Robust component-based car detection and counting in aerial imagery based on the mean-shift colour space clustering

[a]Yueh Ouyang, [a]Samir Sahli, [a*]Yunlong Sheng and [b]Daniel A. Lavigne
[a]COPL, University Laval, Quebec (QC) G1K 7P4, CANADA,
[b] Defence Research and Development Canada-Valcartier, Quebec (QC) G3J 1X5 , CANADA

ABSTRACT

In the aerial images of 11.2 cm/pixel resolution the car components that can be seen are only large parts of the car such as car bodies, windshields, doors and shadows. Furthermore, these components are distorted by low spatial resolution, low color contrast, specular reflection and viewpoint variation. We use the mean shift procedure for robust segmentation of the car parts in the geometric and color joint space. This approach is robust, efficient, repeatable and independent of the threshold parameters. We introduce a hierarchical segmentation algorithm with three consecutive mean-shift procedures. Each is designed with a specific bandwidth to segment a specific car part, whose size is estimated *a priori*, and is followed by a support vector machine in order to detect this car part, based on the color features and the geometrical moment based features. The procedure starts with the largest car parts, which are then removed from the segmented region lists after the detection to avoid over-segmentation of large regions with the mean-shift using smaller bandwidth values. Finally we detect and count the cars in the image by combining the detected car parts according to the spatial relations. Experiment results show a good performance.

Keywords: Car detection, Aerial image, Components, Mean shift clustering, Image segmentation.

1. INTRODUCTION

The vehicle detection in the aerial imagery has drawn recently attention of the research community for civilian and military applications, such as road network detection, military reconnaissance and traffic surveillance for urban planning [1~5]. Car detection in aerial images, however, is a challenge as only the large parts of cars such as hood, roof, trunk, windshield and shadow can be detected in the aerial images at 11.2 cm/pixel resolution and because these components are distorted by low colour contrast and partial reflection.

We present a component-based method for vehicle detection in aerial imagery with mean shift segmentation [6] and the support vector machine (SVM) classification. The mean shift algorithm is used for low-level image colour segmentation. We used the SVM [7] to classify the car body parts, windshields and shadows by the radiometric color and geometric features based on the raw moments of the segmented regions. The mean-shift clustering is non-parametric. However, the sole parameter required, the bandwidth (size of the kernel windows), can affect the quality of the segmentation. Therefore, we introduce a hierarchical segmentation algorithm with three consecutive mean-shift procedures. Each is designed with a specific bandwidth to segment a specific car part, whose size is estimated *a priori*. The segmentation is followed by a support vector machine to detect the specific car part, based on the color features and the geometrical moment based features. The hierarchical procedure starts with the largest car parts, which are then removed from the segmented region lists after the detection to avoid over-segmentation of the large regions by the mean-shift with the smaller bandwidth values. Finally we detect and count the cars in the image by combining the detected car parts according to the spatial relations. Experiment results show a good performance.

2. MEAN-SHIFT SEGMENTATION

The mean-shift algorithm has been proposed by Fukunaga in 1975 for nonparametric feature space probability density gradient estimation and analysis. The density of data point distribution in a *d*-dimensional feature space can be estimated within bounded and radially symmetric windows as

*sheng@phy.ulaval.ca. Tel: 418 656 2131 3908

Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VII, edited by Daniel J. Henry,
Proc. of SPIE Vol. 7668, 76680H · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.852613

$$f(\mathbf{x}) = \frac{c}{nh}\sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{x}-\mathbf{x_i}}{h}\right\|^2\right) \tag{1}$$

where $k(\mathbf{x})$ is called the profile of the kernel $K(\mathbf{x}) = ck(\|\mathbf{x}\|^2)$, $h > 0$ is the bandwidth (window size) of the kernel, and c is a normalization constant. The gradient of the density function can be estimated by

$$\nabla f(x) = \frac{2c}{nh^{d+2}}\sum_{i=1}^{n}(\mathbf{x_i}-\mathbf{x})g\left(\left\|\frac{\mathbf{x}-\mathbf{x_i}}{h}\right\|^2\right) = \frac{2c}{nh^{d+2}}\sum_{i=1}^{n}\left(\frac{\sum_{i=1}^{n}\mathbf{x_i}g\left(\left\|\frac{\mathbf{x}-\mathbf{x_i}}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{\mathbf{x}-\mathbf{x_i}}{h}\right\|^2\right)} - \mathbf{x}\right)\cdot\sum_{i=1}^{n}g\left(\left\|\frac{\mathbf{x}-\mathbf{x_i}}{h}\right\|^2\right) \tag{2}$$

