

Image Averaging for Improved Iris Recognition

Karen P. Hollingsworth, Kevin W. Bowyer, and Patrick J. Flynn

University of Notre Dame

Abstract. We take advantage of the temporal continuity in an iris video to improve matching performance using signal-level fusion. From multiple frames of an iris video, we create a single average image. Our signal-level fusion method performs better than methods based on single still images, and better than previously published multi-gallery score-fusion methods. We compare our signal fusion method with another new method: a multi-gallery, multi-probe score fusion method. Between these two new methods, the multi-gallery, multi-probe score fusion has slightly better recognition performance, while the signal fusion has significant advantages in memory and computation requirements.

1 Introduction

The field of iris recognition is an active and rapidly expanding area of research [1]. Many researchers are interested in making iris recognition more flexible, faster, and more reliable. Despite the vast amount of recent research in iris biometrics, all past published work has relied on still iris images. Zhou and Chellapa [2] reported that using video can improve face recognition performance. We postulated that employing similar techniques for iris recognition could also yield improved performance. There is some prior research in iris recognition that uses multiple still images; for example, [3–6]. However, no researchers have published on how to use additional information available in iris video.

One problem with single still images is that they usually have a moderate amount of noise. Specular highlights and eyelash occlusion reduce the amount of iris texture information present in a single still image. With a video clip of an iris, however, a specular highlight in one frame may not be present in the next. Additionally, the amount of eyelash occlusion is not constant throughout all frames. It is possible to obtain a better image by using multiple frames from a video to create a single, clean iris image.

Zhou and Chellapa suggested averaging to integrate texture information across multiple video frames to improve face recognition performance. By combining multiple images, noise is smoothed away, and relevant texture is maintained. This paper presents experiments demonstrating how signal-level fusion of multiple frames in a video improves iris recognition performance.

2 Related Work

Video has been used effectively to improve face recognition. A recent book chapter by Zhou and Chellapa [2] surveys a number of methods to employ video in

face biometrics. As an example of a specific algorithm, Thomas et al. [7] use multiple frames from a video clip to improve performance of a face matcher. Thomas et al. use temporal continuity in video to improve face detection accuracy. They use both a rank-based average and score-based average when combining matching scores from different frames.

In contrast, literature in iris biometrics reports results exclusively on still images. Some iris biometric research has used multiple still images, but all such research uses score-level fusion, not signal-level fusion. The information from multiple images has not been combined to produce a better image. Instead, these experiments typically employ multiple enrollment images of a subject, and combine matching results across multiple comparisons.

Du et al. [3] showed that using three enrollment images instead of one increased their rank-one recognition rate from 98.5% to 99.8%. Ma et al. [4] also used three templates of a given iris in their enrollment database, and took the average of three scores as the final matching score. Krichen et al. [5] performed a similar experiment, but used the minimum match score instead of the average. Schmid et al. [6] present two methods for fusing Hamming distance scores. They compute average Hamming distance and also log-likelihood ratio. They find that in many cases, the log-likelihood ratio outperforms the average Hamming distance. In all of these cases, information from multiple images was not combined until after two stills were compared and a score for the comparison obtained. Thus, these researchers use score-level fusion.

Another method of using multiple iris images is to use them to train a classifier. Liu et al. [8] used multiple iris images for a linear discriminant analysis algorithm. Roy and Bhattacharya [9] used six images of each iris class to train a support vector machine. Even in training these classifiers, each still image is treated as an individual entity, rather than being combined with other still images to produce an improved image.

3 Data

We used an Iridian LG EOU 2200 camera to acquire iris data. The LG 2200 takes video data of a subject's eye, and the accompanying software selects specific images from the video to use for iris recognition. During acquisition, the eye is illuminated by one of three infrared LEDs: one above the lens, one to the left, and one to the right. While taking data, the camera cycles through the LEDs, picking one eye image when the top LED is lit, one image when the left LED is lit, and one image while the right LED is lit. Next, the software selects the best of the three selected images to use for recognition. The chosen frame must pass Iridian's quality tests. Our LG 2200 camera uses special software that allows us to save the entire set of three images, rather than just saving the best of the three.

