Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **4**, July **2016**

Quasiconformal-Matrix-Based Multikernels Learning for Sensory Data Classification

Xiaodan Xie*
School of Automation Science and Electrical Engineering
Beihang University, Beijing, China
xiexiaodanbeijing@126.com
*Corresponding author

Bohu Li
School of Automation Science and Electrical Engineering
Beihang University, Beijing, China
bohulibeijing @126.com

Xudong Chai
Beijing Simulation Center, Beijing, China
xudongchaibeijing @126.com

Received December, 2015; revised March, 2016

Abstract. *Sensory data classification is a crucial task in many information processing. The vector data or matrix data are the most common data format, especially matrix data, for example, image data. We propose Gaussian kernel constructing method adaptive to the distribution of the input data for classification. In the current matrix based feature extraction, the matrix image is transformed to the vector, and the procedure of transforming will increase the ability of a large saving space and computing. And secondly the different geometrical structures under the different kernel function will bring the different class discrimination of the input data in the feature space. The performance of kernel will be increased under the adaptive choosing the parameters of kernel. We present the matrix Gaussian kernel, Quasiconformal matrix Gaussian kernel, and Quasiconformal matrix multikernels, and we implement the three kinds of kernels on different classifiers on ORL and Yale image databases. The experimental results show that the proposed kernels perform better than the traditional kernel functions.*
Keywords: Matrix data, Gaussian matrix kernel, Quasiconformal matrix Gaussian kernel, Image classification.

1. **Introduction.** Kernel-based image classification systems are widely studied in many areas [1], and kernel learning is to improve the linear learning methods. The data distribution in the space-based nonlinear feature space is easy to classification, and the data geometrical structure is determined by the kernel function. The discriminative ability of the data could be even worse if an inappropriate kernel is used. In the previous work [2, 3], researchers optimized the parameters of kernel function, but these methods are not effective through only choosing the parameter from a set of discrete values. The geometry structure of data distribution in the kernel space is not be changed only through the changing the parameters of kernel. Xiong proposed a data-depend kernel for kernel optimization [4], and Amari presented support vector machine classifier through modifying the kernel function [5]. Moreover, multiple kernel learning methods are developed, for

example, Sparse Multiple Kernel Learning [6], Large Scale Multiple Kernel Learning[7], lp-Norm Multiple Kernel Learning[8]. And kernel learning method is applied to hyperspectral remote sensing imagery classification [1], and the feasibility and excellent performance were also reported in the relative works[9, 10]. Sparse kernel learning [11, 12], Multiple sparse kernel coding [13, 14, 15], and some other methods were proposed based on kernel learning method, such as fuzzy kernel learning[17], Semisupervised learning[17], online learning [18], and these methods are widely used in object tracking[19, 20], machine vision[21] and other applications areas. Moreover, multiple kernel learning methods are developed, for example, Sparse Multiple Kernel Learning [22], Large Scale Multiple Kernel Learning [23], lp-Norm Multiple Kernel Learning[24].In order to improve the performance of kernel-based classification system, many methods of optimizing the kernel parameters of the kernel function are developed in recent years (Huang [25], Wang [26] and Chen [27]). However, choosing the parameters for kernel just from a set of discrete values of the parameters doesnt change the geometrical structures of the data in the kernel mapping space. In order to overcome the limitation of the conventional KDA, we introduce a novel kernel named quasiconformal kernel which were widely studied in the previous work [28, 29, 30, 31], where the geometrical structure of data in the feature space is changeable with the different parameters of the quasiconformal kernel. The optimal parameters are computed through optimizing an objective function designed with the criterion of maximizing the class discrimination of the data in the kernel mapping space.The currently kernel functions used in kernel learning method are vector-based function. All images must be transformed to vectors, and these vectors must be saved for the kernel-based learning system. And in the thus image classification system, the original image and transformed vector must be saved for the kernel learning. The input of the kernel function is vector or an [N×1] matrix. For an [m×n] matrix, the matrix must be transformed to a vector of [M×1], where $M = m \times n$. Thus, the one image matrix and vector must be saved, that is, for a [m×n] pixels of image, the save space is m×n×2 of pixels for the traditional kernel learning, but only m×n of pixels of saving space for the matrix-kernel learning. So, for the proposed matrix-kernel learning will save 50% of memory resources.

