

# Research and Implementation of a License Plate Recognition Algorithm Based on Hierarchical Classification

Haofei Gao<sup>1</sup>, Wanrong Sun<sup>1</sup>, Xinran Liu<sup>2</sup>, Ming Han<sup>2</sup>

<sup>1</sup>School of Electronic Engineering, Xidian University, Xi'an, China; <sup>2</sup>School of Life Sciences and Technology, Xidian University, Xi'an, China.

Email: [gaohaofei8906@gmail.com](mailto:gaohaofei8906@gmail.com)

Received October 2013

## ABSTRACT

This paper proposed an improved method for license plate recognition based on hierarchical classification. First, the method of feature extraction and dimension reduction is presented by finding the optimal wavelet packet basis in the process of wavelet packet decomposition and K-L transform. Then the recognition algorithm is introduced based on feature extraction and hierarchical classification. Finally, the principles and procedures of using support vector machines, Harris corner detection algorithm and digital character classification are explained in detail. Simulation results indicate that the presented recognition algorithm performs well with higher speed and efficiency in recognition.

## KEYWORDS

Wavelet Packet; K-L Transform; SVM; Recognition; License Plate

## 1. Introduction

License plate recognition (LPR) system is an important component of the intelligent transportation systems in the modern society. Being more and more popular, LPR is the hot issue in the area of image processing, computer vision and pattern recognition. The steps of License plate recognition include image acquisition, image digital processing, license plate location and extraction, character segmentation and character recognition. Character recognition algorithm is the significant part.

So far, there have been many methods of LPR, such as template matching, feature matching, artificial neural networks, support vector machines (SVM), et al. In the so many techniques, SVM is favored. SVM takes advantages of generalization and simple structure. SVM also has the outstanding effect in solving issues of small sample, nonlinear and high-dimension space [1]. However, due to the limitations of low decision-making speed and classification accuracy, the conventional SVM method has a worse efficiency. In addition, the more feature dimensions, the more difficult to complete the recognition. Therefore, academics have made improvement continuously in speed and accuracy to optimize recognition performance. However, these methods which based on

the more complicated mathematical and structural model lead to increasing complexity. Literatures [2] and [3] use hidden Markov to do recognition, but the probability model needs lots of prior knowledge. Literature [4] presents a tree of SVM decision based on prior knowledge with one certain character being used frequently. The system is relatively limited, and the recognition efficiency is not satisfied when classifiers are too much in the case of one-to-one.

Based on the above considerations, this paper proposes an improved license plate recognition algorithm based on hierarchical classification. Through the rational application of wavelet packet decomposition and SVM, the algorithm has rational analysis and allocation for the layers of recognition decision. Test results show that the improved algorithm based on hierarchical classification takes advantages of high speed and accuracy, and performs well with high efficiency and robustness.

## 2. Feature Extraction

Distributed characteristics of characters' structure in the four spaces of wavelet decomposition can mainly reflect the structure characteristics of Chinese, numbers and letters, which can be used as basis for license plate cha-

racter recognition. In **Figure 1(a)**, only the low-frequency part of the signal will be decomposed in the conventional orthogonal wavelet decomposition. When all the wavelet coefficients are used as features, the decomposition has lots of dimensions and low flexibility. **Figure 1(b)** shows that wavelet packet decomposition can make continuous decomposition in the wavelet spaces of high-frequency part of signal so as to obtain abundant information in time-frequency domain. First, in order to overcome the shortcomings of wavelet transform, this paper uses wavelet packet decomposition which has better characteristics in time-frequency domain to finish feature extraction. Then find the optimal wavelet packet basis in the process of wavelet packet decomposition in order to reduce the dimension of feature vectors. At last, the coefficients of the optimal basis will be reduced in dimension by K-L transform. So feature is extracted.

### 2.1. Principle of Wavelet Packet Transform

From the perspective of multi-resolution analysis, wavelet transform decomposes space  $L^2(\mathbb{R})$  into direct sum of all the subspaces  $W_j (j \in \mathbb{Z})$  based on different scale factor  $j$ , that is  $L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j$ . Given wavelet function  $\psi(t)$  and orthogonal scaling function  $\phi(t)$ , there are following recurrence relations.

$$\begin{cases} \psi_{2l}(x) = \sum_{k=-\infty}^{+\infty} h_k \psi_l(2x-k) \\ \psi_{2l+1}(x) = \sum_{k=-\infty}^{+\infty} g_k \psi_l(2x-k) \end{cases}, \quad (1)$$

Where  $h_k$  and  $g_k$  are filter coefficients, and is defined as wavelet packet on the basis of orthogonal scaling function  $\phi(t)$ . Equation (1) shows that the wavelet decomposition is a special situation of wavelet packet decomposition.

