

Predicting Identification Errors in a Multibiometric System Based on Ranks and Scores

Emanuela Marasco, Arun Ross, Carlo Sansone

Abstract—The goal of a biometric identification system is to determine the identity of the input biometric data. In such a system, the input probe (e.g., a face image) is compared against the labeled gallery data (e.g., face images in a watch-list) resulting in a set of ranked scores pertaining to the different identities in the gallery database. The identity corresponding to the best score is then associated with that of the probe. The aim of this work is to predict identification errors and improve the recognition accuracy of the biometric system. The method utilizes the rank and score information generated by the identification operation in order to validate the output. Further, we demonstrate the proposed predictor can be effectively applied in multimodal scenarios. Experiments performed on two multimodal databases show the effectiveness of our framework in improving identification performance of biometric systems.

I. INTRODUCTION

The goal of a biometric *identification* system is to determine the identity of the input biometric data. In such a system, the input probe (e.g., a face image) is compared against a labeled gallery data (e.g., face images in a watch-list) resulting in a set of ranked scores pertaining to the different identities in the gallery database. The identity corresponding to the best score is then typically associated with that of the probe. In large-scale identification systems, the feature space of the identities in the gallery may significantly overlap resulting in the degradation of identification accuracy. Moreover, in adverse operational scenarios the input data is often noisy, and the similarity between the probe and the associated gallery data is substantially reduced thereby impacting overall recognition accuracy.

This paper concerns itself with the possibility of automatically determining if the decision rendered by a biometric identification system is correct or not. In the work proposed here, we use both rank and score information in order to predict errors and, subsequently, improve the performance of a biometric system. In the proposed methodology, the probability that the output decision is reliable is estimated by a pattern classifier referred to as a *predictor*. Its role is to detect potentially erroneous decisions. Further, we propose three fusion mechanisms based on the trained predictor that can extend the benefits of the proposed scheme into a multimodal

scenario. In particular, a predictor-based voting strategy, a predictor-based serial fusion scheme and a predictor-based Borda Count method are presented and compared against other common approaches to *rank-level* fusion.

The idea of marginalizing potentially incorrect decisions in a pattern recognition system was used by Chow [1] to define an optimum rejection rule. In the pattern recognition and machine learning literature, several techniques have been proposed to predict the reliability of a classification decision rendered by a pattern recognition system (e.g., [2]). However, such methods have been sparingly used in the biometric literature until recently. Kryszczuk et al. [3] presented a method in which classifier decisions and the corresponding reliability information are combined to predict and correct verification decisions. Kryszczuk et al. [4] later proposed a framework for probabilistic error rectification based on credence estimation which was used to eliminate unreliable verification decisions.

Quality has been generally considered a good indicator for failure prediction. However, a quality-based scheme cannot be applied unless real-time assessment about the quality of the input data is possible; also Scheirer *et. al* [5] have shown cases in which ‘poor quality’ images produce better matching scores than ‘high quality’ images. An alternative to quality-based analysis was presented by Scheirer and Boulton [6] in which they proposed the idea of post-recognition failure prediction. They defined two types of error, viz., Failure Prediction False Accept Rate (FPFAR) and Failure Prediction False Reject Rate (FPFRR) which were used to generate the Failure Prediction Receiver Operator Characteristic (FPROC) curve. The failure prediction features were derived from similarity scores and were designed to capture information that are not explicit in the raw score. Specifically, they extracted the differences between scores and the DCT coefficients based on transforming the top n scores. Their scheme predicts the failure of individual modalities and also assigns weights to individual modalities during fusion.

The novelty of our work is two-fold: (a) it addresses the problem of incorrect decisions in an *identification* system, and (b) it utilizes this information to improve the recognition accuracy in a multibiometric scenario.

The paper is organized as follows. In Section 2, we describe the sources of information that can be exploited during the biometric recognition process. Section 3 presents the proposed approach. Section 4 reports the experimental procedure and some comparative results against traditional schemes for rank-level fusion. Section 5 presents our conclusions.

