

# Pattern Classification using Rectified Nearest Feature Line Segment

Hao Du and Yan Qiu Chen \*

Department of Computer Science and Engineering,  
School of Information Science and Engineering,  
Fudan University,  
Shanghai 200433, China

**Abstract.** This paper proposes a new classification method termed Rectified Nearest Feature Line Segment (RNFLS). It overcomes the drawbacks of the original Nearest Feature Line (NFL) classifier and possesses a novel property that centralizes the probability density of the initial sample distribution, which significantly enhances the classification ability. Another remarkable merit is that RNFLS is applicable to complex problems such as two-spirals, which the original NFL cannot deal with properly. Experimental comparisons with NFL, NN(Nearest Neighbor), k-NN and NNL (Nearest Neighbor Line) using artificial and real-world datasets demonstrate that RNFLS offers the best performance.

## 1 Introduction

Nearest Feature Line (NFL) [1], a newly developed nonparametric pattern classification method, has recently received considerable attention. It attempts to enhance the representational capacity of a sample set of limited size by using the lines passing through each pair of the samples belonging to the same class. Simple yet effective, NFL shows good performance in many applications, including face recognition [1] [2], audio retrieval [3], image classification [4], speaker identification [5] and object recognition [6].

On the other hand, feature lines may produce detrimental effects that lead to increased decision errors. Compared with the well-known Nearest Neighbor (NN) classifier [7], NFL has obvious drawbacks under certain situations that limit its further potential. The authors of [8] pointed out one of the problems – extrapolation inaccuracy, and proposed a solution called Nearest Neighbor Line (NNL). This extrapolation inaccuracy may lead to enormous decision errors in a low dimensional feature space while a simple NN classifier easily reaches a perfect correct classification rate of 100%. Another drawback of NFL is interpolation inaccuracy. Distributions assuming a complex shape (two-spiral problem for example) often fall into this category, where, by the original NFL, the interpolating parts of the feature lines of one class break up the area of another class and severely damage the decision region.

---

\* Corresponding author. Email: chenqy@fudan.edu.cn

In this paper, a new nonparametric classification method, *Rectified Nearest Feature Line Segment* (RNFLS), is proposed that addresses both of the above-mentioned drawbacks and significantly improves the performance of NFL. The original NFL can conceptually be viewed as a two-stage algorithm – building representational subspaces for each class and then performing the nearest distance classification. We focus mainly on the first stage. To overcome extrapolation inaccuracy, *Nearest Feature Line Segment subspace* (NFLS-subspace) is developed. For the interpolation inaccuracy, the “territory” of each sample point and each class is defined, and we obtain *Rectified Nearest Feature Line Segment subspace* (RNFLS-subspace) from NFLS-subspace by eliminating those feature line segments trespassing the territory of other classes. As a result, RNFLS works well for all shapes of sample distribution, which is a significant improvement.

Another remarkable advantage of RNFLS is that it centralizes the probability density of the initial sample distribution. We show, in an experiment, that the decision region created by RNFLS gets closer to the one built by using the optimal Bayesian rule, bringing the correct classification rate higher. Comparisons with NN, k-NN, NFL, NNL using artificial and real-world datasets demonstrate that the proposed RNFLS method offers remarkably superior performance.

## 2 Background

### 2.1 The Nearest Feature Line Method

The Nearest Feature Line (NFL) [1] method constructs a feature subspace for each class, consisting of straight lines passing through every pair of the samples belonging to that class. The straight line passing through samples  $x_i, x_j$  of the same class, denoted by  $\overline{x_i x_j}$ , is called a *feature line* of that class. All the feature lines of class  $\omega$  constitute an *NFL-subspace* to represent class  $\omega$ , denoted by  $\mathcal{S}_\omega = \{\overline{x_i^\omega x_j^\omega} | x_i, x_j \in \omega, x_i \neq x_j\}$ , which is a subset of the entire feature space.

