

# Decision-Tree Based Hybrid Filter-Wrapping Method for the Fusion of Multiple Feature Sets

Cuicui Zhang<sup>1</sup>, Xuefeng Liang<sup>1</sup>(✉), and Naixue Xiong<sup>2</sup>

<sup>1</sup> IST, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan  
zhang@vision.kuee.kyoto-u.ac.jp, xliang@i.kyoto-u.ac.jp

<sup>2</sup> School of Computer Science, Colorado Technical University, Colorado Spring, USA  
nxiong@coloradotech.edu

**Abstract.** This paper proposes a decision-tree based hybrid filter-wrapping method to solve multiple-feature recognition problems. The decision tree is constructed by various feature sets. Each tree node comprises all possibilities of individual feature combinations: the original features, the serial fusion of original features, and the parallel fusion of original features. In order to generate the best discriminate feature set, a two-stage feature searching algorithm is developed. The first stage is a kind of feature filtering method to find out the local optimal individual features in each level of the tree using a LDA-motivated discrimination criterion. The second stage is a global optimal feature vector generation based on a kind of forward wrapping method. In contrast to literature feature fusion methods which considered filter and wrapping separately, our method combines them together. Furthermore, since our method takes all possibilities of feature combinations into consideration, it is more likely to generate the best discriminate feature set than other feature fusion methods. In addition, our method also compensates discrimination if some portions of original features are missing. The effectiveness of our method is evaluated on a 3D dataset. Comparative experimental results show that our method can impressively improve the recognition accuracy and has better performance than existing methods.

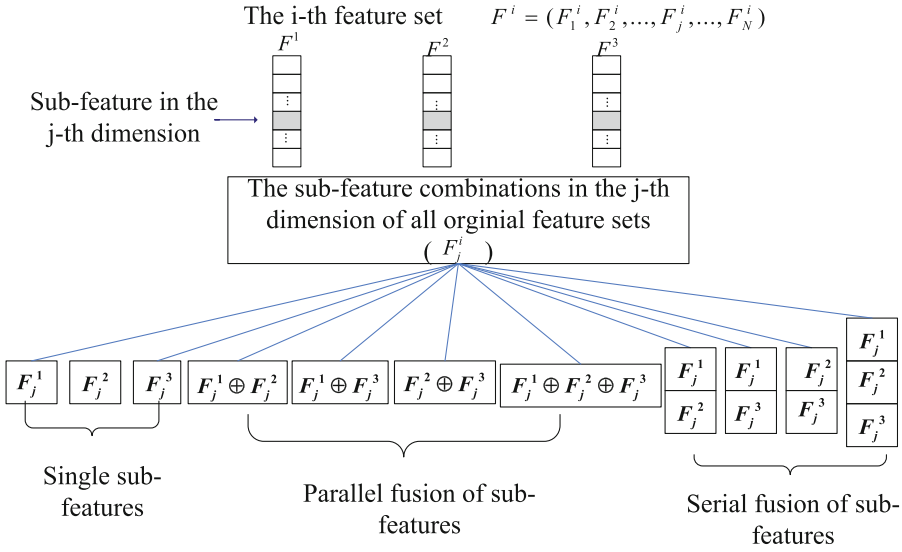
**Keywords:** Decision-tree · Hybrid filter-wrapping · Feature fusion · Face recognition

## 1 Introduction

Information fusion is promising for a large extent of applications in modern image/video retrieval or pattern recognition systems. The multimodality fusion strategies are often categorized by feature fusion, match score fusion, and decision fusion [13]. Compared with match score fusion and decision fusion, feature fusion shows obvious advantages [4]. Wisely integrating multiple feature vectors extracted from the same pattern or from different patterns will not only enhance the discrimination of the features, but also eliminate redundancies.