Emanuela Marasco and Carlo Sansone are with Dipartimento di Informatica e Sistemistica, Università degli Studi di Napoli Federico II, Via Claudio, 21 I-80125 Napoli, Italy {emanuela.marasco, carlosan}@unina.it

Arun Ross is with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, PO Box 6109 Morgantown, WV, USA arun.ross@mail.wvu.edu

This work was performed when Marasco was a Visiting Scholar at West Virginia University. Ross was supported by US NSF CAREER Grant No. IIS 0642554

II. BACKGROUND

In a typical biometric system, the sensor acquires the biometric data of a user from which a representative feature set is extracted. This feature set is matched against the feature set stored in the database of the system. The decision taken by the system is based on the match scores generated during the matching process [7]. In an identification system, these scores are transformed to ranks in order to determine potential matching identities. Since we consider a multibiometric scenario in our work, the use of score and rank information in a fusion framework is discussed below.

A. Match Score Information

Match scores are commonly used to consolidate the decisions rendered by multiple biometric classifiers since they are easy to access (compared to, say, feature sets which can be proprietary). However, the scores output by different biometric matchers may not be homogeneous and can conform to different scales [8]. Thus, before integration, they may have to be transformed into a common domain via an effective normalization scheme. Such a scheme may rely on a fixed set of parameters estimated from the impostor and genuine score distributions [7] [9]. A number of match score fusion schemes have been proposed and studied in the literature [10].

B. Rank Information

At the rank level, each biometric matcher orders the candidate identities in the gallery according to their similarities to the given probe and transforms this ordering into a set of N integer values or *ranks*. A fusion scheme at this level consolidates the rankings provided by multiple biometric matchers in order to obtain a consensus rank for each identity in the gallery [8]. If we consider an input image having low quality, the genuine score as well as the impostor scores are likely to be low [11] [12]. The use of such a score (for a genuine user) during the fusion process may confound a fusion algorithm. The rank, on the other hand, is a relatively stable statistic and does not require normalization; combining this rank with other ranks (for the genuine class) in a judicious manner can result in a correct classification.

C. Hybrid Rank-Score Information

The use of both ranks and match scores [9] is expected to be more reliable and has been demonstrated to increase the recognition accuracy of a multibiometric system [13]. For a given probe image, a $N \times C$ score matrix $S = [s_{ik}]$ can be generated where s_{ik} represents the similarity score computed by the k^{th} modality matcher C_k after comparing the probe against the i^{th} entry in the gallery database, $i = 1 \dots N$ and $k = 1 \dots C$. For each modality, the corresponding scores can be sorted in decreasing order. So a $N \times C$ rank matrix $R = [r_{ik}]$ can be generated where r_{ik} is the rank assigned to the i^{th} identity in the database by the matcher C_k . Thus, the output of each matcher, C_k , can be viewed as a two-tupled entry (s_{ik}, r_{ik}) , $i = 1 \dots N$. The approach presented by Brunelli and Falavigna [9] generated a mapping from

(s_{ik}, r_{ik}) into the set $\{0,1\}$ where the entries corresponding to the correct identity label is mapped to 1 and to 0 otherwise. A HyperBF network was used to approximate this mapping. This method has some drawbacks. Firstly, a network-based framework requires a large amount of training examples to tune the free parameters involved. Secondly, it requires the availability of all classifiers. Finally, when a new user is added, the network has to be trained again.

Recently, Nandakumar et al. [13] proposed a scheme that utilizes both ranks and scores to perform fusion in identification systems. They defined a hybrid rank-score fusion rule based on a combination of score and rank statistics. This approach, however, requires an explicit estimation of the genuine and impostor distributions, and a large dataset is required to accurately estimate the score distributions.

III. PROPOSED APPROACH: A RATIO-BASED ANALYSIS

As stated earlier, a generic identification system compares the input biometric data to all the known identities stored in the database and outputs a set of similarity scores. The scores are then sorted in decreasing order to form a ranking list in which the lowest rank is assigned to the highest similarity [14]. Let $\mathbf{G} = [G_1, G_2, \dots, G_N]$ be the *gallery* set, composed by N biometric samples belonging to N different subjects. Let $\mathbf{P} = [P_1, P_2, \dots, P_M]$ be the *probe* set, composed by M *unknown* samples belonging to subjects that are presumed to be in the *gallery*. Given a single probe image, N comparisons of that probe against the gallery are performed and N similarity scores are generated [15].

The present study is based on computing the ratio of scores corresponding to rank 1 and the other ranks. The vector of these ratios is treated like a feature vector and used for training a pattern classifier. Such a classifier is used to learn the relationship between the ratios and the *posterior* probabilities of the *correct* and *error* classes. Here, the term “correct class” is used to indicate that the rank-1 identity is indeed the correct identity of the probe; the term “error class” is used to indicate that the rank-1 identity does not correspond to the correct identity of the probe. Thus, the classifier (*predictor*) is used to learn the decision boundary between the correct identification region and the erroneous one.

