Pool Adjacent Violators Based Biometric Rank Level Fusion

Nanang Susyanto[1]

Abstract: We propose a new method in rank level fusion for biometric identification. Our method is based on the pool adjacent violators (PAV) algorithm after the ranks have been transformed to the approximated scores. We then show that our method outperforms various approaches that commonly used in biometric rank level fusion on NIST BSSR1 multimodal database.

Keywords: rank level fusion, pool adjacent violators, biometric fusion, biometric identification, multimodal.

1 Introduction

Biometric fusion is a combination of several biometric systems or algorithms that aims to improve the performance of the individual system or algorithm. It can be divided into six categories [RNJ06]: multi-sensor, multi-algorithm, multi-instance, multi-sample, multi-modal and hybrid. Several studies show the performance improvement by combining information from multiple traits or algorithms [LWJ03, RJR02, RNJ06, Ul06]. For instance, Lu et al. [LWJ03] combine three different feature extractions (Principle Component Analysis, Independent Component Analysis and Linear Discriminant Analysis) while Prabhakar and Jain [PJ02] in the fingerprint biometric field use the left and right index fingers to verify an individual's identity.

Biometric fusion can be done at the sensor, feature, match score, rank and decision levels either for verification or identification. In this paper, we will focus on the rank level for closed identification problem in the sense that the unknown person is assumed to be one of the people in a given enrollment set. This scenario is suitable for combining ranked identities from commercial biometric devices that may only produce the ranked identities of the users instead of matching scores because of a security reason. This means that ranks from multiple biometric classifiers of every unknown person in a given enrollment set are transformed to a new rank and this new rank is used to assign the identity of that unknown person.

There are several methods that are commonly used in biometric rank level fusion: Borda count, weighted Borda count, maximum rank method, Bucklin majority voting [Po14], and some nonlinear weighted ranks [KS11]. While maximum rank takes the highest rank amongst all the matchers, the remaining methods use weight to represent the contribution of each classifier. The present paper uses the pool adjacent violators (PAV) algorithm to

---

[1] Faculty of Mathematics and Natural Sciences (FMNS), Department of Mathematics, Universitas Gadjah Mada, Sekip Utara Yogyakarta, Indonesia, nanang_susyanto@ugm.ac.id

compute the likelihood ratio (LR) of any rank after it has been transformed to its approximated similarity score for every classifier and combine the classifiers by summing their individual LRs up to get the final score. This final score will represent the combined similarity score. The rest of this paper is organized as follows. Section 2 gives a detailed explanation how the proposed method woks. Several examples using NIST BSSR1 database are provided in Section 3. Finally, this paper will be closed by our conclusions in Section 4.

2   PAV-based Method

This section will explain how our proposed method is built. In principle, there are two steps: (1) transforming ranks to their approximated similarity scores and (2) applying the PAV to these transformed scores.

2.1   Transforming Ranks to Approximated Scores

Let x be an unknown subject that belongs to the enrollment set $E = \{x_1, \ldots, x_n\}$. Of course, the original similarity scores of x and all elements E contain much more information than the ranked identity of all elements E with respect to the closeness to x. While Susyanto et al. [Su16a, Su16b] use a modified empirical distribution function to transform similarity scores to their corresponding uniformly distributed scores, which are only a scale of their ranks, to model dependence between classifiers, we will work on the other direction, i.e, approximating the uniformly distributed similarity scores from their ranks. Suppose that there are n_{train} identities in the enrollment set in *training* data. Since the rank-i has to have the i-highest probability for every $i = 1, \ldots, n_{\text{train}}$, we set it to have probability $(n_{\text{train}} + 1 - i)/n_{\text{train}}$. It means that the estimated probabilities run from $1/n_{\text{train}}$ to 1, which is already shown in [Su16a, Su16b] that they are uniformly distributed. Below is an example how the approximated similarity scores of the training set with subjects s_1, s_2, s_3, s_4, and s_5 are obtained from the ranks.

