Robust component-based car detection in aerial imagery with new segmentation techniques

[a] Ouyang, Yueh, [a]Duval, Pierre-Luc, [a*] Sheng, Yunlong, and [b]Lavigne, Daniel A.

[a]COPL, University Laval, Quebec (QC) G1K 7P4, CANADA
[b] Defence Research and Development Canada-Valcartier, Quebec (QC) G3J 1X5 , CANADA

ABSTRACT

Several new techniques are introduced to the component-based vehicle detection in the aerial imagery. The shape-independent tricolour attenuation model based on the spectral power density difference between the regions lighted by direct sunlight and/or diffuse skylight is used to identify cast shadows. The simple linear iterative clustering (SLIC) performs local clustering for superpixels, which were merged by a statistical region merging (SRM) method based on the independent bounded difference inequality theorem. The car body parts were found with Support Vector Machine based on the radiometric and geometric features of the segmented regions. All the algorithms used in this approach require minimum human intervention, providing a robust detection.

Keywords: Car detection, components-based detection, superpixel, image segmentation, Support vector machine, Tricolour attenuation model.

1. INTRODUCTION

Vehicle detection in aerial images is an interesting topic in digital image processing and plays an important role for the civilian and military uses, such as road network detection, military reconnaissance, and traffic surveillance, which gathers the traffic information including traffic statistics for urban planning[1-5]. The component-based object detection has been used successfully in many applications including the vehicle detection for ground vehicles and in aerial images[6]. As a whole object appearance can change easily by the changes in view angle, in object scale, orientation and illumination and by the occlusion, the component-based approach can provide a robust detection by first detecting small and characteristic parts of object and then combining object parts using an object model. Detecting small object patches may be faster and less sensitive to the distortions than that for whole object. In addition, the object model can be flexible. In the application to vehicle detection in the aerial images, the component-based approach can allow detection of a class of vehicle, as detecting and counting the civil cars, instead of a specific vehicle.

However, at 11.2 cm/pixel resolution in the aerial image only the large car parts as car bodies (hood, roof, and trunk), windshields and shadows can be seen. Car bodies appear with various sizes, shapes and colors. Their images are distorted by low color contrast, specular reflection and noise, especially for the dark-color cars. The windshields and door windows appear as dark narrow regions, which can be confused with shadows and dark color car bodies. The shape of shadows varies constantly with the sunlight direction. In spite the difficulties in the detection, combining the car parts according their specific spatial relationship provides a robust tool to car detection with the detection redundancies[6]. In the previous implementation of the component-based car detection, the mean shift (MS) algorithm was used for low-level image colour segmentation, and the Support Vector Machine (SVM) was used for car-part detection. The MS segmentation depends on the bandwidths of the Gaussian windows, which are parameters to choose, a too large window size can lead to missing small regions and a too small window size can lead to too small segmented regions. Three cascade region detection processes using MS + SVM in each process with decreasing bandwidths in the MS were used to detect first shadows, then car bodies and finally windshields. As the combining of the detected components, using double paring and regrouping process based on the spatial relations among the car parts, was powerful and robust, the civil cars in the parking lot were detected with high accuracy and completeness.

In this paper we report two significant improvements carried out in the previous component-based vehicle detection in the aerial images reported in [6]. First, the shadow detection will no longer depend on the shape of shadows. The shape-independent tricolour attenuation model (TAM)[7], based on the spectral power density difference between the regions

lighted by direct sunlight and/or diffuse skylight, will be used to identify the cast shadows. Second, the simple linear iterative clustering (SLIC)[8] will perform a local clustering of pixels in the spatial and rang joint space, resulting in a dense over-segmentation of image, and in superpixels, which can be merged into regions by a statistical method based on the independent bounded difference inequality theorem[9]. We then found the car parts with Support Vector Machine[10] based on the radiometric and geometric features of the segmented regions. The combination of car bodies, windshields, and shadows according to their spatial relationship, will be used to determine and to count the cars. All the precedent algorithms require minimum human intervention, providing a robust detection.

