Feature Extraction Using Histogram Entropies of Euclidean Distances for Vehicle Classification

Ming Bao, Luyang Guan, Xiaodong Li, Jing Tian, and Jun Yang

Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China
{baoming, guanluyang, lxd, tian, j_yang}@mail.ioa.ac.cn

Abstract. This paper presents a novel method for feature extraction based on the generalized entropy of the histogram formed by Euclidean distances, which is named distributive entropy of Euclidean distance (DEED in sort). DEED is a nonlinear measure for learning feature space, which provides the congregate and information measure of learning samples space. The ratio of between-class DEED to within-class DEED (J_{rd}) is used as a new nonlinear separability criterion for optimizing feature selection. Experiments on vehicle classification show that the proposed method has better performance on all the datasets than the fisher linear discriminant analysis.

1 Introduction

Feature extraction is a crucial preprocessing step for pattern recognition. It can be viewed as a process that extracts effective features from the original measurements though some functional transformations [1]. Feature extraction for classification aims to select features leading to large between-class distance and small within-class variance in the feature vector space, at the same time preserve class separability as much as possible. Various linear mapping based criteria have been proposed for evaluating the effectiveness of features [2]. Fisher linear discriminant analysis (FLDA) is a popular and powerful linear classification technique, which clusters patterns of the same class and separates patterns of different classes by maximizing the criterion function. As a measure of class separability, the Fisher criterion is defined by the ratio of the between-class variance to the within-class variance. FLDA works mostly for linearly separable classes and provides second-order statistics of data only. However, many pattern classifications are not linearly separable and features are highly nonlinear functions. Since it is difficult to capture a nonlinear relationship with a linear mapping, the basis problem is to find a proper nonlinear mapping function for the given data.

The rationale of performing a nonlinear mapping goes back to Cover's theorem on the separability of patterns, which states that a complex pattern–classification problem cast in a high-dimensional space nonlinearly is more likely to be linearly separable than in a low-dimensional space [3]. Many neural network architectures apply this idea for a linear solution obtained in the feature space [4]. Other nonlinear feature extraction approaches can be found in the Ref. [5]. In this paper, we address the feature extraction problem from an information theoretic perspective. The generalized entropies of histograms formed by Euclidean distances are employed for classification.

Y. Wang, Y. Cheung, and H. Liu (Eds.): CIS 2006, LNAI 4456, pp. 1085–1096, 2007.

The histogram entropy concept has been utilized in the image processing [6], which reflects the statistical information content of an image and hence its structure in the most general sense. A histogram of a measurement provides the basis for an empirical estimate of the probability density function. In this paper, we propose an entropy-based nonlinear mapping method for effective feature extraction. As an alternative criterion, the ratio of the between-class histogram entropy to the within-class histogram entropy is maximized to increase class separability.

The rest of this paper is organized as follows. In section 2, the distributive entropy of Euclidean distance (DEED) is defined and applied to a novel nonlinear separability criterion. Experiment results from the tracked vehicle and wheel vehicle classification problem are presented in Section 3, comparing the performance of linear and nonlinear discriminant analysis. Finally, conclusions are summarized in Section 4.

2 Distributive Entropy of Euclidean Distance

2.1 The Definition and Properties of DEED

Assume a matrix is formed by $m \times n$ dimension vectors, all vectors can be mapped to certain points in the Euclidean space. The Euclidean distances between any points can be used to compute a histogram. The Shannon entropy of this histogram is defined as distributive entropy of Euclidean distance (DEED). The value of DEED provides the uncertainty information concerning feature vectors. For a mutual classification problem, the distributive entropy of Euclidean distance between any sample and the sample mean of the same class is defined as within-class DEED (WCDEED). The distributive entropy of Euclidean distance between any sample in one class and the sample mean of other class is defined as between-class DEED (BCDEED). Theoretically, separability of features in training samples will be enhanced when the mapped points in Euclidean space are more consistently convergence to the center point. However, if the distribution of all mapped points in Euclidean space is diffusive, it should be difficult to achieve a good classification performance by using such training samples.