with $g(\mathbf{x})=-k'(\mathbf{x})$. The second term in the right-hand side of Eq. (2) is the difference between the weighted mean of the data points $\mathbf{x_i}$ and $\mathbf{x}$ and is defined as mean shift vector $\boldsymbol{m}(\mathbf{x})$. According to Eq. (2) $\boldsymbol{m}(\mathbf{x})$ is proportional to the gradient of the data point density function. The formulation in Eq. (2) is valid for arbitrary kernels. A kernel profile $g(\mathbf{x})$ can be, for instance, a uniform window as a uniform sphere of radius h in the d-dimensional space. In this case Eq. (2) estimates the gradient of the probability density function $f(\mathbf{x})$, which is estimated by Eq. (1) with the Epanechnikov kernel profile as

$$k(x) = \begin{cases} 1-x & 0 \le x \le 1 \\ 0 & x > 1 \end{cases} \tag{3}$$

This kernel minimizes the square error between $f(\mathbf{x})$ and its estimate. The mean-shift algorithm is a powerful tool for defining the modes of the probability density function in the feature space, where the modes are located among the zeros of the gradient $\nabla f(\mathbf{x}) = 0$. Starting from a point $\mathbf{x}$ the recursive mean-shift procedure first computes the mean-shift vector with the kernel window $G(\mathbf{x})$ and then translates $G(\mathbf{x})$ by the mean-shift vector to compute the new mean-shift vector. The updating of the mean-shift continues until converging to a mode center, where the mean-shift vector amplitude is below a threshold. The convergence is guarantied for infinitesimal mean-shift steps. The mean-shift procedure is nonparametric in the sense that it depends only on the applied data, and does not require a priori knowledge of the number of modes and their shapes. The only parameter to choose is the bandwidth (window size) h, which can be critical.

The advantage of the mean-shift algorithm for color image segmentation is that the feature space can be represented in a joint domain of the spatial location space and the L*u*v* color range space as a 5-dimensional space. In both domains the Euclidean distance metric is assumed. However, the metric distances in the location and range domains are of different physical nature. Therefore, a multivariable kernel is defined as a product of two radially symmetric kernels as

$$K(\mathbf{x}) = \frac{c}{h_s h_r} k\left(\left\|\frac{\mathbf{x^s}}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x^r}}{h_r}\right\|^2\right) \tag{4}$$

where $\mathbf{x^s}$ and $\mathbf{x^r}$ are measured in the location space and range space, respectively, and h_s and h_r are two independent bandwidths to be chosen. In the implementation of the mean-shift segmentation, the windows of bandwidth (h_s, h_r) are first tessellated over the entire feature space of space-range joint domain. Then, the mean-shift is run in all the windows in parallel until the convergence, resulting in a set of maxima to which the windows are converged. All the data points traversed by a window at its successive locations on the path towards the mode center (i.e. all the points within the basin

of attraction of the corresponding mode) are assigned to this mode. Finally all the mode centers that are closed than h_s in the spatial domain and than h_r in the range domain are regrouped together and delineated as a segmented region.

We note that the kernel bandwidths (sizes) and the initial tessellation of the kernel windows could alter the segmentation results, especially for the data points close to boundaries. Those points could be visited by several windows that went along divergent paths towards different modes, resulting in incertitude of the segmentation, which jeopardizes accuracy of the segmentation as shown in Fig. 1 which was obtained with application of the mean-shift algorithm to an aerial image of 11.2 cm/pixel resolution with default values in the algorithm of bandwidths in spatial space h_s=5 pixels, and that in range space h_r=10 pixels and the minimum size of the regions M =20 pixels. In addition, a close look at this result shows that many apparently homogeneous regions in Fig. 1 are in fact over-segmented into small regions. When the spatial bandwidth h_s is too small, there will be more over-segmentation. When h_s is too large, the small regions such as the windshield will be missed in the segmentation.

