Object Classification in Static Images with Cluttered Background using Statistical Feature Based Neural Classifier

[1]B. Nagarajan and [2]P. Balasubramanie

[1]Department of Computer Applications, Bannari Amman Institute of Technology,
Sathyamangalam, Erode District, Tamil Nadu, India

[2]Department of Computer Science and Engineering, Kongu Engineering College,
Perundurai, Erode District, Tamil Nadu, India

Abstract: Object classification in static images is a difficult task since motion information in no longer usable. The challenging task in object classification problem is the removal of cluttered background containing trees, road views, buildings and occlusions. The goal of this study is to build a system that detects and classifies the car objects amidst background clutter and mild occlusion. This study addresses the issues to classify objects of real-world images containing side views of cars with cluttered background with that of non-car images with natural scenes. The threshold technique with background subtraction is used to segment the background region to extract the object of interest. The background segmented image with region of interest is divided into equal sized blocks of sub-images. The statistical features are extracted from each sub-block. The features of the objects are fed to the back-propagation neural classifier. Thus, the performance of the neural classifier is compared with various categories of block size. Quantitative evaluation shows improved results of 83.8%. A critical evaluation of our approach under the proposed standards is presented.

Key words: Object classification, background segmentation, statistical features, neural classifier

INTRODUCTION

Object detection and classification are necessary components in an artificially intelligent autonomous system. Especially, object classification plays a major role in applications such as security systems, traffic surveillance system, target identification, etc. We expect these artificially intelligent autonomous systems to venture onto the street of our world, thus requiring detection and classification of car objects commonly found on the street. In reality, these classification systems face 2 types of problem: objects of same category with large variation in appearance and the objects with different viewing conditions like occlusion, complex background containing buildings, people, trees, road views, etc. This study tries to bring out the importance of the background elimination with statistical based feature extraction method of varying sub-block size for object classification. Since, cdynamic motion information is no longer usable for static images, background elimination becomes a more difficult task. Thus, background removed and statistical features of squared sub-blocks of the images are fed to the neural classifier. The objects of interest being a car and non-car images are classified.

Image understanding is a major area where researchers design computational systems that can identify and classify objects automatically. Identification and classification of vehicles has been a focus of investigation over last decades (Hsieh *et al*., 2006; Shan *et al*., 2005; Sun *et al*., 2006). A new approach to object detection that makes use of a sparse, part-based representation is proposed by Agarwal *et al*. (2004). This study gives very promising results in the detection of vehicles from a group of non-vehicle category of natural scenes. Nagarajan and Balasubramanie (2007) have proposed their research based on wavelet and moment invariant features towards object classification with cluttered background. Said *et al*. (2000, 2001) has presented their research on texture features that automatically distinguishes between face objects. The statistical features along with texture parameters form the basis for facial recognition. Papageorgiou and Poggio (1999) utilized appropriate global statistical features for classification to detect the car objects. The advantage of such approach is that it has some self-learning ability.
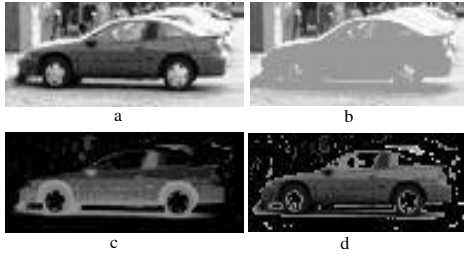
Corresponding Author: B. Nagarajan, Department of Computer Applications, Bannari Amman Institute of Technology,
Erode District, Sathyamangalam-638 401, Tamil Nadu, India

Fig. 1: a) A sample occluded image with natural background denoted as I (x, y). b) The small regions are removed by filling the holes. c) Image difference obtained by subtraction (a) by (b) denoted as d (x, y). d) Image obtained by mapping function f (x, y)

BACKGROUND REMOVAL AND MAPPING FUNCTION

The overall complexity increases for the natural images as the object of interest is lying on the background region. In object classification problem, it is essential to distinguish the object of interest and the background. Segmentation of object is done through background subtraction technique. This method is more suitable when the intensity levels of the objects fall outside the range of levels in the background.

