Xutong NIU, Rongxing LI & Morton O'Kelly

TRUCK DETECTION FROM AERIAL PHOTOGRAPHS

Xutong NIU [a], Rongxing LI [a], Morton O'Kelly [b]

[a] Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University
470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210
niu.9@osu.edu
[b] Department of Geography, The Ohio State University, 1036 Derby Hall, 154 North Oval Mall, Columbus, OH 43210

Commission II, WG II/1

KEY WORDS: Mean Shift Image Segmentation, Feature Extraction, Truck Detection, Hopfield Neural Network

ABSTRACT:

This paper presents a method for extraction and recognition of trucks from aerial photographs. A new method, mean shift image segmentation, is employed to extract truck shadow and boundary information. After a conversion of the boundaries into vector data, an enclosing minimum area rectangle is fit to each boundary. These rectangles are input as candidates into a designed Hopfield Neural Network. Trucks in the photographs are then recognized by comparison with predefined truck models.

1. INTRODUCTION

Nowadays, remotely sensed images are playing a more and more important role in GIS data collection and application. With a view to facilitating this transition, the United States Department of Transportation (USDOT) Research and Special Programs Administration (RSPA) and the National Aeronautics and Space Administration (NASA) have collaborated to establish the National Consortia on Remote Sensing in Transportation (NCRST). The NCRST consortia are pursuing research on application of remote sensing and spatial information technologies to transportation issues in four focus areas: (a) environmental assessment, (b) infrastructure management, (c) traffic flow and (d) hazards, safety and disaster assessment (USDOT and NASA, 2001). The consortium on traffic flows is particularly interested in improving the efficiency of the transportation system at national, state, and local levels by integrating remotely sensed traffic flow data collected from airborne or satellite platforms with traditional data collected from the ground.

This paper is based on the research results of the project "Freight and Inter-modal Flow Analysis", a consortium project on traffic flows. This project is trying to measure the impact of freight movement surrounding large ports and/or inter-modal facilities on a statewide scale.

In order to derive traffic flow from the photo sequence, the trucks must first be detected. This work has usually been performed manually. This paper presents a method for automatically extracting and recognizing trucks from sequential aerial photographs. Region-based information about truck shadows and boundaries is extracted using a mean shift procedure-based method. A minimum area rectangle is then fit to each identified possible region. Taking these rectangles as truck candidates, a two-layer Hopfield Neural Network algorithm is applied to achieve truck recognition.

2. METHODOLOGY

2.1 Feature Extraction

Feature extraction is an active area of research in computer vision and mapping. In the past, feature extraction has been divided into several components and stages. Examples include edge detection, contour derivation, and shape modeling. Later models, such as the deformable contour model (Kass et al., 1987), treat these problems in a general unified manner. However, no criteria exist for testing models and validating parameters.

The Digital Mapping Laboratory of Carnegie Mellon University (CMU) attempted to combine all available image, scene, and domain knowledge in their feature extraction methods. For example, they used the vanishing point technique for building detection, and they incorporated automatic feature extraction and stereo matching into a semi-automatic system (McKeown et al., 1996). In another study (Gruen and Li, 1996), a semi-automatic method for linear-feature extraction from digital images was presented. This procedure combined human identification of features with high precision line extraction performed by a computer for GIS data capture.

2.2 Mean Shift Image Segmentation

Shadows and top area information about the trucks are very important information for truck detection. In this paper, mean shift procedure-based image segmentation (Comaniciu et al., 2002) is used to identify the area information.

The mean shift procedure was proposed by Fukunaga and Hostetler in 1975. It is a density estimation-based non-parametric clustering approach. The rationale behind this approach is that the feature space can be regarded as the empirical probability density function (p.d.f.) of the represented parameter. Dense regions in the feature space thus correspond to local maxima of the p.d.f., which are the modes of the unknown density (Comaniciu et al., 2002). Once the location of a mode is determined, the cluster associated with it is delineated based on the local structure of the feature space.

