AUTOMATIC VEHICLE DETECTION IN AERIAL IMAGE SEQUENCES OF URBAN AREAS USING 3D HOG FEATURES

S. Tuermer [a, *], J. Leitloff [a], P. Reinartz [a], U. Stilla [b]

[a] Remote Sensing Technology Institute, German Aerospace Center (DLR), 82230 Wessling, Germany
(sebastian.tuermer, jens.leitloff, peter.reinartz)@dlr.de
[b] Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, Arcisstrasse 21, 80333 Munich, Germany
stilla@tum.de

Commission III, WG III/5

KEY WORDS: Vehicle detection, Aerial images sequences, Urban areas, HoG features, 3K camera system

ABSTRACT:

With the development of low cost aerial optical sensors having a spatial resolution in the range of few centimetres, the traffic monitoring by plane receives a new boost. The gained traffic data are very useful in various fields. Near real-time applications in the case of traffic management of mass events or catastrophes and non time critical applications in the wide field of general transport planning are considerable. A major processing step for automatically provided traffic data is the automatic vehicle detection. In this paper we present a new processing chain to improve this task. First achievement is limiting the search space for the detector by applying a fast and simple pre-processing algorithm. Second achievement is generating a reliable detector. This is done by the use of HoG features (Histogram of Oriented Gradients) and their appliance on two consecutive images. A smart selection of this features and their combination is done by the Real AdaBoost (Adaptive Boosting) algorithm. Our dataset consists of images from the 3K camera system acquired over the city of Munich, Germany. First results show a high detection rate and good reliability.

1. INTRODUCTION

1.1 Motivation

Within recent years automatic traffic monitoring using aerial images has become an essential part of several valuable applications (Stilla et al., 2005) (Stilla et al., 2009). Some of them are in the field of traffic research and should provide efficiently planned and highly optimized road networks, e.g. control of traffic signals (Hickman and Mirchandani, 2008). This leads to less environmental pollution, a reduction of CO_2 emissions and a saving of resources. Another field of application can be found in the case of mass events, which have to be managed (Kühne et al., 2005). For this purpose a complete overview of the existing traffic situation is useful. This includes data about travel times (Kurz et al., 2007a), traffic flow and density as well as smart controlling of the parking situation. Catastrophes and disaster management also require data about the current traffic situation. In an emergency, scenarios can be analysed faster, better reactions can be initiated, and the emergency crews can act more efficiently. To satisfy these needs traditional methods of gathering traffic information are induction loops or stationary video cameras, but utilization is limited due to their inherently fixed location and sparse distribution.

To overcome the drawback of current sensor systems, an airborne camera system has been developed at the German Aerospace Center (DLR) (Reinartz et al., 2006). The new system is called 3K camera system and is used in the context of the VABENE (V̲erkehrsma̲nagement b̲ei Großere̲reignissen un̲d Katastrophe̲n) project. Goal of this project is to provide traffic information in case of mass events and catastrophes. A small but important part of this comprehensive project is the automatic vehicle detection.

1.2 Related Work

We can split the methods for vehicle detection from optical images mainly in three groups according to the platform of the sensor. The field with definitely the highest amount of research activity are stationary video cameras which provide side view images. Schneiderman and Kanade (2000) use wavelet features and AdaBoost. Also She et al. (2004) are detecting cars by the use of Haar wavelets features in the HSV colour space. Negri et al. (2008) use Haar features and HoG features which are formed to a strong cascading classifier by boosting. Kasturi, et al. (2009) uses a simple background subtraction, which is only working for video data. An overview on the work for stationary cameras can be found in Sun (2006).

The next group considers satellite imagery which provide a reduced spatial resolution (lowest pixel size is 0.5 m) and mainly use single images, not time series. Promising results have 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 which either use 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 further try to match it with extracted edges of the end of a Bayesian network. A similar way is suggested by Hinz (2003a, 2003b), he makes the approach more complex and added additional parameters like the position of the sun. Kozempel and Reulke (2009) provide a

[*] Corresponding author

very fast solution which takes four special shaped edge filters trying to represent an average car.