An object with natural background is shown in Fig. 1. Initially morphological operations are applied to suppress the residual errors with help of open and close pair statements (Li *et al*., 2004; Richord *et al*., 2005). The small regions are removed by filling the holes. Then, the image subtraction is applied with the previous result. Thus, the object is segmented from the background.

A mapping function (1) is used to restore the object of interest from that of the subtracted image.

$$f(x, y) = \begin{cases} 0, \text{ if } d(x, y) = 0 \\ I(x, y), \text{ Otherwise} \end{cases} \quad (1)$$

where, f (x, y) is the transformed image, d (x, y) is image difference after fill operation and I (x, y) is the original image.

STATISTICAL FEATURES

Statistical functions such as mean, median, standard deviation and higher moments are most common to characterize data set, which have been used as pattern features in many applications (Papageorgiou and Poggio, 1999; Said *et al*., 2000, 2001). Higher moments can be used to classify the actual shape of the distribution function. Depend on the statistical analysis of the vector, we extract 6 features as follows:

Feature (1): Mean or the average, which is defined as:

$$\overline{x} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i \quad (2)$$

where, $\overline{x}$ is the mean value and N is the total number of data values, x_i.

Feature (2): The standard deviation: The best known measure of the spread of the distribution is the simple variance and it is defined as:

$$var \equiv \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2 \quad (3)$$

The standard deviation is a well-known measure of deviation from its mean value and is defined as the square root of the variance.

$$\sigma \equiv \sqrt{var} \quad (4)$$

Feature (3): Smoothness is measured with its second moment as:

$$smoothness = 1 - \frac{1}{(1 + var)} \quad (5)$$

Feature (4): The skewness, or third moment, is a measure of asymmetry of distribution defined as:

$$skew \equiv \frac{1}{N} \sum_{i=1}^{N} \left[\frac{x_i - \overline{x}}{\sigma} \right]^3 \quad (6)$$

Feature (5): Energy is calculated as:

$$energy = \sum_{i} \sum_{j} P_d^2(i, j) \quad (7)$$

The gray level co-occurrence matrix P_d for a displacement vector d= (d_x, d_y) is defined in (8). The entry (I, j) for P_d is the number of occurrences of the pair of gray levels i, j which are a distance d apart.

$$P_d(i, j) = \left| \left\{ \left((r, s), (t, v) \right) : I(r, s) = i, I(t, v) = j \right\} \right| \quad (8)$$

Feature (6): Entropy is calculated as:

$$\text{entropy} = -\sum_i^i \sum_j^j P_d(i, j) \log P_d(i, j) \qquad (9)$$

It should be noted that spatial gray level co-occurrence estimated image properties are related to the second order statistics of image.

Thus 6 statistical measures of texture are calculated for every sub-block of an image. The feature vector is populated with multiples of 6 with that of number of sub-blocks in an image.

BUILDING A NEURAL CLASSIFIER

A binary Artificial Neural Network (ANN) classifier is built with back-propagation algorithm that learns to classify an image as a member or nonmember of a class.

The number of input layer nodes is equal to the dimension of the feature space obtained from the moments and invariant features. The number of output nodes is usually determined by the application (Khotanzand and Chung, 1998) which is 1 (either Yes/ No) where, a threshold value nearer to 1 represents Yes and a value nearer to 0 represents No. The neural classifier is trained with different choices for the number of hidden layer. The final architecture is chosen with single hidden layer shown in Fig. 2 that results with better performance.

The connections carry the outputs of a layer to the input of the next layer have a weight associated with them. The node outputs are multiplied by these weights before reaching the inputs of the next layer. The output neuron (10) will be representing the existence of a particular class of object.

