

A Comparison of Methods for Character Recognition of Car Number Plates

Lihong Zheng¹, Xiangjian He¹ and Yuanzong Li²

¹Department of Computer Systems
University of Technology, Sydney
PO Box 123, Broadway 2007, Australia

²College of Mechanical Engineering
Taiyuan University of Technology
Taiyuan, Shanxi, 030024, China

Abstract

Number plate is the unique identification of a vehicle. Automatic Number Plate Recognition (ANPR) is a system to capture the number plate of a moving vehicle and then to recognize it. Real-time recognition of number plate of moving vehicles is critical in numerous applications such as surveillance of moving vehicles, tracking and stopping vehicles associated with known suspects. It can be used not only for security and traffic management purposes but also for safety and general information gathering applications. Aiming at general ANPR's objectives, namely fast processing speed, high accuracy and low cost, many researches focus on recognizing the characters on number plate image quickly and accurately. The goal of ANPR research is to develop algorithms and methods that can recognize number plate accurately and quickly at any one time, day or night, in almost any weather conditions. In this paper, the general process of image recognition system is extracted. This process includes preprocessing, feature extraction, classification, and optimization. Following this processing chain, each stage is discussed in detail and a few of well-known approaches used are represented. After analysis of the merits and pitfalls of existing systems, a combined recognition system is proposed for ANPR system. Using several efficient classifiers such as Template Matching and Support Vector Machines (SVMs), the combined system can improve the processing ability not only in accuracy but also in speed.

Keywords: Automatic number plate recognition, template matching, Support Vector Machines

1 Introduction

A number plate is the unique identification of a vehicle. Automatic Number Plate Recognition (ANPR) is designed to locate and recognize the number plate of a moving vehicle automatically.

Over last decade, there have been many studies on automatic number plate recognition. Real-time recognition of the number plate of a moving vehicle has become more and more important in many applications such as traffic control and vehicle identification. In ANPR, the main parts are moving vehicle detection, number plate location and character recognition. Since last two decades, there have been various commercial ANPR products around the world such as SeeCar in Israel [1], VECON in Hongkong [2], LPR in USA [3], the ANPR in UK [4], IMPS in Singapore [5], and the CARINA in Hungary [6]. Even though there are so many successful applications, there are still several problems for character recognition of number plates. Following three issues must be considered. Firstly, the recognition system must be able to handle various sizes, fonts, spaces and alignments of the characters in the number plates. Secondly, the recognition system must be robust to changes in illumination and colours used. Thirdly, the recognition system must be able to distinguish the obscured characters in real-life images due to rust, mud, peeling paint, and fading colour.

In order to resolve the problems above, an effective method and algorithms are needed urgently. It must classify and recognize the characters in number plate accurately and credibly. It will have a general adaptability for different conditions. It should have good tolerance for noise.

This paper includes three main sections. In Section 2, we review and compare the techniques for character recognition including character classification techniques. We further discuss the various classification techniques and make our conclusions in Section 3.

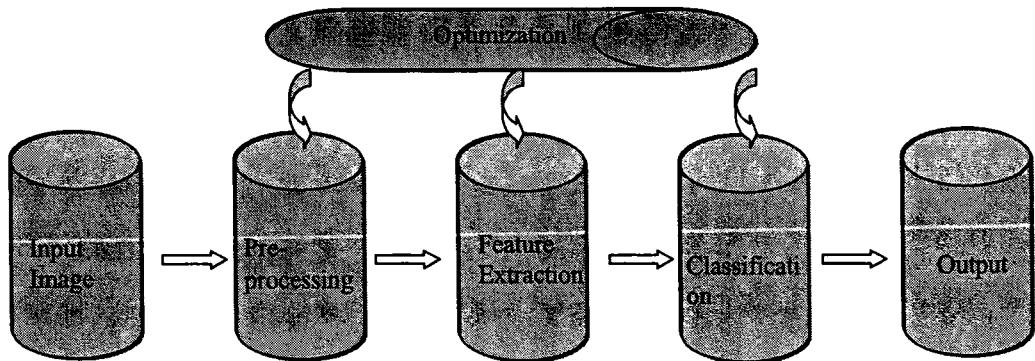


Fig. 1 Character Recognition Process

2 Existing Character Recognition Techniques

As a component of an ANPR system, a successful character recognition on a number plate is to correctly assign each character on the number plate to a letter from A to Z or a digit from 0 to 9. In the following, we will introduce the basic process of a character recognition system and review some well-known techniques for character recognition.

