OCULAR MOTION CLASSIFICATION FOR MOBILE DEVICE PRESENTATION ATTACK DETECTION

A Dissertation
IN
Electrical and Computer Engineering
and
Computer Science

Presented to the Faculty of the University
of Missouri–Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
JESSE LOWE

B. S., University of Missouri, Kansas City, USA, 2014

Kansas City, Missouri
2020

OCULAR MOTION CLASSIFICATION FOR MOBILE DEVICE PRESENTATION ATTACK DETECTION

Jesse Lowe, Candidate for the Doctor of Philosophy Degree

University of Missouri–Kansas City, 2020

ABSTRACT

As a practical pursuit of quantified uniqueness, biometrics explores the parameters that make us who we are and provides the tools we need to secure the integrity of that identity. In our culture of constant connectivity, an increasing reliance on biometrically secured mobile devices is transforming them into a target for bad actors. While no system will ever prevent all forms of intrusion, even state of the art biometric methods remain vulnerable to spoof attacks. As these attacks become more sophisticated, ocular motion based presentation attack detection (PAD) methods provide a potential deterrent. This dissertation presents the methods and evaluation of a novel optokinetic nystagmus (OKN) based PAD system for mobile device applications which leverages phase-locked temporal features of a unique reflexive behavioral response. Background is provided for historical and literary context of eye motion and ocular tracking to provide context to the objectives and accomplishments of this work. An evaluation of the improved methods for sample processing and sequential stability is provided with highlights for the presented improvements to the stability of convolutional facial landmark localization, and automated spatiotemporal feature extraction and classification models. Insights gleaned from this work are provided to elucidate some of the major challenges of mobile ocular motion feature extraction, as well as additional future considerations for the refinement and application of OKN motion signatures as a novel mobile device based PAD method.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "OCULAR MOTION CLASSIFICATION FOR MOBILE DEVICE PRESENTATION ATTACK DETECTION," presented by Jesse Lowe, candidate for the Doctor of Philosophy degree, and hereby certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Reza Derakhshani, Ph.D., Committee Chair
Department of Computer Science & Electrical Engineering

Yugyung Lee, Ph.D., Co-discipline Advisor
Department of Computer Science & Electrical Engineering

Praveen Rao, Ph.D.
Department of Computer Science & Electrical Engineering

Ahmed Hassan, Ph.D.
Department of Computer Science & Electrical Engineering

Zhu Li, Ph.D.
Department of Computer Science & Electrical Engineering

CONTENTS

ILLUSTRATIONS

TABLES

ACKNOWLEDGEMENTS

For the opportunities, support, and guidance I would like to thank my advisor Dr. Reza Derakhshani who helped me discover my path to becoming an engineer. His patience and wisdom were critical to the discoveries and achievements I've made in both academic pursuits and personal growth.

I also want to thank the members of my committee Dr. Yugyung Lee, Dr. Praveen Rao, Dr. Ahmed Hassan, and Dr. Zhu Li, for giving their time to advise, listen, and instruct me through this process.

To those who friends have sat alongside me so many days, Dr. Ajita Rattani, Ahmad Mohammad, Ala-Addin Nabulsi, Mark Nguyen, and Narsi Reddy, the challenges we faced together and triumphs we've shared will remain in my heart forever.

A special thanks to Eyeverify (dba ZOLOZ) for funding and supporting my research in a variety of ways.

Finally and with utmost sincerity I want recognize the contributions and unending support of my mom and dad for which gratitude simply is not enough. Through the years of my life they have shown me the meaning of dedication, support, and love. I cherish the bond we share.

CHAPTER 1

INTRODUCTION

1.1 Motivation

Data security is inherently dependent on reliable methods of access control [1]. Establishing a secure system requires reliable methods of user authorization and identity verification [2, 3]. Biometrics are generally considered one of the most secure methods of identity verification, but these systems can be vulnerable to spoof attacks [1, 4–6]. These vulnerabilities open users to considerable personal and professional risks which negatively impact user experience [7, 8]. Presentation Attack Detection (PAD) seeks to mitigate some of the potential risks by incorporating the temporal characteristics of behavioral response to enhance the integrity of the sample collection process [8–10].

Spoof detection using PAD is an independent verification process intended to augment rather than replace traditional morphological biometric methods [8]. While behavior can be considered a type of biometric, PAD methods typically take a generalized approach to facilitate subject independent operation [11]. Security protocols are only useful when there is sufficient user participation, so it is important for an additional layer of the process to add as little complexity as possible to maintain the quality of the user experience [12, 13]. Many existing PAD methods rely on user compliance with instructional prompts or action sequences that add significant durations to the sample collection. Optokinetic nystagmus (OKN) is a reflexive behavior which manifests reliably in normal healthy populations [14–16]. Ocular reflexive responses are virtually imperceptible to the user and don't require compliance or conscious effort [16]. Utilizing OKN as a method of PAD involves the display of a suitable visual stimulus and sequential response sample collection [17]. Computer Vision (CV) and Machine Learning (ML) analysis of the visible features of the face and eyes are a critical part of classification of the response characteristics. While recognition and segmentation of facial and ocular features is a well-researched field, even simple detection is not considered a solved problem [18–22]. Optokinetic response classification requires precise localization of ocular features and facial landmarks for reliable sequential motion estimation [23]. Processing pipelines, libraries, and classification models used to process the response samples should be suitable for mobile device operation, which add non-trivial size and complexity constraints. Based on the literature and preliminary feasibility analysis of the relevant technology, it is hypothesized that the OKN response can be reliably elicited and detected using the display and imaging sensor of a mainstream mobile device.

1.2 Overview

Interactions with mobile devices now comprise a considerable amount of most users secure computing [24]. A broad array of applications and functionality, combined with high-speed mobile networks, offer remote access to wide array of data including sensitive business transactions, financial information, and personal communication [6,25]. These highly portable devices enable more fluid and natural interaction with the digital universe, but also present an enticing target for intrusion and theft [24]. Creating a reliable method to secure these devices is an increasing necessity which has garnered a significant interest in academic and industrial research [1, 5, 26–33].

OKN is a neurological pathway driven response to specific configurations of moving visual stimulus that presents a promising target for motion sequence behavioral classification. Highly structured, yet slightly eccentric, the motions generated by the underlying control system are a prime candidate for subject agnostic behavioral PAD [14, 15, 34–36]. Recent improvements in visible wavelength (VW) based gaze angle estimation methods mean the motion features can be extracted from the same sensors already broadly in use for mobile biometric applications [18, 19, 22, 37, 38]. Due to the fact that OKN is a response to specifically structured moving edges and textures, it's not a behavior commonly observed in daily interaction, making covert recording of the characteristic response difficult [15, 16, 36, 39]. Additionally, as the response is sensitive to specific velocity and acceleration parameters of the visual stimulus, the pattern of animation can be modulated to create an extreme improbability of a pre-recorded response aligning with anticipated behavior [14–17, 35, 40].

OKN is controlled by a reflexive neurological pathway that expresses in a characterizable way among the general population [14–16]. Motions generated are virtually imperceptible to an individual observing the stimulus as integration of visual information is suspended during the larger eye movements produced as part of the response [16, 34, 41]. Despite a broad interest in eye motion and gaze approximation applications in the literature, the collection of the precise sequential motions of the eye necessary to differentiate subtle characteristics remain an unsolved challenge to the broad implementation of this method [23,42–44].

Most of the successful methods currently utilized to secure mobile devices employ authentication credentials hashed from user-specific biological features [1, 28, 30, 31]. Some recent device designs have incorporated additional specialized sensors to more efficiently collect this biometric data [45–47], but even with these advancements mobile devices remain susceptible to subversion via reproductions of likeness or biological traits [48–50]. Since the traits which are used to generate feature hashes that provide secure authentication are typically highly visible elements of personal identity, keeping these features secret is largely impractical [1, 2, 8]. Fortunately, if the methods used to authenticate the user's identity are augmented with the capacity of validating that the features are being presented to the system in real time, it's not necessary to secure the identifying features [11,51]. Systems with this capability can differentiate a direct source from a spoof attack, meaning identifying features compromised by unauthorized collection or replication would be rejected. This layer of knowledge protects from scenarios of persons presenting duplicated features or visual representations as authorized credentials. Methods which are designed to accomplish this task are referred to broadly as liveness detection methods [11,26,51]. Liveness methods are based on temporal factors such as changes in color, texture, or orientation of the input over some period of observed time [9,10,52–54]. These changes are used to verify that the presented credentials are generated from features of a living human rather than a replicated source. Liveness authentication is a broad field which includes discrete methods like PAD, but also encompasses continuous authentication methods. While some potential applications exist for persistent liveness verification via ocular motion analysis, this dissertation will focus primarily on PAD applications [6,8,11,55].

Biometrics is an applied study of quantified uniqueness which incorporates a broad spectrum of scientific disciplines [1,5,26,56,57]. Understanding data derived from a biological source requires investigation of the underlying characteristics and functions associated with the structures being evaluated [29, 32, 44, 58–61]. Collection of useful features from those structures relies on engineering to analyze the capabilities, and the limitations, of imaging and sensing devices used to acquire the data. Once that data has been collected, the tools and algorithms of computer science are essential to developing reliable methods for processing and decision making [32]. Behavioral biometric analysis carries on even further into the understanding of the neurological and psychological processes which underlie the dynamic thoughts, states, and emotions that govern human actio [23,44,62–65].

While not a primary biometric in and of themselves, the main goal of a PAD system is to improve access control when coupled with a suitable morphological feature authentication system [8, 11, 55]. Sequential inputs are important to behavioral methods as single image frames lack descriptive power for complex behavioral traits like pose variation and ocular motion [17,44,66]. These cues can typically be easily detected by a human observer when collected from video sources of sufficient sampling rate, however, there are several layers of complexity for computational detection due to appearance variations and intrasubject response variability [20, 64, 67, 68]. This dissertation presents a review of the relevant literature, an explanation of the collection and processing of participant data, and an evaluation of the methods employed in the investigation of an OKN based spatiotemporal PAD.

1.3 Human Motion and Behavioral Based Classification

Action recognition and human motion processing is a complex but highly promising area in ML research [43, 62, 69–71]. Motion tracking applications are widely used in the entertainment industry, sports, fitness, and biomedical research [72, 73]. Head movements, gesture control, and eye motion have also become a critical part of future innovations in virtual and augmented reality devices like headsets, controllers and room scale tracking systems for immersive interactive simulations [70, 74]. Applications of motion based behavioral classification are used to improve public safety by monitoring areas with high traffic or crime rates and identify potential disruptive or dangerous behavior before it occurs [75–77]. Trainers and coaches commonly utilize motion tracking hardware and software that can provide insights into the biomechanical efficiency of an athlete's technique and provide alternative training strategies to improve sports performance [78, 79].

Motion tracking is a computationally complex combination of feature extraction, change detection and sequence prediction [41, 80–82]. From the broad capabilities and adoption of the technology, it is clear that there are many potential applications of human motion and behavioral recognition. While utilizing this technology can be complex and computationally intensive, it has the potential to streamline interactions with the now ubiquitous world of mobile computing devices by augmenting vulnerable biometric safety applications [37, 70, 81, 83, 84]. Some of the specific aspects of eye tracking differ from other human motion tracking scenarios, but the same general methods can be applied from other areas of the discipline [70]. Eye tracking has a wide array of potential applications that range from assistive devices [80] and touch less interfaces [70]to adaptive user interfaces and predictive content delivery [13, 85] .

Most commercially available mobile devices employ powerful multi-core embedded processors, integrated Graphics Processing Units (GPU), and high-resolution imaging sensors [46]. Edge processing in time critical computational functionality can be accelerated by GPU deployment of ML models, providing fast and secure local device sample processing [86]. Similar hardware configurations for accelerated edge computing are used in autonomous vehicle to track and extrapolate human gait patterns of nearby pedestrians and cyclists to avoid accidents [87]. The power and sensing capabilities provided by these devices, along with their wide user adoption, make them an excellent practical target for the development of new human interaction based applications [13, 88].

1.4 Summary of Contributions

This dissertation presents a novel method of PAD based on reflexive behavioral features. Methods presented here are targeted for mobile device applications, but the techniques are also suitable for other properly equipped platforms. Several key discoveries and contributions related to this work are highlighted below.

- Collection and processing of a pilot study of human subject responses to visual stimulus generated and recorded using a mobile device.

Recordings provide the first known large scale high-resolution collection which captures ocular dynamics of generic device interactions and specific behavioral responses in the presences of prescribed visual stimulus. Consistent responses were noted across the participant population providing verification of the stimulus rendered by the mobile device display to suitably evoke the OKN reflex under the conditions which were used in the study.

- A customized mobile device based data collection application was designed and tested to automate and standardize the collection process.

  Sequences of images cropped from the captured video were used to build response sequences based on the ROI of the eye center provided by facial landmark model localization. Measurements and gaze angle estimates were obtained from the video samples using state-of-the-art libraries and feature extraction methods.

- Generation and adaptation of temporal feature extraction methods based on refined ocular feature localization libraries.

  Proposed improvements to the sequential feature localization which were implemented on a task focused branch of the library used in the pilot study. Samples processed using this branch provide more precise gaze estimation. Subsequent processing work developed a set of key features from sequences of localizataions generated by the improved library. Multiple types of feature processing and classification models where trained and tested. LSTM based ML models trained using these features demonstrate the potential for gaze approximation based applications, such as OKN based PAD, for mobile device biometrics.

- Proposed refinement method for gaze estimation using deep learning based facial landmark localization and alignment methods.

  MTCNN models were deployed in extended testing in an attempt to improved sequential stability by use localization, ROI segmentation, and pose variance normalization. Sequential and recurrent processing models were generated to test the stability of extracted features, and several configurations were tested. Results and possible avenues of improvement in this method are provided.

- An optical flow based motion detection and context stabilization framework for gaze feature extraction from selfie video.

  Results of initial testing provide validation for a potential context stabilized optical flow based motion feature extraction method for ocular and biological structures. Motion of larger groups of pixels in refined regions of interest provided overall better performance in the detection of changes in structural conformation than other tested edge and threshold based methods.

- Spatiotemporal automated feature extraction and classification methods for use with high-resolution biological behavioral data.

  Methods developed in this investigation applied spatial transforms to sequential imaging data obtained from front facing device cameras. These sequences contain stimulus elicited reflexive ocular response, which were treated as a single 3D volume containing feature data of both behavior and structure. Models trained using this method contain a set of spatiotemporal kernels which describe the relationships between common elements of the response dynamics. These models demonstrate the potential future applications for rich causal relationship analysis in other biological and ML fields.

CHAPTER 2

RETROSPECTIVE, CONTEXTUAL AND CONTEMPORARY ANALYSIS

2.1 Liveness and Biometrics

Biometric methods of user authentication are an applied set of fundamental image and signal processing techniques with some innovative solutions specific to processing structured data derived from human biology. Feature extraction and template matching are among the most common areas of refinement among academic and industrial research, but the tools utilized to accomplish those objectives are highly varied. Techniques like deep learning based machers are becoming increasingly common, providing methods of extracting and identifying all sorts of novel biological structures. Biometrically equipped devices allow for easy and relatively secure device access, and as adoption of the technology increases so does the interest in development. Strong global markets for biometric and biomedical computer vision applications have motivated and funded substantial research in the field. The popularity of biometric authentication, coupled with investigating solutions to the current limitations, have also generated a variety of published industrial and academic resources. Many of these studies focus on improving the detection and classification of morphological features in image and sensor data, however, subversion detection and prevention has arisen as a distinct branch of active research. Liveness detection is an increasingly popular area within the branch, and has received extensive treatment in the literature.

2.2 Methodology of Literature Search

At the preliminary stages of the topical investigation a mapping-based approach was employed to optimize the search for relevant source literature. Publications and resources from multiple disciplines were evaluated to establish an index of relevant search terms and criteria. Contextual and semantic methods, which provide content matching suggestions based on frequency and co-occurrence of search terms, were employed to expand the search criteria across the relevant disciplines. Queries were formulated based on, but not limited to, keywords such as behavioral biometrics, *eye tracking, gaze estimation, human motion classification, liveness, ocular reflex, mobile biometrics, ocular image segmentation, optokinetic nystagmus, selfie biometrics, sequence classification, and spoof detection*.

Documents sourced by direct query of keywords were leveraged to provide additional associated documents via reference management tools *Mendeley* and *Zotero*. Library resources and databases such as *EBSCOhost, Google Scholar, JSTORE,* and *PubMed* provided additional direct sources, contextual cross-references, and citation statistics.

2.3 History of Eye Tracking

Eye tracking systems for both academic and industrial applications been the subject of substantial practical and theoretical research in the fields of engineering, computer vision and psychology for over a century. Modern methods provide versatility, but many of the discoveries made in the early research have paved the way for the current state-of-the-art devices and applications. This section provides a brief overview of the field in its historical context to provide a foundation to understanding the tools used and created by our research.

Developed in the late 1800s, the first practical eye tracking methodologies utilized mechanical systems to amplify and record ocular motions. These devices were both archaic and unwieldy, but provided new and interesting observations. While devices like Edmund Delabarre's contraption required the use of an anesthetic to numb the subject's eye in order to attach the mechanical apparatus to the cornea, it allowed a stylus to record the subject's eye motion directly on a rotating cylinder. Despite the limited capabilities, being able to observe the subtle behaviors and properties of the eye led to countless discoveries and insights into the mechanics of disease and the treatment of visual impairment. Cost and complexity were still limiting factors however, and adoption of even the crude measurement devices was limited. Most experimental processes of the period were still conducted via observation with reflective devices or optics. These crude methods made it difficult to observe and measure some of the small but important motions responsible for refining foveal alignment, as rapid small scale motions are typically filtered out by the perceptual systems of the human observer. Despite some limits and setbacks, promising findings spurred a variety of dialogues and publications which drew considerable interest in the field. More laboratories began their own work, and toward the turn of the century many researchers started noticing several types of uniform behaviors. I would take several years for observational data to provide the proof, but these findings indicated a more subconscious origin of fixation control than would have initially been assumed if the eyes were primarily under voluntary control.

Some of the most notable findings of the late 19th century included Dr. Louis Javal's observations of the regular high velocity displacements, which he called saccades, when performing reading related visual search tasks. Realizing these movements were too rapid and rhythmic to likely be the result of conscious control, which likely would involve a smooth scanning over the line of text, Javal and his colleges began to theorize about the possibility of a subcortical pathway which controlled a greater part of ocular motion. Despite the meaningful content of the visual space containing linearly sequenced elements, the participants gaze only fixed briefly before rapidly jumped toward some other element. Each of these jumps resulted in a non-trivial angular shift, causing the traversal of a variable linear distance. Given that most voluntary muscular actions contain some degree of gross motion in the initial stages combined with more granular control as the action nears the target, most theories at the time would have indicated that the behavior he observed was the result of a muscle memory and learned behavior. Javal's work in ocular motion focused on pathologies of binocular coordination and alignment which grew from a familiar issue of strabismus, a developmental issue that causes the eyes to have difficulty aligning retinal projections. His observations led him to believe that if some individuals lacked the ability to synchronize their eye motions, the system that controlled this function was one that arose from hereditary physical structures in the brain, rather than some defect of learned behavior or diminished muscular control. Voluntary motions of most muscle groups are the result of integrated neurological activations in the motor cortex which result in modulation of a reflex pathway to suspend the normal force balance of agonist and antagonist muscle groups. While eye gaze can be consciously directed, many common ocular motions are generated by involuntary reflexive pathways. While the technology to objectively measure the virtually imperceptible characteristics of saccadic behavior didn't exist until some years later, subsequent research conducted by Erdmann and Dodge produced remarkably accurate measurements of saccadic latency with only slight improvements on the original visual observation method. Measurements of the duration of fixation between saccades indicated that the short stops allowed time for integration of visual content, with the eyes pausing only long enough to generate recognition of the target content and then move on. Related studies noted that this integration appeared to be suspended while the eye was in motion during a saccade, but not when following a moving target.

Current research still builds on the work of pioneers like Delabarre, Javal, and Erdmann and Dodge, but it was the improvement of measurement and logging devices built in the early to mid-20th century that ultimately provided the biggest step forward for ocular motion research. Scientific advances provided several powerful research tools, most notably those related to photographic equipment, optics, and electrical measurement. More sensitive films allowed for reduced exposure times while maintaining detail, and video cameras became a common part of most behavioral laboratory space. Techniques utilized in imaging studies took advantage of the reflective characteristics of the eye's surface by noting the change of apparent glint produced by a fixed light source relative to the angular characteristics of the subject's gaze. The process of calculating ocular motion through this reflective source methodology is somewhat complicated by the elongated anatomical characteristics of the eye, however, the ease of implementation combined with relatively high-quality gaze localization has seen this technique largely persisting into the modern era of ocular motion research.

Response time has always been a major item of interest relative to the function and control of the visual system. As materials and electronics improved, researchers began work in measuring the electrophysiological responses of the eye's muscular system. Recording motions with electromyography (EMG) enabled improved understanding of the temporal characteristics, but the devices are invasive, expensive, and required the subject to hold completely still to reduce noise from other muscular movements. Of the more useful developments resulting from the early work were the simplified configurations of EMG that focused on measuring periocular musculature responsible for movement in the eye lid to note blink events. Knowing when blink events are present, and implementing the proper filtering is a major part of stable eye tracking applications.

Improved camera systems played a major part in the advancement of both spatial and temporal analysis in optokinetic research. Film based collections tended to be limited in low light conditions, and since many studies were conducted in relatively dark environments, traditional cameras produced under exposed images that lacked clarity. Since the measurement calculations using images are based on changes in position, this became especially challenging when dealing with the millisecond level events common in motion research. Low light conditions are often necessary when observing pupillary dynamics or using distant illumination, so researchers were desperate for new methods.

Toward the end of the 20th centry digital camera technology would gradually replace and supersede most other eye tracking tools. Charge-coupled device (CCD) based image sensors had limited spatial resolution, but their infrared sensitivity meant they could capture structural details of the eye without the need for visible light illumination. Improvements were made over several decades in the imaging device resolution, photo sensitivity, and noise characteristics, resulting in improved capture rates and image quality. By the 1990's, digital imaging devices had become notably smaller, lighter, and cheaper which improved their utility in research applications. Fully digital ecosystems meant experimental results could be recorded directly to a computer and processed by computer vision and pattern recognition systems virtually instantaneously.