2. SUPERPIXEL SEGMENTATION

Superpixels are a set of sub-regions that are local and coherent clusters of pixels, preserving the homogeneity and the most image information in the regions. Initial idea of the superpixels instead of pixels is to reduce the complexity of subsequent image processing tasks. The use of superpixels for image segmentation usually results in an over-segmentation of the image, so that a process for merging the superpixels to regions should follow. Ren and Malik[11] proposed using the Normalized Cut[12] to recursively segment an image with contour and texture cues. This method can adapt the image structure by changing the superpixel shapes and avoiding merging salient contours of the image to produce high quality superpixels. However, the computation complexity is high for large images. Levinshtein et al[13] developed Turbopixel algorithm, which generates superpixels by progressively growing seeds in the image plane with geometric flow, which relies on local image gradients. Turbopixel produces uniform size, compact and adaptive-to-boundary superpixels. However, in low contrast aerial images the approach may fail detecting region edges.

2.1 SLIC superpixel detection

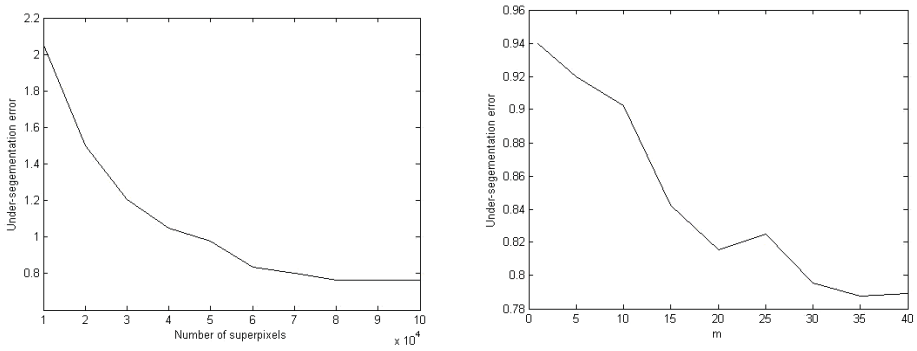
Recently Achanta et al[8] developed the simple linear iterative clustering (SLIC) method. The SLIC may overperform other algorithms such as Normalized Cut, Turbopixel, Mean-shift[14], and Quick-shift[15] by better quality and higher computational efficiency. The superpixels are described by average colors in regions as radiometric features and centroids of regions as geometric features. The SLIC performs a local clustering of pixels in the *lab* color space combined with the *xy* space as a 5D space. In the superpixels detections one usually needs only to decide a desired number of superpixels K as the input parameter, just like the number K of clutters in the K-mean clustering algorithm. Thus, for an image of N pixels, the approximate size of each superpixel is $S = \sqrt{N/K}$ pixels. Usually one prefer the roughly equally sized superpixels, so that the superpixel centers are first set at every grid of interval S. Then, the superpixel centers are moved to the lowest gradient position in the color space for avoiding the superpixels to contain salient edges of the image. After that, each pixel in the image is associated to the nearest cluster center, whose search area of size S overlaps this pixel. After all the pixels are associated, the new centers of the clusters are computed as the centroids of the newly formed superpixels, according to the average distance, *i. e.* the first order moments in the 5D space. The clustering proceeds iteratively until the cluster center converging. There are three points, in which the SLIC is different from the K-mean cluttering. First, SLIC localizes pixel search in an area of ($2S$, $2S$), rather than in the whole image plane. Second, SLIC starts the clustering from the centroids which are distant from salient edges. Third, SLIC uses a distance measure D_s defined as follows:

