Challenging Ocular Image Recognition

V. Paúl Pauca, Michael Forkin, Xiao Xu, Robert Plemmons, and Arun Ross

Department of Computer Science, Wake Forest University, Winston-Salem, NC 27109;
Department of Mathematics, Wake Forest University, Winston-Salem, NC 27109;
Department of Computer Science & Electrical Engineering, West Virginia University, Morgantown, WV 26506;

ABSTRACT

Ocular recognition is a new area of investigation targeted at overcoming the limitations of iris recognition performance in the presence of non-ideal data. There are several advantages for increasing the area beyond the iris, yet there are also key issues that must be addressed such as size of the ocular region, factors affecting performance, and appropriate corpora to study these factors in isolation. In this paper, we explore and identify some of these issues with the goal of better defining parameters for ocular recognition. An empirical study is performed where iris recognition methods are contrasted with texture and point operators on existing iris and face datasets. The experimental results show a dramatic recognition performance gain when additional features are considered in the presence of poor quality iris data, offering strong evidence for extending interest beyond the iris. The experiments also highlight the need for the direct collection of ocular imagery.

1. INTRODUCTION

Biometric recognition is the science of recognizing the identity of a person based on the physical or behavioral attributes of the individual, such as face, fingerprints, voice, and iris. Biometric systems operate under the premise that many of these attributes or traits are distinctive to an individual and that they can be reliably acquired and represented in a digital format, enabling automatic decision-making for identity management purposes. With many applications ranging from picture tagging to secure access control, biometrics has become a viable technology for large-scale identity management systems and law enforcement.

Biometrics based on the use of the iris as a trait is perhaps one of the most reliable to date. The iris, an annular colored membrane surrounding the pupil, exhibits a very rich texture due to numerous structures that renders it essentially random from individual to individual. Several studies have established the uniqueness of the iris texture across individuals, under predominantly constrained conditions. The popularity of the iris as a biometric also stems from its availability for remote and noninvasive assessment, facilitating the development of standoff acquisition systems based upon machine vision.

However, non-ideal data resulting from challenging imaging conditions can significantly degrade the performance of iris recognition. Indeed current recognition systems generally accept for processing only those images that have passed a predetermined quality threshold, requiring acquisition of the iris under highly constrained imaging and illumination conditions. Images not meeting the required quality threshold are rejected by the iris recognition system. Figure 1 shows sample iris images that satisfy (a) and do not satisfy (b-e) typical quality thresholds. As a result, large numbers of images acquired under relaxed imaging constraints or of lower quality can remain unutilized.

Ocular recognition is an emerging field recently proposed as a way to enhance traditional iris recognition techniques in the presence of non-ideal data and several preliminary studies have explored information contained in extended regions around the iris. Usher, Tosa, and Friedman explored the use of retinal vasculature obtained under coherent light illumination. Derakhshani and Ross investigated the potential of blood vessel structure in the conjunctiva as an additional biometric, when the iris is “off-axis” with respect to the imaging device.

*This work is sponsored under IARPA BAA 09-02 through the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0013. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of IARPA, the Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.
Additional global and local features found in an extended region around the eye, a so-called periocular region, have also been considered by Park, Ross, and Jain,\(^9\) where a 77\% rank-1 recognition accuracy was established on RGB data collected with a commercial Canon SLR camera. Near infrared data is typically required for high accuracy iris recognition. The use of skin texture around the eye has also been recently considered by Miller et al.,\(^10\) where recognition rates of nearly 90\% were obtained from FRGC and FERET datasets. The appearance of skin texture, however, can be strongly affected by the direction from which it is viewed and illuminated.\(^11\) Further study may be needed on its implication for biometric recognition.

In this work, we seek to further establish the use of information in the ocular region as a viable alternative for enhancing iris recognition in non-ideal data resulting from challenging imaging scenarios. Our approach is two fold:

- First, we developed the Challenging Ocular Image Recognition (COIR) database of metadata associated with existing face and iris image and video datasets, e.g. MBGC, FRGC, etc.
- Second, we empirically characterize the performance of iris versus ocular recognition on large non-ideal datasets in the COIR database, noting their weaknesses, advantages, and constraints.

