Chapter 12
Fusion of Face and Iris Biometrics

Ryan Connaughton, Kevin W. Bowyer, and Patrick J. Flynn

Abstract This chapter presents a system which simultaneously acquires face and iris samples using a single sensor, with the goal of improving recognition accuracy while minimizing sensor cost and acquisition time. The resulting system improves recognition rates beyond the observed recognition rates for either isolated biometrics.

12.1 Introduction

The practice of using more than one biometric modality, sample, sensor, or algorithm to achieve recognition, commonly referred to as *multi-biometrics*, is a technique that is rapidly gaining popularity. By incorporating multi-biometrics into the recognition process, many of the shortcomings of traditional single-biometric systems can be alleviated, and overall recognition accuracy can be improved. Multi-biometrics can inherently increase system robustness by removing the dependency on one particular biometric approach. Further, a system that utilizes more than one biometric feature or matcher may be more difficult to deliberately spoof [17]. Systems that make use of multiple biometric features can also provide redundancy that may lower failure-to-acquire rates. Though multi-biometrics offers many potential advantages over traditional biometric systems, inefficient system design can greatly increase sensor cost, computation time, and data acquisition time.

While research into multi-biometrics has received a large increase in attention over recent years, the task of fusing multiple biometric modalities from a single sensor remains an understudied challenge. Due to a lack of available multimodal data, many current experiments in multi-biometrics create "chimeric" datasets, in

R. Connaughton (✉) • K.W. Bowyer • P.J. Flynn
University of Notre Dame, Notre Dame, IN, USA
e-mail: rconnaug@nd.edu; kwb@cse.nd.edu; flynn@cse.nd.edu

M.J. Burge and K.W. Bowyer (eds.), *Handbook of Iris Recognition*, Advances in Computer Vision and Pattern Recognition, DOI 10.1007/978-1-4471-4402-1_12,

which samples of one biometric modality from one set of subjects are arbitrarily paired with a second biometric modality from a separate set of subjects in order to simulate a multi-biometric scenario [1]. This approach, though useful for preliminary experimentation, may mask unknown dependencies between modalities. Further, chimeric datasets simulate a multi-biometric scenario in which samples of each modality are acquired independently. In practice, it is much more desirable to simultaneously acquire multiple modalities from a single sensor if possible for cost and usability reasons.

This chapter presents a system which simultaneously acquires face and iris samples using a single sensor, with the goal of improving recognition accuracy while minimizing sensor cost and acquisition time. The resulting system improves recognition rates beyond the observed recognition rates for either isolated biometric.

12.2 Characteristics of Multi-biometric Systems

The term *multi-biometrics* encompasses a wide range of fusion techniques, and its precise meaning is somewhat inconsistent in the literature [1, 16, 21]. In the simplest, traditional single-biometric system, one sensor images a particular body part (i.e., iris, face, or fingerprint) to produce a single image. The image is then processed and matched against a gallery using a specific algorithm to obtain a verification or identification result. A multi-biometric system aims to improve recognition rates (or address some other drawbacks of traditional systems) by providing redundancy at one or more of the steps in this recognition process.

In general, there are five types of multi-biometric systems [15]:

1. *Multi-sample*: Multi-sample systems collect and process multiple images of the same biometric. Such systems benefit from some of the advantages of multi-biometrics while minimizing sensor cost.
2. *Multi-instance*: Similar to multi-sample, multi-instance systems collect and process images of several distinct instances of the same biometric trait. Examples of multi-sample systems include systems that consider multiple fingerprints or both irises for recognition. Alternatively, multi-instance systems may collect multiple images of the same trait with some controlled variation; for example, a system may collect face images with smiling and neutral expressions.
3. *Multi-sensor*: A multi-sensor system images the same biometric trait using more than one sensor. Multi-sensor systems may be considered implicitly multi-sample as well. The incorporation of multiple sensors naturally leads to an increase in system cost, but this approach may help to address a particular bias or shortcoming in a specific sensor by obtaining a cross-sensor consensus.
4. *Multi-algorithm*: Multi-algorithm systems use more than one matching algorithm on the same biometric sample and then fuse the results to improve system performance. Because this approach can make use of the same biometric sample for each matcher, multi-algorithm systems can be cost-effective and help to reduce algorithmic biases.

5. *Multi-modal*: Multimodal systems consider more than one biometric trait, or modality, in the recognition process. Ko [7] suggests that multi-modal fusion benefits the most when the biometric modalities are orthogonal. Modalities can be considered orthogonal when the match performance of one modality does not predict the performance of the other. In the ideal scenario, all of the biometrics would be orthogonal, simultaneously imaged with the same sensor, and captured at high quality.

While these five classifications can be used to describe many multi-biometric approaches, there are naturally some systems which represent hybrids of more than one multi-biometric approach. Nonetheless, it is useful to have some method of categorizing multi-biometric systems, and understanding the advantages and disadvantages associated with each approach is crucial to good system design.

12.3 Levels of Fusion

In multi-biometric systems, the term *fusion* is often used to describe the process of combining information from more than one source in the recognition process. The previous section described the stages at which multi-biometric systems may use redundancy to improve performance; fusion is used to combine the results of the redundancy so that a single output can be produced. There are five levels at which fusion can occur in a multi-biometric system:

1. *Signal level*: Using signal-level fusion, multiple samples may be combined together to create one superior sample. An example of signal fusion is a super-resolution technique which combines multiple images of the same iris to achieve a higher-quality image.
2. *Feature level*: In a system that uses feature-level fusion, matching features are first extracted from each biometric sample, and fusion is used to condense all of the features into a single biometric signature.
3. *Score level*: With score-level fusion, the match scores are combined to produce a final result. Examples include a multi-sample approach in which each sample is matched separately and the resulting scores are fused or a multi-algorithm approach in which the same sample is matched using multiple matchers and the results of all of matchers are combined.
4. *Rank level*: Similar to score-level fusion, rank-level fusion combines match rankings, rather than the actual scores, into a final ranking to determine the best match.
5. *Decision level*: Decision-level fusion applies a matcher to each biometric sample (or the same matcher to multiple samples) to obtain a Boolean response indicating whether or not each comparison is a match. The outputs are then fused using Boolean operators, a voting scheme, or some similar method.