$$D_s = d_{lab} + \frac{m}{S} d_{xy}, \tag{1}$$

where d_{lab} is the *lab* distance in CIELAB color space and d_{xy} is the spatial distance in the image. The variable m is used to control the compactness of superpixels. The larger the value of m, the higher weight for the spatial distance d_{xy} and the more compact the clusters. In the application to image segmentation we observed that the smaller the size S of the superpixels, the smaller the size of segmented regions, and that the use of larger size superpixels can miss segmenting small regions. On the other hand, over-segmentation by too small segments will require heavy merging tasks. In an image of N =1000 x1000 pixel size we chose k = 66666 so that S = 3.87 pixels. Our experiments showed that the value of S smaller than that value will lead to failure in producing a good segmentation. In fact, the optimal values of K and m have been determined by an under-segmentation error testing, which is a measure of errors in the image segmentation compared with the ideal segmentation. For a set of superpixels $\{s_j\}$ and idea segment regions $\{g_1, \ldots\ldots g_M\}$, the under-segmentation error is defined as:

$$U = \frac{1}{N}\left[\sum_{i=1}^{M}\left(\sum_{[s_j | s_j \cap g_i > B]} |s_j|\right) - N\right], \tag{2}$$

where $|s_j|$ is number of pixels in superpixel s_j, $s_j \cap g_i$ is the intersection of the superpixel s_j and the ideal segmentation g_i, and B is the required minimum number of overlapped pixels, which is set usually to be 5% of $|s_j|$. Thus, the under-segmentation errors for segmenting the car windshields were investigated. U decreased as the superpixel number K and the compactness parameter m increased. For small K and small m, the errors descanted quickly. Then, the descent rates slowed down after K > 60000 and m>35. We chose K = 66666 and m = 40 for S = 3.87 with the consideration for the computation efficiency.



(a) (b)

Fig. 1. Under-segmentation errors as a function of (a) Number of superpixels K as m=40; (b) Compactness control parameter m as K = 66666.

2.2 Superpixel merging

The statistical region merging (SRM) algorithm was proposed by Nock et al to merge the over-segmenting SLIC superpixels into regions. The SRM does not require setting the threshold, and therefore, can adapt various imaging conditions[16]. The algorithm is based on the independent bounded difference inequality theorem as: Let $x_1, \ldots xi \ldots\ldots, x_n$ be n independent random variables and the vector $\boldsymbol{x}$ differ from $\boldsymbol{x'}$ only in a x_i coordinate, and for a real-valued function $f(x)$ the difference

$$\sup_{x_1,\cdots,x_n,x_i} \left| f(x_1,\cdots,x_{i-1},x_i,x_{i+1},\cdots,x_n) - f(x_1,\cdots,x_{i-1},x'_i,x_{i+1},\cdots,x_n) \right| \le c_i,$$

for i =1,.....,n. Then, for any $\tau > 0$, $f(x)$ and its expectation $E\{f\}$ the probability

$$\Pr[(f - E\{f\}) \ge \tau] \le \exp(-2\tau^2 \Big/ \sum_{i=1}^{n} c_i^2) \tag{3}$$

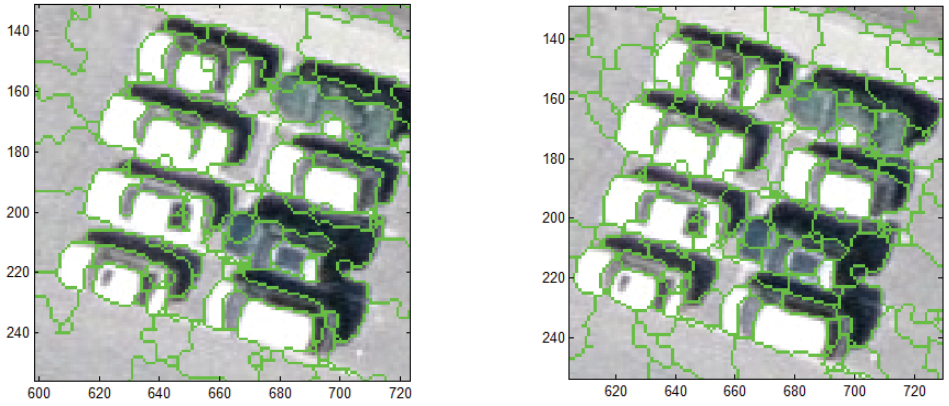
From the independent bounded difference inequality theorem for a given pair of two superpixels s and s' in an image of a single color band with the individual average intensity in each superpixel expressed as $I(s)$ and $I(s')$, respectively, the merging predicate for s and s' can be derived as:

