AUTOMATIC ANALYSIS OF TRAFFIC SCENARIO FROM AIRBORNE THERMAL INFRARED VIDEO

W. Yao [a, *], S. Hinz[a], U. Stilla [b]

[a] Remote Sensing Techbology, [b] Photogrammetry and Remote Sensing,
Technische Universitaet Muenchen, Arcisstr.21, 80290 Munich, Germany
(wei.yao, stefan.hinz, stilla)@bv.tum.de

Commission III, WG III/5

KEY WORDS: Airborne Thermal IR Video, Video Registration, Trajectory Grouping, Movement Estimation

ABSTRACT:

Traffic-related data analysis plays an important role in urban and spatial planning. Infrared video cameras have capabilities to operate at day and night and to acquire the scene sampled with video frame rate, but at the cost of geometric resolution. In this paper, an approach for the estimation of vehicle motion and the assessment of traffic activity from airborne IR video data is presented. This strategy is based on the separate handling of detection and tracking vehicle in the video, which differs from the common method developed to extract the object motion. The reason for it is that static vehicles are also intended to be detected. A single vehicle detector is firstly applied to find the vehicles in the image frames of video successively. Sensor movement is compensated by co-registering the image sequence under the selected geometric constraint. Afterwards, a progressive grouping concept considering temporal coherence and geometric relation is designed to recover the vehicle trajectories and classify them into static, moving and uncertain type. Image matching and the topology of trajectory are integrated into grouping process to aid the verification. Testing the algorithm on an IR video of urban area show us a promising result that 83% of moving vehicles are successfully extracted which is able to serve as basis for traffic density analysis.

1. INTRODUCTION

Traffic-monitoring systems rely on sensors to acquire traffic information. In last decade many ground-based sensors, e.g. loops detectors, bridge sensors and stationary cameras have been widely used and extensively studied (Hu et al., 2004). Airborne-video data acquisition for traffic-parameter estimation has been explored as an alternative to conventional data-collection methods, because it may enable us to cover a relatively broad area and potentially derive additional parameters such as travel time, relative velocity, vehicle trajectory, etc (Shastry & Schowengerdt, 2005; Cohen & Medioni, 1998; Reinartz et al., 2006). Thermal IR cameras provide us the possibility of night vision and sensing the traffic situation at day and night. Moreover, temperature is an important feature for recognition tasks and also gives an important cue for the activity of cars; temperature can be remotely sensed by IR-cameras sensitive in the 3-5µm or 8-12µm spectral band. But the activity of cars cannot only be restricted to their movement, and they have to be counted based on need of application. Independent of the color or type all vehicles appear similar with respect to their size and outer conditions (Stilla et al., 2004, Ernst et al., 2005). Often they will appear as cold spots on the warmer road surface, but for active vehicles parts of them may appear as hot spot (e.g. exhaust).

Since the videos are taken from a moving platform the simple optical flow estimation cannot be used to detect object motion, we have to distinguish the sensor movement from true object movement in the scene in order to characterize traffic activity. A number of approaches have recently been proposed to automatically detect vehicles or vehicle queues in the IR images of dense city areas or to detect moving objects and estimate vehicle movement from airborne IR video. Stilla & Michaelsen ~~(2002) have developed a method of de~~tecting single vehicle in the aerial IR images of urban areas based on spot-filtering. In Hinz & Stilla (2006) a detector for extracting single vehicles and vehicle queues combing global and local context is introduced, we can hardly get information about time from the single image. Concerning IR video, Kirchhof & Stilla (2006) have applied planar homograph as geometric tool to co-register the video data and attempted to detect and track moving objects by analyzing the motion channel. Michaelsen & Stilla (2004) have analyzed and accessed different methods for pose estimation from an oblique airborne video in order to optimize processing chain for specific scene reconstruction.

All previous works mentioned above can be regarded as the foundations and components of traffic monitoring system from airborne platforms. In this work we will give an integrated contemplation. A strategy of accessing traffic scene and estimating vehicle motion from airborne IR video is designed. The complete processing chain is implemented in the context of separation of detection from tracking of vehicles, which will be presented and discussed in following sections. Image matching is only used to aid verification of the vehicle instance to recovered trajectory.