During classification, a query point  $q$  is classified to class  $\omega$  if  $q$  assumes the smallest distance to  $\mathcal{S}_\omega$  than to any other  $\mathcal{S}_{\omega'}$ , ( $\omega \neq \omega'$ ). The distance from  $q$  to  $\mathcal{S}_\omega$  is

$$d(q, \mathcal{S}_\omega) = \min_{\overline{x_i x_j} \in \mathcal{S}_\omega} d(q, \overline{x_i x_j}). \quad (1)$$

$$= \min_{\overline{x_i x_j} \in \mathcal{S}_\omega} \|q - p_{ij}\| \quad (2)$$

where  $p_{ij}$  is the projection point of  $q$  onto line  $\overline{x_i x_j}$ .

The projection point can be computed by

$$p_{ij} = (1 - \mu)x_i + \mu x_j, \quad (3)$$

where

$$\mu = \frac{(q - x_i) \cdot (x_j - x_i)}{(x_j - x_i) \cdot (x_j - x_i)}. \quad (4)$$

## 2.2 Shortcomings

NFL extends the samples of one class by adding the straight lines linking each pair. A good argument for doing this is that it adds extra information to the sample set. The extra information, however, is a double-edged sword. When a straight line of one class trespasses into the territory of another class, it will lead to increased error probability. There are two types of trespassing, causing two types of inaccuracies: *extrapolation inaccuracy* and *interpolation inaccuracy*.

Fig.1(a) shows a classification problem in which the *extrapolation inaccuracy* occurs. The query  $q$ , surrounded by four “cross” sample points, is in the territory of the “cross” class, leading to the expectation that  $q$  should be classified to the “cross” class. But the extrapolating part of feature line  $\overline{x_1x_2}$  makes the distance from  $q$  to  $\overline{x_1x_2}$  smaller. Thus,  $d(q, \mathcal{S}_{circle}) < d(q, \mathcal{S}_{cross})$ , and NFL will assign  $q$  the label “circle”, not “cross”. This is very likely to be a decision error. Similarly, the *interpolation inaccuracy* caused by the interpolating part of a feature line is illustrated in Fig.1(b).

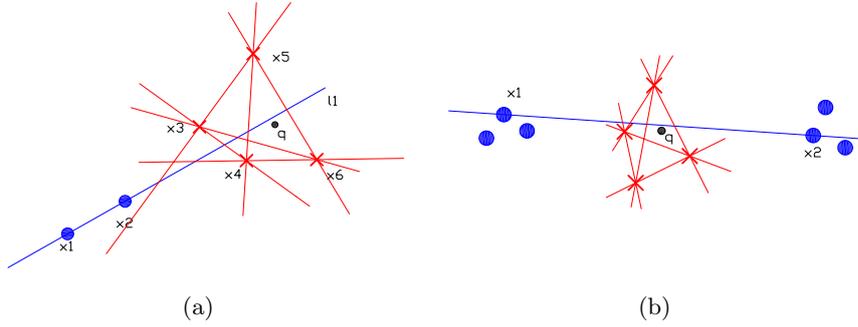


Fig. 1. (a)Extrapolation inaccuracy. (b)Interpolation inaccuracy.

The above inaccuracies are drawbacks that limit the applicability of NFL. In the following section we pursue a more systematic approach in which a new feature subspace for each class is constructed to avoid both drawbacks. The original advantage of NFL that linearly extending the representational capacity of the original samples is retained in our method.

## 3 Rectified Nearest Feature Line Segment

### 3.1 Using Feature Line Segments

To avoid extrapolation inaccuracy, we propose to use line segments between pairs of the sample points to construct a *Nearest Feature Line Segment subspace* (NFLS-subspace) instead of the original NFL-subspace to represent each class.

Let  $X^\omega = \{x_i^\omega | 1 \leq i \leq N_\omega\}$  be the set of  $N_\omega$  samples belonging to class  $\omega$ . The NFLS-subspace ( $\tilde{\mathcal{S}}_\omega$ ) representing class  $\omega$  is

$$\tilde{\mathcal{S}}_\omega = \{\widetilde{x_i^\omega x_j^\omega} | 1 \leq i, j \leq N_\omega\}, \quad (5)$$

where  $\widetilde{x_i^\omega x_j^\omega}$  denotes the line segment connecting point  $x_i^\omega$  and  $x_j^\omega$ . Note that a degenerative line segment  $\widetilde{x_i^\omega x_i^\omega}$  ( $1 \leq i \leq N_\omega$ ), which is a point in the feature space, is also a member of  $\tilde{\mathcal{S}}_\omega$ .