Shannon entropy is a strictly convex function, which reaches a maximum value when all probabilities are equal. Any approach to uniform the probability distributions will increase the entropy. Because DEED is a measure of histogram uniformity, it also shares the properties of entropy. That means the value of DEED should be large if the mapped points uniformly converge around a certain point, and the value of DEED should be small if the mapping points exhibit non-uniform convergence. Furthermore, the DEED measure can be extended to the mutual classification problem due to the additivity property of Shannon entropy. We propose DEED-based criterion function as follows:

$$J_{rd} = BCDEED / WCDEED . \tag{1}$$

Hence, the larger ratio of between-class DEED (BCDEED) to within class DEED (WCDEED) is, the better separability of training samples will be. Unlike FLDA, equation (1) is an efficient criterion via nonlinear mapping, because it contains high order statistical information and improves the discriminative capability. In addition, as a nonparametric method, the DEED-based algorithm provides weighting information in training procedure for further application.

2.2 DEED-Based Algorithm

Consider a matrix $W \in R^{m \times n}$ representing all the vectors in a set of $m \times n$ dimensional vectors $u_i = (x_{i,1}, x_{i,2}, \cdots x_{i,n})$, it is normalized by L_2. Let $\bar{u} = (\bar{x}_1, \bar{x}_2, \cdots \bar{x}_n)$ denote the mean of u, it is convenient to obtain the Euclidean distance between each vector and the mean vector $\bar{u}$ of W and express in terms of array δ.

$$\delta_k(u_k, \bar{u}) = (u_k - \bar{u})(u_k - \bar{u})^T, \ k = 1, 2 \cdots m. \tag{2}$$

Set $\delta_{\max} = \max(\delta)$, $\delta_{\min} = \min(\delta)$, then $\delta_k \in [\delta_{\min}, \delta_{\max}]$. Given a constant N $(N << m)$, we get:

$$\Delta\delta = (\delta_{\max} - \delta_{\min}) / N. \tag{3}$$

There are N intervals of the form Φ_i give by

$$\Phi_i = \begin{cases} [\delta_{\min} + (i-1) \times \Delta\delta, \delta_{\min} + i \times \Delta\delta] & i = 1, 2, \cdots N-1 \\ [\delta_{\min} + (N-1) \times \Delta\delta, \delta_{\max}] & i = N \end{cases}. \tag{4}$$

We have various estimates of a density function, and obtain histograms of Euclidean distances belong to each intervals of Φ_i.

If p_i is defined as the number of samples belong to interval Φ_i, the sum of p_i (i=1, 2, …, N) is m. The probability of Φ_i is derived as follows when $m \to \infty$,

$$P_i = p_i / m, \qquad i = 1, 2, \cdots, N, \tag{5}$$

and

$$\sum_{i=1}^{N} P_i = 1. \tag{6}$$

Using equations (5) and (6), we may calculate the distributive histogram of Euclidean distance, and employ it to derive the DEED,

$$E(P) = -\sum_{i=1}^{N} P_i \log_2 P_i. \tag{7}$$

Given a confidence coefficient α, we denote the maximum Euclidean distance under confidence interval as $\hat{\delta}_{\max}$, and ignore those samples out of the confidence interval. Then the modified DEED can be obtained by multiplying the $f(\hat{\delta}_{\max})$ (for simplicity, let $\hat{f}(\hat{\delta}_{\max}) = \hat{\delta}_{\max}$) to the right side of equation (7),

$$\hat{E}(P) = -\hat{\delta}_{\max} \sum_{i=1}^{N} P_i \log_2 P_i \ . \tag{8}$$

The parameter $\hat{E}(P)$ contains the information of both the distribution and scatter range of mapped points in Euclidean space. We also abbreviate the name of modified distributive entropy of Euclidean distance $\hat{E}(P)$ to distributive entropy of Euclidean distance DEED for reason of convenience.

