Variation in Accuracy of Textured Contact Lens Detection Based on Sensor and Lens Pattern

James S. Doyle, Kevin W. Bowyer, Patrick J. Flynn
Department of Computer Science and Engineering
University of Notre Dame
{jdoyle6,kwb,flynn}@nd.edu

Abstract

Automatic detection of textured contact lenses in images acquired for iris recognition has been studied by several researchers. However, to date, the experimental results in this area have all been based on the same manufacturer of contact lenses being represented in both the training data and the test data and only one previous work has considered images from more than one iris sensor. Experimental results in this work show that accuracy of textured lens detection can drop dramatically when tested on a manufacturer of lenses not seen in the training data, or when the iris sensor in use varies between the training and test data. These results suggest that the development of a fully general approach to textured lens detection is a problem that still requires attention.

1. Introduction

Textured contact lens detection is an important problem in preventing spoofing in iris recognition systems. A number of approaches have been reported in the literature in recent years, many reporting correct classification rates of over 95% on their experimental dataset. 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. Some commercial iris biometrics systems also claim to have a method for detecting the presence of textured contact lenses to prevent a spoofing attempt [8]. In all results reported to date, the contact lens manufacturer(s) represented in the test data have also been represented in the training data. Also, only one previous paper has considered images acquired from more than one iris sensor. In practice, it is desirable that algorithms developed with image data from one sensor may be migrated to work with another sensor. And, importantly, In many applications, it is not reasonable to assume that the algorithm will have been developed using image data from all manufactures of contact lenses that will be experienced in operation.

This paper outlines a potential textured lens detection algorithm and highlights the potential issues with textured lens detection systems. Related work is outlined in Section 2. A detailed description of the problem is presented in Section 3. Section 4 describes the dataset used in this work and how the problem was addressed. Results of the experiment are presented in Section 5. Finally, concluding remarks are given in Section 7.

2. Related Work

As early as 2003, Daugman [3] proposed using a Fourier transform to detect the highly periodic fake iris pattern that was prevalent in textured lenses manufactured at that time. Newer lenses have multiple layers of printing, making the Fourier response less pronounced, and textured lens detection by this method less reliable. Additionally, not all textured lenses necessarily use a dot-matrix style printing method.

Lee et al. [11] 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 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 False Reject Rate of 0.33% on the data, but suggest that the dataset may be too small to draw generalized conclusions. This work is substantially different from the classifier trained on texture features approaches described in this paper and used by the

remaining citations below.

He et al. [6] propose training a support-vector machines on texture features in a gray-level co-occurrence matrix (GLCM). They constructed a dataset of 2000 genuine iris images from the SJTU v3.0 database and 250 textured lens images, of which 1000 genuine and 150 textured are used for training. They report a correct classification rate of 100% on the testing data.

Wei et al. [15] 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 and BATH for genuine iris images and a special acquisition for textured contact lenses. Each dataset contained 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. [7] 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 and ICE v1.0.

Zhang et al. [16] 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.

Doyle et al. [4] present an analysis of modified 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 trained classifiers is constructed. A dataset of 1000 images from each of the three classes is used for training, and a dataset of 400 images per class is used to test. A correct classification for the three class problem is 71% but increases to 98% when detecting textured lenses alone.

Kohli et al. [10] 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.

3. Problem Definition

Textured contact lenses are designed to alter the appearance of the wearer's eye, giving it a different color and/or texture. Unfortunately, they also greatly reduce the amount of genuine iris texture visible to iris recognition systems, greatly increasing the chance of a false non-match. Accordingly, these images should be rejected at the time of segmentation, before a template is generated for them.

Soft lenses have additionally been shown[1] to negatively impact the performance of an iris biometrics system. Users who wear soft lenses may experience a higher false non-match rate than users who do not wear soft contact lenses. A first step in correcting this lower true accept rate is determining which subjects are wearing soft contact lenses.

Sample images of an iris with no contact lens, a soft contact lens, and a textured lens appear in Figures 1 and 2. Figures 1(c) and 2(c) show how a textured lens obscures genuine iris texture.