In this paper, we proposed a novel Quasiconformal-Matrix-Based Multikernels Learning for sensory data classification. In this model, firstly the data is to compute the kernel matrix directly using the matrix data but without transforming the matrix data to vector data, and secondly the parameter of quasiconformal kernel function is adjusted through solving the eigenvalue equation, which is adaptive to the input data distribution. Based on the fact that, the inappropriate selection of kernels will decrease the performance because the geometrical structures of the data in the feature space will not be changeable, our proposed quasiconformal matrix-kernel function differs from the traditional kernel function, and the structure of data distribution will be changed through changing the parameter of quasiconformal kernel. The parameters are solved through solving the constrained optimization equation, and the optimization equation is designed by the maximizing the class discriminative ability. So, the performance of kernel-based classification learning is improved including recognition accuracy and efficiency.

2. Algorithm.

2.1. Motivation.

As the discussion, among these machine learning methods, kernel learning is a feasible and effective nonlinear feature extraction methods. In the practical application, the data features of the specific tasks are generally chosen, for example, for object recognition, image detection, pyramids and HOG feature selection. Polynomial kernel function, Gaussian kernel function, RBF kernel function should be used in many areas. In fact, there are a number of kernel functions, and no optimal kernel function can be adapted to all the applications. For example, in the kernel-based image processing, each kind of kernel function can only be used to describe the characteristics of the data effectively, such as texture, color, edge, etc.. The researchers began to have a significant interest in the multi kernel function of the combination of basic nuclear functions. The multiple kernel learning has the different kernel representation for the different feature subspace. Multiple kernel learning is an effective method to solve the above problems. Multiple kernel learning method can choose the kernel function according to the different characteristics of the data, and then combine them. Multiple kernel learning is a feature extraction method of combining many features, and is better to describe the data features than the single feature extraction method. The multiple kernels learning can not only improve the efficiency of feature extraction, but also adjust the weights of the adaptive selection and the basic kernel function. Multiple kernel learning not only preserves the nonlinear mapping characteristics of kernel functions, but also shows the possibility of using different kernel functions and also shows the possibility of a unified framework for the classical method of nonlinear feature extraction.

Kernel trick is an effective method to solve the nonlinear problems, and system performances including such as recognition accuracy, prediction accuracy are largely increased by the nonlinear kernel mapping. The performance of kernel-based system is largely influenced by the function and parameter of kernel. Optimizing the parameters is not effective to promote the kernel-based learning system owing to the unchanged data structure with the changing of the parameter of kernel function. No a universal single kernel is very effective way to detecting intrinsic information for the complicate sample data in the input data space. Quasiconformal mapping-based kernel is to change the data structure in the feature space through adjusting the parameter of quasiconformal function. In order to improve the performance of kernel learning, we combine the idea of the quasiconformal kernel and multiple kernel combination. So, in this paper we present a framework of quasiconformal mapping-based multiple kernels learning. The learning system is improved adaptively the data structure in the kernel mapping from the fact that quasiconformal multiple kernels are combined to more precisely characterize the data for improving performance on solving complex visual learning tasks. We use the Gaussian kernel as the basic kernel in the proposed quasiconformal mapping-based multiple kernels.

2.2. Definitions.

2.2.1. *Quasiconformal matrix kernel.* Firstly, we proposed the matrix Gaussian kernel for image computing. Supposed that n samples $I_p^q(I_p^q \in \mathbb{R}^{m\times n}), (p = 1, 2, ..., L, q = 1, 2, ..., n_p$ where n_p denotes the number of samples in the pth class and L denote the number of the classes. Matrix Gaussian kernel $K_g(I_x, I_y)$ is defined as,

$$K_g(I_x, I_y) = exp\left(-\frac{\sum_{j=1}^{n}(\sum_{m}^{i=1}(I_{ij}^x - I_{ij}^y)^2)}{2\sigma^2}\right)(\sigma > 0) \tag{1}$$

where I_x, I_y are the different images, and the following work is to proof of $K_g(I_x, I_y)$ is a kernel function. Kernel function can be defined in various ways. In most cases, however, kernel means a function whose value only depends on a distance between the input data, which may be vectors. As the kernel function, a symmetric function is a sufficient and necessary condition that its Gram matrix is positive semi-definite. Given a finite data set $I = \{I_1, I_2, ..., I_N\}$ in the input space and a function $k(.,.)$, the $n \times n$ matrix K with elements $K_{ij} = k(I_i, I_j)$ is called Gram matrix of $k(.,.)$ with respect to $I_1, I_2, ..., I_N$. $K_g(I_x, I_y)$ is a symmetric function. The kernel matrix is computed with kernel function

