Attack of Mechanical Replicas: Liveness Detection with Eye Movements

Oleg V. Komogortsev, *Member*, *IEEE*, Alexey Karpov, *Member*, *IEEE*, and Corey D. Holland, *Student Member*, *IEEE*

***Abstract*—This paper investigates liveness detection techniques in the area of eye movement biometrics. We investigate a specific scenario, in which an impostor constructs an artificial replica of the human eye. Two attack scenarios are considered: 1) the impostor does not have access to the biometric templates representing authentic users, and instead utilizes average anatomical values from the relevant literature; 2) the impostor gains access to the complete biometric database, and is able to employ exact anatomical values for each individual. In the current work, liveness detection is performed at the feature- and match score-levels for several existing forms of eye movement biometric, based on different aspects of the human visual system. The ability of each technique to differentiate between live and artificial recordings is measured by its corresponding false spoof acceptance rate (FSAR), false live rejection rate (FLRR), and classification rate (CR). The results suggest that eye movement biometrics are highly resistant to circumvention by artificial recordings when liveness detection is performed at the feature-level. Unfortunately, not all techniques provide feature vectors that are suitable for liveness detection at the feature-level. At the match score-level, the accuracy of liveness detection depends highly on the biometric techniques employed.**

***Index Terms*—Biometrics, liveness detection, spoofs, attack vectors, eye movements, pattern analysis, security and protection.**

I. Introduction

LIVENESS DETECTION is an important problem in the biometric domain, due to the fact that it is relatively simple to create convincing replicas of many existing biometrics. For example, commercial iris identification systems can be spoofed by high-resolution images of the eye printed on paper, with a hole to present the intruder's pupil, bypassing liveness detection mechanisms [1, 2]. There are further examples of fingerprint scanners being spoofed by common household items like gelatin [3], and face detection systems spoofed by printed images of the face [4-6].

Manuscript received February 23, 2015. This work is supported in part by NSF CAREER Grant #CNS-1250718 and NSF GRFP Grant #DGE-11444666, and NIST Grant #60NANB12D234.

O. V. Komogortsev is with the Computer Science Department, Texas State University, San Marcos, TX 78666 USA (phone: 512-245-0349; fax: 512-245-8750; e-mail: ok11@txstate.edu).

A. Karpov is with the Computer Science Department, Texas State University, San Marcos, TX 78666 USA (e-mail: ak26@txstate.edu).

C. D. Holland is with the Computer Science Department, Texas State University, San Marcos, TX 78666 USA (e-mail: ch1570@txstate.edu).

While liveness detection has been researched extensively in a number of fields, there is almost no research on the topic of liveness detection as it relates to the field of eye movement biometrics. The idea of liveness detection via eye movements is appealing because eye movements can be captured in tandem with iris information using a single image sensor [7]. In fact, the hardware in many existing iris recognition devices is capable of supporting modern video-oculography techniques to produce an eye movement signal. This provides an opportunity to increase the accuracy [8] and counterfeit-resistance of existing iris recognition devices.

A. Related Work in Liveness Detection

In the field of iris recognition, several liveness detection methods have been proposed and evaluated, involving: frequency spectrum analysis of the iris image; analysis of light reflected from the spherical corneal surface (front wall of the eye); detection of corneal moistness; pupil dilation; and the quality-related features of the captured image [9-13].

In the field of fingerprint recognition, Coli and colleagues provide a survey of existing liveness detection techniques and their performance [14]. Shuckers et al. achieved 90-100% correct classification rate using the statistics of ridges and moisture to measure the liveness of fingerprints [15, 16]. Ghiani et al. applied liveness detection by local binary patterns, pores, power spectrum analysis, wavelet energy signatures, valleys wavelets, and curvelets to achieve equal error rates of 7-13% [17]. Recent competitions indicate that fingerprint-based biometrics are still susceptible to spoofing [18].

In the field of face recognition, liveness detection methods can be roughly categorized as analysis of motion, texture, and detection of life signs. Competition results of various algorithms in this area indicate a high accuracy of detection for 2D spoofing attacks [19], particularly in regard to the PRINT-ATTACK dataset [4].

B. Related Work in Eye Movement Biometrics

In the current work, we explore the liveness detection properties of four existing eye movement biometric techniques, based on various aspects of the human visual system. These techniques include: oculomotor plant characteristics (OPC), complex oculomotor behavior (COB), complex eye movement patterns (CEM-P), and complex eye movement behavior (CEM-B) biometrics.

In 2011, Holland and Komogortsev [20] described complex eye movement pattern (CEM-P) biometrics. CEM-P compares

average and aggregate measures of the eye movement scanpath with a distance function to estimate the similarity of two recordings. Biometric features include: fixation count, average fixation duration, average saccadic amplitude, average saccadic velocity, average saccadic peak velocity, velocity waveform indicator (Q), scanpath length, scanpath area, regions of interest, inflection count, amplitude-duration coefficient, and main sequence coefficient.

In 2012, Komogortsev et al. [21] described oculomotor plant characteristic (OPC) biometrics. OPC utilizes mathematical models of the oculomotor plant to estimate the non-visible anatomical properties of the eye from the measurable properties of eye movements. Feature vectors are compared between recordings using the Hotelling's *T*-square test to obtain a measure of similarity. Biometric features vary according to the parameters of the oculomotor plant model employed.

In 2013, Holland and Komogortsev [22] described complex eye movement behavior (CEM-B) biometrics. CEM-B compares the distribution of basic eye movement features throughout a recording using statistical techniques, such as the Kolmogorov-Smirnov test, to obtain a measure of similarity between recordings. Biometric features include: fixation start time, fixation duration, fixation centroid, saccade start time, saccade duration, saccadic amplitude, saccadic velocity, and saccadic peak velocity.

In 2013, Komogortsev and Holland [23] described complex oculomotor behavior (COB) biometrics. COB examines the relative amount of corrective eye movements that occur in response to artifacts or otherwise aberrant behavior of the human visual system. Feature vectors are compared with a distance function to obtain a measure of similarity between recordings. Biometric features include: uncorrected, corrected, and multi-corrected saccadic dysmetria, express saccades, dynamic saccades, and compound saccades.