It has been suggested in the literature that systems which incorporate fusion at an early stage of the recognition process (e.g., signal- or feature-level fusion) have the potential to be more effective than systems which use fusion later in the pipeline [17]. Despite this, many researchers believe that score-level fusion offers the best trade-off between potential performance gain and ease of implementation [11].

12.4 Related Work

The fusion of face and iris modalities is a biometric approach that has gained increasing attention over the past decade, likely due to the popularity of the individual modalities as well as the natural connection between them. Despite this recent trend, very few studies have been done on fusion of face and iris biometrics from a single sensor.

The most common method of multi-biometric fusion is score-level fusion. Zhang et al. approach the problem of fusing face and iris biometrics under near-infrared lighting using a single sensor [24]. Frontal face images are acquired using a 10-megapixel CCD camera. Eye detection and face alignment are performed using Local Binary Pattern histogram matching as described in Li et al. [9]. The eigenface algorithm and Daugman's algorithm are used to perform face and iris recognition, respectively, and score-level fusion is accomplished via the sum and product rules after min–max normalization. Numerous other score-level fusion approaches have been tested on chimeric datasets. Chen and Te Chu use an unweighted average of the outputs of matchers based on neural networks [4]. Wang et al. test weighted average, linear discriminant analysis, and neural networks for score-level fusion [22].

Another common approach to biometric fusion is feature-level fusion through concatenation. Rattani and Tistarelli compute SIFT features for chimeric face and iris images and concatenate the resulting feature vectors [14]. The number of matching SIFT features between two vectors (measured by Euclidean distance) is used as a match score for that comparison. Son and Lee extract features for face and iris images based on a Daubechies wavelet transform [18]. Concatenation is used to form a joint feature vector, and Euclidean distance between feature vectors is used to generate match scores.

The Multiple Biometrics Grand Challenge (MBGC) provided a collection of face and iris data to researchers in order to provide a standard test bed for comparing matching and fusion techniques [12, 13]. The MBGC data included a subset of the near-infrared videos used in the experiments being presented in this chapter, as well as face stills, high-quality color face video, iris stills, and iris video. In general, results showed that fusion of face and iris biometrics offered improved accuracy over either biometric alone. The near-infrared videos released as part of the MBGC are also used by Yang et al. [23]. Yang et al. investigate the use of SIFT features to perform alignment between the partial faces present in the dataset in order to facilitate face matching, but do not incorporate these results into a multi-biometric experiment.

The work presented in this chapter differs from previous work in the fusion of face and iris biometrics in several facets. First, this chapter uses only genuine multimodal data, rather than chimeric data for experimentation. Additionally, the fusion is accomplished using a single sensor. Though Zhang et al. also use a single sensor, the authors also manually acquire each image to guarantee high-quality face and iris samples. In the experiments presented in this chapter, an on-the-move and at-distance sensor is used to acquire data for a high-throughput scenario. The resulting dataset consists of a much wider range of sample quality with incomplete data for some subjects, making the dataset a practical but challenging test bed for fusion experiments. These experiments also differ from work presented on the MBGC data; the near-infrared videos used in the MBGC dataset were manually selected to guarantee the presence of a subject in the field of view, whereas in the experiments shown in this chapter, this process is done automatically. Finally, this work uses multi modal, multi-sample, and multi-instance approaches to improve system accuracy and robustness.

12.5 Approach

To facilitate the fusion of face and iris biometrics from a single sensor, the Iris on the Move (IOM) sensor was selected for data acquisition. The IOM, shown in Fig. 12.1, is a sensor designed for high-throughput stand-off iris recognition [10]. The IOM features a portal which subjects walk through at normal walking pace.

Fig. 12.1 Picture of the Iris on the Move sensor designed by Sarnoff Corporation. The IOM was used for all probe data collection (Picture reprinted from [2] with permission from Elsevier)

Fig. 12.2 Example of corresponding frames from the IOM as the subject passes through the in-focus region of the portal. The *left* image shows a frame from the *top* camera, the *middle* image shows a frame from the *middle* camera, and the *right* shows a frame from the *bottom* camera

As a subject passes through the portal, the subject is illuminated with near-infrared (NIR) LEDs, and frontal video is captured by an array of three vertically arranged, fixed-focus cameras equipped with NIR filters. The presence of multiple cameras allows the system to handle a larger range of subject heights, and the sensor can be extended to include more than three cameras to support an even larger range of subject heights. Though the sensor is intended for iris image acquisition, the face is typically captured as well. While the sides of the portal help to direct subjects into the field of view of the cameras, it is possible for subjects to stray partially out of the video frames, leading to frames with partial faces or only one iris visible. Figure 12.2 shows corresponding frames from each of the three IOM cameras while a subject passes through the in-focus region of the IOM. Each frame captured by one of the IOM cameras is a 2048 by 2048 pixel gray scale image. A typical iris acquired by the system is approximately 120 pixels in diameter.

The general steps used in this work to combine face and iris biometrics from the IOM sensor are outlined in Fig. 12.3. As previously described, when a subject passes through the IOM portal, three videos are collected, with one video coming from each of the IOM cameras. In a preprocessing step, the corresponding frames of the three videos are stitched together to create one virtual video. Next, a series of detection phases are used to locate whole faces and eyes in each frame. Matching is then performed on each face sample and iris sample independently, and the results are fused using several different techniques.

