

Multi-Traffic Scene Perception Based on Supervised Learning

Lisheng Jin, Mei Chen, Yuying Jiang, Haipeng Xia

Abstract—Traffic accidents are particularly serious at a rainy day, night without street lamp, overcast, rainy night, foggy day and many other low visibility conditions. Present vision driver assistance systems are designed to perform under good-natured weather conditions. Classification is a methodology to identify the type of optical characteristics for vision enhancement algorithms to make them more efficient. To improve machine vision in bad weather situations, a multi-class weather classification method is presented based on multiple weather features and supervised learning. Firstly, underlying visual features are extracted from multi-traffic scene images, and then the feature was expressed as an eight-dimensions feature matrix. Secondly, five supervised learning algorithms are used to train classifiers. The analysis shows that extracted features can accurately describe the image semantics and the classifiers have high recognition accuracy rate and adaptive ability. The proposed method provides the basis for further enhancing the detection of anterior vehicle detection during nighttime illumination changes, as well as enhancing the driver's field of vision in a foggy day.

Index Terms—Underlying visual features, supervised learning, intelligent vehicle, complex weather conditions, classification

I. INTRODUCTION

Highway traffic accidents bring huge losses to people's lives and property. The advanced driver assistance systems (ADAS) play a significant role in reducing traffic accidents. Multi-traffic scene perception of complex weather condition is a piece of valuable information for assistance systems. Based on different weather category, specialized approaches can be used to improve visibility. This will contribute to expand the application of ADAS.

Little work has been done on weather related issues for in-vehicle camera systems so far. Payne *et al.* propose classifying indoor and outdoor images by edge intensity [1]. Lu *et al.* propose a sunny and cloudy weather classification method for single outdoor image[2]. Lee and Kim propose intensity curves arranged to classify four fog levels by a neural network [3]. Zheng *et al.* present a novel framework for recognizing different weather conditions[4]. Milford *et al.* present vision-based simultaneous localization and mapping in changing outdoor environments [5]. Detecting critical changes

of environments while driving is an important task in driver assistance systems[6]. Liu *et al.* propose a vision-based skyline detection algorithm under image brightness variations[7]. Fu *et al.* propose automatic traffic data collection under varying lighting conditions [8]. Fritsch *et al.* use classifiers for detecting road area under multi-traffic scene [9]. Wang *et al.* propose a multi-vehicle detection and tracking system and it is evaluated by roadway video captured in a variety of illumination and weather conditions[10]. Satzoda *et al.* propose a vehicle detection method on seven different datasets that captured varying road, traffic, and weather conditions[11].

II. PROBLEM STATEMENT

A. Impact of complex weather on driver

Low visibility conditions will bring the driver a sense of tension. Due to variations of human physiological and psychological, driver's reaction time is different with the different driver's ages and individuals. The statistics show that driver's reaction time in complex low visibility weather conditions is significantly longer than on a clear day. In general, the driver's reaction time is about 0.2s ~ 1s. If the driver needs to make a choice in complex cases, driver's reaction time is 1s ~ 3s. If the driver needs to make complex judgment, the average reaction time is 3s~ 5s.

The overall stopping distance can be defined as $d = d_r + d_b$. It includes the distance $d_r = t_r v_0$ that means the driver's reaction time and the stopping distance $d_b = v_0^2 / 2a$. Hereby v_0 describes the initial velocity, t_r donate the reaction time and a represent deceleration rate.

As shown in Fig. 1, when the initial braking speed is 100km/h, if the driver's reaction time is 1.5s, 3s, 5s respectively, the braking distance is 93.11m, 134.77m, 190.33m respectively [12]. This means that if driver's response delay one second, it may lead to serious traffic accidents. The data is obtained on the dry road, the friction coefficient is 1.0, deceleration rate is $7.5m/s^2$. The mean deceleration originate from an example of the Bavarian police, taken from their website <http://www.polizei.bayern.de/verkehr/studien/index.html/31494> on 28th October 2011.

B. Enhancing the driver's field of vision in foggy day and night

Weather understanding plays a vital role in many real-world applications such as environment perception in self-driving cars. Automatic understanding weather conditions can enhance traffic safety. For instance, Xu *et al.* summary image defogging algorithms and related studies on image restoration and enhancement [13]. Gallen *et al.* propose a nighttime visibility estimation method in the presence of dense fog[14]. Gangodkar *et al.* propose a vehicles detection method under complex

This work is supported by the National Natural Science Foundation (no. 51575229), National key R & D project of China (no. 2017YFB0102600), and Electric intelligent vehicle innovation team of the Science and Technology Department of Jilin Province. (Corresponding author: Yuying Jiang).

L. Jin is with the Transportation College of JiLin University, Changchun 130022, China (e-mail: jinls@jlu.edu.cn).

M. Chen is with the Transportation College of JiLin University, Changchun 130022, China. (e-mail: llyjks@163.com).

Y. Jiang is with the China-Japan Union Hospital of Jilin University, Changchun 130033, China(e-mail: jiangyy@jlu.edu.cn).

H. Xia is with the Transportation College of JiLin University, Changchun 130022, China (e-mail: xhp774380200@163.com).

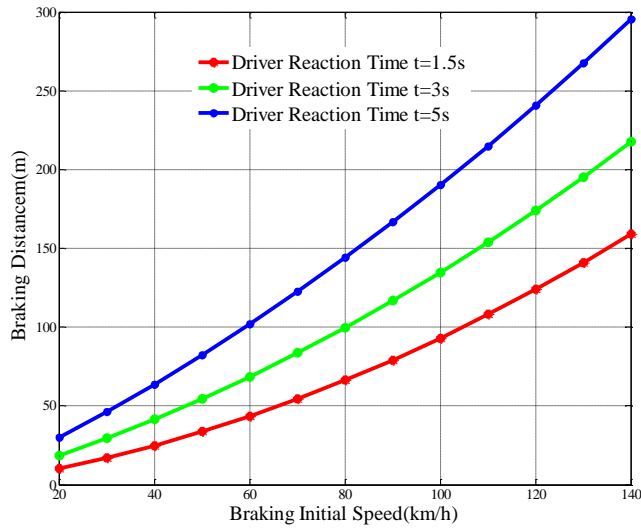


Fig. 1 Different braking distance caused by different reaction time at different brake initial velocity.

outdoor conditions[15]. Chen *et al.* propose night image enhancement method in order to improve nighttime driving and reduce rear-end accident [12]. Kuang *et al.* present an effective nighttime vehicle detection system based on image enhancement [16]. Yoo *et al.* present an image enhancement algorithm for low-light scenes in an environment with insufficient illumination [17]. Jung propose a image fusion technique to improve imaging quality in low light shooting [18]. Zhou *et al.* present global and local contrast measurements method for single-image defogging [19]. Liu *et al.* present single image dehazing by using of dark channel model[20]. Pouliand and Reinhard present a novel histogram reshaping technique to make color image more intuitive [21]. Arbelot *et al.* present a framework that uses the textural content of the images to guide the color transfer and colorization[22]. In order to improve visibility, Xiang *et al.* propose a improved EM method to transfer selective colors from a set of source images to a target image [23].