2.3 Validation in Simulation

Consider two classes of overlapping, two dimensional, Gaussian-distributed patterns labeled as 1 and 2. Let C_n denote the set of events in which a random vector X belongs to patterns labeled n. We have the conditional probability density function,

$$f_x(x / C_n) = \frac{1}{2\pi\sigma^2} \exp(-\frac{1}{2\sigma^2} \|\bar{x} - \bar{u}\|^2) \ , \tag{9}$$

where $\bar{u}$ is the mean of two dimension vectors, σ^2 is variance. In two-class problem, $n = 1,2$. Five sets of data with different separabilities can be obtained by changing the parameters $\bar{u}$ and σ^2 . Assume three conditions of equal probability, the costs of correct classifications are zero, and an equal cost to all misclassifications, we determine the optimum decision boundary using the likelihood ratio test as shown in Table 1.

Table 1. Parameters of numerical simulation

	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5
C_1 [1]	[(0,0); 2]	[(0,0); 1]	[(0,0); 1]	[(0,0); 1]	[(0,0); 1]
C_2 [1]	[(2,0) ; 4]	[(0,2) ; 4]	[(0,3) ; 6]	[(0,4) ; 8]	[(0,5) ; 10]
Decision boundary[2]	(-2,0); 3.68	(-0.67,0); 2.34	(-0.6,0); 2.54	(-0.57,0); 2.71	(-0.55,0); 2.86
Correct classification probability [3]	0.7428	0.8164	0.8763	0.9145	0.9385

1. The mean value and variance of Gaussian model.
2. The center and the radius of Bayesian decision boundary.
3. The average correct classification probability of 20 samples set.

The simulation results show the efficiency of J_{rd} as a separability criterion, and a large criterion value corresponds to the enhanced performance of classification.

3 Experiment Results and Discussions

In this section, we apply the separability criterion of J_{rd} to the classification of ground vehicles. The data set consists of 3250 samples of 5 types of wheeled vehicle and 4250 samples of 9 types of tracked vehicles, which is collected from four field experiments. The sampling rate is 1000 Hz, and the ground vehicles are classified by tracked and wheeled vehicles.

3.1 Various Features Extraction of Ground Vehicles

Some features of ground vehicle have been obtained by analyzing the noise signal of ground vehicles, e.g. 1) most noise energy of vehicle are spread at frequencies in the range of 0-500 Hz; 2) the tracked vehicles is distinctive from wheeled vehicle due to its stronger harmonic component; 3) there is more energy in lower frequency and the fundamental frequency varies with the status of running vehicle. These three features were adopted for the tracked and wheeled vehicle classification.

3.1.1 Non-uniform Subband Energy Feature

A filterbank with 25 bands was designed to filter the noise signal. The features were derived from energies of each band and represented by a 25-dimension vector. A second order IIR peaking filter was determined by the equation (10) [7], [8].

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b_0 + b_1 \cdot z^{-1} + b_2 \cdot z^{-2}}{1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2}} \quad (10)$$

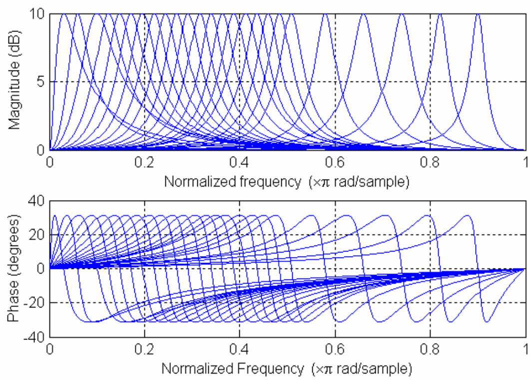

Fig. 1. Non-uniform subbands filters

The frequency and phase response of filter bank are shown in Fig. 1. Frequency bands dividing are determined by equation (11). From Fig. 1, the frequency between the 50Hz -200Hz is partitioned to 18 bands, and the frequency below 50 Hz was divided into 2 bands, while the other five bands were assigned to the frequency from 250 Hz to 500 Hz. Obviously, the frequency between 50-200 Hz was analyzed more carefully than other frequency bands.