Existing feature fusion methods can be divided into two categories: serial feature fusion and parallel feature fusion [13]. Serial fusion groups multiple feature vectors into a new vector, and is doomed to deal with high dimensional feature vectors. Recently, several methods of this kind were proposed. For example, Canonical Correlation Analysis (CCA) in [6, 12] was used to fuse audio-visual features in which the representations of projected audio and visual features in their own subspaces were able to preserve the correlation conveyed by the original audio and visual feature spaces. However, the applications of CCA are merely limited to the situation that the covariance matrix of the feature space is nonsingular. Another method presented in [1] used Partial Least Squares (PLS) regression for feature pair fusion. Although the fused feature set using this method may have better discrimination, its (the feature vector) dimension is also multiplied. This obviously leads to more difficulties in afterward processing. Parallel feature fusion is achieved by combining multiple feature vectors into a new vector in parallel. The popular parallel feature fusion methods include Probabilistic fusion methods [10], Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) [3], sum rule [7], Fisher Linear Discriminate Analysis (FLDA) [5, 8, 9], Radial Basis Function Neural Network (RBFNN) [18], and so on. These feature fusion methods simply concatenate or integrate several heterogeneous features together [4]. Nevertheless, systems adopting such feature fusion may not always have a better performance due to unbalanced combination of multiple feature sets. Considering the limitations of current feature fusion methods, we are trying to develop a new approach to overcome aforementioned drawbacks.

In supervised learning, feature selection algorithms are usually grouped into two categories: filtering and wrapping. Filters use an evaluation function that relies solely on properties of the data, thus are independent on any particular algorithm. Wrappers use the inductive algorithm to estimate the value of a given subset. Wrapper methods are widely recognized as a superior alternative in supervised learning. However, even for algorithms that exhibit a moderate complexity, the number of executions that the search process requires results in a high computational cost. This paper proposes a hybrid filter-wrapping method based on a decision tree to combine the advantages of filter and wrapper together while avoiding their shortcomings meanwhile. The decision tree is constructed by hybrid feature combinations of original feature sets. For a modality system with  $n$  feature sets of the same length  $N$ , the  $i$ -th ( $i = 1, 2, \dots, N$ ) level of the decision tree has  $[n + 2(C_n^2 + C_n^3 + \dots + C_n^n)]^{i-1}$  tree nodes, and each node at the same level possesses the same children including  $n$  sub-features in the  $i$ -th dimension of the original feature sets,  $C_n^2 + C_n^3 + \dots + C_n^n$  serial fusion of these sub-features, and  $C_n^2 + C_n^3 + \dots + C_n^n$  parallel fusion of these sub-features. An example consisting of three feature sets is shown in Fig. 1. The advantage of using decision-tree is that all possibilities of feature combinations can be taken into consideration. However, if the best feature set searching exhausts all possibilities, it requires extremely high computational cost. In order to do it efficiently and effectively, a two-stage feature searching algorithm is proposed. The first stage is



**Fig. 1.** An example of individual feature combinations: the combination of the  $j$ -th features of multiple features sets

a filter seaming method to find out the local optimal individual features in each level of the tree based on a LDA-motivated discrimination criterion. The second stage is a forward wrapping method to generate the global optimal feature set. However, our method differs from the classical wrappers that usually combine and visit all features. Through doing experiments, we find that some features are redundant or even downgrading the performance. Here, PCA + LDA is used to generate features in much lower dimensional feature space. This scheme does not only guarantee the accuracy improvement, also reduces the computational cost significantly.

The advantages of our method are as follows: 1. The two-stage feature searching algorithm, which combines filter and wrapping together, not only finds out the local best individual features, but also generates a global optimal feature set; 2. Our method takes all possibilities of feature combinations into account: the original features, the serial fusion of original features, and the parallel fusion of original features. This characteristic makes our method more likely to generate the best feature vector than literature feature fusion methods; 3. Once some portions of original features are missing, the proposed algorithm can self-adjust the weights among original features so that it can compensate feature missing problem.

The remaining of this paper is organized as follows: Sect. 2 presents the construction of the decision tree, and the two-stage feature searching algorithm in details. In Sect. 3, comparative experiments are reported and the performance of our method is analyzed. Finally, Sect. 4 concludes this paper.

## 2 Decision-Tree Based Hybrid Filter-Wrapping Method for Feature Fusion

The method can be roughly divided into four steps: 1. the construction of the decision-tree; 2. the local optimal feature selection based on a proposed Maximal Classifiable Criterion (MCC); 3. the sorting of the feature vector according to the MCC score; and 4. the forward global optimal feature selection.