The distance from a query point  $q$  to an NFLS-subspace  $\tilde{\mathcal{S}}_\omega$  is defined as

$$d(q, \tilde{\mathcal{S}}_\omega) = \min_{x_i x_j \in \tilde{\mathcal{S}}_{\omega_k}} d(q, \widetilde{x_i x_j}) \quad (6)$$

where

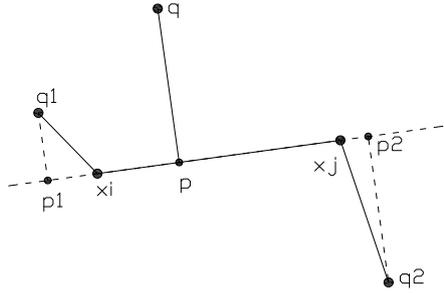
$$d(q, \widetilde{x_i x_j}) = \min_{y \in \widetilde{x_i x_j}} \|q - y\|. \quad (7)$$

And to calculate  $d(q, \widetilde{x_i x_j})$ , there are two cases. If  $x_i = x_j$ , the answer is simply the point to point distance,

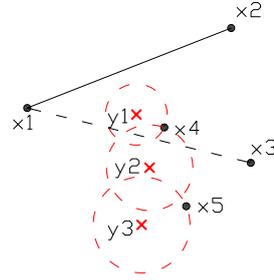
$$d(q, \widetilde{x_i x_i}) = \|q - x_i\|. \quad (8)$$

Otherwise, the projection point  $p$  of  $q$  onto  $\widetilde{x_i x_j}$  is to be located first by using Equ.(3) and Equ.(4). Then, different reference points are chosen to calculate  $d(q, \widetilde{x_i x_j})$  according to the position parameter  $\mu$ . When  $0 < \mu < 1$ ,  $p$  is an interpolation point between  $x_i$  and  $x_j$ , so  $d(q, \widetilde{x_i x_j}) = \|q - p\|$ . When  $\mu < 0$ ,  $p$  is a “backward” extrapolation point on the  $x_i$  side, so  $d(q, \widetilde{x_i x_j}) = \|q - x_i\|$ . When  $\mu > 1$ ,  $p$  is a “forward” extrapolation point on the  $x_j$  side, so  $d(q, \widetilde{x_i x_j}) = \|q - x_j\|$ . Fig.2 shows an example.

In the classification stage, a query  $q$  is classified to class  $\omega_k$  when  $d(q, \tilde{\mathcal{S}}_{\omega_k})$  is smaller than the distance from  $q$  to any other  $\tilde{\mathcal{S}}_{\omega_i}$  ( $\omega_i \neq \omega_k$ ).



**Fig. 2.** Distance (solid lines) from feature points to feature line segment  $\widetilde{x_i x_j}$ .



**Fig. 3.** The territory of “cross”-samples shown in dashed circle.

### 3.2 Rectifying the Feature Line Segment Subspace

The next step is to rectify the NFLS-subspace to eliminate interpolation inaccuracy. Our motivation is to have the inappropriate line segments removed from the NFLS-subspace  $\tilde{\mathcal{S}}_{\omega_k}$  for each class  $\omega_k$ . The resulting subspace denoted by  $\tilde{\mathcal{S}}_{\omega_k}^*$  is a subset of  $\tilde{\mathcal{S}}_{\omega_k}$  termed *Rectified Nearest Feature Line Segment subspace* (RNFLS-subspace).