Rather than processing still images we wanted to use multiple frames from video to improve performance. We split the NTSC video signal from the camera so that the analog video feed could be sent both to the computer running the

Iridian software, but also to a separate computer so that we could digitize and save the entire video clip instead of six images. These videos were digitized using a Daystar XLR8 USB digitizer and the resulting videos were stored in a high bit rate (nearly lossless) compressed MP4 format.

We collected a data set of 1061 videos during the spring of 2008. Each video clip was long enough for the Iridian software to capture two sets of three images. We captured videos of both left and right eyes for each subject; however, we treat each individual eye as a separate “subject”. We have not yet considered how fusing left and right iris information adds to performance. We captured videos of 296 different eyes. The first video of each subject was used as the gallery, and the remaining videos from each subject were used as the probe. For each subject, we have between one and five iris videos in the probe set. Our total probe set contains 765 videos. For any two videos from the same subject in our data set, they were acquired at least one week apart, and possibly up to two months apart.

LG sells iris cameras that are newer than the LG 2200. However, all of those cameras were designed for saving single images, not video. The videos used in this research were captured with the same camera as the iris videos released in 2008 for the Multiple Biometrics Grand Challenge (MBGC) evaluation [10]. We are not aware of any other data sets containing iris videos instead of iris stills.

4 Creating Average Images and Templates

4.1 Selecting Frames and Preprocessing

The auto-focus on the LG 2200 camera continually adjusts the focus in attempts to find the best view of the iris. Thus, some frames have good focus, while others suffer from severe blurring. To simplify our experiments, we refrained from developing our own quality metrics to select frames. Instead, we decided to use the six frames from the video that corresponded to the six frames that the Iridian software selected to save.

The raw video frames were not pre-processed like the still images that the Iridian software saved. We do not know what preprocessing is done by the Iridian system, although it appears that the system does contrast enhancement and possibly some deblurring. Some differences between the stills and the video frames may be due to differences in the digitizers used to save the signals. We used the Matlab `imadjust` function to enhance the contrast in each frame. This function scales intensities linearly such that 1% of pixel values saturate at black (0), and 1% of pixel values saturate at white (255).

Our next step was to segment each frame. Our segmentation software uses a Canny edge detector and a Hough transform to find the iris boundaries. The boundaries are modeled as two non-concentric circles. A description of the segmentation algorithm is given in [11].

We did not want segmentation problems to bias our results, so we discarded a few images where the circles clearly did not match the boundaries of the iris.

In addition, we discarded a few images because the iris was not visible in the image (e.g. the frame showed a blink, or the bridge of the nose).

4.2 Signal Fusion

For each video, we now had about six frames selected and segmented. We wanted to create an average image consisting only of iris texture. In order to align the irises in the six frames, we transformed the raw pixel coordinates of the iris area in each frame into dimensionless polar coordinates. In the polar coordinates, the radius r ranged from zero (adjacent to the pupillary boundary) to one (adjacent to the limbic boundary). The angle θ ranged from 0 to 2π . This yielded an “unwrapped” iris image for each video frame selected.

In order to combine the six unwrapped iris images, we wanted to make sure they were aligned correctly with each other. Rotation around the optical axis induces a horizontal shift in the unwrapped iris texture. We tried three methods of alignment. First, we identified the shift value that maximized the correlation between the pixel values. Second, we tried computing the iris codes and selecting the alignment that produced the smallest Hamming distance. Third, we tried the naive assumption that people would not actively tilt their head while the iris video was being captured. The first two approaches did not produce any better recognition results than the naive approach. Since the naive approach worked well, we used it in our subsequent experiments.

Parts of the unwrapped images contained occlusion by eyelids and eyelashes. Our iris software implements some eyelid and eyelash detection, and masks those areas. We computed an average unwrapped image from unmasked iris data in the six original images, using the following algorithm. For each (r, θ) position, first find how many of the corresponding pixels in the six unwrapped images are unmasked. If zero or one of the corresponding pixels from the six images are unmasked, mask that position in the average image. Otherwise, compute an average pixel value based on unmasked pixel values of the corresponding frames. (Therefore, the new pixel value could be an average of between two and six pixel intensities, depending on mask values.) To compute an average, we first summed original pixel values, then divided by the number of pixels, then rounded to the nearest unsigned, 8-bit integer.

