

A Morphological Neural Network Approach for Vehicle Detection from High Resolution Satellite Imagery

Hong Zheng¹, Li Pan², and Li Li¹

¹ Research Center for Intelligent Image Processing and Analysis,
School of Electronic Information, Wuhan University
129 Luoyun Road, Wuhan, Hubei 430079, China
zhenghong@21cn.com

² School of Remote Sensing Information & Engineering,
Wuhan University, 129 Luoyun Road, Wuhan, Hubei 430079, China
li.pan@126.com

Abstract. This paper introduces a morphological neural network approach to extract vehicle targets from high resolution panchromatic satellite imagery. In the approach, the morphological shared-weight neural network (MSNN) is used to classify image pixels on roads into vehicle targets and non-vehicle targets, and a morphological preprocessing algorithm is developed to identify candidate vehicle pixels. Experiments on 0.6 meter resolution QuickBird panchromatic data are reported in this paper. The experimental results show that the MSNN has a good detection performance.

1 Introduction

With the development of traffic there is high demand in traffic monitoring of urban areas. Currently the traffic monitoring is implemented by a lot of ground sensors like induction loops, bridge sensors and stationary cameras. However, these sensors partially acquire the traffic flow on main roads. The traffic on smaller roads – which represent the main part of urban road networks – is rarely collected. Furthermore, information about on-road parked vehicle is not collected. Hence, area-wide images of the entire road network are required to complement these selectively acquired data. Since the launch of new optical satellite systems like IKONOS and QuickBird, this kind of imagery is available with 0.6-1.0 meter resolution. Vehicles can be observed clearly on these high resolution satellite images. Thus new applications like vehicle detection and traffic monitoring are raising up. This paper intends to study the vehicle extraction issue from high resolution satellite images.

Some vehicle detection methods have been studied using aerial imagery [1][2][3][4]. In the existing methods, two vehicle models are used. They are explicit model and appearance-based implicit model. The explicit model describes a vehicle as a box or wire-frame representation. Detection is carried out by matching the model "top-down" to the image or grouping extracted image features "bottom-up" to create structures similar to the model.

Few research on vehicle detection from high-resolution satellite imagery with a spatial resolution of 0.6-1.0m has been reported [5][6]. At 0.6-1.0 meter resolution,

vehicle image detail is too poor to detect a vehicle by model approaches. Thus, it is necessary to develop specific approaches to detect vehicles from high resolution satellite imagery.

Morphological shared-weight neural network (MSNN) combines the feature extraction capability of mathematical morphology with the function-mapping capability of neural networks in a single trainable architecture. It has been proven successful in a variety of automatic target recognition (ATR) applications [7][8][9]. Automatic vehicle detection belongs to ATR research, thus, in this paper the MSNN is employed to detect vehicle targets.

In this paper, we concentrate the vehicle detection on roads and parking lots, which can be manually extracted in advance. In order to reduce searching cost and false alarm, a morphology based preprocessing algorithm is developed. The algorithm automatically identifies candidate vehicle pixels which include actual vehicle pixels and non-target pixels similar to vehicle pixels. Some of sub-images centered at those pixels are selected as the vehicle and non-vehicle training samples of the MSNN. The trained MSNN is tested on real road segments and parking lots. The performance results are also discussed in this paper.

The paper is organized as follows. In Section 2, the details of our vehicle detection approach are described. In Section 3, experimental results are given and conclusions are provided in Section 4.

2 Vehicle Detection Approach

The vehicle detection is carried out by an MSNN classification method. Before describing the vehicle detection approach, we briefly introduce the MSNN architecture as follows.

2.1 MSNN Architecture

Before describing the MSNN architecture, we provide brief definitions of some gray scale morphological operations. A full discussion can be found in [10]. The basic morphological operations of erosion and dilation of an image f by a structuring element (SE) g are

$$\textit{erosion} : (f \ominus g)(x) = \min \{ f(z) - g_x(z) : z \in D[g_x] \} \quad (1)$$

$$\textit{dilation} : (f \oplus g)(x) = \max \{ f(z) - g_x^*(z) : z \in D[g_x^*] \} \quad (2)$$

where $g_x(z) = g(z - x)$, $g(z) = -g(-z)$ and $D[g]$ is the domain of g . The gray-scale hit-miss transform is defined as

$$f \otimes (h, m) = (f \ominus h) - (f \oplus m^*) \quad (3)$$

It measures how a shape h fits under f using erosion and how a shape m fits above f using dilation. High values indicate good fits.

