Confidence Based Rank Level Fusion for Multimodal Biometric Systems

Hossein Talebi(✉) and Marina L. Gavrilova

University of Calgary, 2500 University Dr. NW, Calgary T2N 1N4, Canada
{htalebi,mgavrilo}@ucalgary.ca

Abstract. Multimodal biometric systems have proven advantages over single biometric systems as they are using multiple traits of users. The intra-class variance provided by using more than one trait results in a high identification rate. Still, one of the missing parts in a multimodal system is inattention to the discriminability of each rank list for each specific user. This paper introduces a novel approach to select a combination of rank lists in rank level so that it provides the highest discrimination for any specific query. The rank list selection is based on pseudo-scores lists that are created by combination of rank lists and resemblance probability distribution of users. The experimental results on a multimodal biometric system based on frontal face, profile face, and ear indicated higher identification rate by using novel confidence based rank level fusion.

Keywords: Multimodal biometrics · Rank level fusion · Rank list selection · Resemblance probability distribution

1 Introduction and Background

The undeniable need for higher security has resulted in a tremendous growth in biometric systems. The integration of biometric system with government identity management systems as well as consumer products has created the demand for more accurate systems.

Biometric systems can be categorized into unimodal and multimodal systems [1]. Although single biometric systems have been widely used in access control and identity management systems, there are some issues regarding their performance. The low inter-class and high intra-class variance caused by the use of one biometric as well as the non-universality, sensitivity to noise and data quality issues have shifted the attention toward multimodal biometric systems [2]. In a multimodal biometric system, different biometrics of a person are captured in order to provide uncorrelated information about the identity. These multiple evidences help the system to infer about the identity of the query with higher confidence [1].

One of the important aspects of a multimodal biometric system is the fusion of information from different biometric traits. The fusion can be done in pre-mapping stage where biometric traits have not matched against the training

G. Azzopardi and N. Petkov (Eds.): CAIP 2015, Part I, LNCS 9256, pp. 211–222, 2015.
DOI: 10.1007/978-3-319-23192-1_18

samples as well as post-mapping stage where the comparison is done and the recognition results need to be fused [3].

Post-mapping fusion can take place on three levels: score level, decision level, and rank level [1]. Score level deals with systems that provide score values to show the proximity of a query to the samples in the database. On decision level, each biometric classifier independently provides the final decision about the identity of the user. These levels of fusion are mostly used in commercial biometric systems which solely provide the final decision [4]. Unlike score level and decision level, some biometric devices provide a ranking of users as their output. The output of such systems is similar to the scores in case that it contains a list of possible identities, although it only provides a ranking of identities and lacks the rich score information. The lack of scores obscures the information on how confident each classifier is about the results. On the other hand, rank level fusion does not require normalization of scores which can be computationally expensive and also the inappropriate selection of normalization method can degrade the recognition rate [2]. Rank level fusion is a relatively new approach which has not been studied much compared to others. Several rank-level fusion methods have been proposed and developed in the past decade using different fusion techniques and biometric traits [5–7].

Lee et al. [8] compared rank-level methods, such as Borda count and Bayes fuse, and score level, such as sum rule and binary classification, based on fingerprint and face biometric traits. They concluded that the binary classification outperforms other methods and in case of lack of scores, Bayes fusion has an advantage over Borda count. Monwar et al. [9] showed that using more sophisticated approaches with rank information can result in a higher recognition rate than score level. They proposed a fuzzy rule based inference system for rank fusion. The comparison of fuzzy rank fusion with other rank, score, and decision level fusions demonstrated not even a better accuracy but a faster system performance.

The quality of biometric data can significantly impact the classifiers confidence and the recognition result, especially when there is a possibility that image quality degrades due to ambient conditions and acquisition device. Marasco et al. [10] investigated the stability of rank-level fusion and analyzed the performance of rank-level and score level fusion in presence of biometric data with low quality. Their experiments with low quality face images and syntactically degraded fingerprints revealed that both rank and score are unstable while the degradation is significant, although rank is more stable than scores in case of small degradations. Abaza and Ross [11] proposed a modification for highest rank and Borda count fusion by incorporating a quality factor. Alam et al. [12] utilized a quality measure which did not require any prior modeling of the noise and degradation factors of input data. They tried to extract the quality measure by considering a measure of deviation of scores from the mean and then utilizing this confidence measure for highest rank and Borda count methods. They developed a multimodal system for face and voice biometrics and validated the improvement impact of confidence measure by comparing their proposed method against highest rank and Borda count.

