Indian Journal of Science and Technology, Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/105475, December 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Vehicle Detection and Classification in Aerial Images

Dmitry Sincha, Mikhail Chervonenkis and Pavel Skribtsov*

PAWLIN Technologies Limited, Dubna, Russia; dps@pawlin.ru, mach@pawlin.ru, pvs@pawlin.ru

Abstract

Background Nowadays the problem of vehicle detection and classification on aerial images received from UAV (Unmanned Aerial Vehicles) has become important because of development of UAV technology and image analysis methods. **Methods** The paper describes a multilevel detection of vehicles and their subsequent classification. This method can be used for search of moving and stationary vehicles. In this work we propose new approaches such as: image segmentation into superpixels by SEEDS method, trainable five-level cascade detector of combined superpixels-regions, which uses technology of artificial neutral networks. Characteristics of regions are built based on their geometric and texture features (HoG and LPB descriptors) and directly from the image patch using technology of nonlinear autoencoders. Additional cascade of the detector uses data of moving objects in the image. Similar responses of the detector are combined and classified by color and type of the vehicle. **Findings** For training these algorithms a largest image dataset was compiled from different sources. The results of tests of detection and classification showed high accuracy. **Improvements** Algorithm is fast enough to allow on-board usage. The proposed method can be applied for road traffic monitoring, analysis of parking lots occupancy and other similar tasks.

Keywords: Drone, Image Dataset, Object Classification,Road Traffic Monitoring, UAV, Unmanned Aerial Vehicles Vehicle Classification,Vehicle Detection

1. Introduction

Great attention is given to design of algorithms and programs for vehicle (car) detection with the help of UAV due to development of UAV technology and methods of video analytics. The works[1-6] and number of others describe studies on vehicle detection with sufficiently acceptable accuracy. Use of UAV for automatic monitoring of road traffic can provide sufficient increase of traffic capacity, optimization of actions on the removal of traffic accidents consequences, loss reduction related to accidents. UAVs have onboard video surveillance system.

The work[1] has suggested using the image segmentation into superpixels in challenges of vehicle monitoring for selection of moving vehicles only. Segmentation by the superpixels technique helps them to compression of received frames for their fast sending over the wireless lines. The works[2-4] present image segmentation into superpixels with their following integration into regions. Vehicle detection in these works is conducted on the base of analysis of region features; such approach allows using information about shape of the region for detection of the object.

Evaluation of motion parameters of object of interest (during traffic monitoring of vehicles) will allow automatic (without operator) selection of road segments with lower traffic capacity or traffic accidents what happened[3].

Systems implemented on board of UAV are based on use of computer vision, methods of processing and analysis of the images. Different methods of image processing are presented in[6,7], providing objects of interest detection with aid of templates, motion cues, etc. Issues of viewing conditions enlargement are considered in[8]. On the base of similar technologies a large number of studies are conducted, particularly on selection and vehicle detection on the roads[9-11].

Number of works suggest building of object features on the base of image texture, first of all HoG-descriptors (Histogram of Oriented Gradients)[4,12]. For example good results were shown by approach based on HoG-descriptors, suggested in[13,14] for solution of the pedestrian detection task. Such approach also has been used in this work.

*Author for correspondence

This work is a further development of the work[4]. The work introduces new approaches to development of training set, selection of features for building of detector cascade, combination of approximate responses and classification of vehicles.

2. Concept Headings

The purpose of the article is a development of methods (algorithm) of vehicle detection and classification on the base of analysis of video data stream obtained from UAV. Algorithm is designed for implementation on board of UAV, but also can be applied on land based server. Detection and classification of vehicles is applicable for tasks of road traffic monitoring and analysis of parking lots occupancy.

2.1 Algorithm Scheme

Functional flowchart of developed vehicle detector is presented in Fig1. Description of the algorithm is given below.