$$P(s,s') = \begin{cases} true & if \quad |I(s) - I(s')| \le g\sqrt{\dfrac{1}{2Q}\left(\dfrac{\ln 6NN_b(s)}{|s|} + \dfrac{\ln 6NN_b(s')}{|s'|}\right)} \\ false & otherwise \end{cases} \tag{4}$$

where N is the total number of pixels in the image, $|s|$ is number of pixels in the superpixel s, g is the maximum value of intensity, Q is a parameter for statistical complexity and N_b is the upper-bound number described as:

$$N_b(s) = (|s|+1)^{\min\{g,|s|\}}$$

If the two superpixels have the same expectation value in intensity, from Eq. (4) we can infer that their merging predicate will be true. For a color image, the merging predicates are computed in the R, G, and B band respectively to form a set of three merging sub-predicates. Thus, the merging predicate in a color image is true only when all the three merging sub-predicates are true. The parameter Q is to control the statistical complexity of the optimal segmented image. In fact, Q controls coarseness of the segmentation. Using a set of Q values, we can build hierarchical segmentations of an image. However, only the homogeneity in color is considered in this merging predicate. Sometimes over-merging can still occur as we observed in many cases. For instance, in Fig. 2(a), hoods, roofs and trunks in the two white color cars are over-merged. To avoid the over-merging we added a contour constrain into the SRM, that the ratio of the circumference to the square root of the area of a merged region must be less than 6.5 in order to avoid merging car components to a region of too complex shape. The statistical merging result with contour constrain is shown in Fig. 2(b).



(a) (b)

Fig. 2 Superpixels merging with (a) statistical merging predicate; (b) merging predicate with contour constrain

3 SHADOW DETECTION

Shadows are omnipresent in aerial images in the sunny days and provide additional clues about the shape and position of its casting object. Hence, detecting or removing shadows are a basic important step for aerial image processing. Shadow detection methods can be classified into property-based and physics-based. Property-based techniques identify shadows through shadow features, such as color, edge, histograms, texture, color ratios, geometry property moments and gradient. In practice, most methods use more than one combined features. Following the feature extraction, the threshold method, Bayesian approach, SVM and adaptive boosting etc. are usually used for classification of shadows in the property-based techniques. However, one notes that the features based on the shape of shadows can be not reliable, as the shape of shadows vary constantly with direction of the sun light projection, so that we have to ensure the training and testing set images were taken in the same sunny conditions, or to build and use a mass training image set, which includes all possible shapes of the shadows cast by objects.

Physics-based techniques are independent of the shape and position of shadows. From some prior knowledge on lighting, scene conditions and camera calibration, the shadows are found after physical derivation. Based on the assumptions of Lambertian reflectance, approximate Planckian lighting and narrowband camera sensors, Finlayson proved that pixels with several different lighting could form a straight line in a 2D log-chromaticity space. The direction of this straight line is referred to as an invariant direction. If we project these pixels onto a line perpendicular to invariant direction, the pixels will converse to a point, and if we project all image pixels onto the same line, we could get an L1-chromaticity intrinsic reflectivity image that is free of shadow. The invariant direction depends on the camera sensor. The images shunted by the same camera have the same invariant direction. The invariant direction of image could be obtained via minimizing the entropy of a greyscale image. It is easy to remove and detection shadow with Finlayson's intrinsic image method. However, the method is only suitable for the images obtained from a fairly narrowband camera. The method is also sensitive to noise because noise would break the narrowband camera assumption. It fails to get intrinsic images from our aerial images.

