Korean Journal of Remote Sensing, Vol.24, No.5, 2008, pp.473~481

The Application of BP and RBF Neural Network Methods on Vehicle Detection in Aerial Imagery

Jae-Young Choi, Hyoung-Jong Jang, and Young-Kyu Yang†

College of IT, Kyungwon University, Seongnam-si, Gyeonggi-do, 461-701, KOREA

Abstract : This paper presents an approach to Back-propagation and Radial Basis Function neural network method with various training set for automatic vehicle detection from aerial images. The initial extraction of candidate object is based on Mean-shift algorithm with symmetric property of a vehicle structure. By fusing the density and the symmetry, the method can remove the ambiguous objects and reduce the cost of processing in the next stage. To extract features from the detected object, we describe the object as a log-polar shape histogram using edge strengths of object and represent the orientation and distance from its center. The spatial histogram is used for calculating the momentum of object and compensating the direction of object. BPNN and RBFNN are applied to verify the object as a vehicle using a variety of non-car training sets. The proposed algorithm shows the results which are according to the training data. By comparing the training sets, advantages and disadvantages of them have been discussed.

Key Words : *Vehicle detection, Aerial imagery, Mean shift, Shape description, Back propagation, Radial basis function.*

1. Introduction

Interesting of the object detection has been increased in the computer vision area because this item is such a significant part of our life. Especially, Detection and extraction of vehicle objects in high resolution satellite imagery are required in many transportation applications.

However, it is still not as easy as it may seem to be. Numerous imaging conditions can affect the performance of a technique, including noise, illumination, and size of vehicle in the input scene.

This paper deals with automatic detection of vehicle in 0.25 meter resolution aerial imagery. Mean-shift clustering algorithm extracts the candidate region with symmetric characteristic of car shape. By using both geometric and radiometric properties, the ambiguous regions are eliminated from the candidate list and the cost of processing for detected region of interest (ROI) is reduced in the subsequent step. To make the features from the ROI, we obtain the edge strengths and represent it as a spatial histogram using the orientation and distance between edge and the center of region. It is called log-polar shape descriptor. It consists of the 12 angles and 5 distances. The maximum strengths of direction and phase

Received October 4, 2008; Revised October 12, 2008; Accepted October 19, 2008.
† Corresponding Author: Young-Kyu Yang (ykyang@kyungwon.ac.kr)

symmetry information help to compensate the orientation of the shape context. In our last stage, Back-propagation (BP) neural network and Radial Basis Function (RBF) neural network are used for verifying the object as a vehicle. This paper uses a variety of training sets in order to evaluate the results which depend on the non-vehicle training set. We believe that the study is useful to adapt neural network to detection of vehicle.

The remainder of this paper is organized as follows. We first introduce related work on vehicle detection in Section 2. Section 3 describes the proposed candidate object extraction approach. In this section, we present the region clustering method using mean-shift and measurement of phase symmetry. Section 4 shows the shape description and compensation of detected object. In section 5 we evaluate a variety of training data for BP and RBF neural networks. The experiment results are given in Section 6 and concluding remarks are made in Section 7.

2. Related Works

Object detection become challenging due to time-consuming manual detection. Several papers suggested the use of feature vectors from image region. Lowe's Scale Invariant Feature Transform (SIFT) algorithm is a popular algorithm for extracting the salient features from object (Lowe, 2004). Invariant features are robust to image scale, rotation, and partially invariant to changing viewpoints, and change in illumination (Bay *et al*, 2006). However, if the object or region is too small to detect, it cannot be known a priori if it will be represented by any key point.

Generalized Hough Transform (GHT) is used for extract road side. A certain rectangular band width is examined in search position where GHT response is high. The number of edge pixels inside the search band is taken to be the likelihood of the presence of a vehicle in that area. They use the mask which consists of four elongated operators with first-derivative of Gaussian cross sections in recent work (Chellappa *et al*, 1998; Moon *et al*, 2002).

Hinz introduced Model based approach (Hinz *et al*, 2003). The method uses an explicit model that consists mainly of geometric features and radiometric properties. The vehicle is modeled as 3D object by a wire frame representation. They also use vehicle queue model to find stands on a road. In this case, detection relies on matching the model. If there is sufficient support of the model in the image, a vehicle is assumed to be detected.

Neural network or Support Vector Machine (SVM) algorithm is applied to detect the vehicle from low resolution image (Zhao *et al*, 2003). In this work, they only address detection of cars aligned with road direction. This approach may be sufficient for their data.