The basic process of character recognition consists of following four steps in general. There are *preprocessing*, *feature extraction*, *classification*, and *optimization* as shown in Figure 1.

2.1 Preprocessing

Preprocessing is necessary for any accurate image recognition. It is designed to reduce noise and produce a good quality set of image data for recognition. Good quality refers to several aspects such as brightness, contrast and less noise points. In this stage, noise filtering, binarization, normalization, and segmentation are the four basic operations.

Noise filtering

Due to practical environment, a lot of physical phenomena can induce noise, such as shadows, wet or dirty number plate, wear and tear, and so forth. Noise may inundate the useful image information and it may hence lead to an error in character recognition. Hence, noise filtering becomes important for the success of character recognition.

Binarization

As a preprocessing step to extract interested region, a synthesized image with a wide dynamic range should be binarized. Thresholding is a

simple method for binarization. Since some information will somehow be lost during binarization, an adaptive thresholding technique is employed by Chang [7]. There is a tradeoff between the size of mask window and image quality. A large window may fail to binarize characters and a small window may generate many pieces of noise fragments. Chang uses a local optimal threshold value for each image pixel so as to avoid the problem originating from non-uniform illumination.

Normalization

Besides noise effect introduced by system disturbances and digital processing, many indeterminate factors may introduce warping. The frame of a number plate may be tilted due to the fixed position of camera. Tilting correction helps reduce the variation of characters represented on a number plate. Liu [8] proposed a method to amend the angle for tilting correction. Based on projection-density histograms, the baseline of number plate is found first and then the tilted angle is calculated.

Segmentation

Image segmentation is to distinguish objects from background. It is the process of dividing the image into some constituted parts so that each constituted part can be extracted for further processing. The segmentation is usually based on discontinuity and similarity of the gray level within the image. Four popular approaches are: thresholding methods, edge-based methods, region-based methods, and connectivity-preserving relaxation methods [9].

Thresholding is the earliest technique for image segmentation. The segmentation is performed by splitting an area that has grey levels in the same range from the other part of the image. Thresholding technique makes decisions based on local pixel information. It is effective when the grey levels of the interested object fall outside the

range of grey levels in the background. Edge-based technique relies on contour detection. The need to connect together broken contour lines makes this technique prone to failure in the presence of blurring. A region-based technique divides an image into smaller parts and merges the neighboring areas that have the same features. Adjacent areas are merged according to some criteria such as homogeneity or sharpness of region boundaries. A connectivity-preserving relaxation-based technique usually refers to as the active contour model. The main idea is to start with some initial boundaries represented in the form of spline curves, and iteratively modify them by applying various shrinking or expansion operations. Time consuming and information loss are its main pitfalls.

The above-mentioned four operations for preprocessing are not necessarily to be applied for all applications. For some applications, some of these operations may be eliminated.

2.2 Feature extraction

In this subsection, we firstly talk about what the features are and how many kinds of features there are. Then, some techniques used for feature extraction are introduced and the merits and pitfalls of these techniques are given.

Features

A feature is defined as a representation of an object in an image. A feature is a characteristic that can distinguish the objects of different types. A good set of features enables the recognition system to discriminate different classes of objects effectively, to reduce feature redundancy, and to be robust to noise and deformation. The selection of features strongly influences the classification results. Features should also be easily computed, and robust and insensitive to various distortions and variations in the images. Generally speaking, features can be categorized into physical features and mapping features. The physical features, such as geometric (or structural) and statistical features, have clear physical meanings. Some examples of structural or geometric features are points, lines, area, perimeter, centroid, length and width, rectangularity, aspect ratio, circularity, stroke and so forth. Statistical feature is based on the intensities and intensity variations of pixels. Five statistical features, which are mean, variance, average deviation, skew and kurtosis, are used commonly [10]. The mapping features have no physical meanings. An example of mapping features is the coefficients of a Fourier Transform or Wavelet Transform.

An advantage of physical features is that no redundant features are included for character classification. Only an optimal set of features are collected. An advantage of the mapping features is that they make classification easier because clear boundaries will be obtained between object classes. But, they increase the computational complexity.

