AUTOMATIC VEHICLE TRAJECTORY EXTRACTION FOR TRAFFIC ANALYSIS FROM AERIAL VIDEO DATA

Jiří Apeltauer[a], Adam Babinec[b], David Herman[b], Tomáš Apeltauer[a]

[a] Faculty of Civil Engineering, Brno University of Technology, Czech Republic - (apeltauer.j, apeltauer.t)@fce.vutbr.cz
[b] RCE systems s.r.o., Svatopluka Čecha 1d, 612 00 Brno, Czech Republic - (adam.babinec, david.herman)@rcesystems.cz

KEY WORDS: traffic monitoring, vehicle detection, vehicle tracking, aerial imagery, UAV, particle filter

ABSTRACT:

This paper presents a new approach to simultaneous detection and tracking of vehicles moving through an intersection in aerial images acquired by an unmanned aerial vehicle (UAV). Detailed analysis of spatial and temporal utilization of an intersection is an important step for its design evaluation and further traffic inspection. Traffic flow at intersections is typically very dynamic and requires continuous and accurate monitoring systems. Conventional traffic surveillance relies on a set of fixed cameras or other detectors, requiring a high density of the said devices in order to monitor the intersection in its entirety and to provide data in sufficient quality. Alternatively, a UAV can be converted to a very agile and responsive mobile sensing platform for data collection from such large scenes. However, manual vehicle annotation in aerial images would involve tremendous effort. In this paper, the proposed combination of vehicle detection and tracking aims to tackle the problem of automatic traffic analysis at an intersection from visual data. The presented method has been evaluated in several real-life scenarios.

1. INTRODUCTION

A detailed analysis and evaluation of traffic flow is essential for a precise design and development of transport infrastructure. The current primary sources of traffic statistics are measurement stations based on induction loops and ultrasonic sensors, which count vehicles that pass a given point on the road. These conventional solutions typically provide data only in the form of basic frequency statistics, i.e. Annual Average Day Traffic (AADT). However, these systems have obvious shortcomings due to fixed installation, a very limited field of view and the type of sensors utilized.

Intersections are the most limiting factor to the capacity of the whole transport network, and therefore it is necessary to pay close attention to their design. Complex road junctions are of various types and can take up a large area, which is difficult to monitor in its entirety. For that reason, the standard monitoring systems are often placed only at the entrances and exits. However, this can be very restrictive in situations where the information about the behaviour of traffic participants during the passage through the intersection is crucial (e.g. the evaluation of the intersection design, a detailed comparison with simulated data, etc.). Furthermore, deploying fixed cameras or other on-ground counters over such wide range typically requires massive investment (which is far from realistic) and so alternatives are needed.

Aerial video surveillance using a wide field-of-view sensor has provided new opportunities in traffic monitoring over such extensive areas. In fact, unmanned aircraft systems equipped with automatic position stabilization units and high resolution cameras could be the most effective choice for data acquisition in sufficient quality. UAVs, unlike satellites or airplanes, are able to collect visual data from low altitudes, and therefore provide images with adequate spatial resolution for further traffic inspection, i.e. vehicle detection and tracking.

In this paper, we are concerned with tracking multiple vehicles in order to obtain detailed and accurate information about the vehicles' trajectories in the course of their passage through the intersection. The visual data for the analysis are captured by a camera

mounted on an UAV. The operating time of a UAV is about twenty minutes due to its rapid consumption of battery power. The data is not analysed on board but recorded on a memory card and subsequently post-processed on a high-performance computer.

The remaining part of the present paper is organized as follows: Section 2 provides an overview of the related research conducted in the field of traffic surveillance and aerial image processing. Section 3 outlines the architecture of the proposed vehicle detection and tracking system. In section 4 our vehicle detection algorithm is described. A utilized vehicle tracking framework is presented in section 5. Section 6 provides insight into evaluation experiments and their results. Section 7 discusses the performance and applicability of the proposed system and possible future improvements.

2. RELATED WORK

Over the past few years, aerial image processing has become a popular research topic because of increased data availability. Aerial images can cover a large area in a single frame, which makes them attractive for monitoring and mapping tasks. Therefore, the utilization of UAVs operating in low-altitude for traffic inspection has been a major research interest in the past decade; an introduction to the current trends can be found in this brief survey paper (Lee and Kwak, 2014). Generally speaking, the task can be divided into two essential parts: vehicle detection and vehicle tracking.