In recent decades, studies of ocular motion largely focus on either user interaction or biologically focused analysis of neurophysiological pathways directing reflexive responses and visual attention. Neuroscientific and psychological studies of the ocular dynamics often integrate EEG, fNIRS, or fMRI to measure neurological responses that associate cognitive states and brain regional activations with saccades, pursuits, and fixations. Low latency visual response mechanisms like the Vestibulo Ocular Reflex (VOR) and Optokinetic Nystagmus (OKN) have been extensively investigated using these methods, but new discoveries and insights are still being made. Consumer research enables improvements in user interface and experience by understanding how the presentation of elements or products impacts behavior or interest. Many of these studies are done using wearable head mounted tracking systems or specialized hardware trackers, but as the accuracy of image-based gaze tracking methods improve there is increasing interest in the potential of consumer grade webcams and mobile integrated camera devices.

2.4 State of the Art

2.4.1 Infrared

Imaging of the eye is complex for a variety of reasons that relate at least in part to the special interaction of light sources with the tissues that compose the lens, iris, and cornea. Translucent properties of these tissues and the mucous membranes that cover them create *specular reflections* of incident light which is both an advantage and a disadvantage in imaging based tracking. Reflections can be a problem for most photographic applications, as localized high intensity reflections can obscure important features. Many hardware based commercial and research grade eye tracking devices have been designed with the take advantage of these reflective properties by approximating gaze angle using the interaction of light with the eye's physical structure. Anterior portions of the eye are given an ovoid curvature by the slight budge of the aqueous humor that fills the cornea. Glints are small points of bright light formed by reflections of focused sources on the corneal surface. Fixed sources at some distance from the surface follow characteristic patterns of projection relative to the curvature of the anterior surface. These patterns can be mapped out to determine the change in gaze angle relative to the apparent change in linear position of the reflected glint pattern Fig. 1. Rather than a single source, trackers commonly use an array of sources to improve tracking accuracy.

Visual pigments which are used in human photoreceptor cells as part of the process of converting light into nerve impulses are sensitive only to a narrow band of wavelengths.

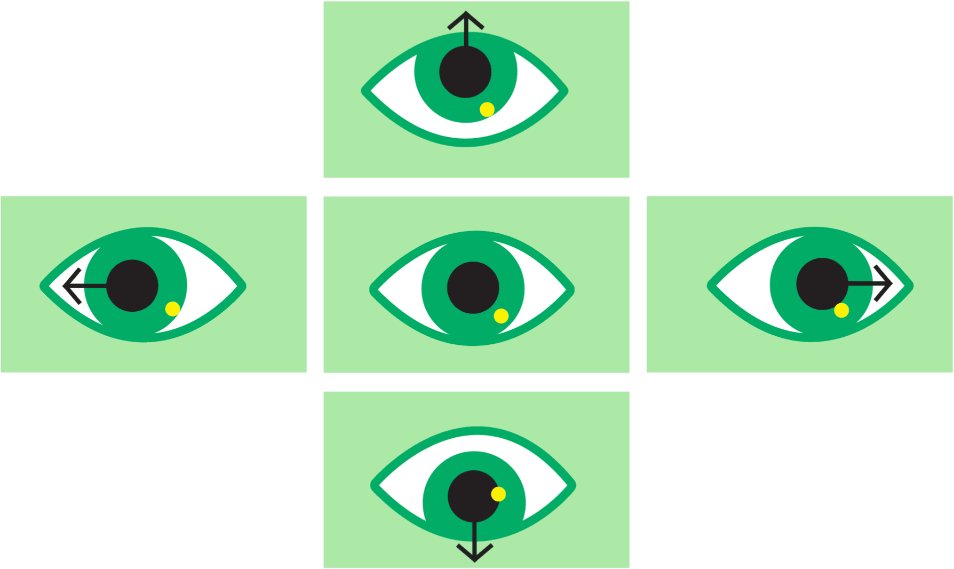

Figure 1: Glint based eye tracking provides an estimation of the gaze angle based on the movement of the reflection of an IR source.

Using matched infrared (IR) imaging and light sources provides better focal characteristics when designing the lens systems, and allows glints to be tracked relative to the easily segmented dark pupil center. Modern cameras offer frame rates and imaging resolutions sufficient to accurately and non-invasively record normal ocular motion. Pupillary dynamics aren't impacted by the use of IR illumination, which is a key advantage for the study of ocular response dynamics. Heat generated by high intensity IR sources can potentially damage the cells of the retina by causing thermal damage to the tissues. Illuminators used in most studies generally have intensities that fall far below exposure guidelines, and thus are generally considered safe, but exposure to all radiation sources should be minimised. Precautions are always merited and protocols should consider all relevant risks in human studies.

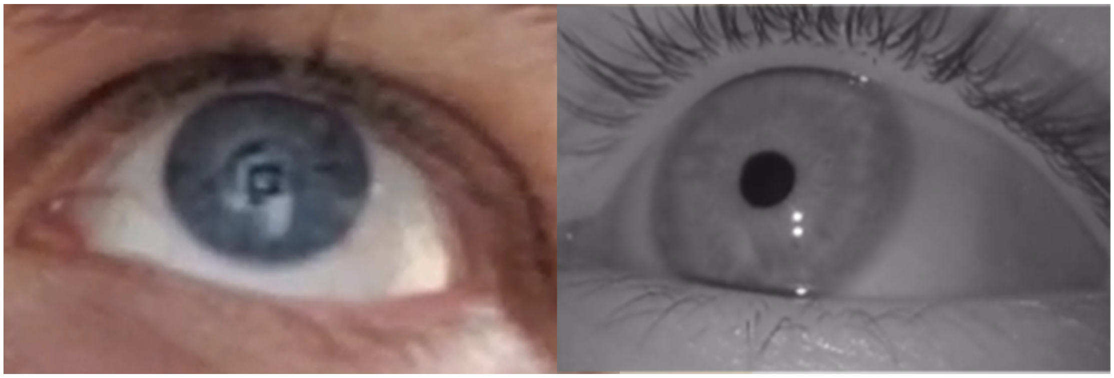

Figure 2: Images of the eye collected in both Visible (left) and Infrared (right) Wavelengths. Iris pigmentation drastically impacts visible imaging of iris features.

Head mounted tracking solutions provide precise measurements of ocular response dynamics by mounting miniature cameras and illuminators on a secured platform up close to the eye. Desktop trackers are less invasive, but generally less precise. They are designed to work at considerable distance using reflective glint-based tracking. Facial pose can influence the apparent angle of gaze, so many desktop trackers have started to include some facial pose detection and compensation in their software.

Structures of the iris are easily resolved in IR as the pigments that compose the eye coloring are highly reflective outside the visible spectrum Fig. 2. For this reason, many studies of identification and behavioral response related to the eyes have focused on film and sensor data collected in IR. This is ideal in a laboratory situation as optimal parameters can be selected and confounding variables are easily controlled. In the case of collections in the wild with user devices, key conditions can vary based on use case and environment. Given that the intensity of IR is typically less than visible light in most conditions, the exposure characteristics and detail of images captured are more consistent. Most objects and materials emit or reflect some amount of light outside the visible spectrum so there are still a variety of sources for interference. Most IR imaging sensors are based on either CCD or complementary metal oxide semiconductor (CMOS) designs that provide good sensitively and reasonable levels of noise. Even CCD and CMOS sensors with embedded color pattern filters still convert some amount of the invisible spectrum, resulting in washed out images with unappealing color anomalies due to the additive characteristics of the invisible spectrum. To maintain visual appeal of the images captured with these cameras, most lens systems are equipped with cutoff filters to limit IR and ultraviolet wavelengths. Only a few mobile devices come equipped with IR sensitive sensors, and even in these cases the interface may not allow direct access to the sensor data.

2.4.2 Visible Wavelengths

Visible Wavelength (VW) cameras generate representations of the color content of the subject through use of filter arrays for red, green and blue wavelengths. Capturing color data can be useful for distinguishing some of the unique features of the eye as well as localizing some features in the periocular space. Miniaturization of the camera technology ultimately gave way to the inclusion of VW imaging sensors on most mobile devices. Many of these devices have more than one camera, or at least allow the camera orientation to be changed to the front of the device. Front facing cameras have been commonly used for identification based biometric applications, selfies for social media, and video conferencing, but they are also becoming tools for behavioral analysis and user interaction research.

Until quite recent generations of mobile devices the research applications of integrated VW cameras have been limited as the front facing sensors lacked sufficient spatial or temporal resolution to provide quality features. Improvements in the sensor technology include more precise optics, better light sensitivity, and increasingly dense imaging arrays. User behavior and preference ultimately influence the selection and specifications of mobile device sensors. Most of the recent changes appear to benefit eye tracking applications, as trends in application use have begun to dictate the inclusion of sensors with improved video capture performance.

All gaze analysis methods require the imaging device capture rate to meet a minimal event sampling frequency. Capturing temporal values related to saccade velocity or dwell time can only reliably be accomplished within a proportion of the frame capture interval. Some short duration ocular events can achieve velocities of nearly 1000 °/s resulting from accelerations up to 100,000°/s^2 [67]. Characterizing events like these requires a sampling frequency well beyond the range of most existing mobile device cameras. Consumer demand for high quality video capture and streaming rates has spurred integration of sensors that can reach 30-60 frames per second on some models. Front facing devices tend to lag behind rear cameras, but trends in the adoption of video conferencing and live streaming are driving improvements. Future devices may soon provide front facing sensors with sampling frequencies matching current commercial hardware-based trackers.

2.5 Facial Feature Localization

Understanding the topology of the structures surrounding the eye is an important part of measuring ocular movement. Ideally, the method best suited to provide the necessary landmarks is one which can operate with standard imaging sensors present on most devices. Specialized depth imaging sensors have recently been developed that are capable of mapping the face using structured light or laser scanning, but so far only high-end devices have begun to integrate them [46]. Libraries like Dlib [89] use pattern recognition methods to detect and label key parts of the face in an image. Dlib uses a feature set based on the Histogram of Oriented Gradients (HOG) and classifications derived from a Support Vector Machine (SVM) to return an array of localisations based on the distribution of gradient directions. More traditional pattern recognition based systems like this are typically selected because they tend to be lightweight enough not to compromise performance in mobile applications, but with the increasing operability of on device acceleration deep learning based localization models are becoming increasingly popular. The Dlib localization method utilizes Active Shape Models (ASM) to provide a dense array of estimated morphological markers. ASMs are a class of coordinate deformations which provide a systematic constraint for facial landmarks to improve the accuracy of detection systems [90]. Mesh models are also being developed which aim to provide better 3D structural matching than ASM systems. The *eos* library [91] uses morphable mesh regression to facilitate a better feature fit.

2.6 Facial Pose Detection

Reliable assessments of facial pose are necessary to fully understand the motion context when observing visual response of a user in free space. Most facial structures are composed of soft tissues that deform with motion or changes in expression, making them a poor choice for assessing relative distance. A fixed frame of reference is necessary for measuring change in gaze angle relative to facial orientation. Displacement calculations are based on linear translation of detected landmarks between keyframes. Change in relative position provides the baseline for deriving velocity using an assumed feature scale and frame interval. Relative precision of the response kinetics is dependent on the segmentation quality of ocular features, and the stability of anchor points. Regions where facial tissues connects closely with the skull, such as the eye corners, brow, and bridge of the nose are typically selected.

2.6.1 Deep Learning Based Models

MTCNN is a multi-stage Convolutional Neural Network (CNN) based facial bounding box detector and feature localization model which is currently one of the top performing method of facial detection and pose approximation for VW images [92]. While the array of markers provided by the refinement stages of this model are far less dense than some other ASMs, the model performs far better in terms of sequential stability. Like most state-of-the art facial landmark and feature localization methods MTCCNs are trained with and therefor perform inference operations on small images. Resolution limitations of these feature extraction methods are primarily related to the computational complexity and memory constraints of processing high resolution image arrays. Even in cases where resources might allow the processing of high-definition images, faults tend to increase in the pattern recognition performance. Losses resulting from large scale inputs are widely credited to a phenomenon known as the curse of dimensionality, a result of oversaturation that leads to diminishing capability to derive feature significance.

2.7 Artifact Based Attack Detection

Scale plays a factor in pattern recognition in terms of the resolving power [93] respective to the spatial features, as sufficient detail must be provided to distinguish boundries of adjacent elements. In terms of behavior this is represented by the temporal sampling rate with respect to the period of some sampled event, and the required ¡2N sampling frequency as detailed by Nyquest [94]. Utility of additional data density therefor must be balanced when conducting a search for feature patterns, as irrelevant or highly correlated elements of the sample can hinder performance of a classifier by diluting feature significance. Native resolution images and videos collected using standard sensor modes on most mobile devices need to be downsampled to work in these models, meaning some localization errors are likely to occur due to a loss of input resolution. Errors of this type are integrated into all feature localizations in the case of single samples and compounded in the case of sequential analysis.

2.8 Liveness

No system will ever prevent all forms of intrusion, but support systems such as liveness detection are commonly used to combat common subversion [11,26,51,52,95]. Liveness based methods have been proposed as a method to provide protection from some sophisticated attacks by verifying that samples originate in real time. Leveraging sequential sources, these methods utilize the temporal features derived from the motion of physiological structures. A wide variety of studies have been successfully demonstrated the benefits of implementation for liveness detection utilizing estimates of physiological signals like pulse and respiration from video based sensing [96,97]. While eye motion has been studied in some contexts as it relates to liveness and PAD, no known system integrates OKN as a part of the behavioral classification. Comparison to state of the art PAD methods is difficult as a result of this factor, but comparable methods of ocular behavior classification will serve as a baseline.

2.8.1 Presentation Attack Detection

OKN response based PAD leverages the a parameterized animated visual sequence to elicit a designated response within a specific temporal window. A lead or lag in the temporal characteristic response, or incorrect displacement (phase) provide an assessment of input synchronization. Significant deviation of a presented response provides the classifier with an indication the sample may not originate from a live source. Extracting ocular motion features from a static image or screen based attacks will result in large skew toward a particular gaze angle estimate, whereas a replay-based attack would likey distribute a more variable estimate that drifts from the mean of responses for the given stimulus parameters. In the case of sensitive applications, where more robust detection is required, randomized adjustments to stimulus parameters, such as velocity, direction or sequence of onset can be employed to limit the possibility of the correct response parameters being presented for a more sophisticated simulated or reproduction based attack.

One critical factor to establishing trust in any secure transaction is related to verifiable presence of the authorized user [3, 98–100]. Presentation attacks occur when an impostor attempts to mimic an authorized user by means of presenting duplicated or synthesized biometric data [6, 8, 11]. Means and methods of constructing and presenting this data are ever-evolving, comprising a vast array threats which plague developers and users alike [101–104]. At the most basic level, there will always be some set of features that separate samples collected from a live user and even the most advanced spoof [104]. Determining what those factors are for some specific collection method, and prescribing a process for reliable detection, is the primary focus of PAD [6, 8, 11].

PAD systems are designed to serve as an enhancement to pure appearance based biometric authentication methods [11, 55]. Confirming a sample originates from a live source adds an additional layer of confidence to the template matching and identity verification process. Behavioral based liveness detection and PAD methods are meant to augment rather than replace other identity technologies [11]. Utilizing behavioral information gleaned by processing the response of the user to targeted stimuli adds a robust temporal characteristics which make real samples substantially more difficult to generate or replicate [104]. Detection of spoof attacks via this analysis of response based parameters adds a dimension not employed by most traditional methods [65, 103–105].

2.9 Behavioral Classification and Reflexive Motion

Reflexive responses make up a large part of human motor activities [106]. Even normal voluntary motions such as walking or picking up an object are composed of delicately attenuated reflexive responses [107]. For the purposes of behavioral biometrics, most reflexive responses are impractical to elicit under normal circumstances, but ocular methods hold promise of utilizing unconscious pathway responses to prescribed stimulus to insure the biometric features being presented are originating at the time of request [68, 74]. Reflexes of the eye's motor system used in visual stabilization during head movements are a promising candidate for response based classification, as they manifests in a highly predictable way to specific patterns of moving visual stimulus [17,39,41]. This particular response has been observed to be consistent and characterizable, but its applications for biometric authentication have not yet been fully investigated [14, 101, 108]. One major issue associated with implementation of biometrics with behavioral features is the complication of user compliance, which can lead to significant reduction in usability and overall user experience [11, 13, 18, 109]. OKN responses are generated by a distinct neurological circuit and thus requires no instructions or user training.

Pathways involved in the OKN response are distinct from other conscious processes of ocular control, but share some of the same connections used by the VOR.

Visual integration, the neruological process responsible for perception, is suspended during saccadic displacements as part of maintaining balance. As a result of the integration suspension the fast phase saccade events, which regularly occur as part of the OKN cycle, are virtually imperceptible to the user [13, 14, 110].Passive methods of PAD provide benefits over interactive methods due to their transparency, ease of integration, and independence from user compliance [74, 111]. As with all behaviorally derived metrics, some variation can be observed based on individual factors [14]. Observations made in our study indicate that the high-level motion signature of stimulus induced OKN can be used to generate subject independent models capable of discriminating between simulated video based attacks and genuine response samples. In this paper we explore the potential applications of a novel PAD method for mobile devices based on the classification of characteristic time and phase locked OKN response to sequential stimulus, especially for biometric modalities which use the front facing cameras [26] of mobile devices.

2.10 Usability and Adoption

In our culture of constant connectivity, mobile devices are increasingly becoming a primary means of online interaction and physical access control. Mobile based biometrics are facilitating a burgeoning array of diverse smart and connected devices such as remote door locks, electronically secured safes and even user specific smart firearms. Many applications employ protocols which provide secure remote transactions, but convenience driven user practices tend to subvert these additional features due to undesirable complexit [13, 112]. A recent study of user attitudes and interaction with mobile device security indicated users prefer devices security features to be activated out-of-the-box, and despite expressing concerns about security (68%) most fail to engage with the features provided [12]. While user attitudes are subject to change, next generation biometric tools will need to focus on integration with user behavior by compensating for obstacles that may limit their implementation and utilization [103, 113].

Usability of a system decreases in proportion to complexity of implementation [112]. For any security measure to be effective, it must gain acceptance and be consistently utilized [114]. This process works best when the time and effort required to comply remain low. PAD methods promote adherence by providing added security while maintaining usability. As such one of the primary goals advantages of reflexive system such as OKN is ease of adaptation and integration via virtually transparent functionality.

2.11 Optokinetic Nystagmus

Evoked OKN is complex pathway driven cyclic two phase reflexive response. Motion of the eyes in OKN occur in response to specific stimuli and consist of an initial slow phase, or smooth pursuit, followed by a fast phase, or saccadic return [110]. Among individuals with normal neurophysiology, moving fields or arrays of high contrast colors, values, or textures predictably elicit the response [16, 115–117]. Visual integration is suspended during the large scale translations which accompany the return saccade, making the response virtually imperceptible to the subject, however, the oscillatory motion of the eye is readily distinguishable to an outside observer [107]

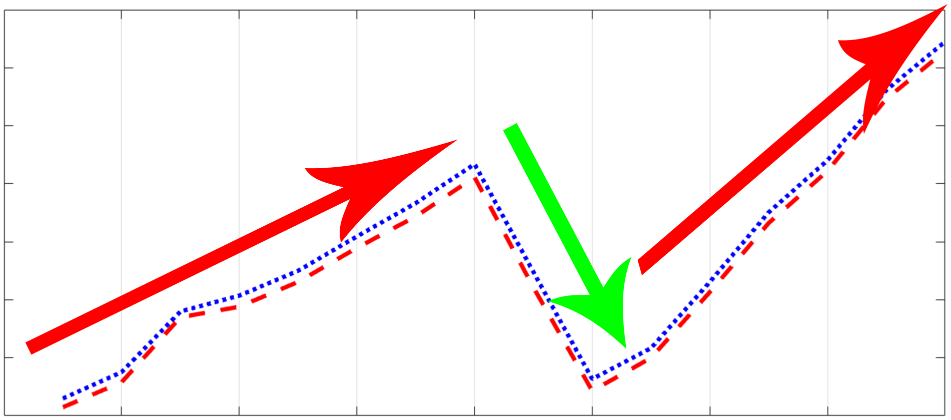

Figure 3: Motion paths of typical OKN response sequence showing two pursuit (red arrow) motions and one return saccade(green arrow).

2.12 Saccades and Smooth Pursuit

Saccadic eye movements are a common target for analysis due to their large displacements and high velocity [67]. When viewing static scenes, most visual search activities consist of a brief dwell time, also referred to as gaze or fixation, followed by a rapid ballistic motion which orients the central visual field to a new point of interest [107]. Control systems of the eye adjust position based on visual content presented to the densely populated high visual acuity region of the retina known as the fovea which occupies the central 10°of the visual field [66, 115]. New gaze points typically occur several degrees away from the previous point of fixation. These gaze points are selected by the brain's visual attention system, and are based on features such as movement or content which is predicted to enhance the viewer's understanding of the scene [107]. Saccades are most predictable when interacting with highly structured visual information such as text, basic geometry, or high frequency patterns and textures [39, 118]. Occasionally an overshot or underestimate by the visual control system can cause an error in the end position of the eye relative to the intended target, in these cases a subsequent corrective saccade, typically of lower magnitude, is initiated to compensate for the initial error [40, 67]. Visual responses to dynamic environments differ significantly from static scenes.

In the presence of moving visual targets, ocular control systems prioritize tracking over a detailed structural analysis [34, 40]. Smooth pursuit movements of the OKN response typically begin with a ballistic acceleration designed to catch up to moving visual stimuli as it slips from the central visual field [67, 107]. A sequence of a normal OKN cycle is shown in Fig. 3. Once the target is back in the central field, the ocular control system adapts to match the target velocity [115]. Smooth pursuit motions are regulated by systems which are difficult to modulate by conscious control and as such rarely occur without the presence of a moving target [115, 117].

2.13 Clinical Practice

Much of the existing scientific work surrounding measurement and evaluation of stimulus evoked OKN focuses on the application of the associated technology to the study of motor pathway related diseases and other neurological pathologies [119–123]. In clinical practice, latency based observations of induced optokinetic events are utilized to diagnose brain and nervous system disorders such as brain lesions or concussion [34]. Clinical applications of OKN response measurement rely on large projection surfaces that range from computer monitors and televisions to room scale devices containing rotating banded cylinders [67, 110, 117].

