MOTION COMPONENT SUPPORTED BOOSTED CLASSIFIER FOR CAR DETECTION IN AERIAL IMAGERY

Sebastian Tuermer[a], Jens Leitloff[a], Peter Reinartz[a], Uwe Stilla[b]

[a] Remote Sensing Technology Institute, German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
sebastian.tuermer@dlr.de, jens.leitloff@dlr.de, peter.reinartz@dlr.de
[b] Photogrammetry and Remote Sensing, Technische Universitaet Muenchen (TUM)
Arcisstrasse 21, 80333 Munich, Germany
stilla@tum.de

Commission III/5

KEY WORDS: Vehicle detection, AdaBoost, HoG features, Aerial image sequence, Motion mask

ABSTRACT:

Research of automatic vehicle detection in aerial images has been done with a lot of innovation and constantly rising success for years. However information was mostly taken from a single image only. Our aim is using the additional information which is offered by the temporal component, precisely the difference of the previous and the consecutive image. On closer viewing the moving objects are mainly vehicles and therefore we provide a method which is able to limit the search space of the detector to changed areas. The actual detector is generated of HoG features which are composed and linearly weighted by AdaBoost. Finally the method is tested on a motorway section including an exit and congested traffic near Munich, Germany.

1 INTRODUCTION

Already within the last century the impact and the significance of mobility and especially individual traffic has increased enormously (Banister et al., 2010). The phenomenon results in overloaded streets and highways. Further this leads to environmental pollution, wast of resources and finally threatens humans' quality of life (Ouis, 2001).
To adequately overcome this problem, scientists worldwide are working on smart solutions. They all need data of realistic traffic scenarios which can be analyzed and evaluated. Final goal are strategies to improve the current traffic situation. Mainly two applications should be named in the real-time case, mass events and catastrophes. Manager of mass events will be able to canalize the usual high volume of traffic. This results in a higher security level. Also emergency teams and rescue crews are supported by traffic data in the event of a disaster. They will be able to choose the fastest ways reaching the affected area and can see in detail where to set up a control room or a collection point. Due to these important applications there are some other procedures of gathering traffic information besides the optical ones. For instance induction loops, light barriers, radar based methods or floating car solutions. But all of these methods are not suitable for monitoring a wide area consistently.
We present a method for extracting vehicles in sequential aerial imagery. The method uses HoG features and Boosting as machine learning algorithm. The focus lies on the motion mask which affords detection of moving objects faster and more reliable.

2 RELATED WORK

Methods for vehicle detection in optical images often belong to one of three groups according to the platform of the sensor. The field with definitely the highest amount of research activity during the last years are stationary video cameras which provide side view images or at least oblique view images. Further property is a quite high imaging frequency in comparison to the other groups. The use of wavelet coefficients as features and AdaBoost can be seen in (Schneiderman and Kanade, 2000). Also (She et al., 2004) are detecting cars by the use of Haar wavelets features in the HSV color space. A combination of Haar and HoG features which are formed to a strong cascading classifier by Boosting presents (Negri et al., 2008). In (Kasturi et al., 2009) a simple background subtraction is done which is only working for video data. An overview on the work for stationary cameras can be found in (Sun et al., 2006).
The next group considers satellite imagery which provide a reduced spatial resolution (highest resolution is often max 0.5 m) and mainly use single images, not time series. An approach which uses simple features based on shape and intensity presents (Eikvil et al., 2009). Using segmented images and applying a maximum likelihood classification can be observed in (Larsen et al., 2009). Promising results have also been achieved by (Leitloff et al., 2010). They use Haar-like features in combination with AdaBoost.
The last group of approaches deals with airborne images. At this step we first suggest a further separation in explicit or implicit models. Approaches based on explicit models are for example given in (Moon et al., 2002) with a convolution of a rectangular mask and the original image. Also (Zhao and Nevatia, 2003) offer an interesting method by creating a wire-frame model and try to match it with extracted edges at the end of a Bayesian network. A similar way is suggested by (Hinz, 2003a) (Hinz, 2003b), the author makes the approach more mature and added additional parameters like the position of the sun. (Kozempel and Reulke, 2009) provide a very fast solution which takes four special shaped edge filters trying to represent an average car. Another approach of (Reilly et al., 2010) shows a method which is based on background subtraction. The background is computed by a 10 frame median image.
Finally implicit modeling is used by (Grabner et al., 2008), they take Haar-like features, HoG features and LBP (local binary patterns). All these features are passed to an on-line AdaBoost training algorithm which creates a strong classifier.
Another approach using aerial data and trying to have benefit of

