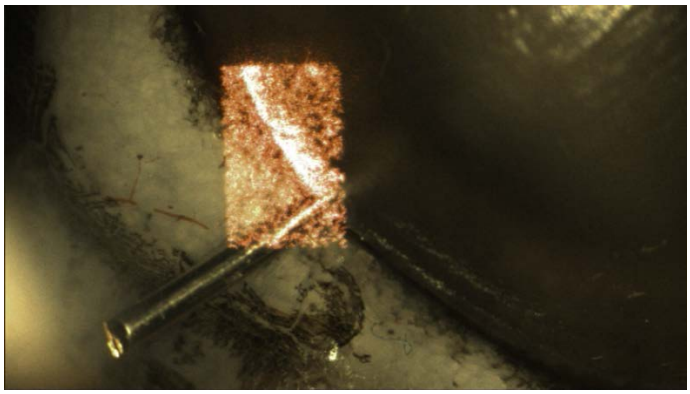
5.4.1 OCT MARS v1

To test the OCT MARS version 1 system on animal tissue, we procured ex-vivo pig eyes from a local butcher. The eyes were harvested and transported to our lab on the same day, maximizing the quality and stability of the tissue. Excess tissue (skin, etc.) was removed and the eyes were stored refrigerated in isotonic saline to be available for 1-2 days of scanning.

Figures 5-10 and 5-11 show the results from the initial ex-vivo pig eye tests. Figure 10 shows the OCT-MARS v1 system scanning an excised pig eye. Figure 5-11 shows the enface OCT overlay as seen through the microscope. The overlay is displayed at the correct scale and depth, and clearly shows the needle as it enters the cornea.



5-10: Assembled OCT_MARS v1, scanning a pig's eye

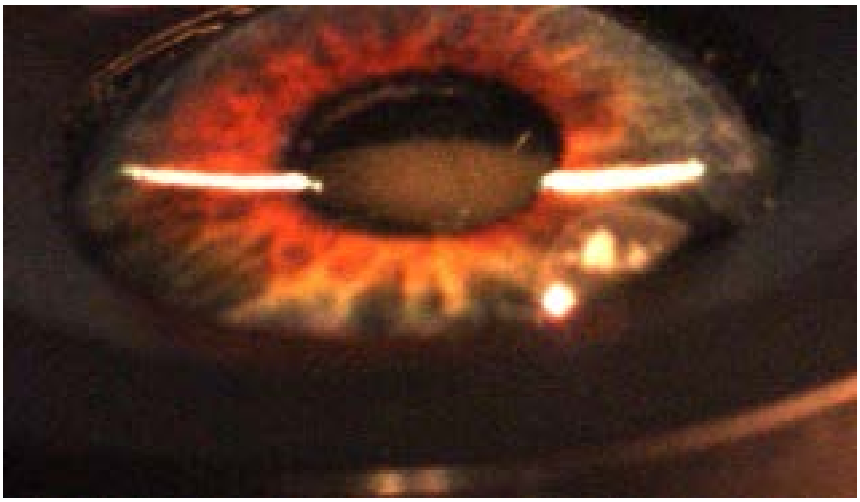


5-11: Through-scope image of the OCT-MARS semi-calibrated overlay, showing a needle entering a pig's eye