2.14 Applications

Smaller devices such as hand held cylinders are commonly used for eliciting OKN in diagnostic settings. While full visual field stimulation is more desirable for measuring some clinical metrics, a limited pathway response can be observed for stimulus in the primary visual field [16, 17]. Limited research has been conducted on the generation and analysis of ocular motion responses using a mobile device [18, 74, 88, 113], but no know studies have incorporated an OKN based collection protocol. Over the past decade, significant innovations in image processing and machine learning have facilitated a new wave of analysis aimed toward combining behavior with standard models to create a more robust system. These innovations enable motion signatures to be extracted, with relatively high accuracy, from non-ideal sources such as front facing cameras on mobile devices [124, 125]. Combining physiological morphology based ocular biometrics with behavioral response features facilitates enhanced resistance to emerging attack vectors by adding response based PAD to the sample collection process [11, 63].

2.15 Secure Computing

In secure computing trusted access is granted by providing the appropriate credentials at the appropriate time. Potentially fraudulent or deceptive behavior can be detected by discrepancy or irregularities in this process of authentication. While secure data transmission and strong encryption have been improved by advancements in computing technology, even increased algorithmic complexity can't secure data indefinitely. Secure credentials or other identifying factors are part of establishing trust in secure computing methods [3, 98, 100]. Sensitive systems restrict critical functions through the use of credentialed authentication for privileged users. Passwords controlled systems are relatively easy to implement and therefor remain popular.

Passwords still serve as a primary method of access control for most commercial remote accesses and secure storage resources. Identity is often closely related to access control for restricted areas or systems access, thus biometrically secured systems have become increasingly popular for high security operations. Biometrics is an applied engineering discipline which focuses on image and signal processing methods to provide quantification of unique identifying features which can be used as secure access credentials [20,28,44,64,126,127]. While passwords can be changed if lost or stolen, biometrics are based on structural features that last a lifetime. Facial recognition is perhaps the most well-known optically-acquired biometric modality, but ocular authentication methods are becoming increasingly popular [30, 54, 101, 103, 104, 128–130]. Mobile device front facing *selfie* imaging sensors have improved significantly over the past decade, and now provide advanced VW photography capabilities and exceptionally high-quality imagery. These advancements have enabled the collection of highly detailed features from both the ocular and periocular regions [30, 32, 59, 125, 131]. As evidence of the power of VW front facing cameras, studies have demonstrated their capabilities in imaging conjunctival and episcleral vasculature, as well as peri-ocular features without the need for additional sensors [29, 32].

2.16 Mobile Device Security

Mobile computing dominates the digital access ecosystem [12,24,27]. In the ubiquitous and inherently complex world of mobile device security, developers and manufacturers are engaged in constant efforts to fight theft, fraud, and address privacy concerns. Finding a reliable way to secure these platforms from unauthorized access has been a principal focus of many biometrics related applications over the past decade [28, 30, 56, 63, 99, 109, 126, 132]. With each new generation of devices, developers have introduced new sensors [46], methods [133], and features that bring with them significant complications of design and cost. Despite their best efforts, many of these devices suffer from significant faults that resourceful and determined bad actors can leverage to gain virtually unfettered access to sensitive personal information, communications, media, and finances. Establishing secure access to all this sensitive information means becoming increasingly reliant on tools such as biometrics to establish the trust needed for reliable remote transactions [27,28,47,54,62,84,124,127]. Consequently, reliance on these methods has transformed them into a target for bad actors in the never ending digital arms race [6, 12, 24, 26, 63, 132]. Enhanced detection of illicit behavior is an economic imperative, as even with state of the art processes incidents and costs are rising unsustainably. Predictions of transactional fraud on digital platforms, 60% of which can be attributed to unauthorized access of mobile computing devices [134], estimate the annual economic impact to exceed $6 trillion by 2021 [24, 135].

2.17 Susceptibility of Biometric Systems

Many authentication methods that rely exclusively on morphological features have flaws that can be exploited by adequately motivated individuals. Even sate of the art biometric methods struggles to provide secure operation, and are commonly the target of spoof attacks [27, 30, 54, 57, 103, 104, 128, 136]. Presentation attack detection (PAD) methods classify samples utilizing features and artifacts specific to low effort screen and print based attack vectors. These methods serve as a deterrent to most intrusion attempts, but increasingly sophisticated attacks are now emerging [8, 48–50, 52, 55, 105, 130, 133].

2.18 Challenges of Ocular Kinetics Capture

OKN methods face many of the same challenges as other biometric and biometric adjacent methodologies for mobile device applications. The following section provides an overview of some challenges faced in our data collection and evaluation processes, as well as some measures which were implemented to mitigate their impact.

2.18.1 Resolution and Optical Quality

Compact devices like smartphones and tablets are at the vanguard of miniaturization. Due to the highly valued internal real-estate, a device's front facing camera is likely to reside in an incredibly small space. While some mobile devices now come equipped with infrared (IR) sensors designed capture structures and features outside the visible range, VW sensors are the most common due to factors of cost for consumer grade devices. As users increasingly rely on their mobile devices as part of social media and video streaming platforms, high density front facing VW imaging sensors are increasingly common. Packing tiny cameras in tight spaces generates demand for precision optics which become increasingly difficult to fabricate at scale without introducing substantial image aberration. To compensate for the distortion generated by these ultra-compact lens sensor systems, images are corrected using integrated pre-processing methods. These methods attempt to improve visual appeal by smoothing and denoising the input without specific concern for the morphological features. While the visual appeal is typically improved, the resulting images can be substantially degraded as it applies to machine vision applications. To facilitate machine vision applications, most mobile operating systems provide developers with libraries that enable the acquisition of raw images that improve segmentation and retain critical structural details. Images captured from VW devices in raw mode typically contain a luminance channel which contains the bulk of relevant features stored as a grayscale array. Using this data allows an image processing pipeline to generate high quality feature localization while minimizing processing complexity.

2.18.2 Stability and Sampling Rate

Mobile device operating systems utilize an asynchronous command queue framework to maintain a responsive user interface. Execution of application directives is intentionally restricted to contravene blocking sequential operations, and commands are instead based on a callback process which democratizes system resources. As a result of this architecture, commands issued to capture data from cameras, sensors, or any other hardware can have variable execution time. Reliability of the motion data derived from these imaging sensors rests on the consistency of the interval between frame captures. Inconsistency can lead to error in the estimated displacement velocity and response kinetics for time-locked or phase-locked operations.

OKN based liveness methods like rely on stability of temporal responses for the quality of stimulus onset detection. Critical faults and undesirable outcomes can result if the processes of display or collection are interrupted. False rejections of live samples, caused by loss of synchronous screen space visualization and imaging capture can lead to degraded user experience.

Alternative access methods and third party hardware interface libraries can be implemented to counteract the issue of frame rate variability in some cases. However, due to the dependence on stable screen space visualization, these solutions may produce additional collection instability by altering the stimulus rendering process priority. While a there are a variety of solutions to this dilemma, our experimental collection application was modified to provide a frame capture stability estimate based on timestamps provided by the camera interface. When combined with the display frame rate estimate, the offset was calculated to determine the nearest frame to the stimulus start event.

2.18.3 Temporal Resolution

Due to physical limits of the sensing device hardware such as read rates, transfer bandwidth and shutter speed, constraints are introduced by firmware and software control. Encoding formats used in conjunction with most integrated front facing camera sensors place an upper limit of 29.97 to 30 frames per second. These limitations on the frame

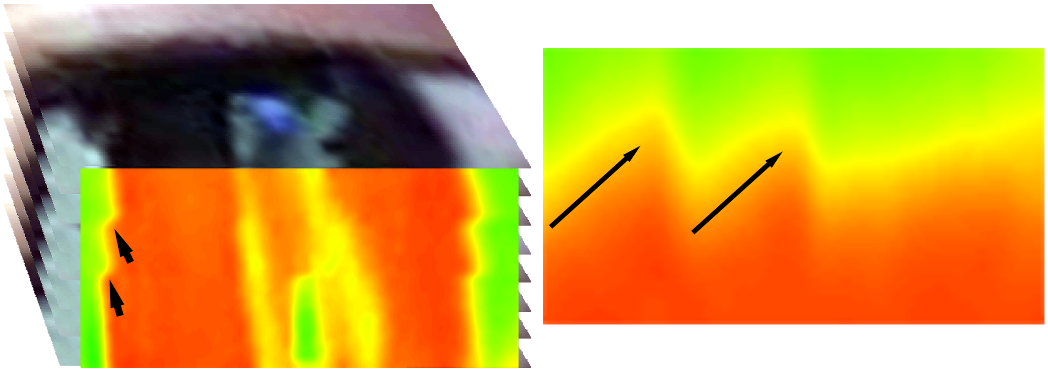

Figure 4: A sample sequence of stacked registered images collected from a mobile device camera. A motion based heat map (right) generated from the changes (arrows) to the iris border during the OKN response.

rate place a cap on the motion characteristics which can be derived from the video sequence data using mobile device hardware.These constraints limit the capture interface and thereby the total temporal sampling rate which can be reliably achieved by a mobile sensor systems. While some of the events that make up the OKN cycle are far fast to be captured at this sampling rate, slower motions that make up the pursuit and small scale displacements fall within the limits of a front facing system. Motion signatures of gaze response sequences can be visualized frames extracted from capture videos as seen in Fig. 4.

2.18.4 Illumination Conditions

Changes in external illumination condition, such as indoor versus outdoor lighting environments generate substantial complexity for machine vision algorithms designed to extract and localize biological features. While many methods have been implemented to solve these issues relative to ocular and periocular features for biometric applications [33], [59]-[64], no known methods address the application of these techniques for changes in optokinetics resulting from variable lighting source positions and intensities.

This issue can be particularly problematic as it relates to the generation of high contrast stimulus on mobile device displays. Limited screen brightness results in constrained relative contrast which can be generated by these devices in outdoor illumination conditions. To assess the feasibility of real world implementation, our collection protocol implemented illumination conditions intended to simulate an outdoor environment.

2.19 Feature Localization

Ocular features localization with VW imaging devices presents several unique challenges. OKN response detection relies on high precision estimates of the iris center for reliable classification. Mobile device users engage with their devices in a variety of non-ideal scenarios which introduce significant variability in the parameters of the device as an imaging platform.

Detection of the small angular changes resulting from smooth pursuit based motion requires adequate spatial resolution. Constraints are therefore required to maintain a distance from the user which maintains the ratio of scale required for stable sequential estimation. Factors required to compute this ratio are dependent on temporal factors like sampling rate and the spatial resolution of the imaging sensor.

Changes in distance, angle or field of view as a result of movement of the device or adjustment in user posture increase the complexity of generating precise ocular feature localization. One aspect explored in this study as a potential factor to mitigate the impact of device or user motion is minimization of the required stimulus duration.

CHAPTER 3

EXPERIMENTAL DESIGN

3.1 Preliminary Assessment

Preliminary investigations focused on the assessment of mobile device screens to generate visual stimulus within the required parameters to evoke the OKN response. Multiple methods were implemented to generate the visualizations. Prototypes were first constructed using Adobe™After Effects video editing software by plotting the linear motion of a pre-rendered grid for each sample sequence. Later this process was updated by generating the frame sequences programaticly by generating a square wave with fixed 50% duty cycle in the length or width of the target frame. The frames were generated by replicating the output vector, rendered as images by plotting the resulting matrix as an image, and saving the sequence of images as the offset of the square wave was adjusted to generate the displacement of the grid. Four planar stimulus motion types were selected to mimic the rotation of a cylinder spinning clockwise and counter clockwise in both the horizontal and vertical orientation.

3.2 Hardware

Experimental collection recorded responses to stimulus presented on the screen of the mobile device. Preliminary testing of stimulus parameters was accomplished via analysis of motion information collected utilizing the open source Pupil Pro head mounted

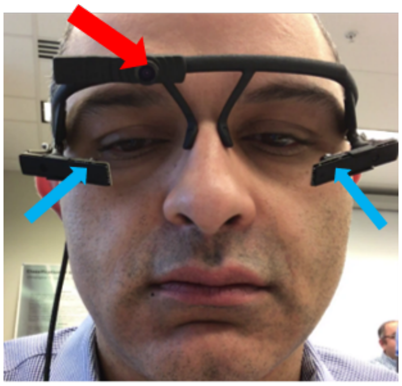

Figure 5: Pupil Pro Head Mounted Eye Tracker. World tracker camera (red) and IR eye tracking cameras (blue) are used to provide gaze estimates and visual overlay using the *Pupil Capture* application.

binocular eye tracking platform Fig. 5 from Pupil Labs [137]. The tracking platform also provided a ground truth for validation of ocular kinetic estimates derived from the recordings by software-based methods such as *Drishti*.

3.3 Software

Mobile device based OKN liveness detection requires precise and reliable feature segmentation and localization methods that operate with images captured using VW sensors. For viability assessment, our implementation targeted methods suitable for on device execution. Libraries and software methods with mobile deployable versions were assessed relative to their computational complexity, to insure realistic execution times. As this study was facilitated in part by a grant provided to investigate the potential of OKN for industrial implementation, the tool chain selected for the prototype was chosen to operate alongside applications already in use or development by our primary sponsor. Development of our OKN application took advantage of the highly optimized ocular feature localization methods implemented in the *Drishti* library [138].

3.4 Ocular Feature Localization

3.4.1 Drishti

The *Drishti* library combines an Aggregate Channel Feature (ACF) object detection library based on fast feature pyramid detection [139] using an OpenGL framework expansion . This part of the library is mainly about finding the main features that compose the face and generating bounding boxes for future search refinement.

3.4.2 Iris Ellipse Fitting

General refinement of high priority landmarks of the eye region include the ellipses created on the upper and lower eyelid. Segmentation of the eye itself is done by selecting the region enclosed by the intersection of two curves that generate one convex enclosed space. This process builds on from both Cascade Pose Regression [140] and XGBoost [141] regression.

3.4.3 Ensemble Regression Trees

Face landmark refinement is executed using Ensemble Regression Trees [142], which provide a fast methods alignment of the eye model based on a modified implementation of Dlib [89]. Global alignment is provided via the use of line indexed features, normalized pixel differences, and principle component analysis (PCA) dimensionality reduction.

3.4.4 Output

Localized features are generated for a single frame, or a sequences of frames depending on the input. Output is formatted as a structured json file which contains location and motion information for each eye and several other facial landmarks. Angular gaze estimates in a polar coordinate system are also generated based on normalized pose and distance as given by the apparent scale and ocular features.

3.4.5 Deep Learning Based Models

Additional testing and performance comparison of the *Drishti* model, and other feature extraction tools used, was conducted with cutting edge facial detection, alignment, and landmark localization tools. MTCNN [92] and Google Mediapipe [143] were applied to all sequence data, and stored as arrays, images, and video sequences for comparison and further evaluation. Architecture of these models was unaltered from published states to facilitate reproducibility.

3.5 Data Collection

While eye tracking based biometrics have been investigated previously [19,23,42, 66,101,144], no known studies have investigated the suitability of OKN based response using stimulus response for mobile devices. To investigate this potential novel application, we devised and conducted a study [145] which has been outlined in this section with the intention of establishing the viability of eliciting and analyzing the OKN response using the screen and selfie camera of a mainstream mobile device. Volunteers for the study consisted of 45 healthy adult participants recruited from the student body. Participants were provided with a brief explanation of the study objectives, but no specific instructions were given prior to the start of the assessment to prevent behavioral bias. Applicants were screened for ocular pathology, disorder or neurological disease via a self-assessment questionnaire to limit risk factors associated with the collection protocol. All applicants were required to provide written consent to participate.

Response based data in the form of video recordings were collected via the integrated front facing camera of the mobile device. In the case of ocular biometrics, as it pertains to liveness assessment, segmentation of small anatomical features relies on detailed image data. To facilitate an optimal assessment of our method's feasibility, image sequences were captured at Full High Definition (FHD) resolution in portrait orientation (1080x1920) to maximize the resolution of the ocular region of interest (ROI). Subject records and identification numbers were randomized at time of collection, and anonymized to preserve privacy.

3.6 Collection Parameters

Stimulus generated for display on the mobile device utilized simplified patterns monochromatic high contrast elements of variable spatial frequency and translation velocity. A total of 5 distinct of stimuli were employed including horizontal, vertical, and oblique gratings. An illustration of the structural and motion based parameters utilized is shown in fig. 6. Several methods were investigated to generate the visual stimulus, including direct playback of pre-rendered media files stored on the local device. This method produced inconsistent results due to access times and video player startup routines resulting in a lag of up to 150ms, well beyond acceptable tolerance. To maintain timing factors critical to this method, it was ultimately necessary to construct a custom application which utilized built-in features of the user interface to render the desired stimulus directly to the screen space via the integrated graphics processor.

Rates at which the rendered elements moved across the visual space were iterative refined to evoke a sufficiently strong OKN response. It was observed that the desired response was produced for stimuli with a displacement velocities greater than 6° per second. Including breaks and rests periods, our data collection protocol was designed not to exceed 30 minutes per session, and high velocity stimulus were avoided to minimize discomfort [115]. Regular monochromatic bands occupying approximately 2° of the visual space were displayed full screen in portrait orientation Fig. 6. Scale of the stimulus was selected to optimize visual frequency for the estimated 10° field of view provided by a mobile device screen at a distance of 20-25cm.

The first full scale collection attempted to elicit a strong onset based kinetic response by introducing a blank screen prior to the introduction of the target motion stimulus. Changes in illumination intensity of the screen space may generate some additional information in the form of pupillary dynamic response commonly seen as a change in pupil diameter as a response to direct light level stimulus. Distinct changes in pupillary dynamics have also been demonstrated to occur when the subject is engaged in searching

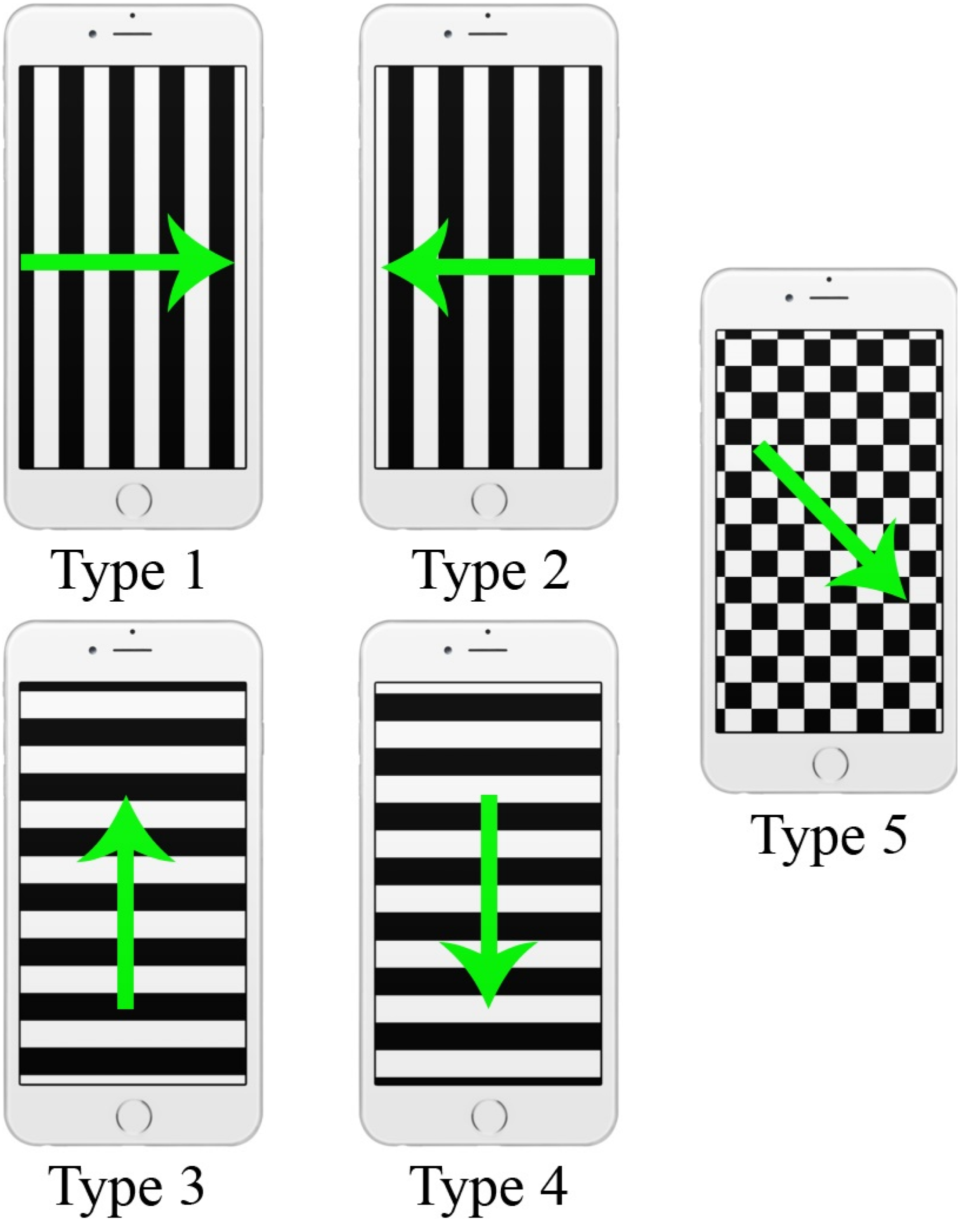

Figure 6: Visualization of the 5 types of stimulus used in the collection. Arrows denote the direction of flow for the continuous pattern.

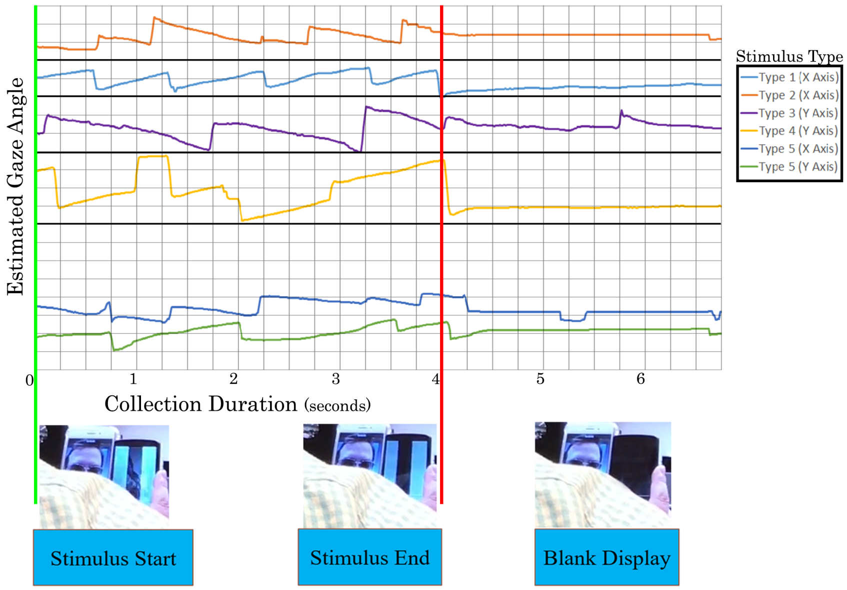

Figure 7: Collection sequence timeline showing a typical response for each major stimulus class

behavior for specific items in the visual scene, or when presented with an interesting or provocative stimulus.

