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 consolidates the output of multiple biometric classifiers to render a decision about the identity of an individual. We consider the problem of designing a fusion scheme when 1) the number of training samples is limited, thereby affecting the use of a purely density-based scheme and the likelihood ratio test statistic; 2) the output of multiple matchers yields conflicting results; and 3) 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: 1) 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 2) the design of a dynamic selection algorithm that unifies the constituent classifiers and fusion schemes in order to optimize both verification accuracy and computational cost. The case study in multiclassifier face recognition suggests that the proposed algorithm can address the issues listed above. 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, face verification, match score fusion, support vector machine (SVM).**

I. Introduction

THE paradigm of information fusion, which 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 evidence in order to deduce or verify human identity is often referred to as *multibiometrics*. While biometric 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 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 suboptimal fusion rules; the use of nonparametric 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 overtraining 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 fusion 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

Manuscript received November 15, 2009; revised June 14, 2010; accepted June 16, 2010. Date of publication July 08, 2010; date of current version August 13, 2010. This work was supported in part by research grants from the Army Research Laboratory (Award W911NF-10-02-0021). The work of A. Ross was also supported by NSF CAREER Grant IIS 0642554. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Davide Maltoni.

M. Vatsa and R. Singh are with the Indraprastha Institute of Information Technology (IIIT) Delhi, New Delhi 110078, India (e-mail: mayank@iiitd.ac.in; rsingh@iiitd.ac.in).

A. Noore and A. Ross are with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26505 USA (e-mail: afzel.noore@mail.wvu.edu; arun.ross@mail.wvu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2010.2056683

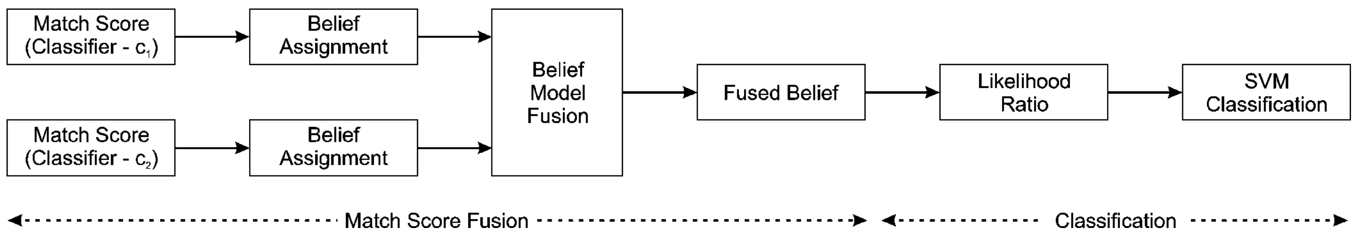

Fig. 1. Block diagram illustrating the steps involved in the proposed sequential match score fusion algorithm.

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 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 minimum/maximum (min/max) normalization can be used for most of the probe cases when multiclassifier biometric output is not highly conflicting. One way to improve the verification accuracy, without increasing the computational cost, 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, biclassifier biometric system, the dynamic selection 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 (LBF) [3] and neural network architecture-based 2-D log polar Gabor transform (2DG-NN) [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, a belief model is used for fusion and finally, the likelihood ratio test statistic and SVM are used for classification. 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 $\Theta = \{\theta_{\text{gen}}, \theta_{\text{imp}}\}$, where θ_{gen} represents the genuine class and θ_{imp} represents the impostor class. The first step in the sequential fusion algorithm is to transform match scores into belief assignments. 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]. The multivariate Gaussian density function [7] can be written as