For a given input probe, let ρ_j denote the ratio of the rank-1 score to that of the rank- j score. Thus, the vector $(\rho_2, \rho_3, \dots, \rho_{d+1})^t$, $d \in \{1 \dots N - 1\}$, is used as input to the classifier. Typically, the rank-1 similarity score is expected to be significantly higher than the other scores (for a genuine match at rank-1). However, there are situations when the rank-1 score may be comparable to that of other scores associated with the nearby ranks thus suggesting the *possibility* of an error. In this work, we confirm this notion and, further, exploit it to improve recognition accuracy.

The architecture of the proposed approach is shown in Fig. 1.

A. Error prediction in multibiometric identification system

In a multibiometric identification system, the output of K different biometric modality matchers C_1, C_2, \dots, C_K have to

Algorithm for training the unimodal predictor

Let $\mathbf{G} = [G_1, G_2, \dots, G_N]$ be the *gallery* set.

Let $\mathbf{P} = [P_1, P_2, \dots, P_M]$ be the *probe* set.

- 1) For each probe, generate N similarity match scores $s_i, i = 1 \dots N$ by comparing that probe against the gallery.
- 2) Sort the match scores in decreasing order.
- 3) Based on the previous sorted match scores, assign a rank R_i to each enrolled identity.
- 4) Compute the ratio ρ_j between the score corresponding to rank-1 and the score corresponding to rank- j .
- 5) Label the ratio score vector as *correct* if rank 1 is assigned to the correct identity by the unimodal matcher; otherwise label it as an *error*.
- 6) Use the labeled ratio score vectors as feature vectors to train a supervised classifier.

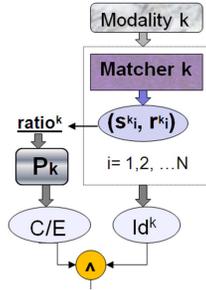


Fig. 1. Error Prediction in a unimodal identification system. Here, s_i^k and r_i^k denote the score and rank, respectively, assigned to the i^{th} identity in the gallery by the k^{th} matcher; P_k denotes the classifier used to predict if the rank-1 identification is correct (C) or not (E) based on the vector of score ratios ($ratio^k$). The output of the matcher, Id^k , is accepted or rejected based on the predictor.

be consolidated. The information observed at the score level can be represented as a $N \times K$ matrix $\mathbf{S} = [s_n^k]$, where s_n^k represents the match score output when the probe image is compared against the n^{th} gallery image, using the k^{th} classifier, $k = 1, \dots, K; n = 1, \dots, N$. This score matrix can be converted to a rank matrix $\mathbf{R} = [r_n^k]$ where r_n^k represents the rank of the n^{th} gallery image with respect to the probe as assessed by the k^{th} modality matcher.

B. Predictor-based Majority Voting

In the majority voting scheme, the outputs of the K classifiers are examined and the most commonly occurring output is selected as the final output. Thus, for a given probe, K unimodal matchers are employed and the winner is the identity to which the majority of matchers have assigned a rank value equal to one. The majority vote will result in an ensemble decision [16]:

$$\arg \max_{i=1 \dots N} \sum_{k=1}^K d_{ik} \cdot v_k \tag{1}$$

where the binary variable d_{ik} is 1 if the k^{th} matcher outputs identity i in rank-1, and the binary variable v_k is

1 if the identification is deemed to be *correct* by the k^{th} predictor. Fig. 2 denotes this scheme. The majority vote scheme will assign an identity to the probe only if the output of at least $\lfloor \sum_{k=1}^K v_k \rfloor + 1$ unimodal systems correspond to the same identity and are deemed to be correct by v_k . If a majority is not possible, then the proposed mechanism attempts to use the rank-1 accuracy of individual classifiers to make the decision. According to this design, when the unimodal outputs are K different identities, the output from the overall system will correspond to the identity output from the unimodal system with the highest accuracy (as assessed using training data before deployment of individual matchers). Those contributions considered as errors by the predictor module are excluded from the final decision.