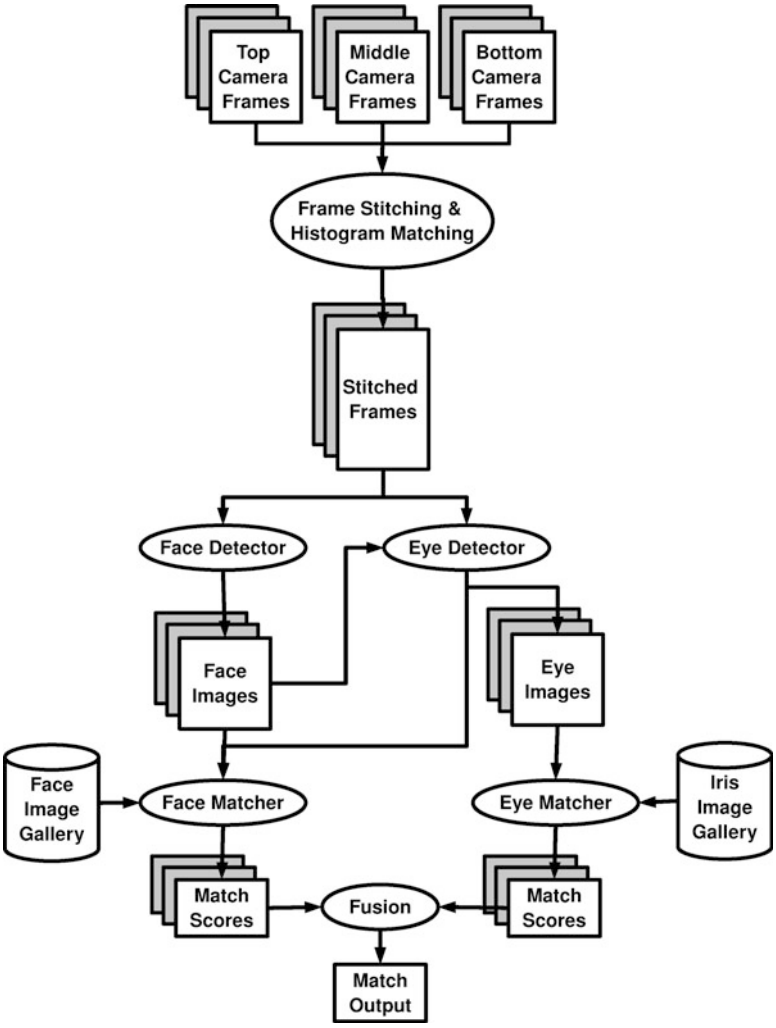

Fig. 12.3 A diagram of the pipeline used in the proposed multi-biometric system

12.5.1 Preprocessing

In order to increase the likelihood of a whole face being captured for each subject, the three videos from each IOM acquisition are "stitched" together to combine corresponding frames. As can be seen in Fig. 12.2, there is significant vertical overlap between the top and middle cameras, as well as between the middle and bottom cameras. Due to imperfect calibration of the individual cameras, some horizontal misalignment between the cameras is also present.

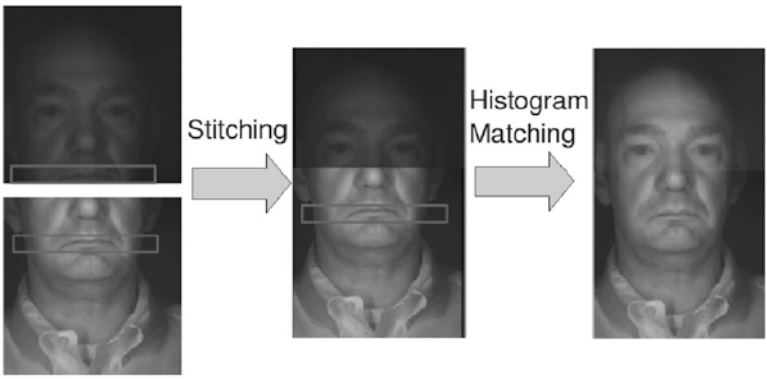

Fig. 12.4 An example of the progression during alignment between corresponding frames from the *top* to *middle* camera. The *top left* image is the frame from the *top* camera with the template marked as a *rectangle*. The *bottom left* image is the frame from the *middle* camera, with the matched region indicated. The *middle* image is the composite image, with the frame from the *top* camera cropped and padded. The overlapping region is indicated. The *right* image shows the final stitching results after histogram matching. A similar approach is used to stitch the frame from the *bottom* camera to the *middle* frame

A template-matching approach is taken to determine the desired translation to align frames from adjacent cameras. Specifically, the bottom portion of the top frame is cropped and used as a template. This template is then matched against the upper half of the middle frame, and the best match is selected as the desired alignment. This process is repeated for the bottom camera, where the template is created from the top portion of the bottom frame and matched against the lower half of the middle frame.

Finally, noticeable illumination differences were observed between corresponding frames from different cameras. To account for this discrepancy, histogram matching is used to match the top and bottom frame to the illumination observed in the middle frame. Figure 12.4 shows the intermediate and final results of the stitching procedure for an example frame.

12.5.2 Face Detection

Once the frame stitching is completed, the next step in the preprocessing phase is to detect a face in each frame. To accomplish this task, the OpenCV implementation of the Viola-Jones cascade face detector is used [3, 20]. The detector was trained on whole faces and thus may or may not detect faces which lie only partially within the field of view of the camera.