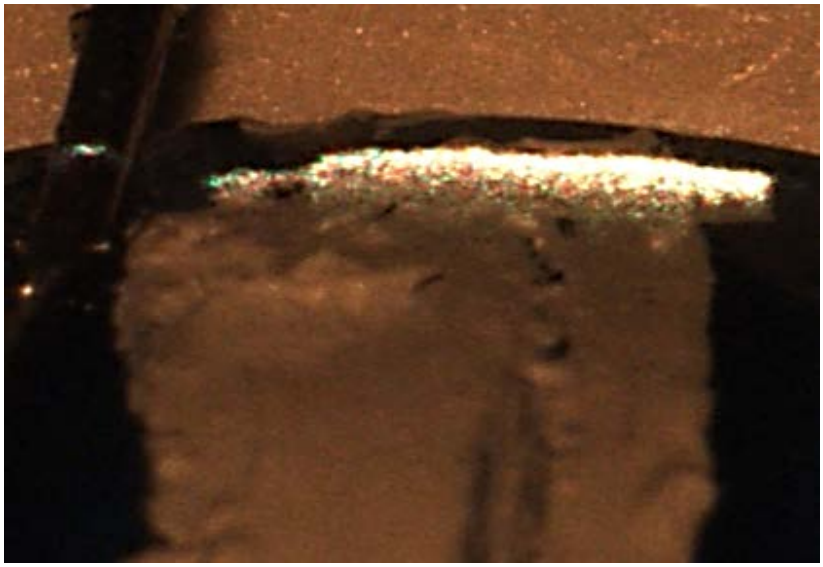
5.4.2 OCT MARS v2

Figures 5-12 and 5-13 show tests of the fully calibrated OCT MARS v2 system. Figure 5-12 shows the calibration procedure of the OCT-MARS v2 design. The tops of the needles can be seen in the overlaid OCT image. Figure 5-13 shows the performance of the system on the anterior chamber phantom, where the overlay highlights the underside of the iris structure in the phantom. Figure 5-14 shows an overlay on one of our custom phantoms.

5-12: OCT MARS v2 with 5 pin calibration phantom. The top surfaces of the pins are highlighted by the overlay.



5-13: OCT MARS v2 with anterior chamber phantom. The overlay is highlighting the underside of the iris, with the pupil opening



5-14: OCT-MARS v2 overlay of paint phantom. The paint stripe is clearly highlighted, along with top of the inserted needle.

5.5 LIMITATIONS OF THIS STUDY

As already stated, because of the limitations of our custom OCT scanner (see section 3.6 and section 5.3), we were unable to perform meaningful preclinical trials with clinicians to determine the performance of the system for actual procedures on pig eyes, which had been intended when this thesis was proposed. Replacing the OCT scanner was prohibitive in both time and monetary expenditure. Instead, effort was focused on expanding Aim 2 into stereomicroscope studies using ophthalmology inspired tools and OCT images (see Experiment 3 in section 4.4). In future work, we plan to expand Experiment 3 with studies on the performance of surgical professionals in our stereo microscope tasks.

6 CONCLUSION

6.1 OCT-MARS IMPLEMENTATION

6.1.1 Current Progress

Significant progress was made on implementing and understanding our *in situ* microsurgical augmented reality system. A modified, refocused scanned laser projector serves as an excellent image source for creating an in-situ image through a microscope. The OCT-MARS was built in two iterations, with version 1 capable of projecting flat, fronto-parallel (with respect to the microscope objective) images. The OCT-MARS version 2 extended the projection capability to tilted images (along two axes) by exploiting the small depth of field of the microscope to provide significant tolerance for the focus of the projection system image. The blur induced by the optics of the microscope exceeds any blur created by tilting the screen in the projection system out of the plane of focus. This optical design was verified both theoretically and using qualitative image analysis.

In the course of building our system, we found that modifications made to the scanner to allow for larger working distance were detrimental to the resolution of the system. Increasing the working distance and scan depth required an increased waist size, reducing resolution and increasing noise.

6.1.2 Future work

Future work on the OCT-MARS will likely be based on the inclusion of adaptive optics. Adaptive optics, such as a phase only spatial light modulator (POSLM) can be used to reposition and reshape the projected image during a surgical procedure. Future work will also require replacing the OCT system, possibly including theoretical advancements and novel optical design for a new scan head capable of quality imaging subject to the distance constraints of in-situ surgical guidance (see section 6.3). Further work in real-time analysis and rendering of the data will also be needed.

6.2 PERCEPTION OF TILT UNDER MAGNIFICATION

6.2.1 Current Progress

We presented an algorithm for creating a printable scale-invariant image that provides stereo information but not pictorial cues. This image allowed subsequent experiments to investigate the difference in tilt perception between stereo and pictorial cues. We found that human subjects can accurately perceive tilt under magnification, while using only stereo information. We also found that human subjects can use ophthalmology-inspired tools to match slants viewed under magnification based on a variety of stimuli, including OCT data of pig corneas acquired in our lab.

6.2.2 Future work

Experiment 3B will be expanded with additional subjects to make the study comparable to Experiment 3A. We also want to repeat Experiment 3B using skilled subjects (surgeons or other medical professionals) instead of naïve campus volunteers. Subjects will be recruited from the Ophthalmology department at the University of Pittsburgh Medical Center (UPMC). We will collect feedback on the usefulness of the OCT-MARS and our experiments along with subject data.

