A real-time optical airborne road traffic monitoring system

Gellért Máttyus, Franz Kurz, Dominik Rosenbaum, and Oliver Meynberg

German Aerospace Center (DLR)
Remote Sensing Technology Institute
gellert.mattyus@dlr.de
franz.kurz@dlr.de
dominik.rosenbaum@dlr.de
oliver.meynberg@dlr.de

Abstract. In the last years, the Remote Sensing Technology Institute of the German Aerospace Center has been developing a real-time airborne monitoring system which can be used during natural disasters, mass events and huge traffic jams. This system consists of a cost-effective optical camera system, onboard computers for real-time processing and a datalink for transmitting the traffic information and aerial images to the ground station. The traffic information is automatically extracted from the images by detecting the cars and tracking them within a few frames. Results are the position and the speed of each cars, which are written into a database on the ground. In this paper we give a description of the overall system with focus on the automatic vehicle detection.

1 Introduction

Measuring the road traffic, e.g. density, flow, are important for urban planning, traffic management and authorities like police, emergency, fire service etc. The conventional method for acquiring traffic informations are systems installed on the ground, e.g. induction loops, camera systems. The DLR investigates new solutions to extend existing sensors with an airborne system which can work independently from the ground infrastructure and can provide high flexibility. Such a system can be used during disasters and mass events, when the infrastructure is damaged and/or the traffic shifts to alternative routes without ground based traffic sensors, e.g. smaller streets, unpaved/unsealed roads. Additionally, the presented airborne camera system provides also images which can give additional important information about the conditions of the monitored area.

In the chapter 2 we describe briefly the entire system, in the chapter 3 we give a detailed description about the automatic car detection process. Later we write our test results, conclusions and future work.

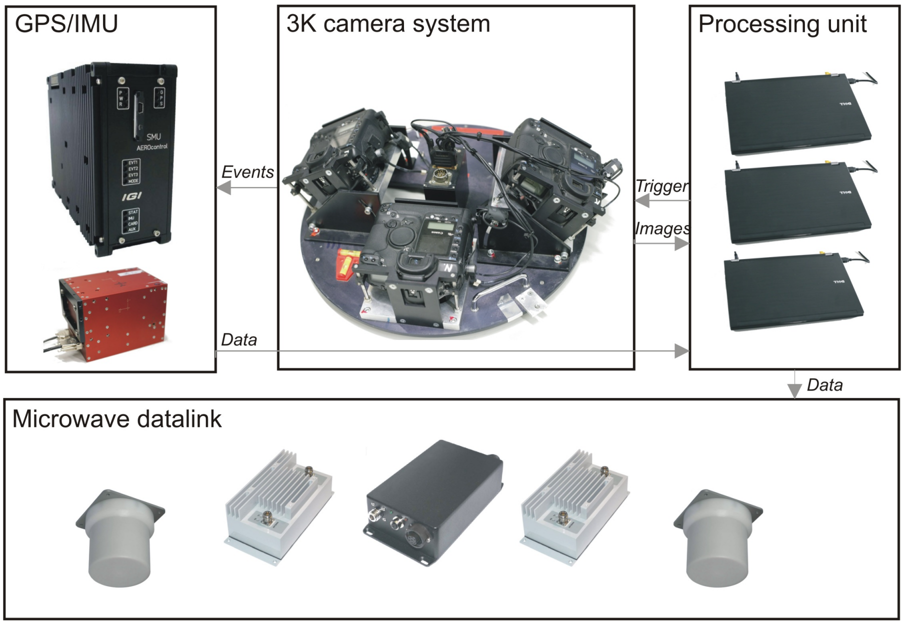


Fig. 1. The system overview

2 System Description

Our system is equipped with three commercial cameras. The acquired images are orthorectified with the use of a GPS, IMU and a Digital Elevation Model. On the orthorectified image the streets are extracted with the use of a street database, on the road areas in the images the cars are detected and tracked within a few frames and their speeds are calculated. The traffic data together with the orthorectified image are downlinked to the mobile ground station where they are written to a database which can be accessed on the internet. The overview of the system, with the hardware components, is shown on the figure 1. The described system is operational on two of the research aircrafts of the German Aerospace Center, the Cessna™ Grand Caravan and the Dornier™ Do228. [1]

