Ocular and Iris Recognition
Baseline Algorithm

Yooyoung Lee
Ross J. Micheals
James J. Filliben
P. Jonathon Phillips
Hassan A. Sahibzada
Ocular and Iris Recognition Baseline Algorithm

Yooyoung Lee
Ross J. Micheals
P. Jonathon Phillips
Hassan A. Sahibzada

Information Access Division
Information Technology Laboratory

James J. Filliben
Statistical Engineering Division
Information Technology Laboratory

National Institute of Standards and Technology
Gaithersburg, MD 20899

November 2011
Due to its distinctiveness, the human eye is a popular biometric feature used to identify a person with high accuracy. The “Grand Challenge” in biometrics is to have an effective algorithm for subject verification or identification under a broad range of image and environmental conditions. As a response to the challenge, this paper presents baseline performance results derived from an enhanced version of VASIR (Video-based Automated System for Iris Recognition), as well as initial performance results based on a broader ocular recognition system. We describe the details of the VASIR procedure and demonstrate its superiority over the IrisBEE (Iris Biometric Evaluation Environment) baseline algorithm. We examine the relationship between VASIR performance and image quality scores. Finally, for less-constrained imaging conditions, we provide a comparison of iris and ocular recognition results.

1. Introduction

Human physiological or behavioral characteristics can be integrated into a system to identify a person; such a system is called a Biometric system. Specifically, the human eye is a highly distinctive feature that can provide a promising security application which we call iris recognition.

Iris recognition is a method to identify a person which focuses on and makes use of the distinctive and unique patterns of the human iris.

Ocular recognition is a multi-components method which makes use of a variety of features in and around the human eye: eyebrows, eyelashes, eyelids, eye shape, sclera, iris, pupil, etc. Iris recognition makes use of some of the same multiple components, but only as a means to an end—to navigate and to extract the iris itself, from which all subsequent analytic comparisons are done. On the other hand, ocular recognition treats these components as features onto themselves which become part of the comparative process.

A challenge of current iris recognition systems is often limited by significant input and environmental restrictions related to image acquisition. For example, an iris recognition system might perform well only for ideal imaging conditions which were captured with devices specially designed for the system.

With more factors at its disposal, ocular recognition has the potential for higher comparative power—bringing that potential to a practical reality is the challenge that this paper addresses.

The broad goal of this paper is to encourage the development of robust ocular/iris recognition algorithms that yield accurate biometric results for a broad range of image and environmental conditions; e.g., a near-infrared (NIR) video frame captured at a distance with an Iris On the Move (IOM) 1 system which we call “distant-image-based” contrasted with a classical image acquisition system such as the LG 2200 EOU which we call “classic-image-based” iris recognition.

Ocular/iris recognition under challenging imaging conditions is a relatively new research subject which still needs to overcome a number of issues. The main purpose of this paper is to provide a performance reference for ocular/iris recognition to address current problems.

As part of the description of ocular systems, this paper describes a particular iris recognition system VASIR (Video-based Automated System for Iris Recognition) [1] which was developed to address, via a rigorous and structured design and analytic methodology, such image and environmental challenges.

To establish additional robustness in the VASIR system, we present a modified Hamming distance (HD) matching methodology for improving False Reject Rates (FRR).

To assess an ocular/iris system’s performance, we follow the same framework that was used for the VASIR project:
1) Is VASIR robustly superior to IrisBEE?
2) Can VASIR’s results be used as a state-of-the-art baseline for both classic-image-based and distant-image-based ocular/iris recognition?
3) How does VASIR performance change with the image quality?

1 The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology.
The VASIR system was evaluated using ICE (Iris Challenge Evaluation) 2005, MBGC (Multiple Biometrics Grand Challenge), and FOCS (Face and Ocular Challenge Series)—subset from MBGC—datasets provided by the National Institute of Standards and Technology (NIST) [2][3].

Because of the existing test methodology developed in the VASIR project, the paper as a whole will be structured so as to present VASIR iris results first followed by corresponding ocular results.

Section 2 gives a brief overview of IrisBEE and VASIR. Section 3 discusses a new matching method. Section 4 introduces ocular recognition. Section 5 describes the ICE, MBGC, and FOCS datasets. Section 6 describes detailed performance results vis-à-vis the three above-mentioned questions.

2. IrisBEE and VASIR

Most iris-based systems currently in existence are for controlled classic-image-based iris recognition. IrisBEE, based on Masek’s algorithm [2][4], was originally developed at NIST to be used as a reference implementation for such a classic-image-based iris recognition.