$$
\begin{pmatrix}
enr. & s_1 & s_2 & s_3 & s_4 & s_5 \\
s_1 & 1 & 2 & 3 & 4 & 5 \\
s_2 & 1 & 4 & 3 & 5 & 2 \\
s_3 & 2 & 5 & 1 & 3 & 4 \\
s_4 & 5 & 4 & 3 & 2 & 1 \\
s_5 & 5 & 2 & 3 & 4 & 1
\end{pmatrix}
\mapsto
\begin{pmatrix}
enr. & s_1 & s_2 & s_3 & s_4 & s_5 \\
s_1 & 1 & 4/5 & 3/5 & 2/5 & 1/5 \\
s_2 & 1 & 2/5 & 3/5 & 1/5 & 4/5 \\
s_3 & 4/5 & 1/5 & 1 & 3/5 & 2/5 \\
s_4 & 1/5 & 2/5 & 3/5 & 4/5 & 1 \\
s_5 & 1/5 & 4/5 & 3/5 & 2/5 & 1
\end{pmatrix}
\tag{1}
$$

When we are working in a testing data that contains more that n_{train} in its enrollment set, then we map all ranks greater than n_{train} to 0. Mathematically, whenever the training data T with n_{train} identities in its enrollment set is given and the subjects $x_1, \ldots, x_n$ has rank $r_1, \ldots, r_n$, respectively, with respect to the unknown subject x, the approximation of the uniformly distributed scores $s_1, \ldots, s_n$ will be

$$
s_i = \max\left\{\frac{n_{\text{train}} + 1 - i}{n_{\text{train}}}, 0\right\}
\tag{2}
$$

for every $i = 1, \ldots, n$. For example, if $x_1, x_2, x_3, x_4, x_5, x_6$, and x_7 in the testing set has ranks $2, 3, 1, 4, 5, 6$, and 7, respectively, with respect to x_3 then the approximated similarity scores using training data (1) is

$$\begin{pmatrix} enr. & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ x_3 & 4/5 & 3/5 & 1 & 2/5 & 1/5 & 0 & 0 \end{pmatrix}. \tag{3}$$

2.2 PAV-based Naive Bayes Fusion

Once we have had the approximated similarity scores of the training data, we can split them into *genuine* and *impostor* scores. A genuine score is the score obtained by comparing a pair of biometric samples originating from the same person while an impostor scores obtained by comparing a pair of biometric samples stemming from different people. In our example given by (1), the elements on the main diagonal of the second matrix are genuine while the elements off the main diagonal are impostor. The key of our method is computing the loglikelihood (LLR) of the approximated similarity scores using independence assumption. Even though it is not realistic, some experimental results show that its performance is still promising [TV13]. To do that, we need to compute the individual LLRs of the classifiers. The most common methods are Kernel Density Estimation (KDE), Logistic Regression (Logit), Histogram Binning (HB), and Pool Adjacent Violators (PAV); see [ASV12] for a brief explanation of these methods.

In this paper, we choose the PAV method because of its optimality [ZE02]. For every classifier $k = 1, \cdots, d$, (d is the number of classifiers), PAV sorts and assigns a posterior probability of 1 and 0 to the k-th component of genuine and impostor scores, respectively, in a given training set. It then computes the non-monotonic adjacent group of probabilities and replaces it with average of that group. This procedure is repeated until the whole sequence is monotonically increasing which estimates the posterior probability $P(H_1|(\cdot))$ of the k-th component of genuine and impostor scores where H_1 correspond to a genuine score. By assuming

$$P(H_1) = \frac{n_{\text{gen}}}{n_{\text{gen}} + n_{\text{imp}}},$$

the corresponding LR_ks of genuine and impostor scores can be computed according to the Bayesian formula by

$$\widehat{\text{LR}}_k(\cdot) = \frac{P(H_1|(\cdot))}{1 - P(H_1|(\cdot))} \times \frac{n_{\text{imp}}}{n_{\text{gen}}} \tag{4}$$

where n_{gen} and n_{imp} are the number of genuine and impostor scores, respectively. This gives a numerical function that maps score to its $\widehat{\text{LR}}_k$ so that for every score from the k-th classifier, its corresponding $\widehat{\text{LR}}_k$ value can be computed by interpolation. Finally, the final approximated similarity score is just the sum of all $\widehat{\text{LR}}_k$s for $k = 1, \ldots, d$.