### 2.1 Construction of the Decision Tree

Feature fusion is performed after feature extraction using PCA [15] and LDA [2]. Here, PCA is used for dimension reduction and LDA is used to generate feature vectors in the lower dimensional space. LDA is implemented via scatter matrix analysis. For an  $n$ -class problem, the within-class and between-class scatter matrices  $S_w$ ,  $S_b$  are calculated as follows:

$$S_w = \sum_{i=1}^n Pr(C_i)\Sigma_i. \quad (1)$$

$$S_b = \sum_{i=1}^n Pr(C_i)(\mathbf{m}_i - \mathbf{m}_0)(\mathbf{m}_i - \mathbf{m}_0)^T, \quad (2)$$

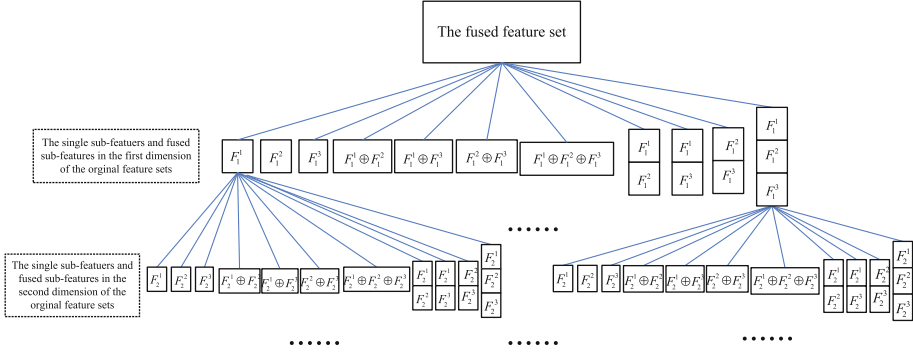
where  $Pr(C_i)$  is the prior class probability and usually replaced by  $1/n$  with the assumption of equal priors. Please refer to [2] for more details.

The decision tree is constructed by the sub-features in the same dimension of the original feature sets. The node at the  $j$ -th level comprises the sub-features in the  $j$ -th dimension of original feature sets and the new serial and parallel fused sub-features. Thus, the height of the tree equals to the number of dimensions  $N$  of the original feature sets. For a modality system of  $n$  different feature sets, each tree node at  $j$ -th level has  $n + 2(C_n^2 + C_n^3 + \dots + C_n^n)$  children nodes including the  $n$  original sub-features in the  $(j+1)$ -th dimension, the  $C_n^2 + C_n^3 + \dots + C_n^n$  serial fusion of these sub-features, and the  $C_n^2 + C_n^3 + \dots + C_n^n$  parallel fusion of these sub-features. All the tree nodes at the same level have the same children. Figure 1 shows an example. There are three heterogeneous feature sets:  $F^1$ ,  $F^2$ , and  $F^3$ . Each tree node at  $j$ -th level has 11 children, including 3 original sub-features in the  $(j+1)$ -th dimension, 4 parallel fused sub-features and 4 serial fused sub-features. The decision tree is shown in Fig. 2.

The serial fusion sequentially connects all original sub-features in the same dimension to generate a new feature. The parallel fusion often linearly combines all sub-features by a weighted sum rule defined in Eq. (3).

$$V = w_1X_1 + w_2X_2 \dots + w_iX_i \dots + w_nX_n, \quad (3)$$

where  $X_i$  denotes the  $i$ -th original feature set. In [11, 14, 16], the weights are determined by the recognition accuracy of each original feature set. In [9], the weights are defined by  $trace(S_b./S_w)$ , where  $S_b$  and  $S_w$  are defined in Eqs. (1) and (2), respectively. Through conducting these two weight calculation methods, we find that the latter is not always proportional to the former one. For example, the  $trace(S_b./S_w)$  of three feature images: maximum principal curvature image



**Fig. 2.** An example decision-tree for a modality system of 3 feature sets. The nodes at the  $j$ -th level represent the combinations of the individual sub-features in the  $j$ -th dimension: the original sub-features, the parallel fusion of these sub-features, and the serial fusion of these sub-features. Each node at the same level has the same children. The height of the tree equals to the dimensions of the original feature sets.