In the acquisition phase of IrisBEE, an image of the subject’s eye region is taken through classical image acquisition system (e.g., LG2200 EOU)—such an image is captured with careful image quality controls.

In contrast to the IrisBEE’s algorithm, VASIR was developed with less constrained video-based iris recognition in mind. VASIR also robustly allows performing classic-image to classic-image (CC), distant-image to classic-image (DC), and distant-image to distant-image (DD) iris recognition. In CC case, a classic-image is matched against other classic-images of the same person that were captured by the same device (LG2200 EOU system). DC means the video-frame captured at a distance (IOM system) is compared to classic-images, captured by a different camera. In DD case, the extracted iris region of distant-images is matched to other frames either from the same video or different video sequences of the same person.

Figure 1 illustrates the procedural components of VASIR. The VASIR image acquisition step includes a facial-visible NIR video captured while a person walks through a portal, in addition to a classical still iris image. Four additional steps are used to identify a person from less constrained video at a distance. First, VASIR detects the eye region within each frame. Second, the system automatically extracts an eye region sub-image. Then the algorithm measures the quality of the detected eye region sub-images by calculating the edge density within the images. Finally, VASIR selects the best quality image out of all frames—a procedure we call Automated Image Quality Measurement (AIQM) [1].

Figure 1: VASIR procedure

To facilitate comparing two iris images, Daugman’s polar coordinate-based system [5] is then employed to normalize the iris region. The 1D-Log-Gabor Filter is used to process the feature extracting and feature encoding of the normalized iris images, as introduced by Yao et al [6]. VASIR was developed with the normalization and feature extraction based on Masek’s algorithm [4].

3. Template Matching

The Hamming distance (HD) counts the number of positions (in two templates of the same size) for which the binary digits are different. The HD can be used to decide whether or not two iris templates are of the same person. A noise mask [4] helps to exclude the insignificant bits of a template using the equation:

$$\text{HD} = \sum_{i=1}^{N} (T_i \oplus Q_i) \cap (T_m_i \cap Q_m_i)$$

$$N - \sum_{i=1}^{N} (T_m_i \cup Q_m_i)$$

where $T$ and $Q$ are two bit-wise templates and $Tm$ and $Qm$ are the corresponding noise masks. $N$ is the total bits of a template.

For rotational inconsistencies between two iris templates, one template is two bit-wise shifted left or right and the HD value is selected from successive shifts [11].

However, this method may lead to matching failure because the pupil and iris may not be segmented to the exact circle boundaries—especially when the image has significant noise from motion blur or inconsistent lighting conditions. In reality, the iris outer boundary is not uniquely-defined due to its gradual gradient change from

[Image 353x479 to 544x709]
the iris to the sclera. Figure 2 illustrates a case of less-than-optimal segmentation of the pupil (a) and (b) iris boundaries.

We therefore propose a new matching method in which the template is shifted not only left and right (horizontal) bit-wise, but also upward and downward (vertical); the Hamming distance values for these X and Y shifts are indicated by HD_X and HD_Y, respectively. Figure 3 shows the detailed HD shifting method.

Further, we select successive shift from the values of X or Y (HD_XorY). Finally, the template can also shift in two directions at the same time—HD_XandY. For instance, when one template shifts by two bits to the left, it can also shift one bit up. The new matching method can help to cover both the rotational inconsistency and the segmentation inaccuracy.

For CC case, the matching scores are ordered by XorY, XandY, X, and then Y methods. XorY also leads with the best results followed by the Y, X, and XandY for DC. Interestingly, Y is better for DD, (opposite to the results of CC) and then XorY, X, XandY follow in order. Overall, XorY is better because XorY is best or second best in all three scenario cases.

4. Ocular Recognition

For ocular recognition, we took the high resolution NIR distant-images and used the iris location to extract an ocular region surrounding the eye of size 750 × 600 pixels. For comparison with HD_XorY, we performed direct cosine correlation—described in below—on the ocular regions to attain a similarity score:

\[
\frac{\sum_{i=1}^{n} T_i \times Q_i}{\sqrt{\sum_{i=1}^{n} (T_i)^2} \times \sum_{i=1}^{n} (Q_i)^2}
\]

The ocular region includes the entire eye, eyelashes, part of the nose, and sometimes the eyebrow. In addition we provide results from a fusion algorithm where we normalize and combine the scores from cosine correlation and HD_XorY to produce a similarity score for each pair of ocular regions.