The design of efficient and robust vehicle detection methods in aerial images has been addressed several times in the past. In general, these methods can be classified into two categories depending on whether explicit or implicit model is utilized (Nguyen et al., 2007). The explicit model approach uses a generic 2D or 3D model of vehicle and the detection predominately relies on geometric features like edges, lines and surfaces (Zhao and Nevatia, 2001, Moon et al., 2002). Kozempel and Reulke (Kozempel and Reulke, 2009) provided a very fast solution which is based on four special shaped edge filters aimed to represent an average vehicle. These filters, however, have to be pointed in a cor-
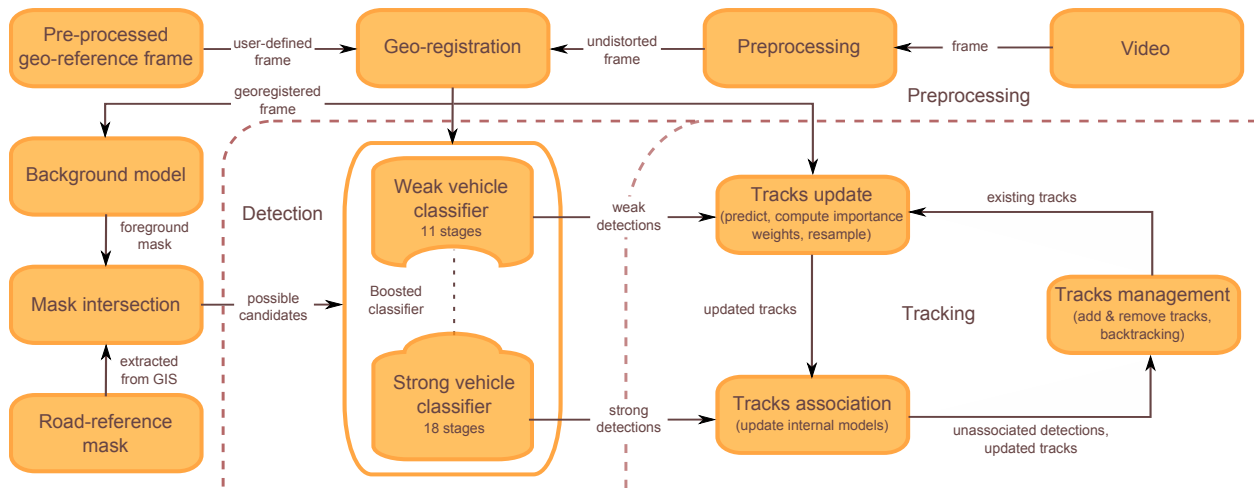
Figure 1. System architecture – the Boosted classifier is divided into two parts, the weak and the strong classifier. The detections taken from the strong classifier are used for initialization of new targets and/or for the update of the appearance models of existing targets when an association occurs. The detections returned by the weak classifier are used as clues in the tracking procedure.

rect direction according to the street database. In case of implicit approaches, the internal representation is derived through collecting statistics over the extracted features like histogram of oriented gradients (HoG), local binary patterns (LBP), etc. The detection for candidate image regions is performed by computing the feature vectors and classifying them against the internal representation (Nguyen et al., 2007, Sindoori et al., 2013, Lin et al., 2009, Tuermer et al., 2010). Among the main generic disadvantages of these approaches are the need for a huge amount of annotated training data, a lot of miss detections of rotated vehicles, and computational expensiveness during the training phase (the features are usually passed to a cascade classifier training algorithm).

To achieve real-time performance, Gleason and Nefian (Gleason et al., 2011) employed a two-stage classifier. The first stage performs a fast filtration based on the density of corners, colour profiles, and clustering. The second stage is more complex: it computes HoG and histogram of Gabor coefficients as features for binary classifier. Similar preprocessing phase is often used in order to make the detection faster and more reliable (Moranduzzo and Melgani, 2014). Another strategy for elimination of false positive detections is restricting the areas for vehicle detection by application of street extraction techniques (Pacher et al., 2008, Tuermer et al., 2010). In contrast to the above mentioned approaches, which take information from a single image only, Truemer et al. (Tuermer et al., 2011) tried to enhance the performance by incorporating temporal information from motion analysis into the detection process. In (Xiao et al., 2010), the authors also employed a motion analysis by a three-frame subtraction scheme; moreover, they proposed a method for track associations by graph matching and vehicle behaviour modelling. Next to the region or sliding window based method in (Cheng et al., 2012), they also designed a pixel-wise detector of vehicles which employs dynamic Bayesian network in the classification step. (Selvakumar and Kalaivani, 2013) presents a brief comparative study of detection techniques.