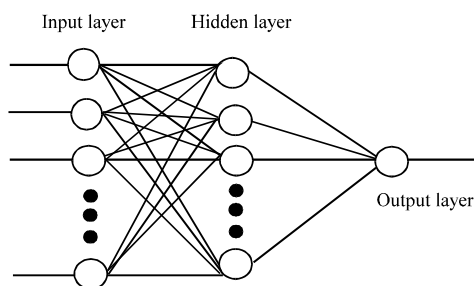$$O_j^l(k) = f\left(\sum_{m=0}^{Nl-1} w_{jm}^l O_m^{l-1}\right) \qquad (10)$$

Fig. 2: The 3 layer neural architecture

PROPOSED WORK

This study addresses the issues to classify objects of real-world images containing side views of cars amidst background clutter and mild occlusion. The objects of interest to be classified are car (positive) and non-car (negative) images taken from University of Illinois at Urbana-Champaign (UIUC) standard database. The image data set consists of 1000 real images for training and testing having 500 in each class. The sizes of the images are uniform with the dimension 100×40 pixels.

The proposed framework consists of 3 methods followed by background removal as given in section Method-I: 10 Blocks of size 20×20 each, Method-II: 40 Blocks of size 10×10 each Method-III: 160 Blocks of size 5×5 each. Six statistical features are calculated from each single block of sub-image using equations mentioned in section-II. Data normalization is applied for the statistical features, which are the deviated from its mean by standard deviation. This process improves the performance of the neural classifier. The overall flow of the framework is shown in Fig. 3.
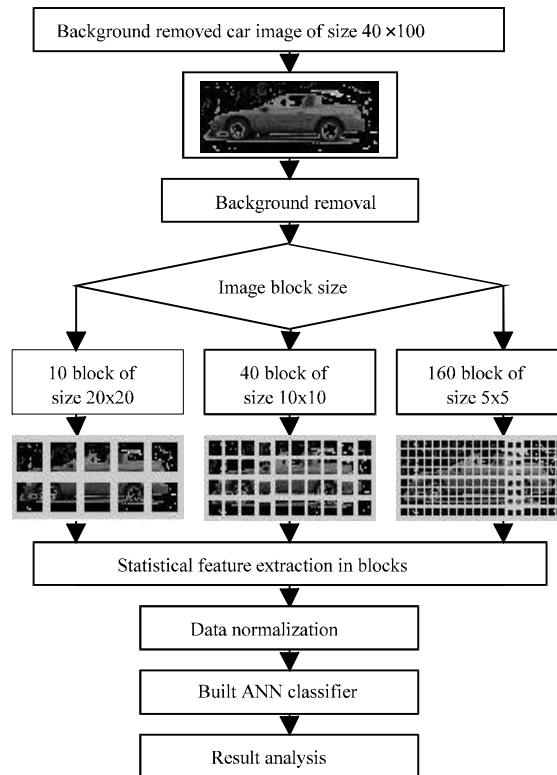
Fig. 3: The description of the proposed work

IMPLEMENTATIONS

We trained our methods with different kinds of cars against a variety of background, partially occluded cars of positive class. The negative training examples include images of natural scenes, buildings and road views. The training is done with 400 images (200 positive and 200 negative) against all the methods. The testing of images are done with 1000 images (500 positive and 500 negative) taken from the same image database.

The feed-forward network for learning is done for 10 blocks of size 20×20 namely method-I, 40 blocks of size 10×10 namely method-II and 160 blocks of size 5×5 namely method-III, respectively. The input nodes for method-I is 60 (10 blocks ×6 features), method-II is 240 (40 blocks ×6 features) and method-III is 960 (160 blocks ×6 features), respectively. Optimal structure validation is done and the structure given below performs well and leads to better results. Thus the optimal structure (Fig. 2) of the neural
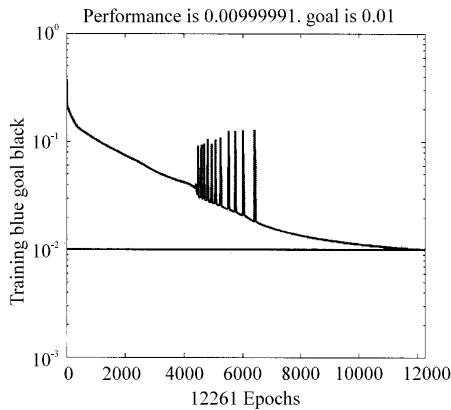
Performance is 0.00999991. goal is 0.01