C. Motivation & Hypothesis

The foundation for eye movement biometrics was formed in 1971, when Noton and Stark [24] found that the eye movements exhibited by a subject during the initial viewing of a pattern were repeated in 65% of subsequent viewings. The potential biometric applications of eye movements were largely ignored, however, until 2004, when Kasprowski and Ober [25] applied voice recognition techniques to the eye movement signal. Over the past decade, the field of eye movement biometrics has expanded at a rapid pace, but there are still many aspects that lack definition.

The potential liveness detection properties of human eye movements have yet to be investigated by rigorous experimentation. This paper builds on our previous research [26] on the liveness detection properties of eye movement biometrics to include: 1) additional models of the oculomotor plant; 2) an extended database of 173 living and artificial recordings; 3) recordings conducted on equipment with properties resembling those of commercial iris recognition devices; 4) new statistical methods that allow for more accurate liveness detection; and 5) additional forms of eye movement biometric.

II. Biometric Attack Vector

The potential attack vectors for eye movement biometrics are limited, and may consist of physical or graphical representations. For a physical representation, the impostor must construct a robotic and anatomically convincing model of the human eye, and for a graphical representation, the impostor may utilize a graphics-generated model of the human eye presented on a display medium, such as a phone. Assuming that both representations of an artificial eye can be calibrated[1] by the eye tracking system, and bypass existing liveness detection techniques based on image analysis, the aim of our algorithms is to identify these impostor recordings based on features of the eye movement signal, including its variability.

Mathematical modeling of the oculomotor plant simulates both representations, utilizing several models of varying complexity to represent different classes of artificial eye that an impostor could employ. As a result, the presented baselines are not contaminated by the inherent noise associated with eye tracking equipment; however, the artificial eye movement signal is generated in such a way as to contain typical abnormalities and artifacts that occur naturally within the human visual system.

A. Human Visual System

The human visual system is composed, primarily, of two major components: the oculomotor plant and the brainstem control. The oculomotor plant encompasses the eye globe, six extraocular muscles, and a variety of surrounding tissues, ligaments, and fluids. The brainstem control is responsible for the generation and transmission of a neuronal control signal, sent to each of the extraocular muscles to produce the many and varied types of human eye movement [27].

For our purposes, two basic types of eye movement are of particular importance, due to the ease with which they can be evoked and replicated in humans via pre-programmed stimuli; these are: fixations and saccades. "Fixations occur when the eye globe is held in a relatively stable position, such that the fovea remains centered on an object of interest, providing heightened visual acuity; saccades occur when the eye globe rotates quickly between points of fixation, with very little visual acuity maintained during rotation." [28]

In addition to these basic eye movements, the human visual system exhibits a number of corrective eye movements that occur naturally due to saccadic dysmetria or otherwise aberrant behavior [23, 27]. These corrective behaviors often follow an off-target saccade with one or more small-amplitude saccades in the direction of the target.

It is assumed that an artificial eye would be capable of exhibiting basic eye movements, such as fixations and saccades, interspersed with corrective eye movements at natural frequencies and amplitudes found in the human visual system.

[1] Calibration is the process of detecting the position of the pupil and corneal reflection for a selected number of points, presented on the screen in known locations, in order to accurately interpolate the coordinates of future gaze locations.

B. Oculomotor Plant Mechanical Models

Mathematical models of varying complexity can represent the mechanical functions of the oculomotor plant to simulate the dynamics of human eye movement; that is, in addition to the eye movement signal, the employed models simulate the physical muscle mechanics involved in the generation of eye movements. By optimizing the parameters of a given model to the recordings of a specific individual, it is possible to generate artificial recordings using the model. These false recordings are used as the biometric attack vector in the context of eye movement biometrics. Three distinct models were utilized in the current work, as follows:

- **Model I** is Westheimer's second-order model [29]. This model represents the eye globe and corresponding viscoelasticity by single linear elements for inertia, friction, and stiffness.
- **Model II** is Robinson's fourth-order model [30]. This model uses a more realistic neuronal control signal, in the pulse-step form.
- **Model III** is Komogortsev and Khan's fourth-order model [31, 32]. This model extends Bahill's model [33], representing each extraocular muscle and their internal forces individually, with a separate pulse-step neuronal control signal provided to each muscle.

To obtain the model parameters of a given individual, the raw positional signal of an eye movement recording is classified into fixations and saccades. Each recorded saccade is matched by a simulated saccade, generated via an oculomotor plant mathematical model (OPMM), with the goal of minimizing the absolute error between the measured and simulated saccade trajectories. Then, each saccade in a given recording generates a unique set of model parameters, also known as oculomotor plant characteristics (OPC).

C. Spoofing Strategies

There are two degrees of attack that an impostor could potentially employ when generating an artificial eye movement recording. The naïve approach assumes that the impostor does not have access to the biometric database, while the sophisticated approach assumes that the impostor has gained access to one or more OPC vectors from the database.

- **Approach A** is the naïve approach, and assumes that the impostor does not have access to the biometric database, and instead generates artificial eye movements using measured or derived OPC values reported in the relevant research literature. Two artificial recordings are generated for each stimulus, and tested against the entire biometric database.
- **Approach B** is the sophisticated approach, and assumes that the impostor has gained full access to the biometric database, and is able to generate artificial eye movements specific to an individual using the previously extracted OPC values now stored in the database. An artificial recording is generated for each live recording, and tested against the entire biometric database.

In the current work, we apply **Approach A to Models I, II, and III**, and we apply **Approach B only to Model III**, as this is the only model that can support individualized replication. Individualization of Models I and II lead to eye movement trajectories that are very different from normal, and can therefore be easily rejected due to their abnormalities. This means that two spoof recordings were generated with each model for Approach A, and an equal number of artificial recordings to live recordings for Approach B.

