Chapter VI
Two Novel Facial Feature Extraction Methods

ABSTRACT

In this chapter, we introduce two novel facial feature extraction methods. The first is multiple maximum scatter difference (MMSD) which is an extension of a binary linear discriminant criterion, i.e. maximum scatter difference. The second is discriminant based on coefficients of variances (DCV) which can be viewed as a generalization of N-LDA. At last, we give a summery of the chapter.

6.1 MULTIPLE MAXIMUM SCATTER DIFFERENCE

The maximum scatter difference (MSD) discriminant criterion (Song, Zhang, Chen, & Wang, 2007) presented in Section 3.4 is a binary discriminant criterion for pattern classification. Because MSD utilizes the generalized scatter difference rather than the generalized Rayleigh quotient as a class separability measure, it avoids the singularity problem when addressing the SSS problems that trouble the Fisher Discriminant Criterion. Further, experimental studies demonstrated that MSD classifiers based on this discriminant criterion have been quite effective on face recognition tasks (Song et al., 2007). The drawback of the MSD classifier is that, as a binary classifier, it cannot be applied directly to multiclass classification problems such as face recognition. This means that multiple recognition tasks have to be divided into a series of binary classification problems using one of three implementation strategies: one-vs-rest, one-vs-one, or directed-acyclic-graph (Hsu...
Experiments have shown that MSD classifiers are not very effective when using the first strategy, while using the latter two strategies requires the training of \( l(l-1)/2 \) MSD classifiers for a \( l \)-class recognition problem. The efficiency of such an approach will greatly be affected by any increase in the number of classes. Ultimately, then, like all binary classifiers, MSD classifiers are not suitable for large-scale pattern recognition problems.

To address the problem, this section generalizes the classification-oriented binary criterion to its multiple counterpart—multiple maximum scatter difference (MMSD) discriminant criterion for facial feature extraction (Song, Liu, & Yang, 2006; Song, Zhang, Mei, & Guo, 2007). The MMSD feature extraction method based on this novel discriminant criterion is a new subspace-based feature extraction method. Unlike most conventional subspace-based feature extraction methods that derive their discriminant vectors either in the range of the between-class scatter matrix or in the null space of the within-class scatter matrix, MMSD computes its discriminant vectors in both subspaces. MMSD is theoretically elegant and easy to calculate. Extensive experimental studies conducted on the benchmark database, FERET, show that MMSD outperforms many state-of-the-art facial feature extraction methods including nullspace LDA (N-LDA) (Chen, Liao, Ko, Lin, & Yu, 2000), direct LDA (D-LDA) (Yu & Yang, 2001), Eigenface (Turk & Pentland, 1991), Fisherface (Belhumeur, Hespanha, & Kriegman, 1997), and complete LDA (Yang, Frangi, Yang, Zhang, & Jin, 2005).

### 6.1.1 Related Works

Almost all facial feature extraction methods have been developed from Fisher linear discriminant (FLD). FLD seeks an optimal linear transformation \( W^* \in \mathbb{R}^{d \times r} \) from a high-dimensional input space \( \mathbb{R}^d \) into a low-dimensional feature space \( \mathbb{R}^r \), by observing the following Multiple Fisher Discriminant Criterion,

\[
\max_{W \in \mathbb{R}^{d \times r}} \frac{W^T S_B W}{W^T S_W W} \quad (6.1)
\]

Here \( S_B \) and \( S_W \) are respectively the between- and within-class scatter matrices. The transformation matrix \( W^* \) is called the Fisher discriminant matrix and its column vectors are called Fisher discriminant vectors. It has been proved that the matrix whose \( i \)th column vector is an eigenvector of \( S_W^{-1} S_B \) corresponding to the \( i \)th largest eigenvalue is an optimal solution for the model (6.1) when \( S_W \) is nonsingular. Unfortunately, in SSS problems such as face recognition, the dimension of an input space \( d \) is usually greater than the number of training samples \( N \). As
Overview
David Zhang, Fengxi Song, Yong Xu and Zhizhen Liang (2009). Advanced Pattern Recognition Technologies with Applications to Biometrics (pp. 1-23).
www.igi-global.com/chapter/overview/4273?camid=4v1a