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

where $\mathbf{x}$ is a vector with d components, $\boldsymbol{\mu}$ is the mean vector, and $\boldsymbol{\Sigma}$ is the covariance matrix. Let $C(\mathbf{x}, i)$ be the conditional joint density of d match scores and $i \in \Theta$. $C(\mathbf{x}, i)$ is computed using

$$C(\mathbf{x}, i) = \sum_{j=1}^{M(i)} W(i,j) p(\mathbf{x}, \boldsymbol{\mu}(i,j), \boldsymbol{\Sigma}(i,j)) \quad (2)$$

where $\boldsymbol{\mu}(i,j)$, $\boldsymbol{\Sigma}(i,j)$, and $W(i,j)$ are the mean vector, covariance matrix, and weight factor, respectively, corresponding to the jth mixture component in the conditional joint density. Also, $\sum_{j=1}^{M(i)} W(i,j) = 1$ and $M(i)$ is the number of mixture components used to model the density. A recursive algorithm [29] is used to estimate the parameters of the mixture model.

Let $\mathbf{x} = (x_1, \ldots, x_d)$ be the match score vector, where x_n is the match score computed by the nth biometric classifier or matcher. To mitigate the effect of curse-of-dimensionality and for faster computation, we assume independence among constituent matchers and compute the marginal density $C_n(x_n, i)$ of the nth classifier. The belief assignment m_n for the nth classifier is computed using

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

where $\alpha_n(i)$ is the verification accuracy prior of the nth classifier that is used as the ancillary information to estimate the beliefs. With the help of (3), the belief assignments for individual biometric classifiers are computed. For example, in a two-class two-classifier biometric system, we compute $\{m_{c_1}(\theta_{\text{gen}}), m_{c_1}(\theta_{\text{imp}})\}$ and $\{m_{c_2}(\theta_{\text{gen}}),\ m_{c_2}(\theta_{\text{imp}})\}$.

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

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

Here $i,\, k \in \Theta$, $i \neq k$, and w_1 and w_2 are the belief model weight factors $(0 \leq w_1, w_2 \leq 1)$. m_{c_1} and m_{c_2} denote the belief assignments of classifier 1 and classifier 2, respectively, computed using (3). $\mathbf{m}_{\text{fused}}$ is a vector with values $\{m_{\text{fused}}(\theta_{\text{gen}}), m_{\text{fused}}(\theta_{\text{imp}})\}$[1] representing the fused belief. In (4), the first term denotes the degree of conflict between the classifiers and the formulation effectively combines the beliefs of multiclassifier match scores.

B. Classification

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

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

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 criteria [10] for decision making.

III. Dynamic Selection of Constituent Biometric Classifiers and Fusion Algorithms

When encountering a good quality gallery-probe pair,[2] an efficient classifier can verify the identity without the need for fusion. For cases when the two biometric classifiers have minor conflicts, the sum rule with min/max normalization [18] can effectively fuse the match scores and yield correct results with much less time complexity. The sequential fusion rule is used to perform fusion when individual classifiers are prone to generate conflicting or ambiguous decisions, i.e., cases involving uncertainties. 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 in the constituent 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 a multiclassifier 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 selection algorithm. The algorithm is explained in the context of face recognition but it can be easily generalized to any multibiometric scenario.

Input to the dynamic selection algorithm is a quality vector which is a quantitative representation of biometric information pertaining to the gallery-probe pair. In the context of 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. A face image I of size $n \times n$ is decomposed into three levels of the RDWT, i.e., $j = 1, 2, 3$. Let $i = A, H, V, D$ represent the approximation, horizontal, vertical, and diagonal subbands, respectively. The RDWT decomposition can be written as

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

The image quality score Q is computed using (7).

$$Q = \frac{\sum_i a_i b_i}{\sum_i b_i}, \quad (7)$$

where

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

and

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

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

- Image properties such as brightness and contrast can be encoded using the visual activity level which is computed

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

[2] The term *gallery-probe pair* is used to denote that, in the verification mode, a probe is compared against a gallery.