(MPCI), average edge image (AEI), and range image (RI) have a order as  $RI > AEI > MPCI$ , however the recognition accuracy of them are  $RI > MPCI > AEI$ . Therefore, we can say that either the  $trace(S_b./S_w)$  or the recognition accuracy alone is not adequately robust to determine weights for fusion. In this paper, a new weight calculation method which combines the two weight calculation method together is proposed as follows:

$$w_{(i,j)} = \frac{1}{2} \times (nTrace_{(i,j)}(S_b./S_w) + dA_{(i,j)}), \tag{4}$$

where  $w_{(i,j)}$  is the weight of the sub-features in the  $j$ -th dimension of the  $i$ -th feature set,  $Trace_{(i,j)}$  is the trace of the  $S_b./S_w$ , calculated from these sub-features, and the  $dA_{(i,j)}$  (short for *dimensionalAccuracy* $_{(i,j)}$ ) is the recognition accuracy of these sub-features. As  $trace_{(i,j)}$  and  $dA_{(i,j)}$  are of different data ranges, we, therefore, normalize the trace to the range of [0,1] by

$$nTrace_{(i,j)}(S_b./S_w) = \frac{trace_{(i,j)}}{\max(trace_{(1,j)}, trace_{(2,j)}, \dots, trace_{(n,j)})}. \tag{5}$$

And  $w_{(i,j)}$  is normalized by

$$w_{(i,j)} = \frac{w_{(i,j)}}{w_{(1,j)} + w_{(2,j)} + \dots + w_{(n,j)}}. \tag{6}$$

In [11], feature fusion is performed on the whole feature sets using Eq. (3), and  $w_i$  is given by the recognition accuracy of each original feature set. In [9], the trace of  $S_b./S_w$  is also generated from the whole feature set. In our work, feature fusion is performed on the sub-features in the same dimension of the original feature sets sequentially. Thus, we calculate the weight for each dimensional sub-features rather than on the whole original feature set. This weight determination at microlevel lets the fused feature more fit to the data.

## 2.2 Local Optimal Feature Selection

The local optimal feature selection is to find out the best sub-features at each level of the decision tree from the original sub-features and the fused sub-features. Selection is performed based on a classification criterion defined in Eq. (4). The feature with the maximal classifiable criterion score (MCC) is considered as the best feature in this level. Since the mechanism of parallel fusion differs from that of serial fusion, MCC is defined under two conditions:

- (1) for original sub-features in the  $j$ -th dimension of the  $i$ -th feature set, and the parallel fusion of these original sub-features, MCC is defined by

$$\max(n\text{Trace}_{(i,j)}(S_b./S_w) \times dA_{(i,j)}); \quad (7)$$

- (2) for the serial fusion of these original sub-features, MCC is defined by

$$\frac{\sum_{i=1}^n n\text{Trace}_{(i,j)}(S_b./S_w) \times dA_{(i,j)}}{\sum_{i=1}^n dA_{(i,j)}}. \quad (8)$$

The reason we define a different MCC for serial fusion is that a new serial fused feature is created by linking original sub-features sequentially. Then, the trace of the serial fused feature is the integration of traces of those original sub-features. As each original sub-feature performs with its own accuracy that is varied from others, weighting these traces with respect to their accuracies will more fit to the new feature.

## 2.3 Sorting of the Feature Vector

The local optimal sub-feature at each tree level form a feature vector going through the tree from root to leaves. Elements in this feature vector are then sorted in a descending order according to the MCC score to form the feature set  $F'$ .

## 2.4 Global Optimal Feature Selection

The global optimal feature selection aims at eliminating the feature redundancy and removing noisy features from the feature set  $F'$  generated in Sect. 2.3. It is a wrapper seeming method which selects features forward from the top to the bottom. The selected sub-feature at  $j$ -th dimension of  $F'$  will be inserted into the final discriminate feature set  $\hat{F}$  sequentially. If this feature does not improve the recognition accuracy of  $\hat{F}$ , then it will be eliminated from  $\hat{F}$ . Since the sub-features in each dimension is independent with each other, feature elimination here will not affect the discriminative power of the subsequent feature integration. Please see Fig. 3 for demonstration. After this procedure, the best discriminate feature set  $\hat{F}$  is obtained for recognition.

