On the Dynamic Selection of Biometric Fusion Algorithms

Mayank Vatsa, *Member, IEEE,* Richa Singh, *Member, IEEE,* Afzel Noore, *Member, IEEE,* and Arun Ross, *Member, IEEE*

***Abstract*—Biometric fusion involves consolidating the output of two or more biometric classifiers in order to render a decision about the identity of an individual. We consider the problem of designing a fusion scheme when (a) the number of training samples is limited, thereby undermining the use of a purely density-based scheme and the likelihood ratio test statistic; (b) the output of multiple matchers report conflicting results; and (c) the use of a single fusion rule may not be practical due to the diversity of scenarios encountered in the probe dataset. To address these issues, a dynamic reconciliation scheme for fusion rule selection is proposed. In this regard, the contribution of this paper is two-fold: (a) the design of a sequential fusion technique that uses the likelihood ratio test-statistic in conjunction with a support vector machine classifier to account for errors in the former; and (b) the design of a dynamic reconciliation algorithm that unifies the constituent classifiers and fusion schemes to optimize both verification accuracy and computational cost. The case study on multi-classifier face recognition suggests that the proposed algorithm can address the uncertainty associated with component matchers. Indeed, it is observed that the proposed method performs well even in the presence of confounding covariate factors thereby indicating its potential for large-scale face recognition.**

***Index Terms*—Biometrics, match score fusion, face verification, support vector machine.**

I. Introduction

THE paradigm of information fusion, that entails the consolidation of evidence presented by multiple sources, has been successfully used to enhance the recognition accuracy of biometric systems. The use of multiple pieces of evidences in order to deduce or verify human identity is often referred to as *multibiometrics*. While fusion can be accomplished at several different levels in a biometric system [18] - viz., data-level, feature-level, score-level, rank-level, and decision-level - fusion at the match score level has been extensively studied in the literature. Fusion at the match score level involves combining the match scores generated by multiple classifiers (or matchers) in order to render a decision about the identity of the subject. There are several different schemes for performing score level fusion based on different models. These include density-based fusion schemes (where the model is based on estimating density functions for the genuine and impostor score distributions); transformation-based fusion schemes (where the model is based on estimating normalization functions); and classifier-based fusion schemes (where the model is a classifier).

While match score fusion has been demonstrated to be effective [18], [22], its matching performance is compromised under several scenarios:

1) Density-based score fusion schemes [18] which use the likelihood ratio test to formulate the fusion rule can be affected by the use of incorrect density functions for the genuine and impostor scores. The use of parametric methods of density estimation can be based on the assumption of incorrect models (e.g., Gaussian densities for both genuine and impostor scores) that can lead to sub-optimal fusion rules; the use of non-parametric methods, on the other hand, is affected by the availability of a small number of training samples (especially genuine scores) thereby impacting the feasibility of designing an effective fusion rule.
2) Classifier-based fusion schemes [2] are susceptible to over-training on one hand and classifier bias on the other [4], [27]. Further, a pure data-driven approach will not be able to accommodate scenarios that are not represented in the training data. For example, when conflicting scores from multiple matchers are presented to the fusion classifier, then, in the absence of sufficient training samples representing such a scenario, an incorrect decision may be regularly rendered.

Training and using a single fusion rule - whether it be the simple sum rule or the likelihood ratio-based fusion rule - on the *entire* probe dataset may not be appropriate for the reasons stated above. Further, component classifiers can render "conflicting" decisions that can impact the performance of transformation-based schemes such as the simple sum rule. To address these issues and subsequently improve the verification performance of a biometric system, we propose a sequential fusion algorithm which combines a density-based fusion scheme with a classifier-based scheme. *The first contribution lies in using a support vector machine (SVM) classifier in conjunction with the likelihood ratio test statistic*. The likelihood ratio aspect of the algorithm helps in modeling the underlying class distribution using simple Gaussian mixture models; the statistical and geometrical properties of SVM [14], [15], [23] ensures that there is a "correction" of the decision rendered by the likelihood ratio test statistic. By employing a simple model to characterize the genuine and impostor density functions, the requirement for a large number of training samples is avoided.

The sequential nature of the proposed fusion algorithm

M. Vatsa and R. Singh are with IIIT Delhi, India email: {mayank, rsingh}@iiitd.ac.in
A. Noore and A. Ross are with the Lane Department of CSEE, West Virginia University, USA e-mail: {afzel.noore, arun.ross}@mail.wvu.edu.

makes it computationally expensive. The fusion algorithm may not be required if the probe image is of high quality and exhibits sufficient biometric information useful for recognition using only one biometric classifier. Further, simple fusion rules such as sum rule with min/max normalization can be used for most of the probe cases when multi-classifier biometric information is not highly conflicting. One way to improve the verification accuracy without increasing the computational cost too much is to develop a context switching scheme that dynamically selects the most appropriate classifier or fusion algorithm for the given probe. *The second contribution of this work is the design of an algorithm for the dynamic selection of constituent unimodal biometric classifiers or match score fusion algorithms that not only improves the verification accuracy but also decreases the computational cost of the system.* In a two-class, bi-classifier biometric system, the reconciliation algorithm uses quality information (not based on match scores) to select one of four options: (1) first biometric classifier only, (2) second biometric classifier only, (3) sum rule with min/max normalization, and (4) sequential match score fusion. The selected option is then used to render the final decision.

The performance of the proposed algorithm is evaluated in the context of a face recognition application to mitigate the effect of covariate factors such as pose, expression, illumination, and occlusion. Match scores computed from two face recognition algorithms, namely local binary pattern [3] and neural network architecture based 2D log polar Gabor transform [20], are fused and the verification performance is compared with existing match score fusion algorithms. Experiments indicate that the proposed fusion architecture efficiently improves the verification performance without increasing the computational cost.

II. Proposed Sequential Match Score Fusion Algorithm

Fig. 1 shows the steps involved in the proposed fusion algorithm that consists of two steps: (1) match score fusion and (2) classification. First, the match scores are transformed into belief assignments using density estimation schemes. In the next step, belief model is used for fusion and finally, statistical likelihood ratio and SVM are used for classification. Description of the fusion algorithm uses two-class bi-classifier approach and throughout the paper we use c_1 to represent the first biometric classifier and c_2 to represent the second biometric classifier.