2. DETECTION OF VEHICLES IN EACH FRAME

In order to detect vehicles in single images, we have implemented an existing automatic approach (Stilla & Michaelsen, 2002). Because of low contrast and noisy characteristic of IR image, many parking cars along two margins of the road have failed to be detected. An improvement upon this method has been made towards the increasing the completeness of vehicle detection.

The algorithm of single vehicle detection used here is based on following two assumptions:

* Corresponding author.

(1) The image is searched for cold-spots (black) which represent the vehicles themselves.
(2) For a moving vehicle, a hot-spot presenting the warmed engine bonnet must exist around a cold-spot.

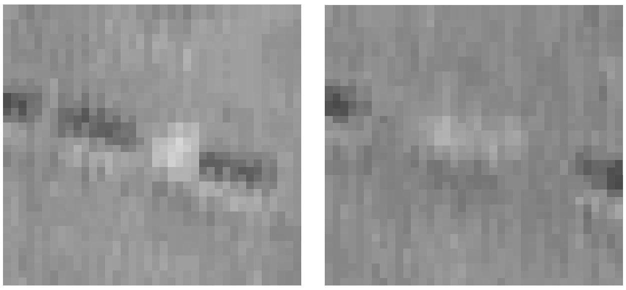


Figure.1 Examples of thermal IR-image section

Since we want to estimate traffic parameters in this project, the detection of moving vehicles is brought into focus. While applying this algorithm for vehicle detection, some imaginable problems emerge instantly, e.g. a car had started just recently, the engine bonnet has not warmed up yet; the hot spot of vehicle is missing. (Fig.1) Therefore it is at first not possible to distinguish still vehicles from moving vehicles. It is only about the detection of vehicles in this step.

As many vehicles can not adequately be distinguished from road surface in the thermal IR spectrum, the thresholding method used by the old toolbox can not segment the region for vehicle hypothesis correctly. Consequently we decided to improve the segmentation process by jointly considering the regions obtained by spot-filtering operation. This process not only provided clues for vehicle position delivered to image segmentation but also can generate hypotheses for vehicle region itself whose features were analyzed to verify the vehicle hypotheses of weak contrast. Finally, detection results from both routes have to be fused and comprise the result of vehicle detection from single IR images. (Fig. 2)

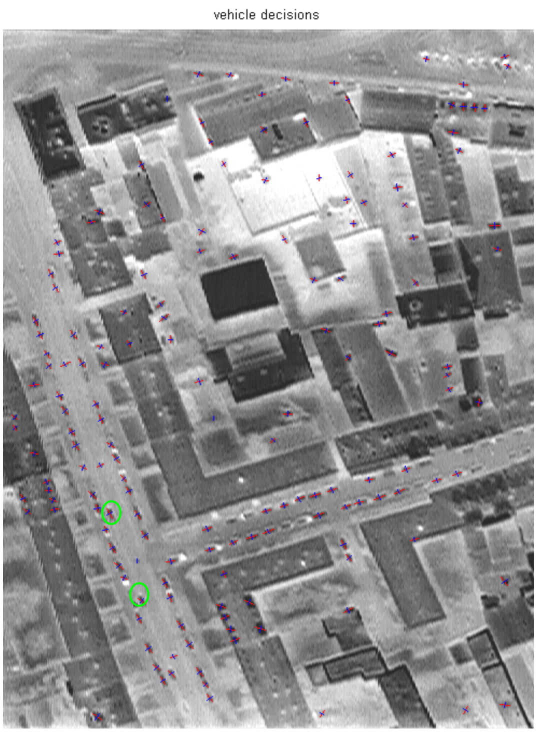


Figure.2 Result of vehicle detection in one image

A few errors emerged during the detection:

— Oversegmentation, which means that the completeness of vehicle detection has increased while the correctness reduced.