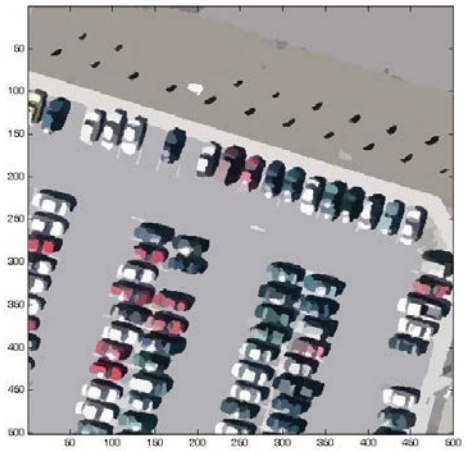

Fig. 1 Mean-shift segmentation of an aerial image of 11.2 cm/pixel resolution with the Epanechnikov kernel of the bandwidth (h_s=5, h_r=10 and M=20)

3. COMPONENTS DETECTION WITH MULTIPLE MEAN-SHIFT AND HIERARCHICAL SVM CLASSIFICATION

The car components that appear in the aerial images of 11.2 cm/pixel resolution are the large parts of the car as roof and hood, windshield, door windows and shadows only. Furthermore, these components are distorted by low spatial resolution, low color contrast, specular reflection and viewpoint variation. The mean-shift segmentation is robust, repeatable and does not require presetting parameters such as threshold. However, the choice of the bandwidths is still critical to the quality of the segmentation. The technique proposed for the mean-shift with adaptive bandwidths could be computationally expensive. Thus, we used the top-down information to preselect the bandwidths. As the aerial image resolution is known *a priori* we estimate sizes of the components of a passenger car as: car size: 40 x 20; car bodies: roof: 15 x 15, hood: 18 x 10 and truck: 16 x 8; windshield: 17 x 6 and shadow: 40 x 10 pixels. We observed in the experiments that the regions were better segmented with the mean-shift algorithm when the kernel window size was roughly the half of the region size. We empirically selected 3 bandwidths as 4 pixels for segmenting windshield, 7 for car bodies (roof, hood and trunk) and 12 for shadows. We applied 3 mean-shift procedures with the 3 bandwidths respectively. Each mean-shift procedure is followed by a binary-class SVM classification to detect one specific car component. As the mean-shift with to small kernel window size can introduce over-segmentation of the large regions, resulting in noisy regions to be discriminated against by the SVM, we organized the detection process in a hierarchical manner starting from the largest region and finishing with the smallest region, i.e. sequentially detecting shadows, car bodies and windshields. Moreover, after each detection process we removed the detected regions from the list of regions in the next step classification in order to easy the works of the SVM for detecting the next smaller components.

Aerial images of 1000 x1000 pixel size, as shown in Fig. 2, were used in the experiment. In general, one pass of the mean-shift segmentation can generate over 2000 regions. Radiometric and geometric features were used for describing the segmented regions. The RGB color model is suitable for color display, but not for color analysis because of its high

correlation among R, G and B components. Besides, the distance in RGB color space does not represent the perceptual difference for human eye. In our experiment, the intensity I, hue angle H and saturation S obtained by the HSI transformation from the RGB components as :

$$I = 0.299R + 0.587G + 0.114B \tag{5}$$

were used as the radiometric parameters. The area and circumference of the region were chosen as the size features. The circumference of a region A, denoted by $\beta(A)$, was obtained by first eroding A by B, denoted as by $A \Theta B$, and then account the pixels in the difference between A and its erosion as

$$\beta(A) = A - (A \Theta B) \tag{6}$$

where B was 3x3 square structuring element. The aspect ratio and the retangleness of the region are important features permitting distinguishing car component from natural objects. Car component can be approximated as a rectangle with the two sides lengths determined by the second order moments as

$$a = 2\left(\frac{m_{20} + m_{20} + [(m_{20} - m_{02})^2 + 4m_{11}^2]^{1/2}}{m_{00}/2}\right)^{1/2}$$
$$b = 2\left(\frac{m_{20} + m_{20} - [(m_{20} - m_{02})^2 + 4m_{11}^2]^{1/2}}{m_{00}/2}\right)^{1/2} \tag{7}$$

so that the aspect is a/b and the rectangleness is the ratio (area of the region)/ab.