1. Image pre-processing and ROI selection.

The algorithm receives input flow of shots and performs pre-processing (cutting, scaling, contrasting - MSR modified for 3-channal image), conducts simple but rapid image segmentation by unconventional algorithm and assembles the region of road (ROI) out of segments by hierarchical algorithm. Optionally ROI is formed on the base of navigation information with the help of on-board GIS (Geographic Information System).

2. Selection of regions in the image.

Regions in the image must be selected in the way that vast majority of vehicles is presented at least by one region, which is sufficiently close to the image of the corresponding vehicle. As a proximity criterion $Sim(A,B)$ of two areas A and B in the image was selected the next:

$$Sim(A,B) = \frac{S(A \cap B)}{\min(S(A), S(B))} \qquad (1),$$

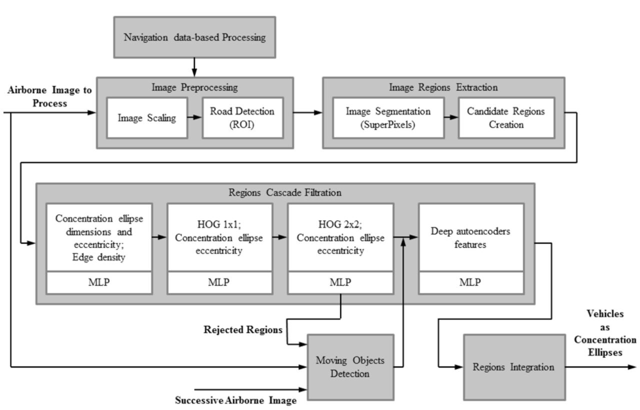

Figure 1. Functional flowchart of the vehicle detector.

Where: $S(A)$ = square of the area in pixels. Selected regions may intercross against each other.

Regions are built based on image segmentation into superpixels. Outcome of image segmentation process is a number of connected non-crossing areas – segments:

$$X = \bigcup_i S_i,$$

Where: X = set of pixels of the fragmented image, S_i = segment, i = integer index and $S_i \cap S_j = \emptyset$ at $i \neq j$.

Segmentation is performed so that individual segment would be sufficiently uniform while different, but neighboring segments would differ in color and brightness characteristics. If all segments are limited in size they are called superpixels.

In the capacity of regions individual superpixels and associations of two or three neighboring superpixels are selected, satisfying simple restrictions on square of region. Each selected region corresponds to hypothesis that this region determines a vehicle image. The works[2-4] presented different methods of segmentation for vehicle tracking: FHS (Felzenszwalb-Huttenlocher Segmentation)[15], SEEDS (Superpixels Extracted via Energy-Driven Sampling)[16], SEEDS Revised [17,18], SLIC (Simple Linear Iterative Clustering)[19], Model Based Clustering (MBC)[20], Quick Shift[21]. The works[2-4] come to conclusion that only FHS и SEEDS methods provide required quality and performance, for this reason only FHS и SEEDS methods were studied in this work.

3. Vehicle detection

Five-level cascade detector calculates features of regions and filters them by characteristics. All levels of this detector are tuned with the use of massive training set with vehicle marking on images. The features are: size, eccentricity, HoG, HoG for 16 cells, LPB, outputs of non-linear autoencoder. Regions are integrated past filtration if they relate to the same vehicle and vehicle descriptor is formed.

4. Detected vehicles are classified by color and type.
5. There is a subsystem of training sets forming and training of classifiers with the use of them.
6. Moving vehicles detection

Adjacent (with predefined step) images are compared and transformation is calculated which connects the images. One image is brought to another one by this transformation. Then difference of images is determined. Most bright points (significant distinctions) are found thereon. For these points velocity vectors are found along the optical flow for initial images and homographic correction is made for transformation. Similar in position and velocity points are put together into response and the vehicle descriptor is formed for them without type and color.

7. The rule of recount of pixel coordinates, velocities and size into global is formed (taking into account performed transformations and information from the navigation system). Data in the vehicle descriptor is recorded in two systems of coordinates – local and global.