— Smokestacks and house corners are falsely extracted due to lack of context information, so we can use GIS-data in order to restrict the search of vehicles only on roads and/or parking lot.

All of test images have been processed successively to obtain the (row, col) coordinates of vehicles.

3. PROCESSING OF IMAGE SEQUENCES

Our approach of processing airborne infrared video acts as the preparatory step to obtain static and dynamic information of vehicle for performing general analysis of traffic density, which consists of two major sub-steps. The first sub-step is egomotion estimation and video stabilization, comprising the co-registration of every two images in the video, it is important for separating sensor movement from vehicle movement. In Sect.3.2 all of the vehicles detected in each frame of video by last step will be transformed into a unified coordinate frame, e.g. the domain of the last image in video. Primary product of this step is synthetic image mosaic in which the motion field induced by the displacement of the platform has been cancelled. The overall processing chain is depicted in Fig.3.

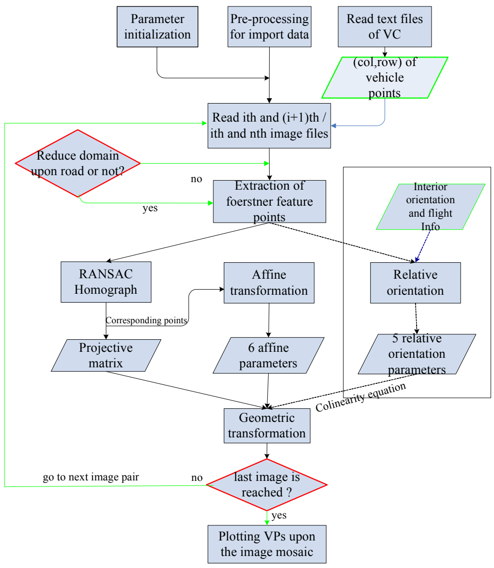


Figure.3 Workflow of approach for processing image sequences

3.1 Image sequence stabilization

To be able to establish the geometric relation between every two images, Image stabilization has been chosen to perform this task. Image stabilization consists of registering the two images and computing the geometric transformation **T** that warps the image I_1 such that it aligns with the reference image I_0. We can utilize three geometric tools listed below to realize the Co-registration:

(1) ***Planar homograph***
(2) ***Affine model***
(3) ***Perspective model*** (not implemented in this work)

There are diverse fashions to implement co-registration within the frame of whole image sequences. We use different results generated by various combination of geometric tools and transformation orders to compare and demonstrate their ability for selecting the best one.

Being independent of geometric tools, the co-registration procedure can be divided into two main processes:

(1) **Initial relative orientation**

(a) Pre-processing steps e.g. initialize some empirical parameters for operators (e.g. offset of the search window, threshold for gray value matching etc.) and read the coordinates of vehicles. Since these initial parameters have a great influence over the quality of co-registration, they should be selected carefully

(b) Feature points are extracted with Förstner-operator. The image domain can be reduced upon road region, so that it can better fit the prerequisite of planar homograph, but at the expense of the quantity of robust feature points. Therefore, this restriction is ignored in the case of directly transformation order due to low overlapping rate.

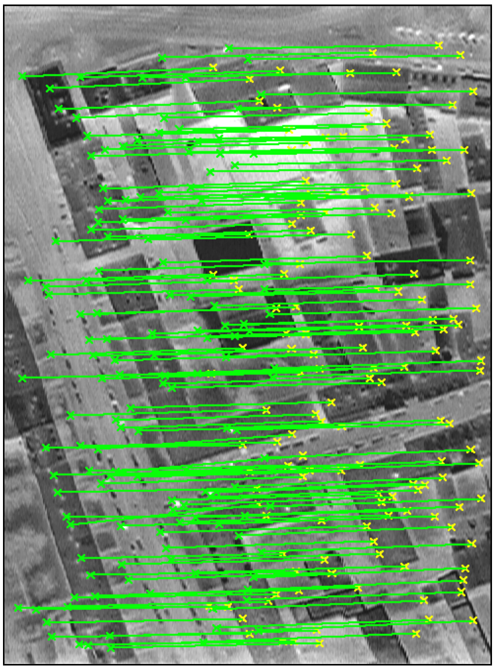


Figure.4 Correspondence found by relative orientation RANSAC

(c) Compute a projective transformation matrix between two images based on RANSAC. Feature points from both input images are read to serve as candidates of corresponding points based on gray-value matching. Once the initial matching is complete, a randomized search algorithm (RANSAC) is used to determine the transformation matrix of homograph. Because of the nature of RANSAC operator hardly yields the same result on every call.