MSNN is composed of two cascaded sub-networks: feature extraction (FE) sub-network and feed-forward (FF) classification sub-network. The feature extraction sub-network is composed of one or more feature extraction layers. Each layer is composed

of one or more feature maps. Associated with each feature map, is a pair of structuring elements – one for erosion and one for dilation. The values of a feature map are the result of performing a hit-miss operation with the pair of structuring elements on a map in the previous layer (see Fig. 1). The values of the feature maps on the last layer are fed to the feed-forward classification network of the MSNN [11][12].

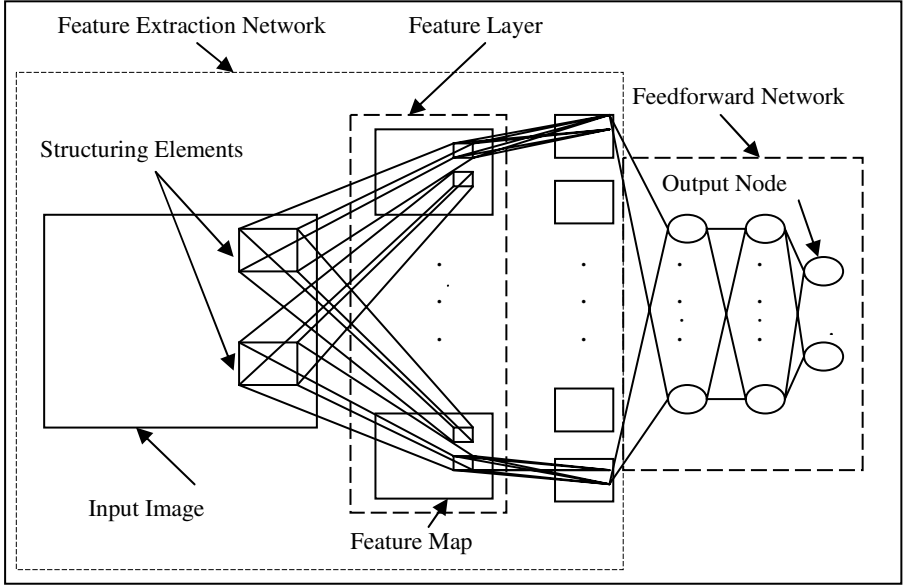


Fig. 1. The architecture of the morphology shared-weight neural network

2.2 Vehicle Detection Using MSNN

2.2.1 Morphology Preprocessing

In order to reduce searching cost and false alarm, a morphology based preprocessing algorithm is developed. In the algorithm, some morphological operations are used to enhance vehicle targets. These morphological operations are gray-scale top-hat and bottom-hat transforms, which are defined as

$$\text{top-hat: } T-HAT(f) = f - (f \circ g) \quad (4)$$

$$\text{bottom-hat: } B-HAT(f) = (f \bullet g) - f \quad (5)$$

where $f \circ g$ and $f \bullet g$ means opening operation and closing operation respectively, i.e.

$$\text{opening: } f \circ g = (f \ominus g) \oplus g \quad (6)$$

$$\text{closing: } f \bullet g = -((f \ominus g) \oplus g) \quad (7)$$

From empirical observation, the width of most vehicles on QuickBird images generally is less than or equal to 4 meters, and the length is not more than 6 meters. Thus the SE used is a disc with radius $r = 3$. Bright vehicles are smoothed out by the morphological opening operation and dark vehicles are smoothed out by the morphological closing operation. As a result, vehicles generally have a high value either on the top-hat image or the bottom-hat image. By setting a threshold on the top-hat image or the bottom-hat image, almost all vehicle pixels are detected and non-target pixels most similar to the vehicle pixels are also extracted. The threshold is obtained automatically using Ostu method [13]. In the Ostu method, pixels of a given image are represented in L gray Levels $[1, 2, \dots, L]$. The number of pixels at level i is denoted by n_i , and the total number of pixels by $N = n_1 + n_2 + \dots + n_L$. Then the dichotomisation of pixels into two classes C_0 and C_1 , which denote respectively pixels with $[1 \dots k]$ and $[k+1 \dots L]$. The method determines the threshold by determining the grey level that maximizes the between-class variance of the gray level histogram.