the temporal component, similar to our idea, is (Benedek et al., 2009). Their aim is not only the detection of cars but all moving objects. To realize this idea a three layer Markov random field model is introduced.
A comprehensive overview and evaluation of airborne sensors for traffic estimation can be found in (Hinz et al., 2006) and (Stilla et al., 2004).

3 METHOD

In general, the method is developed for airborne, high resolution frame camera systems with high imaging frequency. The workflow of our method is shown in Fig. 1. Following subsections give explanations to parts of the workflow or refer to related literature for detailed information.

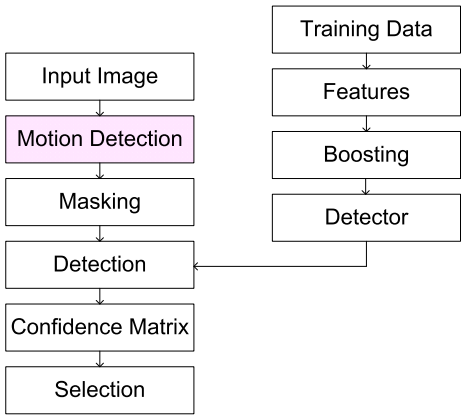

Figure 1: Workflow of proposed car detection method

3.1 Color Space

For our purpose we decided to use a color space which is technically oriented. That means per definition the color space is a linear transformation of the RGB color space. The utilized color space is named I1I2I3 and meets, according to (Ohta et al., 1980) and own tests, the requirements of the proposed method (Sec. 3.2) very well. Which is mainly the quality of the resulting difference image. Mathematically expressed the transformation is shown in Eq. 1:

$$\begin{bmatrix} I1 \\ I2 \\ I3 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & -1/2 \\ -1/4 & 1/2 & -1/4 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

where R, G, B are the red, green, blue channels and I1, I2, I3 are the resulting channels of I1I2I3 color space model.

3.2 Motion Detection

The idea of the motion mask is based on turning all available information to account which is delivered by our camera system. To reach that aim a usual way of motion detection is processing a difference image. A difference image shows all pixels which have changed in comparison to the other image. One possibility is to calculate the difference image with the current image and its background image. Unfortunately the problem is that we do not have an image without foreground objects.
A solution of this problem offers the use of three images and a subtraction of each (Dubuisson and Jain, 1995). In detail, we calculate the difference of the current image and the previous image, and the difference of the current image and the subsequent image as well. The two resulting difference images are linked with the Boolean AND. The approach expressed in formulas can be seen in Eq. 2 where the first difference image D_1 is calculated (Rehrmann and Birkhoff, 1995):

$$D_1(t_1, t_2, x, y) =$$

$$\begin{cases} 1, \text{ if } & |I_{I1}(t_2, x, y) - I_{I1}(t_1, x, y)| \\ & + |I_{I2}(t_2, x, y) - I_{I2}(t_1, x, y)| \\ & + |I_{I3}(t_2, x, y) - I_{I3}(t_1, x, y)| > d_{min} \\ 0, \text{ else} \end{cases} \tag{2}$$

where the functions of the images are $I_{I1}(t, x, y)$, $I_{I2}(t, x, y)$ and $I_{I3}(t, x, y)$. The parameter t is a discreet time whereas x and y are the position in the image for the three different channels $I1$, $I2$, $I3$ of the color space. The parameter d_{min} is a threshold which is necessary for excluding intensity changes of pixels due to camera noise, various illuminations or the different illustration geometry.
Subject to the condition that we have 3 consecutive images the next step is linking the two difference images which is depicted in Eq. 3:

$$D_2(t_1, t_2, t_3, x, y) =$$

$$\begin{cases} 1, \text{ if } & D_1(t_1, t_2, x, y) = 1 \\ & \wedge D_1(t_2, t_3, x, y) = 1 \\ 0, \text{ else} \end{cases} \tag{3}$$

with $D_1(t_1, t_2, x, y)$ difference image of previous and current image and $D_1(t_2, t_3, x, y)$ difference image of current and consecutive image.