2.1 Camera system

The Camera System consists of three commercial Canon EOS 1Ds Mark III™ cameras with 50mm Zeiss™ lenses, 21 MPixel resolution and an IGI™ GPS/IMU which provides the accurate position and orientation of the plane. [1]. The cameras are mounted on a ZEISS™ aerial shock mount ready for the DLR airplanes,

[1] Our institute develops a similar, but smaller and compacter system for the Lange Antares H2 electrical glider plane

Height above ground	Coverage	Ground Sampling Distance
$500m$	$1280m \times 240m$	$6.5cm$ nadir
$1000m$	$2560m \times 480m$	$13cm$ nadir
$1500m$	$3840m \times 720m$	$19.5cm$ nadir
$3000m$	$7680m \times 1440m$	$39cm$ nadir

Table 1. Coverage and ground sampling distance on different heights

see the figure 1. One camera faces the nadir direction, the other two are tilted by maximum 32°. The system can be installed in the plane along- or across track. The across track installation provides a field of view of 104°, the along track 26°. The ground sampling distances(GSD) of the acquired images can be found in the table 1. The cameras are triggered by the onboard PCs via the USB connection. This can be programmed to exposure at a given geographical position. The flash output of each camera is wired into the GPS/IMU, which registers the position and orientation for each image.

2.2 Onboard Processing

The airplane is equipped with 4 PCs(Intel™ Core i7 2600), one for each camera and one extra for handling the transmission and providing the user interface for the operator.
The processing steps are:

- Orthorectification, the image is transformed from the image coordinate system into a world coordinate system(i.e. UTM).
- Transforming the roads on the image to straight strips.
- Detecting the cars on the road strips
- Tracking the cars for 2 or 3 frames

Orthorectification The GPS/IMU registers the location and orientation of the image, the cameras are calibrated, thus each pixel can be projected on a Digital Elevation Model (DEM) in the UTM system. This process can be parallelized massively, since each pixel is calculated independently. This process runs on the GPU (Nvidia™ GeForce 9800 GTX) written in CUDA™. The GPU implementation compared to the simple C version running on the CPU is approximately 50 times faster.

Road Extraction Using a road database e.g. Navteq, OpenStreetMap, the roads on the orthorectified image are transformed to straight image strips on which the car detection is executed aligned in x-axis. This accelerates the whole process chain, since the car detector is rotation variant and it also eliminates several false positives, e.g. building roofs and cars which are not in traffic (i.e. parking cars).

Car Detection The car detector is based on the Viola-Jones face detection method [2] extended with tilted Haar like features [3]. The detection is processed on the image strips converted to grayscale images. The detection is more detailed in chapter 3.

Car Tracking The tracking is based on shape based matching around the location of the car where the best match is choosen as the next position of the car [8]. This is done for two or three frames. On the orthorectified image each object motion is converted to motion in meters. Thus the speed is simply calculated, since the accurate time stamp is given by the GPS/IMU. The cameras can achieve a frame rate of 5 Hz, which enables an unambiguous alignment of the detected cars, since in this short time (200 ms) the motion of the car is approximately straight.

2.3 Data Transmission

The real-time[2] traffic and image data have to be accessible on a central server. For fulfilling this requirement a relatively high speed data link is needed between the airplane and the ground station. The first alternative is a SRS C-band microwave link, this link can reach a speed of 10 Mbits upto 70 km of distance. The second alternative is a laser link developed at DLR, which can provide a transmission speed upto 1 Gbps to 120 km distance. The ground station of both links track the airplane and vice versa.

3 Car Detector

3.1 Related works

The Remote Sensing Technology Institute of DLR in cooperation with the Remote Sensing Department of the University of Technology Munich has long time experience researching in the field of traffic and car detection on aerial and satellite images.
Stefan Hinz et al. [4] has used features of a vehicle's upper surface and a 3D model of the car for predicting the shadow region for detecting cars on aerial images.

Leitloff et al. [6] has detected vehicles on satellite images by using Haar-like features with AdaBoost combined with a model and extraction of vehicle queues.