By fusing the results from the iris alone and the entire ocular region, we examine whether we can better perform classification with the aid of the extra information contained in the ocular region as compared to the iris alone. If a fusion algorithm performs better than either the ocular
or iris algorithm by itself, then one would expect that the ocular region provides us with something extra that we don’t get with the iris alone.

5. Datasets

5.1. Iris Challenge Evaluation (ICE) 2005

The ICE 2005 data was collected at the University of Notre Dame in cooperation with NIST. The near-infrared iris still images are 640x480 in resolution and were captured by a LG EOU 2200 acquisition system.

<table>
<thead>
<tr>
<th>Position</th>
<th>No. of Images</th>
<th>No. of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Eye</td>
<td>1,528</td>
<td>120</td>
</tr>
<tr>
<td>Right Eye</td>
<td>1,425</td>
<td>124</td>
</tr>
</tbody>
</table>

The total number of iris images is 2,953 and only 112 subjects exist for both right and left out of the total 132 subjects. In the original ICE2005 dataset, the image “246260.tif” was mistakenly included in the left image set instead of in the right images [2], and consequently was set aside in our study.

5.2. MBGC (Multiple Biometrics Grand Challenge) and FOCS (Face and Ocular Challenge Series)

Our VASIR performance study used two different data types included in the MBGC dataset.

The near-infrared facial video (NIR_Face_Video or distant-image) were captured by an Iris On the Move (IOM) system; each frame is 2048x2048 pixels and the iris diameter is ~100 pixels. The near-infrared iris still images (NIR_Iris_Still or classic-image) were collected by an LG2200 EOU and are 640x480 in resolution and the iris diameter is ~200 pixels [12].

<table>
<thead>
<tr>
<th>IOM System</th>
<th>No.</th>
<th>LG2200 EOU</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total video</td>
<td>628</td>
<td>Left</td>
<td>4,025</td>
</tr>
<tr>
<td>Both eyes visible</td>
<td>598</td>
<td>Right</td>
<td>4,013</td>
</tr>
<tr>
<td>Total frames</td>
<td>11,341</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The detailed numbers for distant (IOM) and classic (LG2200) iris images can be found in Table 2.

Figure 5 shows the differences between distant-image and classic-image in the MBGC dataset.

FOCS dataset is a subset of the MBGC dataset, which contains 9,592 images (left and right); each image is 750x600 in resolution.

6. Experiments

In this section, we describe which performance metrics are applied and how our different experiments are designed and executed to evaluate VASIR performance.

Biometric systems are conventionally evaluated by the True Accept Rate (TAR)—known as Verification Rate (VR)—where two biometric samples are correctly decided to be the same person and the True Reject Rate (TRR) where the system correctly decided that two samples (genuine and imposter) are not the same person. The False Accept Rate (FAR) is defined as the rate where people are accepted as the same identity while in reality they are different and False Reject Rate (FRR) is where two samples from the same person are rejected. The TRR and FRR can be computed as FRR=1-TAR and TRR=1-FAR, respectively. The TAR and FAR are normally obtained by varying an acceptance decision threshold $\lambda$.

The Equal Error Rate (EER) is defined as the rate where FAR($\lambda$) and FRR($\lambda$) are the same value (i.e., intersect on a FAR and FRR vs. $\lambda$ plot). The Receiver Operator Characteristic (ROC) curve is a plot of TAR against FAR.

In terms of Verification (1:1), an image matches if its matching score has satisfied a certain threshold $\lambda$, whereas Identification (1:Many) normally indicates that an image matches if its score is the best. Our results are based on verification.
6.1. **Comparison of VASIR and IrisBEE**

Is VASIR robustly superior to IrisBEE? In order to answer this question, the ICE 2005 evaluation protocol [2] was applied to measure VASIR’s performance for comparison purposes. An algorithm compares a query image $Q_i$ to a target image $T_i$ to produce the similarity scores for matching and non-matching. The two criteria for the comparison are:

1) Verification Rate (VR) at False Accept Rate (FAR) of 0.001 (for which higher is better) and
2) Equal Error Rate (EER) value—for which smaller is deemed as better.

For this experiment, only classic-image to classic-image (CC) matching protocol is allowed since the ICE 2005 dataset were collected by a LG EOU 2200 system. Note that the images for the subjects were captured on different days.