Previous works tried to improve the accuracy of the rank level fusion by considering different factors and using different approaches. Aiming at a same goal, this paper introduces a novel approach to user based rank list selection based on the confidence of rank lists. The main contribution of this paper is an adaptive selection of rank lists for any queries based on the rank lists performance for that specific query. This approach does not require score information to evaluate the confidence of classifiers. It calculates the confidence by recovering pseudo-score lists using the similarity of each user's resemblance probability distribution with the rank lists. The most important feature of this approach is the adaptive selection of rank lists based on their performance for each user. This system is advantageous since it can be adopted for different queries based on how different rank lists are performing. By modifying the rank lists using the novel confidence based rank list selection (CBRLS) approach, the fusion method is able to select the most confident rank lists and reach a higher identification rate.

This paper is organized as follows: Section 2 introduces the novel confidence based rank list selection by explaining the resemblance probability distributions, rank list confidence, and the whole cascade rank list selection approach. Section 3 validates the applicability of the confidence based rank list selection by providing experimental results. Section 4 concludes the paper.

2 Methodology

This section details the novel confidence based rank list approach. Fig. 1 shows a flowchart of the novel confidence based rank level fusion method. The fusion starts with rank lists that are created by the classifiers. Each rank list is converted to a pseudo-score list using the resemblance probability distributions. For each biometric, the confidence of the pseudo-score list is calculated and a fraction of each list is selected based on its confidence. Then, the algorithm finds the next rank list that provides the most confidence value for each of the restricted lists. The algorithm continues until some stopping criteria is met. This procedure results in a cascade of rank lists working together to provide the highest confidence to the list of users. This approach is novel in terms of confidence calculation from rank lists and also rank list selection to improve the confidence of rank lists. The rest of this section explains the detail of the proposed multimodal biometric system and the novel fusion approach.

2.1 Feature Extraction

The importance of a discriminative feature extraction in any machine learning systems is undeniable. Without features that provide discrimination between different classes of objects, it is impossible to reach a high recognition rate. Fisher Linear Discriminant Analysis (FLDA) [13] is a feature extraction and dimensionality reduction approach. In its essence, it projects data to a linear subspace that provides high inter-class variance as well as low intra-class similarity. It also

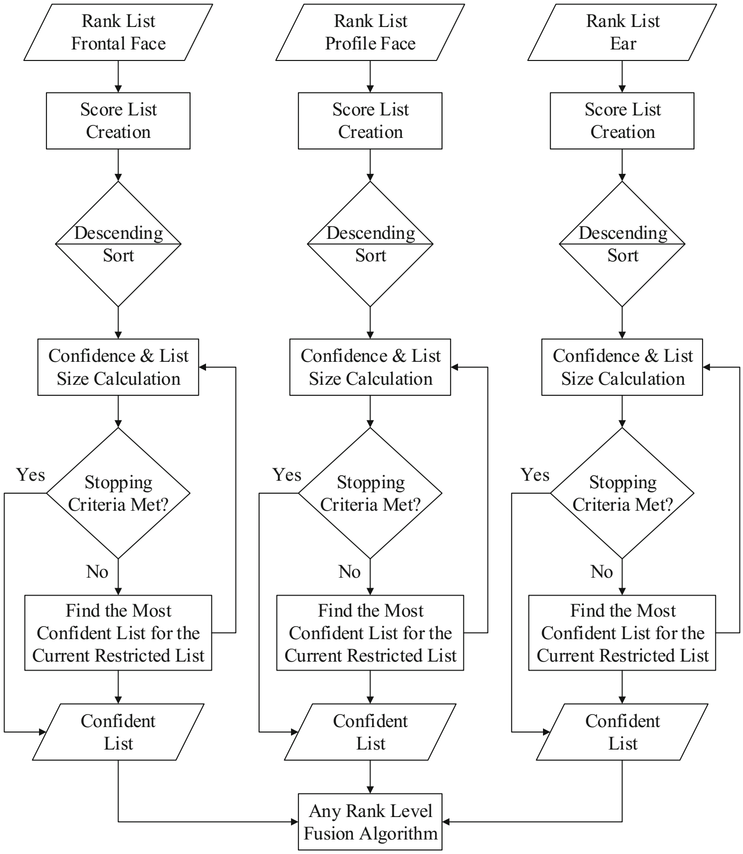


Fig. 1. Flowchart of the confidence based rank list selection (CBRLS) approach.

reduces the computation time by dimensional reduction with the least effect on the discriminatory information.