4. Experimental Method

4.1. Data Set

Two datasets were constructed for the evaluation of textured lens detection. Dataset I consists of a training set of 3000 images and a verification set of 1200 images. All images were acquired with an LG 4000 [12] iris camera. Both the training set and the verification set are divided equally into three classes: (1) no contact lenses, (2) soft, non-textured contact lenses, and (3) textured contact lenses. Classes (1) and (2) are balanced between male and female, and represent a variety of ethnicities. Category (3) images are predominantly from Caucasian males. Dataset II consists of a training set of 600 images and a verification set of 300 images. All images were acquired with an IrisGuard AD100 [8] iris camera. Again, the dataset is balanced across the three categories in the same manner.

All textured contact lenses in the dataset came from three major suppliers of textured lenses: Johnson&Johnson [9], Ciba Vision [2], and Cooper Vision [14]. Multiple colors were selected for each manufacturer and some lenses were also toric lenses designed to correct for astigmatism.

Both datasets are segmented using a commercially available iris biometrics SDK to extract center and radius for circles defining the pupillary boundary and the limbic boundary. Segmentations for the training sets were inspected visually by overlaying circles defined by the segmentation algorithm. Ill-fitting circles were cor-

(a) No Lens - Full (b) Soft Lens - Full (c) Textured Lens - Full

(d) No Lens - Pupil (e) Soft Lens - Pupil (f) Textured Lens - Pupil

(g) No Lens - Iris (h) Soft Lens - Iris (i) Textured Lens - Iris

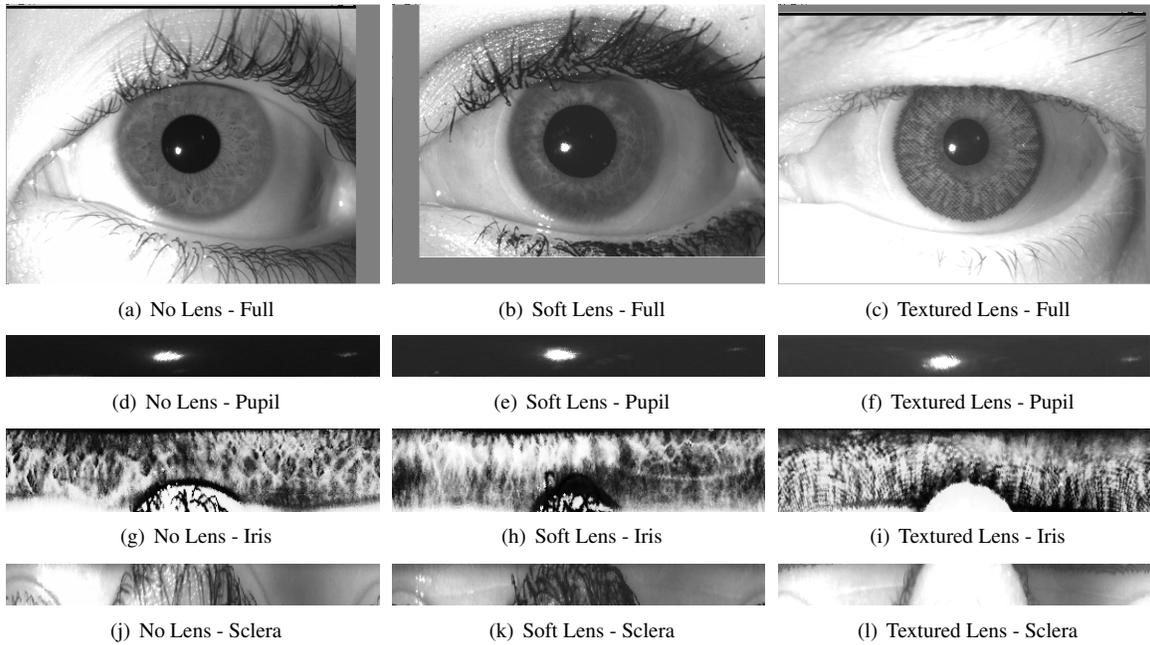(j) No Lens - Sclera (k) Soft Lens - Sclera (l) Textured Lens - Sclera

Figure 1. Sample LG400 images for the three classes showing the original images and the unrolled sections from which the features were extracted. The no lens images were taken from sample 05629d33. The soft lens images were taken from sample 05675d5684. The textured lens images were taken from sample 04261d2211.