8. Vehicle tracking and refinement of their features

Tracking filters are built (based on of Kalman filters) for the newly found vehicle descriptors. These filters live several shots and make predictions of the features (velocity, coordinates, size, etc.) of their vehicle. The filters integrate descriptors related to predictions. The filter is removed if prediction is not confirmed for a long time. The features are specified if prediction is confirmed. Using filters that have found confirmation of prediction, output vehicle descriptors set is formed (in global coordinates).

9. Possible application of obtained descriptors is detection of the road situations in traffic monitoring.

Velocity histogram is formed upon which road situation (RS) is recognized – traffic jam, norm, empty, speeding in each direction. RS hazard assessment is performed. Additionally presence of proper vehicle is recorded with a vehicle indication list (type + color + direction). On the base of current task for UAV decision on warning of operator is made.

3. Results

3.1 Image Dataset

A large-scale image dataset was prepared for training. Training dataset of the marked images and video records of road traffic, received from UAV was compiled from different sources:

1) Several video records, obtained with the help of UAV in Russia (Izhevsk City and its south and east outskirts, Agryz - Mozhga route, Izhevsk - Agryz route).
2) Video record, Bandung City, West Java, Indonesia[25].
3) Marked dataset of vehicle images – KIT AIS Data Set. Cities of Germany [26].
4) Marked dataset of vehicle images – VEDAI (Vehicle Detection in Aerial Imagery). USA, Utah [27].
5) Dataset of vehicle images – DLR 3K. Munich, Germany. The author: Institute of Transportation Systems[28,29].

Aggregation of specified data from noted sources allowed creating of world largest dataset of aerial vehicle images. There are images obtained in different environments (urban lands, countryside, forests, winter, summer, etc.). All images were transformed into the same scale approximately 22 cm/pixel and resolution 864 x 468 pixels (or less). Each image with a size of more than 864 x 468 pixels was split into overlapped subimages of 864 x 468 size. More than 44000 shots and over 60000 marked vehicle images in these shots are presented in the dataset.

3.2 Image Segmentation

Two methods of image segmentation were studied – FHS [15] and SEEDS [16]. Quality of segmentation was evaluated on the base of these images. Parameters of segmentation and number of pixels in the region were adjusted. Expression (1) was selected as a criterion of marking and region fitting. It is found that SEEDS provides coverage of marking with quality close to FHS and it is enough to build regions as association of not more than 2 superpixels while FHS also requires regions consist of 3 superpixels. Since SEEDS is much faster than FHS, SEEDS method was selected for segmentation. Obtained regions were covered at preset threshold of 0.5; an example of marking is from .91 to .99 (depending on data set).

3.3 Cascade Detector

The cascade region detector is used to classify the regions corresponding to vehicles. In this paper, we research a modification of the cascade detector described in[4].

The detector comprises of the following stages:

1) Classification based on parameters of the size and shape of a region's concentration ellipse.
2) Classification based on the HoG-descriptor calculated in a rectangular area delineating a region. This is a common approach to the construction of objects detectors[13,14]. In this case, a more suitable delineating rectangle oriented at either 0 ° or 45 ° is chosen.
3) Classification based on HoG-descriptor and LPB-descriptor 4 x 4 calculated on 16 adjacent rectangles.

4) Classification based on LPB-descriptor 8 x 8 calculated on 64 adjacent rectangles with orientation axes aligned with the region's concentration ellipse.
5) Classification based on outputs of neurons of the hidden layer of nonlinear autoencoder. For training the autoencoder we used the KL (Kullback–Leibler) regularization[22]. Visual representations of hidden neurons are shown in Figure. 2. Sample of training fragments of images containing vehicles is shown in Figure. 3.

Each stage of the cascade uses MLP (Multilayer Perceptron) decision rule as a learning rule. Each stage is trained to ensure high recall (close to .99). Each cascade stage is trained so as to ensure high recall (close to .99).