Metadata collected in the COIR database allows us to carefully sieve through the large amounts (GB) of existing data, collected specifically for face and iris recognition, to glean a fundamental understanding of ocular recognition as a function of ocular features as well as the external factors (illumination, occlusions, etc.) directly affecting recognition performance. In addition to leveraging the reuse of imagery from the existing face and iris datasets, the collected metadata and associated recognition results can help identify and guide future efforts for the collection of a dedicated ocular dataset.

There are several issues that must be addressed in attempting to offer a clear definition of ocular recognition and addressing all these issues is beyond the scope of this work. The size and boundaries of the ocular region remains an open issue to be addressed. As seen in Figure 2, the size and bounds of the ocular region are not easily defined. Features of interest may include eye shape, eyelashes, skin tags, bone geometry, and eyebrows, in addition to any iris texture available. Some of these features, such as, wrinkles, may also be age dependent.
Our focus in this paper is to characterize the performance of iris versus ocular recognition on RGB imagery taken under significantly more relaxed conditions than is usual for iris recognition. We do not exploit any information about the imaging device, the ocular regions used for recognition are of different sizes both in terms of pixels and optical resolution, and there is no full control of the illumination. The underlying dataset consists of imagery extracted from MBGC 1.0 and 2.0. In the remainder of this paper we propose bounds for the ocular region, introduce the COIR database for the collection of such data from the existing face and iris datasets, and present preliminary results on the characterization of iris versus ocular recognition performance.

2. OCULAR REGION AND THE COIR DATABASE

Current face and iris datasets contain thousands of images and video taking more than 300 GB of space. Much of this data has been collected at laboratories at the University of Notre Dame, University of Texas at Dallas, and West Virginia University, under a variety of imaging conditions and using various commercial grade cameras. Figure 3 shows a sample of face and iris imagery from the MBGC and WVU datasets. As can be appreciated, face images contain significantly more ocular feature information than the iris imagery at the expense of a loss of definition of the iris itself. Most of the collected iris imagery is of good quality in terms of resolution of the iris but contain little additional ocular feature information. Though a great amount of existing iris imagery is not of interest for studying ocular recognition, a subset of such imagery can be used for studies of iris recognition when the iris is occluded or shows an off-angle gaze as seen in Figure 3. This representative sample of face and iris images also highlights the need for the collection of a dedicated set of ocular imagery where quality (with respect
to resolution, noise, gaze direction, occlusions, and illumination) and imaging factors affecting performance can be studied in isolation.

In this work, we propose defining the ocular region in terms of the coordinates of a bounding box around the region of interest as shown in Figure 4. In particular, we define:

- **biocular region**: the set of image coordinates specifying the rectangular area from the top of the eyebrows to the tip of the nose, and from the left to the right temple.

- **(left/right) monocular region**: the set of image coordinates specifying the rectangular area from the top of the eyebrow to the middle of the cheekbone, and from the temple to just past the inside eye corner.

![Figure 4. Proposed boundaries for the ocular region. (a) Biocular region. (b) Left monocular region. (c) Right monocular region.](image)

Clearly, the above definitions contain ambiguities that can make the relative size of the ocular region change from person to person. For example determining precisely the middle of the cheekbone or temple locations can be a subjective task since there are no common landmarks in these regions from person to person (whereas tip of the nose is more easily determined). However, regardless of the size and exact location, the ocular regions should contain minimally the entire eyebrows, eyes, and bridge of the nose when appropriate.

Defining the ocular region in terms of image coordinates (instead of a cropped image region) is crucial to facilitate the study of ocular recognition. Image coordinates are stored as part of metadata associated with an image and can be used in conjunction with additional metadata, such as interocular distance, to estimate the relative size of the ocular region (independent of number of pixels) in the given image. The ocular region can then be refined appropriately if needed without incurring any additional data collection.