Given n data points $\mathbf{x}_i$, $i = 1,…,n$ in the d-dimensional space $\mathbf{R}^d$, the multivariate kernel density estimator with kernel $K(\mathbf{x})$ and a symmetric positive definite d x d bandwidth matrix $\mathbf{H}$, computed at the point $\mathbf{x}$ is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x}-\mathbf{x}_i) \qquad (1)$$

where

$$K_{\mathbf{H}}(x) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x}) \qquad (2)$$

$$K(\mathbf{x}) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \qquad (3)$$

and H is proportional to the identity matrix $\mathbf{H} = h^2\mathbf{I}$; h > 0, which is the bandwidth parameter.

Then, the mean shift is

$$m_{h,K}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i K(\mathbf{x}-\mathbf{x}_i)}{\sum_{i=1}^{n} K(\mathbf{x}-\mathbf{x}_i)} - \mathbf{x}, \qquad (4)$$

which is the difference between the weighted mean (using the kernel K for weights), and x, the center of the kernel (window).

Thus the relationship between the gradient of the density estimator and the mean shift can be derived (Comaniciu et al., 2002) as

$$m_{h,K}(\mathbf{x}) = \frac{1}{2}h^2 c \frac{\hat{\nabla}f_{h,K}(\mathbf{x})}{\hat{f}_{h,K}(\mathbf{x})} \qquad (5)$$

This expression shows that the mean shift vector always points to the direction of maximum increase in the density.

Motivated by this relationship, in 2002 Comaniciu et al. presented a mean shift procedure-based image segmentation scheme. This scheme provides an autonomous solution for image segmentation. In their paper, the feature space of one image is typically composed of two domains. The first domain is the spatial domain, which represents the two-dimensional lattice space of p-dimensional vectors (pixels), where $p = 1$ for gray-level images, $p = 3$ for color images, and $p > 3$ for the multi-spectral case. The second is the range domain, which represents gray level, color, or spectral information. For both domains, Euclidean metric is assumed. In the joint spatial-range domain of dimension $d = p + 2$, these location and range vectors have to be normalized to compensate for their quantitative differences.

Since the spatial domain is involved in the estimation, results are not only related to pixel information, but also to the spatial relationship between pixels. Based on these considerations, the authors have decided to use mean shift image segmentation for the feature extraction part of this project.

2.3 Truck Models

After image segmentation, Canny edge detection (Canny, 1986) is used to extract the boundaries of the detected area. These boundaries then are vectorized into polygons.

To simplify the recognition problem, the complicated shape of the front part of a truck is disregarded. The main truck body is simplified to a rectangle. From the aerial camera, the top of a truck and its shadow are always visible, except when the shadow disappears at high noon (Figure 1).
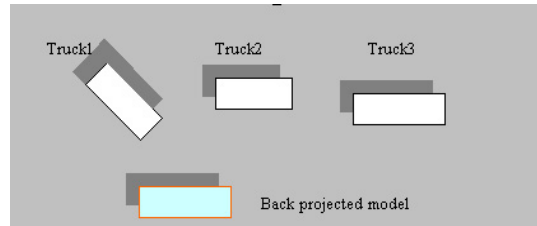

Figure 1. Simplified truck image features and back projected model

In order to fit the detected boundaries to the truck model, we need to convert the boundaries into rectangular shapes.
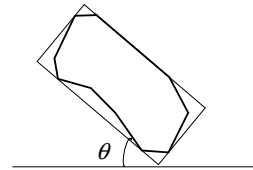

Figure 2. Detected truck boundary and its minimum area-enclosing rectangle

Figure 2 shows a truck boundary and its minimum area-enclosing rectangle. θ is the angle between the longer side of the rectangle and the horizontal line. If θ is known, the bounding rectangle can be computed efficiently from boundary points. The transformation from the x and y coordinate system that is used to define the boundary to a new coordinate system that has axes α and β parallel to the two sides of the bounding rectangle is (Sonka et. al., 1999)

$$\alpha(x,y) = x\cos\theta + y\sin\theta$$
$$\beta(x,y) = -x\sin\theta + y\cos\theta \qquad (6)$$