3.3 Features

We use HoG features (Dalal and Triggs, 2005) to differentiate cars from other objects. A reason for this choice is a test where Haar and HoG features are compared with regard to their car detection capability (Tuermer et al., 2011). HoG features are created by quantize gradient magnitudes to a histogram. The particular bin is chosen according to the gradient orientation. A detailed explanation of these features and how the feature extraction works can be found in (Tuermer et al., 2010).

3.4 Training

The training creates the custom classifier. We pass the extracted features of more than 400 car samples to the machine learning algorithm. This algorithm is part of the Boosting group (Freund and Schapire, 1997) (Freund and Schapire, 1999) and is named Real AdaBoost. Boosting is a method which builds a strong classifier by a weighted linear combination of weak classifiers. In our case a weak classifier is a threshold applied on a feature which is able to classify more accurate than 50 percent object of interest or not object of interest. The procedure of weighting and re-weighting is graphically explained in Fig. 2. The formula of the composite strong classifier H can be expressed as Eq. 4 shows:

$$H(X) = \boldsymbol{sign}(a_1 h_1(x) + a_2 h_2(x) + a_3 h_3(x)) \tag{4}$$

where a_i are weightings and h_i are weak classifiers.

3.5 Detection

The ordinary detection is done by sliding the previously generated classifier over the whole search image and applying it at every pixel position. A method which is time consuming and susceptible to mistakes. Alternatively, the proposed innovative

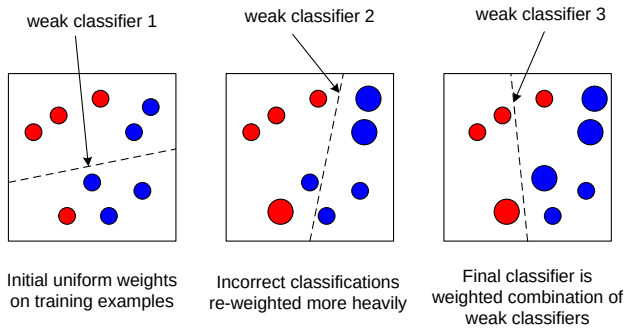

Figure 2: Boosting Schema

method just applies the detector where the motion mask is true. An additional graphical explanation can be found in Fig. 3. The response obtained from the classifier is a confidence value which has information how reliable the detection candidate is. Sometimes applying a threshold to the confidence matrix is necessary to adjust the result to the respective requirement. On the one hand it could be useful to detect all cars in the image and accept false positives as consequence. On the other hand it could be necessary to obtain correct detections only and accept false negatives.

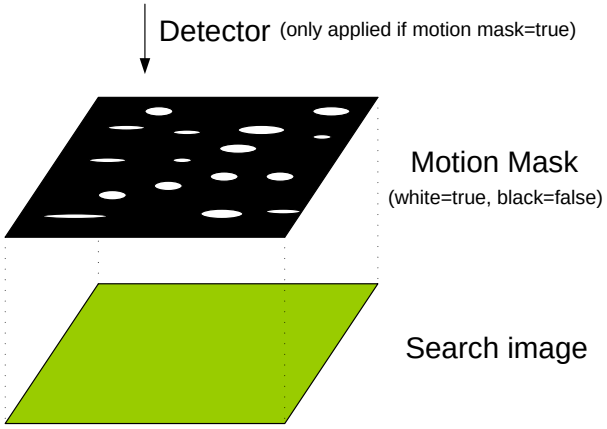

Figure 3: Functional description of the motion mask

4 CAMERA SYSTEM

The utilized aerial test data are acquired from the 3K camera system, which is composed of three off-the-shelf professional SLR digital cameras (Canon EOS 1Ds Mark II). These cameras are mounted on a platform which is specially constructed for this purpose. A picture of the cameras and the platform is shown in Fig. 4. Furthermore a calibration was done (Kurz et al., 2007) to enable the georeferencing process which is supported by GPS (Global Positioning System) and INS (inertial navigation system). The system is designed to deliver images with maximum 3 Hz recording frequency combined into one burst. A burst consists of 2 to 4 images and is necessary because otherwise the camera would not be able to write the data to the memory card. After one burst a pause of 10 seconds follows. Depending on the flight altitude a spatial resolution up to 15 centimeters (at 1000m altitude) is provided. For further information about the 3K camera system please refer to (Reinartz et al., 2010).