### 2.1 The COIR Database

As previously mentioned, the COIR database consists of metadata associated with existing face and iris image and video datasets. A key goal is to leverage existing face and iris imagery towards the study of ocular recognition. Figure 5 shows the relationship between the COIR database, the underlying raw and video imagery (Original
Image Datasets) and the user interface. The data in Original Image Datasets is proprietary and can be obtained directly from the institutions performing the data collection via appropriate licensing mechanisms. The COIR database stores metadata associated with each image and video frame in the Original Image Datasets.

The user interface to the COIR database is implemented through two separate mechanisms: viewer and researcher. A viewer’s job is to provide metadata associated with specific tasks through a web-based interface (see Figure 6). A researcher queries the COIR database, using a MATLAB interface, to obtain metadata and select imagery fitting desired constraints. Neither viewers nor researchers can obtain the raw imagery itself from the COIR database. Other important features of this design include:

- Approved accounts: a user must request an account with viewer or researcher privileges by email directly from the COIR database administrator,
- Data security: Users do not have direct access to the underlying image data, which is protected by specific license agreements,
- Simple interface: viewers insert metadata through a web interface while researchers obtain and insert metadata through a secure API.

The back-end for the COIR database was implemented in MySQL and was designed to be easily integrated with existing datasets. Currently, MBGC v.1, MBGC v.2, and WVU multispectral data are available. The metadata stored for each image and video frame includes:

- Image coordinates of ocular/biocular region
- Presence of occlusions such as hair and eyeglasses
- Gaze position (on-angle/off-angle)
- Image orientation (portrait/landscape)
- Interocular distance in pixels, and
- coordinates of iris centers

The MATLAB interface for researcher access includes the following functions:

- `setUser(username, password)` initiates a secure connection to the COIR database.
- `getDatabases()` - returns a cell array of databases currently integrated into the system.
- `getImageList(database, region)` returns the names of all images in the specified database that contain metadata for the specified ocular region.
• `getImageData(filename, database, region, [method])` returns the metadata for image filename in given database and ocular region. If a method is specified then stored metadata corresponding to a user-task is also returned.

• `sendImageData(filename, database, region, method, result)` uploads researcher-defined metadata pertaining to image filename in given database and ocular region. The user must also specify a method and numeric value to be stored.

These functions can be used to quickly create image subsets satisfying a number of imaging requirements, such as resolution, occlusion type (e.g. hair or glasses), etc.

3. IRIS VS. OCULAR RECOGNITION USING NON-IDEAL DATA

The COIR database has been used to empirically characterize the performance of iris versus ocular recognition on large non-ideal datasets. Moreover, we seek to obtain a performance baseline against which new ocular image recognition algorithms can be compared.

3.1 Iris and Ocular Datasets

Table 1 shows specific characteristics of the four datasets selected for our study. The Iridian dataset consists of ideal iris imagery taken under near IR illumination. We consider this set to be ideal because most of its characteristics (illumination, iris diameter, etc.) are optimal for iris recognition. PortalChallenge consists of imagery showing face/upper body for each subject in a neutral background and roughly with the same illumination and camera-to-subject distance. MBGC-1 and MBGC-2 are subsets of imagery obtained from VisibleFaceStills in MBGC 2.0. These images differ significantly in resolution (camera to subject distance) and illumination conditions. It is important to notice that as a result the corresponding ocular regions can vary significantly in pixel size and resolution. For the MBGC data sets in Table 1 we chose images having between 30 and 250 pixels across the iris, however they were normalized to have between 70 and 90 pixel irides during matching.
<table>
<thead>
<tr>
<th>Subjects</th>
<th>Iridian</th>
<th>PortalChallenge</th>
<th>MBGC-1</th>
<th>MBGC-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images/subject</td>
<td>23</td>
<td>115</td>
<td>40</td>
<td>107</td>
</tr>
<tr>
<td>Total images</td>
<td>1–15</td>
<td>1–2</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Illumination</td>
<td>similar</td>
<td>similar</td>
<td>varying</td>
<td>varying</td>
</tr>
<tr>
<td>iris diameter</td>
<td>220 pixels</td>
<td>25–30 pixels</td>
<td>30–250 pixels</td>
<td>30–250 pixels</td>
</tr>
<tr>
<td>wavelength</td>
<td>Near IR</td>
<td>Visible</td>
<td>Visible</td>
<td>Visible</td>
</tr>
</tbody>
</table>

Table 1. Datasets used for characterization of iris vs. ocular recognition.