Figure 6 shows the comparison between IrisBee's and VASIR's results by plotting the ROC curve. For the left eye, at FAR of 0.001, VASIR has a VR of 88.5% while IrisBEE is only 85.0%. For the right eye, VASIR’s VR is 90.3% while IrisBEE’s VR is 85.2%. Hence for both eyes, VASIR is an improvement over IrisBEE. For the EER criterion—(not shown), VASIR’s EER is better with ~4.4%, while IrisBEE’s EER is nearly twice larger.

The results show that even though VASIR was developed primarily as a tool for distant-image-based iris recognition, the system is robust for classic-image-based iris recognition as well.

6.2. **Baseline for Ocular/Iris Recognition**

Can VASIR’s results be used as a state-of-the-art baseline for both classic-image-based and distant-image-based ocular/iris recognition? We present the performance results based on the MBGC/FOCS dataset using VASIR for the purpose of providing a baseline both under less constrained distant-image-based and constrained classic-image-based ocular/iris recognition. They can be used as a reference for subsequent evaluation or for comparison with custom algorithms. These results may also provide an opportunity for other researchers to advance their ocular/iris recognition algorithms and to promote their technology.

For distant-image data, no eye region could be detected in 41 images—out of the total 9,592 images (false detection rate: 0.43%). For classic-image dataset, 4,025 images for the left and 4,013 images for the right eye were evaluated.

Furthermore, we provide a prototype of our experiment design showing how to evaluate performance results.

For non-matching scores, the query $Q_i$ is compared to all targets $T_i$ where the target subject is a different person. This full matching (one to all others) can be time-consuming and is not necessary since the non-matching scores are sufficiently consistent for virtually all query subjects. We therefore suggest randomly selecting a subsample of targets, all of a different person for getting the non-matching results; we found 50 targets to be a sufficient subsample. This not only reduces time cost but also allows us to characterize the performance for analysis purposes more easily. Our results show that in comparison full non-matching and random non-matching are nearly equivalent.

In Figure 7 (plots of the left and right ROC curves), we compare the performance results of CC, DC, and DD matching cases using HD_XorY matching method based on the results as shown in Figure 4.
The detailed performance rates are described in Table 3. The results show that CC matching has the best performance; DC and DD follow next as expected.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>VR at FAR = 0.001</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFT</td>
<td>RIGHT</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>0.837</td>
<td>0.067</td>
</tr>
<tr>
<td>DC</td>
<td>0.105</td>
<td>0.309</td>
</tr>
<tr>
<td>DD</td>
<td>0.118</td>
<td>0.326</td>
</tr>
</tbody>
</table>

Figure 8 shows the ROCs using the Iris, Ocular and Fusion algorithms on the NIR distant-imager. It is interesting to note that just cosine correlation on the ocular regions performs better for recognition purposes as compared to the iris recognition using the HD,Xor,Y distance measure.

<table>
<thead>
<tr>
<th>Method</th>
<th>VR at FAR = 0.001 for DD case</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFT</td>
<td>RIGHT</td>
</tr>
<tr>
<td>Fusion</td>
<td>0.289</td>
</tr>
<tr>
<td>Ocular</td>
<td>0.203</td>
</tr>
<tr>
<td>Iris</td>
<td>0.118</td>
</tr>
</tbody>
</table>

6.3. Image Quality Scores and Performance

How does VASIR performance change with image quality? Although several metrics[13][14] were used to address one or another aspect of image quality metrics, the edge density—as defined by the Sobel operator—was our primary image quality metric of choices. After first using the edge density metric to determine focus, the metric was again applied to arrive at a final order categorization of the quality of specific images.

The Sobel operator is designed to respond to edges running vertically and horizontally relative to the pixel...
grid. The $Dx$ and $Dy$ are applied separately to the input image to measure the gradient component in each orientation. The gradient magnitude is given by the following equation:

$$QS = \sqrt{Dx^2 + Dy^2}$$

which a higher score is considered better image quality in our evaluation.

Quality score by Flynn et al [10, 15]—obtained by measuring a pair of samples $q$ and $t$ from the ICE2006—is defined as,

$$QS(q,t) = \min(QS(q), QS(t))$$

To calculate a quality measure $QS(q,t)$ for a set of match scores $M$ and a set of non-match scores $N$, decision threshold $\lambda$ and image quality threshold $\tau$ are given as parameters. Therefore, $VR(\lambda, \tau)$ is computed as,

$$VR(\lambda, \tau) = \frac{|\{M \cap SS(q,t) \geq \lambda \land QS(q,t) \geq \tau\}|}{|\{M \cap QS(q,t) \geq \tau\}|}$$

$M$ is the set of match similarity scores and $SS(q,t)$ is the similarity score where the two images $q$ and $t$ are of the same person. The $VR$ is the fraction of similarity score $SS(q,t)$ greater than or equal to a decision threshold $\lambda$, where the quality score $QS(q,t)$ is greater than or equal to a threshold $\tau$.