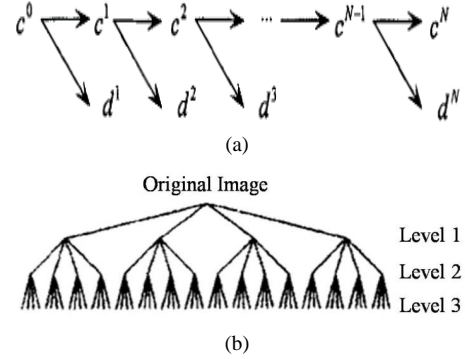
### 2.2. Find the Optimal Wavelet Packet Basis

Based on a certain standard, select the optimal wavelet packet basis from the collection of wavelet packet basis. The wavelet packet decomposition coefficients of the optimal basis will be used as identification feature. The essential aspect is how to find the optimal wavelet packet basis.

According to the literature [5], there follows the basic criterion. The mean vectors of two categories  $\omega_i$  and  $\omega_j$  are  $m_i$  and  $m_j$  with the within-class scatter  $S_\omega$ .

$$\begin{cases} J_F^{(i,j)} = \frac{\|m_i - m_j\|^2}{\text{tr}(S_\omega^{-1})} \\ J_F = \sum_{i=1}^{c-1} \sum_{j=i+1}^c \sqrt{\frac{J_F^{(i,j)}}{T_{\max} - T_{\min}}} \end{cases}, \quad (2)$$

where  $J_F$  is defined as the criterion value, and  $[T_{\min}, T_{\max}]$



**Figure 1. Wavelet decomposition and wavelet packet decomposition. (a) Structure diagram of wavelet decomposition; (b) Structure diagram of wavelet packet decomposition in 3 levels.**

is the range of  $J_F^{(i,j)}$ . Character is decomposed by wavelet packet basis into levels of  $N$ .  $U_l^k$  ( $l = 1, \dots, N; k = 1, \dots, 2^N$ ) means the  $k$ -th node in the layer  $l$ . The child node of  $U_l^k$  is  $U_l^k(n)$ , while the child node of  $U_l^k(n)$  is  $U_l^k(n)(m)$  ( $n, m = 1, 2, 3, 4$ ).  $\alpha_l^k(n)$  is defined as the ratio of  $U_l^k(n)$  to  $U_l^k$ .  $\alpha_l^k(n) > 1$  indicates that the child nodes are more useful to classification, otherwise the parent nodes are more useful.  $\alpha = \sum (\alpha_l^k(n) - 1)$  is total incremental value of  $J_F$ . Nodes meets the conditions of decomposition when  $\alpha > 0$  and nodes of  $\alpha_l^k(n) > 1$  are not less than 2. Combined the parameters with conditions, the algorithm of feature extraction is following [6].

1) The image of license plate character is normalized to size of  $32 \times 16$ .

2) The character image decomposed by wavelet packet decomposition in there levels and find  $J_F$  of each node.

3) Find the optimal wavelet packet basis  $Z$ . If the node  $U_l^k$  does not meet the conditions of decomposition, the child node  $U_l^k(n)$  will be verified whether the node meets the conditions. When node  $U_l^k(n)$  does not meet the conditions,  $U_l^k$  will join in the optimal wavelet packet basis  $Z$ . If  $U_l^k(n)$  meets the conditions and the nodes of  $\alpha_l^k(n) > 1$  are not less than 2,  $U_l^k(n)(m)$  will join in  $Z$ . Meanwhile, if  $U_l^k$  meets the conditions at the beginning, all  $U_l^k(n)$  will join in  $Z$  and delete  $U_l^k$  from  $Z$ .

4) Repeat step 3 until each node of  $Z$  is checked. Then the optimal wavelet packet basis can be obtained with their coefficients being recognized characters.

### 2.3. Use K-L Transform to Get Eigenvectors

The dimension of coefficients of the optimal wavelet packet basis is large, which is a disadvantage of improving recognition rate and speed. It is indispensable to reduce dimension of feature vectors, and optimize recognition effect through reducing redundant information. K-L transform means the method for analysis of principal

component. The purpose is to remove the correlation between the data, as well as transform and project the data from the original  $R$ -dimension space to the  $M$ -dimension space ( $R \gg M$ ) with the minimum distortion based on the criterion of mean square error [7].

