Iris Recognition using Signal-level Fusion of Frames from Video

Karen Hollingsworth, Tanya Peters,
Kevin W. Bowyer, *Fellow, IEEE,* and Patrick J. Flynn, *Senior Member, IEEE*

Abstract—No published prior work has shown any advantage of the use of video over still images. We take advantage of the temporal continuity in an iris video to improve matching performance using signal-level fusion. From multiple frames of a frontal iris video, we create a single average image. For comparison, we re-implement three score-level fusion methods (Ma et al., Krichen et al., and Schmid et al.). We find that our signal-level fusion of N images performs better than Ma's or Krichen's score-level fusion methods of N Hamming distance scores. Our signal-level fusion performs comparably to Schmid's log-likelihood method of score-level fusion, and our method achieves this performance using less computation time. We compare our signal fusion method with another new method: a multi-gallery, multi-probe method involving score-level fusion of N^2 Hamming distances. The multi-gallery, multi-probe score fusion has slightly better recognition performance, while the signal fusion has significant advantages in memory and computation requirements.

Index Terms—Iris biometrics, iris code, image averaging, noise reduction, signal-level fusion, score-level fusion, iris video.

I. INTRODUCTION

THE field of iris recognition is an active and rapidly expanding area of research [2]. Many researchers are interested in making iris recognition more flexible, faster, and more reliable. Despite the vast amount of recent research in iris biometrics, past published work has relied mainly on still iris images. Zhou and Chellappa [3] 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, [4]–[8]. However, no researchers have published techniques focusing on the use of additional information available in iris video.

There are drawbacks to using single still images. 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

Authors are with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, 46556.
E-mail: kholling,tpeters,kwb,flynn at cse.nd.edu

iris image. A second difficulty with still images is that lighting differences can cause an increased Hamming distance score in a comparison between two stills. By combining information from multiple frames of a video, we can reduce variations caused by changes in lighting.

Zhou and Chellappa 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. In this paper, we present a method of averaging frames from an iris video. Our experiments demonstrate that that our signal-level fusion of multiple frames in an iris video can improve iris recognition performance.

We perform image fusion of iris images at the pixel level. Our experiments show that the traditional segmentation and unwrapping of the iris can be used as a satisfactory method of image registration. We compare two methods of pixel fusion: using the mean and using the median.

There have been a number of papers discussing score-level fusion for iris recognition, but there has not been any work done with signal-level fusion for iris recognition. Since we are the first to propose the use of signal-level fusion for iris recognition, we show that this type of fusion can perform comparably to score-level fusion. We focus on reimplementing multiple score-level fusion techniques to show that signal-level fusion can achieve at least as good recognition rates as score-level fusion. Our experiments show that our method achieves superior recognition rates to some score-level fusion techniques suggested in the literature. Additionally, our signal-fusion method has a faster computation time for matching than the score-level fusion methods.

The fusion method proposed in this paper involves a pixel-by-pixel average. This method has the advantage of being simple, but can come at the expense of reduced contrast. There are a number of other possible methods for performing image fusion [9]. Such methods have potential to yield further performance improvements, although such improvements would come at a cost of increased computational complexity. An in-depth comparison of these other ideas could easily be the topic of another full paper and would be a good topic of future research. For brevity, we focus this paper on comparing a pixel-level average image fusion method to various score-fusion methods.

II. RELATED WORK

A. Video

Video has been used effectively to improve face recognition. A recent book chapter by Zhou and Chellappa [3] surveyed a number of methods to employ video in face biometrics. In contrast, there is very little research using video in iris biometrics. In an effort to encourage research in iris biometrics using unconstrained video, the U.S. government organized the Multiple Biometric Grand Challenge [12]. The data provided with this challenge included two types of near infrared iris videos: (1) iris videos captured using an LG 2200 camera, and (2) videos containing iris and face information captured using a Sarnoff Iris on the Move portal [13].

There has been a small amount of work published using the MBGC data. First, some preliminary results were presented at a workshop [14]. In addition, two conference papers using MBGC iris videos were published in the most recent International Conference in Biometrics. The first paper was our initial version of this research [1]. The second paper by Lee et al. [15] presented methods to detect eyes in the MBGC portal videos and measure the quality of the extracted eye images. They compared portal iris videos to still images. At a false accept rate of 0.80%, they achieved a false reject rate of 43.90%.

A recent journal paper by Zhou et al. [16] also presented some results on the MBGC iris video data. Zhou et al. suggested making some additions to the traditional iris system in order to select the best frames from video. First they checked each frame for interlacing, blink, and blur. They used interpolation to correct deinterlacing, and they discarded blurry frames and frames without an eye. Selected frames were segmented in a traditional manner and then assigned a confidence score relating to the quality of the segmentation. They further evaluated quality by looking at the variation in iris texture, the amount of occlusion, and the amount of dilation. They divided the iris videos into five groups based on quality score, and showed that a higher quality score correlated with lower equal error rate.[1]

Our work differs from Lee's [15] and Zhou's [16] in that we use videos for both gallery and probe sets. Also, we compare the use of stills and the use of videos directly, while theirs do not. In addition, their papers focus on selecting the best frame from a video to use for subsequent processing. In contrast, the main focus of this work is to how to combine information from multiple frames using signal-level fusion.

B. 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.

---

[1]Lee et al. [15] and Zhou et al. [16] both investigate quality of video frames. A number of papers have investigated quality of still images including Vatsa et al. [17], Belcher and Du [18], and Proença and Alexandre [19].

Du [4] showed that using three enrollment images instead of one increased their rank-one recognition rate from 98.5% to 99.8%. The paper reported, "We randomly choose three images [of] each eye from the database to enroll and used the rest [of the] images to test. We did [this] multiple times and the average identification [accuracy] rate is 99.8%. If two images are randomly selected to enroll, ... the average identification accuracy rate is 99.5%. If one image is randomly selected to enroll ... the average identification accuracy is 98.5%." In another paper [5], Du et al. used four enrollment images instead of three.

Ma et al. [6] 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. [7] performed a similar experiment, but used the minimum match score instead of the average. Schmid et al. [8] presented two methods for fusing Hamming distance scores. They computed average Hamming distance and also a log-likelihood ratio. They found that in many cases, the log-likelihood ratio outperformed 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 used score-level fusion.

