VEHICLE DETECTION FROM AN IMAGE SEQUENCE COLLECTED BY A HOVERING HELICOPTER

Fatemeh Karimi Nejadasl[a] and Roderik C. Lindenbergh[b]

[a]Department of Molecular Cell Biology, Leiden University Medical Center
P.O. Box 9600, 2300RC Leiden, The Netherlands, fkariminejadasl@lumc.nl
[b]Dept. of Remote Sensing, Delft University of Technology
Kluyverweg 1, 2629 HS, The Netherlands, r.c.lindenbergh@tudelft.nl

Working Groups I/2, III/1, III/4, III/5

KEY WORDS: Foreground and background identification, vehicle detection, temporal profile, sequential method, maximum frequency

ABSTRACT:

This paper addresses the problem of vehicle detection from an image sequence in difficult cases. Difficulties are notably caused by relatively small vehicles, vehicles that appear with low contrast or vehicles that drive at low speed. The image sequence considered here is recorded by a hovering helicopter and was stabilized prior to the vehicle detection step considered here. A practical algorithm is designed and implemented for this purpose of vehicle detection. Each pixel is identified firstly as either a background (road) or a foreground (vehicle) pixel by analyzing its gray-level temporal profile in a sequential way. Secondly, a vehicle is identified as a cluster of foreground pixels. The results of this new method are demonstrated on a test image-sequence featuring very congested traffic but also smoothly flowing traffic. It is shown that for both traffic situations the method is able to successfully detect low contrast, small size and low speed vehicles.

1 INTRODUCTION AND TEST DATA DESCRIPTION

Traffic is a problem of all large cities and is continuously analyzed by both authorities and researchers. Driving behavior is the most influential element in traffic and still less is known about it. This is due to the lack of instruments to track many vehicles for a long period of time without their awareness in taking part in an experiment (Ossen, 2008).

For the purpose of studying driving behavior in real traffic situations (Ossen et al., 2006; Ossen, 2008), a freeway is observed by a camera mounted below a hovering helicopter. The helicopter flies in the range of 300 to 500 m above the freeway and records image sequences for a period of maximum half an hour. The camera used for this purpose has a black and white visual sensor, a frequency of 15 frames per second and a resolution of 1392×1040 pixels. An area of 300 to 500 meter on the ground was covered with a spatial resolution of 20 to 50 centimeter. In this paper, a test data set is considered of an image sequence with 1501 images. The image sequence is previously stabilized on the road area which is the region of interest in this research.

In this paper, vehicles and background are identified from the image sequence of the stabilized road area. The specific problems of the vehicle detection in our data set are caused by slow moving vehicles,vehicles that appear with low contrast and small vehicles with a low amount of detail.

A vehicle here is extracted as a cluster of pixels (blob). First pixels are divided into two groups, one group consists of background pixels and the other group corresponds to moving objects. The division is performed based on the temporal profile of each pixel. The foreground pixels are grouped as a blob.

The paper is organized as follows. In Section 2, a brief review of related literature is presented. In Section 3, the foreground and background identification method is sketched and results are presented in Section 4. Conclusions are drawn in Section 5.

2 RELATED WORK

Vehicles can be detected by model-based methods, a 2D or 3D shape or/and an intensity template for a vehicle. The objective is to find this template back in each image considered, (Ballard, 1981; Ferryman et al., 1995; Tan et al., 1998; Pece and Worrall, 2002; Hinz, 2003; Zhao and Nevatia, 2003; Dahlkamp et al., 2004; Pece, 2006; Ottlik and Nagel, 2008). The disadvantage of model-based methods is the high dependency on geometric details of the considered object, which in our case would require that vehicles should appear in the image sequence with many details and with clear boundaries. In our data set, the shape and appearance of cars are simple are lacking detail. The similarity of some of the vehicles to road stripes, moreover, may cause a failure for the model-based methods.