Methods for feature extraction

In many real world problems, features are scattered onto various nonlinear subspaces. The larger a set of features is, the more training samples are needed. Hence, a small but the most powerful set of features reduces the time for recognition. *Feature extraction* is to find an "optimal" set of features which are invariant under certain transformation. Feature extraction is to reduce the computation complexity, to reduce the feature dimension (i.e., the number of features needed), and to achieve high recognition rate. Selected features are expected efficient for distinguishing various classes. Supervised feature extraction methods are common in order to extract invariant features and reduce the amount of features. Supervised approaches require prior knowledge. There are two kinds of supervised methods, which are *linear feature extraction* and *nonlinear feature extraction*.

Linear feature extraction techniques include Principal Component Analysis (PCA) [11], Linear Discriminant Analysis (LDA) [12], and Independent Component Analysis (ICA) [13]. Nonlinear feature extraction techniques include Kernel PCA (KPCA)[14], Multi-Dimensional Scaling (MDS) [15], and Artificial Neural Network (ANN) such as nonlinear auto-associative network and Self-Organizing Map (SOM) [16].

PCA calculates the eigenvectors of the covariance matrix of the original inputs and transforms the original input vector into a lower dimensional one of which the components are uncorrelated. PCA is not effective for analyzing a complex problem. PCA mixture model overcomes the limitation of PCA by using a combination of PCAs to improve the performance. LDA tries to find a linear transform to map such that a suitable criterion of class separability is maximized. The transformation is obtained via maximizing the ratio of between-group to the within-group in any particular data set. Although LDA-based algorithms outperform PCA-based algorithms in many applications, traditional LDA algorithms cannot provide reliable and robust solutions since their separability criteria are not directly related to their classification accuracy in the output space

and often result in misclassification. ICA attempts to achieve statistically independent components in the transformed vectors. ICA is more appropriate for non-Gaussian distributions since they do not rely on the values of the second-order derivatives of image brightness function.

KPCA and MDS are both applied mapping methods to transform the complicated feature space into a simpler feature space. Kernel PCA first maps input data into some new feature space F typically via a nonlinear function Φ and then performs a linear PCA in the mapped space. MDS aims to project a multi-dimensional dataset onto a two or three dimensional space such that the distance matrix in the original d -dimensional feature space is preserved as faithfully as possible in the projected space. But MDS does not give an explicit mapping function. So, when a new pattern is added, a new mapping needs to be sought. ANNs can be used directly for feature extraction in an unsupervised mode. The defects of ANNs are that the training speed is low and the classification accuracy may be lost due to overfitting problem.

2.3 Classification

After optimal feature subset is selected, a classifier can be designed for character recognition. This process is called Optical Character Recognition (OCR). Roughly, there are three different approaches [15]. The first approach is the simplest and the most intuitive approach which is based on the concept of similarity. Template matching is an example of this approach. The second one is a probabilistic approach. It includes methods based on Bayes Decision Rule, the Maximum Likelihood or the Density Estimator. Three well-known methods for this approach are K-Nearest Neighbour (KNN) [17], Parzen Window Classifier [18] and Branch-and-Bound methods (BnB) [19]. The third approach is to construct decision boundaries directly by optimizing certain error criterion. Examples are Fisher's Linear Discriminant, Multilayer Perceptrons – an ANN-based method [20-22], Decision Tree [23] and Support Vector Machine (SVM) [24]. Three most commonly used classification techniques for ANPR are: Template Matching methods, ANN-based methods and SVM-based methods. In the following, these three typical classifiers are introduced in detail.

Template matching

Template matching is the simplest method for image recognition. In this method, a template for

the pattern to be recognized is obtained and stored firstly. Then the pattern from the unknown candidates is recognized based on a similarity measure. For example, distance based method or correlation method are often used to find the most similar candidate to the template. It determines the similarity between a candidate and the template. As an example, the similarity can be computed as the least square error between the candidate and the template

Template matching has some disadvantages. It is time consuming. The more classes of objects there are, the more templates must be stored. Furthermore, the template matching may fail when the input image is distorted due to image processing, viewpoint change, and so forth. It is difficult to find a good match because many indeterminate factors can create warping. Template matching may also fail if there are changes in the size of objects being identified, changes in contrast and changes in the brightness [25]. Template matching does not work well either if the image containing candidates is broken or defiled by noise heavily [26-28].