Video-based moving object tracking is one of the most popular research problems in computer vision. However, it is still a challenging task due to the presence of noise, occlusion, dynamic and cluttered backgrounds, and changes in the appearance of the tracked object, which are all very common in aerial images. Numerous tracking approaches have been presented in recent years; a detailed survey can be found in (Yilmaz et al., 2006). Our goal

is to obtain the trajectories of targets over time and to maintain a correct, unique identification of each target throughout. Continuous tracking of multiple similar targets becomes tricky when the targets pass close to one another, which is very common at intersections. One of the early attempts to deal with the occlusions for traffic surveillance was proposed by Koller et al. (Koller et al., 1993), employing an explicit occlusion reasoning coupled with Kalman filters. However, to speed up the process of tracking and to accommodate the non-Gaussianness nature of the problem, a group of sequential Monte Carlo methods, also known as Particle Filters, is utilized (Rothrock and Drummond, 2000, Danescu et al., 2009, Hue et al., 2002). Particle filters can be discriminatively trained for specific environment and different objects-to-be-tracked tasks, as demonstrated by Hess and Fern in (Hess and Fern, 2009). Current approaches in vehicle tracking from aerial or satellite imagery aim at off-line optimization of data association, e.g. by deploying bipartite graph matching (Xiao et al., 2010, Reilly et al., 2010) or by revising temporal tracking correspondence as done by Saleemi and Shah in (Saleemi and Shah, 2013) by maintaining multiple possible candidate tracks per object using a context-aware association (vehicle leading model, avoidance of track intersection) and applying a weighted hypothetical measurement derived from the observed measurement distribution.

3. SYSTEM OVERVIEW

This paper proposes a method for detection and tracking of vehicles passing through an intersection for a detailed traffic analysis. The results are used for evaluation of the design of intersection and its contribution in the traffic network. The output from the analysis needs to be in the orthogonal coordinate system of the analysed intersection; therefore the transformation between the reference image and the intersection's coordinate system is known. For simplicity's sake the interchange types of intersections are not addressed and the analysed area is approximated by a plane. Figure 1 depicts the overall design of the system, which can be divided into three main parts: preprocessing, vehicle detection, and tracking.

In the preprocessing step, the acquired image is undistorted and geo-registered against a user-selected reference frame. The methods for image undistortion have been addressed in literature several times (Mallon and Whelan, 2004, Wang et al., 2011, Beauchemin and Bajcsy, 2001). In our case, radial and tangential

distortions are employed. The perspective transformation model is used in the geo-registration process. First, local ORB features (Rublee et al., 2011) are extracted both from the acquired undistorted frame and the reference frame. The features are then matched based on their descriptor distances and cross-validated, forming pairs of points which are used for estimation of geometrical transformation. Robustness of the algorithm is achieved utilizing RANSAC procedure (Fischler and Bolles, 1981).

Due to its outstanding detection performance, the boosting technique introduced by Viola and Jones was adopted for the vehicle detection (Viola and Jones, 2001, Lienhart and Maydt, 2002). In order to increase robustness to changes in illumination and to accelerate the training phase, Multi-scale Block Local Binary Patterns (MB-LBP) was employed. The searching space is restricted to the intersection of the motion and street masks with aim to considerably decrease false positive rate and computational demands. The detections which are not associated with existing tracks are added to the tracker as new targets.

For tracking, a sequential particle filter has been adopted. However, due to exponential complexity in the number of tracked targets, the system utilizes a set of fully independent Bootstrap particle filters (Isard and Blake, 1998), one filter per vehicle, rather than the joint particle filter (Khan et al., 2003). A target is represented by a gradually updated rectangular template. To further improve the tracker's robustness to cluttered background, a weak vehicle classifier with high positive detection rate is introduced to generate possible candidates for each frame. In fact, the weak classifier is obtained as an earlier stage of the robust one. Therefore, the acquired detections naturally contain a lot of false alarms; however, the probability of true positives is also not inconsiderable. Thanks to the high frame rate of input video, it seems to be beneficial to assume that the true positive detections and the predicted target states are very close to each other. The application of this assumption can effectively eliminate false alarms and can help avoid the tracking failures. The following sections provide detailed explanations of the most important parts of the whole system.