Stauffer and Grimson (1999) modeled the background PDF as a mixture of Gaussians: usually three to five Gaussian models are enough for a complex background with illumination variation and background movement such as swaying trees. The value of each pixel in time, intensity or color, is a mixture of Gaussians. The parameters of the mixture of Gaussians model are weight, mean and standard deviation of each Gaussian model. They are estimated in an adaptive way. For every new image, the new observation, a pixel value, only updates the Gaussian parameters it belongs to. If a new observation does not belong to any Gaussian model, it constructs a new Gaussian model. The last Gaussian model, the Gaussian with the smallest weight, is combined with the Gaussian model with the second smallest weight. As a result, this pixel is assumed to belong to a moving object. Each parameter of the Gaussian model is updated as a combination of its previous value and the value of the new pixel. The weight of the pixel value is set by a learning parameter. A larger learning value increases the chance of wrongly modeling the object as background. A smaller value, however, cannot be used for very fast changing backgrounds. After background pixel identification, object pixels are connected to reconstruct the blob. In Stauffer and Grimson (2000), blobs are tracked using a Kalman

filter with multiple models. Every time a pool of blobs is evaluated against a pool of models. The model that explains the blobs best is used as a tracking result.

Pece (2002) assumed a mixture of probability likelihood models for both background and object clusters. The probability model of a cluster is the multiplication of position and gray-level PDF models. The background position is assumed to have a uniform distribution. The background is subtracted from each image to construct the difference image. The gray levels of the difference-image follow a Laplacian distribution (two-sided exponential). The object model for the position has a Gaussian or a top-hat distribution and the gray level has a uniform distribution. Expectation maximization is used to estimate the parameters of the PDF models. Clusters are analyzed for merging, splitting or making a new cluster. The current position of each cluster initializes the position of the cluster in the next image. The expectation maximization calculates the new location of the clusters. The camera should be fixed.

In Elgammal et al. (2002), a PDF model of a pixel value, color or intensity, is modeled as a mixture of nonparametric kernel models. The kernel model assumes a Gaussian shape, where the mean is the value of the pixel in the previous image and the the variance is defined as $\frac{1}{0.68\sqrt{2}}$ of the median value of the pixel temporal profile. A set of a recent pixel values is used for estimation of the PDF model of the pixel value. The pixel is considered as an object pixel if the pixel probability is less than a predefined threshold. The pixels assigned as background are grouped as a region. The region is only considered as background if the maximum probability of the region is larger than a predefined threshold. Otherwise the region is classified as an object. Remaining falsely detected background pixels are removed by a test against another predefined threshold value based on the product of the probability values of pixels in the connected region: if this product is lower than the threshold, the region is considered an object. The images should be obtained without camera motion.

Inspired by the works of Stauffer and Grimson (1999); Pece (2002); Elgammal et al. (2002), we have designed and implemented a practical approach to identify vehicles in the data set that also consists of vehicles, which are either moving slow, or are small, or are having low contrast. Such vehicles are, in existing approaches, easily grouped as belonging to the background.

3 VEHICLE DETECTION

Differences between a specific image and consecutive images can be used to initialize the motion detection. The illumination variation is considered negligible between consecutive images. The difference images have values around zero for a background and values larger than zero for locations where moving objects are present. At those locations where vehicles overlap, usually values near zero occur as well. The region after and before the overlap area shows significant differences (Figure 1).

Non-zero areas in a difference image will be relatively small in case either slow vehicles or low contrast vehicles are present. As a consequence, a low contrast, small sized or slow vehicle will be easily discarded as noise.

The solution to the slow motion problem is to use a background image, an image without any vehicles, instead of consecutive images to highlight the vehicles (Figure 2). The problem however is how to construct such background image considering illumination variation.
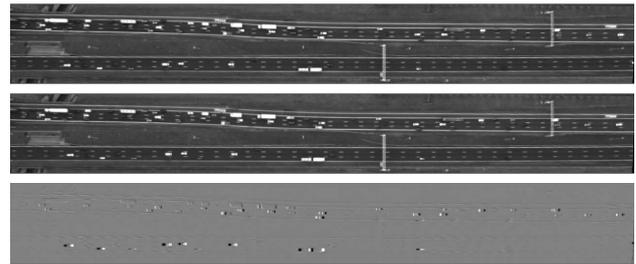
Figure 1: The difference image (bottom) is obtained by subtracting the middle from the top image. The top and middle images are consecutive.
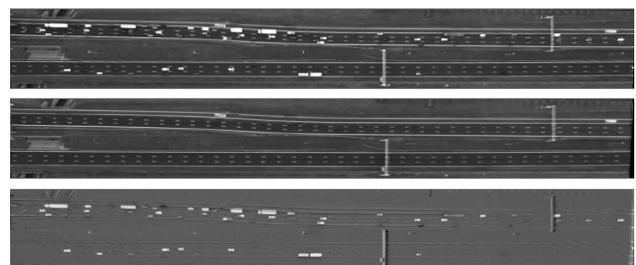
Figure 2: The difference image (bottom) is obtained by subtracting the background image (middle image) from the top image.