12.5.3 Eye Detection

The purpose of the eye detection phase is twofold. The primary goal is to detect any eyes present in each frame for iris matching. However, the locations of the eyes that are detected in the faces produced by the face detector are also used for an alignment phase during face matching. A template-matching approach is adopted for eye detection. The template used to search for eyes in each frame is based on the specular highlights generated by the reflection of the IOM LEDs.

The eye detection is completed in two phases. First, the template matching is performed on the upper left and upper right quadrants of each face detected by the face detector. This approach guarantees that each detected face will have two eye locations estimated as well.

Because it is possible for eyes to be detected in frames where whole faces were not present (or in frames where the face detector failed to detect the face), a second round of template matching is performed on any stitched frame where a face was not detected. In these frames, the location of the partial face can be crudely estimated by computing the sums of the rows and columns of the image and comparing these sums to appropriate thresholds. This partial face detection step is not required, but reduces the likelihood of false eye detections by limiting the search space to the region of the image that is likely to contain the eyes. An example of a face region being estimated in this manner is shown in Fig. 12.5. Once the partial face region has been estimated, the template matching is performed twice to identify the two best eye locations. Finally, the detected eyes are cropped from the corresponding location in the *original* frames to remove any possible artifacts caused by the histogram matching in the stitching phase. In cases where the detected eye is located in the overlapping region between two cameras, the eye is cropped from *both* camera frames.

12.5.4 Face Matching

In this work, Colorado State University's implementation of the eigenface algorithm is used for face matching [5, 19]. To achieve alignment with the training set, the probe face images are normalized using the eye centers detected by the eye detector. The Mahalanobis cosine metric is used to compute the distance between two feature vectors. Using this metric, match scores can range from -1.0 to 1.0, with -1.0 being a perfect score. The output of the face matcher stage of the pipeline is a distance for every comparison between each probe face image and gallery face image.

12.5.5 Iris Matching

For the iris matcher, a modified version of Daugman's algorithm is used to compare each probe iris image to the gallery [6]. The normalized fractional Hamming

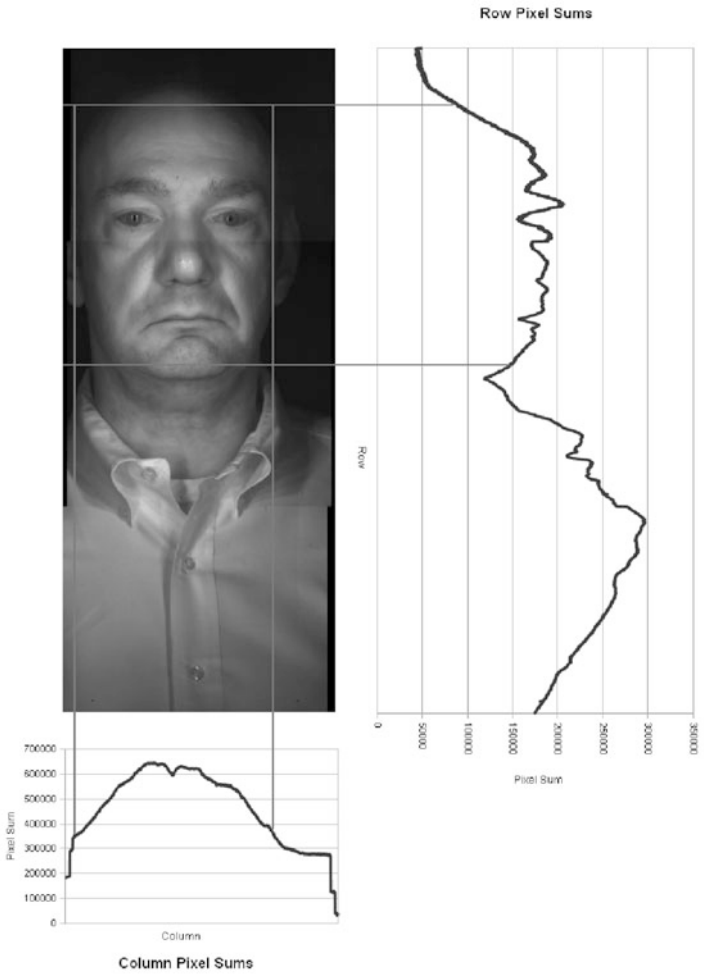

Fig. 12.5 Example of the image projection technique used to estimate the location of the face during eye detection. The graphs on the *right* and *bottom* of the image represent the summations of the pixel values in each *row* or *column*, respectively. The projected lines represent the face boundaries determined using appropriate thresholds

distance, referred to simply as the Hamming distance in the rest of this work, ranges from 0.0 to 1.0, with 0.0 being a perfect match. The Hamming distance is normalized to adjust low Hamming distances that occur for comparisons that used relatively few bits. The output of the iris matcher stage of the pipeline is a Hamming distance for every comparison between each probe eye image and gallery iris image.

12.5.6 Fusion

In this framework, the problem is multi-sample (i.e., several faces from each video), multimodal (i.e., both iris and face samples from each video), and multi-instance (i.e., both left and right irises from each video). Consequently, there are many methods which could be used to combine the face and iris biometrics from each video. Several fusion techniques are considered at both the score and rank levels.

The first method considers only one biometric modality in the fusion process and makes use only of the multi-sample and multi-instance dimensions of the problem by taking the minimum score for a given modality. For example, in the MinIris approach, the minimum score for all of the iris comparisons from a given video is reported as the best match. Similarly, the MinFace approach takes the minimum score for all of the face comparisons from a given video to determine the best match. Equations (12.1) and (12.2) express the MinIris and MinFace fusion rules, respectively, for a given probe video:

$$\text{MinIris} = \text{Min}\{I_{i,j} | i = 1 \ldots n, j = 1 \ldots G\} \tag{12.1}$$

$$\text{MinFace} = \text{Min}\{F_{i,j} | i = 1 \ldots m, j = 1 \ldots G\} \tag{12.2}$$

where n and m are the number of irises and faces detected in the video, respectively; G is the number of gallery subjects; $I_{i,j}$ is the Hamming distance between the i-th iris and the j-th gallery subject; and $F_{i,j}$ is the score for the comparison between the i-th face and the j-th gallery subject.