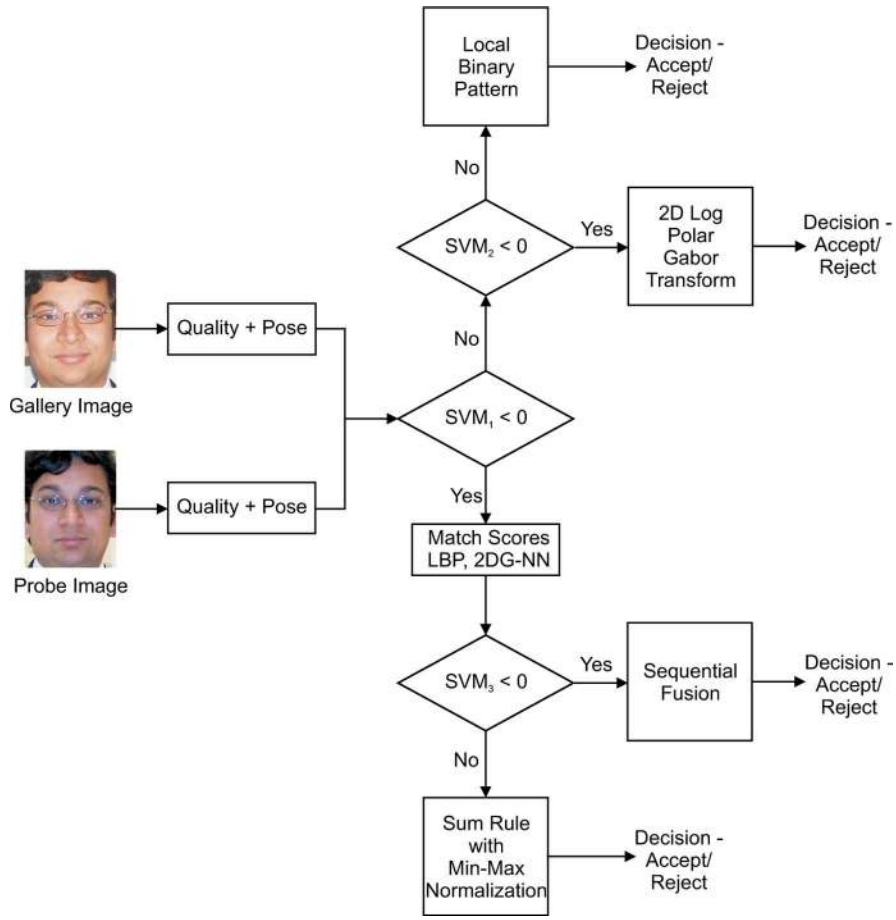

Fig. 2. Dynamic selection of biometric classifiers and fusion algorithms in the context of a face recognition application.

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

using (10), shown at the bottom of the page. Activity level A is then normalized in the range $[0,1]$ and used as the second element in the quality vector. A higher activity level represents properly illuminated and contrast normalized image.

- 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 a 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. The output of the algorithm is the pose angle θ which serves as the third element in the quality vector.

Fig. 3 shows examples of the image quality vector on the LFW face database [9]. In the dynamic selection algorithm, if the quality of gallery-probe pair is high then the constituent classifiers are used; if not, the fusion rules are chosen. The proposed algorithm uses three SVMs to select from the two classifiers

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

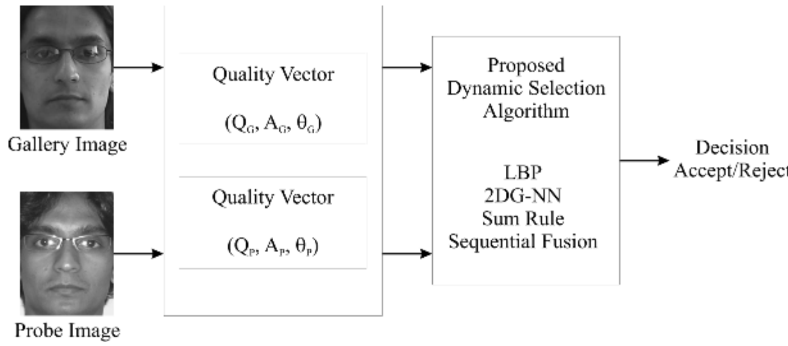

Fig. 4. Illustrating the steps involved in match score fusion of a multiclassifier face recognition system.