For every pixel, a time series of gray level observations can be constructed by analyzing the image sequence, which is called a temporal profile. The histogram of gray levels in each profile shows a distinct peak for the gray level of the background. The shape of the histogram is represented by a single Gaussian when the illumination variation is gradual and when the most of the time the pixel is not occupied by vehicles. The tail of the Gaussian shows the gray value of the foreground (Figure 3).
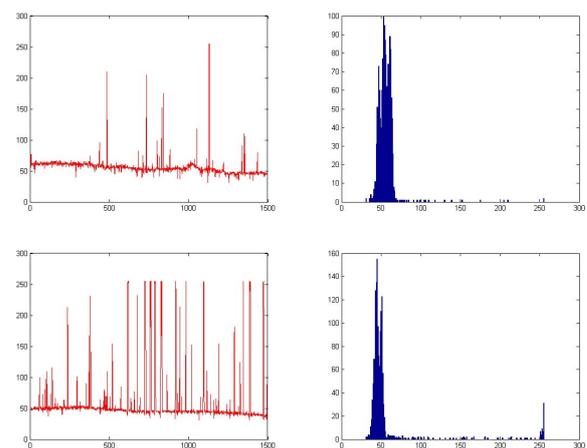
Figure 3: Top: the left graph is the temporal profile of a selected pixel gray level as a function of time, that is, the gray level (y-axis) is given as a function of the image number in the image sequences (x-axis). The top right graph shows the histogram of gray levels of the same profile. Bottom: similarly for another pixel.

The distribution of the gray levels in the profile of a particular pixel, therefore, identifies which pixel value belongs to the background and which one to the foreground. The most reliable way to determine the frequency distribution of the gray levels is by

processing all available observations, i.e. all images including a particular pixel. This operation requires a relatively long time however.

We have developed an efficient and fast way to identify background pixels on the basis of observed gray level frequencies without processing all available observations. This procedure is based on a sequential search on the gray level temporal profile for each pixel.

All pixels in the first image are assigned to the background class. Next, for each pixel, the gray level in the next image is compared to the background value. If the value is within ϵ gray levels of the background gray-level, the pixel is classified as background and the background value is updated to the value of this pixel. Here $\epsilon = \pm 6$ is used for an image gray-level range of $0 - 255$. The frequency of the background is updated by one. If the value falls outside an ϵ gray level interval, the pixel is classified as a foreground (1st identified object). Then the following image is processed, i.e. the following observation for any given pixel is analyzed in the same way as in the previous step.

The comparison with the current background gray level and statistics is done exactly as in the previous step. If the new observation falls within a +/- 6 gray level interval of the value associated with the 1st identified object, the pixel value and its probability are updated in the same way as for the background. If it falls outside this range a new object with a frequency of one is created (last identified object). The following observation is compared with the background and last identified object gray-levels. When a gray level is observed which falls within the ϵ range of either background or last identified object, the corresponding gray-level and frequency are updated. If the gray-level falls outside both ranges, a new object is identified with a frequency of one. The two previously identified objects are retained but not used for the analysis of subsequent observations. When a new observation falls outside the gray level ranges associated with the background and the last identified object, the oldest object in the list is removed.

The gray level of the last identified object is also compared with the gray level of its predecessor. If it falls inside the range associated with the latter, the frequencies are added and the predecessor is removed. The frequencies of the current background and last identified objects are compared and the larger frequency is assigned to the background class and the lower one to the last identified object. The corresponding gray levels are exchanged as well. This procedure corrects the initial assignment of all pixels to the background class. Moreover, it prevents erroneously assigned background gray levels from propagating to other images. Likewise, the procedure improves progressively the estimate of the last-identified-object gray level. Figure 4 shows the result of the background/foreground pixel identification.

The details of this algorithm is described in Algorithm 1.