The FLDA [13] subspace for each biometric is created using the vectorised images of biometric traits accompanied with their corresponding classes. The FLDA algorithm provides the basis for the subspace and each biometric sample can be projected as a point to the FLDA subspace. The similarity of the samples can be considered as the Euclidean distance between FLDA feature vectors.

2.2 Resemblance Probability Distribution (RPD)

The FLDA projects training data into a feature space where each biometric sample becomes a point. For a specific biometric trait, each user's samples cluttered in this subspace. The configuration of all data points in the feature space creates the training data set. Based on the distances between data points of different classes, there are different probabilities to misclassify each class during the test phase. Formally speaking, if there are n different classes, namely $c_1, ..., c_n$ and each class c_i contains m training samples $s_{i,1}, ..., s_{i,m}$, then the distance from class c_i to c_j is defined as:

$$D_{b_j}(c_i, c_{i'}) = \begin{cases} 1/m \sum_{k=1}^{m} \min_{l \in \{1,...,m\}} \|(f_{s_k^{c_i,b_j}} - f_{s_l^{c_{i'},b_j}})\|, & \text{if } i \neq i' \\ 1/m \sum_{k=1}^{m} \min_{l \in \{1,...,m\}-\{k\}} \|(f_{s_k^{c_i,b_j}} - f_{s_l^{c_{i'},b_j}})\|, & \text{if } i = i' \end{cases} \tag{1}$$

The distance of class c_i to all the classes in the training data set can be used to create the Resemblance Probability Distribution (RPD) [14] as follows:

$$PRD(c_i) = \frac{\max(D_{c_i,b_j}) \times \mathbf{1} - D_{c_i,b_j}}{\|\max(D_{c_i,b_j}) - D_{c_i,b_j}\|} \tag{2}$$

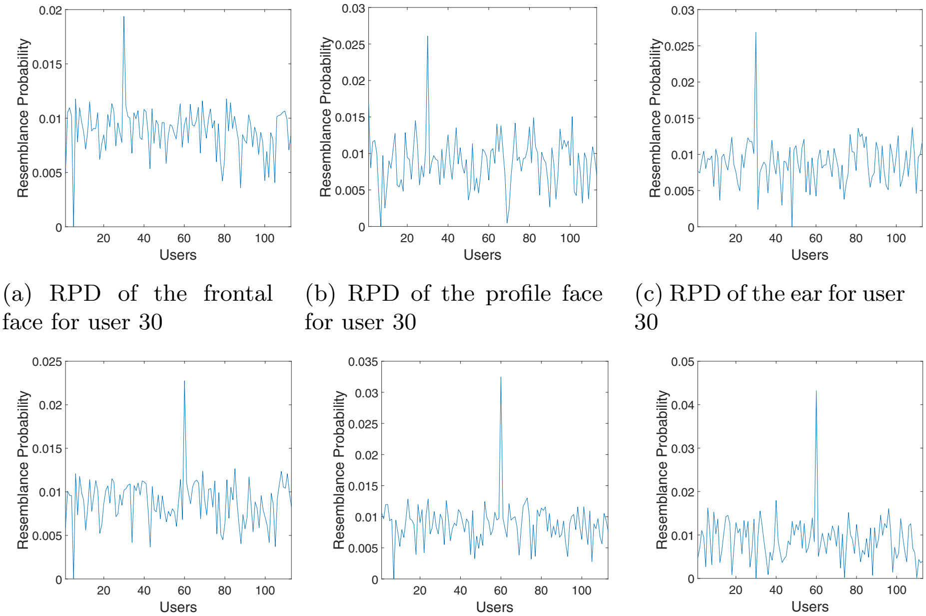
where D_{c_i,b_j} is a vector representing the distance of c_i from all the other classes and $\mathbf{1}$ is a vector of ones, the same size as D_{c_i,b_j}. The Resemblance Probability Distribution (RPD) [14] for each class shows how different classes resemble that specific class and manifests the probability of misclassifying a certain class with other classes.