Using this method, we obtained 296 average images from the gallery videos. We similarly obtained 765 average images from the probe videos. An example average image is shown in Figure 1. On the top of the figure are the six original images, and on the bottom is the average image fused from the six original signals.

4.3 Creating an iris code template

Our software uses one-dimensional log-Gabor filters to create the iris code template. The log-Gabor filter is convolved with rows of the image, and the corresponding complex coefficients are quantized to create a binary code. Each complex coefficient corresponds to two bits of the binary iris code – either “11”, “01”,

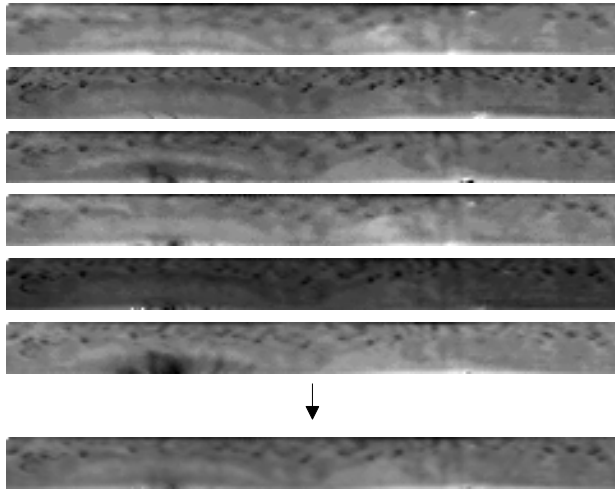


Fig. 1. From the six original images on the top, we created the average image shown on the bottom.

“00”, or “10” – depending on whether the complex coefficient is in quadrant I, II, III, or IV of the complex plane.

Complex coefficients near the axes of the complex plane do not produce stable bits in the iris code, because a small amount of noise can shift a coefficient from one quadrant to the next. We use fragile-bit masking [12,13] to mask out complex coefficients near the axes, and therefore improve recognition performance.

5 Experiments

5.1 Comparison to Previous Work

In biometrics, it has been found that enrolling multiple images improves performance [14–16]. Iris recognition is no exception. Many researchers [4–6] enroll multiple images, obtain multiple Hamming distance scores, and then fuse the scores together to make a decision. However, the different researchers have chosen different ways to combine the information from multiple Hamming distance scores.

Let N be the number of gallery images for a particular subject. Comparing a single probe image to the N gallery images gives N different Hamming distance scores. To combine all of the N scores into a single score, Ma et al. [4] took the average Hamming distance. We will call this type of experiment an N -to-1-average comparison. Krichen et al. [5] also enrolled N gallery images of a particular subject. However, they took the minimum of all N different Hamming distance scores. We call this type of experiment an N -to-1-minimum comparison.

Table I: Signal-fusion Compared to Previous Methods

First Author	Method	d'	EER	FRR@FAR=0.001
Baseline	no fusion	4.37	1.64×10^{-2}	4.51×10^{-2}
Ma	score fusion: N-to-1 average	5.08	6.93×10^{-3}	1.40×10^{-2}
Krichen	score fusion: N-to-1 minimum	5.44	6.63×10^{-3}	1.43×10^{-2}
This work	signal fusion: 1-to-1	6.06	3.88×10^{-3}	7.61×10^{-3}

Schmid et al. [6] enrolled N gallery images of a particular subject and also took N images of a probe subject. The N gallery images and N probe images were paired in an arbitrary fashion and compared. Thus they obtained N different Hamming distance scores. Finally, they averaged the N scores into one single score. We will call this type of experiment an N -to- N -average comparison. Schmid also combined the N different Hamming scores using the log-likelihood ratio. We will call this type of experiment an N -to- N -loglikelihood comparison.