(a) No Lens - Full (b) Soft Lens - Full (c) Textured Lens - Full

(d) No Lens - Pupil (e) Soft Lens - Pupil (f) Textured Lens - Pupil

(g) No Lens - Iris (h) Soft Lens - Iris (i) Textured Lens - Iris

(j) No Lens - Sclera (k) Soft Lens - Sclera (l) Textured Lens - Sclera
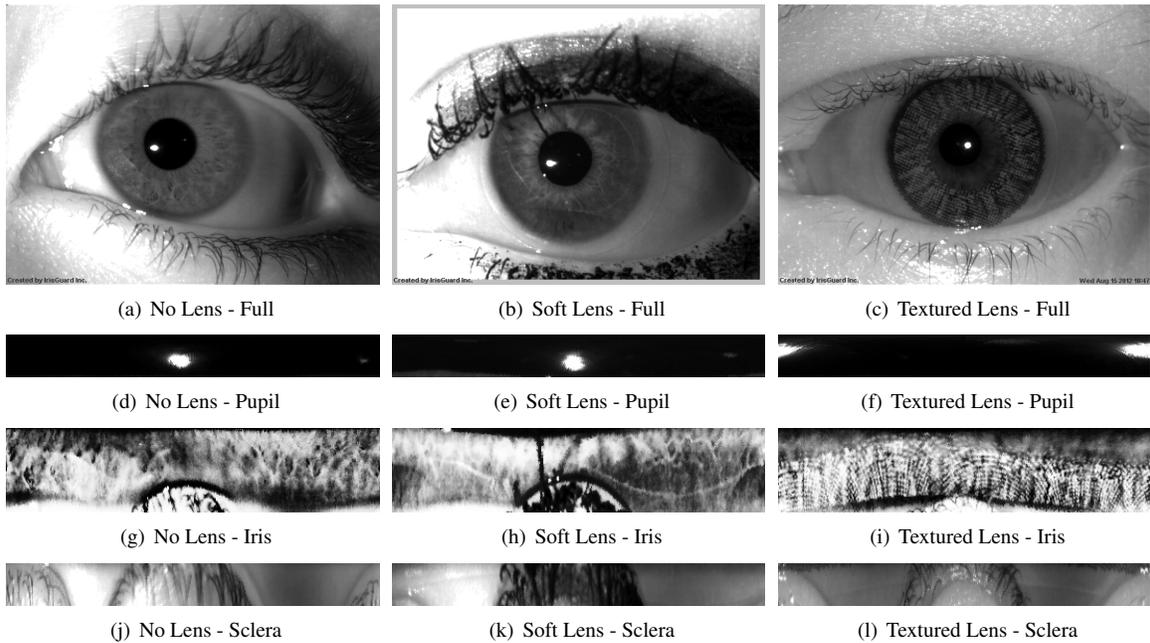
Figure 2. Sample AD100 images for the three classes showing the original images and the unrolled sections from which the features were extracted. The no lens images were taken from sample 05629d932. The soft lens images were taken from sample 05675d1366. The textured lens images were taken from sample 04261d3849.
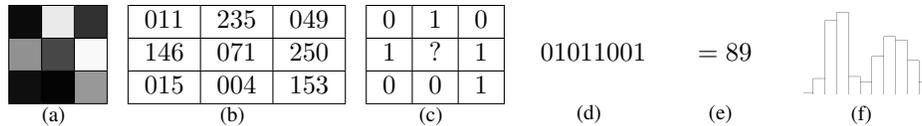
011	235	049
146	071	250
015	004	153

0	1	0
1	?	1
0	0	1

01011001   = 89

(a)   (b)   (c)   (d)   (e)   (f)

Figure 3. A $3x3$ neighborhood of pixels (a) with values (b) can be converted into a Binary Pattern by comparing perimeter values against the center value. A larger perimeter value gets replaced with a 1 and a smaller or equal perimeter value is replaced with a 0 in (c). Starting with the upper left neighbor and working clockwise, a binary number (d) is formed. The binary value is converted to decimal format for a final scalar value (e) and added into the histogram of values for the entire image (f).

rected. The verification set segmentation was not verified, to better simulate a real-world iris biometrics system. The data is also divided into 10 subject-disjoint and class-balanced folds for training evaluation.