A search is made for the minimum and maximum of α and β over all boundary points (x, y). After the values of $\alpha_{min}, \alpha_{max}, \beta_{min}, \beta_{max}$ are found, then we define the bounding rectangle, where $l_1 = (\alpha_{max} - \alpha_{min})$ and $l_2 = (\beta_{max} - \beta_{min})$ are its length and width.

In our case, the value of θ is unknown. But it can be obtained by using the following steps:

1) For each value of θ from 0 to 180 degrees, calculate the values of corresponding length, width, and area of the bounding rectangle;
2) Find the minimum of the area values and take this rectangle as the preferred bounding rectangle.

In implementation of this method, the convex hull of the boundary points is computed first to reduce computational time. Then the above steps are carried out to find the proper result.

2.4 Neural Network Object Recognition

A Hopfield neural network method (Li et al., 1998; Tu and Li, 1999) for automatic truck recognition is used to recognize trucks from the extracted candidates.

The recognition problem is treated as an optimization problem, where the correct answer is given when a global minimized energy state is reached. The energy function is

$$E = -A\sum_i\sum_k\sum_j\sum_l C_{ikjl}V_{ik}V_{jl} + B\sum_i(1-\sum_k V_{ik})^2 + C\sum_i\sum_k\sum_{l\neq k}V_{ik}\times V_{il} \quad (7)$$

$$+D\sum_k(1-\sum_i V_{ik})^2 + E\sum_k\sum_i\sum_{j\neq i}V_{ik}\times V_{jk}$$

The neuron state, V_{ik}, converges to 1.0 if the model feature i perfectly matches the candidate image feature k. Otherwise, it is equal or close to 0. Thus, the first term measures the similarity between the model and image features. The second term, $\sum_i(1-\sum_k V_{ik})^2$, implies that the final states of neurons in the same row add up to 1 to enforce a unique match within a row, and the third term, $\sum_i\sum_k\sum_{l\neq k}V_{ik}\times V_{il}$, confirms that there is at most one neuron that has a value greater than 0 in each row. The fourth term, $\sum_k(1-\sum_i V_{ik})^2$, implies that the final states of neurons in the same column add up to 1 to enforce a unique match within a column, and the fifth term, $\sum_k\sum_i\sum_{j\neq i}V_{ik}\times V_{jk}$, confirms that there is at most one neuron that has a value greater than 0 in each column. Combining the second term, $\sum_i(1-\sum_k V_{ik})^2$, with the third term, $\sum_i\sum_k\sum_{l\neq k}V_{ik}\times V_{il}$, gives a solution that forces each model feature to match only one candidate image feature. Similarly, combining the fourth term, $\sum_k(1-\sum_i V_{ik})^2$, with the fifth term, $\sum_k\sum_i\sum_{j\neq i}V_{ik}\times V_{jk}$, gives a solution that guarantees each candidate image feature will match only one model feature. The determination of coefficients A, B, C, D and E depends on how strictly the unique matching condition should be implemented.

The solution to Equation (7) is a set of neuron states, V_{ik}, and is achieved by a numerical approach to solving the partial differential equations derived from Equation (2) (Li et al., 1998).

3. EXPERIMENTS

3.1 Data

Georeferenced 1:9,600 scale aerial photographs acquired in southern Franklin County, Ohio are used to test the feature extraction method. Figure 3 and Figure 5 show a parking lot with more than thirty cargo stacks, and an intersection of highway I-270 and a local road, respectively. At first, the segmentation method was used to extract the shadows and cargo top areas. The results are shown in Figure 4(a) and Figure 6. The boundaries of these areas are then extracted using Canny Edge detection (Figure 4(b) and Figure 7). Because there were still some gaps existing in the boundaries of the areas, morphological erosion and dilation operations were applied to the boundaries and, finally, the closed boundaries were obtained. Figure 8 and Figure 9 show the enclosing rectangle overlapped on the original images. It can be seen that these rectangles match with the trucks in the images very well.