3   Experimental Results

This section gives the comparison between the proposed method and the existing methods in rank level fusion (Borda count, weighted Borda count, maximum rank method, Bucklin majority voting [Po14], and some nonlinear weighted ranks [KS11]) on NIST BSSR1 database [Na04]. The NIST-BSSR1 database has three different set:

- NIST-Multimodal: Two fingerprints and Two face matchers applied to 517 subjects,

- NIST-Face: Two face matchers applied to 3000 subjects,

- NIST-Finger: Two fingerprints applied to 6000 subjects.

We will use the same protocol as used in [KS11] (Protocol 1 and 2) and an additional protocol (Protocol 3). The exp(1) and exp(2) are the methods proposed in [KS11]. Note that the maximum rank, the Borda count, and the Bucklin majority voting methods do not need training data while our proposed method needs training data as the weighted Borda count and nonlinear weighted ranks [KS11]) do.

3.1   Multi-instance Test: Protocol 1

In this experiment, we use the NIST-Finger database containing 6000 subject where the first 1000 subject were used for training our proposed method and the rest were used for testing. The comparison of our method with the other methods is presented in Table 1. We can see that the rank-1 of our method jumps from the best existing methods (exp(2): 89.56%) to 94.44%.

Tab. 1: Performance (in %) From NIST-Finger Database (6000 Subjects). The bold face in every row is the best one.

	Highest Rank	Borda	Weighted Borda	Bucklin	exp(1)	exp(2)	Proposed
rank-1	82.57	85.65	87.74	74.58	89.34	89.56	**94.44**
rank-2	94.43	86.68	89.04	88.23	93.98	94.42	**95.00**
rank-3	94.48	87.33	89.74	93.65	95.22	95.20	**95.34**

3.2   Multi-modal and Multi-algorithm Test: Protocol 2

Using the same protocol as in [KS11], we put the first 100 subject of the NIST-Multimodal database for training and the rest (417) for testing. We can see from Table 2 that our proposed method outperforms the other methods even it attains 100% recognition rate at rank-1.

Tab. 2: Performance (in %) From NIST-Multimodal Database (517 Subjects). The bold face in every row is the best one.

	Highest Rank	Borda	Weighted Borda	Bucklin	exp(1)	exp(2)	Proposed
rank-1	80.66	91.68	94.39	88.78	98.84	99.28	**100.00**
rank-2	96.32	93.81	95.55	98.84	99.42	99.76	**100.00**
rank-3	100.00	94.97	96.32	99.81	100.00	100.00	**100.00**

3.3   Multi-modal and Multi-algorithm Test: Protocol 3

In order to make a larger database for testing, we make a virtual database by taking the first image of every person in NIST-Face database and the fist 3000 subjects in NIST-Finger. As the results, our virtual database contains 3000 subjects in which every subject has 2 scores from face comparisons, 1 score from left-index finger comparison, and 1 score from right-index finger comparison. By using the same training data as in Protocol 2, we can see from Table 2 that the highest rank, Borda count, weighted Borda count, and Bucklin methods do not perform better than the exp(1) and exp(2) methods. Therefore, we will only compare our method with the exp(1) and exp(2) methods. The cumulative match curve is provided by Figure 1 thats shows clearly that the proposed method does outperform the exp(1) and exp(2). The recognition rate at rank-1 of the exp(1), exp(2), our proposed method are 95.73%, 91.97%, and 98.87%, respectively.
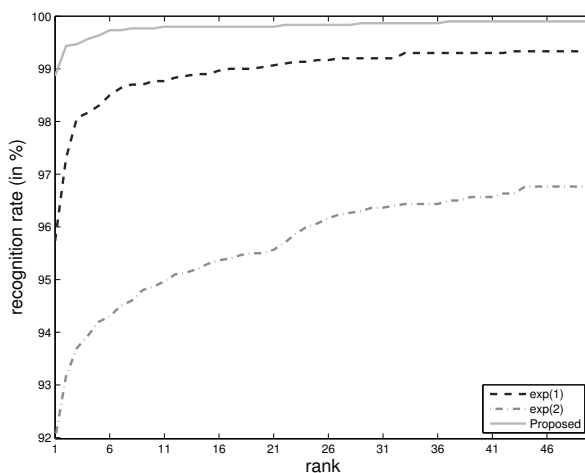