Above approaches have their advantages and disadvantages. Most of the previous work regards a vehicle as a 2D pattern. Vehicle detection in aerial images is relatively constrained by the resolution. Therefore, we need a variety of information of vehicle in order to extract outstanding characteristics.

3. Candidate Object Extraction

This section describes a proposed approach to detect the candidate object which is similar with car structure. With respect to the detection of a vehicle, we can consider following conditions. One is that car has symmetric structure as geometric aspect and another is that size of car is approximately constant length in same resolution.

Proposed method measures the pixels in an input aerial image, and clusters the pixels which consist of car-structure as a geometric aspect and have a similar value of car as a radiometric point of view. After that, we make the shape description using the log-polar histogram. From the symmetric information and strength of gradient, we can estimate the momentum of object and compensate its direction. That is, the histogram for shape description is shifted in order to rotate the object. Finally, we evaluate the shape description to verify the shape of object as a vehicle using Back-propagation and Radial basis function with a variety of training data as shown in Fig. 1.

Mean-shift clustering algorithm is designed to find modes, centers of the regions of high concentration, of data represented as arbitrary dimensional vectors. The major steps in the computation of the algorithm as follows (Fukunaga, 1975; Comaniciu, 1997). First, choose the radius r of the search window. Second, choose the initial location of the window. Third, compute the mean shift vector and translate the search window by that amount. Finally, Repeat till convergence.

The mean-shift vector is described in Eq. (1).

If y_j is instead of x, and $\{y_j\}_{j=1,2,..}$ denotes the

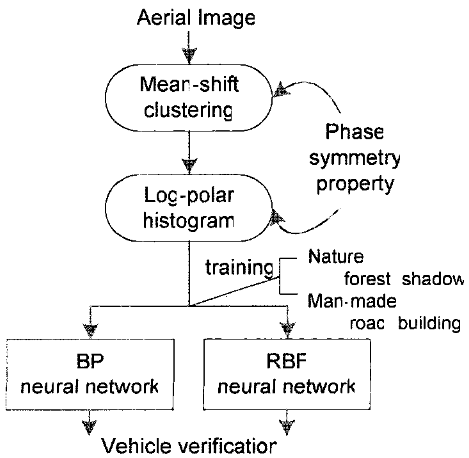


Fig. 1. Flow chart of the suggested algorithm.

sequence of successive locations of the kernel $G(x)$, the equation can be the weighted mean at y_j.

$$m(x) = \left[\sum_{i=1}^{n} x_i g\left(\frac{\|x - x_i\|^2}{h}\right)\right] \Big/ \left[\sum_{i=1}^{n} g\left(\frac{\|x - x_i\|^2}{h}\right)\right] - x \quad (1)$$

In case of the color image clustering like our application, the RGB color image is mapped into the $L^*u^*v^*$ color space model. The mean-shift method clusters this multi-dimensional data set by associating each point to a peak of the data set's probability density.

In our approach, we use geometric (symmetric structure of car shape) property as well as radiometric (intensity of pixel level) characteristics. During the probability density function estimates a density in color space, each point which is peak of data set is examined as local maximum of phase symmetry.

Symmetry is an important property we identify the structure of objects. In this work, the Wavelet Transform is used to obtain local frequency information. Because we are interested in phase information in signals, the method uses wavelets based on complex valued Gabor functions to modulate sine and cosine waves by Gaussian (Kovesi, 1997). Let I denote the signal and M_n^e and M_n^o denote the even-symmetric (cosine) and odd-symmetric (sine) wavelets at a scale n, then the responses of each quadrature pair of filters as in Eq. (2).

$$[e_n(x), o_n(x)] = [I(x)*M_n^e, I(x)*M_n^o] \quad (2)$$

At each point x in a signal, the response vector for each scale of filter will be obtained. For example, at a point of symmetry, the absolute value of the even-symmetric filter output will be large and another output will be small. This produces a function that varies between ± 1 and varies linearly with phase deviation. In case of multiple scales, the difference of the values of the even and odd filter responses at each scale is weighted by the magnitude of the filter response vector at each scale n. This equation is shown in Eq. (3). The amplitude of the transform at a given wavelet scale is $A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2}$ and the phase is $\phi_n(x) = atan2(e_n(x), o_n(x))$.

$$Sym(x) = \frac{\sum_n \lfloor A_n(x)[\,|\cos(\phi_n(x))| - |\sin(\phi_n(x))|\,] - T \rfloor}{\sum_n A_n(x) + \varepsilon} = \frac{\sum_n \lfloor [\,|e_n(x)| - |o_n(x)|\,] - T \rfloor}{\sum_n A_n(x) + \varepsilon} \quad (3)$$

The factor T is a noise compensation term representing the maximum response generated from noise in signal. This 1D equation can extend to 2D by applying the 1D analysis in multiple orientations and forming a weighted sum of the result. This can be used as a line and blob detector. Phase symmetry is an illumination and contrast invariant measure of symmetry in an image.