A. Match Score Fusion

For a two class problem, let $\Theta = \{\theta_{gen}, \theta_{imp}\}$, where θ_{gen} represents the genuine hypothesis and θ_{imp} represents the impostor hypothesis. The first step in the sequential fusion algorithm is to transform match scores into belief assignment. Since the belief functions are a generalized form of probability theory that can perform fusion in presence of uncertainty and imperfect data, probabilistic approach can be effectively used as the basis of the fusion algorithm. A multivariate density estimation technique is used to compute belief assignments induced by the match scores because previous literature has shown the usefulness of mixture models in biometrics [18]. Multivariate Gaussian density [7] in d dimensions can be written as,

$$p(\mathbf{x}, \mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^t\ \mathbf{\Sigma}^{-1}\ (\mathbf{x}-\mu)\right] \tag{1}$$

where $\mathbf{x}$ is a vector with d components, μ is the mean vector, and $\mathbf{\Sigma}$ is the covariance matrix. Let $C_n(\mathbf{x}, i)$ be the conditional joint density of d match scores obtained from the n^{th} classifier and $i \in \Theta$. $C_n(\mathbf{x}, i)$ is computed using Equation 2.

$$C_n(\mathbf{x}, i) = \sum_{j=1}^{M_n(i)} W_n(i,j)\ p(\mathbf{x},\ \mu_n(i,j),\ \mathbf{\Sigma}_n(i,j)) \tag{2}$$

where $\mu_n(i,j)$, $\mathbf{\Sigma}_n(i,j)$ and $W_n(i,j)$ are the mean vector, covariance matrix, and weight factor respectively corresponding to the j^{th} mixture component in the conditional joint density. Also, $\sum_{j=1}^{M_n(i)} W_n(i,j) = 1$ and $M_n(i)$ is the number of mixture components used to model the density. Further, a recursive algorithm [29] is used to estimate the parameters of the mixture model.

Let $\mathbf{x} = (x_1, ..., x_n)$ be the match scores computed from $n (= c_1, c_2)$ biometric classifiers or matchers. Belief assignment of the n^{th} classifier, m_n, is computed using Equation 3.

$$m_n(i) = \frac{\alpha_n(i) C_n(x_n, i)}{\sum_i\ \alpha_n(i) C_n(x_n, i)} \tag{3}$$

where $C_n(x_n, i)$ is the marginal density and $\alpha_n(i)$ is the verification accuracy prior of the n^{th} classifier that is used as the ancillary information to attune the beliefs. With the help of Equation 3, the belief assignments of each biometric classifier are computed. For example, in a two classifier biometric system, we compute $\{m_{c_1}(\theta_{gen}),\ m_{c_1}(\theta_{imp})\}$.

The belief assignments of each biometric classifier are then fused using the proportional conflict redistribution rule [6]. In this rule, redistribution of the conflicts is performed only to the elements which are involved in each conflict and is done according to the proportion/weight of each classifier. The belief assignments of classifier c_1 and c_2 are fused using Equation 4.

$$m_{fused}(i) = m_{c_1}(i)\, m_{c_2}(k) + w_1 \frac{m_{c_1}^2(i)\, m_{c_2}(k)}{m_{c_1}(i) + m_{c_2}(k)} + w_2 \frac{m_{c_2}^2(i)\, m_{c_1}(k)}{m_{c_2}(i) + m_{c_1}(k)} \tag{4}$$

Here $i, k \in \Theta$, $i \neq k$, and w_1 and w_2 are the belief model weight factors ($0 \leq w_1, w_2 \geq 1$). m_{c_1} and m_{c_2} denote the belief assignments of classifier 1 and classifier 2 respectively computed using Equation 3. $\mathbf{m}_{fused}$ is a vector with values $\{m_{fused}(\theta_{gen}),\ m_{fused}(\theta_{imp})\}$[1] representing the

[1] $m_{fused}(\theta_{gen}) = m_{c_1}(\theta_{gen})\, m_{c_2}(\theta_{imp}) + w_1 \frac{m_{c_1}^2(\theta_{gen})\, m_{c_2}(\theta_{imp})}{m_{c_1}(\theta_{gen}) + m_{c_2}(\theta_{imp})} + w_2 \frac{m_{c_2}^2(\theta_{gen})\, m_{c_1}(\theta_{imp})}{m_{c_2}(\theta_{gen}) + m_{c_1}(\theta_{imp})}$,

$m_{fused}(\theta_{imp}) = m_{c_1}(\theta_{imp})\, m_{c_2}(\theta_{gen}) + w_1 \frac{m_{c_1}^2(\theta_{imp})\, m_{c_2}(\theta_{gen})}{m_{c_1}(\theta_{imp}) + m_{c_2}(\theta_{gen})} + w_2 \frac{m_{c_2}^2(\theta_{imp})\, m_{c_1}(\theta_{gen})}{m_{c_2}(\theta_{imp}) + m_{c_1}(\theta_{gen})}$

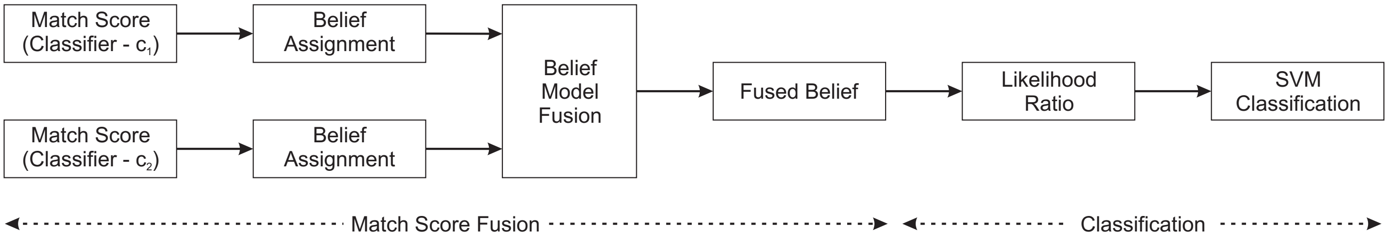

Fig. 1. Block diagram illustrating the steps involved in the proposed sequential match score fusion algorithm.

fused belief. In Equation 4, the first term denotes the degree of conflict between the classifiers and the formulation effectively combines the beliefs of multi-classifier match scores even with conflicts.

B. Classification

For decision making, first the fused belief assignments induced from match scores are converted into the likelihood ratio $R = \left\{\frac{m_{fused}(\theta_{gen})}{m_{fused}(\theta_{imp})}\right\}$. Next, the likelihood ratio is used as input to the SVM classifier for decision making (Equation 5). Utilizing the SVM with likelihood ratio for decision making ensures that the algorithm is less prone to over-fitting and addresses the non-linearity in the biometric match scores.

$$\text{Decision} = \begin{cases} \text{accept} & if \quad \text{SVM}(R) \geq t \\ \text{reject} & \text{otherwise} \end{cases} \tag{5}$$

Here t is the decision threshold chosen for a specific false accept rate (using the concept of SVM regression). The advantage of this approach is its control over the false accept and false reject rates, and it also satisfies the Neyman-Pearson theorem [10] for decision making.