4. DETECTION

To detect vehicles, Viola and Jones's AdaBoost algorithm was utilized for the selection of appropriate features and construction of a robust detector (i.e. binary classifier). In their original paper, the authors used HAAR features for object representations; however, HAAR features have poor robustness to illumination changes and lead to a high false alarm rate (Ju et al., 2013). To alleviate these challenges and accelerate the learning phase, MB-LBP has been employed (Liao et al., 2007). Comparing with the original LBP calculated in a 3x3 pixel neighbourhood, MB-LBP is computed from average values of block sub-regions; therefore, MB-LBP is more robust since it encodes not only microstructures but also macrostructures, and provides more complete image representation. According to (Ju et al., 2013), MB-LBP features have a comparable hit rate to HAAR features, but a significantly smaller false alarm rate, which is crucial in the proposed system. Classification is performed in multi-scale and the obtained detections are grouped together with respect to their spatial similarity. The number of neighbours is used as a measure of confidence for determination of further addition of an unassociated vehicle as a new target to the tracker.

In order to eliminate false positives and accelerate detection, we restricted the searching area only to road surface. In our case, this information can be easily extracted from GIS because of the implicit requirement of geo-registration process. Moreover, the

foreground mask is generated by background subtraction method based on Gaussian Mixture model (Zivkovic, 2004), and it is subsequently intersected with the acquired street mask. Afterwards, several morphological operations (erosion and dilation) are performed on the result of the intersection to reduce noise. This strategy reduces the false positive rate of the detector significantly.

The detector was trained on a hand annotated training dataset with 20,000 positive and 20,000 negative samples taken from aerial videos. The size of the samples is 32×32 pixels. The positive samples contain cars of different types, colours, and orientations. The negative samples were created from the surroundings of the analysed intersection, as well as from the road surface with the emphasis on horizontal traffic signs. Examples of both positive and negative samples are shown in Figure 2.
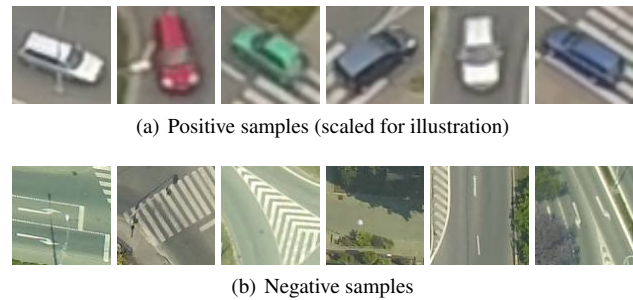

(a) Positive samples (scaled for illustration)

(b) Negative samples

Figure 2. Examples of samples used to train vehicle detector.

5. TRACKING

Over the last few years, particle filters, also known as sequential Monte Carlo methods, have proved to be a powerful technique for tracking purposes owing to their simplicity, flexibility, and the ability to cope with non-linear and non-Gaussian tasks (Rothrock and Drummond, 2000). Yet, particle filters may perform poorly when the posterior density is multi-modal. To overcome this difficulty, each target can be tracked by an individual particle filter as used in (Schulz et al., 2001, Danescu et al., 2009). In contrast to the approaches based on the extension of the state-space to include all targets, this approach considerably reduces the number of particles which are necessary for reliable tracking. Next, these individual particle filters can easily interact through the computation of the particle weights to handle the tracking scenario constraints (e.g. two vehicles cannot occupy the same place).

For individual trackers, a subtype of sequential importance sampling filter, Bayesian bootstrap particle filter, has been employed. Bootstrap filter uses transition density as a proposal density and performs resampling step in each iteration. In what follows, let $MPF = (C, \{\mathcal{X}_i\}_{i \in C}, \{M_i\}_{i \in C}, \mathcal{W}, t, \mathcal{E})$ denote a simplified particle representation of the $|C|$ tracked vehicles, where C represents the index set, $\mathcal{X}_i$ represents the set of particles and M_i is set of internal representations which belong to the i-th target. $\mathcal{W}$ stands for a function which computes the importance weight, t represents a transition function, and $\mathcal{E}$ returns the estimated state of the target as a single measurement.

5.1 Target representation and update

Target is represented by a rectangular descriptor template consisting of 4 channels: the sum of absolute response of Scharr operator and 3 colour channels (red, green, blue); all extracted from the processed image. This type of representation is able to carry both spatial and colour information, and its computation is very fast. To further emphasise the central part of the template, where the vehicle is to be present, we deploy a circular weighted mask.

The template is updated over the period of tracking only if one or both of the following events happen:

- The strong classifier yields a new detection in place where a significant overlap with the template occurs.

- The *heat* condition c_h of weak classifier is fulfilled, which is as follows:

$$c_h(\mathbf{x}) = true \Leftrightarrow \sum_{d \in \mathcal{D}_{weak}} f(\mathbf{x}, \mathbf{x}_d, 1.5) > T_{heat}, \quad (1)$$

where $\mathcal{D}_{weak}$ is a set of current detections of the weak classifier, $\mathbf{x}_d$ is a position of the detection d in the geo-registered image, $f(\mathbf{x}, \boldsymbol{\mu}, \sigma)$ represents the value of a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ expressed at center $\mathbf{x}$ of the tracked object, $\boldsymbol{\Sigma}$ is diagonal 2×2 matrix with values of σ^2 at its diagonal, and $T_{heat} = 4$ is the predefined threshold.