4.2. Feature Extraction

The segmentation divides each iris image into three regions: (1) pupil, (2) iris, and (3) sclera. Examples of extracted regions appear in Figures 1 and 2.

Modified Local Binary Pattern analysis (described in Figure 3 and similar to [13]) is applied to each region of each image at multiple scales to produce feature values. The pupil, iris, and sclera all have significantly different appearances, and as such the binary pattern analysis is performed separately for each region. Unlike traditional LBP, this method does not decompose the image into blocks and independently analyze each block to construct a large feature vector. Instead, the extracted region is treated as one large block. The kernel size for the binary pattern analysis is scaled from 1 to 20 in increments of 1 for a total of 20 different feature sets for each of three regions and 60 feature sets overall.

4.3. Model Training

Sixteen different classifiers, intentionally sampling a variety of different classifier technologies [5], were explored as possible approaches to train models on the feature sets. Each of the feature sets described in Section 4.2 is treated as an independent dataset for the purposes of model training.

For data-mining algorithms that had tunable parameters, a parameter sweep was conducted with reasonable values. The predefined folds for each dataset are used to evaluate the performance of each trained model by cross-fold evaluation. If a classifier yielded a correct classification rate (CCR) of 100% on all 10 folds, a model was built using all training data. This process resulted in an ensemble of trained models to be evaluated on the verification set. The total number of trained models was 472 for Dataset I and 983 for Dataset II.

5. Novel Sensor Experimental Results

Two experiments were performed to evaluate the correct classification rate of the constructed model ensembles on Dataset I and II. Both the intra-sensor and inter-sensor cases are considered.

5.1. Intra-Sensor Validation

The performance of the ensembles built for both training datasets is evaluated on the corresponding verification sets. For each image of each verification set, a prediction and a confidence is output by each of the model ensembles. A final prediction for each image is decided by taking the maximum of the sum of confidences for each ensemble for each class. Both datasets yield similar accuracy for detecting textured lenses in the Inter-sensor experiment. Note that each classifier in the ensemble was achieving 100% correct classification on the training set.

For Dataset I, the final ensemble resulted in a CCR of over 65% on the three-class problem. Nearly 83% of the images containing a textured contact lens were correctly detected. The confusion matrices can be found in Figure 4.

For Dataset II, the final ensemble resulted in a CCR of 71% on the three-class problem. The accuracy of detecting instances of textured contact lenses was quite high. Nearly 96% of the images containing a textured contact lens were correctly detected. The confusion matrices can be found in Figure 5.

5.2. Inter-Sensor Validation

The performance of the ensembles built for both training datasets is evaluated on the other's verification sets. For each image of each verification set, a prediction and a confidence is output by each of the model ensembles. A final prediction for each image is decided by taking the maximum of the sum of confidences for each ensemble for each class. Both datasets perform about equally in the Inter-sensor experiment.

For Dataset I models on Dataset II data, the final ensemble resulted in a CCR of 53% on the three-class

	None	Soft	Textured	Total
None	258	172	33	463
Soft	35	162	1	198
Textured	107	66	366	539
Total	400	400	400	1200

(a) 3-Class Problem

	None/Soft	Textured	Total
None/Soft	627	34	661
Textured	173	366	539
Total	800	400	1200

(b) 2-Class Problem (Textured/Non-Textured)

Figure 4. Intra-sensor confusion matrices for the 3-class problem and the 2-class problem on Dataset I. For each matrix, the true class is the column label and the resulting classification is the row label.