Assume that  $X = \{x_1, x_2, \dots, x_M\}$  is a sample set (where  $x_i \in R^N$ ),  $M$  is the total number of training samples, and  $N$  is the dimension of each training sample. There is a matrix based on total scatter matrix of the training sample set  $X$  in Equation (3):

$$E = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(x_i - \bar{X})^T = \frac{1}{N} AA^T, \quad (3)$$

where  $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i, (i=1, 2, \dots, n)$ . A correlation matrix  $R$

can be expressed with  $R = AA^T$  on the basis of generated matrix  $A$ . Eigen values of  $R$  which can be expressed with  $\lambda_i (1 \leq i \leq N)$  are solved and ranged, that is  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ . The corresponding eigenvectors in the sequence are  $v_i (i=1, 2, \dots, N)$ . Taking the first  $m$  maximum, a matrix  $U$  of K-L transform concludes the corresponding  $m$  eigenvectors, that is  $U = (v_1, v_2, \dots, v_m)$ . The value of  $m$  can be determined through the quotient of sum of the largest eigen values to sum of total eigen values, that is  $\theta = \sum_{i=1}^m \lambda_i / \sum_{i=1}^N \lambda_i$ . Take  $\theta$  as large as possible, while  $m$  as small as possible. The situation can keep image information to the greatest extent with minimum dimension of the feature vectors. According to the simulation and experiment, the variation of  $\theta$  with  $m$  is obtained in Figure 2. When the value of  $m$  is about 30, the feature vectors in low-dimension space can reflect the situation of the characteristics in the original high-dimension space. So this paper selects  $m$  to be 30 [8].

Actually, feature extraction is a process of extracting the decomposed coefficients of the optimal wavelet packet basis through wavelet packet decomposition in three levels of the identified character. In the extraction, not only does the high-frequency component information increase to improve the recognition rate, but also K-L transform reduce the dimension to improve recognition rate, so as to optimize recognition effects.

### 3. Recognition Algorithm Based on Hierarchical Classification

Figure 3 shows the recognition algorithm based on hierarchical classification which mainly includes feature extraction and character recognition. In the process of feature extraction, the pretreated characters decomposed by wavelet packet decomposition, and find the optimal wavelet packet basis. Then the wavelet coefficients of the optimal wavelet packet basis are reduced in dimension by K-L transform. So there are the eigenvectors to be

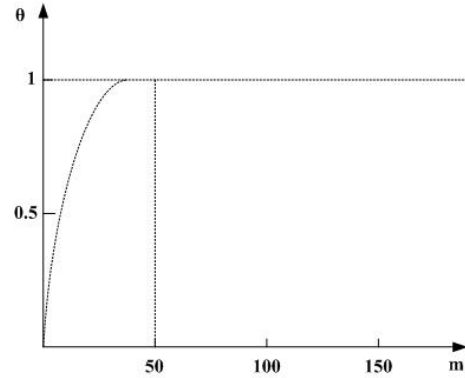


Figure 2. Variation of  $\theta$  with  $m$ .

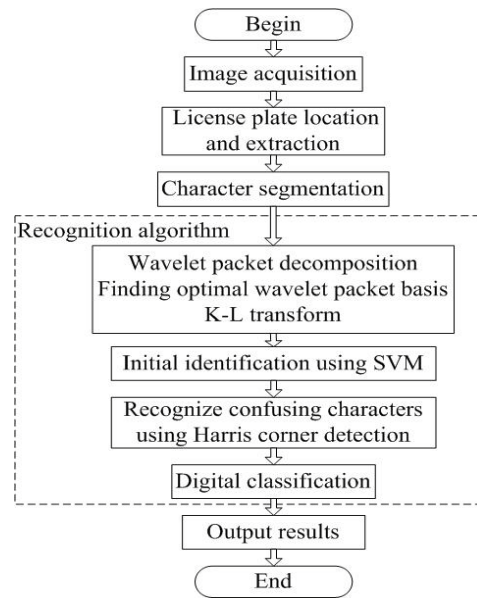


Figure 3. License plate recognition algorithm based on hierarchical classification.

recognized. As to the design of character recognition classifier, support vector machines can be used as the first identification. Next, confusing pictographic characters are secondarily classified taking advantage of Harris corner detection algorithm. Ultimately, all the digits and letters are classified according to the feature of arrangement of the license plate characters. The improved PLR algorithm based on hierarchical classification can effectively improve the accuracy and recognized speed, and the effect is impressive to confusing characters.