III. Reconciliation of Constituent Biometric Classifiers and Fusion Algorithms

With good quality gallery-probe pair[2], any efficient classifier can verify the identity without the need for fusion. For cases when the two biometric classifiers have minor conflicts, sum rule with min-max normalization [18] can effectively fuse the match scores and yield correct results with very less time complexity. The sequential fusion rule is required to perform fusion when individual classifiers are prone to generate conflicting or ambiguous decisions, i.e., cases with uncertainties and imperfection. In our previous research, we introduced an adaptive framework that reconciles match score fusion algorithms to improve the verification performance both in terms of accuracy and time [24]. The concept behind the framework is to dynamically select an optimal fusion algorithm for the given probe image. *In other words, the algorithm selects a complex fusion algorithm only when there is uncertainty or imperfection in the match scores otherwise it selects a simple fusion algorithm*. In this paper, we extend the framework to reconcile constituent biometric classifiers (e.g. two face recognition algorithms in multi-classifier system) with the proposed sequential fusion algorithm and the sum rule in order to optimize both verification accuracy and computational time. Fig. 2 illustrates the steps involved in the proposed dynamic reconciliation algorithm. The algorithm is explained in context to face recognition but it can be easily transformed for any multimodal biometrics scenario.

[2]The term *gallery-probe pair* is used to represent that in verification mode, a probe is matched with a gallery.

Input to the reconciliation algorithm is a quality vector which is a quantitative representation of biometric information pertaining to the gallery-probe pair. In context to face recognition, the quality vector consists of quality score, visual activity level and pose of the face image. The quality vector $[Q, A, \theta]$ is computed using the following approach:

- To encode the facial edge information and noise present in the image, a redundant discrete wavelet transformation (RDWT) based quality assessment algorithm [25] is used that provides both frequency and spatial information. Face image I of size $n \times n$ is decomposed into three levels of RDWT, i.e. $j = 1, 2, 3$. Let $i = A, H, V, D$ represents the approximation, horizontal, vertical and diagonal subbands. The RDWT decomposition can be written as,

$$[I_{Aj}, I_{Hj}, I_{Vj}, I_{Dj}] = RDWT(I) \tag{6}$$

Image quality score Q is computed using equation 7.

$$Q = \frac{\sum_i a_i b_i}{\sum_i b_i}, \tag{7}$$

where

$$a_i = \sum_{j=1}^{3} ln \sqrt{\left(\frac{\mu_{ij} - \sum_{j=1}^{3} \sum_{x,y=1}^{n,n} I_{ij}(x,y)}{\sigma_{ij}}\right) / n^2} \tag{8}$$

and

$$b_i = \sum_{j=1}^{l} ln \sqrt{\left(\frac{1}{1 + \sum_{x,y=1}^{n,n} \nabla I_{ij}(x,y)}\right) / n^2}. \tag{9}$$

Here, μ_{ij} and σ_{ij} are the mean and standard deviation of the RDWT coefficients of the i^{th} subband and the j^{th} level respectively, and ∇ denotes the gradient operator. Finally, the quality score, Q, is normalized in the range of $[0, 1]$ using min-max normalization [18] (0 represents the worst quality and 1 as best) and used as the first element of the quality vector.

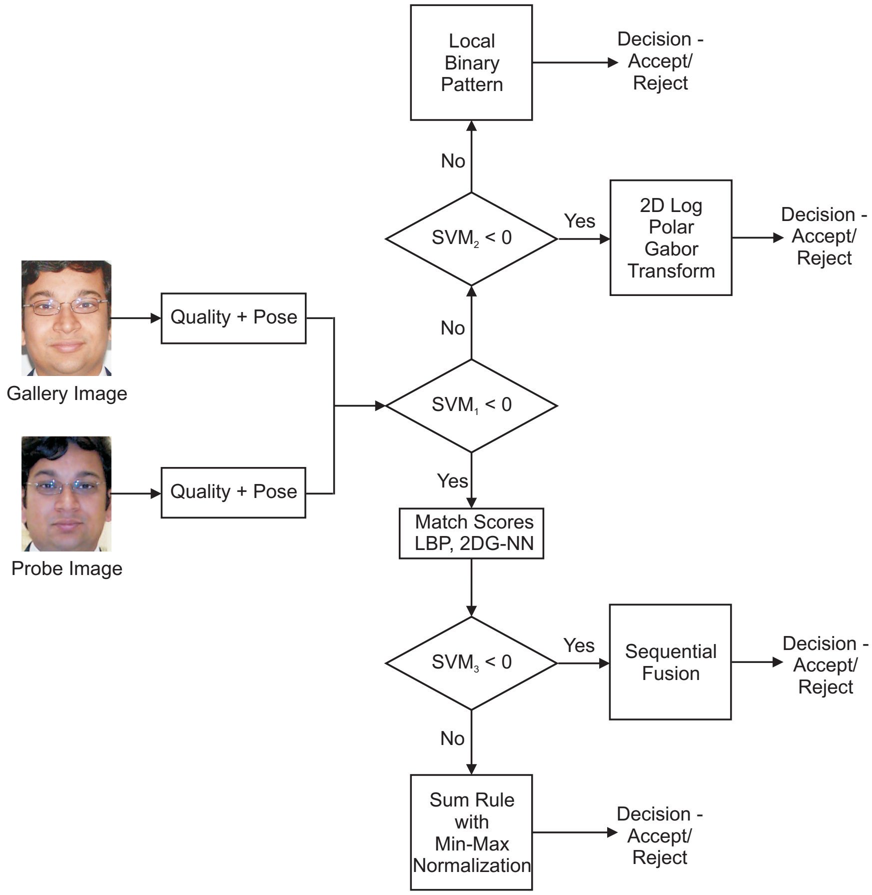

Fig. 2. Reconciling biometric classifiers and fusion algorithms in context to face recognition application.

- Image properties such as brightness and contrast can be encoded using visual activity level which is computed using Equation 10. Activity level A is then normalized in the range of $[0, 1]$ and used as the second element in the quality vector. Higher value of activity level represents properly illuminated and contrast normalized image.

$$A = \sqrt{\frac{1}{n^2}\left[\sum_{i=0}^{n-1}\sum_{j=1}^{n-1}\{(I(i,j) - I(i,j-1)\}^2 + \sum_{j=0}^{n-1}\sum_{i=1}^{n-1}\{(I(i,j) - I(i-1,j)\}^2\right]} \quad (10)$$

- In face recognition, pose variations can reduce the amount of overlapping biometric features required for recognition. Therefore, it is important to include the head position or angle as pose parameter in the quality vector. In this research, a fast single view algorithm [13] is used for estimating the pose of a face image. This algorithm [13] starts with active appearance model for landmark feature extraction. A statistical anthropometric model, in combination with *pose from orthography and scaling iterations* scheme, uses these features for pose angle estimation. The output of the algorithm is pose angle θ which serves as the third element in quality vector.

