



University of Kentucky
UKnowledge

Theses and Dissertations--Electrical and
Computer Engineering

Electrical and Computer Engineering

2019

Spatial Augmented Reality Using Structured Light Illumination

Ying Yu

University of Kentucky, yyu226@g.uky.edu

Digital Object Identifier: <https://doi.org/10.13023/etd.2019.384>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Yu, Ying, "Spatial Augmented Reality Using Structured Light Illumination" (2019). *Theses and Dissertations--Electrical and Computer Engineering*. 143.
https://uknowledge.uky.edu/ece_etds/143

This Doctoral Dissertation is brought to you for free and open access by the Electrical and Computer Engineering at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Electrical and Computer Engineering by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Ying Yu, Student

Dr. Daniel L. Lau, Major Professor

Dr. Aaron Cramer, Director of Graduate Studies

Spatial Augmented Reality Using Structured Light Illumination

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the

College of Engineering

at the University of Kentucky

By

Ying Yu

Lexington, Kentucky

Co-Directors: Dr. Daniel L. Lau, Professor of Electrical and Computer Engineering

and Dr. J. Todd Hastings, Professor of Electrical and Computer Engineering

Lexington, Kentucky 2019

Copyright© Ying Yu 2019

ABSTRACT OF DISSERTATION

Spatial Augmented Reality Using Structured Light Illumination

Spatial augmented reality is a particular kind of augmented reality technique that uses projector to blend the real objects with virtual contents. Coincidentally, as a means of 3D shape measurement, structured light illumination makes use of projector as part of its system as well. It uses the projector to generate important clues to establish the correspondence between the 2D image coordinate system and the 3D world coordinate system. So it is appealing to build a system that can carry out the functionalities of both spatial augmented reality and structured light illumination.

In this dissertation, we present all the hardware platforms we developed and their related applications in spatial augmented reality and structured light illumination. Firstly, it is a dual-projector structured light 3D scanning system that has two synchronized projectors operate simultaneously, consequently it outperforms the traditional structured light 3D scanning system which only include one projector in terms of the quality of 3D reconstructions. Secondly, we introduce a modified dual-projector structured light 3D scanning system aiming at detecting and solving the multi-path interference. Thirdly, we propose an augmented reality face paint system

which detects human face in a scene and paints the face with any favorite colors by projection. Additionally, the system incorporates a second camera to realize the 3D space position tracking by exploiting the principle of structured light illumination.

At last, a structured light 3D scanning system with its own built-in machine vision camera is presented as the future work. So far the standalone camera has been completed from the a bare CMOS sensor. With this customized camera, we can achieve high dynamic range imaging and better synchronization between the camera and projector. But the full-blown system that includes HDMI transmitter, structured light pattern generator and synchronization logic has yet to be done due to the lack of a well designed high speed PCB.

KEYWORDS: spatial augmented reality, dual-projector, structured light, multi-path, face paint

Author's signature: _____ Ying Yu

Date: _____ September 30, 2019

Spatial Augmented Reality Using Structured Light Illumination

By

Ying Yu

Co-Director of Dissertation: Daniel L. Lau

Co-Director of Dissertation: J. Todd Hastings

Director of Graduate Studies: Aaron Cramer

Date: September 30, 2019

ACKNOWLEDGMENTS

I would like to thank all these people who helped and supported me toward the completion of my PhD study.

First of all, I am so grateful to Dr. Daniel L. Lau who guided me into the area of computer vision, more specifically the 3D imaging. He led me to a few interesting and practical projects where I can stick to my personal interest, hardware development. Also he taught me how the textbook knowledge is applied to real world engineering projects, how the scientific articles should be written, and how I can learn carrying out academic research independently. More importantly, he spent a great amount of time working with me on concrete tasks in our projects instead of just sitting down and giving general instructions, which impressed and motivated me to work harder.

Secondly, I want to thank all other faculty members in my committee, including Dr. Todd Hastings, Dr. Ruigang Yang, Dr. Caicheng Lu and Dr. Robert Heath. They provided valuable advices on how I should proceed with the research and pointed out the weaknesses which I can improve in my future career.

Thirdly, I would like to thank Matthew P. Ruffner who cooperated with me in almost every project I was involved. He saved me a lot of time by inspiring me in hardware and software debugs as well as designing PCBs for our projects.

Finally, I want to thank my parents and my wife. Without their long-term patience and encouragement, I could not have persisted till the day when I obtained my degree.

TABLE OF CONTENTS

Acknowledgments	iii
Table of Contents	iv
List of Figures	vi
List of Tables	xii
Chapter 1 Introduction	1
1.1 Spatial Augmented Reality	6
1.2 Structured Light Illumination	9
1.3 Our Contribution	17
Chapter 2 Dual-projector Structured Light 3D Scanning System	29
2.1 Dual-frequency pattern scheme	31
2.2 Dual-Projector Scanning	35
2.3 Calibration	44
2.4 Experiments	45
2.5 Summary	53
Chapter 3 Detecting and Removing Multi-path in Structured Light Scanning	
Using A Dual-projector System	55
3.1 Bimodal Multi-Path	57

3.2	Dual-projector system setup	62
3.3	Experiments	67
3.4	summary	70
Chapter 4 AR Face Paint and Headset Tracking		72
4.1	HDMI Receiver	74
4.2	Synchronization Scheme for SAR	78
4.3	Experiments	82
4.4	Summary	88
Chapter 5 Conclusions and Future Work		90
Bibliography		106
Vita		126
Education		126
Professional Experience		126
Publications		126

LIST OF FIGURES

1.1	AR applications: A yellow virtual line in NFL broadcast [1](left), A rear view camera incorporated with AR [2](right)	2
1.2	IR constellation points scatter on the surface the Oculus Rift [3](left) and HTC lighthouse(right)	4
1.3	AR sandbox(left) and AR pool table [4](right)	7
1.4	Illustration of an SLI scanner	11
1.5	A taxonomy of SLI 3D surface imaging techniques	12
1.6	Wrapped phases of low (left), medium (center) and high (right) frequencies SLI scans	15
1.7	Unwrapped phase of the scene	16
1.8	input-output curves of gamma encoding (left) and decoding (right) . . .	20
1.9	System configuration of the dual-projector structured light scanning system	22
1.10	Experimental results of the dual-projector scanner that show the improvement in terms of increased SNR and the removal of occlusion and multi-path	23
1.11	The scanned wedge (top) and the 3D reconstruction without any treatment for multi-path (bottom-left), replace the multi-path with foreground path (bottom-center), replace the multi-path with background path (bottom-right)	24
1.12	An screenshot of our customized EDID recognized by Windows 10	26
1.13	An 2D human face before (left) and after (right) face paint	28

2.1	Example structured light scanner architectures illustrating: (a) a traditional single projector, single camera scanner; (b) dual-projector scanner with projectors in opposing positions in order to minimize scan occlusions; (c) dual-projector scanner with projectors in complementary positions for increasing projected image brightness; and (d) dual-projector scanner with projectors in stacked orientation for detecting multi-path.	31
2.2	Block diagram of HDMI [5]	37
2.3	System diagram of FPGA controller showing the GPIO interface to the machine vision camera and HDMI outputs to the component projectors.	40
2.4	Timing diagram of the signals for camera synchronization where “RGB” represents the pixel buss of the HDMI signal where the pixels associated with each component scan frame labeled as “Frame 0” to “Frame $N - 1$ ”, “V” denotes the positions in time of vertical syncs, LINEs 1-4 represent the GPIO signals between FPGA and camera, and the letters A , B , and C are specific events corresponding to the handshake between camera and FPGA during a scan.	41
2.5	The light output for the red, green, and blue LEDs in a (top) TI DLP3010EVM-LC and a (bottom) Optoma ML750ST projector. Vertical dashed lines indicate vertical syncs for 120Hz video with the green LED of the TI DLP3010EVM-LC projector overlapping vertical syncs where the Optoma ML750ST keeps the color patterns inside the syncs. Shown as gray boxes are the moments in time when the camera is exposing its sensor.	41
2.6	Illustration showing dual-projector scanner calibration.	44

2.7	One sample CalTag in our system	45
2.8	The (left) dual-projector scanner and the (right) scanning target composed of a 22.9" diameter tire with steel rim, both painted with matte white; a plastic makeup artist face/head; a non-slip area rug pad; and an occluding 4" wide, 0.25" thick piece of hobby wood.	46
2.9	Phase gradient of the reference plane for projector 1 obtained when only projector 1 is scanning (left), and while both projectors are operating (right)	48
2.10	3D reconstruction with (left) only the projector at the bottom; (center-left) with only the projector at the top; (center-right) by merging previous two using weighted averaging scheme; (right) after removing the pixels with multi-path	49
2.11	The raw and gradient images (left) the unit frequency phase from projector 1; (center-left) the high frequency phase from projector 2; (center-right) the predicted unit frequency of projector 2; (right) unwrapped high frequency phase of projector 2.	52
2.12	3D reconstructions (left) before and (right) after bimodal multi-path removal.	53
3.1	Illustration of bimodal multi-path for (left) time-of-flight imaging, (center) conventional structured light imaging, and (right) our dual-projector structured light imaging.	57

3.2	Illustration of the change in direction and magnitude, on the (B^R, B^I) -axis, in the (blue) observed complex vector $B^R + jB^I$ created by the superposition of (red) complex vectors from multi-path fore and background objects for $k = 1$, $k = 8$, and $k = 12$	59
3.3	Illustration of the change in (vertical axis) magnitude of the observed complex vector, $B^R + jB^I$, as a function of the (horizontal axis) scaling factor k	59
3.4	The dual-projector scanner we built for multi-path	67
3.5	The scanned wedge (top) and the 3D reconstruction without any treatment for multi-path (bottom-left), replace the multi-path with foreground path (bottom-center), replace the multi-path with background path (bottom-right)	68
3.6	The scanned teeth mold (top) and the 3D reconstruction without any treatment for multi-path (bottom-left), 3D reconstruction after processing with the Algorithm 2 (bottom-right)	69
3.7	The magnitude maps of the wedge with multi-path (top-left), foreground path (top-center), background path (top-right) and the magnitude maps of the teeth mold with multi-path (bottom-left), foreground path (bottom-center), background (bottom-right)	70
4.1	The block diagram of the HDMI receiver [6]	75
4.2	The process of EDID handshake [7]	76
4.3	The flowchart of the EDID handshake	77

4.4	The distinct duty cycles of positive and negative polarities	79
4.5	Uncertain delay (top) and replicates of frames (bottom) happen in graphics card	80
4.6	Camera triggers (signal in yellow) recorded for a sequence of 8 SLI patterns in SAR mode	81
4.7	The internal timing of HDMI at a resolution of 800×600	81
4.8	The hardware platform of face paint system	83
4.9	Face paint with red dots (left) and the same face painted with dark hair (right)	84
4.10	The unit composite pattern (top-left) with its horizontal/vertical fringe components (top-center/top-right) and the high frequency composite pattern (bottom-left) with its horizontal/vertical fringe components (bottom-center/bottom-right)	85
4.11	The mapping between the master camera and the slave camera	87
4.12	The experimental setup of the position tracking	88
4.13	The 3D reconstruction of the scanned scene including the positions of both master and slave cameras (left and center), 3D reconstruction of the scanned scene just recording one slave camera position and orientation(right)	88
5.1	The latest version of PCB that has been made (left) and the system diagram of the camera (right)	95

5.2	The PCB layout of the CCC showing high speed traces (red) and PDN (blue). Power pours on one internal plane are shown in light gray, while the all-over ground pour on the other internal plane is not shown	97
5.3	A sequence of consecutive six images taken when the non-de	100
5.4	Data from the image sensor to host PC.	102
5.5	A sample image with fixed pattern noise(left) and the same image after the removal of fixed pattern noise(right)	104

LIST OF TABLES

1.1	List of popular industrial DLP light engines available on the market [8] .	18
-----	--	----

Chapter 1 Introduction

With recent advances in computer hardware and software, virtual reality (VR) [9, 10, 11] has drawn more and more attention [12]. According to Burdea, virtual reality is a simulation in which computer graphics is used to create a realistic-looking world. Moreover, the synthetic world is not static but responds to the user's input, which defines one key feature of virtual reality, real-time interactivity [13]. The most common examples of VR devices are these various kinds of headsets that are sold as commercial products, such as the Samsung Gear VR, HTC Vive, Oculus Rift, etc.

Augmented reality (AR) is another technology to VR that dates back to 1968 when Ivan Sutherland developed a head-mounted display system that could show the user a virtual drawing [14]. According to Azuma [15, 1], a typical AR system can be defined as a system that complements the the real world with computer-generated information or virtual objects. In a more general sense, an augmented reality system enhances the users' experience by enabling the users to see, hear, smell or touch something that does not actually exist. Unlike VR systems that completely remove the real environment, AR systems blend the real and virtual worlds together.

Augmented Reality has a wide range of applications including entertainment [16]; medical [17]; education [18]; military [19]; and others. The first functioning AR system was known as virtual fixtures and was developed at the US Air Force Research Armstrong Laboratory [20]. The objective of the system was to train pilots to improve their performance in direct and remotely manipulated tasks. In sports, AR has been

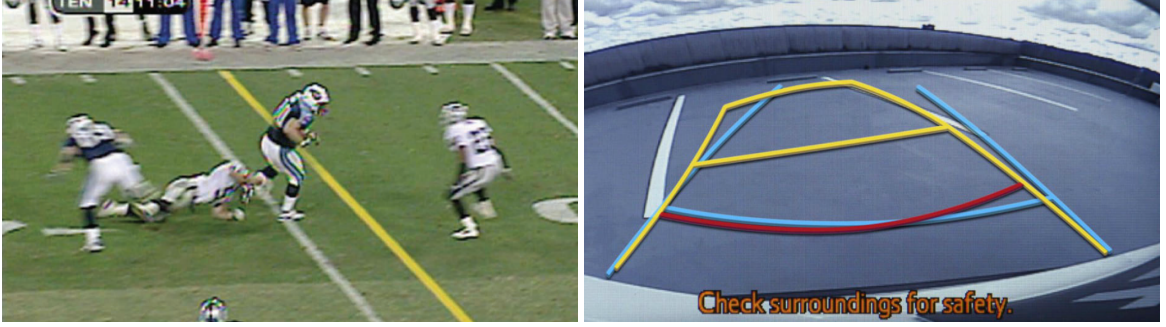


Figure 1.1: AR applications: A yellow virtual line in NFL broadcast [1](left), A rear view camera incorporated with AR [2](right)

used to superimpose virtual objects like first-down lines on a football field as well as highlight a hockey puck with streaks when in high-speed motion from a slap shot [21]. Figure 1.1 (left) shows the illustration of a virtual yellow line during an NFL broadcast showing the first down marker. Some similar applications can also be found in other sports including tennis [22], car racing [23], and so on.

Another common application of AR is the backup camera, equipped in most luxury vehicles that, when moving in reverse, illustrates the path of the left and right side of the vehicle as illustrated in Fig. 1.1 (right). This outline helps the driver get a better idea about how far the obstructions are on the basis of the physical size of the vehicle and the anticipated position of the vehicle according to the current orientation of the steering wheel. Thus the chances of getting the vehicle scratched or any other undesired accidents are reduced [24].

Wearable head-mounted displays for AR/VR applications have become especially popular with the release of the Oculus Rift and HTC Vive. These headsets have been specifically marketed for video games and include a pair of vibrant OLED displays with a combined resolution of 2160×1200 and a refresh rate of 90 Hz [25]. Augmented

reality specific headsets include the Microsoft HoloLens with a see-through optical head-mounted display made of three layers of glasses [26] and the Magic Leap which reportedly features a six-layer waveguide display [27].

In either VR or AR headsets, accurate positional tracking of the headset is of utmost importance because the system needs to know the exact location and orientation of the headset as the user moves in the virtual or real environment [28] so that the host processing unit can render the current field-of-view. Furthermore, position tracking can help the users of VR/AR headsets overcome the motion sickness which is a common but unpleasant experience for them [29].

Because of the importance for accurate head tracking, the Oculus Rift employs a set of sensors including a gyroscope, an accelerometer, and a magnetometer to track the orientation of the headset [30]. To obtain location data, a constellation positional tracking system uses an array of invisible near-infrared (NIR) LEDs located under the external plastic of the headset with an array of desktop NIR cameras tracking these LEDs like stars forming constellation.

The HTC Vive uses a positional tracking system called Lighthouse, which involves two laser emitters placed at predefined locations that periodically transmit laser stripes that sweep both horizontally and vertically in their independent directions [31]. Meanwhile, an array of photosensors mounted on the headset and controllers capture the laser beam at a particular sweep angle with respect to the emitter and measuring the time differences, the location and orientation of the headset can be determined in a certain 3D space [32].

Another VR head tracker used by the Open Source Virtual Reality (OSVR) is an

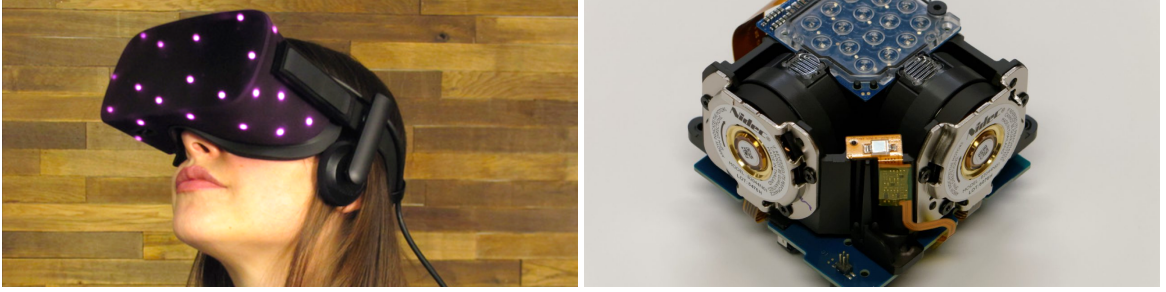


Figure 1.2: IR constellation points scatter on the surface the Oculus Rift [3](left) and HTC lighthouse(right)

electromagnetic (EM) tracking system from Polhemus [33]. Polhemus uses an EM source to generate a tuned electromagnetic dipole field such that by measuring the induced voltages on three orthogonal coils, placed within the given electromagnetic field, the position and the orientation of the coils can be derived following Faraday's law [34, 35]. The Polhemus tracking system is a proven technology that has been used commercially, for example, the fastrak system used for position and orientation tracking in the VR haptic workstation, the patriot system used in Volkswagen spray paint gun color and trim exhibit, the liberty latus system used for helicopter trainer etc [36]. However, the EM tracking system is not perfect, its cost is too expensive comparing to other tracking techniques mentioned earlier, more importantly, the accuracy will be seriously degraded by metal and electronic equipment located in the environment, because they can cause magnetic field distortions [37].

Of the wearable, head-mounted, AR devices, the Microsoft HoloLens includes not only a head-mounted display and sensing hardware but also all necessary computing hardware for its AR applications [38]. Specifically, the Microsoft HoloLens is equipped with a set of surprisingly powerful processing components including an Intel Cherry Trail SoC, which contains the CPU and GPU, and a holographic processing unit

(HPU) that runs as a co-processor integrating data from the sensors and managing tasks like spatial mapping, gesture recognition, voice recognition etc [39]. The capability and the users' feedback of the HoloLens are so appealing that the U.S. army has decided to arm the soldiers with a ruggedized version of HoloLens called integrated visual augmentation system (IVAS) in the future. This special version of HoloLens looks nothing different from a commercial version except for a thermal camera mounted on top which is made by FLIR [40]. The thermal camera can enable the vision for soldiers at night and help them see through the smoke and bushes. Moreover, the IVAS is capable of identifying allies and enemies at night, recognizing the shape of a person. At last, the IAVS is harder to spot while the regular night vision goggles glow in the dark [41].

The Magic Leap headset includes a 6-core CPU and a 256-core GPU from NVIDIA, as well as a linear resonant actuator haptic device [42]. Google has also developed its own Google glass which looks nothing more than an ordinary pair of eye glasses in terms of size and shape, but Google managed to fit a variety of hardware into a small box mounted on the right hand side of the glass, including a dual-core CPU from Texas Instruments, a 5MP camera, a 3-axis gyroscope, a 3-axis accelerometer, a 3-axis magnetometer, an ambient light sensing and proximity sensor, plus a Himax liquid crystal on silicon (LCoS) display which projects videos directly into wearer's eye through a series of complex procedures [43]. Other notable AR headsets include the M300 from Vuzix, Moverio BT-300 from EPSON, focals from North and so on [44, 45].

1.1 Spatial Augmented Reality

Regardless of the tracking technology, there are applications for augmented reality such as medical surgery where the wearing of headsets is prohibitively distracting from the task at hand. To alleviate the need to wear headsets, spatial augmented reality is a potential alternative where digital projectors are used to fuse real objects with virtual content. Projector-based AR systems are generally referred to as spatial augmented reality (SAR) [46, 47]. One typical application of SAR is the AR sandbox, which was developed at the University of California Davis and consists of a traditional child's sandbox with a projector mounted above and looking down on to the sand. A depth sensor is also positioned above the sand and looking down to collect the topographic shape of the surface of the sand. From the topographic data, different colors of light are projected on to and in registration with the sand [48]. The projection changes responsively to changes in topography. Inspired by this invention, various AR sandboxes have been created to benefit geodesign [49] and pilot training [50].

Unlike the AR Sandbox at the University of California Davis which runs on Linux OS with a special software package [51], the AR Sandbox developed here at the University of Kentucky runs on Windows. As such, it takes advantage of the latest Microsoft Kinect V2 camera with twice the field of view of the Kinect V1, so the size of the sandbox can potentially be built twice as large as the one at UC Davis. In addition, the depth video is fed to the GPU directly which greatly increases the speed of data processing, the projected images update in real time at 30 fps. Figure 1.3 (left) shows the AR sandbox that we built at University of Kentucky.

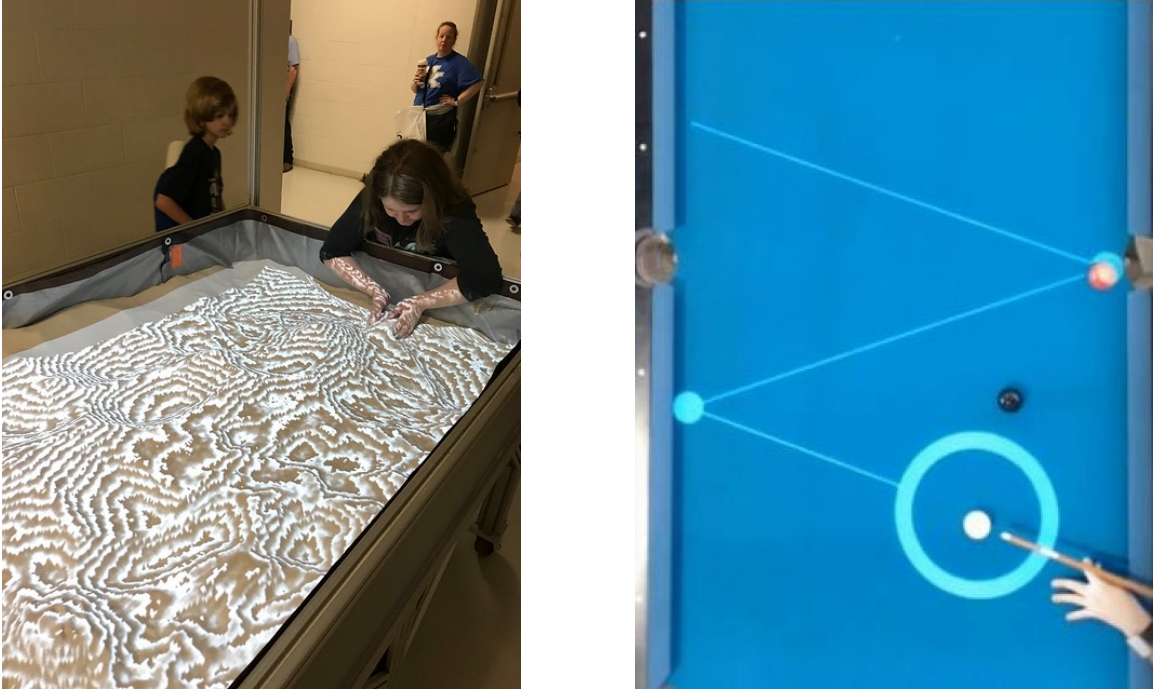


Figure 1.3: AR sandbox(left) and AR pool table [4](right)

Inspired by the AR Sandbox, the AR pool table uses a projector to display visible lines that direct the player on how to strike the cue ball as seen in Fig. 1.3 (right). Here, a 2D color camera is used to monitor the positions of the balls on the table [52]. Like the AR Sandbox, various iterations of the AR pool table have been built with some displaying the trajectory of the target ball after being hit. One step further, robotic arms have even been made to play [53].

For more mundane applications, the Puppy Cube, by Puppyrobot [54], is a smart projector with integrated camera to provide a touch sensing interface that turns any flat surface into a touchscreen. With a pre-installed Android OS, the smart projector promises to make everyday life more convenient and wonderful. Similar inventions have been proposed by Touchjet [55], Sony [56] and a research group at Carnegie Mellon University [57].

Looking into the future of AR smart tables, a potentially strong application of SAR that has not yet been explored is for the teaching of human anatomy to first year medical school students where the simultaneous teaching of anatomy and exposure to human cadavers have long been considered essential to medical education. Dissected cadaver-based anatomy is especially necessary for students intending to be surgeons [58]; however, there are instances of medical student training using digital only cadavers [59, 60, 61]. Although there are case studies that show the digital only cadavers are pretty effective in educating medical students [62], and Panchaphongsaphak *et al.* [63] even developed a multimodal VR system that functions as a 3D touch sensitive device using a six-axis force-torque sensor and a tangible object representing the shape of an anatomical structure, debates on whether the seemingly obsolete cadaver-based gross anatomy classes should be completely replaced by the promising AV/VR-based classes have arisen.

Despite the difficulties of carrying out cadaver-based anatomy classes such as high costs and potential hazard and preservation, some people still deem it as a better way to help students gain visual-spatial ability [64] as well as other vital experiences that a qualified doctor should have as there is concern that future doctors or surgeons will start practical examinations without any previous experience regarding real cadaver dissection [65]. Noticing an opportunity to bring the benefits of digital augmentation together with physical cadaver dissection, we hereby envision a SAR cadaver table that ingeniously brings together real cadavers with digital projectors.

Imagining a prosection class where an experienced instructor demonstrates the dissection of a real body in front of students [66], the proposed cadaver table could

project color onto otherwise colorless bodies, made colorless by the chemicals used in the body’s preservation. Arrows and other location markers projected onto the points of interest would allow the instructor to accurately identify points of interest regardless of the student’s particular point of view around the table. Considering minimally invasive surgery where only a limited size of incision is allowed [67], the indirect access to the operation area restricts the surgeon’s vision to what the endoscope sees. Now it might be possible to solve this problem with the recent advances in AR technology, e.g. projecting virtual organs, bones or tissues on the surface of the body so that the surgeon can better perform the operation [68].

In the AR cadaver table application, existing AR tables are inadequate due to their reliance on real-time RGB+D cameras, which have depth accuracy of about a centimeter when accuracy under a millimeter are required. So supposing the cadaver needs to be dissected, existing AR tables may not accurately define where to cut or even how deep. Given the condition that the projection accurately displays the organs and their locations to direct the surgeons to do any following actions, the technique that is employed to collect this data becomes crucial. So to pursue an AR cadaver table effectively, we need to develop a much higher accuracy scanner.

1.2 Structured Light Illumination

A potential means of providing the necessary 3D resolution for an AR cadaver table is through the use of structured light illumination (SLI), which is a means of 3D scanning based on projecting a series of striped or structured patterns onto a target and then reconstructing a 3D model based on the warping of the stripes as measured by a

camera [69, 70, 71]. A traditional SLI system consists of one camera, one projector, and one image data processing instrument, usually a personal computer [72] where, by projecting a series of striped patterns, the computer is able to reconstruct a 3D surface based on the warping of the stripes as recorded by the camera. Specifically, SLI relies on the same triangulation principal of stereo-vision but where correspondences are calculated between camera and projector instead of camera to camera. Since the projector actively emits energy into the scene, SLI is commonly referred to as active triangulation [73] instead of stereo-vision's passive triangulation [74]. As an active technique, SLI is not affected by and, in fact, prefers low textured surfaces where passive techniques cannot derive correspondences.

When describing this correspondence matching of SLI, we assume a camera and projector configuration as illustrated in Fig. 1.4 where a camera is oriented above the projector, each having its line-of-sight oriented in the direction of the Z-axis. In this configuration and assuming a pin-hole lens model for both camera and projector, it can be shown that a 3D world coordinate, (X_w, Y_w, Z_w) , can be derived for any camera pixel, $I_c[r, c]$ with discrete-space row (r) and column (c) coordinates, that corresponds to the projector pixel, $I_p(x, y)$ with continuous-space row (y) and column (x) coordinates, according to:

$$X_w = a[r, c]Z_w + b[r, c], \quad (1.1)$$

$$Y_w = c[r, c]Z_w + d[r, c], \text{ and} \quad (1.2)$$

$$Z_w = (e[r, c]y + f[r, c]) / (g[r, c]y + 1), \quad (1.3)$$

where $a[r, c]$, $b[r, c]$, \dots , $g[r, c]$ are scalar constants unique to each camera pixel and

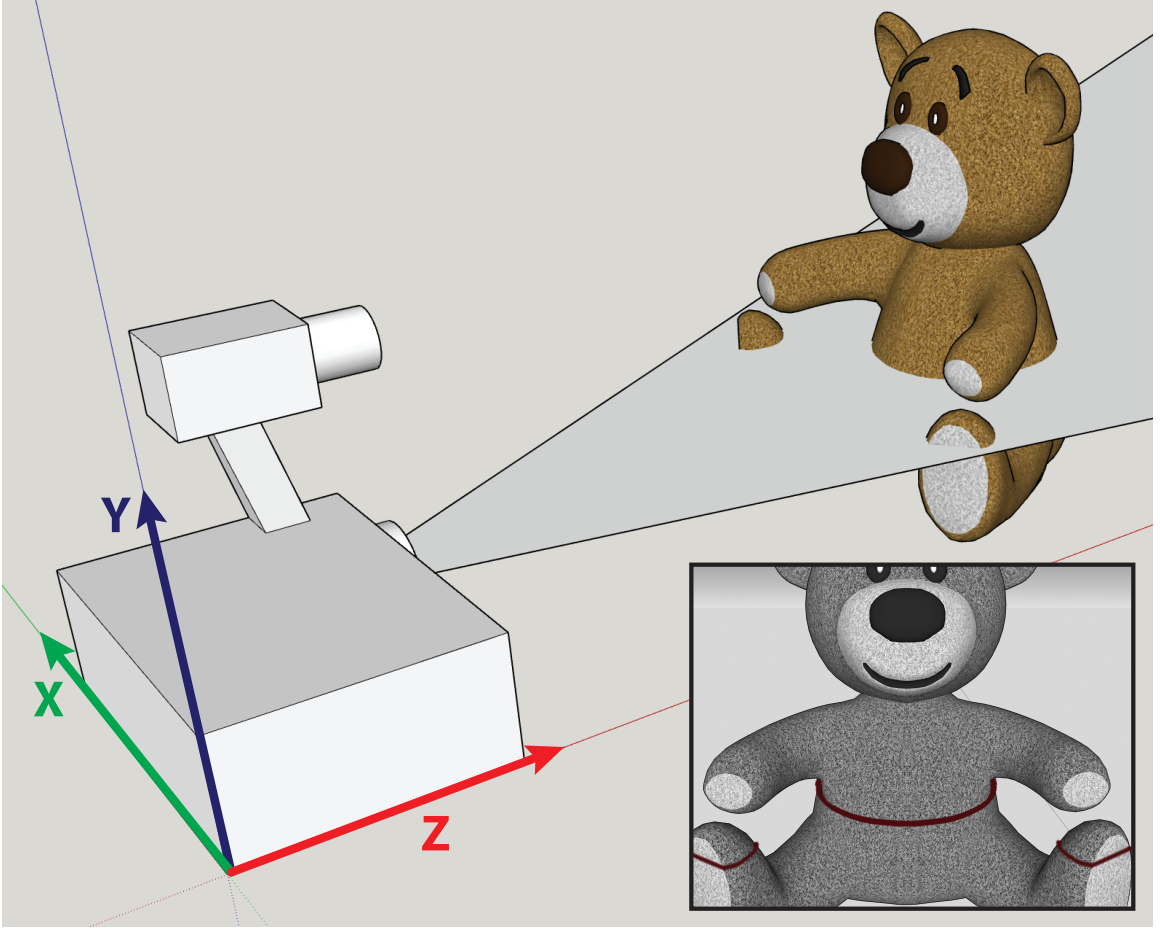


Figure 1.4: Illustration of an SLI scanner

derived during scanner calibration.