$$\begin{cases} F_c(1) = 15 \\ F_c(2) = 30 \\ F_c(3) = 50 \\ F_c(i) = 50 + (i-2)\times 12 & 3 \le i \le 19 \\ F_c(i) = 250 + (i-19)\times 40 & 20 \le i \le 24 \end{cases} . \tag{11}$$

3.1.2 Modified Mel Frequency Cepstral Coefficient Feature

Cepstral analysis is an effective analysis method for harmonic signal. Mel frequency cepstral coefficient (MFCC) is a feature widely used in the context of speech recognizing because it considers the hearing mechanism. The energy of recorded signal concentrate in the frequency range of 0-500 Hz, therefore we design a new nonlinear mapping function,

$$F_{smel} = 900 \times \log_{10}(1 + f_n / 300) . \tag{12}$$

Using equation (12), we divide F_{smel} into 25 bands uniformly in the frequency range of interest. The center frequency is determined by mapping the center of each band in F_{smel} to a linear frequency. According to the center frequency in linear frequency, the triangle filter banks of modified MFCC can be designed. The relationship between F_{smel} and linear frequency are shown in Fig. 2, and the 25 triangle filters are shown in Fig. 3. The 25 dimensions features were obtained using equations (13)-(15). Similar to the hearing mechanism, the high frequency components were analyzed in a large scale and the low frequency components were analyzed in a small scale. The harmonic characteristics can be achieved by the cepstral analysis.

$$X(k) = \sum_{n=0}^{N} x(n) e^{-jwn} , \quad k = 1 : N , \tag{13}$$

$$\hat{X}(l) = \sum_{k=0}^{N/2} | X(k) |^2 \, M_l(k) , \tag{14}$$

where $M_l(k)$ corresponds to the triangle filter of the *k*th band.

$$c(i) = \sqrt{\frac{2}{L}} \sum_{m=1}^{l} \log_{10}(\hat{X}(l)) \cos[\frac{i\pi}{L}(m - 0.5)] , \ i = 1 \cdots L . \tag{15}$$

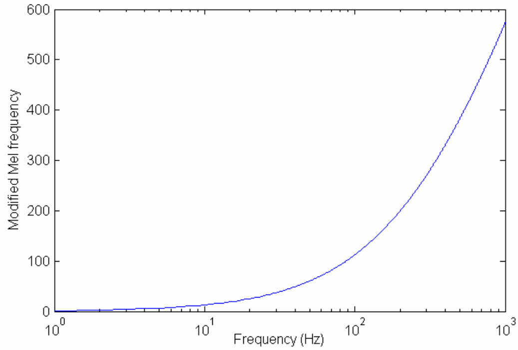

Fig. 2. Relationship of modified Mel frequency and linear frequency

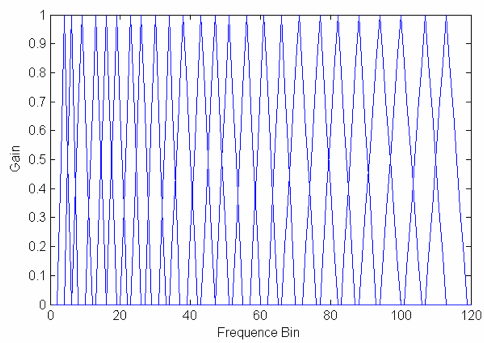

Fig. 3. Modified Mel frequency triangle filters

3.1.3 The Coefficient of Wavelet Package

The feature of wavelet package was selected by using typical wavelet kernel function 'db6'. We used 5-scale wavelet analysis of the recorded signal from vehicles and achieved 32 dimensions vector. The wavelet algorithm for features extraction is provided in the Wavelet Toolbox of Matlab.