As there are two kinds of light sources in the outdoor scenes: sun light and sky background light coming from the scattered sun light and according to the Rayleigh scattering theory, the sky light has higher saturation in shorter wavelength, and as the direct sun light is absent in the shadows cast by the objects, the shadow regions will have higher saturation in blue. Polidorio utilized this simple feature to detect shadows in the color aerial images. Tian et al[7] proposed a more involved model, referred to as the tricolor attenuation model (TAM) for shadow detection based on the similar idea that in aerial images attenuation of the red band is lager than that of the blue band. The non-shadow regions are illuminated by day light that hybrids the sun light and the skylight, and the shadow regions are illuminated only by sky light, which is equal to the day light subtracted by sun light. Thus the attenuations of the signals in R, G and B color bands by the shadow casting R_{NS}-R_S can be written as:

$$R_{NS} = \sigma S(\lambda_R) E_{day}(\lambda_R) Q_R \quad and \quad R_S = \sigma S(\lambda_R)(E_{day} - E_{sun})(\lambda_R) Q_R \tag{5}$$

$$R_{NS} - R_S = \sigma S(\lambda_R) E_{sun}(\lambda_R) Q_R \tag{6}$$

where R_{NS} is that from non-shadow regions and R_S is that from shadow regions, E_{day} and E_{sun} is the spectral density function of the Planckian lighting of daylight and sunlight, respectively, σ is exposure time, S is surface reflecting function and Q_R is the value of the camera response function in R band at wavelength = 650 nm. G_{Ns}- G_S, and B_{NS}-B_S can written in a similar way as

$$\begin{bmatrix} \Delta R \\ \Delta G \\ \Delta B \end{bmatrix} = \begin{bmatrix} \sigma S(\lambda_R) E_{sun}(\lambda_R) Q_R \\ \sigma S(\lambda_G) E_{sun}(\lambda_G) Q_G \\ \sigma S(\lambda_B) E_{sun}(\lambda_B) Q_B \end{bmatrix}$$

In order to eliminate the surface reflection and the camera response we have from Eq. (1) :

$$\sigma S(\lambda_R) Q_R = R_{NS} / E_{day}(\lambda_R)$$

so that the attenuation vector is normalized by the corresponding signals in the B band that

$$\begin{bmatrix} \Delta R \\ \Delta G \\ \Delta B \end{bmatrix} = \begin{bmatrix} \dfrac{E_{day}(\lambda_B) E_{sun}(\lambda_R)}{E_{day}(\lambda_R) E_{sun}(\lambda_B)} \cdot \dfrac{R_{NS}}{B_{NS}} \\ \dfrac{E_{day}(\lambda_B) E_{sun}(\lambda_G)}{E_{day}(\lambda_G) E_{sun}(\lambda_B)} \dfrac{G_{NS}}{B_{NS}} \\ 1 \end{bmatrix} \Delta B = \begin{bmatrix} 1.31 \dfrac{R_{NS}}{B_{NS}} \\ 1.19 \dfrac{G_{NS}}{B_{NS}} \\ 1 \end{bmatrix} \Delta B \tag{7}$$

where the factors before R_{NS}/B_{NS} and G_{NS}/B_{NS} in the second vector in Eq. (7) can be computed by the Plank radiation law with the day light temperature of 5500 K and the sun light temperature of 6500 K, resulting in the third vector in Eq. (7). Thus, using the signals from a given non-shadow region as reference, we can measure the attenuation vector of a pixel and decide it belong or not to shadow, if the attenuation vector is larger or smaller than that computed by Eq. (7). As one does not know in priori the non-shadow area in the images, the original TAM uses an expensive iterative assumptions and checking process to determine the non-shadow regions. In the aerial images, the non-shadow reference area can be easily identified, by choosing the largest homogenous regions, corresponding to parking lot or field that are always displayed in the aerial scenes, as ideal non-shadow reference regions. Therefore, the TAM process is simplified with the original 12 steps reduced to 3 steps only.