Rosenbaum et al. presented a real-time car detection system described in [8] using a three stage approach, first Gentle AdaBoost with Haar-like features, second an interesting operator for the output confidence image of the first stage and third a Support Vector Machine (SVM) for a number of statistical values calculated from geometric and radiometric properties in the confidence image and in all channels of the RGB image.

[2] For the whole system the time limitation is 5 minutes, from the image exposure until the registration in the central data base on the ground.

Other research groups have also worked in this field.

Nguyen et al. in [5] use online boosting for car detection on aerial images where the output of the detector is labeled manually and is used to update the parameters for the classifier.

Mattyus et al. [7] presented a real-time capable algorithm which can track moving objects on aerial videos. They register the frames of the camera to each other, build a background model, detect the blobs of moving pixels and track them with Kalman filter.

3.2 Description

The detector is based on the object detection framework proposed by Viola and Jones in [2]. We first train the detector off-line with an image data set containing manually labeled samples of cars and areas without cars in the heading direction. The result of the training is a binary classifier, which can classify a sub image as car or non-car. In the on-line detection phase a sliding window is scanned across the image at multiple scales and locations with, each sub-window is classified as car or non-car and the overlapped detections are grouped together.

The three main ideas of [2], also utilized in our system, are the followings:

Using Haar-like features with integral images The Haar-like features are simple features based on the difference of the sum of the intensity values at two image areas, see figure 2 for the used features introduced in [3] and the figure 3 for examples of the selected features by the training. The number of the possible Haar-like features is much higher than the number of pixels in the image, so the information in the image is overrepresented, but during the training the important features are selected. Using the integral images one feature can be calculated extremely efficient, by a few memory accesses and arithmetic operations, thus it can run fast on the processor.

Combining the haar-like features to a strong classifier by Boosting The Haar-like features are weak-classifiers, they provide a decision slightly better than by chance (0.5 probability). By boosting they can be combined together to form a strong classifier. We use the Gentle AdaBoost algorithm described in [9].

Train and detect in a cascade structure On most of the images, the background covers much higher percent of the image as the objects which the detector is searching for. These background areas are mostly easily distinguishable from the objects, e.g. homogeneous regions on the roads, in the sense of Haar-like features this means they can be classified by using a small number of Haar-like features, but there are subwindows which are hard to classify correctly, e.g. roof of bus stops and buildings. An important part of a fast object detector is dividing the classification to stages. Each stage rejects the object or passes it to next stage. The image subwindow which passes the last stage is classified as object at the end. The figure 4 shows the detector at different stages. This structure

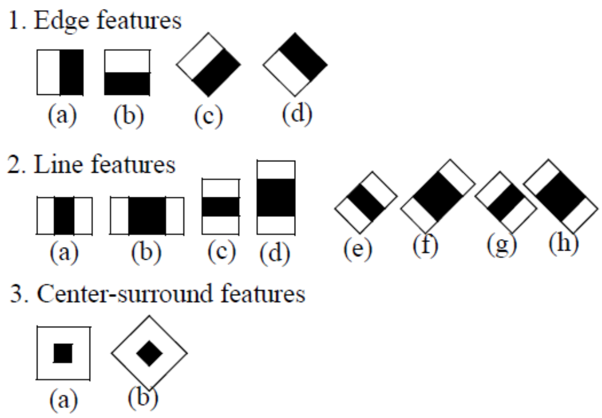


Fig. 2. The used Haar-like features[3]

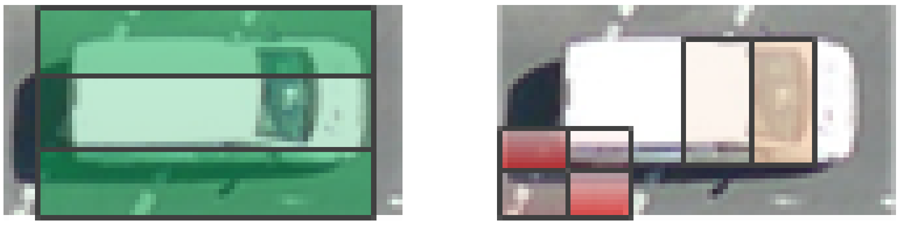


Fig. 3. The first Haar-like features of the car detector

is not only essential for the rapid speed, but for the training process. During the training the negative samples for the first stage are chosen randomly from manually labeled image areas which do not contain cars in the desired direction. The second and consecutive stages are trained with negative samples which were classified wrong (i.e. false positives) by the earlier stages. This has the effect that the last classifier stages concentrate on the hard negative samples and so these classifiers contain significantly more Haar-features as the first stages.