and the two fusion algorithms. In this research, we use LBP [3] and 2DG-NN [20] based face recognition algorithms as the constituent classifiers, and the sum rule with min/max normalization and the proposed sequential fusion as the two fusion algorithms. As shown in Fig. 2, the first SVM, denoted as SVM_1, is used to select between the classifiers and the fusion rules. If the classifiers are selected, then the second SVM, denoted as SVM_2, is used to choose between LBP and 2DG-NN face recognition algorithms. If the option pertaining to fusion rules is selected, then the match scores from LBP and 2DG-NN are computed and the third SVM, denoted as SVM_3, is used to select between the sum rule and sequential fusion. The dynamic selection algorithm is divided into two stages: training the SVMs and dynamic selection of algorithms for every probe instance.

1) Training SVMs: Three SVMs are independently trained using the labeled training database. The training procedure is explained as follows.

a) 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 ith 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 the gallery-probe pair is of high quality and can be correctly matched using individual classifiers and -1 is assigned to the pair that requires match score fusion. At the end of the training stage, a nonlinear decision hyperplane is learned that can select between the individual classifiers and match score fusion.

b) 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 ith training gallery-probe pair and $y_{1i} \in (+1, -1)$. In this case, $+1$ indicates the gallery-probe pair that can be matched using the LBP classifier and -1 is assigned to the data that requires matching using the 2DG-NN classifier. A nonlinear decision hyperplane is learned that can select either the LBP or the 2DG-NN.

c) SVM_3 is trained using the labeled training data $\{\mathbf{x_3}_i, y_{3i}\}$. Here, $\mathbf{x_3}_i$ is the ith training data vector that contains match scores and verification accuracy priors pertaining to the two classifiers, 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 an output of $\mathrm{SVM}_3 > 0$ indicates the use of the sum rule and $\mathrm{SVM}_3 \leq 0$ indicates the use of the sequential fusion algorithm.

2) Dynamic Selection of Algorithms: For probe verification, the trained SVMs are used to dynamically select the most appropriate algorithm depending on the quality vector.

a) The quality vectors pertaining to both the gallery and probe images are provided as input to the trained SVMs. The SVM_1 classifier selects between using a single classifier and fusion.

b) Depending on the classification result of the 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, nonlinear, 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 to mitigate the effect of specific covariates. For example, the neural network architecture-based 2DG-NN algorithm [20] can tolerate variations in expression, illumination, and occlusion whereas local facial features can handle pose and expression variations. It is our assertion that the performance of a face recognition system can be greatly enhanced if information from multiple algorithms is fused and a final decision is obtained using the fused information. In this section, we use the sequential fusion and dynamic selection algorithms to fuse the match scores computed from a nonlinear 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 dynamic selection algorithms. First, the face region from the input image is detected using the triangle-based face detection algorithm [21]. The size of the detected face image is normalized to 128 × 96. Next, the following algorithms are used for feature extraction and matching.

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	≥ 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

- **Neural Network Architecture-based 2-D Log Polar Gabor Transform:** The face image is transformed into polar coordinates and phase features are extracted using the neural network architecture-based 2DG-NN [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 LBF features are extracted to generate a feature vector [3]. Matching of two LBP feature vectors is performed using the weighted χ^2 distance measure.

A. *Face Databases Used for Evaluation*

To evaluate the performance on a large database with challenging intraclass variations, we combined images from multiple face databases to create a heterogeneous database of more than 116 000 images pertaining to 1194 subjects. Table I 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] 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] is comprised 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 single database available in the public domain which encompasses such a wide range of intraclass variations. The images are partitioned into two nonoverlapping 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 for training) and the remaining images (over 110 000) are used as the test data to evaluate the verification performance of the algorithms. Fig. 5 shows sample images in the training dataset and the 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 (GARs) over these trials at different false accept rates (FARs).