III. Methodology

Eye movement recordings were collected from two disjoint subject pools on high- and low-resolution eye tracking systems. The high-resolution eye tracking system resembles the state-of-the-art in current eye tracking technology, while the low-resolution eye tracking system applies video-oculography techniques and resembles hardware found in current iris recognition devices. The collected eye movement datasets are available as part of the EMDB database [34, 35].

A. Participants

High-resolution recordings [35] were conducted on a subject pool of 32 participants (26 male, 6 female), with ages ranging from 18 – 40, average age 23 (SD = 5.4). 29 of the subjects performed 4 recordings each, and 3 of the subjects performed 2 recordings each, generating a total of 122 unique eye movement recordings.

Low-resolution recordings [34] were conducted on a subject pool of 173 participants (117 male, 56 female), with ages ranging from 18 – 49, average age 23 (SD = 5.3). 170 of the subjects performed 2 recordings each, and 3 of the subjects performed 1 recording each, generating a total of 343 unique eye movement recordings. Note, the last 6 participants, each of which performed 2 recordings, were excluded from the sophisticated spoofing approach due to scheduling difficulties.

B. Apparatus & Software

High-resolution recordings were taken with a desktop mounted EyeLink 1000 commercial eye tracking system [36], with a sampling rate of 1000 Hz, vendor-reported spatial accuracy of 0.5°, average calibration accuracy of 0.7° (SD = 0.5°), and average data validity of 95% (SD = 5%). Stimuli were presented on a flat screen monitor positioned at a distance of 685 mm from each subject, with dimensions of 640×400 mm, and screen resolution of 2560×1600 pixels.

Low-resolution recordings were taken with the PlayStation Eye Camera [37], using a modified version of the open-source ITU Gaze Tracker software [7], with a sampling rate of 75 Hz and average calibration accuracy of 1.1° (SD = 0.8°). Average data validity is the percentage of gaze points reported by the eye tracking system that contained valid eye movement data, and is unreportable in this instance as it was not possible to detect when the eye tracker began tracking an area of the image other than the subject pupil. Stimuli were presented on a flat screen monitor positioned at a distance of 540 mm from each subject, with dimensions of 375×302 mm, and screen resolution of 1280×1024 pixels.

In both cases, the pupil was illuminated by infrared LED to improve eye tracking accuracy, and a chin rest was employed to improve stability. Stimulus presentation was consistent across both devices, with only minor changes required to accommodate varied screen dimensions. Algorithms and data analysis were implemented and conducted in MATLAB.

C. Procedure

For each recording session, eye movements were evoked using a horizontal saccade stimulus, in which a small white dot jumps back and forth across a plain black background, eliciting a fixed-amplitude saccade with each jump. The distance between jumps was set to correspond to 30° of the visual angle, due in part to screen constraints, complications separating low-amplitude saccades (less than 1°), and variation in the dynamics of high-amplitude saccades (greater than 50°). Subjects were instructed to follow the white dot with their eyes, with 100 horizontal saccades elicited per session.

After the data collection process, OPC values were extracted from each live recording. According to the methods described in Section II, artificial recordings were generated for three oculomotor plant models and two spoofing strategies. The live and artificial recordings were then combined into a single dataset for each eye tracking systems, and liveness detection was conducted at the feature- and match score-level. For comparison, liveness detection was also conducted using the techniques proposed by Komogortsev and Karpov in 2013.

Previous techniques developed by Komogortsev and Karpov perform principal component analysis on biometric feature vectors to obtain the eigenvector, and compare the maximum eigenvalue against a fixed threshold to determine the liveness of a given recording [26]. In the original paper, liveness thresholds ranged from 1000 to 6000 in increments of 100. On a sample of 122 recordings from 32 subjects, this technique was able to achieve a maximum classification rate of 93% and a minimum equal error rate of 5% in previous studies using 10-fold cross validation.

Feature-level liveness detection utilized a regression SVM [38] with RBF kernel (gamma = 1) and leave-one-out cross-validation for the selection of training and testing sets. Biometric feature vectors were calculated for each recording session, and each feature vector was labeled as live or spoof. Feature vectors were classified, one at a time, after training the SVM on all other recordings (exclusively). An ideal threshold was then selected to maximize classification rate. For both the OPC and CEM-B techniques, which utilize multi-dimensional feature vectors, the eigenvector obtained by principal component analysis was passed to the SVM.

Score-level liveness detection applied biometric matching techniques across all recordings, and utilized leave-one-out cross-validation for the selection of training and testing sets. Biometric matching was conducted according to the standards suggested by each technique, with match score-level information fusion applied by a regression SVM [38] with RBF kernel (gamma = 1); however, instead of labeling and matching according to the identity of the subject, each recording was labeled and matched as either a live or spoof recording. In the case of the OPC technique, the Cramér-von Mises variant of feature comparison was used to generate match scores for individual features preceding information fusion. For each recording, biometric match scores were generated against all other recordings in the database, after training the biometric algorithms on all other recordings (exclusively). An ideal threshold was then selected to minimize classification error.

Leave-one-out cross-validation performed a number of repetitions, n, equal to the total number of recordings in a given dataset. For Approach A, there were 2 spoof recordings for each model, with 122 high-resolution recordings (n = 122 + 2 = 124) and 343 low-resolution recordings (n = 343 + 2 = 124). For Approach B, there was 1 spoof recording for each live recording, with 122 high-resolution recordings (n = 122 + 122 = 244) and 343 low-resolution recordings, less 24 recordings that could not be replicated by the mathematical model (n = 343 + 343 – 24 = 662). Due to the exponential growth in time and space complexity under score-level liveness detection, it was necessary to reduce the size of the SVM training set of the low-resolution recordings for spoofing technique III-B. For each iteration of cross-validation, exactly ¼ of the available subject pool was selected for the training set. This lead to the same number of total comparisons, but in each case the training set for the SVM was reduced from 438,244 match scores to 27,390 match scores, making computation tractable.