Finally implicit modelling 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 the AdaBoost training algorithm which creates a strong classifier.

A comprehensive overview and evaluation of airborne sensors for traffic estimation can be found in Hinz et al. (2006) and Stilla et al. (2004).

The next chapter has technical details of the currently used optical sensor system. Chapter 3 explains the fast pre-processing and the presented method for car detection. The experimental results with urban area imagery are shown in chapter 4. Gained results a going to be discussed and evaluated in chapter 5. Also a prospect of future work can be found here. Finally chapter 6 gives a conclusion of the 3D HoG feature vehicle detection method.

2. SENSOR SYSTEM

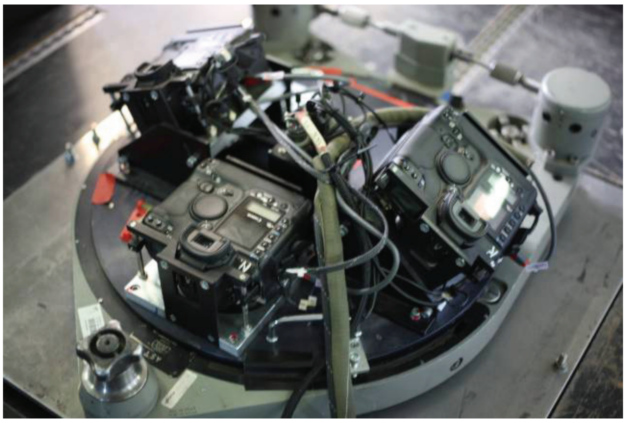


Figure 1: 3K camera system

As previously mentioned, the utilized 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. Figure 1 shows the ready for flight installed camera system. A calibration was done (Kurz et al., 2007b) to enable the georeferencing process which is supported by GPS and INS. The system is designed to deliver images with maximum 3 Hz recording frequency combined into one burst, which consists of 2 to 4 images. After one burst a pause of 10 seconds follows. Depending on the flight altitude a spatial resolution up to 15 centimetres (at 1000m altitude) is provided. The acquired images are processed on board the plane in real time and the extracted information is sent without further delay to the ground station. The processing step includes ortho-rectification followed by car detection and tracking. The received data are ready to use for instantaneous analysis of the current traffic situation. For further information about the 3K camera system please refer to Reinartz et al. (2010).

3. METHOD

The proposed technique uses 2D and 3D HoG features in combination with AdaBoost to detect vehicles in aerial images. Figure 2 depicts the workflow of the presented method. In this section we give detailed explanations of the theory behind our approach. Experimental results are shown in Section 4.

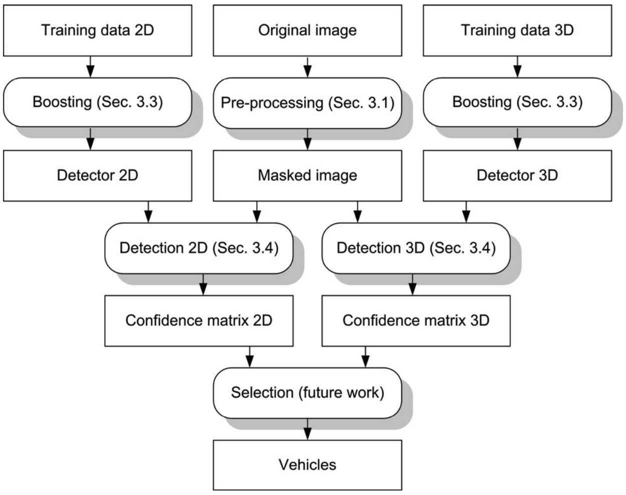


Figure 2: Workflow of proposed method

3.1 Pre-processing

Pre-processing is often used to limit the search space for the detector. This is usually done by fast and simple methods. The final goal is a faster detection of the vehicles in the image due to fewer regions that have to be examined.