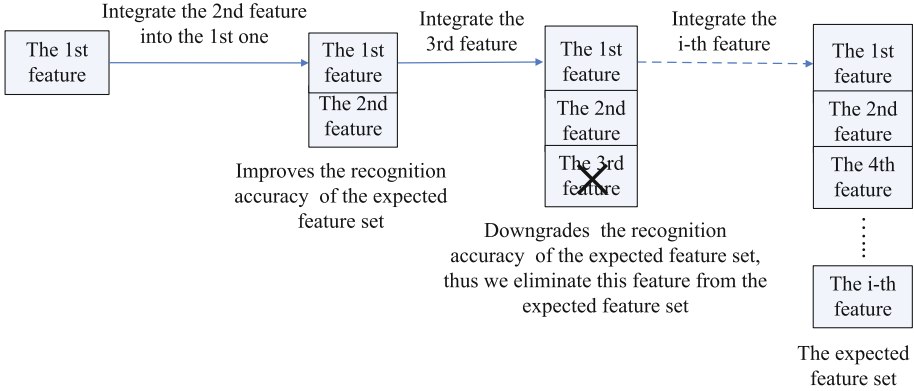


Fig. 3. The forward global feature selection

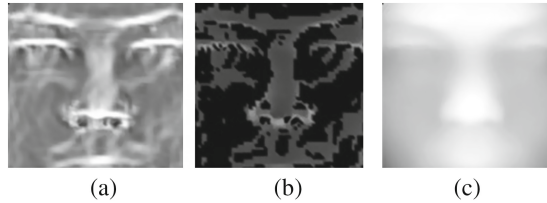
### 2.5 Handing with Feature Missing

Apart from intralmodal fusion where multiple features from a single modality are integrated, our method can also be applied to intermodal fusion (multimodal fusion), where complementary features from multiple patterns are integrated. Such a case is Audio Visual Automatic Speech Recognition, where fusing visual and audio cues can lead to improved performance. One motivation of multimodal fusion is to solve missing feature problems. If some portion of a feature vector is missing, other feature vectors should compensate to guarantee the recognition accuracy. This can be done in the first stage of our feature searching algorithm. Local feature selection assigns specific weights to different sub-features according to their discriminative powers, higher discriminative power larger weight. There usually exists a key feature in the system, such as the RI feature in our experimental system. However, when it is missing in some tree levels, our method will take other kinds of features into consideration, such as MPCl. Therefore, our algorithm can still maintain the performance in case of feature missing.

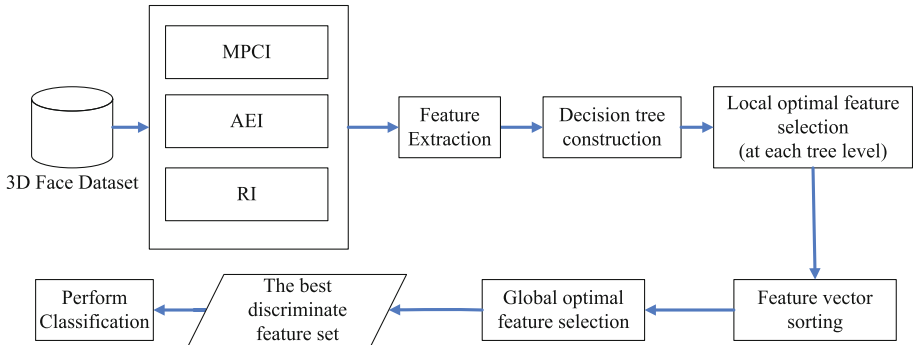
## 3 Experimental Results and Analysis

### 3.1 3D Dataset and Feature Extraction

The experiment is performed on the 3D dataset consisting of 38 face classes. Each face class possesses 10 samples, 9 of which are of various head poses and 1 of different lighting conditions. For each person, 6 samples are selected randomly and put into the training subset while the other 4 samples are put into the test subset. Here, 3D face models are represented by triangular meshes. In order to reduce data size, mesh simplification [19] is used here. The goal of mesh simplification is to use fewer vertices and fewer triangles to reconstruct the 3D face model while preserving the shape and feature points as much as possible. Data ratio is defined by the percent of the data after mesh simplification to



**Fig. 4.** Three kinds of facial feature images (a) MPCCI, (b) AEI, (c) RI



**Fig. 5.** Our feature fusion method for 3D face recognition

that before mesh simplification. For each 3D face data, mesh simplification is performed at 6 levels to get 6 kinds of data files: the file with 100 % data (the original data file), 50 % data, 25 % data, 12.5 % data, and 6.25 % data.

After data normalization, three kinds of facial feature images are generated using the method in [17]: the maximal principal curvature image (MPCI), the average edge image (AEI), and the range image (RI) (see Fig. 4). Feature extraction based on PCA [17] and LDA [2] is performed to generate three feature sets to build up the decision tree. The procedure is shown in Fig. 5.