Figure 7 shows a sample of ocular images excluded from MBGC-1 and MBGC-2. On the left, the iris fits our diameter requirements but exhibits an unacceptable signal-to-noise ratio. On the right, the resolution of the ocular region is below the iris diameter criterion specified in Table 1.

3.2 Iris Recognition

While the performance of iris recognition on high quality data is well understood, less is known regarding how iris recognition performs with challenging non-ideal data. Several confounding issues that must be addressed include image resolution, eye gaze, occlusions, reflections, eye color (due to usage of RGB data instead of near IR data), and segmentation. We have chosen two iris recognition algorithms to provide a performance baseline: Libor Masek’s implementation of Daugman’s algorithm and an executable developed by Iridian Technologies, Inc.

- Libor Masek’s code. Libor Masek’s implementation\textsuperscript{13} is freely available and provides several pieces upon which many of the most recent implementations have been built. However, we chose Libor Masek’s original implementation since it is widely accessible to the research community.

- Iridian’s executable: This executable was developed several years ago by Iridian Technologies, Inc. and a copy was obtained by Wake Forest under a Cooperative Research and Development Agreement (CRADA). In addition, a fairly large iris dataset was obtained for testing purposes. The Iridian executable is the strongest implementation of Daugman’s algorithm we have tested.

3.3 Ocular Recognition using SIFT

The scale-invariant feature transform (SIFT) algorithm was originally proposed by Lowe\textsuperscript{14} for describing and detecting local image features showing partial invariance to translation, scale, and orientation changes. Applications in computer vision include object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, and video tracking. Recently, SIFT has also been used for face recognition\textsuperscript{15} as well as iris recognition.\textsuperscript{16} Features chosen by SIFT correspond to locations on the image resulting from maxima and minima of a difference of Gaussians function applied to a series of smoothed and re-sampled images. Key parameters controlling performance include diameter of the Gaussian function, magnification, and number of scales.
We have used SIFT to encode information in the ocular region. We use as input images cropped from the existing data sets, as defined by metadata obtained from the COIR database. Figure 8 shows the process employed for processing all input imagery. Input ocular images were normalized (by upsampling or downsampling) to be 500 pixels in height, resulting in irides of 70 to 90 pixels in diameter for the MBGC-1 and MBGC-2 data sets (see Table 1). SIFT parameters were tuned to achieve appropriate separation of the frequency distributions of true negative and true positive scores. In particular, we selected $4 \times 4$ spatial sections, 16 orientations, and a magnification factor of 5 (number of octaves). A smaller subset of MBGC data was used for training and this set did not overlap with the subset used for testing. As shown in Figure 8 input imagery was preprocessed using histogram equalization and multiple images per subject were used in the gallery. Keypoint matching was performed using the ratio of distance from the closest neighbor to the distance of the second-closest neighbor as well as the image distance between keypoints in the images being matched. Finally, the max-rule was used to obtain a single match score from the comparison of the probe to the set of gallery images.

4. RESULTS AND DISCUSSION

4.1 Segmentation Performance

Segmentation is an important consideration in the study of iris and ocular recognition. Without segmentation of the iris it is not possible to enroll and appropriately encode the information. For the Iridian dataset (300 near-IR images), Libor Masek’s implementation was able to segment roughly 75% of these images. On the other hand, the Iridian executable was able to segment 100% of the images. The Iridian dataset is our control dataset and is mainly used to highlight the efficacy of the iris algorithms on ideal data.