For this purpose we take a three channel RGB colour image and smooth it with a Gaussian filter. Afterwards, we perform regiongrowing that uses the distance of each pixel to the pixels in the von Neumann neighbourhood. Equation 1 shows the algorithm which calculates the Euclidean distance of two pixels. Pixels that have a distance below the predefined threshold are accumulated to one region.

$$D = \sqrt{\frac{1}{3}\left((r_{ia} - r_{ib})^2 + (g_{ia} - g_{ib})^2 + (b_{ia} - b_{ib})^2\right)} \qquad (1)$$

where D = Euclidean distance
r_{ia}, g_{ia}, b_{ia} = red, green, blue channel ∈ pixel a ∈ image i
r_{ib}, g_{ib}, b_{ib} = red, green, blue channel ∈ pixel b ∈ image i

Finally the resulting segmentation allows excluding large homogeneous regions, which often appear as road surface.

3.2 Histogram of Oriented Gradients

The Histrogram of Oriented Gradients offers a way to describe typical parts of the vehicle. For this reason the gradient directions and their magnitude are calculated and finally saved in a histogram.

The basic work has been done by Lowe (1999) with the introduction of the SIFT operator. The process of creating HoG features (Dalal and Triggs, 2005) begins with a gradient filtering of the image; e.g. using the Sobel operator. Afterwards, the number of bins of the histogram has to be chosen adequately. Too many bins increase the calculation time while an insufficient number of bins will lead to distinctive features. Finally each bin contains the accumulated magnitudes of the gradient vectors with the corresponding direction. The obtained histogram is finally normalized.

For a fast calculation of the HoG features it is recommended to use the integral histogram (Porikli, 2005). The integral histogram is an accumulation of all magnitudes of the current pixel and the previous pixels within the same bin over the whole image (Equation 2).

$$H(x,y)=\sum_{i=0}^{i\leq x}\sum_{j=0}^{j\leq y} I(x,y) \tag{2}$$

where H(x,y) = Integral Histogram
I(x,y) = magnitudes of gradients with same orientation
x, y are the pixel coordinate

The integral histogram allows for every region of interest to calculate the HoG feature very fast by two subtractions and one addition for each bin (as Figure 3 and Equation 3 show).

$$\begin{aligned} F(x,y,w,h) &= H(x,y)+H(x+w,y+h) \\ &-H(x,y+h)-H(x+w,y) \end{aligned} \tag{3}$$

where F(x,y,h,w) = one bin from HoG feature F
w, h are width and height of HoG feature F

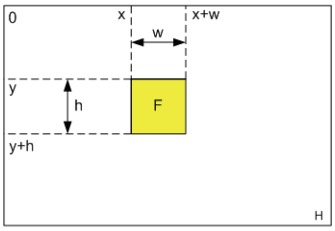


Figure 3: Integral Histogram and sum of one bin in yellow

So far the presented method is applied to 2D data (single image). Our goal is to enhance this method by extending the 2D features by a further dimension. This can be realized if two consecutive images are used for feature extraction (Viola et al., 2005). This should especially improve the detection of moving vehicles, because in the moving case the first image shows a car and on the same position in the next image there is a road instead (Figure 4).

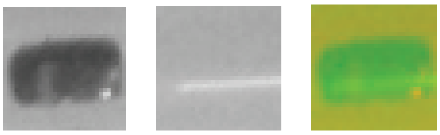


Figure 4: (a) car in image1 (b) same location in image2 (c) sample1 as red channel and sample2 as green channel

We give the 2D and the 3D feature to the training that is depicted in the next section.

3.3 Training

The previous step delivers a lot of different features which are passed to the boosting algorithm (Freund and Scharpire, 1997), which performs feature selection and reduction. Finally we receive a strong classifier which consists of as many features as necessary to classify the tested negative and positive samples correctly (break by thresholds of detection rate and false positives). The training for the 2D and the 3D detector is processed separately.

3.4 Detection

Classification is performed by sliding the detector over the original image, which is masked as Section 3.1 describes. All the areas that are masked are examined by the detector. The detector works in a cascading way that means, if the first hierarchical step is sure that the examined location is not a vehicle the next hierarchical steps are not applied. This cascading detector construction helps saving computation time for the detection, as well as pre-processing does. Result after applying the detector is a confidence matrix for each processed pixel. Further step is a manual threshold applied to the confidence values to get detected cars.