Fig.2(a) shows a road segment, and Fig.2(b)-(d) show its top-hat image, the bottom-hat image and their binary images after thresholding. From Fig.2(b)-(c), it can be seen that both bright vehicles and dark vehicles are enhanced after morphology preprocessing. As a result, these vehicles are labeled as white after thresholding. However, some noise like bright lane marks and tree shadow are also enhanced and mixed with vehicles. In order to further discriminate vehicle target pixels and non-vehicle target pixels, MSNN is introduced to implement pixel classification.

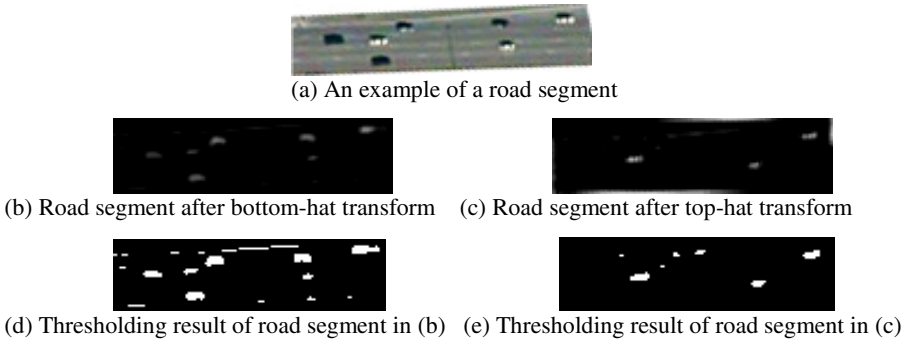


Fig. 2. An example of the morphology preprocessing algorithm

2.2.2 Network Training and Classification Testing

After the morphology preprocessing, the candidate vehicle pixels are obtained (see Fig.2(d)-(e)). Based on these candidate pixels, some sub-images centered at these pixels are selected as the vehicle and non-vehicle training samples of the MSNN. During training, test sub-images provide the input to the first feature extraction layer and the final output is a classification of “vehicle” or “non-vehicle”. This method of training is called the “class-coded” mode of operation. While the network outputs values of 0 to 1 representing the confidence that an input represents a vehicle or non-vehicle, the returned result is an actual classification.

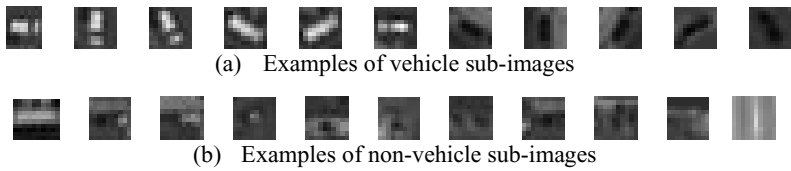


Fig. 3. Examples of training sub-images

Training data consists of a set of sub-images, which contain bright vehicles, dark vehicles, varying views of the “vehicle” and different “background”. Fig.3 shows some examples of training sub-images.

Several parameters specify and/or affect network training. The regularization parameter indicates the reliability of the training set, with a value of zero indicating that the set is completely reliable and a value approaching infinity indicating less reliability. The learning rate and momentum constant are used to adjust the speed of convergence and stability while reaching a desired error size.

Weights for the feature extraction operation are user-initialized, while the initial feedforward weight matrices are populated by a random number generator. All FE and FF weights are learned by back propagation. A signal completes its forward pass and then the correction its backward pass at the end of each training epoch, before the next input begins processing. A weight correction is the function of the learning and momentum parameters, the local gradient of the activation function, and the input signal of the neuron.

After learning, the trained weights are used to implement pixel classification, which includes the feature extraction and feedforward classifications. Feature extraction is performed over the entire image rather than on a sub-image. The resulting feature maps centered at the candidate vehicle pixels with subimage-sized windows are input into the feedforward network for classification, and output value represents the attribution of the candidate vehicle pixel, i.e., vehicle pixel or non-vehicle pixel.

3 Experimental Results

QuickBird panchromatic data set used in our study was collected from Space Imaging Inc. web site. The data set contains different city scenes. A total of 15 road segments and 5 parking lots segments containing over 1000 vehicles were collected. Most vehicles in the images are around 5 to 10 pixels in length and around 3 to 5 pixels in width. Since the vehicles are represented by a few pixels, their detection is very sensitive to the surrounding context. Accordingly, the collected images consist of a variety of conditions, such as road intersections, curved and straight roads, roads with lane markings, road surface discontinuity, pavement material changes, shadows cast on the roads from trees, etc. These represent most of the typical and difficult situations for vehicle detection.