**Territory** We begin with the definitions of two types of territories. One is *sample-territory*,  $T_x \in \mathbb{R}^n$ , that is the territory of a sample point  $x$ ; the other is *class-territory*,  $\mathcal{T}_\omega \in \mathbb{R}^n$ , that is the territory of class  $\omega$ . Suppose the sample set  $X$  is  $\{(x_1, \theta_1), (x_2, \theta_2), \dots, (x_m, \theta_m)\}$ , which means  $x_i$  belongs to class  $\theta_i$ . The *radius*  $r_{x_k}$  of the sample-territory  $T_{x_k}$  is,

$$r_{x_k} = \min_{\forall x_i, \theta_i \neq \theta_k} \|x_i - x_k\|. \quad (9)$$

Thus,

$$T_{x_k} = \{y \in \mathbb{R}^n \mid \|y - x_k\| < r_{x_k}\}. \quad (10)$$

The class-territory  $\mathcal{T}_{\omega_k}$  is defined to be

$$\mathcal{T}_{\omega_k} = \bigcup_{\theta_i = \omega_k} T_{x_i}, \quad (x_i, \theta_i) \in X. \quad (11)$$

In Fig.3, the points denoted by ‘‘circle’’ and ‘‘cross’’ represent the samples from two classes. Each of the ‘‘cross’’-points ( $y_1, y_2, y_3$ ) has its own sample-territory as shown by the dashed circle. The union of these sample-territories is  $\mathcal{T}_{cross}$ .  $\mathcal{T}_{circle}$  is obtained in a similar way.

**Building RNFLS-subspace** For class  $\omega_k$ , its RNFLS-subspace  $\tilde{\mathcal{S}}_{\omega_k}^*$  is built from the NFLS-subspace  $\tilde{\mathcal{S}}_{\omega_k}$  by having those line segments trespassing the class-territories of other classes removed. That is

$$\tilde{\mathcal{S}}_{\omega_k}^* = \tilde{\mathcal{S}}_{\omega_k} - \tilde{\mathcal{U}}_{\omega_k}, \quad (12)$$

where ‘‘-’’ is the set difference operator, and

$$\begin{aligned} \tilde{\mathcal{U}}_{\omega_k} &= \{\widetilde{x_i x_j} \mid \exists \omega_y, \omega_k \neq \omega_y \wedge \widetilde{x_i x_j} \in \tilde{\mathcal{S}}_{\omega_k}^* \wedge \widetilde{x_i x_j} \cap \mathcal{T}_{\omega_y} \neq \emptyset\} \\ &= \{\widetilde{x_i x_j} \mid \exists (x_y, \theta_y) \in X, \widetilde{x_i x_j} \in \tilde{\mathcal{S}}_{\omega_k}^* \wedge \omega_k \neq \theta_y \wedge d(x_y, \widetilde{x_i x_j}) < r_{x_y}\}. \end{aligned} \quad (13)$$

**Classifying using RNFLS-subspaces** To perform classification using RNFLS-subspaces is similar to using NFLS-subspaces, since the only difference between an RNFLS-subspace and an NFLS-subspace is  $\tilde{\mathcal{S}}_{\omega_k}^* = \tilde{\mathcal{S}}_{\omega_k} - \tilde{\mathcal{U}}_{\omega_k}$ , where, except for some removed line segments,  $\tilde{\mathcal{S}}_{\omega_k}^*$  is still a set consisting of line segments. The distance measure from a query point to the RNFLS-subspace remains the same.

### 3.3 Analyzing the Centralization Property

In many real-world pattern recognition problems, samples from one class tend to scatter around a certain center point because of systematic error and random noise. Gaussian distribution is an example. Two scattered classes may overlap each other, causing decision errors. Compared with the original ideal sample distribution without noise, the NFLS-subspace of each class has an impressive property – distribution centralization, which can be viewed as the converse of scattering. With the help of NFLS-subspace, the distribution overlapping is reduced, and the probability distribution grows closer to the original. And so, we get a higher correct classification rate.