4. EXPERIMENTAL RESULTS

The used data, acquired by the 3K camera system, shows a street in the western inner city of Munich with a ground sampling distance of 15 cm (Figure 5(a)). The scene is orthorectified and can be overlaid with vector data from street databases (i.e. Navteq). This also helps limiting the search space but due to the low accuracy of this data, all near to the road positioned houses remain.

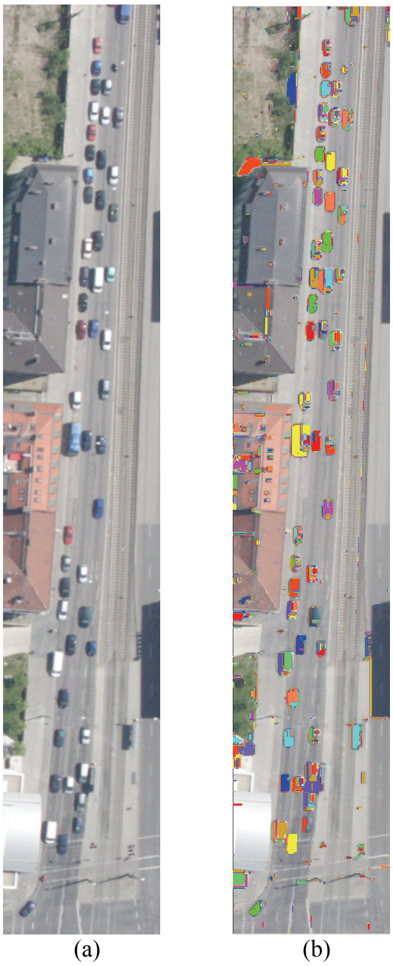


Figure 5: (a) Original image (b) pre-processed image with masked regions

4.1 Pre-processing

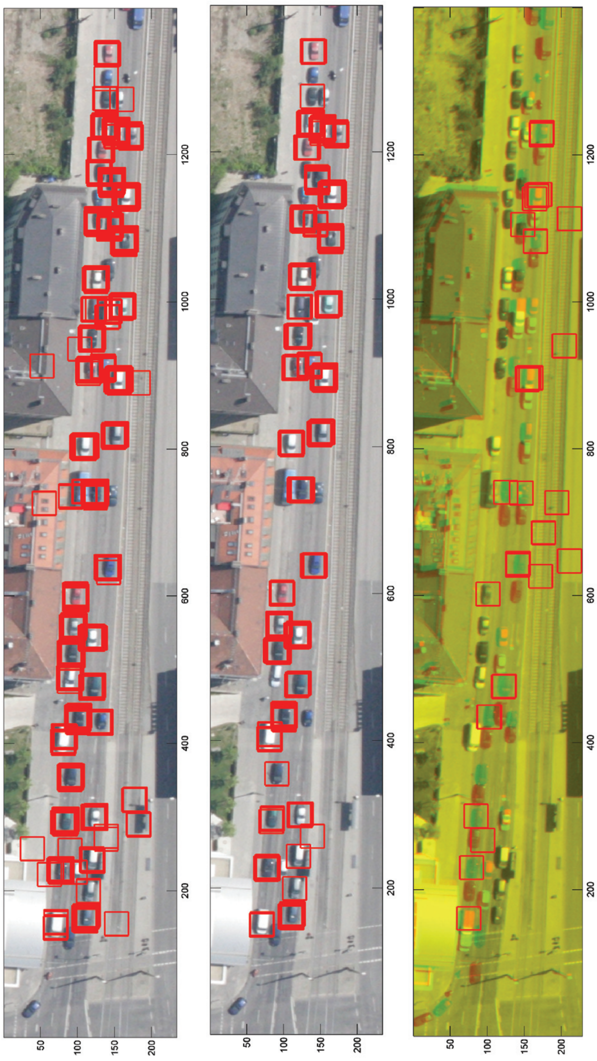
Figure 5(a) shows the original image and Figure 5(b) shows the image after the pre-processing step. The upper left corner of each region is used as starting point for the regiongrowing. Parameters for the method described in 3.1 are Gaussian filter of size 5 by 5 pixels (results in sigma 0.87) and tolerance value for the distance between neighbour pixels of 5. Finally the search space after pre-processing is only 14 percent of the original image.