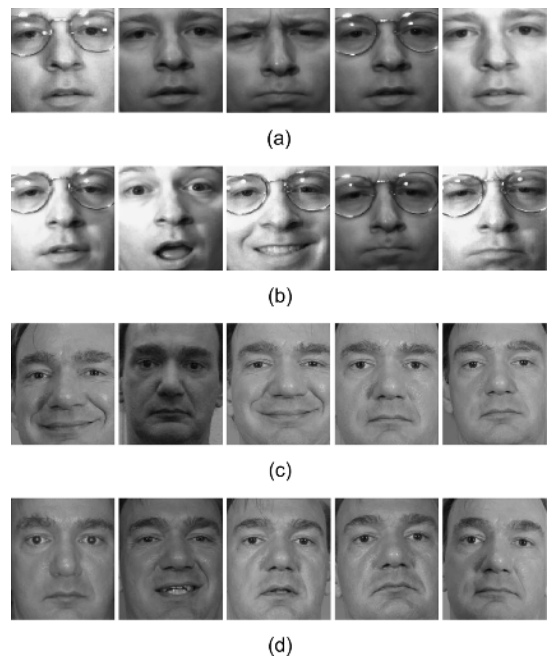

Fig. 5. Illustrating the examples of nonoverlapping training and gallery-probe datasets: (a) training images from the Equinox database; (b) gallery-probe images from the Equinox database; (c) training images from the Notre Dame database; and (d) gallery-probe images from the Notre Dame database.

B. *Performance Evaluation*

The training data is first used to train the proposed fusion algorithm and dynamic selection 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 labeled training match scores where labels are *genuine* and *impostor*. Unimodal classifier precision on the 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 the SVM kernel parameter (γ in RBF kernel[5]), that provide the best verification performance on the training data are chosen for testing. Similarly, the

[3]Available: http://amp.ece.cmu.edu/projects/FaceAuthentication/download.htm

[4]Available: http://www.equinoxsensors.com/products/HID.html

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

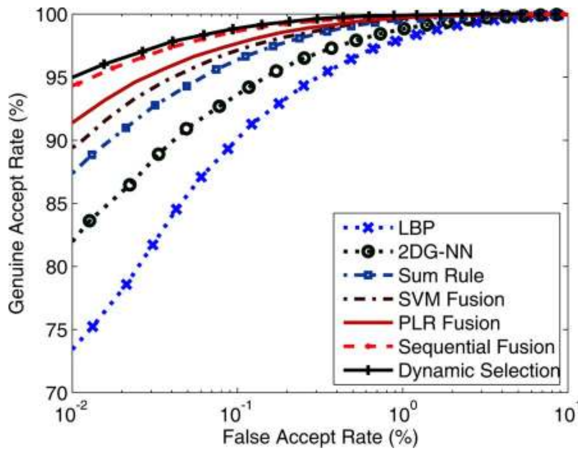

Fig. 6. ROC curves of the constituent face matchers, the proposed sequential fusion scheme, the dynamic selection algorithm, and some existing fusion algorithms.

