IEEE Access
The journal for rapid open access publishing

Received July 6, 2015, accepted August 14, 2015, date of publication September 14, 2015, date of current version October 5, 2015.

Digital Object Identifier 10.1109/ACCESS.2015.2477470

Robust Detection of Textured Contact Lenses in Iris Recognition Using BSIF

JAMES S. DOYLE, JR., (Student Member, IEEE), AND KEVIN W. BOWYER, (Fellow, IEEE)
Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
Corresponding author: K. W. Bowyer (kwb@nd.edu)

ABSTRACT This paper considers three issues that arise in creating an algorithm for the robust detection of textured contact lenses in iris recognition images. The first issue is whether the accurate segmentation of the iris region is required in order to achieve the accurate detection of textured contact lenses. Our experimental results suggest that accurate iris segmentation is not required. The second issue is whether an algorithm trained on the images acquired from one sensor will well generalize to the images acquired from a different sensor. Our results suggest that using a novel iris sensor can significantly degrade the correct classification rate of a detection algorithm trained with the images from a different sensor. The third issue is how well a detector generalizes to a brand of textured contact lenses, not seen in the training data. This paper shows that a novel textured lens type may have a significant impact on the performance of textured lens detection.

INDEX TERMS Biometrics, machine learning, image processing, image classification, image texture analysis.

I. INTRODUCTION

Textured (or ''cosmetic'') contact lenses prevent an iris recognition system from imaging the natural iris texture. Therefore, automatic detection of textured contact lenses is an important anti-spoofing technique for iris recognition systems. At least one commercial iris recognition system claims to have a method for detecting the presence of textured contact lenses [16]. However, to our knowledge there is no published evaluation of the algorithm used or its accuracy. A number of approaches have appeared in the literature in recent years, many reporting correct classification rates of over 95% on experimental datasets [8], [12], [13], [21], [37]–[39]. These approaches are based on computing texture features from the iris image and training a classifier to distinguish the case of no textured lens versus the case of textured lens.

This paper makes contributions on three aspects related to automatic detection of textured contact lenses in iris recognition images. One aspect is whether accurate iris segmentation is needed in order for textured lens detection to be effective. This question is important because the presence of textured contact lenses can make accurate iris region segmentation more difficult. All previous research on textured lens detection has assumed that an accurate iris region segmentation is available. Our results show that an accurate iris segmentation is not required in order to achieve high accuracy in detection of textured contact lenses. A second aspect is whether textured lens detection generalizes to images obtained with a different iris sensor. This question is important because large-scale and long-term iris recognition applications will have to deal with images acquired from different sensors. Our results indicate that current textured lens detection algorithms do not necessarily generalize well to use with images from a novel iris sensor. A third aspect is how well textured lens detection generalizes to a brand of lenses not seen in the training data. This question is important because any deployed iris recognition system will eventually be confronted with novel brands of textured contact lenses. Our results suggest that a textured lens detection algorithm trained on images of only one brand of textured lenses may be very brittle, but that training on a larger number of brands of textured lenses improves generalization. The dataset that we use to explore these issues contains images from a larger number of different manufacturers of textured lenses than any other published work.

This paper extends the state-of-the-art in textured lens detection in several ways. This is the first paper to consider whether or not the iris must be accurately segmented in order to detect the presence of textured lenses. The dataset used in evaluating the effects of novel contact lenses contains lenses from more different manufacturers than any other publicly available dataset. This paper also adds additional support to

IEEE Access
The journal for rapid open access publishing

the cross-sensor effects that have been considered in some previous work [9], [38].

The remainder of this paper is organized as follows. Related work is outlined in Section II. Section III describes the dataset and method used in this work. Results of the experiment are presented in Section IV. A comparison to previous results using LBP is offered in Section V. Finally, concluding remarks are given in Section VI.

II. LITERATURE REVIEW

Approaches to detection of fake irises, whether they are printed images of genuine irises, textured contact lenses, or model eyes, can be broken down into three major categories: 1) pattern recognition on single iris images; 2) exploiting some biological trait to detect liveness; and 3) analyzing some physical property of the iris.

A. PATTERN RECOGNITION APPROACHES

As early as 2003, Daugman [7] (building on his previous work [17]) proposed analysis of the 2D Fourier power spectrum to detect the highly periodic fake iris pattern that was prevalent in 'dot-matrix'-style textured lenses manufactured at that time. Some lenses have multiple layers of dot-matrix printing, or are printed via another technique that does not produce the regular dot pattern. This reduces or eliminates the high-power response resulting from the constant spacing of the dots on the lens. Textured lens detection by this method may no longer be reliable.

He et al. [12] propose training a support-vector machine on texture features in a gray-level co-occurrence matrix (GLCM). They constructed a dataset of 2,000 genuine iris images from the Shanghai Jiao Tong University (SJTU) v3.0 database and 250 textured lens images, of which 1,000 genuine and 150 textured are used for training. They report a correct classification rate of 100% on the testing data.

Wei et al. [37] analyze three methods for textured contact lens detection: measure of iris edge sharpness, characterizing iris texture through Iris-Textons, and co-occurrence matrix (CM). Two class-balanced datasets are constructed using CASIA [2] and BATH [34] images for genuine iris images and a special acquisition for textured contact lens images. Each dataset contains samples of a single manufacturer of textured contact lenses. Correct classification rates for the three methods and two datasets vary between 76.8% and 100%.

He et al. [13] use multi-scale Local Binary Patterns (LBP) as a feature extraction method and AdaBoost as a learning algorithm to build a textured lens classifier. They acquire a custom dataset of 600 images with 20 different varieties of fake iris texture, a majority of which are textured contact lenses. A training set of 300 false iris images is combined with 6000 images from the CASIA Iris-V3 [2] and ICE v1.0 [35] data sets.

Zhang et al. [39] investigated the use of Gaussian-smoothed and SIFT-weighted Local Binary Patterns to detect textured lenses in images acquired with multiple iris cameras. They constructed a dataset of 5000 fake iris images with 70 different textured lens varieties. They report a correct classification rate of over 99% when training on heterogenous data, but this drops to 88% when different sensors are used for training and testing sets.

Galbally et al. [10] propose a fake iris classifier based on quality metrics. Twenty-two quality features are extracted from the iris image and combined into a feature vector by a Sequential Floating Forward Selection (SFFS) algorithm. The final feature vector is used to classify an image as either a real or a fake iris.

Kohli et al. [21] perform an analysis of the effects of various types of contact lenses on the performance in a commercial iris biometrics system. They investigate four techniques for contact lens detection and present ROC curves demonstrating an improvement when lens detection is used to filter probe images.

Doyle et al. [8] present an analysis of local binary pattern texture extraction to classify an iris image as no lens, transparent lens, or textured lens. Several machine learning algorithms are investigated and an ensemble of classifiers is constructed. A dataset of 1,000 images from each of the three classes is used for training, and a dataset of 400 images per class is used to test. The correct-classification rate for the three-class problem (textured lenses, clear lenses, no lenses) is 71% but increases to 98% when detecting textured lenses alone. Further analysis on larger datasets is offered in [9].

Yadav et al. [38] compare previous textured lens detection algorithms and multiple proposed algorithms over a common dataset, extending the work in [8], [9], and [21]. A combined dataset with 11,670 images from four different sensors, representing four manufacturers of textured lenses, is used to evaluate the existing techniques and the proposed techniques. The proposed algorithms are shown to outperform the previous methods. Additionally, analysis of the impact of textured lenses, and the benefit of their detection, is presented in the form of ROC curves.