4.2 Training

For the creation of the 2D detector we use a training database of 446 positive samples and 700 negative samples. Each sample has a size of 32 by 32 pixel and same resolution as test image sample. A window for creating 2D HoG features in the quadratic sizes 4, 8, 12 and 16 is moved over them. In this experiment we take 8 bins for each feature.

The same procedure is done with the 3D HoG features, these features consist of 16 bins because each of the two image layers has 8 bins. The 3D training data are 29 positive and 700 negative samples. One positive training sample is displayed in Figure 4(c).

4.3 Detection



(a) (b) (c)

Figure 6: (a) Detection result without threshold (b) detection result with confidence threshold set manually (c) Image1 in red, image2 in green and detected cars by 3D HoG detector in red rectangle (manual threshold 0.9)

Sliding the hierarchical 2D detector with a 32 by 32 rectangle over the image results in the detected vehicles displayed in Figure 6(a). For every pixel position in the image we receive a confidence value on which a threshold can be applied to suppress low values. Figure 6(b) shows the result with confidence values above 0.87.

$$precision\ rate = \frac{true\ positives}{true\ positives\ +\ false\ positives} \tag{4}$$

$$recall\ rate = \frac{true\ positives}{true\ positives\ +\ false\ negatives} \tag{5}$$

The precision rate according to Equation 4 is 97 percent in the case of Figure 6(b). The recall rate (Equation 5) is 80 percent. Of course, the prerequisite for this assumption is a successfully applied method to suppress multi detections.

To enhance this 2D result we use the 3D detector which is thought to detect moving vehicles better. Figure 6(c) shows a result that is supposed to explain the idea behind this technique. Cars that are moving do appear in green (image1) and in red (image2). The missing cars after 2D detection can now be substituted with the 3D detection.

5. DISCUSSION AND FUTURE WORK

The proposed pre-processing step shows its capability to limit the search space for the following classification. It is correct that the first hierarchical step of the cascading detector can exclude those regions as well. To gain clarity about this fact an exhaustive testing has to be made. Additionally important is the robustness of the operation, taking care that regions of vehicles are not excluded from further processing.

The achieved result of the 2D detector shows still some false positives and some false negatives. Therefore the next step is finding an adequate method to select final vehicles correctly without setting a threshold manually. The information we have is the position of the rectangle including double and triple detections and the confidence values. Developing a method utilising this information for further improvement will be the next implementation step.

Also, the 3D detector that should support the 2D detector needs further development. Reason for the poor detection rate is probably the low quantity of training samples. Another point that has to be reviewed is the impact of the road markings, which appear in the second image of 3D training data (Figure 4(b)). These markings are very different for each image and return completely different features. Hence the sensitive boosting returns bad rates for the trained detector due to highly heterogeneous training data.

6. CONCLUSIONS

Detection of vehicles from aerial images is a challenging problem due to the great diversity of vehicles and its restricted resolution. The presented approach shows the beginning of an innovative processing chain for vehicle detection. Starting with a very fast pre-processing which is able to exclude large homogeneous regions of multi channel images without exclusion of useful information. Afterwards we proposed the idea of generating a robust detector which is build on 2D and 3D HoG features in combination with a training algorithm. These 3D HoG features show additional enhancement and are able to support classification solely based on 2D HoG features.

REFERENCES

Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society,* San Diego, CA, USA, Vol. 1, pp. 886 – 893.

Freund, Y., Schapire, R. E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, Vol. 55 (1), pp. 119-139.

Hickman, M. D. , Mirchandani, P. B., 2008. Airborne Traffic Flow Data and Traffic Management. *Symposium on the Fundamental Diagram: 75 Years (Greenshields 75 Symposium), July 8-10, 2008*, Woods Hole, MA, USA.

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

Hinz, S., 2003b. Integrating Local and Global Features for Vehicle Detection in High Resolution Aerial Imagery. In: *Photogrammetric Image Analysis (PIA), Int. Arch. Photogram. Rem. Sens. Spatial Inform. Sci.,* Munich, Germany, Vol. 34 (3/W8), pp. 119-124.