Another method of using multiple iris images is to use them to train a classifier. Liu et al. [20] used multiple iris images for a linear discriminant analysis algorithm. Roy and Bhattacharya [21] used six images of each iris class to train a support vector machine. Even in training these classifiers, each still image was treated as an individual entity, rather than being combined with other still images to produce an improved image.

III. DATA

We used the Multiple Biometric Grand Challenge (MBGC) version 2 iris video data [12] in our experiments. The videos in this data set were acquired using an Iridian LG EOU 2200 camera (Figure 1). To collect iris videos using the LG2200 camera, the analog NTSC video signal from the camera was digitized using a Daystar XLR8 USB digitizer and the resulting videos were stored in a high bit rate (nearly lossless) compressed MP4 format.

The MBGCv2 data contains 986 iris videos collected during the spring of 2008. However, three of the videos in the data set contain less than ten frames. We dropped those three videos from our experiments and used the remaining 983 videos. The data includes videos of both left and right eyes for each subject; we treated each individual eye as a separate "subject" in our experiments. There are a total of 268 different eyes in these videos. We selected the first video from each subject to include in the gallery set and put the remaining 715 videos in our probe set. For each subject, there were between one and seven iris videos in the data set. Any two videos from the same subject were acquired between one week and three months apart. The MBGC data is the only set of iris videos publicly available.

Fig. 1. The Iridian LG EOU 2200 camera used to acquire iris video sequences.

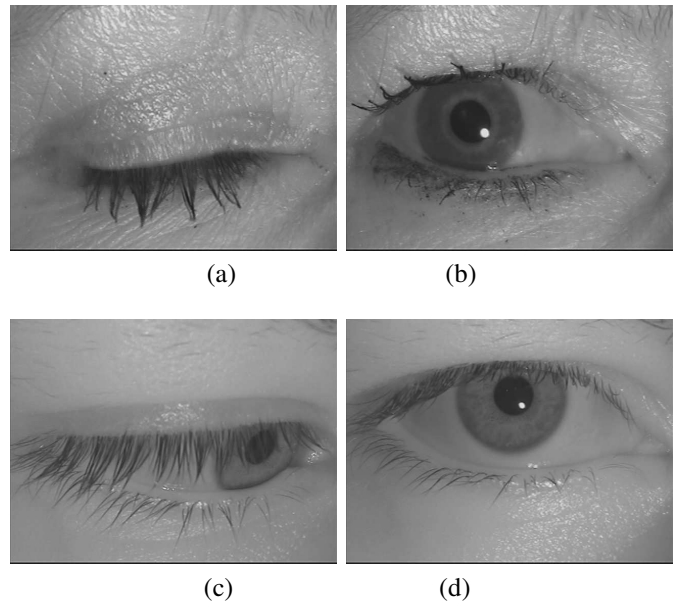(a)   (b)

(c)   (d)

Fig. 2. The frames shown in (a) and (c) were selected by our frame-selection algorithm because the frames were in focus; however, these frames do not include much valid iris data. In our *automated* experiments presented in this paper we kept frames like (a) and (c) so that we could show how our software performed without any manual quality checking. In our *semi-automated* experiments we manually replaced frames like (a) and (c) with better frames from the same video like (b) and (d). We expect that in the future, we may be able to develop an algorithm to detect blinks and off-angle images so that such frames could be automatically rejected.

IV. AVERAGE IMAGES AND TEMPLATES

A. Selecting Frames and Preprocessing

Once each iris video was acquired, we wanted to create a single average image that combined iris texture from multiple frames. The first challenge was to select focused frames from the iris video. The auto-focus on the LG 2200 camera continually adjusts the focus in attempts to find the best view of the iris. Some frames have good focus, while others suffer from severe blurring due to subject motion or illumination change.

We used a technique described by Daugman with a filter proposed by Kang to select in-focus images. As described by Daugman in [22], a filter can be applied to an image as a fast focus measure, typically in the Fourier domain. By exploiting Parseval's Theorem, we were instead able to apply the filter within the image domain, squaring the response at each pixel. We summed the responses over the entire image, applying the filter to non-overlapping pixels within the image and then averaged the response over the number of pixels the kernel was applied to. The kernel described by Kang and Park [23] was applied to each frame, and the ten with the highest scores were extracted from the video for use in the image averaging experiments.

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. Differences between the stills and the video frames are likely due to differences in the digitizers used to save the signals. We used the Matlab imadjust function [24] 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 [25]. Our segmentation algorithm is designed to work for frontal iris images acquired from co-operative subjects. A possible area of future work would be to obtain a segmentation algorithm that could work on off-angle irises and test our image-averaging technique on that type of iris images.

Our segmentation and eyelid detection algorithms are not as finely tuned as commercial iris recognition software. To make up for this limitation, we ran two types of experiments for this paper. The first type of experiments uses the data obtained from the completely automated frame selection, segmentation, and eyelid detection algorithms. We also ran a second set of experiments that included manual steps in the preprocessing. We manually checked all 9830 frames selected by our frame-selection algorithm. A few of the frames did not contain valid iris information; for example, some frames showed blinks. We also found some off-angle iris frames. We replaced these frames with other frames from the same video (Figure 2). In total, we replaced 86 (0.9%) of the 9830 frames. Next we manually checked all of the segmentation results and replaced 153 (1.6%) incorrect segmentations (Figure 3). We corrected the eyelid detection in an additional 1765 (18%) frames (Figure 4).