Fig. 4: The performance graph of neural network training for Method-I: 10 Blocks of size 20×20

Performance is 0.00999764. goal is 0.01
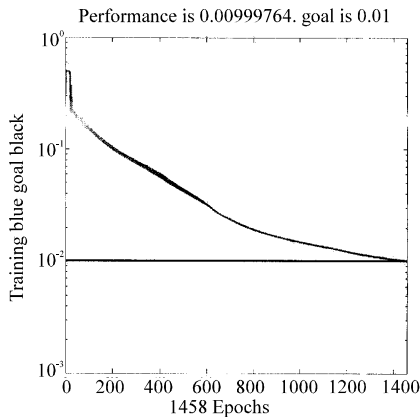
Fig. 5: The performance graph of neural network training for Method-II: 40 Blocks of size 10×10

Performance is 0.00996651. goal is 0.01
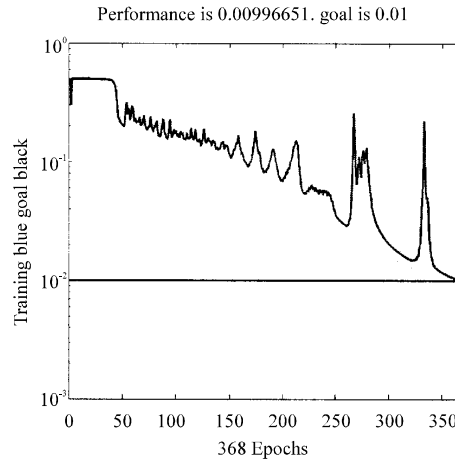
Fig. 6: The performance graph of neural network training for Method-III: 160 Blocks of size 5×5

Table 1: Parameters for training of the neural classifier

Parameters	Method-I	Method- II	Method-III
Learning Rate	0.5	0.5	0.5
Performance Goal	0.01	0.01	0.01
No. of Epochs taken to meet the performance goal	12261	1458	368
Time taken to learn	141.26 Secs	20.64 Secs	24.20 Secs

classifier for method-I is 60-20-1, method-II is 240-10-1 and method-III is 960-25-1, respectively.

The various parameters for the neural classifier training for all the methods are given in Table 1. The Performance graph of the neural classifier for method-I method-II and method-III are shown in Fig. 4-6, respectively.

DISCUSSION

In object classification problem, the four quantities of results category are given.

True Positive (TP): Classify a car image into class of cars.

True Negative (TN): Misclassify a car image into class of Non-cars.

False Positive (FP): Classify a non-car image into class of non-cars.

False Negative (FN): Misclassify a non-car image into class of cars.

The objective of any classification is to maximize the number of correct classification denoted by True Positive Rate (TPR) and False Positive Rate (FPR) where by minimizing the wrong classification denoted by True Negative Rate (TNR) and False Negative Rate (FNR).

Fig. 7: Sample results of the neural classifier of the category car images with cluttered background and mild occlusion

Fig. 8: Sample results of the neural classifier of the category non-car images containing trees, road view, bike, wall, buildings and persons

Table 2: Comparison of experimental methods

Threshold for Classification: 0.7	Classifying positive images (car images)		Classifying negative images (non-car images)	
	TPR	TNR	FPR	FNR
Method-I	78.6%	21.4%	84.4%	15.2%
10 Blocks of size 20 × 20	**Method-I** Overall Classification Accuracy (TPR+FPR)/2 is 81.5%			
Method-II	78.4%	21.6%	88.4%	11.6%
40 Blocks of size 10 × 10	**Method-II** Overall Classification Accuracy (TPR+FPR)/2 is 83.4%			
Method-III	74.8%	25.2%	92.8%	7.2%
160 Blocks of size 5 × 5	**Method-III** Overall Classification Accuracy (TPR+FPR)/2 is 83.8%			

$$TPR = \frac{\text{Number of true positive} \ (TP)}{\text{Total number of positive in data set} \ (nP)}$$

$$TNR = \frac{\text{Number of true negative} \ (TN)}{\text{Total number of negative in data set} \ (nN)}$$

$$FPR = \frac{\text{Number of false positive} \ (FP)}{\text{Total number of positive in data set} \ (nP)}$$

$$FNR = \frac{\text{Number of false negative} \ (FN)}{\text{Total number of negative in data set} \ (nF)}$$