Additionally, to prevent undesirable swaps between the targets, the template update is disabled if multiple targets overlap. On the event of template update, the values of the template are altered by weighted average of the former template and new template extracted from the currently processed frame, where the former template has the weight of 0.95 and new the template has the weight of 0.05. This way we achieve plasticity of the target representation over the period of time while still being stable enough to keep the tracker from drifting away from the target.

5.2 Target state and motion model

In order to cover every movement of the target, a weak motion modelling was utilized. This approach brings an enhanced capability to overcome high perturbations and therefore may be more robust in situations where the movement modelling is cumbersome (typically at intersections). The particle approximation of the i-th target state consists of $\mathcal{X}_{(i)} = \{(x, y, s)|x, y, s \in \mathbb{R}\}$ where (x, y) is the location of the target in the intersection coordinate system, and s is the size of the rectangular bounding box. The components of the state are assumed to follow independent Gaussian random walk models with variances $(\sigma_x^2, \sigma_y^2, \sigma_s^2)$. The estimated target state is represented by the highest-weighted particle (*maximum a posteriori*), i.e. $\mathcal{E}(i) = \underset{p \in \mathcal{X}_{(i)}}{\mathrm{argmax}}(\mathcal{W}(p, i))$.

5.3 Importance weight computation

The proposed importance weight computation is composed of two parts: appearance similarity and AdaBoost attraction factor, and it is defined as:

$$\mathcal{W}(p, i) = e^{App(p,i) \cdot Ada(p,i)} \quad (2)$$

Let $\varphi(p, I)$ be a function which returns the descriptor template for image patch of interest in the current registered frame I parametrised by particle p and resized to the size of internal representation $T \in M_{(i)}$. Then, the appearance similarity between the internal representation and the obtained template is computed as a sum of weighted absolute differences:

$$App(p, i) = \\ \frac{\sum_{(x,y) \in T} (1 - \mathrm{absdiff}(\varphi(p, I)_{(x,y)}, T_{(x,y)}))W_{(x,y)}}{n|T|} \quad (3)$$

where W is a circular weighted mask, n denotes the number of descriptor template channels and $|.|$ returns the number of pixels

of the template (all intensity values of channels are normalized to the interval of $[0, 1]$).

AdaBoost attraction factor substantially helps to overcome the situations in which the background is heavily cluttered. In such cases, the measure of the appearance similarity is not discriminative enough and the tracker may be prone to failure. To alleviate this difficulty, the detections produced by the weak classifier are used as clues during the tracking. Let $\mathcal{D}_{weak}$ be a set of detections returned by the weak classifier; then the attraction factor is defined as follows:

$$Ada(p, i) = \sum_{d \in \mathcal{D}_{weak}} f(\mathbf{x}, \mathbf{x}_d, 1.5), \quad (4)$$

where $\mathbf{x}$ is a position of the evaluated particle p, $\mathbf{x}_d$ is a position of the detection d, and function $f(\mathbf{x}, \boldsymbol{\mu}, \sigma)$ is the same as in the Equation 1.

5.4 Tracking termination

Tracking of the i-th target is automatically terminated when $\mathcal{E}(i)$ is outside the road mask, or when there was no association with any detection produced by the strong classifier for the predefined amount of time (frames). The results are recorded only for instances for which the input and output lanes are known. If the target reached the output lane and the input lane is unknown, backward tracking is utilized from the first detection in order to try to determine the input lane. The backward tracking algorithm is basically the same as the forward version; the frames are processed in a reversed order and the target is only tracked until the input lane is discovered, or until any termination condition is met.

6. EXPERIMENTS

The system presented in this paper has been evaluated on two sequences of video data captured by action camera GoPro Hero3 Black Edition mounted on a UAV flown at the height of approx. $100\,\mathrm{m}$ above the road surface. The video was captured with the resolution of $1920\,\mathrm{px} \times 980\,\mathrm{px}$ at $29\,\mathrm{Hz}$. Due to utilization of ultra-wide angle lens, the diagonal field of view was $139.6°$. The spatial resolution of the captured scene was approximately $10.5\,\mathrm{cm/px}$. In the course of data acquisition the UAV was stabilized around a fixed position in the air.