Volunteers were instructed to direct their attention to the mobile device screen during the active collection phase of each collection set. A sequence sample shown in Fig. 7. At the conclusion of each collection sequence, a brief rest period was added to mitigate discomfort and reduce the impact of fatigue.

3.7 Experimental Setup

All collections took place inside a semi-secluded cubicle sized collection booth constructed to provide privacy to our volunteers and limit distractions and interruptions

Fig. 8. An enclosure made from opaque materials such as black felt cloth and plywood was constructed around the booth to control illumination conditions. A customized lighting solution was employed inside the booth based on digitally adjustable Philips Hue™Smart Bulbs. A total of 4 bulbs were arranged above, and in front of the participant to provide maximum coverage and adjustable illumination during response recording. A diffuser was placed in front of the bulb located directly in front of the collection booth, this helped eliminate discomfort while viewing the mobile device. Using the application provided by the manufacture, the intensity was set to maximum of approximately 800 lux and an emission color temperature of 6500 °K was selected to simulate direct sunlight. The primary purpose of this configuration was to provide optimal imaging quality from the camera device by insuring adequate illumination for the collection. Use of dedicated lighting provided consistent illumination conditions, reducing the complexity of feature extraction and ocular segmentation in the processing stages. Bright, direct conditions were chosen to minimize the impact of pupillary responses as segmentation of the iris can suffer in low VW conditions when the pupil is too heavily dilated due to low light. Outdoor color temperatures and intensity also helped evaluate the viability of limited screen brightness in the creation of adequate stimulus contrast.

3.7.1 Mobile Device Constraints

To provide as reasonable of a simulation of real world conditions under the collection constraints, participants were asked to engage with the device as they would in normal use. An adjustable mount was used to secure the mobile device, ensure the appropriate orientation, and minimize occlusion and vibration during the collection. These constraints were implemented to maximize the viability of the collected response data by removing the additional degrees of freedom typical in a hand held mobile device collection. Reducing the complexity in this way provides significant value in the processing stage, as the stability of capture sequences motion artifacts like blur, smearing and perspective distortion. No restraints or chin bar were used, allowing normal head movements and orientations. Volunteers were requested to sit comfortably, and move normally, but we asked to attempt to maintain a distance of roughly 20-30cm from the device during the collection Fig. 8.

3.7.2 Hand Held Collections

Some additional collections of response data were collected for additional application testing. These samples were collected by the sponsor of the study, and provided for evaluation of methods only. While the large scale, fixed device, collections utilized the generated collection application discussed in the next section the additional collections relied on an industrial testing application. Parameters of this collection follow on from the protocols discussed, but conditions were unconstrained. The additional 6 degrees of freedom Fig. 9 hand held collections introduce have proven to provide substantial additional challenge for motion feature extraction. Scale and pose variation require more precise levels of sequential precision, as the noise from both camera and visual target can easily be integrated into tracking samples. Many of the key improvements discussed in

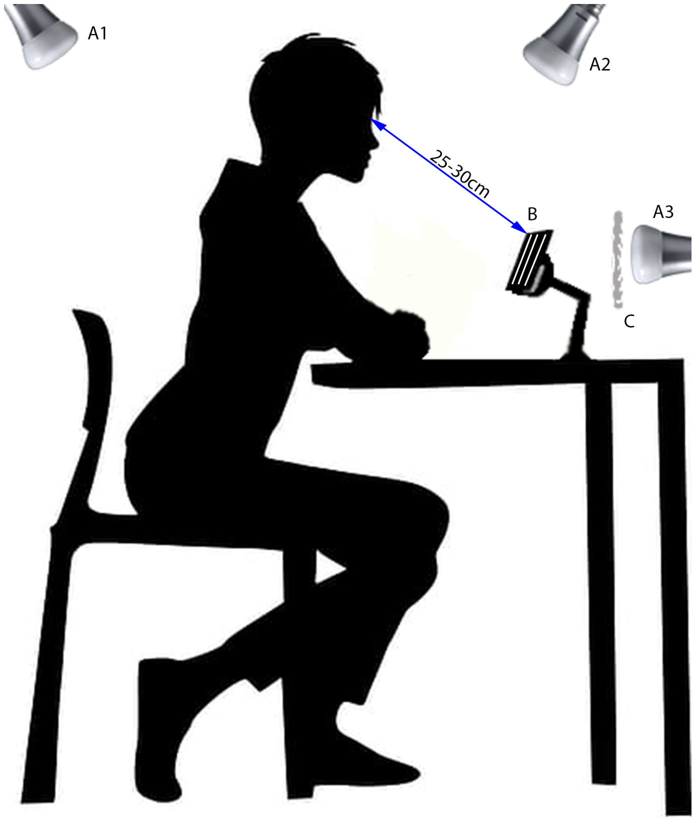

Figure 8: A diagram of the collection cubicle as used in the 45 participant study. Relative positions of Phillips Hue Lights (A[1-3]), Collection Device and Hands Free Mount (B), Diffuser (C).

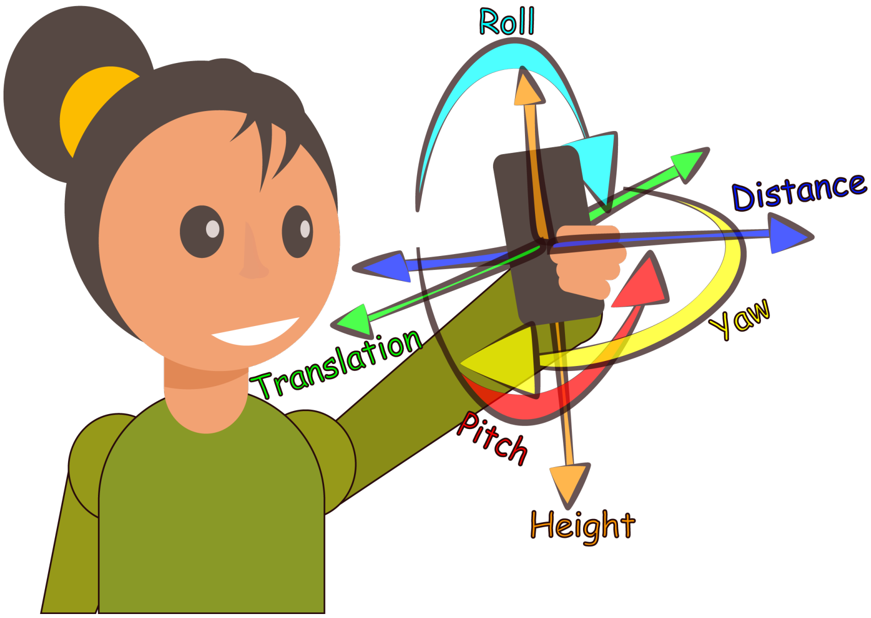

Figure 9: Six additional degrees of freedom introduced by hand held selfie response collection.

this dissertation focus on understanding and overcoming the challenges observed from these additional collections. Alternative stimulus parameters were also used in some of these samples, further confounding the direct comparison. All reported performance data reflects the use of the full scale data, but some results of the methods will be discussed as they apply to the hand held collections.

3.7.3 Collection Application

Stimulus presentation and response video recording were accomplished using the display and front facing camera of a stock Apple™iPhone 7. A customized application was designed and implemented on the device to automate the collection sequence. Stimului were rendered to the screen space using programmatically generated animations which employed the native layer rendering functions treating the contrasting display elements as discrete UI regions. Procedural generation allows the potential for parametric variability application to bypass limitations related to decompression and storage performance. splaying the animations on the device allowed for the maximum resolution and rendering rate provided by the device hardware. Frame capture from the front facing VW imaging sensor utilized direct device access, but required that a frame buffer be created each time the collection was initialized. Startup time for that process was variable in our application testing, so a delay to the display sequence was implemented using a callback the primary function que when the allocation and startup process was complete. Acquisition of the frame sequence must occur synchronously with the display of the intended visual stimulus. Achieving consistent and reliable results while running these two processes simultaneously requires the camera's frame is treated a priority. When initialized using a fixed frame rate mode, built in buffers can help compensate for potential capture delay or frame drop that might occur when acquiring frames on demand. Device firmware protection creates access restrictions that mean control of the camera's frame capture is treated as a protected process. Dereferencing of the software object generated by the cameras library resulted in an error which the interpreter couldn't address, so initiating the capture process from the applications main thread required the camera object to be implicitly passed by reference to the initialization process call. Developing an interlaced rendering and capture process with milisecond level operational consistency required utilizing thread control functions to more precisely orchestrate the the asynchronous operation structure of the iOS command que. Reliability is a critical factor to ensuring the goals of the study can be achieved using the resulting data. Risk factors associated with the collection protocol must be taken into consideration with respect to the quality of the collection application as failure of the device to collect usable sequences would result in wasted time, resources, and potential unnecessary risk to participants. Before the application could be utilized in the collection of response information from study participants, system reliability testing was performed to assure all operations occurred within the selected tolerance for repeated sample collections in a testing environment. Integrity of the frame capture intervals were authenticated by external sensors and device logs. Verification of replay frame rate was carried out utilizing the Pupil Pro eye tracker camera operating at 120Hz. Synchronization and frame display stability were confirmed by remote activation of the replay function with the device tethered to the host computer and an event monitor function of the Pupil capture software. Start time of the remote replay trigger was logged by the external video collection to determine average latency from the application trigger to the first full frame displayed on the device. using the timestamps returned by the camera object process. A fault state was implemented in the collection routine which generated a diagnostic message for events where frame capture intervals exceeding the explicit maximum of 35ms were observed. Intervals were measured using timestamps supplied by the camera device. This fault state was not encountered in the data collection, but could be triggered by forcing a display overlay. Collection stability, stimulus presentation velocity, and onset synchronization were tested. Results of stability testing are reported in the next chapter.

Collection of each participant's response data was initialized by selecting a marked icon in the collection application. The total stimulus response collection process was subdivided into 5 rounds for each set of the 5 classes of motion stimuli. Variation in the sequence of stimulus presentation was accomplished by the use of a shifted sequences which allowed each of the stimuli to be presented in each of the possible ordered sets. The method employed for our process utilized the Latin square configuration, a single shift sequence, to provide each stimulus type before and after each of the other types. Video collection was initialized by the collection app for each stimulus presentation. A single video file was saved to the device memory for each of the 25 visual presentations per collection round. To assure that a sufficient baseline was collected before each response, collections began 4.5 seconds before the stimulus display was initialized. Each stimulus was presented for 4.5 seconds, followed by a post stimulus blank display recovery period of 1 second. A brief 10 second rest period was added between each set of 5 response collections to reduce fatigue.

The data collection process consisted of four sessions conducted for each volunteer. Brief break periods were added between each session to avoid discomfort and minimize fatigue based artifacts. Each session was constructed of 5 ordered sequences, with each group order rotated 5 times, resulting in 100 sequences per participant for the full collection.

3.7.4 Human Subject safety and Biometric Data Security

All data were collected under the supervision of the University of Missouri Kansas City (UMKC) Institutional Review Board (IRB) protocol ID 17-360. The information provided here describes the provisions and steps used for data collection process, additional participate selection and safety criteria was established as part of the approved protocol.

Considerations were given to the possible impact of stimulus sequences used in the collection replicating a flashing type visual effect. Strobe type flashing is known to induce seizure activity in some susceptible individuals when the frequency falls inside the range of 5 to 30Hz. Participants were screened for risk factors or histories of photosensitive epilepsy, and made aware of the risk factors. Stability of applications and hardware used in the collection was tested, and fault conditions were implemented to limit unintentional and erroneous generation of strobe type events.

Head mounted eye trackers like the Pupil Pro device used in the response verification portion of the study are equipped with LED based IR illuminators for improved image quality and glint tracking. Some risk factors are known related to the exposure of tissues to high intensity light, however, based on information provided by the manufacturer of the diodes used in the device the total level of exposure falls far under established safe levels. Total time of use for the Pupil devices was limited to 1 hour to avoid any fatigue or discomfort.

3.7.5 Hand Held Collection

Some additional data were collected for use in evaluation of unconstrained mobile device collections. Allowing the participant to hold and position the mobile device during the collection adds several additional degrees of freedom relative to the

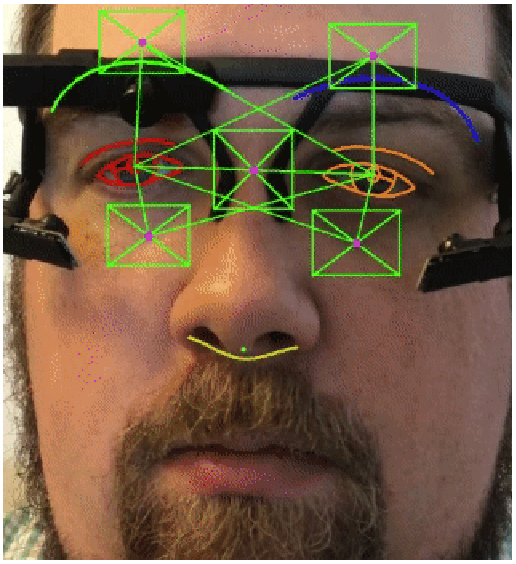

Figure 10: Patch Tracking Stabilization

3.8 Data Processing

Videos collected using the collection application were copied to a secure local storage device. Randomized alpha numerical sequences assigned to each of the collections for file storage labeling were eventually removed and replaced with simplified designations to maintain participant anonymity. Video sequences of the recorded participant responses were processed offline using the Drishti library [138]. Structured *json* files output from the library were imported and processed within the Matlab™ environment.

Gaze values and other facial landmark measurements provided by the main library and the subsequent *gaze.improvements* branch were compiled into a single ordered sequential cell database. Displacement values from the base library were calculated using an apparent scale derived distance which was implemented to normalize pixel level linear distance valuesFig. 10. Positions for each sample were measured from the iris center, point 26 in the index of landmark values, to the mean of three anchor points along the innermost points of the ocular ellipse given as points 7, 8 and 9 of the index. Values from the *improvements* branch provided direct normalized changes of iris position relative to the elements of the tracked patches of the face rig. Output from the models were processed and statistically analyzed to provided an array of values and measurements for the iris center. An explanation of the formulation and utility of the most significant values calculated, such as displacement velocity, acceleration, absolute magnitude, displacement,inflection, skewness, and kurtosis are given below. Feature vectors composed of these values were used to train some of the early ML models used in PAD classification.

- Velocity
  - Sequential change in absolute position derived by translation of the target feature. In visual localization this value is derived by the relative distance of the target feature from some other feature or array of features that serve as anchors.
  - Velocity is most useful for classifying slower motion ocular events. This feature is sometimes used in online classification methods, but there are limits to the accuracy that can be obtained with a single pass approach.
- Acceleration
  - Standard derivation of change in relative velocity. Errors in feature localization can highly impact the stability of featured at this level. Adaptive and

    contextual filtering were applied in some experiments, however, they tend to insert inconsistent delay for onset detection.

  - Discontinuity in feature localization can be a factor in calculating acceleration, as blink events will sometimes register as saccade type events.
  - Acceleration is the main method of identifying larger saccadic displacements as onsets are defined by rapid changes in velocity.

- Magnitude
  - Measured relative to the horizontal line generated between each of the pupil centers and the vertical median line of the detected facial pose.
  - Provided as an absolute value of combined binocular motion vector.
  - Captures motion of the combined planar displacement, more sensitive to total motion but less specific than other features.
- Displacement
  - Scale normalized moving window aggregate of total travel.
  - Variable window scale can be used for onset threshold in noisy samples.
  - Integrates significant error when used with sequences collected in low light conditions or with large camera motions.
- Inflection
  - A binary vector indicating the regions of detected saccades in the extracted feature sample.

  - Samples considered part of a saccade are marked as zero for the duration indicated by the extraction algorithm.
  - Intended to serve as a mask for removing saccades when training smooth motion tracking classification.

- Skewness

$$[moment] = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu^3}{\sigma^3} = \frac{E[X-\mu^3]}{(E[(X-\mu)^2])^{3/2}} = \frac{\kappa^3}{\kappa^{3/2}} \tag{3.1}$$

  - Skeweness provides a parametric assessment of the deviation of the distribution in some sample or set of samples from a centralized mean.
  - Samples with slanted distributions indicate a lack of symmetry in some values with respect to the population making up the represented set.

- Kurtosis

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E[X-\mu^4]}{(E[(X-\mu)^2])^2} = \frac{\mu^4}{\sigma^2} \tag{3.2}$$

  - Used to characterize some outlier events from the general set of extracted motion features.
  - Indicates the general moment of the sample relative to the mean supplying an assessment of the likelyhoold of a sample or region of a sample to contain values with lower relevance due to sampling error.

3.9 Deep Learning

Intelligence, be it natural or artificial, has a proven elusive to define, and is still a subject of contention for philosophers and researchers alike. One of the more broadly endorsed definitions describes intelligence as ”the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience” [146]. This definition both confines and highly generalizes intelligence by requiring an intelligent system to simultaneously possess the constrained capacity for contextual solutions, and the unconstrained novelty of abstraction. As a result, all modern computational methods (even the advanced and highly celebrated emerging techniques in ML), are woefully under equipped to be defined as intelligent. Humans, as a species, are typically accepted to manifest intelligence in some form. This isn’t as trivial of a statement as it seems, as the processes that underlie consciousness and sentient self-awareness aren’t fully understood. The statement that humans possess adaptability (manifesting as a capacity to contextually adjust behavior to a wide variety of stimuli), is perhaps more precise as it is one of the chief contributing traits of animal survival. In biological systems, intelligence (or adaptability), is understood to originate from physiological phenomena [106, 146]. Within neurological tissues, the structures which generate high order adaptive capabilities are sequential layers of processing units which generate responses to sensory stimuli by varying their sensitivity to intensity and frequency [106]. Morphologically, these adjustments are accomplished by direct modification of the physical member cell. These networks have a huge numbers of processing elements, organized in densely packed structures, and possess various modes of connectivity between elements. One structure might rely on direct connections between several sequential units to process and integrate sensory data, while another parallel network transforms that output by selectively inhibiting elements such that only high priority information is propagated to the output. These structure have been shown to possess the ability to process massive amounts of sensory data, separating the essential from the trivial, with remarkable speed, accuracy, and reliably [107].

Attempts to computationally model intelligent behavior such as large scale pattern and behavior recognition have resulted in a series of architectural modification to the standard Artificial Neural Network (ANN) model. One of the major factors that needed to be addressed was the decay of gradients which occurs in long chains of back propagation for updates of weights used to define features the networks extracted from sample data. Any sufficiently complex fully connected ANN will eventually suffer from the impact of gradient decay, making the rate at which the values are modified so small the impact can't be differentiated in subsequent iterations. Given that the power of functional approximation is linked to the number of sequential layers utilized in the architecture, gradient decay is a primary limiting factor to the application of ANNs. Complex recognition tasks like feature extraction systems used in biometrics require especially deep structures to adequately model the feature space seen in morphologically diverse structures like human facial and ocular features. Ultimately it took several generations of computational hardware advancement, and the re-purposing of graphics acceleration hardware, to overcome the limitations of traditional ANNs. DL based adaptation utilize greedy module level optimization that allow updates to be derived from the total output activation [147].

3.9.1 Convolution

For the most part we as humans interact and interpret the world through vision. For machines, digital images approximate this sensory capacity and offer a discretized representation of the physical world. Virtually unsurpassed in information density, pictures now represent the bulk of data in existence. With the proliferation of mobile devices and other handheld electronics the bulk of computer processing power will soon be spent processing visual data. While there are a variety of time tested ways to process and extract significant features from images, the recent innovations in Convolutional Neural Networks (CNNs) have come to represent the standard. CNNs have a relatively specific function, but a broad array of applications in DL [148]. The convolution of an input can take place in arbitrarily dimensions and can be greatly accelerated via GPGPU parallelization where it can be processed as a multiplication by transformation in the Fourier domain [148–150]. While there are some special cases, most of the widely successful CNN implementations so far have focused on 2 dimensional data. Higher dimensionality is easily generalized by comparing additional dimensions as channels of an image, such as frames of a video stacked to generate a volume as seen in Fig. 11.

CNNs output can be viewed as the region wise dot product of a filter (kernel) matrix with dimensions [m x n] in the 2D space, or [m x n x k] in the 3D space, evaluated on a region of the same scale. The filter kernels are symmetrical in most cases [m = n] but the temporal dimension k, can adjusted independently of the spatial structure. Values of the filters are initialized as a zero-mean Gaussian distribution at the start of the initial training to encourage an optimal distribution of kernel values relative to the feature space.

