Perceptual Nearest Neighbors for Classification

Wen Guihua #1, Wen Jun *2, Jiang Lijun #1

South China University of Technology,Guangzhou, China
1 crghwen@scut.edu.cn

* *Hubei Institute for Nationalities,Ensi, China*
2 wenj_64@163.com

***Abstract*—Finding nearest neighbors plays a fundamental role in many artificial intelligence tasks, such as manifold learning, data mining, and information retrieval, etc. Directly applying this idea to perform classification is simple and often results in good performance on complex data types. However existing classifiers apply a well designed measure to find nearest neighbors. They still can not be comparable with human being in many complex cases such as on noisy, sparse or high dimensional data. This paper proposes a quite different but much interesting approach that utilizes Lipschitz function to define a simple topological transformation for modeling Gestalt laws of psychology from data and then designs a new measure to evaluate the quality of the discovered Gestalts. Subsequently, the nearest neighbors are selected from higher quality Gestalts, from which a new classifier is proposed that has much better classification performance.**

***Index Terms*—Classification, Gestalt laws, nearest neighbors, topological transformation**

I. Introduction

Finding nearest neighbors plays an fundamental role in many artificial intelligence tasks[1], such as manifold learning[2], data mining[3], [4], and information retrieval[5]. Manifold learning requires finding nearest neighbors to build the neighborhood graph that should faithfully represent the underlying data manifold[6], but currently found nearest neighbors are often deformed due to noise in data which in turn lead to drastically incorrect low-dimensional embedding[7]. The k-nearest neighbors (KNN) classifier directly applies this idea to perform classification. It is the one of most important classification approaches due to its high classification accuracy in problems with unknown and non-normal distributions[3], [10] and wide applications[8]. This approach finds the k nearest neighbors of the query in the training set, and then predicts the class label of the query as the most frequent one occurring in the neighbors[10]. However it heavily depends on the found nearest neighbors. On high-dimensional data, noisy data, and small data, its performance is severely influenced [9], under such cases it can not be comparable with human being. This is because all existing approaches of finding nearest neighbors heavily depends on some carefully selected measures[3], [11]. When the training data set is noisy and imbalanced, or particularly on manifold, the selected neighbors by these measures are often conflict with human perception, because they only consider the local properties rather than topological properties of the neighborhood. Their selected nearest neighbors often can not be able to organize an meaningful and regular group. According to Gestalt theory, grouping is the main process in human visual perception[12]. Whenever previously formed visual objects have one or several characteristics in common, they easily get grouped and form a new, larger visual object, a gestalt, satisfying properties of vicinity, similarity, continuity of direction, closure, pregnancy, etc., such as shown in Fig.1 from [13]. These gestalt grouping laws were directly related to the geometric statistics of the natural world. Therefore, we take a quite different approach that applies Lipschitz function to model Gestalt visual perceptual laws to find k nearest neighbors. Fig.2 and Fig.3 illustrates the superiority of our novel approach on artificial data and high dimensional image data from olivettifaces database at ATT.

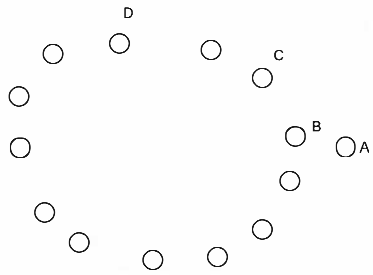

Fig. 1. Illustration of Gestalt visual perceptual laws. We intend to see the circle composing of points such as B, C, and D while to regard A as independent point. In this way, D has more probability than A to be selected as a nearest neighbor of the point B, even if Euclidean distance between B and A is much smaller than that between B and D.Existing approaches have difficulties in finishing the task even if the much complex distance such as geodesic distance is applied.