Fig. 3 shows examples of image quality vector on LFW face database [9]. In the reconciliation algorithm if quality of gallery-probe pair is high then the constituent uniclassifier algorithms can be used, otherwise the fusion rules are chosen. The proposed algorithm uses three SVMs to reconcile two uniclassifiers and two fusion algorithms. In this research, we use local binary pattern (LBP) [3] and 2D log polar Gabor transform (2DG-NN) [20] based face recognition algorithms as uniclassifiers and sum rule with min/max normalization and the proposed sequential fusion as two fusion algorithms. As shown in Fig. 2, first SVM, denoted as SVM_1, is used to select uniclassifiers or fusion rules. If uniclassifiers are selected then the second SVM, denoted as SVM_2, is used to choose between LBP and 2DG-NN. If the option pertaining to fusion rules is selected then match scores from LBP and 2DG-NN are computed and the third SVM, denoted as SVM_3, is used to select between sum rule or sequential fusion. The reconciliation algorithm is divided into two stages: training SVMs for reconciliation and dynamic selection for every query instance.

Training SVMs for Reconciliation: Three SVMs are independently trained using the labeled training database. The training procedure is explained below.

(1) SVM_1 is trained using the labeled training data $\{\mathbf{x_1}_i, y_{1i}\}$. Here, $\mathbf{x_1}_i$ is the quality vector belonging to the i^{th} training gallery-probe pair, i.e. $\{Q_{Gi}, A_{Gi}, \theta_{Gi}, Q_{Pi}, A_{Pi}, \theta_{Pi}\}$. $y_{1i} \in (+1, -1)$ is the respective label such that +1 is assigned when gallery-probe pair is of high quality and can be correctly matched using uniclassifiers and -1 is assigned to the data

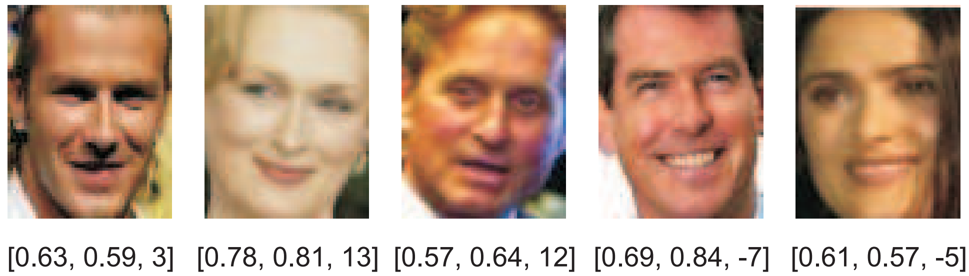

Fig. 3. Illustrating examples of quality vector on images from LFW database [9].

that requires match score fusion. At the end of the training, a non-linear decision hyperplane is learned that can perform classification for selecting either uniclassifier algorithms or match score fusion.

(2) SVM_2 is trained using the labeled training data $\{\mathbf{x_2}_i, y_{2i}\}$ where, $\mathbf{x_2}_i$ is the quality vector belonging to the i^{th} training gallery-probe pair and $y_{1i} \in (+1, -1)$. In this labeling, +1 belongs to gallery-probe pair that can be matched using LBP classifiers and -1 is assigned to the data that requires matching with 2DG-NN classifier. In this training, a non-linear decision hyperplane is learned that can perform dynamic selection of either LBP or 2DG-NN.

(3) SVM_3 is trained using the labeled training data $\{\mathbf{x_3}_i, y_{3i}\}$. Here, $\mathbf{x_3}_i$ is the i^{th} training data vector that contains match scores and verification accuracy priors pertaining to two uniclassifiers, and $y_i \in (+1, -1)$ is the label such that +1 belongs to match scores that should be fused using the sum rule with min-max normalization and -1 belongs to the match scores that should be fused using the sequential fusion algorithm. The SVM is trained such that the output of $SVM_3 > 0$ denotes the use of sum rule otherwise the sequential fusion algorithm is used.

Dynamic Selection at Probe level for Reconciliation: For probe verification, the trained SVMs are used to dynamically select the most appropriate algorithm depending on the quality vector.

1) The quality vectors pertaining to both the gallery-probe images are provided as input to the trained SVMs. The SVM_1 classifier selects between uniclassifier and fusion.
2) Depending on the classification result of SVM_1 classifier, SVM_2 and SVM_3 are used to select one of the four options: (1) LBP, (2) 2DG-NN, (3) Sum rule with min/max normalization, and (4) sequential fusion.

IV. Reducing the Effect of Covariate Factors in Face Recognition using Match Score Fusion

There are several global, local, non-linear, appearance-based, texture-based, and feature-based face recognition algorithms [11], [26], [28]. These algorithms independently attempt to reduce the effect of covariate factors such as expression, illumination, pose, and occlusion on the recognition performance. However, most of the existing algorithms are optimized for particular covariates only. For example, neural network architecture based 2D log polar Gabor transform algorithm [20] can tolerate variations in expression, illumination, and occlusion whereas local facial features can handle pose and expression variations. It is our hypothesis that the performance of a face recognition system can be greatly enhanced if information from multiple algorithms are fused and a final decision is obtained using the fused information. In this section, we use the sequential fusion and reconciliation algorithms to fuse the match scores computed from a non-linear face recognition algorithm and a local facial feature based algorithm to mitigate the effect of covariate factors.

As shown in Fig. 4, two face classifiers (c_1 and c_2) are used for feature extraction and matching. The match scores computed using these classifiers are combined using the proposed sequential fusion and reconciliation algorithms. First, the face region from the input image is detected using the triangle based face detection algorithm [21] (the size of detected face image is 128×96) and then following algorithms are used for feature extraction and matching.

- **Neural Network Architecture based 2D Log Polar Gabor Transform:** The face image is transformed into polar coordinates and phase features are extracted using the neural network architecture based 2D log polar Gabor transform [20]. These features are matched using Hamming distance to generate the match scores.
- **Local Binary Pattern:** The face image is divided into several regions and weighted Local Binary Pattern features are extracted to generate a feature vector [3]. Matching of two LBP feature vectors is performed using weighted χ^2 distance measure algorithm.