The foreground pixels are connected by a 8-neighborhood connected component to cluster a group of foreground pixels, called a blob. The blob here represents the vehicle. Very small blobs are assumed to be caused by noise and are removed by morphological opening with a 3×3 structural element (Gonzalez and Woods, 2007). Figure 5 represents the extracted blobs.

4 RESULTS AND DISCUSSIONS

Algorithm 1 is applied on the test data consisting of 1501 images. Figure 6 shows the identified vehicles in yellow. Two regions

Algorithm 1: Near real-time foreground and background pixel identification. V and P represent respectively a value and a probability. The subscripts b, f, $f1$, and $f2$ denote background and three reference objects respectively. **0** and **1** are matrices with all elements equal to zero and one respectively. The subscript one for image I and classified image BF indicates the first image.

Input: Stabilized image-sequence of the road area ($\{I\}$) initialized by $BF_1 \leftarrow \mathbf{1}$, $V_b \leftarrow I_1$, $P_b \leftarrow \mathbf{1}$, $V_f, P_f, V_{f1}, P_{f1}, V_{f2}, P_{f2} \leftarrow \mathbf{0}$

1   **for** $f = image\ number$ **do**
2   **for** $y = row$ **do**
3   **for** $x = column$ **do**
4   **if** $I(x,y) \in [V_b(x,y) \pm \epsilon]$ **then**
5       $V_b(x,y) = I(x,y,f)$
6       $P_b(x,y) = P_b(x,y) + 1$
7       $BF(x,y,f) = 1$
8   **else**
9   **if** $I(x,y,f) \in [V_f(x,y) \pm \epsilon]$ **then**
10        $V_f(x,y) = I(x,y,f)$
11        $P_f(x,y) = P_f(x,y) + 1$
12        $BF(x,y,f) = 0$
13   **else**
14        $V_{f1}(x,y),\ P_{f1}(x,y) \rightarrow V_{f2}(x,y),\ P_{f2}(x,y)$
15        $V_f(x,y),\ P_F(x,y) \rightarrow V_{f1}(x,y),\ P_{f1}(x,y)$
16        $V_f(x,y) = I(x,y,t),\ P_f(x,y) = 1$
17        $BF(x,y,f) = 0$
18   **if** $V_f(x,y) \in [V_{f2}(x,y) \pm \epsilon]$ **then**
19       $P_f(x,y) = P_{f2}(x,y) + P_f(x,y)$
20       $V_{f2}(x,y) = 0,\ P_{f2}(x,y) = 0$
21   **if** $P_b(x,y) < P_f(x,y)$ **then**
22       $V_f(x,y),\ P_f(x,y) \leftrightarrow V_b(x,y),\ P_b(x,y)$
23       $BF(x,y,f) = 1$

Output: New image-sequence with background and foreground identified($\{BF\}$)
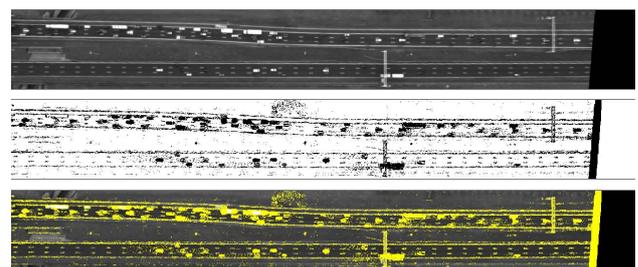
Figure 4: Background/foreground pixel identification. From top to bottom: 1) an arbitrary image, 2) discrimination of the background (white) and foreground (black) pixels, and 3) the detected foreground pixels highlighted.

from the upper and lower part of the road are considered in the zoom-in.