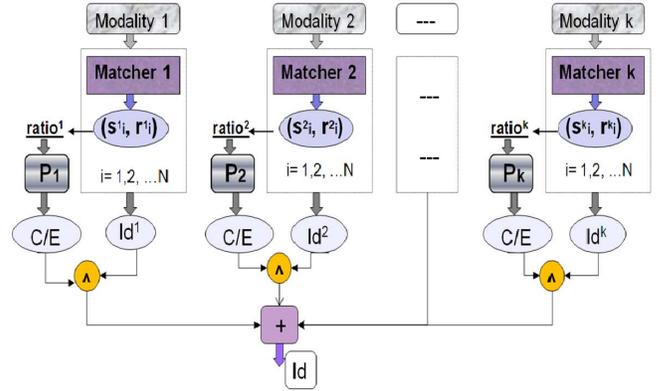


Fig. 2. Predictor-based Majority Voting.

C. Predictor-based Serial Fusion Scheme

In the serial scheme, the decisional process is split into two successive stages [17]. The subject to be authenticated submits the first biometric modality to the system which is processed and matched against all the templates present in the gallery. If the resulting identity is labeled to be correct by the predictor module, the input biometric trait is associated to the current identity, otherwise the system suspends the decision and an additional processing stage is performed. In the second stage, $K-1$ additional biometric modalities are automatically requested and a voting strategy involving $K-1$ unimodal matchers is adopted in the second stage. The predictor-based serial combination framework is shown in Fig. 3. It can be formulated as follows:

$$Id_m = \begin{cases} Id_u, & \text{if } v_u = 1 \\ \arg \max_{i=1 \dots N} \sum_{k=1}^{K-1} d_{ik} \cdot v_k, & \text{if } v_u = 0 \end{cases} \tag{2}$$

where Id_m is the output of the multimodal system and Id_u is the output of the unimodal system at the first stage.

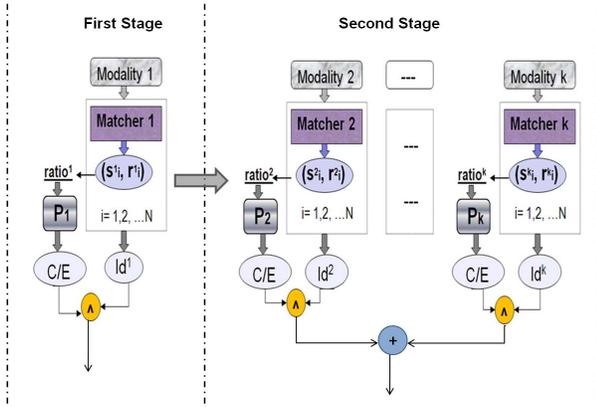


Fig. 3. Predictor-based Serial Fusion: the first stage is based on the unimodal system and the error predictor for this modality while the second stage consists of a predictor-based majority voting scheme which uses $K-1$ modalities.

D. Predictor-based Borda Count Fusion

In the *Borda Count* model, the rank for each identity in the database is calculated as the weighted sum of the individual ranks assigned by the K matchers:

$$R_i = \sum_{k=1}^K w_k r_{ik}, \quad i = 1, 2, \dots, N. \quad (3)$$

This method assigns a higher weight to the ranks provided by the more accurate matcher. Therefore, it is useful when different biometric matchers exhibit significant differences in their accuracies. A training phase has to be used to determine the weights. This is achieved by computing the weight w_k (using the training set) as the ratio between the number of correct identifications as detected by the predictor and the total number of test probes. The weight factor based on the predictor reduces the effect of inaccurate decisions due to potentially incorrect matcher output.

IV. PERFORMANCE EVALUATION

A. Datasets

The performance of the proposed strategy was evaluated on two databases. The first is the West Virginia University (WVU) multimodal biometric database. A subset of this database pertaining to the fingerprint (left thumb [FL1], right thumb [FR1], left index [FL2], right index [FR2]) and face modalities of 240 subjects was used in our experiments. Five samples per subject for each modality were available. Table I provides the details of the database. For the *face* modality, frontal images were collected in a controlled scenario. For the *fingerprint* modality, images were collected using an optical biometric scanner, without explicitly controlling the quality [18]. The entire dataset was divided into five sets: the first sample of each identity was used to compose the *gallery* and the remaining four samples of each identity were used as *probes* (P_1, P_2, P_3, P_4). The VeriFinger software was used for generating the fingerprint scores and the VeriLook software was used for generating the face scores.

