

Likelihood Ratio in a SVM Framework: Fusing Linear and Non-Linear Face Classifiers

Mayank Vatsa, Richa Singh, Arun Ross, and Afzel Noore
Lane Department of Computer Science and Electrical Engineering
West Virginia University

{mayankv, richas}@csee.wvu.edu, {arun.ross, afzel.noore}@mail.wvu.edu

Abstract

The performance of score-level fusion algorithms is often affected by conflicting decisions generated by the constituent matchers/classifiers. This paper describes a fusion algorithm that incorporates the likelihood ratio test statistic in a support vector machine (SVM) framework in order to classify match scores originating from multiple matchers. The proposed approach also takes into account the precision and uncertainties of individual matchers. The resulting fusion algorithm is used to mitigate the effect of covariate factors in face recognition by combining the match scores of linear appearance-based face recognition algorithms with their non-linear counterparts. Experimental results on a heterogeneous face database of 910 subjects suggest that the proposed fusion algorithm can significantly improve the verification performance of a face recognition system. Thus, the contribution of this work is two-fold: (a) the design of a novel fusion technique that incorporates the likelihood ratio test-statistic in a SVM fusion framework; and (b) the application of the technique to face recognition in order to mitigate the effect of covariate factors.

1. Introduction

In general, it has been well established that a carefully designed match score fusion algorithm can improve the performance of a multibiometric system [1], [12]. A major problem with existing match score fusion algorithms occurs when different classifiers generate conflicting results on the same input biometric data. While evidence-based fusion algorithms [2], [16] can address these conflicting cases, they are typically *ad hoc* in their approach and have a large computational complexity.

This paper presents a match score fusion algorithm to address these challenges and improve the verification performance of a multibiometric system. *The novelty of this algorithm lies in its incorporation of the likelihood ratio test-*

statistic [10] in a SVM fusion framework. The likelihood ratio aspect of the algorithm makes it robust to uncertainties in the component matchers; the use of a SVM ensures that the algorithm is less prone to over-fitting thereby permitting it to handle conflicting match scores. The performance of the proposed hybrid fusion algorithm is evaluated in the context of a face recognition application. Match scores computed from linear face recognition algorithms (i.e., Principal Component Analysis (PCA), Independent Component Analysis (ICA), Fisher Discriminant Analysis (FDA)) and their non-linear counterparts (i.e., Kernel PCA, Kernel ICA, Kernel FDA) are fused, and the verification performance is compared with existing match score fusion algorithms. Experiments indicate that the proposed fusion algorithm efficiently utilizes both statistical and learning basis, and improves the verification performance without increasing the computational cost.

2. Incorporating the Likelihood Ratio Test Statistic in a SVM Fusion Framework

The proposed fusion algorithm is implemented in three steps: (1) transforming match scores into likelihood ratios, (2) integrating the verification prior, and (3) applying support vector machine (SVM) fusion. This section presents an overview of 2ν -support vector machine followed by the proposed fusion algorithm.

2.1. Overview of 2ν -Support Vector Machine

Support vector machine [15] based match score fusion has been widely used in the multibiometrics literature. SVM is a pattern classifier that constructs non-linear hyperplanes in a multidimensional space. In this research, we use dual ν -SVM (2ν -SVM) [3]. 2ν -SVM is an attractive alternative to SVM that offers much more natural setting for parameter selection with reduced computational complexity. A brief overview of 2ν -SVM is presented here.

Let $\{\mathbf{x}_i, y_i\}$ be a set of N data vectors with $\mathbf{x}_i \in \mathbb{R}^d$, $y_i \in \{+1, -1\}$, and $i = 1, \dots, N$. \mathbf{x}_i is the i^{th} data

vector that belongs to a binary class y_i . According to Chew *et al.* [3], the objective of training 2ν -SVM is to find the hyperplane that separates the two classes with the widest margins, i.e., $\mathbf{w}\varphi(\mathbf{x}) + b = 0$ to minimize,

$$\begin{cases} \frac{1}{2}\|\mathbf{w}\|^2 - \sum_i C_i(\nu\rho - \psi_i) \\ \text{subject to } y_i(\mathbf{w}\varphi(\mathbf{x}_i) + b) \geq (\rho - \psi_i), \quad \rho, \psi_i \geq 0 \end{cases} \quad (1)$$

where $\varphi(\mathbf{x})$ is the mapping function used to map the data space to the feature space and provide generalization for the decision function that may not be a linear function of the training data. ρ is the position of the margin, ν is the error parameter, $C_i(\nu\rho - \psi_i)$ is the cost of errors, \mathbf{w} is the normal vector, b is the bias, and ψ_i is the slack variable for classification errors. The error parameter ν can be calculated using Equation (2).