#### 3.1. Initial Recognition Using SVM

As the youngest member of statistical theory, SVM takes advantage of the principle of minimum structural risk, performs a particular merit in solving problem of small sample, nonlinear and high-dimension space and exhibits a tremendous favorable prospect [9]. To linear problems, the kernel function is a dot product of two vectors. As to

nonlinear problems, SVM defines nonlinear mappings which will project input vectors from a lower-dimension space to a higher one. Then an optimal hyperplane is constructed in this high-dimension space. Not only does the optimal hyperplane separate the two types properly, but also the interval between types reaches to the maximum. As the important concept of SVM, maximize the interval is to control the capability of generalization [10]. Actually, the structural problem of the optimal hyperplane is to solve a quadratic optimization problem with a specific constraint. The optimal decision function is presented in Equation (4) [11].

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^L y_i \cdot a_i \cdot K(x, x') + b \right\}, \quad (4)$$

where  $\text{sgn}(\bullet)$  is a sign function,  $L$  is the number of training samples, and  $a_i \geq 0$  is defined as the Lagrange multiplier and  $K(\bullet, \bullet)$  is a kernel function.

### 3.1.1. Choose a Kernel Function

Generally, kernel functions are defined as

$K(x, x') = \varphi(x) \bullet \varphi(x')$ , where  $x$  and  $x'$  are vectors in low-dimension space, while  $\varphi(x)$  and  $\varphi(x')$  are the transformed vectors. One form called radial basis kernel function is forceful in locality which is similar to the characteristics of human visual [12]. Taking into the factors, the form of radial basis function (the other name is RBF or Gauss kernel function) is chosen to be trained. RBF can be shown as:

$$K(x, x') = e^{-\gamma \|x - x'\|^2}. \quad (5)$$

### 3.1.2. Train and Recognize Vectors Using LIBSVM

LIBSVM is an integrated software developed by Professor Chih-Jen Lin from Taiwan. LIBSVM is used in various fields such as pattern recognition, regression analysis, and estimation of probability distribution [13]. The application consists mainly of interface functions such as svm-scale, svm-train, svm-predict and so on. Training and recognition based on LIBSVM involves the following steps [14].

1) Training samples of the character are extracted and converted to the required format. In order to facilitate the calculation and prevent one feature to be too large or too small, the function of svm-scale can scale data to the proper range, generally, [0,1] or [1,1].

2) Radial basis function is used as the kernel function.

3) Use cross-validation method to train test samples repeatedly and find the optimal parameters. The so-called cross-validation is to divide data into  $k$  sets. When one set participates in training, the other  $k-1$  sets take the test to obtain  $k$  parameters ( $C$  and  $\zeta$ ). Compared with the other results, the one set of parameters which have the best effect is elected as the optimal parameters.

4) Use the optimal parameters and RBF to train the training sets to get SVM model.

5) Do license plate characters recognition taking advantage of the above-mentioned model.

## 3.2. Recognize Confusing Characters Using Harris corner Detection

Due to the similar outline and interference by substantive reasons, confusing characters such as {2, Z}, {0, D, Q, C, U}, {8, B, R}, {5, 8, S}, {4, A} and {C, G} will affect the recognition accuracy of the final results. So a further identify to similar characters is a pivotal process which affects the recognition results. Because of scale invariance and rotation which can inhibit light and noise impacts, Harris corner detection algorithm is a good choice to have a further recognition to similar characters.

Harris corner detection is presented on the basis of Moravec operator by Harris. C. G and Stephen. M. J in 1988. Using a first-order partial derivatives to describe the gray change, Harris corner detection has a matrix  $M$  associated with autocorrelation function.