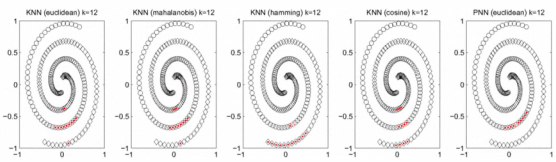

Fig. 2. Finding nearest neighbors by the different approaches. In figure, a query q is indicated by the red circle whereas the found nearest neighbors are marked by the red dot. It can be observed that the selected nearest neighbors by any single measure, shown as the left four figures, are from the different classes, while the found ones by our novel approach completely belong to the same class, satisfying good continuation Gestalt law. This is exactly what we want.

978-1-4244-6439-5/10/$26.00 ©2010 IEEE

Fig. 3. Finding nearest neighbor for the query image q by our novel approach which can find nearest neighbors that are most suitable to human perceptual laws. (A)Fourty face images randomly sampled from olivettifaces database are presented, where the image q will be taken out as the query image and the rest are taken as training samples. (B)Nearest neighbors found by Euclidean distance. It can be observed that only one neighbor is found while the found nearest neighbors seems without any logic order. (C) All images similar to the query q are found from training samples by our novel approach. They are also sorted in the smile way from left to right side.

II. New approach to finding nearest neighbors

It is empirically investigated that Gestalt corresponds to organized structure which stresses the concept of organization and on a whole that is orderly, rule-governed, and nonrandom. This organized structures can keep constant even when the data changes in a certain range due to noise[14]. These stable properties can be described by topological invariance[15] and then applied to perform pattern recognition. It makes sense that in normal visual perception, the extraction of topological properties serves as the starting point of object perception. However in real life, it is hard for us to recognize the topological properties directly from data. A new mathematical model is required. In real world, if two samples are of the same class, a gradual change sequence must exist between them[16]. To make simple reasoning, if two samples are nearest neighbors, they should be of the same class and can be induced from each other by topological transformation. We apply Lipschitz function to model this idea. A function $f : X \to X$ on a metric space (X, d) is called a Lipschitz function if there exists a constant α such that[17]

$$d(f(x), f(y)) \leq \alpha d(x, y), x, y \in X$$

where X is a set of data samples,together with a metric d, that is a non-negative, symmetric function $d : X \times X \to \Re$, which fulfills $d(x, y) = 0, x = y$ and the triangle inequality $d(x, y) + d(y, z) < d(x, z)$, α is the small Lipschitz constant. It can be proved that Lipschitz function is topological transformation. The advantage of the method is that it can be applied without knowledge of data samples since d is treated as a 'black box'. For any query sample x, it is expected that its nearest neighbors should have the same topological properties. Namely, if y is a nearest neighbor of x, it can be transformed topologically from x:

$$y = f(x) = \min(d(x, \overline{X})), x, y \in X$$

where $d(x, \overline{X})$ is all distances between x and any element of $\overline{X}$, and $\overline{X} = X - \{x\}$. Namely Lipschitz function here simply takes the form of nearest neighbor function. In this

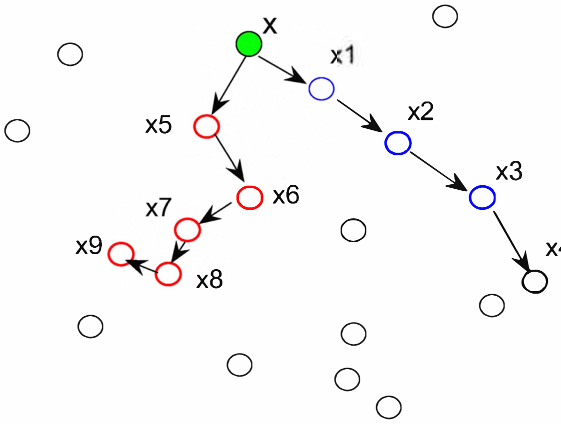

Fig. 4. How our novel approach works. For any query x, many topological sequences such as $S_1 =< x, x1, \cdots, x4 >$ and $S_2 =< x, x5, \cdots, x9 >$ can be constructed by Lipschitz function. Subsequently, a measure is designed to evaluate the quality of these topological sequences in terms of Gestalt laws. The samples in the topological sequence with higher quality are selected with higher superiority as the nearest neighbors for the query x.