6.3 FUTURE WORK: PRECLINICAL VALIDATION

Work on validating the OCT-MARS for clinical tasks will require replacing the PSI OCT scanner. Advances in swept-source OCT systems since the inception of this project have made

available scanners with significantly better resolution and scan depth at the same cost level as our PSI system. This may allow the production of images of sufficient quality useful in clinical settings. Once an updated system is functioning, the research can move ahead with the studies originally planned, described in section 5.1.

6.4 EVENTUAL CLINICAL IMPORTANCE

Once fully developed, the general approach of in-situ display of real-time OCT images may have a significant effect on ophthalmological surgery. In-situ virtual image displays have the advantage of creating true 3D images that do not rely on stereo rendering to create depth. We have shown that basic 3D structure can be perceived under stereo magnification, which is promising for the use of virtual images in augmented reality. This is an excellent launch point for further development in 3D virtual image generation and further research on human perception of magnification.

7 WORKS CITED

[1] M. Wojtkowski, "High-speed optical coherence tomography: basics and applications," *Applied Optics,* vol. 49, no. 16, pp. D30-D61, 2010.

[2] K. Zhang and J. U. Kang, "Real-time intraoperative 4D full-range FD-OCT based on the dual graphics processing units architecture for microsurgery guidance," *Biomedical optics express,* vol. 2, no. 4, pp. 764-770, 2011.

[3] S. I. Tomarev, "Eyeing a new route along an old pathway," *Nature medicine,* vol. 7, no. 3, pp. 294-295, 2001.

[4] R. JMJ, " Leading causes of blindness worldwide," *Bull Soc Belge Ophtalmol,* vol. 283, pp. 19-25, 2002.

[5] M. C. Leske, A. Heijl, M. Hussein, B. Bengtsson, L. Hyman and E. Komaroff, "Factors for glaucoma progression and the effect of treatment: the early manifest glaucoma trial," *Archives of ophthalmology,* vol. 121, no. 1, pp. 48-56, 2003.

[6] L. Kagemann, G. Wollstein, H. Ishikawa, R. A. Bilonick, P. M. Brennen, L. S. Folio and J. S. Schuman, "Identification and assessment of Schlemm's canal by spectral-domain optical coherence tomography," *Investigative ophthalmology & visual science,* vol. 51, no. 8, pp. 4054-4059, 2010.

[7] R. Sharma, A. Sharma, T. Arora, S. Sharma, A. Sobti, B. Jha and T. Dada, "Application of anterior segment optical coherence tomography in glaucoma," *Survey of ophthalmology,* vol. 59, no. 3, pp. 311-327, 2014.

[8] A. L. Coleman, "Advances in glaucoma treatment and management: surgery," *Investigative ophthalmology & visual science,* vol. 53, no. 5, pp. 2491-2494, 2012.

[9] D. W. F. Van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," *International Journal of Virtual Reality,* vol. 9, no. 2, p. 1, 2010.

[10] R. Wen, W. L. Tay, B. P. Nguyen, C. B. Chng and C. K. Chui, "Hand gesture guided robot-assisted surgery based on a direct augmented reality interface," *Computer methods and programs in biomedicine,* vol. 116, no. 2, pp. 68-80, 2014.

[11] X. Kang, M. Azizian, E. Wilson, K. Wu, A. D. Martin, T. D. Kane and R. Shekhar, "Stereoscopic augmented reality for laparoscopic surgery," *Surgical endoscopy,* vol. 28, no. 7, pp. 2227-2235, 2014.

[12] F. Volonté, N. C. Buchs, F. Pugin, J. Spaltenstein, B. Schiltz, M. Jung and P. Morel, "Augmented reality to the rescue of the minimally invasive surgeon. The usefulness of the interposition of stereoscopic images in the Da Vinci™ robotic console," *The International Journal of Medical Robotics and Computer Assisted Surgery,* vol. 9, no. 3, pp. e34-e38, 2013.

[13] B. C. Becker, R. A. MacLachlan, G. D. Hager and C. N. Riviere, "Handheld micromanipulation with vision-based virtual fixtures," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011.

[14] J. P. Ehlers, P. K. Kaiser and S. K. Srivastava, "Intraoperative optical coherence tomography using the RESCAN 700: preliminary results from the DISCOVER study," *British Journal of Ophthalmology,* 2014.

[15] L. Shen, O. Carrasco-Zevallos, B. Keller, C. Viehland, G. Waterman, P. Desouza and J. A. Izatt, " Novel microscope-integrated stereoscopic display for intrasurgical optical coherence tomography," in *SPIE BiOS* , 2015.