Komulainen et al. [23] apply the Binarized Statistical Image Feature (BSIF) developed by Kannala and Rahtu [20] to the problem of cosmetic lens detection. In their work, they use the 2013 release of the Notre Dame Cosmetic Lens Database. Instead of unrolling the iris region, as in the original work with this dataset [9], [38], the feature extraction is performed on the Cartesian image. BSIF is shown to outperform LBP; LBP had an average CCR of 94.01% and BSIF had an average CCR of 98.42%. Additionally, BSIF was shown to generalize slightly better than LBP in the "leave-1-out" experiment defined by the dataset.

Menotti et al. [28] propose using deep representations for iris spoof detection. A modification to the standard convolution neural network (convnet) was created (spoofnet) and achieved close to state-of-the-art correct classification rates on two publicly-available fake iris datasets (98.93% accuracy on Biosec [32], 98.63% on MobBIOfake [33]) and did

achieve state-of-the-art accuracy for one publicly available fake iris dataset (99.84% on Warsaw [5]).

This paper extends our previous work in textured lens detection [8], [9], [38] in several ways. Our previous work focused on LBP as a feature extraction technique, however we instead use the BSIF feature extraction [20] in this work. The dataset that is used in this work contains additional lens manufacturers that were not present in the previous versions of the dataset. Additionally, we evaluate the correct detection rate of novel lens manufacturers when an increasing number of lens manufacturers are used for training and whether segmentation is necessary for accurate detection of textured lenses.

B. BIOLOGICAL APPROACHES

Park [30] describes a countermeasure to textured lenses in iris biometrics. Park proposes exploiting the natural hippus[1] movements of the human iris to determine if the acquired samples are of fake or real irises. The proposal involves capturing multiple images of the same subject eyes at the time of acquisition and comparing the pupil-to-iris ratio of the multiple samples. The natural hippus contractions should result in changes in the pupil to iris ratio between the different samples. To improve the natural hippus, visible light LEDs added to an iris camera are proposed.

Puhan et al. [31] extend the work of Park [30] by proposing a method by which the Park detection would fail to recognize a textured lens. Their spoofing method involves the use of a textured lens that does not fully occlude the genuine iris texture near the pupillary boundary. By doing so, the majority of the iris texture would be blocked, but would still allow for the hippus movement to be detected. Puhan et al. also propose a countermeasure by which such attacks could be detected.

Pacut and Czajka [29] describe three methods for detecting printed irises: frequency spectrum (FS), controlled light reflection (CLR), and pupil dynamics (PD). The FS method uses frequency analysis of the image to classify the image as either genuine or printed. The CLR method uses an iris camera supplemented with additional near-IR diodes that produce additional reflections detectable in real irises. The PD method uses a visible light illuminant to constrict the iris while a near-IR video sensor records the eye.

Lee et al. [25] propose a method for fake iris detection involving capturing the iris under two different wavelengths of near-IR illumination and checking the reflectance ratio between the sclera and iris portions of the image. The iris and the sclera should have different reflectance ratios when under different illuminations. When there is no observed difference between the two reflection ratios, the iris and sclera are assumed to be made of the same material, and therefore it is a fake. This method was shown to perform well against irises printed on paper, plastic eyes, and textured contact lenses.

[1]Hippus, also known as pupillary athetosis, is spasmodic, rhythmic, but regular dilating and contracting pupillary movements between the sphincter and dilator muscles. From https://en.wikipedia.org/wiki/Hippus, accessed June 2nd, 2015.

Bodade and Talbar [1] describe a system for fake iris detection by using an external illuminant to produce a pupillary constriction. Fake irises will be unchanged in the presence of the extra light, but a true iris will have a smaller pupil to iris ratio under the new illumination. They report 99.45% and 100% accuracy on two datasets of fake iris images.

Huang et al. [14] create a 2-camera NIR face acquisition system, from which iris images are extracted, with the capability to illuminate the iris with visible light in order to force a pupillary constriction. Unlike other methods, the pupil to iris ratio is not the only measure of dilation in this work. An SVM is trained using small patches of iris texture as well.

Kanematsu et al. [19] present a method for fake iris detection by determining the brightness of the iris before and after a pupillary light reflex. They show that there is a significant difference between live irises and fake irises, enough to perfectly segment their dataset.

Czajka [6] presents a liveness measure based on pupil dynamics. Short videos (< 3 seconds) of the eye are acquired under changing illumination. The Kohn and Clynes [22] model of pupil dynamics is used to model expected pupil dilation under various illumination changes. Specifically, the iris responds more quickly to a dark-to-light illumination change than a light-to-dark illumination change. Overall results as a liveness detection mechanism are positive and the author offers a fair assessment of the failure modes of this particular approach.

C. PHYSICAL APPROACHES

Lee et al. [24] suggest that the Purkinje images will be different between a live iris and a fake iris. They propose a novel iris sensor with structured illumination to detect this difference in Purkinje images between a known model of the human eye and an observed fake iris texture. They report results on a dataset of 300 genuine iris images and 15 counterfeit images. They report a False Accept Rate and a False Reject Rate of 0.33% on the data, but suggest that the dataset may be too small to draw generalized conclusions.

Hughes and Bowyer [15] document a prototype stereo iris sensor for textured lens detection. The iris is idealized as a planar torus located posterior to the cornea. When captured with a stereo sensor, the iris is seen as a flat surface. Contact lenses rest on the surface of the convex cornea. Therefore, if a subject is wearing a textured lens the stereo sensor will not see a flat surface but rather a curved surface. This technique approaches detecting textured lenses as distinguishing whether the imaged iris texture lies on a flat surface or a spherical surface in 3D.

III. EXPERIMENTAL METHOD

A. DATASET

The Notre Dame Contact Lens Detection 2015 (NDCLD'15) Dataset[2] is used in this paper. It also defines the leave-*n*-out

[2]Available by request at http://www3.nd.edu/~cvrl/CVRL/Data_Sets.html.

experiments, where $n = \{1, 2, 3, 4\}$. Segmentation information from a commercially-available iris biometrics SDK is also supplied.

1) ACQUISITION

The IrisAccess LG 4000 sensor [27] captures images of both irises simultaneously. All images have a resolution of 640×480 pixels. Two banks of infrared LEDs, one on each side of the sensor, illuminate the eyes. Images can be captured under either "direct" or "cross" illumination, referring to which bank of LEDs illuminates the eye. The illumination options may be used to obtain images with reduced specular highlighting; choice of illuminator is a part of automated image selection for the sensor. The raw iris image data appears to undergo some displacement, to place the pupil near the center of the 640×480 output image, padding with a constant gray level if necessary.

The IrisGuard AD100 [16] sensor captures images of both irises simultaneously. Two types of LEDs allow for near-IR and visible-light illumination of the eyes. All images have a resolution of 640×480 pixels.

All iris images were captured in a windowless indoor lab under consistent lighting conditions. Subjects were supervised during acquisition to ensure proper acquisition procedures were followed. Human subjects participated under the terms of protocols approved by the University Human Subjects Institutional Review Board. Before any biometric information is captured, participants self-report information such as ethnicity, gender, and whether or not the participant is wearing contact lenses. This information is captured for each acquisition session.

2) COMPOSITION

A well-constrained database of 7300 images was constructed to evaluate contact lens detection under various experimental scenarios. The main dataset is composed of 6000 images for model training and 1200 images for model evaluation. Images were acquired using either an IrisAccess LG4000 or an IrisGuard AD100 sensor; both sensors are equally represented. The dataset is composed of images from one of three equally-represented classes: *No Lens*, *Soft Lens*, and *Textured Lens*. Images in the *No Lens* class were acquired while the subject was not wearing any type of contact lens. Images in the *Soft Lens* class were acquired while the subject was wearing a clear soft contact lens which may or may not contain a support boundary, lettering, or other small markings, and may be either toric[3] or non-toric. Images in the *Textured Lens* class were acquired while the subject was wearing a textured/cosmetic soft contact lens with an opaque printing designed to alter the visual appearance of the iris texture. Hard lenses are not represented in this dataset. The distribution of images in the Training Set can be seen in Table 1 and the distribution of images in the Verification Set can be seen in Table 2. The training set and the verification set are subject disjoint; subject eyes appearing in the training set are not part of the verification set.