C. Flow Framework

In this work, firstly, owing to classify multi-traffic scene road images, underlying visual features (color features, texture features, edge features) are extracted from multi-traffic scene images, and then the features expressed as eight-dimensions feature matrix. The traffic scene classification problem is becoming the supervised learning problems. Secondly, BP neural network, support vector machine, probabilistic neural network, S_Kohonen network and extreme learning machine algorithms are used to train classifiers. In order to achieve weather images automatic classification, the main steps are shown in Fig.2.

This paper is organized as follows. An experimental image set is constructed and global underlying visual features are extracted in Section III. Five supervised learning classification algorithms are introduced in Section IV. Comparison and analysis of the five supervised learning classification methods are illustrated in Section V. Finally, we conclude the paper in Section VI.

III. CONSTRUCT AN EXPERIMENTAL IMAGE SET AND EXTRACT UNDERLYING VISUAL FEATURE

Image feature extraction is the premise step of supervised learning. It is divided into global feature extraction and local feature extraction. In the work, we are interested in the entire image, the global feature descriptions are suitable and conducive to understand complex image. Therefore, multi-traffic scene perception more concerned about global features, such as color distribution, texture features.

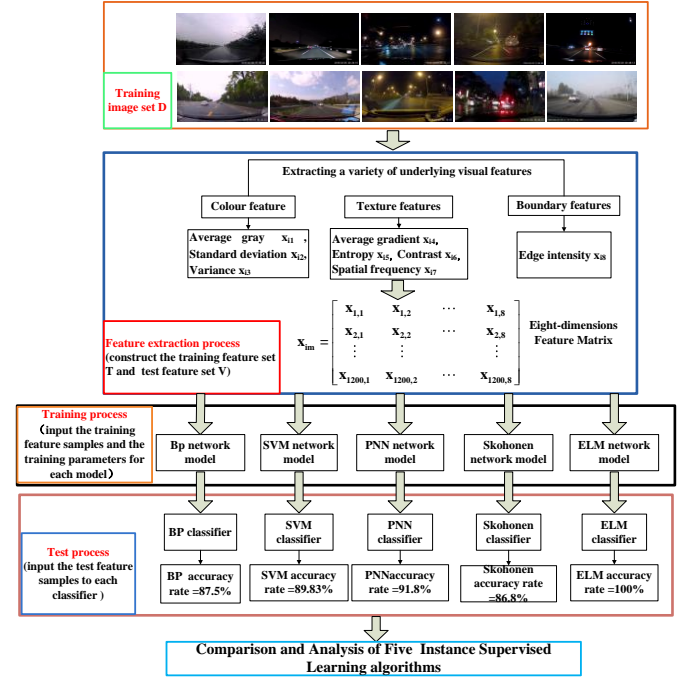


Fig.2 Multi-traffic scene classification algorithm flow framework diagram

Image feature extraction is the most important process in pattern recognition and it is the most efficient way to simplify high-dimensional image data. Because it is hard to obtain some information from the $M \times N \times 3$ dimensional image matrix. Therefore, owing to perceive multi-traffic scene, the key information must be extracted from the image.

A. Create an experimental image set

In the work, there are 1200 images are collected by use of driving recorder and the image set D is established for training and test. There are 10 categories traffic scene images are classified in the work, 120 images are chosen from each category at random. The camera system provides images with a resolution of 856×480 pixels. The images are obtained respectively under rainy day, night without street lamp, night with street lamp, overcast, sunlight, rainy night, foggy day and other low visibility road environment images, the classification labels is 1-10, sub-sample set is shown in Fig. 3. The image set $D = \{(Img_1, y_1), (Img_2, y_2), \dots, (Img_N, y_N)\}$, $N \in \{1, 2, \dots, 1200\}$, the category label can be expressed as $y = \{y_1, y_2, \dots, y_N\}$, $y_i \in \{1, 2, 3, \dots, 10\}$ $i = 1, 2, 3, \dots, N$.



Fig. 3 Sub-sample of 10 categories traffic road scene (from left to right, from top to bottom, the labels are 1-10)

B. Underlying visual features extraction

In order to train classifier, underlying visual features are extracted that can describe color distribution and structure of image. Such as, color features, texture features[24-26], and edge features. Han *et al.* propose a road detection method by extracting image features[27]. Zhou *et al.* propose a automatic detection of road regions by extract distinct road feature [28]. Bakhtiari et al. propose a semi automatic road extraction from digital images method[29]. Chowdhury et al propose a novel texture feature based multiple classifier technique and applies it to roadside vegetation classification[30]. In this work, average gray x_{i1} , standard deviation x_{i2} , variance x_{i3} , average gradient x_{i4} , entropy x_{i5} , contrast x_{i6} , spatial frequency x_{i7} and edge intensity x_{i8} are extracted and they are as shown in formula(1). The feature sample set with labels can be expressed as $C=[x_{i1}, x_{i2}, \dots, x_{im}]$, $i=1, 2, 3 \dots N$, $m=1, 2, 3 \dots 8$, where, i represents the number of images in the image set. The process that extracted eight underlying visual features is simple, time saving and eight underlying visual can comprehensively describe the visual information of the image.

$$x_{im} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,m} \\ x_{2,1} & x_{2,2} & \dots & x_{2,m} \\ \vdots & \vdots & & \vdots \\ x_{i,1} & x_{i,2} & \dots & x_{i,m} \end{bmatrix} \quad (1)$$

Where, $i=1, 2, 3 \dots 1200, m=1, 2, 3 \dots 8$.

The underlying visual features are described as follows:

1) Average Gray

The average gray can reflect the average brightness of image. According to the distribution of visual effect, average gray value between 100 and 200 belongs to optimal visual. The formula of average gray AG can be expressed as follows:

$$AG = k * P_k \quad (2)$$

where, $P_k = \frac{N_k}{w * h}$, k represents gray value of the input image, $k \in (0 \dots 255)$, N_k indicates the number of pixels with a gray value k , w represents image width, h represents image height, P_k represents frequencies histogram of relative gray value.