(d) Homograph prepares the corresponding points for affine transformation. Matched input points filling the condition of RANSAC are output as additional product, namely corresponding points for affine relation. (Fig. 4)

(2) **Refinement of the orientation**

After initial relative orientation, the estimated parameters should be refined. The least squares bundle adjustment is the classical method and usually delivers the best results. Depending on the scene characteristics also some simple transformations, e.g. affine transformation, yield similar result, especially when the baseline between the images is small.

3.2 Transformation of vehicle points

After every two images are co-registered and connected by the transformation matrix, we have to sequentially project vehicle coordinates into reference image domain using homogeneous transformation matrices.

For our case, e.g. the vehicles of image 45 have to be transformed 50 times, into last image, namely image 95, die vehicles of 46.image 49 times, etc.; or the vehicles of image 45 are transformed directly into image 95. Both need the loop operations of same number.

Finally, all of the vehicles points detected from different images sampled temporally have been transformed into the coordinate frame of the last image which serves as the reference frame in our experiment, and also been plotted on the mosaicked image sequence being free of camera motion, leading to yielding the stabilized map of vehicle detection. In this map, the moving vehicles are supposed to build a trajectory, whereas the still vehicles ought to accumulate together like a point cluster. Then, we can analyze and measure vehicle trajectories on the basis of this result image, if we have plotted a cross sign at each position of transformed vehicles.

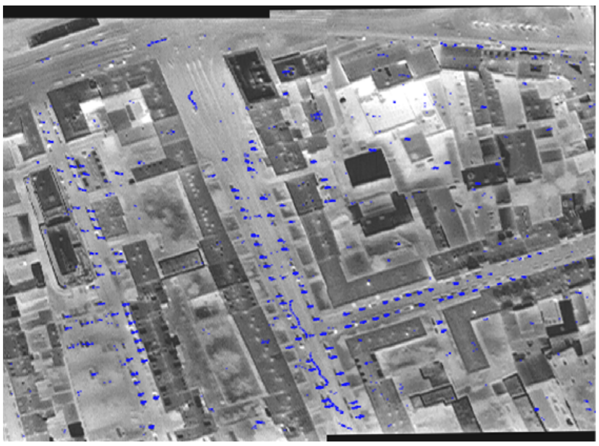


Figure.5 Stabilized map of vehicle detection by sequentially affine transformation

After accomplishing transformations of vehicle points and analyzing the experiment results of different configurations, we can estimate the vehicle activity based on trajectory plotted on the best result image — sequentially affine transformation (Fig.5), where it can be seen that parking cars accumulate nearly in the same place and slightly shift; trajectories of moving cars are obvious.

4. AUTOMATIC CHARACTERIZATION OF VEHICLE MOVEMENT

In this step, the objective of our approach is to automate the analysis process of interpreting and inferring vehicle movement by means of information acquired by airborne IR camera. The stabilized map of vehicle detection is generated by last two steps; where vehicles detected from the single IR frames are projected into coordinate system of the reference frame and depicted as blue cross. Multiple instances of one vehicle entity, corresponding to different discrete time tags of recording, tend to build and describe the temporal behavior of the vehicle trajectory. In order to characterize and analyze the traffic activity, it is required to reconstruct the trajectories of vehicles in this map, and to label them as moving or static. Our strategy to perform this task features a progressive operational concept and split and merge of trajectory based on temporal coherence and geometric relation. The correspondence relation between detected single vehicles of each frame is to be re-established here. We do not use image difference and matching to characterize the moving object just as normal methods, but rather perform detection and tracking of vehicle separately.