The first sequence was captured near Netroufalky construction site in Bohunice, Brno, Czech Republic. The second sequence was captured at the site of roundabout junction of Hamerska road and Lipenska road near Olomouc, Czech Republic.

The evaluation was carried out against ground truth annotated by hand consisting of trajectories of all vehicles that both fully entered and exited the crossroad area during the evaluation sequence. As high level evaluation metrics we used relative number of missed targets $\mathbf{NMT_r} = \frac{\mathbf{NMT}}{|L|}$, relative number of false tracks $\mathbf{NFT_r} = \frac{\mathbf{NFT}}{|L|}$, average number of swaps in tracks $\mathbf{ANST}$ and temporal average of measure of completeness $\mathbf{MOC_a}$ which is defined as follows:

$$\mathbf{MOC_a} = \frac{\sum_{k=0}^{n_{video}} Comp(k)}{n_{video}}. \quad (5)$$

$\mathbf{NMT}$ and $\mathbf{NFT}$ are defined as in (Gorji et al., 2011) but with respect to the whole evaluation sequence, $|L|$ is a number of ground truth tracks, n_{video} is the number of images in the evaluation sequence, $\mathbf{ANST}$ and $Comp(k)$ are described in (Gorji et al., 2011) as well.

(a) Netroufalky

(b) Olomouc

Figure 3. Scenes used for evaluation.

The spatial precision of the algorithm was evaluated using root mean square error (**RMSE**) of track position averaging over all valid tracks. The estimated track $\mathcal{E}_i$ is considered corresponding to the ground truth l_j at given time moment k if the track $\mathcal{E}_i$ is the nearest track to the truth l_j at the moment k and vice versa, and the distance between them is less than threshold $t_{dist} = 3\,\mathrm{m}$. The estimated track is valid if it corresponds to the same ground truth for the whole period of the vehicle presence in the analysed area.

Sequence	Netroufalky	Olomouc
True Tracks #	165	168
Estimated Tracks #	173	153
Valid Estimated Tracks #	143	115
NMT$_\mathbf{r}$	0.054	0.197
NFT$_\mathbf{r}$	0.052	0.163
ANST	0.049	0.066
MOC$_\mathbf{a}$	0.841	0.667
RMSE $[m]$	1.045	1.102

Table 1. The results of the evaluation.

A user of the traffic analysis tool may want to inspect and fix invalid tracks. To indicate the necessary effort that is needed to fix the invalid tracks the following graph shows the dependency of the number of missed true tracks (true tracks that were not assigned to any valid estimated track) on the rate of adjustments needed to the best partially matching estimated tracks.
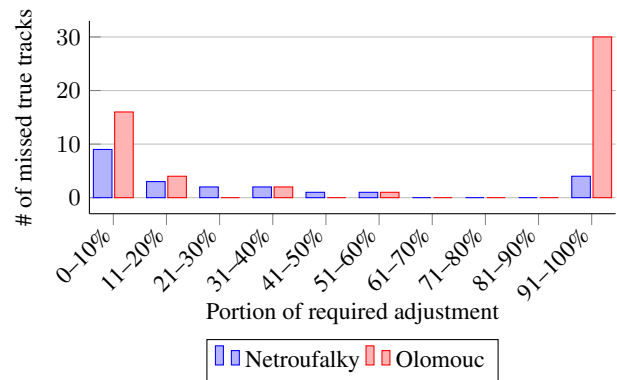

Figure 4. Histogram of missed true tracks, according to required adjustment to best matching fragments.

7. CONCLUSION

In this paper, we proposed a system for vehicles' trajectories extraction from aerial video data captured by a UAV. The functionality of the system was demonstrated on two extensive hand annotated data sets. Our approach shows sufficient performance for automatic extraction of vehicles' trajectories for further traffic inspection. Moreover, we illustrated that the manual effort needed for the correction of most of the missed trajectories to get more accurate results is negligible. Several questions remain open for future research. It would be interesting to handle the road junctions with grade separations, as well as using data fusion from more UAVs.

ACKNOWLEDGMENT

REFERENCES

Beauchemin, S. S. and Bajcsy, R., 2001. Modelling and removing radial and tangential distortions in spherical lenses. In: *Multi-Image Analysis*, Lecture Notes in Computer Science, Vol. 2032, Springer Berlin Heidelberg, Berlin, pp. 1–21.

Cheng, H.-Y., Weng, C.-C. and Chen, Y.-Y., 2012. Vehicle detection in aerial surveillance using dynamic bayesian networks. *IEEE Transactions on Image Processing*, 21(4), pp. 2152–2159.