2) Standard Deviation

The image standard deviation denotes the discrete situation of each pixel's gray value relative to the average gray-value of the images. In general, the larger the variance, the more abundant the gray layer of the image, and the better the definition. According to the distribution of visual effect, the standard deviation value between 35 and 80 is the optimal visual. The formula to determine the standard deviation of the image can be expressed as follows:

$$SD = \sqrt{\frac{\sum_{i=1}^w \sum_{j=1}^h (A_{ij} - \bar{A}_{ij})^2}{w * h}} \quad (3)$$

where, A_{ij} represents gray value of the image at the pixel point (i, j) , \bar{A}_{ij} represents average gray value of the image.

3) Variance

The variance is the square of the standard deviation and it represents the degree of discrete of the image pixels. If the standard deviation results are not obvious, variance can enlarge the distinction between features. The formula to determine the variance of the image can be expressed as follows:

$$V = SD^2 \quad (4)$$

4) Average Gradient

The average gradient is an important technical characteristics indicator of the image structure. The average gradient of the images can reflect the details and image definition. In general, the larger images average gradient, the more abundant the images marginal information in the images, and the clearer the image will be. The average gradient formula for gray images is as follows:

$$AG = \frac{\sum_{i=1}^w \sum_{j=1}^h \sqrt{\frac{(A_{ij} - A_{(i+1)j})^2 + (A_{ij} - A_{ij+1})^2}{2}}}{w * h} \quad (5)$$

where, A_{ij} represents gray value of the image at the pixel point (i, j) , $A_{(i+1)j}$ represent gray value of the image at the pixel point $(i+1, j)$, A_{ij+1} represent gray value of the image at the pixel point $(i, j+1)$.

5) Entropy

The entropy describes the gray value distribution. It is independent of the position of the pixels. It means that the position of pixels has no influence on the entropy within an image. Information entropy of a clear image is greater than the information entropy of an unclear image. Furthermore, the color information entropy can distinguish the different multi-traffic scene images. The calculating formula of the image information entropy is as follows:

$$EN = -\sum_{k=0}^{255} P_k \log_2(P_k) \quad (6)$$

6) Contrast

Contrast describes the variation of image values in image space. In general, the better the image resolution, the larger the image contrast will be. The contrast of clear images is usually larger than the contrast of unclear images. The contrast of the narrow sense is the main factor to decide different texture structure that can be used for image classification and segmentation problems[31]. The contrast features are significant as a global textures description to distinguish multi-traffic scene image. Contrast varies widely depending on the lighting conditions of the different scenes. So, contrast can be used as the typical feature to distinguish multi-traffic scene image. Its formula is as follows:

$$C = \frac{SD}{\sqrt{\sum_{k=1}^{255} (k-AG)^4 * P_k}} \quad (7)$$

where, SD represents image standard deviation, AG represents average gradient, $P_k = \frac{N_k}{w * h}$, k represents variance, V represent variance,

represents gray value of the input image.

7) Spatial Frequency

Spatial frequency is a texture feature that reflects the overall activity of an image spatial domain and it describes the variation of image values in image space. Its formula is as follows:

$$SF = \sqrt{\frac{\sum_{i=1}^w \sum_{j=1}^h (A_{ij} - A_{i(j-1)})^2}{w * h} + \frac{\sum_{i=1}^w \sum_{j=1}^h (A_{ij} - A_{(i-1)j})^2}{w * h}} \quad (8)$$

where, A_{ij} represents gray value of the image at the pixel point (i, j) , $A_{i(j-1)}$ represent gray value of the image at the pixel point $(i, j-1)$, $A_{(i-1)j}$ represent gray value of the image at the pixel point $(i-1, j)$.

8) Edge Intensity

Edge intensity characterizes the edge of the image. It can be known that the image edge intensity can be used as the typical feature to distinguish multi-traffic scene image. The aim of extracting edge intensity is to identify points that the image brightness changes sharply or discontinuity in a digital image. Edge feature extraction is a fundamental work of image processing and feature detection in computer vision field. The formula is as follows:

$$ED = \sqrt{P_{ij}^2 + Q_{ij}^2} \quad (9)$$

where, $P_{ij} = \frac{1}{2} \sum_{i=1}^w \sum_{j=1}^h (A_{ij+1} - A_{ij} + A_{(i+1)(j+1)} - A_{(i+1)j})$ represents the horizontal edge intensity at the pixel point (i, j) , and

$Q_{ij} = \frac{1}{2} \sum_{i=1}^w \sum_{j=1}^h (A_{ij} - A_{(i+1)j} + A_{i(j+1)} - A_{(i+1)(j+1)})$ represents the vertical edge intensity.

TABLE I
EIGHT FEATURES OF TEN CATEGORIES TRAFFIC SCENE IMAGES IN Fig. 3

Category	Average gray	Standard deviation	Variance	Average gradient	Entropy	Contrast	Spatial frequency	Edge intensity
1	122.8828	91.25563	8327.59	1.995528	10.7268	83.88179	12.78693	20.63685
2	17.87396	33.86915	1147.119	1.623863	6.736342	17.2878	12.28752	17.46855
3	24.59817	36.83159	1356.566	1.718391	9.402654	18.97007	11.41997	18.10727
4	21.39146	27.05671	732.0653	2.00427	11.21858	11.77669	11.51215	21.25563
5	9.856643	29.97591	898.5554	1.532995	4.961755	12.24682	13.01507	16.31867
6	132.6913	87.28431	7618.551	4.863127	11.59447	77.96682	20.23549	50.18328
7	150.4952	68.61536	4708.068	4.26789	13.40175	54.51677	17.54234	45.13826
8	46.82204	37.92524	1438.324	2.48249	13.09253	21.79842	12.16807	26.44314
9	37.91887	45.99895	2115.903	3.789411	12.92101	22.16752	15.77728	40.08121
10	138.1328	84.92857	7212.862	2.209985	10.65098	76.0236	11.95648	22.95146