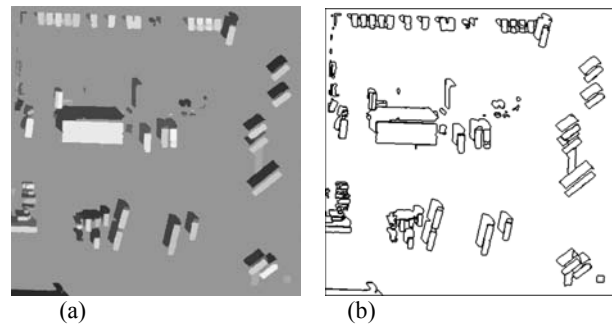Figure 3. Parking lot with cargo stacks

(a)   (b)

Figure 4. Segmentation result of the parking lot (a) and (b) the edge detection results

Figure 5. The Intersection of Highway I-270 and Alum Creek Road, Columbus, Ohio

353

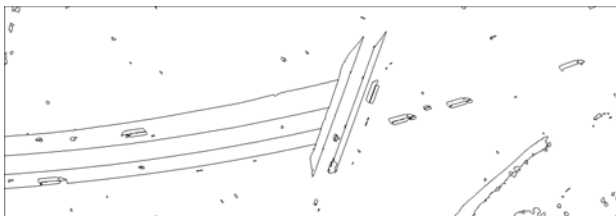Figure 6. Segmentation result of the highway image

Figure 7. Edge detection result of Figure 6

Figure 8. Minimum area enclosing rectangles on the parking lot image
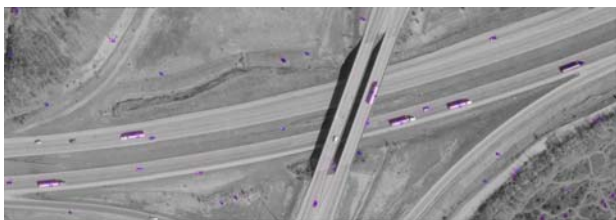

Figure 9. Minimum area enclosing rectangles on the highway image

3.2 Neural Network Recognition Result

A two-layer Hopfield neural network was used to recognize the trucks among the extracted bounding rectangles in Figure 7 (Tu and Li, 1999). According to the final neuron states, there are two categories of features that may match the trucks:

1) Two polygons match with the model, one with the truck top and the other with the shadow of the model.
2) In the region pattern layer there are no polygon pairs matching with the top and/or shadow of the model. However, there are polygon edges matching with those of the model top in the line pattern layer. This may be due to

a loss of shadow information caused, for example, by a gray value difference or occlusion.

In the experiment, two truck models were created to fit the different sizes of the cargo stacks in the parking lot or trucks on the roads. Most trucks (29 out of 34) in the parking lot (Figure 10) were recognized. Five out of six trucks were extracted in the highway image (see Figure 11).

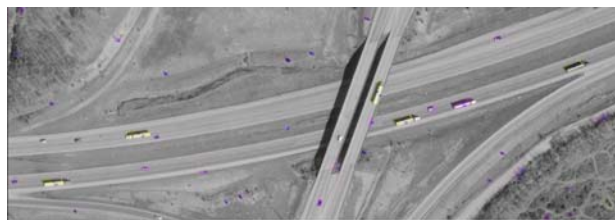Figure 10. Recognition result for the parking lot image

Figure 11. Recognition result for the highway image

4. DISCUSSION

Overall, this process detected nearly all trucks in several trial runs. In a total of 34 trucks in the parking lot, 29 trucks were extracted (85% hit rate). One building and some small cars were mistakenly identified as trucks; two trucks were missed. In the highway image, only one truck was missed.