way, for any point x_0, many topological sequences from data X can be constructed by Lipschitz function. As we aim to find the k perceptual nearest neighbors, we start by selecting k nearest neighbors from X with Euclidean distance, denoted as $\Theta = \{x_1^1, \cdots, x_1^i, \cdots, x_1^k\}$, then we apply Lipschitz function to construct the i-th topological sequence for each $< x_0^i, x_1^i >$ where $x_1^i \in \Theta$ and $x_0^i = x_0$, denoted as S_i.

$$S_i =< x_0^i, x_1^i, x_2^i, \cdots, x_j^i, \cdots >$$

$$x_j^i = f(x_{j-1}^i), x_j^i, x_{j-1}^i \in X, 0 < j \leq |X|, x_1^i \in \Theta$$

Let $X = X - S_i$, construct another topological sequence from X for x_0 until $X = \emptyset$,shown as Fig.4. In this way X can be represented by union of topological sequences,namely $X = \cup S_i$. Subsequently, we will find the perceptual nearest neighbors through selecting the optimal topological sequences. Each topological sequence has different properties, so that they need some ordering by a measure. The designed measure obviously should be consistent with perceptual organization laws. We designed it as follows based on the distribution of the topological sequence.

$$E(S_i) = \frac{\sigma(S_i^*) \times \mu(S_i^*)}{L(S_i^*) \times i}$$

$$S_i^* = \{r_j | r_j = \frac{d(f(x_j^i), f(x_{j-1}^i)) - d(x_j^i, x_{j-1}^i)}{d(x_j^i, x_{j-1}^i)}, x_j^i, x_{j-1}^i \in S_i\}$$

$$\mu(S_i^*) = \frac{1}{L(S_i^*)} \sum_{r_j \in S_i^*} r_j$$

$$\sigma(S_i^*) = [\frac{1}{L(S_i^*)} \sum_{r_j \in S_i^*} (r_j - \mu(S_i^*))^2]^{\frac{1}{2}}$$

$\mu(S_i^*)$ and $\sigma(S_i^*)$ denote respectively the average value and the variance of the relative change rate of two adjoint distances. The smaller are these two values, the better. $L(S_i)$ represents the length of the topological sequence, indicating the stability of the topological sequence.The longer is it, the better. The factor *i* in $E(S_i)$ is applied to perform the tradeoff between the nearest neighbors determined by Euclidean distance and perceptual nearest neighbors. $E(S_i)$ evaluates the quality of the topological sequence, such as good continuation, uniform, and stability. The smaller is it, the better. Now we can apply it to define the perceptual nearest neighbors by sorting $E(S_i)$ in the descending order and then selecting *k* neighbors from the first sequence to the last sequence until *k* nearest neighbors are selected.

Algorithm 1 **find-pnn**(X, x_0, k, α)

/* This algorithm is to find *k* perceptual nearest neighbors for x_0 from samples data X. α is a Lipschitz constant*/

Step 1. Select *k* nearest neighbors from X with Euclidean distance for x_0, denoted as $\Theta = \{x_1^1, \cdots, x_1^i, \cdots, x_1^k\}$

Step 2. Apply Lipschitz function to construct the topological sequence,denoted as S_i,for each $< x_0^i, x_1^i >$ from X, where $x_0^i = x_0$, $x_1^i \in \Theta$,and $L(S_i) \leq k$.

$$S_i =< x_0^i, x_1^i, x_2^i, \cdots, x_j^i, \cdots >$$

Step 3. Compute the quality of each topological sequence by $E(S_i)$.