The upper lanes in the road image shows congested traffic while the lower lanes have fluid traffic. To evaluate the performance of the identification method on these two different type of traffic, seven images are selected and the results are displayed in
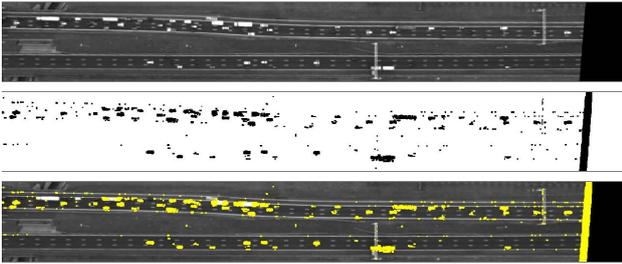
Figure 5: Vehicle identification. From top to bottom: 1) an arbitrary image 2) extracted blobs (black), and 3) highlighted extracted blobs.
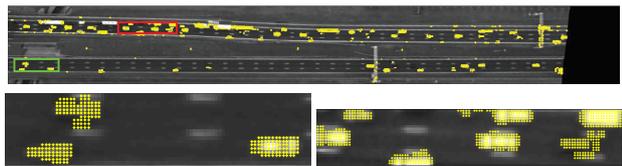
Figure 6: Identified vehicles in the image 100 (top), the zoomed area depicted by a green rectangle (bottom left), the zoomed area depicted by a red rectangle (bottom right).

Figure 7. The statistics come in Table 1 for fluid traffic and in Table 2 for congested traffic. All validation is based on careful visual inspection of the results.
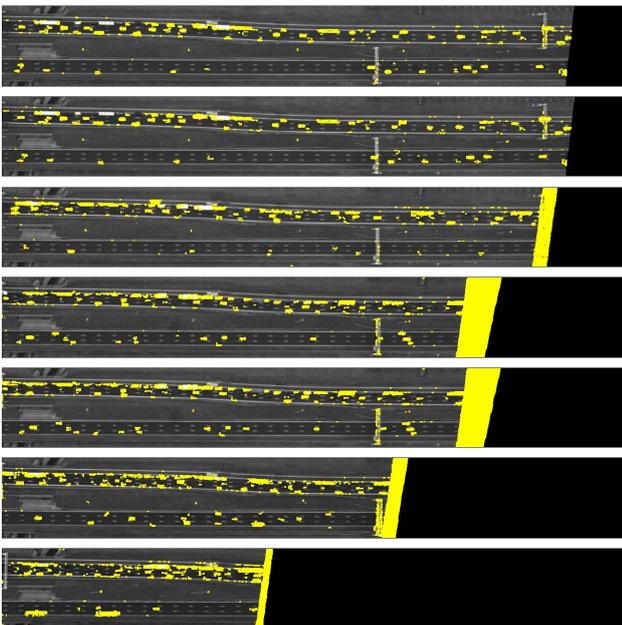
Figure 7: Identified vehicles: top to bottom: image number 100, 101, 248, 500, 501, 1000 and 1500.

In each table, the total number of vehicles, the number of identified vehicles, missing vehicles, wrongly identified vehicles, mixed vehicles and vehicles which are identified as more than one vehicle are listed. Only wrongly detected vehicles on the road area are counted, vehicles found by the algorithm on road lines and a gantry were discarded.

Vehicles which were missed in one image can be identified in an other image. For example in images 101 and 501 one of the vehicles which were not detected in the images 100 and 500 respectively, were identified. In image 101, one vehicle leaves the

Table 1: Identification results for smoothly flowing traffic. N, T, TP, FN, FP, Mx and Dj are respectively the image number, the total number of vehicles, the number of true positives (correctly identified vehicles), false negatives (not identified vehicles), false positives (wrongly identified vehicles), mixed vehicles and disjoint regions.

N	T	TP	FN	FP	Mx	Dj
100	18	17	1	0	0	1
101	17	17	0	0	0	2
248	9	9	0	1	0	1
500	19	17	2	1	1	2
501	19	18	1	0	0	0
1000	17	15	2	1	1	0
1500	6	6	0	0	0	1

Table 2: Identification results for very congested traffic. The labels are as identified in Table 1.

N	T	TP	FN	FP	Mx	Dj
100	59	50	9	3	4	4
101	59	50	9	1	4	1
248	64	59	5	3	20	5
500	68	57	11	1	9	4
501	68	57	11	1	6	3
1000	52	47	5	6	3	3
1500	35	30	5	10	2	3

area. Therefore the total number of correctly detected vehicles is the same for both images 100 and 101.

The vehicle identification method works well in case of moving traffic. The results is however less good in the case of congested traffic. The main problem in this case is the mixing of vehicles which sometimes happens when vehicles are too close. This problem cannot be solved until the vehicles start to separate. Another problem is less serious: sometimes one vehicle in reality leads to more than one detected vehicle. This problem occurs mainly on trucks with very slow movement. The number of disjoint vehicles is relatively low however.