In order to suppress the disadvantages of template matching method, several modified methods have been proposed. Yamaguchi [25] proposed a method by comparing the correlation values of images. He calculates the correlation values between candidate image and different template images, and compared the differences of these correlation values. The creditable matching result was decided as the one with bigger values. This method is not affected by the magnitude of the correlation value. It can sufficiently accommodate variances in the correlation values created by fluctuations in the environment.

Feature matching [29] is another method based on template matching. It is obvious that all pixels forming a character are not equally important to recognize the character. So, we need not use all the pixels on the objects to form the object templates. Feature matching uses only some critical pixels on the objects to represent the objects. This reduces storage space and the computational time can hence be reduced.

Artificial Neural Networks

Since 1990s, ANNs have increasingly been used as classic pattern classifiers in digital processing. An ANN consists of a number of neurons in various levels, which are an input level, one or more hidden levels, and an output level. The mapping from the inputs to the outputs is nonlinear, and one needs not know the exact inner process between the input level and the output level. ANNs are used to learn from samples. The

learning process can run in parallel. ANNs can perform well even when the image data set contains noise or is incomplete.

Many kinds of ANNs have been used for character recognition. Chang [7] applied the Kohonen Self-organizing neural network to recognize the number on a license plate. Although its overall rate of success is 93.7%, this method is very time consuming and is not suitable for real-time application. Fahmy [30] proposed the Bidirectional Associative Memories (BAM) neural network for number plate recognition. Fahmy's method is only appropriate to a small range of number patterns. Nijhuis et al [31] combined the Fuzzy Theory [32] with a neural network for number recognition of Dutch license plates. This method uses the Fuzzy Theory for segmentation and a cellular neural network for feature extraction and recognition. Coetze [21] used Binarization Linear Transform (BLT) to reduce the input dimension of a neural network. In his work, six small parallel and smaller neural networks were created and each network is used to recognize six characters.

Although ANN has had some success for the number recognition, it requires a huge amount of training samples to adjust weights between neurons in every level, and it cannot guarantee to obtain the global extremum of the risk function (squared training error) but local extrema. Furthermore, ANN is time consuming during its learning process.

Support Vector Machine

As a statistical approach and an approach of ANNs, SVM [14] has gained an increasing attention in the last decade. It works well even when the dimension of input space is very high. SVM represents an $m \times n$ image using $m \times n$ 2-dimensional vectors. Each vector consists of two components, of which the first component is related to the grey value of corresponding pixel and the second component has value of either -1 or +1. In order to separate any two object classes, a subset of these vectors is found through a training process. Each vector in this subset is called a support vector for the classification of the two classes. SVM constructs an optimal hyperplane so as to maximize the "margin" between two classes.

The SVM embodies the Structural Risk Minimization principle employed by most of neural networks for object recognition [14]. With the help of kernel function, $K(x, x_i) = \Phi(x) \cdot \Phi(x_i)$, unseparable classes can be transformed into a high-dimensional feature space. Then, one can do a linear separation there. So, a SVM corresponds

to a linear method in a very high dimensional feature space that is not linearly related to the input space. Classification is achieved by realizing a linear or non-linear separation surface in the input space. It can be depicted from the figures above. Aiming at reducing the computational complexity and obtaining the higher accuracy, various kinds of SVMs have been proposed in recent years, such as SSVM (Smooth SVM), ASVM (Active SVM), LS_SVM (Least square SVM), L_SVM (Lagrangian SVM) and R_SVM (Reduced SVM) [33]. The main advantages of using SVMs for object recognition are listed as follows. SVMs classify data without knowing the exact distribution function about the underlying process. Thus they have good generalization ability. SVMs use less parameters than ANNs, and the classification error has nothing to do with the dimension of the feature space. Furthermore, training SVM consists of a quadratic programming problem that can be solved efficiently and for which we are guaranteed to find global extremum. Applying similar structure of ANNs, SVMs take less training time due to only small training set (support vectors) is used.

2.4 Optimization

Optimization means to best meet the criteria set for character recognition. The most commonly used criterion for evaluation of classification algorithms is accuracy. Another criterion is the stability. An algorithm is regarded as 'stable' if it always produces a similar result for various sets of input data. Other criteria are the cost (e.g. the cost of acquiring data or the cost of classification errors) and the complexity (e.g. the computational complexity of the algorithm or the syntactic complexity of the induced rules).