[3]Toric lenses are often constructed such that they do not freely rotate around the optical axis as is the case with non-toric lenses.

TABLE 1. Image distribution of the base Training Set.

	No	Soft	Textured	*Total*
LG4000	1,000	1,000	1,000	3,000
AD100	1,000	1,000	1,000	3,000
Total	2,000	2,000	2,000	6,000

TABLE 2. Image distribution of the base Verification Set.

	No	Soft	Textured	*Total*
LG4000	200	200	200	600
AD100	200	200	200	600
Total	400	400	400	1,200

TABLE 3. Image distribution of the *Textured Lens* group by manufacturer.

	J&J	Ciba	Cooper	ClearLab	Eyedea
LG4000	250	250	250	250	250
AD100	250	250	250	250	250
Total	500	500	500	500	500

For the training set, ten images from each subject eye were selected from the *No Lens* and *Soft Lens* classes. For the *Textured Lens* class, more images are selected from each subject eye, due to the limited number of subjects available to wear textured contact lenses. Between 36 and 192 images are selected from each subject eye in the *Textured Lens* class, dependent upon how many different brands of textured lenses were worn by that subject. The subject breakdown of the dataset can be found in Tables 5, 6, and 7.

For the *No Lens* and *Soft Lens* classes in the verification set, ten images were selected from each subject eye not represented in the training set. For the *Textured Lens* class, more images are selected from each subject eye, due to the limited number of subjects available to wear textured contact lenses. Between 35 and 65 images are selected from each subject eye, dependent upon how many different brands of textured lenses were worn by that subject.

The image distributions in each combination of Sensor and Class are balanced between Right Eye and Left Eye. The image distributions in each Sensor for *No Lens* and *Soft Lens* are additionally balanced between Male and Female; the *Textured Lens* subject pool was predominately Male. The majority of the dataset is Caucasian subjects, but African and Asian are also represented.

All textured contact lenses in the NDCLD'15 base dataset came from five major suppliers of textured lenses: Johnson&Johnson [18], Ciba Vision [3], Cooper Vision [36], Clearlab [4] and United Contact Lens [26]. Multiple colors were selected for each manufacturer and some lenses were also toric lenses designed to correct for astigmatism. The distribution of images per lens manufacturer can be found in Table 3.

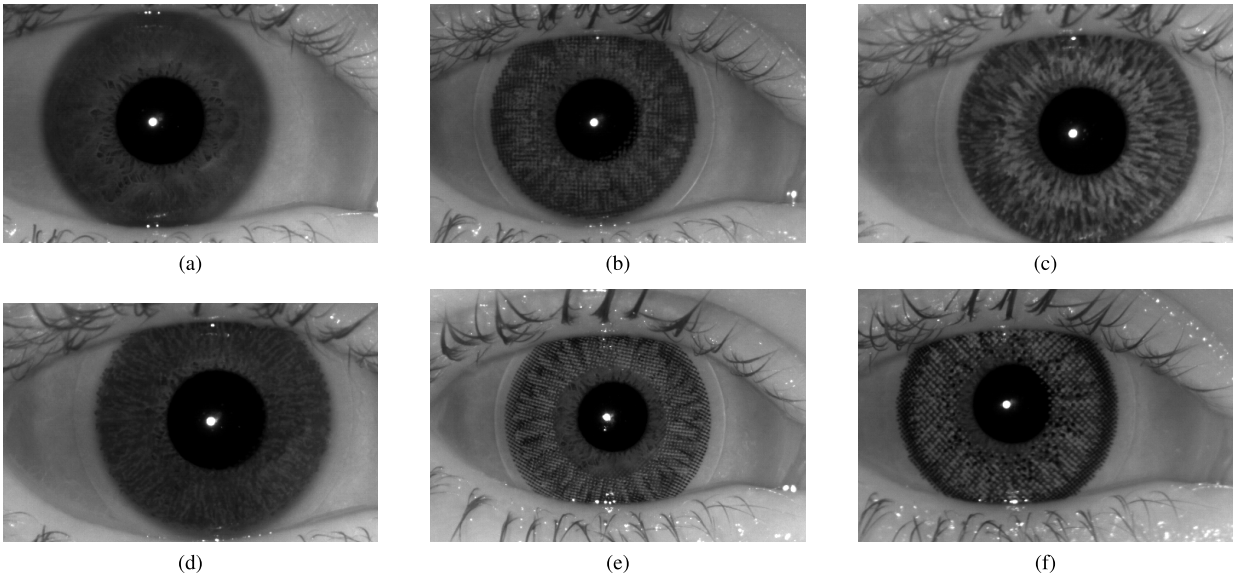

FIGURE 1. Cropped AD100 sample images for the five textured lens manufacturers represented in this work in the same eye. These images are samples only, they may or may not be part of the dataset. (a) No Lens. (b) Ciba. (c) Cooper. (d) J&J. (e) UCL. (f) Clear Lab.

TABLE 4. Image distribution of the *Textured Lens* group by Leave *n* Out.

	Leave-1-Out	Leave-2-Out	Leave-3-Out	Leave-4-Out
Training	2,000	1,500	1,000	500
Testing	5 * 400	5 * 400	5 * 400	5 * 400
Permutations	5	10	10	5

The database also defines multiple datasets for leave-n-out experimentation, where $n = \{1, 2, 3, 4\}$. The number of images in each of the arrangements can be found in Table 4. In order to create a dataset with the proper number of cosmetic lens images (250) from each manufacturer, another 100 cosmetic images were added to the database. This accounts for the apparent discrepancy in the first paragraph of the dataset description. (6,000 training + 1,200 testing + 100 extra = 7,300 images.)

Sample images of each textured lens manufacturer can be seen for the AD100 sensor in Figure 1 and for the LG4000 sensor in Figure 2.

B. SEGMENTATION

All images were segmented using a commercially-available iris biometrics SDK to extract center and radius for circles defining the pupillary boundary and the limbic boundary. The segmentation divides each iris image into three regions: (1) pupil, (2) iris, and (3) sclera/periocular. Details about the specific implementation of the algorithm are not available as the software is closed-source. The software outputs center point (x, y) and radius of two circles only. More accurate segmentation representations (ellipses, snakes) and mask (eyelid/eyelash occlusion, spectral highlights) information are not available with this software.

Segmentations for the Training Set were inspected visually by overlaying circles defined by the segmentation algorithm.

TABLE 5. Subject distribution of the *AD100* images.

	Training	Verification
No	200	40
Soft	200	40
Textured	10	4

TABLE 6. Subject distribution of the *LG4000* images.

	Training	Verification
No	200	40
Soft	200	40
Textured	10	4

TABLE 7. Subject distribution of the *Combined* images.

	Training	Verification
No	284	80
Soft	272	80
Textured	10	4

Ill-fitting circles were corrected by taking a "best-fit" circle from four mouse clicks each around the limbus and the pupillary circles. The Verification Set segmentation was not visually inspected or adjusted to better simulate an unsupervised "real-world" iris biometrics system.

C. FEATURE EXTRACTION

Binarized Statical Image Feature (BSIF) analysis [20][4] is applied at multiple scales to produce feature vectors. The kernel size for the BSIF pattern analysis is $s = \{3, 5, 7, 9, 11, 13, 15, 17\}$[5] for a total of

[4]Source code for BSIF is generously provided by the University of Oulu Center for Machine Vision Research. http://www.ee.oulu.fi/~jkannala/bsif/bsif_code_and_data.zip

[5]The runtime of the BSIF feature increases with larger kernel sizes but no runtime performance analysis has been performed in this work.