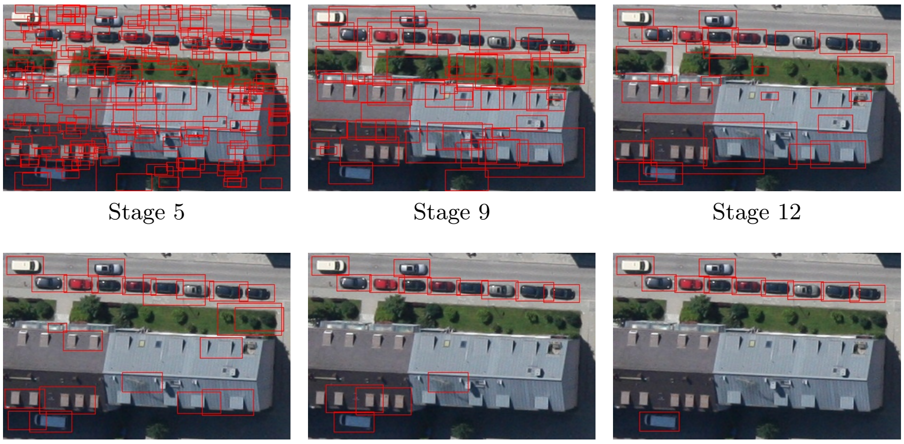
The false positive rate of the final cascade is

$$F = \prod_{i=1}^{K} f_i \tag{1}$$

where K is the number of stages, f_i is the false positive rate of the classifier of the ith stage. The detection rate is

$$D = \prod_{i=1}^{K} d_i \tag{2}$$

where d_i is the detection rate of the classifier of the ith stage. Thus a 21 stage classifier with $d_i = d = 0.995$ and $f_i = f = 0.5$ provides a theoretical detection rate of $D \approx 0.9$ and a false positive rate of $F \approx 5 \times 10^{-7}$.



Stage 5 Stage 9 Stage 12

Stage 15 Stage 18 Stage 21

Fig. 4. The detector output at different cascade stages

4 Results

4.1 Data sets and training

The detector was trained on samples from two image sequences, one taken over the German city Cologne during clouded weather without rain, the other taken over Munich by sunny weather. The figure 5 shows positive samples, the figure 6 shows regions from which negative samples are chosen. The car samples, manually labeled on orthorectified images, are croped, rotated to x-axis direction, resized to 24×12 and converted to greyscale.

The detector was trained with 2000 positive samples. For each stage negative samples were chosen randomly, from the manually labeled regions, and evaluated by the earlier stages until 2000 false positives were found, these samples were used for the training. The training process lasted approximately 20 hours on an 8 core Intel™ Xeon E5410 processor. The final cascade has 21 stages, each with a detection rate $d_i = 0.995$ and false positive rate $f_i = 0.5$.

4.2 Results

The detector was evaluated on the same image sequences, but on more images from the scene. For the Munich sequence a new image sequence over the same

Fig. 5. Positive training samples

Fig. 6. Regions for negative training samples

	Cologne scene	Munich scene	$\sum$
No. ground truth (G)	1885	4985	6870
No. Correctly detected (N_D)	1811	4379	6190
No. False positives (N_F)	35	82	117
Detection rate ($D = N_D/G$)	96.1%	87.84%	90.1%
Correctness ($C = \frac{N_D}{N_D+N_F}$)	98.11%	98.16%	98.14%
Wrong detection rate ($W = 1 - C$)	1.89%	1.83%	1.85%

Table 2. Detection results

area were also chosen. The quantitative results are in the table 2. We have evaluated the detector for the cars in traffic on the road strips of major roads. The performance of the whole system is effected by the other process steps (e.g. a car turning from one street to an other might not be detected due to the road extraction, a street can be missed due to inaccuracy in the street database or the orthorectification). For detections on road strips see the figure 7. On these images the correct detections are marked with green, the false positives with red and the missing detections with black. The figure 8 shows the detections on an ortho image in an urban scenario, the detections are marked here with red color.