Optimization is not regarded as a separate step in our image processing chain but as a set of auxiliary techniques, which support all steps in the chain. In the preprocessing step (see Fig. 1), optimization guarantees that each input pattern has the best quality. In the step for feature extraction, optimal feature subsets are obtained using some optimization techniques such as ranking, correlation criteria or relevance [37]. In the classification step, after all the characters have been recognized, a syntax checking method is used to check whether the classification results match the prior knowledge on the structure of a number plate, such as the number of letters, the number of digits, the order of characters and so forth. Any mismatched results must be re-

processed. This optimization process appeared in the classification step reduces the possibility of misclassification.

3 Discussion and Conclusions

Classification is a critical part in the processing chain of character recognition system. The overall recognition accuracy and speed are decided by the classifier applied. After comparing the three approaches for classification in Subsection 2.3, one can see that template matching approach provides a simple and fast processing, ANN approach may produce accurate results but is time consuming for a large number of training samples. SVM approach is more suitable to the problems with large random variations in data. It has been demonstrated in various papers that the SVM-based classification methods are more accurate than other comparable methods. Kim [34] proposed a SVM-based character recognizer for license plate recognition. The recognition rate of Kim's module is about 97.2%. Zheng [35] compared several types of classifiers and found that SVM approach had the highest accuracy for printed text and handwriting identification in noisy document images. Zhao [36] found the same conclusion after comparing several classifiers for recognizing handwritten number. SVMs have considerable potential for classification. To summarize, we believe that SVM approach provides a better accuracy and is better for recognizing the characters in an incomplete image or an image with heavy noise while template matching is simple and fast. Every different classifier may have its own region in the feature space where it performs the best. Different classifiers trained on the same data may not only differ in their global performance, but they also may show strong local differences. Furthermore, we may have different feature sets, different training sets, different classification methods or different training sessions. Instead of selecting the best method and discarding the others, one can combine various techniques, thereby taking advantages of all the attempts to learn from the data. A combination of classifiers can obviously reduce computational complexity, improve their generalization ability, tolerate the failure of individual classifiers, increase training efficiency, and hence improve the classification accuracy and overall performance.

We can hence conclude that, in order to develop a system for character recognition that is able to handle different types of uncertainty, a combination of various classifiers will be a good selection. In this combined system, various

classifiers are executed in parallel and the outputs of these classifiers are combined to form the final classification results.

References

- [1]"<http://www.htsol.com/AboutUs.html>."
- [2]"<http://www.asiavision.com.hk/>."
- [3]"<http://www.perceptics.com/license-plate-reader.html>."
- [4]"<http://www.ivsuk.com/anpr.asp>."
- [5]"<http://www.singaporegateway.com/optasia/imsps>."
- [6]"<http://www.arhungary.hu/>."
- [7]S.-L. Chang, L.-S. Chen, Y.-C. Chung, and S.-W. Chen, "Automatic license plate recognition," Intelligent Transportation Systems, IEEE Transactions on, vol. 5, pp. 42-53, 2004.
- [8]Z.-Q. Liu, J.-H. Cai, and R. Buse, Handwriting recognition: soft computing and probabilistic approaches. New York: Springer, 2003.
- [9]B. Chazelle, "Application Challenges to Computational Geometry," Princeton University, Princeton, TR-521-96 April 1996.
- [10]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes: The Art of Scientific Computing with IBM PC or Macintosh: Cambridge University Press, 1986.
- [11]K. Sohara and M. Kotani, "Application of Kernel Principal Components Analysis to pattern recognitions," SICE 2002. Proceedings of the 41st SICE Annual Conference, vol. 2, pp. 750 - 752, 2002.
- [12]S. S. Kajarekar, B. Yegnanarayana, and H. Hermansky, "A study of two dimensional linear discriminants for ASR," Acoustics, Speech, and Signal Processing, IEEE International Conference on, vol. 1, pp. 137-140, 2001.
- [13]Y. Huang and S. Luo, "Genetic algorithm applied to ICA selection," Proceeding of the International Joint Conference on Neural Networks, vol. 1, pp. 704 - 707, 2003.
- [14]B. Schölkopf and A. J. Smola, Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond: The MIT Press, 2001.
- [15]A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. Volume: 22, pp. 4 - 37, Jan. 2000.
- [16]D. R. Hush and B. G. Horne, "Progress in supervised neural networks," Signal