In our signal-fusion method, we take N frames from a gallery video and do signal-level fusion, averaging the images together to create one single average image. We then take N frames from a probe video and average them together to create a single average image. Thus, we can call our proposed method a signal fusion-1-to-1 comparison.

One automatic advantage of the signal fusion method is that storing a single, average-image iris code takes only a fraction of the space of the score-fusion methods. Instead of storing N gallery templates per subject, the proposed method only requires storing one gallery template per subject.

In order to compare our method to previous methods, we have implemented the N -to-1-average and the N -to-1-minimum methods. Implementing the N -to- N -average and N -to- N -loglikelihood methods remains future work. For our experiments, we let $N = 6$. For each of these methods, we used the same data sets. Figure 2 shows a decision error threshold curve with results from these experiments. As an additional baseline, we graph the DET curve for a single-gallery, single-probe experiment (No Fusion). The proposed signal fusion method has the lowest false accept and false reject rates for the majority of the decision error threshold curve. There is a small region on the lower right part of the graph where the score fusion average method performs about the same as the proposed method. In this region, the false reject rate is smaller than 2.6×10^{-3} . False reject rates that small are based on 2 or fewer of the 765 possible genuine comparisons in our data set. Thus, we would need a larger amount of data to decide on performance in that region of the graph. (Since we have 225675 impostor comparisons in our data set, we would also need more data if we wanted to determine performance for false accept rates less than 4.4×10^{-6} .)

We conclude that on our data set, the signal-fusion method generally performs better than the previously proposed N -to-1-average or N -to-1-minimum methods. In addition, the signal fusion takes $1/N^{th}$ of the storage and $1/N^{th}$ of the matching time.

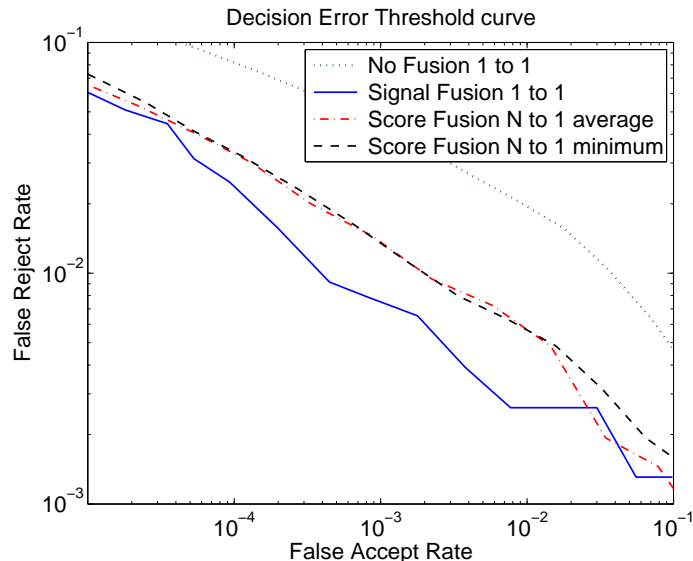


Fig. 2. The proposed signal-fusion method has better performance than using a multi-gallery approach with either an “average” or “minimum” score-fusion rule.

5.2 Comparing Signal-Fusion to Large Multi-Gallery, Multi-Probe Methods

The score-fusion experiments listed in subsection 5.1 compared our signal-fusion method to previously-published methods. Ma’s and Krichen’s methods both used multiple gallery images of a subject compared to a single probe image of a subject. Each of those score-fusion methods fused N Hamming distance scores to create the final score. We also wished to consider the situation where for a single comparison, there are N gallery images available, N probe images available, and therefore N^2 Hamming distance scores to fuse. We would expect that the fusion of N^2 scores would perform better than the fusion of N scores. Although this multi-gallery, multi-probe fusion is a simple extension of the methods listed in subsection 5.1, we do not know of any published work that uses this idea for iris recognition.