Fig. 2 illustrates process of candidate object (region) extraction. As geometric property of vehicles, there is a bilateral symmetry even though a vehicle has its shadow. Therefore, we calculate the phase symmetry value and apply it to the mean shift process to detect object (or ROI) which has car-like structure.

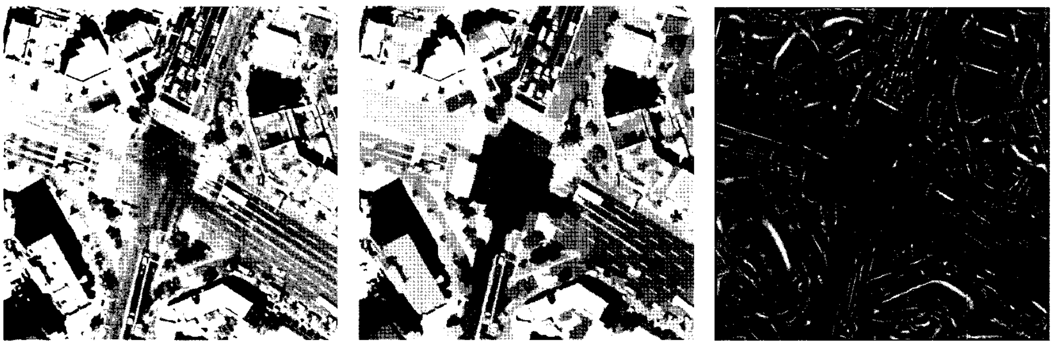
The advantage of using above fusion method is that it is able to eliminate the ambiguous blobs. As a result of this, the cost of next step is also reduced because it does not need to check whole blobs in order to verify the vehicle.

4. Shape Description and Compensation

In order to describe the shape of detected object (or region) as vehicle, we treat a region as a point set, and assume that the shape of an object is represented by discrete pixel set which has the information such as geometric relationship between pixels and the different value among its adjacency pixels.

A conventional approach of matching between a given model and target object is an exhaustive search on whole region or histograms for every possible points. In case the search should be done at different orientations, the whole process should be repeated as many times as the number of directions.

In this section, we propose the shape description that is flexible in rotation of object. As mentioned in the beginning of Section 3, the size of a vehicle is approximately constant length and the shape of one is almost uniform. Therefore we can estimate the geometric shape of vehicle by measuring the distance and orientation between the center of object and its surrounding edges.



(a) (b) (c)

Fig. 2. Region extraction. (a) Input aerial image, (b) Clustering of blob using Mean-shift alfoirthm, (c) Phase symmetry output

Consider an edge of image $I(x,y)$ with an object and its neighborhood (the size of search mask is 25×25 and the center of mask is located at a centroid of region in this experiment), the orientation θ is defined by $\theta(x,y) = \arctan\lfloor d_y(x,y)/d_x(x,y)\rfloor$ where d_x, d_y are the distance from the center of region. Distance r is also obtained by $r = \sqrt{d_x^2 + d_y^2}$.

We use a log-polar coordinate system which has five bins for log r and twelve bins for θ because the descriptor should be more sensitive to differences in nearby pixels in Cartesian spaces (Belongie *et al*, 2002). Fig. 3(c) depicts the diagram of log-polar histogram bins. Each bins of log-polar diagram is represented as a 5×12 arrays as shown in Fig. 3(d). For normalizing the histogram H, each accumulated bin $H(i)$ is divided by the number of detected pixel in each sector.

The direction of object can be estimated from the log-polar histogram information treated in above and will be used for compensating the orientation of the shape context. In our approach, we rotate the inspect region which contains a object and its neighborhood by the value of compensated orientation. First, we accumulate each distance along with same direction (one of twelve directions). Second, an accumulated value competes with another in order to get the strength of each direction. The maximum strength direction will be symmetry axis. After the symmetry axis of the inspect region is obtained, its shape context bin is shifted from direction of symmetry to the vertical axis by the difference of angle.