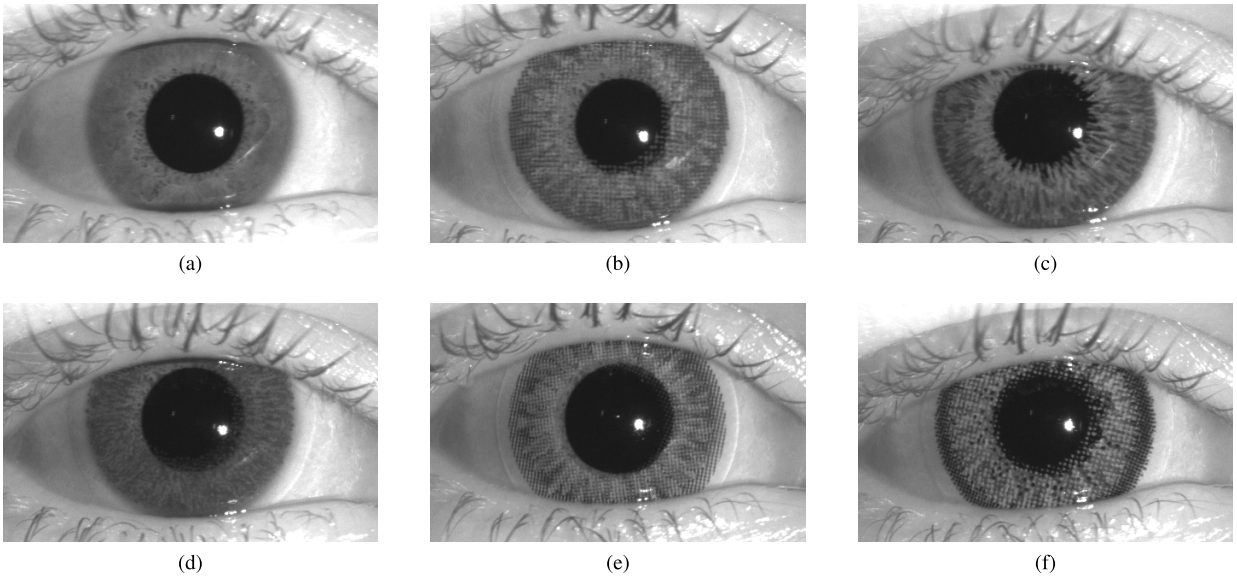

FIGURE 2. Cropped LG 4000 sample images for the five textured lens manufacturers represented in this work in the same eye. These images are samples only, they may or may not be part of the dataset. (a) No Lens. (b) Ciba. (c) Cooper. (d) J&J. (e) UCL. (f) Clear Lab.

8 different feature vector sets. The kernel depth was held constant at 8-bits resulting in a feature vector of length 256.

Three different applications of BSIF are evaluated in this work: *Whole Image*, *Best Guess* and *Known Segmentation*. In *Whole Image*, the BSIF feature vector is calculated for the entirety of the image. In *Best Guess*, the kernel is evaluated inside a set torus, eliminating the need for a segmentation algorithm, while also excluding eyebrow and other noise from the majority of images. The pupillary boundary is defined by the average center point of all pupillary circles in the verified training set and the pupillary circle radius is defined as the average pupillary radius from the verified training set. The limbic boundary is defined by the average center point of all limbic circles in the verified training set and the limbic circle radius is defined as the average limbic radius from the verified training set, plus a delta of 30 pixels. The distributions of centers and radii for the verified combined training set can be found in Figure 3. The segmentation provided by the dataset[6] is used in *Known Segmentation* to limit the scope of the BSIF kernel to only the localized iris texture in each image. The known iris radius is increased by 30 pixels to include the contact lens boundary which is usually located just outside the limbic boundary in the sclera.

Both the LG4000 and AD100 cameras appear to perform some iris localization during the acquisition process which positions the iris roughly in the center of the image, which can be seen in the box plots in Figure 3. However, the apparent size of the iris can vary, as seen visually in Figures 1 and 2 and plotted in Figure 3. Furthermore, the presence of a textured contact lens yields less accurate segmentations versus the presence of a soft lens or the absence of any contact lens.

[6]Segmentation information is provided by a commercial matcher. Training set segmentation was manually verified and corrected, verification set was not inspected.

D. MODEL TRAINING

Six different classifiers were explored as possible approaches to train models on the feature sets. The specific classifiers were Naïve Bayes, Logistic, Multilayer Perceptron, Simple Logistic, SMO, and LMT. Implementations of these algorithms were provided by Weka [11].

The images in the dataset were down-sampled by 50% in each direction to facilitate evaluation of BSIF scales above $s = 17$. Applying the original kernel sizes of BSIF on the reduced data simulates BSIF kernel sizes of $s = \{6, 10, 14, 18, 22, 26, 30, 34\}$. Combined with the BSIF kernels applied to the original-scale data yields a total of 16 different feature vector sets.

Models were trained over the single-sensor portions of the dataset and on the two-sensor combined dataset. When data from both sensors was used, the source sensor was not used as a feature. A separate ensemble of models is constructed for each combination of sensor, data arrangement, and feature extraction method. The ensemble is 6 classifiers by 16 scales for 96 trained models in the ensemble.

E. MODEL EVALUATION

Each ensemble of models is evaluated using the verification set defined in Section III-A.2. The single-sensor training also allows for the evaluation of a novel sensor (i.e., AD100 models evaluated on LG4000 verification images) and the leave-n-out verification set allows for the examination of the effect of a novel textured lens.

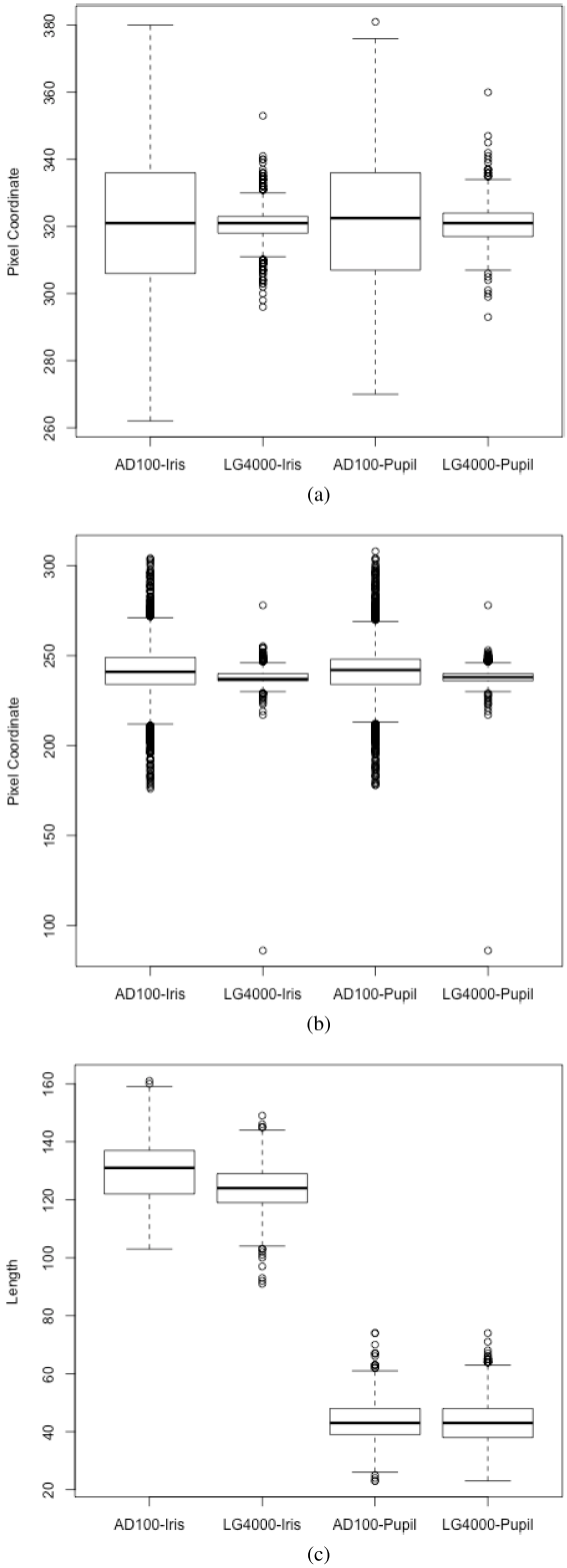

FIGURE 3. Center points and radii for both the pupillary and limbic segmentation circles for the combined dataset. (a) X. (b) Y. (c) R.

