International Journal of u- and e- Service, Science and Technology
Vol.9, No. 2 (2016), pp.319-330
http://dx.doi.org/10.14257/ijunesst.2016.9.2.31

Research on Vehicle Type Recognition based on Datasurveying

Zhao jun[1] and LIU Zejun[2]

*(1.Jiangsu food & pharmaceutical science college Huai'an 223004 China;
2. Faculty of A r chitectur e and CivilEngineering , H uaiy in I ns titute of Technology , Huaian 223001, China;)*

Abstract

In the field of transportation, many of the vehicle amount of raw data value which can be accompanied by a large number of unknown information, and these information can provide people with more help. This paper aimed at the master data, and by using the method of survey of data modeling and prediction. First this paper introduces the data survey methods and the characteristics of the model, and then applied to the vehicles recognition judgment in their ability to recognize and to compare the analysis. The conclusion by principal component analysis and discriminant analysis method bus and van in four categories, the two on the training set and testing set of identification is quite high (high purity), so the four classified into two categories, the key to deal with opel (has been) and class and Po (saab) class. After the processing of raw data, and finally compares the model results.

Keywords*: Data surveying; Vehicle type recognition; Discriminant analysis; Clustering analysis; Classification tree; random*

1. Introduction

The establishment of a harmonious society is the theme of The Times. Traffic is the social material, social workers, social resources and so on various factors in an orderly and efficient communication and mobile platform and channel, how to build a reasonable and accurate and efficient transportation, transportation, logistics network is the core content of the transportation sector today. As a major to meet the needs of transportation industry and the research topic: the vehicle type recognition, for further enhance the intelligence and automation level of the vehicle traffic network has realistic and far-reaching significance.

Science and technology in modern society and the world today, with the rapid development of the economy, transportation industry as the relationship between the people living in many fields, although great progress has been made, but still can not fully meet the needs of people living in transportation, and in real life constantly appeared in a series of traffic problems has risen as an urgent problem to be, such as traffic accidents, traffic congestion, car congestion and environmental pollution, and so on, the rapid development of social economy, the number of cars in China increased rapidly. Since 2002, China's auto industry began to enter into explosive growth stage, the auto industry has completed the process of growing up, is gradually realize the huge across from weak to strong. As a rapid huge growth in the number of vehicles in our country, to raises many new problems in the transportation, such as overloaded vehicles, speeding and running red lights, stop to violate the rules such as surge phenomenon, even vehicle deaths and injuries, such as the crime occur repeatedly. According to statistics, China's traffic accident rate compared with this, the United States, Germany, France and other countries, is very big, and traffic accidents frequently occur, caused great damage to our economy

Intelligent, traffic system in our country is relatively late start, and slow development and traffic management means the blade is relatively backward, I have very big one part

ISSN: 2005-4246 IJUNESST

or only rely on artificial on-site maintenance and management, so the traffic law enforcement personnel of the intensity of labor is larger, and the efficiency is not high. Whether in big cities or in many medium and small cities and towns, traffic operation condition, and how to take the management way has become a very important factors influencing the economic development of our country. Under the condition of the trend is more and more intelligent traffic management system, both in terms of monitoring or not parking charge system, the vehicle type identification is very important, so the vehicle type recognition is a hot research topic in recent years.

Modern city requirements management modernization, the traffic modernization and modern living standards. With the development of social economy, the original manual management, the way to charge more and more not adapt to the requirement of modern road traffic safety and the efficient demand, in the face of automatic charging, speeding record, road usage monitoring, we can not help but reflect on how to guarantee the road transportation in clear. First to establish a standard set of vehicle type recognition system is imperative.