TABLE II
NORMALIZED DATA OF EIGHT UNDERLYING VISUAL FEATURES IN TABLE I

Category	Average gray	Standard deviation	Variance	Average gradient	Entropy	Contrast	Spatial frequency	Edge intensity
1	0.787402	0.912031	0.86024	0.179792	0.714914	0.914055	0.267747	0.177267
2	0.084166	0.165838	0.076538	0.109	0.353028	0.122763	0.233997	0.118452
3	0.129198	0.204358	0.099398	0.127005	0.59483	0.142753	0.175369	0.130309
4	0.107723	0.077256	0.031238	0.181457	0.759513	0.057278	0.181599	0.188754
5	0.030475	0.115214	0.049409	0.091693	0.192095	0.062865	0.283164	0.097106
6	0.853089	0.860392	0.782853	0.725984	0.793602	0.843772	0.771112	0.725762
7	0.97232	0.617641	0.465193	0.612609	0.9575	0.56513	0.589111	0.632107
8	0.278029	0.218579	0.108321	0.272544	0.929457	0.17636	0.225925	0.285054
9	0.218406	0.323561	0.182275	0.521473	0.913902	0.180746	0.469831	0.538229
10	0.88953	0.82976	0.738575	0.220639	0.708038	0.820682	0.211626	0.220235

IV. INTRODUCTION OF SUPERVISED LEARNING CLASSIFICATION ALGORITHMS

In Section III, each image will be transformed into a learning bag by extracting eight features. After extracted global features, machine learning classification approaches come into operation. In recent ten years, a variety of pattern recognition methods have been proposed and proved is useful. Maji *et al.* propose additive kernel svms for classification [32]. A histogram intersection kernel and support vector machine classifiers are presented for image classification [33]. A deep neural networks image detection was presented in[34]. A review paper about fault and error tolerance in neural network was presented in[35]. Another new related method was presented in[36]. A BP-NN and improved-adaboost algorithm was presented[37]. In this section, five supervised learning algorithms will be introduced to solve the multi-traffic scene classify problem.

A. Back Propagation Neural Network classifier

BP network was presented by Rumelhart and McClland. It is a multi-layer feed forward network that was trained by error back propagation algorithm. Currently, it is one of the most widely used neural network model. Its learning rule is constantly adjusting the whole network weight and threshold through the reverse propagation in order to get the minimum network error square sum. In our work, let m , M and n respectively stand for the number of the input layer nodes, hide layer nodes and output layer nodes. The process of multi-traffic scene perception using BPNN is shown as follows.

The main steps of achieving images automatic classification

(1) Feature extraction	Extracting the underlying visual features from image set D and construct feature set C.
(2) Construct training set and test set	Feature set C are divided into two parts in the proportion of 50% to construct training set T and test set V.
(3) Training process	Import hidden layer nodes, training feature set T and other parameters to the network. The Output is a classifier F.
(4) Test process	Test set V is used to test classifier F and final we obtain the accuracy, calculating time and other evaluation index.

In our case, the classes correspond to weather situations which we divide into {clear weather, light rain, heavy rain, night without street lamp, overcast, rainy night, foggy day}. Thus, the problem of classification can be thought of as finding some function f that maps from descriptor space C into the classes F.

In this section, BP network is used to train a classifier. The sigmoidal function is chosen by test common kernels function. The number of iterations is 10000, the learning rate is 0.1 and the target value is 0.00004. The specific method include: Firstly, a total of 60 images are randomly selected from each category road environment image. Secondly, in order to construct the training feature set T, eight global underlying visual features are extracted from 600 images. Thirdly, in order to construct the test feature set V, eight underlying visual

features were extracted from the remaining 600 images. Test result is shown in Fig.4. The X axis represents the 600 test images, and the Y axis represents the 10 categories traffic scene. BP network recognition accuracy rate can reach 87.5% when the number of hidden neurons at 240.

The recognition rate represents the label of the actual test image coincides with the label of predicted test image, it means the classification is correct. The accuracy rate is calculated as follows.

$$Accuracy = \frac{\sum_{i=1}^{i=600} num(simlabel_i - testlabel_i = 0)}{num(testlabel_i)} \times 100\% \quad (10)$$

which, the $testlabel_i$ represents actual category label of the test image and $simlabel_i$ represents the predicted category label of test image after test process.

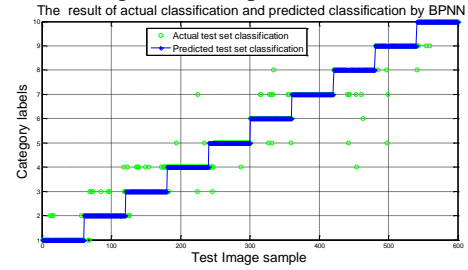


Fig.4 The result of the actual classification and predicted classification by BPNN (accuracy =87.5%, elapsed time =1.398 seconds)

B. Support vector machine classifier

The support vector machine was first proposed by Cottes and Vapnik. The most classic DD-SVM [38] and MILES algorithm [39] are proposed by Chen for image classification. Based on statistical learning theory (SLT), SVM can automatically find the support vector. The support vector can distinguish different image category and it can maximize the interval between class and class. As svm is simple, fast, powerful and robust, we decided to use SVM as our learning and classification method.

There are many toolboxes for implementing SVM, such as LSSVM, SVMlight, Weka, SVMlin, SVM_SteveGunn LIBSVM-FarutoUltimate, LS-SVMLab Toolbox, OSV SVM Classifier Matlab Toolbox. LIBSVM package developed by Professor Lin Chih-Jen of Taiwan University in 2001. Because LIBSVM package is a simple, fast and effective SVM toolbox, LIBSVM was used for classifying image in this section. Radical basis function is chosen as kernel function. We set the scale factor $g = 1$ and the penalty factor $c = 2$. The specific method is as same as the BP network in Section IV-A. Test result is shown in Fig. 5.

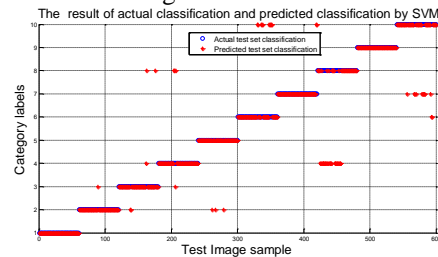


Fig. 5 The result of the actual classification and predicted classification by SVM (accuracy = 89.833% (539/600), elapsed time =0.334 seconds)

C. Probabilistic Neural Network Classifier

Probabilistic neural network (PNN) was first proposed by Dr. D. F. Specht. In principle, although BP network as same as PNN are calculated by neurons, the models are different. A newff function is used to create the BP network and a newpnn function is used to create a probabilistic neural network. There are some advantages for image classification by using PNN. Firstly, PNN is training fast. The training time is only slightly larger than the time of read the data. Secondly, no matter how complex the classification problem is, as long as there is enough training data, PNN can get the optimal solution under the bayesian criterion. Thirdly, PNN allow to increase or decrease training data and it is no need to re-train.