TABLE I
WVU MULTIMODAL BIOMETRIC DATABASE

Biometric	Subjects	Samples	Scores
Face	240	5 per subject	Gen 1200×4 Imp $240 \times 239 \times 25$
Fingerprint	240	5 per finger	Gen $(1200 \times 4) \times 4$ Imp $(240 \times 239 \times 25) \times 4$

TABLE II
THE BIOSECURE DATABASE: DEVELOPMENT SET

Biometric	Subjects	Samples	Scores
Face	51	4 per subject	Gen 204×3 Imp $51 \times 50 \times 16$
Fingerprint	51	4 per subject	Gen $(204 \times 3) \times 3$ Imp $(51 \times 50 \times 16) \times 3$

The second database is a subset of the BioSecure multimodal database. This database contains 51 subjects in the Development Set (training) and 156 different subjects in the Evaluation Set (testing). For each subject, four biometric samples are available over two sessions: session 1 and session 2. The first sample of each subject in the first session was used to compose the gallery database while the second sample of the first session and the two samples of the second session were used as probes (P_1, P_2, P_3). For the purpose of this study, we used the face and three fingerprint modalities, denoted as *fnf*, *fo1*, *fo2* and *fo3*, respectively [19]. The details about the number of match scores per person are reported in Tables II and III.

B. Evaluation Procedure

First, we performed a preliminary analysis to understand the distribution of the ratios between scores as a function of the ranks (i.e., the ρ_k 's) for the correct and error classes. This was used to determine the dimension of the vector of ratios (i.e., d) that is suitable for error prediction. The number d was empirically derived for each modality in the individual databases considered in this work. Next, the proposed algorithm was evaluated on the two databases. Since the number of identification errors made by some of the biometric matchers is low, the negative class cannot be efficiently represented. This affects the training of the predictor. In order to maximize the amount of available data, the training and testing was performed by adopting the *leave-one-out* strategy. The classifier was trained by using the

TABLE III
THE BIOSECURE DATABASE: EVALUATION SET

Biometric	Subjects	Samples	Scores
Face	156	4 per subject	Gen 624×3 Imp $156 \times 155 \times 16$
Fingerprint	156	4 per subject	Gen $(624 \times 3) \times 3$ Imp $(156 \times 155 \times 16) \times 3$

samples provided by all but one of the identities in the gallery and its performance was tested on the excluded identity [9].

C. Results

As Fig. 4 and Fig. 5 show, in the space of ratios, the distributions of the misclassified identities are reasonably separated from those that were correctly recognized.

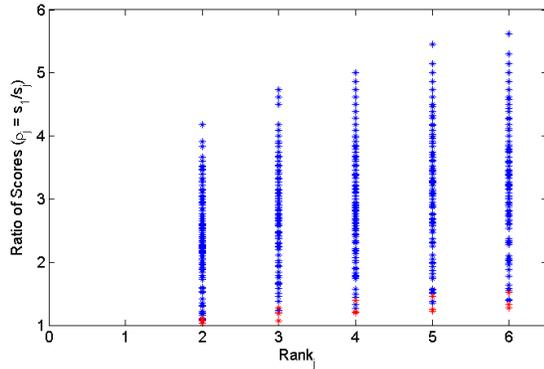


Fig. 4. The distribution of the ratios between scores in terms of ranks of all the users in the WVU database for the face modality, where the gallery set is composed by the first sample of each subject and the probe set by the fifth sample. Red points represent rank-1 misclassifications.

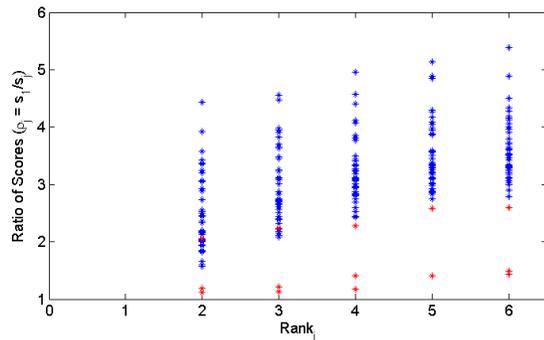


Fig. 5. The distribution of the ratios between scores in terms of ranks of all the users of the Development Set in the Biosecure database for the face modality, where the gallery set is composed by the first sample of each subject and the probe set by the second sample. Red points represent rank-1 misclassifications.