For the PortalChallenge dataset we used only the red channel from each of the 150 images. Masek’s implementation was able to segment 19% of the images while the Iridian executable was able to segment 46% of them. Figure 9 shows some examples of successful and unsuccessful segmentations obtained using both algorithms. The top row shows failed as well as successful segmentations obtained with Masek’s code. Notice that the poor quality of the iris on the top left image makes the inner iris boundary difficult to discern. The bottom row shows similar results using the Iridian executable. The Iridian executable has a strong segmentation routine and is able to better identify partial irides.

For the MBGC-1 dataset, segmentation performance of Masek’s code and the Iridian executable was very low, namely 4% and 7% respectively. The segmentation performance on MBGC-2 was similarly low and thus not fully recorded. Table 2 summarizes these segmentation performance results. These numbers were calculated from the ratio of the number of images successfully segmented over the total number of images in the dataset. Each image was manually inspected to evaluate performance.

†Several modifications of Masek’s implementation exist that can produce significantly improved segmentation results. We chose the original Masek implementation to avoid confusion
4.2 Iris Recognition Performance

Figure 10 shows receiver operating characteristic (ROC) curves for the performance of both Masek’s code and the Iridian executable on the Iridian and PortalChallenge datasets. As can be expected, both implementations perform very well on ideal data (left). The Iridian executable is clearly superior and contains no false matches. Though not as robust, Masek’s code has a true positive rate of nearly 90% for a false positive rate of $10^{-3}$. The recognition performance over the PortalChallenge dataset is not nearly as good (right). Notice that the maximum value of each curve corresponds to the maximum amount of imagery that each algorithm can segment appropriately (see Table 2). This recognition rate is dramatically lower than recognition rates obtained with SIFT, showing strong evidence for the inclusion of ocular features other than the iris in the recognition process.

4.3 Ocular Recognition Performance

Figure 11(a) shows cumulative match curves (CMC) for the performance of SIFT on the MBGC-1 and MBGC-2 datasets. Note that the rank-1 accuracy on MBGC-1 and MBGC-2 is 95% and 93%, respectively. These results include the use of image distance between keypoints, in addition to the ratio of distances between the closest and second-closest neighbors. Note that the gallery and probe image subsets are disjoint and the images for these subsets were selected randomly.

Figure 11(b) shows the receiver operating characteristic (ROC) curve for the performance of SIFT on MBGC-2. The curves show performance obtained with matching based on the ratio of distances between near neighbors only (curve labeled Raw Match Score) and with the addition of image distance between keypoints (curve labeled...

---

Table 2. Overall segmentation performance of the iris recognition algorithms.

<table>
<thead>
<tr>
<th></th>
<th>near IR iris</th>
<th>RGB PortalChallenge</th>
<th>MBGC-1</th>
<th>MBGC-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Libor Masek</td>
<td>75%</td>
<td>19%</td>
<td>4%</td>
<td>-</td>
</tr>
<tr>
<td>Iridian Inc.</td>
<td>100%</td>
<td>46%</td>
<td>7%</td>
<td>-</td>
</tr>
</tbody>
</table>
(a) Iris recognition on near IR iris data.  
(b) Iris recognition on PortalChallenge RGB data.

Figure 10. Performance of iris recognition.

(a) CMC curves for ocular recognition on MBGC-1 and MBGC-2  
(b) ROC curve for ocular recognition on MBGC-2

Figure 11. Performance of ocular recognition on MBGC-1 and MBGC-2 datasets.

Adjusted Match Score). The Raw Match Score ROC curve shows a true positive rate of 80% can be achieved for an error rate of $10^{-3}$ (EER = 11.4%) while the Adjusted Match Score curve shows a true positive rate of 80% can be achieved for an error rate of $10^{-4}$ (EER = 10.3%). These performance results are dramatically higher than those obtained with iris recognition on the same non-ideal data.

One of the key advantages of SIFT over Masek’s code and the Iridian executable is that segmentation of specific features is not required. SIFT selects features that it deems to be robust with respect to scale, rotation and other transformations, and does not require the same number of features for matching. Figure 12 shows salient features typically selected by SIFT in the ocular region. Notice that the largest concentration of keypoints (marked with red dots) is around the eyes, followed by the eyebrows, and the skin (such as on the nose bridge and cheeks). Figure 12 also shows robustness of SIFT with respect to variations in illumination type. The images on the left and right columns differ on the type of illumination used, yet SIFT is able to find a sufficient number of matching keypoints. It is also important to notice that different illumination conditions such as type and angle can affect the way skin reflects light, as can be seen on the specular reflections and patterns found on the skin in between the eyebrows and around the eyes. These differences can significantly affect the use of skin reflectance as a biometric under visible light.