In relation to both feature- and score-level liveness detection we are primarily concerned with three metrics: false spoof acceptance rate (FSAR), false live rejection rate (FLRR), and classification rate (CR) [26]. In both cases, the OPC technique was omitted from the low-resolution recordings, due to the fact that it would take prohibitively long to re-extract biometric templates from the spoof recordings.

$$\mathrm{FSAR} = \frac{\text{Misclassified Spoof Recordings}}{\text{Total Spoof Recordings}} \quad \textbf{(1)}$$

$$\mathrm{FLRR} = \frac{\text{Misclassified Live Recordings}}{\text{Total Live Recordings}} \quad \textbf{(2)}$$

$$\mathrm{CR} = \frac{\text{Correctly Classified Recordings}}{\text{Total Recordings}} \quad \textbf{(3)}$$

$$\mathrm{EER} = (\mathrm{FSAR} = \mathrm{FLRR}) \quad \textbf{(4)}$$

False spoof acceptance rate (FSAR), shown in Equation 1, is the ratio of misclassified spoof recordings to the total number of spoof recordings. False live rejection rate (FLRR), shown in Equation 2, is the ratio of misclassified live recordings to the total number of live recordings. Classification rate (CR), shown in Equation 3, is the ratio of correctly classified recordings to the total number of recordings (either live or spoof). The equal error rate (EER), shown in Equation 4, is the rate at which false spoof acceptance rate and false live rejection rate are equal, and does not necessarily correspond to the threshold at which the classification rate is maximized.

IV. Results

Results are labeled according to the oculomotor plant model and spoofing strategies employed. For example, "III-B" indicates that the results pertain to Model III and Approach B. An overview of the accuracy of previous liveness detection techniques [26] is provided in Table I. An overview of feature-level liveness detection accuracy is provided in Table II. An overview of match score-level liveness detection is provided in Table III. The following section provides further explanation and analysis of the obtained results.

V. Discussion

The results presented in the current paper represent a major improvement on our previous research [26] in the area of liveness detection. Where, in the previous paper, the best performing attempt at liveness detection achieved only a 5% equal error rate [26], using the naïve approach and the simplest model on a high-resolution eye tracking system, the techniques presented in this paper have shown that liveness detection at the feature-level is capable of achieving 0% equal error rate (see Table II) even when the impostor has gained access to the biometric database and employs complex models of the human visual system.

A. Comparison to Previous Techniques

Where our previous research [26] examined only the liveness detection properties of the OPC biometric, using broad thresholds and random splitting, the current paper has applied those techniques to several additional forms of eye movement biometric, using fine-grained threshold selection and leave-one-out cross validation. While this has shown that our previous techniques are capable of achieving higher accuracy than indicated by previous studies, utilizing the maximum eigenvalue ignores much meaningful information, and in some cases is clearly the wrong approach, leading to equal error rates which exceed 50% (see Table I).

In comparison, the techniques presented in this paper are more accurate at both the feature- and match score-levels. For example, utilizing the full eigenvector for classification by support vector machine leads to near-perfect accuracy in the OPC and CEM-B biometrics (see Table II). Further, previous techniques fail to achieve acceptable accuracy on all but the simplest models, and have a tendency toward high false spoof acceptance rate (see Table I).

B. Liveness Detection

There are obvious and immediate differences between the accuracy of liveness detection at the feature- and match score-levels. While liveness detection at the feature-level achieves *almost* perfect accuracy under all tested conditions (see Table II), match score-level detection does substantially worse (see Table III). This may be due to any of several factors.

It is possible that the differences in accuracy may be due to overfitting at the feature-level, resulting from an overall smaller knowledge base. That is, at the feature-level there are (N-1) training vectors for the SVM at each iteration, whereas there are $(N-1)^2$ training vectors at the match score-level, where N is the total number of live and spoof recordings in the dataset. This could indicate that match score-level detection rates are a more accurate representation of the liveness detection capabilities of eye movement biometrics; however, the only way to confirm this would be to apply feature-level detection to a much larger subject pool.

More likely, however, is the fact that information is simply lost when the biometric feature vectors are converted into a relevant match score. Essentially, by combining groups of biometric feature vectors into a single value, noise is added to the system. This makes it more difficult for the system to differentiate between recordings. For example, at the feature-level, there is no crossover in the information provided by live recordings and spoof recordings, but match scores may contain information from both.

High-Resolution Recordings (HR)					
Biometric	***Spoof***	***FSAR***	***FLRR***	***CR***	***EER***
OPC	I-A	100%	1%	98%	4%
	II-A	100%	1%	98%	18%
	III-A	100%	1%	96%	80%
	III-B	29%	32%	70%	31%
COB	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	0%	1%	99%	1%
	III-B	2%	1%	99%	2%
CEM-P	I-A	100%	1%	98%	2%
	II-A	100%	1%	98%	11%
	III-A	100%	1%	96%	51%
	III-B	2%	61%	69%	51%
CEM-B	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	100%	1%	96%	50%
	III-B	14%	66%	61%	46%
Low-Resolution Recordings (LR)					
Biometric	***Spoof***	***FSAR***	***FLRR***	***CR***	***EER***
COB	I-A	100%	0%	99%	97%
	II-A	100%	0%	99%	97%
	III-A	100%	0%	99%	97%
	III-B	2%	26%	86%	21%
CEM-P	I-A	100%	0%	99%	13%
	II-A	100%	0%	99%	26%
	III-A	100%	0%	99%	29%
	III-B	1%	30%	84%	29%
CEM-B	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	100%	0%	99%	57%
	III-B	0%	1%	99%	1%

Table I. Liveness detection accuracy of previous techniques.

C. Biometric Attack Vectors

When examining the spoofing strategies employed, there is obviously very little difference at the feature-level. With the exception of spoofing strategy III-B on the low-resolution recordings, feature-level liveness detection was able to accurately distinguish between live and spoof recordings (see Table II). For this reason, it is necessary to examine classification accuracy at the match score-level to identify differences in spoofing models and approaches.

At the match score-level, there is an obvious tendency for the naïve spoof approach (A) to result in a relatively high percentage of false spoof acceptance, where the sophisticated spoof approach (B) results in a relatively high percentage of false live rejection (see Table III), at the optimal classification rate. Under the naïve approach, this occurs because there are a substantially smaller number of spoof recordings than live recordings, which leads to a strategy that maximizes classification rate by minimizing the live rejection rate. As such, this can be seen as an artifact of the methodology, not necessarily the techniques. The classification rates obtained under the sophisticated approach, in which the number of live and spoof samples are equal, more accurately resemble real-world usage. Overall, however, equal error rate is a better indicator of accuracy than classification rate, because it is not affected by the relative amount of live and spoof samples.