According to different types of pattern schemes that are used to compute correspondences between the captured images by the camera and the projected patterns, a number of different structured light techniques have been developed. Geng [75] summarized these SLI techniques as illustrated in Fig 1.5 where, at the very top level, SLI techniques are classified into single and multi-shot methods. Single-shot methods are faster than multi-shot because they only have one image to process, but their reconstruction results are not as impressive as multi-shot in terms of spatial accuracy [76]. As the accuracy is of top priority in our case of AR table, the single-shot

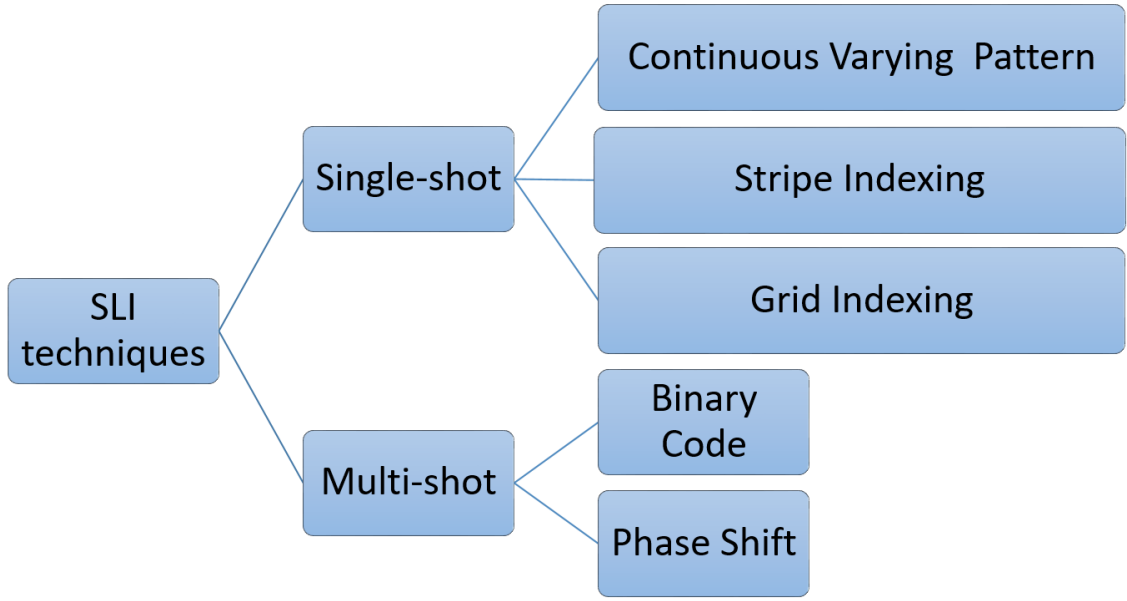


Figure 1.5: A taxonomy of SLI 3D surface imaging techniques

method is not within the scope of our discussion.

Multi-shot schemes are called sequential projection techniques which sequentially project a series of patterns onto the scene, then obtain the 3D depth data by capturing and analyzing all the images shot under the projection of each unique pattern [73]. Typical sequential techniques include binary coding and phase shifting. Binary coding [70, 77] projects a sequence of black and white stripes where the widths of stripes vary between patterns. By putting all the projections in a complete sequence together, each point on the object surface will have a unique binary code[78]. Given all the binary codes of point cloud in the image and the triangulation principle, the shape of the 3D object surface can be reconstructed. However, binary coding suffers from low spatial resolution unless the number of patterns in a sequence is high. Considering large binary numbers introduce much larger computational cost, binary coding

techniques are not widely used [75]. Alternatively, gray-level patterns are essentially an improved binary coding technique developed to reduce the number of patterns in a sequence while not losing its high resolution [79].

Phase shifting methods (PSMs) are known as a kind of fringe projection in which a set of phase-shifted patterns are projected onto the object surface, and then by capturing and processing a series of images with deformed fringe patterns, the 3D coordinates of the object surface can be derived [80]. Various phase shifting methods have been proposed including phase measuring profilometry (PMP) [81], trapezoidal pattern strategy [82], 2+1 pattern strategy [83], triangular pattern strategy [84], etc.

Phase Measuring Profilometry (PMP) [85] originated from the classical optical interferometry. It has the advantage of high robustness, high depth accuracy, as well as resistance to both ambient light and reflection variation [86, 87]. In PMP, the component patterns are defined by the set, $\{I_n^p : n = 0, 1, \dots, N-1\}$, according to:

$$I_n^p(x^p, y^p) = \frac{1}{2} + \frac{1}{2} \cos \left(2\pi \frac{n}{N} - 2\pi y^p \right). \quad (1.4)$$

where (x^p, y^p) is the column and row coordinate of a pixel in the projector, I_n^p is the intensity of that pixel in a projector with dynamic range from 0 to 1, and n represents the phase-shift index over the N total patterns.

On the camera side, the light intensity of the image captured under the projection of PMP patterns is expressed as

$$I_n(x, y) = \alpha(x, y) \cdot [A + B \cos(2\pi f_\phi y^p + \phi(x, y) - 2\pi n/N)], \quad (1.5)$$

where (x, y) is the image coordinates of a pixel in the camera's point of view, $\alpha(x, y)$ is the reflectance variation. $\phi(x, y)$ is the phase distortion caused by the presence of

the object, which corresponds to the depth value of the object surface. Therefore, as long as we can determine the $\phi(x, y)$, the depth coordinate of the object surface can be somehow derived. From the captured images which are under the projection of sinusoidal patterns, we can have $\phi(x, y)$ calculated by

$$\phi(x, y) = \arctan \frac{\sum_{n=1}^N I_n(x, y) \sin(2\pi n/N)}{\sum_{n=1}^N I_n(x, y) \cos(2\pi n/N)}. \quad (1.6)$$

The phase value $\phi(x, y)$ calculated in Eq. 1.6 is wrapped within the range of $(-\pi, \pi]$ by the arctan function, whatever the frequency of the sinusoidal wave f_ϕ is. In order to obtain the non-ambiguous phase value from the wrapped phase, a phase unwrapping procedure needs to be carried out [88]. The higher the frequency f_ϕ is, the higher signal to noise ratio (SNR) of the the depth data have after the non-ambiguous phase unwrapping [89]. Phase unwrapping techniques are commonly utilized in applications related to interferometry [90, 91, 92, 93], in the field of 3D shape measurements, there are two principal groups of phase unwrapping techniques: spatial and temporal.

The basic idea of spatial phase unwrapping is that the phase unwrapping at any pixel is conducted by comparing its phase to the phases of its neighboring pixels [94], but this has the disadvantage of effectively diffusing phase error between pixels and it fails to obtain the correct unwrapped phase when there are discontinuities on the surface of the object [95, 96]. On the other hand, temporal phase unwrapping, as initially proposed by Huntley and Saldner [97], employs one or more sequences of fringe patterns where the phase of each pixel is studied along the time axis [98]. The phase of each pixel is, therefore, independently derived from other pixels so that the noise at certain pixels does not affect its neighbors. More importantly, surface discontinuities

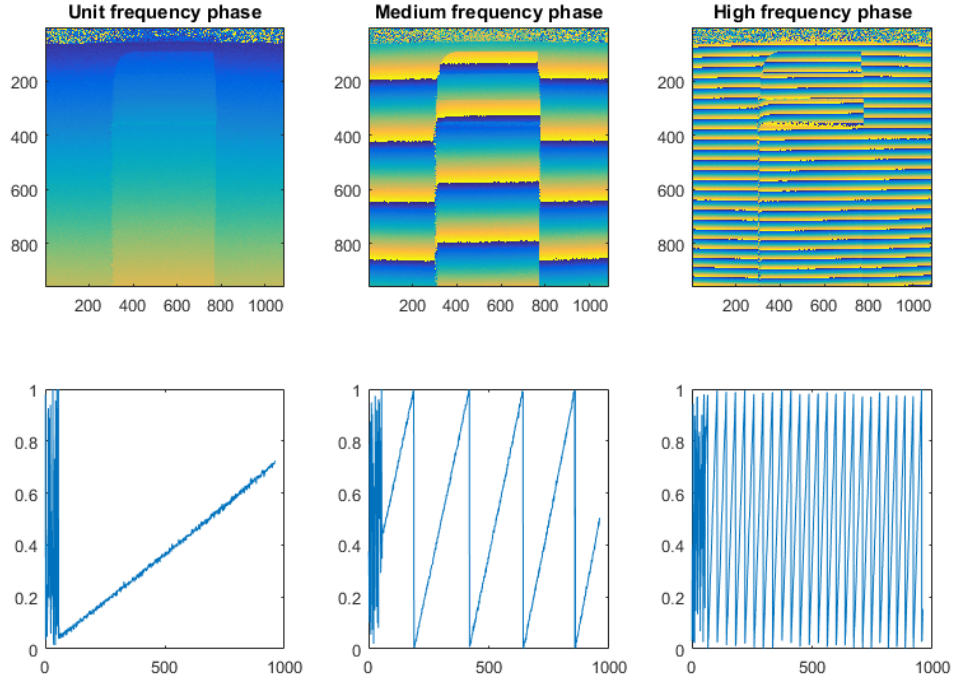


Figure 1.6: Wrapped phases of low (left), medium (center) and high (right) frequencies SLI scans

are no longer an issue [99, 100]. Liu *et al.* [101] categorized the temporal phase unwrapping algorithms into three representative classes according to the different fringe sequences being used: linear sequences [97]; exponential sequences [96]; and reversed exponential sequences [102]. Figure 1.6 shows the phase maps of a scanned scene and the phase of pixels along the 900th column when the spatial frequency is 1 (left), 6 (center), and 36 (right). Figure 1.7 presents the result of the unwrapped phase that is derived from the three wrapped phases of Fig. 1.6 where the noisy phase seen in the unit frequency image is replaced with a much cleaner phase map derived from unwrapping the higher frequency phases.

In sum, SLI is a highly accurate means of 3D scanning, but as a scanning process, SLI can be somewhat inconvenient because the target object must remain still during

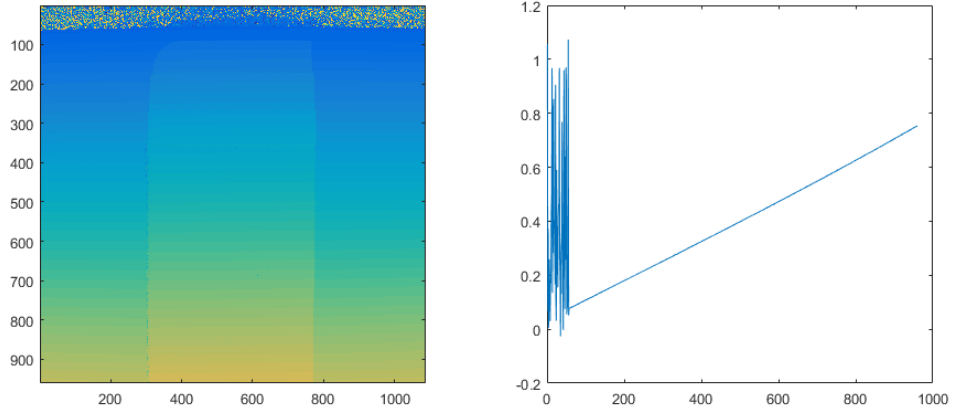


Figure 1.7: Unwrapped phase of the scene

scanning. Any potential motion of the target must be minimized by, most likely, running the camera and projector at such high speeds as to make the target appear motionless. But high speed hardware incurs a substantial cost in hardware. Custom light engines are typically used by manufacturers such as Vialux [103], Greenlight Optics [104] and Ajile light industries [105], which may charge as much as \$10,000 for a DLP projector [106]. In the case of an AR table, there is already a digital projector from which to scan. So adding SLI scanning means simply inserting a machine vision camera onto the AR table’s canopy and then borrowing the PC’s HDMI signal to insert structured light patterns.

The downside to this borrowing of the projector is that the human user loses the illusion of augmented reality during scanning, and for this reason, any SLI solution must also minimize this inconvenience. So in order to provide the high resolution of structured light to an AR table while also minimizing the inconvenience of scanning, this dissertation proposes a custom FPGA controller that reads the HDMI signal from the PC and inserts structure light patterns as requested by the user through a

machine vision camera’s GPIO interface. By doing so, we can minimize the number of video frames stolen from the user used for scanning so that any HDMI projector can be instantly converted into a structured light scanner by simply plugging in the FPGA. Because of our system’s reliance on commodity projectors, we pay particular attention on achieving frame rates and scan resolutions as good as, if not better than, custom light engines at 120 Hz. While this, at first, seems like a substantial sacrifice when industrial light engines can achieve frame rates up to 4,300 Hz [107], few cameras can do the same. As such, we can maximize the available bandwidth between camera and computer at 120 Hz by employing higher pixel counts.

1.3 Our Contribution

Having understood the principles and advantages of structured light illumination in 3D measurement, it would be of great interest to build an affordable, easy-to-use, and state of the art SLI 3D scanner. As the essential component of any SLI scanner, the industrial and scientific light engines are usually either quite expensive or with low brightness as shown in Tab. 1.1. So as an alternative to industrial light engines, there are commodity projectors available that have a large diversity, ranging from pico projectors to theater projectors. Using commodity projectors, therefore, would allow us to assemble scanners of different scales, with fields of view ranging from several millimeters to several meters, but since most commodity projectors receive video streams from HDMI, we need a way to synchronize a machine vision camera to the HDMI signal as well as identify what SLI pattern is being captured by the camera at any moment in time.

As industrial light engines can project patterns slaved to the camera, our HDMI controller will have to provide a means of slaving the camera to the HDMI signal while synchronizing the camera exposures to specific patterns without these captured images containing duplicates of a particular pattern. What's more, industrial light engines do not indicate the beginning of the stored pattern sequence, which brings great inconvenience to users. Our HDMI controller, on the other hand, utilizes a dedicated enable/reset bit to tell the user when the first frame of the sequence comes out. So the user can capture a sequence of SLI patterns without having to reshuffle the captured frames of video to match the intended sequence. Another limitation of the industrial light engine is that they have limited memory in which to store pre-defined patterns, whereas, our HDMI controller contains an FPGA which generates






	Brightness (lumen)	Price (dollar)	Photo
TI DLPDLCR2010EVM	25	520	
TI DLP3010EVM-LC	125	830	
TI DLPLCR4500EVM	150	1350	
CEL5500-LED	120	8000	
Vialux STAR-07	1100	9700	

Table 1.1: List of popular industrial DLP light engines available on the market [8]

SLI patterns on the fly, so the FPGA can generate an infinite number of patterns.

To further exploit our HDMI controller, we devise a new system structure where video streams from a host PC pass through our controller before they feed the projector. In this way, any patterns designed on a host PC can be used for SLI scanning without reconfiguration of the FPGA. Furthermore, the FPGA can read the incoming HDMI signal to extract specific pixels from the signal which act as a de facto handshake with the camera, telling the camera when to trigger an exposure and avoid duplicate frames. So when a sequence of SLI patterns encoded with index byte are passed from PC, the HDMI controller knows which pattern it is by reading the first pixel. Thus we manage to synchronize the PC video with the camera.

Also in order to guarantee that our FPGA can work with any commodity graphics card, we implemented an Extended Display Identification Data (EDID) emulator, which is an indispensable part for the HDMI receiver in that it tells the source device to stream videos in the expected format. EDID is a data structure standardized by the video electronics standard association (VESA) which contains important information such as the manufacturer name, the product ID, the supported timings, the display size and the luminance characteristics and so on [7]. Normally the EDID data is stored in a serial programmable read only memory (PROM) or a serial electrically erasable PROM (EEPROM) that is part of the display system. Upon the physical connection between the HDMI source and sink devices, the EDID is transferred from sink to source via I2C bus.

Another issue when performing SLI with a commodity projector is the nonlinear gamma effect [108] where human eyes do not perceive the light in a linear manner

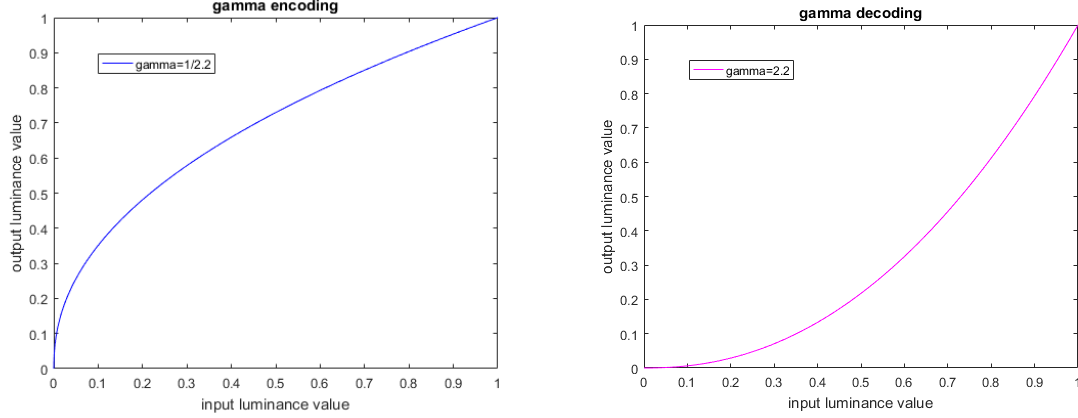


Figure 1.8: input-output curves of gamma encoding (left) and decoding (right)

but are more sensitive to the changes in darker tones. So the digital images are often encoded in a way to optimize the usage of the bandwidth [109]. The gamma encoding and decoding can be generalized in the form of:

$$V_{out} = AV_{in}^{\gamma}, \quad (1.7)$$

where the V_{in} and V_{out} are the input and output luminance values respectively, A is a constant which equals to 1 in common case. When $\gamma < 1$, the power-law nonlinearity is called gamma encoding, while $\gamma > 1$, it is called gamma decoding. Figure 1.8 demonstrates the gamma nonlinearity, both the encoding and decoding. From the power-law expression of gamma, if the encoding γ_{en} is the reciprocal of the decoding γ_{de} , the eventual outcome of the image would be identical to the original, assuming that the image is encoded first and then get decoded by the display system.

Finally the most difficult issue for using commodity projectors is how to set the camera's exposure so as to capture undistorted SLI patterns. Specifically, Digital light processing (DLP) projectors use a digital micromirror device (DMD) to either

direct light toward or reflect light away from the projection screen [110] and are the only spatial light modulators for 120 Hz projection rates. Since these spatial light modulators are binary, the RGB channels of the DLP projector have to be projected in a time-multiplexed manner, and as a result, the timing of the camera exposure should be carefully treated to make sure that the exposure covers all the RGB information within a frame. Otherwise, the camera captures artifacts from the way the DLP represents continuous shades of color.

With all the factors above taken into consideration, this dissertation describes the development of a number of FPGA-based projector controllers to build industrial-quality 3D scanners using commodity projectors addressing the specific problems of camera-to-projector synchronization, gamma distortion, and EDID spoofing. In other words, our projector controller connects an off-the-shelf machine vision camera to any HDMI projector to instantly turn them into a structured light 3D scanner. With our projector controller, many hobbyists and those who are interested in SLI will be able to build their own structured light scanners, and as we will demonstrate, any SAR table can instantly add high resolution 3D scanning to their feature lists.

To demonstrate the capabilities of our FPGA-based SLI controller, Chapter 2 introduces a novel dual-projector scanner as shown in Fig. 1.9. The scanner manages to run two projectors simultaneously with their own PMP patterns, make up the gamma correction that is already built-in the projector but in fact is unnecessary, because the SLI patterns generated inside FPGA do not have any gamma distortions. In addition, a robust synchronization scheme is proposed to coordinate the timing between the camera and projectors. Three actions including triggering the camera,

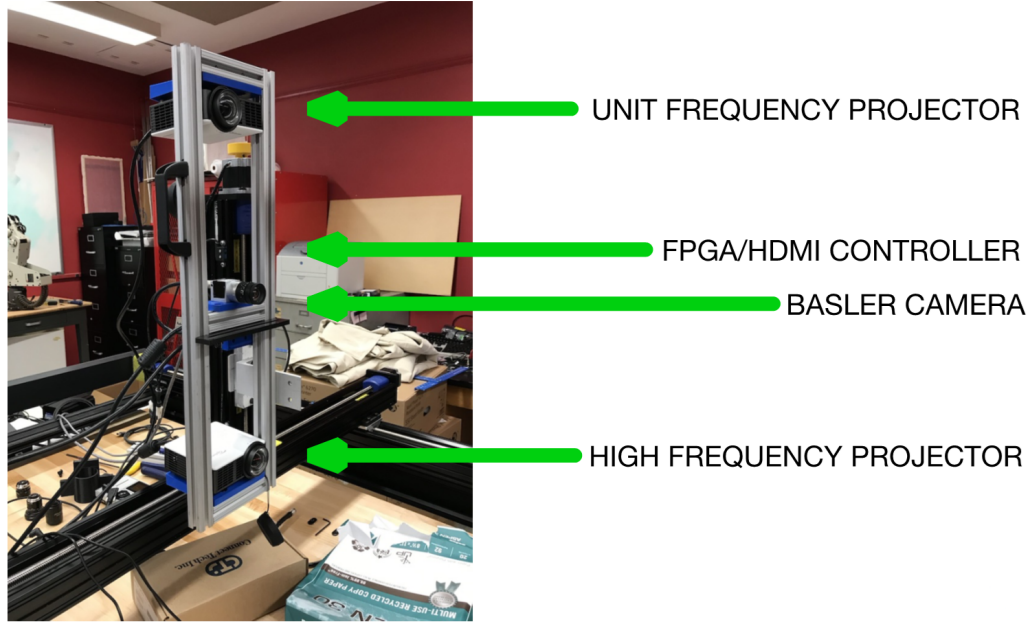


Figure 1.9: System configuration of the dual-projector structured light scanning system

updating the pattern, ending the camera exposure, should be executed in a precise order and timing. Each one of these actions is closely related to and impacts the other two. In the end, a special feature of DLP projectors and its influence in SLI scanning are discussed, subsequently a solution is provided to guarantee reliable 3D scans. It is the discovery that the DLP projectors divide the time in multiple slots within a frame, in each slot, only one of the red, green or blue color component is being projected. To avoid bending effect and inconsistencies among images, the camera exposure needs to be carefully configured so that every exposure covers the same amount of these time slots at the same position for all the frames.

The results of the related experiments as shown in Fig. 1.10 prove that the dual-projector setup improves the SNR without any interference between two projectors and minimize the influences of occlusion and multi-path. The left and center-left

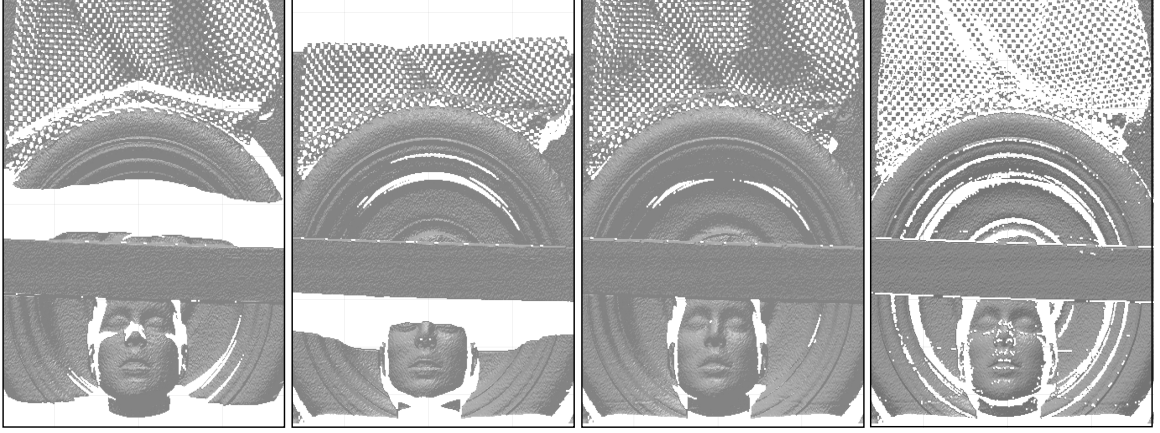


Figure 1.10: Experimental results of the dual-projector scanner that show the improvement in terms of increased SNR and the removal of occlusion and multi-path

figures show the occlusions while only one projector is running; the center-right figure is the result after putting the results of each projector together, where occlusion is minimized; the right figure is the result without the pixels that suffer from multi-path interference.

In Chapter 3, we look into the issue of multi-path as this is a major issue that has received very little attention in the SLI literature. Multi-path is the occurrence where a pixel of the image sensor captures light reflecting from two or more different paths due to the complex geometry of the object surface [111]. Our discussion is narrowed down to bimodal multi-path which is the simplest but most representative instance, where the number of paths is strictly restrained as two. Up to now, most research about multi-path has been conducted in the area of 3D imaging using time-of-flight technique, several analysis models and algorithms to remove multi-path in ToF images have been proposed [112, 113, 114]. By modifying the dual-projector SLI scanner described in Chapter 2 , we made another dual-projector SLI scanner exclusively for multi-path study. The reliability of our FPGA controller has been

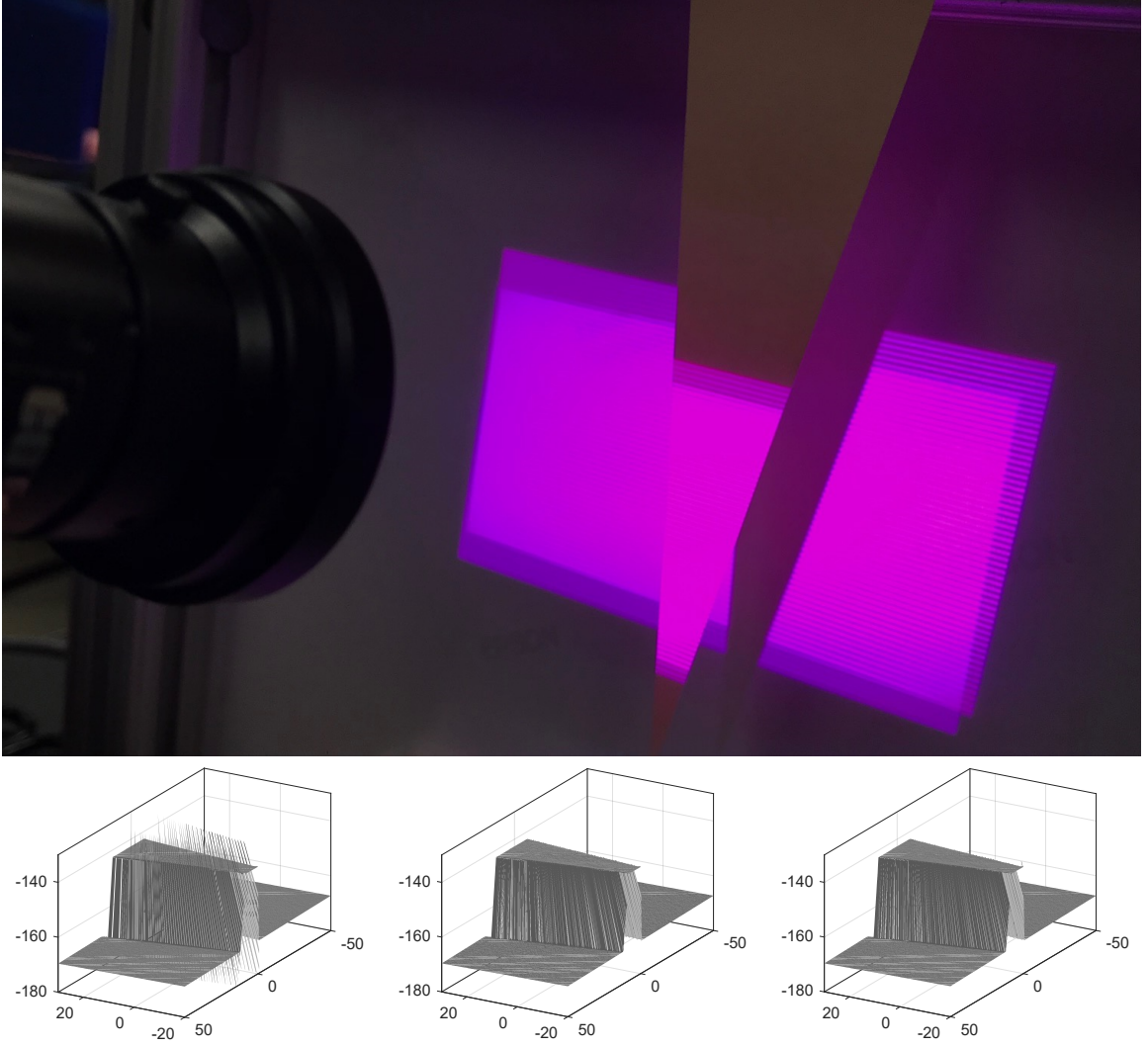


Figure 1.11: The scanned wedge (top) and the 3D reconstruction without any treatment for multi-path (bottom-left), replace the multi-path with foreground path (bottom-center), replace the multi-path with background path (bottom-right)

thoroughly tested and proved to be satisfactory as 488 SLI patterns that contains 61 distinct spatial frequencies are scanned without any dropping or duplication of frames. Experiments are conducted with particular setups for multi-path with results verifying our analysis and expectation.

As an illustration of the improvements made in Chapter 3 removing multi-path affects near discontinuities in depth, Fig. 1.11 (top) shows the scanning of a thin

wedge target placed about an inch in front of the plain white board. When the dual-projector system scans, multi-path would inevitably be found at the edges of the wedge since the corresponding pixels see projected light from both the wedge and the white board behind. Figure 1.11 presents the results of 3D reconstruction where, on the left, is the 3D view of the wedge without any particular treatment for multi-path pixels and where many noisy spikes can be seen. In the center, Fig. 1.11 shows the view of the wedge with the detected multi-path pixels replaced by their foreground-only components while, on the right, are the detected multi-path pixels replaced by their background components.

In Chapter 4, we focus on how the HDMI video passes from PC to the projector through our FPGA controller in order to unlock the capability of our system for SAR applications. The HDMI does not automatically start the transmission of video once a cable is connected between the source and sink as it requires an initial handshake to let the PC source know what kind of display device is connected so as to issue the video with the appropriate format. As the sink device in the system, the FPGA is not an actual display device like a PC monitor or projector. So an EDID emulator is implemented to reply the request from the source and then fulfill the handshake such that the PC sees it like a real display device. Figure 1.12 illustrates that Windows 10 recognizes the FPGA as if it is a DVI display.

In addition to emulating the EDID, we propose a top-left pixel encoding scheme that allows us to implement a GPU to FPGA handshake that enables our FPGA to identify specific targeted frames coming from the PC among many unpredictable unrelated frames, triggering the camera at the right moment and skipping any dupli-

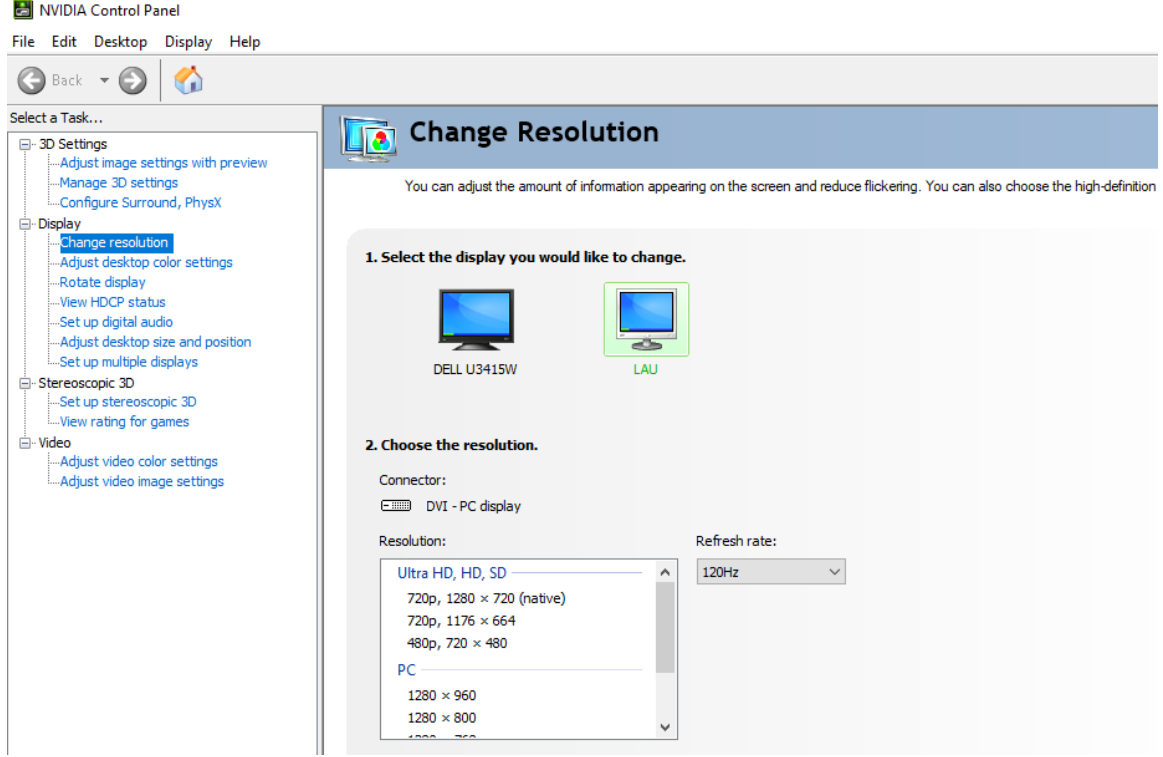


Figure 1.12: An screenshot of our customized EDID recognized by Windows 10

ates. So a user can implement their own structured light encoding scheme knowing that the camera will captured each and every frame in the sequence once and only once. We even use a second GPO bit from FPGA to camera to indicate when the top-left pixel is equal to zero, which a programmer can use to further ensure that the captured images correspond to specific frames in the SLI sequence. For SAR applications, this handshake also gives the programmer the ability to use the camera for traditional human-computer interaction with the AR environment.