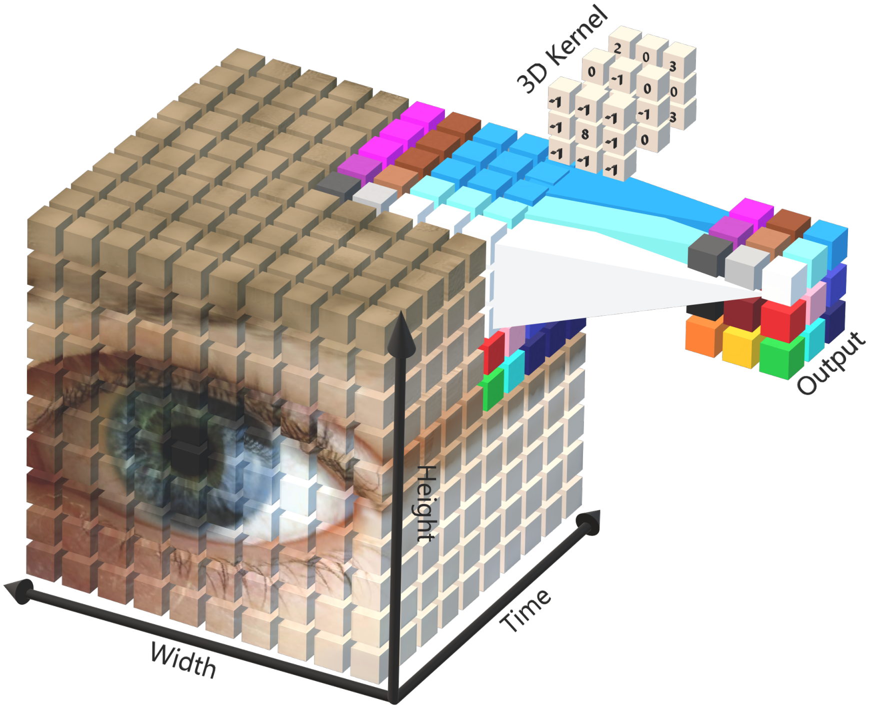

Figure 11: A visualization of convolution in a 3D space. Kernels such as the one shown above the 3D array are adapted to extract features in both the spatial and temporal domain. Stride motion of the kernel is visualized by color coding relative to the partial output array.

The output of each convolution operation is a single scalar value which is stored in an output tensor with a size less than or equal to the scale of the input array or tensor. At each step a new region of the image is then selected and the filter evaluated. The location of the new region is based on the scale of the *stride* with which the center of the filter moves across the input array. Selection of a stride less than the scale of the filter can generate adjacent evaluations of overlapping spatial or temporal windows, allowing a more robust search which allow for the more precise detection and localization of target features. As the filter continues to make its way across the image, each value is passed forward to the processes which will prepare the output.

Dimensionality reduction between convolutional layers is necessary to maintain learning rates and maintain issues with gradient optimization (such as vanishing or exploding gradients). Pooling is one of the most common module elements used to accomplish the necessary consolidation of values. The values generated from convolution have a high spatial correlation, so despite the fact that much of the spatial data is lost, performance isn't highly impacted in terms of feature localization. The value of the input matrix is divided, evaluated, and aggregated by the function, as prescribed by the user and then passed through a linear or non-linear activation function on its way to the next module. The goal of any given output in ML is to provide the most critical data relative to that factor. Toward that goal, the information accumulated by an individual neuron in an ANN should discriminate between real, significant information, and diffuse noise. As in biological networks, this is accomplished by establishing a threshold of activation [106, 151]. Traditional ANNs use activation functions such as hyperbolic tangent functions, which can require large amounts of sequential input to reach saturation. To overcome the issues inherent in DL, several methods of activation have been developed, or repurposed for the challenges of deep structures. One of the more commonly utilized methods of insuring rapid adaptation to the input has been the Rectified Linear Unit (ReLU) [152], which when properly applied can account for many common issues of poor kernel response in image and signal processing. As mentioned before the relationship between over-fitting and training rate is a major issue in DL. It can be difficult to optimize a system when training a new structure can takes weeks or months, so any algorithmic (as compared with hardware based) method which allows the system to start learning faster can provide a huge advantage. ”ReLUs have the desirable property that they do not require input normalization to prevent them from saturating. If at least some training examples produce a positive input to a ReLU, learning will happen in that neuron [149].”

3.9.2 Recurrent Methods

In early experiments a variety of recurrent configurations were applied to the features derived from the library feature extraction pipeline. Some of these results are discussed in the next chapter. Attempts were made when training these models to keep the total number of parameters low. This was partially to maintain feasibility for the deployment of the models on a mobile platform, but also to avoid over fitting. Smaller models are also forced to learn more robust representations, leading to better performance in the wild.

Recurrent Convolutional Nerual Network (RCNN) models Given that precision plays a large factor in the quality of OKN classification models, the resolution and depth of the input sequences generated by the collection created a training batch scale limitation for the commercial grade graphics hardware available for the experiments. Tensors of ocular ROIs were eventually generated using an automated extraction process. This reduced the scale of the input sequence enough to allow larger batches to be loaded into the GPU memory, but the

3.9.3 LSTM

Sequential models like the Long Short Term Memory (LSTM) build off earlier Time Delay Neural Networks (TDNN) and but have several unique additions like non-saturating activation functions that allow for training with substantially large and diverse data sets. Structural improvements like regulated side memory states and parallel operational execution also allow for better acceleration with modern GPGPU hardware.

Sequences of temporal features generated by our processing pipeline were classified using a recurrent deep learning approach based on the Long Short Term Memory (LSTM) model. LSTM models were principally designed to overcome limitations of vanishing and exploding gradients that commonly occur in traditional Recurrent Neural Networks (RNN) when processing samples with long sequence lengths. Like RNNs, LSTMs utilize feedback provided from previous time steps as part of the input to subsequent states, but improvements in their architecture Fig. 12 allow LSTMs to selectively retain contextually relevant aspects of past elements in working memory. This selective memory aids in the detection of patterns that contain some variation in the frequency or

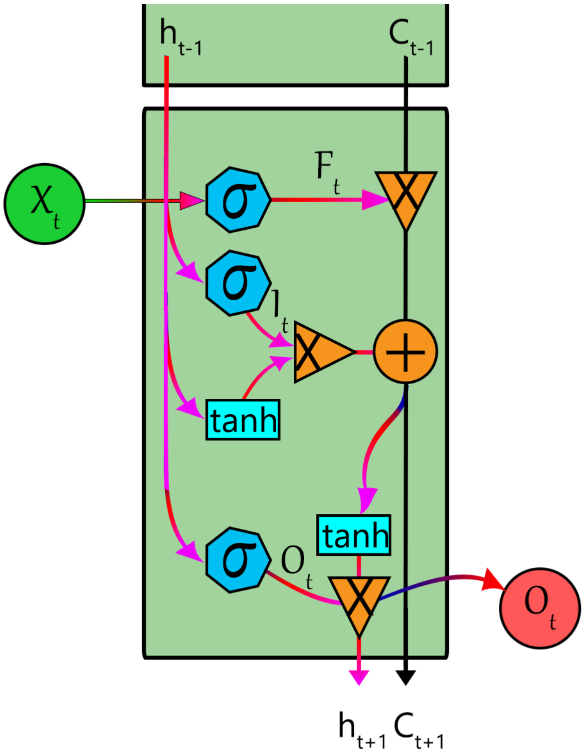

Figure 12: Structured model of a processing cell as used in recurrent Long Short Term Memory (LSTM) Networks

duration of key features. LSTM models have been shown in several studies to perform exceptionally well with the detection and classification of behavioral patterns and anomalies embedded in sequential sources [69, 153].

Training the model involves supplying both the features of the current time step and sequential state memory to the recurrent LSTM cell. Processing occurs in four stages called forget, input, update and output. A diagram of the generalized LSTM architecture is included in fig. 9. Each of these stages employ some pointwise operation and either a sigmoidal (σ) or hyperbolic tangent ($\tanh$) activation. The action generated by these nonlinear activations vary at each gate.

Sequence processing begins when new sample elements, along with elements retained from a past iteration, are passed to the LSTM cell forget gate. The value of this sigmoid activation determines to what extent the elements from the past time step are retained based on the data provided by the past state and current time step. Operation of the forget gate

$$Ft = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \tag{3.3}$$

provides an update F_t to the cell's state C_{t-1} by assessing the hidden state h_{t-1} and the current input X_t . If an activation passed to the cell state at this stage is set to zero, the element is removed from the network's memory and effectively forgotten. Since we've decided what to keep, we now want to determine what new information we want to add based on the hidden state and current value. Cell state value candidates are first selected from the input and the update is calculated and stored. An index of the input

$$I_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f)(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{3.4}$$

and the vector containing the values of the update candidate

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, X_t] + b_C) \tag{3.5}$$

are once again calculated based on the hidden state and current input.

Updated values are then stored as part of the new cell state

$$C_t = F_t * C_{t-1} + I_t * \widetilde{C}_t \tag{3.6}$$

by combining with the output of the previous step of the memory pipeline. At the final stage of processing, the cell's state is saved and the output is calculated. A mask

$$O_t = \sigma(W_O \cdot [h_{t-1}, X_t] + b_i) \tag{3.7}$$

is first calculated and then combined with the activation as

$$h_t = O_t * \tanh(C_t) \tag{3.8}$$

providing a scaled and filtered version of combined data to be used as the sequence output, or as the hidden state of the next iteration.

Providing an adaptive sequential memory allows the architecture to develop a high level approximation of the underling function, leading to a model with increased capacity to generalize sequential events with variable frequency characteristics. By learning what elements of the past sequential data are most relevant to carry to the next iteration, the model performs a type of automated temporal filtering and feature extraction. Limiting the calculations to a memory state and the current input also allows the model to avoid the pitfalls of long chain derivatives that can occur with other recurrent methods. Learning features in this way aids in sorting out temporally shifted or jittered elements which occur due to behavioral variations. As initial assumptions are generated, selected sequential features are committed to a semi-persistent memory allowing the model to retain and contextualize longer duration events such as smooth pursuit motions while retaining the ability to recognize high amplitude saccades.

Our experiments generated two potentially viable LSTM based models

- Multi-class model designed to identify spoof data as a specific class and differentiate specific response types for the full set of samples.

- A binary output model which differentiates samples of single response from spoof samples.

Both models were trained using 2 second samples from before and after the onset of stimulus. Two LSTM layers were utilized in sequential mode with a hidden unit size of 25 and 15 respectively. A dropout layer of 20% was incorporated for the output of each layer to minimize over fitting. In the multi-class model the output from the second LSTM layer is provided to a fully connected layer containing one node for each of the 6 output classes (5 response classes and one negative sample class). A softmax layer provides the output for a classification based on the activation state of the nodes in the fully connected layer. A comparison of the results obtained from the trained models and other traditional pattern classification methods is provided in chapter 4.

3.10 Quality of Approximation

Video data are essentially an array of snapshots that provided some structural or temporal representation of some discrete set of states. Extracting only the most relevant features from complex structures moving through space and time requires an understanding of syntax and context. As it relates to OKN, extracting behavioral information relies more on detection reliability and continuity of feature localization than on structural accuracy. In the case of biological structures, localization of feature in planar space is ill posed due at least in part to the variation in pose and lighting conditions. Differences in individual structural morphology further confound the segmentation of features based on 2D projections of the 3D facial structures. Refinement of representations in ML are usually achieved by leveraging high entropy of distributions. Large sets of labeled samples are required for supervised ML techniques which can become problematic for sequential data due to the effort and cost involved in annotation. Even when labeled data sets are available, some of the loss integrated into feature updates will be directly related to annotation inconsistency or human error. Implementing batch normalization, response pooling, or drop out can help avoid overfitting, but even with fine tuned models significant ambiguity is present in the sequential approximations generated from the segmented features. Ocular behavior classification performance is directly related to the precision of localization models. Errors resulting from unstable sequences of key points generated by naive treatment of the frame sequences become embedded in the temporal features based on the approximations of gaze angle. Obtaining the desired performance means treating context as a sort of ground truth such that the loss functions that govern landmark placement will prioritize stabilizing jittered features as part of the error minimization. A variety of leading cost and loss functions are implemented in adaptation training in an attempt to achieve this objective.

The flow of structural elements has far more significance than the morphology of the specific elements. Much of the significant information generated by analyzing optical flow can be used to spatiotemporally constrain what must be uniform translation or deformation of the face. While treating each frame as an independent element is much faster when using optimized feed forward pattern recognition systems, single frame approximations tend to perform greedy minimizations that produce less than desirable results. Larger elements, which are less likely to undergo instantaneous change can be used to keep landmark systems honest, but given that the error can occur at any stage of the system, regression is the best approach for reliable fitment of complex structural motion. Simple landmark mesh deformation constraints can be used, but pose or illumination variations are difficult to account for using this method.

Recurrent ML models provide the learner with some concept of time and are structured to encourage more robust understanding of causal structure. Most recurrent approaches, like RCNN and CNN-LSTM models, still rely on features extracted from the individual frames of a sequence, independently extracting features from each fame. While some aspects of localization error can still be learned by propagating error to the feature extraction layers, vanishing gradients tend to lead to slow or failed convergence. Truly temporally dependent ML systems have outputs contingent on both all sequential states. Additional memory is required to store the relevant contextual elements at each processed time step, meaning a recurrent network will likely require more resources and have greater computational complexity than an temporally agnostic models. Resource constraints limit the use of some fully temporal models on mobile devices. Some observations relevant to the applications of these models is provided in the discussion in chapter 5.

CHAPTER 4

METHODS AND RESULTS

4.1 Overview

This chapter provides an explanation of some models, methods and designs used in the classification of the data collected. Quantitative performance is reported for some models, and qualitative assessments of each method are provided in the next chapter. Parameters related to the data collection, processing, and feature extraction process are provided for the purposes of explanation and critical analysis. For clarity, only the most relevant results from each of the experiments detailed are included in this section. Some additional considerations and possible future work related to these topics are discussed chapter 6.

As the processing methods rely on temporal considerations, the results will be reported with respect to the sequence duration used to produce the features, localization approximations, and ML model used in the analysis. A variety of metrics have been applied to results of the models developed in the study. Accuracy of the classification is among the most commonly reported metric, as the main focus of the work was related to liveness of the sample provided to the classifier.

Descriptions of the software and processing methods utilized are provided for comparison where appropriate to the discussion of experimental results. Reporting based on experimental results is intended to highlight both the strengths and shortcomings of the proposed design. Some instances of less than desirable results are given with the intent of encouraging the continued exploration and related work. In some cases, additional processing steps were required to rescale the image sequences for reasons of both computational and algorithmic constraints. It is noted that in some cases the reduction in scale exceeds the required spatial resolution to accurately identify the target response.

Segmentation accuracy is assessed only as it applies to methods of using the full image resolution or in experiments related to localization refinement. Some enhanced methods of feature extraction and landmark localization tested to provide a competitive baseline are reported separately at the end of this chapter.

4.2 Collection Application Performance

Feasibility of this novel OKN based liveness detection for mobile deployment is demonstrated by eliciting and collecting response data directly on the device. Processing and training of the experimental models was conducted offline, however, all libraries used for localization and training have been extensively tested for mobile deployment. Classification results reported here reflect models trained using evoked OKN features derived from mobile device video sequences recorded in our collection.

Positive samples extracted for each stimulus class were collected for a duration of 2 seconds (60 frames) from the onset of stimulus. Extending or shifting this collection window was noted to increase the incidence of blink events in the resulting sequence.

The automated data collection application used in our experiments was designed to begin each recording 4.5 seconds before the onset of stimulus. Starting the recording early allows conformation of frame collection latency and insures the image collection pipeline is fully engaged before the stimulus begins. As the primary goal of our experiment was to determine the efficacy of temporal feature extraction for identifying specific motion types, the video sequences collected before the stimulus were selected to act as negative samples. Since these sequences contain normal eye motion collected in real time on the same device, they present a distinctly nuanced challenge that mimics a highly sophisticated attack. Subsamples of these videos were extracted using a random window approach. In each sample case, an equal number of negative and positive samples were selected from each user. A total of 9000 samples were generated from 4500 video sequences containing 20 responses per subject to each of the 5 stimulus classes. Samples utilized in our binary response classification experiments focused on detection of response to the oblique class of visual stimulus which was designed to elicit a simultaneous vertical and horizontal OKN response. This dataset was comprised of a total of 1120 samples, 800 of which were utilized for training and 320 reserved for testing. Multiple experiments were conducted to determine the efficacy mobile device collections for OKN based PAD using the library derived and hand crafted features described in this section. Additional experiments were also performed using direct processing of extracted sample sequences using automated feature extraction and classification methods. and explanation is presented in the following section of the top performing model and its associated parameters and selection criteria. Experiments conducted in this study were designed to evaluate the visual stimulus parameters of OKN sequential and response. Estimates of the ocular movements generated by the visual stimuli were provided by specialized feature extraction and localization methods. Several factors which have been noted to impact the quality of those sequential response estimates are reported.

4.3 Classification

Multiple experiments were conducted using a variety of traditional and cutting edge techniques. An array of ML methods were trained using library extracted landmarks and motion estimates which were processed through hand crafted feature reduction and statistical distributions based methods. Several classes of automated sequential representation DL classification models were trained using both raw image and model extracted ROI sequences. Performance evaluations provided in this chapter focuis mainly on models trained using data derived from gaze sequence estimates provided by specialized ocular segmentation and localization libraries, but viability estimates are provided based on the preliminary testing of the more advanced techniques.

The array of motion features extracted from gaze estimates, as described in the previous chapter, highlight pursuit type motions which have been deemed more reliably observable at the capture rate of the mobile device. Movements analyzed in the sample videos sequences were the result of parametrically derived visual stimulus displayed on the mobile device used to collect the samples, and more natural motions generated by undirected visual search of the blank display. In some cases, a simple single frame display attack was simulated for verification of reliable motion extraction, however, sequential classification used features derived from natural motion as negative samples for detection of the specific OKN event onset.

DL based approaches which required limited additional pre-processing, but substantially more preparation and computing resources. Classification performance for these models is reported with respect to the optimization parameters employed in their training. Some factors of computational complexity and memory capacity that limit their implementation on the target platform in the current stage of development, but insights gained may help improve future analytical and model based methods.

4.3.1 LSTM Classifiers

Subsets of response samples were selected from the full array of sequences and used to construct several specific classifiers. Each of the five stimulus types were tested against a randomly selected, equally sized subset of negative normal response (spoof) samples. Time shifted samples were used for verification of onset detection. Top performing models of phase-locked OKN onset achieved a detection rate of nearly 98% for combined data from a single stimulus type, and a rate of nearly 95% when applied to reserved test data. A confusion matrix of the combined output of this classifier is provided in Fig. 13. Further results are listed in 1 for the most promising of our experimental configurations.

Models for multi-class detection, where response samples from each stimulus type were considered as independent classes, proved slightly more difficult to train due to a high negative sample ratio and subsequent imbalance based degeneracy. A random subsampling approach was employed to offset the imbalance in several trials, however, the

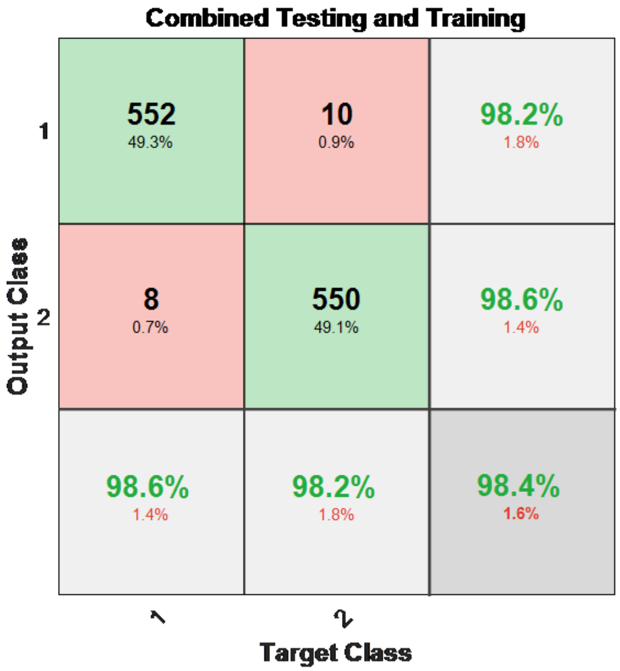

Figure 13: Combined results confusion matrix for single class best performing LSTM model trained using extracted motion features.

results of the testing applied to the remainder of negative examples indicates more samples may be required to model the variance observed with undirected ocular motion.

Estimated total processing time for a single sample conducted off-line was 650ms. Total run time of the deployed pipeline on a mainstream mobile device is estimated between 2.4 to 2.7 seconds as parallel execution of collection and processing would allow feature localization and sequential classification to occur alongside of the sample collection.

4.4 Optical Flow Based Tracking

Selected samples from the full scale collection, and subsequent hand held collections were processed using Gunnar-Farneback Optical Flow (OF) algorithm. Resulting arrays of flow vectors were used to generate maps of the resulting segmented regions of the eye and limbic boundary. Images indicating the general result are indicated in Fig. 14.