There is a further tendency for detection accuracy to decrease linearly with model complexity (see Tables I, II, and III). This is notable in the false spoof acceptance rates and equal error rates of the naïve approach. Westheimer's second-order model, having the least complexity (and thereby the least realistic signal), produced the most easily distinguished spoof recordings, whereas Komogortsev and Khan's fourth-order model, with its representation of individual muscle properties, performed more successful attacks at the match score-level than any other oculomotor plant model.

D. Eye Movement Biometrics

There were several notable differences in the liveness detection properties of the considered biometric techniques. Of the considered biometrics, the CEM-P technique was overall the most accurate in distinguishing live recordings from spoof recordings. While COB and OPC provided similar accuracy at the feature-level, CEM-P outperformed both at the match score-level. Despite the fact that CEM-B performed relatively well at the match score-level, with similar accuracy to the CEM-P technique, the inability to perform liveness detection at the feature-level is seen as a major shortcoming.

The computational overhead associated with the OPC technique makes it unsuitable for real-time liveness detection, and intractable for experimentation on large datasets. At the same time, however, this can be seen as an advantage to the liveness detection properties of eye movements in general. It is only possible to generate artificial eye movements by modeling the human visual system, whether at the software level or even with a hardware counterpart. Existing models of the oculomotor plant fail to represent the human visual system with perfect accuracy, and tradeoffs must be made. Further, the time requirements of even these computationally efficient linear models make it infeasible for a potential impostor to generate more than a few targeted artificial recordings in a realistic timeframe with modern technology.

E. Eye Tracking System

With few exceptions, the decreased spatial accuracy and sampling rate of the low-resolution recordings reduced the overall accuracy of liveness detection. This, in itself, is not surprising; however, some interesting connections can still be made. Under the naïve spoofing approach, in which the impostor does not have access to the biometric database, feature-level liveness detection maintained perfect accuracy on the low-resolution system, and low equal error rates even at the match-score level.

High-Resolution Recordings (HR)					
Biometric	***Spoof***	***FSAR***	***FLRR***	***CR***	***EER***
OPC	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	0%	0%	100%	0%
	III-B	0%	0%	100%	0%
COB	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	0%	0%	100%	0%
	III-B	0%	0%	100%	0%
CEM-P	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	0%	0%	100%	0%
	III-B	0%	0%	100%	0%
CEM-B	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	0%	0%	100%	0%
	III-B	2%	0%	99%	2%
Low-Resolution Recordings (LR)					
Biometric	***Spoof***	***FSAR***	***FLRR***	***CR***	***EER***
COB	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	0%	0%	100%	0%
	III-B	13%	0%	94%	13%
CEM-P	I-A	0%	0%	100%	0%
	II-A	0%	0%	100%	0%
	III-A	0%	0%	100%	0%
	III-B	40%	0%	80%	40%
CEM-B	I-A	0%	2%	98%	2%
	II-A	0%	2%	98%	2%
	III-A	0%	2%	98%	2%
	III-B	35%	3%	81%	35%

Table II. Feature-level liveness detection accuracy.

There is, however, a dramatic reduction in accuracy on the low-resolution system when coupled with the sophisticated spoofing approach. When we consider that the low-resolution recordings, taken at 75 Hz, contain less than $1/13^{th}$ of the information present in the high-resolution recordings, taken at 1000 Hz, it is perhaps surprising that the differences in detection accuracy are not more drastic.

In some instances, we even notice that detection accuracy *increases* on the low-resolution system (see Table III), particularly with regard to spoofing strategy III-A. In this case, it may be that the artificial recordings failed to accurately reflect the inherent noise and lower accuracy of the low-resolution system. This suggests that, even if an oculomotor plant model were to achieve perfect reproduction of human eye movements, software-level attacks must also account for the inherent properties of the capture device as well. While this would not be the case for functional physical replicas, the construction of such a replica poses its own difficulties.

F. Vulnerability Analysis

While we have demonstrated that it is possible to identify spoof eye movement recordings with high accuracy, it is still necessary to demonstrate that spoofed recordings present a viable threat to eye movement biometrics, in order to justify the need for liveness detection techniques. Attack vectors are typically targeted at biometric verification systems in an attempt to gain access to the resources of a specific individual. For this reason, we are primarily interested in the rate at which spoof recordings are accepted as genuine users in a verification scenario.

To this end, biometric error rates were calculated on the high- and low-resolution data sets using the standard biometric techniques for OPC [21], COB [23], CEM-P [20], and CEM-B [22] biometrics, as described in the relevant references, without the introduction of spoof recordings. For the OPC technique, information fusion was performed using the Hotelling's *T*-square test; for the COB technique, information fusion was performed using likelihood ratio; for the CEM-P technique, information fusion was performed using linear combination; and for the CEM-B technique, information fusion was performed using a 50-tree random forest.

This was done to identify the acceptance threshold at which false acceptance rate (the rate at which impostor match scores exceed the acceptance threshold) and false rejection rate (the rate at which genuine match scores fall below the threshold) are equal. Equal error rates and acceptance thresholds were averaged over 20 random partitions, in which half of all subjects were used for training and half were used for testing, without subject overlap. It should be noted that equal error rate used in this context is different than the equal error rate used elsewhere in this paper, as this term is used in relation to biometric accuracy rather than liveness detection accuracy.

Spoof recordings were then introduced to identify the false spoof acceptance rate at the equal error rate of the biometric system. Biometric match scores were generated for comparisons for each spoof recording to every live recording. Match scores were averaged over 20 random partitions, and the average match scores for spoof recording to live recording comparisons were used to calculate the false spoof acceptance rate with and without feature-level liveness detection.