Fig. 2 demonstrates the resemblance probability distribution for three biometrics of two users with identity 30 and 60. The RPDs for all three biometrics have their maximum values for the identity 30 or 60 which demonstrates the fact that they have the most resemblance to their actual identities. The RPDs have different values for other identities which demonstrates the independence of different biometrics [15].

2.3 Confidence Based Rank List Selection (CBRLS)

A rank list only provides a relative ordering of users based on their similarities to the queried sample. A rank list does not provide any information about how confident is the classifier about the ranking. On the other hand, score list can provide information about how confident is the classifier based on the scores assigned to each user. In order to calculate the confidence of each classifier about the rank list, there is a need to recover scores information. Here, we propose a novel method to convert rank lists to pseudo-score lists using the resemblance probability distributions. The confidence of each classifier can be calculated based on its pseudo-score list. Then, the novel rank list selection is used to create a cascade of rank lists and increase the confidence of rank lists.

Converting from Rank List to Pseudo-score List. Since the rank list and resemblance probability distributions are both generated from a same feature space and samples, it can be considered that the rank list should be similar to the resemblance probability distribution of its true identity or its neighbors in the feature space. In order to find how similar the rank list is to each resemblance probability distribution, we consider the rank list as a probability distribution and used the Bhattacharyya distance[16] to find the similarity of two probability



(a) RPD of the frontal face for user 30

(b) RPD of the profile face for user 30

(c) RPD of the ear for user 30

(d) RPD of the frontal face for user 60

(e) RPD of the profile face for user 60

(f) RPD of the ear for user 60

Fig. 2. Resemblance probability distribution for three biometrics of two users.

distributions. The Bhattacharyya distance of two discrete probability distributions P_1 and P_2 over a same domain X is defined as:

$$D_B(P_1, P_2) = -ln(BC(P_1, P_2)) \tag{3}$$

where,

$$BC(P_1, P_2) = \sum_{x \in X} \sqrt{P_1(x)P_2(x)} \tag{4}$$

The similarity measure based on the Bhattacharyya distance is defined as:

$$S_B(P_1, P_2) = 1 - D_B(P_1, P_2) \tag{5}$$

The rank list of each classifier can be converted to a probability distribution by dividing each element by the sum of all the ranks.

The transformation of a rank list to a pseudo-score list is done using algorithm 1. The algorithm calculated the Bhattacharyya similarity of the rank list with all the resemblance probability distributions and then multiply the similarity of each RPD of each user to the rank of that user in the rank list to create a pseudo-score list (PSL). At last, the pseudo-score list (PSL) is normalized by dividing by each pseudo-score by the sum of all the pseudo-scores.

for *each normalized rank list* NL_j **do**
  for *each enrolled identity* I_i **do**
    $PSL_i^j = (N - Rank_i^j) * S_B((1 - NL_j), RPD_i^j)$
  end
  $NPSL^j$=Normalize(PSL^j)
end

Algorithm 1. The process of creating pseudo-score list from rank list containing N users using resemblance probability distributions.

Confidence of Lists. The confidence measure is considered as how probable is that the current list misclassifies the actual user as its neighbors. For the normalized pseudo-score list $NPSL^j$ of biometric j, the confidence measure is defined as:

$$CM(NPSL^j) = \sum_{i=1}^{N-1} e^{\frac{-(i-1)^2}{N}} (SortedNPSL_i^j - SortedNPSL_{i+1}^j) \tag{6}$$

where N is the number of users registered to the system and $SortedNPSL^j$ is the decreasing sorted list of $NPSL^j$. This equation considers the confidence of each list as a weighted sum of differences of consecutive sorted pseudo-scores. The more the distance between the consecutive users' normalized sorted pseudo-scores, the higher is the confidence of that list. The weighting factor has larger values for score difference of users at the top of the list and the value decreases as it reaches to less probable identities. This weighting makes sense, since users at the beginning of the sorted list have more effect on the final decision. The confidence measure ranges from 0 to 1. The confidence measure is equal to 1, when the probability of first user is 1 and the rest is equal to 0. In this case, rank list is completely confident about its choice. The confidence measure is equal to 0 when the probabilities of all users are equal to a same number. In this case, the rank list cannot distinguish between users.