The classification was accomplished using three different classifiers: a *Support Vector Machine (SVM)*, a *Decision Tree* and a *Bayesian* classifier. Since the SVM classifier gave the best results on both databases, only its performances are being reported in this paper. Further, the classification performance was observed as a function of d , i.e., the number of ratios used to construct the feature vector. The face modality in the WVU database required $d = 5$; the FL1, FR1 and FR2 modalities required $d = 7$ and the fingerprint FL2 modality required $d = 10$ (see Figure 6). For the Biosecure dataset, all the 3 fingerprint modalities required $d = 3$ while the face required $d = 5$ (see Figure 7).

Tables IV, V, VI and VII compare the results of the proposed scheme against other schemes. We compared the

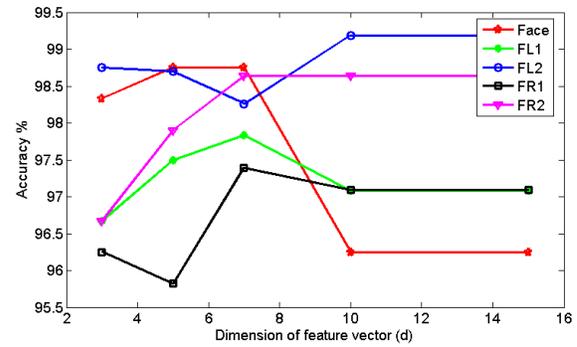


Fig. 6. Performance of the prediction scheme using a Support Vector Machine trained on the WVU data.

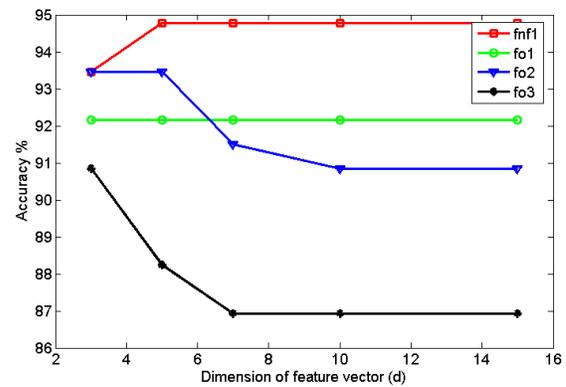


Fig. 7. Performance of the prediction scheme using a Support Vector Machine trained on the Biosecure data.

TABLE IV
PERFORMANCE OF THE TRADITIONAL FUSION SCHEMES ON THE FOUR PROBE SETS IN THE WVU DATABASE.

Probe	Highest Rank	Borda Count	Pure Majority Voting
P1	91.67%	97.22%	100.00%
P2	88.33%	95.56%	99.44%
P3	90.56%	96.11%	97.78%
P4	93.33%	96.67%	99.44%
Avg	90.97%	96.39%	99.17%

TABLE V
PERFORMANCE OF THE PREDICTOR-BASED FUSION SCHEMES ON THE FOUR PROBE SETS IN THE WVU DATABASE

Probe	Predictor-based Majority Voting	Predictor-based Serial	Predictor-based Borda Count
P1	100.00%	100.00%	97.22%
P2	100.00%	99.44%	96.11%
P3	100.00%	99.44%	96.11%
P4	100.00%	98.89%	97.22%
Avg	100.00%	99.44%	96.67%

performance of our methods against the *Highest Rank* and *Borda Count* approaches [10] as well as the *pure* Majority Voting Scheme in which the predictor for each modality was not used (ties were broken randomly). From these tables it is evident that the *predictor-based majority voting* which uses the predictor for each modality, outperformed the other traditional approaches. Moreover, the serial scheme also improved the correct identification rate since, in the second stage, it is able to handle those cases that are classified as *errors* in the first stage. We also observed that the improvement in performance was especially significant in the case of the BioSecure Database where traditional rank-level fusion schemes did not perform very well.