The next type of fusion method considered is rank-level fusion and can incorporate face, iris, or both modalities into the decision process. A Borda count is used to determine a best match across the desired biometric modalities. In a Borda count, the scores for all comparisons from a given sample are sorted such that the first rank corresponds to the best score for that sample. Each sample then casts votes for the top v ranked subjects, where the weight of each vote is inversely proportional to rank number. Each sample votes in this manner, and the gallery subject with the most votes is taken to be the best match. In these experiments, the BordaIris method considers only the iris scores to perform fusion, and the BordaFace method considers only face scores. The BordaBoth method allows both face and iris samples to vote, with v votes being cast by each iris and face sample.

Two vote weighting schemes are tested for the BordaIris, BordaFace, and BordaBoth fusion methods. In the linear approach, the vote weight is linearly proportional to the rank; specifically, the weight associated with the rank-n match is described by the equation

$$\text{VoteWeight}_n = v + 2 - n \tag{12.3}$$

and v represents the total number of votes cast by each biometric sample. In the exponential approach, the weight of the vote is exponentially related to the

rank. Specifically, the weight associated with the rank-n match is described by the equation

$$\text{VoteWeight}_n = 2^{v-n} \tag{12.4}$$

The third fusion method again uses score-level fusion, implementing a weighted summation of the iris and face scores. The summation rule can be expressed as Eq. (12.5) for a given probe video:

$$\text{SumScore}_k = \frac{\alpha * \sum_{i=1}^{n} (1 - \text{FNorm}_{i,k}) + \beta * \sum_{j=1}^{m} (1 - \text{INorm}_{j,k})}{\alpha * n + \beta * m} \tag{12.5}$$

where n and m are the number of irises and faces detected in the video, respectively; $\text{INorm}_{j,k}$ is the normalized Hamming distance between the j-th iris and the k-th gallery subject; and $\text{FNorm}_{i,k}$ is the normalized score for the comparison between the i-th face and the k-th gallery subject. Each face and iris score is normalized using min-max normalization, according to the expression

$$\text{Score}' = \frac{\text{Score} - \text{Min}}{\text{Max} - \text{Min}} \tag{12.6}$$

where Min and Max are the minimum and maximum possible values for each score metric, so that all normalized scores fall between 0.0 and 1.0, with 1.0 representing a perfect match. In Eq. (12.5), α and β are coefficients used to weight the face and iris biometrics, respectively. In the presented work, $\alpha = 1 - \beta$ for simplicity. In Eq. (12.5), SumScore_k represents the final match score for the given probe video with gallery subject k; the best match score can be determined by finding the maximum SumScore_k for all k. SumIris is the special case where $\alpha = 0$ and $\beta = 1$, which corresponds to summing only the iris scores to determine the best match. Similarly, SumFace is the case where $\alpha = 1$ and $\beta = 0$ and equates to summing only the normalized face scores.

12.6 Experiments

The previously described multi-biometric system was tested on a probe dataset of 1,886 IOM video sets. Note that here a video "set" refers to the corresponding videos from each of the three IOM cameras, so the dataset is comprised of 5,658 videos in total. The 1,886 videos spanned 363 unique subjects, with an average of about five videos per subject. The most frequently occurring probe subject had 15 videos in the probe set, and the least frequently occurring had one probe video.

The iris gallery contained one left eye and one right eye for each of the 363 gallery subjects. The gallery images were acquired using the LG IrisAccess 4000 (LG4000) [8], a high-quality iris acquisition camera, and the gallery was manually

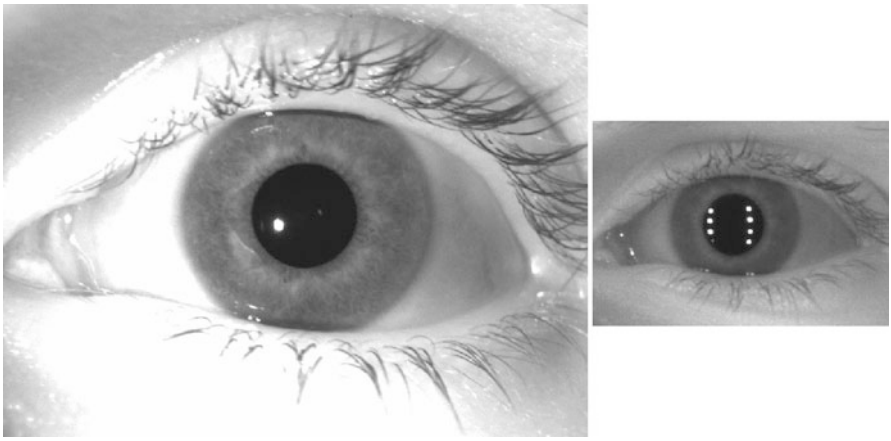

Fig. 12.6 Images of the same iris image using the LG4000 (*left*) and the IOM (*right*). The IOM image shown on the right represents a well-focused IOM iris image

screened for good quality and segmentation. For comparison, Fig. 12.6 shows an example of an image of the same iris acquired from both the LG4000 and the IOM.

The face gallery contained one full face image for each of the 363 subjects. The gallery images were acquired using the IOM. Each of the 363 subjects in the study had an additional IOM video set acquired in which the presence of a whole face was verified manually. The frames were stitched using the process previously described, and then, the best frame was manually selected, and the coordinates of the eye centers were manually annotated for alignment. The PCA training was performed on the face image gallery.