The simplest case to show the centralization property is when the distribution is uniform in a two-dimensional feature space. Suppose that the sample points of class  $\omega$  are uniformly distributed in a disk  $D$  whose radius is  $R$  and the center is at  $O$ , as shown in Fig.4. For the NFLS-subspace of the class, consider a small region  $M(a, r)$  ( $a \leq R$ ), that is a round area with an arbitrarily small radius  $r$  and distance  $a$  from  $O$ . Let  $N_a^\omega$  be the probability of a randomly selected feature line segment of class  $\omega$  passing through  $M(a, r)$ .

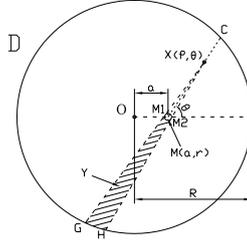


Fig. 4. Calculating  $N_a^\omega$  for a uniform sample point density on a disk.

**Proposition 1** Given an arbitrarily small  $r$ ,  $N_a^\omega$  is decreasing on  $a$ .

*Proof* We calculate  $N_a^\omega$  in a polar coordinate system by choosing the center of  $M(a, r)$  as pole and  $\overrightarrow{OM}$  as polar axis. For a line segment  $\widetilde{XY}$  passing through  $M(a, r)$ , given one endpoint  $X(\rho, \theta)$  in  $D$ , the other endpoint  $Y$  has to appear in the corresponding  $\square_{M_1M_2HG}$ , as shown in Fig.4. Thus we obtain

$$\begin{aligned} N_a^\omega &= \iint_D \frac{1}{\pi R^2} A(\rho, \theta) \rho d\rho d\theta \\ &= \int_0^{2\pi} \int_0^{|MC|} \frac{1}{\pi R^2} A(\rho, \theta) \rho d\rho d\theta \end{aligned} \quad (14)$$

where  $A(\rho, \theta)$  is the probability that the randomly generated endpoint  $Y$  appears in  $\square_{M_1 M_2 HG}$ ,

$$A(\rho, \theta) = \frac{1}{\pi R^2} \left[ \frac{1}{2}(2r + |GH|) \cdot |MG| + o(r) \right] \quad (15)$$

According to Equ.(14) and (15)

$$N_a^\omega = \frac{2r(R^2 - a^2)}{(\pi R^2)^2} \int_0^{2\pi} \sqrt{R^2 - a^2 \sin^2 \theta} \cdot d\theta + o(r). \quad (16)$$

Thus, for a fixed  $r$ ,  $N_a^\omega$  gets smaller when  $a$  gets larger.

Proposition 1 indicates that the distribution of line segments in the NFLS-subspace is denser at the center than at the boundary if the original sample points distribution is under a uniform density. A Gaussian distribution can be viewed as a pile-up of several uniform distribution disks with the same center but different radius. It is conjectured that this centralization property also applies to the Gaussian case, and can be extended to classification problems in which the overlapping is caused by noise scattering of two or more classes under similar distribution but different centers. It reverses the scattering and achieves a substantial improvement.

## 4 Experiment Results and Discussions

The performance of the RNFLS method is compared with four classifiers - NN, k-NN, NFL and NNL - using two artificial datasets as well as a group of real-world benchmarks widely used to evaluate classifiers. The results on these datasets, representing various distributions and different dimensions, demonstrate that RNFLS possesses remarkably stronger classification ability than the other four methods.