Step 4. Sort S_i in descending order in terms of $E(S_i)$ as

$$\{S_{\sigma(1)}, S_{\sigma(2)}, \cdots, S_{\sigma(i)}, \cdots, S_{\sigma(|X|)} | E(S_{\sigma(i)}) \leq E(S_{\sigma(i+1)})\}$$

Step 5. Combine $S_{\sigma(i)}$ as follows and then take *k* first elements from S as *k* nearest neighbors.

$$S = [S_{\sigma(1)} S_{\sigma(2)} \cdots S_{\sigma(i)} \cdots S_{\sigma(|X|)}]$$

This algorithm is designed based on the topological transformation defined by Lipschitz function. The constructed topological sequence is a manifold varying with the application of the Lipschitz constant. The time complexity of this algorithm is $O(k^2 N)$.

III. Designed new classifier

To validate our novel idea of finding *k* nearest neighbors, we apply it to design a new classifier that improves from the local mean classifier (LMC)[10], [9]. LMC is more robust to outliers than classic KNN and has wider applications. Let X^i be a training sample set from class $\omega_i, 1 \leq i \leq n_c$, where n_c is the total number of classes in the whole training samples. The newly designed classifier is depicted as follows.

Algorithm 2 **PNN**(X, x_0, k, α)

Step 1. Select *k* nearest neighbors for x_0 from each class $X^i \subseteq X$ using **find-pnn** algorithm, denoted as $\Theta(\omega_i, k)$

Step 2. Compute the local mean vector for each class ω_i using *k* nearest neighbors as follows:

$$y_i = \frac{1}{k} \sum x \in \Theta(\omega_i, k)$$

Step 3. Classify x_0 into class ω_i if

$$\omega_i = arg \min_{1 \leq i \leq n_c} \{|x_0 - y_i|\}$$

The PNN classifier differs from LMC in that it applies **find_pnn** to find *k* nearest neighbors instead of using a measure. Therefore it increases the time complexity as the same as that of **find_pnn**.

IV. Experimental results

A. Experimental setup

To validate our novel idea and classifier, we compare PNN with KNN and LMC through experiments on benchmark data sets. The accuracy rate is taken as the measure of performance of all compared classifiers, which usually is used as the most effective measure of the performance of a classifier[9], [10]. In experiments, *k* takes the value over the range of $[3, 6, 9, \cdots, 30]$, α takes the values from $[0.1, 0.2, \cdots, 2.0]$, and Euclidean distance is taken in all compared classifiers. When classifying, For each data we performed ten times ten-fold cross validations. On each partition, the parameters are determined for each compared classifier are performed through ten-fold cross validations on the training samples, and then applied to perform the classification over the testing samples. Finally the average of ten times is reported.

B. On artificial data sets

Generally the performance of a classifier is severely influenced by the outliers or noise, particularly in small training sample size situations[9]. To compare PNN with KNN and LMC in these cases, we perform the experiments on two benchmark artificial data sets: two spiral pattern data [18], ring norm data set[19]. Using artificial data,we can easily control the number of the available samples and add noise according to the experimental purpose. It can be observed from Fig.5 (A∼B) that on two spiral pattern data, PNN performs obviously much better than LMC does at any case. The gap between their average accuracies is up to 2.76%. It outperforms KNN at many cases while it loses the superiority on some cases. This is because KNN outperforms LMC on this data set whereas our novel idea is not yet applied to find *k* nearest neighbors for KNN classifier. On ring norm data, as the variance of noise increases, the accuracy rates of PNN and LMC decrease slowly but PNN always outperforms LMC at any noise variance with up to 4.21% of the average accuracy. By the way, KNN performs much worse in all cases on this data. This means that the proposed approach is stronger to resisting in noise disturbance and has wider applications.