$$\nu = \frac{2\nu_+\nu_-}{\nu_+ + \nu_-}, \quad 0 < \nu_+ < 1, \quad \text{and} \quad 0 < \nu_- < 1 \quad (2)$$

where ν_+ and ν_- are the error parameters for the positive and negative classes, respectively. Error penalty C_i is calculated as,

$$C_i = \begin{cases} C_+, & \text{if } y_i = +1 \\ C_-, & \text{if } y_i = -1 \end{cases} \quad (3)$$

where,

$$C_+ = \left[n_+ \left(1 + \frac{\nu_+}{\nu_-} \right) \right]^{-1}, \quad C_- = \left[n_- \left(1 + \frac{\nu_-}{\nu_+} \right) \right]^{-1} \quad (4)$$

Here, n_+ and n_- are the number of positive and negative training samples, respectively. Finally, 2ν -SVM training [3] can be formulated as,

$$\max_{(\alpha_i)} \left\{ -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\} \quad (5)$$

where,

$$0 \leq \alpha_i \leq C_i, \quad \sum_i \alpha_i y_i = 0, \quad \text{and} \quad \sum_i \alpha_i \geq \nu \quad (6)$$

$i, j \in 1, \dots, N$, α_i, α_j are the Lagrange multipliers and $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function. One example of kernel function is the Radial Basis Function (RBF) kernel as shown in Equation (7).

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad \gamma > 0 \quad (7)$$

2.2. Proposed Match Score Fusion Algorithm

This section presents the proposed match score fusion algorithm using 2ν -SVM. Traditionally, SVM-based fusion techniques use the match scores directly in the fusion process. However, match scores may provide inaccurate or fuzzy information due to large intra-class variations or matcher limitations. By transforming the match scores into likelihood ratios [5], the uncertainties associated with them can be handled better [10]. Therefore, the input to the fusion algorithm comprises of pertinent and discriminatory attributes such as likelihood ratios induced from match scores and the verification prior of individual matchers (i.e., precision of the matcher). The procedure of transforming match scores into likelihood ratios is described as follows:

For a two class biometrics problem, i.e. $\Theta = \{\text{genuine}, \text{impostor}\}$, the match scores corresponding to all N classifiers are first computed, i.e. $\mathbf{x} = (x_1, x_2, \dots, x_N)$, and then the densities of the genuine and impostor scores ($f_{gen}(\mathbf{x})$ and $f_{imp}(\mathbf{x})$, respectively) are estimated. In the proposed SVM fusion algorithm, it is assumed that the distribution of match scores is a Gaussian distribution, i.e.,

$$f_j(x_i, \bar{\mu}_{ij}, \bar{\sigma}_{ij}) = \frac{1}{\bar{\sigma}_{ij}\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left\{ \frac{x_i - \bar{\mu}_{ij}}{\bar{\sigma}_{ij}} \right\}^2 \right] \quad (8)$$

where $\bar{\mu}_{ij}$ and $\bar{\sigma}_{ij}$ are the mean and standard deviation of the i^{th} classifier corresponding to the j^{th} element of Θ . While this is a very strong assumption, it does not impact the performance of the fusion system in the context of this application.

We compute the likelihood ratio $F_i = \frac{f_{gen}(x_i)}{f_{imp}(x_i)} a_i$ pertaining to each classifier where a_i is the verification prior. The resultant value F_i is used as input to the SVM fusion algorithm. *Further, utilizing the SVM classifier for fusion addresses the limitations of the likelihood test-statistic if the input data does not conform to the Gaussian assumption (which is usually the case).*

In the training phase, likelihood ratios induced from the match scores and their labels are used to train the 2ν -SVM for fusion. Let the labeled training data be represented as $Z_i = (F_i, y)$, where, i represents the i^{th} classifier. For each match score, the class label $y \in \Theta$ (or $y \in (+1, -1)$; here, +1 represents the genuine class and -1 represents the impostor class). N SVMs are trained using these labeled training data; one for each classifier. During the training phase, error parameters ν_+ and ν_- for each SVM can be computed either heuristically or by using the number of training samples. In this SVM formulation, we use the following equation to compute the error parameters.

$$\nu_+ = \frac{n_+}{n_+ + n_-}, \quad \nu_- = \frac{n_-}{n_+ + n_-} \quad (9)$$

The training data is mapped to a higher dimensional feature space such that $Z \rightarrow \varphi(Z)$ where $\varphi(\cdot)$ is the mapping function. The optimal hyperplane which separates the data into two different classes in the higher dimensional feature space can be obtained as the solution of Equation (5).