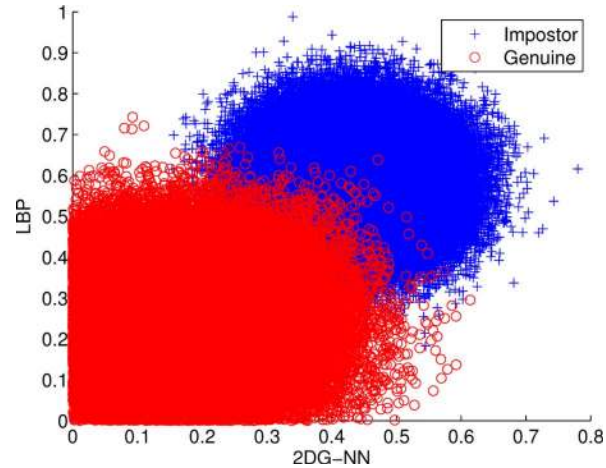

Fig. 7. Scatter plot shows that there is limited correlation between match scores obtained from LBP and 2DG-NN face recognition algorithms. Thus, score level fusion is expected to improve the matching performance.

dynamic selection algorithm is trained using the labeled data as described in Section III. The training set is also used to train the LBP and 2DG-NN face recognition algorithms. Further, the performance of the sequential fusion algorithm is compared against the sum rule with min/max normalization [18], SVM fusion [2], and product-of-likelihood-ratio (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 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 $\sim 5\%$ to 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 dynamic selection 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 technique. The results and their analysis are summarized as follows:

- The scatter plot in Fig. 7 and experimental results show that the match scores obtained from the 2DG-NN and LBP algorithms can be fused to significantly improve the verification accuracy. Further, covariate analysis in Table II suggests that the 2DG-NN algorithm provides good performance inspite of variations in expression, illumination, and occlusion whereas the 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 intrapersonal variations. Furthermore, during cross validation trials, we observed that the difference between minimum and maximum half total error rates $(\mathrm{HTER} = \mathrm{FAR} + \mathrm{FRR}/2)$ [5] for the sum rule is very large (Table III). This shows that the sum rule with min/max normalization is not able to handle disparities in the training-testing datasets.
- Tables II and III 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 the PLR fusion has the advantage of generalization whereas the SVM fusion algorithm can handle the nonlinearities in the match score.
- The sequential fusion algorithm effectively improves the verification accuracy. The algorithm transforms the match scores into probabilistic entities. Multiclassifier 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-based SVM classifier. Further, the t-test at 95% confidence suggests that the sequential fusion algorithm is significantly different than the other fusion algorithms. The HTER test also shows that the sequential fusion is *stable* across all cross validation trials.
- If the classifiers are in agreement (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, Fig. 8(b) and (c) shows sample cases when the two classifiers are in conflict but the proposed sequential fusion algorithm correctly accepts the subjects while the other fusion algorithms (sum rule, SVM fusion, and PLR fusion) provide incorrect results. Finally, there are few cases [sample shown in Fig. 8(d)] when both the classifiers reject a genuine subject. In such cases, the fusion algorithms cannot do much to improve the performance and, therefore, a 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

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	Dynamic Selection
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**

TABLE III
COMPARISON OF FUSION ALGORITHMS IN TERMS OF COMPUTATION TIME AND HTER

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]
Dynamic Selection	1.9	[4.35, 1.99]

2-GB RAM in a MATLAB environment, the proposed algorithm requires around 3.6 s for facial feature extraction, matching, fusion, and decision-making, whereas existing fusion algorithms require 1.7 to 2.8 s.

- The dynamic selection algorithm that unifies the LBP and 2DG-NN recognition algorithms, sum rule, and sequential match score fusion algorithm yields the best verification accuracy. Although the t-test at 95% confidence suggests that the dynamic selection algorithm is not significantly different from the sequential fusion scheme, the advantage of the former is computational time and stability (HTER test). As shown in Tables II and III, the computational cost of the dynamic selection algorithm is similar to that of the sum rule but it provides a relative performance gain of more than 60%.
- For cases in which the quality of the gallery-probe pairs are good and pose variation is minimum, the 2DG-NN algorithm is selected. The LBP technique is selected when images are of good quality and have pose variations. The fusion rules are selected when image quality is poor to moderate, gallery-probe pairs have large variations in pose, or facial features are occluded using cap/hat, scarf, and glasses. Further, the sum rule is chosen when intrapersonal variations are minimal and match scores exhibit minor conflicts. On the other hand, the sequential fusion algorithm is selected for cases with large intrapersonal variations. In the experiments, we observed that when the quality is good ($Q \geq 0.7$ and $A \geq 0.7$) and the difference in gallery-probe pose angles is small ($\leq 10^0$), then both the classifiers are in agreement 98% of the time. Overall, we found that around 38% of the time, LBP or 2DG-NN algorithms are chosen; 44% of the time, sum rule with min/max normalization is selected; and 18% of the time, the sequential fusion algorithm is selected.