To demonstrate the SAR benefits provided by our FPGA controller, two practical SAR applications are demonstrated. The first is an SAR face paint system that uses a machine vision camera and a DLP projector to project digital masks onto a human subject. Specifically, the camera and projector are positioned in front of a human

face with SLI mode used to scan the face at startup and derive a camera pixel to projector pixel mapping. And then in SAR mode, the FPGA triggers the camera on every 8th frame while simultaneously turning off the pixels of pass through video. This means that, as depicted in Fig. 1.13, the camera captures a photo of the human subject's face while not projected any SAR image. This photo is then processed using an open source facial feature detector with the detected features used to digitally warp a supplied mask image, projecting it onto the human subject's face in perfect registration.

As a second demonstration of our HDMI controller for SAR, we implement a novel camera tracking system for AR headset tracking without a direct line of sight between headset and base station. Specifically, a camera-projector pair with fixed position serves as the base station, performing traditional SLI scanning where a the mapping between the camera's 2D image plane and 3D world coordinates can be obtained. While the base station is scanning, a second camera, serving as a mobile headset, takes images of the projected patterns concurrently. Therefore, both cameras have recorded the 2D phases information that is encoded in the SLI patterns. Subsequently, the 2D phases from the second camera are compared with the phases from the base station camera, pixel by pixel. If there is a match, a correspondence between the second camera's pixel and its 3D world coordinate from the first is found. With a minimum of 6 such pixels, the perspective matrix of the second camera can be derived from which the position and orientation of the second camera are extracted. Thanks to its unique setup, our collaborative tracking system opens up the possibility for simultaneous multiple AR headsets tracking so that two or more headsets can display

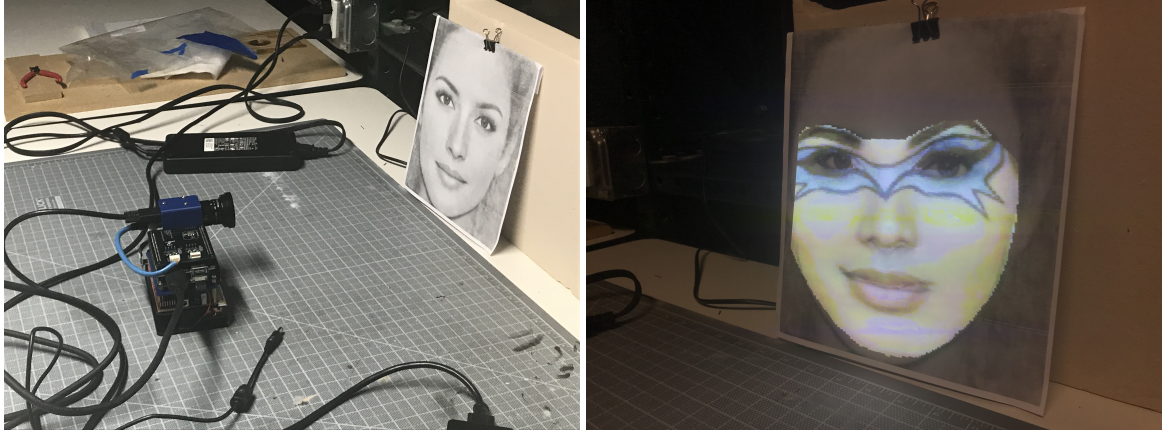


Figure 1.13: An 2D human face before (left) and after (right) face paint

AR information over the scene in perfect registration across headsets.

To conclude this dissertation, Chapter 5 reviews the results of this study and traces out a path for future work combining our HDMI controller with a machine vision sensor. The details on how a machine vision camera that supports high dynamic range (HDR) imaging is built from the bare CMOS image sensor are presented. Based on the LUPA300 image sensor from Cypress and the Zynq-7010 micro-zed board, the proposed smart camera meets of targeted frame rate of 120 fps while performing phase processing inside the unit, transferring phase data over a gigabit ethernet (GigE) interface.

Chapter 2 Dual-projector Structured Light 3D Scanning System

Structured light illumination (SLI) is a means of 3D scanning based on projecting a series of striped or structured patterns onto a target and then reconstructing a 3D model based on the warping of the stripes as measured by a camera [69, 70, 71]. A traditional SLI system consists of one camera, one projector, and one image data processing instrument, usually a personal computer [72]. With evolving electronics, SLI systems have been gaining in performance with faster hardware [115, 116]. Meanwhile, in terms of the algorithm principles, many experimental or practical pattern codecs have been introduced by researchers to achieve high-speed and/or high accuracy 3D reconstruction [116, 117].

For those applications where high speed 3D scanning is critical, one-shot SLI methods exist that rely on constantly projecting a single, static pattern onto the target [118, 119]. While capable of achieving video-rate 3D, the one-shot techniques sacrifice a substantial amount of resolution in depth [120]. In response, many studies have been conducted trying to obtain high speed scans without sacrificing scan quality [121]. Liu *et al.* [122], in particular, proposed a dual-frequency pattern scheme that combines a unit-frequency pattern sequence with a second, high frequency pattern sequence into a single composite pattern sequence. By combining the two sets into a single sequence, Liu *et al.* is able to achieve high spatial resolution with half the number of video frames as traditionally required.

One challenge that traditional SLI systems face is the limited viewing angles

of both the projector and camera that causes issues such as occlusions [123] and multi-path [124]. Occlusions are a particularly common problem associated with complex surfaces, like step edges, where portions of the target surface are not visible to either the projector or camera. To address this issue, Zhang *et al* [125] proposed a scheme that adds a second camera. Alternatively, Jiang *et al.* [126] employed a second projector. The problem with adding a second camera is that it requires twice the computational complexity, having to process a second set of pixels. This both increases the needed bandwidth between the PC and the cameras in order to maintain high frame rates as well as the computational horsepower of the processing element. Adding a second projector doubles the number of recorded frames of video, which doubles the acquisition time as well the number of pixels to be processed.

In this chapter, we also propose a dual-projector scanner scheme but in a far more flexible manner than Jiang *et al* [126]. In particular, we believe a dual-projector scanner can address a range of issues associated with structured light scanning including: (1) minimizing occlusions, (2) increasing projector signal strength, (3) minimizing the number of projector patterns using a novel phase unwrapping scheme, and (4) detecting multi-paths. Principal to this Chapter is an FPGA controller that synchronizes two commodity LED projectors to a machine vision camera, allowing our system to run at 120Hz without any frame loss. Compared to custom light engines such as the TI Lightcrafter, commodity projectors are much brighter and lower cost, a feature especially important when using multiple projectors.

Specific scanner architectures that we introduce are illustrated in Fig. 2.1 where Fig. 2.1(a) shows the traditional signal camera-single projector setup. Figure 2.1(b)

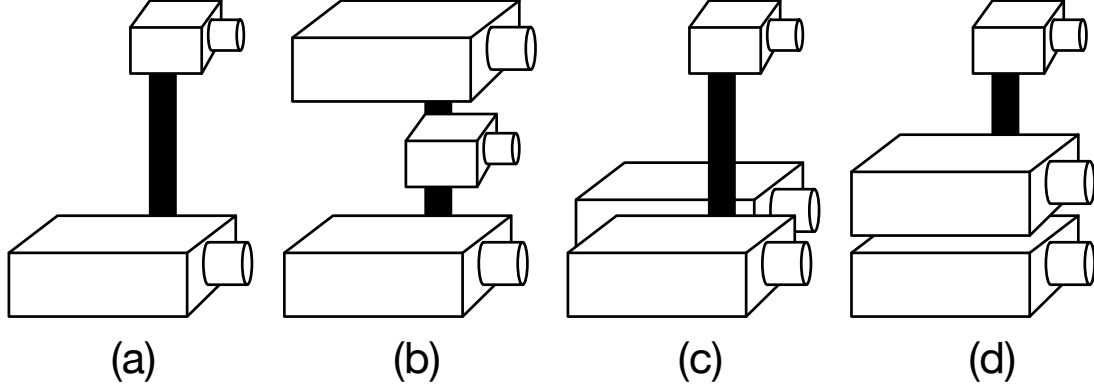


Figure 2.1: Example structured light scanner architectures illustrating: (a) a traditional single projector, single camera scanner; (b) dual-projector scanner with projectors in opposing positions in order to minimize scan occlusions; (c) dual-projector scanner with projectors in complementary positions for increasing projected image brightness; and (d) dual-projector scanner with projectors in stacked orientation for detecting multi-path.

shows our first dual-projector scanner where projectors are placed in opposing positions with the camera placed in-between. Differing from Jiang’s scheme in which two projectors work independently and the discussion focuses on the approach of combining the images of the object under projections from different angles, our system drives both projectors simultaneously using Liu’s dual-frequency pattern scheme [122]. Using the architecture of Fig. 2.1(c) with projectors placed in complementary side-by-side position, we can detect the multi-path effect when scanning particular objects.

2.1 Dual-frequency pattern scheme

In Phase Measuring Profilometry (PMP) [85] using the vertical above-below scanner architecture of Fig. 2.1(a), a series of phase-shifted sinusoidal fringe patterns are projected onto the object to be measured with the pixels of projected PMP patterns

generalized as:

$$I_n^p(x^p, y^p) = \frac{1}{2} + \frac{1}{2} \cos \left(2\pi f y^p - \frac{2\pi n}{N} \right), \quad (2.1)$$

where I_n^p is the intensity of the pixel at the projector coordinate (x^p, y^p) and its range is $[0, 1]$; f is the spatial frequency of the fringe pattern set which is equal to the number of sinusoidal periods from top to bottom in the projection area; and n is the frame index of the pattern in its sequence of N frames and numbered from 0 to $N - 1$.

From the camera's perspective, the object under observation can be seen as:

$$I_n^c(x^c, y^c) = A^c(x^c, y^c) + B^c(x^c, y^c) \cos \left(\phi(x^c, y^c) - \frac{2\pi n}{N} \right), \quad (2.2)$$

where I_n^c is the intensity of the given pixel with camera coordinate (x^c, y^c) ; $A^c(x^c, y^c)$ is the average intensity of the given pixel over the N patterns; and $B^c(x^c, y^c)$ is the observed amplitude of the PMP pattern seen by the camera pixel. Now since every (x^c, y^c) has its own I_n^c, A^c, B^c , the notation (x^c, y^c) will be assumed present and omitted hereafter. Same will apply to (x^p, y^p) as well.

After capturing a sequence of N frames, the constant, B^c , can be derived as:

$$B^c = \frac{2}{N} \left\{ \left[\sum_{n=0}^{N-1} I_n^c \sin\left(\frac{2\pi n}{N}\right) \right]^2 + \left[\sum_{n=0}^{N-1} I_n^c \cos\left(\frac{2\pi n}{N}\right) \right]^2 \right\}^{0.5}. \quad (2.3)$$

with the angle, ϕ , derived as:

$$\phi = \arctan \frac{\sum_{n=0}^{N-1} I_n^c \sin\left(\frac{2\pi n}{N}\right)}{\sum_{n=0}^{N-1} I_n^c \cos\left(\frac{2\pi n}{N}\right)}. \quad (2.4)$$

Having derived ϕ , we can then identify the projector coordinate, y^p , visible to the camera pixel using an appropriate phase unwrapping scheme [97], depending on f .

And once y^p is uniquely identified, we can derive the corresponding world Cartesian coordinate, (X^w, Y^w, Z^w) , using calibrated lens models for the camera and projector.

In the case of the classic pin-hole camera model [89], the 3D world coordinate can be derived from:

$$[X^w \ Y^w \ Z^w]^T = C^{-1}D, \quad (2.5)$$

where C and D are expressed as:

$$C = \begin{bmatrix} m_{11}^c - m_{31}^c x^c, m_{12}^c - m_{32}^c x^c, m_{13}^c - m_{33}^c x^c \\ m_{21}^c - m_{31}^c y^c, m_{22}^c - m_{32}^c y^c, m_{23}^c - m_{33}^c y^c \\ m_{21}^p - m_{31}^p y^p, m_{22}^p - m_{32}^p x^p, m_{23}^p - m_{33}^p y^p \end{bmatrix}^{-1} \quad (2.6)$$

and

$$D = \begin{bmatrix} m_{34}^c x^c - m_{14}^c \\ m_{34}^c y^c - m_{24}^c \\ m_{34}^p y^p - m_{24}^p \end{bmatrix} \quad (2.7)$$

where the parameters, m_{ij}^c and m_{ij}^p , are the scalar values forming the perspective matrices:

$$M^c = \begin{bmatrix} m_{11}^c & m_{12}^c & m_{13}^c & m_{14}^c \\ m_{21}^c & m_{22}^c & m_{23}^c & m_{24}^c \\ m_{31}^c & m_{32}^c & m_{33}^c & m_{34}^c \end{bmatrix} \quad (2.8)$$

and

$$M^p = \begin{bmatrix} m_{11}^p & m_{12}^p & m_{13}^p & m_{14}^p \\ m_{21}^p & m_{22}^p & m_{23}^p & m_{24}^p \\ m_{31}^p & m_{32}^p & m_{33}^p & m_{34}^p \end{bmatrix} \quad (2.9)$$

of the camera and projector, respectively.

The major factor that influences the accuracy of structured light scanning is noise, which can come from various sources such as the camera, the projector, ambient light, etc. Wang *et al.* [127] demonstrated that the signal-to-noise ratio (SNR) is

proportional to the number of patterns in the PMP sequence, N , and the spatial frequency of the pattern set, f . In other words, the accuracy of the 3D scanning can be improved by either increasing the number of patterns in the sequence or utilizing the patterns with high spatial frequency, but high spatial frequency also introduces phase ambiguity. Because pixels with different y^p can possibly share the same phase, ϕ , as long as they have an equal $\cos(2\pi f y^p)$, unit frequency patterns are needed to unwrap the ambiguous high frequency pattern set. Practically, two sets of patterns, one with unit frequency the other with high frequency, are projected sequentially to implement 3D scans.

Noting that (2.3) is the magnitude of the second N -point, Discrete Fourier Transform (DFT) coefficient while (2.4) is the angle of that same coefficient, Liu *et al.* [122] proposed using frequency modulation in time and placing a second set of patterns on the third N -point DFT coefficient instead of performing time-multiplexed scanning. This allowed Liu *et al.* to do in N patterns what others would have done using $2N$ patterns. Specifically, Liu *et al.* define their PMP patterns according to:

$$I_n^p = \frac{1}{2} + \frac{1}{4} \cos(2\pi f_h y^p - \frac{2\pi n}{N}) + \frac{1}{4} \cos(2\pi f_u y^p - \frac{4\pi n}{N}), \quad (2.10)$$

where f_h is the spatial frequency of the added PMP pattern set, f_u is the spatial frequency of the original pattern set which is usually 1. The magnitude, B_1^p is identical to the expression in (2.3) with phase, ϕ_u , derived from (2.4). The magnitude, B_2^p , and phase, ϕ_h , are derived from:

$$B_2^c = \frac{2}{N} \left\{ \left[\sum_{n=0}^{N-1} I_n^c \sin(\frac{4\pi n}{N}) \right]^2 + \left[\sum_{n=0}^{N-1} I_n^c \cos(\frac{4\pi n}{N}) \right]^2 \right\}^{0.5} \quad (2.11)$$

and

$$\phi_u = \arctan \frac{\sum_{n=0}^{N-1} I_n^c \sin(\frac{4\pi n}{N})}{\sum_{n=0}^{N-1} I_n^c \cos(\frac{4\pi n}{N})}. \quad (2.12)$$

Note that, in order to project two spatial frequencies simultaneously, Liu *et al.* sacrifices signal strength using the sinusoidal amplitude of $\frac{1}{4}$ instead of $\frac{1}{2}$ in order for I_n^p to be within the dynamic range, $[0,1]$, of the projector. As such, combining more than two sinusoids is impractical.

2.2 Dual-Projector Scanning

The dual-frequency pattern scheme reduces the effects of noise without increasing scan times by incorporating a second sequence of high frequency patterns. With this same objective, we propose a dual-projector structured light scanning system that not only takes advantage of the dual-frequency pattern scheme but also improves its performance by increasing the projected light energy. We do this by means of routing the two sequences of patterns to two separate projectors, instead of combining them inside of one. Specifically, we define the projected patterns for two separate projectors according to:

$$I_n^{p_1}(x^{p_1}, y^{p_1}) = \frac{1}{2} + \frac{1}{2} \cos \left(2\pi f_1 y^{p_1} - \frac{2\pi n}{N} \right), \quad (2.13)$$

and

$$I_n^{p_2}(x^{p_2}, y^{p_2}) = \frac{1}{2} + \frac{1}{2} \cos \left(2\pi f_2 y^{p_2} - \frac{4\pi n}{N} \right). \quad (2.14)$$

Since these two sets of patterns temporally shift at different rates, they can be decoupled in the camera by means of the DFT along the time, n , axis.

The selection of f_1 and f_2 is application dependent, as we illustrate in Sec. 2.4, but in any case, both equations are intimately linked through the frame index, n , which is the same for both projectors at any moment in time. As such, there is an implication that both projectors are near perfectly synchronized to each other as well as the camera. In order to guarantee this synchronization, we employed a commercially available FPGA with dual-HDMI interface and added a 4-pin I/O camera interface to our selected camera, the USB 3.0 Basler ACE acA1920-150u.

The Basler ACE acA1920-150u, like many machine vision image sensors, incorporates a programmable I/O interface with two pins assigned to input and two to output. Specifically, Line 1 is an opto-coupled input line used by the FPGA to trigger a camera exposure. Line 2 is an opto-coupled output line employed by the camera as an enable bit, allowing the host to tell the FPGA that the camera wants to initiate a scan. Line 3 is a general purpose line, which we used as an input to the camera, letting the FPGA tell the camera that the pattern currently being projected is the first pattern in the SLI sequence. That is, this pin is set high when the projected pattern corresponds to $n = 0$. Finally, Line 4 is a general purpose I/O line used by the camera to tell the FPGA that it is ready for a trigger, i.e. the camera's trigger ready signal. This pin is set high after the camera completes an exposure and readout of its image sensor and is ready to initiate the next exposure.

The system diagram of the FPGA is shown in Fig. 2.3 where there are two input signals from the camera, Line 2 and Line 4, as well as one output signal to the camera, Line 1. While Line 3 is available for testing captured images from the camera and testing which frame corresponds to the first frame in the scan sequence, its

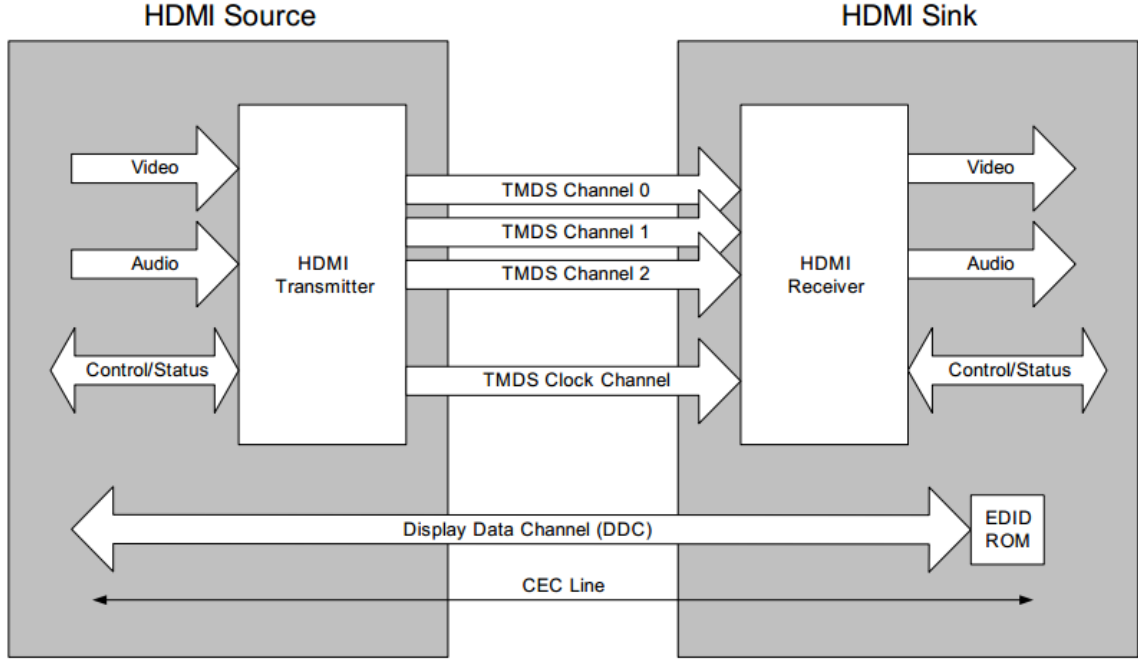


Figure 2.2: Block diagram of HDMI [5]

presence is not essential to this study. Also illustrated in Fig. 2.3 are two HDMI transmitters used to drive the two projectors simultaneously. HDMI stands for High-Definition Multimedia Interface, it is used to carry high quality multi-channel audio data and all standard high definition consumer electronics video formats. HDMI system architecture is defined to consist of sources and sinks. An HDMI cable contains four differential pairs of transition minimized differential signaling (TMDS) data and a display data channel (DDC). Three of these TMDS are video, audio and auxiliary data channels and the last one is clock channel. The DDC is used for configuration and status exchange between the source and sink, in other words, the source reads the sink's enhanced extended display identification data (E-EDID) through DDC channel to discover the sink's configuration and capabilities. Figure 2.2 shows the internal structure of HDMI. Included with the FPGA is an external EEPROM used to store a

user supplied tone correction curve, which the FPGA uses to adjust the output pixel values in order to achieve a linear response from the light engines in the presence of projector gamma distortion.

The system diagram of the FPGA is shown in Fig. 2.3 where there are two input signals from the camera, Line 2 and Line 4, as well as one output signal to the camera, Line 1. While Line 3 is available for testing captured images from the camera and testing which frame corresponds to the first frame in the scan sequence, its presence is not essential to this study. Also illustrated in Fig. 2.3 are two HDMI transmitters used to drive the two projectors simultaneously. HDMI stands for High-Definition Multimedia Interface, it is used to carry high quality multi-channel audio data and all standard high definition consumer electronics video formats. HDMI system architecture is defined to consist of sources and sinks. An HDMI cable contains four differential pairs of transition minimized differential signaling (TMDS) data and a display data channel (DDC). Three of these TMDS are video, audio and auxiliary data channels and the last one is clock channel. The DDC is used for configuration and status exchange between the source and sink, in other words, the source reads the sinks enhanced extended display identification data (E-EDID) through DDC channel to discover the sinks configuration and capabilities. Figure 2.2 shows the internal structure of HDMI.

Included with the FPGA is an external EEPROM used to store a user supplied tone correction curve, which the FPGA uses to adjust the output pixel values in order to achieve a linear response from the light engines in the presence of projector gamma distortion. Most display systems, including computer monitors and digital projectors

have $\gamma > 1$, because in this way, brightness looks better in human eyes. Nonetheless, we do not need such default gamma correction to be existing in the projector for our SLI system. In our case, the projection would be captured by the camera, which unfortunately operates in a different manner from human eyes. Cameras respond to the brightness in a linear way. Since the SLI patterns are generated naturally by the digital circuit without any gamma encoding, the ideal projector in this system should be a projector without any internal gamma compensation. In [128], Zhang classifies the existing methods that are used to calibrate the nonlinearity of the digital fringe projection systems into two categories, active and passive error compensation methods.

The active approach modifies the fringe patterns before they come to the projector [129], on the contrary, the passive approach compensates the phase error after those fringe patterns are captured by the camera [130]. In practice, we utilize the active approach with which we adjust the pixel values of the fringe patterns through a look-up table before they are sent to the projector. Since projectors from different manufacturers or of models have their own unique γ , each distinctive projector should have its unique look-up table. The look-up table can be either hard coded in the FPGA's configuration file or stored in an external memory which would be read by the FPGA. In order to make the system more adaptable, the latter one is selected in our case. Such that, an electrically erasable programmable read-only memory (EEPROM) is used to store the look-up table for γ correction, the EEPROM can be erased and reprogrammed through a simple serial port communication tool whenever it needs, then the FPGA reads the EEPROM to load the look-up table every time it

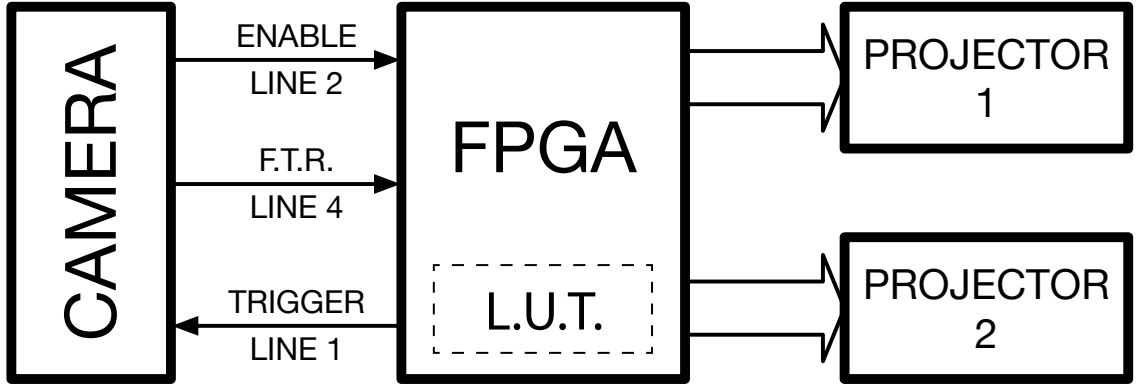


Figure 2.3: System diagram of FPGA controller showing the GPIO interface to the machine vision camera and HDMI outputs to the component projectors.

is powered on.

To see how we use these four pins, Fig. 2.4 shows a timing diagram with four significant events marked by the letters, **A**, **B**, **C**, and **D**. Event **A** is caused by the host PC programming Line 2 to go high, the rising edge of which tells the FPGA that user wants to initiate a scan. Upon receiving this, the FPGA waits for the first following vertical sync, at which it resets its internal frame counter and begins sending the pixels for the $n = 0$ frames to both projectors. The projector continues sending pixels for the first frame for all subsequent vertical syncs until event **B**, when it detects a rising edge on Line 4, the trigger ready signal from the camera.

Event **C** happens on the immediately following vertical sync, when the FPGA triggers a camera exposure using a rising pulse on Line 1. At this same moment, the FPGA increments its frame counter, and begins sending pixels for the $n = 1$ frames. Again, the projector continues to send the $n = 1$ frames until it detects another event **B**, a rising edge on Line 4. This leads to event **C** on the following vertical sync and incrementing the frame counter and sending pixels for the $n = 2$ frames. The process

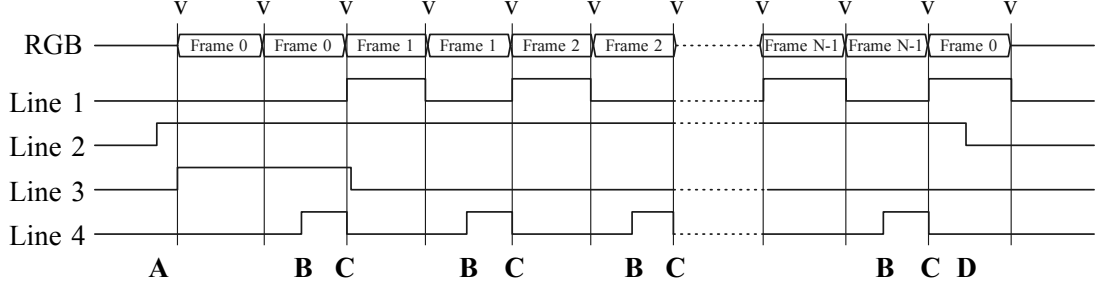


Figure 2.4: Timing diagram of the signals for camera synchronization where “RGB” represents the pixel buss of the HDMI signal where the pixels associated with each component scan frame labeled as “Frame 0” to “Frame $N - 1$ ”, “V” denotes the positions in time of vertical syncs, LINEs 1-4 represent the GPIO signals between FPGA and camera, and the letters **A**, **B**, and **C** are specific events corresponding to the handshake between camera and FPGA during a scan.

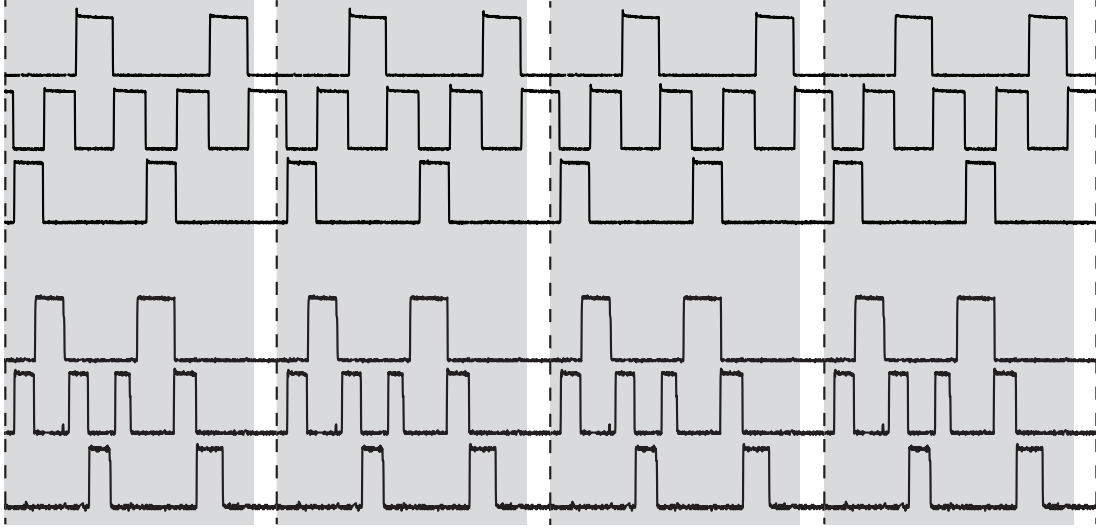


Figure 2.5: The light output for the red, green, and blue LEDs in a (top) TI DLP3010EVM-LC and a (bottom) Optoma ML750ST projector. Vertical dashed lines indicate vertical syncs for 120Hz video with the green LED of the TI DLP3010EVM-LC projector overlapping vertical syncs where the Optoma ML750ST keeps the color patterns inside the syncs. Shown as gray boxes are the moments in time when the camera is exposing its sensor.

of event **B-C-B-C-...** and so on continues until the camera completes grabbing N frames, causing event **D** when the camera toggles Line 2 to tell the FPGA that scanning is complete. At this point, the FPGA sends all zero pixels to the projectors.

Assuming that the camera can complete an exposure and sensor read-out cycle

in between consecutive vertical syncs thereby raising the trigger ready signal, scans can be performed at the video rate of the projectors, in this case 120 Hz. However, if the camera is slow, relative to the projectors, the above process will run at integer divisions of 120 Hz, such as 60, 40, 30, or 20 Hz. But regardless of frame rate, the camera must complete its exposure of the n^{th} frame before the projector begins displaying the $n + 1^{st}$ frame since the FPGA sends the pixels of $n + 1^{st}$ immediately following its triggering of the camera for the n^{th} frame.

Properly exposing the camera sensor in response to a trigger can be a complicated and precise process depending on the timing of the projectors. In particular, DLP projectors produce RGB images by means of modulating the mirrors of a binary spatial light modulator and then using a color filter wheel or flashing LEDs to produce a time-multiplexed sequence of binary color images. The images are then displayed at such high temporal rates as to create an illusion of continuous tone. To avoid any hint of flicker, DLP projectors are further complicated by projecting multiple red light images multiplexed with multiple blue and multiple green.

This complicated timing is observed using a photosensor, placed in the light path of our scanner's projector to measure the instantaneous light output of the DLP light engines. What Fig. 2.5 (top) shows is the voltage output of a photosensor placed in front of a (top) TI DLP3010EVM-LC and a (bottom) Optoma ML750ST DLP projector. Specifically, the top three waveforms illustrate when the red, green, and blue LEDs of the TI DLP3010EVM-LC are turned on in-between vertical syncs (dashed lines) showing that red and blue LEDs are turned on twice each during a single frame of video while green is turned on four times. The bottom three waveforms

show the LED timings for the Optoma ML750ST also having two pulses of light for red and blue and four for green.

What is most interesting to see in the waveforms of Fig. 2.5 is that the Optoma ML750ST projects light for the current frame within the bounds of the vertical syncs while the DLP3010EVM-LC overlaps the green LED over the vertical sync. In order to achieve 120Hz scanning, this slight overlap means that our FPGA must disable the green channel, projecting strictly magenta patterns. The alternative would be to delay the triggering of the machine vision camera slightly beyond a vertical sync; however, it must still complete the current exposure before the next vertical sync. That would imply clipping the fourth pulse of green light, but again, DLP projectors project binary patterns where clipping the last pulse would result in seeing only a partial set of its eight bits. And the bits being clipped may not be the least significant, resulting in catastrophic distortion.

In order to capture all bits of all colors, a programmer could extend the exposure of the camera beyond the following vertical sync, but this would imply losing the next available projector frame since the frame trigger ready won't go high until after the following vertical sync has past. While one could program the FPGA to ignore the frame trigger ready and just send consecutive frames, doing so would place a substantial burden on the PC to keep up with the incoming video frames, especially for long scan sequences. So in summary, the reader should note that some commodity projectors are better than others for applications in SLI.