The results of this experiment are provided in Table IV, indicating the number of spoof attacks which generated a match score above the acceptance threshold, the number of spoof attacks which generated a match score below the acceptance threshold, and the false spoof acceptance rate for a given attack vector against a given biometric technique.

It is worth noting that, in several cases, false spoof acceptance rate increases with the addition of feature-level liveness detection. This is due to the overall lower number of spoof recordings which make it to the decision module. For example, if there are 2 spoof recordings and 1 is accepted as genuine, there is a false spoof acceptance rate of 50%, whereas if there are 10 spoof recordings and 4 are accepted as genuine, there is a false spoof acceptance rate of only 40%, despite the fact that 4 times as many spoofs were accepted. In most cases where false spoof acceptance rate increases, the number of spoof recordings with match scores above the acceptance threshold is actually reduced with the inclusion of feature-level liveness detection, with one notable exception.

High-Resolution Recordings (HR)					
Biometric	***Spoof***	***FSAR***	***FLRR***	***CR***	***EER***
OPC	I-A	57%	1%	98%	10%
	II-A	74%	0%	97%	24%
	III-A	68%	1%	97%	16%
	III-B	9%	57%	79%	26%
COB	I-A	68%	0%	97%	22%
	II-A	69%	0%	98%	23%
	III-A	100%	0%	97%	24%
	III-B	6%	34%	87%	17%
CEM-P	I-A	1%	0%	100%	0%
	II-A	3%	0%	100%	1%
	III-A	100%	0%	97%	19%
	III-B	8%	26%	88%	14%
CEM-B	I-A	1%	0%	100%	0%
	II-A	1%	0%	100%	0%
	III-A	100%	0%	97%	18%
	III-B	4%	35%	88%	16%
Low-Resolution Recordings (LR)					
Biometric	***Spoof***	***FSAR***	***FLRR***	***CR***	***EER***
COB	I-A	96%	0%	99%	28%
	II-A	100%	0%	99%	29%
	III-A	93%	0%	99%	18%
	III-B	2%	8%	96%	5%
CEM-P	I-A	3%	0%	100%	1%
	II-A	11%	0%	100%	2%
	III-A	37%	0%	99%	3%
	III-B	2%	3%	98%	2%
CEM-B	I-A	60%	0%	99%	2%
	II-A	99%	0%	99%	4%
	III-A	100%	0%	99%	5%
	III-B	1%	6%	98%	3%

Table III. Match score-level liveness detection accuracy.

With the use of feature-level liveness detection via CEM-B for low-resolution recordings, the number of III-B spoof recordings with match scores above the acceptance threshold increases. Most likely this is due to the fact that biometric matching trains on a subset of the population. Without feature-level liveness detection, there is a greater population of spoof recordings to train against, and thus the CEM-B algorithm is less likely to generate a high match-score for comparisons of spoof recordings to live recordings. When, presumably the weakest, spoof recordings are removed from the population by feature-level liveness detection, the remaining spoofs are more likely to generate a higher match score against live recordings.

Despite a low spoof acceptance rate, even for targeted attacks, the need for explicit liveness detection techniques is apparent. While human eye movements expose a number of properties that allow for a high degree of counterfeit-resistance, they are not inherently immune to attacks. The research presented in this paper provides a starting point for reducing the potential attack surface.

High-Resolution Recordings (HR)								
Biometric	***Spoof***	***EER***	***Without Liveness Detection***			***Feature-Level Detection***		
			Above Threshold	***Below Threshold***	***FSAR***	***Above Threshold***	***Below Threshold***	***FSAR***
OPC	I-A	18%	8	236	3%	0	0	0%
	II-A		0	244	0%	0	0	0%
	III-A		8	236	3%	0	0	0%
	III-B		208	14676	1%	0	0	0%
COB	I-A	30%	22	222	9%	0	0	0%
	II-A		18	226	7%	0	0	0%
	III-A		21	223	9%	0	0	0%
	III-B		1207	13677	8%	0	0	0%
CEM-P	I-A	38%	45	199	18%	0	0	0%
	II-A		0	244	0%	0	0	0%
	III-A		57	187	23%	0	0	0%
	III-B		1930	12954	13%	0	0	0%
CEM-B	I-A	23%	5	239	2%	0	0	0%
	II-A		8	236	3%	0	0	0%
	III-A		5	239	2%	0	0	0%
	III-B		104	14780	1%	8	237	3%
Low-Resolution Recordings (LR)								
Biometric	***Spoof***	***EER***	***Without Liveness Detection***			***Feature-Level Detection***		
			Above Threshold	***Below Threshold***	***FSAR***	***Above Threshold***	***Below Threshold***	***FSAR***
COB	I-A	41%	26	660	4%	0	0	0%
	II-A		18	668	3%	0	0	0%
	III-A		29	657	4%	0	0	0%
	III-B		1467	112066	1%	1201	12536	9%
CEM-P	I-A	38%	7	679	1%	0	0	0%
	II-A		0	686	0%	0	0	0%
	III-A		42	644	6%	0	0	0%
	III-B		5500	108033	5%	238	445	35%
CEM-B	I-A	31%	38	648	6%	0	0	0%
	II-A		24	662	4%	0	0	0%
	III-A		41	645	6%	0	0	0%
	III-B		103	113430	0%	3044	42370	7%

Table IV. Biometric vulnerability with and without liveness detection.

G. Further Analysis

Many existing iris recognition devices are not designed for video capture, and often have hardware limitations that cap the effective sampling rate between 15 Hz and 30 Hz. To examine liveness detection accuracy under these conditions, recordings from both the high- and low-resolution eye tracking systems were downsampled to a sampling rate of 15 Hz by removing data points at uniform intervals to lower the average time between points.

Again, the OPC technique was omitted from both feature- and match score-level detection at 15 Hz due to time constraints, and the CEM-B technique was inapplicable to feature-level detection. Further, the CEM-B technique failed to extract features once downsampling to 15 Hz was applied (that is, an error occurred in the algorithm itself, unrelated to classification rate). This leaves analysis open to only the CEM-P and COB techniques.