In our experiments, the TAM method was applied to the SLIC superpixels instead of pixels. We first over segmented the input aerial image into the superpixels by the SLIC algorithm. Merging these superpixels with the SRM was necessary for extracting the largest regions in the image and considering them as non-shadow regions for the reference in the TAM. However, the TAM shadow detection was applied to the superpixels before the merging. In a test image, as shown in Fig. 4b, 4696 shadow superpixels were identified by TAM, independently on their shapes and positions, 1630 among

them were belonged to real shadows of the cars. The others were cast by other objects like buildings. There were 467 superpixels of the dark-color car bodies and 563 superpixels of car windshields were detected as shadows. The detection accuracy was 61.28% without taking into account shadows of the buildings and in the forest.

A large part, 467+563, superpixels on the car components was miss-classified as shadows by TAM not only lowers the detection accuracy, but also jeopardize the component-based car detection. In fact, the most difficult step in the component-based car detection in the aerial images is segmenting correctly dark-colored car bodies, windshields and shadows. Thus, global and local thresholding processes based on the non-parametric Ostu method[17] have been applied in order to recuperate the miss-classified car components. Again, we choose the Ostu threshold method, because it does not require an input threshold value, and is therefore independent of the images. Otsu's method computes an optimal threshold, which best separates the ensemble of pixels into to two groups in the histogram.

In the global thresholding the Otsu's method is applied to the ensemble of superpixels, which were classified as shadows by the TAM. The superpixels in the higher intensity group are returned into non-shadow group with the separability factor (SP) calculated as

$$SP = \frac{\sigma_B^2}{\sigma_T^2},$$

where σ_B^2 is the intra-class variance and σ_T^2 is the total variance of the intensity. The superpixels with the SP value higher than 0.55 were classified and moved to non-shadow group. Only those with the SP value less than or equal to 0.55, remained. After the global Ostu thresholding, the total number of shadow superpixels decreased from 4696 to 2495. The number of hit was 1131. The number of miss-classified car components was reduced to 147 for car body and 238 for windshields. The shadow detection accuracy increased to 74.64%.

Then, in the local thresholding we merged the remaining superpixels surviving from the global thresholding by the SRM, and we applied the Otsu's method to each of the merged shadow regions. It is unlikely that the local threshold would miss shadows, but rather recover some non-shadows car components from the shadows. After the local Otsu thresholding, only 38 car body superpixels and 107 windshield superpixels were miss-classified as shadows. Although the number of hitting superpixels decreased to 867. The detection accuracy was increased to 85.67%. The shadow superpixels were processed with morphology close operation to eliminate some small gaps exited in shadow region. The gaps were the noise pixels in shadow region, which were threshold off by local threshold. Finally, the shadow superpixels were merged into shadow region again for car detection. The shadow detection results in the precedent are reported using the number of superpixels. In terms of the segmented regions, among the 92 shadow regions, which were manually selected in testing image, 82 shadow regions were detected correctly. There were 40 shadows which are not cast by cars, and 10 shadow regions missed, as shown in Fig. 5b. Note that a car can cast two shadows. One is beside the car body. Another is in the front or rear side. All the cars shown in Fig. 3b have at least one shadow region. There are also 7 detected shadow regions, which contain a part of windshields.

(a) (b)

Fig. 3 (a) Detected shadows (marked by green) with the tricolor attenuation model (b) Shadows after the successive global and local thresholding on (a)

4 CAR BODY DETECTION

Aerial images of 1000 x1000 pixel size, as shown in Fig. 4, were used in the experiment. For detecting car body components, we first removed the superpixels, which were classified as shadows. Then, we merged the remained superpixels with the SRM. According to the a priori knowledge on the car size of about 20 x 40 pixels, we eliminated the merged regions larger than 1000 pixels in first place. As all the feature parameters including the region areas are normalized by scaling in the SVM, removing the regions of too large areas increases the normalized area feature values and enhances the differences between features of small regions. There were 2299 regions the in test image. The same procedure was applied to the training image, resulting in 2563 regions. For training the SVM all the 435 car bodies in the image were manually selected and labelled by +1 in the training image. The remaining regions are labelled as -1. Radiometric and geometric features were thus used for describing all the regions. The RGB color model is suitable for color display but not for color analysis because of the 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 were used as the radiometric parameters. The area and circumference of the region were chosen as the size features. The circumference of a region was obtained by morphological operation. The aspect ratio and the rectangleness of regions are important features permitting distinguishing car component from natural objects. Car components can be approximated as rectangles with the two side lengths determined by the second order moments as