- Processing Magazine, IEEE, vol. 10, pp. 8 - 39, 1993.
- [17]H.-S. Lim, "An improved kNN learning based korean text classifier with heuristic information," Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on, vol. 2, pp. 731 - 735, 2002.
- [18]J. Byeungwoo and D. A. Landgrebe, "Fast Parzen density estimation using clustering-based branch and bound," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, pp. 950 - 954, 1994.
- [19]P. Franti, O. Virmajoki, and T. Kaukoranta, "Branch-and-bound technique for solving optimal clustering," Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol. 2, pp. 232 - 235, 2002.
- [20]M. Raus and L. Kreft, "Reading car license plates by the use of artificial neural networks," Circuits and Systems, 1995, Proceedings., Proceedings of the 38th Midwest Symposium on, vol. 1, pp. 538 - 541, 1995.
- [21]C. Coetze, C. Botha, and D. Weber, "PC based number plate recognition system," ISIE '98. Proceedings. ISIE '98. IEEE International Symposium on, vol. 2, pp. 602-605, 1998.
- [22]L. De Vena, "Number plate recognition by hierarchical neural networks," IJCNN '93-Nagoya. Proceedings of 1993 International Joint Conference on Neural Networks, vol. 3, pp. 2105 - 2108, 1993.
- [23]G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by Support Vector Machines," Geoscience and Remote Sensing, IEEE Transactions on, vol. 42, pp. 1335 - 1343, June 2004.
- [24]V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1999.
- [25]K. Yamaguchi, Y. Nagaya, and K. Ueda, "A method for identifying specific vehicles using template matching," Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on, pp. 8 - 13, 1999.
- [26]L. Prasad and S. S. Iyengar, "High performance algorithms for object recognition problem by multiresolution template matching," Tools with Artificial Intelligence, 1995. Proceedings, Seventh International Conference on, pp. 362 - 365, 1995.
- [27]R. M. Dufour, E. L. Miller, and N. P. Galatsanos, "Template matching based object recognition with unknown geometric parameters," Image Processing, IEEE Transactions on, vol. 11, pp. 385 - 1396, 2002.
- [28]B. R. Meijer, "Rules and algorithms for the design of templates for template matching," Pattern Recognition, 1992. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on, vol. 1, pp. 760 - 763, 1992.
- [29]C.-C. Han, P.-C. Chang, C.-C. Hsu, and B.-S. Jeng, "An on-line signature verification system using multi-template matching approaches," Security Technology, 1999. Proceedings. IEEE 33rd Annual 1999 International Carnahan Conference on, pp. 477 - 480, Oct. 1999.
- [30]M. M. M. Fahmy, "Automatic number-plate recognition: neural network approach," Vehicle Navigation and Information Systems Conference, 1994. Proceedings, pp. 99 - 101, 1994.
- [31]J. A. G. Nijhuis, M. H. Ter Brugge, K. A. Helmholz, J. P. W. Pluim, L. Spaanenburg, R. S. Venema, and M. A. Westenberg, "Car license plate recognition with neural networks and fuzzy logic," Neural Networks, 1995. Proceedings., IEEE International Conference on, vol. 5, pp. 2232 - 2236, 1995.
- [32]G. J. Klir, *Fuzzy sets and fuzzy logic: theory and applications*: Upper Saddle River, N.J. : Prentice Hall PTR, c1995, 1995.
- [33]S. Zheng, X. Lu, N. Zheng, and W. Xu, "Unsupervised clustering based reduced support vector machines," Acoustics, Speech, and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference on, vol. 2, pp. II - 821-4, 2003.
- [34]K. K. Kim, K. I. Kim, J. B. Kim, and H. J. Kim, "Learning-based approach for license plate recognition," Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop, vol. 2, pp. 614 - 623, 2000.
- [35]Y. Zheng, H. Li, and D. Doermann, "Machine printed text and handwriting identification in noisy document images," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 26, pp. 337 - 353, Mar. 2004.
- [36]B. Zhao, Y. Liu, and S.-W. Xia, "Support vector machine and its application in handwritten numeral recognition," Pattern Recognition, Proceedings, 15th International Conference on, vol. 2, pp. 720 - 723, 2000.
- [37]I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, pp. 1157-1182, 2003.