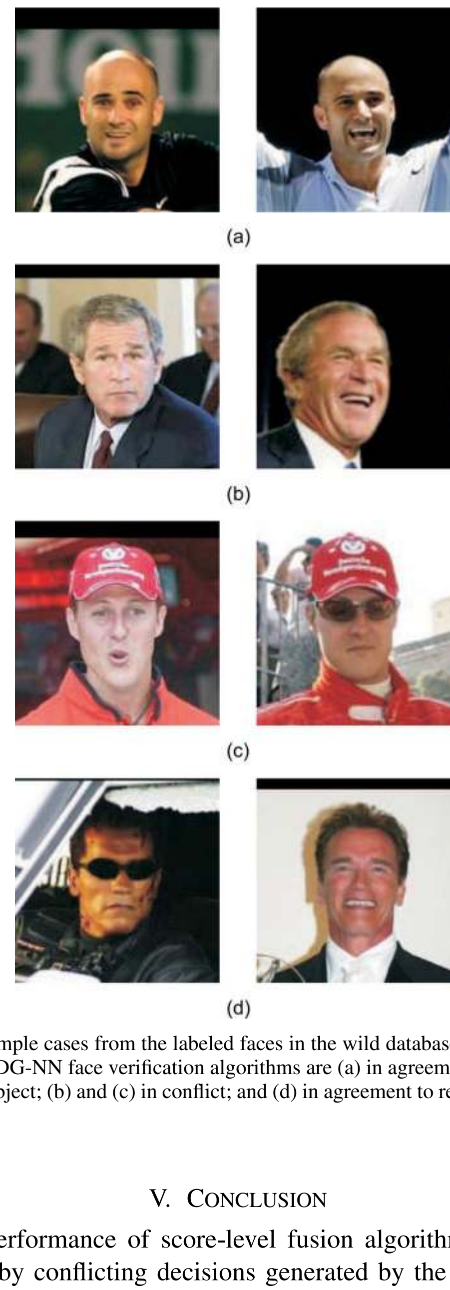

Fig. 8. Sample cases from the labeled faces in the wild database [9] when the LBP and 2DG-NN face verification algorithms are (a) in agreement to accept a genuine subject; (b) and (c) in conflict; and (d) in agreement to reject a genuine subject.

V. CONCLUSION

The performance of score-level fusion algorithms is often affected by conflicting decisions generated by the constituent matchers/classifiers for the same individual. Further, the computational cost of fusion algorithms that 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 an SVM 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 dynamic selection algorithm that unifies the constituent classifiers with the fusion schemes in order to optimize recognition accuracy and computation time. Depending on the quality of the input biometric data, the proposed algorithm dynamically selects between various 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. Experimental results on a heterogeneous face database of 1194 subjects suggest that the proposed algorithms can significantly improve the verification performance of a face recognition system with low computational overhead. In the future, we plan to extend the sequential fusion algorithm to include other parameters in the face quality assessment algorithm [1]. The sequential fusion and dynamic selection algorithms can also be extended to other multimodal scenarios involving face, fingerprint, and iris matchers.

ACKNOWLEDGMENT

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 Laboratory, Dr. A. R. Martinez, Dr. E. G. L. Miller, and Equinox Corporation for granting them access to the face databases used in this research.