Restricted List Creation. The restricted lists are created based on the confidence of each list. If a list has a low confidence value, it means that more users should be on the restricted list. If a list is confident about its ranking, then we can consider a less number of users can be on the restricted list. The number of users in the restricted list of a rank list with N users and the confidence measure of CM is defined as:

$$k = \max\{(1 - CM)(N - 1) + 1, n\} \tag{7}$$

where n is the minimum number of users that the restricted list should maintain. Using threshold value n protects the algorithm from losing the actual user during the restricted list creation due to accidental wrong confidence measure.

Increasing Lists Confidence Using Confidence Based Rank List Selection. For each biometric trait, the rank list is first converted to a pseudo-score list using resemblance probabilities. Then, for each list, the confidence level and the restricted list size are calculated. For each list, based on the restricted list size, the first k users are kept and then the confidence level of all the rank lists for these k users are calculated. The rank list, which provides the highest confidence for the current restricted list is used for the next round. The algorithm continues the same process for all the lists till there is no more changes in the lists and the confidence levels reach to its highest values. In this situation, any rank level fusion approach can be applied to the lists to create the consensus list.

3 Experimental Results

The performance evaluation of the proposed system is essential in order to assure its effectiveness for real-world scenarios. Even though the confidence based rank list selection algorithm is not dependent on any particular biometrics and is able to work with any sets of biometric traits, it has its best performance in case there is the least correlation between selected biometric traits. In order to demonstrate the performance of the system in case of correlated traits, we have selected frontal face and profile face as two of correlated biometric traits and ear as the third one.

Due to inherit cost associated with creating a real multimodal biometric database, most multimodal biometric systems are evaluated based on virtual databases. A virtual database is constructed by pairing of users from different biometrics databases and creating a virtual user that borrows each biometric from a specific database. In this work we have created a virtual database using frontal face, profile face and ear.

Facial Recognition Technology (FERET) database [17] is used for both frontal and profile faces. The collection of this database in done by researchers at George Mason University sponsored by The U.S. Department of Defense. FERET is a widely used database due to the collection of facial images with various positions, illuminations, and expressions. Since we were only interested to frontal and profile faces, we exclude images that were not taken from these angles. The database consist of 1199 subjects with multiple images of each subject. Among all the subjects there are only 974 subjects with profile face images.

For ear, the USTB ear database [18] is used. All the subjects were students from University of Science and Technology Beijing. They gathered three different databases based on three different conditions and configurations. In total there are 216 subjects among 1276 ear images that were taken in different illumination and orientation.

In order to create the virtual database, a user from the FERET database (which consists of frontal and profile faces) is selected and randomly paired with a user from the USTB ear database to form a virtual person. Enrollment and identification databases are created by randomly selecting one or two image(s)

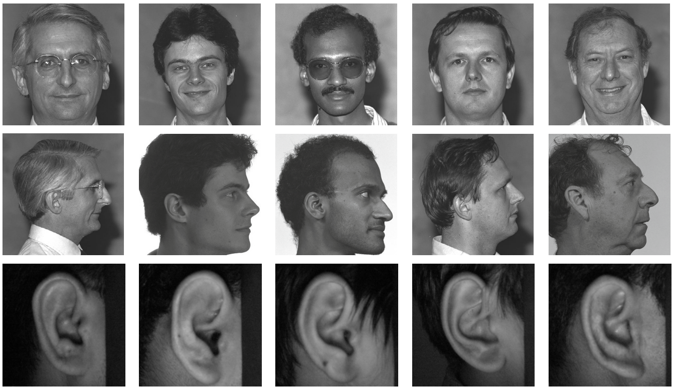


Fig. 3. Samples of virtual multimodal database.