The development of science and research make it possible to have an objective to distinguish the vehicle type. On the one hand, the invention of advanced intelligent instrument provides hardware environment for objective to distinguish the vehicle type, by using the instrument display of information control vehicle index to identify the features should be what kind of vehicle type. Statistical methods, data, survey method, on the other hand, the improvement of the difference in the comprehensive applications, provides software support for the objective to distinguish the vehicle type

Sharply increase the number of vehicles today, information on how to better through have to understand all kinds of vehicles, a quick monitoring and how to a particular location, specific time correct identification and classification of vehicle, and use it as a traffic management fee, all kinds of intersection and statistical basis, and quickly achieve the traffic control is the problem need to be solved at present. In some developed countries, the vehicle type automatic classification technology is already quite mature technology, but because of many reasons for the use of these systems in our country is far not meet the requirements, like recognition rate and recognition index meets the requirement of use, and many other problems. To realize the automation of charge system in our country, and scientific management norms, the vehicle type automatic recognition method and its implementation in the form of study is a must. This thesis mainly is data survey method was applied to the vehicle type recognition, using the data obtained, to distinguish the four given vehicle type. Aim to increase the recognition accuracy of vehicle, for domestic intelligent transportation vehicle control to provide certain reference.

2. Related Works

2.1 Common Data Survey Method

In the actual problem, often encountered many indicators (variables), but in most cases, is a certain correlation between the different characteristics. Index under the premise of more, have a certain correlation between the indexes, then analyses the complexity of the problem must be increased. Method of principal component analysis to according to actual needs will be the original characteristic indexes combined into a new set of linearly independent each other again a few comprehensive characteristic index instead of the original characteristic index, at the same time according to the need to take a number from the comprehensive indicators of as little as possible to convey the original index by as much as possible. Principal component analysis[1] is the purpose of the original feature index more Numbers can be converted to relatively few comprehensive index methodshave nothing to do with each other.

Discriminant analysis is to use most Fisher discriminant method[2-4], the basic idea of Fisher's discriminant is projection, aimed at some point x = P dimension (x1, x2, x3, ,,, xp) to find a can make it down to a one dimensional numerical linear function y (x). Then apply the linear function in the P d space known category and general knowledge category of samples are changed for one dimensional data, according to which the degree of closeness of the attribution of unknown sample points to judge its attribution. The linear function should be able to put the P all point in the dimensional space into a one-dimensional numerical, can maximum limit shrink the same kind of difference between each sample points, and to maximize the expanding various differences between the sample points in a different category, so you may gain higher discriminant efficiency. Here to borrow the idea of a yuan of variance analysis, and is based on the mean square deviation between group and group, all the variance ratio of the maximum principle for judging. Fisher discriminant method actually didn't give the most appropriate classification, weighting method in practical application and weighting method [5]. The general idea is to use a maximum discriminant function method to determine, for us to build the discriminant function, if the first j a discriminant function value maximum, we think we have to belong to the first j a class[5].

Clustering analysis is also called data segmentation, has many goals, but involves the division of a collection of objects or grouped into subsets or "clusters" [6](also called clustering), make each cluster within the object than the correlation between the correlation between other objects in the cluster and more closely. Cluster analysis of the common way of thinking is similar to (to) data as a "distance" between the object distance of one dimension, and object to close apart, belong to different classes of object distance is relatively far away. Clustering is the core of the goal of all is for the similarity between individual object concept of clustering. Clustering method according to the definition of similarity given to try to group objects, obtain inherent law. According to the different definition of the distance, there are many in the clustering method. With the more is k. average clustering analysis. K. average (McQueen show, 1967) is used to solve the problem of clustering of famous one of the simplest unsupervised learning algorithm. It is a class label without any data to acquire the method of cluster and the cluster center, determine the desired number of cluster centers, k. average process repeatedly mobile cluster center to make the cluster-heads variance minimization. It USES the squared Euclidean distance as the similarity measure.

Its intuitive basic idea is: first to determine the initial clustering center, to divide the sample points to its nearest cluster center in categories, and then by iteration, center, according to the moving all kinds of clustering center until no longer changes, the final clustering results are obtained. In fact, the distance between the class and class definition methods will produce different clustering method, for example: the shortest distance, middle distance, the longest distance method, gravity method and average method and so on, have the literature shows that also can make use of the minimum tree method to classify elements. Intuitive clustering analysis model, the conclusion is concise. But one thing to say, it is the sample size is large, clustering conclusion has the certain difficulty.