IV. EXPERIMENTAL RESULTS

Observing the confusion matrices and correct classification rates (CCR) in aggregate allows for generalized conclusions

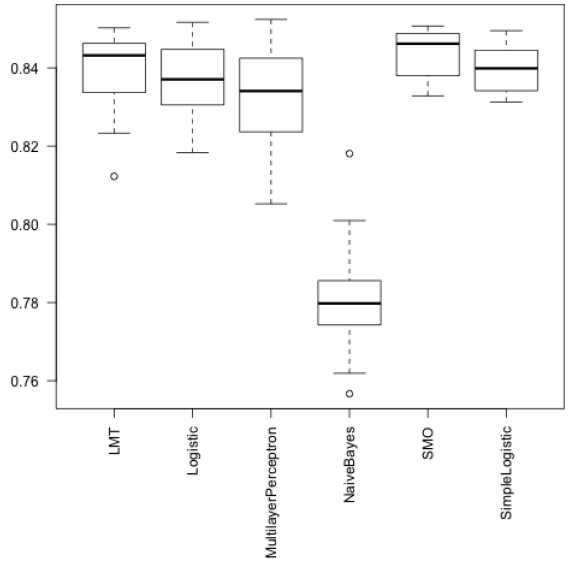

FIGURE 4. Boxplots of the Correct Classification Rate (CCR) of each Weka [11] classifier for Best Guess and Known Segmentation.

to three questions regarding contact lens detection: (1) is segmentation a necessary part of the process? (2) what effect does a novel sensor have on constructed models? and (3) what effect does a novel textured lens have on constructed models?

As previously mentioned there are six classifiers that are used for textured lens detection. With the exception of Naïve Bayes, all classifiers perform at about the same correct classification rate, roughly 84%. The specific rates can be found in Figure 4. While the classifier does not appear to have much impact, the scale at which the BSIF features are applied shows a definite trend. The BSIF code comes with preset scales at $s = \{3, 5, 7, 9, 11, 13, 15, 17\}$ and the average CCR for each s is presented in Figure 5 for both the original-scale and reduced-scale verification data. For the original-scale verification data, the smallest scale $s = 3$ starts with a CCR of 83% and goes up to a CCR of 85% with $s = 17$. For the reduced-scale verification data, the smallest scale $s = 6$ starts with a CCR of 83% and peaks at $s = 18$ with a CCR of 85%.

Figure 6 shows the marginal increase in CCR achieved when an ensemble of n models is used, where $n = \{1, 2, ...96\}$. When $n = 1$ the performance is similar to the average single classifier results shown in Figure 4 and Figure 5. However, as the n increases, the CCR asymptotically approaches the maximum observed CCR for each combination of sensor and segmentation method.

A. VALUE OF KNOWN SEGMENTATION

The three different segmentation scenarios (mentioned in Section III-C) are evaluated and ranked by the average CCR on the verification set. A bar chart of the results can be found in Figure 7. The results presented in this sub-

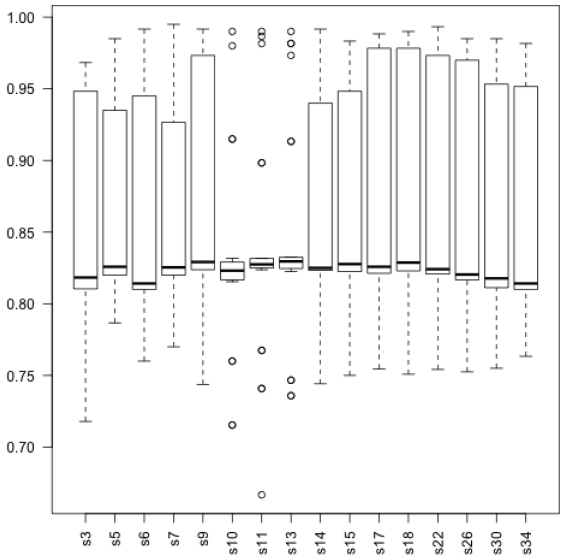

FIGURE 5. Correct classification rate (CCR) for each BSIF feature size. Results are shown for kernel sizes of $s = 3, 5, 7, 9, 11, 13, 15, 17$ and down sampled images resulting in $s = 6, 10, 17, 18, 22, 26, 30, 34$.

section are the average of the homogeneous[7] and combined [8] experiments.

Using the entire image as the region of interest for the BSIF feature extraction technique described in Section III-C yields a surprisingly accurate CCR, on par with using known segmentation with LBP. For the AD100 set, the CCR is 99.5%, for LG4000 the CCR is 99.67%, and combined the CCR is 99.75%. However, using the average segmentation to guess at the location of the iris within the image outperforms this method. Guessing at the true segmentation of the verification sets by using the average center point and radius of the training sets results in a perfect CCR of 100% for AD100 and LG4000. Using the known segmentation maintains the same perfect CCR of 100% for AD100 and LG4000. The combined numbers for best guess and known segmentation are 99.92% and 99.92%, respectively. However, the sensor is not given as part of the feature vector. The independently-trained models could trivially be used instead of the combined classifier if the sensor is known.

Using the known segmentation does not improve over best guess in CCR. The lack of improvement when moving from best guess to known segmentation may motivate an early-reject mechanism when a textured lens is detected as a separate thread while segmentation is being performed.

B. NOVEL SENSOR

The homogeneous sensor case is defined by evaluating the trained models on a dataset from the same sensor, i.e. using LG4000 data to evaluate the performance of a classifier trained using LG4000 images. Accordingly, the heterogeneous sensor case is defined by evaluating the trained models on a dataset from a different sensor, i.e., using LG4000 data to evaluate the performance of a classifier trained using AD100 images. A bar chart of the average results can be found in Figure 8.

[7]The homogeneous sensor case is defined by evaluating the trained models on a dataset from the same sensor.

[8]The combined dataset is defined as the union of the AD100 and LG4000 datasets.