Danescu, R., Oniga, F., Nedevschi, S. and Meinecke, M., 2009. Tracking multiple objects using particle filters and digital elevation maps. In: *2009 IEEE Intelligent Vehicles Symposium*, pp. 88–93.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp. 381–395.

Gleason, J., Nefian, A., Bouyssounousse, X., Fong, T. and Bebis, G., 2011. Vehicle detection from aerial imagery. In: *2011 IEEE International Conference on Robotics and Automation*, pp. 2065–2070.

Gorji, A. A., Tharmarasa, R. and Kirubarajan, T., 2011. Performance measures for multiple target tracking problems. In: *Proceedings of the 14th International Conference on Information Fusion*.

Hess, R. and Fern, A., 2009. Discriminatively trained particle filters for complex multi-object tracking. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 240–247.

Hue, C., Le Cadre, J.-P. and Perez, P., 2002. Sequential monte carlo methods for multiple target tracking and data fusion. *IEEE Transactions on Signal Processing*, 50(2), pp. 309–325.

Isard, M. and Blake, A., 1998. Condensation — conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1), pp. 5–28.

Ju, Y., Zhang, H. and Xue, Y., 2013. Research of Feature Selection and Comparison in AdaBoost based Object Detection System. *Journal of Computational Information Systems*.

Khan, Z., Balch, T. R. and Dellaert, F., 2003. Efficient particle filter-based tracking of multiple interacting targets using an mrf-based motion model. In: *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 254–259.

Koller, D., Weber, J. and Malik, J., 1993. Robust multiple car tracking with occlusion reasoning. Technical Report UCB/CSD-93-780, EECS Department, University of California, Berkeley, Berkeley.

Kozempel, K. and Reulke, R., 2009. Fast vehicle detection and tracking in aerial image bursts. *International Society for Photogrammetry and Remote Sensing*, 38(3), pp. 175 – 180.

Lee, J.-N. and Kwak, K.-C., 2014. A trends analysis of image processing in unmanned aerial vehicle. *International Journal of Computer, Information Science and Engineering*, 8(2), pp. 2 – 5.

Liao, S., Zhu, X., Lei, Z., Zhang, L. and Li, S.-Z., 2007. Learning multi-scale block local binary patterns for face recognition. In: *Advances in Biometrics*, Lecture Notes in Computer Science, Vol. 4642, Springer Berlin Heidelberg, pp. 828–837.

Lienhart, R. and Maydt, J., 2002. An extended set of haar-like features for rapid object detection. In: *Proceedings of 2002 International Conference on Image Processing*, Vol. 1, pp. 900–903.

Lin, R., Cao, X., Xu, X., Wu, C. and Qiao, H., 2009. Airborne moving vehicle detection for video surveillance of urban traffic. In: *2009 IEEE Intelligent Vehicles Symposium*, pp. 203–208.

Mallon, J. and Whelan, P., 2004. Precise radial un-distortion of images. In: *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 1, pp. 18–21 Vol.1.

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

Moranduzzo, T. and Melgani, F., 2014. Automatic car counting method for unmanned aerial vehicle images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(3), pp. 1635–1647.

Nguyen, T. T., Grabner, H., Gruber, B. and Bischof, H., 2007. On-line boosting for car detection from aerial images. In: *IEEE International Conference on Reasearch, Innovation and Vision for the Future*, pp. 87–95.

Pacher, G., Kluckner, S. and Bischof, H., 2008. An improved car detection using street layer extraction. In: *Proceedings of the 13th Computer Vision Winter Workshop*, Ljubljana.

Reilly, V., Idrees, H. and Shah, M., 2010. Detection and tracking of large number of targets in wide area surveillance. In: *11th European Conference on Computer Vision*, Vol. 6313, Springer-Verlag Berlin Heidelberg, Berlin, pp. 186–199.

Rothrock, R. L. and Drummond, O. E., 2000. Performance metrics for multiple-sensor multiple-target tracking. In: *Signal and Data Processing of Small Targets 2000*, Proceedings of SPIE, Vol. 4048, pp. 521–531.

Rublee, E., Rabaud, V., Konolige, K. and Bradski, G., 2011. Orb: An efficient alternative to sift or surf. In: *2011 IEEE International Conference on Computer Vision*, pp. 2564–2571.

Saleemi, I. and Shah, M., 2013. Multiframe many-many point correspondence for vehicle tracking in high density wide area aerial videos. *International Journal of Computer Vision*, 104(2), pp. 198–219.