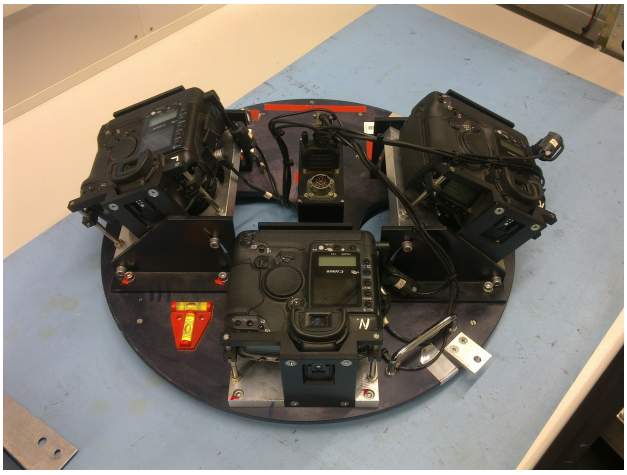

Figure 4: 3K camera system

5 EXPERIMENTAL RESULTS

The experimental results are based on image samples from a motorway in the east of Munich, Germany. Our intention is the detection of cars in two directions only (from right to left and vice versa); note the cars which take the exit have different orientations and are not classified. The search image (Fig. 5) is the second image out of three and thus imaged at time t_2 according to the preceding remarks (Sec. 3.2). To give an impression how helpful the motion mask is, we display the result of a classification without motion mask in Fig. 6.
The next images show the genesis of the motion mask. The result of applying Eq. 2 can be seen in Fig. 7 and Fig. 8. The manual chosen threshold d_{min} amounts 30. But if necessary it can be easily substituted by the automatic Otsu thresholding method (Otsu, 1979). Applying Eq. 3 results in the final motion mask shown in Fig. 9. The remaining search space after applying the mask is depicted in Tab. 1. Finally the result of the proposed detection method is shown in Fig. 10. Where detections of moving vehicles are marked with red rectangles.

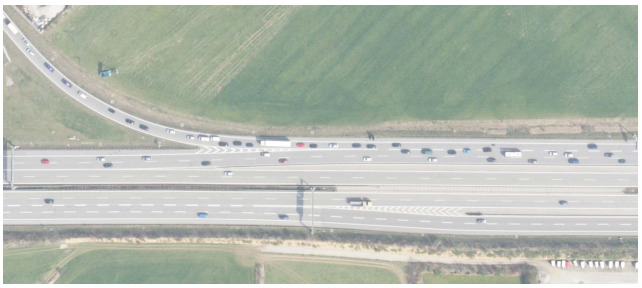

Figure 5: Original 3K image sample

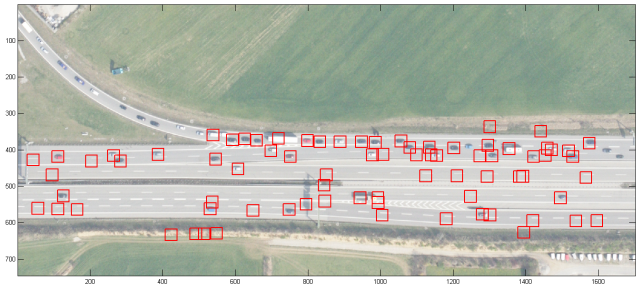

Figure 6: Classification result without motion mask

Figure 7: Difference image of image t_0 and t_1

Figure 8: Difference image of image t_1 and t_2

Figure 9: Boolean AND of the two difference images (Fig. 7, Fig. 8)