According to the a priori knowledge on the car size of about 20 x 40 pixel^2, we eliminated the region of areas larger than 800 pixels in first place. As all the feature parameters including the region areas are normalized by the scaling in the SVM, removing the regions of too large areas increases the normalized area feature values and the differences between smaller regions.

(a) (b)

Fig. 2. Aerial images used as (a) training image (b) testing image

In the first step of process for shadow detection, the images were segmented by mean-shift algorithm with the spatial bandwidth h_s=12 and the range bandwidth h_r=10 units with the minimum pixel number in the regions of M=20 pixels. The mean-shift generated 1779 regions in the image. For training the SVM all the 168 shadow regions in the training image shown in Fig. 2a were carefully selected manually as the positive examples and were labeled by +1. The remaining segmented regions were chosen as negative examples and were labeled as -1. After training the SVM model was applied to testing image, shown in Fig. 2b, with the 857 regions segmented by the same mean-shift algorithm. There were 118 hits that the shadow regions were classified correctly, 6 false-alarms and 13 shadow regions missed, as shown

in Fig. 3a. Once detected the shadow regions were removed from the list of the regions segmented in the next mean-shift procedures with smaller bandwidths h_s according to the locations of the detected region pixels before the detection of the car bodies and windshields.

For detecting car body components, the mean-shift procedure with the spatial bandwidth h_s=7 pixels and the range bandwidth h_r=10 units and M=20 pixels was applied, resulting in 1427 regions in the training image in Fig. 2a. For training the SVM all the 347 car bodies in the image were manually selected and labeled as +1. The remaining regions excluding the shadows are chosen to be labeled as -1. After training the SVM model was applied to testing image shown in Fig. 2b, with the 1406 regions segmented by the same mean-shift algorithm. There were 167 hits that the car bodies were classified correctly, 13 false-alarms and 14 car body regions missed, as shown in Fig. 3b. Once detected the car body regions were removed from the list of the regions segmented in the next mean-shift procedures for detecting windshields.

For detecting windshields, the mean-shift procedure with the spatial bandwidth h_s=4 pixels and the range bandwidth h_r=10 units and M=20 pixels was applied, resulting in 4202 regions in the training image in Fig. 2a. For training the SVM all the 456 windshields in the training image Fig. 2a were manually selected and labelled as +1. The remaining regions excluding the shadows are chosen to be labeled as -1. After training the SVM model was applied to testing image shown in Fig. 2b, with the 2899 regions segmented by the same mean-shift algorithm. There were 160 hits that the windshields were detected correctly, 36 false-alarms and 34 windshields missed, as shown in Fig. 3c.

Fig. 3. Regions (marked by green) classified as (a) shadow, (b) car bodies, and (c) windshield in testing image Fig. 2b

4. COMBINING CAR COMPONENTS WITH SPATIAL RELATIONSHIP

A car is detected when its components segmented by the mean-shift algorithm and classified by the SVM can be combined to a car model according to the known geometrical relationships. We had three classes of components: car bodies, including roofs, hoods and trunks, shadows, and windshields and door windows. The known spatial relations are

(a) Windshields and door windows joint directly to car bodies, sharing common edges and vice-versa
(b) Shadows are located on one same side of the car body for all the cars in an aerial images
(c) Link between car bodies which are separated by a shadow is prohibited
(d) All components of a car are located within a region of 20x40 pixels.
(e) Distances between centroids of car body regions and shadow are in the range of 10 to 20 pixels.

These unique spatial relationships were used to greatly help the car detection. The mean-shift segmentation and the SVM classification resulted in three lists of classified regions as: bodies, windshields and door windows, and shadows. As car windows are directly jointed to car bodies, we first paired windshields and door windows with car bodies. We looked for common edges between any regions in the list of car bodies and any regions in the windshields list by superimposing the dilated regions and looking for overlapping between the dilated regions on one list and the regions on another list. When the number of overlapped pixels was over a threshold value, such as 5 pixels, the two regions are paired. The constraints in the spatial relations (c) and (d) were also used to check the pairs to satisfy the relations. By this operation one car body region can joint several windows and vice-versa.