4.1 Coarse classification

The first step of our strategy is to classify the vehicle region as four different classes of by means of clustering analysis and temporal coherence criterion. The vehicle region map is generated by labelling connected components in the stabilized map of vehicle detection, which can be viewed here as binary image when using an image of single intensity as background. Afterwards initial vehicle regions for trajectory delineation and classification are created; they can serve as trajectory candidates for single vehicle entity. These initial regions are to be undergone the classification process according to a feature measure of region density describing cluster physical compactness. This feature measure is defined as follows:

$$FM = (Maj\_A\_L / R\_A) * (10 / Ind\_Ma - Ind\_Mi) \qquad (1)$$

where *Maj_A_L* is the length of major axis of region; *R_A* is the number of vehicle points included in a vehicle region; *Ind_Ma /Mi* is maximum/minimum index within a region.

If *FM* >=1 and *Real_area* > 1, classified as candidates for moving vehicle; If *FM* <1 and *Real_area* > 1, classified as candidates for static vehicle; If *Real_area* = 1, classified as single vehicle class. A joint consideration with compatibility of temporal index within single vehicle regions is necessary. Because vehicle instances from two vehicles in reality may merge into one initial vehicle region (hybrid class) displayed here, so it has to be delivered to split process further. The result map after this step is showed in Fig.6.

4.2 Refinement of classification results

Due to unavoidable existence of co-registration and detection errors, vehicle points belonging to static category usually do not accumulate in connected cluster. In this intermediate step we merge the green category of vehicle region map generated from last step, and analyze the white points to split them into independent vehicle regions. For analysis of static category, we take these regions as seed point, and then do a search in the close surrounding area, in which red, blue and green vehicle to be analyzed concerning temporal and geometric accordance with the seed region. In order to generate hypotheses for static vehicle, we have to further verify them via image matching; then, those regions confirmed by two operations above will be accepted as static vehicle and aggregated with seed region to build the new green class labelled as one region.

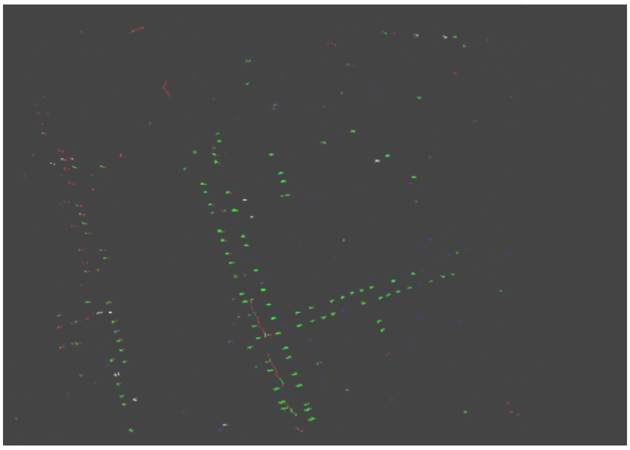


Figure.6 Vehicle region map after first classification, green: static vehicle; blue: single vehicle; red: moving vehicle; white: hybrid vehicle

One usually has to restrict the amount or eccentricity of green region, after or while merging green region with another green class, so as to exclude some ones being lacking of temporal completeness of trajectory or of inordinate trajectory elongation. The advantage of this step is that the problem domain and complexity can be reduced; we can focus on individual vehicle category by sequential processing.

4.3 Grouping and extracting vehicle trajectory

Based on results generated by last step, the green regions supposed to be static vehicle class are relatively fixed and here we focus on the red vehicle region, attempting to group these fragmented regions into reasonable trajectory of moving vehicle. The grouping algorithm is implemented by sequentially searching process based on jointly analyzing geometric relation and temporal coherence, starting from an arbitrary red vehicle region and orientating the search direction towards the major axis. The criterion for testing the compatibility between temporal index and geometry is used and formulated as below: $T_i \cap T_j = \{O\}$ *and* $\min\{T_i\} - \max\{T_j\}$ *or* $\min\{T_j\} - \max\{T_i\}$ mu -st be consistent with the distance between border points of each region: $d_{ij} = \left|R_{i,\min} - R_{j,\max}\right|$ *or* $\left|R_{j,\min} - R_{i,\max}\right|$,

$$\text{i.e,} ThreshV - \Delta t \leq d_{ij} / (T_{i,\min} - T_{j,\max}) \leq ThreshV + \Delta t \qquad (2)$$