V. CONCLUSIONS

In this paper we presented a methodology in which both ranks and scores have been used to improve the identification accuracy of multimodal biometric systems. Ranks and scores have been used for designing a classifier (*predictor*) for each modality in order to detect identification errors. This information has been used to design novel rank-level fusion schemes in a multimodal scenario. The proposed predictor-based techniques performed better than the commonly used rank-level fusion mechanisms. In particular, the predictor-based majority voting resulted in the best accuracy by achieving an average recognition rate of 100% on the WVU dataset and 97.22% on the Biosecure dataset. The improvement in performance is especially significant in the BioSecure database. Since the predictor is based on a training phase, it generalizes very well across identities. Consequently, the predictor does not have to be retrained when a new individual is added to the database. Experiments are underway to determine the robustness of the scheme to variations in quality on the input data. It must be stated that the simple sum of scores results in good identification performance on the database used in our experiments; however, we can consider the methodology proposed in this paper as a promising approach for using both ranks and scores in a systematic way to predict identification errors in biometric systems.

TABLE VI

PERFORMANCE OF THE TRADITIONAL FUSION SCHEMES ON THE THREE PROBE SETS IN THE BIOSECURE DATABASE

Probe	Highest Rank	Borda Count	Pure Majority Voting
P_1	87.18%	96.15%	89.74%
P_2	78.85%	88.46%	83.97%
P_3	74.36%	92.31%	84.62%
Avg	80.13%	92.31%	86.11%

REFERENCES

- [1] T. Chow. On optimum recognition error and rejection trade-off. *IEEE Transaction on Information Theory*, 16:41–46, 1970.
- [2] L. P. Cordella, P. Foggia, C. Sansone, F. Tortorella, and M. Vento. Neural network classification reliability: Problems and applications. *Image Processing and Pattern Recognition*, 5:161–200, 1998.

TABLE VII

PERFORMANCE OF THE PREDICTOR-BASED FUSION SCHEMES ON THE THREE PROBE SETS IN THE BIOSECURE DATABASE

Probe	Predictor-based Majority Voting	Predictor-based Borda Count	Predictor-based Serial
P_1	100.00%	96.15%	100.00%
P_2	94.23%	89.10%	94.87%
P_3	97.44%	92.31%	94.87%
Avg	97.22%	92.52%	96.58%

- [3] K. Kryszczuk, J. Richiardi, P. Prodanov, and A. Drygajlo. Error handling in multimodal biometric systems using reliability measures. In *Proc. 13th European Signal Processing Conference (EUSIPCO)*, 2005.
- [4] K. Kryszczuk, J. Richiardi, and A. Drygajlo. Reliability estimation for multimodal error prediction and fusion. In *7th International Workshop on Pattern Recognition in Information Systems (PRIS)*, 2007.
- [5] W. Scheirer and T. Boult. Predicting biometric facial recognition failure with similarity surfaces and support vector machines. *IEEE Computer Society Workshop on Biometrics*, 2008.
- [6] W. Scheirer and T. Boult. A fusion-based approach to enhancing multimodal biometric recognition system failure prediction and overall performance. *IEEE Int. Conf. on Biometrics Theory Application and Systems*, 2008.
- [7] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, December 2005.
- [8] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters* 24, pages 2115–2125, 2003.
- [9] R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(10), October 1995.
- [10] A. Ross, K. Nandakumar, and A. Jain. *Handbook of MultiBiometrics*. Springer, 2006.
- [11] O. Melnik, Y. Vardi, and C. Zhang. Mixed group ranks: Preference and confidence in classifier combination. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 26(8):973–981, August 2004.
- [12] A. Saranlı and M. Demirekler. A statistical unified framework for rank-based multiple classifier decision combination. *Pattern Recognition*, 34:865–884, 2001.
- [13] K. Nandakumar, A. Jain, and A. Ross. Fusion in multibiometric systems: What about the missing data? *International Conference on Biometrics*, June 2009.
- [14] A. Abaza and A. Ross. Quality-based rank level fusion in biometrics. *Third IEEE International Conference on Biometrics: Theory, Applications and Systems*, September 2009.
- [15] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior. The relation between the roc curve and the cmc. *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pages 15–20, October 2005.
- [16] L.I. Kuncheva. *Combining Pattern Classifiers Method and Algorithms*. Wiley, 2004.
- [17] G.L. Marcialis and F.Roli. Serial fusion of fingerprint and face matchers. *Lecture Note in Computer Science*, 4472:151–160, June 2007.
- [18] S. Crialmeanu, A. Ross, S. Schuckers, and L. Hornak. A protocol for multibiometric data acquisition, storage and dissemination. *Technical Report, West Virginia University*, 2007.
- [19] N. Poh, T. Bourlai, and J. Kittler. A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms. *Pattern Recognition*, 43:1094–1105, 2010.