For each selected road segment image or parking lot image, roads and parking lots were extracted manually in advance and vehicle detection was performed only on the extracted road surfaces. To build the vehicle example database, a human expert manually delineated the rectangular outer boundaries of vehicles in the imagery.

A total of 100 vehicles delineated in this manner from 10 road segments. An image region with size 6×6m can cover most vehicles in the imagery. Hence, sub-images of size 10×10 pixels centered at vehicle centroids were built into the vehicle example database. Taking vehicle orientations into account, each sub-image was rotated every 45° and the resulting sub-images were also collected in the vehicle example database. As a result, the vehicle example database consisted of $100 \times 4 = 400$ sub-image samples. In addition, 400 non-vehicle sub-image samples covering different road surfaces were also collected to build the non-vehicle example database.

After building sample databases, sub-image samples were used to train the MSNN and validate the vehicle detection approach. The MSNN used in our experiments had a 20×20 input and one feature extraction layer with two feature maps. The downsampling rate was 2 (i.e., 10×10 feature maps) and the structuring elements were 5×5. The feed-forward network of the MSNN was composed of a two-node input layer, ten-node hidden layer and a two-node output layer (target and non-target). All weights were initialized with random numbers in [-0.1, 0.1]. The learning rate was 0.002. A logistic function was used as the activation function. The expected outputs for vehicle targets and non-vehicle targets were set to [1, 0] and [0, 1] respectively. With these training parameters, the network was trained for 1600 epochs.

After training, the MSNN was tested on 15 road segments and 5 parking lots. The detection statistical results are shown in Tables 1. Fig. 4 shows some images of vehicle detection results.

Table 1. Vehicle detection results

Site	No. of vehicles	No. of detected vehicles	No. of missing vehicles	No. of false alarm	Detection rate %
Road1	6	5	1	0	83.3
Road2	8	7	1	0	87.5
Road3	11	9	1	1	81.8
Road4	15	13	2	0	86.6
Road5	20	16	3	1	80
Road6	18	15	3	0	83.3
Road7	28	23	5	0	82.1
Road8	63	52	8	3	82.5
Road9	54	41	10	3	75.9
Road10	82	66	12	4	80.4
Road11	114	92	15	7	80.7
Road12	154	125	23	6	81.1
Road13	210	175	29	6	83.3
Road14	268	227	31	10	84.7
Road15	304	234	50	20	76.9
Parking1	7	5	2	0	71.4
Parking2	13	9	4	0	69.2
Parking3	20	13	5	2	65
Parking4	46	28	15	3	60.8
Parking5	90	46	40	4	51.1

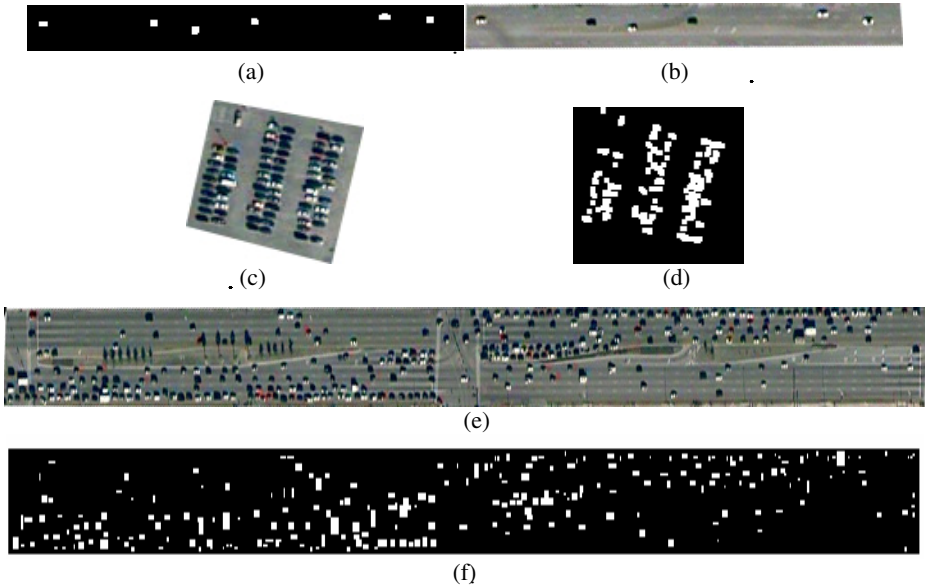


Fig. 4. Vehicle detection results (a)(c)(e) The original images of road segments and parking lots. (b)(d)(f) The binary images of vehicle detection results for images shown in (a)(c)(e).