In the testing phase, the fused score of a multimodal test pattern $[F_i]$, ($i = 1, 2, \dots, N$) is defined as,

$$g(F_{fused}) = \sum_{i=1}^N g(F_i), \quad (10)$$

where,

$$g(F_i) = w_i \varphi(F_i) + b_i. \quad (11)$$

Here, w_i and b_i are the parameters of the hyperplane. The solution of Equation (10) is the signed distance of F_{fused} from the separating hyperplane [15]. Finally, to verify the identity, a decision of *accept* or *reject* is made on the test pattern using a threshold t ,

$$\text{Decision}(F_{fused}) = \begin{cases} \text{Genuine,} & \text{if output of SVM} > t \\ \text{Impostor,} & \text{otherwise.} \end{cases} \quad (12)$$

3. Case Study in Face Recognition

In face recognition, appearance-based algorithms typically use subspace analysis methods [6], [9] such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Fisher Discriminant Analysis (FDA). However, these recognition algorithms do not provide good performance on real world databases because subspace analysis methods do not efficiently encode *non-linearly* distributed biometric data (i.e., the features reside in a nonlinear subspace). To address this limitation, researchers have adopted kernel approaches to subspace analysis that can capture higher-order statistics for better feature extraction [8]. Non-linear subspace analysis methods such as Kernel PCA (KPCA), Kernel ICA (KICA), and Kernel FDA (KFDA) are now used in face recognition and reported results show significant improvement over their linear counterparts.

We analyze the match scores obtained from linear techniques and their non-linear counterparts to determine if match score fusion can enhance the performance of face recognition. Scatter plot of match scores between PCA and KPCA algorithms (Figure 1) shows that there is limited correlation between the match scores generated by these two classifiers. Therefore, there is scope to enhance the face recognition performance by combining the match scores obtained using the PCA and KPCA techniques. A similar observation can be made for ICA and KICA, as well as FDA and KFDA.

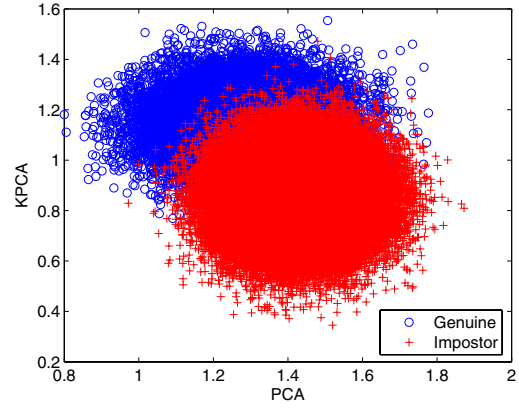


Figure 1. The scatter plot of match scores corresponding to the PCA and KPCA techniques. For these two classifiers, the Pearson's correlation coefficient for genuine scores is 0.57 and that of impostor scores is 0.41.

As shown in Figure 2, the Gaussian pyramid of gallery and probe face images are first generated at three different resolutions (i.e., levels): (G_0, G_1, G_2) . The use of multiple resolution allows the feature extraction algorithm to extract different types of features at various levels in the pyramid. Also, neuroscientists have shown that the human mind can recognize familiar faces even at low-resolution and can perform better analysis with multi-resolution images [14]. Therefore, at each level of the Gaussian pyramid, a linear appearance-based algorithm and its non-linear counterpart are used to compute two match scores (i.e., L_{G_i} and NL_{G_i} where L represents the score emitted by the linear classifier and NL represents that of the non-linear classifier). Thus, six match scores are generated using three levels of the Gaussian pyramid. These match scores are then fused using Equation (10) with $N = 6$ while Equation (12) is used to determine if the fused score corresponds to the 'genuine' or 'impostor' class.