Figure 10: Classification result with motion mask

6 DISCUSSION

The car detection quality with motion mask (Fig. 10) is considerably enhanced compared to the test without motion mask (Fig. 6). This is due to the limited search space where static areas which include mainly no vehicles are excluded. But a consequence of this method is that also cars without or with low velocity are excluded. Now it might be necessary to develop a method which brings the detection methods for static and for moving cars together.
Concerning the detector design, it should be mentioned that there is still room for improvement as fas as the training data is concerned. We can trace the false positives in Fig. 6 back to the fact that the negative training sample database is not sufficient. These false car candidates often look very similar and are very often parts of the road with a small part of road markings. Perhaps it is possible to exclude them by a more intelligent training.
The advantage of the motion mask is not only the improved detection quality, but of course reduction of calculation time as well. A quick look at Tab. 1 shows that in the end only about one percent of the original test image have to be examined. This does not mean that the detector is 100 times faster, because it is a cascading detector and only the application of the first hierarchical level can be spared for all pixel positions. But calculating the motion mask is still faster than calculating all the features of the first hierarchical level of the detector.
Another interesting point in the processing chain of the motion mask itself is that the result in Fig. 8 has obviously much more disturbances than Fig. 7. This can be explained due to a lack of co-registration. The overlay of the images is only done by the use of the geocode and the relative error (image to image) of the georeferencing comes into full account. However the presented method is able to handle these kind of errors dependable. By the way the same result using RGB color space is much more noisy in comparison to the utilized I1I2I3 color space.

Table 1: Limited search space due to motion mask

	remaining search space of original image
$D_1(t_1, t_2, x, y)$ (Fig.7)	2.05 %
$D_1(t_2, t_3, x, y)$ (Fig.8)	6.03 %
$D_2(t_1, t_2, t_3, x, y)$ (Fig.9)	1.01 %

7 CONCLUSIONS AND FUTURE WORK

We present a vehicle detection method which is improved by using additional information provided by the temporal component. Making use of three consecutive images allows to determine the position of a moving car very accurately. The resulting mask shows potential to identify moving objects, which will help to make vehicle detection more reliably in the future. But there is also a catch to progress in the case of slowly moving vehicles. It can be observed that slowly moving vehicles with intent to take the exit of the highway are not captured perfectly. The same applies to non-moving objects. This happens because some pixels still have the same color as the pixels at t_{i-1}. In this case the method needs further development. Benefit of the proposed detection method for moving vehicles is:

- detection runs much faster (up to 37x)
- more robust and reliable
- very high detection quality

Running the tests with a more intelligent training and a extended training database is one point of future work. Furthermore we would like to use test images from more difficult areas near city centers for instance. And finally the detector itself will get an upgrade regarding the ability of being rotation invariant.
Of course the detection can be remarkable improved by using additional information that is not used till this day. The database with positive samples consists only of images that show a car. One idea is to not only us sample chips with just a single car inside, but introduce a training database which regards to the surrounding of the cars. This could be helpful to distinguish if an object is situated on the road or on a roof for for example.

REFERENCES

Banister, D., Browne, M. and Givonia, M., 2010. Transport reviews - the 30th anniversary of the journal. Transport Reviews: A Transnational Transdisciplinary Journal 30, pp. 1–10.

Benedek, C., Sziranyi, T., Kato, Z. and Zerubia, J., 2009. Detection of object motion regions in aerial image pairs with a multilayer markovian model image processing. IEEE Transactions on Image Processing 18(10), pp. 2303 – 2315.

Dalal, N. and Triggs, B., 2005. Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, IEEE Computer Society, San Diego, CA, USA, pp. 886 – 893.

Dubuisson, M. P. and Jain, A. K., 1995. Contour extraction of moving objects in complex outdoor scenes. International Journal of Computer Vision 14(1), pp. 83–105.

Eikvil, L., Aurdal, L. and Koren, H., 2009. Classification-based vehicle detection in high-resolution satellite images. ISPRS Journal of Photogrammetry and Remote Sensing 64, pp. 65–72.

Freund, Y. and Schapire, R. E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), pp. 119–139.

Freund, Y. and Schapire, R. E., 1999. A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence 14(5), pp. 771–780.

Grabner, H., Nguyen, T. T., Gruber, B. and Bischof, H., 2008. On-line boosting-based car detection from aerial images. ISPRS Journal of Photogrammetry and Remote Sensing 63(3), pp. 382 – 396.

Hinz, S., 2003a. Detection and counting of cars in aerial images. In: International Conference on Image Processing (ICIP), Vol. 3, pp. 997–1000.