B. Signal Fusion

For each video, we now had ten frames selected and segmented. We wanted to create an average image consisting only of iris texture. In order to align the irises in the ten
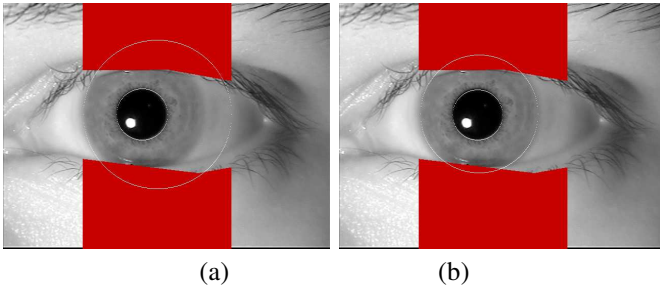
Fig. 3. Our *automated* experiments contain a few incorrect segmentations like the one shown in (a). In our *semi-automated* experiments we manually replaced incorrect segmentations to obtain results like that shown in (b).
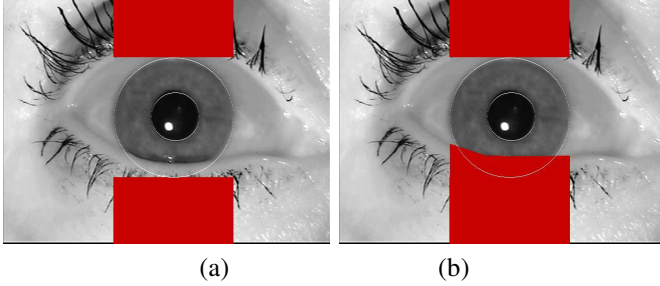

Fig. 4. Our *automated* software did not correctly detect the eyelid in all frames. In our *semi-automated* experiments we manually replaced incorrect eyelid detections to obtain results like that shown in (b).

Fig. 5. From the ten original images on the top, we created the average image shown on the bottom.

frames, we transformed the raw pixel coordinates of the iris area in each frame into normalized polar coordinates. In 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 ten 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 and thus assumed that no shifts were needed. The first two approaches did not produce any better recognition results than the naive approach. This is because the images used in our experiments are frontal iris images from cooperative users. A different method of alignment would be necessary for iris videos with more eye movement. Since the naive approach worked well for our data, we used it in our subsequent experiments.

Parts of the unwrapped images contained occlusion by eyelids and eyelashes. We masked eyelid regions in our image. Then we computed an average unwrapped image from unmasked iris data in the ten original images, using the following algorithm. For each (r, θ) position, we find how many of the corresponding pixels in the ten unwrapped images are unmasked. If a pixel is occluded in nine or ten of the images, we mask it in the average image. Otherwise, an average pixel
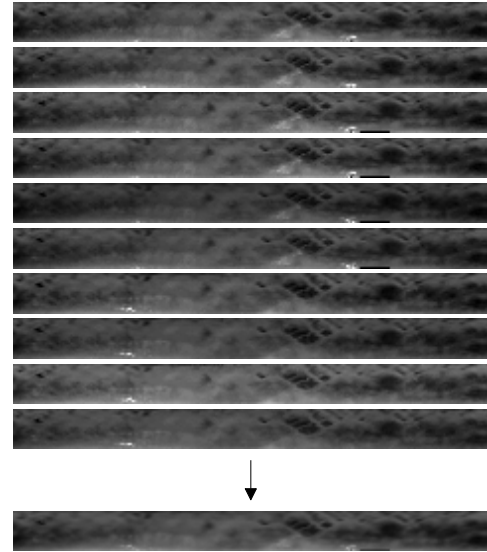
value is based on unmasked pixel values of the corresponding frames. (Therefore, the new pixel value could be an average of between two and ten pixel intensities, depending on mask values.) Section V will give more details on averaging the pixel values.

Using this method, we obtained 268 average images from the gallery videos. We similarly obtained 715 average images from the probe videos. An example average image is shown in Figure 5. On the top of the figure are the ten original images, and on the bottom is the average image fused from the original signals.

C. 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", "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 [26], [27] to mask out complex coefficients near the axes, and therefore improve recognition performance.

V. COMPARISON OF MEDIAN AND MEAN FOR SIGNAL FUSION

Using the basic strategy described in IV-B and IV-C, we needed to determine the best method of averaging pixels. Recall that each (r, θ) position in the new average image is the average of corresponding, unoccluded pixels in the ten original unwrapped iris images. We considered two ideas:

using a median rule to combine the pixel values, or using a mean rule.[2]

To determine which of these two methods was most appropriate for iris recognition, we compared all images in our probe set to all images in our gallery and graphed a detection error tradeoff (DET) curve [28]. Figure 6 shows the result. It is clear from the graphs that using the mean rule for creating the average images produces better recognition performance than using the median rule.

The median is a useful statistic for removing outliers. However, it is possible that many of the extreme outliers in these iris images have already been removed by eyelid detection. Furthermore, since we are averaging only a small number of pixels (ten or fewer), the median statistic may be less useful that if we had more available data. While the median statistic uses information from only one or two pixels, the mean statistic involves information from all available pixels. Therefore, in this context, the mean rule is a better averaging rule than the median.

VI. How Many Frames Should Be Fused In An Average Image?

As described in subsection IV-B, we fuse ten frames together to create an average image. However, ten frames may not be the optimal number of frames to use. Fusing more frames can give a better average. On the other hand, we add the best focused frames first, so as we increase the number of frames, we are fusing poorer quality data. To investigate this trade-off, we ran an experiment varying the number of frames to use in the fusion.

Recall that from each video, we had frames selected, segmented, and unwrapped into normalized polar coordinates. For this experiment, rather than using all ten selected frames to create an average image, we selected the four frames having the highest focus scores and we created an average image. In this manner, we collected a gallery set of four-frame average images, and a probe set of four-frame average images. We compared all gallery images to all probe images and graphed the corresponding DET curve (red dash-dot line, Figure 7).

We repeated this procedure, this time using six of our selected frames to create each average image. The set of six frames from each video was a superset of the set of four frames. We created a gallery set of six-frame average images, and a probe set of six-frame average images, tried all comparisons, and graphed the DET curve on the same axes as the four-frame curve (green solid line, Figure 7).

We repeated the same procedure three more times, using eight, nine, and ten frames. All DET curves are shown together in Figure 7.

With the automated segmentation, each increase in the number of frames fused yielded an increase in performance. With the manually corrected segmentation, this trend holds for four, six, and eight frames. However, the DET curves for eight, nine, and ten frames all overlap, suggesting that we

have approached the limit of the benefit that can be gained by adding frames.