Therefore, PNN is used for classifying image in this section. Radical basis function is chosen as kernel function and we set the distribution density spread is 1.5. The specific method is as same with the BP network that described in section IV-A. Test result is shown in Fig.6.

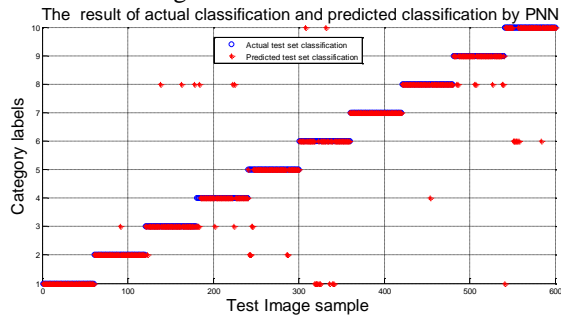


Fig.6 The result of the actual classification and predicted classification by PNN (accuracy =91.8%, elapsed time is 3.636 seconds)

D. S_Kohonen network classifier

SKohonen neural network is a feed forward neural network. Let m, M and n respectively stand for the number of input layer nodes, competitive layer nodes and output layer nodes. When SKohonen network is used for supervised learning, radical basis function is used as kernel function. We set the input node number m= 8, competitive layer node M= 8 and output layer node n=10. The maximum learning rate 0.01, the learning radius is 1.5 and the number of iterations is 10000. The specific method is as same as the BPNN that described in section IV-A. Test result is shown in Fig.7.

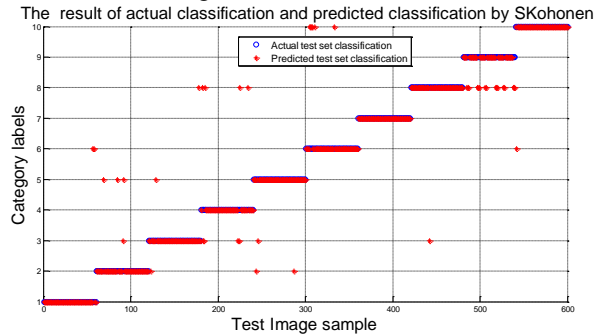


Fig.7 The result of the actual classification and predicted classification by SKohonen (accuracy = 86.8%(521/600) , elapsed time is 2.137 seconds)

E. Extreme Learning Machine classifier

Extreme Learning Machine was first proposed in 2004 by

Huang Guangbin in Nanyang Technological University. Huang et al. propose using extreme learning machine for multiples classification [40]. ELM is a single hidden layer feed forward neural network learning algorithm. Let m, M and n respectively stand for the number of input layer nodes, hide layer nodes and output layer nodes. If $g(x)$ is the activate function of the hidden layer neurons. b_i presents the threshold. There are N different image samples and the feature set can be expressed as $C=[x_{i1}, x_{i2}, \dots, x_{im}]$. Therefore ELM network training model is shown in Fig. 8[41, 42].

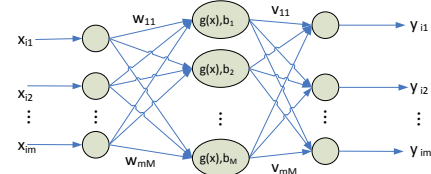


Fig. 8 ELM network training model

Mathematical formula of ELM network model can be expressed as follow:

$$\sum_{i=1}^M v_i g(w_i \cdot x_i + b_i) = o_j, j = 1, 2, 3 \dots N \quad (11)$$

which, $w_i = [w_{i1}, w_{i2}, \dots, w_{mi}]$ represent input weight vector that located between the network input layer node and the hidden layer node. $v_i = [v_{i1}, v_{i2}, \dots, v_{in}]$ represent output weight vector that located between the hidden layer node and the network output node. $o_j = [o_{j1}, o_{j2}, \dots, o_{jn}]$ represent the network predicted output value.

We propose ELM is used for classfying image. The sigmoidal function is used as the kernel function. The number of hidden neurons is 200. The specific method is as same as the BPNN that described in Section IV-A. The test result is shown in Fig. 9.

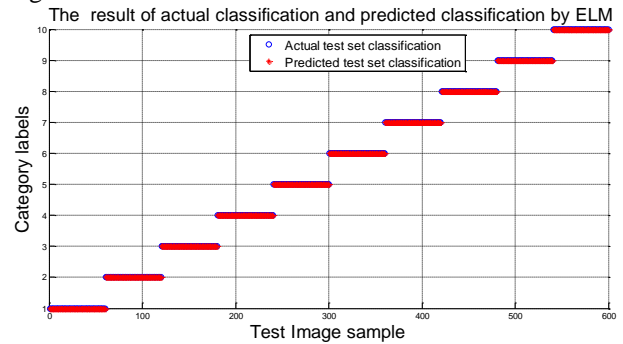


Fig. 9 The result of the actual classification and predicted classification by ELM (elapsed time is 0.431 seconds. accuracy = 100%(600/600))

The number of hidden layer neurons M is the only parameter of ELM. In order to verify the effect of the hidden layer neurons M on the accuracy, there are 600 images are used for training classifier, and the rest 600 images are used for testing. The relationship between the number of hidden nodes and the accuracy is shown in Fig. 10. Accuracy of the ELM algorithm is increased with the increase of the hidden layer neurons M. The prediction accuracy can reach 100% when M is 200. When M more than 200, accuracy is not increased with the number of

hidden neurons. In short, when the number of hidden layer of neurons at 200, the classification result is the best.

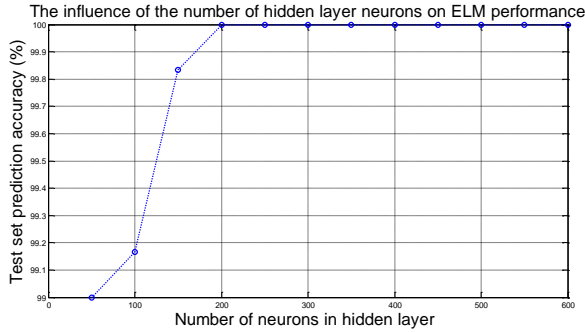


Fig. 10 The influence of the number of hidden layer neurons on ELM performance

V. COMPARISON AND ANALYSIS OF FIVE SUPERVISED LEARNING METHODS

In order to verify the effectiveness of the classification result, BPNN, SVM, PNN, SKohonen and ELM are compared by time and accuracy. Consider the comparison fairness, the experimental image database D, training feature set T and test features set V are the same in the five supervised learning frameworks. The feature extraction process is described in III-A and III-B. The experimental platform includes Intel Core i5 Processor, 8 GB RAM, Windows7 operating system, matlab 2010a test environment.