Hinz, S., 2003b. Integrating local and global features for vehicle detection in high resolution aerial imagery. In: Photogrammetric Image Analysis (PIA), Vol. 34(3/W8), International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 119–124.

Hinz, S., Bamler, R. and Stilla, U., 2006. Editorial theme issue: Airborne und spaceborne traffic monitoring. ISPRS Journal of Photogrammetry and Remote Sensing 61(3-4), pp. 135–136.

Kasturi, R., Goldgof, D., Soundararajan, P., Manohar, V., Garofolo, J., Bowers, R., Boonstra, M., Korzhova, V. and Zhang, J., 2009. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2), pp. 319–336.

Kozempel, K. and Reulke, R., 2009. Fast vehicle detection and tracking in aerial image bursts. In: CMRT09, Vol. 38(3/W4), IAPRS, pp. 175–180.

Kurz, F., Mller, R., Stephani, M., Reinartz, P. and Schroeder, M., 2007. Calibration of a wide-angel digital camera system for near real time scenarios. In: ISPRS Hannover Workshop: High-Resolution Earth Imaging for Geospatial Information.

Larsen, S. O., Koren, H. and Solberg, R., 2009. Traffic monitoring using very high resolution satellite imagery. Photogrammetric Engineering and Remote Sensing 75(7), pp. 859–869.

Leitloff, J., Hinz, S. and Stilla, U., 2010. Vehicle extraction from very high resolution satellite images of city areas. IEEE Trans. on Geoscience and Remote Sensing 48, pp. 1–12.

Moon, H., Chellappa, R. and Rosenfeld, A., 2002. Performance analysis of a simple vehicle detection algorithm. Image and Vision Computing 20(1), pp. 1–13.

Negri, P., Clady, X., Hanif, S. M. and Prevost, L., 2008. A cascade of boosted generative and discriminative classifiers for vehicle detection. EURASIP Journal on Advances in Signal Processing 2008, pp. 1–12.

Ohta, Y.-I., Kanade, T. and Sakai, T., 1980. Color information for region segmentation. Computer Graphics and Image Processing 13, pp. 222–241.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Sys., Man., Cyber. 9(1), pp. 6266.

Ouis, D., 2001. Annoyance from road traffic noise: A review. Journal of Environmental Psychology 21(1), pp. 101–120.

Rehrmann, V. and Birkhoff, M., 1995. Echtzeitfhige Objektverfolgung in Farbbildern. In: Tagungsband 1. Workshop Farbbildverarbeitung, Fachberichte Informatik 15/95, University of Koblenz, pp. 36–39.

Reilly, V., Idrees, H. and Shah, M., 2010. Detection and tracking of large number of targets in wide area surveillance. In: European Conference on Computer Vision (ECCV) 2010.

Reinartz, P., Kurz, F., Rosenbaum, D., Leitloff, J. and Palubinskas, G., 2010. Near real time airborne monitoring system for disaster and traffic applications. In: Optronics in Defence and Security (Optro), Paris, France.

Schneiderman, H. and Kanade, T., 2000. A statistical method for 3d object detection applied to faces and cars. In: IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 746–751.

She, K., Bebis, G., Gu, H. and Miller, R., 2004. Vehicle tracking using on-line fusion of color and shape features. In: International IEEE Conference on Intelligent Transportation Systems, pp. 731–736.

Stilla, U., Michaelsen, E., Soergel, U., Hinz, S. and Ender, J., 2004. Airborne monitoring of vehicle activity in urban areas. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34(Part B3), pp. 973–979.

Sun, Z., Bebis, G. and Miller, R., 2006. On-road vehicle detection: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5), pp. 694–711.

Tuermer, S., Leitloff, J., Reinartz, P. and Stilla, U., 2010. Automatic vehicle detection in aerial image sequences of urban areas using 3d hog features. In: International Archives of Photogrammetry, Remote Sensing and the Spatial Information Sciences, Vol. XXXVIII(Part 3), Paris, France.

Tuermer, S., Leitloff, J., Reinartz, P. and Stilla, U., 2011. Evaluation of selected features for car detection in aerial images. In: Hanover Workshop 2011.

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