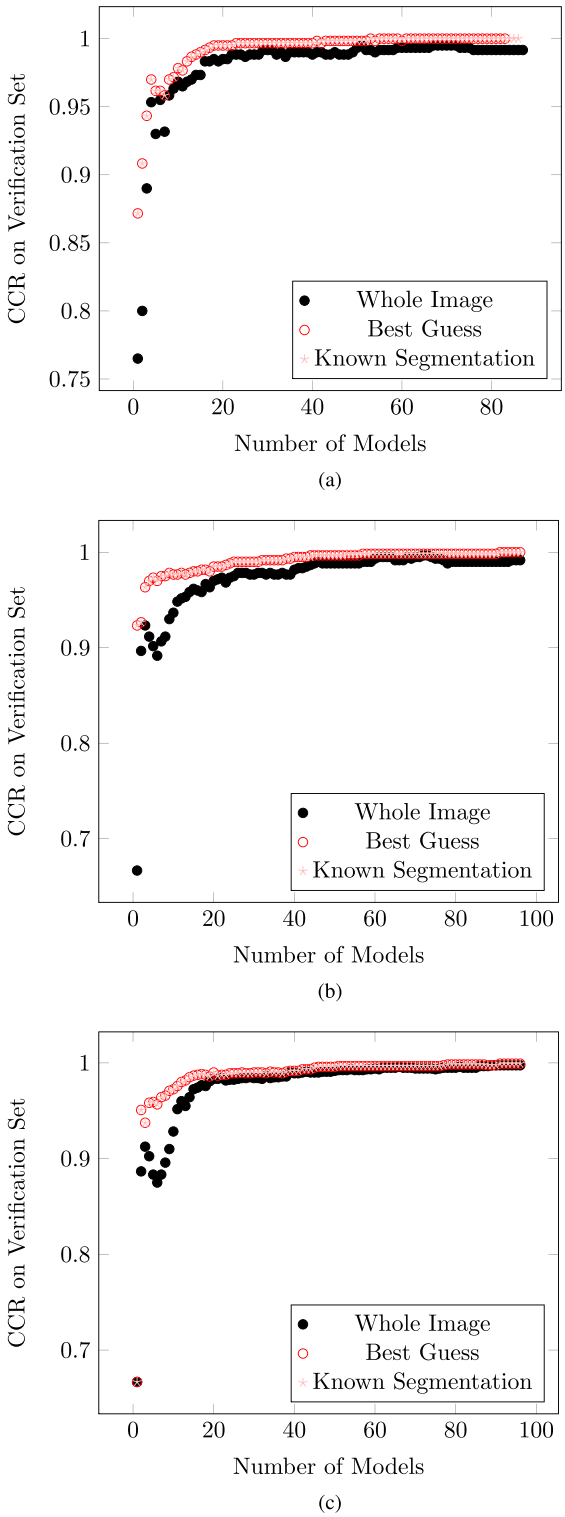

FIGURE 6. CCR as a function of number of models in the ensemble. (a) AD100. (b) LG4000. (c) Combined.

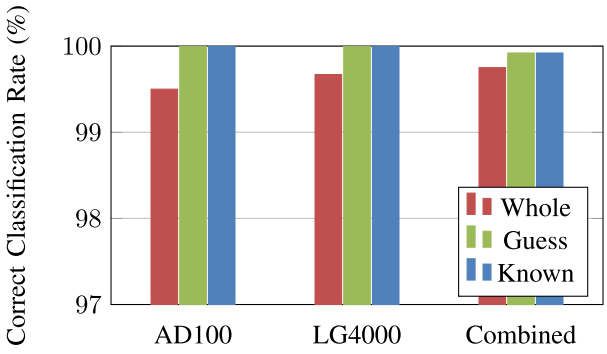

FIGURE 7. Correct Classification Rate trends across the different segmentations. Results for the homogenous sensor experiments are shown. Combined contains images from both sensors, not labeled as to which sensor the image comes from.

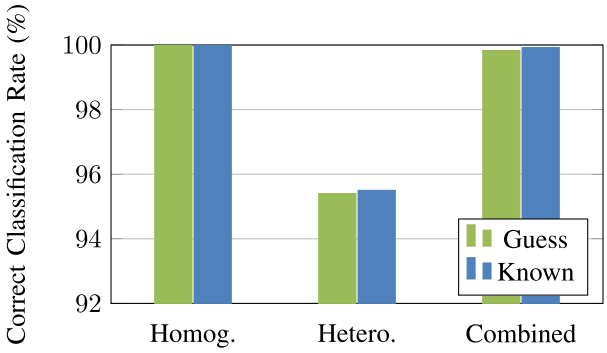

FIGURE 8. Correct Classification Rate trends for Homogeneous and Heterogeneous AD100 and LG4000 verification datasets for Known Segmentation.

Unsurprisingly, the CCR of the homogenous case is higher than the heterogeneous CCR. A drop is observed from 100% in the homogenous case to just over 95% in the heterogenous case. Evaluation of the combined dataset shows that it is possible to correct for the effect of a novel sensor when images from multiple sensors are included in the training step.

The drop in CCR from homogenous sensor to heterogenous sensor implies that there are sensor-specific factors in the detection of textured contact lenses. These may be due to, for example, differences in the near-IR wavelength used and how it interacts with the pigment used in the textured lenses. This result suggests that, for maximum detection accuracy, a textured lens detection algorithm should be trained with sample images from each sensor with which it will be used.

C. NOVEL TEXTURED LENS

The effect of a novel lens on a trained ensemble of models is evaluated by the CCR of a verification set of images containing textured lenses from a different manufacturer or manufacturers than the set of images used to train the ensemble. The CCRs reported here follow the experimental outline of combined sensor evaluations and *Best Guess* segmentation.

For the leave-1-out experiment, the models were trained on data from four of the textured lens manufacturers and tested against the data from the fifth. For instance, an ensemble of models was trained on images from Johnson&Johnson, CibaVision, Cooper Vision, and United Contact Lens and then evaluated using images from ClearLab. This was repeated for each of the five manufacturers represented in this dataset, and for each of the five verification sets. The average CCR across all leave-1-out experiments is 97.65%.

For the leave-2-out experiment, the models were trained on data from three manufacturers and tested on data from the remaining two manufacturers. For instance, an ensemble of models was trained on images from Johnson&Johnson, CibaVision, and Cooper Vision, and then evaluated using images from United Contact Lens and ClearLab. This was repeated for each combination of $\binom{5}{3}$ manufacturers represented in this dataset, and for each of the five verification sets. The average CCR across all leave-2-out experiments is 95.97%.

For the leave-3-out experiment, the models were trained on data from only two manufacturers and tested on data from the remaining three. For instance, an ensemble of models was trained on images from Johnson&Johnson and CibaVision, and then evaluated using images from Cooper Vision, United Contact Lens and ClearLab. This was repeated for each combination of $\binom{5}{2}$ manufacturers represented in this dataset, and for each of the five verification sets. The average CCR across all leave-3-out experiments is 92.59%.

For the leave-4-out experiment, the models were trained on data from a single manufacturer and tested against data from the remaining four. For instance, an ensemble of models was trained on images from Johnson&Johnson and then evaluated using images from CibaVision, Cooper Vision, United Contact Lens, and ClearLab. This was repeated for each of the five

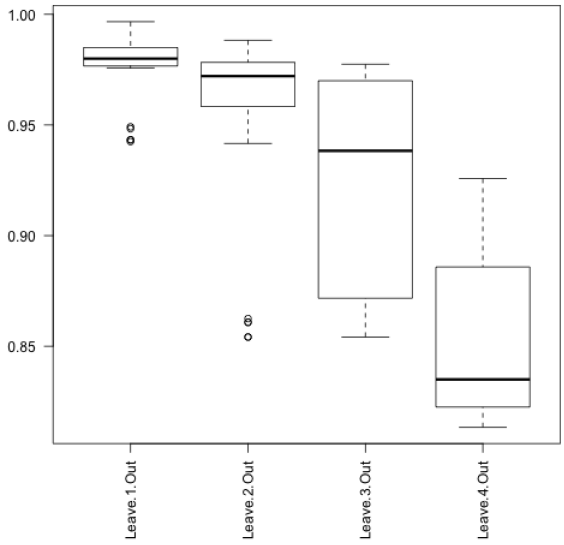

FIGURE 9. Drop in CCR as a function of number of lens manufacturers left out of the training set and used exclusively in the verification set. Results are shown for *Best Guess* segmentation.

manufacturers represented in this dataset, and for each of the five verification sets. The average CCR across all leave-4-out experiments is 85.69%.