TABLE III
THE CORRECT CLASSIFICATION NUMBER OF TEN CATEGORIES TRAFFIC SCENE (60 IMAGES EACH CLASS)

Classifier	1	2	3	4	5	6	7	8	9	10	Time/s	Accuracy(%)
BP	49	57	54	54	54	48	60	50	55	44	1.398	87.5% (525/600)
SVM	60	59	55	56	57	50	58	36	60	48	0.334	89.83% (539/600)
PNN	60	59	56	52	53	48	60	59	53	51	3.636	91.8% (551/600)
SKohonen	60	46	57	59	55	17	60	59	60	48	2.137	86.8% (521/600)
ELM	60	60	60	60	60	60	60	60	60	60	0.431	100% (600/600)

The test results are shown in Table III. The conclusions are as follows.

(1) Accuracy is the most important evaluation index for classification algorithms performance. In Table III, the prediction accuracy rate of SVM classifier and BP neural network is similar (87.5%, 89.83%). However, compared to BPNN classifier, SVM are relatively stable and faster.

(2) The predicted accuracy of ELM is slightly higher than other classifier that indicates ELM has better performance in classification.

(3) The running time of ELM and SVM is respectively 0.431s and 0.334s, which indicate the running speed of ELM and SVM are almost the same, and the time is much higher than PNN, SKohonen and BP neural network classifier.

In addition to the correct rate, error rate also can be used to measure the performance of the algorithm in the traffic scene perception. According to Table III, the result of classification error number of 10 categories traffic scenes is shown in Fig.11.

We can conclude that in terms of image algorithms, PNN and ELM are better than other classifiers in accuracy. In terms of classification correct number, there are three categories traffic scenes are below 50. It indicates the classification effect of Skohonen and BP is poor.

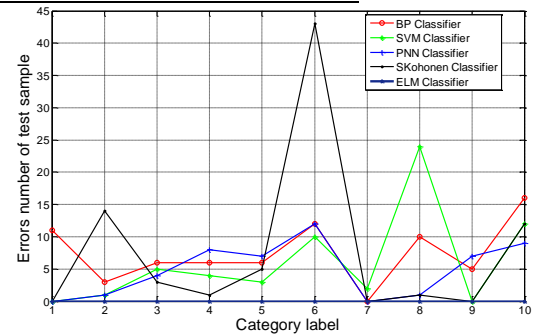


Fig.11 The error number of test samples

The classification correct number of category 6 and category 10 are below 50. This indicate that the extracted features cannot be described the images very well. In Fig. 3, the category 6 and 10 respectively represent overcast and foggy image. The images are blurred and texture features are not obvious. So the image enhancement algorithm can be considered to improve the visibility.

The category 7 represents sunlight image, its classification correct number are above 58 by five classifiers, it indicates that the 8 global underlying visual features can fully describe the images. In summary, ELM classifier has a stable recognition accuracy and performance.

TABLE IV
CONFUSION MATRIX OF TEN CATEGORIES TRAFFIC SCENE BASED ON SVM SUPERVISED LEARNING ALGORITHM

Category	label1	label 2	label 3	label 4	label 5	label 6	label 7	label 8	label 9	label 10	True
label 1	60	0	0	0	0	0	0	0	0	0	60
label 2	0	59	1	0	0	0	0	0	0	0	60
label 3	0	2	55	1	0	0	0	2	0	0	60
label 4	0	0	1	56	0	0	0	3	0	0	60
label 5	0	3	0	0	57	0	0	0	0	0	60
label 6	0	0	0	10	0	50	0	0	0	0	60
label 7	0	0	0	0	0	0	58	0	0	2	60
label 8	0	0	0	24	0	0	0	36	0	0	60
label 9	0	0	0	0	0	0	0	0	60	0	60
label 10	0	0	0	0	0	2	10	0	0	48	60
Predict	60	64	57	91	57	52	68	41	60	50	600

A. Confusion matrix

Confusion matrix is a visualization tool in artificial intelligence field. It is especially suitable for the analysis of supervised learning. The confusion matrix of the rest four supervised learning algorithm is similar with the SVM confusion matrix. Therefore, SVM confusion matrix is discussed here. As shown in Table V. The diagonal data represent the classify correct number of the corresponding category. The last row is SVM classifier prediction results of each category. The last column represents 60 images from each category sample. The conclusions are as follows.

B. Accuracy, Precision and Recall

The system measures the performance can depend on accuracy, precision and recall. The accuracy represents the ratio between the number of correctly classified image and the total number of images. The precision represents the ratio between the number of correctly classify an image and the number of all marked positive. The recall is defined as the rate that the number of a actual positive example is divided into a positive example.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$Recall = \frac{TP}{TP + FP} \quad (13)$$

$$Precision = \frac{TP}{TP + FN} \quad (14)$$

where, TP is the number that positive image is correctly predicted as positive, FP is the number that positive image is predicted as negative, TN is the number that negative image is truly predicted as negative, and FN presents the number that positive image is predicted as negative sample.

From Table IV, we can obtain that category4 and category6, category4 and category8, category7 and category10 are easy to be confused. Because confused category image sample set possess similar characteristics, as shown in Fig. 3. Recall ratio that classify category 6 and category 4 is $50 / (50 + 10) = 83.3\%$. Recall ratio that classify category 8 and category 4 is $36 / (36 + 24) = 60\%$. Recall ratio that classify category 10 and category 7 is $48 / (48 + 10) = 82.8\%$. In order to prove the effectiveness of the five supervised learning algorithm, recall ratio and precision ratio are used to measure the performance of the algorithm. The recall ratio and precision ratio of each category are shown in Table V.