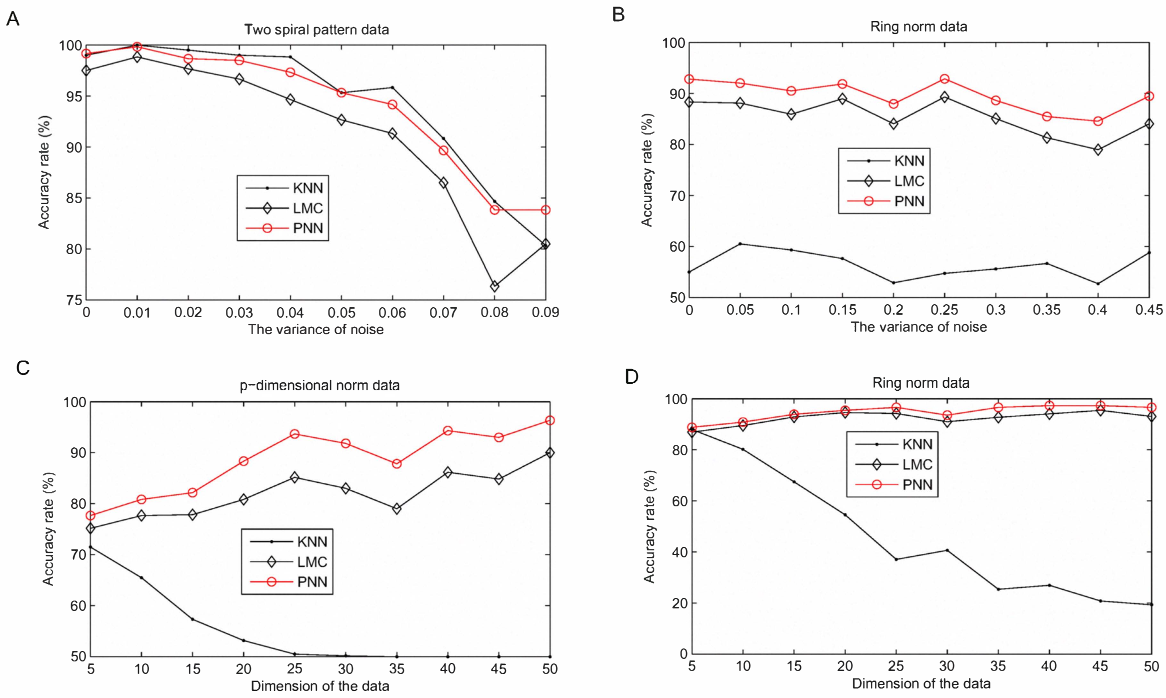

Fig. 5. Evaluation of our novel approach on noisy and high dimensional data. (A∼B) Accuracy rate of the compared classifiers on two spiral pattern and ring norm data sets with 200 points respectively, appended with different random Gaussian noise where the mean of the noise is 0 and the variance is different for different data set, as illustrated as x-coordinate in Figure. (C∼D)Accuracy rate of the compared classifiers on ring norm data and p-dimensional norm data sets with 200 points as a function of the dimensionality from $\{5, 10, \cdots, 50\}$

It is well-known that the curse of the dimensionality is a hard issue for pattern recognition, as in high dimensional data there may be redundant dimensions and exists a high degree of correlation among these dimensions. This may be a severe drawback to clasifiers when dealing with small high dimensional situations where definitely the data are sparsely distributed. To validate the proposed approach with the better ability to deal with these problems, we do experiments on ring norm data set[19] and p-dimensional norm data[10], as they can be generated using different dimensions. It can be observed from Fig.5 (C∼D) that as the dimensionality increases, the accuracy rate of PNN increases quickly than LMC on p-dimensional norm data with the gap up to 6.63%, whereas the accuracy of KNN decreases quickly down to 50% accuracy rate. On ring norm data, the similar trend can be observed. The gap is also up to 2.25%. This suggests that our novel approach may be more robust to the dimensionality and shows a favorable behavior in high dimensions,whereas KNN is the worst performer because its performance is far from the Bayesian performance with the increase of dimensions. To conclude, PNN shows much excellent performance on these artificial data sets that are generated by rules. This is consistent with our idea that PPN works well on data sets having intrinsic structure such as well-formed Gestalts.