$$a = 2\left(\frac{m_{20}+m_{02}+\sqrt{(m_{20}-m_{02})^2+4m_{11}^2}}{m_{00}/2}\right)^{1/2} \quad \text{and} \quad b = 2\left(\frac{m_{20}+m_{02}-\sqrt{(m_{20}-m_{02})^2+4m_{11}^2}}{m_{00}/2}\right)^{1/2}, \tag{8}$$

so that the aspect is a/b and the rectangleness is the ratio (area of the region)/ab, where m_{00} is the zero order moment of region and m_{11}, m_{20}, and m_{02} are the second order moments of the region. Thus, the seven features of region used in SVM are intensity, hue angle, saturation, area, circumference, aspect and rectangleness.

(a) (b)

Fig. 4. Aerial images used as (a) training image; (b) testing image

The SVM is a powerful tool for data classification based on statistical learning theory. This algorithm performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. SVM is closely related to neural networks. In fact, a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network. The classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains a class label and several features. The data instances in the testing set are given only the features. The SVM produces a model to predict the class label value of testing instances.

Given a training set of instance-label pairs ($\mathbf{x}_i$; y_i); i = 1, 2,..., l where $\mathbf{x}_i \in R^n$ and $y_i \in \{1,-1\}^l$, SVM requires the solution of the following optimization problem:

$$\min_{\mathbf{W},b,\xi} \frac{1}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w} + C\sum_{i=1}^{l}\xi_i \qquad (9)$$

$$\text{subject to} \quad y(\mathbf{w}^T\phi(\mathbf{x}_i)+b) \geq 1-\xi_i \quad with \quad \xi_i > 0$$

where $\mathbf{x}_i$ is training feature vector, y_i is the class label, ϕ is the mapping function that maps the feature vectors $\mathbf{x}_i$ into a higher dimensional space. The hyperplane for classification can be written as

$$\mathbf{w}^T\phi(\mathbf{x}_i)+b=0 \qquad (10)$$

where vector $\mathbf{w}$ is a normal vector: it is perpendicular to the hyperplane. The parameter $\frac{b}{\|\mathbf{w}\|}$ determines the offset of the hyperplane from the origin along the $\mathbf{w}$ normal vector. $C > 0$ is the penalty parameter of the error term. The kernel function K is defined as $K(\mathbf{x}_i - \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$. There are four basic kernel functions; linear, polynomial, radial basis function (RBF), and sigmoid. We chose RBF kernel function because the RBF kernel can handle the case when the relation between class labels and features is nonlinear. The RBF kernel function is defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma|\mathbf{x}_i - \mathbf{x}_j|^2) \qquad (11)$$

where γ is kernel parameter. Scaling data and searching the optima values of parameters are the two important steps for SVM. Linearly scaling each feature to the range [0; 1] is necessary for SVM because it can avoid not only the features in greater numeric ranges dominate those in smaller numeric ranges, but also numerical difficulties during the calculation. For scaling feature, the features are divided by the maximum value of this kind of feathers in training data. For example, in the size features used in our case, the maximum area in training instances is 1000 pixel, and then all of the size feature in both training and test data are divided by 1000. The accuracy of an SVM model is largely dependent on the selection of the penalty parameter C and the kernel parameter γ. There are two methods to find the best parameters: a grid search and a pattern search. A grid search scans each parameter across a specified search range using geometric steps. A pattern search starts at the center of the search range and makes trial steps in each direction for each parameter. We use the grid method to find the best parameter because of its simplicity. The training set is divided into 5 subsets of equal size. One subset is tested by SVM with remaining 4 subsets as training set. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. It is a cross-validation method. The cross-validation accuracy are calculated with the parameters increased exponentially such as C = 2^{-5}, 2^{-3},…, 2^5 and γ = 2^{-15}, 2^{-13} ,…., 2^3. The best parameter is picked by the highest value of cross-validation accuracy. In our experiment, the best parameter C=2.6 and γ=10 for the cross-validation accuracy = 95.59%.