TABLE V

RECALL RATIO AND PRECISION RATIO OF 10 CATEGORIES TRAFFIC SCENE BASED ON FIVE SUPERVISED LEARNING ALGORITHM										
Category	BP		SVM		PNN		SKohonen		ELM	
	Recall ratio(%)	Precision ratio(%)	Recall ratio(%)	Precision ratio(%)	Recall ratio(%)	Precision ratio(%)	Recall ratio(%)	Precision ratio(%)	Recall ratio(%)	Precision ratio(%)
label 1	88.33 (53/60)	98.1 (53/54)	100 (59/60)	100 (59/64)	100 (60/60)	85.7 (60/70)	100 (60/60)	71.42 (60/84)	100 (46/60)	100 (46/49)
label 2	95 (57/60)	85.1 (57/67)	98.3 (59/60)	92.18 (59/64)	98.33 (59/60)	90.76 (59/65)	76.67 (46/60)	93.88 (46/49)	100 (46/60)	100 (46/49)
label 3	90 (54/60)	93.1 (54/58)	91.6 (55/60)	96.5 (55/57)	93.33 (56/60)	87.5 (56/64)	95 (57/60)	78.1 (57/73)	100 (57/73)	100 (57/73)
label 4	90 (54/60)	85.7 (54/63)	93.3 (56/60)	61.5 (56/91)	86.67 (52/60)	98.11 (52/53)	98.33 (59/60)	96.72 (59/61)	100 (59/61)	100 (59/61)
label 5	90 (54/60)	84.4 (54/64)	95 (57/60)	100 (57/57)	88.33 (53/60)	100 (53/53)	91.67 (55/60)	94.82 (55/58)	100 (55/58)	100 (55/58)
label 6	80 (48/60)	97.96 (48/49)	83.3 (50/60)	96.2 (50/52)	80 (48/60)	85.71 (48/56)	28.33 (17/60)	100 (17/17)	100 (17/17)	100 (17/17)
label 7	100 (60/60)	80 (60/75)	96.7 (58/60)	85.3 (58/68)	100 (60/60)	98.36 (60/61)	100 (60/60)	100 (60/60)	100 (60/60)	100 (60/60)
label 8	83.3 (50/60)	92.59 (50/54)	60 (36/60)	87.8 (36/41)	98.33 (59/60)	81.94 (59/72)	98.33 (59/60)	100 (59/59)	100 (59/59)	100 (59/59)
label 9	91.67 (55/60)	78.57 (55/70)	100 (55/70)	100 (55/70)	88.33 (53/60)	100 (53/53)	100 (60/60)	100 (60/60)	100 (60/60)	100 (60/60)
label 10	77.33 (44/60)	59.65 (44/46)	80 (48/60)	96 (48/50)	85 (51/60)	96.22 (51/53)	80 (48/60)	60.76 (48/79)	100 (48/79)	100 (48/79)

VI. CONCLUSIONS

Weather recognition based on road images is a brand-new and challenging subject, which is widely required in many fields. Hence, research of weather recognition based on images is in urgent demand, which can be used to recognize the weather conditions for many vision systems. Classification is a methodology to identify the type of optical characteristics for vision enhancement algorithms to make them more efficient.

In this paper, eight global underlying visual features are extracted and five supervised learning algorithms are used to perceive multi-traffic road scene. Firstly, our method extracts

colour features, texture features and boundary feature which are used to evaluate the image quality. Thus, the extracted features are more comprehensive. Secondly, the ten categories traffic scene image are marked as labels 1-10. Owing to the category label represents the whole image, there is no need to mark the specific area or key point of image. Thirdly, by using of five supervised learning that mentioned in Section IV, we can greatly simplify the manual annotation process of feature sample and improve the classifier efficiency. At last, experiments and comparisons are performed on large datasets to verify the effectiveness of the proposed method in Section V. It proved that the proposed eight features not only can accurately describe image characteristics, but also have strong

robustness and stability at the complex weather environment and the ELM algorithm is superior to other algorithms.

In the future, the proposed algorithms will need to be further verified by the larger image set. Integrated learning is a new paradigm in machine learning field. It is worth to be studied improve the generalization of a machine learning system. And visual image enhancement algorithms in fog and night time applied to general image are worth to be further studied.