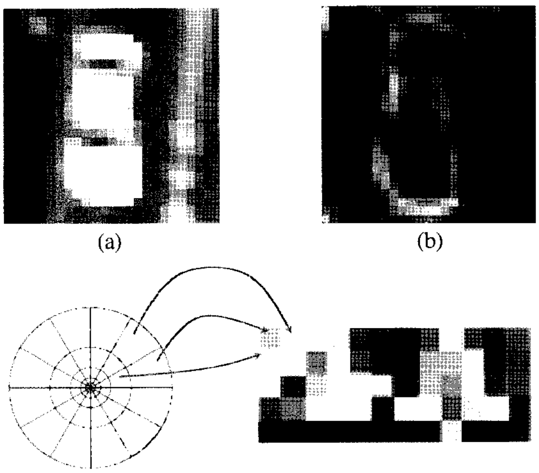

(a) (b) (c) (d)

Fig. 3. Shape description. (a) Input region for shape description, (b) Gradient image, (c) Diagram of log-polar histogram, (d) Shape context represented 5×12 bins.

A compensated shape context bins have same direction with reference one, and are evaluated by the Back-propagation neural network and Radial Basis Function neural network in next stage.

5. Neural Networks

Neural network are adaptive networks which are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems.

Back-propagation is a common method of teaching artificial neural networks how to perform a given task. It is a supervised learning method, and is an implementation of the Delta rule. It requires a teacher that knows, or can calculate, the desired output for any given input. It is most useful for feed-forward networks. Back-propagation requires that the activation function used by the artificial neurons is differentiable. The errors propagate backwards from the output nodes to the inner nodes like the name of algorithm. Back-propagation is used to calculate the gradient of the error of network with respect to the network's modifiable weights as shown in Fig. 4 (Demuth, 1998).

This paper uses 3 layer perceptrons with one input, one hidden, and one output layer. Hidden layer has the same neurons (60 neurons are used in this experiment) to input layer and uses the tan-sigmoid transfer function. Output layer has one neuron and uses the log-sigmoid transfer function. Training

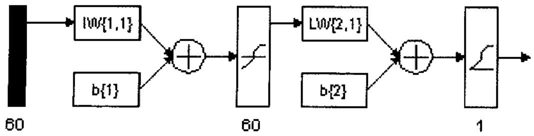


Fig. 4. Architecture of BP neural network used in experiment.

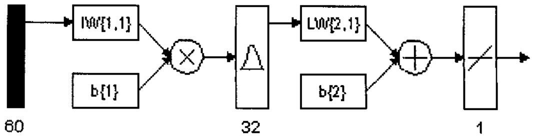


Fig. 5. RBF architecture used in experiment.

function is Levenberg-Marquardt algorithm. Adaptation learning function is gradient descent with momentum weight/bias learning function. Performance function is a mean squared error by Eq. (4).

$$MSE=\frac{1}{mp}\sum_{p=1}^{p}\sum_{m=1}^{m}(\hat{y}_{pj}-y_{pj})^2 \qquad (4)$$

Where m is the number of neurons in output layer and p represents the number of training samples. $\hat{y}_{pj}$ is the expected output value of the network and y_{pj} denotes the actual output value of the network.

Radial Basis Function networks represent a special category of the feed-forward neural networks architecture (Park, 1991). The RBF structure consists of an input layer, a single hidden layer with a radial activation function and an output layer as illustrated in Fig. 5. RBF is a real-valued function whose value depends only on the distance from some other point c, called a center. Any function φ that satisfies the property $\varphi(x)=\varphi(\|x\|)$ is a radial function. The norm is usually Euclidean distance. The transfer function for a radial basis neuron is $\exp(-n^2)$

The standard technique used to train an RBF networks the hybrid approach, which is a two-stage learning strategy. First, an unsupervised clustering algorithm is used to extract the parameters of the radial basis functions, namely the width and the centers. This is followed by the computation of the weights of the connections between the output nodes and the kernel functions using a supervised least mean square algorithm. To tune even further the parameters of the networks, a supervised gradient based algorithm is later applied and it makes use of some of the training patterns presented to the network.

For training both BP and RBF neural network, a large number of training samples are needed. Besides, training a neural network for vehicle detection is very challenging due to the difficulty in collecting typical non-vehicle samples.

Our non-vehicle training samples are divided into four classes - the forest or trees, the shadow between buildings or vehicles, the building or roof of building, and the road or traffic lane. The two former classes can represent nature, whereas the others can be obtained from man-made structure as shown in Fig. 6. We make it as a log-polar histogram for 60 shape description features, and training these features using BP and RBF.