$K_g(I_x, I_y)$, is positive definite. So, the proposed matrix $K_g(I_x, I_y)$ is the kernel function. Gaussian kernel denotes the distribution of similarity between two vectors. Similarly matrix Gaussian kernel also denotes the distribution of similarity between two matrices. Matrix Gaussian kernel views an image as a matrix, which enhances the computation efficiency without influencing the performance of kernel-based method.

Secondly, the quasiconformal matrix kernel function $K_{qg}(I_x, I_y)$ can be denoted as

$$K_{qg}(I_x, I_y) = f(I_x)f(I_y)K_g(I_x, I_y) \tag{2}$$

where $f(I)$ is a positive real valued function I.

$$f(I) = b_0 + \sum_{k=1}^{N_{XM}} b_k exp\left(-\delta \sum_{j=1}^{n}\left(\sum_{i=1}^{m}\left(I_{ij} - \tilde{I_{ij}}\right)^2\right)^{1/2}\right) \tag{3}$$

where $\tilde{I_{ij}}$ is the elements of matrix $\tilde{I}(n = 1, 2, ..., N_{XM})$, and δ is a free parameter, and $\tilde{I_{ij}}$, $1 \le n \le N_{XM}$, are called the expansion matrices (XMs) in this paper, $N_{XM}$is the number of XMs, and $b_i \in R$ is the expansion coefficient associated with $\tilde{I_{ij}}$.

2.2.2. *Quasiconformal matrix multi-kernels.* We extend the quasiconformal kernel to quasiconformal multi-kernels. Different from the single quasiconformal kernel in (9), only the expension paramters need to computed by constrain equation. While on the quasiconformal multi-kernels, the weight parameter and expansion coefficient need to be computed through the equation. The quasiconformal multi-kernels version has the higher ability on describing the data distribution than the quasiconformal kernel.

$$K_{qmk}(I_x, I_y) = q(I_x)\sum_{i=1}^{N_{XM}} d_i k_{g,i}(I_x, I_y)q(I_y) \tag{4}$$

where $k_{g,i}(I_x, I_y)$ is the ith matrix kernel, N_{XM} is the number of candidate basic kernels for combination, $d_i \ge 0$ is the weight for the ith basic kernel, $q(.)$ is the factor function given by

$$f(I) = a_0 + \sum_{k=1}^{N_{XM}} a_k exp\left(-\delta \sum_{j=1}^{n}\left(\sum_{i=1}^{m}\left(I_{ij} - \tilde{I_{ij}}\right)^2\right)^{1/2}\right) \tag{5}$$

The definition of the quasiconformal kernel shows that the geometrical structure of the data in the kernel mapping space is determined by the expansion coefficients with the determinative XMs and the free parameter.

The procedure of the algorithm is shown as follows.

Step 1. Optimize the weights of multiple kernels

On multiple kernels learning, the crucial step is to choose adaptively the weights of multiple kernels through the comprehensive consideration on the computation efficiency and classification accuracy on hyperspectral image classification.

Step 2. Optimize the coefficients of Quasiconformal kernels

This step is to optimized the coefficients of the quasiconformal kernels based on the classification criterion, for example, Fisher criterion. The method is similar to the single kernel based quansiconformal kernel method [2]. On the multiple classes of classification problems, we solve the optimal coefficients in the in the empirical feature space. The detailed information about the kernel optimization can be referred to [2].