Fig. 1: CMC of the exp(1), exp(2), and the Proposed Method

4   Conclusion

We have proposed a new method in biometric rank level fusion via pool adjacent violators (PAV). The method can be done by two main steps: (1) transforming ranks to their approxi-

mation of the uniformly distributed similarity scores and (2) applying the PAV of the transformed scores for every classifier and simply taking the naive Bayes fusion. It has been demonstrated that our proposed method outperforms the Borda count, weighted Borda count, maximum rank method, Bucklin majority voting, and some nonlinear weighted ranks in every scenario using the NIST BSSR1 database.

References

[ASV12] Ali, T.; Spreeuwers, L. J.; Veldhuis, R. N. J.: Forensic Face Recognition: A Survey. In (Quaglia, A.; Epifano, C. M., eds): Face Recognition: Methods, Applications and Technology, Computer Science, Technology and Applications, p. 9. Nova Publishers, 2012.

[KS11] 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, Sept 2011.

[LWJ03] Lu, Xiaoguang; Wang, Yunhong; Jain, A.K.: Combining classifiers for face recognition. In: Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on. volume 3, pp. III–13–16 vol.3, July 2003.

[Na04] National Institute of Standards and Technology: , NIST Biometric Scores Set - release 1, 2004. Available at http://www.itl.nist.gov/iad/894.03/biometricscores.

[PJ02] Prabhakar, Salil; Jain, Anil K.: Decision-level fusion in fingerprint verification. Pattern Recognition, 35(4):861 – 874, 2002.

[Po14] Porter, Melvin P.: Preferential voting and the rule of the majority. National Municipal Review, 3(3):581–585, 1914.

[RJR02] Ross, A.; Jain, A.; Reisman, J.: A hybrid fingerprint matcher. In: Pattern Recognition, 2002. Proceedings. 16th International Conference on. volume 3, pp. 795–798 vol.3, 2002.

[RNJ06] Ross, Arun; Nandakumar, Karthik; Jain, Anil K.: Handbook of Multibiometrics (International Series on Biometrics). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[Su16a] Susyanto, N.; Veldhuis, R. N. J.; Spreeuwers, L. J.; Klaassen, C. A. J.: Fixed FAR Correction Factor of Score Level Fusion. In: The 8th IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS2016), 6-9 September, Buffalo, NY, USA. September 2016.

[Su16b] Susyanto, N.; Veldhuis, R. N. J.; Spreeuwers, L. J.; Klaassen, C. A. J.: Two-step Calibration Method for Multi-algorithm Score-based Face Recognition Systems by Minimizing Discrimination Loss. In: The 9th IAPR International Conference on Biometrics (ICB 2016), 13-16 June, Halmstad. June 2016.

[TV13] Tao, Q.; Veldhuis, R. N. J.: Robust Biometric Score Fusion by Naive Likelihood Ratio via Receiver Operating Characteristics. IEEE Transactions on Information Forensics and Security, 8(2):305–313, February 2013.

[Ul06] Ulery, Brad; Hicklin, Austin; Watson, Craig; Fellner, William; Hallinan, Peter; Gutierrez, Carlos M: Studies of Biometric Fusion NISTIR 7346. 2006.

[ZE02] Zadrozny, Bianca; Elkan, Charles: Transforming Classifier Scores into Accurate Multiclass Probability Estimates. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '02, ACM, New York, NY, USA, pp. 694–699, 2002.