KNN (K - Nearst Neighbor) is the K nearest Neighbor classification[7], the algorithm was first put forward by Cover and Hart, belong to the nonparametric classification algorithm, is widely used in various fields of data exploration and intelligent recognition. Classification thought is: given a classification of the sample X, X is the most similar or a sampler recently K already know the category of the training set of sample, look at the most samples of the K neighbors are belong to which kind of, is the sample X as such. K nearest neighbor algorithm is a kind of typical lazy learning method [10].Design structure model in a given test element classification before the last minute. If and only if the tested group to generalize, so according to the similarity of training tuples stored in classifying tuples.

Specific algorithm steps as follows[8]:

1. Constructtraining sample set X.

2. Determine the initial value of K. For the determination of K value at present there is no a unified way. Different selection of the specific issues of K value is different; Different determine the K value of the same problem, the result is also different. The general method is to determine an initial value, according to the result of the continuous adjusting K values to make to achieve optimal results.

3. In the training sample set, the selection and sample under test or nearest is the most similar K samples. Assumes that the sample points x belongs to n dimensional space, usually by the Euclidean distance to measure the "near abroad" between samples.

4. Given a classification samples of X, with the x1, x2,... Said similar to X or the most recent K samples.

K neighbor does not require explicit rules, so it is higher than other classification methods of classification accuracy. It also has many advantages, such as: there can be noise, that does not require additional data to describe the rules, the training sample is its rules, the consistency of the data does not make the request; Can better avoid the sample amount of imbalance, that is when decisions that category, only with a small amount of adjacent sample is about; When large Numbers of training set is given, the algorithm is still valid. Classification effect is largely dependent on the selection of K value. K value is too large, easy to cause too much noise data and affect the classification accuracy; K value is too small, result in neighbor number is too small will reduce the classification accuracy. For global optimal solutions and local optimal solution of a compromise, a lot of literature put forward the selection of parameter K different method, the literature [9] put forward according to the dynamic context to adjust the K value. Cas Stan proposed an adaptive KNN algorithm, the algorithm for different categories of large difference of sample sizes, method is to set different K, to reduce reliance on sample distribution algorithm. Here a little bit to illustrate, k. neighbor is not suitable for high-dimensional data, here is the problem of "dimension curse" : when the data dimension is quite high, even if the number of sample data is large, the data of scattered points is still relatively sparse, meaning that the vast majority of tested samples near the sample points, will make use of space in the estimates of the number of sample points to build each near neighbour method is not easy to good use.

Logistic regression analysis is a more effective method of processing data, are widely used in various fields, in the research of the relationship between classification variables and a set of independent variables can be done when the Logistic regression analysis[10]. Using Logistic regression to the independent variable, there is noparticularrequirements. In fact, in Logistic model, we can firstly selected by stepwise regression method to the influence degree of the dependent variable is significant variable. Logistic regression coefficients can be interpreted as a unit of corresponding to the change of the independent variables. Of the dependent variable is not normal variable, but logit. Logistic regression parameter estimates usually adopts maximum likelihood (ML), the larger the sample data, the more accurate estimates.

Classification and Regression tree (Classification and Regression Trees) based on statistical theory[11], is a kind of parameter data survey technology. It can automatic screening of variables, in order to reduce the variable dimension, and make full use of the known prior information to deal with non homogeneous relation between data, the data is valid. Classification and regression tree to do is to construct a more accurate classification model to predict, also is the study of variables and the mutual relationship between, through the establishment of the decision tree and decision rules corresponding category projections for type unknown objects, object is through some certain relevant variable

values can be determined its category. It includes two parts: the classification and regression trees. This thesis mainly use classification tree, the idea is: based on the overall sample data, generate a hierarchy tree, leaf node for more, the purpose is to fully reflect the relationship between the data, then the tree under the overtraining data link was given, and then to produce a series of cuts, pruning to subtree, selection of proper size subtree, used for categorical data.