3. **Experimental results.**

3.1. **Experiment setup.** In the practical applications, the kernel function is adaptively chosen subject to the training samples set. The training sample set is constructed by the training images, and the kernel function can be adaptively chosen by kernel machine. In the experiments, we choose the procedural parameters through cross-validation method for the practical application. Its not estimate the algorithms parameter using the entire dataset, but only training dataset. In the algorithm evaluation, the dataset is divided into training and test dataset. The cross-validation method is only implemented in the training dataset. We applied the k-fold cross-validation method. In K-fold cross-validation, the original image sample dataset is randomly partitioned into K subsamples dataset. Of the K subsamples dataset, a single subsample is retained as the validation data for testing the model, and the remaining K-1 subsamples are used as training data. The cross-validation process is then repeated K times (the folds), with each of the K subsamples used exactly once as the validation data. The K results from the folds then can be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. 10-fold cross-validation is used in the experiment. For the detail description, the training dataset is randomly divided to training sub-dataset and test sub-dataset, and the parameters is to train the classifier with training sub-dataset then the test sub-dataset is to test the performance of the parameters. In the practical applications, we choose it with expert experience for some applications. In the most applications, two kinds of kernel functions are applied, including Polynomial kernel and Gaussian kernel. But it depends on the expert experience to choose the kernel. Kernel selection is usually done by minimizing either an estimate of generalization error or some other related performance measure. This paper focuses the improving of Gaussian kernel for image-based system. But in the practical application, how to select the kernels is proposed in the previous work [32]. It establishes the connection between the spectral perturbation stability and the generalization error through minimizing the derived generalization error bound, and then the kernel selection criterion is to guarantee good generalization properties. In the criterion, the perturbation of the eigenvalues of the kernel matrix is efficiently computed by solving the derivative of a newly defined generalized kernel matrix. And the detailed parameter will be solved through optimizing the constrained equations.

3.2. **Results on ORL and Yale databases.** In this section, we implement Kernel Principal Component on ORL and Yale databases to evaluate the performance of Gaussian kernel, matrix Gaussian kernel, Quansiconformal matrix Gaussian kernel. ORL database, developed at the Olivetti Research Laboratory, Cambridge, U.K., is composed of 400 grayscale images with 10 images for each of 40 individuals. The variations of the images are across pose, time and facial expression. The Yale database was constructed at the Yale Center for Computational Vision and Control. It contains 165 grayscale images of 15 individuals. These images are taken under different lighting condition (left-light, center-light, and right-light), and different facial expression (normal, happy, sad, sleepy, surprised, and wink), and with/without glasses. In our experiments, to reduce computation complexity, we resize the original ORL images sized 112×92 pixels with a 256 gray scale to $48{\times}48$ pixels. We randomly select 5 images from each subject, 200 images in total for training, and the rest 200 images are used to test the performance. Similarly, the images from Yale databases are cropped to the size of 100×100 pixels. Randomly selected 5 images per person are selected as the training samples, while the rest 5 images per person are used to test the performance. In the experiments on ORL and Yale databases, the number of iterations is 150 for optimal parameter section. We have the detailed analysis and results on the performance of multiple kernels version of Gassian matrix-based kernel. The results are shown in Table.1 and Table.2. As experimental results on the performance of quasiconformal single kernel leanring and quasiconformal multiple kernels learning on ORL and Yale image databases, we can conclude that, the quasiconformal single kernel learning performs better than basic kernel, and quasiconformal multiple kernels performance than other methods. Quasiconformal single kernel structure changes the data structure in the kernel empirical space. And then, quasiconformal multiple kernels are combined to more precisely characterize the data for improving performance on solving complex visual learning tasks, so the proposed framework outperforms others in the different datasets. The parameter beta is the final optimal parameter through solving the optimization equation. The efficiency will be influenced owing to the iteration solution method, but we can optimize it in the off-line. So in the practical application, the efficiency will be balanced considered compared with the recognition accuracy.

Table 1. Performance on ORL database.

Feature dimension	20	40	60	80	100	120	140
Gaussian kernel	0.61	0.66	0.72	0.77	0.78	0.80	0.82
Matrix Gaussian kernel	0.70	0.75	0.78	0.79	0.79	0.81	0.83
Quasiconformal matrix kernel	0.73	0.77	0.79	0.80	0.81	0.83	0.84
Quasiconformal-Matrix-Based-Multikernels-Learning	0.75	0.79	0.81	0.82	0.83	0.85	0.87

Table 2. Performance on Yale database.