### 3.2 Performance Analysis

In this section, two experiments are performed to evaluate the efficiency of our method. In the first experiment, our method is compared with recognition without fusion (using single feature sets) and recognition using several existing feature fusion methods. The recognition accuracy of three single feature sets are calculated first, which are shown in Table 1. The average recognition accuracy of these feature sets is used to compare with our method in order to preserve the diversity of features' performances (see the second column of Table 2). In [11, 14, 16], a parallel fusion based on a weighted sum rule defined by Eq. (3) was developed. We call these recognition rate based weight calculation method as RRW (see the third column of Table 2). In [17], a modified weight calculation



**Table 1.** Recognition accuracies of three single facial feature images based on PCA and PCA + LDA for data files of six levels, respectively

Data Ratio \ Accuracy(%)	MPCI		AEI		RI	
	PCA	PCA+LDA	PCA	PCA+LDA	PCA	PCA+LDA
100%(500KB)	92.76	98.03	85.53	96.71	92.76	<b>98.68</b>
75%(375KB)	93.43	98.68	88.82	95.39	93.42	<b>99.34</b>
50%(250KB)	92.11	98.03	85.53	96.05	93.42	<b>99.34</b>
25%(125KB)	92.76	97.37	71.71	95.39	94.08	<b>99.34</b>
12.5%(62KB)	82.24	93.42	50.00	82.24	94.08	<b>98.68</b>
6.25%(31KB)	63.16	92.76	32.24	70.39	88.82	<b>99.34</b>

**Table 2.** Accuracies of recognition without fusion and that using four feature fusion methods

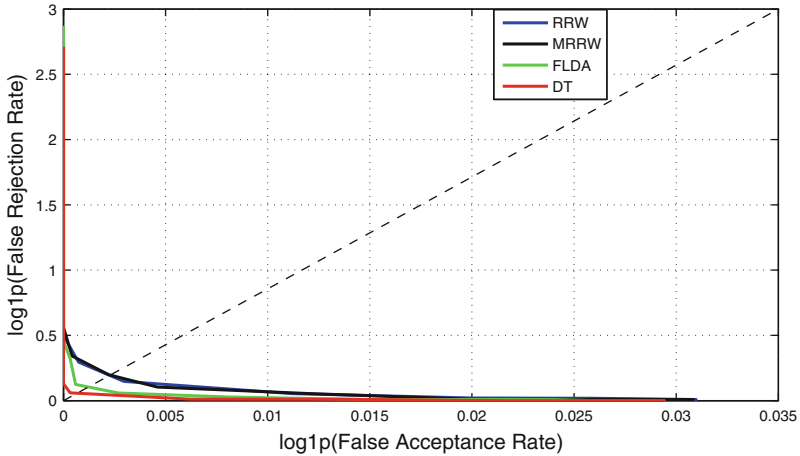
Data Ratio \ Accuracy(%)	Single Feature	Feature Fusion			
		RRW [11]	MRRW [17]	FLDA [9]	DT (our method)
100%(500KB)	97.81	94.74	97.37	<b>98.68</b>	<b>98.68</b>
75%(375KB)	97.80	96.05	98.03	98.68	<b>99.98</b>
50%(250KB)	97.81	98.03	98.68	99.34	<b>99.98</b>
25%(125KB)	97.37	98.03	97.37	98.68	<b>99.98</b>
12.5%(62KB)	91.45	95.39	<b>99.98</b>	98.68	<b>99.98</b>
6.25%(31KB)	87.49	89.47	96.71	98.03	<b>99.34</b>

method was proposed. It was defined as follows:

$$w_i' = (10w_i - \min(\lfloor 10w_1 \rfloor, \lfloor 10w_2 \rfloor, \dots, \lfloor 10w_n \rfloor))^2, \quad (9)$$

where  $w_i$  is the recognition accuracy of the  $i$ -th original feature set ( $i = 1, 2, \dots, n$ ).  $w_i'$  is its new weight. We call this method as MRRW (the modified RRW, see the third column of Table 2). Another feature fusion method gained wide popularity is the one based on fisher linear discriminate analysis (FLDA) [9]. It uses the discriminate criterion defined by  $S_b$  and  $S_w$ . The recognition results of this method are shown in the fourth column of Table 2. The last column represents the recognition results using our method. We call our method 'DT' for short. Table 2 reveals that our method has a better performance than recognition using single features and that using the aforementioned feature fusion methods. It achieves at a remarkable recognition accuracy even when the data file is small. In addition, comparing the performance of RRW and that without fusion, we can see that in some cases, recognition using feature fusion even performs worse than ones without fusion. The reason is that the weights of different features are not balanced well. Compared to the existing feature fusion methods based on simple fusion strategies, our method is more likely to find the best discriminate feature set thanks to its characteristic of taking all the feature combination possibilities into account and its efficient two-stage feature selection algorithm.