VII. REFERENCES

- [1] A. Payne and S. Singh, "Indoor vs. outdoor scene classification in digital photographs," *Pattern Recognition*, vol. 38, no. 10, pp. 1533-1545, Oct 2005.
- [2] C. Lu, D. Lin, J. Jia, and C.-K. Tang, "Two-Class Weather Classification," *IEEE transactions on pattern analysis and machine intelligence*, 2016-Dec-15 2016.
- [3] Y. Lee and G. Kim, "Fog level estimation using non-parametric intensity curves in road environments," *Electron. Lett.*, vol. 53, no. 21, pp. 1404-1406, 2017.
- [4] C. Zheng, F. Zhang, H. Hou, C. Bi, M. Zhang, and B. Zhang, "Active Discriminative Dictionary Learning for Weather Recognition," *Mathematical Problems in Engineering*, 2016 2016, Art. no. 8272859.
- [5] M. Milford, E. Vig, W. Scheirer, and D. Cox, "Vision-based Simultaneous Localization and Mapping in Changing Outdoor Environments," *Journal of Field Robotics*, vol. 31, no. 5, pp. 814-836, Sep-Oct 2014.
- [6] C. Y. Fang, S. W. Chen, and C. S. Fuh, "Automatic change detection of driving environments in a vision-based driver assistance system," *Ieee Transactions on Neural Networks*, vol. 14, no. 3, pp. 646-657, May 2003.
- [7] Y. J. Liu, C. C. Chiu, and J. H. Yang, "A Robust Vision-Based Skyline Detection Algorithm Under Different Weather Conditions," *IEEE Access*, vol. 5, pp. 22992-23009, 2017.
- [8] T. Fu, J. Stipanovic, S. Zangenehpour, L. Miranda-Moreno, and N. Saunier, "Automatic Traffic Data Collection under Varying Lighting and Temperature Conditions in Multimodal Environments: Thermal versus Visible Spectrum Video-Based Systems," *Journal Of Advanced Transportation*, pp. 1-15, 2017 2017, Art. no. Unsp 5142732.
- [9] J. Fritsch, T. Kuehnl, and F. Kummert, "Monocular Road Terrain Detection by Combining Visual and Spatial Information," *Ieee Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1586-1596, Aug 2014.
- [10] K. Wang, Z. Huang, and Z. Zhong, "Simultaneous Multi-vehicle Detection and Tracking Framework with Pavement Constraints Based on Machine Learning and Particle Filter Algorithm," *Chinese Journal of Mechanical Engineering*, vol. 27, no. 6, pp. 1169-1177, Nov 2014.
- [11] R. K. Satzoda and M. M. Trivedi, "Multipart Vehicle Detection Using Symmetry-Derived Analysis and Active Learning," *Ieee Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 926-937, Apr 2016.
- [12] M. Chen, L. Jin, Y. Jiang, L. Gao, F. Wang, and X. Xie, "Study on Leading Vehicle Detection at Night Based on Multisensor and Image Enhancement Method," *Mathematical Problems in Engineering*, 2016 2016, Art. no. 5810910.
- [13] Y. Xu, J. Wen, L. Fei, and Z. Zhang, "Review of Video and Image Defogging Algorithms and Related Studies on Image Restoration and Enhancement," *Ieee Access*, vol. 4, pp. 165-188, 2016 2016.
- [14] R. Gallen, A. Cord, N. Hautière, D. É, and D. Aubert, "Nighttime Visibility Analysis and Estimation Method in the Presence of Dense Fog," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 310-320, 2015.
- [15] D. Gangodkar, P. Kumar, and A. Mittal, "Robust Segmentation of Moving Vehicles Under Complex Outdoor Conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1738-1752, 2012.
- [16] H. Kuang, X. Zhang, Y. J. Li, L. L. H. Chan, and H. Yan, "Nighttime Vehicle Detection Based on Bio-Inspired Image Enhancement and Weighted Score-Level Feature Fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 927-936, 2017.
- [17] Y. Yoo, J. Im, and J. Paik, "Low-Light Image Enhancement Using Adaptive Digital Pixel Binning," *Sensors*, vol. 15, no. 7, pp. 14917-14931, Jul 2015.
- [18] Y. J. Jung, "Enhancement of low light level images using color-plus-mono dual camera," *Opt. Express*, vol. 25, no. 10, pp. 12029-12051, May 15 2017.
- [19] L. Zhou, D.-Y. Bi, and L.-Y. He, "Variational contrast enhancement guided by global and local contrast measurements for single-image defogging," *Journal Of Applied Remote Sensing*, vol. 9, Oct 6 2015, Art. no. 095049.
- [20] Y. Liu, H. Li, and M. Wang, "Single Image Dehazing via Large Sky Region Segmentation and Multiscale Opening Dark Channel Model," *IEEE Access*, vol. 5, pp. 8890-8903, 2017.
- [21] T. Pouli and E. Reinhard, "Progressive color transfer for images of arbitrary dynamic range," *Computers & Graphics-Uk*, vol. 35, no. 1, pp. 67-80, Feb 2011.
- [22] B. Arbelot, R. Vergne, T. Hurtut, and J. Thollot, "Local texture-based color transfer and colorization," *Computers & Graphics-Uk*, vol. 62, pp. 15-27, Feb 2017.
- [23] Y. Xiang, B. Zou, and H. Li, "Selective color transfer with multi-source images," *Pattern Recognition Letters*, vol. 30, no. 7, pp. 682-689, May 1 2009.
- [24] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610-621, 1973.
- [25] O. Regniers, L. Bombrun, V. Lafon, and C. Germain, "Supervised Classification of Very High Resolution Optical Images Using Wavelet-Based Textural Features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3722-3735, 2016.
- [26] G. Tian, H. Zhang, Y. Feng, D. Wang, Y. Peng, and H. Jia, "Green decoration materials selection under interior environment characteristics: A grey-correlation based hybrid MCDM method," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 682-692, 2018.
- [27] X. Han, H. Wang, J. Lu, and C. Zhao, "Road detection based on the fusion of Lidar and image data," *International Journal Of Advanced Robotic Systems*, vol. 14, no. 6, Nov 9 2017.
- [28] H. Zhou, H. Kong, L. Wei, D. Creighton, and S. Nahavandi, "On Detecting Road Regions in a Single UAV Image," *Ieee Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1713-1722, Jul 2017.
- [29] H. R. R. Bakhtiari, A. Abdollahi, and H. Rezaeian, "Semi automatic road extraction from digital images," *Egyptian Journal Of Remote Sensing And Space Sciences*, vol. 20, no. 1, pp. 117-123, Jun 2017.
- [30] S. Chowdhury, B. Verma, and D. Stockwell, "A novel texture feature based multiple classifier technique for roadside vegetation classification," *Expert Systems with Applications*, vol. 42, no. 12, pp. 5047-5055, Jul 15 2015.
- [31] H. Tamura, S. Mori, and T. Yamawaki, "Textural Features Corresponding to Visual Perception," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 6, pp. 460-473, 1978.
- [32] S. Maji, A. C. Berg, and J. Malik, "Efficient Classification for Additive Kernel SVMs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 66-77, 2013.
- [33] J. Wu, "Efficient HIK SVM Learning for Image Classification," *IEEE Transactions on Image Processing*, vol. 21, no. 10, pp. 4442-4453, 2012.
- [34] Z. Yin, B. Wan, F. Yuan, X. Xia, and J. Shi, "A Deep Normalization and Convolutional Neural Network for Image Smoke Detection," *Ieee Access*, vol. 5, pp. 18429-18438, 2017 2017.
- [35] C. Torres-Huitzil and B. Girau, "Fault and Error Tolerance in Neural Networks: A Review," *IEEE Access*, vol. 5, pp. 17322-17341, 2017.
- [36] G. Tian, M. Zhou, and P. Li, "Disassembly Sequence Planning Considering Fuzzy Component Quality and Varying Operational Cost," *IEEE Transactions on Automation Science and Engineering*, vol. PP, no. 99, pp. 1-13, 2017.
- [37] K. Lu, W. Zhang, and B. Sun, "Multidimensional Data-Driven Life Prediction Method for White LEDs Based on BP-NN and Improved-Adaboost Algorithm," *IEEE Access*, vol. 5, pp. 21660-21668, 2017.
- [38] Y. X. Chen and J. Z. Wang, "Image categorization by learning and reasoning with regions," *Journal of Machine Learning Research*, vol. 5, pp. 913-939, Aug 2004.
- [39] Y. Chen, J. Bi, and J. Z. Wang, "MILES: Multiple-Instance Learning via Embedded instance Selection," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1931-1947, Dec 2006.
- [40] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513-529, 2012.
- [41] X. Chen and M. Koskela, "Skeleton-based action recognition with extreme learning machines," *Neurocomputing*, vol. 149, pp. 387-396, Feb 3 2015.
- [42] J. Xin, Z. Wang, L. Qu, and G. Wang, "Elastic extreme learning machine

for big data classification," *Neurocomputing*, vol. 149, pp. 464-471, Feb 3 2015.