[16] P. N. Dayani, R. Maldonado, S. Farsiu and C. A. .. Toth, " Intraoperative use of handheld spectral domain optical coherence tomography imaging in macular surgery," *Retina,* vol. 29, no. 10, p. 1457, 2009.

[17] R. Ray, D. E. Barañano, J. A. Fortun, B. J. Schwent, B. E. Cribbs, C. S. Bergstrom and S. K. Srivastava, "Intraoperative microscope-mounted spectral domain optical coherence tomography for evaluation of retinal anatomy during macular surgery," *Ophthalmology,* vol. 118, no. 11, pp. 2212-2217, 2011.

[18] G. D. Stetten.United States of America Patent 6,599,247, 2003.

[19] J. J. Gibson, The perception of the visual world, Boston: Haughton Mifflin, 1950.

[20] W. P. Purdy, The hypothesis of psychophysical correspondence in space perception, Dissertation Abstracts International, 1958.

[21] J. T. Todd, L. Thaler, T. M. H. Dijkstra, J. J. Koenderink and A. M. L. Kappers, "The effects of viewing angle, camera angle and sign of surface curvature on the perception of 3D shape from texture," *Journal of Vision,* vol. 7, no. 12, p. 1–16, 2007.

[22] J. T. Todd and L. Thaler, "The perception of 3D shape from texture based on directional width gradients," *Journal of Vision,* vol. 5, pp. 1-13, 2010.

[23] Vishwanath and Hibbard, "Seeing in 3-D with just one eye: Stereopsis without binocular vision," *Psychological Science,* vol. 4, no. 9, pp. 1673-1685, 2013.

[24] W. E. L. Grimson, "A computer implementation of a theory of human stereo vision (A.1. Memo No. 565)," M.I.T Artificial Intelligence Laboratory, Cambridge, MA, 1980.

[25] D. Marr and T. Poggio, "A computational theory of human stereo vision.," *Proceedings of the Royal Society,* vol. Series B, no. 204, pp. 301-328, 1979.

[26] J. E. W. Mayhew and J. P. Frisby, "The computation of binocular edges," *Perception,* vol. 9, pp. 69-86, 1979.

[27] J. I. Nelson, "Globality and stereoscopic fusion in binocular vision," *Journal of Theoretical Biology,* vol. 49, pp. 1-80, 1975.

[28] T. O. Aydin, M. Cadik, K. Myszkowski and H.-P. Seidel, "Visually significant edges," *ACM Transactions on Applied Perception,* vol. 7, no. 4, pp. 1-15, 2010.

[29] B. Gillam, T. Flagg and D. Finay, "Evidence for disparity change as the primary stimulus for stereoscopic processing," *Attention, Perception, and Psychophysics,* vol. 36, pp. 559-564, 1994.

[30] B. Gillam and M. J. Pianta, "The effect of surface placement and surface overlap on stereo slant contrast and enhancement," *Vision Research,* vol. 45, pp. 3083-3095, 2005.

[31] A. P. Witkin, "Scale-space filtering," in *Eighth International joint Conference on Artifical Intelligence - Volume 2*, San Francisco, CA, USA, 1983.

[32] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, no. 5, pp. 363-370, 1984.

[33] X. Zabulis and B. Backus, "Starry night: A texture devoid of depth cues," *Jornal of the Optical Society of America*, vol. 11, pp. 2049-2060, 2004.

[34] B. Wu, R. Klatzky and J. Galeotti, "Effects of magnification on depth perception and visually-guided reaching," *Journal of Vision*, vol. 12, p. 220, 2013.

[35] J. Galeotti, A. Sajjad, B. Wang, L. Kagemann, G. Shukla, M. Siegel and G. Stetten, " The OCT penlight: In-situ image guidance for microsurgery," *SPIE Medical Imaging*, pp. 762502-762502, 2010.

[36] S. Horvath, J. Galeotti, M. Siegel and G. Stetten, "Refocusing a scanned laser projector for small and bright images: simultaneously controlling the profile of the laser beam and the boundary of the image," *Applied optics*, vol. 53, no. 24, pp. 5421-5424, 2014.

[37] T. Mathai, J. Galeotti, S. Horvath and G. Stetten, "Graphics processor unit (GPU) accelerated shallow transparent layer detection in Optical Coherence Tomographic (OCT) images for real-time corneal surgical guidance," *Lecture Notes in Computer Science*, vol. 8678, pp. 1-13, 2014.