In a previous paper [1], we used six frames instead of ten, but in that paper, we had a different data set and different frame selection algorithm. The data set in our previous paper was a pre-release version of the MBGCv2 videos. 617 of those videos were included in MBGCv2 and we also had an additional 444 iris videos captured during the same semester that were not included in MBGCv2.

In our previous paper [1], we chose to use the same frames as were selected by the special Iridian software that came with the camera. That frame selection technique picked two frames captured while the top camera LED was lit, two frames captured while the right LED was lit, and two frames captured while the left LED was lit. Therefore that technique guaranteed some lighting differences between the frames selected. Our current frame selection technique does not enforce such a requirement, so the ten frames selected using our current method may have fewer variations between them. With fewer variations between the frames, it makes sense that we could average more frames before losing any important texture in the iris.

We imagine that the optimal number of frames to fuse in creating an average image depends both on the data set and on the frame selection algorithm. For this paper, we decided to use ten frames in creating our average images. Using ten frames gave the best performance using the automated segmentation. The choice between using eight, nine, or ten frames for the manually corrected segmentation was not as clear, but ten frames still gave the best equal error rate, and gave reasonable performance across the whole DET curve.

VII. How Much Masking Should Be Used In An Average Image?

We initially allowed a pixel to be unmasked in the average image if at least two corresponding pixels from the ten frames were unmasked. However, we suspected that a different masking rule could improve performance. We could require that all unmasked pixels in an average image be an average of *ten* unmasked pixel values from the ten frames (instead of an average of at least two pixels). This requirement could result in average images with not much available unmasked data. If any one frame had a large amount of occlusion, the average image would be heavily masked. On the other hand, we could use any unmasked pixel values from the frames in creating the average image, so that an average pixel value could be an average of between one and ten pixel intensities from the frames, depending on mask values in the frames.

We defined a parameter, the *masking level*, to specify how much masking is done in an average image. A masking level of 100% means that we only have unmasked pixels in the average image if *all ten* of the corresponding pixels from our ten frames were unmasked. A masking level of 10% means that the new pixel value could be an average of *between one and ten* pixel intensities, depending on mask values. A masking level of 50% means that we require *at least half* of the corresponding pixels to be unmasked before we compute an

---

[2]To compute the mean, we first summed original pixel values, then divided by the number of pixels, then rounded to the nearest unsigned 8-bit integer.
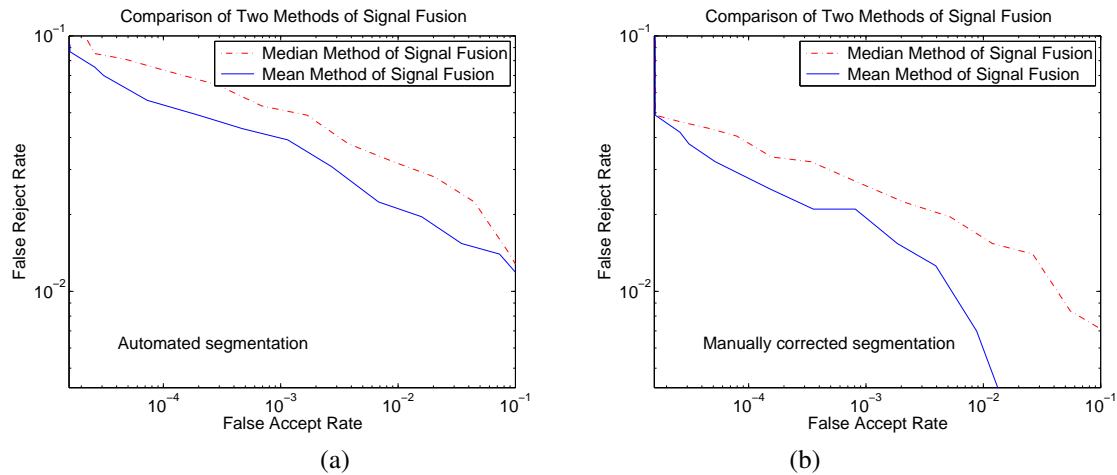
Fig. 6. Using a *mean* fusion rule for fusing iris images produces better iris recognition performance than using a *median* fusion rule. Graph (a) shows this result using automated segmentation. Graph (b) shows the same result using the manually corrected segmentations.
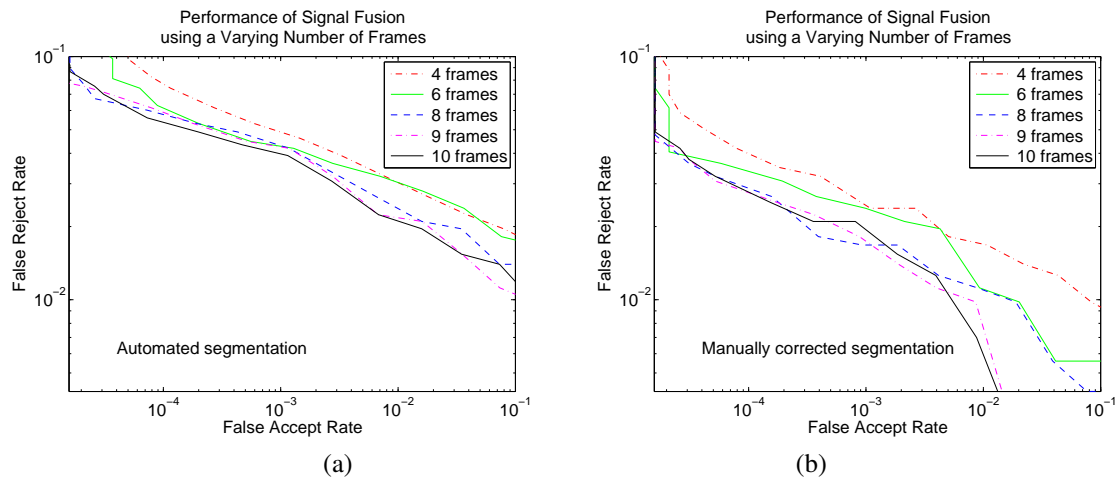
Fig. 7. Fusing ten frames together yields better recognition performance than fusing four, six, or eight frames.

average and create an unmasked pixel in the average image. At this level, the new pixel value could be an average of between five and ten pixel intensities, depending on mask values.