Feature dimension	20	40	60	80	100	120	140
Gaussian kernel	0.715	0.755	0.775	0.780	0.785	0.790	0.795
Matrix Gaussian kernel	0.717	0.758	0.778	0.785	0.789	0.795	0.797
Quasiconformal matrix kernel	0.720	0.760	0.780	0.790	0.793	0.805	0.815
Quasiconformal-Matrix-Based-Multikernels-Learning	0.731	0.770	0.790	0.811	0.815	0.823	0.831

4. **Conclusions.** We present the matrix Gaussian kernel, Quasiconformal matrix Gaussian kernel, and Quasiconformal matrix multi-kernels, and we implement the three kinds of kernels on different classifiers on the ORL and Yale image databases. The experimental results show that the proposed kernels perform better than the traditional kernel functions. The proposed Gaussian kernels views images as matrices, which saves the storage and increase the computational efficiency of feature extraction. The proposed kernels can be used in other areas, such as content-based image indexing and retrieval as well as video and audio classification. In this work, we only consider the Gaussian kernel for the kernel-based learning machine. Other kinds of kernel functions based on matrix will be studied in the future work. Our work only pays attention to the classification problem based on kernel learning. So the experiments show the classification performance, and the criterion of kernel optimization is created by increasing the classification performance. So, the kernel optimization criterion is not adaptive to clustering. The clustering application based kernel optimization is our future research work. On the computation efficiency, the iteration solution method will cost much time.

REFERENCES

[1] C. Chen,W. Li, H. Su, et al., Spectral-Spatial Classification of Hyperspectral Image Based on Kernel Extreme Learning Machine, *Journal of Remote Sensing*, vol. 6, no. 6, pp. 5795-5814, 2014.

[2] J. S. Pan, J. B. Li, Z. M. Lu, Adaptive quasiconformal kernel discriminant analysis, *Journal of Neuro-computing*, vol. 71, no. 13-15, pp. 2754-2760, 2008.

[3] W.S. Chen, P. C. Yuen , H. Jian, et al., Kernel machine-based one-parameter regularized fisher discriminant method for face recognition, *Journal of IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society*, vol. 35, no. 4, pp. 659-669, 2005.

[4] X. Huilin, M. N. Swamy, M. Omair A, Optimizing the kernel in the empirical feature space, *Journal of Neural Networks IEEE Transactions on*, vol. 81, no.3, pp. 260-276, 2005.

[5] S. Amari, S. Wu, Improving support vector machine classifiers by modifying kernel functions, *Journal of Neural Networks*, vol. 12, no. 6, pp. 783-789 1999.

[6] N. Subrahmanya, Y. C. Shin, Sparse Multiple Kernel Learning for Signal Processing Applications, *Journal of IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 32, 2010.

[7] S. Sonnenburg, Rätsch G, Schäfer C, et al., Large Scale Multiple Kernel Learning, *Journal of Machine Learning Research*, vol. 7, pp. 1531-1565, 2006.

[8] M. Kloft, U. Brefeld, et al., l p -Norm Multiple Kernel Learning, *Journal of Machine Learning Research*,vol. 12, np. 12, pp. 953-997, 2011.

[9] Y. R. Yeh , T. C. Lin, Y. Y. Chung, et al., A Novel Multiple Kernel Learning Framework for Heterogeneous Feature Fusion and Variable Selection,*Journal of IEEE Transactions on Multimedia*, vol. 14, no. 3, pp.563-574, 2012.

[10] Y. Haiqin, X. Zenglin, Y. Jieping, et al., Efficient sparse generalized multiple kernel learning, *Journal Neural Networks IEEE Transactions on*, vol. 22, no. 3, pp. 433-46, 2011.

[11] J. C. Nascimento, J. G. Silva, J. S. Marques, et al., Manifold learning for object tracking with multiple nonlinear models, *Journal of IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*,vol. 23, no. 4, pp. 1593-1605, 2014.

[12] S. Xue, O. Abdel-Hamid, H. Jiang, et al., Fast Adaptation of Deep Neural Network Based on Discriminant Codes for Speech Recognition, *Journal IEEE/ACM Transactions on Audio Speech & Language Processing*,vol. 22, 2014.

[13] J. J. Thiagarajan, R. Karthikeyan Natesan, S. Andreas, Multiple kernel sparse representations for supervised and unsupervised learning, *.Journal of IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 23, no. 7, pp. 2905-2915, 2014.