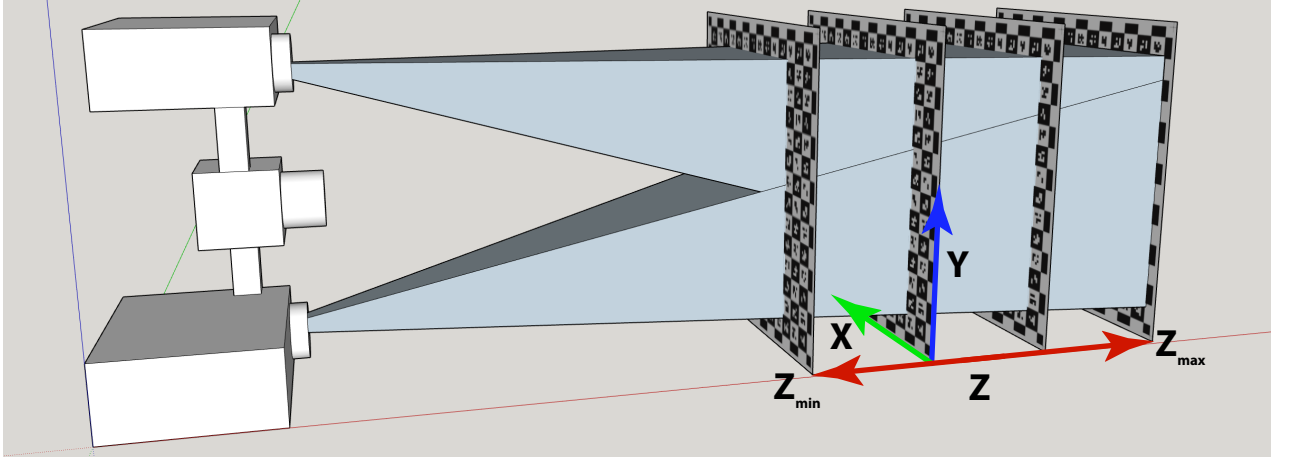


Figure 2.6: Illustration showing dual-projector scanner calibration.

2.3 Calibration

The procedure for calibrating our scanner is illustrated in Fig. 2.6 and involves sweeping a rigid, flat checkerboard pattern in correspondence with the XY -plane and moving in fixed, small-steps through the scan volume from the nearest plane where $Z = Z_{min}$ to the farthest plane at $Z = Z_{max}$. The physical moving of the checkerboard pattern is performed using a motorized rail like the Velmex Bi-slide and VXM controller. These rails can be configured to move with a positional accuracy of 0.076 mm and repeatability of 0.005 mm. The specific checkerboard pattern that we use is the CalTag calibration target introduced by Atcheson *et al.* [131] where, inside each square, is a unique 16-bit codeword identifying the absolute world (X_w, Y_w) coordinates of each square's four corners. Figure 2.7 shows a CALTag we use when calibrating a structured light scanning system.

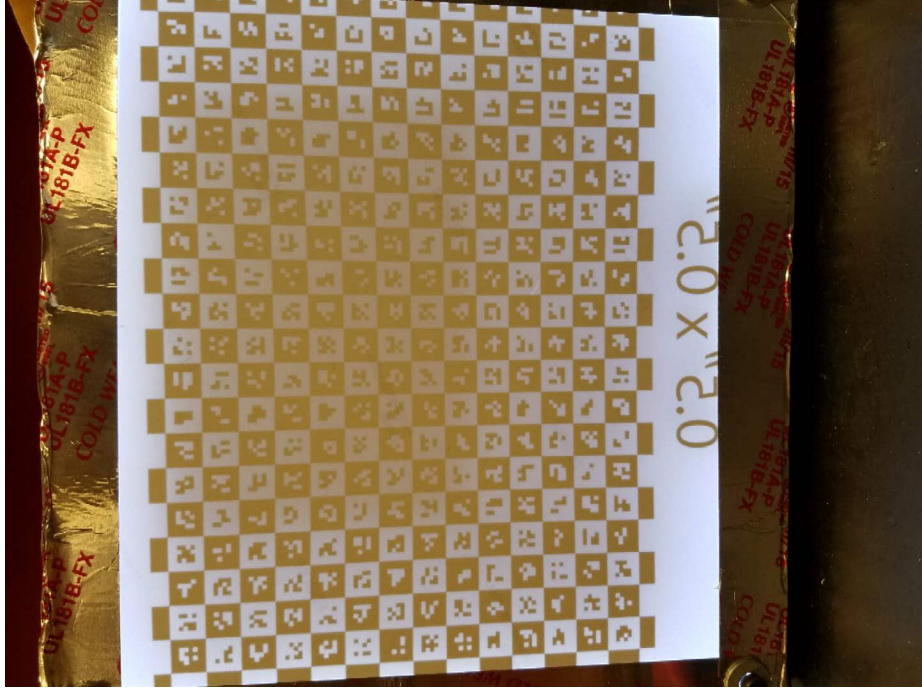


Figure 2.7: One sample CalTag in our system

2.4 Experiments

In this section, a series of experiments and their results are presented to show how the structured light 3D scanning improves with our dual-projector system. For these experiments, we use two Optoma ML750ST projectors paired with a Basler ACE acA1920-150u and an Embedded Micro Mojo V3 Spartan 6 FPGA board with dual-HDMI shield. The generated video is 800 by 600 pixels running at 120 Hz. In order to maintain 120 fps, we sub-windowed the image sensor to its center 1088 by 960 pixels with a 8.1 msec exposure and no trigger delay.

Two projectors operating simultaneously

In this experiment, we demonstrate that we can scan with both projectors simultaneously without cross talk between light engines. Specifically, we employ the scanner

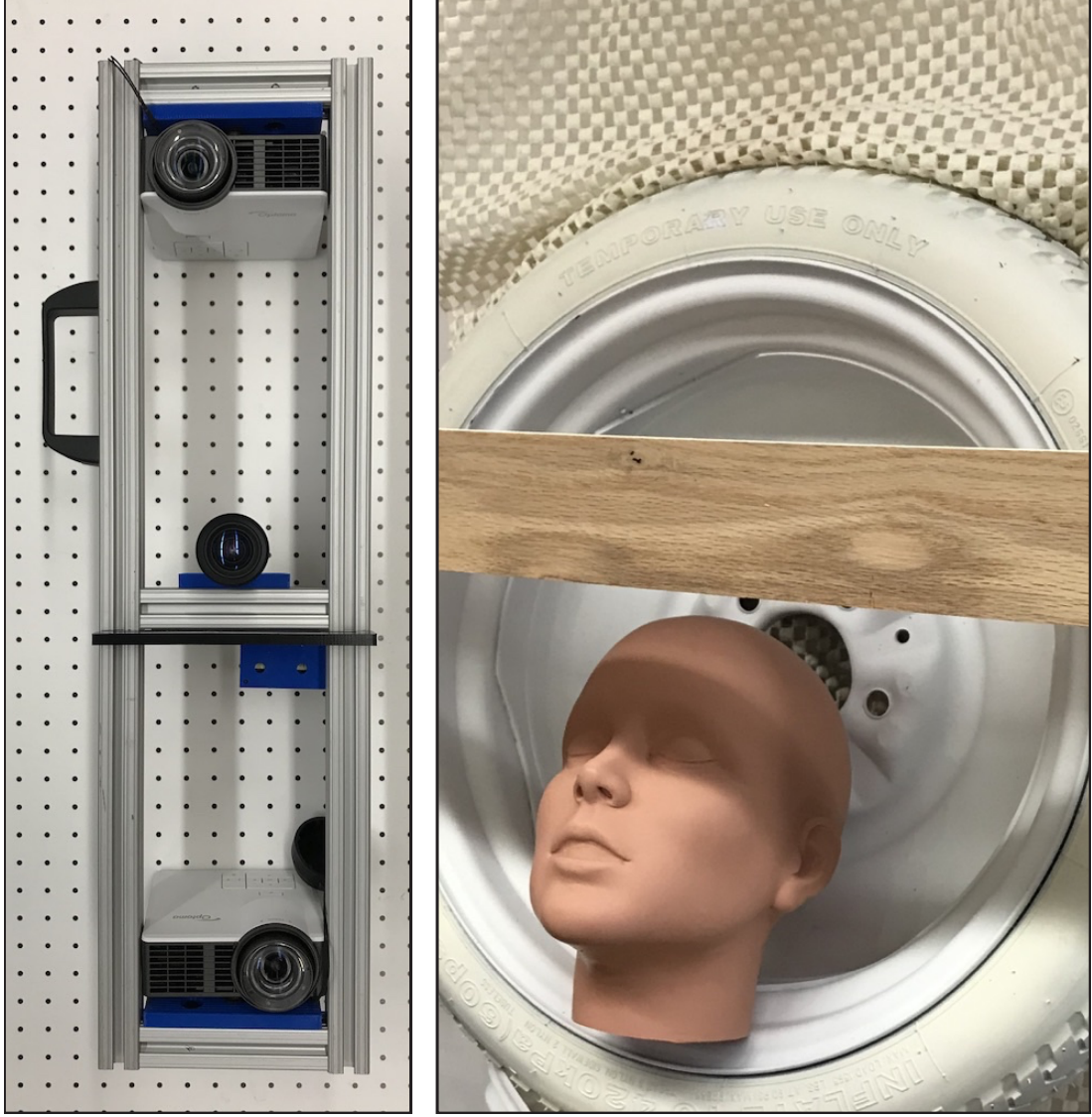


Figure 2.8: The (left) dual-projector scanner and the (right) scanning target composed of a 22.9" diameter tire with steel rim, both painted with matte white; a plastic makeup artist face/head; a non-slip area rug pad; and an occluding 4" wide, 0.25" thick piece of hobby wood.

arrangement illustrated in Fig. 2.8 where the two projectors are arranged in opposition. A single scan is composed of 3 sets of 8 patterns, with each set corresponding to spatial frequencies of 1, 6, and 18. Phase unwrapping is then used to derive a single, unambiguous phase value from 0 to 1 where 0 represents the bottom of the projected image and 1 represents the top. In Fig. 2.9, we show the vertical gradient

of the phase image of a flat reference plane with (left) only one projector turned on during scanning and (right) both projectors turned on during scanning. From visual observation, these two phase gradients are, for all intents and purposes, identical with negligible crosstalk between light engines.

Occlusion

Although an obvious benefit of using two projectors [132], we consider the reduction in occlusions achieved by using two projectors placed in opposition in Fig. 2.8. In this case, we use a 22.9" diameter tire with steel rim, both painted with a matte finish white spray paint; a plastic makeup artist face/head sitting inside the rim; and at the top, a non-slip area rug pad. We further placed a 4" wide by 0.25" thick piece of hobby wood occluding the tire surface so as to create two separate shadows. Now instead of using the complicated ICP merging scheme of Jiang *et al* [126], we propose using a simple weighted averaging scheme where we weight the world coordinates, \vec{x}_1 and \vec{x}_2 , from the two component scans by their corresponding SNR magnitudes B_1^c and B_2^c .

For those pixels whose magnitudes are considered unreliable (below a pre-determined threshold), the world coordinates are discarded, and the second of vectors, if available, are preserved. The scheme can be expressed as:

$$\vec{x}_m = \begin{cases} \vec{x}_1 \frac{B_1^c}{B_1^c + B_2^c} + \vec{x}_2 \frac{B_2^c}{B_1^c + B_2^c}, & \vec{x}_1 \text{ and } \vec{x}_2 \text{ both reliable} \\ \vec{x}_1, & \vec{x}_2 \text{ unreliable} \\ \vec{x}_2, & \vec{x}_1 \text{ unreliable} \end{cases} \quad (2.15)$$

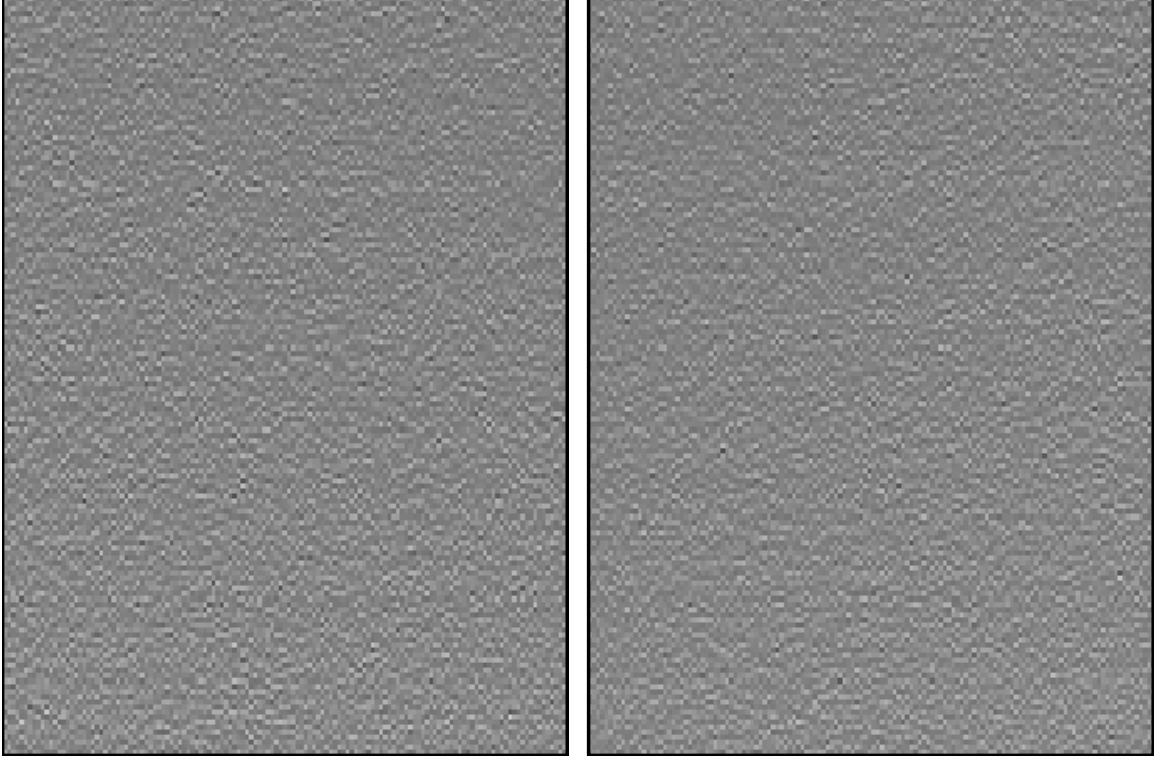


Figure 2.9: Phase gradient of the reference plane for projector 1 obtained when only projector 1 is scanning (left), and while both projectors are operating (right)

in which \vec{x}_m is the merged world coordinate. Shown in Fig. 2.10 is a close up of the reconstructed tire surface, where the (left) and (center-left) columns are the reconstructions from each individual projector. The white areas either above or below the hobby wood are the occlusions. The (center-right) column presents the result of the weighted averaging scheme where both occlusions are removed and the final scan recovers the vast majority of surface areas visible to the camera.

Increasing Signal Strength

As was denoted in (2.10), dual-frequency scanning requires a halving of the amplitudes of the component sinusoids compared to projecting each spatial frequency in temporal order; whereas, using two projectors maintains the signal strength of each. The

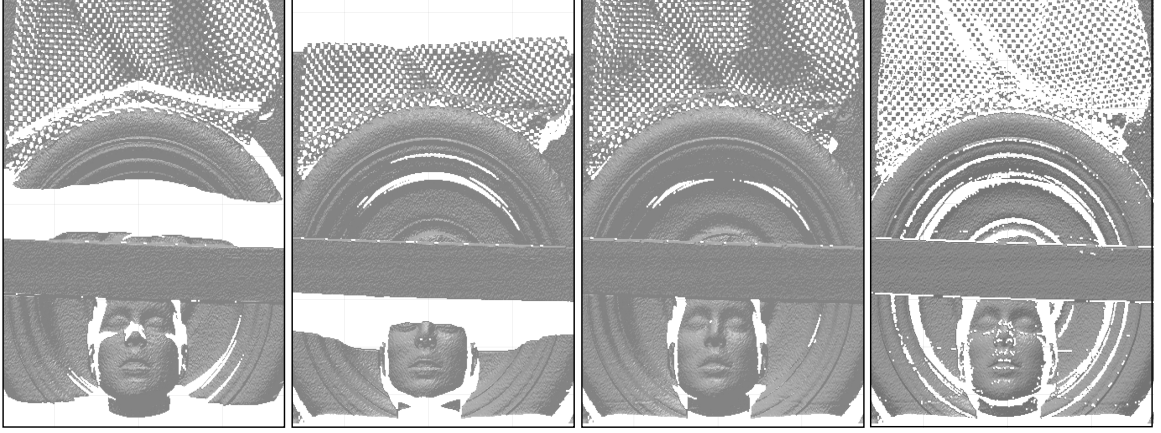


Figure 2.10: 3D reconstruction with (left) only the projector at the bottom; (center-left) with only the projector at the top; (center-right) by merging previous two using weighted averaging scheme; (right) after removing the pixels with multi-path

problem with using two separate projectors, though, is that there is no deterministic relationship between phase values from one projector to the next. As such, the traditional means of phase unwrapping defined according to [132]:

$$\tilde{\phi}_h = (\phi_2 + \lfloor f_h \phi_1 - \phi_2 + 0.5 \rfloor) / f_h, \quad (2.16)$$

where $\tilde{\phi}_h$ is the unwrapped phase that will be used to calculate the world coordinates, f_h is spatial frequency of high frequency patterns, ϕ_1 and ϕ_2 denote the phases of unit frequency patterns and high frequency patterns respectively, cannot be used to unwrap phase in a dual-projector scanner. Instead, the phase relationship between each projector is unique to each pixel of the camera, and as such, we propose phase unwrapping according to:

$$\tilde{\phi}_h = (\phi_2 + \lfloor LUT[x^c, y^c, \phi_1] - \phi_2 + 0.5 \rfloor) / f_h, \quad (2.17)$$

where $LUT[x^c, y^c, \phi_1]$ is built during calibration.

Figure 2.11 illustrates the proposed process of cross-projector phase unwrapping where each figure is split into its (left) raw phase and (right) phase gradi-

ent image. Specifically, Fig. 2.11 (left) shows the raw/gradient phase for the unit-frequency phase scan of the first of two projectors; Fig. 2.11 (center-left) shows the raw/gradient phase for the high frequency phase scan of the second of two projectors; Fig. 2.11 (center-right) shows the predicted low-frequency phase produced by the look-up table, $LUT[x^c, y^c, \phi_1]$; and Fig. 2.11 (right) shows the unwrapped high-frequency phase image.

Detecting Bimodal Multi-path

Bimodal multi-path is an undesirable but common phenomenon in structured light scanning for objects with step edges. Zhang *et al.* [133] define bimodal multi-path as the specific instance of a camera pixel seeing both a fore and background surface. This results in a linear combination of two sinusoids whose superposition results in a magnitude that fluctuates between constructive and destructive interference, depending on the spatial frequency, f . This fluctuation is characterized by the equation:

$$|\vec{A}\vec{B}|^2 = |\vec{A}|^2 + |\vec{B}|^2 + 2|\vec{A}||\vec{B}|\cos(2\pi f(y_a^p - y_b^p)), \quad (2.18)$$

where $|\vec{A}\vec{B}|$ is magnitude of the multi-path pixel seen by the camera; $|\vec{A}|$ and $|\vec{B}|$ are the magnitude components that are reflected from the fore and background surfaces respectively; f is the spatial frequency of the structured light patterns; and y_a^p and y_b^p are the row coordinates for the two paths of the pixel.

While its possible to separate the fore and background components from (2.18) with just three scan frequencies, the presence of noise in the sensor makes it unlikely, but we can use the magnitudes of our three frequencies as a test for multi-path, delet-

ing any pixels that fail. Specifically when no multi-path exists and with ideal optics and taking into account the reduction in magnitude of the projected sinusoids due to projector lens defocus, we expect the ratio of the low to medium frequency magnitude to closely approximate the ratio of the medium to high frequency magnitude. So we propose that if the two ratios are not within a pre-defined threshold, then we delete the pixel from our 3D reconstruction.

To better describe the detection of bimodal multi-path in an SLI scan mathematically, we define four variables as the ratio of low to medium and medium to high frequency magnitudes for each projector:

$$\begin{aligned} R_1^L &= \frac{B_1^L}{B_1^M}, \quad R_1^H = \frac{B_1^M}{B_1^H} \quad \text{and} \\ R_2^L &= \frac{B_2^L}{B_2^M}, \quad R_2^H = \frac{B_2^M}{B_2^H}, \end{aligned} \tag{2.19}$$

where R_1^L and R_2^L are the ratios of low to medium frequency magnitude of projector 1, R_1^H and R_2^H are the ratios of medium to high frequency magnitude of projector 2. B_1^L , B_1^M , B_1^H correlate with magnitudes of low, medium, and high frequency patterns of projector 1, which can be derived from (2.3) by dividing the total N frames into three pieces; B_2^L , B_2^M , B_2^H represent the magnitude of low, medium and high frequency patterns of projector 2, similarly they can be calculated from (2.11).

Ideally magnitude of low, medium and high frequency patterns for one same projector should be very close unless there is multi-path interference at that pixel. So for any good pixels, the value of R_1^L , R_1^H , R_2^L , R_2^H are all supposed to be near 1. In our experiment, we consider pixels that meet the following conditions good, otherwise,

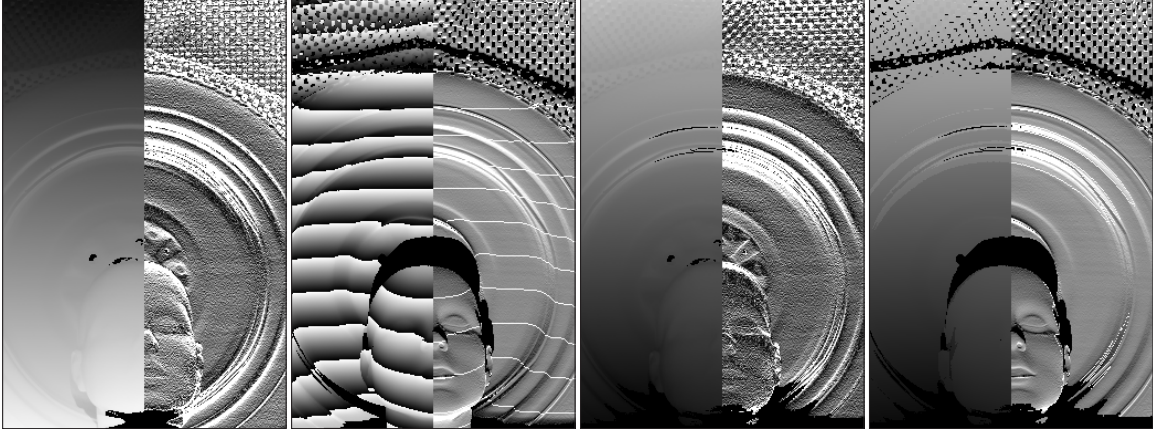


Figure 2.11: The raw and gradient images (left) the unit frequency phase from projector 1; (center-left) the high frequency phase from projector 2; (center-right) the predicted unit frequency of projector 2; (right) unwrapped high frequency phase of projector 2.

we remove the pixels from the reconstruction.

$$\begin{aligned} |R_1^L - 1| < 0.3, \quad |R_1^H - 1| < 0.3, \\ |R_2^L - 1| < 0.3, \quad |R_2^H - 1| < 0.3, \end{aligned} \tag{2.20}$$

Of course, it is not a given that even in the presence of multi-path, that at least one of the three spatial frequencies will land in a valley of the cosine in eqn. (2.18). Having two projectors at two different angles makes it far more likely, however.

Shown in Fig. 2.10 (left) is an illustration of the surface removal achieved with our ratio test for bimodal multi-path. In particular, bimodal multi-path is especially prevalent along the step edges of the occluding hobby wood as well as the inner step edges of the wheel rim. It's also very prevalent in the step edges formed by the holes in the non-slip area rug pad. To better appreciate these artifacts, Fig. 2.12 shows the 3D reconstructions of Fig. 2.10 (left-center) and (left) from a different perspective. What is most striking in Fig. 2.12 (left) are the spikes forward and back along the edges of hobby wood and its shadow caused by the averaging a phase between fore

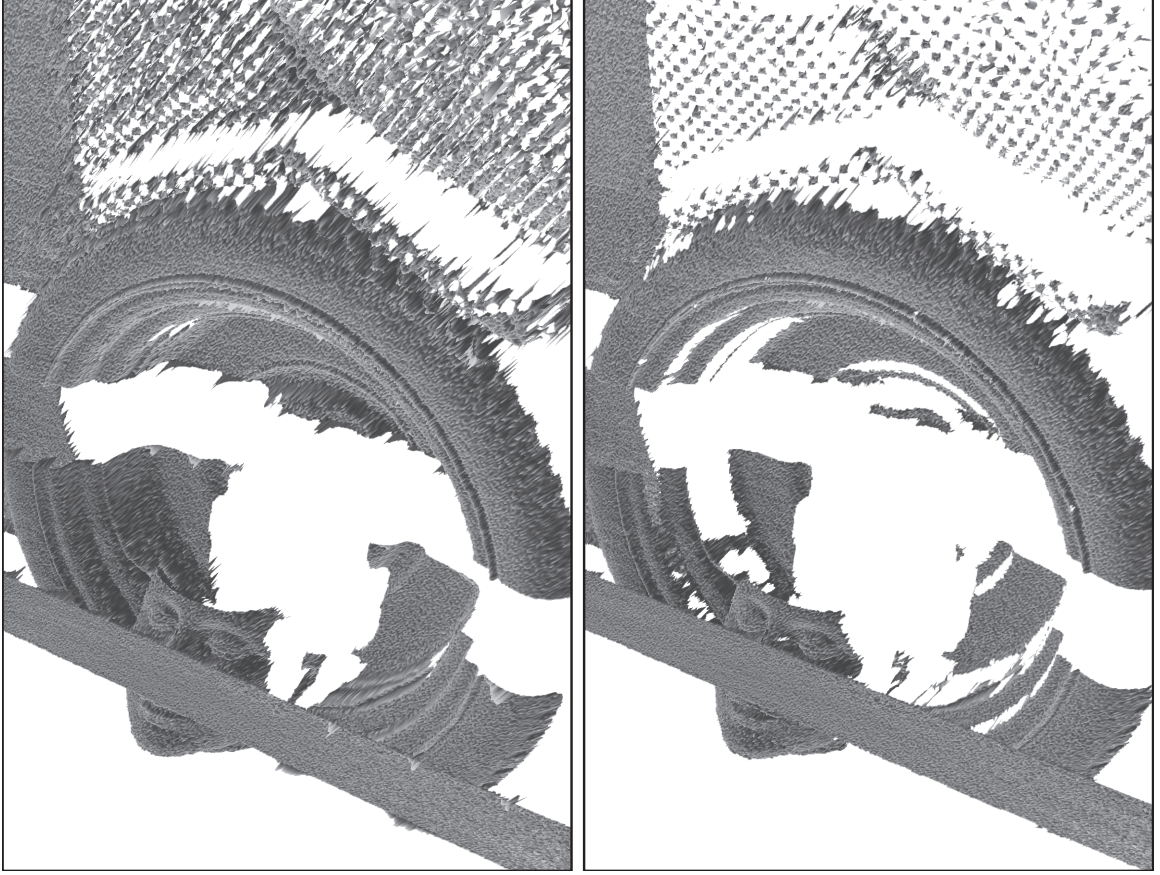


Figure 2.12: 3D reconstructions (left) before and (right) after bimodal multi-path removal.

and background surfaces. These spikes are largely, if not completely, eliminated in Fig. 2.12 (right). While the number of valid scan points have been diminished, it's important to remember that these bimodal multi-path points were never valid to begin with, and their removal is a great improvement in terms of scan accuracy.

2.5 Summary

In this chapter, we have presented a novel dual-projector structured light system to improve the quality of the 3D scanning. Specifically, we have introduced an FPGA-based hardware platform that can be easily extended to a diverse range of scanning

architectures containing multiple projectors. In a head-to-head comparison with single projector scanning, this Chapter has demonstrated the dual-projector scanners inherent advantages for minimizing occlusions, increasing projector signal strength, reducing the number of video frames needed for scanning, and identifying and removing bimodal multi-path. And since one could easily trigger multiple cameras with the introduced FPGA controller, one could easily extend this work to multiple-camera-to-multiple-projector scanners. Ultimately, we hope readers will see this work not as a one-off, novelty scanner but as an entirely new scanning platform opening doors to promises and challenges associated with pattern coding and decoding, improved multi-path reconstruction, noise reduction, and component phase fusion as the proposed algorithms were merely exemplary.

Chapter 3 Detecting and Removing Multi-path in Structured Light Scanning Using A Dual-projector System

The multi-path is the occurrence where one pixel of the image sensor captures light reflecting from two or more different paths due to the complex geometry of the object surface [111]. It is almost inevitable in places where discontinuities split one camera pixel into pieces, or where light interreflections take places before it is collected by the camera [133, 134]. The result of 3D measurement would be seriously compromised if those pixels suffered from multi-path effect are not properly dealt with.

The multi-path issue has drawn much attention in the area of time-of-flight 3D imaging [112, 135, 136], where light reflects off specular surfaces onto neighboring surface points before reflecting back to the camera. Considerable efforts have been made to remove or minimize the influence of multi-path. Freedman *et al.* propose a sparse reflections analysis algorithm that is able to remove any general multi-path interference from ToF images [113]. Naik *et al.* [114] introduce an image formation model to separate the global component of radiance from the direct component based on the Nayar method [137] without any impractical assumption of the sparsity in optical reflections.

The multi-path interference also negatively impacts the structured light 3D scanning. As an active imaging technique, phase measuring profilometry is susceptible to errors and distortions caused by the redirection of the projected light from multiple paths from projector to camera besides the direct path of projector to target to

camera [111]. O’Toole *et al.* develop a model where only indirect reflection paths are extracted to analyze in order to better understand how the unconventional indirect light behaves [124]. Zhang *et al.* report on a constructive and destructive interference cue for bimodal multi-path, and present a practical approach to simultaneously identify and extract the dominant phases/magnitudes by taking advantage of this cue in an intuitive way without any hardware modifications or additional requirements for customized patterns, but it requires scans of the target with a large range of spatial frequencies [133].

Direct illumination represents the lights coming from sources, falling on the target surface, and reflecting in the direction of the viewer [138]. Alternatively, indirect/global illumination may come from many sources, such as diffuse reflection [139, 140], subsurface scattering [141], and so on. In the case of dense multi-path [137], direct light and indirect/global illumination are mixed together. The goal of dense multi-path correction is, therefore, to separate the direct light path from the indirect illumination/global illumination [142, 143].

Bimodal multi-path is a sparse multi-path model that occurs in SLI when a camera pixel sees light from exactly two projector positions, and this type of multi-path affects phase images regardless of frequency. It is most commonly associated with complicated scan objects having step edges and other discontinuities in depth or specular targets that reflect light away from the camera. To do correction here means to separate two light paths and realize two decoupled depth measurements.

In this Chapter, we take advantage of the Zhang’s algorithm to separate those bimodal multi-path pixels’ intensity back into its source formations, including both fore-

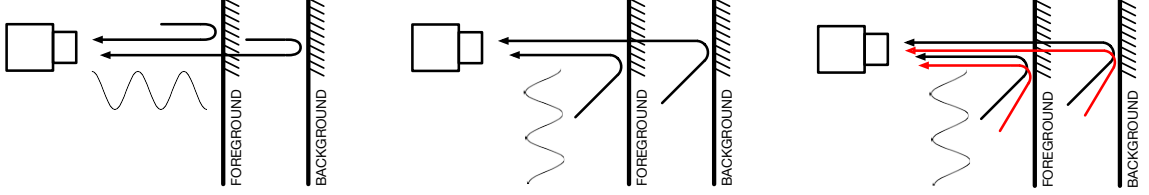


Figure 3.1: Illustration of bimodal multi-path for (left) time-of-flight imaging, (center) conventional structured light imaging, and (right) our dual-projector structured light imaging.

ground path and background path. Since structured light patterns with a wide range of spatial frequencies need to be scanned, the dual-frequency pattern scheme [122] is applied to obtain higher speed. Furthermore, we also demonstrate how the dual-projector setup [132] helps us obtain better results not only theoretically but also experimentally. Unlike the dual-projector scan where two projectors operate sequentially and independently [126], our dual-projector setup synchronize two projectors, so it is twice faster in theory.