3.2 Comparison of the Performance of Class Separability Criteria

The experiments randomly choose two third of tracked vehicle samples as training samples from database, the rest are used for test. Similarly, we obtain the training and test samples of wheeled vehicle. There are 2166 wheeled vehicle samples and 3030 tracked vehicle samples in the training set. After 20 times such selection independently, we create 20 sets of both training and test set for the classification of ground vehicles. We denote the wheel samples as 'W', and the tracked samples as 'T'.

3.2.1 Separability Estimation by the Use of Criterion J_{rd}

Distributive histogram of Euclidean distance and DEED can be calculated from the three features given in Section 3.1, and shown in Figs. 4-6 and Table 2 respectively.

The number of statistical interval is 256. From Table 2, it is seen that the best separability of feature is non-uniform subband feature, and the modified MFCC feature is inferior to the non-uniform sub-band feature. The worst feature for separating class is the wavelet package feature.

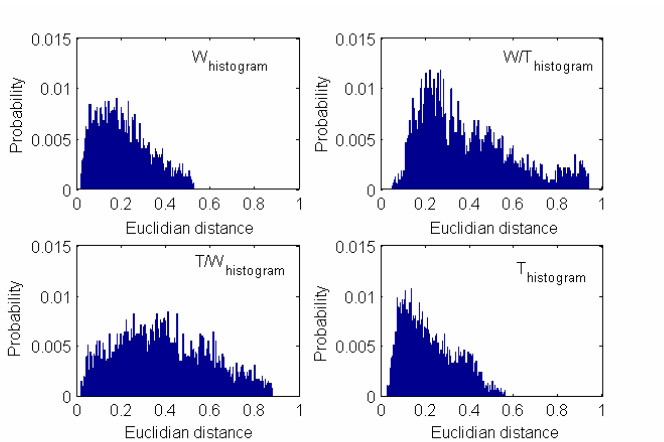

Fig. 4. Distributive histogram of Euclidean distance for the non-uniform subband filter

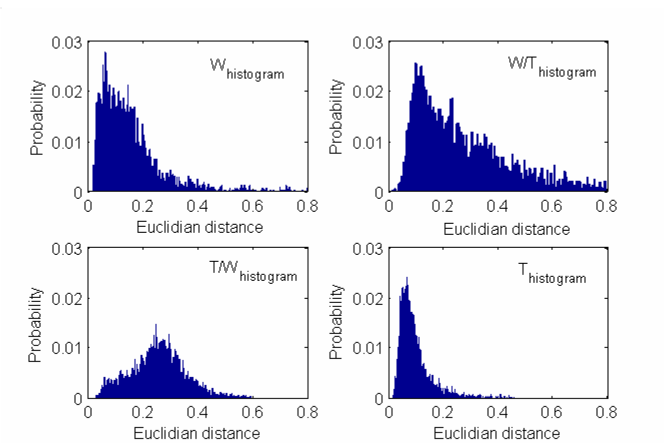

Fig. 5. Distributive histogram of Euclidean distance for the modified MFCC feature

3.2.2 Separability Estimation by the Use of Trace and Determinant Criteria

Consider the FLDA, the between-class scatter matrix S_b and within-class scatter matrix S_w are computed using the feature vectors of training samples and used in the trace criterion J_t and determinant criterion J_d defined by

$$J_t = tr(S_w^{-1} S_b)\,, \tag{16}$$

and

$$J_d = |S_w + S_b| / |S_w| \tag{17}$$

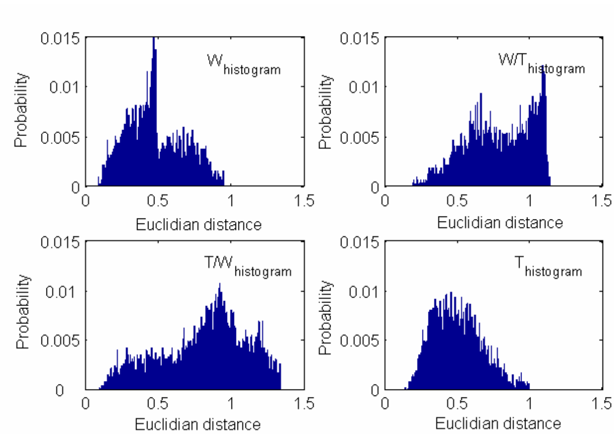

Fig. 6. Distributive histogram of Euclidean distance for the wavelet package feature