C. *On real data sets*

To be practical, we also perform experiments on benchmark real data sets from UCI Repository of machine learning databases[20], where the records with missing values and non-numeric attribute are all removed. Most of them may be noisy, sparse, and imbalanced. It can be observed from Table.I that PNN outperforms the compared classifiers on all data sets. Most of them are obvious, as in which there are stable Gestalt forms. It also can be observed that the standard deviation of the accuracy of PNN is smallest, indicating that its performance is most stable. These results do indicate the significant value of the proposed idea and the classifier.

V. Related Work

On the whole, KNN related classifiers much depend on the selected nearest neighbors. However, no distance function is known to perform consistently well in finding k nearest neighbors, even under some conditions[11]. Particularly, when the training data set is noisy and imbalanced, these approaches are very challenging to find the true neighbors to the query sample, as in such case, the normal samples are easily not representative and the clear geometrical shape of data can not be formed. PNN quite differs from these approaches to finding k nearest neighbors. These existing approaches can not find nearest neighbors organized by Gestalt laws even if the complex measures such as the geodesic distance is applied[6]. PNN is related in spirit partly to the theory

TABLE I
AVERAGE VALUES AND STANDARD DEVIATIONS OF THE ACCURACIES OF KNN, LMC, AND PNN ON REAL DATA SETS

Data	Attributes	Classes	Size	KNN	LMC	PNN
wine	13	3	178	74.62 ± 4.33	75.38 ± 4.13	78.08 ± 5.14
dermatology	34	6	358	88.11 ± 1.73	93.40 ± 1.41	93.77 ± 1.42
diabetes	8	2	768	75.70 ± 2.13	75.83 ± 1.25	76.96 ± 0.74
ionosphere	34	2	351	84.90 ± 2.72	90.96 ± 3.24	92.31 ± 2.56
glass	9	7	214	68.85 ± 4.44	71.97 ± 5.60	74.43 ± 4.65
optdigits	64	10	1797	98.88 ± 0.54	99.21 ± 0.29	99.33 ± 0.29
segmentation	19	7	210	82.06 ± 5.24	83.97 ± 3.47	85.24 ± 3.43
yeast	8	10	1484	59.14 ± 1.54	58.84 ± 1.94	59.46 ± 1.78
yaleface	4135	15	165	61.78 ± 4.78	64.22 ± 4.25	64.89 ± 3.89
iris	4	3	150	96.67 ± 2.40	96.67 ± 1.57	97.33 ± 1.41
			avg	**79.07 ± 2.98**	**81.04 ± 2.71**	**82.18 ± 2.53**

of topological perception[21], [22], which accounts for the amazing ability of human by defining the global properties as topological invariants.This is consistent with Gestalt theory of perception[13]and topological psychology[23]. They discovered the human perception laws whereas concrete mathematical models oriented to solve the problems in information field such as classification are not presented. In computer vision, some computational theories are developed to find automatically gestalts in digital images[12]. They used specific properties of images,instead of taking each image as independent objects for classification. Particularly,they used very little and almost nothing of the Gestalt theory results[12]. This is why little recent work can be available from the literature. Another novel work is the proposal of principle of homology continuity[16]. It applies the relation of training samples in the same class to design artificial neural network. This approach is important but not for finding nearest neighbors, while it does not aim to model Gestalt laws so that for example the topological transformation and the measure on topological transformation are not presented.

VI. CONCLUSION AND FUTURE WORK

This paper proposes a quite different but much interesting approach to finding k nearest neighbors by modeling Gestalt laws of psychology. To the best of our knowledge, PNN is the first one to find nearest neighbors by modeling perceptual Gestalt laws based on topological transformation. It also provides an efficient means to measure the degree of constructed topological sequence consistent with Gestalt laws. The most importance of our work seems that it opens a new direction as a fundamental methodology to develop a quantitative Gestalt theory which is then applied to design all kinds of classifiers. PNN is likely to be even more useful in combination with other methods in data analysis, statistical learning, and information retrieval. Given the broad appeal of traditional approaches, the proposed idea and methodology should find widespread use in more areas of science and engineering due to its simplicity, general applicability, and excellent performance on the noisy,sparse and high dimensional problems.