From the detected 180 car bodies and 196 windshields, we obtained 187 body-windshield pairs. Then, we regrouped the car-body-windshield pairs according to the indices of the regions in the pairs. In this ways the hood and trunk can be

connected with roof through the windshield. The regrouping procedure assembled 68 car body groups from 187 body-windshield pairs. The detected car body-windshield groups were under a merging procedure. The groups with their edges separated by less than 7 pixels and without any shadows in between were merged. That resulted in 60 merged groups from the 68 groups.

We also paired the car bodies with the shadow regions. Note that the car bodies and the shadows could have no common edges because of the possible presence of the door windows and doors depending on the view angle. The pairing was then based on the fact that the distance between the controids of a car body and a shadow should be within a range of 10-20 pixels and that the shadows should be presented on the one same side of the car for all the cars in an image. We obtained 158 car-body-shadow pairs by this procedure. Then, we regrouped the pairs according to the indices of the regions in the pairs. In this way, the pairs of car-bodies and windows which belong to the same car, including the shadows regions which were over-segmented, were regrouped. However, there was also possible in this step that the bodies of two different cars were in the same group because of the irregularity of the shadow shapes. This happened especially for the cars which were line up. We therefore applied a dividing procedure to divide the car groups if their have shadow regions in between. We found finally 69 car body and shadow groups.
Two sets of car groups, one from car body-windshield groups and another from car body-shadow groups were obtained. The next fusion procedure was to find out the car groups from the two sets, which share the same car bodies according the indices of the car body regions. If both the car groups had the lengths below 55 pixels, the two car groups were fused by an union to represent one car. If the length of the union of the two car groups was more than 55 pixels, the longer car group was chosen to represent the car. If both car groups were longer than 55 pixels, we randomly choose one of them to represent the car. If the length of one car group is longer than 55 pixels and the other's is shorter than 55 pixels, the small one is the choice.

Finally we found 69 car groups to present 64 detected cars. There were 2 double hittings and 3 false alarms in this experiment. There were 3 cars in the image shown in Fig. 2b, which were missed in the detection.

Fig. 4 Car detection (marked with green rectangles) in image of Fig. 2b with 3 false alarms.

5. CONCLUSIONS AND FUTURE WORK

We have proposed a component-based car detection algorithm for the aerial imagery. A set of three spatial bandwidths are used in mean-shift procedures to segment three car components of different sizes. The hierarchical SVM classifier directly follows each segmentation to detect the specific car part and increase the classification accuracy. The radiometric features and geometric features based on the moments of segmented regions are used in the SVM. Pairing and regrouping the detected car components with spatial relationship combine car components into the car. In our experiment, the detection accuracy is 95.5 % and the completeness is also 95.5%. However, a small number of car bodies were missed in our experiment, due to the over-segmentation of large regions. In the next, confidence scores will be assigned to the detected car components and to the declaration of car detection, and the false alarms will be eliminated.

REFERENCES

[1] R.Ruskone, L. Guigues, S. Airault, and O. Jamet, "Vehicle detection on aerial images: a structural approach", Proc. of International Conf. On Pattern Recognition, Vienna, Austria, 900-904 (1996).

[2] C. Schlosser, J. Reitberger, and S. Hinz, "Automatic car detection in high resolution urban scenes based on an adaptive 3D-model", Proc. of the 2nd GRSS/ISPRS Joint Workshop on Data Fusion and Remote Sensing over Urban Area, Berlin, Germany, 167-170 (2003).

[3] T. Zhao and R. Nevatia, "Car detection in low resolution aerial image", Proc. of International Conf. Computer Vision, Vancouver, Canada, 710-717 (2001).

[4] U. Stilla, E. Michaelsen, U. Soergel, S. Hinz, H. J. Ender, "Airborne monitoring of vehicle activity in urban areas", International Archives of Photogrammetry and Remote Sensing, 35, 973-979(2004).

[5] H. Zheng and L. Li, "An Artificial Immune Approach for Vehicle Detection from High Resolution Space Image", IJCSNS, 7, 67-72(2007).

[6] Dorin Comaniciu , Peter Meer, Mean Shift: "A Robust Approach Toward Feature Space Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, v.24 n.5, p.603-619(2002)

[7] R.-E. Fan, P.-H. Chen, and C.-J. Lin. "Working set selection using the second order information for training SVM". Journal of Machine Learning Research 6, 1889-1918 (2005).