At the feature-level, the CEM-P technique maintained the same accuracy achieved at 1000 Hz, while the COB technique showed only minor reduction in accuracy (1-3% equal error rate) in all but one case (in which equal error rate increased to 46% with strategy II-A). At the match-score level, the COB technique showed major accuracy loss across all strategies, approaching near random detection rates. This is expected, as COB features become less pronounced at reduced sampling rates. On the other hand, CEM-P was more resistant to these effects, with equal error rates increased by 10-20% under spoof strategies I-A and II-A, while spoof strategy III-A actually showed equal error rates *reduced* by 2-18% in all cases.

To examine the validity of the downsampling technique, the high-resolution recordings taken with the EyeLink 1000 were downsampled to 75 Hz, and liveness detection was compared to the accuracy achieved by the low-resolution PlayStation Eye recordings taken at a native 75 Hz. There was a general tendency for the downsampled high-resolution recordings to outperform the equivalent low-resolution recordings by 1-5% equal error rate. The relatively minor difference suggests that they may be accounted for by the reduced spatial accuracy of the PlayStation Eye camera (1.1° calibration accuracy vs. 0.7° calibration accuracy) or the differences in subject pool.

H. Limitations & Future Research

There are several notable limitations of the current work that must be addressed in future research. First, in the current work we have employed only linear models of the oculomotor plant due to computational constraints and the size of the available datasets; however, the human visual system exhibits a number of non-linear properties [39] that can only be approximated by linear models. To the best of our knowledge, there are no existing models of the oculomotor plant that achieve a perfect representation of the human visual system, but there do exist non-linear models capable of more accurate representation [40]. As technology advances, these models may become more viable as potential attack vectors, and further testing will be necessary to identify what threat, if any, they might pose.

As well, our classification threshold was selected empirically to maximize classification rate, in order to demonstrate the potential liveness detection properties of eye movement biometrics. In real-world usage, the classification threshold would be selected on a subset of available data, and would likely be far from ideal. As such, there is still much work to be done in this area, identifying accuracy and usage conditions in real-world scenarios. Further, a comparison of biometric authentication accuracy with and without attack vectors would be useful in quantifying the inherent spoofability of eye movement biometrics, before the application of anti-spoofing methods; however, such a comparison is beyond the scope of the current paper, and will likely be a topic of future research.

Another limitation of the current work is the purely software-based treatment of biometric attack vectors. Perhaps fortunately, the engineering required to produce a viable physical replica of the oculomotor plant (particularly, such a replica that could physically simulate multiple models of the human eye without loss of accuracy) made its employment in the current study infeasible. At some point in the future it will be necessary to revisit the possibility of the physical attack.

Finally, the relatively sparse treatment of the OPC biometric technique due to the time requirements of extracting biometric templates is seen as a disadvantage of the current study. It remains to be seen whether the OPC technique performs adequately for liveness detection at low sampling rates. This is an area of active research.

VI. Conclusion

This paper has investigated liveness detection techniques in the area of eye movement biometrics. We have investigated a specific scenario, in which an impostor constructs an artificial replica of the human eye. Two attack scenarios were considered, in which the imposter does and does not have direct access to the biometric database.

Liveness detection was performed at the feature- and match score-levels for several existing eye movement biometric techniques. The results suggest that eye movement biometrics are highly resistant to circumvention by artificial recordings when liveness detection is performed at the feature-level. At the match score-level, the accuracy of liveness detection depends highly on the biometric techniques employed.

These tests were repeated on high- and low-resolution recording devices, and while obvious degradation occurred at the match score-level, feature-level liveness detection achieved near perfect accuracy even at very low sampling rates. This suggests that eye movement biometrics could be employed on existing iris recognition devices to improve liveness detection capabilities.

References

[1] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, and J. Ortega-Garcia, "Direct Attacks Using Fake Images in Iris Verification," in *Biometrics and Identity Management*. vol. 5372, B. Schouten, N. C. Juul, A. Drygajlo, and M. Tistarelli, Eds., ed Roskilde, Denmark: Springer Berlin Heidelberg, 2008, pp. 181-190.

[2] L. Thalheim, J. Krissler, and P.-M. Ziegler, "Body Check: Biometric Access Protection Devices and Their Programs Put to the Test," *c't Magazine,* 2002.

[3] J. M. Williams, "Biometrics or… Biohazards?," in *2002 Workshop on New Security Paradigms*, 2002, pp. 97-107.

[4] A. Anjos and S. Marcel, "Counter-Measures to Photo Attacks in Face Recognition: a Public Database and a Baseline," in *International Joint Conference on Biometrics (IJCB)*, Washington, DC, 2011, pp. 1-7.

[5] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera," in *International Conference on Computer Vision (ICCV)*, Rio de Janeiro, 2007, pp. 1-8.

[6] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model," in *Computer Vision - ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, Eds., ed: Springer Berlin Heidelberg, 2010, pp. 504-517.

[7] J. S. Agustin, H. Skovsgaard, J. P. Hansen, and D. W. Hansen, "Low-cost Gaze Interaction: Ready to Deliver the Promises," in *Conference on Human Factors in Computing (CHI)*, Boston, MA, USA, 2009, pp. 4453-4458.

[8] O. V. Komogortsev, A. Karpov, C. D. Holland, and H. Proença, "Multimodal Ocular Biometrics Approach: A Feasibility Study," in *Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Washington DC, USA, 2012, pp. 1-8.

[9] A. Pacut and A. Czajka, "Aliveness Detection for IRIS Biometrics," in *International Carnahan Conferences Security Technology*, Lexington, KY, 2006, pp. 122-129.

[10] J. Galbally, J. Ortiz-Lopez, J. Fierrez, and J. Ortega-Garcia, "Iris Liveness Detection Based on Quality Related Features," in *International Conference on Biometrics (ICB)*, New Delhi, 2012, pp. 271-276.

[11] N. B. Puhan, N. Sudha, and A. S. Hegde, "A New Iris Liveness Detection Method Against Contact Lens Spoofing," in *International Symposium on Consumer Electronics (ISCE)*, Singapore, 2011, pp. 71-74.