After the training the SVM model was applied to testing image shown in Fig. 5a. There are 175 car body regions in the testing image. There were 153 hits for the car bodies (marked by green) classified correctly , 17 false-alarms (marked by red) and 22 car body regions missed (marked by blue), as shown in Fig. 5a. The detection accuracy is 90.00% and the completeness is 87.42%. In the previous implementation of the component-based car detection with the mean shift algorithm for low-level image colour segmentation, and the SVM for car body detection, there were 148 hits for the car bodies (marked by green) classified correctly , 16 false-alarms (marked by red) and 27 car body regions missed (marked by blue), as shown in Fig. 5b. The detection accuracy is 90.24% and the completeness is 84.46%. Thus, the new method detected more 5 car body, increased the detection completeness from 84.46% to 87.42%, and decreased the number of missed car body from 27 to 22.

(a) (b)

Fig. 5. Detected car bodies (marked by green), False alarms (marked by red), and missing (marked by blue) with (a) new segmentation technique and SVM; (b) mean-shift segmentation and SVM

5. CONCLUSIONS

By including the three color attenuation model we significantly increased the reliability of the shadow detection. By using the SLIC superpixels, statistic region mergin (SRM) and global and local Otsu's thresholding we also improved the detection of vehicle bodies. We notice that the SLIC superpixels preserved well edges of regions, even for that between shadow, windshields and dark-color car bodies. This is promising for solving the difficult problem with dark-color vehicle detection. However, over-merging of shadows with windshields and dark car bodies can still occur in the SRM. We selected all the above algorithms for them to have minimum human interventions and therefore more robustness in the detection.

REFERENCES

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

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

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

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

[6] Ouyang Y., Sahli S., Sheng Y. and Lavigne D. A.,"Robust component-based car detection and counting in aerial imagery", Proc. SPIE Defence an security (2001)

[7] Tian J., Sun J., and Tang Y., "Tricolor attenuation model for shadow detection", *IEEE Trans. Imag. Proc*. 18(10), pp 2355-2363, 2009

[8] Achanta R., Shaji A., Smith K., Lucchi A., Fua P., and Susstrunk S., "SLIC superpixel", EPFL Technical Report 149300 (2010)

[9] McDiarmid C., "Concentration," Probabilistic Methods for Algorithmic Discrete Math., M. Habib, C. McDiarmid, J. Ramirez-Alfonsin,and B. Reed, eds., pp. 1-54, Springer Verlag, 1998.

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

[11] Ren X. and Malik J., "Learning a classification model for segmentation", Int. ICCV, 2003.

[12] Shi J. and Malik J., "Normalized cuts and image segmentation". IEEE Trans. PAMI, 22:888–905, 2000

[13] Levinshtein, A., Stere, A., Kutulakos, K., Fleet, D., Dickinson, S., Siddiqi, K.: “Turbopixels: Fast superpixels using geometric ows”, IEEE Trans. PAMI (2009)
[14] Comaniciu D., Meer P., Mean Shift: “A Robust Approach Toward Feature Space Analysis”, IEEE Trans. PAMI, v.24 n.5, p.603-619(2002)
[15] Vedaldi, A., Soatto, S.: “Quick shift and kernel methods for mode seeking” ECCV(2008)
[16] Nock R. and Nielsen F., “Statistical region merging,” IEEE Trans. PAMI, 26(11), 1452–1458(2004).
[17] Otsu N., “A threshold selection method from gray level histograms,” IEEE Trans. Syst., Man, Cybern., vol. SMC-9(1), 62–69(1979).