Table 2. The average DEED of three features used in experiments and criterion J_{rd}

	W	W/T	T/W	T	WT	CDEED	J_{rd}
Non-uniform subban feature	3.8082	7.3273	6.9469	3.4835	7.2917	14.2742	1.9576
Modified MFCC	4.7927	8.5399	4.1159	2.2804	7.0731	12.6558	1.7893
Wavelet package	6.2504	7.6938	9.227	6.3796	12.63	13.9442	1.1041

Note: W: WCDEED of wheeled vehicle,
T: WCDEED of tracked vehicle,
WT: sum of W and T
W/T: BCDEED by the wheeled samples to mean vector of tracked
T/W: BCDEED by the tracked samples to mean vector of wheel
CDEED: the sum of T/W and W/T

The criteria values using the three features are shown in Table 3. It is different from the observation in section 3.2.1 that modified MFCC feature is the best for the separability estimation, whereas the non-uniform subband feature is inferior one. However, the wavelet package feature is the worst one among three kinds of features. Furthermore, we employed the K-mean cluster method to analyze the separability of three features. The clustering results are also shown in Table 3. It is clear that the clustering performance using non-uniform sub-band feature is the best, modified MFCC feature is inferior while the performance using the feature of wavelet package is the worst.

Table 3. The average value of fisher criterion J_t , J_d and cluster results

	J_t	J_d	Correct clustering rate of W (%)	Correct clustering rate of T (%)	Average correct clustering rate of W and T (%)
Non-uniform sub-band feature	1.9347	2.9347	71.45	91.23	81.34
Modified MFCC	2.3569	3.5456	63.16	94.70	78.93
Wavelet package	1.7316	2.7316	70.12	87.48	78.80

3.2.3 Discussion of Three Separability Criteria

In the two-class experiments, the performances of the three separability criteria with three kinds of features are not consistent. Simulation results given in Table 3 indicate that the correct clustering rates of all tracked vehicle features are higher than those of wheeled vehicle features. It is also explained in Figs. 4-6. Take the modified MFCC feature as an example, the Euclidean distance from the feature vector of wheeled vehicle to mean vector of tracked vehicle's feature are spread in the interval of 0-0.2, which is closed to the distance distribution of tracked vehicle feature (0-0.1). That is why the misclassification rate of wheeled samples is high. In addition, the Euclidean distance from tracked vehicle feature vector to mean vector of wheeled vehicle feature are spread in the interval of 0.2-0.4, which less overlap the distance distribution of wheeled vehicle feature (0.1-0.2). It results in a good clustering performance for the tracked vehicle. We can draw a similar conclusion when considering the other two features. Hence, the distributive histogram of Euclidean distance provides a reasonable explanation for the clustering results.

3.3 Validation of the Criterion J_{rd} by the Use of Classification Algorithms

In the two-class ground vehicle classification problem, we adopted the distributive histogram of Euclidean distance and separability criterion J_{rd} . The separability estimation result shows that the non-uniform subband classification feature yields the best separability performance, the inferior one is the modified MFCC feature and the separability of wavelet is the worst. In the case of using FLDA-based J_t and J_d , the modified MFCC feature exhibits the best separability and the non-uniform subband feature is inferior to the modified MFCC feature. The wavelet package is the worst one in terms of separability. However, we have the same observation from cluster analysis as the estimation method using J_{rd} . Next, we will validate it by using three supervised classifiers.