3.1 Bimodal Multi-Path

As described in Chapter 1 section 2, SLI by means of phase measuring profilometry defines the set of projected patterns, $\{I_n^p : n = 0, 1, \dots, N - 1\}$, according to:

$$I_n^p(x^p, y^p) = \frac{1}{2} + \frac{1}{2} \cos \left(2\pi \left(\frac{n}{N} - ky^p \right) \right). \quad (3.1)$$

where (x^p, y^p) is the column and row coordinates of a pixel in the projector, I_n^p is the intensity of that pixel in a projector with dynamic range from 0 to 1, n represents the phase-shift index over the N total patterns, and k is the spatial frequency equal to the number of sinusoidal periods spanning the projector's field of view. For

reconstruction, the captured images are expressed as:

$$I_n^c(x, y, k) = A(k) + B(k) \cos \left(\frac{2\pi n}{N} - \theta \right). \quad (3.2)$$

where (x, y) is the column and row coordinates of a pixel in the camera while $I_n^c(x, y, k)$ is the intensity of the pixel. To simplify the notation, $I_n^c(k)$ will be used henceforth instead of $I_n^c(x, y, k)$. The parameter θ , represents the phase value of the captured sinusoidal pattern. The term $A(k)$ is the averaged pixel intensity across the pattern set that includes the ambient light component, correspondingly, the term $B(k)$ is the intensity modulation of a given pixel and is derived from $I_n^c(k)$, both of them are functions of k as stated by the optical transfer function [144]. They can be expressed in terms of real and imaginary components where:

$$B^R(k) = \sum_{n=0}^{N-1} I_n(k) \cos \left(\frac{2\pi n}{N} \right) \quad \text{and} \quad B^I(k) = \sum_{n=0}^{N-1} I_n(k) \sin \left(\frac{2\pi n}{N} \right) \quad (3.3)$$

such that

$$B(k) = \|B^R(k) + jB^I(k)\| = \left\{ B^R(k)^2 + B^I(k)^2 \right\}^{\frac{1}{2}}, \quad (3.4)$$

which is the magnitude of the observed sinusoids. The phase θ , of the captured sinusoidal pattern is finally derived as:

$$\theta(k) = \angle(B^R(k) + jB^I(k)) = \arctan \left\{ \frac{B^I(k)}{B^R(k)} \right\}, \quad (3.5)$$

which is used to derive the projector row indices using an appropriate phase unwrapping technique [132]. The principal problem of bimodal multi-path occurs when a pixel collects light from both fore and background surfaces such that:

$$B^R(k) = B_f^R(k) + B_b^R(k) \quad \text{and} \quad B^I(k) = B_f^I(k) + B_b^I(k) \quad (3.6)$$

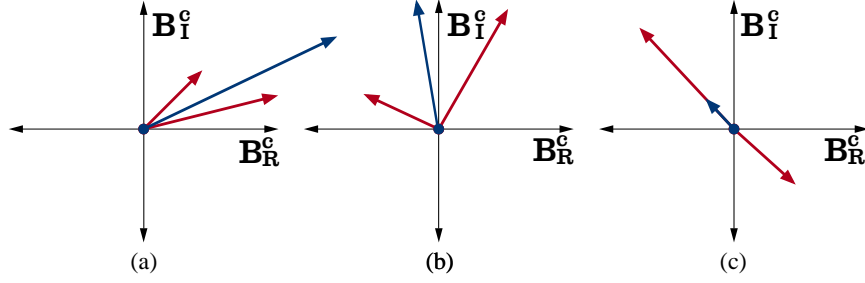


Figure 3.2: Illustration of the change in direction and magnitude, on the (B^R, B^I) -axis, in the (blue) observed complex vector $B^R + jB^I$ created by the superposition of (red) complex vectors from multi-path fore and background objects for $k = 1$, $k = 8$, and $k = 12$.

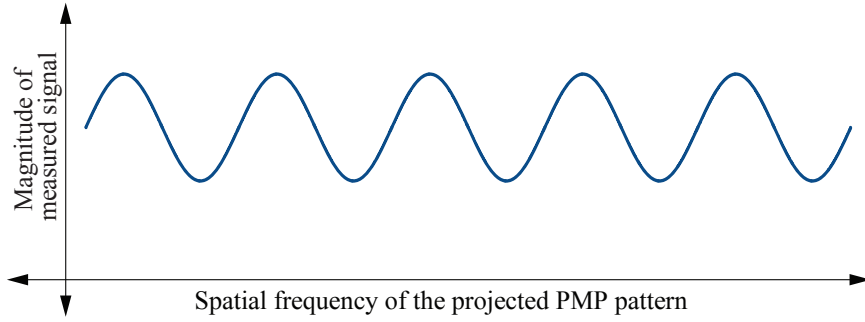


Figure 3.3: Illustration of the change in (vertical axis) magnitude of the observed complex vector, $B^R + jB^I$, as a function of the (horizontal axis) scaling factor k .

where we added the subscripts f and b to distinguish between the foreground and background components on $B^R(k)$ and $B^I(k)$ as illustrated in Fig. 3.1. The magnitude and phase of a subject pixel are then defined according to vector $\vec{B}_{fb}[k]$, with foreground vector $\vec{B}_f(k)$ and background vector $\vec{B}_b(k)$, such that:

$$|\vec{B}_{fb}[k]|^2 = |\vec{B}_f[k]|^2 + |\vec{B}_b(k)|^2 + 2|\vec{B}_f(k)||\vec{B}_b(k)|\cos(2\pi k(y_a^p - y_b^p)) \quad (3.7)$$

where y_a^p and y_b^p are the projector row coordinates for the two paths, $|\vec{B}_{fb}[k]|$ is the measured magnitude of a pixel from the captured images at the given spatial frequency k .

Figure 3.2 shows an example of a foreground vector and a background vector and

their resulted pixel vector at different spatial frequencies, and Fig. 3.3 demonstrates that the magnitude of the measured pixel under multi-path interference changes in a sinusoidal manner.

To now extract $\vec{B}_f(k)$ and $\vec{B}_b(k)$, Zhang et al propose a two-step procedure where the first step finds the parameters $|\vec{B}_f(k)|$, $|\vec{B}_b(k)|$, and $dy = y_a^p - y_b^p$ that minimize the mean-squared error given by:

$$(|\vec{B}_f^*|, |\vec{B}_b^*|, dy^*) = \underset{|\vec{B}_f(k)|, |\vec{B}_b(k)|, dy}{\operatorname{argmin}} \sum_k \{|\vec{B}_{fb}[k]| - |\vec{B}_f(k) + \vec{B}_b(k)|\}^2. \quad (3.8)$$

Note that $|\vec{B}_f(k)|$, $|\vec{B}_b(k)|$ are not independent variables, they are constrained by the zero-frequency scan where $k = 0$, and are defined as:

$$|\vec{B}_{fb0}|^2 = |\vec{B}_f(k)|^2 + |\vec{B}_b(k)|^2 + 2|\vec{B}_f(k)||\vec{B}_b(k)|. \quad (3.9)$$

From this, we get the constraint:

$$|\vec{B}_f(k)| + |\vec{B}_b(k)| = |\vec{B}_{fb0}| \quad (3.10)$$

Here, we define three matrices \mathbf{X} , \mathbf{Y} , \mathbf{M} which will be used later in our dual-projector setup to better demonstrate the properties of Eq. (3.8) as:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \cos(2\pi \cdot 0 \cdot (y_a^p - y_b^p)) \\ 1 & 2 \cos(2\pi \cdot 1 \cdot (y_a^p - y_b^p)) \\ \vdots & \vdots \\ 1 & 2 \cos(2\pi \cdot 60 \cdot (y_a^p - y_b^p)) \end{bmatrix}, \quad (3.11)$$

$$\mathbf{Y} = \begin{bmatrix} |\vec{B}_f(k)|^2 + |\vec{B}_b(k)|^2 \\ |\vec{B}_f(k)| \cdot |\vec{B}_b(k)| \end{bmatrix} \text{ and } \mathbf{M} = \begin{bmatrix} |\vec{B}_{fb}^0|^2 \\ |\vec{B}_{fb}^1|^2 \\ \vdots \\ |\vec{B}_{fb}^{60}|^2 \end{bmatrix}, \quad (3.12)$$

where the size of \mathbf{X} , \mathbf{Y} , \mathbf{M} are 61×2 , 2×1 and 61×1 respectively and $|\vec{B}_{fb}^0|$, $|\vec{B}_{fb}^1|$, $\dots |\vec{B}_{fb}^{60}|$ denote the measured pixel magnitude at spatial frequencies from 0 to 60. So that solving the Eq. (3.8) becomes finding the least squares solution of \mathbf{Y} for

$$\mathbf{XY} = \mathbf{M}. \quad (3.13)$$

And once this is solved, the second step is to obtain the absolute phases, y_a^p and y_b^p , by minimizing the mean squared error given by:

$$(y_a^{p*}, y_b^{p*}) = \underset{y_a^p, y_b^p}{\operatorname{argmin}} \sum_k \{ \vec{B}_{fb}[k] - (\vec{B}_f(k) + \vec{B}_b(k)) \}^2 \quad (3.14)$$

such that $|\vec{B}_f(k)| = |\vec{B}_f^*|$, $|\vec{B}_b(k)| = |\vec{B}_b^*|$, and $y_a^p - y_b^p = dy^*$.

This algorithm reduces the search space in the first step from the three independent variables $|\vec{B}_f(k)|$, $|\vec{B}_b(k)|$, and dy to two, $|\vec{B}_f(k)|$ and dy , so that we can perform an exhaustive search over $|\vec{B}_f(k)|$, $|\vec{B}_b(k)|$, and the phase difference $y_a^p - y_b^p$ along the line $|\vec{B}_b(k)| = |\vec{B}_{fb}^0| - |\vec{B}_f(k)|$ to find the values that minimize the mean-squared error in Eq. (3.8) over all scanned values of k . In this way, we reduce the computing complexity in the first step from $O(n^3)$ to $O(n^2)$ while keeping the second step as $O(n)$ since $y_a^p - y_b^p$ has already been determined from the first step. So the overall algorithm complexity is reduced from $O(n^4)$ to $O(n^2)$ based on the proposed two-step optimization. Algorithm 1 lists the above procedures.

3.2 Dual-projector system setup

A dual-projector system has been developed in Chapter 2 to yield better result for structured light scanning in terms of increased signal strength, minimized occlusion and faster scanning speed, we find that it helps in addressing the multi-path problem as well. As low spatial frequency structured light patterns are more susceptible to sensor noise, higher frequency patterns are often employed to achieve higher accuracy. But the high frequency patterns alone give rise to phase ambiguity, unless the corresponding low frequency patterns come together with high frequency ones to unwrap the ambiguous high frequency phase. With a high frequency pattern set and a low frequency set to be projected, the length of the structured light sequence doubles, so does the time it takes to process a scan.

Algorithm 1 Two Step Optimization Reconstruction Algorithm

```

1: procedure Reconstruction( $\vec{B}_{fb}$ )
2:   for  $|\vec{B}_f(k)| \leftarrow 0$  to  $|\vec{B}_{fb0}|$  do
3:      $|\vec{B}_b(k)| \leftarrow |\vec{B}_{fb0}| - |\vec{B}_f(k)|$ 
4:     for  $dy \leftarrow 0$  to  $2\pi$  do
5:       if best error in Eq. (3.13) then
6:          $|\vec{B}_f^*| \leftarrow |\vec{B}_f(k)|$ 
7:          $|\vec{B}_b^*| \leftarrow |\vec{B}_b(k)|$ 
8:          $|dy^*| \leftarrow |dy|$ 
9:       end if
10:    end for
11:  end for
12:  for  $y_a^p \leftarrow 0$  to  $2\pi$  do
13:     $y_b^p \leftarrow y_a^p - dy^*$ 
14:    if best error in Eq. (3.14) then
15:       $y_a^{p*} \leftarrow y_a^p$ 
16:       $y_b^{p*} \leftarrow y_b^p$ 
17:    end if
18:  end for
19: end procedure

```

To solve this dilemma between the speed and accuracy, Liu *et al.* propose a dual-frequency pattern scheme which combines a high frequency pattern and a low frequency pattern into one composite pattern [122]. Inspired by this scheme, in Chapter 2, we develop a dual-projector structured light scanning system where the dual-frequency patterns are fed to two projectors simultaneously. In such a system, the output patterns of the two projectors are expressed as:

$$I_n^{p_1}(x^{p_1}, y^{p_1}) = \frac{1}{2} + \frac{1}{2} \cos \left(2\pi k_1 y^{p_1} - \frac{2\pi n}{N} \right) \text{ and} \quad (3.15)$$

$$I_n^{p_2}(x^{p_2}, y^{p_2}) = \frac{1}{2} + \frac{1}{2} \cos \left(2\pi k_2 y^{p_2} - \frac{4\pi n}{N} \right). \quad (3.16)$$

where k_1 and k_2 are spatial frequencies of the structured light patterns coming out of the projectors. Note that because the two cosine patterns in Eq. (3.15) have distinct temporal frequencies that are determined by $\frac{2\pi n}{N}$ and $\frac{4\pi n}{N}$, the two spatial frequency components can later be separated applying temporal Fourier transform.

With such a dual-projector setup, the analysis of the bimodal multi-path becomes more complicated because instead of two paths, the dual-projector has four which are foreground from projector 1, $\vec{B}_{f_1}(k_1)$, background from projector 1, $\vec{B}_{b_1}(k_1)$, foreground from projector 2, $\vec{B}_{f_2}(k_2)$ and background from projector 2, $\vec{B}_{b_2}(k_2)$.

Suppose $|\vec{B}_{fb_1}^0|$ and $|\vec{B}_{fb_2}^0|$ denote the magnitude of a subject pixel for projector 1 and projector 2 at $k_1 = 0$ and $k_2 = 0$ respectively, then refer to Eq. (3.10), the constraints between $|\vec{B}_{f_1}(k_1)|$, $|\vec{B}_{f_2}(k_2)|$ and $|\vec{B}_{b_1}(k_1)|$, $|\vec{B}_{b_2}(k_2)|$ can be described as:

$$|\vec{B}_{f_1}(k_1)| + |\vec{B}_{b_1}(k_1)| = |\vec{B}_{fb_1}^0| \text{ and} \quad (3.17)$$

$$|\vec{B}_{f_2}(k_2)| + |\vec{B}_{b_2}(k_2)| = |\vec{B}_{fb_2}^0|. \quad (3.18)$$

We also use $y_a^{p_1}$ and $y_b^{p_1}$ to represent the phases of projector 1 foreground and background paths; likewise, $y_a^{p_2}$ and $y_b^{p_2}$ define the phases of projector 2 foreground and background. Although there are four phase terms which correlate to four paths for each pixel, they only emanate from two points of the surface, the foreground and the background, and in calibration, we recorded the phases of projector 1 and projector 2 at each discrete step along the z axis, which indicates the depth in the light path. So we can lock the two phase difference terms together based on phase measurements made between the projectors using the distance from the camera to the foreground, z_f , and the distance from the camera to the background, z_b , along the z axes. In other words, given z_f and z_b , $y_a^{p_1}$, $y_b^{p_1}$, $y_a^{p_2}$ and $y_b^{p_2}$ are all determined as functions of z , such that

$$y_a^{p_1} = y^{p_1}(z_f), \quad y_b^{p_1} = y^{p_1}(z_b), \quad (3.19)$$

$$y_a^{p_2} = y^{p_2}(z_f), \quad y_b^{p_2} = y^{p_2}(z_b). \quad (3.20)$$

From Eq. (3.13), a new equation for projector 1 can be expressed as:

$$\mathbf{X}_1 \mathbf{Y}_1 = \mathbf{M}_1, \quad (3.21)$$

where \mathbf{X}_1 , \mathbf{Y}_1 , \mathbf{M}_1 are 61×2 , 2×1 and 61×1 matrices that are defined as:

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 2 \cos(2\pi \cdot 0 \cdot (y_a^{p_1} - y_b^{p_1})) \\ 1 & 2 \cos(2\pi \cdot 1 \cdot (y_a^{p_1} - y_b^{p_1})) \\ \vdots & \vdots \\ 1 & 2 \cos(2\pi \cdot 60 \cdot (y_a^{p_1} - y_b^{p_1})) \end{bmatrix}, \quad (3.22)$$

$$\mathbf{Y}_1 = \begin{bmatrix} |\vec{B}_{f_1}(k_1)|^2 + |\vec{B}_{b_1}(k_1)|^2 \\ |\vec{B}_{f_1}(k_1)| \cdot |\vec{B}_{b_1}(k_1)| \end{bmatrix} \text{ and } \mathbf{M}_1 = \begin{bmatrix} |\vec{B}_{fb_1}^0|^2 \\ |\vec{B}_{fb_1}^1|^2 \\ \vdots \\ |\vec{B}_{fb_1}^{60}|^2 \end{bmatrix}, \quad (3.23)$$

and \mathbf{M}_1 is a known vector whose values are obtained from the collected scanning data. The scanning data consists scans with 61 pattern sets that have integer spatial frequencies ranging from 0 to 60. For any given pair of $y_a^{p_1}$ and $y_b^{p_1}$, Eq. (3.21) can be considered as an overdetermined system with respect to two variables $|\vec{B}_{f_1}(k_1)|$ and $|\vec{B}_{b_1}(k_1)|$. By running through the standard procedures of finding least squares solutions to overdetermined system, in which the transpose of \mathbf{X}_1 is multiplied to both sides of Eq. (3.21) as expressed in:

$$\mathbf{X}_1^T \mathbf{X}_1 \mathbf{Y}_1 = \mathbf{X}_1^T \mathbf{M}_1, \quad (3.24)$$

the best fitted solutions for Eq. (3.21) can be found if Eq. (3.24) is converted to the following expression:

$$\mathbf{Y}_1 = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{M}_1, \quad (3.25)$$

note that $|\vec{B}_{f_1}(k_1)|$ and $|\vec{B}_{b_1}(k_1)|$ are not independent, they are constrained by Eq. (3.17).

Similarly, $|\vec{B}_{f_2}(k_2)|$ and $|\vec{B}_{b_2}(k_2)|$ can be determined when $y_a^{p_2}$ and $y_b^{p_2}$ are given, by finding the least squares solution to:

$$\mathbf{X}_2 \mathbf{Y}_2 = \mathbf{M}_2, \quad (3.26)$$

where \mathbf{X}_2 , \mathbf{Y}_2 , \mathbf{M}_2 are the corresponding matrices of \mathbf{X}_1 , \mathbf{Y}_1 , \mathbf{M}_1 with respect to projector 2.

Then we use the determined $|\vec{B}_{f_1}(k_1)|$ and $|\vec{B}_{f_2}(k_2)|$ to search through the (z_{min}, z_{max}) , looking for the minimum weighted summation of the defined as:

$$(z_f^*, z_b^*) = \underset{z_f, z_b}{\operatorname{argmin}} E_1 + E_2, \quad (3.27)$$

where z_{min} and z_{max} define the entire range of z_f and z_b during calibration, E_1 and E_2 represent the least squares error of the vectors for projector 1 and projector 2 that are described as:

$$E_1 = \sum_{k_1} \{ \vec{B}_{fb_1}(k_1) - (\vec{B}_{f_1}(k_1) + \vec{B}_{b_1}(k_1)) \}^2 \text{ and} \quad (3.28)$$

$$E_2 = \sum_{k_2} \{ \vec{B}_{fb_2}(k_2) - (\vec{B}_{f_2}(k_2) + \vec{B}_{b_2}(k_2)) \}^2. \quad (3.29)$$

Algorithm 2 lists the above procedures.

Algorithm 2 Optimization Reconstruction Algorithm

```

1: procedure Reconstruction( $\vec{B}_{fb_1}$  and  $\vec{B}_{fb_2}$ )
2:   for  $z_f \leftarrow z_{min}$  to  $z_{max}$  do
3:     for  $z_b \leftarrow z_f$  to  $z_{max}$  do
4:       if best fitted solution for Eq. (3.21) then
5:          $|\vec{B}_{f_1}^*(k_1)| \leftarrow |\vec{B}_{f_1}(k_1)|$ 
6:          $|\vec{B}_{b_1}^*(k_1)| \leftarrow |\vec{B}_{b_1}(k_1)|$ 
7:       end if
8:       if best fitted solution for Eq. (3.26) then
9:          $|\vec{B}_{f_2}^*(k_2)| \leftarrow |\vec{B}_{f_2}(k_2)|$ 
10:         $|\vec{B}_{b_2}^*(k_2)| \leftarrow |\vec{B}_{b_2}(k_2)|$ 
11:      end if
12:      if best error in Eq. (3.27) then
13:         $z_f^* \leftarrow z_f$ 
14:         $z_b^* \leftarrow z_b$ 
15:      end if
16:    end for
17:  end for
18: end procedure

```

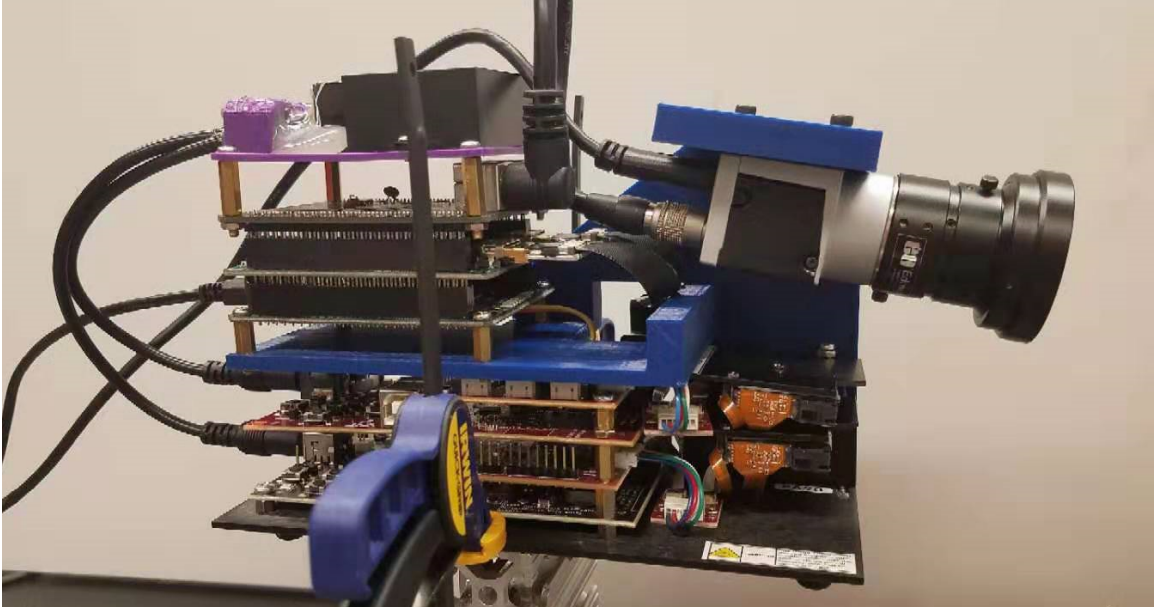


Figure 3.4: The dual-projector scanner we built for multi-path

3.3 Experiments

To implement such a system, we employ two identical Texas Instrument’s DLP2010 projectors, as well as a monocular Basler uc-1500 camera. Besides, we develop an FPGA-based controller to generate the separated structured light patterns feeding the two projectors and synchronize the camera to them.

The experiments are setup in two different manners. In the first one, we place a piece of wedge about one inch in front of the reference plain as shown in Fig. 3.5 (top), so some pixels that see the edge of the wedge also receive light reflecting from the reference plain. Thus bimodal multi-path should be observed at these pixels, because one pixel sees light reflecting from two points at different depths. The spikes in the figure at bottom-left demonstrate how multi-path affect the result of 3D reconstruction when it is not taken care of. The other two figures look much cleaner, because the pixels with multi-path are detected and simply replaced with either the path from

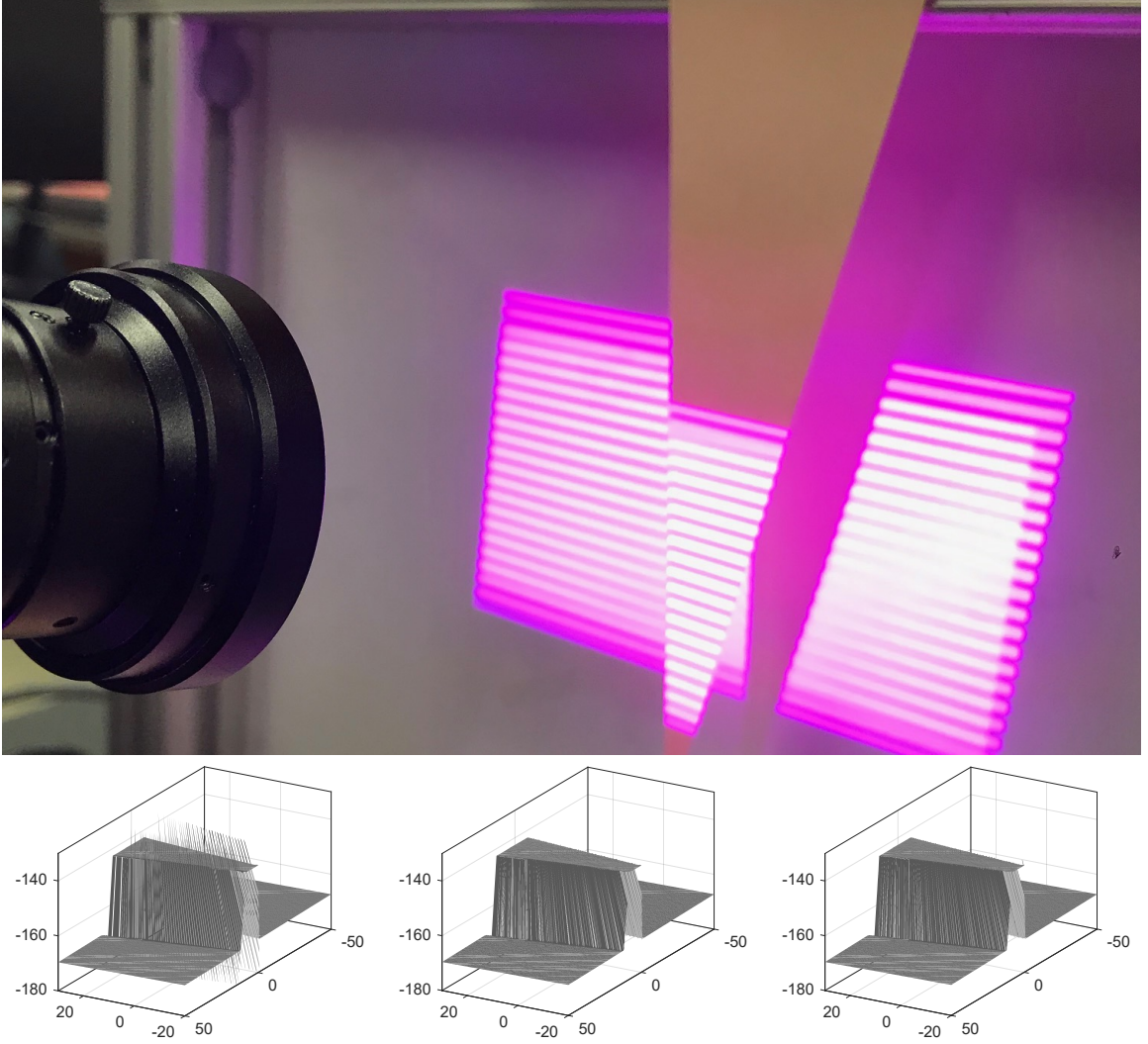


Figure 3.5: The scanned wedge (top) and the 3D reconstruction without any treatment for multi-path (bottom-left), replace the multi-path with foreground path (bottom-center), replace the multi-path with background path (bottom-right)

the foreground (bottom-center) or the path from the background (bottom-right).

In the second experiment, we use a teeth mold which later turns out to be an inappropriate 3D model to give rise to bimodal multi-path as the scanned object as seen in Fig. 3.6 (top). In the bottom-left figure where multi-path pixels are not specially taken care of, and noise in 3D reconstruction is noticeable, while 3D reconstruction after applying the proposed algorithm in the bottom-right figure shows little differ-

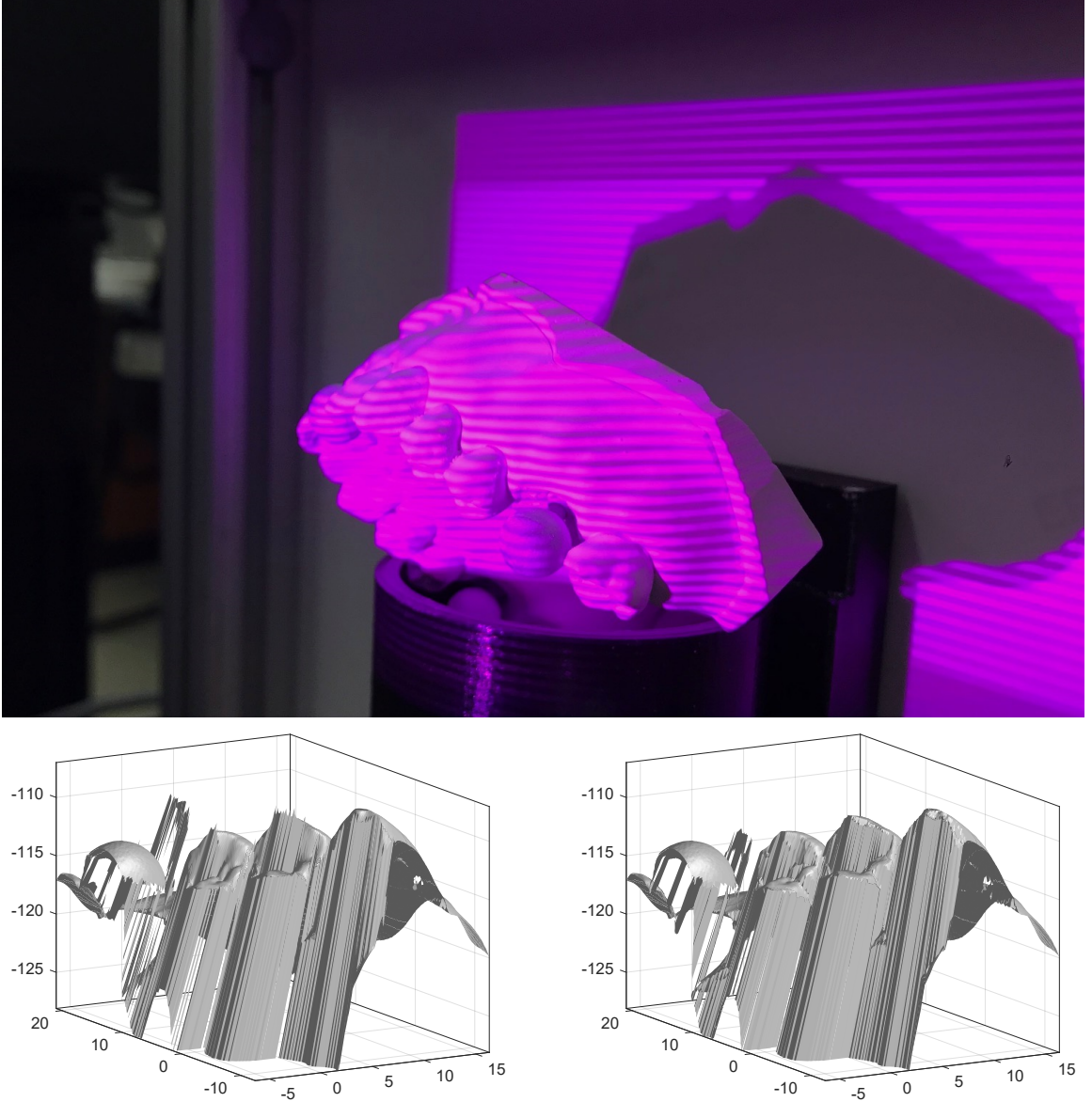


Figure 3.6: The scanned teeth mold (top) and the 3D reconstruction without any treatment for multi-path (bottom-left), 3D reconstruction after processing with the Algorithm 2 (bottom-right)

ence. Because the round and smooth surface of the teeth edges creates more than two light paths, so it goes beyond the scope of bimodal multi-path analysis we propose.

Lastly in Fig. 3.7, the differences of magnitudes in the two target objects among the different paths are presented. Most pixels in these figures do not have multi-path issue so they appear to be constant in all three figures. But for the detected multi-

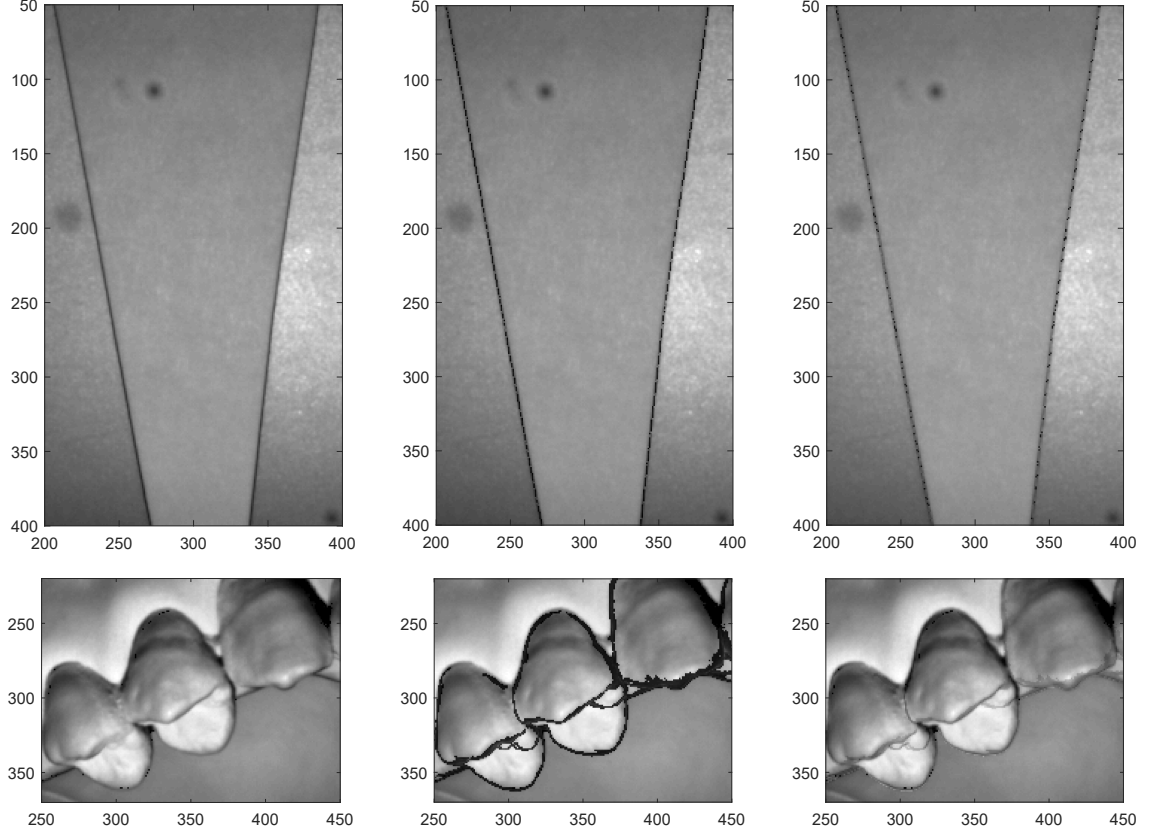


Figure 3.7: The magnitude maps of the wedge with multi-path (top-left), foreground path (top-center), background path (top-right) and the magnitude maps of the teeth mold with multi-path (bottom-left), foreground path (bottom-center), background (bottom-right)

path pixels, their magnitudes of light from foreground path look obviously darker as illustrated in the center figures.