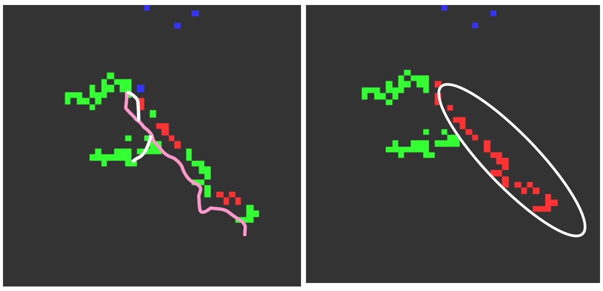
where T_i , T_j are temporal index set of two vehicle region i and j; $R_{i\ \min}$, $R_{i\ \max}$ are border points of each region; d_{ij} is distance between border points (usually max or min temporal index) of each region; *ThreshV* is the threshold relating to assumed vehicle velocity; Δt is allowable deviation affected by detection and co-registration accuracy, $(T_{i,\min} - T_{j,\max})$ can be replaced by $(T_{j,\min} - T_{i,\max})$.

After examining the assumed accordance of temporal coherence with geometric distance, we extend and link the adjacent vehicle regions to create the whole trajectory of vehicle entity, and then we take into account the topology within each trajectory of moving vehicle. It is required to achieve distribution optimization for vehicle trajectory. The graph description (Fig.8) of vehicle regions (Fig.7) is created to support this task, the extraction of the vehicle trajectory amounts to find an optimal path along each connected nodes. Defining an optimality criterion to characterize an optimal path is equivalent to associating to each edge of the graph a cost. Each edge of the graph corresponds to an image match between vehicle instances of two regions. We have also to consider the relations between each node evaluated above, since nodes describing the same vehicle entity are likely to demonstrate the optimal accordance of temporal coherence with geometric configurations. Therefore we assign for each edge connecting region i to j the following cost:

$$c_{ij} = \frac{C_{ij}}{1 + (d_{ij}/(T_{i,\min} - T_{j,\max}) - ThreshV)^2 + (\theta_i - \theta_{ji})^2} \qquad (3)$$

where, C_{ij} is the correlation between regions i and j, and θ_i, θ_{ij} represent the orientation angles of major axis vector of region i and connection vector from region i to j.

Generally, the trajectory of a moving vehicle is assumed to be resolved by adequate temporal resolution, which means it should consist of enough vehicle instances detected from at least 40% of the total frames; otherwise these trajectories have to be assigned to the vehicle object of uncertain status (behavior). Finally, all of vehicle points are grouped to vehicle entities that are classified into three categories regarding movement.



(a) (b)