3.1. Experimental Protocol

For evaluating the performance on a large database with challenging intra-class variations, we combined images from multiple face databases to create a heterogeneous database of 910 subjects. Table 1 lists the databases used and the number of subjects selected from the individual databases. The AR face database [7] contains face images with varying illumination and accessories, while the CMU-AMP database¹ contains images with large expression variations. The FERET database [11] has face images with different variations over a time interval of 3-4 years. The CMU-PIE dataset [13] contains images with variation in pose, illumination and facial expressions. The Notre Dame

¹<http://amp.ece.cmu.edu/projects/FaceAuthentication/download.htm>

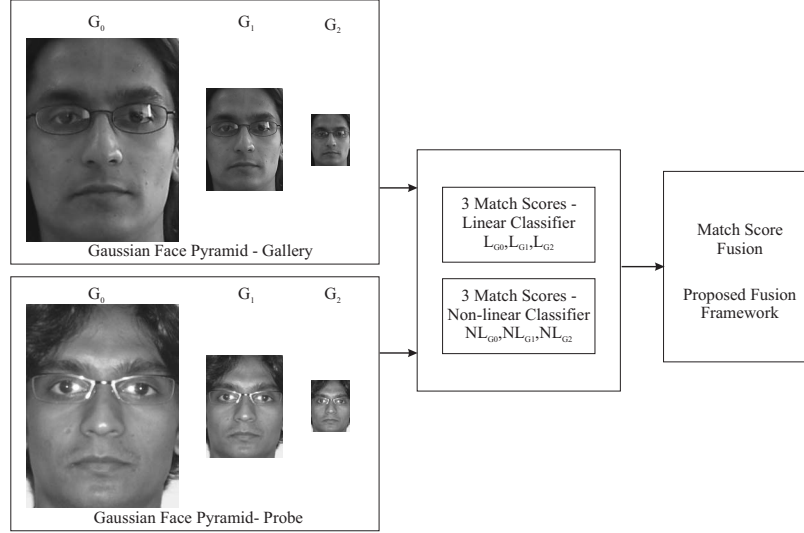


Figure 2. Illustrating the steps involved in match score fusion of linear and non-linear face recognition classifiers.

Table 1. Composition of the heterogeneous face database.

Face Database	Number of Subjects
AR	120
CMU-AMP	13
FERET	300
CMU - PIE	65
Notre Dame	312
Equinox	100
Total	910

face database [4] comprises of images with different lighting and facial expressions over a period of one year. The Equinox database² has images captured under different illumination conditions with accessories and expressions. To the best of our knowledge, there is no single database available in the public domain which encompasses all these types of intra-class variations. We partition the images into two sets: (1) the training dataset is used to train the individual classifiers (i.e., the linear and non-linear techniques) and fusion algorithm, and (2) the gallery-probe dataset (the test set) is used to evaluate the performance of the proposed algorithm. Three images of each subject are randomly selected to compose the training set. The remaining images are used as the test data to evaluate the algorithms. This train-test partitioning is repeated twenty 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). Furthermore, verification accuracies are reported at 0.01% false accept rate (FAR).

²<http://www.equinoxsensors.com/products/HID.html>

3.2. Performance Evaluation

The training data is first used to (a) train the linear and non-linear classifiers; (b) to compute the likelihood ratio and the verification priors; and (c) to train the 2ν -SVM fusion classifier. In all the experiments, we use the radial basis function (RBF) kernel with RBF parameter $\gamma = 4$ for the SVM classifier. The image dimensions of the three levels of the Gaussian pyramid are 128×96 (for G_0), 64×48 (for G_1), and 32×24 (for G_2). The performance of the proposed match score fusion algorithm is also compared with three existing match score fusion algorithms namely, sum rule with min/max normalization [12], product of likelihood ratio (PLR) fusion with Gaussian assumption, and classical SVM fusion [1].

The ROC plot in Figure 3(a) shows the comparative results of the PCA and KPCA face verification algorithms, and the improvement due to match score fusion. The non-linear version of the classifier outperforms its linear version by around 10%. Match score fusion improves the verification accuracy by $\sim 7\%$ - 21% and the proposed fusion approach gives the best results with a 94.02% accuracy. Similarly, as shown in Figures 3(b), and 3(c), the proposed fusion algorithm outperforms existing match score fusion algorithms for ICA and KICA as well as for FDA and KFDA.

Experiments are also performed to evaluate the effect of different covariate factors (viz., expression, illumination, pose) on the performance of face verification. This experiment facilitates the comparative analysis of linear and non-linear subspace methods and the subsequent improvement by employing the proposed match score fusion. The results of this experiment are summarized below:

- From Table 2, the covariate analysis ascertains our previous observation that non-linear algorithms can bet-

Table 2. Covariate analysis of linear and non-linear face recognition algorithms, and the match score fusion algorithms.