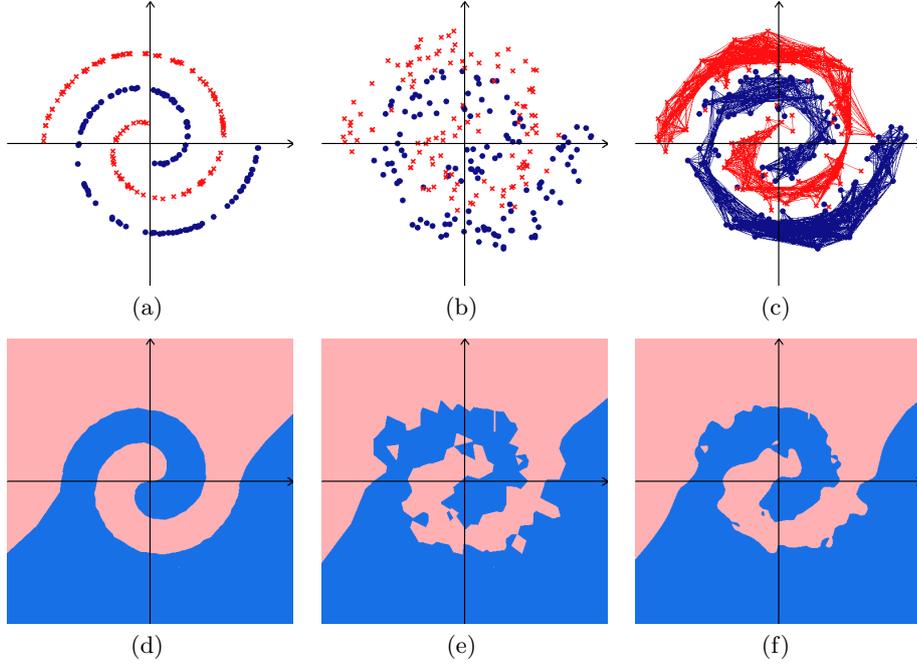
### 4.1 The Two-Spiral Problem

The two-spiral problem is now included by many authors as one of the benchmarks for evaluation of new classification algorithms. The two-spiral curves in a two-dimensional feature space is described as follows

$$spiral1 : \begin{cases} x = k\theta \cos(\theta) \\ y = k\theta \sin(\theta) \end{cases} \quad spiral2 : \begin{cases} x = k\theta \cos(\theta + \pi) \\ y = k\theta \sin(\theta + \pi) \end{cases} \quad (17)$$

where  $\theta \geq \pi/2$  is the parameter. If the probability density of each class is uniform along the corresponding curve, an instance of such distribution is shown in Fig.5(a).

In our experiment, Gaussian noise is added to the samples so that the distribution regions of the two classes may overlap each other, as shown in Fig.5(b). If the prior distribution density were known, according to the optimal Bayesian rule, Fig.5(d) should be the optimal decision region. This, however, can hardly



**Fig. 5.** (a)Two-spiral problem. (b)Two-spiral problem with Gaussian noise. (c)RNFLS subspaces. (d)Bayesian decision region. (e)NN classification result. (f)RNFLS classification result.

be achieved because the only information we have is from a finite number of sample points.

The original NFL is not a good choice for this classification problem. We may imagine how fragmented the decision region is carved up because of its interpolation and extrapolation inaccuracy. The decision region created by NN rule is shown in Fig.5(e). When it comes to RNFLS, Fig.5(c) is the RNFLS-subspaces and Fig.5(f) is the corresponding decision region. Compared with the decision region created by NN, RNFLS produces a much better one in which the boundary is smoother and some incorrect regions caused by isolated noise points is smaller. This significant enhancement can be attributed to the centralization property.

As a concrete test, let  $\theta \in [\pi/2, 3\pi]$  and the Gaussian noise is of a variance  $\sigma = 1.7$  and an expectation  $\mu = 0$ . We produce 500 points according to the well-defined distribution, where 250 belong to class  $\omega_1$  and the other 250 belong to class  $\omega_2$ . Then, half of them are randomly chosen to form the sample set and the remaining half constitute the test set. The classifiers, NN, k-NN(k=3), NFL, NNL and RNFLS, are applied to this task for 10 times, and Table 1 shows the results.

**Table 1.** Performance evaluation on the two-spiral problem using NN, 3-NN, NFL[1], NNL[8] and RNFLS. (CCR: correct classification rate, percentage)

Classifier	CCR (average)	CCR (min)	CCR(max)
NN	83.2	80.4	85.3
k-NN(k=3)	85.3	83.2	87.3
NFL	53.2	49.8	56.7
NNL	72.4	69.0	78.0
<b>RNFLS</b>	<b>86.1</b>	<b>84.0</b>	<b>88.2</b>

## 4.2 Real-World Classification Problems

We test the RNFLS classifier on a group of real-world datasets as listed in Table 2. All of the datasets are obtained from the U.C. Irvine repository [9]. Since we do not deal with the issue of missing data, instances with missing values are removed. For the fairness of the procedure, attributes of the instances are standardized (normalized) by their means and standard deviations before submitted to the classifiers. The performance in CCR is obtained using the leave-one-out procedure.