When we mask too much, we do not have as much iris data in our images from which to make appropriate decisions. With less iris data, and consequently fewer unmasked bits in a comparison, we get fewer degrees of freedom in the nonmatch distribution. To illustrate this phenomenon, we graphed the nonmatch distribution for a range of masking levels (Figure 8). As the masking level increased, the histogram of nonmatch scores got wider, causing an increased false accept rate. In contrast, when we mask too little, we lose the power gained from combining data from a number of different images. The result would be like using too few gallery images in a multi-gallery biometrics experiment.

The optimal masking level depends partly on the quality of the segmentation. We created DET curves showing the verification performance as we varied the masking level used in creating the average images (Figure 9). With our automated segmentation, a higher masking parameter is better to mitigate the impact of segmentation errors. With the manually corrected segmentations, the quality of the segmentation is good enough
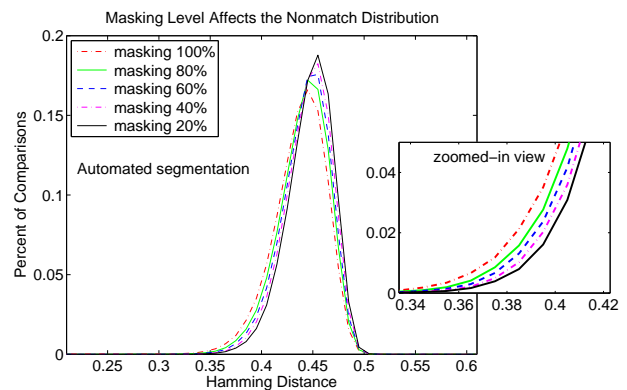
Fig. 8. Too much masking decreases the degrees of freedom in the nonmatch distribution, causing an increased false accept rate. (This graph shows the trend from the automatically segmented images. The manually corrected segmentation produces the same trend.)

for us to use a smaller masking parameter and thus avoid as large an increase in false accept rate. For our current data set and segmentation, we chose to use a masking level of 80% for the automated segmentation experiments, and a masking level
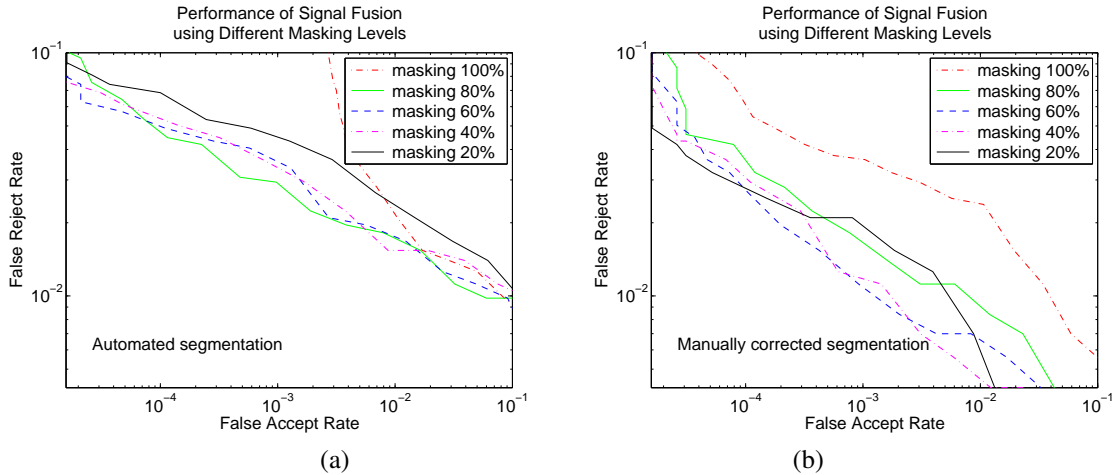
Fig. 9. The amount of masking used to create average images affects performance. When using the manually corrected segmentation, we can use a smaller masking level (masking level = 60%). With the automated segmentation, a higher masking level (masking level = 80%) mitigates the impact of missed eyelid detections.

of 60% when using the manually corrected segmentation.

VIII. COMPARISON TO OTHER METHODS

We now present experiments comparing our method to previous methods. We compare our signal-fusion method to the multi-gallery score-fusion methods described by Ma [6] and Krichen [7]. Then we compare signal-fusion to Schmid's log-likelihood method [8]. Our last experiment compares signal-fusion to a new multi-gallery, multi-probe score-fusion method.

A. Comparison to Previous Multi-gallery Methods

In biometrics, it has been found that enrolling multiple images improves performance [29]–[31]. Iris recognition is no exception. Many researchers [6]–[8] 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. [6] took the average Hamming distance. We will call this type of experiment an N-to-1-average comparison. Krichen et al. [7] 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.

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

TABLE I
SIGNAL-FUSION COMPARED TO PREVIOUS METHODS

Method	d'	EER	FRR@FAR=0.001
no fusion	4.62	1.56×10^{-2}	3.32×10^{-2}
score fusion: N-to-1 avg	5.02	8.62×10^{-3}	1.90×10^{-2}
score fusion: N-to-1 min	5.49	7.55×10^{-3}	1.36×10^{-2}
signal fusion: 1-to-1	6.06	6.99×10^{-3}	1.10×10^{-2}

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 N-to-1-minimum methods. For our experiments, we let N = 10. For each of these methods, we used the same data sets. Table I shows statistics for from these experiments for the manually corrected segmentation, and Figure 10 shows the detection error tradeoff curves. As an additional baseline, we graph the DET curve for a single-gallery, single-probe experiment (No Fusion). The DET curve shows that the proposed signal fusion method has the lowest false accept and false reject rates of all methods shown here.

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.

B. Comparison to Previous Log-Likelihood Method

Schmid et al. [8] 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. They combined the N different Hamming scores using the log-likelihood ratio.