[14] L. Wang, H. Yan, K. Lv, et al., Visual Tracking Via Kernel Sparse Representation With Multikernel Fusion, *Journal Circuits & Systems for Video Technology IEEE Transactions on*, vol. 24, no. 7, pp. 1132-1141, 2014.

[15] S. Ashish, V. M. Patel, C. Rama, Multiple kernel learning for sparse representation-based classification, *Journal of IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 23, no. 7, pp. 3013-3024, 2014.

[16] P. Chiranjeevi, S. Sengupta, Detection of moving objects using multi-channel kernel fuzzy correlogram based background subtraction, *Journal IEEE Transactions on Cybernetics*, vol. 44, no. 6, pp. 870-881, 2014.

[17] M. Kan, D. Xu, S. Shan, et al., Semisupervised Hashing via Kernel Hyperplane Learning for Scalable Image Search, *Journal IEEE Transactions on Circuits & Systems for Video Technology*, vol. 24, no. 24, pp. 704-713, 2014.

[18] X. Hao, S. C. H. Hoi, J. Rong, et al., Online Multiple Kernel Similarity Learning for Visual Search, *Journal Pattern Analysis & Machine Intelligence IEEE Transactions on*, vol. 36, no. 3, pp. 536-549, 2014.

[19] D. H. Kim, H. K. Kim, S. J. Lee, et al., Kernel-Based Structural Binary Pattern Tracking, *Journal IEEE Transactions on Circuits & Systems for Video Technology*, vol. 24, no. 08, pp. 1288-1300, 2014.

[20] G. Yu, H. Rangwala, C. Domeniconi, et al., Predicting Protein Function Using Multiple Kernels, *Journal IEEE/ACM Transactions on Computational Biology & Bioinformatics*, vol. 12, no. 1, pp. 219-233, 2015.

[21] Z. Liu, S. Xu, Y. Zhang, et al., A multiple-feature and multiple-kernel scene segmentation algorithm for humanoid robot, *Journal IEEE Transactions on Cybernetics*, vol. 44, no. 11, pp. 2232-2241, 2014.

[22] N. Subrahmanya, Y. C. Shin, Sparse Multiple Kernel Learning for Signal Processing Applications, *Journal IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 32, 2010.

[23] S. Sonnenburg, G. Rätsch, C. Schäfer C, et al., Large Scale Multiple Kernel Learning *Journal of Machine Learning Research*, 2010, 7(2006), pp. 1531-1565.

[24] M. Kloft, U. Brefeld, S. Sonnenburg, et al. l p -Norm Multiple Kernel LearningJournal Journal of Machine Learning Research, 2011, 12(12), pp. 953-997.

[25] J. Huang, O. C. Yuen, ?. S. Chen , et al. Kernel Subspace LDA with Optimized Kernel Parameters on Face Recognition[C]// Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on. IEEE, 2004, pp. 327-332.

[26] W. Lei C. Kap Luk, Ping X. A criterion for optimizing kernel parameters in KBDA for image retrieval, *Journal of IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society*,vol. 35, no. 3, pp. 556-562, 2005.

[27] W. S Cchen, P. C. Yuen, H. Jian, et al., Kernel machine-based one-parameter regularized fisher discriminant method for face recognition, *Journal IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society*, vol. 35, no. 4, pp. 659-669, 2005.

[28] S. Amari, S. Wu, Improving support vector machine classifiers by modifying kernel functions, *Journal Neural Networks*, vol. 12, no. 6, pp. 783-789, 1999.

[29] G. R. Lanckriet, N. Christianini, P. Bartlett, et al., Learning the Kernel Matrix with Semi-Definite Programming, *Journal of Machine Learning Research*, vol. 5, no. 1, pp. 323-330, 2002.

[30] X. Huilin, M. N. Swamy, M. Omair A, Optimizing the kernel in the empirical feature space, *Journal of Neural Networks IEEE Transactions on*, vol. 81, no. 3, pp. 260–276, 2005.

[31] J. Peng, D. R. Heisterkamp, H. K. Dai, Adaptive quasiconformal kernel nearest neighbor classification, *Journal of IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 26, no. 5, pp. 656-661, 2004.

[32] L. Yong, S. Liao, Kernel selection with spectral perturbation stability of kernel matrix, *Journal Science China*, vol. 57, no. 11, pp. 1-10, 2014.