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

Kasturi, R., Goldgof, D., Soundararajan, P., Manohar, V., Garofolo, J., Bowers, R., Boonstra, M., Korzhova, V., Zhang, J., 2009. Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 31 (2), pp. 319-336.

Kozempel, K., Reulke, R., 2009. Fast Vehicle Detection and Tracking in Aerial Image Bursts. In: *Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluation (CMRT)*, Paris, France, Vol. 38 (3/W4), pp. 175-180.

Kühne, R., Ruhé, M., Bonert, M., 2005. Anwendungen der luftgestützten Verkehrsdatenerfassung bei Großveranstaltungen. *Verkehrswissenschaftliche Tage, 2005-09-19 - 2005-09-20,* Dresden, Germany.

Kurz, F., Charmette, B., Suri, S., Rosenbaum, D., Spangler, M., Leonhardt, A., Bachleitner, M., Stätter, R., Reinartz, P., 2007a. Automatic traffic monitoring with an airborne wide-angle digital camera system for estimation of travel times. In: *Photogrammetric Image Analysis (PIA), Int. Arch. Photogram., Rem. Sens. Spatial Inform. Sci.,* Munich, Germany, Vol. 36 (3/W49B), pp. 83-86.

Kurz, F., Müller, R., Stephani, M., Reinartz, P., Schroeder, M., 2007b. Calibration of a Wide-Angel Digital Camera System for Near Real Time Scenarios. *ISPRS Workshop: High-Resolution Earth Imaging for Geospatial Information,* Hanover, Germany.

Leitloff, J., Hinz, S., Stilla, U., 2010. Vehicle extraction from very high resolution satellite images of city areas. *IEEE Trans. on Geoscience and Remote Sensing,* Vol. PP (99), pp. 1-12.

Lowe, D., 1999. Object recognition from local scale invariant features. In: *International Conference on Computer Vision (ICCV),* Corfu, Greece, Vol. 2, pp. 1150-1157.

Moon, H., Chellappa, R., Rosenfeld, A., 2002. Performance Analysis of a Simple Vehicle Detection Algorithm. *Image and Vision Computing,* Vol. 20 (1), pp. 249-253.

Porikli, F., 2005. Integral histogram: a fast way to extract histograms in Cartesian spaces. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society,* San Diego, CA, USA, Vol. 1, pp. 829-836.

Reinartz, P., Kurz, F., Rosenbaum, D., Leitloff, J., Palubinskas, G., 2010. Near Real Time Airborne Monitoring System for Disaster and Traffic Applications. *Optronics in Defence and Security* (*OPTRO),* Paris, France.

Reinartz, P., Lachaise, M., Schmeer, E., Krauss, T., Runge, H., 2006. Traffic monitoring with serial images from airborne cameras. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 61 (3-4), pp. 149-158.

Schneiderman, H., Kanade, T., 2000. A statistical method for 3D object detection applied to faces and cars. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Hilton Head Island, SC, USA, Vol. 1, pp. 746-751.

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

Stilla, U., Michaelsen, E., Sörgel, U., Hinz, S., Ender, J., 2004. Airborne monitoring of vehicle activity in urban areas. In: *ISPRS Congress, Int. Arch. Photogram. Rem. Sens. Spatial Inform. Sci.,* Istanbul, Turkey Vol. 35 (B3), pp. 973-979.

Stilla, U., Rottensteiner, F., Hinz, S. (Eds.), 2005. *Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms, and Evaluation (CMRT),* Paris, France, Vol. 36 (3/W24).

Stilla, U., Rottensteiner, F., Paparoditis, N. (Eds.), 2009. *Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluation (CMRT)*, Paris, France, Vol. 38 (3/W4).

Sun, Z., Bebis, G., Miller, R., 2006. On-Road Vehicle Detection: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28 (5), pp. 694-711.

Viola, P., Jones, M. J., Snow, D., 2005. Detecting Pedestrians Using Patterns of Motion and Appearance. *International Journal of Computer Vision*, Vol. 63 (2), pp. 153-161.

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