Schulz, D., Burgard, W., Fox, D. and Cremers, A. B., 2001. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In: *IEEE International Conference on Robotics and Automation.*, Vol. 2, pp. 1665–1670 vol.2.

Selvakumar, S. and Kalaivani, S., 2013. Comparative study on vehicle detection techniques in aerial surveillance. *International Journal on Cybernetics & Informatics*.

Sindoori, R., Ravichandran, K. and Santhi, B., 2013. Adaboost technique for vehicle detection in aerial surveillance. *International Journal of Engineering and Technology*, 5(2), pp. 765 – 769.

Tuermer, S., Leitloff, J., Reinartz, P. and Stilla, U., 2010. Automatic vehicle detection in aerial image sequences of urban areas using 3d hog features. *Photogrammetric Computer Vision and Image Analysis*, 28, pp. 50–54.

Tuermer, S., Leitloff, J., Reinartz, P. and Stilla, U., 2011. Motion component supported boosted classifier for car detection in aerial imagery. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, Vol. 38 (3/W22), pp. 215–220.

Viola, P. and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. I–511–I–518 vol.1.

Wang, H., Li, M. and Zhang, L., 2011. The distortion correction of large view wide-angle lens for image mosaic based on opencv. In: *International Conference on Mechatronic Science, Electric Engineering and Computer*, pp. 1074–1077.

Xiao, J., Cheng, H., Sawhney, H. and Han, F., 2010. Vehicle detection and tracking in wide field-of-view aerial video. In: *2010 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 679–684.

Yilmaz, A., Javed, O. and Shah, M., 2006. Object tracking: A survey. *ACM Computing Surveys*.

Zhao, T. and Nevatia, R., 2001. Car detection in low resolution aerial images. In: *Eight IEEE International Conference on Computer Vision*, IEEE, Vancouver, pp. 710–717.

Zivkovic, Z., 2004. Improved adaptive gaussian mixture model for background subtraction. In: *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 2, pp. 28–31 Vol.2.

APPENDIX
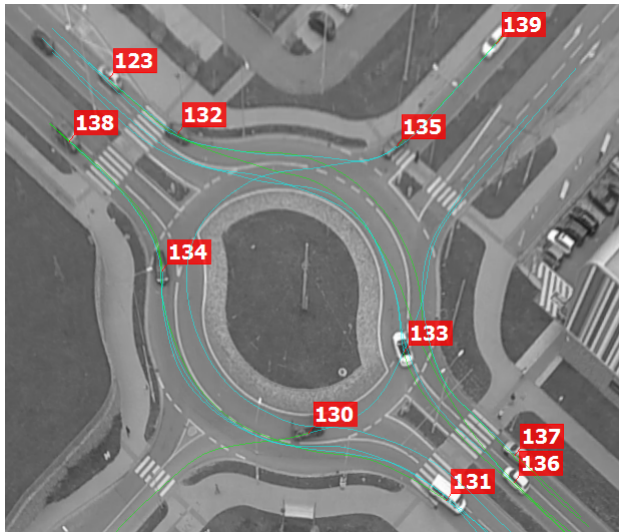

(a) Netroufalky

(b) Olomouc

Figure 5. Illustration of extracted vehicle trajectories in evaluated traffic scenes. The images are converted to grayscale for better readability of overlay graphics: blue-green curves represent vehicle trajectories and red labels with white numbers represent unique identificators of tracked objects.

(a) Geo-registered input image

(b) Overlaid detections: dark blue rectangles represent weak detections, light blue rectangles represent strong detections.
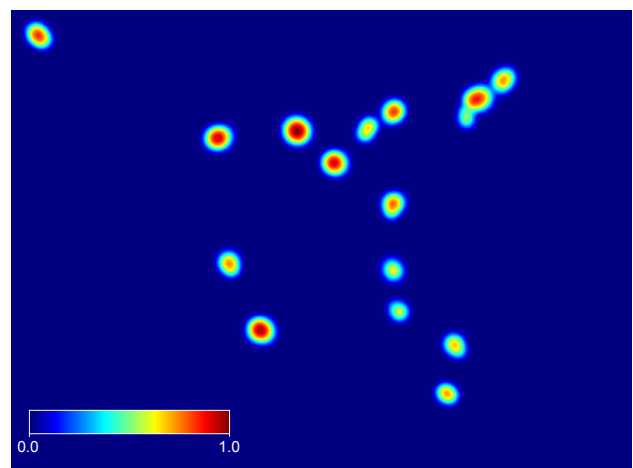
(c) Colourised map representing normalised values of attraction factor expressed over whole area of the scene, according to Equation 4.

Figure 6. Illustration of selected steps of single image analysis.