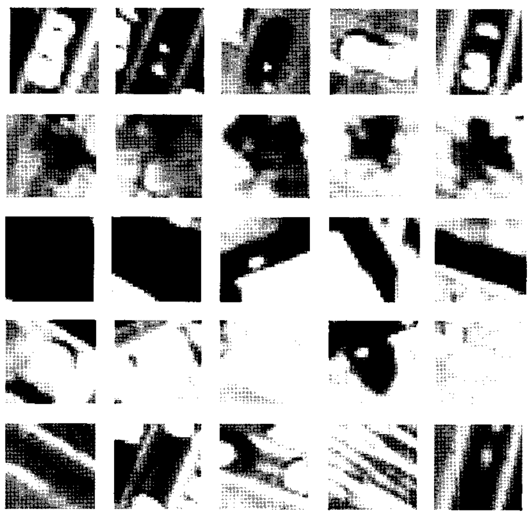


Fig. 6. Training classes (from top row to bottom row; vehicle, Forest, Shadow, Building, Road).

6. Experimental Results

The implemented method is intended to evaluate the rate of vehicle detection according to a variety of training data set in BP and RBF. The input aerial image is 512×512 sub-images from the whole aerial image and has a 0.25 meter resolution. The experiment is performed on Matlab 7.0 platform. In the training or test image, a typical vehicle (medium sedan) is around 15 to 25 pixels in length and around 10 to 12 pixels in width. The evaluation does not refer to real ground truth. That is why we marked all cars in each sub-images by hand.

After extraction of candidate object, false negative alarm (missing detection) occurs in regions where the road or parking lot is darkened by building or tree shadows. Large vehicles are also missing because our method focused on the typical vehicles and examined around 25×25 pixels neighborhood.

In addition to above environment in input image, the detection rate is relative to the setting of parameters during mean-shift algorithm processing. Over-segments cause false negative alarms because the different part of vehicle such as hood, roof, trunk, windshield and shadow is not merged. For these reasons, a small blob is eliminated as a noise or an isolated blob.

For addressing these problems, the suggested method has to be extended by considering the global information in order not to damage a component of a vehicle structure. The more detailed descriptions are used, the more number of features and conditions are needed to cover all types of vehicles.

We can observe from Fig. 7 that ROIs extract several object types even though the system tried to detect object looks like vehicle structure. Most of them are forest (trees), shadow, building, and road (or traffic lane).

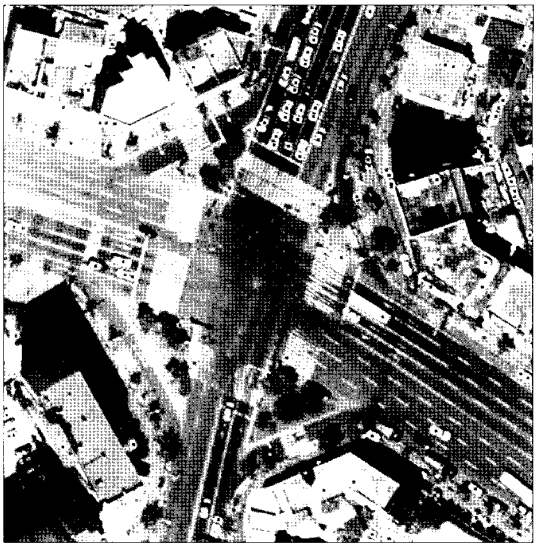


Fig. 7. Extraction of ROI (red dot indicates the center of extracted region).

The experimental results of BP and RBF are summarized in Table 1 and Table 2, respectively. There is always a tradeoff between false alarm rate and mis-detection rate.

The numerical evaluation is performed by the specificity(SP), sensitivity(SE), and accuracy using true positive(TP), true negative(TN), false positive (FP), and false negative(FN) as described in Eq. (5)~(7).

$$SP(specificity) = \frac{TP}{TP + FN} \tag{5}$$

$$SE(sensitivity) = \frac{TN}{TN + FP} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Table 1 displays comparative results. In case of

Table 1. Evaluation of test images using BP

Training data for non-vehicle		SP(%)	SE(%)	Accuracy (%)
Nature	Forest	82.6	58.3	70.2
	Shadow	78.3	58.3	68.1
Man-made	Building	78.3	54.2	66.0
	Road	95.7	54.2	74.5
Mixture		69.6	66.7	68.1

Table 2. Evaluation of test images using RBF

Training data For non-vehicle		SP(%)	SE(%)	Accuracy (%)
Nature	Forest	82.6	16.7	48.9
	Shadow	86.9	70.8	78.7
Man-made	Building	78.3	66.7	72.3
	Road	86.9	58.3	72.3
Mixture		69.6	66.7	68.1

man-made training set, the SP is 78.3% with 74.5% accuracy, whereas the nature case has 69.6% SP and 59.6% accuracy. We can observe from this fact that a training set consists of man-made object in BP algorithm reports higher accuracy rate than one of nature object.