3.4 summary

In this Chapter, we presented a dual-projector hardware platform and a corresponding algorithm to detect and analyze the multi-path interference in SLI 3D scanning. Structured light patterns of 61 different spatial frequencies were projected sequentially, from 0 to 60, the observed magnitude of the subject pixels change in a sinu-

soidal manner along the 488 patterns, which supported the theoretical analysis we proposed. Two exemplary objects, a piece of wedge and a teeth mold were scanned in the experiment part, the result of 3D reconstructions showed that the proposed algorithm only applies to bimodal multi-path, objects that generate multiple reflected paths need further research in the future.

Chapter 4 AR Face Paint and Headset Tracking

Spatial augmented reality is a special sort of AR in which projectors are used to visualize the virtual contents on top of the real world [46, 47]. Typical applications of SAR includes AR sandbox and AR pool table [48, 4]. As the research goes extensive, the accuracy of depth measurement for SAR systems becomes an vital parameter that needs to be taken into consideration. Harrison *et al.* developed a wearable multi-touch human interface system called OmniTouch which is built on a pico-projector and a depth camera [145]. The system projects a graphical user interface (GUI) on almost any surface, and the surface with the projected GUI magically turns out to be virtual touchscreen just like it on any smartphone. But it is different from the real touchscreen which contains a sensor detecting the physical contact between the human finger and the surface of touchscreen, for the OmniTouch system, glitches occur when the finger moves close enough to the surface. It would consider the surface and finger are contacted while they are actually not. Because the depth camera does not provide sufficiently accurate depth measurement. In seeking for a 3D measurement technique with high accuracy, structured light illumination stands out for its simple implementation, low cost, high speed and high accuracy [73]. Additionally, the exploitation of SLI enables the system to realize 3D position tracking which is also an important feature of any VR/AR headset.

As stated in the introduction, our ultimate goal is to develop an HDMI/camera controller that plugs into an existing AR table and instantly adds structured light

scanning. By doing so, the AR table has the ability to carry out high accuracy 3D measurements. Considering the SLI scanning system and SAR system both require the projector, a shared projector should be able to switch between SLI mode and SAR mode. In SLI mode, it projects PMP patterns while in SAR mode it displays virtual information passed from host PC.

In this Chapter, we describe the details of how our FPGA based SAR/SLI controller is implemented, also present an AR face paint system which takes advantage of our controller, as well as a position tracking system differing from all the existing tracking techniques. The face paint system consists of a projector, a camera and an FPGA-based controller. The projector, camera pair are calibrated using SLI in advance, and then the projector switch back and forth between the SAR and SLI mode. In every eight consecutive frames, there is one and only one frame projecting plain white light, during which time an image is taken by the camera. The plain white light can be considered as a special case of SLI pattern, where the spatial frequency equals to 0. By processing the image, the detected human face, if there is one, can be located in the 2D image coordinate system. Because the projector and camera are calibrated, the corresponding location of face in the projector coordinate can be obtained too. In this way, any augmented information can be rendered exactly on top the detected face in the following SAR mode.

The position tracking system consists of a projector, two cameras and an identical FPGA-based controller as we use for face paint. One camera serves as the base station which has a fixed position, it is paired with the projector and the correspondence between the 2D image coordinates and the 3D world coordinates can be determined

through calibration. Next the other camera which is presumed to be a sensor on headset takes a sequence of image containing the whole SLI pattern set. Given the condition that the 3D world coordinates of the patterns is known, a pin hole camera model can then be established to get the current location of the second camera with respect to the base station. This novel position tracking system system outshines in situations where there is no directly line of sight between the headset and the base station. Admittedly, it can not provide orientation data about the headset like like other tracking systems do, because the system only works when the second camera can see the projected patterns.

4.1 HDMI Receiver

To make our system SAR compatible, it is required to be able to operate as an HDMI sink which receives HDMI video streams. We built the HDMI receiver based on the application note from Xilinx [6], Fig. 4.1 illustrates the system block diagram. At the beginning of data flow, serial data streams are converted to the 10-bit HDMI encoded word individually for RGB channels. As the available primitives for the Xilinx Spartan 6 devices don't support 1 to 10 deserialization, the conversion is done in two steps. Firstly, 1 to 5 deserialization is carried out twice sequentially, and then a 5 to 10 gear box is used to combine them into one word.

The phase locked loop (PLL) reproduces a bit rate sample clock based on reference clock from the TMDS clock channel for the incoming serial streams. However, this bit rate sample clock may not be synchronized with any of the three data channels. Therefore, the job of the clock data recovery module is to individually adjust the clock

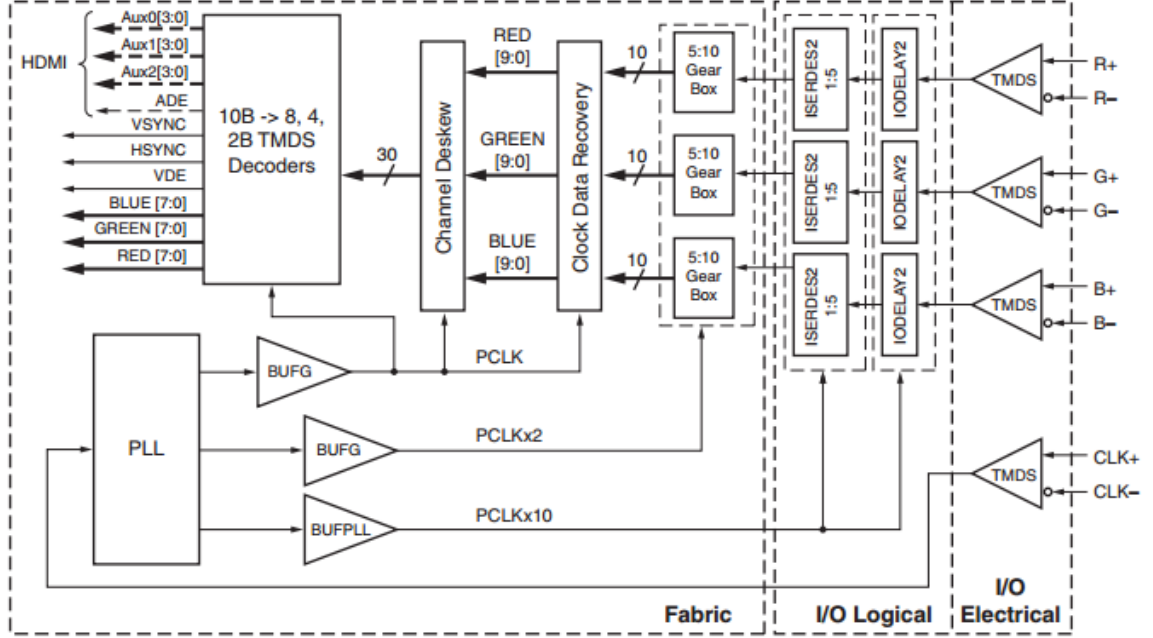


Figure 4.1: The block diagram of the HDMI receiver [6]

phase for each data lane to ensure that the incoming serial bits are sampled at the correct timing. Since the deserialization is not aware of word boundaries, a dedicated logic called word boundary detection is implemented to detect the beginning and the end of a word. The basic idea is to continuously compare the received raw data with the specifically defined control tokens until they match. The channel deskew module applies a FIFO-based logic to synchronize the RGB data lanes between which skews of up to 0.6 pixel time are tolerant.

To make a system AR compatible, it is required to be able to act as an HDMI sink which passes the EDID to the source via DDC. EDID is a data structure standardized by the video electronics standard association (VESA). It contains information such as the manufacturer name, the product ID, the input type, the supported timings, the display size and the luminance characteristics. Usually the EDID data is stored in the display system using a serial programmable read only memory (PROM) or

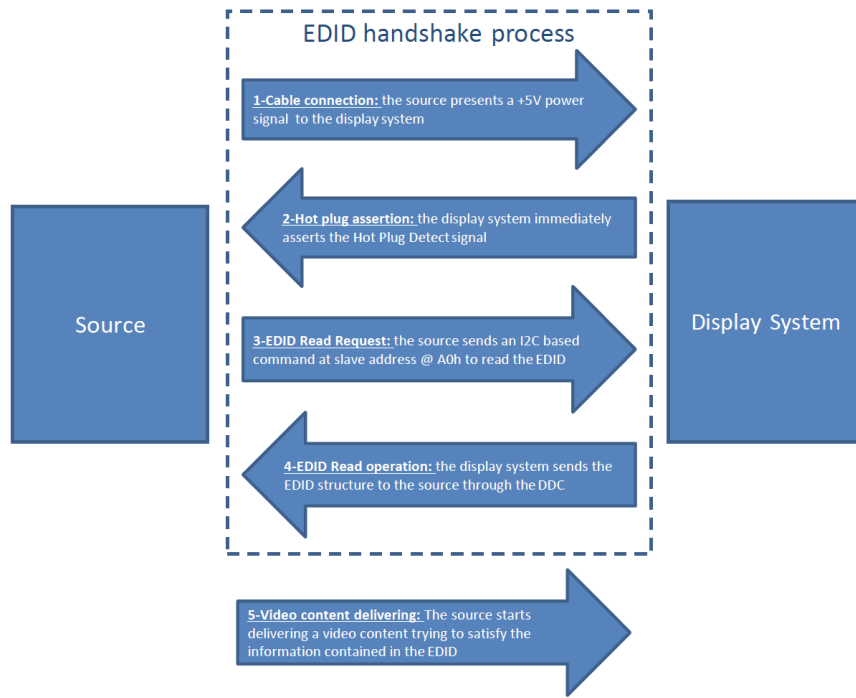


Figure 4.2: The process of EDID handshake [7]

a serial electrically erasable PROM (EEPROM). Three wires in the HDMI cable form the DDC channel, SCL, SDA and DDC/CEC Ground. These wires shall meet the requirements specified in the I2C-bus Specification, version 2.1, Section 15 for Standard-Mode devices. Figure 4.2 shows the process of EDID handshake that occurs at the initial stage upon an HDMI connection.

The EDID emulator is not something included in the application note or any reference design, but without it the HDMI receiver above will not gain any data. It passes the EDID which is basically a data structure standardized by the video electronics standard association (VESA) to the source via DDC. It contains information such as the manufacturer name, the product ID, the supported timings, the display size and the luminance characteristics. Usually the EDID data is stored in the display sys-

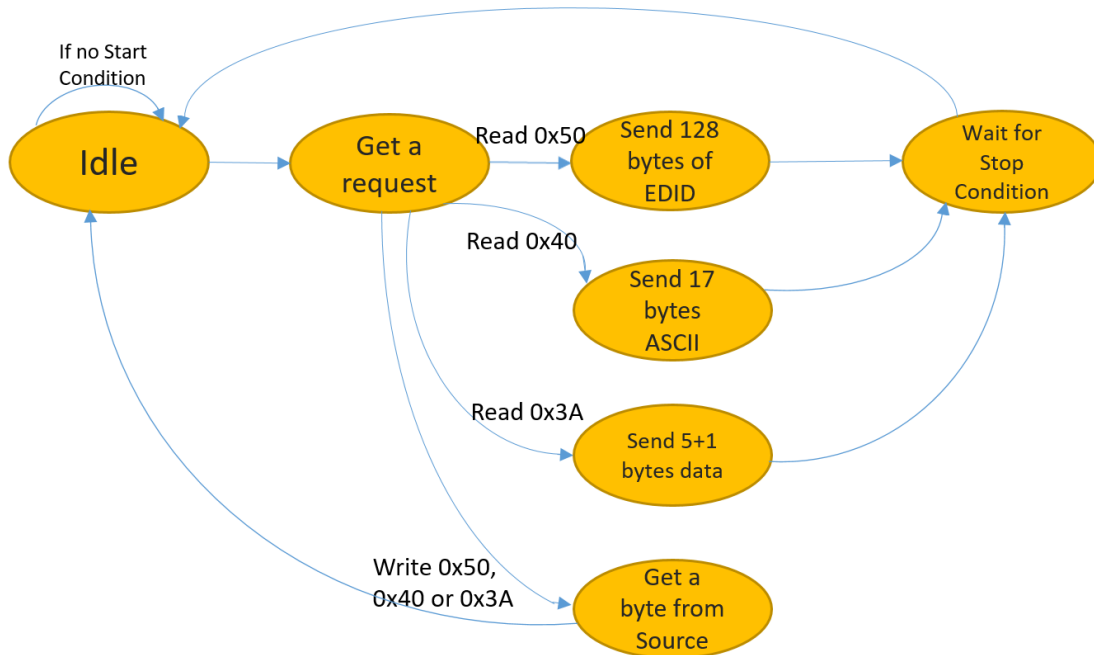


Figure 4.3: The flowchart of the EDID handshake

tem using a serial programmable read only memory (PROM) or a serial electrically erasable PROM (EEPROM). Three wires in the HDMI cable form the DDC channel, SCL, SDA and DDC/CEC Ground. These wires shall support the communication protocols defined in the I2C-bus Specification. Figure 4.2 shows the process of EDID handshake that occurs at the initial stage upon an HDMI connection.

In practice, we found that the establishment of the EDID communication is not that straightforward. Several tricks are used to make the handshake robust as we recorded using a logic analyzer. On plugging in the HDMI cable, test signals are sent from the HDMI source to ensure that the sink does not response to invalid format of I2C signals. And then unknown commands `0x40`, `0x3A` would be issued from the source, certain contents should be replied by the sink as shown in Fig. 4.3. By means of reverse engineering the initial handshake data between the PC and a DELL

2408WFP monitor, replacing the EDID with customized data, we managed to make an emulated HDMI sink device recognized by the source.

4.2 Synchronization Scheme for SAR

The synchronization scheme defines how the system switch between the SAR mode and SLI mode, as well as how the system triggers the camera. A general purpose I/O pin of the camera which is denoted as Line 2 or ENABLE as illustrated in Chapter 2 Fig. 2.3 is used to select which video source is sent to the HDMI transmitter. The Line 2 is a user defined pin that can be programmed with software running on PC, when Line 2 is set to 1, the SLI pattern sets generated inside FPGA are transmitted to the projector; when Line 2 is set to 0, whatever video from PC is transmitted to the projector.

As the HDMI allows both positive and negative polarities of the vertical sync signal as illustrated in Fig. 4.4, and only the rising edge of the vertical sync with a positive polarity should trigger the camera, so a dedicated logic module that is used to identify the polarity of the incoming video from PC is instantiated. The module includes two counters, one increments at every clock cycle when the vertical sync is 0, the other increments at every clock cycle when the vertical sync is 1. By comparing the maximum value of these two counters, the polarity of vertical sync is determined. If it is negative polarity, the falling edge which indicates the beginning of a frame should be sent as the trigger to the camera.

The triggering scheme in SAR mode is quite different from it is in SLI mode in that the instantaneous output of a PC's graphics card is unpredictable [132]. Fig-

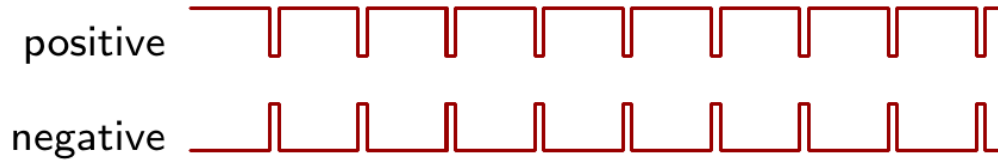


Figure 4.4: The distinct duty cycles of positive and negative polarities

Figure 4.5 illustrates two typical scenarios that take place in PC's graphics card. On top, the demanded video sequence ABCD does not arrive at the output port after issuing the command from PC, until an unknown number of do not care (dc) frames have been transmitted. At the bottom, the frame B is replicated and sent twice, it can randomly happen to any frames in the sequence. In both cases, there is no way to effectively capture the useful video sequence for the camera, a much larger number of frames than necessary need to be captured to ensure that the entire sequence is included.

To solve this issue, we devise an intuitive approach in which we encode a unique index byte in red channel of the top left pixel without changing the overall looking of the frame. As there is a frame buffer which is used to store the current incoming frame and later project it at next frame, by reading the 8-bit pixel value in red channel for the frame being received in the FPGA, a decision can be made on whether or not the next frame to be projected is the target, so as to instruct the FPGA if a camera trigger should be sent. In Fig. 4.6, it is a screenshot of an oscilloscope that monitors the camera triggers issued while the system is operating in SAR mode. In this case, the PC is given a command to output an 8-pattern SLI sequence, they are supposed

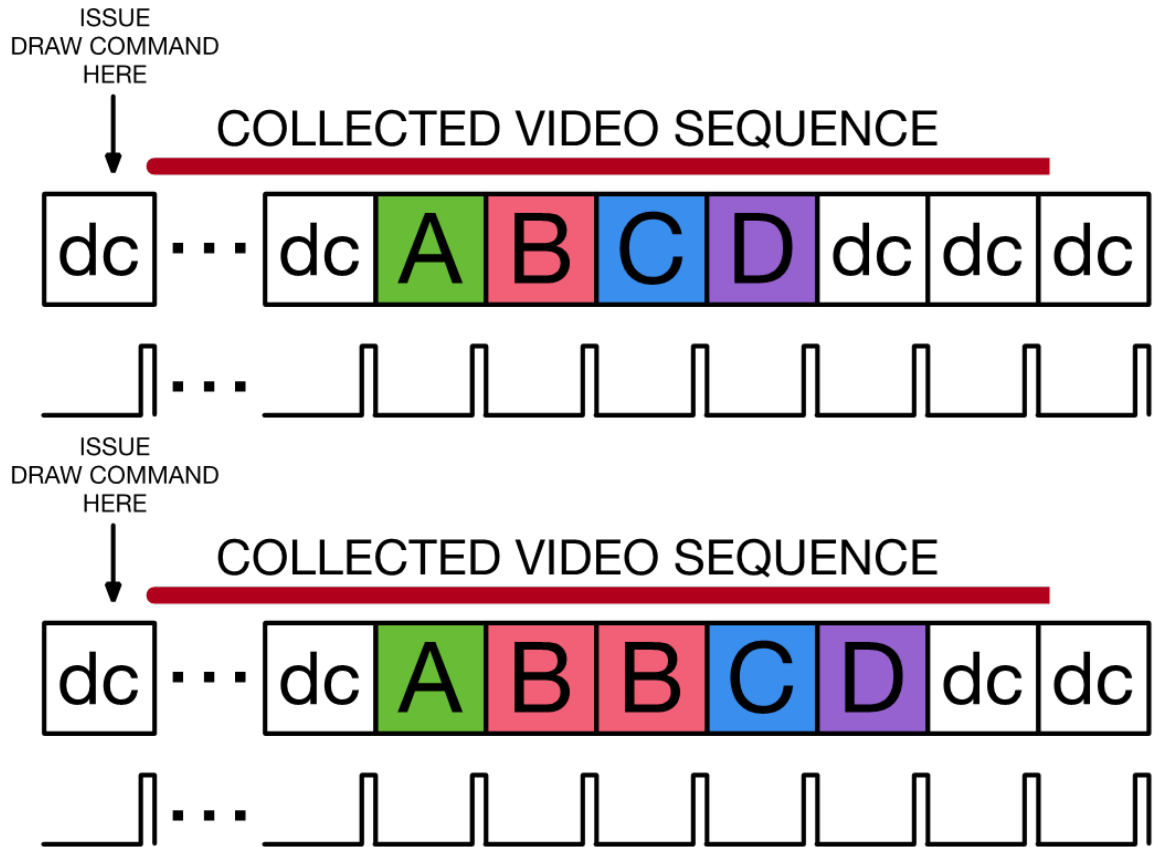


Figure 4.5: Uncertain delay (top) and replicates of frames (bottom) happen in graphics card

to appear at the HDMI port of the graphics card one by one immediately, but the actual triggers to the camera shown on the oscilloscope in the color of yellow are 8 pulses unevenly spread. Again this proves what is illustrated in Fig. 4.5.

The idea of finding the top left pixel involves a deep understanding of how HDMI works. The HDMI bus does not transfer video data all the time, in fact, there are some windows. The video data is transferred only during these windows. Take a video of 800×600 resolution for example, each frame is divided into 600 rows, each row has 800 pixels. All of these 800 pixels are considered a group, each group is operated on as an entity. In other words, any given 800 pixels in a row are transferred without any

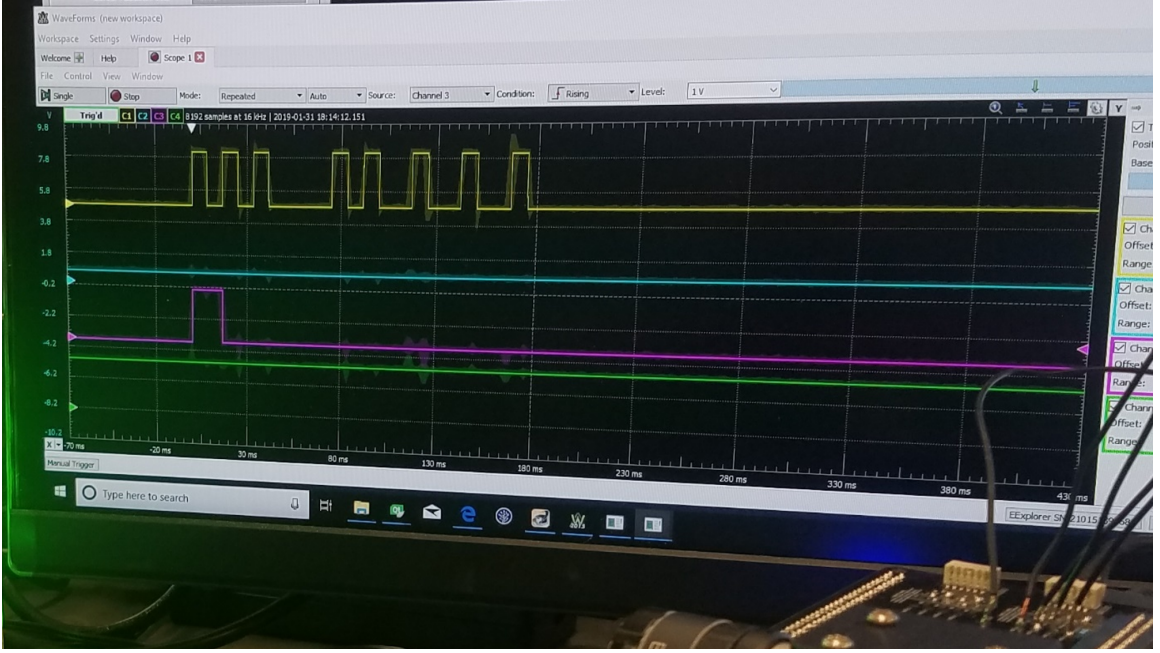


Figure 4.6: Camera triggers (signal in yellow) recorded for a sequence of 8 SLI patterns in SAR mode

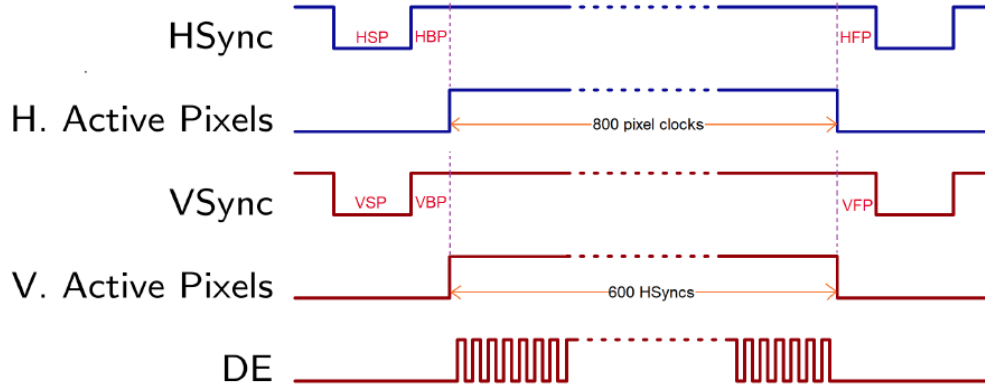


Figure 4.7: The internal timing of HDMI at a resolution of 800×600

interruptions. We call this period horizontal active pixel time. In between any two successive rows, there are intervals during which no video data is being sent. This blank period includes three distinct phases, they are measured in the unit of pixel clock cycles and are usually referred to as horizontal front porch (HFP), horizontal sync pulse (HSP) and horizontal back porch (HBP). Similarly, the HDMI standard

organizes a frame of video in a certain number of rows, which in this case is 600. It treats 600 rows as an entity and between every two successive frames, there is an interval consisting of vertical front porch (VFP), vertical sync pulse (VSP) and vertical back porch (VBP). Instead of pixel clock, they are measured in the unit of row cycle. Figure 4.7 illustrates the principle of the above HDMI timing convention.

Although *HSync*, *VSync* and *DE* are not instantiated in real wires in a standard HDMI bus, by deserializing and decoding the data stream in blue channel of the HDMI bus, they can be extracted from the blue channel. The *DE* signal is a series of windows where each windows represents the transferal of a row of active pixels, it can be expressed as,

$$DE = H.ActivePixels \ \& \ V.ActivePixels, \quad (4.1)$$

the first rising edge of the *DE* after the rising edge of *V.ActivePixels* indicates the beginning of the first row, and the first pixel clock after the beginning of the first row points to the first pixel in a frame. Hence, the target pixel is found.

4.3 Experiments

The face paint system is built with an TI DLP3010EVM-LC projector and a Basler acA1920-150 μ c camera, as shown in Fig. 4.8. During calibration, the correspondence between the camera and projector coordinate systems are established using SLI, then the system runs in a interlaced manner, where a blank frame is sent to the projector for every 7 consecutive video frames from PC. In other words, a blank frame is inserted to replace the last frame in every 8 consecutive frames. It is during this blank frame

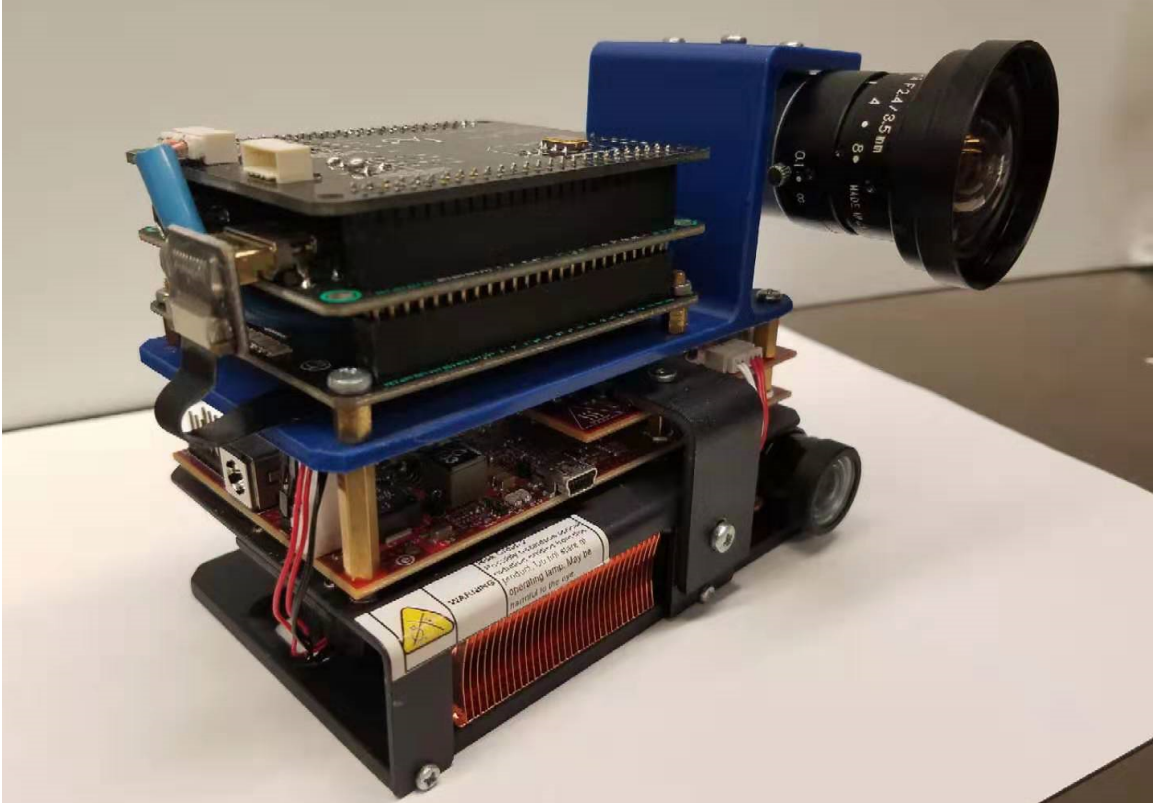


Figure 4.8: The hardware platform of face paint system

that a image is taken by the camera. An open source library for facial recognition is adopted to find out the face in the image taken, after processing the image, the 2D coordinates of the face is derived. Consequently the augmented painting can be projected at the corresponding location in the projector coordinate system. Figure 4.9 illustrates that the augmented red dots (left) and dark hair (right) are both aligned with the human face printed on a wall.

The position tracking system consists of two cameras, one projector and our FPGA controller, among which one camera and the projector form a conventional SLI scanner and serve as the base station, the other camera is mobile and represents the sensor mounted on the headset. In order to differentiate the two cameras, we denote the camera paired with the projector as the master camera and the mobile camera as the

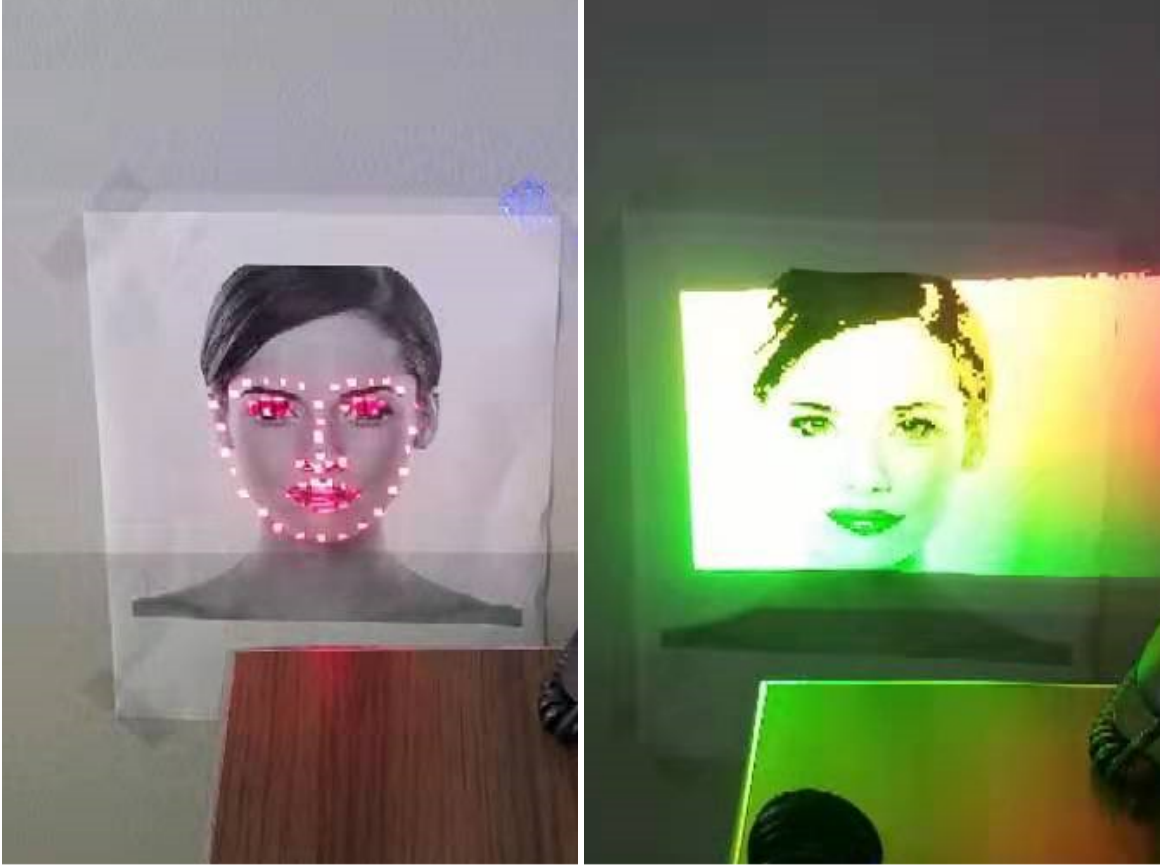


Figure 4.9: Face paint with red dots (left) and the same face painted with dark hair (right)

slave camera henceforth. There is a unique feature of the SLI scanner in the system which is its composite pattern design. Unlike the common SLI pattern sets that are either horizontal or vertical fringes, the FPGA controller in the position tracking system generates composite fringe patterns that are the combination of both the horizontal and vertical fringes as shown in Fig. 4.10. Because in the conventional SLI setup, the positions of camera and projector are deliberately lined up horizontally or vertically, so the projected patterns only need to contain clues leading to X or Y coordinate of the 3D space. Specifically, horizontal fringe patterns are used when the camera and the projector are placed in a perpendicular line; vertical fringe patterns

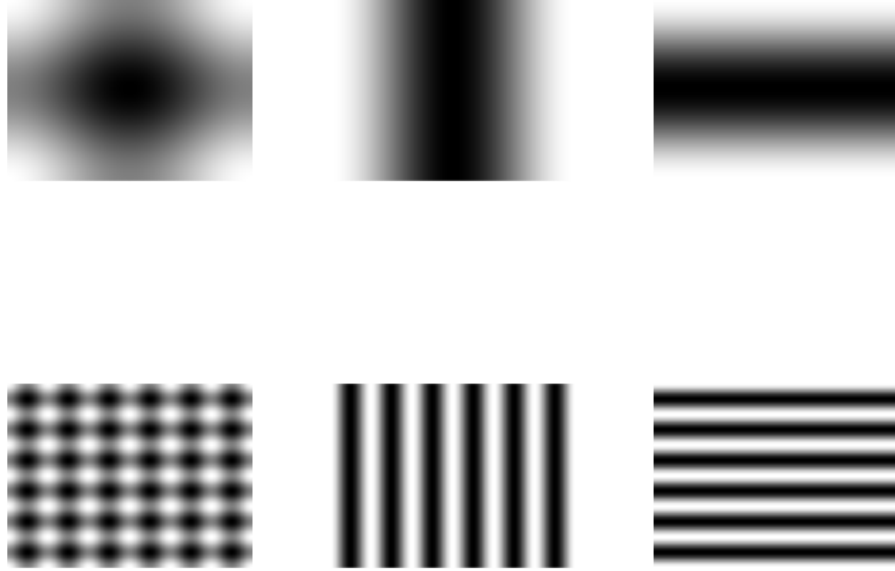


Figure 4.10: The unit composite pattern (top-left) with its horizontal/vertical fringe components (top-center/top-right) and the high frequency composite pattern (bottom-left) with its horizontal/vertical fringe components (bottom-center/bottom-right)

are used when the camera and projector are placed in a flat line. However, in this case the slave camera could be anywhere in the 3D space, both X and Y coordinates of the 3D space need to be found out in the process of 3D reconstruction from the slave camera.