**Table 2.** CCR(%) for NN, 3-NN, NFL, NNL and RNFLS on the real-world datasets

Dataset	#Classes	#Instances	#Attributes	NN	3NN	NFL	NNL	RNFLS
1 iris	3	150	4	94.7	94.7	88.7	94.7	<b>95.3</b>
2 housing	6	506	13	70.8	73.0	71.1	67.6	<b>73.5</b>
3 pima	2	768	8	70.6	<b>73.6</b>	67.1	62.8	73.0
4 wine	3	178	13	95.5	95.5	92.7	78.7	<b>97.2</b>
5 bupa	2	345	6	63.2	65.2	63.5	57.4	<b>66.4</b>
6 ionosphere	2	351	34	86.3	84.6	85.2	87.2	<b>94.3</b>
7 wpbc	2	194	32	72.7	68.6	72.7	54.1	<b>75.8</b>
8 wdbc	2	569	30	95.1	96.5	95.3	64.0	<b>97.2</b>

It can be seen that RNFLS performs well on both two-category and multi-category classification problems in both low and high dimensional feature spaces. This is encouraging since these datasets represent real-world problems and none of them is specially designed to suit a specific classifier. Since one common characteristic of real-world problems is distribution dispersing caused by noise, the centralization property of RNFLS helps improving the correct classification rate.

## 5 Conclusions and Future Work

A new classification method RNFLS is developed. It enhances the representational capacity of the original sample points and constitutes a substantial improvement to NFL. It works well independent of the distribution shape and the feature-space dimension. In particular, viewed as the converse of sample scattering, RNFLS is able to centralize the initial distribution of the sample points and offers a higher correct classification rates for common classification problems.

Further investigation into RNFLS seems warranted. In the rectification process it would be helpful to reduce the runtime-complexity, perhaps using some kind of probability algorithms. It may also be helpful to treat the trespassing feature line segments more specifically, for example, finding a way to cut off a part of a trespasser instead of eliminating the whole feature line segments. Also worth more investigation is the centralization property, which might be of great potential.

## Acknowledgments

The research work presented in this paper is supported by National Natural Science Foundation of China, project No.60275010; Science and Technology Commission of Shanghai Municipality, project No. 04JC14014; and National Grand Fundamental Research Program of China, project No. 2001CB309401.

## References

1. S. Z. Li and J. W. Lu: Face recognition using the nearest feature line method. *IEEE Trans. Neural Networks* **10** (1999) 439-443
2. J. T. Chien and C. C. Wu: Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Trans. Pattern Anal. Machine Intell.* **24** (2002) 1644-1649
3. S. Z. Li: Content-based audio classification and retrieval using the nearest feature line method. *IEEE Trans. Speech Audio Processing* **8** (2000) 619-625
4. S. Z. Li, K. L. Chan and C. L. Wang: Performance evaluation of the nearest feature line method in image classification and retrieval. *IEEE Trans. Pattern Anal. Machine Intell.* **22** (2000) 1335-1339
5. K. Chen, T. Y. Wu and H. J. Zhang: On the use of nearest feature line for speaker identification. *Pattern Recognition Letters* **23** (2002) 1735-1746
6. J. H. Chen and C. S. Chen: Object recognition based on image sequences by using inter-feature-line consistencies. *Pattern Recognition* **37** (2004) 1913-1923
7. T. M. Cover and P. E. Hart: Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* **13** (1967) 21-27
8. W. M. Zheng, L. Zhao and C. R. Zou: Locally nearest neighbor classifiers for pattern classification. *Pattern Recognition* **37** (2004) 1307-1309
9. C. L. Blake and C. J. Merz: UCI repository of machine learning databases. <http://www.ics.uci.edu/~mlearn/MLRepository.html> (1998) [Online]