ACKNOWLEDGMENT

This work was supported by China National Science Foundation under Grants 60973083. Fundamental Research Funds for the Central Universities,SCUT. Guangdong Science and Technology project under Grants 2007B030803006.

REFERENCES

[1] H.Samet,K-Nearest Neighbor Finding Using MaxNearestDist,IEEE Trans. Pattern Anal. Mach. Intell.30,243(2008)
[2] J.B. Tenenbaum and V. de Silva and J.C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction,Science 290,2319(2000)
[3] M.H Tristan and R.Stephane,Tailored Aggregation for Classification,IEEE Trans. Pattern Anal. Mach. Intell.31,2098(2009)
[4] B. J. Frey and D. Dueck, Clustering by Passing Messages Between Data Points,Science ,Science 315,972(2007)
[5] V.Athitsos,J.Alon,S.Sclaroff,and G.Kollios,BoostMap: An Embedding Method for Efficient Nearest Neighbor Retrieval,IEEE Trans. Pattern Anal. Mach. Intell.30,89(2008)
[6] G.Wen,L.Jiang, and J.Wen, Using Locally Estimated Geodesic Distance to Optimize Neighborhood Graph for Isometric Data Embedding,Pattern Recognition 41,2226(2008)
[7] M. Balasubramanian and E.L. Schwartz, The Isomap Algorithm and Topological Stability, Science 295,7(2002)
[8] S. N.Karl and Q. N.Truong,An Adaptable k-Nearest Neighbors Algorithm for MMSE Image Interpolation,IEEE Trans. Image Processing 18, 1976(2009)
[9] Y.Mitani and Y. Hamamoto, A local mean-based nonparametric classifier, Pattern Recognition Letters 27, 1151(2006)
[10] B.Li, Y.W.Chen, and Y.Q.Chen, The Nearest Neighbor Algorithm of Local Probability Centers, IEEE Trans. Syst., Man, Cybern.,2008,38(1): 141-153
[11] H.Wang, Nearest neighbors by neighborhood counting,IEEE Trans. Pattern Anal. Mach. Intell.28, 942(2006)
[12] A.Desolneux, L.Moisan, and J.Morel,Computational gestalts and perception thresholds,Journal of Physiology - Paris 97, 311(2003)
[13] K.Koffka , Principles of Gestalt Psychology, Routledge, London,1955
[14] S.He,Holes, objects, and the left hemisphere,PNAS 105,1103(2008)
[15] L.Chen, S.Zhang, and V.S.Mandyam,Global perception in small brains: Topological pattern recognition in honey,PNAS 100,6884(2003)
[16] S.Wang and X.Chen,Biomimetic (Topological) Pattern Recognition-A new Model of Pattern Recognition Theory and Its Application,Proceedings of the international joint conference on neural network,2258(2003)
[17] U.Luxburg and O.Bousquet,Distance-based Classification with Lipschitz Functions, journal of machine learning research,2004, (5):669-695
[18] S. Singh, 2D spiral pattern recognition with possibilistic measure,Pattern Recognition Lett. 19,131(1998)
[19] L. Breiman,Arcing classifiers, Ann. Statist.26, 801(1998)
[20] A. Asuncion and D.J.Newman, UCI Machine Learning Repository [http://www.ics.uci.edu/ mlearn /MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science.(2007).
[21] L.Chen,Science 218,699(1982)
[22] L.Chen,Visual Cognit 12:553(2005)
[23] K.Lewin, Principles of topological pyschology,(McGraw-Hill Book Company,Inc.,1936)