International Journal of Innovative Computing, Information and Control
Volume 4, Number 4, April 2008
 ISSN 1349-4198
pp. 857–865

FEATURE EXTRACTION BASED ON A LINEAR SEPARABILITY CRITERION

Yong Xu[1] and Fengxi Song[2]

[1]Shenzhen Graduate School
Harbin Institute of Technology
Shenzhen 518100, P. R. China
laterfall@hitsz.edu.cn

[2]New Star Research Inst. of Applied Tech.
Hefei 230000, P. R. China
fx_song@sina.com

Received February 2007; revised August 2007

Abstract. *Though Fisher discriminant analysis (FDA) can perform feature extraction very well on data with simple distributions, it still has a number of shortcomings. For one, it usually fails to extract genuine optimal features from real-world data when such data exhibits an other-than-normal distribution. This is especially so when data exhibits a complex distribution, in which cases FDA is not able to obtain representative features at all. A second shortcoming is that FDA may not produce a sufficient number of transforming axes for the purpose of capturing representative features. This is because the maximum possible rank of the between-class scatter matrix does not exceed L-1, meaning that there are at most L-1 effective transforming axes, where L is the number of sample classes. In this paper, we develop a novel FDA method that deals with both of these problems. The novel FDA method is developed on the basis of the maximization of the distance between two arbitrary samples from two different classes and the minimization of the distance between samples in the same class. Besides this novel FDA method makes available a greater number of effective transforming axes than sample classes; it can extract representative features from data with a normal or complex distribution. We also propose three alternative forms of this novel FDA method. Experiments show that the novel FDA method outperforms a standard FDA.*
Keywords: Computer vision, Fisher linear discriminant analysis (FDA), Effective transforming axis, Feature extraction

1. **Introduction.**

Feature extraction is an important component of a pattern recognition system [1-4]. FDA, a feature extraction method with a simple form, has been widely used in the area of feature extraction [5-18]. FDA is used to maximize the distance between the mean vector of each class and that of the total samples and to minimize the distance between a sample and the mean vector of the class which the sample belongs to. Consequently, the feature of a sample extracted using FDA will be statistically much closer to the mean vector of its own class than those of other classes. If each class has a Gaussian density with a common covariance, FDA can produce the best transforming axes, enabling it to obtain the feature space with the largest linear separability. However, FDA has several weaknesses. Firstly, for real-world data, the condition that each class has a Gaussian density with a common covariance is generally not satisfied. Therefore, FDA cannot obtain optimal transforming axes for a number of real-world applications [13]. Secondly, because the rank of the between-class scatter matrix must be smaller than the number of sample classes [13-15], there must be fewer available effective transforming