$$M = \begin{bmatrix} \left( \frac{\partial I}{\partial X} \right)^2 & \left( \frac{\partial I}{\partial X} \right) \left( \frac{\partial I}{\partial Y} \right) \\ \left( \frac{\partial I}{\partial X} \right) \left( \frac{\partial I}{\partial Y} \right) & \left( \frac{\partial I}{\partial Y} \right)^2 \end{bmatrix}, \quad (6)$$

where  $I(x, y)$  is the brightness value expressed by gray scale, and  $\partial I / \partial X$  is the gradient in the  $X$  direction of the image  $I(x, y)$ , while  $\partial I / \partial Y$  is the gradient in the  $Y$  direction. The two eigenvalues of the matrix  $M$  which are first-order curvatures of autocorrelation can be used to judge the different areas in the image. If both are large, the point is regarded as a corner; if one of the eigenvalues is large and another one is small, the point is located at the edge; if both are small, it is determined in a flat region [15]. On the basis of the matrix  $M$ , corner response function (CRF) is defined as:

$$CRF = \det(M) - k \bullet \text{trace}^2(M), \quad (7)$$

where  $k$  sets 0.04 which is optimal parameter [16],  $\det$  is the determinant of  $M$ , being small at the edges and large at the corner points, and  $\text{trace}$  is the trace of  $M$ , being large at the corner points. Therefore, the local maximum of CRF is the corner point.

Simulation results show that Harris corner detection on similar characters recognition has excellent effect. For instance, D, U and 0 of approximate outline can be seen in **Figure 4**. The number and position of Harris corner clearly distinguished between each other. Experiments on plates containing confusing characters found that the Harris corner detection can be used to resolute similar characters with good results, and the recognition rate can be improved effectively.





Figure 4. Harris corner diagram of D, U and 0.

### 3.3. Digital Classification Algorithm

The plate contains Chinese characters, numbers and letters in China. Character samples can be classified into Chinese network, digital network, alphabet network and alphanumeric network. According to the “Vehicle license of the people’s Republic of China” revised in 2007, the last five characters in the seven ones are called serial numbers, which have three kinds of encoding rules: a) each of serial numbers can use Arabic numerals; b) each of serial numbers can use letters alone, but O or I cannot be used from the 26 letters; c) it is allowed that two letters can be existed in serial numbers, and O or I cannot be used from the 26 letters.

As there are no more than two letters, it can be identified by setting the number of detected letters. The algorithm of digital classification is presented in Figure 5.

In the recognition of the last five characters, if there are two letters been detected in the serial numbers, following characters which are prepared to be identified can be input the digital network, instead of alphanumeric network. For instance, the vehicle license shown in Figure 6 can be recognized with the digital character classification algorithm. A large number of experiments indicate that digital character classification algorithm can improve the recognition speed; because the samples are restricted to some degree, the recognition accuracy is improved as well.

## 4. Experimental Results and Achievement of the System

In this experiment, the vehicle images are 24 true-color real-shots. License plate characters contain digital samples of 10, letter samples of 24, and Chinese character samples of 55. In training samples, there are letter samples of 120, digital samples of 32 and Chinese character samples of 166. The recognition applies LIBSVM library using C-SVC as model and RBF as kernel function. Cross-validation K sets 5, and penalty parameter C sets 8. The recognition results and comparison are proposed in Tables 1 and 2. It can be noticed that the proposed algorithm has advantages of high recognition rate and speed over other identification methods such as artificial neural networks and traditional support vector machine. The system has an outstanding performance.

Table 1. Recognition experiment results.

Character type	Total	Correct	Correct rate (%)
Chinese character	160	152	95.00
Letter	375	361	96.36
Digital	591	587	99.15
Total	1126	1105	98.15

Table 2. Comparison of the presented method with other methods.

LPR method	Recognition rate	Average time (s)
The presented method	1105/1126(98%)	0.71
Traditional SVM	418/431(97%)	0.88
Neural Network	357/376(95%)	0.80

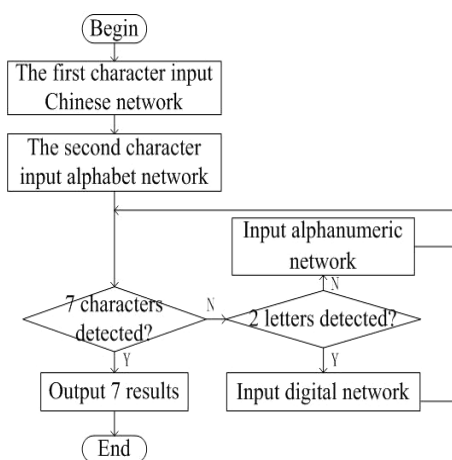


Figure 5. Algorithm of digital classification.

License plate recognition based on hierarchical classification is simulated by Matlab 7.12 (R2011a), designed by Visual C++ 6.0 on C language, proceeded by MFC to build software platform and transplanted to the license plate recognition system with ARM9-S3C3440A as the master chip. The implemental processes of the system are demonstrated in Figure 6.