Random Forests (Random Forests cleans, RF) is Breiman in 2001 put forward a new combination of statistical learning classifier is proposed[12]. It is the method of classification and regression tree and bagging methods, using random characteristic internal nodes of the tree property division. Random forest tree classifier is {h (x, ai), I = 1, 2,... } is a collection of one yuan h (x, ai) classifier is built using the CART algorithm without classification and regression tree pruning, x is the input variables, the ai is independent identically distributed random variables, it decides the growth process of single tree; Obtained by a simple majority vote to predict when the final output of random forest.

2.2 Applicable Conditions of Data Survey Method

Compared with discriminant analysis, Logistic model requires that the dependent variable as the binary classification, based on the relationship between independent variable and dependent variable nonlinear relationship. K-nearest model, classification tree is a parameter identification method, nothing special request to the data distribution, so as to avoid the traditional technology of the model set of difficulties.

K - nearest neighbor is not suitable for high-dimensional data, here is the problem of "dimension curse" : when the data dimension is quite high, even if the number of sample data is large, the data in the scattered points in high dimensional space is still relatively sparse, meaning that the vast majority of tested samples near the sample points, will make use of space in the estimates of the number of sample points to build each near neighbor method is not easy to good use.

Neighbor model, Logistic model and Fisher discriminant analysis need each sample at the same dimension and no any missing data. Classification tree and random forest is not strict with the data distribution, can deal with incomplete or noisy data, has strong fault tolerance ability and the generalization capability, practical to broadly.

By aggregating multiple classifiers to improve the accuracy of recognition, these techniques are called integrated classifier. Integration method is composed of a set of training data base classifier (base classifier), then use to vote for the forecast of each base classifier to classify.

3. Basic Data Processing and Analysis

This article from the network using the vehicle data contains 786 sample records, there are 12 attribute variables [13] (car the outline of the related variables, respectively) and 1 about which belongs to a kind of vehicle type of dependent variable. Data in this paper is mainly to use MATLABR2010, SPSS and R2.12. O for data analysis, four kinds of vehicle types of opel (has been) respectively, and Po (saab), bus (bus), vans (van) four categories, and the corresponding sample respectively: 212, 217, 218, 217. Specific 12 independent variables are shown in the following table.

Table 1. Vehicle Type Data

Independent variables	Minimum	Maximum	Mean	Mode
X(1) tightness	70	120	95	89
X(2) ring	35	61	48	47
X(3)radius ratio	110	320	215	199
X(4)aspect ratio	44	134	89	95
X(5)zoom variance	112	265	169	178
X(6) rotating zoom radius	108	270	189	190
X(7)main shaft deflection	58	134	96	93
X(8)short axis deflection	0	33	9	11
X(9)main shaft kurtosis	0	45	14	17
X(10) short axiskurtosis	147	257	195	183
X(11)rounded corners facing ratio	178	210	198	193
X(12) maximum aspect ratio	2	56	36	32

Table 1 is the main property parameter we identify types of vehicles, some is very important for a certain vehicle may not contain the attribute variables, these may produce adverse effect to our analysis, which embodies the vehicle types of certain variable is missing.

Refer to relevant information can know [14], annular, roundness distance, aspect ratio, aspect ratio, the largest about axis of rectangular, postpone the spindle/short axis scaling error of this a few variables mean is decreasing, namely opel maximum, minimum vail. For variance, ring and the radius ratio, narrow body words, rotating about the short axis of zoom radius, kurtosis is gradually diminishing. Except aspect ratio, the narrow body words and deflection about the spindle/short axis of the four variables, the mean of other variables, opel and saab's average than bus and vall, and can be found that much of opel and saab variables is very close to the mean, like a ring, aspect ratio, the axis of the rectangle, the mean radius of rotating zoom and kurtosis are very close; Relatively bus and van with larger average differences, and their average gap with the opel and saab is relatively larger. This illustrates the opel and saab's appearance is very close to some similarity, the data that is purely based on their shape will be more difficult points, and a large car, bus and the van belongs to the obvious gap in shape, will be relatively easy to distinguish.