12.6.1 Detection Results

Across the entire dataset, 14,829 left irises and 14,711 right irises were detected and successfully segmented, and 9,833 faces were detected with valid eye locations for alignment. In this context, "successful segmentation" simply means that the iris segmentation routine returned pupil and limbic boundaries; it does *not* guarantee correctness. On average, 15.7 ($\sigma = 8.1$) irises, 5.2 ($\sigma = 3.7$) faces, and 20.9 ($\sigma = 20.9$) of either biometric samples were found in each video.

Table 12.1 provides a breakdown of the detection results by frame and video. The 1,886 videos were composed of a total of 28,381 frames. From Table 12.1, it can be seen that while a large number of frames (44.1%) contained no detected features, a much larger percentage of the probe *videos* (99.3%) had at least one biometric feature detected. Further, the majority (80.6%) of the probe videos contained samples of face and both iris features.

Table 12.1 Detailed detection results

Modalities detected	Frame count	Video count
Left iris (only)	1,447 (5.1%)	35 (1.9%)
Right iris (only)	2,104 (7.4%)	46 (2.4%)
Face (only)	900 (3.2%)	2 (0.1%)
Left and right irises (only)	2,495 (8.8%)	209 (11.1%)
Face and left iris (only)	1,411 (5.0%)	34 (1.8%)
Face and right iris (only)	724 (2.6%)	27 (1.4%)
Face, left, and right irises	6,798 (24.0%)	1,522 (80.6%)
None	12,502 (44.1%)	11 (0.6%)

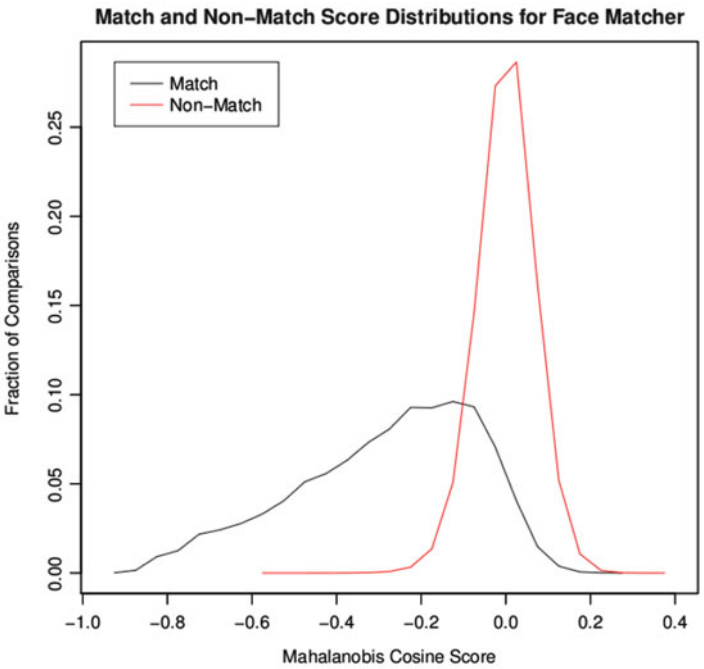

Fig. 12.7 The match and non-match score distributions for the face features from the entire probe dataset

12.6.2 Matching Results

Figure 12.7 shows the match and non-match score distributions for all 9,833 detected faces. The mean match score was −0.281 with a standard deviation of 0.213, while the mean non-match score was 0.000 with a standard deviation of 0.676. If each face were treated independently, the rank-one recognition achieved for the 9,833 probes faces would be 51.6% (5,073/9,833) recognition.

The results from the left and right irises were aggregated, and Fig. 12.8 shows the match and non-match score distributions. The mean match score was 0.398 with

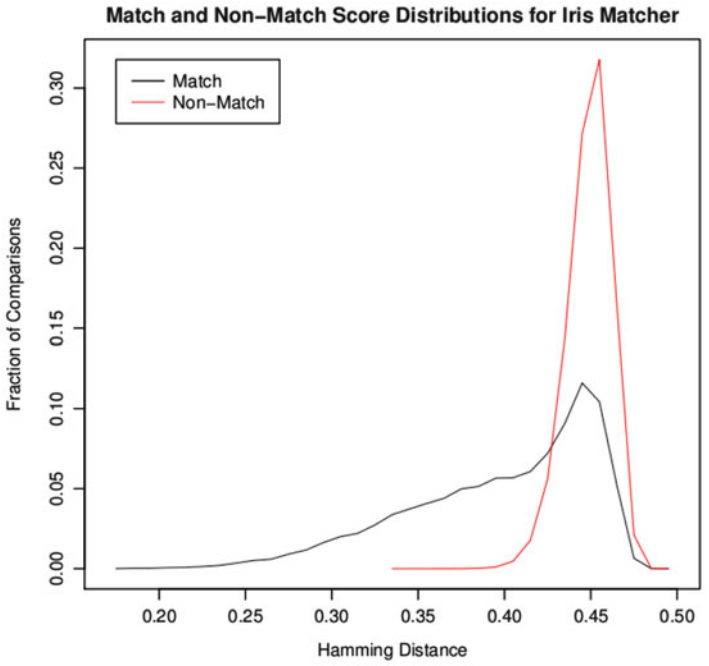

Fig. 12.8 The match and non-match score distributions for the *left* and *right* iris features from the entire probe dataset

a standard deviation of 0.053, while the mean non-match score was 0.449 with a standard deviation of 0.013. Figure 12.8 shows a significant number of match comparisons with fairly high scores. Upon examination of the data, it was found that most of these scores arise from incorrect segmentation. In some cases, these high match scores were caused by severe image defocus. Additionally, there are some false positives from the eye detector (non-eye regions) that contain features that resemble pupil and limbic boundaries according to the segmentation routine. If each iris image were treated independently, the rank-one recognition achieved for all of the probe irises would be 46.6% (13,556/29,112) recognition.