Another three important parameters to evaluate the performance of pattern recognition algorithms are also investigated: the false rejection rate (FRR),



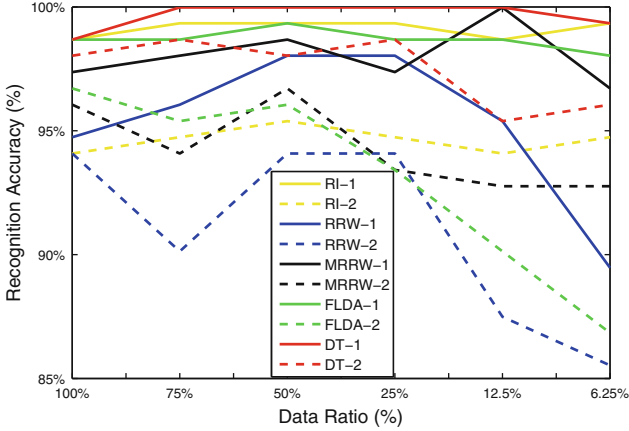
**Fig. 6.** ROC curves of four feature fusion methods

**Table 3.** Recognition accuracies of recognition using single features and feature fusion methods while portions of RI are missing.

Data Ratio	Accuracy(%)		Feature Fusion			
	RI	MPCI	RRW [11]	MRRW [17]	FLDA [9]	DT (our method)
100%(500KB)	94.08	<b>98.03</b>	94.08	96.05	96.71	<b>98.03</b>
75%(375KB)	94.74	<b>98.68</b>	90.13	94.08	95.39	<b>98.68</b>
50%(250KB)	95.39	<b>98.03</b>	94.08	96.71	96.05	<b>98.03</b>
25%(125KB)	94.74	97.37	94.08	93.42	93.42	<b>98.68</b>
12.5%(62KB)	94.08	93.42	87.49	92.76	90.13	<b>95.39</b>
6.25%(31KB)	94.74	92.76	85.53	92.76	86.84	<b>96.05</b>

the false acceptance rate (FAR), and the equal error rate (EER). The ROC curves plot  $\log(\text{FAR})$  against  $\log(\text{FRR})$  of those aforementioned four feature fusion methods in Fig. 6. ROC curves are plotted by  $\log(\text{FAR})$  against  $\log(\text{FRR})$  because it makes the relationship of FAR and FRR clearer than conventional ROC diagrams. We can see that EER of our method approximately equals to 0. It is smaller than that of other feature fusion methods. This suggests that our method obtains a very satisfactory performance not only in recognition accuracy but also in FAR, FRR.

The second experiment is performed to verify that our method can deal with missing feature problems. RI is the key feature in the first experiment since it achieves a higher recognition accuracy than MPCl and AEI (see Table 1). In this experiment, we randomly remove portions (20%) of RI features of some samples (10% samples). Table 3 shows the comparative result. In the first three rows, the performance of our method keeps the same as using MPCl alone. The reason is that the missed portions have seriously downgraded the discrimination of RI, and affected results of other three feature fusion methods. Comparatively, MPCl



**Fig. 7.** Recognition accuracies of the methods in Table 3 with and without missing features, where solid curves represent recognition accuracy without missing features and dot lines represent that method with missing features (Color figure online).

risers to the first priority comparing with other fused features in each dimension. Thus, our method assigns larger weight to MPCFI and selects it alone to perform recognition. That is why our method has the same performance as MPCFI. In the next three rows, our method performs better than other all. Because the mesh simplification in MPCFI is significant (the data ratio is no greater than 25%), and also seriously downgrades the discrimination of MPCFI. Then, our method uses the generated new features under this case. In Table 3, as other feature fusion methods are not capable of rebalancing the weights after feature missing, their performances downgrade significantly. Figure 7 plots the recognition accuracy curves of all methods before and after feature missing. It clearly shows that the degeneration affected by feature missing of our method (red line) is the least.