Furthermore, $FAR(\lambda, \tau)$ is defined as,

$$FAR(\lambda, \tau) = \frac{|\{N \cap SS(q,t) \geq \lambda \land QS(q,t) \geq \tau\}|}{|\{N \cap QS(q,t) \geq \tau\}|}$$

$N$ is the set of non-match similarity scores and $SS(q,t)$ is the similarity score where the two images $q$ and $t$ are of the different persons. The $FAR$ is the fraction of similarity score $SS(q,t)$ greater than or equal to a decision threshold $\lambda$, where the quality score $QS(q,t)$ is greater than or equal to a threshold $\tau$.

Figure 9 shows the distribution of the quality scores and VASIR-based Hamming distances ($XorY$) based on left and right images for distant-image to distant-image (DD). The plot illustrates that the DD dataset consists of poor quality images; the scatter points are primarily found on the left side of the plot. The Hamming distance of non-matching scores is mainly found in the top quarter of the plot. The results show that VASIR’s performance and image quality scores (edge density) are correlated with the coefficient value -0.52 on DD matching, while the non-matching has only -0.11.

Following, the False Reject Rate (FRR) is calculated at six levels of the image quality. The quality score for DD case is calculated with the edge density ranges approximately from 10 to 35.

Figure 10 shows VASIR’s performance on the image quality for distant-video to distant-video (DD) matching.
We define a set of six quality levels (0 to 5) with thresholds \( \{\tau_0, \ldots, \tau_5\} \) that were uniformly divided; Quality level 0 (Q0) is all images; Quality level 1 (Q1) means all images with quality score \( \geq \tau_1 \); the remaining levels are Q2 \( \geq \tau_2, Q3 \geq \tau_3, Q4 \geq \tau_4 \) and Q5 \( \geq \tau_5 \) (best). For each quality level, VASIR’s performance is evaluated with an FRR value with FAR at 0.01, using only samples of all quality levels above the current level. Figure 10 illustrates the DD performance results of the eye images with the matching method by applying HD_XorY.

The plots in Figure 10 show that the FRR in the performance of DD case is reduced for both left and right when low quality images are excluded. It is important to note that the image quality was measured from the datasets “as-is” without localizing the iris region in an image. A possible problem is that the quality score will be high if the image includes a lot of eyebrows and eyelashes.

The positive side of these results is that the image quality (as measured via the edge density) may serve as a predictor of VASIR performance on DD matching. However, there is a need for analyzing quality metrics for good image quality datasets such as classic-image iris images. We are pursuing the investigation of this matter as well.

7. Conclusion

VASIR was evaluated using the ICE 2005 and MBGC/FOCS datasets. Using the ICE 2005 dataset, at FAR of 0.001, VASIR has a VR of 88.5% while IrisBEE is only 85.0% for the left eye. VASIR’s VR is 90.3% while IrisBEE’s VR is 85.2% for the right eye. For the EER criterion (where small is better), VASIR is \(-4.4\% \) for the left and right eyes while IrisBEE’s EER is considerably worse at twice the value. The results show that VASIR is superior to IrisBEE.

The four HD shifting methods in VASIR were evaluated using the MBGC dataset: For classic-image to classic-image (CC), the matching HD scores are ordered by XorY, XandY, X, and then Y. For distant-image to classic-image (DC), XorY also leads with results superior to Y, X, followed by XandY. Interestingly, for distant-image to distant-image (DD), Y is better (opposite to the results of CC) and then XorY, X, XandY in that order. Overall, we choose XorY as best because XorY is best or second best in all three cases.

We also provided a baseline of performance results on CC, DC, and DD ocular/iris recognition under a broad range of image and environmental conditions.

The ocular recognition produced promising results. We showed that straightforward cosine correlation on the ocular regions yields a better ROC curve than the iris regions on the distant-images. When we combined XorY scores with the scores produced by cosine correlation the result was a better ROC curve than either of the two separately. These results show that making use of the ocular region aids in recognition compared to the iris region alone.

Finally, our study shows that image quality is correlated with VASIR performance on DD matching case. Hence, the quality score metrics serve based on edge density as a predictor of VASIR performance. Various quality metrics on VASIR performance are currently in progress.

References