The reason for the mistaken identifications is that the gray level and area of these candidates were very similar those of the truck. They could be removed by further processing, such as including texture information into the neural network model. The truck missed in the highway image was because the truck shadow included part of the truck cab, making it much longer than the truck body.

Thus, this neural network model can still be modified and improved. The results reported here represent progress in an on-going research project. An improved neural network model will use image orientation parameters and an enhanced multi-layer neural network structure to measure the similarity between a 3-D truck model on the ground and the extracted polygons of trucks and shadows in aerial photograph images. This truck extraction method can also be extended to an entire highway corridor or image sequence. Efforts will also be made to automate the matching process.

.

5. CONCLUSION

This paper presents a general framework for truck extraction from aerial photographs. At first, the mean shift procedure-based image segmentation was introduced to aerial photograph processing, and exhibits very good results for feature extraction in the natural scene. Then, a Hopfield Neural Network is used to verify the extraction results.

Since the mean shift procedure is a non-parametric estimation method, this frame can be extended to other kinds of feature extraction from aerial photographs and satellite images.

ACKNOWLEGEMENTS

REFERENCES

Agahi, R., A. V. Gafarian, P. Jagger, L. T. Nguyen, and J. Pahl, 1976. Characteristics of Multilane Traffic Flow from Aerial Data. Report No. DOT-TST-76T-2, U.S. Department of Transportation.

Canny, J. F., 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), pp. 679-698.

Cheng, Y., 1995. Mean Shift, Mode Seeking, and Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8), pp. 790-799.

Comaniciu, D. and P. Meer, 2002. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (5), pp. 603-619.

Daganzo, C., 1997. *Fundamentals of Transportation and Traffic Operations*. Elsevier Science, New York.

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

Gruen, A. and H. Li, 1996. Linear Feature Extraction with LSB-Snakes from Multiple Images, *International Archives of Photogrammetry and Remote Sensing*, 31(B3), pp. 266-272.

Kass, M., A. Witkin and D. Terzopoulos, 1987. Snakes: Active Contour Models, *International Journal of Computer Vision*, 1, pp. 321-331.

Li, R., 1998. Potential of High-resolution Satellite Imagery for National Mapping Products, *Journal of Photogrammetric Engineering and Remote Sensing,* 64(12), pp. 1165-1169.

Li, R., W. Wang and H.-Z. Tseng, 1997. Geometric Constraints In Image Sequences and Neural Networks for Object Recognition, Project Report, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, p. 78.

McKeown, D. M. et al., 1996. Research in the Automated Analysis of Remotely Sensed Imagery: 1994-1995. CMU-CS-96-101, Digital Mapping Laboratory, Carnegie-Melon University, Pittsburgh.

Mintzer, O. W., ed., 1983. *Manual of Remote Sensing*, 2nd Edition, Volume 2. *Interpretations and Applications*, American Society of Photogrammetry, Falls Church, VA, pp. 1955-2109.

Sonka, M., V. Hlavac, and R. Boyle, 1998. *Image Processing, Analysis, and Machine Vision.* Second Edition, PWS Publishing, p. 258.

Treiterer, J., 1975. Investigation of Traffic Dynamics by Aerial Photogrammetry Techniques. The Ohio State University, Engineering Experiment Station.

Tu, Z. and R. Li, 1999. A Multilayer Hopfield Neural Network for 3-D Object Recognition. *Proceedings of International Mobile Mapping Workshop*, Bangkok, Thailand, April 21-23, 1999, pp. 7A.3.1-6.

US Department of Transportation and NASA, 2001. Remote Sensing and Spatial Information Technologies in Transportation Synthesis Report. Available at: http://scitech.dot.gov/reeng/sensmsrm/rmtsense/rmtsense.html (accessed 14 Nov. 2001).