Figure.7 Vehicle region map: before (a) and after (b) trajectory grouping

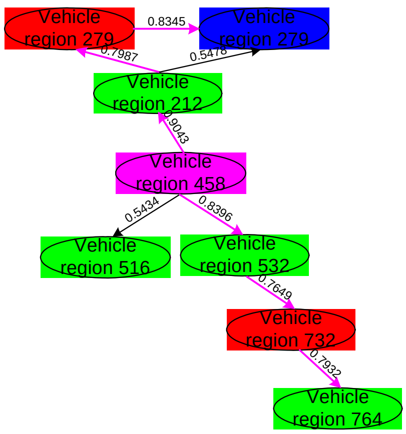


Figure.8 Graph representation and analysis of vehicle regions in Fig. 7

5. RESULTS

We used the IR video data captured over a dense build-up area, including two main roads with moving and parking vehicles on them, to test our algorithm proposed above. Fig.9 illustrates the classification result of vehicle movement concerning movement - three categories: red: moving vehicle; green: static vehicle; blue: uncertain. It can be seen that most of the parking vehicles along road sides located in the centre area covered by the IR video are detected and classified into the static class, and a large proportion of the static class vehicles are also represented by the parking vehicles, which is reasonable for this test scene. Due to similar appearance with vehicle, some green class vehicles are falsely detected among the building roofs, which correspond to chimneys in reality. Only 5 moving vehicles with distinct moving trajectory have been found. Although the number of extracted moving vehicle is much fewer compared to static class, it has represented essential dynamical information for traffic analysis. The uncertain class contains either vehicle anomaly generated in the detection step or vehicle entities whose trajectory can not be resolved by available temporal resolution, e.g. vehicles located in the margin of image mosaic.

Afterwards, we try to derive the velocity of moving vehicles based on their trajectories. Basic information concerning infrared video dataset used here can be acquired in advance:

— Pixel size (GSD) of the image: 0.5 m;

— The whole test area is covered by 51 images in all, FPS = 25 image/sec, so the duration of flight △t = 2.04 s.

The length of car's trajectories is obtained via sample pixel coordinates, which we have selected and read out from Fig.10 empirically, here the both trajectory curves are approximated with 5 sample points.

It yields following velocities:

V1 = 58 km/h
V2 = 51km/h
V3 = 42 km/h
V4 = 36km/h
V5 = 41 km/h

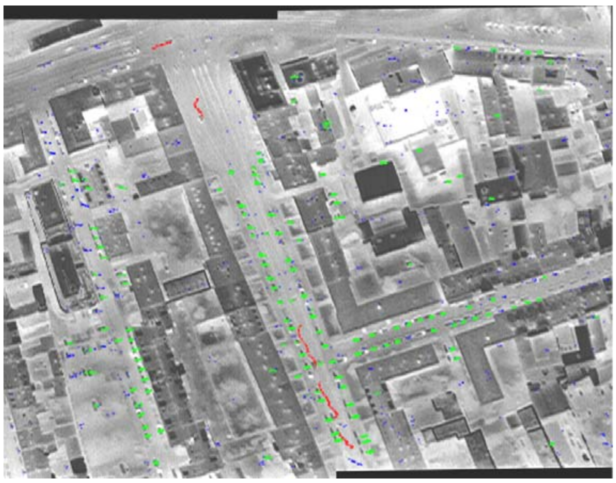


Figure.9 Classification result of vehicle movement

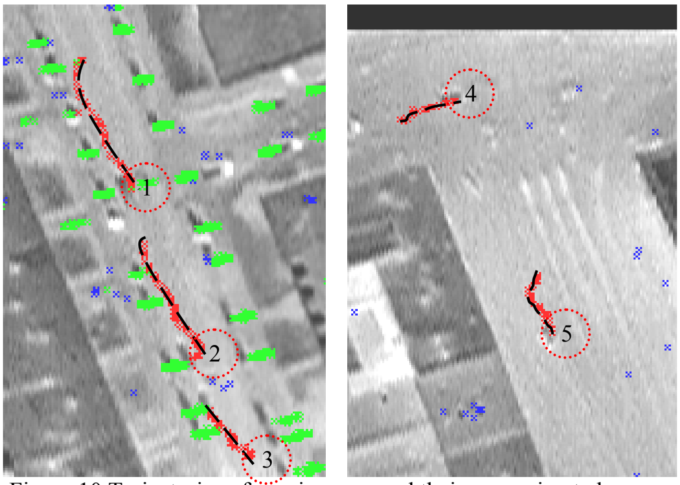


Figure.10 Trajectories of moving cars and their approximated curves

The detection and classification of static and moving vehicles for movement indication is evaluated in terms of completeness and correctness against reference data, respectively. Due to the lack of simultaneously captured optical images, the reference data used for this evaluation has been manually acquired from the same data set as used for extraction. Hence, one has to keep in mind that the above mentioned values refer to the capabilities of a human operator working with such kind of imagery. The evaluation does not refer to real "ground truth".