	None	Soft	Textured	Total
None	44	25	4	73
Soft	14	19	0	33
Textured	42	56	96	194
Total	100	100	100	300

(a) 3-Class Problem

	None/Soft	Textured	Total
None/Soft	102	4	106
Textured	98	96	194
Total	200	100	300

(b) 2-Class Problem (Textured/Non-Textured)

Figure 6. Inter-sensor confusion matrices for the 3-class problem and the 2-class problem on Dataset II data using Dataset I models. For each matrix, the true class is the column label and the resulting classification is the row label.

	None	Soft	Textured	Total
None	80	55	4	139
Soft	19	39	2	60
Textured	1	6	94	101
Total	100	100	100	300

(a) 3-Class Problem

	None/Soft	Textured	Total
None/Soft	193	6	199
Textured	7	94	101
Total	200	100	300

(b) 2-Class Problem (Textured/Non-Textured)

Figure 5. Intra-sensor confusion matrices for the 3-class problem and the 2-class problem on Dataset II. For each matrix, the true class is the column label and the resulting classification is the row label.

	None	Soft	Textured	Total
None	83	48	94	225
Soft	215	250	136	601
Textured	102	102	170	374
Total	400	400	400	1200

(a) 3-Class Problem

	None/Soft	Textured	Total
None/Soft	596	230	826
Textured	204	170	374
Total	800	400	1200

(b) 2-Class Problem (Textured/Non-Textured)

Figure 7. Inter-sensor confusion matrices for the 3-class problem and the 2-class problem on Dataset I data using Dataset II models. For each matrix, the true class is the column label and the resulting classification is the row label.

problem, a significant drop in performance over the intra-sensor validation. The accuracy of detecting instances of textured contact lenses also dropped to 66%. The confusion matrices can be found in Figure 6.

For Dataset II models on Dataset I data, the final ensemble resulted in a CCR of 42% on the three-class problem, a significant drop in performance over the intra-sensor validation. The accuracy of detecting instances of textured contact lenses also dropped to nearly 64%. The confusion matrices can be found in Figure 7.

6. Novel Lens Experimental Results

A third experiment was performed to evaluate the effects of a previously unseen textured lens. Models were trained on images of two of the three textured lens manufacturers, and then the models were evaluated with images of lenses from the third manufacturer. All images were acquired with the LG4000 camera and come from Dataset I. A constant set of no lens and soft lens images, a subset of Dataset I, was also used to preserve the three-class problem as presented in Section 5. For all three lens manufacturers, the introduction of a novel lens in the testing set reduced the correct classification rate from the observed rate of 100%. The performance hit ranges from a minimum of 4% to a maximum of 43%, an extremely significant degradation. Full results can be seen in Figure 10. Sample images for each of the three textured lens manufacturers can be found in Figures 8 and 9.

On average, the novel Cooper Vision lens was the least affected of the three in the novel lens analysis. A visual inspection of the lenses as presented in Figure 9 may offer some insight into why this is the case. The Ciba Vision lens exhibits a clear dot-matrix pattern.
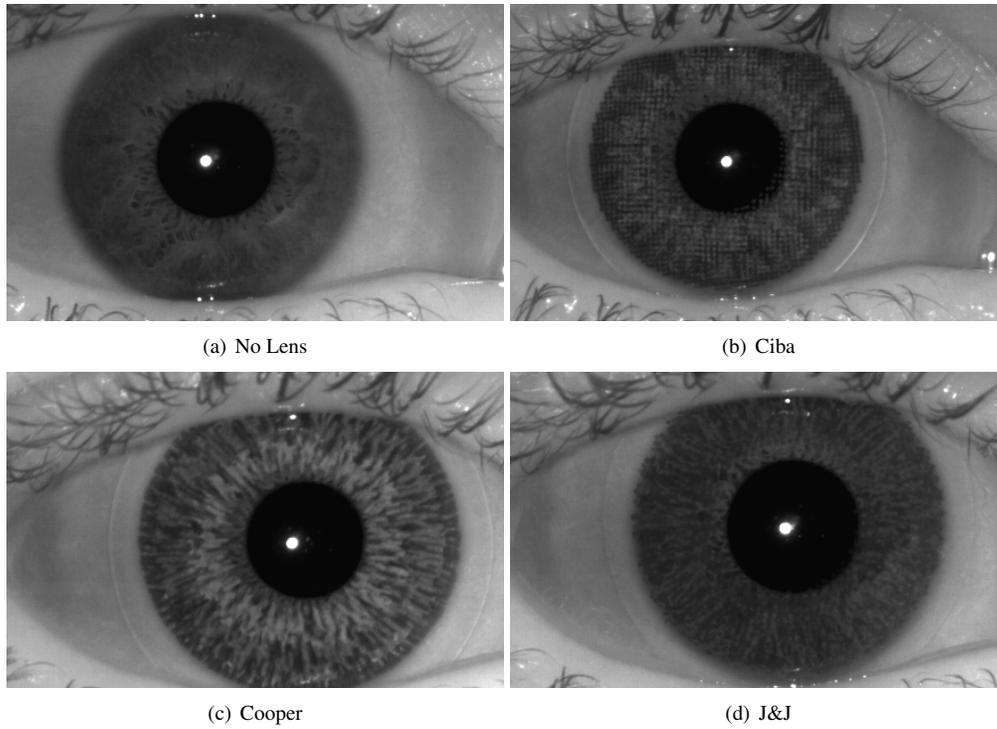
(a) No Lens