[12] H. Zhang, Z. Sun, and T. Tan, "Contact Lens Detection Based on Weighted LBP," in *International Conference*

on Pattern Recognition (ICPR), Istanbul, 2010, pp. 4279-4282.

[13] X. He, Y. Lu, and P. Shi, "A New Fake Iris Detection Method," in *Advances in Biometrics*. vol. 5558, M. Tistarelli and M. S. Nixon, Eds., ed: Springer Berlin Heidelberg, 2009, pp. 1132-1139.

[14] P. Coli, G. L. Marcialis, and F. Roli, "Vitality Detection from Fingerprint Images: A Critical Survey," in *Advances in Biometrics*. vol. 4642, S.-W. Lee and S. Z. Li, Eds., ed: Springer Berlin, 2007, pp. 722-731.

[15] B. Tan and S. Schuckers, "Liveness Detection for Fingerprint Scanners Based on the Statistics of Wavelet Signal Processing," in *Computer Vision and Pattern Recognition Workshop (CVPRW)*, 2006, pp. 1-8.

[16] R. Derakhshani, S. Schuckers, L. A. Hornak, and L. O'Gorman, "Determination of Vitality from a Non-Invasive Biomedical Measurement for Use in Fingerprint Scanners," *Pattern Recognition,* vol. 36, pp. 383-396, 2003.

[17] L. Ghiani, G. L. Marcialis, and F. Roli, "Experimental Results on the Feature-Level Fusion of Multiple Fingerprint Liveness Detection Algorithms," in *Multimedia and Security Workshop (MMSec)*, Coventry, UK, 2012, pp. 157-163.

[18] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers, "LiveDet 2011 - Fingerprint Liveness Detection Competition 2011," in *International Conference on Biometrics (ICB)*, New Delhi, 2011, pp. 208-215.

[19] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda*, et al.*, "Competition on Counter Measures to 2-D Facial Spoofing Attacks," in *International Joint Conference on Biometrics (IJCB)*, Washington, DC, 2011, pp. 1-6.

[20] C. D. Holland and O. V. Komogortsev, "Biometric Identification via Eye Movement Scanpaths in Reading," in *International Joint Conference on Biometrics (IJCB)*, Washington, D.C., 2011, pp. 1-8.

[21] O. V. Komogortsev, A. Karpov, L. R. Price, and C. R. Aragon, "Biometric Authentication via Oculomotor Plant Characteristics," in *International Conference on Biometrics (ICB)*, New Delhi, India, 2012, pp. 1-8.

[22] C. D. Holland and O. V. Komogortsev, "Complex Eye Movement Pattern Biometrics: Analyzing Fixations and Saccades," in *IAPR International Conference on Biometrics (ICB)*, Madrid, Spain, 2013, pp. 1-8.

[23] O. V. Komogortsev and C. D. Holland, "Biometric Authentication via Complex Oculomotor Behavior," in *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Washington, DC, USA, 2013, pp. 1-8.

[24] D. Noton and L. W. Stark, "Scanpaths in Eye Movements during Pattern Perception," *Science,* vol. 171, pp. 308-311, 1971.

[25] P. Kasprowski and J. Ober, "Eye Movements in Biometrics," in *European Conference on Computer Vision (ECCV)*, Prague, Czech Republic, 2004, pp. 248-258.

[26] O. V. Komogortsev and A. Karpov, "Liveness Detection via Oculomotor Plant Characteristics: Attack of Mechanical Replicas," in *IEEE/IAPR International Conference on Biometrics (ICB)*, Madrid, Spain, 2013, pp. 1-8.

[27] R. J. Leigh and D. S. Zee, *The Neurology of Eye Movements*, 4 ed. Oxford, NY, USA: Oxford University Press, 2006.

[28] C. D. Holland and O. V. Komogortsev, "Biometric Verification via Complex Eye Movements: The Effects of Environment and Stimulus," in *Fifth Internation Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Washington DC, USA, 2012, pp. 1-8.

[29] G. Westheimer, "Mechanism of Saccadic Eye Movements," *Archives of Ophthalmology,* vol. 52, pp. 710-723, 1954.

[30] D. A. Robinson, "Models of the Saccadic Eye Movement Control System," *Biological Cybernetics,* vol. 14, pp. 71-83, 1973.

[31] O. V. Komogortsev and J. I. Khan, "Eye Movement Prediction by Kalman Filter with Integrated Linear Horizontal Oculomotor Plant Mechanical Model," in *Eye Tracking Research & Applications (ETRA) Symposium*, Savannah, GA, USA, 2008, pp. 229-236.

[32] O. V. Komogortsev and J. I. Khan, "Eye Movement Prediction by Oculomotor Plant Kalman Filter with Brainstem Control," *Journal of Control Theory and Applications,* vol. 7, pp. 14-22, 2009.

[33] A. T. Bahill, "Development, Validation, and Sensitivity Analyses of Human Eye Movement Models," *Critical Reviews in Bioengineering,* vol. 4, pp. 311-355, 1980.

[34] O. V. Komogortsev. (2012). *Eye Movement Biometric Database v2*. Available: http://cs.txstate.edu/~ok11/embd_v2.html

[35] O. V. Komogortsev. (2012). *Eye Movement Biometric Database v1*. Available: http://cs.txstate.edu/~ok11/embd_v1.html

[36] EyeLink. *EyeLink 1000 Eye Tracker*. Available: http://www.sr-research.com/

[37] Sony. *PlayStation Eye Camera*. Available: http://us.playstation.com/ps3/accessories/playstation-eye-camera-ps3.html

[38] T. Joachims. *SVM-Light Support Vector Machine*. Available: http://svmlight.joachims.org/

[39] C. Quaia, H. S. Ying, and L. M. Optican, "The Viscoelastic Properties of Passive Eye Muscle in Primates. III: Force Elicited by Natural Elongations," *PLoS ONE,* vol. 5, pp. 1-19, 2010.

[40] C. Quaia and L. M. Optican, "Dynamic Eye Plant Models and the Control of Eye Movements," *Strabismus,* vol. 11, pp. 17-31, 2003.