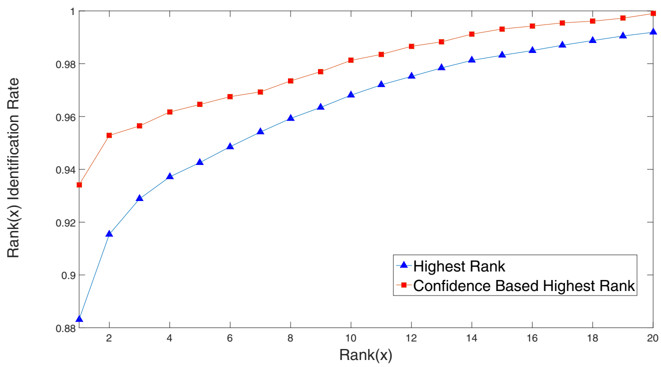
Table 1. First rank identification rates of rank level fusions for k=1 and k=2

Fusion Method	Identification % for k=1	Identification % for k=2
Highest rank	88.32%	88.51%
Confidence based highest rank	93.42%	93.79%
Borda count	90.92%	91.35%
Confidence based Borda count	95.71%	96.12%
Logistic regression	92.32%	92.74%
Confidence based logistic regression	96.89%	97.41%

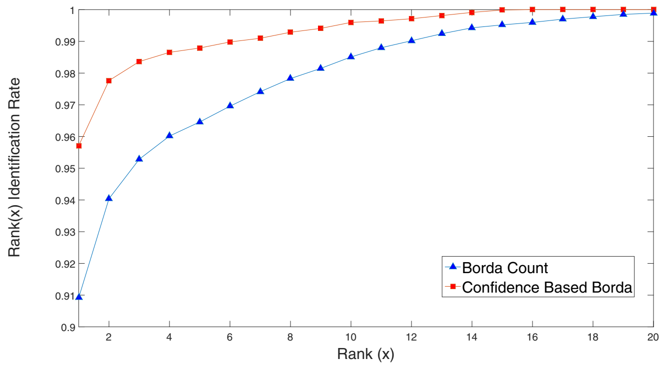
(depending on the availability) from each biometric of a virtual users for enrollment and taking the rest for the identification phase. Fig. 3 shows five samples of virtual users created using this approach. By following this approach, we have created 2,210 virtual persons in 10 different virtual databases where each database contains 216 individuals. Test queries for identification phase are created by pairing of test images from biometrics of each virtual person. For the performance evaluation of proposed method 98,500 test queries have been created.

Rank lists are created using the k-nearest neighbors classifier [19] for k equals 1 and 2. Table 1 demonstrates that there is a minor improvement in identification rate by using $k = 2$. Since there is not much differences between the identification rates for $k = 1$ and $k = 2$, we have considered $k = 1$ for the rest of experiments.

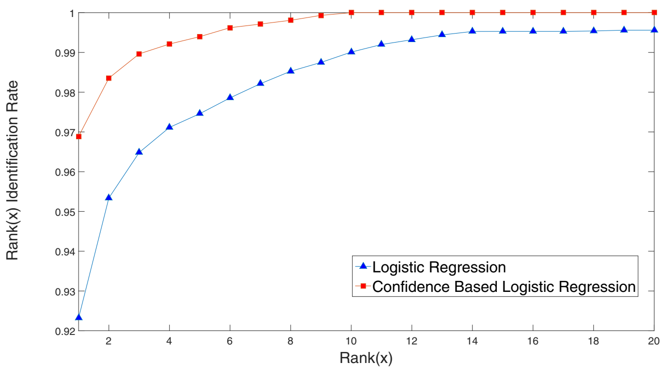
To demonstrate the performance of our fusion method, we used cumulative match characteristic (CMC) curve, which is a commonly used metric for evaluation of biometric systems. CMC is used to show the identification rate of the system at different ranks. The identification rate is the portion of times the system recognizes the true identity of the users. The CMC value for rank k is the



(a) CMC curves for highest rank with and without CBRLS.



(b) CMC curves for Borda count with and without CBRLS



(c) CMC curves for logistic regression with and without CBRLS

Fig. 4. CMC curves for multimodal biometrics rank fusion with and without the use of confidence based rank list selection.

portion of times the true identity is among the first k users listed as the final result of the system.

Fig. 4 shows the CMC curves for highest rank, Borda count, and logistic regression with and without confidence based rank list selection (CBRLS). The comparison of three CMC curves for fusion without CBRLS reveal that logistic regression outperforms Borda count and highest rank while Borda count performs better than highest rank. The comparison of CMC curves in each subfigure shows that CBRLS increases the first rank identification rate of highest rank by 5.1%, Borda count by 4.8%, and logistic regression by 4.6%. The highest first rank identification rate obtained using the system is 96.9% for CBRLS with logistic regression.