5 CONCLUSIONS

In this paper, vehicles were detected from the image-sequence stabilized on the road area as recorded by a hovering helicopter. It turns out to be possible to extract vehicles, as blobs, in several difficult situations. Previous problems with the vehicle detection of i) small size vehicles observed with scarce detail, ii) vehicles with low speed, and iii) low contrast vehicles, could be largely solved. A very low amount of memory is needed for the processing. Because each pixel is only compared to the previous value of the background, the procedure is done in a sequential way for every image. Besides, the procedure starts without any extra assumptions and more importantly, there is a mechanism to recover the background when a pixel is wrongly classified as a vehicle. Moreover, the background value is always updated to the current gray level. As a result, this method also works for image sequences representing heavy traffic conditions and for image sequences that are exposed to gradual illumination variations. By looking at a temporal profile of a pixel, it has been observed that illumination variations of this pixel gray-level for the background

could be large, but these background illumination variations are negligible if just a few consecutive images are considered. A sudden large illumination variation, however, cannot be handled by this method.

Although in general only part of a very low contrast vehicle can be extracted by this method, tracking of such a vehicle using our algorithm is still possible, compare (Karimi Nejadasl et al., 2006). Vehicles, which are too close to each other, are grouped as one blob, thus their tracking is unreliable. When these vehicles start to take a distance from each other, they are identified as separate blobs and then tracked reliably. However, if a vehicle stays very long in one place it is classified as background. When the vehicle starts to move, updating the background value requires more images to be processed.

ACKNOWLEDGEMENTS

References

Ballard, D. H., 1981. Generalizing the hough transform to detect arbitrary shapes. Pattern Recognition 13(2), pp. 111–122.

Dahlkamp, H., Pece, A. E., Ottlik, A. and Nagel, H. H., 2004. Differential analysis of two model-based vehicle tracking approaches. DAGM LNCS 3175, pp. 7178.

Elgammal, A., Duraiswami, R., Harwood, D. and Davis, L. S., 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of the IEEE 90(7), pp. 1151–1163.

Ferryman, J. M., Worrall, A. D., Sullivan, G. D. and Baker, K. D., 1995. A generic deformable model for vehicle recognition. BMVC95 1, pp. 127–136.

Gonzalez, R. C. and Woods, R. E., 2007. Digital Image Processig. 3 edition edn, Prentice Hall.

Hinz, S., 2003. Integrating local and global features for vehicle detection in high resolution aerial imagery. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 34(Part 3/W8), pp. 119–124.

Karimi Nejadasl, F., Gorte, B. G. H. and Hoogendoorn, S. P., 2006. Optical flow based vehicle tracking strengthened by statistical decisions. ISPRS Journal of Photogrammetry and Remote Sensing 61(3-4), pp. 159–169.

Ossen, S. J. L., 2008. Longitudinal driving behavior: theory and empirics. PhD thesis.

Ossen, S. J. L., Hoogendoorn, S. P. and Gorte, B. G. H., 2006. Interdriver differences in car-following: a vehicle trajectory-based study. Transportation Research Record 1965, pp. 121–129.

Ottlik, A. and Nagel, H. H., 2008. Initialization of model-based vehicle tracking in video sequences of inner-city intersections. International Journal of Computer Vision.

Pece, A. E., 2006. Contour tracking based on marginalized likelihood ratios. Image and Vision Computing 24, pp. 301–317.

Pece, A. E. and Worrall, A. D., 2002. Tracking with the em contour algorithm. European Conference Computer Vision pp. 3–17.

Pece, A. E. C., 2002. Generative-model-based tracking by cluster analysis of image differences. Robotics and Autonomous Systems 39(3-4), pp. 181–194.

Stauffer, C. and Grimson, W., 1999. Adaptive background mixture models for real-time tracking. IEEE Int'l Conf. on Computer Vision and Pattern Recognition.

Stauffer, C. and Grimson, W. E. L., 2000. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Tan, T., Sullivan, G. and Baker, K., 1998. Model-based localisation and recognition of road vehicles. International Journal of Computer Vision 27(1), pp. 5–25.

Zhao, T. and Nevatia, R., 2003. Car detection in low resolution aerial images. Image and Vision Computing 21(8), pp. 693–703.