## 5. Conclusion

This paper proposed an advanced license plate recognition system based on an improved hierarchical classification algorithm on the basis of wavelet packet decomposition and support vector machines. As the significant part of recognition, the improved algorithm provides an efficient and easy-going way. This paper presented a multi-level model based on classification which is different from other recognition algorithms. In feature extraction, find optimal wavelet packet basis in the process of wavelet packet decomposition to obtain eigenvectors. Then the dimension of eigenvectors is reduced by K-L transform to optimize the performance. In recognition, the characters are identified by SVM. Harris corner detection



(a)



(b)

**Figure 6. Implement of the license plate recognition system based on hierarchical classification. (a) Designed by MFC; (b) PLR system on ARM9.**

is used to correct confusing characters. Finally, the method is optimized in recognition speed by the algorithm of digital classification decision. Experimental results indicate that, compared with systems based on other recognition algorithms, the performance of this presented PLR system is developed effectively and the system enhances a better stability, robustness and efficiency.

## REFERENCES

- [1] V. N. Vapnik, "The Nature of Statistical Learning Theory," Springer, New York, 1995.  
<http://dx.doi.org/10.1007/978-1-4757-2440-0>
- [2] J. T. Chien and C. P. Liao, "Maximum Confidence Hidden Markov Modeling for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 8, 2008, pp. 606-616.  
<http://dx.doi.org/10.1109/TPAMI.2007.70715>
- [3] G. B. Singh and H. Song, "Using Hidden Markov Models in Vehicular Crash Detection," *IEEE Transactions on Vehicular Technology*, Vol. 58, No. 3, 2009, pp. 1119-1128. <http://dx.doi.org/10.1109/TVT.2008.928904>
- [4] W. J. Wei, H. F. Kan and Z. L. Zhu, "The Image Classification Using a Neural Network by Zernike Moment," *IEEE National Aerospace and Electronics Conference*, 1992, pp. 254-259.
- [5] S. Sunny and P. S. David, "Feature Extraction Methods Based on Linear Predictive Coding and Wavelet Packet Decomposition for Recognizing Spoken Words in Malayalam," *2012 International Conference on Advances in Computing and Communications (ICACC)*, Cochin, Kerala, 2012, pp. 27-30.  
<http://dx.doi.org/10.1109/ICACC.2012.7>
- [6] T. Stutz and A. Uhl, "Efficient and Rate-Distortion Optimal Wavelet Packet Basis Selection in JPEG2000," *IEEE Transactions on Multimedia*, Vol. 14, No. 2, 2012, pp. 264-277.  
<http://dx.doi.org/10.1109/TMM.2011.2177644>
- [7] J. F. Al-Asad, A. Moghadamjoo and Y. Leslie, "Ultrasound Image Denoising through Karhunen-Loeve (K-L) Transform with Overlapping Segments," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'09)*, 2009, pp. 318-321.
- [8] D. Xie and J. Q. Zhu, "Research on Method of Main Reducer Assembly Quality Evaluation Based on K-L Transform and Support Vector Machine," *2011 International Conference on Electric Information and Control Engineering (ICEICE)*, 2011, pp. 677-680.
- [9] Q. Zhao, J. L. Cao and Y. L. Hu, "Joint Optimization of Feature Selection and Parameters for Multi-class SVM in Skin Symptomatic Recognition," *International Conference on Artificial Intelligence and Computational Intelligence*, 2009, pp. 407-411.
- [10] Wang, S. P. Patrick, "Pattern Recognition, Machine Intelligence and Biometrics," Higher Education Press, Beijing, 2011.
- [11] J. A. K. Suykens, "Least Squares Support Vector Machines," World Scientific, River Edge, 2002.
- [12] N. Cristianini and J. S. Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods," Publishing House of Electronics Industry, Beijing, 2004.
- [13] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines".  
<http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [14] C. H. Lin, J. C. Liu and C. H. Ho, "Anomaly Detection Using LibSVM Training Tools," *International Conference on Information Security and Assurance*, 2008, pp. 166-171.
- [15] Z. Y. Ye, Y. J. Pei and J. H. Shi, "An Improved Algorithm for Harris Corner Detection," *2nd International Congress on Image and Signal Processing*, 2009, pp. 1-4.
- [16] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proceedings of the 4th Alvey Conference*, Manchester, England, 1998, pp. 147-151.