Clustering of displacements relative to normalized pixel scale indicate the mean ocular motion phase for large regular gratings is approximately 5 pixels or 2.5°. This finding corresponds with tracking data provided from conformation testing from the Pupil tracking software, and prevailing literature. Segmentation of the flow field vectors using texture and contrast base edge detection and best fit, scale constrained, circular Hough transforms resulted in motion vectors related directly to the limbic boundary of the iris. Limbic only vector based OF field data was used to perform a background agnostic displacement vector sequence, but ROI jitter was found to embed substantial additional motion features in extended testing. Ultimatly, optical flow fields were used primarialy to

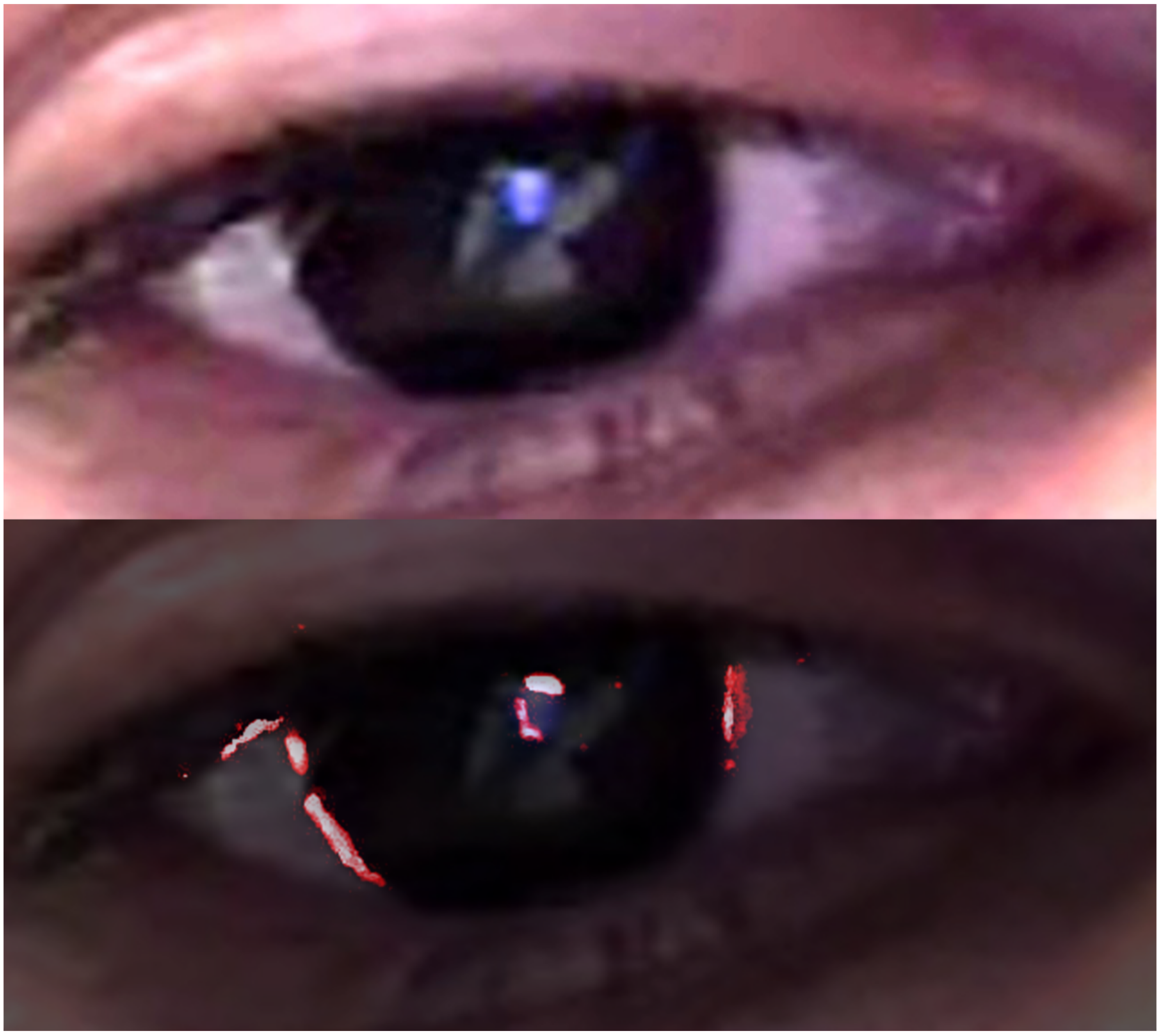

Figure 14: An example images of a sequence used to compute the optical flow of the iris from a full cycle of the OKN response. The input image (above) has been superimposed with a heat map of the flow (below).

stabilize sequential estimates provided by facial landmark detection systems. Bidirectional flow Kanade-Lucas-Tomasi algorithms were tested to provide optimization for the regional motion vectors, and compared as an absolute difference in localization for the detected landmarks relative to the tracked region of pixels in the target sequence. While results from the testing aren't definitive, due to lack of ground truth motion features, OF provided the best results for landmark stabilization for the samples and methods applied in these experiments when additional sensors or viewpoint data isn't available.

4.5 Facial Pose Variance

While the mobile device was securely mounted in the large scale collection to minimize some factors of pose variation, some amount of re-positioning and drift were noted during most collections. While the angular changes weren't sizable in the majority of cases, the apparent pose of the head is a necessary component of gaze angle estimation. Using the patch tracking approach from the gaze improvements branch of the Drishti library, ocular phase was reported relative to the rig of tracked points. Instantaneous displacements reported by this method were equivalent to linear iris translation rather than angular approximations relative to the device screen space. Sequential and recurrent methods used features derived from these relative displacements to classify smooth pursuit events based on windowed sequences with fixed velocities occurring before high velocity events. Window widths were varied between 3 and 5 samples with a stride of 1 to 3 samples. In the large scale collection a width of 3 frames with a stride of 2 frames was used to provide faster detection of the higher displacement onset and saccade events.

All approximations for the sample displacements were normalized with respect to a scale estimate derived using the average iris diameter of 11mm and interpupillary distance of 62mm. Selected frame sequences were labeled using inference annotations provided by the pretrained MTCNN described in [92]. Performance of this model was evaluated for use on both mounted and hand held selfie sequences. Sequences were stabilized using full frame OF methods as described in the previous section. Landmark localizations used to compare total accuracy were derived using the 64 point high-quality *dlib* detector combined with *eos*, and deformable facial face mesh and pose estimation of Google MediapipeFig. 15 [143]. Visualization of the dense array as detected by Mediapipe can be compared against the 5 point detector shown in Fig. 17. It should be noted that comparison of landmark localization only provides a relative assessment of performance and not overall accuracy. Generalizing the temporal stability utilizing mean displacement of sequential arrays of localized landmarks was deemed the best method of evaluation.

4.6 Convolutional Models

The combination of the three major elements above, convolution, pooling, and activation compose what is widely called a module or layer in DL. The label layer can be a bid confusing as it is use by different sources to refer to elements of different scale. In general, the notion of a module originates from considering the elements normally connected to generate one functional subunit of a DNN. In most cases, a DNN is constructed using several modules of the same type connected in series such as the toy robot network shown in Fig. 4 which consists of two convolutional modules connected in sequence.

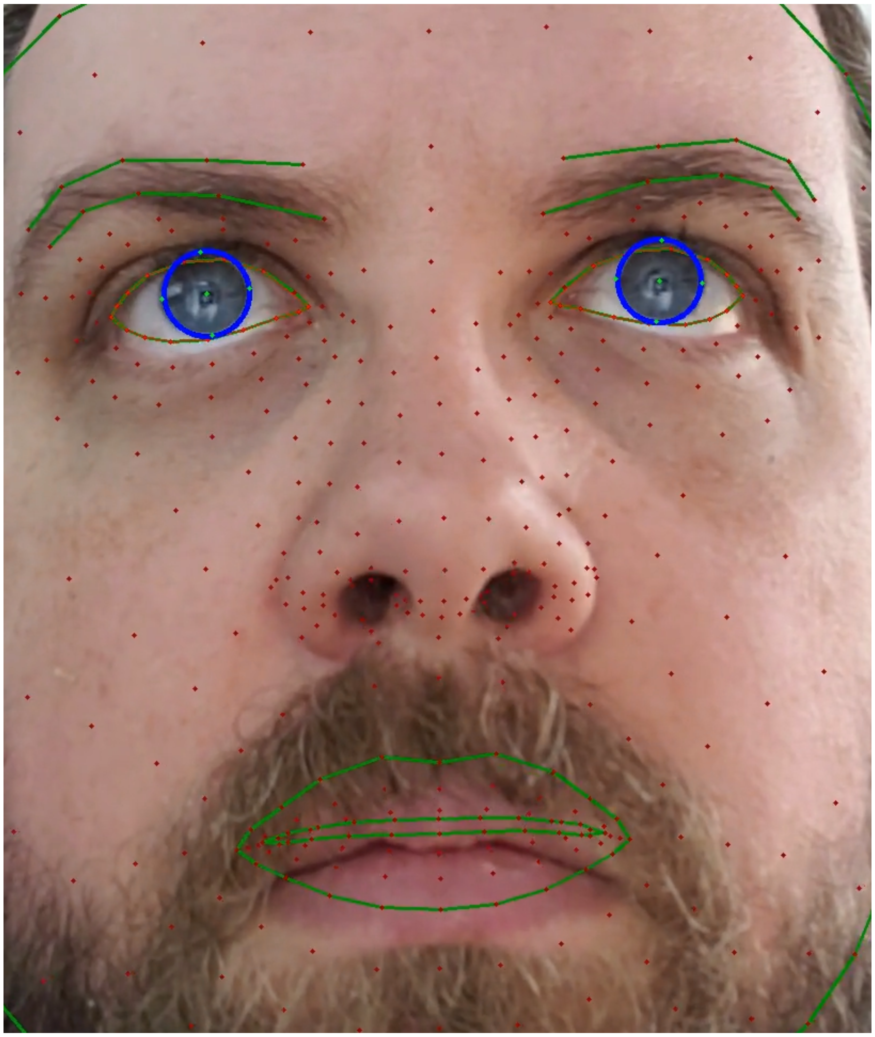

Figure 15: Face Mesh Generated by the Google Mediapipe library.

Additional parameters are often used between modules, such as dropout, and regularization functions, these parameters are invaluable for some types of data to avoid the constant scourge of data memorization. Not every DL software platform offers the ability to implement these operations at every module or layer type, it is important to choose a platform which allows the functions best for your application (more on platforms later).

Sequential precision, a measure of the total residue of alignment defects using a geometric transform based on the mesh model fit parameters, is noted to likely provide a more realistic assessment of total utility as part of an ocular motion based PAD system. Precision as measured in the selected subset of 25 sequences is reported in the table 2.

4.7 Spatiotemporal Models

A model consisting of stacked 3DCNN layers was constructed for end to end image sequence classification. Separate models were used for full frame sequences and a range of MTCNN derived ocular sequences cropped from each sample. Due to memory limitations, batches of full frame images were extremely limited, and training stalled after only a few epocs. Learning rate adjustments and different scheduling methods were used, but results were mixed. Ultimately an additional GPU was used to augment the memory allocated. Visually representing the input data and model structure of a 3DCNN is difficult, but a rough approximation of the model is included inFig. 18.

A volumetric representation of the input data tensor, based on a crop of the eye band, is included in Fig. 16. The top performing model from this selection had an equal error rate of 92% for stimulus based PAD from selected oblique responses. Data used to

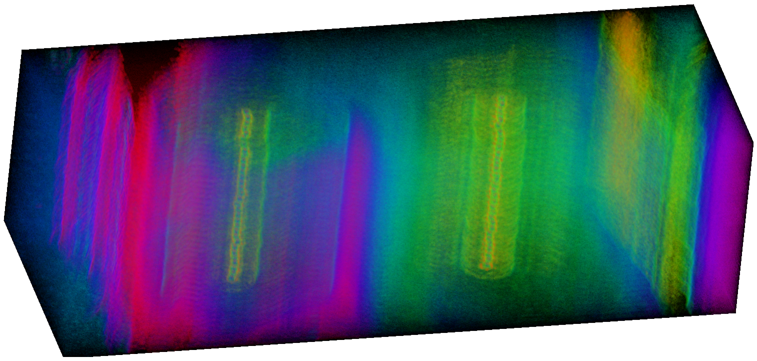

Figure 16: A typical input tensor composed of a cropped video sequence.

train these models was extracted based on sequential estimates of the MTCNN model as part of the pre-processing pipeline.

4.7.1 Model Architecture

General parameters and design for the model used in this study are derived from experimental modification of designs and ideas proposed in other human motion based studies as seen in [154, 155]. Input parameters are an important part of design for most CNN based methods, due to the fact that input structure is directly linked with the total number of free parameters and the effective receptive field of each feature extraction kernel. Once the input size is selected, only images of equal or lesser size can be processed using the pipeline and representations derived. Each stack of images used contains some ocular motion as can be seen in the volume projection displayed in Fig. 16.

While generating a model with variable input scale is possible through the use of padding operations, a uniform scale is generally desirable for targets structures with uniform characteristics such as the eye. While the scale of the input images was variable

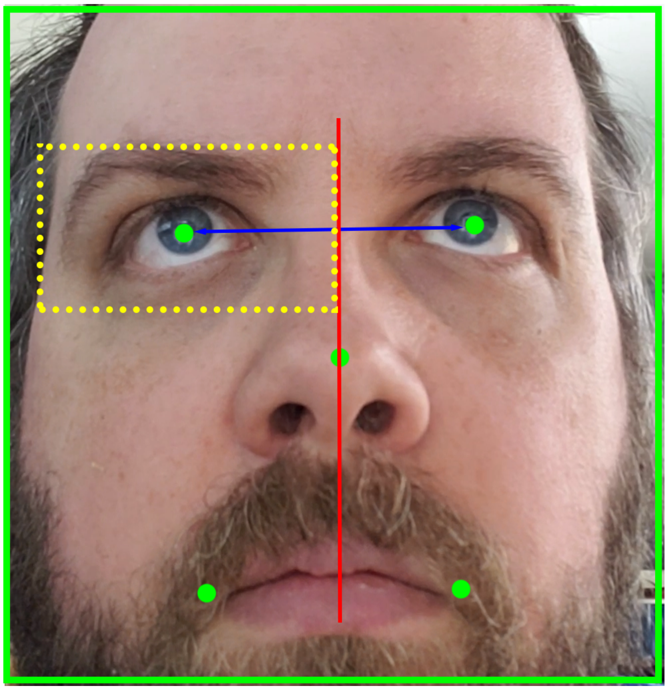

Figure 17: MTCNN detection framework applied to a sample image. Green elements indicate the face detection frame, and 5 point landmarks. The red mid-line is used as the inner boundary for the ROI as shown in the yellow dotted box.

in the extracted ROI samples, the deviation in total scale was less than 13% or 213 ± 28 by 134 ± 14. Image sequences were stacked to the nearest relative center pixel of the largest image in each sequence. The resulting tensors were rescaled to 190x120 pixels. A total of 3600 sequences were generated using the MTCNN ROI extraction method Fig. 17.

General tensor stacking size is relative to the engine for the layer input. In the case of 3D Convolutional layers the input tensor has a dimension (batch size, sequence length, channels, height, width) for torch's 3DCNN layer input. An additional tensor dimension is necessary for for sequence number in the tensorflow interface which adds the normal input direction to indicate which dimension of the input should be processed first.

This information is given as models of the 3DCNN provided here were built and tested in both frameworks, with performance differences relative to acceleration parameters in each layer library.

Batch normalization was employed along with a max pooling approach to minimize the impact of small batches relative to a high free parameter density. Dropout was assigned to each pooling layer in initial trials, but was removed due to poor learning performance. Rectified Linear Units (ReLU) were chosen as the unit activation function for each module to offset activation saturation due to high feature density. ADAM was chosen as the primary optimizer for all tested architectures based on adjacent literature and a search of top performing models on kaggel and modelzoo.

Final model parameters were selected through iterative implementation to achieve the smallest viable architecture by adding layers only when training of the previous design stalled for more than 50 epochs. Models were trained for 500 epochs for 0.1, 0.7,and 0.025 selected for the initial learning rate. No fixed scheduler was added for initial viability testing. The structure of the top performing multi-class model is supplied below.

Visual representations of the true dimensionality of a 3DCNN are difficult even with advanced animation tools as each kernel generates both a spatial and temporal array for the convolutional output of the sequence. No known visualization standards have been proposed, but a volume representation has been observed as most common for publication. A conceptual rendering of the small feed forward 3DCNN used in this experiment is included in Fig. 18.

Early stopping criteria of increasing loss on 5 or more validations and long term

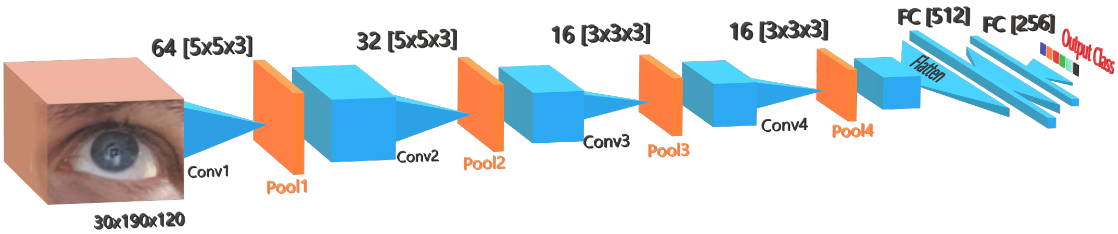

Figure 18: A representation of the 3DCNN model similar to the one used in this study.

snapshots logging the last 50 model states were implemented during training. Due to a very large number of free parameters, models of this class are prone to quickly overfit, resulting in diminishing performance for validation and reserve testing samples. Performance of the best model was determined by selecting the top snapshot on reserved testing samples, which provided a multi-class temporal classification of 92.2% when trained on data derived from the primary large scale collection. Adaptation of the model through transfer learning on hand held data collected relative to new stimulus parameters suffered significant overfitting related errors. Removing the parameters by popping all layers but Conv1 and Conv2, and allowing adaptation resulted in validation loss early in the training. Further discussion is presented relative to this result in chapter 5.

4.8 Additional Stimulus Parameters

Preliminary investigation of the optokinetic response using mobile device displays indicated the viability of ocular reflex inducing stimulus on the mobile display of the testing device. While the stimulus used in the testing was generated using best case parameters for the scale and viewing distance, the phase of gaze displacement remains relatively small for most participants. Additionally, the user experience of these visualizations is generally considered poor relative to visual design and integration standards. Several additional animations related to more intuitive iconography of processing and loading animations were generated and tested with the assistance of the sponsor. All attempts were made to provide at least 5 °visual field stimulation, or about 70% of the device display width and moving edges or textures greater than 6 °per second relative to a 25cm viewing distance. Moving 3D targets proved more difficult to reliably induce motion at smaller scales and lower rotational velocities. Animations which utilized larger portions of the screen while introducing arrays of dense edges greater than 0.1 °in width proved to generate the largest total onset related displacement. Small segment lengths with overlapping lines and edges which incorporated both translation and rotation appeared to generate the most reliable large scale rhythmic ocular motion in the samples collected. Additional studies related to the optimal outcomes for hand held device PAD via ocular motion are ongoing.

4.9 Occlusions and Segmentation Faults

Occasional occlusions of the user's face occurred in several samples. In most cases, these samples were detected due to failure of the Drishti segmentation engine. Cases included subjects covering their face with their hands, turning away from the device, and looking too far up or down. Most of these events occurred at later stages of the collection period, potentially indicating fatigue. In cases where the occlusion or fault took place during the spoof phase, the data was replaced using another randomly selected spoof sample. No occlusions were observed among the response samples, but some sample data contained signal issues or frame corruption, requiring replacement with a duplicate presentation from a random collection sequence. In total 37 sequences were replicated to balance the sample distribution. The complete data of five subjects were removed from evaluation due to some combination of collection software errors, participation issues, long eye closures, and high frequency blink events.

Detection failures resulting in non-continuous sequences of frames containing features occurred in approximately 8 % of the samples processed using the MTCNN framework. Samples processed by this method were discarded where discontinuities of more than 3 frames occurred. Some scale and geometry issues were encountered with the facial landmarking process, with the system detecting non-planar geometries for some samples. In total 218 samples were removed due to sequential discontinuities, and 188 more samples were removed due to failed planar geometry.

Table 1: Results of Model Performance for Tested Architectures

	Training	Testing	Combined
FC ANN	67.0%	58.9%	62.3%
TDNN	73.2%	68.4%	71.1%
SAE	81.1%	75.1%	77.3%
MC LSTM	93.8%	72.7%	87.0%
Binary LSTM (Single Stimulus)			
Type 1 (Horizontal Right)	94.4%	91.0%	92.8%
Type 2 (Horizontal Left)	95.7%	95.4%	95.6%
Type 3 (Vertical Up)	85.1%	87.9%	86.4%
Type 4 (Vertical Down)	95.4%	87.3%	92.4%
Type 5 (Oblique)	99.3%	95.4%	98.4%

Table 2: MTCNN Sequential Feature Localization and Segmentation Precision

	Left Eye	Right Eye	Nose	Left Mouth	Right Mouth
Max Offset	11.4	12.1	10.1	16.7	14.3
Mean Offset	4.4	5.2	5.1	6.2	6.5
Standard Dev.	3.2	3.4	4.1	5	4.8
Library Match	0.87	0.82	0.87	0.79	0.77
Scale Dev.	0.03	.05	0.09	0.07	0.07

CHAPTER 5

DISCUSSION

A variety of factors impact the application of liveness detection methods for mobile device biometrics in the visible spectrum. Quality issues with images collected from mobile cameras range from motion blur and smearing to over exposure and random sensor noise. Impacts of motion during capture are among the most problematic due to the potential for additive velocity when the user moves the camera and their head or eyes in opposite directions. Appearance based biometric methods require clear visible features and often utilize averaged values from registered frame sequences to reduce noise. Reliable registration of ocular features is dependent on relatively low amplitude motions while the samples are being collected, so simultaneous collection of OKN with this signal processing strategy is likely incompatible.

PAD is inherently a process that operates on sequential phenomena, and is therefore highly dependent on temporal characteristics and sampling frequency. Unstable or variable frame rates during the collection process of some additional samples in a hand held collection were indicated to generate sequences which were unusable without dynamic resampling. Compression and extrapolation effects resulting from this processing were similar to those observed in some early experiments with dynamic warping and wavelet decomposition and the resulting smoothing had a blurring effect on critical temporal features.

Multi-modal approaches often require additional sensors, and some models display issues when processing images from ethnically diverse populations [137]. Ocular feature localization methods, such as those used in our motion extraction pipeline have been demonstrated to largely alleviate this concern. Additionally, given the reflexive nature of the OKN response, no known cultural factors influence the expression of the target behavior. Reporting of experimental results involving the detection of print and screen based presentation attacks have been omitted from this publication, as discrimination of ocular motion samples extracted from these sources can be accomplished using frequency based threshold. Additionally, comparison of most state of the art methods is impractical, as sequences used as spoof in our experimental process meet the criteria of live samples in existing metrics.

5.1 User Experience

User experience factors remain a pressing concern for stimulus development. While no user *discomfort* was recorded in our collection, some users reported the structure and motion of the visual elements produced a hypnotic effect over the duration of their visit. This was substantiated by an increase in blink events in some participants, and an increase in OKN onset delay for most samples in the latter half of each session. Blink events play some role in the process of maintaining visual attention, and their presence in the samples collected was expected. Delay in total onset of the OKN response is a known physiological response factor. This delay is thought to act as a feature when the start of stimulus is properly synchronized with the visual collection. Removal of discontinuities caused by blinks requires several additional steps, and can degrade the quality of samples. Blink events also have an undesirable *reset* delay effect on OKN onset.

An increase in velocity was generally considered to offset the impact of delay, but diminishing returns were noted in terms of comfort and visual appeal. Alternative visualizations generated in the follow up process provided some insight into the possibility of smaller and more diverse animations and motion fields, but differentiating the responses to smaller animations remains an ongoing endeavor with multiple possible avenues of optimization including distance based dynamic rescaling. While edges and textures are still a requirement for reliable tracking behavior, large moving arrays with punctuated motion patterns produced more a phase-locked characteristic response when analyzed with respect to overall deflection in the initiation and departure time windows. Adding additional interest points into the acceleration profile of the motion stimulus provides additional freedom for parameter randomization, but there are limitations on the frequency and separation of the acceleration adjustment events due to the requirement of maintaining a maximum total duration.