We tested two ideas: we took the average of all N^2 scores, and also the minimum of all N^2 scores. We call these two methods the (1) multi-gallery, multi-probe, average method (MGMP-average) and the (2) multi-gallery, multi-probe, minimum method (MGMP-minimum). The MGMP-average method produces genuine and impostor Hamming distance distributions with small standard deviations. Since this method has the lowest standard deviation for the impostor distribution, this method is good for reducing false accepts. Therefore, this particular method has the lowest FRR at 0.001 FAR. Using the “minimum” rule for score-fusion produces smaller Hamming distances than the “average” rule. However, both the genuine and impostor distributions are affected. We found that

Table II: Signal-fusion Compared to a Multi-gallery, Multi-probe Method

Method	d'	EER	FRR @ FAR=0.001
MGMP-average	5.53	2.98×10^{-3}	4.69×10^{-3}
Signal fusion	6.06	3.88×10^{-3}	7.61×10^{-3}

for these two multi-gallery, multi-probe methods, the “average” score-fusion rule works better than the “minimum” rule.

We next compared the MGMP-average method to the signal fusion method. The signal-fusion method presented in this subsection is unchanged from the previous subsection, but we are presenting the results again, for comparison purposes. The comparison between the MGMP-average method and the signal fusion method had mixed results. MGMP-average method had lower EER and a lower FRR at FAR=0.001. However, the signal-fusion method had a better value of d' , suggesting that the signal-fusion method does achieve good separation between the genuine and impostor distributions of scores [17]. Statistics for both methods are shown in table II.

Based on the equal error rate and false reject rate, we must conclude that the multi-gallery, multi-probe average method that we present in this section achieves the best recognition performance of all of the methods considered in this paper. However, the signal-fusion performs well, while taking only $1/N^{th}$ of the storage and $1/N^2$ of the matching time.

5.3 Computation Time

In this subsection, we compare the different methods presented in this paper in terms of processing time. We have three types of methods to compare: (1) the multi-gallery, multi-probe approaches (both MGMP-average and MGMP-minimum) which require N^2 iris code comparisons before fusing values together to create a single score; (2) the multi-gallery approaches (Ma and Krichen) which compare N gallery iris codes to one probe before fusing scores together; and (3) the signal-fusion approach which first fuses images together, and then has a single iris code comparison.

For this analysis, we first define the following variables. Let P be the preprocessing time for each image, I be the iris code creation time, and C be the time required for the XOR comparison of two iris codes. Let N be the number of images of a subject in a single gallery entry for the multi-gallery methods. Let A be the time required to average N images together (to perform signal-fusion). Finally, suppose we have an application such as in the United Arab Emirates where each person entering the country has his or her iris compared to a watchlist of one million people [18]. For this application, let W be the number of people on the watchlist. Expressions for the computation times for all three methods are given in terms of these variables in Table III.

From Daugman’s work [17], we can see that typical preprocessing time for an image is 344 ms. He also notes that iris code creation takes 102 ms and

Table III: Processing times for different methods

Method	Gallery Preprocessing	Probe Preprocessing	Comparison to Watchlist	Total Time
MGMP	NP+NI = 2.676 s	NP+NI = 2.676 s	WCN ² = 360 s	365.35 s
Multi-gallery	NP+NI = 2.676 s	P+I = 0.446 s	WCN = 60 s	63.12 s
Signal fusion	NP+A+I = 2.169 s	NP+A+I = 2.169 s	WC = 10 s	14.34 s

an XOR comparison of two iris codes takes 10 μ s. Throughout this paper, we have used six images for all multi-gallery experiments. The time to compute an average image from six preprocessed images is 3 ms. Lastly, we know that the United Arab Emirates watchlist contains one million people. By substituting these numbers in for our variables, we found the processing time for all of our three types of methods. These numeric values are also presented in table III.

From this analysis it is clear, that although a multi-gallery, multi-probe method may have some performance improvements over the signal fusion method, it comes at a high computational cost.