The principle of obtaining the position and orientation data of the slave camera can be described in two steps. At first, the based station performs an SLI scan at a certain direction in the 3D space, so the mapping between the 2D coordinates of the master camera's image plane (X_m, Y_m) and 3D world coordinates (X_w, Y_w, Z_w) can be established. Moreover, for each pixel in the master camera, its two phases P_x, P_y

can be extracted by decomposing the composite pattern into their horizontal and vertical components. Meanwhile, the FPGA controller sends the identical triggers to the slave camera as it triggers the master camera. Assuming at least part of the projection falls into the slave camera's field of view, there are some pixels in the slave camera (X_c, Y_c) will hold a valid pair of P_x and P_y . Then (P_x, P_y) can be used as the index to link one of those subject pixels in the slave camera to its 3D world coordinate (X_w, Y_w, Z_w) . As long as there are 6 or more valid pixels can be extracted from the slave camera, the perspective matrix

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \quad (4.2)$$

is derived through this equation,

$$\begin{bmatrix} X_s \\ Y_s \\ 1 \end{bmatrix} = \mathbf{M} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4.3)$$

In practice, there are much more valid pixels in slave camera can find their matching (P_x, P_y) in master camera, so the least square solution is used to calculate the \mathbf{M} . Figure 4.11 illustrates the process described above. Once the perspective matrix is determined, the rotation matrix \mathbf{R} and translation matrix \mathbf{T} can be derived by matrix decomposition according to Li [146]. It is the \mathbf{R} and \mathbf{T} that eventually tell the orientation and position of the slave camera in 3D space.

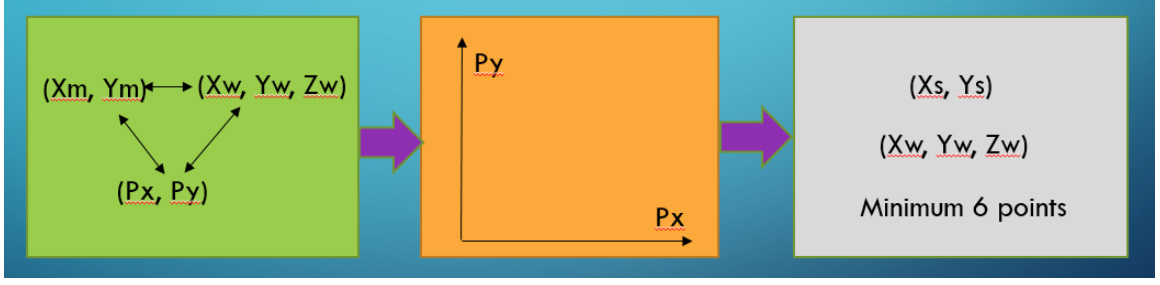


Figure 4.11: The mapping between the master camera and the slave camera

Noticing that the during the whole process the slave camera is not required to see the base station, our position tracking scheme can be employed in applications where there is no direct line of sight between the base station and the headset. Figure 4.12 shows our experimental setup where the master camera and the projector are fixed in a perpendicular line and the slave camera are placed at five different positions along the horizontal rail pointed by the arrow.

Figure 4.13 presents the experimental results of the slave camera's position tracking, the figures on the left and at the center are the same 3D reconstruction of the projected space at different view angle, in which the red circle represents the master camera and the blue circles represent the 5 different positions the slave camera has been placed. In both figures, the 5 blue circles are roughly located in a straight line which are consistent with the experimental setup. The figure on the right picks one of the 5 slave camera positions, replaces the circle with a special camera icon. It not only illustrates the position but also the orientation of the camera.

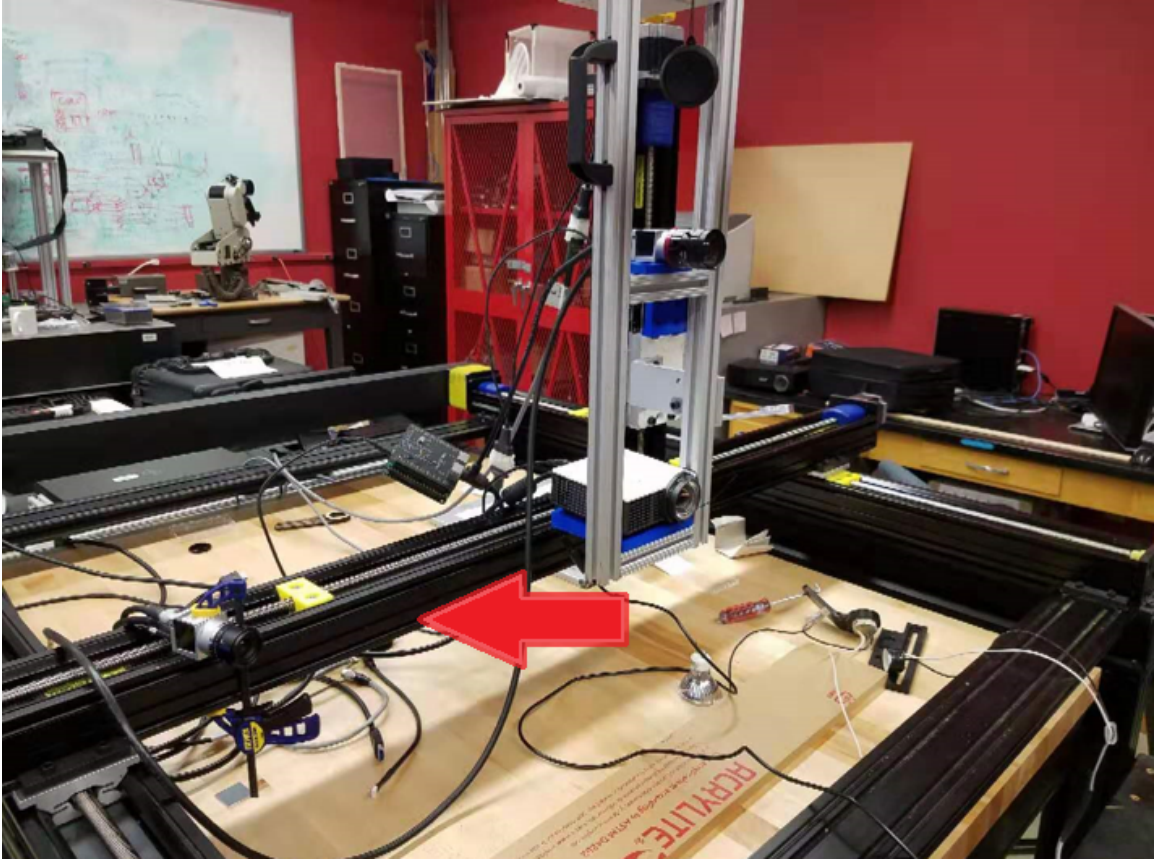


Figure 4.12: The experimental setup of the position tracking

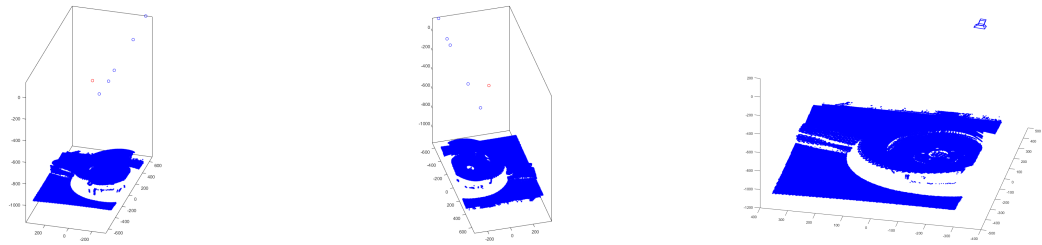


Figure 4.13: The 3D reconstruction of the scanned scene including the positions of both master and slave cameras (left and center), 3D reconstruction of the scanned scene just recording one slave camera position and orientation(right)

4.4 Summary

In this Chapter, we demonstrated an FPGA-based multi-functional SAR/SLI controller and its practical application in face paint and position tracking. The face

paint system operated at 120Hz and constantly took an image every 0.067 second for the software tool installed on PC to track the location of a human face if there is any. Meanwhile, virtual dots or figures were projected onto the face. The experimental results turned out to be good in terms of finding the correct location of the face in the projector coordinate system, but as the location of the face changed in the projected scene, noticeable delay would be observed before the projector finds the updated location of the face. A possible solution is to add another depth camera which runs much faster to collect the location data in real time. The 3D position tracking system provided a whole new idea for VR/AR headset tracking with SLI, it supported position tracking without direct line of sight between the base station and the headset, and the orientation of the headset can be derived as well.

Chapter 5 Conclusions and Future Work

This dissertation documents the evolution of an FPGA-based HDMI input-output module for adding structured light scanner to a spatial AR table. The proposed controller specifically addresses the issues associated with using a commodity digital projector for structured light scanning including properly synchronizing the camera's exposure to the on-off-on timing of the LEDs associated with producing the illusion of continuous shades of gray with a binary digital mirror device. As detailed in Chapter 2, we introduced a novel dual-projector scanner that simultaneously fed two projectors with their own PMP patterns. Because commodity projectors assume that input digital images are encoded with power-law parameter gamma due to the fact that human eyes perceive the light in a non-linear manner, the SLI patterns generated inside our FPGA are pre-calibrate to produce linear responses in the projectors.

In addition to generating SLI video, a robust synchronization scheme was proposed to coordinate the timing between the camera and projectors. Three actions including triggering the camera, updating the pattern, and ending the camera exposure should be executed in precise order and timing. Each one of these actions is closely related to and impacts the other two. In the end, a special feature of DLP projectors and its influence in SLI scanning were discussed, subsequently a solution was provided to guarantee reliable 3D scans. And given how DLP projectors reproduce RGB video, the camera exposure needs to be carefully configured so that every exposure coincides with specific time slots of the projected video frames. The results of the

related experiments proved that the dual-projector setup improves the SNR without any interference between two projectors and minimize the influences of occlusion and multi-path.

In Chapter 3, we looked into the issue of multi-path as this is a major issue that has received very little attention in the SLI literature. Multi-path is the occurrence where a pixel of the image sensor captures light reflecting from two or more different paths due to the complex geometry of the object surface [111]. Specifically, bimodal multi-path is the case of exactly two paths most commonly associated with a discontinuities in depth where a pixel captures a portions of a fore and background surface. By modifying the dual-projector SLI scanner described in Chapter 2, we introduced a dual-projector SLI scanner customized for multi-path analysis. The reliability of our FPGA controller was thoroughly tested for this purpose and proved to be satisfactory as 488 SLI patterns that contains 61 distinct spatial frequencies were scanned without any dropping or duplication of frames. Experiments were conducted with particular setups for multi-path with results verifying our analysis and expectation.

In Chapter 4, we focused on how the HDMI video passes from PC to the projector through our FPGA controller in order to unlock the capability of our system for SAR applications. The HDMI does not automatically start the transmission of video once a cable is connected between the source and sink as it requires an initial handshake to let the PC source know what kind of display device is connected so as to issue the video with the appropriate format. As the sink device in the system, the FPGA is not an actual display device like a PC monitor or projector. So an EDID emulator was implemented to reply the request from the source and then fulfill the handshake

such that the PC sees it like a real display device.

In addition to emulating the EDID, we proposed a top-left pixel encoding scheme that allows us to implement a GPU to FPGA handshake that enables our FPGA to identify specific targeted frames coming from the PC among many unpredictable unrelated frames, triggering the camera at the right moment and skipping any duplicates. So a user can implement their own structured light encoding scheme knowing that the camera will capture each and every frame in the sequence once and only once. We even used a second GPO bit from FPGA to camera to indicate when the top-left pixel is equal to zero, which a programmer can use to further ensure that the captured images correspond to specific frames in the SLI sequence. For SAR applications, this handshake also gives the programmer the ability to use the camera for traditional human-computer interaction with the AR environment.

To demonstrate the SAR benefits provided by our FPGA controller, two practical SAR applications were demonstrated. The first was an SAR face paint system that uses a machine vision camera and a DLP projector to project digital masks onto a human subject. Specifically, the camera and projector were positioned in front of a human face with SLI mode used to scan the face at startup and derive a camera pixel to projector pixel mapping. And then in SAR mode, the FPGA triggers the camera on every 8th frame while simultaneously turning off the pixels of pass through video. This means that, the camera captures a photo of the human subject's face while not projected any SAR image. This photo is then processed using an open source facial feature detector with the detected features used to digitally warp a supplied mask image, projecting it onto the human subject's face in perfect registration.

As a second demonstration of our HDMI controller for SAR, we implemented a novel camera tracking system for AR headset tracking without a direct line of sight between headset and base station. Specifically, a camera-projector pair with fixed position serves as the base station, performing traditional SLI scanning where the mapping between the camera’s 2D image plane and 3D world coordinates can be obtained. While the base station is scanning, a second camera, serving as a mobile headset, takes images of the projected patterns concurrently. Therefore, both cameras have recorded the 2D phases information that is encoded in the SLI patterns. Subsequently, the 2D phases from the second camera are compared with the phases from the base station camera, pixel by pixel. If there is a match, a correspondence between the second camera’s pixel and its 3D world coordinate from the first is found. With a minimum of 6 such pixels, the perspective matrix of the second camera can be derived from which the position and orientation of the second camera are extracted. Thanks to its unique setup, our collaborative tracking system opens up the possibility for simultaneous multiple AR headsets tracking so that two or more headsets can display AR information over the scene in perfect registration across headsets.

In conclusion, three FPGA-based hardware platforms and their individual applications in SLI and/or SAR have been presented in this dissertation. We believe this work will benefit the whole SLI/SAR community in that it provides great potential for researchers to quickly build up their own 3D scanner or SAR platform so they can focus more on designing SLI patterns or devising related new algorithms, etc. We specifically describe a path for using a commodity projector in place of an industrial light engine, achieving a maximum HDMI frame rate of 120 fps without

gamma distortion and with accurate synchronization between camera and projector. Furthermore, our 2D pattern sets will enable high speed scanning for SAR table systems, unlocking 3D resolutions an order of magnitude greater than RGB+D sensors, albeit with a reduction in temporal frequency. Applications such as the cadaver table can now be assembled that fuse the two imaging systems for both high spatial and temporal resolution.

Given the fact that our FPGA-based controller has made it possible to easily assemble an SLI 3D scanner or SAR system with commodity projectors and off the shelf cameras, our future work includes integrating the HDMI controller aspects of the FPGA into the FPGA controller of the machine vision camera. An added incentive to develop such a customized camera are its capability of high dynamic range (HDR) imaging and its flexibility for projector synchronization. Up to now, we have made a working camera that meets our requirement, but the PCB prototype, as shown in Fig. 5.1 (left), that connects the camera and the FPGA controller still has some defects. So if a perfect PCB can be successfully designed and made in the future, we will be able to build a highly integrated SLI/AR system that no one else has developed before.

Unlike the previously introduced SLI/SAR systems that are based on conventional FPGAs, this customized camera system is built with a Zynq-7010 system on a chip (SoC) from Xilinx. The Zynq-7000 series SoCs are not just FPGAs but complete embedded systems that contain an ARM Cortex-A9 MPCore-processor and a 7 series FPGA device [147]. With the powerful computational capabilities of the ARM processor, the Zynq-7000 is able to fulfill high-end application oriented tasks.

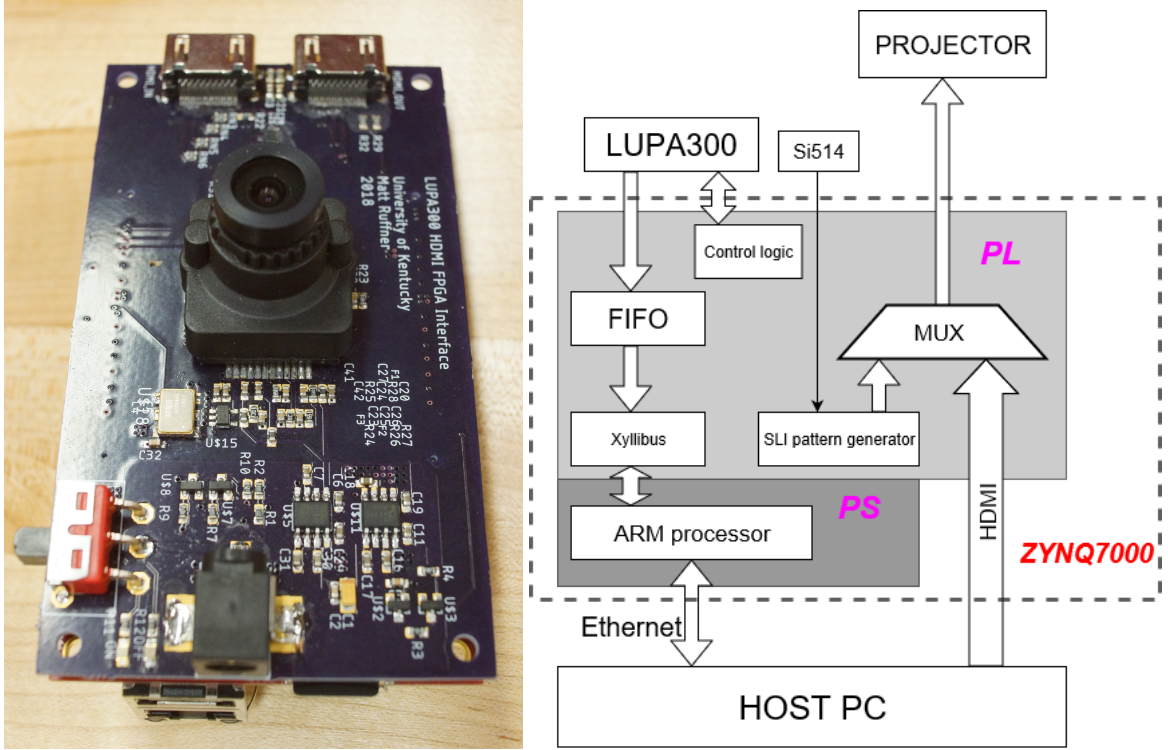


Figure 5.1: The latest version of PCB that has been made (left) and the system diagram of the camera (right)

Specifically, an Avnet MicroZed system on a module (SoM) featuring a Zynq-7010 is adopted in our design, it provides a software interface for collecting and processing image data as well as receiving, generating and transmitting HDMI video for SLI/SAR. An Ethernet port on the MicroZed enables high speed transfer of video frames back to the host PC.

Figure 5.1 (right) shows the overall system diagram that includes everything except the projector and the host PC. The notation PS stands for processing system which is the dual-core ARM processor while the notation PL represents the programmable logic which is essentially an FPGA, almost every module that has been used in previous SLI/AR systems is instantiated here, including an SLI pattern generator, an HDMI transceiver that passes the video stream from host PC to projector,

an external programmable crystal Si514 that is used to generate clocks at some particular frequencies for HDMI and a modified control logic synchronizing the CMOS image sensor and the projector. What's more, there is a 16-bit FIFO bridges the image data flow between the image sensor and the ARM processor, because they belong to different clock domains.

Since the ARM processor runs the Linux OS, middleware running on the ARM cores works as an interface with the ARM core's OS can see the incoming image data flow as a standard I/O file and manipulate it with some common I/O functions. The middleware we picked is the Xillibus [148], it provides C code examples of reading from the Linux device file FIFO that accompany their Verilog IP core. Similarly, the video server C++ code running on the MicroZed's A9 processor reads image pixel values at a very high speed from the device file, so as not to miss any, and then sends them over a network socket via Ethernet. Once this network socket has been published, a remote host may open the stream and view the frames of video read from the image sensor. In this way, remote SLI scans and/or SAR projections can be controlled from a remote location.

The chosen image sensor for the system is the LUPA300 from ON Semiconductor [149]. Interfacing components of the system to the MicroZed is achieved with a custom carrier card (CCC) PCB design as shown in Fig. 5.2. This design is responsible for interfacing the LUPA300 to the MicroZed, which requires custom power supply circuitry and start-up sequencing. The circuitry to realize this comprises the Power Delivery Network (PDN) of the system. Designing a robust, effective PDN requires adequate component selection and informed PCB routing techniques in order

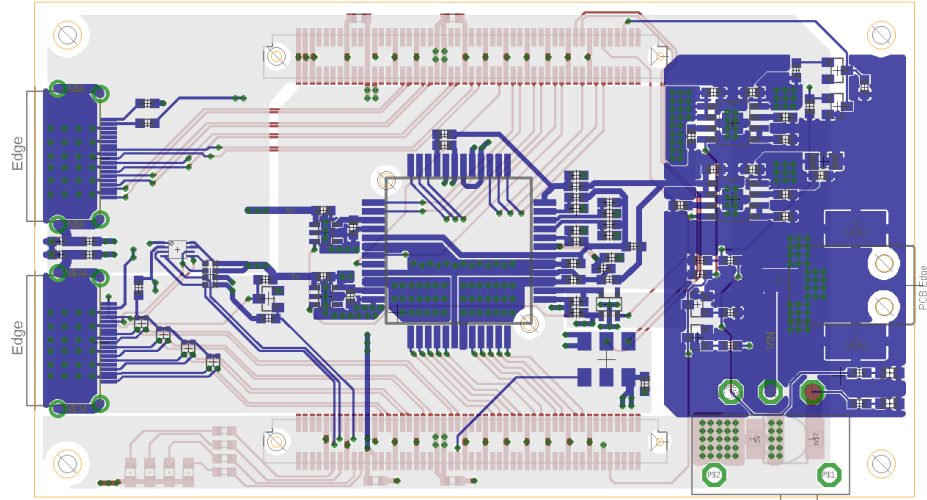


Figure 5.2: The PCB layout of the CCC showing high speed traces (red) and PDN (blue). Power pours on one internal plane are shown in light gray, while the all-over ground pour on the other internal plane is not shown

to avoid image sensor data corruption due to interference.

Providing a noise free power supply for sensitive devices such as image sensors, or other mixed signal and analog devices, is critical for correct device operation. Noise introduced to supply lines for technology such as analog to digital converters (ADC) can directly induce error on the conversion result. In the case of the high-speed ADCs used in the LUPA300 sensor, this power supply noise manifests itself as sinusoidal stripes in the final image, presumably due to a bias in the ADC circuitry. Filtering supply inputs to sensitive components, as well as the output of regulators, helps reduce noise on the power supply. Utilizing separate regulators is another technique, especially when the voltage rail has high instantaneous current usage. Use of ferrite beads helps ensure high frequency noise immunity of the system, and is used for isolating sensitive analog supplies from the shared voltage rail.

In an effort to create a PDN with low output ripple and high noise immunity,

regulators were sourced that featured internal noise suppression and filtering. The regulators chosen also require minimal external passive components and the ability to supply sufficient current for the system. These regulators are 2A ultra-low noise regulators from Analog Devices [150]. The high current rating allows them to be a common supply line for the 2.5V supplies of the LUPA300 sensor. These regulators are supplied from the 5V input to the CCC PCB. This 5V input is from a wall adapter and is first filtered with an inductor capacitor low pass circuit, as well as several decoupling capacitors, before being supplied to the input of the 2A regulators.

Bypass capacitors eliminate ripple in power supply lines due to fast switching DC signals with high di/dt . Minimizing the distance these high speed switching currents have to travel before reaching the capacitor is also beneficial to the noise reduction process. It is best practice to place the bypass capacitors as close as possible to the power connection on the IC of interest. Placing multiple capacitors in parallel across the supplies of high speed devices is also a common practice, as the smaller a capacitor value is, the higher frequency noise it will be effective at conducting. Bypass capacitors on the power rails help to mitigate these effects but cannot suppress them entirely. Ferrite beads are effective for filtering high frequency power supply noise and cleanly sharing similar voltage supply rails while preserving high frequency isolation. The component acts like a resistor, which impedes high frequency noise and dissipates it as heat [151].

When designing a circuit board for high speed applications, the introduction of multiple layers for power and ground planes is an industry accepted method [152, 153]. In applications with advanced complexity and very high routing density, such as PC

motherboard designs, stack-ups can be seen with 6, 8 and 12 layers. For the purpose of this design, a four layer board stack-up was chosen, with a solid ground plane directly under the high speed traces of the HDMI signals. This type of design has multiple advantages over a 2 layer configuration including the availability of extra copper for dedicated power and ground planes. A solid ground plane is not only helpful for optimizing return current paths of high speed digital signals [154], it is also essential to minimizing radiated noise power from these signals by means of microstrip routing techniques [152].

The LUPA300 was chosen due to its ability to operate at greater than the 120Hz (the rate of 3D ready projectors) and its non-destructive readout (NDR) capability. Using an image sensor which supports NDR enables HDR imaging applications [155]. The HDR imaging for SLI has been extensively studied [156, 157, 158], because the SLI 3D measurement in the situations where objects have a wide variations of surface reflectivity becomes quite difficult. Special treatment is necessary for this kind of scenarios, especially when phases are major depth cue that is used to reconstruct the 3D surface. Lin *et al.* proposed a technique called adaptive digital fringe projection that adjust the pixel-wise intensity of the projected fringe patterns [159]. Cai *et al.* applied light field imaging technology to implement HDR, where a special camera that is able to record both the position and direction of the light is employed [160].

Zhang *et al.* [161] presented a technique that takes multiple shots of fringe images and each of them has unique exposure. Considering dark fringe images have good quality in bright area while bright fringe images show good performance in dark area, a perfect image with good overall quality can then be obtained by synthesising multi-

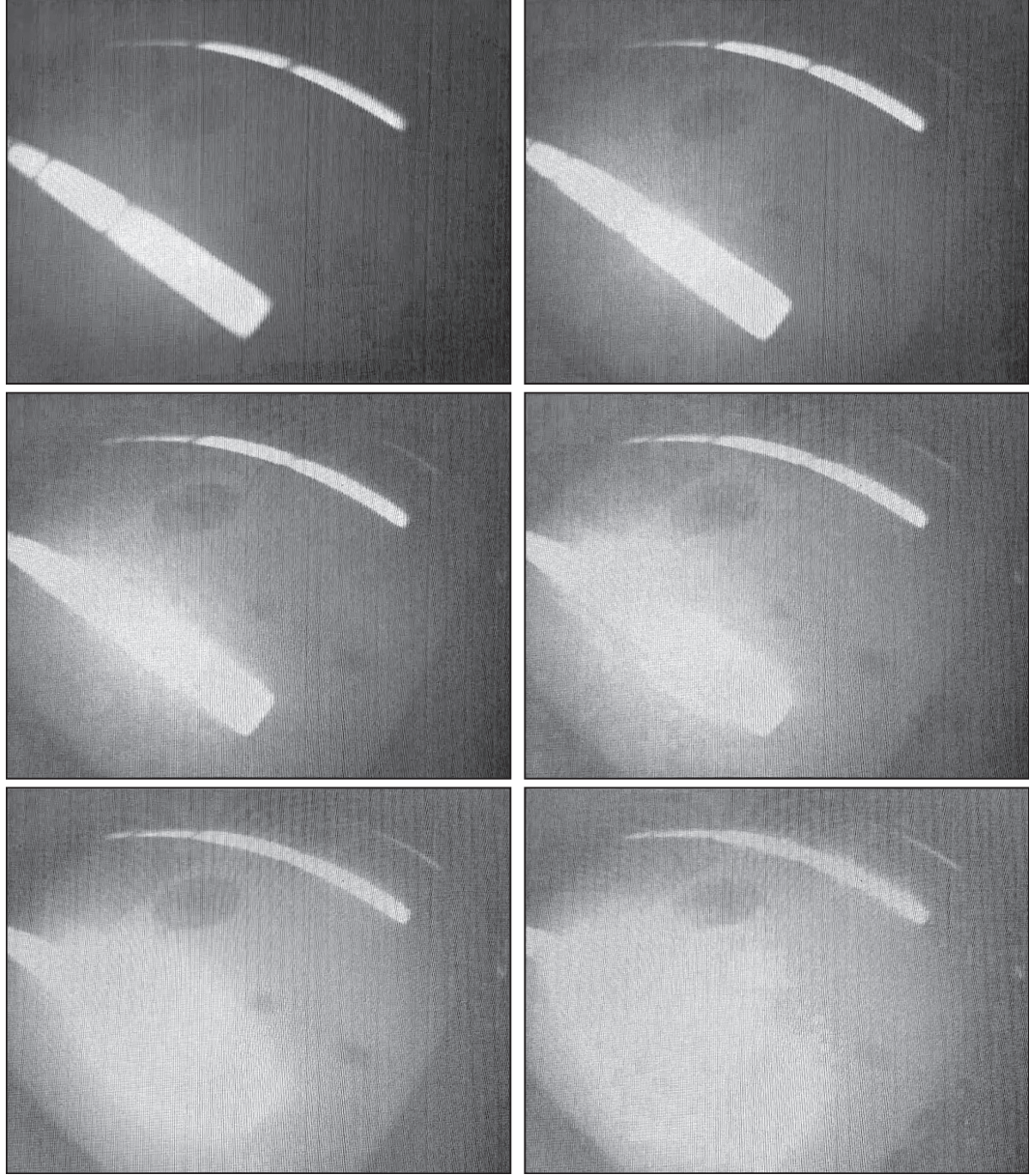


Figure 5.3: A sequence of consecutive six images taken when the non-de

ple images with different exposures. However, Zhang *et al.* do not have any dedicated synchronization scheme, they have no other way except to manually switch the fringe patterns and the 23 different camera exposures from PC, which is so inconvenient and time consuming. With our proposed customized camera based on LUPA300, the predicament can be solved.

To understand the NDR, the fundamentals of how an CMOS image sensor works need to be introduced. Every pixel in a sensor has the identical initial voltage before exposure, for those pixels receive light of stronger intensity, their voltages would drop faster, and there is a threshold which is 0. Once the pixel voltage drops to 0, it will not go any further even if the exposure has not completed. Consequently, if two pixels are both read a voltage of 0, the actual light intensities at them can not be determined. Apparently this is not a desirable situation, the users always want the obtained pixel voltage to be below the initial value and above 0. Therefore, shorter exposure is required for those pixels that receive light of strong intensity, and longer exposure is preferred for those receiving light of weak intensity. Normally, an image sensor refreshes the voltage of every pixel after each exposure. But non-destructive readout does not do this, it only refresh the voltages after an exposure if the NDR bit in a special register is deasserted. Figure 5.3 shows a sequence of six consecutive images that are taken from our system when the NDR bit is asserted. The exposures of them range from 4 ms to 20 ms with a step of 4 ms.

Retrieving image data from the LUPA300 image sensor is a joint effort of FGPA and PC programming. After powering on, the PL delays until the PS boots, after which an 80Mhz clock from the PL is fed to the LUPA300. After this, the PL enables the voltage regulators to the LUPA300. This sequence of clock before power is necessary to reliably enable the image sensor. The CCC PCB provides the connections necessary for the FPGA software to configure and read data from the image sensor. Image data is then written to a Xillybus IP core. The underlying physical connections between the ARM processor and FPGA on the MicroZed in combination with kernel

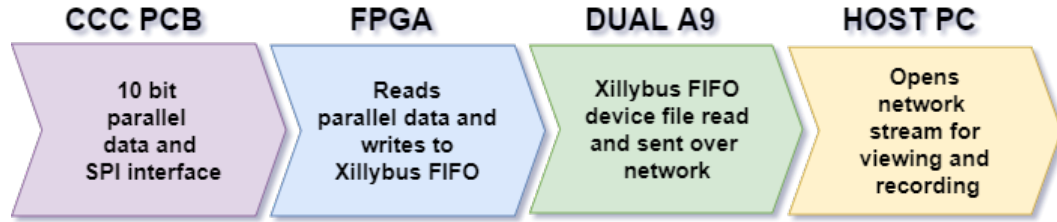


Figure 5.4: Data from the image sensor to host PC.

drivers on the ARM side, enable the FIFO to be read as if it were a Linux device file.