12.6.3 Fusion Results

The results of the iris and face matchers were combined using each of the methods previously described. The rank-one recognition rates achieved by each fusion approach are shown in Table 12.2. In the fusion methods based on Borda counts, the number of votes given to each sample was varied between 1 and 363 (though all samples were given the same number of votes for any given fusion experiment), and the best results for each approach are presented. Similarly, results from the optimal tested values of α and β are presented.

Table 12.2 Rank one recognition rates for fusion approaches

Approach	Fusion parameters	Rank one (raw)
MinIris		86.7% (1,635/1,886)
MinFace		62.6% (1,180/1,886)
BordaIris-linear	$v = 3$	86.4% (1,629/1,886)
BordaIris-exponential	$v = 20$	86.8% (1,637/1,886)
BordaFace-linear	$v = 3$	58.9% (1,110/1,886)
BordaFace-exponential	$v = 5$	59.3% (1,118/1,886)
BordaBoth-linear	$v = 10$	91.7% (1,729/1,886)
BordaBoth-exponential	$v = 10$	92.0% (1,735/1,886)
SumIris	$\alpha = 0.0, \beta = 1.0$	87.8% (1,656/1,886)
SumFace	$\alpha = 1.0, \beta = 0.0$	61.3% (1,156/1,886)
SumBoth	$\alpha = 0.3, \beta = 0.7$	93.2% (1,757/1,886)

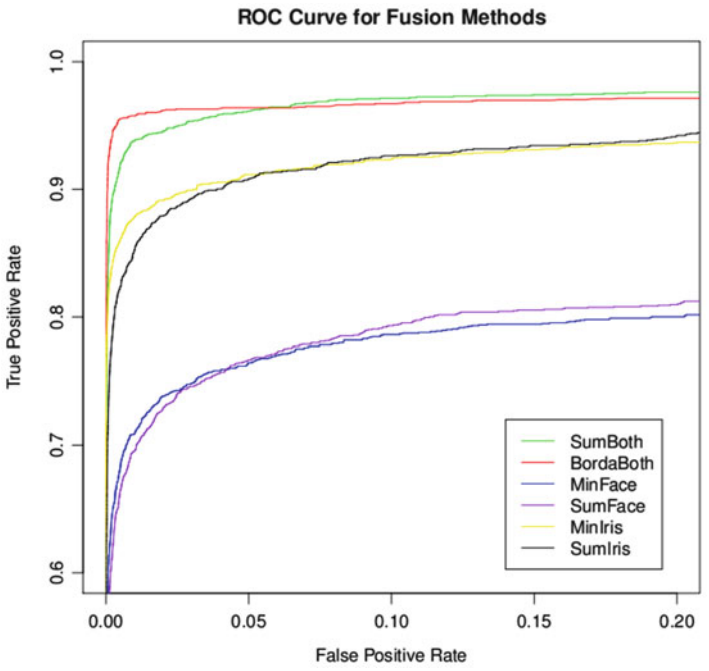

Fig. 12.9 ROC curves for the various fusion methods using the optimal tested parameters for each. The BordaBoth method shown is the BordaBoth-exponential method

In summary, the best single-modality fusion approach was the SumIris approach, which achieved an 87.8% rank-one recognition rate. The SumBoth approach achieved the overall highest recognition rate (93.2%), and all multi-modal fusion approaches achieved higher recognition rates than the fusion methods based on a single modality.

Figure 12.9 shows the ROC curves for the best SumBoth and BordaBoth approaches, as well as the MinIris, MinFace, SumFace, and SumIris results

for comparison. From this graph, it is clear that the BordaBoth and SumBoth approaches outperform the single-modality fusion methods. Interestingly, while SumBoth achieved the highest rank-one recognition rate, Fig. 12.9 shows that the BordaBoth fusion technique performs better at false-positive rates less than 0.06.

In general, the videos that failed to match correctly typically had relatively few face and iris features detected. While the iris proved to be the more accurate of the two modalities in the multi-sample fusion scenarios, Fig. 12.8 indicates that many of the iris features detected are of poor quality, represent false detections from the eye detector, or failed to segment correctly. While the fusion techniques in these experiments were able to overcome these challenges when enough samples were present, videos in which a small number of faces and iris are detected are much less likely to be correctly matched.

12.7 Conclusions

This chapter presents an investigation into the fusion of face and iris biometrics from a single sensor, a surprisingly understudied problem in current literature. The previously described multi-biometrics framework utilizes multi-sample, multi-instance, and multimodal fusion techniques to improve recognition rates from a single sensor. The multi-biometric system is tested on a non-chimeric dataset of over 1,886 videos spanning 363 subjects. This represents one of the largest genuine multimodal experiments that has been conducted to date. Face and iris biometric samples extracted from videos produced from the Iris on the Move sensor were combined using several different fusion methods. In these experiments, the combination of face and iris biometrics via match score summation yielded a 5.4% increase in recognition rate over the best single-modality approach that was tested, while a modified Borda count approach performed best at lower false-positive rates (< 0.06).

The multi-biometrics system proposed exploits the face information collected by the IOM, a sensor that is intended for iris recognition purposes, with no modifications to the sensor and no increase in probe data acquisition time. The resulting system is less likely to experience failures to acquire, and the use of multiple modalities could allow the system to identify subjects with incomplete gallery data. This approach could be extended to operate on other stand-off iris sensors, which often detect the face as a preliminary step to iris image acquisition.