A. Face Databases used for Evaluation

To evaluate the performance on a large database with challenging intra-class variations, we combined images from multiple face databases to create a heterogeneous database of more than *116,000* images pertaining to 1194 subjects. Table 1 lists the databases used and the number of subjects selected from the individual databases. The CMU-AMP database[3] contains images with large expression variations while the CMU-PIE dataset [19] contains images with variations in pose, illumination and facial expressions. The Equinox database[4] has images captured under different illumination conditions with accessories and expressions. The AR face database [12]

[3]http://amp.ece.cmu.edu/projects/FaceAuthentication/~download.htm
[4]http://www.equinoxsensors.com/products/HID.html

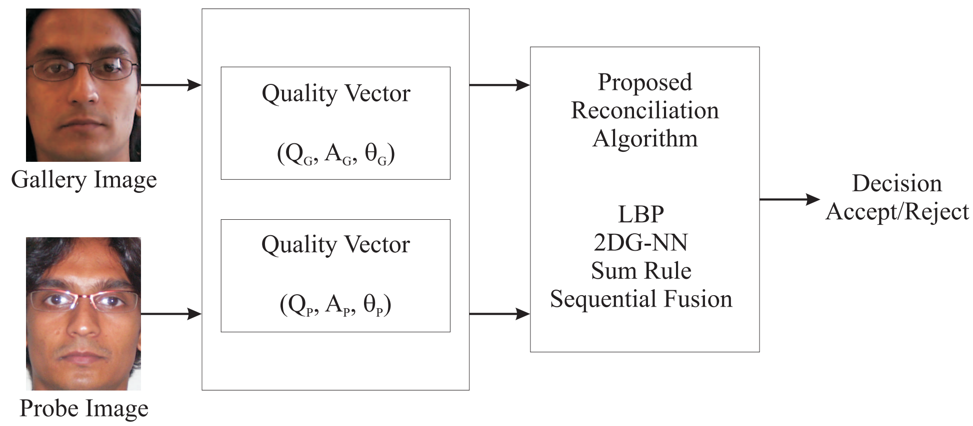

Fig. 4. Illustrating the steps involved in match score fusion of multi-classifier face recognition.

contains face images with varying illumination and accessories, and the FERET database [17] has face images with different variations over a time interval of 3-4 years. The Notre Dame face database [8] comprises of images with different lighting and facial expressions over a period of one year. The Labeled Faces in the Wild database [9] contains real world images of celebrities and popular individuals (this database contains images of more than 1600 subjects from which we selected 294 subjects that have at least 6 images). To the best of our knowledge, there is no database available in public domain which encompasses such wide range of intra-class variations. The images are partitioned into two non-overlapping sets: (1) the training dataset is used to train the individual classifiers (i.e., 2DG-NN, LBP, SVM classifiers) and the fusion algorithms, and (2) the gallery-probe dataset (the test set) is used to evaluate the performance of the fusion algorithms. The training set comprises of randomly selected five images of each subject (i.e. 5970 images in training) and the remaining images (over 110,000) are used as the test data to evaluate verification performance of the algorithms. Fig. 5 shows sample images in training dataset and gallery-probe dataset. This train-test partitioning is repeated 10 times (cross validation) and Receiver Operating Characteristics (ROC) curves are generated by computing the genuine accept rates (GAR) over these trials at different false accept rates (FAR).

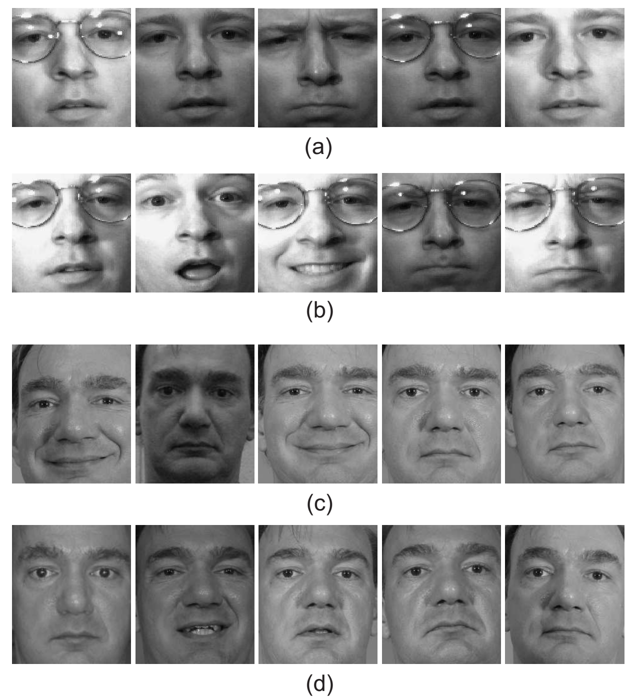

Fig. 5. Illustrating the examples of non-overlapping training and gallery-probe datasets: (a) training images from the Equinox database, (b) gallery-probe images from the Equinox database, (c) training images from the Notre Dame database, and (d) gallery-probe images from the Notre Dame database.

B. Performance Evaluation

The training data is first used to train the proposed fusion algorithm and reconciliation algorithm. For the sequential fusion algorithm, verification accuracy priors, density estimation parameters, belief model weights w_1 and w_2, and SVM parameters are computed using the training data. Note that in sequential fusion algorithm training, we use the labelled training match scores where labels are *genuine* and *impostor*. Unimodal classifier precision on training dataset is used as the verification accuracy prior. To compute other fusion parameters, we perform experiments with all possible combinations of parameters, i.e. training or optimization of parameters is performed globally. The values of parameters, including SVM kernel parameter (γ in RBF kernel[5]), that provide the maximum verification performance on training data are chosen for testing. Similarly, the reconciliation algorithm is trained using labelled data as described in Section 3. The training database is also used to train the LBP and 2DG-NN face recognition algorithms. Further, the performance of sequential fusion algorithm is compared with Sum rule with min-max normalization [18], SVM fusion [2], and PLR fusion [16] with recursive algorithm for density estimation [29].

The ROC plot in Fig. 6 shows comparative results of the LBP and 2DG-NN face verification algorithms, and the improvement due to match score fusion algorithms. The 2DG-NN

[5]RBF parameter $\gamma = 8$ shows the best performance.

TABLE I
COMPOSITION OF THE HETEROGENEOUS FACE DATABASE OF 1194 SUBJECTS.

Face Database	Number of Subjects	No. of Images per Subject	Covariates
CMU-AMP	13	75	Expression
CMU - PIE	65	≥ 600	Pose, Illumination and Expression
Equinox	90	$\geq$200	Illumination, Expression and Occlusion (glasses)
AR	120	≥ 26	Illumination, Expression, and Occlusion
FERET	300	≥ 6	Pose, Illumination, Expression and Occlusion
Notre Dame	312	$min - 6$, $max - 227$	Pose, Illumination and Expression
Labeled Faces in the Wild	294	≥ 6	Pose, Illumination, Expression and Occlusion

classifier yields around 82% verification accuracy at 0.01% FAR and outperforms the LBP classifier by around 9%. The performance of face verification improves by $\sim$ 5-13% when match scores are fused using the fusion algorithms. Among all the fusion algorithms, the proposed sequential fusion approach yields an accuracy of 94.36% and the reconciliation algorithm yields the best verification accuracy of 94.98%.

