Multi-Modal Biometrics based on Data Fusion

Hongxun Yang1,a*, Eason Sun2, Cheng Cheng3, Anthony H Ding4
1Shanghai World Foreign Language Academy, Shanghai, China
2Shanghai Qibao Dwight Highschool, Shanghai, China
3Shanghai Starriver Bilingual School, Shanghai, China
4Shanghai Highschool International Division Shanghai, China
*a*Corresponding author frank_yhx_4314@126.com

Abstract—With the development of intelligent application, biometrics recognition technology has been widely concerned and applied in many fields of the real world, such as access control and payment. The traditional biometrics are usually based on single modality data of the subjects, but they are limited by the feature information capacity and the bottleneck in recognition accuracy. In this paper, a multi-modal biometric recognition framework is presented, which utilizes a multi-kernel learning algorithm to fuse heterogeneous information of different modal data. In order to extract complementary information from them, we combine the kernel matrix to form the mixed kernel matrix, and then give the final classification results. The experimental results on multiple biometric datasets show that our method can obtain higher recognition accuracy compared with the existing single mode and multi-mode fusion methods.

1. INTRODUCTION

Biometric identification [1] technology uses feature analysis to identify physiological or behavioral data, such as face recognition [2], fingerprint recognition [3], and it has been widely used in the areas such as security. Recently, with the continuous progress and development of the society, the requirement of recognition accuracy is getting higher. In many real-world applications, the biometric information in the complex environment may be disturbed or damaged. Therefore, the single modal based biometric identification have been difficult to meet the needs of the accuracy and applicability [4]. Compared with the single modal based biometric method, the multi-modal biometric identification [5] method can meet the above requirements and improve the recognition performance.

Multi-modal biometric recognition collects multiple kinds of the features from the same individual, and then uses feature extraction and fusion to get the final recognition result [6]. Through the fusion of different feature information, the recognition rate can be improved effectively, and the influence of noise can be reduced. The biometric information of multi-mode is not easy to be obtained simultaneously, so it has relatively high security [7].

For more than twenty years, many multi-modal biometric fusion algorithms have been proposed. For example, Jain [8] have made many outstanding contributions in the field of multimodal biometric identification. Brunelli et al. [9] proposed a method for facial and speech fusion to realize multi-biological feature fusion recognition. Multimodal biometric recognition is a hot area of research, but some challenges have not been well addressed, restricting their application.
In this paper, we proposed a novel multi-modal based biometric method, which fuses the features from different modalities by multi-kernel combination. The proposed method can extract the complementary information among different modalities, and its calculation efficiency is high. It is suitable for the task with two or more biometric modalities. The experimental results on multi-modal datasets of palmprint and knuckle show that our method is superior to the single modal method and the existing fusion methods.

The rest of the article contains the following sections. We introduce the related work about data fusion in section II. The framework for multi-modal biometric is presented in section III. The section IV gives the results on several datasets. Finally, the conclusion is drawn in section V.

2. RELATED WORK
According to the different fusion levels, multi-mode biometric fusion is generally divided into four fusion methods, namely sensor layer fusion, feature layer fusion, fractional layer fusion and decision layer fusion.

Sensor layer fusion [10], also known as pixel-level fusion, it refers to the direct processing and fusion of the biometric information collected by the sensor. Because the original image data is processed and fused, there is a large gap between different biometric data, making it difficult to be fused. Feature layer fusion [11] refers to preprocessing the feature information of different organisms before feature extraction to generate corresponding feature coding. Finally, according to different fusion methods, the feature template is updated and splicing for subsequent identity authentication or recognition. Matching level fusion [12] refers to the fusion strategy based on the score layer to obtain a matching score after feature extraction, and then make subsequent identification and authentication. Decision level fusion [13] judges the result of the whole multi-modal recognition based on the combination of the output of each modality.

This paper mainly focuses on the feature layer fusion, which considers the advantages of the matching layer and the sensor layer. On the one hand, it has better compression data performance and improves the fusion speed. On the other hand, it retains as much feature information as possible, so as to make the identification performance more accurate and efficient. The main technical difficulty is the mismatch between dimensional disasters and feature spaces.

3. THE PROPOSED MULTI-MODAL BIOMETRIC RECOGNITION
The common multi-modal learning method based on the kernel method is used here to jointly utilize data information of multiple modals, so that it can obtain better experimental results than using a single modal. This kernel-based multi-modal learning method can be easily embedded in traditional SVM classifiers for high-dimensional pattern classification without the need for additional steps. In addition, unlike other combinations that can only handle one data type (i.e., numeric data type), our approach can combine multiple data types, such as numeric data, strings, and graphs.