Classifiers	Covariate	Verification Accuracy					
		Linear Classifier	Non-linear Classifier	Sum Rule Fusion [12]	PLR Fusion [10]	SVM Fusion [1]	Proposed Fusion
PCA vs. KPCA	Expression	63.44	73.35	80.80	91.34	88.45	96.37
	Illumination	62.98	72.92	80.47	90.95	88.03	95.29
	Pose	61.36	70.01	78.62	86.83	85.98	92.71
	Overall	62.17	72.36	79.35	89.21	87.69	94.02
ICA vs. KICA	Expression	65.53	77.78	84.92	90.65	91.17	94.05
	Illumination	63.01	76.24	83.10	88.06	88.99	91.13
	Pose	61.95	75.30	80.49	87.24	87.36	89.94
	Overall	63.21	76.43	83.41	88.38	89.32	91.67
FDA vs. KFDA	Expression	75.79	79.61	88.06	91.57	92.48	96.91
	Illumination	75.88	79.93	88.51	91.13	92.31	97.06
	Pose	71.04	75.87	83.94	86.69	89.05	93.42
	Overall	74.80	78.59	86.77	89.01	91.26	95.71

ter encode the facial features compared to their linear counterparts.

- This table also indicates that variation in pose causes a large reduction in verification accuracy compared to expression and illumination variations. Furthermore, among the linear algorithms, FDA yields the best verification accuracy; and among the non-linear algorithms KFDA outperforms KPCA and KICA.
- Matcher correlation analysis and experimental results show that linear and non-linear appearance-based face verification algorithms provide complementary information, and that match score fusion can significantly improve the verification accuracy.
- In all cases, the proposed approach for SVM match score fusion yields the best accuracy. Further, t -test at 95% confidence shows that the proposed fusion algorithm is significantly different than the other fusion algorithms.
- The time complexity of the proposed fusion approach is also reasonable when compared with existing fusion algorithms. Let the number of support vectors be M and the number of unimodal classifiers be N . For probe verification, time complexity of the proposed fusion approach is $O(NM)$. On the other hand, time complexity of sum rule is $O(N)$, PLR fusion is $O(N)$, and classical SVM fusion is $O(NM^2)$. On a 2 GHz Pentium Duo Core processor with 2 GB RAM under MATLAB environment, the proposed algorithm requires around 1 second for fusion and decision-making, whereas the classical SVM fusion requires around 1.5 seconds, PLR fusion requires 0.5 seconds, and the sum rule requires 0.2 seconds.

4. Conclusion

The contribution of this work is two-fold: (a) the design of a novel fusion technique that incorporates the likelihood ratio test-statistic in a SVM fusion framework; and (b) the application of the technique to face recognition in order to mitigate the effect of covariate factors. The proposed fusion algorithm first transforms the match score into a likelihood ratio and further attunes it using the verification prior. A 2ν -SVM based match score fusion algorithm is then used for information fusion. The case study using linear and non-linear face recognition classifiers suggests that the proposed fusion algorithm is a cost effective alternative to existing fusion algorithms that 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. Currently, we are investigating methods to generalize the fusion algorithm by allowing non-Gaussian match score distributions in the framework.

5. Acknowledgment

The authors would like to thank CVRL University of Notre Dame, NIST, Robotics Institute CMU, CMU AMP Research Lab, Dr. A. R. Martinez, and Equinox Corporation for granting us access to the face databases used in this research. Ross was supported by NSF CAREER grant number IIS 0642554.

References

- [1] J. F. Aguilar, J. O. Garcia, J. G. Rodriguez, and J. Bigun. Discriminative multimodal biometric authentication based