From Table 1, it can be seen that the detection rates (number of detected vehicles/number of vehicles) for road segments are from 75.9% to 87.5%, and average detection rate is 82%. The detection rates vary with the complexity of road surfaces, as well as the false alarm. The false alarms are due to vehicle-like “blobs” present in some of complex urban scenes, such as the presence of dust and lane markings (see Fig. 4). Some of these “blobs” are very hard to distinguish from actual vehicles, even to a trained eye. Most missing detections occur when the vehicles have a low contrast with the road surface or vehicles are too close.

For the vehicle detection on parking lots, the detection rates are not high. It is because the vehicles are too close to separate due to the resolution limit. How to detect vehicles on parking lots is still an open issue.

4 Conclusions

In this paper, we focus on the issue of vehicle detection from high resolution satellite imagery. We present a morphology neural network approach for vehicle detection from 0.6 meter resolution panchromatic QuickBird satellite imagery. A MSNN was introduced in our approach and was found to have good vehicle detection performance. Further work could include more training samples, better pre-processing method such as adaptive image enhancement and filtering, and introducing more information like edge shapes to improve the detection rate.

Acknowledgement

This work was supported by the National Nature Science Foundation of China (Grant No. 40571102).

References

- [1] Ruskone R., Guigues L., Airault S., and Jamet O.: Vehicle Detection on Aerial Images: A Structural Approach. In: Proceedings of International Conference On Pattern Recognition, Vienna, Austria (1996) 900-904.
- [2] Zhao T. and Nevatia R.: Car Detection in Low Resolution Aerial Image. In: Proceedings of International Conference on Computer Vision, Vancouver, Canada (2001) 710-717.
- [3] Schlosser C., Reitberger J., and Hinz S.: Automatic Car Detection in High Resolution Urban Scenes Based on An Adaptive 3D-model. In: Proceedings of the 2nd GRSS/ISPRS Joint Workshop on Data Fusion and Remote Sensing over Urban Area, Berlin, Germany (2003) 167-170.
- [4] Stilla U, Michaelsen E, Soergel U, Hinz S, Ender HJ.: Airborne Monitoring of Vehicle Activity in Urban Areas. In: Altan MO (ed) International Archives of Photogrammetry and Remote Sensing, 35(B3) (2004) 973-979.
- [5] Sharma G.: Vehicle Detection and Classification in 1-m Resolution Imagery. Ohio State University, Master of Science thesis (2002).
- [6] Gerhardinger A., Ehrlich D., Pesaresi M.: Vehicles Detection from Very High Resolution Satellite Imagery, In: Stilla U, Rottensteiner F, Hinz S (Eds), International Archives of Photogrammetry and Remote Sensing, Vol. XXXVI, Part 3/W24 (2005) 83-88.
- [7] Won Y. : Nonlinear Correlation Filter and Morphology Neural Networks for Image Pattern and Automatic Target Recognition. Ph.D. Thesis, University of Missouri, Columbia, Miss (1995)
- [8] Won Y, Gader PD, Coffield P.: Morphological Shared-Weight Networks with Applications to Automatic Target Recognition. IEEE Trans. on Neural Networks, 8 (1997) 1195–1203.
- [9] Khabou M.A., Gader P.D. and Keller J.M.: LADAR Target Detection using Morphological Shared-weight Neural Networks. Machine Vision and Applications, 11 (2000) 300-305.
- [10] Serra J.: Image Analysis and Mathematical Morphology, Vol. 2, Academic Press, New York, N.Y (1988).
- [11] Gader PD, Won Y, Khabou MA. : Image Algebra Networks for Pattern Classification. In: Proceedings of SPIE Conference on Image Algebra and Morphological Image Processing, 2300 (1994) 157–168.
- [12] Gader PD, Miramonti JR, Won Y, Coffield P.: Segmentation Free Shared-Weight Networks for Automatic Vehicle Detection. Neural Networks, 8 (1995) 1457–1473.
- [13] Ostu N.: A Threshold Selection Method from Gray Level Histograms, IEEE Transactions on System, Management, Cybernet, 9 (1979) 62-66.