We give a brief summary of the log-likelihood method here. A more detailed description can be found in [8]. Let $X_1, X_2, ..., X_N$ be a sequence of N iriscodes representing a
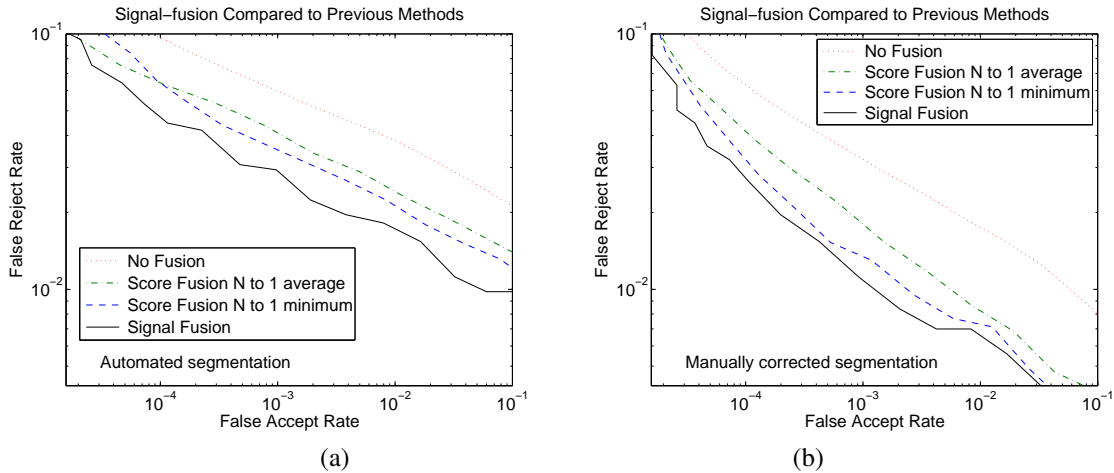
Fig. 10. The proposed signal-fusion method has better performance than using a multi-gallery approach with either an "average" or "minimum" score-fusion rule.

single subject in the gallery. Let $Y_1, Y_2, ..., Y_N$ be a sequence of N iriscodes representing a single subject as a probe. Let $d = [d_1, d_2, ..., d_N]$ be a vector of N Hamming distances formed from these two iriscode sequences. The impostor hypothesis H_0 states that the vector d is Gaussian distributed with common unknown mean for all entries m_0, and unknown covariance matrix C_0. The genuine hypothesis, H_1 states that the vector d is Gaussian distributed with a common unknown mean m_1 and unknown covariance matrix C_1. Denote $p(d|H_i)$ the conditional probability density function for the vector d under hypothesis H_i. The log-likelihood ratio test statistic is

$$l_N = (1/N)\log[p(d|H_1)/p(d|H_0)]. \qquad (1)$$

The statistic l_N can be computed as a function of $m_0, m_1, C_0, C_1, d,$ and N. The values m_0, m_1, C_0, C_1 are obtained using training data, and a vector of Hamming distances d is obtained using testing data. Fractional Hamming distance scores are bounded between zero and one, but log-likelihood test statistics have a wider range. In our experiments we obtained scores between -1.99 to 44.60. Low scores are from impostor comparisons and high scores are from genuine comparisons.

The log-likelihood method requires both training and testing data, so we split our gallery and our probe each in half. We used the first half of the gallery videos (gallery-set-A) and the first half of the probe videos (probe-set-A) for training and obtained a set of maximum-likelihood parameters. Next we compared the second half of the gallery videos (gallery-set-B) and the second half of the probe videos (probe-set-B); applying the the maximum-likelihood parameters to the resulting Hamming distance vectors gave us log-likelihood scores from the test data B.

Of course, it would be better to have as many scores as possible from our data, so we repeated the experiment, this time using set B to train the maximum-likelihood parameters and set A to test. We obtained log-likelihood scores from test data A. We combined all log-likelihood scores and created a DET curve representing the performance of the log-likelihood method.

TABLE II
SIGNAL-FUSION COMPARED TO LOG-LIKELIHOOD METHOD

Method	d'	EER	FRR@FAR=0.001
log-likelihood	3.90	2.65×10^{-3}	9.20×10^{-3}
signal fusion	6.06	6.99×10^{-3}	1.10×10^{-2}

The curves showing performance of the log-likelihood method in comparison with the signal fusion method are shown in Figure 11, and corresponding statistics for the manually corrected segmentations are in Table II. The log-likelihood method has a lower equal error rate, but the signal fusion method performs better at smaller false accept rates. In addition, the signal fusion takes $1/N^{th}$ of the storage and $1/N^{th}$ of the matching time.

C. Comparing Signal-Fusion to Large Multi-Gallery, Multi-Probe Methods

The previous subsections compared our signal-fusion method to previously-published methods. 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 and N probe images available, and all N^2 possible Hamming distance scores are computed and fused. 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 VIII-A, 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 impostor Hamming distance distributions with small standard deviations. Using the "minimum" rule for score-fusion produces smaller Hamming distances
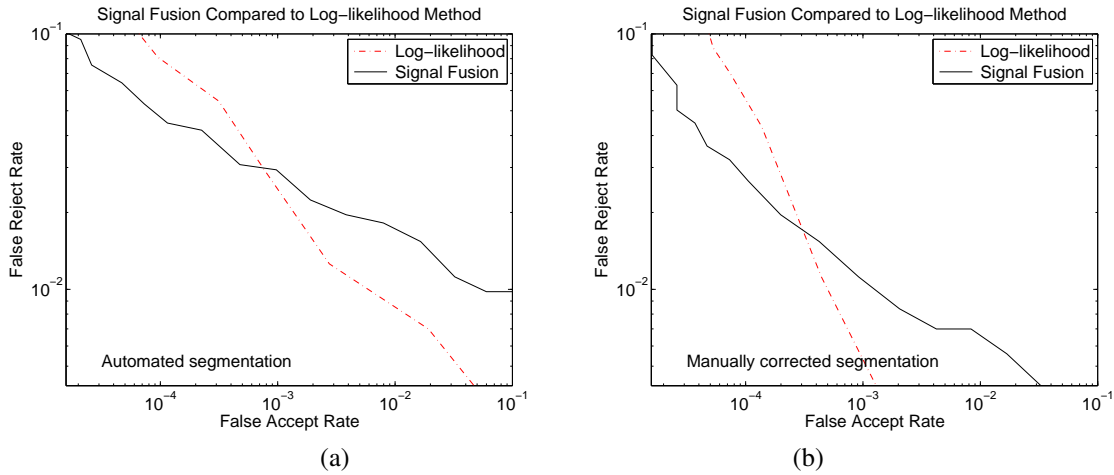
Fig. 11. Signal fusion and log-likelihood score fusion methods perform comparably. The log-likelihood method performs better at operating points with a large false accept rate. The proposed signal-fusion method has better performance at operating points with a small false accept rate.