Acknowledgements Datasets used in this work were acquired under funding from the National Science Foundation under grant CNS01-30839, by the Central Intelligence Agency and by the Technical Support Working Group under US Army Contract W91CRB-08-C-0093. The authors were supported by a grant from the Intelligence Advanced Research Projects Activity.

References

1. Bowyer, K.W., Chang, K.I., Yan, P., Flynn, P.J., Hansley, E., Sarkar, S.: Multi-modal biometrics: an overview. In: Presented at the Second Workshop on Multi-Modal User Authentication (MMUA 2006), Toulouse (2006)
2. Bowyer, K.W., Hollingsworth, K., Flynn, P.J.: Image understanding for iris biometrics: a survey. Comput. Vis. Image Underst. **110**, 281–307 (2008)
3. Bradski, G., Kaehler, A.: Learning OpenCV. O'Reilly Media, Sebastopol (2008)
4. Chen, C.H., Te Chu, C.: Fusion of face and iris features for multimodal biometrics. In: Zhang, D., Jain, A. (eds.) Advances in Biometrics. Lecture Notes in Computer Science, vol. 3832, pp. 571–580. Springer, Berlin/Heidelberg (2005)
5. Colorado State University: Evaluation of Face Recognition Algorithms. http://www.cs.colostate.edu/evalfacerec/algorithms5.html (2010)
6. Daugman, J.: How iris recognition works. In: 2002 International Conference on Image Processing, Rochester, vol. 1, pp. 33–36 (2002)
7. Ko, T.: Multimodal biometric identification for large user population using fingerprint, face and iris recognition. In: 34th Applied Imagery and Pattern Recognition Workshop, Washington, DC, pp. 218–223 (2005)
8. LG Iris: LG Iris Products and Solutions. http://www.lgiris.com/ps/products/irisaccess4000.htm (2010)
9. Li, S.Z., Chu, R., Ao, M., Zhang, L., He, R.: Highly accurate and fast face recognition using near infrared images. In: International Conference on Biometrics (ICB 2006), Hong Kong, pp. 151–158 (2006)
10. Matey, J., Naroditsky, O., Hanna, K., Kolczynski, R., LoIacono, D., Mangru, S., Tinker, M., Zappia, T., Zhao, W.: Iris on the move: acquisition of images for iris recognition in less constrained environments. Proc. IEEE **94**, 1936–1947 (2006)
11. Morizet, N., Gilles, J.: A new adaptive combination approach to score level fusion for face and iris biometrics combining wavelets and statistical moments. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Porikli, F., Peters, J., Klosowski, J., Arns, L., Chun, Y., Rhyne, T.M., Monroe, L. (eds.) Advances in Visual Computing. Lecture Notes in Computer Science, vol. 5359, pp. 661–671. Springer, Berlin/Heidelberg (2008)
12. National Institute of Standards and Technology (NIST): Portal Challenge Problem – Multiple Biometric Grand Challenge, Preliminary Results of Version 2. http://biometrics.nist.gov/cs_links/face/mbgc/2009/PORTAL_V2_FINAL.pdf (2009)
13. Phillips, P.J., Flynn, P.J., Beveridge, J.R., Scruggs, W.T., O'Toole, A.J., David, B., Bowyer, K.W., Draper, B.A., Givens, G.H., Lui, Y.M., Sahibzada, H., Scallan Iii, J.A., Weimer, S.: Overview of the multiple biometrics grand challenge. In: Proceedings of the Third International Conference on Advances in Biometrics, ICB '09, pp. 705–714. Springer, Berlin/Heidelberg (2009)
14. Rattani, A., Tistarelli, M.: Robust multi-modal and multi-unit feature level fusion of face and iris biometrics. In: Tistarelli, M., Nixon, M. (eds.) Advances in Biometrics. Lecture Notes in Computer Science, vol. 5558, pp. 960–969. Springer, Berlin/Heidelberg (2009)
15. Ross, A.: An introduction to multibiometrics. In: 15th European Signal Processing Conference (EUSIPCO), Poznan, pp. 20–24 (2007)
16. Ross, A., Jain, A.K.: Multimodal biometrics: an overview. In: 12th European Signal Processing Conference (EUSIPCO), Vienna, pp. 1221–1224 (2004)
17. Ross, A.A., Nandakumar, K., Jain, A.K.: Handbook of Multibiometrics. Springer Science and Business Media. Springer, New York (2006)
18. Son, B., Lee, Y.: Biometric authentication system using reduced joint feature vector of iris and face. In: Kanade, T., Jain, A., Ratha, N. (eds.) 6th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA'03). Lecture Notes in Computer Science, vol. 3546, pp. 513–522. Springer, Berlin/Heidelberg (2005)

19. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '91), Lahaina, pp. 586–591 (1991)
20. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, vol. 1, pp. 511–518 (2001)
21. Volner, R., Bores, P.: Multi-biometric techniques, standards activities and experimenting. In: 2006 International Baltic Electronics Conference, Tallinn, pp. 1–4 (2006)
22. Wang, Y., Tan, T., Jain, A.K.: Combining face and iris biometrics for identity verification. In: 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA'03), pp. 805–813. Springer, Berlin/Heidelberg (2003)
23. Yang, J., Liao, S., Li, S.: Automatic partial face alignment in nir video sequences. In: Tistarelli, M., Nixon, M. (eds.) Advances in Biometrics. Lecture Notes in Computer Science, vol. 5558, pp. 249–258. Springer, Berlin/Heidelberg (2009)
24. Zhang, Z., Wang, R., Pan, K., Li, S., Zhang, P.: Fusion of near infrared face and iris biometrics. In: Lee, S.W., Li, S. (eds.) Advances in Biometrics. Lecture Notes in Computer Science, vol. 4642, pp. 172–180. Springer, Berlin/Heidelberg (2007)