The decreasing trend as a function of number of lens manufacturers used in the training set can be seen in Figure 9.

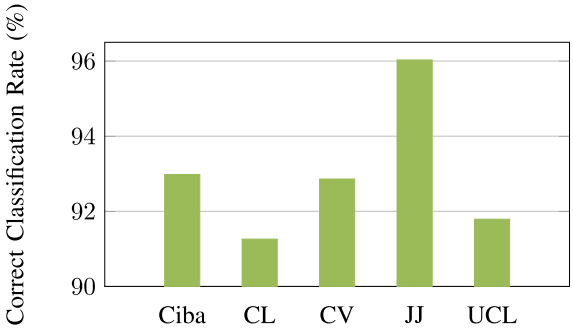

FIGURE 10. Contribution to CCR of each manufacturer. Lens mentioned on X axis was included in training set. Results are shown for *Best Guess* segmentation.

Figure 10 shows the average CCR for each lens manufacturer when that lens type is included in the training set. This chart clearly shows that the models trained with certain lens manufacturers generalize better to a novel lens type. Models trained on images acquired with Johnson&Johnson lenses performed several percentage points higher than when other lens manufacturers are used.

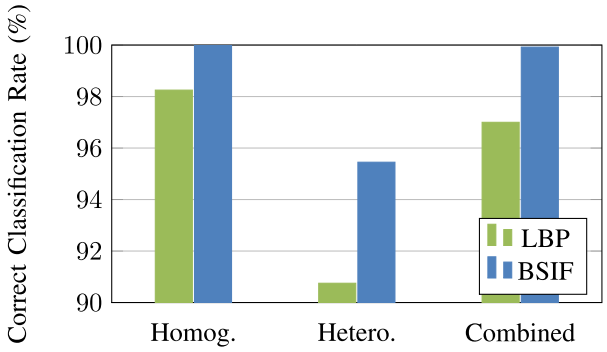

FIGURE 11. Correct Classification Rate comparison for LBP and BSIF under identical experimental frameworks using similarly-structured datasets.

V. COMPARISON WITH LBP

This same experimental framework is examined in [38] using a similarly-structured[9] dataset. Due to the similarities of experimental design and dataset construction, direct comparison between the application of BSIF and LBP in this problem space can be made. Figure 11 highlights the relative performance of LBP and BSIF for the homogeneous sensor, heterogeneous sensor, and combined sensor cases. In all cases, BSIF texture extraction technique is superior to LBP texture extraction. A comparison of the difference

[9]The NDCLD'13 and NDCLD'15 datasets are structured the same, but the NDCLD'15 dataset has the same number of images from the LG4000 and AD100 sensors.

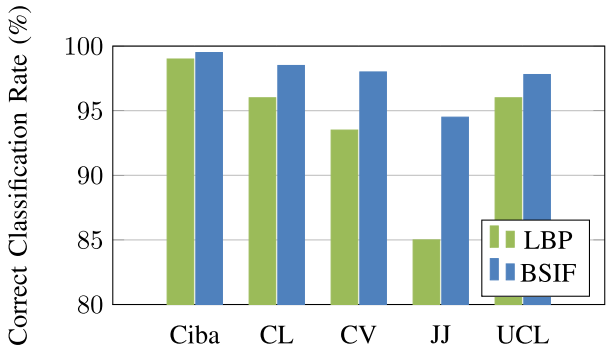

FIGURE 12. Correct Classification Rate comparison for LBP and BSIF under identical experimental frameworks using similarly-structured "leave-1-out" datasets.

in performance between BSIF and LBP on the *leave-1-out* experiment is offered in Figure 12.

Again, the same experimental framework is examined in [9] for the "leave-1-out" experimentation, allowing for a comparison between the relative performance of BSIF and LBP for each lens manufacturer. In every case, the BSIF feature generalized better to the "leave-1-out" experiment than did LBP. For the case of Johnson&Johnson, the BSIF feature greatly outperformed LBP.

VI. CONCLUSIONS AND DISCUSSION

The work in this paper investigates three different issues that arise in the construction of a robust algorithm for detecting iris recognition images that contain textured contact lenses. Three major conclusions can be drawn from the results of these experiments.

A. IS ACCURATE IRIS SEGMENTATION REQUIRED?

Our results suggest that an exact segmentation of the iris region is not required in order to achieve accurate detection of textured contact lenses in iris images. The CCR for both *Best Guess* and *Known Segmentation* are roughly equivalent. Systems that rely on detection of textured lenses may be able to detect them without requiring computationally expensive segmentation algorithms. Additionally, the accuracy of iris segmentation is reduced when textured lenses are present in the image and therefore it is preferable to eliminate the requirement that the image must first be segmented before classification. The evidence for this is summarized in Figure 7.

B. DOES ACCURACY DEGRADE FOR NOVEL SENSOR?

Due to sensor-specific factors, trained models do not generalize with the same accuracy to different sensors when trained on only a single sensor, as is shown in Figure 8. When data from multiple sensors were used for training, the CCR regained most of the loss from the heterogenous evaluation. However, the introduction of a novel sensor into a working biometrics system may still require additional models to be trained in order to maintain a high detection rate of textured lenses.

C. DOES ACCURACY DEGRADE FOR NOVEL LENS TYPE?

The CCR of a trained textured lens detector drops slightly when a type of textured lens that it has not previously seen is introduced into the verification dataset. However, the more manufactures that are observed in the training set, the more robust the models are to novel lens manufacturers. If training on only one lens manufacturer, the CCR on novel lenses is about 86%. This increases dramatically to almost 98% when data from four manufacturers is used in training. Therefore, a trained classifier does reliably generalize to a manufacturer of textured contact lenses that was not represented in the training data.

D. FINAL REMARK

One possibly surprising result emerging from this work concerns the texture filters LBP and BSIF. For this particular problem of textured contact lens detection, BSIF appears to offer substantially better performance. BSIF accuracy is generally higher, and generalizes better. (Figures 11 and 12.)

As a final overall conclusion, provided that the detection algorithm is trained with images from the same sensor used in testing, and that there are no lens types seen in testing that were not seen in training, textured lens detection appears to be a solved problem. In these restricted conditions, which may be approximated in some practical situations, accuracy close to 100% may be achieved. In less controlled conditions, accuracy may drop for a type of lens that was not represented in the training data. However, it appears that training on a large number of different lens types can give some confidence that this method generalizes with reasonably high accuracy.