TABLE III
SIGNAL-FUSION COMPARED TO A MULTI-GALLERY, MULTI-PROBE
METHOD

Method	d'	EER	FRR @ FAR=0.001
MGMP-average	5.32	5.47×10^{-3}	1.17×10^{-2}
MGMP-minimum	6.51	1.60×10^{-3}	3.08×10^{-3}
signal fusion	6.06	6.99×10^{-3}	1.10×10^{-2}

than the "average" rule. However, both the genuine and impostor distributions are affected. Based on the DET curves (Figure 12), we found that for these two multi-gallery, multi-probe methods, the "minimum" score-fusion rule works better than the "average" rule for this data set.

We compared the MGMP methods 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. Statistics for the signal fusion and the MGMP methods are shown in Table III. The error rates for signal fusion in Table I and Table III are the same because we are running the same algorithm on the same data set.

Based on the equal error rate and false reject rate, we conclude that the multi-gallery, multi-probe minimum 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.

D. 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 [32]. 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 IV.

The multi-gallery, multi-probe methods must do preprocessing and iris code creation for N images to create one gallery entry. Thus, the gallery preprocessing time for one gallery subject is NP+NI. They also preprocess and create N iris codes for a probe subject, so the probe preprocessing time is also NP+NI. To compare a single probe entry to a single gallery entry takes CN^2 time because there are N^2 comparisons to be done. To compare a probe to the entire watchlist takes WCN^2 time. Similar logic can be used to find expressions for the time taken for the other two methods. All such expressions are presented in Table IV.

From Daugman's work [22], we can see that typical preprocessing time for an image is 344 ms. He also notes that iris code creation takes 102 ms and an XOR comparison of two iris codes takes 10 μs. Throughout this paper, we have used ten images for all multi-gallery experiments. The time to compute an average image from ten preprocessed images is 5 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 IV.

A graph of the total computation time for these methods over a number of different sizes of watchlist is shown in
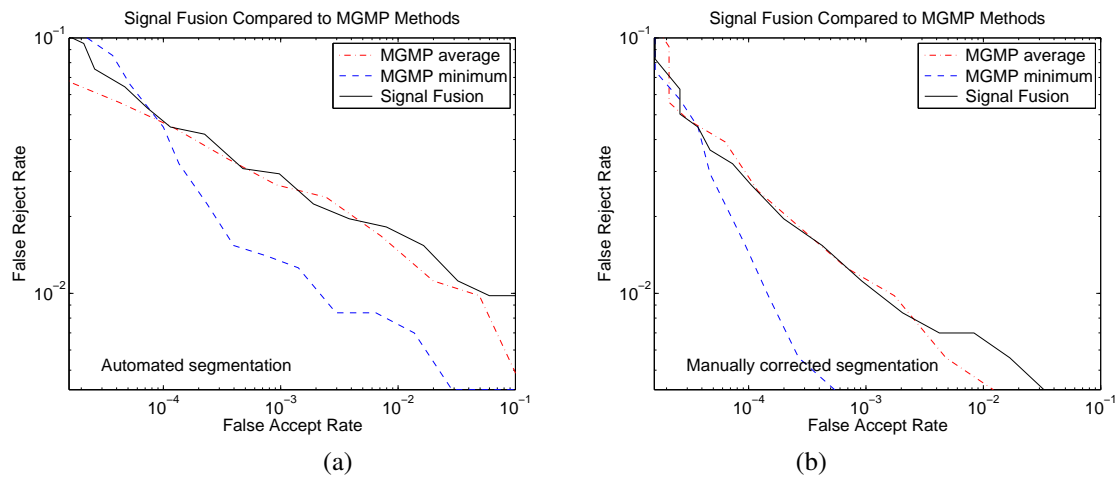
Fig. 12. The MGMP-minimum achieves the best recognition performance of all of the methods considered in this paper. However, the signal-fusion performs well, while taking only $1/N^{th}$ of the storage and $1/N^2$ of the matching time.

TABLE IV
PROCESSING TIMES FOR DIFFERENT METHODS

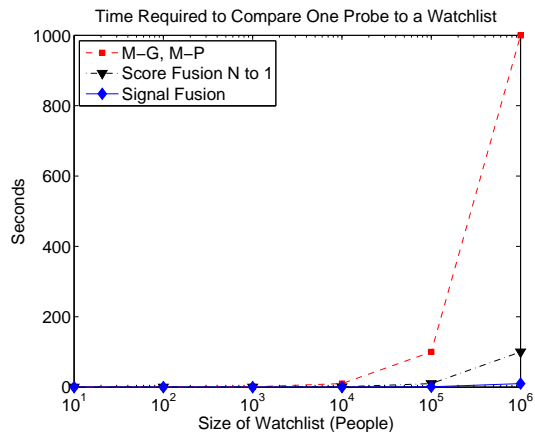| Method	Gallery Preprocessing	Probe Preprocessing	Comparison to Watchlist	Total Time
MGMP	NP+NI = 4.46 s	NP+NI = 4.46 s	WCN^2 = 1000 s	1008.9 s
Multi-gallery	NP+NI = 4.46 s	P+I = 0.446 s	WCN = 100 s	104.9 s
signal fusion	NP+A+I = 3.55 s	NP+A+I = 3.55 s	WC = 10 s	17.09 s

Fig. 13. Even though a large multi-gallery, multi-probe experiment achieves better recognition performance, it comes at a cost of much slower execution time. The proposed signal fusion method is the fastest method presented in this paper, and it achieves better recognition performance than previously published multi-gallery methods.

Figure 13. 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.