(b) Ciba

(c) Cooper

(d) J&J

Figure 8. Enhanced and cropped AD100 sample images for the three textured lens manufacturers in this work in the same eye.
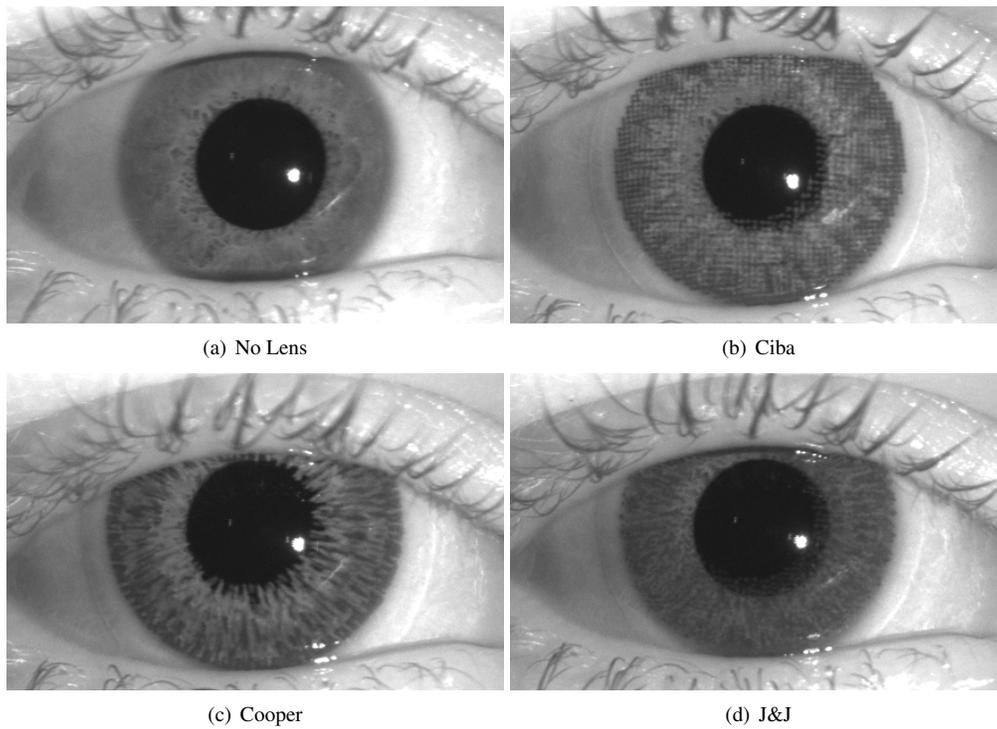

(a) No Lens

(b) Ciba

(c) Cooper

(d) J&J

Figure 9. Cropped LG 4000 sample images for the three textured lens manufacturers in this work in the same eye.