Before introducing the kernel combination method, the single kernel SVM algorithm is briefly introduced. The main idea of SVM algorithm is as follows. First, by using a kernel function, the linear non-separable samples are mapped from their original space to the feature space of higher or even infinite dimensions, so that the samples are more likely to be linearly separable in the higher dimensional space. Then, the maximum edge hyperplane is found in the high dimensional space to divide different types of samples.

Now, we will specifically introduce the steps of the method of fusing multi-modal data. Suppose that we are given n training samples and each of them is of M modalities. Let denote a feature vector of the m-th modality of the i-th sample, and its corresponding class label be . Multiple kernel based SVM solves the following primal problem:

\[
\min_{w^{(m)}, b, \epsilon_i} \frac{1}{2} \sum_{m=1}^{M} \beta_m \| w^{(m)} \|^2 + C \sum_{i=1}^{n} \epsilon_i \\
\text{subject to } y_i (\sum_{m=1}^{M} \beta_m (w^{(m)^T} \phi^{(m)} (x_i^{(m)}) + b)) \geq 1 - \epsilon_i, \epsilon_i \geq 0, i = 1, 2, ..., n
\]
where \( \Psi^{(m)} \) and \( \Phi^{(m)} \) denote the normal vector of hyperplane, the kernel-induced mapping function, and the combining weight on the m-th modality, respectively. The dual form of multiple kernel SVM can be represented as below:

\[
\begin{align*}
\max_\alpha \sum_{i=1}^n \alpha_i &= \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_{m=1}^M \beta_m k^{(m)}(x_i^{(m)}, x_j^{(m)}) \\
\text{s.t.} \sum_{i=1}^n \alpha_i y_i &= 0, 0 \leq \alpha_i \leq C, i = 1, 2, ..., n
\end{align*}
\]

where \( k^{(m)}(x_i^{(m)}, x_j^{(m)}) = \phi^{(m)}(x_i^{(m)})^T \phi^{(m)}(x_j^{(m)}) \) is the kernel function for the two training samples on the m-th modality, and n is the number of training samples.

![Diagram of multi-modal biometric method with multi-kernel method](image)

Figure 1. The framework of multi-modal biometric method with multi-kernel method

For a test sample \( x = \{x^{(1)}, x^{(2)}, ..., x^{(M)}\} \), we first denote the kernel between new test sample and each training sample on the m-th modality, which is formulated as \( k^{(m)}(x_i^{(m)}, x^{(m)}) = \phi^{(m)}(x_i^{(m)})^T \phi^{(m)}(x^{(m)}) \). Then, the decision function for the predicted label can be obtained as below:

\[
f(x^{(1)}, x^{(2)}, ..., x^{(M)}) = \text{sign}(\sum_{i=1}^n y_i \sum_{m=1}^M \beta_m k^{(m)(x_i^{(m)}, x^{(m)})} + b)
\]

We interpret \( k(x_i, x_j) = \sum_{m=1}^M \beta_m k^{(m)(x_i^{(m)}, x_j^{(m)})} \) as a mixed kernel between the multimodal training samples \( x_i \) and \( x_j \), and \( k(x_i, x) = \sum_{m=1}^M \beta_m k^{(m)(x_i^{(m)}, x^{(m)})} \) as a mixed kernel between the multimodal training sample \( x_i \) and the test sample \( x \), so the multi-kernel based SVM can be naturally embedded into the conventional single-kernel SVM. All in all, our approach is essentially to combine multiple kernels into a single kernel.

It is worth noting that the multi-kernel learning method we use does not jointly optimize the weight \( \beta_m \) and other SVM parameters (e.g., \( \alpha \)) in an iterative manner. On the contrary, we constrain...
and through cross-validation on the training samples by grid search to find the best $\beta_m$ value. After we get the optimal $\beta_m$ value, we use them to combine multiple kernels into a mixed kernel, and then use this mixed kernel to perform standard SVM. Noteworthy, our method can be conveniently implemented using traditional SVM solvers, for example, LIBSVM.

Figure 1 shows the framework of multi-modal biometric method with multi-kernel method. The framework used in this paper mainly contains the following steps. First, we obtain the image or other types of the data from each modality. Second, we compute the kernel matrix for each modality. The element of kernel matrix is computed by the kernel function mentioned above. Then, the combined kernel matrix can be easily obtained by the linear combination of the kernel matrices of different modalities. Third, we extract the feature from the combined matrix with SVM. Finally, the classification decision is obtained.