The programs are written in C++ with the use of the Halcon™ and OpenCV image processing and computer vision libraries. The detection algorithm has taken maximum 800 ms for one image. The execution time for the detection heavily depends on the image content (i.e. how many streets are on the image). An exact comparison of different methods is difficult, because there is no free open dataset. Since the image quality and GSD has a high impact on the final result the different methods can not be compared across different datasets. The detection results also depend on the training dataset, e.g. how good they cover the different lighting conditions and street structures.

Nguyen et al. in [5] works on images with GSD of $8 - 12$ cm, which is significantly higher than the images on which we work, GSD of $14 - 22$ cm. Thus the results can not be accurately compared, though our algorithm by using a cascade structure enables the automatic selection of hard negative samples. This elimininates the manual labeling iterations used in [5] and accelerates the detection process.

The SVM based method described in [8] uses a large set of features, which are not necessarily size invariant. In this paper, we do not investigate if the method is usable for Ground Sampling Distances different from the tested without retraining the detector. The detector described in this paper works for different GSDs since it uses only Haar-like features which are size invariant until the object has at least the size of the training window size.

We have tested the system multiple times, also at real world conditions, e.g. The Reunification Day of Germany 2011 where the acquired datas were delivered to the Bonn Police Department. The maximal execution times for one image of the process chain and the downlink time for the traffic datas are in the table

Process	Worst case time
Image acquisition	$0.2s$
Orthorectification	$0.3s$
Road extraction	$0.2s$
Car detection	$1s$
Car tracking	$2s$
Downlink to the ground	$2s$
$\sum$	$5.7s$

Table 3. Execution times (maximum) per image for the system

3. The transmission time of the images to the ground station depends on the connection speed.

5 Conclusions and future work

During the tests we could show that our system can measure the road traffic, process ortho images and transmit them to the ground in real-time.

We experienced that a new scene (i.e. a city which is not covered by the training set) and different lighting conditions (e.g. cloudy weather, sunset) decrease the detection quality (i.e. lower detection rate and higher false positive rate). But a new training with a data set updated with the new images can improve the quality to almost the level of that of the old data set. Thus for a good detector samples from multiple areas, under different lighting conditions have to be used for the training.

For the application of this system some constraints have to be considered. First, because we use an optical sensor, clear weather and daylight is needed. Second, the system operates on aerial images, thus cars covered by buildings, bridges, tunnels, etc. can not be measured.

5.1 Future work

The sub-window classification of the false positive detections is challenging, even for a human, see the training images on the figure 5. For some sub-windows the correct classification is virtually not possible at this GSD. Therefore our future plan is extending the classification based on the local visual data (i.e. the sub-window) with context data (i.e. the scene around the car). This would be a next detector stage which could reject the false positives provided by the current detector. Our institute targets also developing a real-time Digital Elevation Model (i.e. 3D representation of the terrain's surface) processor. This depth information could be utilized for rejecting false positive detections, especially at building roofs and facades.

We plan to add an additional detector trained for trucks and buses, since this vehicle class is important for an accurate traffic simulation. We also work on the assignment of detected vehicles to lanes on the road, this information can