REFERENCES

[1] R. Bodade and S. Talbar, "Dynamic iris localisation: A novel approach suitable for fake iris detection," in *Proc. Int. Conf. Ultra Modern Telecommun. Workshops (ICUMT)*, 2009, pp. 1–5.
[2] Chinese Academy of Sciences Center for Biometrics and Security Research. (Jun. 2015). *CASIA Iris Databases*. [Online]. Available: http://www.cbsr.ia.ac.cn/english/IrisDatabase.asp
[3] CibaVision. (Apr. 2013). *FreshLook Colorblends*. [Online]. Available: http://www.freshlookcontacts.com
[4] Clearlab. (Jan. 2014). *Eyedia Clear Color Elements*. [Online]. Available: http://www.clearlabusa.com/eyedia-clear-color.php
[5] A. Czajka, "Database of iris printouts and its application: Development of liveness detection method for iris recognition," in *Proc. 18th Int. Conf. Methods Models Autom. Robot. (MMAR)*, 2013, pp. 28–33.
[6] A. Czajka, "Pupil dynamics for iris liveness detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 726–735, Apr. 2015.
[7] J. Daugman, "Demodulation by complex-valued wavelets for stochastic pattern recognition," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 1, no. 1, pp. 1–17, 2003.
[8] J. S. Doyle, K. W. Bowyer, and P. J. Flynn, "Automated classification of contact lens type in iris images," in *Proc. 6th IAPR Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.
[9] J. S. Doyle, K. W. Bowyer, and P. J. Flynn, "Variation in accuracy of textured contact lens detection based on sensor and lens pattern," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep./Oct. 2013, pp. 1–7.
[10] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "A high performance fingerprint liveness detection method based on quality related features," *Future Generat. Comput. Syst.*, vol. 28, no. 1, pp. 311–321, 2012.
[11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.
[12] X. He, S. An, and P. Shi, "Statistical texture analysis-based approach for fake iris detection using support vector machines," in *Proc. Int. Conf. Adv. Biometrics*, 2007, pp. 540–546.
[13] Z. He, Z. Sun, T. Tan, and Z. Wei, "Efficient iris spoof detection via boosted local binary patterns," in *Proc. Int. Conf. Adv. Biometrics*, 2009, pp. 1080–1090.
[14] X. Huang, C. Ti, Q.-Z. Hou, A. Tokuta, and R. Yang, "An experimental study of pupil constriction for liveness detection," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2013, pp. 252–258.
[15] K. Hughes and K. W. Bowyer, "Detection of contact-lens-based iris biometric spoofs using stereo imaging," in *Proc. 46th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2013, pp. 1763–1772.
[16] IrisGuard. (Apr. 2013). *AD100 Camera*. [Online]. Available: http://www.irisguard.com/uploads/AD100ProductSheet.pdf
[17] A. K. Jain, R. M. Bolle, and S. Pankanti, Eds., *Biometrics: Personal Identification in Networked Society*. New York, NY, USA: Springer-Verlag, 1999.
[18] Johnson & Johnson. (Apr. 2013). *ACUVUE2 Colours*. [Online]. Available: http://www.acuvue.com/products-acuvue-2-colours
[19] M. Kanematsu, H. Takano, and K. Nakamura, "Highly reliable liveness detection method for iris recognition," in *Proc. Annu. Conf. Soc. Instrum. Control Eng. (SICE)*, 2007, pp. 361–364.
[20] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, 2012, pp. 1363–1366.
[21] N. Kohli, D. Yadav, M. Vatsa, and R. Singh, "Revisiting iris recognition with color cosmetic contact lenses," in *Proc. 6th IAPR Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–7.
[22] M. Kohn and M. Clynes, "Color dynamics of the pupil," *Ann. New York Acad. Sci.*, vol. 156, no. 2, pp. 931–950, 1969.
[23] J. Komulainen, A. Hadid, and M. Pietikäinen, "Generalized textured contact lens detection by extracting BSIF description from Cartesian iris images," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Sep./Oct. 2014, pp. 1–7.
[24] E. C. Lee, K. R. Park, and J. Kim, "Fake iris detection by using Purkinje image," in *Proc. 6th IAPR Int. Conf. Biometrics*, 2006, pp. 397–403.
[25] S. J. Lee, K. R. Park, and J. Kim, "Robust fake iris detection based on variation of the reflectance ratio between the IRIS and the sclera," in *Proc. Biometrics Symp., Special Session Res. Biometric Consortium Conf.*, 2006, pp. 1–6.
[26] United Contact Lenses. (Jan. 2014). *Cool Eyes Opaque*. [Online]. Available: http://www.unitedcontactlens.com/contacts/opaque-lenses.html
[27] LG. (Oct. 2011). *LG 4000 Camera*. [Online]. Available: http://www.lgiris.com
[28] D. Menotti *et al.*, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 864–879, Apr. 2015.
[29] A. Pacut and A. Czajka, "Aliveness detection for IRIS biometrics," in *Proc. 40th Annu. IEEE Int. Carnahan Conf. Secur. Technol.*, Oct. 2006, pp. 122–129.
[30] K. R. Park, "Robust fake iris detection," in *Articulated Motion and Deformable Objects*. Berlin, Germany: Springer-Verlag, 2006, pp. 10–18.
[31] N. B. Puhan, S. Natarajan, and A. S. Hegde, "Iris liveness detection for semi-transparent contact lens spoofing," in *Advances in Digital Image Processing and Information Technology*. Berlin, Germany: Springer-Verlag, 2011, pp. 249–256.
[32] 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*. Berlin, Germany: Springer-Verlag, 2008, pp. 181–190.
[33] A. F. Sequeira, J. C. Monteiro, A. Rebelo, and H. P. Oliveira, "MobBIO: A multimodal database captured with a portable handheld device," in *Proc. VISAPP*, 2014, pp. 133–139.
[34] University of Bath/SmartSensors. (Jun. 2015). *University of Bath Iris Database*. [Online]. Available: http://www.smartsensors.co.uk/irisweb/
[35] University of Notre Dame. (Jun. 2015). *Iris Challenge Evaluation 2005*. [Online]. Available: http://www3.nd.edu/~cvrl/CVRL/Data_Sets.html
[36] Cooper Vision. (Apr. 2013). *Expressions Colors*. [Online]. Available: http://coopervision.com/contact-lenses/expressions-color-contacts

[37] Z. Wei, X. Qiu, Z. Sun, and T. Tan, "Counterfeit iris detection based on texture analysis," in *Proc. 19th Int. Conf. Pattern Recognition (ICPR)*, 2008, pp. 1–4.
[38] D. Yadav, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, and K. W. Bowyer, "Unraveling the effect of textured contact lenses on iris recognition," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 5, pp. 851–862, May 2014.
[39] H. Zhang, Z. Sun, and T. Tan, "Contact lens detection based on weighted LBP," in *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, 2010, pp. 4279–4282.

JAMES S. DOYLE, JR. received the B.S. degree in computer engineering from Purdue University, West Lafayette, IN, in 2007, and the M.S. and Ph.D. degrees in computer science and engineering from the University of Notre Dame, South Bend, IN, in 2011 and 2015, respectively. He is currently a Lead Software Engineer with MITRE Corporation, Clarksburg, WV. His research interests include iris biometrics, pattern recognition, and computer vision.

KEVIN W. BOWYER (F'98) is currently the Schubmehl-Prein Professor of Computer Science and Engineering with the University of Notre Dame and serves as the Chair of the Department of Computer Science and Engineering. His research interests range broadly over computer vision and pattern recognition, including data mining, classifier ensembles, and biometrics. He was a recipient of the 2014 Technical Achievement Award from the IEEE Computer Society, with the citation for pioneering contributions to the science and engineering of biometrics.

Over the last decade, he has made numerous advances in multiple areas of biometrics, including iris recognition, face recognition, and multibiometric methods. His research group has been active in support of a variety of government-sponsored biometrics research programs, including the Human ID Gait Challenge, the Face Recognition Grand Challenge, the Iris Challenge Evaluation, the Face Recognition Vendor Test 2006, and the Multiple Biometric Grand Challenge. He has edited the recent book entitled *Handbook of Iris Recognition* with Dr. M. Burge.

Prof. Bowyer is a fellow of IAPR, and a Golden Core Member of the IEEE Computer Society. He serves as the General Chair of the 2015 IEEE International Conference on Automatic Face and Gesture Recognition. He served as the General Chair of the 2011 IEEE International Joint Conference on Biometrics, the Program Chair of the 2011 IEEE International Conference on Automatic Face and Gesture Recognition, and the General Chair of the IEEE International Conference on Biometrics Theory Applications and Systems in 2007, 2008, and 2009. He has served as the Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and the IEEE BIOMETRICS COMPENDIUM, and also serves on the Editorial Board of the IEEE ACCESS.