4. Modeling and Analysis

And according to the proportion of the total sample first randomly is divided into training set and testing set. Training set is used to establish model, the test set is used to estimate the generalization error of the established model. Will be 786 vehicles contour data samples through random sampling randomly divided into 592 training samples, 251 test samples. As the dependent variable as class variables, this article mainly discusses the classification model: the model, the model to carry on the comprehensive consideration, comparing with the results of intelligent recognition rate has reached what we want.

We know that the more independent variables in the case, general multicollinearity between the variables will be obvious, this article variables for 12, the following from the principal component analysis to identify the level of collinearity among variables.

Table 2. First Four Main Components Analysis

Main component	Eigenvalue	Total variance ratio %	Cumulative variance ratio %
1	8.42	58.2	58.2
2	4.23	17.3	75.5
3	1.32	12.2	87.7
4	1.03	5.1	93.8

Can be seen from table 2, the fourth principal component of accumulated variance reached 93.8%, there is serious multicollinearity phenomenon between variables. So in the analysis process, should try to eliminate the influence of multiple collinearity between independent variables.

The ideas of Fisher linear discriminant analysis is a projection, K (class) P d data projection to a certain direction, to make the projection as far as possible between class and class are separate, namely to find the classification of the practical problems of projection line, the basic idea is based on the variance within every class of small as possible, and can make the variance between the class as far as possible big benchmark to determine the discriminant function. Since there are four kinds of vehicle types, so should be four discriminant functions. By the relevant materials [13], because in shape, the appearance of the bus and the van this 2 class is more special, belong to large cars, appearance is also relatively large, so the use of classifier to identify they have higher accuracy is obviously. And two classes on the contrary, because of their similarity in shape, lead to classifier cannot be accurately identified, so the rest of this article mainly studies the classification of the opel and saab problem.. Fisher linear discriminant result in the following table.

Table 3. Fisher Linear Discriminant Classification Results

	Vehicle type	1	2	Total
Training set number	1	115	48	31
	2	36	104	140
Percentage %	1	78.4	21.6	100
	2	32.1	67.9	100
Test set number	1	43	18	61
	2	24	49	73

Percentage %	1	68.3	31.7	100
	2	35.7	64.3	100

One for opel, 2 for saab, as can be seen from the table, using the Fisher linear discriminant method, the training set of performances of opel and saab are as follows: 78.4% and 67.9%; To predict concentration, identify the correct rates were 68.3% and 64.3%, respectively.

The mean K_ (McQueen show, 1967) is used to solve the problem of clustering of famous one of the simplest unsupervised learning algorithm, belongs to the non-parametric statistical method. For the purpose of this article in the sample, because only two categories, opel and saab, so in K.M eans algorithm set K = 2, USES the Euclidean distance, the strategy of random initial cluster centers, using the mathematical models of K - Means classifying training set and testing set respectively, get the results shown in the table below.

Table 4. K-Means Algorithm Classification Results

	Vehicle type	1	2	Total
Training set number	1	56	49	105
	2	58	48	106
Percentage %	1	58.4	41.6	100
	2	59.2	40.8	100
Test set number	1	28	24	52
	2	42	29	71
Percentage %	1	58.6	41.4	100
	2	56.7	43.3	100

K nearest neighbor classification thought is: given a classification of the sample x, x is the most similar or a sampler recently K already know the category of the training set of sample, according to most of K neighbor sample is belong to which kind of, to determine the x into which category. K nearest neighbor algorithm is a kind of typical lazy learning method. Take K=4, the classification results are obtained as follows.