On the other hand, heavy shadows or large photometric differences in the ocular region can significantly affect the gradient information exploited by the SIFT algorithm. Figure 13 shows images of the same person where only a single keypoint was found to be matching. In these situations preprocessing with methods such as multiscale retinex maybe done to correct the photometric differences, before SIFT is applied.
5. CONCLUSIONS AND FUTURE WORK

We have proposed defining the ocular region in terms of the image coordinates of a bounding box that includes minimally the entire eyebrows, eyes, and bridge of the nose when appropriate (biocular region). When used in conjunction with additional metadata, such as eye to eye distance, the image coordinates can be used to estimate and easily change the relative size of the ocular region, regardless of the available resolution.

The COIR database of metadata was developed to compensate for the lack of a dedicated ocular dataset and to leverage the large amount of existing face and iris image and video data already available. The COIR database consists of a MySQL back-end for metadata storage and a front-end that allows for user interaction in two different modalities: viewer and researcher. A viewer is presented images and associated tasks to be completed via a web-interface. The metadata collected in this way is stored in the back-end. A researcher can query the COIR database through a secure API for imagery fitting specific constraints, thus obtaining metadata needed for a specific study. A MATLAB interface is currently implemented for researcher type access. No proprietary image data is accessible through the viewer or researcher interface modalities.

The COIR database was used to empirically characterize the performance of iris versus ocular recognition on large non-ideal datasets. Two iris recognition algorithms were considered, namely Masek’s code and an executable provided to us by Iridian Technologies Inc. Segmentation failures seem to be a key factor associated with recognition failures with either Masek’s code or the Iridian executable. Segmentation errors seem to happen most often when the algorithm cannot determine the boundary between the iris and the pupil. This is most likely due to the lack of contrast between the two regions when using data collected in the visible electromagnetic spectrum (RGB data). Indeed, there appears to be more segmentation failures on subjects with dark irides than on subjects with lighter colored irides. Other factors affecting segmentation include shadows, occlusions and
unsuitable lighting conditions. Even if segmentation were possible, the iris quality in this dataset is poor due to the fact the images were collected in the visible spectrum, have low resolution, and were taken in non-ideal lighting conditions. Thus, good iris recognition performance is not expected.

For ocular recognition, we used a MATLAB implementation of Lowe’s SIFT algorithm. Performance results obtained with ocular recognition are dramatically better than those obtained with iris recognition on the same non-ideal data. SIFT was able to reliably match ocular features without the need for segmentation. Moreover, during matching SIFT was able to deal with ocular regions that differed both in number of pixels and scale of the underlying features. SIFT features appear to be robust to illumination type (such as fluorescent or natural light) as long as shadows or photometric differences did not significantly affect the underlying gradient information. Methods such as multiscale retinex may be used for photometric correction, prior to the recognition task.

Our future work in ocular recognition will address the proposed definition of the ocular region to facilitate the automatic collection of metadata. In addition, we will also investigate the development of robust methods for the fusion of iris and ocular recognition to adaptively optimize performance based on specific quality metrics of the input data.

6. ACKNOWLEDGMENTS

The authors wish to thank Professor Vijayakumar Bhagavatula at Carnegie Mellon University for his insightful comments and suggestions on this work. We are also grateful to Ryan Barnard at Wake Forest University for implementing the video frame tagging component of the COIR database front-end. Finally, we thank Wake Forest University graduate student Edison Muño and Carnegie Mellon University graduate students Andres Rodriguez, Naresh Boddeti, Eui Seok Hwang, Seungjune Jeon, Qi Wu, and Kathy Brigham for kindly providing much of the metadata for the COIR database.

REFERENCES