5.2 Complexity

Complexity of sequential models increases radically when utilizing both CNN and 3DCNN architecture. Training required modification of some built in layer functions, and the addition of a second GPU with shared memory access. Initially, models were trained in Matlab™, but limitations in the ability to modify some architecture parameters led to development and training using python. Models reported for most of the advanced architectures discussed in chapter 4 were built using pytorch and tensorflow DL libraries. Models were mostly implemented using keras based scripting to simplify the process of model construction due the presence of many ready made layers and transforms. Modeling the 3DCNN was among the most difficult process, and resulted in the most computationally and run-time intensive network. Multiple memory issues were encountered, despite 64gb of system memory and two 8gb of GPUs. Batch size was limited to 24 samples when images were used directly. Processing images at full resolution was known not to be a feasible mobile deployment option, but even with MTCNN cropped ROIs, the batch was limited when training on a single GPU.

Ultimately, it appears that a multiple step pipeline as used in early experiments is likely the only reliable method for mobile deployment at the current development stage. Attempting to generate sequential models of this depth causes increased delay when compared to a parallel feature localization strategy. Several emerging methods of sparse processing may provide a significant reduction in overall model parameters by pruning, convergence, and ablation, but with large arrays fully connected 3DCNN layers require substantially more resources than are available on most mobile devices.

5.3 Dimensional Constraints

High-definition sources can potentially contain a wealth of features that can be difficult to extract at a smaller scale, but devising ways to utilize that information content can present its own challenges. In the case of the data acquired in this study, collection of samples with sufficient target spatial resolution was a requirement which can be directly correlated with the viability of a visible spectrum tracking application. Implementing and testing the stimulus parameters required a baseline assumption of localization-based discrimination, as the protocol necessary to test the primary hypothesis was based not only on the presence of a distinct behavioral signature, but the ability to reliably distinguish between those signatures. An upper limit exists for the accuracy and precision of any gaze approximation method given some inherent sample characteristics. Feature localization plays a major role in the calculation of behavioral patterns and ocular kinetics, but the impact of noise can never be disregarded.

Drishti's gaze approximation pipeline was initially designed to improve segmentation and feature localization for ocular feature extraction in mobile biometric authentication. Sequential precision wasn't a primary consideration in the initial design process, and as such each sample frame is processed using a naive approach. Certain reasonable assumptions are used in the localization and refinement of key features which reduce the runtime and preserve device resources. Some of these assumptions can have an undesirable impact which can become highly problematic when applied to sequences of images were consistency of localization is a direct measure of overall performance. Reflections can contribute to errors in segmentation of pupillary and iris regional boundaries, but this tends to hold true for nearly all ocular processing pipelines. Additional pre-processing steps can be implemented to stabilize the performance, but since all these methods should run reliably on the device computational costs become a mounting factor.

While our models provide some degree of error tolerance, significant jitter in the sequence outputs rendered the base library impractical for ocular motion feature extraction. It was observed in our experiments that ocular motion estimation relies more on the consistency of sequential localization than the accuracy of feature segmentation. To address this issue, a branch of the main Drishti code was modified to implement a multi-frame texture based patch tracking feature. Contextual cues provided by processing small groups of frames in a single batch provided a substantial improvement in the stability of sequential ocular feature localization. While this method could likely be improved, adding some degree of persistence appears to be an essential step in processing motion based feature estimates. The lower noise threshold of the improved model allowed for a minimal pre-processing approach, allowing motion features derived from Drishti's positional estimates to be sent directly to the classifier. While samples collected in our protocol are largely ideal due to our controlled collection environment, future implementations may benefit from implementing a feature localization confidence threshold to alleviate tracking errors due to segmentation faults.

Color stimulus was not fully investigated, but some indications exist that color patterns may provide advantages when implemented with the appropriate parameters. Given the density of color receptors in the primary visual field, colored edges or bands may have a greater total level of detail than strictly monochrome visualizations. Patterns of color dots and bands of either blue or red are commonly used in functional medicine, with patterns available on video sharing sites, but the velocities of most of these stimulus patterns were determined to be far slower than those known to induce OKN. Residual motion illusions appeared more common in some testing of color based visualizations, but this effect required stimulus durations higher than 10 seconds and would likely be infeasible for mobile devices due to poor user experience.

In additional testing and alternative stimulus design several attempts were made to utilize the visual compensation response to draw gaze away from blur and warp effects. With proper textures, some warp effects did generate ocular motion, but the distribution of angles was inconsistent for the samples collected. Animations employing fovea and motion blur mostly impacted kinetics by reducing fixation time in search type patterns. Moving a focused disc approximately 2cm in diameter within a Gaussian blurred screen space did generate several consistent displacements, but it was determined that pursuits accounted for only a small fraction of the total observed motions. Convergence and looming illusions used in some testing generated an undesirable user experience when velocities were sufficient to generate reliable gain. Motion sensations were observed in several tests, but attempts to induce convergence motions resulted in the greatest degree of user feedback for hand held collections. Projections of textures and gratings on complex shapes like curves, cubes and spheres met with poor performance at scale, and generally taxed the rendering methods used in the mobile device test applications.

Front facing camera lens systems are typically calibrated to generate a wide field of view while maintaining an optimal depth of field. As a result of their small size, and compact design, these lenses generate substantial distortion which requires extensive post processing to correct. At normal viewing distance, the ROI for a single eye occupies approximately 1/10th of the front facing camera's field of view, or 200x160 pixels per eye, limiting the spatial resolution and angular fidelity. While some experiments indicate angular fidelity of less than 1°could be possible with spatiotemporal techniques, these methods remain impractical for mobile edge computing due to memory and model size constraints. Future hardware, including higher resolution sensors and better lens systems may facilitate a solution to this dilemma. Temporal super resolution approach might also increase localization fidelity, but without advancement in the technology, substantial libraries of high resolution video responses would be needed to train such as system.

Several of the methods used in this study were intended primarily to circumvent accuracy related issues with gaze approximation by focusing instead on stable sequential estimation of frame to frame localized features relative to the image target. This approach generated reliable features for the relative ocular motion patterns, but differentiation between stimulus classes seems to rely on phase and angle based differences in the induced pattern of displacement. Implementation of a facial pose estimator was demonstrated to potentially provide a path to generating more accurate gaze estimates.

More nuanced individual responses when recording motions using high accuracy gaze systems, potentially indicating the use of OKN or other ocular motion kinetics as a biometric authentication mechanism. Further study of this application may be warranted.

CHAPTER 6

CONCLUSIONS

PAD methods are an essential component of reliable mobile device biometric systems. Results of the experiments conducted in this study indicate that a subject independent PAD method based on reflexive behavioral signatures may help protect against sophisticated emerging attacks like deep fakes, and may even help deter higher effort attack vectors like masks, digital puppets, and 3D printed structure based attacks. While the library dependency of the feature extraction method used in the early testing is cumbersome due to several additional steps, the feature localization provided sufficient stability for reliable OKN onset based PAD classification. Spatiotemporal methods proved more difficult to train, but required only roughly aligned images to generate stable temporal features. Transfer learning from larger pretrained models is typically seen as a key advantage for advanced DL systems, but given the unique constraints of the application few models were applicable given the disparity in training parameters. Even advanced models like Google Mediapipe, which *integrate* optical flow to achieve improved temporal stabilization for face mesh fitting, are still remarkably sensitive to noise. Training with a more broad data set has advantages over small set testing, and while the data collected in this study has proven significant, the total number of sequence examples remains relatively small in comparison to the behavior being analysed.

Top performing models generated in this study indicate the feasibility of OKN based PAD when using a mobile device as the collection platform. Samples from each of the major motion classes where identified reliably for the conditions present in the collection, with the top performing model providing approximately 98% accuracy when processing extracted sequential features. DL methods proved more computationally intensive and time consuming to train than the hand crafted feature models. This was expected, but due to limitations in the memory capacity and processing hardware, only a limited subset of the data could be processed reliably. Results from those configurations showed promise by demonstrating a 92% classification accuracy for multi-class response detection. Based on these results, it appears an automated spatiotemporal feature extraction method for behavioral classification using ocular response data is possible, but further refinement is necessary for generating an approach that fits both classification accuracy and computational complexity constraints.

Some of the most potentially significant discoveries related to the advancement in ocular reflex based PAD generated by this study are related to the parameters of visual stimulus and feature localization. General objectives of most face detection and localization models dictate that the relative level of precision required for high accuracy gaze localization has not been achieved by any currently known model. While high accuracy models aren't needed to generate a reliable PAD model, they are likely necessary to generate better multi-class discriminate features. High accuracy gaze models also have substantial practical applications for other user interface applications.

Utilizing the device display and integrated front facing camera to elicit OKN generates a variety of time-locked and phase-locked response characteristics. While the bulk of data processing was implemented off-line, at least some of the feature extraction pipeline and model architectures were optimized for the target platform. Testing of model deployment for the more advanced models was infeasible due to memory limitations of the target platforms, but results of the precision testing for the sequential optimization methods indicate the goal is likely linked to a viable avenue of investigation. Implementation of these methods will likely remain limited due to cost factors and complexity, however, our experiments indicate the feasibility of a spatiotemporal PAD method without the need for additional sensors.

Results also indicate that with small sample collections, an array of addition stimulus parameters and animation types may be viable for generated general optokinetc responses. In the preliminary case studies conducted on these stimuli, library and sequential processing based assessments show the presence of intended motion signatures. Based on our findings we believe this work represents a potentially significant step for mobile device liveness assessment and biometric security and a provides a road map for improvement and advancement of other gaze based technologies and mobile applications.

REFERENCE LIST

[1] S. Guennouni, A. Mansouri, and A. Ahaitouf, "Biometric Systems and Their Applications," in *Visual Impairment and Blindness [Working Title]*, 2019. [Online]. Available: https://www.intechopen.com/online-first/biometric-systems-and-their-applications

[2] R. Moskovitch, C. Feher, A. Messerman, N. Kirschnick, T. Mustafić, A. Camtepe, B. Löhlein, U. Heister, S. Möller, L. Rokach, and Y. Elovici, "Identity theft, computers and behavioral biometrics," in *2009 IEEE International Conference on Intelligence and Security Informatics, ISI 2009*, 2009. doi: 10.1109/ISI.2009.5137288. ISBN 9781424441730

[3] M. Zloteanu, N. Harvey, D. Tuckett, and G. Livan, "Digital identity: The effect of trust and reputation information on user judgement in the sharing economy," *PLoS ONE*, 2018. doi: 10.1371/journal.pone.0209071

[4] Z. Akhtar, G. Kumar, S. Bakshi, and H. Proenca, "Experiments with ocular biometric datasets: A practitioner's guideline," *IT Professional*, 2018. doi: 10.1109/MITP.2018.032501748

[5] L. M. Mayron, "Biometric Authentication on Mobile Devices," *IEEE Security and Privacy*, 2015. doi: 10.1109/MSP.2015.67

[6] S. M and P. G, “Mobile Device Security: A Survey on Mobile Device Threats, Vulnerabilities and their Defensive Mechanism,” *International Journal of Computer Applications*, 2012. doi: 10.5120/8960-3163

[7] S. B. Gould, “Computers at risk: Safe computing in the information age,” *Government Information Quarterly*, 1991. doi: 10.1016/0740-624x(91)90010-6

[8] R. Ramachandra and C. Busch, “Presentation attack detection methods for face recognition systems: A comprehensive survey,” *ACM Computing Surveys*, 2017. doi: 10.1145/3038924

[9] A. Khodabakhsh, R. Ramachandra, K. Raja, P. Wasnik, and C. Busch, “Fake Face Detection Methods: Can They Be Generalized?” in *2018 International Conference of the Biometrics Special Interest Group, BIOSIG 2018*, 2018. doi: 10.23919/BIOSIG.2018.8553251. ISBN 9783885796763

[10] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, “Face anti-spoofing using patch and depth-based CNNs,” in *IEEE International Joint Conference on Biometrics, IJCB 2017*, 2018. doi: 10.1109/BTAS.2017.8272713. ISBN 9781538611241

[11] A. Husseis, J. Liu-Jimenez, I. Goicoechea-Telleria, and R. Sanchez-Reillo, “A survey in presentation attack and presentation attack detection,” in *Proceedings - International Carnahan Conference on Security Technology*, 2019. doi: 10.1109/CCST.2019.8888436. ISBN 9781728115764. ISSN 10716572

[12] N. Clarke, J. Symes, H. Saevanee, and S. Furnell, “Awareness of mobile device security a survey of user’s attitudes,” *International Journal of Mobile Computing and Multimedia Communications*, 2016. doi: 10.4018/IJMCMC.2016010102

[13] S. Riihiaho, “Usability Testing,” in *The Wiley Handbook of Human Computer Interaction Set*, 2017. ISBN 9781118976005

[14] C. M. Knapp, F. a. Proudlock, and I. Gottlob, “OKN asymmetry in human subjects: a literature review.” *Strabismus*, vol. 21, no. 1, pp. 37–49, 2013. doi: 10.3109/09273972.2012.762532. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23477776

[15] J. Waddington and C. M. Harris, “Human optokinetic nystagmus: A stochastic analysis,” *Journal of Vision*, 2012. doi: 10.1167/12.12.5

[16] J. M. Furman, “Optokinetic Nystagmus,” in *Encyclopedia of the Neurological Sciences*, 2014. ISBN 9780123851574

[17] J. Turuwhenua, T. Y. Yu, Z. Mazharullah, and B. Thompson, “A method for detecting optokinetic nystagmus based on the optic flow of the limbus,” *Vision Research*, vol. 103, pp. 75–82, oct 2014. doi: 10.1016/j.visres.2014.07.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0042698914001758

[18] C. Song, A. Wang, K. Ren, and W. Xu, “EyeVeri: A secure and usable approach for smartphone user authentication,” in *Proceedings - IEEE INFOCOM*, 2016. doi: 10.1109/INFOCOM.2016.7524367. ISBN 9781467399531. ISSN 0743166X

[19] R. Bednarik, T. Kinnunen, A. Mihaila, and P. Franti, “Eye-movements as a biometric,” *Image Analysis, Proceedings*, 2005. doi: 10.1007/11499145_79

[20] V. Cantoni, M. Musci, N. Nugrahaningsih, and M. Porta, “Gaze-based biometrics: An introduction to forensic applications,” *Pattern Recognition Letters*, 2018. doi: 10.1016/j.patrec.2016.12.006

[21] S. Kaymak, “Real-time appearance-based gaze tracking,” Ph.D. dissertation, Queen Mary University of London, 2015. [Online]. Available: http://qmro.qmul.ac.uk/xmlui/handle/123456789/8949

[22] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, “MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. doi: 10.1109/TPAMI.2017.2778103

[23] C. Holland and O. V. Komogortsev, “Biometric identification via eye movement scanpaths in reading,” in *2011 International Joint Conference on Biometrics, IJCB 2011*, 2011. doi: 10.1109/IJCB.2011.6117536. ISBN 9781457713583

[24] R. Anderson, C. Barton, R. Böhme, R. Clayton, C. Gañán, T. Grasso, M. Levi, T. Moore, and M. Vasek, “Measuring the Changing Cost of Cybercrime,” in *The 2019 Workshop on the Economics of Information Security, Boston*, 2019.

[25] A. L. Fantana, S. Ramachandran, C. H. Schunck, and M. Talamo, “Movement based biometric authentication with smartphones,” in *Proceedings*

- *International Carnahan Conference on Security Technology*, 2016. doi: 10.1109/CCST.2015.7389688. ISBN 9781479986910. ISSN 10716572

[26] A. Rattani, R. Derakhshani, and A. Ross, "Introduction to selfie biometrics," in *Advances in Computer Vision and Pattern Recognition*, 2019.

[27] Guodong Guo and H. Wechsler, "Mobile biometrics," in *Mobile Biometrics*, 2017.

[28] A. Rattani and R. Derakhshani, "A Survey Of mobile face biometrics," *Computers and Electrical Engineering*, 2018. doi: 10.1016/j.compeleceng.2018.09.005

[29] V. Gottemukkula, S. Saripalle, S. P. Tankasala, and R. Derakhshani, "Method for using visible ocular vasculature for mobile biometrics," *IET Biometrics*, vol. 5, no. 1, pp. 3–12, 2016. doi: 10.1049/iet-bmt.2014.0059

[30] Guodong Guo and H. Wechsler, "Mobile biometrics," in *Mobile Biometrics*, 2017, pp. 1–5.

[31] S. Wang and J. Liu, "Biometrics on mobile phone," in *Recent Application in Biometrics*, 2011.

[32] A. Rattani and R. Derakhshani, "Ocular biometrics in the visible spectrum: A survey," *Image and Vision Computing*, vol. 59, pp. 1 – 16, 2017. doi: https://doi.org/10.1016/j.imavis.2016.11.019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885616302165

[33] D. LYON, "Biometrics, identification and surveillance," *Bioethics*, vol. 22, no. 9, pp. 499–508, 2008. doi: https://doi.org/10.1111/j.1467-8519.2008.00697.x.

[Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8519.2008.00697.x

[34] A. A. Tarnutzer and D. Straumann, "Nystagmus," *Current Opinion in Neurology*, 2018. doi: 10.1097/WCO.0000000000000517

[35] C. Valmaggia and I. Gottlob, "Role of the stimulus size in the generation of optokinetic nystagmus in normals and in patients with retinitis pigmentosa," in *Klinische Monatsblatter fur Augenheilkunde*, vol. 221, no. 5, 2004. doi: 10.1055/s-2004-812864. ISBN 0023-2165 (Print)\n0023-2165 (Linking). ISSN 00232165 pp. 390–394.

[36] S. Garbutt, Y. Han, A. N. Kumar, M. Harwood, C. M. Harris, and R. J. Leigh, "Vertical optokinetic nystagmus and saccades in normal human subjects," *Investigative Ophthalmology and Visual Science*, 2003. doi: 10.1167/iovs.03-0066

[37] A. T. Duchowski, "A breadth-first survey of eye-tracking applications," *Behavior Research Methods, Instruments, & Computers*, vol. 34, no. 4, pp. 455–470, 2002. doi: 10.3758/BF03195475. [Online]. Available: http://www.springerlink.com/index/10.3758/BF03195475

[38] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "It's written all over your face: Full-face appearance-based gaze estimation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017. doi: 10.1109/CVPRW.2017.284 pp. 2299–2308.

[39] C. Yo and J. L. Demer, “Two-dimensional optokinetic nystagmus induced by moving plaids and texture boundaries: Evidence for multiple visual pathways,” *Investigative Ophthalmology and Visual Science*, vol. 33, no. 8, pp. 2490–2500, 1992.

[40] J. Waddington and C. M. Harris, “Human optokinetic nystagmus and spatial frequency,” *Journal of Vision*, 2015. doi: 10.1167/15.13.7

[41] N. Wade and B. Tatler, *The Moving Tablet of the Eye: The Origins of Modern Eye Movement Research*. Oxford University Press, 2010. ISBN 9780191584954. [Online]. Available: https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780198566175.001.0001/acprof-9780198566175

[42] Y. Zhang and M. Juhola, “On Biometrics with Eye Movements,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1360–1366, 2017. doi: 10.1109/JBHI.2016.2551862

[43] O. V. Komogortsev and I. Rigas, “BioEye 2015: Competition on biometrics via eye movements,” in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems, BTAS 2015*, 2015. doi: 10.1109/BTAS.2015.7358750. ISBN 9781479987764

[44] I. Rigas and O. V. Komogortsev, “Biometric recognition via probabilistic spatial projection of eye movement trajectories in dynamic visual environments,” *IEEE Transactions on Information Forensics and Security*, 2014. doi: 10.1109/TIFS.2014.2350960

[45] W. Gutfeter and A. Pacut, "Face 3D biometrics goes mobile: Searching for applications of portable depth sensor in face recognition," in *Proceedings - 2015 IEEE 2nd International Conference on Cybernetics, CYBCONF 2015*, 2015. doi: 10.1109/CYBConf.2015.7175983. ISBN 9781479983223

[46] A. Support, "About Face ID advanced technology," 2018.