Table 5. K neighbor Algorithm Classification Results

	Vehicle type	1	2	Total
Training set number	1	105	48	153
	2	55	90	145
Percentage %	1	68.3	31.7	100
	2	62.1	37.9	100
Test set number	1	39	25	64
	2	36	47	83
Percentage %	1	64.8	35.2	100
	2	60.2	39.8	100

The training samples and testing samples total recognition rate were 68.3% and 62.1% respectively. Training samples of the correct recognition rate is 69.8% and 61.5%; Test samples the correct recognition rate of 60.3% and 57.9%, respectively. Actually, from the above several kinds of models together k. neighbor model, its effect is not so ideal, the vehicle outline data acquired with the variables about too much.

Logistic regression analysis, as a kind of effective data processing method, the regression model is the dependent variable is a binary classification observations and white variables (factors) of the relationship between important multivariate analysis model of automobile contour Logisic regression variables related to data classification, classification results, as shown in table 6. It is concluded that the training sample and test sample recognition rate were 71.4% and 66.7% respectively. Training samples of the correct recognition rate is 71.8% and 70.9%; Test samples the correct recognition rate of 68.3% and 65.2%, respectively.

Table 6. Logistic Regression Analysis Classification Results

	Vehicle type	1	2	Total
Training set number	1	123	34	157
	2	25	117	142
Percentage %	1	71.8	28.2	100
	2	29.1	70.9	100
Test set number	1	41	22	63
	2	21	58	79
Percentage %	1	68.3	31.7	100
	2	34.8	65.2	100

This paper mainly discusses the issue of classification, according to the establishment of the theory of the third chapter after classification trees need to be pruned to reduce the complexity of the model, in order to avoid excessive fitting. Input parameters for classification and regression tree: the rule of GINI purity function is not divided, with MATLABR2010 get the following results of classification and regression tree figure 1 for tree pruning process.

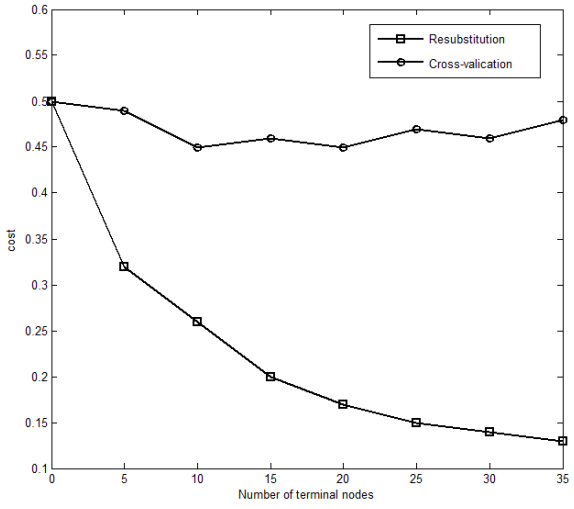

Figure 1. Tree Pruning Process

Observation above can be found that the optimal classification and regression tree only selected the X1, X10, X18, X15, X17 X11 as bifurcate variables, respectively: the tightness, about the maximum length of the rectangle, rounded surface ratio, about the short axis deflection, about the short axis deflection, a scale error in the spindle. This shows the importance of these variables than other more important. This paper established the champion tree to classify the training sample and test sample, get the classification results, as shown in the table below.

Table 7. Champion Tree Classification Results

	Opelcorrect ratio	Saab correct ratio	Total correct ratio
Training set	0.73	0.69	0.71
Test set	0.52	0.55	0.53

From the conventional wisdom in the reality we know, four opel vehicles category, and bo, buses and trucks in the shape of the gap between the last two obviously, classification identified is easy, but before the appearance is very similar, tell some relatively difficult, this article is to compare the pure after two, mainly for the former two are classified.

Can know from the above analysis, the predicted effect of several methods for the training sample is better than the predicted effect of test sample; Discriminant analysis, Logistic, classification tree model of three ability for the forecast of the training sample is very close, and prediction of test sample, the effect of the classification tree is poor, the other two close to can reach 66%; K. average method of the model prediction effect is the worst, the training set and testing set of sample recognition correct rate is very low.