6 Conclusions

We perform fusion of multiple biometric samples at the signal level. Our signal fusion approach utilizes information from multiple frames in a video. This is the first published work to use video to improve iris recognition performance. Our experiments show that using average images created from six frames of an iris video performs very well for iris recognition. Average images perform better than (1) experiments with single stills and (2) experiments with six gallery images compared to single stills. Our proposed multi-gallery, multi-probe method achieves slightly better recognition performance than our proposed signal-fusion method. However, the matching time and memory requirements are lowest for the signal-fusion method, and the signal-fusion method still performs better than previously published multi-gallery methods.

Acknowledgement

This research is supported by the National Science Foundation under grant CNS01-30839, by the Central Intelligence Agency, by the Intelligence Advanced Research Projects Activity, and by the Technical Support Working Group under US Army contract W91CRB-08-C-0093. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of our sponsors.

References

1. Kevin W. Bowyer, Karen P. Hollingsworth, and Patrick J. Flynn. Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding*, 110(2):281–307, 2008.
2. Wenyi Zhao and Rama Chellapa, editors. *Face Processing: Advanced Modeling and Methods*, chapter 17: Beyond one still image: Face recognition from multiple still images or a video sequence by S.K. Zhou and R. Chellapa, pages 547–567. Elsevier, 2006.
3. Yingzi Du. Using 2D log-Gabor spatial filters for iris recognition. In *SPIE 6202: Biometric Technology for Human Identification III*, pages 62020:F1–F8, 2006.
4. Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. Efficient iris recognition by characterizing key local variations. *IEEE Transactions on Image Processing*, 13(6):739–750, Jun 2004.
5. Emine Krichen, Lorène Allano, Sonia Garcia-Salicetti, and Bernadette Dorizzi. Specific texture analysis for iris recognition. In *Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA 2005)*, pages 23–30, 2005.
6. Natalia A. Schmid, Manasi V. Ketkar, Harshinder Singh, and Bojan Cukic. Performance analysis of iris-based identification system at the matching score level. *IEEE Transactions on Information Forensics and Security*, 1(2):154–168, Jun 2006.
7. Deborah Thomas, Kevin W. Bowyer, and Patrick J. Flynn. Strategies for improving face recognition from video. *Workshop on Motion and Video Computing*, Feb 2007.
8. Chengqiang Liu and Mei Xie. Iris recognition based on DLDA. In *Int. Conf. on Pattern Recognition*, pages 489–492, Aug 2006.
9. Kaushik Roy and Prabir Bhattacharya. Iris recognition with support vector machines. In *Proc. Int. Conf. on Biometrics*, pages 486–492, Jan 2006.
10. P. Jonathon Phillips, Todd Scruggs, Patrick J. Flynn, Kevin W. Bowyer, Ross Beveridge, Geoff Givens, Bruce Draper, and Alice O’Toole. Overview of the multiple biometric grand challenge. In *Proc. Int. Conf. on Biometrics (ICB2009)*, 2009.
11. Xiaomei Liu, Kevin W. Bowyer, and Patrick J. Flynn. Experiments with an improved iris segmentation algorithm. In *Proc. Fourth IEEE Workshop on Automatic Identification Technologies*, pages 118–123, Oct 2005.
12. Karen P. Hollingsworth, Kevin W. Bowyer, and Patrick J. Flynn. All iris code bits are not created equal. In *Proc. IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2007.
13. Karen P. Hollingsworth, Kevin W. Bowyer, and Patrick J. Flynn. The best bits in an iris code. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. accepted for publication.
14. Kevin W. Bowyer, Kyong I. Chang, Ping Yan, Patrick J. Flynn, Earnie Hansley, and Sudeep Sarkar. Multi-modal biometrics: an overview. In *Second Workshop on Multi-Modal User Authentication*, May 2006.
15. Kyong I. Chang, Kevin W. Bowyer, and Patrick J. Flynn. An evaluation of multi-modal 2D+3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, Apr 2005.
16. P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, and William Worek. Preliminary Face Recognition Grand Challenge results. In *Int. Conf. on Automatic Face and Gesture Recognition (FG 2006)*, Apr 2006.
17. John Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.
18. John Daugman. United Arab Emirates deployment of iris recognition, <http://www.cl.cam.ac.uk/~jgd1000/deployments.html>, accessed Jan 2009.