[38] M. Freeman, M. Champion and S. Madhavan, "Scanned Laser Pico-Projectors: Seeing the Big Picture (with a Small Device," *Opt. Photon. News*, vol. 20, pp. 28-34, 2009.

[39] S. A. Self, "Focusing of spherical Gaussian Beams," *Applied Optics*, vol. 22, no. 5, pp. 658-661, 1983.

[40] B. Gillam, "Judgements of slant on the basis of foreshadowing," *Scandinavian Journal of Psychology*, vol. 11, pp. 31-34, 1970.

[41] B. B. Mandelbrot, "Fractal aspects of the iteration of z ◊ λz(1 – z) for complex λ and z," *Annals of the New York Academy of Sciences*, vol. 357, pp. 249-259, 1980.

[42] W. Sierpinski, "Sur une courbe dont tout point est un point de ramification," *Computational Rendus Academy of Science Paris*, vol. 16, pp. 302-305, 1915.

[43] J. Galeotti, K. Macdonald, J. Wang, S. Horvath, A. Zhang and R. Klatzky, "Generating an image that affords slant perception from stereo, without pictorial cues," *Displays*, 2016.

[44] V. A. Billock, "Neural acclimations to 1/f spatial frequency spectra in natural images transduced by the human visual system," *Physica D*, vol. 137, pp. 379-391, 2000.

[45] D. Knill, D. Field and D. Kersten, "Human discrimination of fractal images," *Journal of the Optical Society of America*, vol. 7, pp. 1113-1123, 1990.

[46] M. Haidekker, Advanced Biomedical Image Analysis, New York: John Wiley & Sons, 2011.

[47] H. Tan, H. Z. Tan, N. I. Durlach, W. M. Rabinowitz and C. M. Reed, "Information transmission with a multifinger tactual display," *Perception & Psychophysics*, vol. 61, no. 6, pp. 993-1008, 1999.

[48] L. D. Harmon and B. Julesz, "Masking in visual recognition: Effects of two-dimensional filtered noise," *Science,* vol. 180, pp. 1194-1197, 1973.

[49] D. C. Knill, "Discrimination of planar surface from texture: Human and ideal observers compared," *Vision Research,* vol. 38, pp. 1683-1711, 1998.

[50] D. Knill and J. A. Saunders, "Do humans optimally integrate stereo and texture information for judgements of surface slant?," *Vision Research,* vol. 43, pp. 2539-2558, 2003.

[51] J. T. Todd, L. Thaler and T. M. H. Dijkstra, "The effects of field of view on the perception of 3D slant from texture," *Vision Research,* vol. 45, pp. 1501-1517, 2005.

[52] J. T. Todd, J. C. Christensen and K. M. Gukes, "Are discrimination threasholds a valid measure of variance for judgements of slant from texture?," *Journal of Vision,* vol. 10, no. 2, pp. 1-18, 2010.

[53] C. Kaernbach, "Adaptive threshold estimation with unforced-choice tasks," *Perception & Psychophysics,* vol. 63, no. 8, pp. 1377-1388, 2001.

[54] D. M. Shaffer, E. McManama and F. H. Durgin, "Manual anchoring biases in slant estimation affect matches even for near surfaces," *Psychonomic bulletin & review,* vol. 22, no. 6, pp. 1665-1670, 2015.

[55] F. K. Freyberger, M. Kuschel, B. Farber, M. Buss and R. L. Klatzky, "Tilt Perception by Constant Tactile and Constant Proprioceptive Feedback through a Human System Interface," in *WHC*, 2007.

[56] T. S. Rowe and R. J. Zawadzkib, "Development of a corneal tissue phantom for anterior chamber optical coherence tomography (AC-OCT)," in *Proc. of SPIE Vol 8583*, 2013.

[57] N. K. Ravichandran, R. E. Wijesinghe, M. F. Shirazi, K. Park, M. Jeon, W. Jung and J. Kim, "Depth enhancement in spectral domain optical coherence tomography using bidirectional imaging modality witha single spectrometer," *Journal of Biomedical Optics,* vol. 21, no. 7, p. 076005, 2016.

[58] B. Gillam and C. Ryan, "Perception, orientation disparity, and anisotropy in stereoscopic slant perception," *Perception,* vol. 21, pp. 427-439, 1992.