Since we use all types of object for test set in contrast with training set which uses only one of four types, it is no wonder SE is low.

On testing the four classes, SP and accuracy is high when roads compete with vehicle. The reason is that the weight of neuron is well defined through competition of the similar log-polar histograms.

Experimental results using RBF are reported that a man-made training set has 78.3% SP and 76.6% accuracy while nature one has 74.0% SP and 55.3% accuracy as shown in Table 2.

The man-made training method is also higher than nature one like BP case.

The learning duration of RBF neural network is shorter than BP case in spite of the performance has a little bit difference. Therefore if we would use a huge data, the use of RBF instead of BP neural network would be more useful to overcome faster the complexity of the condition.

7. Conclusion

The result of whole system relies on the training set in learning step. Therefore we intend to evaluate how different training set affects the detection performance. In order to accomplish comparison of training set, a hybrid approach for vehicle detection from aerial image is introduced in this paper.

For extraction of the initial candidate vehicle, Mean-shift clustering algorithm is used for detecting the dense and symmetric region. Fusing geometric and radiometric characteristics helps to reject ambiguous blobs and saves the cost of subsequent processing. To make the features from extracted object (or ROI), we apply the log-polar shape descriptor to estimate the direction of object. The log-polar method is able to avoid the sensitivity of differences in nearby pixels. From the measurement of the log-polar histogram, we can compensate the angle of object by using the presumptive direction. Their detection is very sensitive to the surrounding context because the vehicles are represented by a few pixels.

Since different training types may affect the results of detection, we are trying to analyze the results according to the training classes which are roughly divided into two parts as nature and man-made. Furthermore, these are subdivided into forest, shadow, building, and road.

Our work will be continued by refining the process and making the robust and flexible shape context bin to verify a variety of vehicles.

We believe that the analysis of neural network according to the training classes can also be useful for other object detection as well as transportation application.

Acknowledgement

This research was supported by the Ministry of Knowladge Economy, Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement)(IITA-2008-C1090-0801-0040).

References

Bay, H., T. Tuytelaars, and L.V. Gool, 2006. SURF: Speeded up Robust Features, *In: European Conf. Computer Vision*, 404-417.

Belongie, S., J. Malik, and J. Puzicha, 2002. Shape Matching and Object Recognition using Shape Contexts, *IEEE trans. Pattern Analysis and Machine Intelligence*, 24(24): 509-522.

Chellappa, R., P. Burlina, L. S. Davis, and A. Rosenfeld, 1998. SAR/EO Vehicular Activity Analysis System Guided by Temporal and Contextual Information, *Proc. 1994 ARPA Image Understanding Workshop*, 615-620.

Comaniciu, D. and P. Meer, 1997. Robust Analysis of Feature Spaces: Color image Segmentation, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 750-755.

Demuth, H. and M. Beale, 1998. Neural Network Toolbox for Use with Matlab, *The Mathworks Inc.*

Fukunaga, K. and L. D. Hostetler, 1975. The Estimation of the Gradient of a Density Function with Applications in Pattern Recognition, *IEEE Trans. Information Theory*, 21(1): 32-40.

Hinz, S., 2003. Integrating Local and Global Features for Vehicle Detection in High Resolution Aerial Imagery, *Photogrammetry Remote Sensing Spatial Information*, 34(3W/8), 119-124.

Kovesi, P., 1997. Symmetry and Asymmetry from Local Phase, *10th Australian Joint Conf. on Artificial Intelligence*, 185-190.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints, *Int'l J. Computer Vision*, 60(2): 91-110.

Moon, H., R. Chellappa, and A. Rosenfeld, 2002. Performance Analysis of a Simple Vehicle Detection Algorithm, *Image and Vision Computing*, 20(1): 1-13.

Park, J. and J. W. Sandberg, 1991. Universal Approximation using Radial Basis Functions Network, *Neural Computation*, 3: 246-257.

Zhao, T. and R. Nevatia, 2003. Car Detection in Low Resolution Aerial Images, *Image and Vision Computing*, 21: 693-703.