## 4 Conclusion

In this paper, a hybrid filter-wrapping method is proposed based on decision-tree for the fusion of multiple feature sets. This work has four contributions: 1. It combines two kinds of feature selection methods: filter and wrapper together using a two-stage feature searching algorithm to ensure efficiency and effectiveness of finding the best discriminate feature set. This strategy develops the advantages of filters and wrappers while avoiding their shortcomings meanwhile; 2. Our method takes all possibilities of feature combinations into account including the original features, the serial fusion of original features, and the parallel fusion of original features. This characteristic makes our method more likely to find out the optimal feature vector than literature feature fusion methods which treat serial fusion and parallel fusion separately; 3. To ensure the performance of parallel fusion, we designed a new weight calculation method which takes both

the trace and the dimensional recognition accuracy into consideration; 4. The two-stage algorithm has an ability to solve missing feature problems.

Our feature fusion method has been evaluated by 3D face recognition. Experimental results show that our method outperforms methods using single features and other feature fusion approaches in [9,11,17], respectively. The EER of our method approximately equals zero, which shows that our method achieves at a very satisfactory performance.

**Acknowledgement.** This work is supported by: JSPS Grants-in-Aid for Scientific Research C (No.15K00236).

## References

1. Barker, M., Rayens, W.: Partial least squares for discrimination. *J. Chemometr.* **17**(3), 166–173 (2003)
2. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(7), 711–720 (1997)
3. Fang, Y., Tan, T., Wang, Y.: Fusion of global and local features for face verification. *IEEE Conference on ICPR*, pp. 382–385 (2002)
4. Fu, Y., Cao, L., Guo, G., Huang, T.: Multiple feature fusion by subspace learning. In: *Proceedings of ACM International Conference on Image and Video Retrieval*, pp. 127–134 (2008)
5. Gan, J., Liang, Y.: A method for face and iris feature fusion in identity authentication. *IJCSNS* **6**(2), 135–138 (2006)
6. Kim, T., Wong, S., Cipolla, R.: Tensor canonical correlation analysis for action classification. In: *IEEE Conference on CVPR* (2007)
7. Lee, C., Wang, S.: A gabor filter-based approach to fingerprint recognition. In: *IEEE Workshop on Signal Processing Systems*, pp. 371–378 (1999)
8. Liu, Z., Liu, C.: Fusion of color, local spatial and global frequency information for face recognition. *Patterns Recogn.* **43**, 2882–2890 (2010)
9. Patra, A., S. Das, B.: Dual space based face recognition using feature fusion. In: *International Conference on Visual Information Engineering*, pp. 155–160 (2006)
10. Rao, K., Rajagopalan, A.: A probabilistic fusion methodology for face recognition. *J. Appl. Sig. Process.* **17**, 2772–2787 (2005)
11. Ross, A., Govindarajan, R.: Feature level fusion using hand and face biometrics. In: *Proceedings of SPIE Conference on Biometric Technology for Human Identification*, pp. 196–204 (2004)
12. Sargin, M., Yemez, Y., Erzin, E., Tekalp, A.: Audiovisual synchronization and fusion using canonical correlation analysis. *IEEE Trans. Multimedia* **9**(7), 1396–1403 (2007)
13. Sun, Q., Zeng, S., Liu, Y., Heng, P., Xia, D.: A new method of feature fusion and its application in image recognition. *Pattern Recogn.* **38**, 2437–2448 (2005)
14. Sun, Y., Tang, H., Yin, B.: The 3d face recognition algorithm fusing multi-geometry features. *J. Acta Autom. Sinica* **34**(12), 1483–1489 (2008)
15. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)

16. Xiang, Y., Su, G.: Multi-parts and multi-feature fusion in face recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2008, pp. 1–6 (2008)
17. Zhang, C., Uchimura, K., Zhang, C., Koutaki, G.: 3d face recognition using multi-level multi-feature fusion. In: The 4th Pacific Symposium on Image and Video Technology, pp. 21–26 (2010)
18. Zhang, H., Sun, X., Zhao, L., Liu, L.: Image fusion algorithm using RBF neural networks. In: Corchado, E., Abraham, A., Pedrycz, W. (eds.) HAIS 2008. LNCS (LNAI), vol. 5271, pp. 417–424. Springer, Heidelberg (2008)
19. Zhou, Y., Zhang, C., He, P.: Feature preserving mesh simplification algorithm based on square volume measure. *Chin. J. Comput.* **32**(2), 203–212 (2009)