In: Stilla U, Gamba P, Juergens C, Maktav D (Eds) JURSE 2011 - Joint Urban Remote Sensing Event --- Munich, Germany, April 11-13, 2011

Vehicle Detection in Aerial Images using Boosted Classifier with Motion Mask

Sebastian Tuermer, Jens Leitloff and Peter Reinartz
Remote Sensing Technology Institute (IMF)
German Aerospace Center (DLR)
Wessling, Germany
{sebastian.tuermer, jens.leitloff, peter.reinartz}@dlr.de

Uwe Stilla
Photogrammetry and Remote Sensing
Technische Universitaet Muenchen (TUM)
Munich, Germany
stilla@tum.de

***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 offers 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 limiting 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 near Munich including an exit and congested traffic.**

I. Introduction

Already within the last century the impact and the significance of mobility and especially individual traffic has increased enormously [1]. The phenomenon results in overloaded streets and highways. Further this leads to environmental pollution, wast of resources and finally threatens humans' quality of life [2].

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 build up a control room or a collection point. Due to this 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. Focus is on the motion mask which affords detection of moving objects faster and more reliable.

II. 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 [3]. Also [4] 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 [5]. In [6] 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 [7].

The next group considers satellite imagery which provide a reduced spatial resolution (lowest 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 [8]. Using segmented images and apply a maximum likelihood classification can be observed in [9]. Promising results have also been achieved by [10]. 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 [11] with a convolution of a rectangular mask and the original image. Also [12] 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 [13] [14], the author makes the approach more mature and added additional parameters like the position of the sun. [15] provide a very fast solution which takes four special shaped edge filters trying to represent an average car.

Finally implicit modeling is used by [16], 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.

978-1-4244-8657-1/11/$26.00 ©2011 IEEE

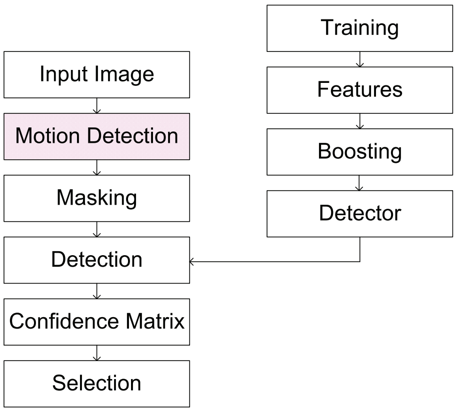

Fig. 1. Workflow

Another approach using aerial data and trying to have benefit of the temporal component, similar to our idea, is [17]. 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 [18] and [19].

III. Method

In general, the method is developed for airborne, high resolution frame camera systems with low 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.

A. 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 new color space is named I1I2I3 and meets, according to [20] and own tests, our requirements very well. 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.

B. 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 [21]. 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 [22]:

$$D_1(t_1, t_2, x, y) =$$
$$\begin{cases} 1, \text{ if} & |I_{I1}(t_2,x,y) - I_{I1}(t_1,x,y)| \\ & + |I_{I2}(t_2,x,y) - I_{I2}(t_1,x,y)| \\ & + |I_{I3}(t_2,x,y) - I_{I3}(t_1,x,y)| \; > d_{min} \\ 0, \text{ else} & \end{cases} \tag{2}$$

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.

C. Features

We use HoG features [23] to differentiate cars from other objects. The 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 [24].

D. 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 [25] [26] 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:

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

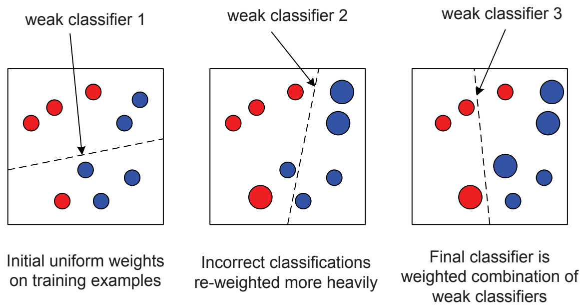

Fig. 2. Boosting Schema

where a_i are weightings and h_i are weak classifiers.

E. Detection

The detection is done by sliding the previously generated classifier over the masked image and apply it for the position of every pixel. The response of the classifier is a confidence value. The application of a threshold to the confidence matrix is sometimes necessary for adjusting 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.

IV. Camera System

The utilized aerial 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 calibration was done [27] 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 centimeters (at 1000m altitude) is provided. The acquired images are processed on board of 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 [28].

V. Experimental Results

The experimental results are based on the image sample (Fig. 3) which is imaged at time t_1 according to the preceding remarks (Sec. III-B). The result of applying Eq. 2 can be seen in Fig. 4 and Fig. 5. The manual chosen threshold d_{min} amounts 30. Next step with Eq. 3 leads to Fig. 6 and we received our final motion mask. The remaining search space after applying the mask is depicted in Tab. I. Finally the result of the entire detection method shows Fig. 7. Where detections of moving vehicles are marked with red rectangles.