The experiments choose 20 independent training sets for each feature. The first classifier is K nearest neighbor (KNN) classifier. Let k = 3 and the number of reference samples be that of training samples. The second one is a three-layer back- propagation (BP) neural network with 25 input nodes, 12 hide layers and 2 output nodes. The third one is a support vector machine (SVM) classifier. The kernel function of SVM is a radial basis function (RBF). Classification results are shown in the table 4.

It is observed from Table 4 that, in the case of using KNN classifier the classification performance of modified MFCC feature is the best, non-uniform subband feature is inferior, and the wavelet package feature is the worst one. This result is the same as the separability estimation result using FLDA-based criteria. Secondly, in the case of using SVM and BP algorithms, the non-uniform subband feature exhibits the best classification performance. The modified MFCC feature yields inferior performance and the wavelet package feature is the worst one for classification. This result is consistent with the separability estimation using the new criterion J_{rd} .

Table 4. Classification results of three kinds classifier

	Correct classification probability of wheeled vehicle (%)			Correct classification probability of tracked vehicle (%)			Average correct classification probability of wheeled and tracked vehicles (%) [1]		
	KNN	SVM	BP	KNN	SVM	BP	KNN	SVM	BP
Non-uniform subband feature	95.7	97.20	97.30	96.90	98.90	97.00	96.30	98.05	97.15
Modified MFCC	96.5	96.60	94.20	96.80	98.00	95.70	96.65	97.30	94.95
Wavelet package	93.4	95.10	96.18	92.20	96.20	88.39	92.80	96.65	92.29

[1] The classification results of KNN classifier accord well with those using J_t and J_d , whereas the classification results from BP and SVM algorithms accord well with those using J_{rd}

Experiment results show that the Fisher linear criteria J_t and J_d are suitable to use with KNN classifier as explained in Ref. [2]. The proposed criterion J_{rd} is suitable to use in the BP and SVM classification experiments. The separability criterion J_{rd} is a nonlinear parameter based on the DEED analysis, which preserves most classification information after the feature has been transformed to the high dimension space.

4 Conclusion

In this paper, we have considered the use of a new optimization criterion based on the histogram entropy of Euclidean distance for classification. A nonlinear parameter DEED is defined for pattern classification. With help of the DEEE, a criterion function J_{rd} for feature extraction can be set up. It has been shown that the larger J_{rd} (ratio of between-class DEED to within-class DEED) is, the better separability of learning samples will be. Because the entropy is an invariant for the nonlinear transform, the DEED is able to preserve most classification information.

Experiment results show that the proposed criterion can improve the classification performance of the extracted features compared to other linear Fisher criteria used in pattern recognition. The DEED-based separability estimation criterion is better than FDLA in the case of using neural network classifier. It should be noted that DEED is an information measure of distributive histogram of Euclidean distance, which provides more information for further study on dynamic learning theory. Still more research about separability criterion as well as improved methods for an optimal feature extraction are necessary.

References

1. Fukunaga, K.: Introduction to statistical pattern recognition, 2nd edn. Academic Press, Boston (1990)
2. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley, New York (2000)
3. Covet, T.M.: Geometrical and statistical properties of systems of linear inequalities with applications to pattern recognition. IEEE Trans. Electronic Computers 14, 326–334 (1965)
4. Haykin, S.: Neural Networks: A comprehensive foundation, 2nd edn. Prentice-Hall, Englewood Cliffs (1999)
5. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, Berlin (1995)
6. Moore, C.J.: Medical image processing: the characterization of display changes using histogram entropy. Image Vision Computing 4, 197–202 (1986)
7. Robert, B.J.: The Equivalence of Various Methods of Computing Biquad Coefficients for Audio Parametric Equalizers. Wave Mechanics, Inc, Burlington VT, http://www.wavemechanics.com
8. White, S.A.: Design of a Digital Biquadratic Peaking or Notch Filter for Digital Audio Equalization. J. Audio Eng. Soc. 34, 479–483 (1986)