5. Research Projects

1. Senior Visiting Scholar Program Sponsored by Education Department of Jiangsu Province(2014FX033);

2. Construction, Science and Technology Project Sponsored by Housing and Construction Department of Jiangsu Province(2014JH20);

3. Science and Technology Project Sponsored by Huai'an Municipality（HAS2014025-3,HAS2014021-2）

6. Conclusion

In this paper, the method of using data survey for modeling and forecasting. First this paper introduces the methods of survey data and the characteristics of the model, and then applied to the vehicle recognition judgment of their ability to recognize and has carried on the comparative analysis. The conclusion by principal component analysis and discriminant analysis method bus and van in four categories, the two on the training set and testing set of identification is quite high (high purity), so the four classified into two categories, the key to deal with opel (has been) and class and Po (saab) class. After the processing of raw data, finally compares the model results.

Reference

[1] Z. Cai, H. Hu and Z. L. H. Huang, "Automatic recognition algorithm of vehicle type based on video-content analysis", Information Science and Management Engineering II (3 Volume Set), **(2014)**, pp. 62, 217.

[2] Z. X. Zhang, T. Li and T. Xiang, "A Novel Vehicle Type Recognition Based on Hierarchical Matching", Applied Mechanics and Materials, **(2014)**.

[3] Y. Wang, X. Wu and X. Li, "Vehicle Type Recognition in Sensor Networks Using Improved Time Encoded Signal Processing Algorithm", Mathematical Problems in Engineering, **(2014)**.

[4] X. Su, C. Zhang and L. Mei, "Vehicle face recognition using weighted visual patches", Multimedia and

Expo Workshops (ICMEW), IEEE International Conference, IEEE, **(2014)**.
[5] P. Melin and O. Castillo, "A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition", Applied Soft Computing, **(2014)**.
[6] S. J. Press and S. Wilson, "Choosing between logistic regression and discriminant analysis", Journal of the American Statistical Association, vol. 73, no. 364, **(2008)**, pp. 699-705.
[7] H. Emami, M. Fathi and K. Raahemifar, "Real Time Vehicle Make and Model Recognition Based on Hierarchical Classification", International Journal of Machine Learning & Computing, vol. 4, no. 2, **(2014)**.
[8] B. Cyganek and M. Woźniak, "Vehicle Logo Recognition with an Ensemble of Classifiers", Intelligent Information and Database Systems. Springer International Publishing, **(2014)**, pp. 117-126.
[9] I. Zafar, E. A. Edirisinghe and S. Acar, "Two-dimensional statistical linear discriminant analysis for real-time robust vehicle-type recognition", Electronic Imaging, International Society for Optics and Photonics, **(2007)**.
[10] T. Li, S. Zhu and M. Ogihara, "Using discriminant analysis for multi-class classification", IEEE 13th International Conference on Data Mining, IEEE Computer Society, **(2003)**.
[11] C. Y. Chen and F. Ye, "Particle swarm optimization algorithm and its application to clustering analysis", Networking, Sensing and Control, IEEE International Conference, IEEE, **(2004)**.
[12] R. Baumgartner, L. Ryner and W. Richter, "Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis", Magnetic Resonance Imaging, vol. 18, no. 1, **(2007)**, pp. 89-94.
[13] A. U. Peker, O. Tosun and H. L. Akin, "Fusion of map matching and traffic sign recognition", Intelligent Vehicles Symposium Proceedings, IEEE, **(2014)**.
[14] M. Z. Siam, "Practical Design of an Automatic License Plate Recognition Using Image Processing Technique", methods, vol. 4, no. 4, **(2014)**.

Authors

Zhao jun, She received the B.Eng degree in Computer Engineering from NanJing University of Science & Technology in 1998, and the Master's degreein Computer EngineeringfromJiangSu University in 2003. She is currently researching on the Network security, data mining

LIU Zejun, He recrived the B.Eng degree in Water conservancy engineering from Hohai University in 2003, and the Master's degree in Water conservancy engineering from Hohai University in 2008. He is currently researching on the Water conservancy engineering and Teaching research