IX. FUTURE WORK

One recent area of interest in iris biometrics is performance on less-cooperative data. Researchers have collected data simulating less-cooperative acquisitions environments. As an example, the UBIRIS database was collected using methods aimed "to minimize the requirement of user cooperation" [33].

The method proposed in this paper was designed to be applied to video. Unfortunately, there are no less-cooperative iris video data publicly available yet. The portal data from MBGC may be termed "less-cooperative video data"; however, those videos typically have fewer than twenty frames and contain only one or two images of sufficient quality for iris matching.

One possible area of future work could be to obtain some lower-quality iris videos and apply image averaging to such videos. Lower-quality data may require some changes to the current technique. In our current technique, we fuse ten focused frames to create an average image. If the only frames available have poor focus, we might need to combine fewer frames, to preserve all available texture. We could design a system that automatically adjusted the number of frames fused based on the focus scores.

Videos of less-cooperative subjects may not have any frontal iris images. In such a situation, we could model the boundaries of the iris as an ellipse and apply an off-axis gaze correction technique like the method proposed by Schuckers et al. [34]. Whether image averaging would work on gaze-corrected images is still an open question.

Poorer data might also necessitate a different method of aligning unwrapped images. With our current data, aligning images using Hamming distance or correlation did not improve performance, but with more challenging data, a complex alignment approach could be beneficial.

X. 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 ten 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 ten gallery images compared to single stills. Our proposed multi-gallery, multi-probe minimum 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.

Acknowledgment

References

[1] Karen P. Hollingsworth, Kevin W. Bowyer, and Patrick J. Flynn. Image averaging for improved iris recognition. In *Proc. Int. Conf. on Biometrics (ICB2009)*, 2009.

[2] 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.

[3] Wenyi Zhao and Rama Chellappa, 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. Chellappa, pages 547–567. Elsevier, 2006.

[4] Yingzi Du. Using 2D log-Gabor spatial filters for iris recognition. In *Proc. SPIE 6202: Biometric Technology for Human Identification III*, pages 62020:F1–F8, 2006.

[5] Ying Du, Robert W. Ives, Delores M. Etter, and Thad B. Welch. Use of one-dimensional iris signatures to rank iris pattern similarities. *Optical Engineering*, 45(3):037201–1 – 037201–10, 2006.

[6] 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.

[7] Emine Krichen, Lorène Allano, Sonia Garcia-Salicetti, and Bernadette Dorizzi. Specific texture analysis for iris recognition. In *Proc. Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA 2005)*, pages 23–30, 2005.

[8] 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.

[9] Zhong Zhang and Rick S. Blum. A categorization of multiscale-decompoition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE*, 87(8):1315–1326, Aug 1999.

[10] Deborah Thomas, Kevin W. Bowyer, and Patrick J. Flynn. Strategies for improving face recognition from video. In *Proc. Workshop on Motion and Video Computing*, Feb 2007.

[11] Deborah Thomas, Kevin W. Bowyer, and Patrick J. Flynn. Multi-factor approach to improving recognition performance in surveillance-quality video. In *Proc. IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept 2008.

[12] 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.

[13] J. R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W. Y. Zhao. Iris on the Move™: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, 94(11):1936–1946, 2006.

[14] P. Jonathon Phillips. MBGC presentations and publications. http://face.nist.gov/mbgc/mbgc_presentations.htm, Dec 2008.

[15] Yooyoung Lee, P. Jonathan Phillips, and Ross J. Michaels. An automated video-based system for iris recognition. In *Proc. Int. Conf. on Biometrics (ICB2009)*, 2009.

[16] Zhi Zhou, Yingzi Du, and Craig Belcher. Transforming traditional iris recognition systems to work in nonideal situations. *IEEE Transactions on Industrial Electronics*, 56(8):3203–3213, 2009.

[17] Mayank Vatsa, Richa Singh, and Afzel Noore. Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, 38(4):1021–1035, Aug 2008.

[18] Craig Belcher and Yingzi Du. A selective feature information approach for iris image-quality measure. *IEEE Transactions on Information Forensics and Security*, 3(3):572–577, Sept 2008.

[19] Hugo Proença and Luís Alexandre. Toward noncooperative iris recognition: A classification approach using multiple signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):607–612, Apr 2007.

[20] Chengqiang Liu and Mei Xie. Iris recognition based on DLDA. In *Proc. Int. Conf. on Pattern Recognition*, pages 489–492, Aug 2006.

[21] Kaushik Roy and Prabir Bhattacharya. Iris recognition with support vector machines. In *Proc. Int. Conf. on Biometrics*, pages 486–492, Jan 2006.

[22] John Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[23] Byung Jun Kang and Kang Ryoung Park. Real-time image restoration for iris recognition systems. *IEEE Transactions on Systems, Man, and Cybernetics–Part B*, 37(6):1555–1566, Dec 2007.

[24] The MathWorks™. Image processing toolbox documentation. http://www.mathworks.com/access/helpdesk/help/toolbox/images/index.html. accessed June 2009.

[25] 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.

[26] 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.

[27] 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*, 31(6):964–973, Jun 2009.

[28] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The DET curve in assessment of detection task performance. In *Proc. 5th European Conference on Speech Communication and Technology*, pages 1895–1898, 1997.

[29] Kevin W. Bowyer, Kyong I. Chang, Ping Yan, Patrick J. Flynn, Earnie Hansley, and Sudeep Sarkar. Multi-modal biometrics: an overview. In *Proc. Second Workshop on Multi-Modal User Authentication*, May 2006.

[30] 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.

[31] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, and William Worek. Preliminary Face Recognition Grand Challenge results. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG 2006)*, Apr 2006.

[32] John Daugman. United Arab Emirates deployment of iris recognition, http://www.cl.cam.ac.uk/˜jgd1000/deployments.html, accessed Jan 2009.

[33] Hugo Proença and Luís A. Alexandre. UBIRIS: A noisy iris image database. http://iris.di.ubi.pt/.

[34] Stephanie A. C. Schuckers, Natalia A. Schmid, Aditya Abhyankar, Vivekanand Dorairaj, Christopher K. Boyce, and Lawrence A. Hornak. On techniques for angle compensation in nonideal iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, 37(5):1176–1190, Oct 2007.