We can see from the above, the framework used in this paper has the following two advantages. For one hand, it is applicable to data with different dimensions. For example, if the image sizes of face image and the palmprint are different, which is often the case in real-world, the framework can deal with them. Therefore, the method in this paper has a wide range of applications. As mentioned above, this kernel combination method can provide a convenient and effective method for fusing various data from different modalities. Through this multi-kernel learning method, useful information of multiple modal data will be used jointly to obtain better experimental results.

4. Experiment

4.1. Datasets

In this paper, we mainly investigate the multimodal problem in biometric identification. Naturally, in the real world, we can identify by different characteristics. For example, Face recognition and palmprint recognition have been widely used for identification. Since each kind of feature has specific identification information, it is of great significance to focus on multimodal biometrics. It should be noted that it is difficult to obtain the real multimodal biometric images collected from the same volunteers, so we combine three datasets (palm print, knuckle and face) to form the multi-modal biometric dataset.

4.1.1. PolyU multispectral palmprint Database [14]: Multi-spectral palmprint images were collected from 250 volunteers, including 195 males and 55 females. The age distribution is from 20 to 60 years old. We collected samples in two separate sessions. In each session, 6 images for each palm are provided. In total, the database contains 6,000 images from 500 different palms with same illumination.

4.1.2. PolyU FKP Database [15]: FKP images were collected from 165 subjects, including 125 males and 40 females. The images are also collected in two separate sessions. In each session, the subject was asked to provide 6 images for each of the left index finger, the left middle finger, the right index finger, and the right middle finger.

4.1.3. Extended YaleB Database [16]: The extended YaleB database is composed of 2414 face images of 38 subjects, where each person has 59–64 near frontal images under different illumination conditions.

4.1.4 Multi-modal biometric Database: It should be noted that it is difficult to obtain the real multi-modal biometric images collected from the same volunteers, so we combine the above three databases to form the multi-modal biometric dataset. This dataset used in our experiment has the common subjects. There are 12 images in each subject, and subject has 4 images from each modality. We believe that the samples of the three modalities combined come from the same person. The following experiments are repeated 10 times, and the average results are reported.
4.2. Single modality
For the sake of comparison, we conduct the single modality recognition experiment. Several classic classification algorithms are compared on the single modality in the biometric task, including SVM, K-Nearest-Neighbors (KNN), and Fully-Connected-Neural-Networks (NN). We perform the above methods on each modality. The results are given in Table I. For the next sub-section, we will perform the multi-modal recognition methods on these modalities.

| Method   | SVM | KNN | NN |
|----------|-----|-----|----|
| Palm print | 92.93 | 93.56 | 92.75 |
| Knuckle   | 91.73 | 92.78 | 92.33 |
| Face      | 92.81 | 93.26 | 93.71 |

4.3. Multi-modality
We compare MKL method with the following methods in the multi-modal biometric task, including Low-rank discriminant embedding (LRDE) [17], deep generalized canonical correlation analysis (DGCCA) [18], Multi-view subspace clustering (MVSC) [19], and multi-view discriminant analysis (MVDA) [20]. For all these methods, we repeat 10 times and report the mean performance. Classification performance is evaluated in terms of accuracy (ACC), sensitivity (SEN) and specificity (SPC).

| Method   | LRDE | DGCCA | MVSC | MVDA | MKL |
|----------|------|-------|------|------|-----|
| ACC      | 94.58 | 96.03 | 95.83 | 95.28 | 96.26 |
| SEN      | 92.76 | 89.92 | 90.71 | 93.69 | 94.78 |
| SPE      | 95.31 | 97.69 | 96.82 | 97.21 | 98.13 |
| AUC      | 92.85 | 96.83 | 93.23 | 97.53 | 97.85 |

As can be seen from above Tables, we can find the multi-modal method often outperforms the single modal method. The reasons maybe that the multi-modal method can extract more complementary features for classification. More importantly, the multi-kernel used in our method achieve the high recognition accuracy among the methods in the indices of ACC, SEN, SPE and AUC. The results show that the proposed framework can effectively fuse the different modalities for classification.

5. CONCLUSION
This paper focuses on multi-mode biometric recognition and proposes a fusion recognition framework. We introduce multi-kernel to the fusion framework, which can fuse data with different dimensions and mine complementary information between different modalities. Experiments show that multi-modal fusion can effectively improve the recognition accuracy, and our method has a higher recognition accuracy than other fusion methods.