Experiments are also performed to evaluate the effect of covariate factors (viz. expression, illumination, pose, and occlusion) on the performance of face verification. This experiment facilitates the comparative analysis of face verification algorithms and the subsequent improvement by deploying the proposed match score fusion. The results and their analysis are summarized below:

- Scatter plot in Fig. 7 and experimental results show that the match scores obtained from 2DG-NN algorithm (non-linear) and LBP algorithm (local features) can be fused to significantly improve the verification accuracy. Further, covariate analysis in Table 2 suggests that 2DG-NN algorithm provides good performance inspite of variations in expression, illumination, and occlusion whereas LBP algorithm can better tolerate variations in expression and pose. Covariate analysis also indicates that variations in pose and occlusion cause a larger reduction in verification accuracy compared to expression and illumination variations.
- In our experiments, we observed that the sum rule with min-max normalization is not able to handle most of the conflicting cases which are caused due to intra-personal variations. Furthermore, during cross validation trials, we observed that the difference between minimum and maximum Half Total Error Rates (HTER = $\frac{FAR+FRR}{2}$) [5] for sum rule is very large (Table 3). This shows that the sum rule with min-max normalization is not able to handle disparities in the training-testing datasets.
- Tables 2 and 3 suggest that the PLR fusion yields better performance compared to SVM fusion both in terms of accuracy and stability across different cross validation trials. We also observed that PLR fusion has the advantage of being generalized whereas SVM fusion algorithm can handle the non-linearities in the match score.
- The sequential fusion algorithm effectively improves the verification accuracy. The algorithm transforms the match scores into probabilistic entities thereby making them robust to sensor noise and matcher limitations. Multi-classifier match score fusion is performed using the proportional conflict redistribution rule that can handle uncertainties in the biometric match scores. Finally, a decision is made using the likelihood ratio induced SVM classifier that satisfies the Neyman-Pearson theorem [10]. Further, t-test at 95% confidence suggests that the sequential fusion algorithm is significantly different than other fusion algorithms. The HTER test also shows that the sequential fusion is *stable* across all cross validation trails whereas the HTERs pertaining to other fusion algorithms vary considerably.
- If the classifiers are in accordance (for example, Fig. 8(a) shows a case when both LBP and 2DG-NN accept the subject), all the fusion rules provide correct results. Further, Figs. 8 (b) and (c) show sample cases when two classifiers are in conflict but the proposed sequential fusion algorithm correctly accepts the subjects and existing fusion algorithms (sum rule, SVM fusion, and PLR fusion) provide incorrect results. Finally, there are few cases (shown in Fig. 8(d)) when both the classifiers reject a genuine subject. In such cases, fusion algorithms cannot do much to improve the performance, therefore 100% accuracy is not achieved.
- The time complexity of the proposed fusion approach is also reasonable when compared with existing fusion algorithms. On a 2 GHz Pentium Duo Core processor with 2 GB RAM under MATLAB environment, the proposed algorithm requires around 3.6 seconds for facial feature extraction, matching, fusion and decision-making, whereas existing fusion algorithms require 1.7-2.8 seconds.
- The reconciliation algorithm that unifies LBP and 2DG-NN recognition algorithms, sum rule, and sequential match score fusion algorithm yields the best verification accuracy. Although, t-test at 95% confidence suggests that the reconciliation algorithm is not significantly different from the sequential fusion, the advantage of the reconciliation algorithm is computational time and stability (HTER test). As shown in Tables 2 and 3, computational cost of the reconciliation algorithm is similar to sum rule but it provides the relative performance gain of more than 60%[6].
- For cases in which the quality of the gallery-probe pair is good and pose variation is minimum, the 2DG-NN algorithm is selected. The LBP technique is selected for feature extraction and matching when images have pose variations but the quality is good. The fusion rules are selected when image quality is moderate to poor,

[6]relative performance gain = $\frac{accuracy_{max} - accuracy_{min}}{100 - accuracy_{min}} \times 100$

TABLE II
COVARIATE ANALYSIS OF FACE RECOGNITION ALGORITHMS AND MATCH SCORE FUSION ALGORITHMS.

Covariate	Verification Accuracy (%) at 0.01% FAR						
	LBP	2DG-NN	Sum Rule [18]	SVM Fusion [2]	PLR Fusion [16]	Sequential Fusion	Reconciliation
Expression	88.10	88.03	91.58	93.63	93.95	95.32	**95.91**
Illumination	84.52	86.26	91.80	94.86	95.43	96.83	**96.86**
Pose	73.14	70.12	82.31	85.74	85.96	89.07	**89.52**
Occlusion	65.52	83.10	88.76	89.12	90.14	94.51	**94.76**
Overall	73.42	82.01	87.39	89.44	91.43	94.36	**94.98**

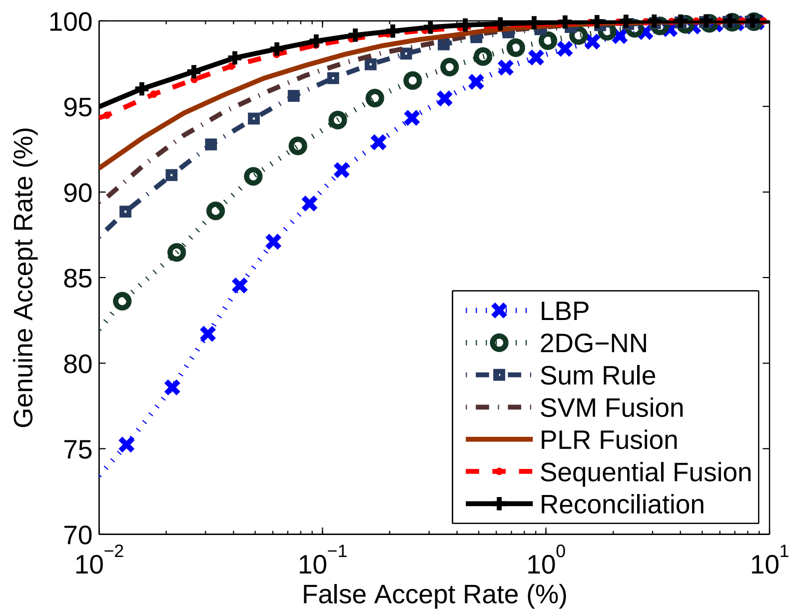

Fig. 6. ROC of the constituent unimodal face matchers, the proposed sequential fusion, reconciliation algorithm, and comparison with existing fusion algorithms.

gallery-probe pairs have large variations in pose, or facial features are occluded using cap/hat, scarf and glasses. Further, sum rule is chosen when intra-personal variations are not very large and match scores exhibit minor conflict. On the other hand, sequential fusion algorithm is selected for cases with large intra-personal variations. In the experiments, we observed that when the quality is good ($Q \geq 0.7$ and $A \geq 0.7$) and difference in gallery probe pose angles is less ($\leq 10^0$), about 98% of times both the constituent uniclassifier algorithms are in accordance. Overall, we found that around 38% times LBP or 2DG-NN algorithms are chosen, 44% times sum rule with min/max normalization is selected and 18% times sequential fusion algorithm is selected.