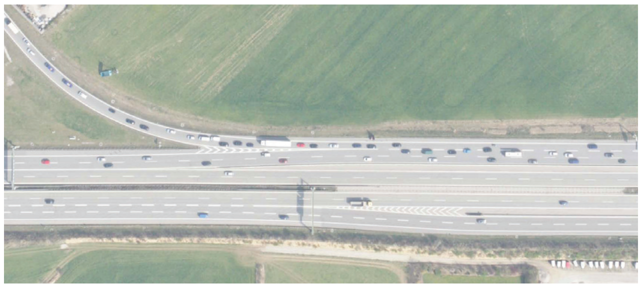

Fig. 3. Original 3K image sample

Fig. 4. Difference image of image t_0 and t_1

Fig. 5. Difference image of image t_1 and t_2

Fig. 6. Boolean AND of the two difference images

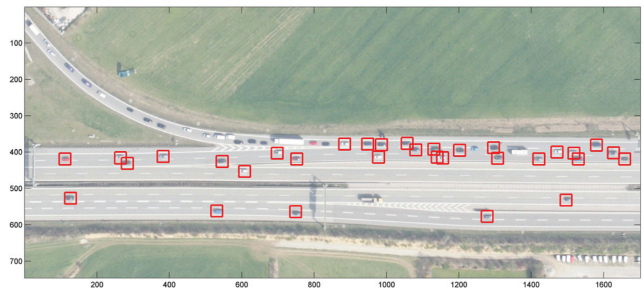

Fig. 7. Result of the detector applied with motion mask

TABLE I
LIMITED SEARCH SPACE DUE TO MOTION MASK

	remaining search space of original image
$D_1(t_1, t_2, x, y)$ (Fig.4)	2.05 %
$D_1(t_2, t_3, x, y)$ (Fig.5)	6.03 %
$D_2(t_1, t_2, t_3, x, y)$ (Fig.6)	1.01 %

VI. DISCUSSION

Obviously the result in Fig. 5 has much more disturbances than Fig. 4. 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 as well. By the way the same result using RGB color space is much more noisy in comparison to the utilized I1I2I3 color space.

VII. CONCLUSIONS

We presented a vehicle detection method which gains improvement by using additional information. The possession of three consecutive images allows to determine the position of a moving car very accurately. The resulting mask shows potential to identify moving objects, this will help to make vehicle detection in future more reliable. 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. Same situation is for not 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)
- is more robust and reliable

REFERENCES

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

[2] D. Ouis, "Annoyance from road traffic noise: A review," *Journal of Environmental Psychology*, vol. 21, no. 1, pp. 101–120, 2001.

[3] H. Schneiderman and T. Kanade, "A statistical method for 3d object detection applied to faces and cars," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, Jun. 13–15, 2000, pp. 746–751.

[4] K. She, G. Bebis, H. Gu, and R. Miller, "Vehicle tracking using on-line fusion of color and shape features," in *International IEEE Conference on Intelligent Transportation Systems*, Oct. 3–6, 2004, pp. 731–736.

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

[6] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang, "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*, vol. 31, no. 2, pp. 319–336, February 2009.

[7] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 694–711, 2006.

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

[9] S. O. Larsen, H. Koren, and R. Solberg, "Traffic monitoring using very high resolution satellite imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 75, no. 7, pp. 859–869, 2009.

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

[11] H. Moon, R. Chellappa, and A. Rosenfeld, "Performance analysis of a simple vehicle detection algorithm," *Image and Vision Computing*, vol. 20, no. 1, pp. 1–13, 2002.

[12] T. Zhao and R. Nevatia, "Car detection in low resolution aerial image," *Image and Vision Computing*, vol. 21, no. 8, pp. 693–703, Jul. 7–14, 2003.

[13] S. Hinz, "Detection and counting of cars in aerial images," in *International Conference on Image Processing (ICIP)*, vol. 3, Sep. 14–17, 2003, pp. 997–1000.

[14] ——, "Integrating local and global features for vehicle detection in high resolution aerial imagery," in *Photogrammetric Image Analysis (PIA)*, vol. 34, no. 3/W8. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2003, pp. 119–124.

[15] K. Kozempel and R. Reulke, "Fast vehicle detection and tracking in aerial image bursts," in *CMRT09*, vol. 38, no. 3/W4. IAPRS, 2009, pp. 175–180.

[16] H. Grabner, T. T. Nguyen, B. Gruber, and H. Bischof, "On-line boosting-based car detection from aerial images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 3, pp. 382 – 396, 2008.

[17] C. Benedek, T. Sziranyi, Z. Kato, and J. Zerubia, "Detection of object motion regions in aerial image pairs with a multilayer markovian model image processing," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2303 – 2315, October 2009.

[18] S. Hinz, R. Bamler, and U. Stilla, "Editorial theme issue: Airborne und spaceborne traffic monitoring," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 61, no. 3-4, pp. 135–136, December 2006.

[19] U. Stilla, E. Michaelsen, U. Soergel, S. Hinz, and J. Ender, "Airborne monitoring of vehicle activity in urban areas," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, no. Part B3, 2004, pp. 973–979.

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

[21] M. P. Dubuisson and A. K. Jain, "Contour extraction of moving objects in complex outdoor scenes," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 83–105, 1995.

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

[23] N. Dalal and B. Triggs, "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, 2005, pp. 886 – 893.

[24] S. Tuermer, J. Leitloff, P. Reinartz, and U. Stilla, "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. XXVIII, no. Part 3, Paris, France, September 2010.

[25] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[26] ——, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, no. 5, pp. 771–780, September 1999.

[27] F. Kurz, R. Mller, M. Stephani, P. Reinartz, and M. Schroeder, "Calibration of a wide-angel digital camera system for near real time scenarios," in *ISPRS Hannover Workshop: High-Resolution Earth Imaging for Geospatial Information*, no. 3/W49B, 2007-05- 29 - 2007-06-01 2007.

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