The values of nP and nN used as testing samples are 500 and 500, respectively. Most classification algorithm includes a threshold parameter for classification accuracy which can be varied to lie at different trade-off points between correct and false classification. The comparison of results of the proposed methods is shown in Table 2 which is obtained with an activation threshold value of 0.7.

Classified images of category car and non-car as resultant sample images are shown below in the Fig. 7 and 8, respectively.

It is evident that the classifier with 160 blocks of size 5×5 (Method-III) is showing improved overall results of 83.8% of classification accuracy comparatively with that of 40 blocks of size 10×10 (Method-II) and 10 blocks of size 20×20 (Method-I). Method-II is also comparatively, good in classification with the accuracy of 83.4%.

CONCLUSION

Thus an attempt is made to build a system that classifies the objects amidst background clutter and mild occlusion is achieved to certain extent. Thus the goal to classify objects of real-world images containing side views of cars with cluttered background with that of non-car images with natural scenes is presented. Comparing the results in Table 2, the performance of the proposed method with 160 blocks of size 5×5 with statistical features after background removal gives a satisfactory classification rate of 83.8%. There by the average classification accuracy is slightly improved by 4.2% compared with the wavelet based method proposed by Nagarajan and Balasubramanie (2007). The limitation of this method is the object with a high degree of occlusion for classification. Further work extension can be made to improve the performance of the classifier system with the inclusion of feature selection process. This complete research is implemented using neural network and image processing toolbox of Matlab 6.5.

REFERENCES

Agarwal, S., A. Awan and D. Roth, 2004. Learning to detect objects in images via a sparse, part-based representation. IEEE. Trans. Pattern Anal. Machine Intell., 26 (11): 1475-1490.

Hsieh, J.W. *et al*., 2006. Automatic traffic surveillance system for vehicle tracking and classification. IEEE. Trans. Intell. Trans. Syst., 7 (2): 175-187.

Khotanzand, A. and C. Chung, 1998. Application of multi-layer perceptron neural networks to vision problem. Neural computing and applications. Springer-Verlag London Limited, pp: 249-259.

Li, L. *et al*., 2004. Statistical modeling of complex backgrounds for foreground object detection. IEEE. Trans. Image Proc., 13 (11): 1459-1472.

Nagarajan, B. and P. Balasubramanie, 2007. Wavelet feature based neural classifier system for object classification with complex background. Iccima 07. IEEE Comp. Soc. Press, 1: 302-307.

Papageorgiou, C.P. and T. Poggio, 1999. A trainable object detection system: Car detection in static images. A.I. Memo No. 1673, Massachusetts Institute of Technology: MIT.

Said, E.E. *et al*., 2000. Neural network face recognition using statistical feature extraction. 17th national radio science conference. Minufiya University, Egypt, C31: 1-8.

Sun, Z. *et al*., 2006. Monocular precrash vehicle detection: Features and classifiers. IEEE. Trans. Image Proc., 15: 2019-2034.

Richord, J.R. *et al*., 2005. Image change detection algorithms: A systematic survey. IEEE. Trans. Image Proc., 14 (3): 294-306.

Said, E.E. *et al*., 2001. Neural network face recognition using statistical feature and skin texture parameters. 18th National Radio Science Conference. Mansoura University, Egypt, C5: 1-8.

Shan, Y. *et al*., 2005. Vehicle Identification between non-overlapping cameras without direct feature Matching. 10th IEEE. Int. Conf. Comp. Vision, pp: 1550-1558.