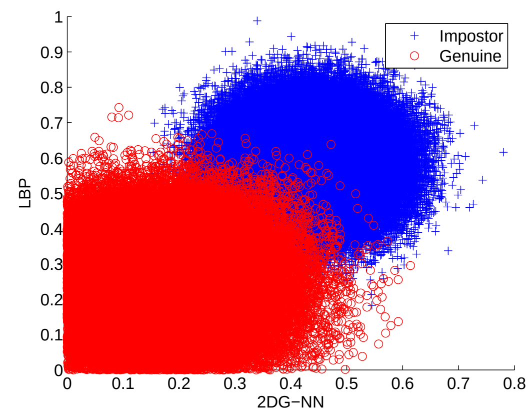

Fig. 7. Scatter plot shows that there is limited correlation between match scores obtained from LBP and 2DG-NN face recognition algorithms and hence match score fusion can improve the performance.

V. CONCLUSION

The performance of score-level fusion algorithms is often affected by conflicting decisions generated by the constituent matchers/classifiers. Further, the computational cost of fusion algorithms that can address conflicting scores increases drastically. This paper presents algorithms to optimize both verification accuracy and computation time. We first proposed a sequential fusion algorithm by incorporating the likelihood ratio test statistic in a support vector machine framework in order to classify match scores originating from multiple matchers. The proposed fusion algorithm takes into account the precision and uncertainties of individual matchers. We also presented a reconciliation algorithm that unifies the constituent classifiers (or matchers) with the fusion schemes in order

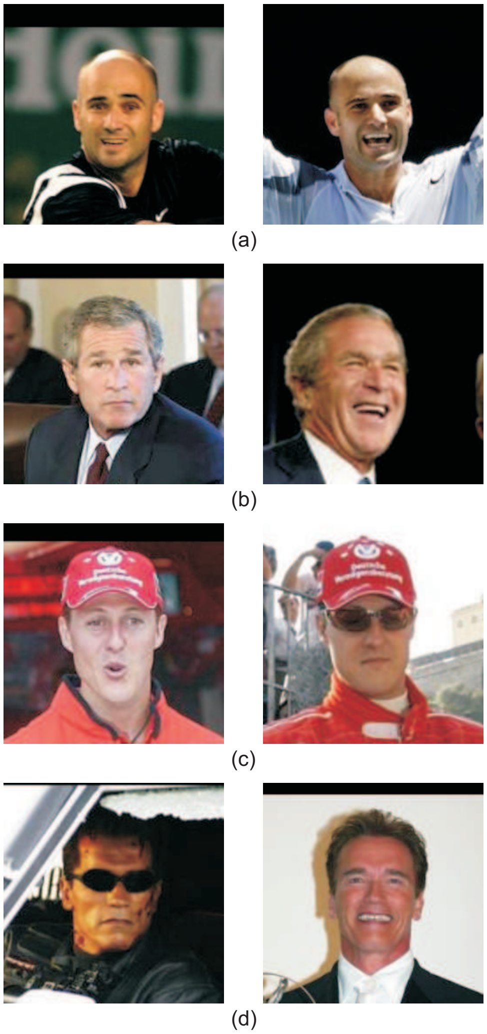

Fig. 8. Sample cases from the Labeled Faces in the Wild database [9] when both LBP and 2DG-NN face verification algorithms (a) are in accordance to accept the genuine subject, (b) and (c) are in conflict, and (d) are in accordance to reject the genuine subject.

TABLE III
COMPARISON OF FUSION ALGORITHMS IN TERMS OF COMPUTATION TIME AND HALF TOTAL ERROR RATE.

Algorithms	Average Time (seconds)	HTER [Max., Min.]
LPB [3]	0.7	[20.34, 9.51]
2DG-NN [20]	0.9	[14.61, 5.28]
Sum rule [18]	1.7	[9.02, 3.37]
SVM Fusion [2]	2.8	[7.49, 2.95]
PLR Fusion [16]	2.5	[7.33, 2.41]
Sequential Fusion	3.6	[5.81, 2.12]
Reconciliation	1.9	[4.35, 1.99]

to optimize recognition accuracy and computational time. Depending on the quality of input biometric data, the proposed reconciliation algorithm selects among the unimodal classifiers and fusion rules to recognize an individual. The resulting algorithms are used to mitigate the effect of covariate factors in face recognition by combining the match scores obtained from two face recognition algorithms: the local binary pattern encoding scheme and 2D log polar Gabor transform based encoding scheme. Experimental results on a heterogeneous face database of 1,194 subjects suggest that the proposed algorithms can significantly improve the verification performance of a face recognition system with low computational overhead. In future, we plan to extend the sequential fusion algorithm to include more parameters in the face quality assessment algorithm [1]. The sequential fusion and reconciliation algorithms can also be extended for multimodal biometrics, e.g. match score fusion of face, fingerprint and iris biometrics.

VI. ACKNOWLEDGEMENT

The authors would like to acknowledge the reviewers for their constructive and insightful comments. The authors would also like to thank CVRL University of Notre Dame, NIST, Robotics Institute CMU, CMU AMP Research Lab, Dr. A.R. Martinez, Dr. E.G.L. Miller and Equinox Corporation for granting us access to the face databases used in this research. Vatsa, Singh and Noore were supported in part through research grants from the National Institute of Justice (Award No. 2003-RC-CX-K001) and NSF-CITeR. Ross was supported by NSF CAREER Grant No. IIS 0642554.