Classifier	Ciba	Cooper	J&J
Naive Bayes	75.34%	90.59%	72.67%
Bagging	62.74%	94.52%	72.07%
LogitBoost	72.13%	92.86%	67.22%
JRip	61.35%	95.24%	67.28%
J48	65.28%	91.83%	60.05%
Rand. Forest	57.38%	95.97%	64.11%

Figure 10. Classifier performance when a novel lens is used to evaluate the trained models. All images were acquired using the LG4000 sensor. The lens manufacturer at the top of the column represents the novel lens. For instance, data in the first column is trained on samples of Cooper Vision and Johnson&Johnson lenses and then evaluated on images from Ciba Vision.

The Johnson&Johnson lenses are, in contrast, streaks of texture oriented towards the center-point of the lens. The Cooper Vision lens appears to be somewhere between the Ciba Vision and Johnson&Johnson lenses in its printing technique.

7. Discussion

One major result of this work is to highlight the sensitivity of textured contact lens detection to the composition of the training data. A classifier that achieves 100% correct detection of textured lenses when trained on a mixture of three lens types may fall to below 60% when trained on just two of the lens types and encountering the third as a new lens type. All previous work in this area has considered only the experimental paradigm in which the same type of textured contact lens was present in both the training and the test data.

Another major result of this work is to highlight the sensitivity of textured contact lens detection to the composition of the training data in the sense of the iris sensor involved. A classifier trained on images from one type of iris sensor may have a significant degradation in performance when applied to images from a different type of sensor.

Further experimentation will examine more manufacturers of textured lenses. More advanced features than the simple binary patterns evaluated in this work may yield higher correct classification rates and me more general across different sensors and lens types. Future work will also simplify features and classifiers to only those actually needed.

References

[1] S. E. Baker, A. Hentz, K. W. Bowyer, and P. J. Flynn. Degradation of iris recognition performance due to non-cosmetic prescription contact lenses. *Computer Vision and Image Understanding*, 114(9):1030–1044, 2010.

[2] CibaVision. Freshlook colorblends, April 2013. `http://www.freshlookcontacts.com`.

[3] J. Daugman. Demodulation by complex-valued wavelets for stochastic pattern recognition. *Int'l Journal of Wavelets, Multiresolution and Information Processing*, 1(01):1–17, 2003.

[4] J. Doyle, K. Bowyer, and P. Flynn. Automated classification of contact lens type in iris images. In *Proc. of the IAPR 6th Int'l Conf. on Biometrics*. IEEE, 2013.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18, nov 2009.

[6] X. He, S. An, and P. Shi. Statistical texture analysis-based approach for fake iris detection using support vector machines. In *Proc. of the 2007 Int'l Conf. on Advances in Biometrics*, pages 540–546, 2007.

[7] Z. He, Z. Sun, T. Tan, and Z. Wei. Efficient iris spoof detection via boosted local binary patterns. In *Advances in Biometrics*, pages 1080–1090. Springer, 2009.

[8] IrisGuard. Ad100 camera, April 2013. `http://www.irisguard.com/`.

[9] Johnson&Johnson. Acuvue2 colours, April 2013. `http://www.acuvue.com/`.

[10] N. Kohli, D. Yadav, V. M., and R. Singh. Revisiting iris recognition with color cosmetic contact lenses. In *Proc. of the IAPR 6th Int'l Conf. on Biometrics*. IEEE, 2013.

[11] E. C. Lee, K. R. Park, and J. Kim. Fake iris detection by using purkinje image. In *Proc. of the IAPR Int'l Conf. on Biometrics*, pages 397–403, 2006.

[12] LG. Lg 4000 camera, October 2011. `http://www.lgiris.com`.

[13] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[14] C. Vision. Expressions colors, April 2013. `http://coopervision.com/`.

[15] Z. Wei, X. Qiu, Z. Sun, and T. Tan. Counterfeit iris detection based on texture analysis. In *Proc. of the 19th Int'l Conf on Pattern Recognition*, pages 1–4. IEEE, 2008.

[16] H. Zhang, Z. Sun, and T. Tan. Contact lens detection based on weighted lbp. In *Proc. of the 20th Int'l Conf. on Pattern Recognition (ICPR)*, pages 4279–4282. IEEE, 2010.