REFERENCES

[1] M. Abdel-Mottaleb and M. Mahoor, "Algorithms for assessing the quality of facial images," *IEEE Comput. Intell. Mag.*, vol. 2, no. 2, pp. 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 Recognit.*, vol. 38, no. 5, pp. 777–779, 2005.

[3] T. Ahonen, A. Hadid, and M. Pietikinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[4] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1065–1074, Sep. 1999.

[5] S. Bengio and J. Mari'ethoz, "A statistical significance test for person authentication," in *Proc. Odyssey: The Speaker and Language Recognition Workshop*, 2004, pp. 237–244.

[6] J. Dezert and F. Smarandache, "Introduction to the fusion of quantitative and qualitative beliefs," *Inf. Security*, vol. 20, pp. 9–49, 2006.

[7] R. O. Duda, P. E. Hart, and D. Stork, *Pattern Classification*. Hoboken, NJ: Wiley, 2001.

[8] P. J. Flynn, K. W. Bowyer, and P. J. Phillips, "Assessment of time dependency in face recognition: An initial study," in *Proc. Audio- and Video-Based Biometric Person Authentication*, 2003, pp. 44–51.

[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 University of Massachusetts, Amherst, Tech. Rep., 2007.

[10] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer, 2005.

[11] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*. New York: Springer, 2005.

[12] A. R. Martinez and R. Benavente, The AR Face Database Computer Vision Center, Tech. Rep., 1998.

[13] P. Martins and J. Batista, "Monocular head pose estimation," in *Proc. Int. Conf. Image Analysis and Recognition*, 2008, pp. 357–368.

[14] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[15] S. Mukherjee and V. Vapnik, Multivariate Density Estimation: An SVM Approach Cambridge, MA, Tech. Rep., 1999.

[16] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio based biometric score fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 342–347, Feb. 2008.

[17] P. J. Phillips, H. Moon, S. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

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

[19] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[20] R. Singh, M. Vatsa, and A. Noore, "Face recognition with disguise and single gallery images," *Image Vis. Comput.*, vol. 27, no. 3, pp. 293–304, 2009.

[21] S. K. Singh, D. S. Chauhan, M. Vatsa, and R. Singh, "A robust skin color based face detection algorithm," *Int. Tamkang J. Sci. Eng.*, vol. 6, no. 4, pp. 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 Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 450–455, Mar. 2005.

[23] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: 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," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 1, pp. 3–13, Jan. 2009.

[25] M. Vatsa, R. Singh, A. Noore, and S. K. Singh, "Quality induced fingerprint identification using extended feature set," in *Proc. IEEE Int. Conf. Biometrics: Theory, Applications, and Systems*, 2008, pp. 1–6.

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

[27] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," in *Proc. 21st Int. Conf. Machine Learning*, 2004, pp. 903–910, ACM.

[28] W.-Y. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Survey*, vol. 35, no. 4, pp. 399–458, 2003.

[29] Z. Zivkovic and F. V. D. Heijden, "Recursive unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 651–656, May 2004.

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

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 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. His research has been funded by the U.S. Department of Justice, the Department of Defense, Unique Identification Authority of India (UIDAI), and the Department of Information Technology, India. He is a recipient of the FAST award by the Department of Science and Technology, India. 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 West Virginia University, Morgantown, WV.

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 CDEFFS, 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. Her research has been funded by the U.S. Department of Justice, the Department of Defense, Unique Identification Authority of India (UIDAI), and the Department of Information Technology, India. She was the recipient of six best paper and best poster awards.

Afzel Noore (S'84–M'87) received the Ph.D. degree in electrical engineering from West Virginia University, Morgantown, WV.

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, *Open Nanoscience Journal*, and the *International Journal of Multimedia Intelligence and Security*. 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 and best poster 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.

From 1996 to 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. WV. 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 *Handbook of Multibiometrics* and coeditor of *Handbook of Biometrics*.

Dr. Ross is a recipient of NSF's 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.