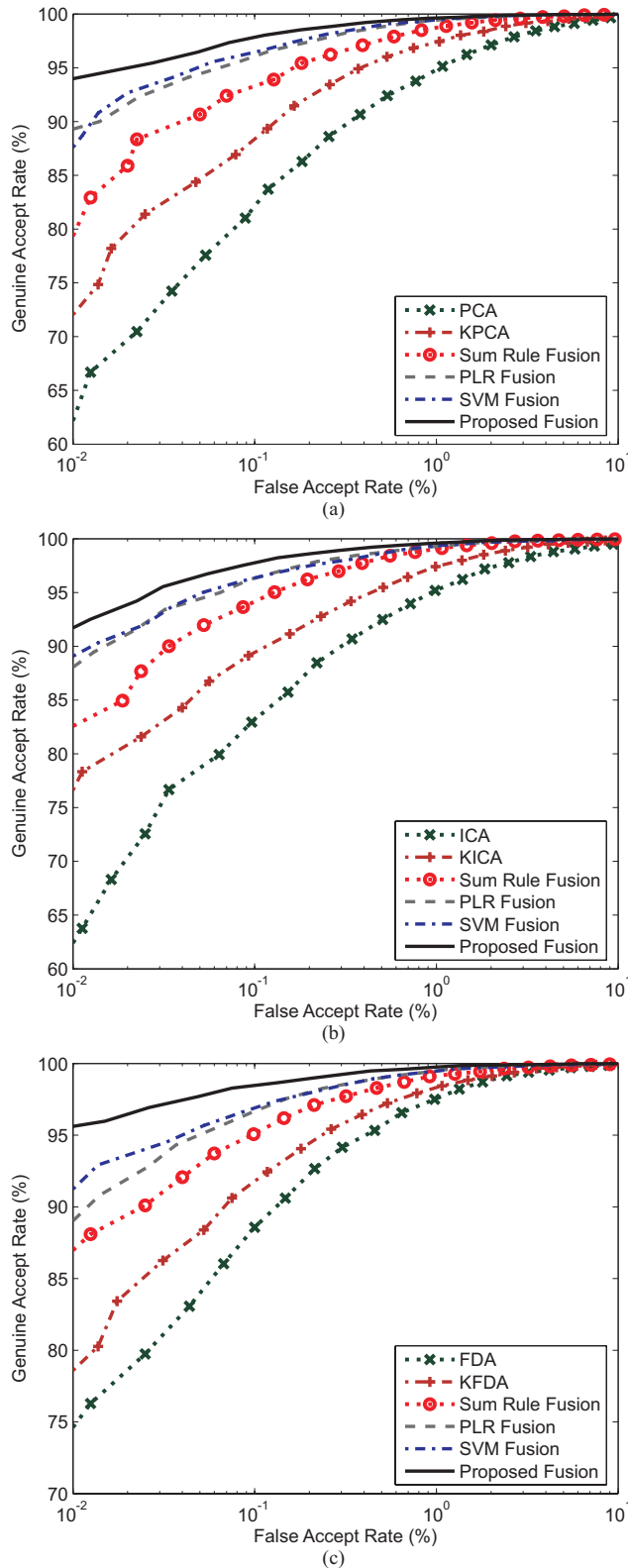


Figure 3. Comparing the verification performance the proposed fusion algorithms with existing fusion algorithm. (a) PCA and KPCA, (b) ICA and KICA, and (c) FDA and KFDA.

- on quality measures. *Pattern Recognition*, 38(5):777–779, 2005. 1, 4, 5
- [2] E. S. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert conciliation for multi modal person authentication systems by bayesian statistics. In *Proceedings of Audio- and Video-Based Biometric Person Authentication*, pages 291–300, 1997. 1
- [3] H. G. Chew, C. C. Lim, and R. E. Bogner. An implementation of training dual- ν support vector machines. *Optimization and Control with Applications, Qi, Teo, and Yang (Editors)*, 2004. 1, 2
- [4] 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. 4
- [5] E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses*. Springer, New York, 3rd edition, 2005. 2
- [6] J. Li, S. Zhou, and C. Shekhar. A comparison of subspace analysis for face recognition. In *Proceedings of International Conference on Multimedia and Expo*, volume 3, pages 121–124, 2003. 3
- [7] A. R. Martinez and R. Benavente. The AR face database, 1998. Computer Vision Center, Technical Report. 3
- [8] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Miller. Fisher discriminant analysis with kernels. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing IX*, pages 41–48, 1999. 3
- [9] B. Moghaddam. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):780–788, 2002. 3
- [10] 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. 1, 2, 5
- [11] 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. 3
- [12] A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of multi-biometrics*. Springer, New York, 1st edition, 2006. 1, 4, 5
- [13] 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. 3
- [14] P. Sinha, B. J. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: 19 results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006. 3
- [15] V. Vapnik, S. Golowich, and A. Smola. Support vector method for function approximation, regression estimation and signal processing. *Advances in Neural Information Processing Systems*, 9:281–287, 1997. 1, 3
- [16] M. Vatsa, R. Singh, A. Noore, and M. Houck. Quality-augmented fusion of level-2 and level-3 fingerprint information using DS_m theory. *International Journal of Approximate Reasoning*, doi:10.1016/j.ijar.2008.01.009, 2008. 1