Evaluation criteria	static vehicle	moving vehicle
Correct decisions	112	5
False alarms	15	0
Missing decisions	26	1
Completeness [%]	81.2%	83.3%
Correctness [%]	88.2%	100%

Table.1 Evaluation of vehicle movement indication

Details of the numerical evaluation of the test video are summarized in Table.1. The velocity of moving vehicles derived by analyzing the trajectories is also not able to be verified strictly. However, the values lie in the expected velocity limit allowed in the city area, and so are plausible.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have addressed issues related to automatic analysis of airborne IR video for vehicle movement indication. Vehicles are automatically detected and distinguished with respect to their status. The velocity of vehicles can be derived subsequently as an important parameter for traffic density analysis. The proposed algorithm is driven by a progressive operation concept. Following compensation of sensor movement, vehicle instances from stabilized map are examined and grouped based on temporal coherence and geometric relation, to construct the vehicle trajectory. This process is realized by analyzing the vehicle region map and supplemented by image matching and trajectory topology. An implementation of algorithm on the test data has delivered us promising result, especially for the moving vehicle. Future work can be put on improving vehicle detector and exploiting thermal information of IR image to support motion analysis. Considerably more test data are required to verify the statistical evaluation of the algorithm performance.

ACKNOWLEDGMENT

The authors are grateful to FGAN-FOM Research Institute for Optronics and Pattern Recognition for providing video data.

REFERENCES

Angel, A., Hickman, M., Mirchandani, P., Chandnani,D., 2002. Methods of traffic data collection, using aerial video. The IEEE 5th International Conference on Intelligent Transportation Systems, 2002. Proceedings, Page(s):31 - 36

Cohen, I., Medioni,G., 1998. Detection and Tracking of Objects in Airborne Video Imagery. In: Workshop on Interpretation of Visual Motion, CVPR98, Santa Barbara

Ernst, I., Hetschler, M., Lehmann, S, Lippok, A., Ruhé, M., 2005. Use of GIS Methodology for Online Urban Traffic Monitoring. In: Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI, Part 8/W27, URBAN2005

Hinz, S., Stilla, U., 2006. Car Detection in Aerial Thermal Images by Local and Global Evidence Accumulation. In: Pattern Recognition Letters, 27: 308-315

Hu, W., Tan, T., Wang, L., Maybank, S., 2005. A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, 35(3):435-441

Kirchhof, M., Stilla, U., 2006. Detection of moving objects in airborne thermal videos. ISPRS Journal of Photogrammetry and Remote Sensing, 61(3-4):187-196

Michaelsen, E., Stilla, U., 2003. Estimation of vehicle movement in urban areas from thermal video sequences. 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and data fusion on urban areas, URBAN 2003. 105-109

Michaelsen, E., Stilla, U., 2004. Pose estimation from airborne video sequences using a structural approach for the construction of homographies and fundamental matrice. In: Fred A, Caelli T, Duin RPW, Campilho A, Ridder Dde (eds) Structural, syntactic, and statistical pattern recognition: S+SSPR, LNCS 3138, Berlin: Springer, 486-494

Paska, E., Toth, C., 2004. A Performance Analysis on Vehicle Detection from Remotely Sensed Imagery, Proceedings of the ASPRS Annual Conference, Denver, May 23-28

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, 61(3-4):149-158

Shastry, AC., Schowengerdt, RA., 2005. Airborne Video Registration and Traffic-Flow Parameter Estimation. IEEE Transactions on Intelligent Transportation Systems, 6(4):391-405

Stilla, U., Michaelsen E., 2002. Estimating vehicle activity using thermal image sequences and maps. Symposium on geospatial theory, processing and applications. International Archives of Photogrammetry and Remote Sensing. Vol. 34, Part 4

Stilla, U., Michaelsen, E., Soergel, U., Hinz, S, Ender, HJ., 2004. Airborne monitoring of vehicle activity in urban areas. International Archives of Photogrammetry and Remote Sensing. Vol 35, Part B3, 973-979