Fig. 7. Detections results on road strips

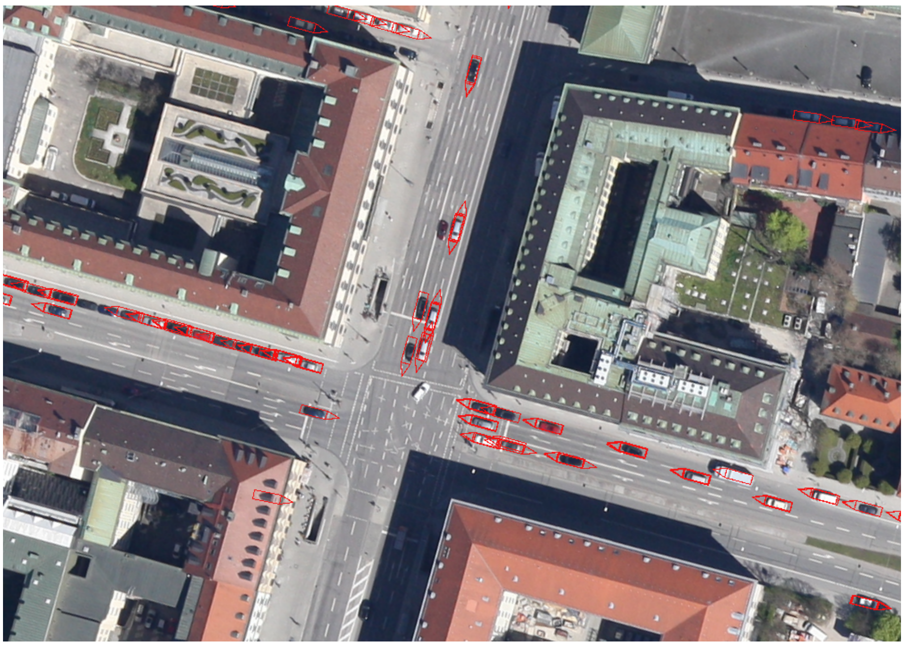


Fig. 8. Detections results on ortho image

improve the traffic analysis and simulation, especially by separating the vehicles in traffic from the parking ones.

The current system heavily depends on the orthorectification based on the accurate position and orientation measurement of the IGI™ IMU/GPS, which component has a fairly high price. A system providing similar functionality but without the IGI™ could provide a more cost effective solution, though this modification can affect the whole process chain. A possible direction for this change is detecting the cars on the original image,- without any orientation information-, instead of the road strips.

References

1. Franz Kurz, Sebastian Tuermer, Oliver Meynberg, Dominik Rosenbaum, Hartmut Runge, Peter Reinartz, Jens Leitloff. *Low-cost optical Camera System for real-time Mapping Applications*, Photogrammetrie Fernerkundung Geoinformation, Jahrgang 2012 (2), pp. 159-176. DOI: 10.1127/1432-8364/2012/0109. ISSN 1432-8364.
2. Paul Viola and Michael J. Jones. *Rapid Object Detection using a Boosted Cascade of Simple Features*, IEEE CVPR, 2001.
3. Rainer Lienhart and Jochen Maydt. *An Extended Set of Haar-like Features for Rapid Object Detection*, IEEE ICIP 2002, Vol. 1, pp. 900–903, Sep. 2002.
4. Stefan Hinz, Albert Baumgartner. *Vehicle Detection in Aerial Images Using Generic Features, Grouping and Context*, Pattern Recognition 2001, Lecture Notes on Computer Science 2191.
5. Thuy Thi Nguyen, Helmut Grabner, Barbara Gruber and Horst Bischof. *On-line boosting for car detection from aerial images*, Proc. of The 5th IEEE International Conference on "Research, Innovation and Vision for the Future (RIVF'07), Hanoi, Mar. 2007
6. Jens Leitloff, Stefan Hinz and Uwe Stilla. *Vehicle Detection in Very High Resolution Satellite Images of City Areas*, IEEE Transactions on Geoscience and Remote Sensing, 2010, Vol 48, pages 2795-2806.
7. Gellert Mattyus, Csaba Benedek and Tamas Sziranyi. *Multi target tracking on aerial videos*, ISPRS Istanbul Workshop 2010 on Modeling of optical airborne and spaceborne Sensors, WG I/4, Oct. 11-13, IAPRS Vol. XXXVIII, part 1/W4.
8. Dominik Rosenbaum, Jens Leitloff, Franz Kurz, Oliver Meynberg, Tanja Reize. *Real-Time Image Processing for Road Traffic Data Extraction from Arial Images*, ISPRS Commission VII Symposium - Vienna (Austria) Juni 2010
9. Jerome Friedman, Trevor Hastie, and Robert Tibshirani, *Additive logistic regression: A statistical view of boosting*, Ann. Stat., vol. 28, no. 2, pp. 337-407, 2000.