This process is visualized in Fig. 5.4.

The synchronization module in the PL of the FPGA synchronizes the LUPA300 with a vertical sync signal either from the PC or the SLI pattern generator. To start a scan, the PC issues a command to the PS via Ethernet. The command consists of two parameters: exposure time and trigger delay. Both are sent to the PL via the Xillybus. The PL then begins to trigger LUPA300 exposures according to these parameters.

The CCC PCB also provides the physical connections between the LUPA300 image sensor's serial peripheral interface (SPI) port and the FPGA, so that a register value can be configured within the sensor. SPI register upload can be done at system initialization or during operation of the sensor to change programmable gain and offset on the fly. There are a total of 16 registers in the sensor, and configuration values are composed of 12 bits. Each SPI transaction is made up of 2 bytes, with the 4 most significant bits of the first byte sent representing the sensor register address. The upload process as part of system initialization consists of first configuring the image sensor over the SPI bus and then reading image data over the parallel data output. Another Xillybus FIFO facilitates this register configuration, having data written to

it by the video server application running on the MicroZed, and the FPGA reading FIFO data.

Aside from an intuitive firmware level interface, the Xillibus also provide 32 general purpose registers which can be accessed from both the PS and PL sides. The values of them are read by the PL constantly and compared to a previous copy of the register values to determine whether or not it should re-upload the new values to the internal registers of LUPA300. In this way, we managed to change the parameters like exposure, subwindow setting, clock granularity, etc. on the fly. As stated earlier, one of the advantages of the customized camera over an off the shelf camera is its better potentials in synchronization with projectors.

Common off-the-shelf cameras take advantage of frame trigger ready for synchronization, the signal indicates the instantaneous status of the camera, when the camera is exposing, it is low, otherwise it is high. Because the exposure is configured in advance by the software for these cameras. So the FPGA has no idea what the exposure is for the next frame, it can only monitor the status of the frame trigger ready to measure the exposure of the current frame. In other words, the FPGA is unable to predict the exposure of the incoming frame, which makes it hard for the FPGA to synchronize the projector especially when the exposure needs to be randomly varying over consecutive frames.

The LUPA300 addresses this problem by supplying an alternative slave mode. Technically, the image sensor has two working modes: master and slave. The master mode runs the image sensor with a fixed exposure and starts the frame only according to the system clock. Once the system clock is provided, the start and end of a frame

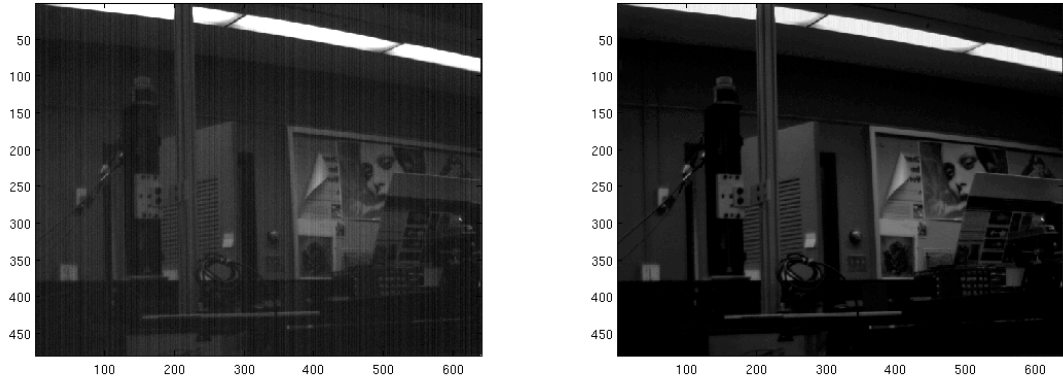


Figure 5.5: A sample image with fixed pattern noise(left) and the same image after the removal of fixed pattern noise(right)

is set. Obviously this is not the mode that can be synchronized with the projector as it is more like the free run mode in an off the shelf camera. In contrast, the slave mode uses an external input pin to receive signal that is able to control the start and end of a frame, as well as the exposure. So by sending a programmable control signal from the FPGA to that specific pin, the FPGA can predict the incoming frames, consequently a more flexible synchronization can be achieved.

Figure 5.5 (left) shows a 640×480 sample image taken in room 675 FPAT, there are noticeable regularly arranged fixed pattern noises (FPN) [162], which is unavoidable to nearly all the image sensors. By some simple processing afterwards, the FPN can be removed as in Fig. 5.5 (right). The NDR has been tested to be working as well, we can see the captured video repeatedly change from dark to bright then back and forth by refresh the pixel voltage once every four frames. Unfortunately, the CCC PCB has not made the HDMI transceiver work, it is so far more like a stand alone camera but not a highly integrated SLI/SAR system. We hope some efforts in the future will accomplish the exiting hardware platform for SLI and SAR.

Copyright© Ying Yu, 2019.

Bibliography

- [1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.
- [2] A. Li, “Augmented reality: what does it mean for ux,” 2016.
- [3] “How virtual reality positional tracking works.” <https://venturebeat.com/2019/05/05/how-virtual-reality-positional-tracking-works/>.
- [4] “AR projector system acts like a billiards coach.” <https://www.play-billiards.com/portfolio-item/ar-projector-system-acts-like-a-billiards-coach/>.
- [5] HDMI Licensing, LLC, *High-Definition Multimedia Interface Specification*, November 2009. Version 1.4.
- [6] B. Feng, “Implementing a tmds video interface in the spartan-6 fpga,” December 13, 2010.
- [7] Video Electronics Standards Association (VESA), *E-EDID standard*, February 2000. Rev. 1.
- [8] “Digi-key electronics.” <https://www.digikey.com/>.

- [9] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of CAVE," *Proceedings ACM SIGGRAPH Conference on Computer Graphics*, pp. 135–142, 1993.
- [10] N. Singh and S. Singh, "Virtual reality: a brief survey," *International Conference on Information Communication and Embedded Systems*, pp. 1–6, 2017.
- [11] M. Carrozzo and F. Lacquaniti, "Virtual reality: a tutorial," *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, vol. 109, no. 1, pp. 1–9, 1998.
- [12] Wikipedia, "Virtual reality," 2019.
- [13] G. C. Burdea and P. Coiffet, *Virtual Reality Technology*. Wiley-Interscience, 2003.
- [14] I. Sutherland, "A head-mounted three dimensional display," *Proc. Fall Joint Computer Conf.*, pp. 757–764, 1968.
- [15] R. T. Azuma, "A survey of augmented reality," *Presence-Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.
- [16] M. Hawkins, "Augmented reality used to enhance both pool and air hockey," 2011.
- [17] T. Sielhorst, M. Feuerstein, and N. Navab, "Advanced medical displays: a literature review of augmented reality," *Journal of Display Technology*, vol. 4, no. 4, 2008.

- [18] S. learning, “<https://smarterlearning.me/2011/11/10/augmented-reality-in-education/>,” 2011.
- [19] G. L. Calhoun, M. H. Draper, M. F. Abernathy, F. Delgado, and M. Patzek, “Synthetic vision system for improving unmanned aerial vehicle operator situation awareness,” *Proceedings of SPIE Enhanced and Synthetic Vision*, vol. 5802, pp. 219–230, 2005.
- [20] L. B. Rosenberg, “Virtual fixtures: perceptual tools for telerobotic manipulation,” *Proceedings of IEEE Virtual Reality Annual International Symposium*, pp. 76–82, 1993.
- [21] Wikipedia, “Foxtrax,” 2019.
- [22] Wikipedia, “Hawk-eye,” 2019.
- [23] stype.tv, “Nascar on fox sports with augmented reality,” 2018.
- [24] “How to use your car’s backup camera correctly: reverse with ease.” <https://drivinglife.net/how-to-use-your-cars-backup-camera/>.
- [25] “Oculus rift vs. /uppercaseHTC vive.” <https://www.digitaltrends.com/virtual-reality/oculus-rift-vs-htc-vive/>.
- [26] “Microsoft holoLens.” https://xinreality.com/wiki/Microsoft_HoloLens.
- [27] “Magic leap one teardown reveals 6-layer display and video projectors.” <https://blogs.windows.com/devices/2015/04/30/build-2015-a-closer-look-at-the-microsoft-hololens-hardware/>.

- [28] J. P. Rolland, Y. Baillot, and A. A. Goon, “A survey of tracking technology for virtual environments,” 2001.
- [29] “How to avoid motion sickness caused by VR headsets while gaming.” <https://www.windowscentral.com/how-avoid-motion-sickness-caused-by-vr-headsets-while-gaming>.
- [30] P. R. Desai, P. N. Desai, K. D. Ajmera, and K. Mehta, “A review paper on Oculus rift- a virtual reality headset,” *International Journal of Engineering Trends and Technology*, vol. 13, no. 4, pp. 175–179, 2014.
- [31] “SteamVR HTC vive in-depth lighthouse tracking system dissected and explored.” <https://pcper.com/2016/04/steamvr-htc-vive-in-depth-lighthouse-tracking-system-dissected-and-explored/>.
- [32] D. C. Niehorster, L. Li, and M. Lappe, “The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research,” *i-Perception*, vol. 8, no. 3, 2017.
- [33] Wikipedia, “Open source virtual reality,” 2019.
- [34] “Polhemus.” <https://polhemus.com/applications/electromagnetics/>.
- [35] “What are electromagnetic tracking systems?.” <https://3dcoil.grupopremo.com/blog/electromagnetic-tracking-systems-virtual-reality/>, 2017.
- [36] “Motion tracking case studies.” <https://polhemus.com/motion-tracking/case-studies/#fastrak>.

- [37] M. Ikits, J. D. Brederson, C. D. Hansen, and J. M. Hollerbach, “An improved calibration framework for electromagnetic tracking devices,” *Proceedings IEEE Virtual Reality*, pp. 63–70.
- [38] “HoloLens 2.” <https://www.microsoft.com/en-us/hololens>.
- [39] “Build 2015: a closer look at the Microsoft HoloLens hardware.” <https://blogs.windows.com/devices/2015/04/30/build-2015-a-closer-look-at-the-microsoft-hololens-hardware/>.
- [40] “Army uses modified Microsoft HoloLens 2 for real-life game of call of duty.” <https://www.digitaltrends.com/wearables/us-army-reveals-modified-microsoft-hololens-2/>.
- [41] “HoloLens on the battlefield: how the army uses AR.” <https://www.techspot.com/news/79542-hololens-battlefield-how-army-uses-ar.html>.
- [42] Wikipedia, “Magic leap,” 2019.
- [43] Wikipedia, “Google Glass,” 2019.
- [44] “The best augmented reality glasses 2019.” <https://www.wareable.com/ar/the-best-smartglasses-google-glass-and-the-rest>.
- [45] Wikipedia, “Smartglasses,” 2019.
- [46] C. Menk and R. Koch, “Truthful color reproduction in spatial augmented reality applications,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 2, pp. 236–248, 2013.

- [47] R. Raskar, K. Low, and G. Welch, “Shader lamps animating real objects with image-based illumination,” 2000.
- [48] “Augmented reality sandbox.” <https://arsandbox.ucdavis.edu/>.
- [49] A. Afrooz, H. Ballal, and C. Pettit, “Implementing augmented reality sandbox in geodesign, a future for geodesign,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-4, pp. 5–12, 2018.
- [50] T. L. Woods, S. Reed, S. Hsi, J. A. Woods, and M. R. Woods, “Pilot study using the augmented reality sandbox to teach topographic maps and surficial processes in introductory geology labs,” *Journal of Geoscience Education*, vol. 64, no. 3, pp. 199–214, 2016.
- [51] “Hardware.” <https://arsandbox.ucdavis.edu/instructions/hardware-2/>.
- [52] H. Uchiyama and H. Saito, “Ar supporting system for pool games using a camera-mounted handheld display,” *Advances in Human-Computer Interaction*, vol. 2008, 2008.
- [53] M. Greenspan, J. Lam, M. Godard, I. Zaidi, S. Jordan, W. Leckie, K. Anderson, and K. Dupuis, “Toward a competitive pool-playing robot,” *Computer*, vol. 41, no. 1, pp. 46–53, 2008.
- [54] “Puppy Cube.” <http://www.puppyrobot.com/en/puppycube>.

- [55] “Pond touchscreen projector.” <https://shop.touchjet.com/products/touch-pond>.
- [56] “Sony’s crazy projector that turns any surface into a touchscreen is going on sale this year.” <https://www.theverge.com/circuitbreaker/2017/2/27/14737824/sony-xperia-touch-projector-android-touchscreen-mwc-2017>.
- [57] “This projector light bulb turns your whole desk into a touchscreen.” <https://www.popularmechanics.com/technology/gadgets/a27124/desk-touchscreen-projector/>.
- [58] J. C. McLachlan, J. Bligh, P. Bradley, and J. Searle, “Teaching anatomy without cadavers,” *Medical Education*, vol. 38, no. 4, pp. 418–424.
- [59] “How VR is revolutionizing the way future doctors are learning about our bodies.” <https://www.ucsf.edu/news/2017/09/408301/how-vr-revolutionizing-way-future-doctors-are-learning-about-our-bodies>.
- [60] “Immersive and interactive virtual reality to improve learning and retention of neuroanatomy in medical students: a randomized controlled study.” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5878950/>.
- [61] “Virtual reality gives gross anatomy students new insight on the human body.” <http://www.buffalo.edu/ubnow/stories/2019/07/digital-anatomy.html>.

- [62] “Blow out your knee? Hope your surgeon’s got a VR headset.” <https://arstechnica.com/science/2019/08/blow-out-your-knee-hope-your-surgeons-got-a-vr-headset/>.
- [63] B. Panchaphongsaphak, R. Burgkart, and R. Riener, “Three-dimensional touch interface for medical education,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 3, pp. 251–263.
- [64] R. S. Lufner, A. C. Zumwalt, C. A. Romney, and T. M. Hoagland, “Effect of visual-spatial ability on medical student’ performance in a gross anatomy course,” *Anatomical Sciences Education*, vol. 5, no. 1, pp. 3–9.
- [65] J. Older, “Anatomy: a must for teaching the next generation,” *The Surgeon*, vol. 2, no. 2, pp. 79–90.
- [66] S. R. Williams, K. L. Thompson, A. J. Notebaert, and A. R. Sinning, “Prosection or dissection: which is best for teaching the anatomy of the hand and foot?,” *Anat. Sci. Educ.*, vol. 12, no. 2, pp. 173–180.
- [67] Wikipedia, “Minimally invasive procedure,” 2019.
- [68] L. T. D. Paolis and G. Aloisio, “Augmented reality in minimally invasive surgery,” *Advances in Biomedical Sensing, Measurement, Instrumentation and Systems*, pp. 305–320.
- [69] G. Indebetouw, “Profile measurement using projection of running fringes,” *Appl. Opt.*, vol. 17, 1978.

- [70] J. L. Posdamer and M. D. Altschuler, "Surface measurement by space-encoded projected beam systems," *Comput. Graph. Image Processing*, vol. 18, no. 1, pp. 1–17, 1982.
- [71] K. L. Boyer and A. C. Kak, "Color-encoded structured light for rapid active ranging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 1, pp. 14–28, 1987.
- [72] S. Siva Gorthi and P. Rastogi, "Fringe projection techniques: whither we are?," *Opt. Laser Eng.*, vol. 48, no. 2, pp. 133–140, 2010.
- [73] K. Liu, *Real-time 3-D reconstruction by means of structured light illumination*. PhD thesis, University of Kentucky, 2010.
- [74] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of the 7th Intl. Joint Conf. on Artificial Intelligence*, pp. 674–679, Aug. 1981.
- [75] J. Geng, "Structured-light 3d surface imaging: a tutorial," *IEEE Intelligent Transportation System Society*, 31 March 2011.
- [76] Y. Wang, *Novel approaches in structured light illumination*. PhD thesis, University of Kentucky, 2010.
- [77] R. J. Valkenburg and A. M. McIvor, "Accurate 3d measurement using a structured light system," *Image Vision Comput.*, vol. 16, no. 2, pp. 99–110, 1998.

- [78] I. Ishii, K. yAMAMOTO, K. Doi, and T. Tsuji, “High-speed image acquisition using coded structured light projection,” *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 925–930.
- [79] W. Krattenthaler, K. J. Mayer, and H. P. Duwe, “3d-surface measurement with coded light approach,” *Proceedings of the 17th Meeting of the Austrian Association for Pattern Recognition on Image Analysis and Synthesis*, vol. 12, pp. 103–114, 1993.
- [80] V. Yalla, *Optimal phase measuring profilometry techniques for static and dynamic 3D data acquisition*. PhD thesis, University of Kentucky, 2006.
- [81] V. Srinivasan, H. Liu, and M. Halioua, “Automated phase-measuring profilometry of 3d diffuse objects,” *Appl. Opt.*, vol. 23, no. 18, pp. 3105–3108, 1984.
- [82] P. S. Huang, S. Zhang, and F. P. Chiang, “Trapezoidal phase-shifting method for 3-d shape measurement,” *Proc. SPIE*, vol. 5606, 2004.
- [83] S. Zhang and S. Yau, “High-speed three-dimensional shape measurement system using a modified two-plus-one phase shifting algorithm,” *Opt. Eng.*, vol. 46, no. 11, pp. 113603–1 – 113603–6, 2007.
- [84] P. Jia, J. Kofman, and C. English, “Intensity-ratio error compensation for triangular-pattern phase-shifting profilometry,” *J. Opt. Soc. Am. A*, vol. 24, pp. 3150–3158, 2007.

- [85] Srinivasan, H.C., H. C. Liu, and M. Halioua, “Automated phase measuring profilometry: a phase mapping approach,” *Appl. Opt.*, vol. 24, pp. 185–188, 1985.
- [86] M. Halioua and H. Liu, “Optical three-dimensional sensing by phase measuring profilometry,” *Opt. Lasers Eng.*, vol. 11, pp. 185–215, 1989.
- [87] G. C., Hassebrook, L. G., and D. L. Lau, “Composite structured light pattern for three-dimensional video,” *Opt. Express*, vol. 11, no. 5, 10 March 2003.
- [88] T. R. Judge and P. J. Bryanston-Cross, “A review of phase unwrapping techniques in fringe analysis,” *Opt. Lasers Eng.*, vol. 21, pp. 199–239, 1994.
- [89] J. Li, L. G. Hassebrook, and C. Guan, “Optimized two-frequency phase measuring profilometry light sensor temporal noise sensitivity,” *J. Opt. Soc. Am.*, vol. 20, pp. 106–115, 2003.
- [90] O. Marklund, “An anisotropic evolution formulation applied in 2-d unwrapping of discontinuous phase surfaces,” *IEEE Trans. Image Processing*, vol. 10, pp. 1700–1711, Nov 2001.
- [91] L. C. Graham, “Synthetic interferometer radar for topographic mapping,” *Proc. IEEE*, vol. 62, no. 6, pp. 763–768, 1974.
- [92] M. Hedley and D. Rosenfeld, “A new two-dimensional phase unwrapping algorithm for mri images,” *Magnetic Resonance in Medicine*, vol. 24, no. 1, pp. 177–181, 1992.

- [93] S. M. Pandit, N. Jordache, and G. A. Joshi, "Data-dependent systems methodology for noise-insensitive phase unwrapping in laser interferometric surface characterization," *J. Opt. Soc. Am. A*, vol. 11, no. 10, pp. 2584–2592, 1994.
- [94] C. Zuo, L. Huang, Q. C. M. Zhang, and A. Asundi, "Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review," *Opt. Lasers Eng.*, vol. 85, pp. 84–103, 2016.
- [95] X. Su, G. von Bally, and D. Vukicevic, "Phase-stepping grating profilometry: utilization of intensity modulation analysis in complex objects evaluation," *Opt. Commun.*, vol. 98, pp. 141–150, 1993.
- [96] J. M. Huntley and H. O. Saldner, "Shape measurement by temporal phase unwrapping: comparison of unwrapping algorithms," *Meas. Sci. Technol.*, vol. 8, pp. 986–992, 1997.
- [97] J. M. Huntley and H. Saldner, "Temporal phase-unwrapping algorithm for automated interferogram analysis," *Appl. Opt.*, vol. 32, no. 17, pp. 3047–3052, 1993.
- [98] H. Zhao, W. Chen, and Y. Tan, "Phase unwrapping algorithm for the measurement of three-dimensional object shapes," *Appl. Opt.*, vol. 33, pp. 4497–4500, 1994.
- [99] H. O. Saldner and J. M. Huntley, "Temporal phase unwrapping: application to surface profiling of discontinuous objects," *Appl. Opt.*, vol. 36, no. 13, pp. 2770–2775, 1997.

- [100] L. Paulsson, M. Sjodahl, J. Kato, and I. Yamaguchi, “Temporal phase unwrapping applied to wavelength-scanning interferometry,” *Appl. Opt.*, vol. 39, no. 19, pp. 3285–3288, 2000.
- [101] Y. Liu, D. Huang, and Y. Jiang, “Flexible error-reduction method for shape measurement by temporal phase unwrapping: phase averaging method,” *Appl. Opt.*, vol. 51, no. 21, pp. 4945–4953, 2012.
- [102] J. M. Huntley and H. O. Saldner, “Error-reduction methods for shape measurement by temporal phase unwrapping,” *J. Opt. Soc. Am. A*, vol. 14, pp. 3188–3196, 1997.
- [103] “Vialux.” <https://www.vialux.de/en/index.html>.
- [104] “Dlp projection optics.” <https://www.greenlightoptics.com/page/dlp-projection-optics>.
- [105] “Ajp-4500 structured light dmd projector.” <https://ajile.ca/ajp-4500/>.
- [106] “Vialux dlp.” <https://www.vialux.de/en/star-07.html>.
- [107] T. Instruments, “Ti dlp lightcrafter 4500 evaluation module user’s guide,” July, 2017.
- [108] Wikipedia, “Gamma correction,” 2019.
- [109] C. A. Poynton, *Digital video and HDTV: algorithms and interfaces*. Morgan Kaufmann, 2003.
- [110] Wikipedia, “Digital light processing,” 2019.

- [111] Y. Wang, K. Liu, Q. Hao, X. Wang, D. L. Lau, and L. G. Hassebrook, “Robust active stereo vision using Kullback-Leibler divergence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 548–563, 2012.
- [112] A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, and L. V. Streeter, “Separating true range measurements from multi-path and scattering interference in commercial range cameras,” *Proceedings of SPIE*, vol. 7864, no. 1, pp. 4–10, 2011.
- [113] D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt, “SRA: fast removal of general multipath for ToF sensors,” *European Conference on Computer Vision*, vol. 8689, no. 1, pp. 234–249, 2014.
- [114] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, S. Kang, and S. Naik, “A light transport model for mitigating multipath interference in time-of-flight sensors,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 73–81, 2015.
- [115] Y. Gong and S. Zhang, “Ultrafast 3-d shape measurement with an off-the-shelf dlp projector,” *Opt. Express*, vol. 18, no. 19, pp. 19743–19754, 2010.
- [116] G. Zhan, H. Tang, K. Zhong, Z. Li, Y. Shi, and C. Wang, “High-speed fpga-based phase measuring profilometry architecture,” *Opt. Express*, vol. 25, no. 9, pp. 10553–10564, 2017.
- [117] S. Zhang and P. S. Huang, “High-resolution, real-time three-dimensional shape measurement,” *Opt. Eng.*, vol. 45, no. 123601, 2006.

- [118] S. Y. Chen, Y. F. Li, and Z. J., “Vision processing for realtime 3-d data acquisition based on coded structured light,” *IEEE Transaction on Image Processing*, vol. 17, pp. 167–176, 2008.
- [119] M. E. Deetjen, A. A. Biewener, and D. Lentink, “High-speed surface reconstruction of a flying bird using structured light,” *J. Expe. Biology*, vol. 220, pp. 1956–1961, 2017.
- [120] Z. H. Zhang, “Review of single-shot 3d shape measurement by phase calculation-based fringe projection techniques,” *Opt. Lasers Eng.*, vol. 50, pp. 1097–1106, 2012.
- [121] W. H. Su and H. Liu, “Calibration-based two-frequency projected fringe profilometry: a robust, accurate, and single-shot measurement for objects with large depth discontinuities,” *Opt. Express*, vol. 14, no. 20, pp. 9178–9187, 2006.
- [122] K. Liu, Y. Wang, D. Lau, Q. Hao, and L. G. Hassebrook, “Dual-frequency pattern scheme for high-speed 3-d shape measurement,” *Opt. Express*, 1 March 2010.
- [123] J. Lin, K. Jiang, and M. Chang, “A novel solution for camera occlusion in stereo vision technique,” *Adv. Mech. Eng.*, vol. 2013, pp. 1–8, 2013.
- [124] M. O’Toole, J. Mather, and K. N. Kutulakos, “3d shape and indirect appearance by structured light transport,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1298–1312, 2016.

- [125] S. Zhang and S. T. Yau, “Three-dimensional shape measurement using a structured light system with dual cameras,” *Opt. Eng.*, vol. 47, no. 1, pp. 0136041–01360412, 2008.
- [126] C. Jiang, B. Lim, and S. Zhang, “Three-dimensional shape measurement using a structured light system with dual projectors,” *Appl. Opt.*, vol. 57, no. 14, pp. 3983–3990, 2018.
- [127] Y. Wang, D. L. Lau, and L. G. Hassebrook, “Fit-sphere unwrapping and performance analysis of 3d fingerprints,” *Applid Optics*, vol. 49, no. 4, pp. 592–600, 2010.
- [128] S. Zhang, “Active versus passive projector nonlinear gamma compensation method for high-quality fringe pattern generation,” *Proceedings of SPIE*, vol. 9110, pp. 911002–911002–10, 2014.
- [129] P. S. Huang, C. Zhang, and F. P. Chiang, “High-speed 3-d shape measurement based on digital fringe projection,” *Opt. Eng.*, vol. 42, pp. 163–168, 2003.
- [130] T. Hoang, B. Pan, B. Nguyen, and Z. Wang, “Generic gamma correction for accuracy enhancement in fringe projection profilometry,” *Opt. Lett.*, vol. 35, no. 2, pp. 1992–1994, 2010.
- [131] B. Atcheson, F. Heide, and W. Heidrich, “Caltag: high precision fiducial markers for camera calibration,” *Proceedings of the Vision, Modeling, and Visualization Workshop*, 2010.

- [132] Y. Yu, D. L. Lau, and M. P. Ruffner, “3d scanning by means of dual-projector structured light illumination,” *Proc. SPIE 10932*, vol. 10932.
- [133] Y. Zhang, D. L. Lau, and Y. Yu, “Causes and corrections for bimodal multipath scanning with structured light,” *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019.
- [134] V. Couture, N. Martin, and S. Roy, “Unstructured light scanning to overcome interreflections,” *International Conference on Computer Vision*, pp. 1895–1902, 2011.
- [135] J. P. Godbaz, M. J. Cree, and A. A. Dorrington, “Closed-form inverses for the mixed pixel/multipath interference problem in AMCW lidar,” *Proceedings of SPIE*, vol. 8296, no. 1, pp. 8–15, 2012.
- [136] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Ramesh, “Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles,” *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 1–10, 2013.
- [137] S. Nayar, G. Krishnan, M. Grossberg, and R. Raskar, “Fast separation of direct and global components of a scene using high frequency illumination,” *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 935–944, 2006.
- [138] D. Burke, A. Ghosh, and W. Heidrich, “Bidirectional importance sampling for direct illumination,” *Rendering Techniques*, vol. 5, pp. 147–156, 2005.

- [139] F. Sillion and C. Puech, “A general two-pass method integrating specular and diffuse reflection,” in *ACM SIGGRAPH Computer Graphics*, vol. 23, pp. 335–344, ACM, 1989.
- [140] T. Ritschel, T. Grosch, and H.-P. Seidel, “Approximating dynamic global illumination in image space,” in *Proceedings of the 2009 symposium on Interactive 3D graphics and games*, pp. 75–82, ACM, 2009.
- [141] J. Jimenez, K. Zsolnai, A. Jarabo, C. Freude, T. Auzinger, X.-C. Wu, J. von der Pahlen, M. Wimmer, and D. Gutierrez, “Separable subsurface scattering,” in *Computer Graphics Forum*, vol. 34, pp. 188–197, Wiley Online Library, 2015.
- [142] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin, “Phasor imaging: A generalization of correlation-based time-of-flight imaging,” *ACM Transactions on Graphics (ToG)*, vol. 34, no. 5, p. 156, 2015.
- [143] M. O’Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos, “Temporal frequency probing for 5d transient analysis of global light transport,” *ACM Transactions on Graphics (ToG)*, vol. 33, no. 4, p. 87, 2014.
- [144] C. S. Williams and O. A. Becklund, *Introduction to the optical transfer function*. Bellingham, Washington: SPIE Press, 2002.
- [145] C. Harrison, H. Benko, and A. Wilson, “OmniTouch: wearable multitouch interaction everywhere,” in *Proc. of the 24th annual ACM symposium on user interface software and technology*, pp. 441–450, Oct. 2011.

- [146] S. Li, *Per-Pixel calibration for RGB-Depth natural 3D reconstruction on GPU*.
PhD thesis, University of Kentucky, 2016.
- [147] Wikipedia, “Xilinx,” 2019.
- [148] “Xillibus.” <http://xillybus.com/>.
- [149] “Lupa300.” <https://www.onsemi.com/products/sensors/image-sensors-processors/image-sensors/lupa300>.
- [150] “Adp7158 datasheet and product info.” <https://www.analog.com/en/products/adp7158.html>.
- [151] Analog Devices, Inc., *Ferrite Beads Demystified*, 2018.
- [152] Maxim Integrated Products, Inc., *General Layout Guidelines for RF and Mixed-Signal PCBs*, 2011.
- [153] T. Williams, *The Circuit Designer’s Companion*. Burlington: Elsevier Science, 2nd ed.. ed., 2004.
- [154] “Eight ways to reduce noise on your next pcb,” *Electronic Engineering*, p. 63, Apr 01 1999. Copyright - Copyright Miller Freeman plc Apr 1, 1999; Last updated - 2011-07-21.
- [155] H. Lin, J. Gao, G. Zhang, X. Chen, Y. He, and Y. Liu, “Review and comparison of high-dynamic range three-dimensional shape measurement techniques,” *Journal of Sensors*, vol. 2017, pp. 1–11, 2017.

- [156] G. Liu, X. Liu, and Q. Feng, “3D shape measurement of objects with high dynamic range of surface reflectivity,” *Applied Optics*, vol. 50, no. 23, pp. 4557–4565, 2011.
- [157] H. Jiang, H. Zhao, and X. Li, “High dynamic range fringe acquisition: a novel 3-d scanning technique for high-reflective surfaces,” *Optics and Lasers in Engineering*, vol. 50, no. 10, pp. 1484–1493, 2012.
- [158] S. Feng, L. Zhang, C. Zuo, T. Tao, Q. Chen, and G. Gu, “High dynamic range 3D measurement with fringe projection profilometry: a review,” *Measurement Science and Technology*, vol. 29, no. 12, pp. 1–19, 2018.
- [159] H. Lin, J. Gao, Q. Mei, Y. He, J. Liu, and X. Wang, “Adaptive digital fringe projection technique for high dynamic range three-dimensional shape measurement,” *Optics Express*, vol. 24, no. 7, pp. 7703–7718, 2016.
- [160] Z. Cai, X. Liu, X. Peng, Y. Yin, A. Li, J. Wu, and Z. Gao, “Structured light field 3d imaging,” *Optics Express*, vol. 24, no. 18, pp. 324–334, 2016.
- [161] S. Zhang and S. Yau, “High dynamic range scanning technique,” *Optical Engineering*, vol. 48, no. 3, 2009.
- [162] W. Guo, Z. Wei, X. He, and C. Qi, “Design and implementation of a high speed CMOS imaging system,” in *3rd International conference on advanced computer theory and engineering*, pp. 469–472, Aug. 2010.

Vita

Ying Yu

Education

- Master of Science in Engineering, Electronic Information School, Wuhan University, Wuhan, China. 2008.
- Bachelor of Science in Engineering, Electronic Information School, Wuhan University, Wuhan, China. 2006.

Professional Experience

- Research Assistant, University of Kentucky, Lexington, KY, USA. 2016-2019.
- Teaching Assistant, University of Kentucky, Lexington, KY, USA. 2012-2015.
- IC Design Engineer, ZTE Corp., Shenzhen, China. 2010-2011.
- R&D Engineer, Shenzhen Nenglian Electronics Co. Ltd., Shenzhen, China. 2008-2010.

Publications

1. 3D scanning by means of dual-projector structured light illumination, Ying Yu, Daniel L. Lau and Matthew P. Ruffner, Proceedings of SPIE Photonics West, 2019.

2. Structured light smart camera for spatial augmented reality applications, Matthew P. Ruffner, Ying Yu and Daniel L. Lau, Proceedings of SPIE Photonics West, 2019.
3. Causes and corrections for bimodal multi-path scanning with structured light, Yu Zhang, Daniel L. Lau and Ying Yu, Computer Vision and Pattern Recognition (CVPR), 2019.
4. How to develop a low-cost 3D scanner using structured light, Ying Yu, Daniel L. Lau, Vision System Design, June, 2019.
5. Dual-projector structured light 3D shape measurement, Ying Yu, Daniel Lau, Matthew Ruffner and Kai Liu, Applied Optics, 2019. (submitted)