REFERENCES

[1] M. Abdel-Mottaleb and M. Mahoor. Algorithms for assessing the quality of facial images. *Computational Intelligence Magazine, IEEE*, 2(2):10 –17, may 2007.
[2] J. F. Aguilar, J. O. Garcia, J. G. Rodriguez, and J. Bigun. Discriminative multimodal biometric authentication based on quality measures. *Pattern Recognition*, 38(5):777–779, 2005.
[3] T. Ahonen, A. Hadid, and M. Pietikinen. Face description with local binary patterns: application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.
[4] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz. Fusion of face and speech data for person identity verification. *IEEE Transactions on Neural Networks*, 10(5):1065–1074, 1999.
[5] S. Bengio and J. Mari'ethoz. A statistical significance test for person authentication. In *Proceedings of Odyssey: The Speaker and Language Recognition Workshop*, pages 237–244, 2004.
[6] J. Dezert and F. Smarandache. Introduction to the fusion of quantitative and qualitative beliefs. *Information and Security*, 20:9–49, 2006.
[7] R. O. Duda, P. E. Hart, and D. Stork. *Pattern Classification*. John Wiley and Sons, 2001.
[8] P. J. Flynn, K. W. Bowyer, and P. J. Phillips. Assessment of time dependency in face recognition: an initial study. In *Proceedings of Audio- and Video-Based Biometric Person Authentication*, pages 44–51, 2003.
[9] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments, 2007. University of Massachusetts, Amherst, Technical Report.
[10] E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses*. Springer, New York, 3rd edition, 2005.
[11] S. Z. Li and A. K. Jain. *Handbook of Face Recognition*. Springer, 2005.
[12] A. R. Martinez and R. Benavente. The AR face database, 1998. Computer Vision Center, Technical Report.
[13] P. Martins and J. Batista. Monocular head pose estimation. In *Proceedings of International Conference on Image Analysis and Recognition*, pages 357–368, 2008.

[14] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(8):1778 – 1790, aug. 2004.

[15] S. Mukherjee and V. Vapnik. Multivariate density estimation: an svm approach. Technical report, Cambridge, MA, USA, 1999.

[16] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain. Likelihood ratio based biometric score fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):342–347, 2008.

[17] P. J. Phillips, H. Moon, S. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[18] A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of multibiometrics*. Springer, New York, 1st edition, 2006.

[19] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[20] R. Singh, M. Vatsa, and A. Noore. Face recognition with disguise and single gallery images. *Image and Vision Computing*, 27(3):293–304, 2009.

[21] S. K. Singh, D. S. Chauhan, M. Vatsa, and R. Singh. A robust skin color based face detection algorithm. *International Tamkang Journal of Science and Engineering*, 6(4):227–234, 2003.

[22] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain. Large scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:450–455, 2005.

[23] V. Vapnik. *The nature of statistical learning theory*. Springer, 1995.

[24] M. Vatsa, R. Singh, and A. Noore. Unification of evidence theoretic fusion algorithms: A case study in level-2 and level-3 fingerprint features. *In IEEE Transactions on Systems, Man, and Cybernetics-A*, 39(1):3–13, 2009.

[25] M. Vatsa, R. Singh, A. Noore, and S. K. Singh. Quality induced fingerprint identification using extended feature set. In *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, 2008.

[26] H. Wechsler. *Reliable Face Recognition Methods: System Design, Implementation and Evaluation*. Springer, 2006.

[27] B. Zadrozny. Learning and evaluating classifiers under sample selection bias. In *Proceedings of the twenty-first international conference on Machine learning*, pages 903–910. ACM, 2004.

[28] W.-Y. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Survey*, 35(4):399–458, 2003.

[29] Z. Zivkovic and F. V. D. Heijden. Recursive unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):651–656, 2004.

Mayank Vatsa (S'04, M'09) received the M.S. and Ph.D. degrees in computer science in 2005 and 2008, respectively from the West Virginia University, Morgantown, USA.

He is currently an Assistant Professor at the Indraprastha Institute of Information Technology (IIIT) Delhi, India. He has more than 90 publications in refereed journals, book chapters, and conferences. His areas of interest are biometrics, image processing, computer vision, uncertainty principles, watermarking, and information fusion.

Dr. Vatsa is a member of the IEEE, Computer Society and Association for Computing Machinery. He is also a member of the Golden Key International, Phi Kappa Phi, Tau Beta Pi, Sigma Xi, Upsilon Pi Epsilon, and Eta Kappa Nu honor societies. He was the recipient of six best paper and best poster awards.

Richa Singh (S'04, M'09) received the M.S. and Ph.D. degrees in computer science in 2005 and 2008, respectively from the West Virginia University, Morgantown, USA.

She is currently an Assistant Professor at the Indraprastha Institute of Information Technology (IIIT) Delhi, India. Her areas of interest are biometrics, pattern recognition, machine learning, and granular computing. She has more than 90 publications in refereed journals, book chapters, and conferences.

Dr. Singh is a member of the IEEE, Computer Society and the Association for Computing Machinery. She is also a member of the Golden Key International, Phi Kappa Phi, Tau Beta Pi, Upsilon Pi Epsilon, and Eta Kappa Nu honor societies. She was the recipient of six best paper and best poster awards.

Afzel Noore (M'03) received the Ph.D. degree in electrical engineering from the West Virginia University, Morgantown, USA.

He was a Digital Design Engineer with Philips, India. From 1996 to 2003, he was the Associate Dean for Academic Affairs and Special Assistant to the Dean with the College of Engineering and Mineral Resources, West Virginia University, where he is currently a Professor with the Lane Department of Computer Science and Electrical Engineering. His research has been funded by NASA, the National Science Foundation, Westinghouse, General Electric, Electric Power Research Institute, the U.S. Department of Energy, the U.S. Department of Justice, and the U.S. Department of Defense.

He serves on the editorial boards of Recent Patents on Engineering and the Open Nanoscience Journal. He has over 100 publications in refereed journals, book chapters, and conferences. His research interests include computational intelligence, biometrics, software reliability modeling, machine learning, hardware description languages, and quantum computing. Dr. Noore is a member of the Phi Kappa Phi, Sigma Xi, Eta Kappa Nu, and Tau Beta Pi honor societies. He was the recipient of six best paper awards.

Arun Ross (S'00, M'03) received the B.E. (Hons.) degree in computer science from the Birla Institute of Technology and Science, Pilani, India, in 1996, and the M.S. and Ph.D. degrees in computer science and engineering from Michigan State University, East Lansing, in 1999 and 2003, respectively.

Between 1996 and 1997, he was with the Design and Development Group of Tata Elxsi (India) Ltd., Bangalore, India. He also spent three summers (2000-2002) with the Imaging and Visualization Group of Siemens Corporate Research, Inc., Princeton, NJ, working on fingerprint recognition algorithms. He is currently an Associate Professor in the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown. His research interests include pattern recognition, classifier fusion, machine learning, computer vision, and biometrics. He is actively involved in the development of biometrics and pattern recognition curricula at West Virginia University. He is the coauthor of the Handbook of Multibiometrics and coeditor of the Handbook of Biometrics.

Dr. Ross is a recipient of NSFs CAREER Award and was designated a Kavli Frontier Fellow by the National Academy of Sciences in 2006. He is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.