ACKNOWLEDGMENT
This work is completed in the scientific research project and writing tutoring.

REFERENCES
[1] L.-q. Zhu and S.-y. Zhang, “Multimodal biometric identification system based on finger geometry, knuckle print and palm print,” Pattern Recognition Letters, vol. 31, no. 12, pp. 1641–1649, 2010.
[2] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 815–823.
[3] E. Marasco and A. Ross, “A survey on antispooﬁng schemes for ﬁngerprintrecognition systems,” ACM Computing Surveys (CSUR), vol. 47, no. 2, pp. 1–36, 2014.

[4] Y. Tong, F. W. Wheeler, and X. Liu, “Improving biometric identiﬁca-tion through quality-based face and ﬁngerprint biometric fusion,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. IEEE, 2010, pp. 53–60.

[5] S. Aleem, P. Yang, S. Masood, P. Li, and B. Sheng, “An accurate multi-modal biometric identiﬁcation system for person identiﬁcation via fusion of face and ﬁnger print,” World Wide Web, vol. 23, no. 2, pp. 1299–1317, 2020.

[6] J. Aravinth and S. Valarmathy, “Multi classiﬁer-based score level fusion of multi-modal biometric recognition and its application to remote biometric authentication,” The Imaging Science Journal, vol. 64, no. 1, pp. 1–14, 2016.

[7] Z. Akhtar, B. Biggio, G. Fumera, and G. L. Marcialis, “Robustness of multi-modal biometric systems under realistic spoof attacks against all traits,” in 2011 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS). IEEE, 2011, pp. 1–6.

[8] A. Nagar, K. Nandakumar, and A. K. Jain, “Multi-biometric cryptosystems based on feature-level fusion,” IEEE transactions on information forensics and security, vol. 7, no. 1, pp. 255–268, 2011.

[9] R. Brunelli and D. Falavigna, “Person identiﬁcation using multiple cues,” IEEE transactions on pattern analysis and machine intelligence, vol. 17, no. 10, pp. 955–966, 1995.

[10] K.-f. Zhang, H.-q. Yuan, and P. Nie, “A method for tool condition monitor-ing based on sensor fusion,” Journal of Intelligent Manufacturing, vol. 26, no. 5, pp. 1011–1026, 2015.

[11] N. Bodla, J. Zheng, H. Xu, J.-C. Chen, C. Castillo, and R. Chellappa, “Deep heterogeneous feature fusion for template-based face recognition,” in 2017 IEEE winter conference on applications of computer vision (WACV). IEEE, 2017, pp. 586–595.

[12] Q. Zhu and Y. Xu, “Multi-directional two-dimensional pca with matchingscore level fusion for face recognition,” Neural Computing and Applications, vol. 23, no. 1, pp. 169–174, 2013.

[13] J. Li, T. Qiu, C. Wen, K. Xie, and F.-Q. Wen, “Robust face recognition using the deep c2d-cnn model based on decision-level fusion,” Sensors, vol. 18, no. 7, p. 2080, 2018.

[14] D. Zhang, W.-K. Kong, J. You, and M. Wong, “Online palmprint identiﬁcation,” IEEE Transactions on pattern analysis and machine intelligence, vol. 25, no. 9, pp. 1041–1050, 2003.

[15] L. Zhang, L. Zhang, D. Zhang, and H. Zhu, “Online ﬁnger-knuckle-print veriﬁcation for personal authentication,” Pattern recognition, vol. 43, no. 7, pp. 2560–2571, 2010.

[16] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fish-erfaces: Recognition using class speciﬁc linear projection,” IEEE tran. on pattern analysis and machine informatics engineering,” an International Journal (IEIJ), vol. 1, no. 1, p. 40, 2013.

[17] J. Li, Y. Wu, J. Zhao, and K. Lu, “Low-rank discriminant embedding for multiview learning,” IEEE transactions on cybernetics, vol. 47, no. 11, pp. 3516–3529, 2016.

[18] A. Benton, H. Khayrallah, B. Gujral, D. A. Reisinger, S. Zhang, and R. Arora, “Deep generalized canonical correlation analysis,” arXiv preprint arXiv:1702.02519, 2017.

[19] H. Gao, F. Nie, X. Li, and H. Huang, “Multi-view subspace clustering,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 4238–4246.

[20] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, “Multi-view discriminant analysis,” IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 1, pp. 188–194, 2015.