The experimental results are indicator of improvement of rank level fusions by incorporating confidence based rank list selection (CBRLS). The usage of CBRLS is not limited to the experimented fusion methods and it can be applied to any rank level fusion approach to improve the identification rate.

4 Conclusion

In this paper, a novel user based rank list selection for fusion of biometrics in rank level was introduced. The proposed algorithm recovered scores information by combining the rank list and the resemblance probability distribution of users. It calculated the confidence of each list based on pseudo-scores and selects rank lists based on the confidence they provide for any specific test query. The most important features of this approach are recovering the scores which are valuable information and also selecting the rank lists based on their performance for each user. The experimental results show the ability of the confidence based rank list selection to improve the identification rate for three well-known rank level fusions, i.e. highest rank, Borda count, and logistic regression. The highest first rank identification rate of the system was 96.9% for logistic regression with confidence based rank list selection which had improved the solo logistic regression by 4.6%. The confidence based rank list selection can be used with any rank level fusion method.

Acknowledgments. The authors would like to thank NSERC DISCOVERY program grant RT731064, URGC, NSERC ENGAGE and Alberta Ingenuity for partial support of this project.

References

1. Ross, A.A., Nandakumar, K., Jain, A.K.: Handbook of multibiometrics. Springer Science & Business Media (2006)
2. Jain, A.K., Flynn, P., Ross, A.A.: Handbook of biometrics. Springer Science & Business Media (2007)
3. Revett, K.: Behavioral biometrics: a remote access approach. John Wiley & Sons (2008)

4. Prabhakar, S., Jain, A.K.: Decision-level fusion in fingerprint verification. Pattern Recognition **35**(4), 861–874 (2002)
5. Mowar, M., Gavrilova, M.L.: Multimodal biometric system using rank-level fusion approach. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics **39**(4), 867–878 (2009)
6. Bhatnagar, J., Kumar, A., Saggar, N.: A novel approach to improve biometric recognition using rank level fusion. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp 1–6. IEEE (2007)
7. Kumar, A., Shekhar, S.: Personal identification using multibiometrics rank-level fusion. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews. **41**(5), 743–752 (2011)
8. Lee, Y.-J., Lee, K.-H., Jee, H., Gil, Y.H., Choi, W., Ahn, D., Pan, S.B.: Fusion for multimodal biometric identification. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 1071–1079. Springer, Heidelberg (2005)
9. Monwar, M.M., Gavrilova, M., Wang, Y.: A novel fuzzy multimodal information fusion technology for human biometric traits identification. In: 2011 10th IEEE International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), pp. 112–119. IEEE (2011)
10. Marasco, E., Abaza, A., Cukic, B.: Why rank-level fusion? and what is the impact of image quality?
11. Abaza, A., Ross, A.: Quality based rank-level fusion in multibiometric systems. In: IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, BTAS 2009, pp. 1–6. IEEE (2009)
12. Alam, M.R., Bennamoun, M., Togneri, R., Sohel, F.: Confidence-based rank-level fusion for audio-visual person identification system. In: 3rd International Conference on Pattern Recognition Applications and Methods, 2014, pp. 608–615 (2014)
13. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 711–720 (1997)
14. Talebi, H., Gavrilova, M.: Prior resemblance probability of users for multimodal biometrics rank fusion. In: IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015). IEEE (2015)
15. Monwar, M., Gavrilova, M.: Fes: a system for combining face, ear and signature biometrics using rank level fusion. In: Fifth International Conference on Information Technology: New Generations, 2008, pp. 922–927. IEEE (2008)
16. Bhattacharyya, A.: On a measure of divergence between two multinomial populations. Sankhyā: The Indian Journal of Statistics, 401–406 (1946)
17. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The feret database and evaluation procedure for face-recognition algorithms. Image and Vision Computing **16**(5), 295–306 (1998)
18. USTB ear database, china. http://www.ustb.edu.cn/resb/ (accessed May 11, 2008)
19. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory **13**(1), 21–27 (1967)