[47] J. Cui, H. Zhang, H. Han, S. Shan, and X. Chen, "Improving 2D face recognition via discriminative face depth estimation," in *Proceedings - 2018 International Conference on Biometrics, ICB 2018*, 2018. doi: 10.1109/ICB2018.2018.00031. ISBN 9781538642856

[48] L. Li, Z. Xia, X. Jiang, Y. Ma, F. Roli, and X. Feng, "3D face mask presentation attack detection based on intrinsic image analysis," *IET Biometrics*, 2020. doi: 10.1049/iet-bmt.2019.0155

[49] S. Liu, B. Yang, P. C. Yuen, and G. Zhao, "A 3D Mask Face Anti-Spoofing Database with Real World Variations," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2016. doi: 10.1109/CVPRW.2016.193. ISBN 9781467388504. ISSN 21607516

[50] S. Jia, G. Guo, and Z. Xu, "A survey on 3D mask presentation attack detection and countermeasures," *Pattern Recognition*, 2020. doi: 10.1016/j.patcog.2019.107032

[51] Z. Akhtar, C. Michelon, and G. L. Foresti, "Liveness detection for biometric authentication in mobile applications," *2014 International Carnahan Conference on Security Technology (ICCST)*, pp. 1–6, 2014. doi: 10.1109/CCST.2014.6986982

[52] A. A. Patil and S. A. Dhole, "Image Quality (IQ) based liveness detection system for multi-biometric detection," in *Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016*, 2017. doi: 10.1109/INVENTIVE.2016.7823297. ISBN 9781509012855

[53] F. Pala and B. Bhanu, "Iris Liveness Detection by Relative Distance Comparisons," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017. doi: 10.1109/CVPRW.2017.95. ISBN 9781538607336. ISSN 21607516

[54] T. Edmunds and A. Caplier, "Face spoofing detection based on colour distortions," *IET Biometrics*, 2018. doi: 10.1049/iet-bmt.2017.0077

[55] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric Face Presentation Attack Detection With Multi-Channel Convolutional Neural Network," *IEEE Transactions on Information Forensics and Security*, 2020. doi: 10.1109/TIFS.2019.2916652

[56] S. Trewin, C. Swart, L. Koved, J. Martino, K. Singh, and S. Ben-David, "Biometric authentication on a mobile device: A study of user effort, error and task disruption," in *ACM International Conference Proceeding Series*, 2012. doi: 10.1145/2420950.2420976. ISBN 9781450313124

[57] A. Ross and A. K. Jain, “Multimodal biometrics: An overview,” in *European Signal Processing Conference*, 2015. ISBN 9783200001657. ISSN 22195491

[58] R. R. Jillela and A. Ross, “Segmenting iris images in the visible spectrum with applications in mobile biometrics,” *Pattern Recognition Letters*, 2015. doi: 10.1016/j.patrec.2014.09.014

[59] N. Reddy, D. F. Noor, Z. Li, and R. Derakhshani, “Multi-frame super resolution for ocular biometrics,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2018. doi: 10.1109/CVPRW.2018.00086. ISBN 9781538661000. ISSN 21607516

[60] M. S. Nixon and J. N. Carter, “Automatic recognition by gait,” *Proceedings of the IEEE*, 2006. doi: 10.1109/JPROC.2006.886018

[61] D. P. Benalcazar, D. Bastias, C. A. Perez, and K. W. Bowyer, “A 3D Iris Scanner from Multiple 2D Visible Light Images,” *IEEE Access*, 2019. doi: 10.1109/ACCESS.2019.2915786

[62] A. Mahfouz, T. M. Mahmoud, and A. Sharaf Eldin, “A behavioral biometric authentication framework on smartphones,” in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, ser. ASIA CCS ’17. New York, NY, USA: Association for Computing Machinery, 2017. doi: 10.1145/3052973.3055160. ISBN 9781450349444 p. 923–925. [Online]. Available: https://doi.org/10.1145/3052973.3055160

[63] A. Alzubaidi and J. Kalita, “Authentication of smartphone users using behavioral biometrics,” *IEEE Communications Surveys and Tutorials*, 2016. doi: 10.1109/COMST.2016.2537748

[64] P. Kasprowski and J. Ober, “Eye movements in biometrics,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3087, pp. 248–258, 2004. doi: 10.1007/978-3-540-25976-3_23

[65] M. Smiatacz, “Liveness measurements using optical flow for biometric person authentication,” *Metrology and Measurement Systems*, 2012. doi: 10.2478/v10178-012-0022-y

[66] I. Rigas, O. Komogortsev, and R. Shadmehr, “Biometric recognition via eye movements: Saccadic vigor and acceleration cues,” *ACM Transactions on Applied Perception*, 2016. doi: 10.1145/2842614

[67] K. Holmqvist and R. Andersson, *Eye-tracking: A comprehensive guide to methods, paradigms and measures*. CreateSpace Independent Publishing Platform, 11 2017. ISBN ISBN-13: 978-1979484893

[68] D. Liu, B. Dong, X. Gao, and H. Wang, “Exploiting eye tracking for smartphone authentication,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015. doi: 10.1007/978-3-319-28166-7_22. ISBN 9783319281650. ISSN 16113349

[69] L. Wang, Y. Xu, J. Cheng, H. Xia, J. Yin, and J. Wu, "Human Action Recognition by Learning Spatio-Temporal Features with Deep Neural Networks," *IEEE Access*, 2018. doi: 10.1109/ACCESS.2018.2817253

[70] K. Pfeuffer, M. J. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt, "Behavioural biometrics in VR," in *Conference on Human Factors in Computing Systems - Proceedings*, 2019. doi: 10.1145/3290605.3300340. ISBN 9781450359702

[71] Z. Chen and B. E. Shi, "Using Variable Dwell Time to Accelerate Gaze-Based Web Browsing with Two-Step Selection," *International Journal of Human-Computer Interaction*, 2019. doi: 10.1080/10447318.2018.1452351

[72] T. Yu, J. Zhao, Z. Zheng, K. Guo, Q. Dai, H. Li, G. Pons-Moll, and Y. Liu, "DoubleFusion: Real-Time Capture of Human Performances with Inner Body Shapes from a Single Depth Sensor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. doi: 10.1109/TPAMI.2019.2928296

[73] H. Joo, T. Simon, and Y. Sheikh, "Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018. doi: 10.1109/CVPR.2018.00868. ISBN 9781538664209. ISSN 10636919

[74] Y. Zhang, W. Hu, W. Xu, C. T. Chou, and J. Hu, "Continuous Authentication Using Eye Movement Response of Implicit Visual Stimuli," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018. doi: 10.1145/3161410

[75] I. Bouchrika, J. N. Carter, and M. S. Nixon, “Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras,” *Multimedia Tools and Applications*, 2016. doi: 10.1007/s11042-014-2364-9

[76] D. Seckiner, X. Mallett, P. Maynard, D. Meuwly, and C. Roux, “Forensic gait analysis — morphometric assessment from surveillance footage,” *Forensic Science International*, vol. 296, 03 2019. doi: 10.1016/j.forsciint.2019.01.007

[77] A. R. Hawas, H. A. El-Khobby, M. Abd-Elnaby, and F. E. Abd El-Samie, “Gait identification by convolutional neural networks and optical flow,” *Multimedia Tools and Applications*, 2019. doi: 10.1007/s11042-019-7638-9

[78] J. Pansiot, “Markerless Visual Tracking and Motion Analysis for Sports Monitoring,” *Biomedicine*, 2009.

[79] S. Barris and C. Button, “A review of vision-based motion analysis in sport,” *Sports medicine (Auckland, N.Z.)*, vol. 38, pp. 1025–43, 02 2008. doi: 10.2165/00007256-200838120-00006

[80] O. Younis, W. Al-Nuaimy, M. H. Alomari, and F. Rowe, “A hazard detection and tracking system for people with peripheral vision loss using smart glasses and augmented reality,” *International Journal of Advanced Computer Science and Applications*, 2019. doi: 10.14569/ijacsa.2019.0100201

[81] A. Godil, R. Bostelman, W. Shackleford, T. Hong, and M. Shneier, “Performance Metrics for Evaluating Object and Human Detection and Tracking Systems,” *Nistir 7972*, 2014.

[82] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, “Tracking without bells and whistles,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019. doi: 10.1109/ICCV.2019.00103. ISBN 9781728148038. ISSN 15505499

[83] M. L. Rusch, M. C. Schall, P. Gavin, J. D. Lee, J. D. Dawson, S. Vecera, and M. Rizzo, “Directing driver attention with augmented reality cues,” *Transportation Research Part F: Traffic Psychology and Behaviour*, 2013. doi: 10.1016/j.trf.2012.08.007

[84] J. Xu, J. Min, and J. Hu, “Real-time eye tracking for the assessment of driver fatigue,” *Healthcare Technology Letters*, vol. 5, no. 2, pp. 54–58, 2017. doi: 10.1049/htl.2017.0020

[85] H. . Drewes, “Eye Gaze Tracking for Human Computer Interaction,” *Thesis*, 2010. doi: 10.1007/3-540-44589-7

[86] A. Ahmed and E. Ahmed, “A survey on mobile edge computing,” in *Proceedings of the 10th International Conference on Intelligent Systems and Control, ISCO 2016*, 2016. doi: 10.1109/ISCO.2016.7727082. ISBN 9781467378079

[87] B. V. Neal E. Boudette, “Tesla Self-Driving System Faulted by Safety Agency in Crash,” New York, p. B1, sep 2017. [Online]. Available: https://www.nytimes.com/2017/09/12/business/self-driving-cars.html

[88] M. Tonsen, J. Steil, Y. Sugano, and A. Bulling, “InvisibleEye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, pp. 1–21, sep 2017. doi: 10.1145/3130971

[89] D. E. King, “Dlib-ml: A machine learning toolkit,” *Journal of Machine Learning Research*, 2009.

[90] S. Milborrow and F. Nicolls, “Locating facial features with an extended active shape model,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008. doi: 10.1007/978-3-540-88693-8-37. ISBN 3540886923. ISSN 03029743

[91] P. Huber, “eos: A lightweight header-only 3D Morphable Face Model fitting library in modern C++11/14.” 2018. [Online]. Available: https://github.com/patrikhuber/eos

[92] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “MTCNN,” *IEEE Signal Processing Letters*, 2016. doi: 10.1109/LSP.2016.2603342

[93] E. A. da Silva and G. V. Mendonca, “Digital Image Processing,” in *The Electrical Engineering Handbook*, 2005. ISBN 9780121709600

[94] Sergios Theodoridis and K. Koutroumbas, *Pattern Recognition (Fourth Edition)*. Boston: Academic Press, 2009. ISBN 9781597492720

[95] S. Thavalengal, T. Nedelcu, P. Bigioi, and P. Corcoran, “Iris liveness detection for next generation smartphones,” *IEEE Transactions on Consumer Electronics*, 2016. doi: 10.1109/TCE.2016.7514667

[96] G. Heusch and S. Marcel, “Pulse-based features for face presentation attack detection,” in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems, BTAS 2018*, 2018. doi: 10.1109/BTAS.2018.8698579. ISBN 9781538671795

[97] C. Huang, H. Chen, L. Yang, and Q. Zhang, “BreathLive: Liveness Detection for Heart Sound Authentication with Deep Breathing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018. doi: 10.1145/3191744

[98] D. Gefen, “E-commerce: The role of familiarity and trust,” *Omega*, 2000. doi: 10.1016/S0305-0483(00)00021-9

[99] M. Galterio, S. Shavit, and T. Hayajneh, “A review of facial biometrics security for smart devices,” *Computers*, vol. 7, p. 37, 06 2018. doi: 10.3390/computers7030037

[100] Y. D. Wang and H. H. Emurian, “An overview of online trust: Concepts, elements, and implications,” *Computers in Human Behavior*, 2005. doi: 10.1016/j.chb.2003.11.008

[101] O. V. Komogortsev and A. Karpov, "Liveness detection via oculomotor plant characteristics: Attack of mechanical replicas," in *Proceedings - 2013 International Conference on Biometrics, ICB 2013*, 2013. doi: 10.1109/ICB.2013.6612984. ISBN 9781479903108

[102] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proceedings of 2009 International Conference on Image Analysis and Signal Processing, IASP 2009*, 2009. doi: 10.1109/IASP.2009.5054589. ISBN 9781424439867

[103] O. Kahm and N. Damer, "2D face liveness detection: An overview," in *Proceedings of the International Conference of the Biometrics Special Interest Group, BIOSIG 2012*, 2012. ISBN 9783885792901

[104] A. Rattani and N. Poh, "Biometric system design under zero and non-zero effort attacks," in *Proceedings - 2013 International Conference on Biometrics, ICB 2013*, 2013. doi: 10.1109/ICB.2013.6612999. ISBN 9781479903108

[105] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein, "A framework for liveness detection for direct attacks in the visible spectrum for multimodal ocular biometrics," *Pattern Recognition Letters*, vol. 82, pp. 232–241, 2016. doi: 10.1016/j.patrec.2015.11.016

[106] J. Martini & Nath, *Fundamentals of Anatomy and Physiology*. Delmar Publishers, 2009, vol. 7th. ISBN 0766804984

[107] D. Purves, G. Augustine, D. Fitzpatrick, L. Katz, A.-S. LaMantia, J. McNamara, and M. Williams, *Neuroscience. 2nd edition*. Sinauer Associates, 2001. ISBN 10: 0-87893-742-0

[108] N. S. Latman and E. Herb, "A field study of the accuracy and reliability of a biometric iris recognition system," *Science and Justice*, 2013. doi: 10.1016/j.scijus.2012.03.008

[109] H. Saevanee, N. Clarke, S. Furnell, and V. Biscione, "Continuous user authentication using multi-modal biometrics," *Computers and Security*, 2015. doi: 10.1016/j.cose.2015.06.001

[110] G. C. Van Die and H. Collewijn, "Control of human optokinetic nystagmus by the central and peripheral retina: Effects of partial visual field masking, scotopic vision and central retinal scotomata," *Brain Research*, vol. 383, no. 1-2, pp. 185–194, sep 1986. doi: 10.1016/0006-8993(86)90019-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0006899386900193

[111] A. Ceccarelli, L. Montecchi, F. Brancati, P. Lollini, A. Marguglio, and A. Bondavalli, "Continuous and transparent user identity verification for secure internet services," *IEEE Transactions on Dependable and Secure Computing*, 2015. doi: 10.1109/TDSC.2013.2297709

[112] J. M. Carroll and M. B. Rosson, "Usability engineering," in *Computing Handbook, Third Edition: Information Systems and Information Technology*. CRC Press, jan 2014, pp. 32–1–32–22. ISBN 9781439898567

[113] V. Vaitukaitis and A. Bulling, “Eye gesture recognition on portable devices,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp ’12*, 2012. doi: 10.1145/2370216.2370370. ISBN 9781450312240

[114] S. B. Gould, “Computers at risk: Safe computing in the information age,” *Government Information Quarterly*, 1991. doi: 10.1016/0740-624x(91)90010-6

[115] U. Büttner and O. Kremmyda, “Smooth pursuit eye movements and optokinetic nystagmus,” *Neuro-Ophthalmology*, vol. 40, pp. 76–89, 02 2007. doi: 10.1159/000100350

[116] S. J. Farooq, F. A. Proudlock, and I. Gottlob, “Torsional optokinetic nystagmus: normal response characteristics.” *Br J Ophthalmol*, vol. 88, no. 6, pp. 796–802, 2004. doi: 10.1136/bjo.2003.028738

[117] C. S. Konen, R. Kleiser, R. J. Seitz, and F. Bremmer, “An fMRI study of optokinetic nystagmus and smooth-pursuit eye movements in humans,” *Experimental Brain Research*, 2005. doi: 10.1007/s00221-005-2289-7

[118] A. Li and Q. Zaidi, “Three-dimensional shape from non-homogeneous textures: Carved and stretched surfaces.” *Journal of Vision*, vol. 4, no. 10, pp. 860–878, 2004. doi: 10.1167/4.10.3. [Online]. Available: ali@sunyopt.edu

[119] M. R. Dursteler and R. H. Wurtz, “Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST,” *Journal of Neurophysiology*, 1988. doi: 10.1152/jn.1988.60.3.940

[120] M. Fujiwara, C. Ding, L. Kaunitz, J. C. Stout, D. Thyagarajan, and N. Tsuchiya, "Optokinetic nystagmus reflects perceptual directions in the onset binocular rivalry in Parkinson's disease," *PLoS ONE*, 2017. doi: 10.1371/journal.pone.0173707

[121] K. V. Haak, F. W. Cornelissen, and A. B. Morland, "Population receptive field dynamics in human visual cortex," *PLoS ONE*, 2012. doi: 10.1371/journal.pone.0037686

[122] L. Pizzamiglio, R. Frasca, C. Guariglia, C. Incoccia, and G. Antonucci, "Effect of Optokinetic Stimulation in Patients with Visual Neglect," *Cortex*, 1990. doi: 10.1016/S0010-9452(13)80303-6

[123] M. Strupp, O. Kremmyda, C. Adamczyk, N. Böttcher, C. Muth, C. W. Yip, and T. Bremova, "Central ocular motor disorders, including gaze palsy and nystagmus," *Journal of Neurology*, 2014. doi: 10.1007/s00415-014-7385-9

[124] F. Jan, "Segmentation and localization schemes for non-ideal iris biometric systems," *Signal Processing*, vol. 133, pp. 192 – 212, 2017. doi: https://doi.org/10.1016/j.sigpro.2016.11.007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168416303206

[125] M. Karakaya, "Deep learning frameworks for off-angle iris recognition," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems, BTAS 2018*, 2018. doi: 10.1109/BTAS.2018.8698565. ISBN 9781538671795

[126] A. L. Fantana, S. Ramachandran, C. H. Schunck, and M. Talamo, "Movement based biometric authentication with smartphones," in *Proceedings - International Carnahan Conference on Security Technology*, 2016. doi: 10.1109/CCST.2015.7389688. ISBN 9781479986910. ISSN 10716572

[127] K. Sundararajan and D. L. Woodard, "Deep Learning for Biometrics," *ACM Computing Surveys*, 2018. doi: 10.1145/3190618

[128] Z. Akhtar, S. Kale, and N. Alfarid, "Spoof attacks in mutimodal biometric systems," *2011 International Conference on Information and Network Technology*, 2011.

[129] R. N. Rodrigues, L. L. Ling, and V. Govindaraju, "Robustness of multimodal biometric fusion methods against spoof attacks," *Journal of Visual Languages and Computing*, 2009. doi: 10.1016/j.jvlc.2009.01.010

[130] A. Pinto, W. R. Schwartz, H. Pedrini, and A. D. R. Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Transactions on Information Forensics and Security*, 2015. doi: 10.1109/TIFS.2015.2395139

[131] N. Reddy, A. Rattani, and R. Derakhshani, "OcularNet: Deep Patch-based Ocular Biometric Recognition," in *2018 IEEE International Symposium on Technologies for Homeland Security, HST 2018*, 2018. doi: 10.1109/THS.2018.8574156. ISBN 9781538634431

[132] Y. Wang and T. P. Jung, "A collaborative brain-computer interface for improving human performance," *PLoS ONE*, 2011. doi: 10.1371/journal.pone.0020422

[133] K. Patel, H. Han, and A. K. Jain, "Secure Face Unlock: Spoof Detection on Smartphones," *IEEE Transactions on Information Forensics and Security*, 2016. doi: 10.1109/TIFS.2016.2578288

[134] H. Bleau, "2017 Global Fraud & Cybercrime Forecast," 2017. [Online]. Available: https://www.rsa.com/en-us/blog/2016-12/2017-global-fraud-cybercrime-forecast

[135] Cybersecurity Ventures, "2019 CyberVentures Cybercrime Report," *Herjavec Group*, 2019.

[136] A. Liu, X. Li, J. Wan, S. Escalera, H. J. Escalante, M. Madadi, Y. Jin, Z. Wu, X. Yu, Z. Tan, and Others, "Cross-ethnicity Face Anti-spoofing Recognition Challenge: A Review," *arXiv preprint arXiv:2004.10998*, 2020.

[137] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *UbiComp 2014 - Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014. doi: 10.1145/2638728.2641695. ISBN 9781450330473

[138] D. Hirvonin, "Drishti: Real time eye tracking for embedded and mobile devices in C++11." [Online]. Available: https://github.com/elucideye/drishti/tree/master

[139] P. Dollar, R. Appel, S. Belongie, and P. Perona, “Fast feature pyramids for object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014. doi: 10.1109/TPAMI.2014.2300479

[140] P. Dollar, P. Welinder, and P. Perona, “Cascaded pose regression,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2010. doi: 10.1109/CVPR.2010.5540094. ISBN 978-1-4244-6984-0. ISSN 10636919 pp. 1078–1085. [Online]. Available: http://ieeexplore.ieee.org/document/5540094/

[141] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. doi: 10.1145/2939672.2939785. ISBN 9781450342322

[142] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014. doi: 10.1109/CVPR.2014.241. ISBN 9781479951178. ISSN 10636919

[143] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, W. Chang, W. Hua, M. Georg, and M. Grundmann, “Mediapipe: A framework for building perception pipelines,” *ArXiv*, vol. abs/1906.08172, 2019.

[144] C. D. Holland and O. V. Komogortsev, “Complex eye movement pattern biometrics: Analyzing fixations and saccades,” in *Proceedings - 2013 International Conference on Biometrics, ICB 2013*, 2013. doi: 10.1109/ICB.2013.6612953. ISBN 9781479903108. ISSN 15566013

[145] J. Lowe and R. Derakhshani, “Optokinetic response for mobile device biometric liveness assessment,” *Image and Vision Computing*, p. 104107, 2021. doi: https://doi.org/10.1016/j.imavis.2021.104107. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885621000123

[146] L. S. Gottfredson, “Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography,” *Intelligence*, vol. 24, no. 1, p. 13–23, 1997. doi: 10.1016/S0160-2896(97)90011-8. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0160289697900118

[147] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–44, 05 2015. doi: 10.1038/nature14539

[148] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in *ISCAS*, 2010, p. 253–256.

[149] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, p. 1097–1105.

[150] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, “Deep convolutional neural network for image deconvolution,” in *Advances in neural information processing systems*, 2014, p. 1790–1798.

[151] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, p. 85–117, 2015.

[152] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, p. 807–814.

[153] C. Dai, X. Liu, and J. Lai, “Human action recognition using two-stream attention based LSTM networks,” *Applied Soft Computing*, vol. 86, p. 105820, 2020. doi: https://doi.org/10.1016/j.asoc.2019.105820. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494619306015

[154] A. Diba, M. Fayyaz, V. Sharma, M. M. Arzani, R. Yousefzadeh, J. Gall, and L. Van Gool, “Spatio-temporal Channel Correlation Networks for Action Classification,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018. doi: 10.1007/978-3-030-01225-0_18. ISBN 9783030012243. ISSN 16113349

[155] S. Ji, W. Xu, M. Yang, and K. Yu, “3D Convolutional neural networks for human action recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013. doi: 10.1109/TPAMI.2012.59

VITA

Jesse Lowe

Jesse Michael Lowe was born on November 6, 1981, in Fort Madison, Iowa. His family relocated frequently and he was educated at home in his early life with the help of various tutors. He later attended private school, graduating from Kansas City based Blue Ridge Christian School in 2000. His life was significantly altered when his father was involved in an industrial accident that same year suffering a brain injury, spinal fracture, and various other injuries. Over the following years, his experience with his father's recovery fostered an interest in medicine, biomedical technology, and neuroscience which he pursued by enrolling at Longview Community College in Lee's Summit Missouri. He later transferred to the University of Missouri-Kansas City where in 2014 he received his Bachelor of Science in Biology.

His initial goal to attend medical school shifted during the last year of his undergraduate studies when he came to a realization that he likely had more to offer by leveraging his technical skills to help improve medical technology than he would by practicing medicine. After consulting with several trusted mentors and teachers, he was encouraged to pursue graduate study in engineering. In 2015 he was accepted as a member of University of Missouri-Kansas City CIBIT Lab where he began work on a Ph.D. in Electrical and Computer Engineering.

Mr. Lowe was extensively involved in teaching alongside his work as research assistant throughout most of his graduate studies. During this time he developed an interest in education and developing tools to enhance learning. In April 2019 he was recognized as an Outstanding Doctoral Student by the University of Missouri-Kansas City Department of Electrical and Computer Engineering. Upon completion of his degree requirements, Mr. Lowe plans to pursue his various reseach interests and seek areas where he can share his skills and knowledge to help develop new tools, treatments, and devices.

Mr. Lowe is a member of the IEEE, the Society for Neuroscience, the Eta Kappa Nu (IEEE-HKN) Honor society, and the Phi Kappa Phi Honor society.