An example of detected vehicles is shown in Figure 4.

Figure. 2. Visual representations of hidden neurons.

Figure 3. Sample of training fragments of images containing vehicles.

Figure 4. An example of detected vehicles.

3.4 Moving Vehicles Detection

Detection uses a comparison of adjacent frames. For adjacent frames (I_1, I_2) optimized transformation T is built, which allows aligning of images. Transformation T has a view of rigid transformation or homograph transformation.

Rigid transformation sufficiently describes the image change connected with optically stabilized camera move. Homographic transformation describes the image change in more general case on the assumption that area is plain. On this stage one-channel image is used for better calculating speed. Transformation T is found sufficiently quickly and exactly by the method based on forming of the optical flow. Let us denote $T(I)$ as a shot I under transformation T. Image of the view $J = abs(I_2 - T(I_1))$ (abs – per-pixel taking of absolute value) is analyzed. The majority of points without move have low intensity on J. The result of smoothing J and thresholding is a map J_1, in which pixels with nonzero intensity corresponds with points of proper motion of objects in the image.

There is a simple method becomes available for calculation of the points number in which movement takes place for any rectangular segment of map J_1. Local maximums J_1 define supporting points for evaluation of motion speed of the objects. Motion speed is evaluated through detected displacement of supporting points and also by optical flow method. Described above approach has a fast implementation.

3.5 Combination of Responses

In the process of detection several regions may correspond to one vehicle. Combination of responses is carried out with the use of the proximity criterion (1). At that in combined region each superpixel *SP* will have weight *W(SP)* which is calculated as a frequency of superpixel appearance.

3.6 Results of Vehicle Detection

The experiments on the marked dataset result to the following quality data of vehicle detection (without considering movement): Precision = 0.758; Recall = 0.85; F1 measure = 0.801. Confusion matrix is presented in tables 1 and 2.

3.7 Classification of Vehicles

Method detection of vehicles described above opens the way to their automatic classification. Classification may be carried out for different classes of vehicles depending on application task and data available for training. Classification by type and color was implemented in this work.

Table 1. Confusion matrix (moving vehicles detection is off)

	Negatives (marked as not a vehicles)	Positives (marked as a vehicles)
Negatives (Detected)	0.84	0.15
Positives (Detected)	0.16	0.85

Table 2. Confusion matrix (moving vehicles detection is on)

	Negatives (marked as not a vehicles)	Positives (marked as a vehicles)
Negatives (Detected)	0.8	0.05
Positives (Detected)	0.2	0.95

3.7.1 Classification by Type

Classification of vehicles by type can be quite a challenge because the images of different types of vehicles may be similar in aerial photography case. Classification was performed based on the data of size of recognized vehicle. At that three types of vehicles were recognized:

1) Heavy (heavy trucks and buses).
2) Light.
3) Middle.

More detailed classification is problematic for now because representatives of different types of vehicle in many cases appear difficult to distinguish (for example: track/bus, van/SUV/crossover). Features for detection were: sizes of axes (a, b) of the concentration ellipse built for regions with account of weights, eccentricity of the ellipse and values $\sqrt{ab}$, $\sqrt{a^2+b^2}$. Classification was provided by use of the cascade of linear classifiers those parameters were tuned in accordance with the training set. Mean accuracy of detection 0.71 was achieved under the test set.

3.7.2 Classification by Color

Classification by color of vehicles is also not a simple task. Taking into account small size of vehicle image (as 6 pixels in width) and variability of color rendering subject to conditions of shooting (lighting, etc.) detection of wide range of vehicle colors is problematic. For this reason limited range of colors was selected:

1) Dark: Black, dark-grey, dark-blue, dark-brown
2) Gray
3) Light
4) White
7) Red
8) Green
9) Yellow
10) Indigo/Blue
12) Violet

Classification was performed in the following way:

Initial image I is transformed from RGB color model into another color model HSV (Hue, Saturation, Value/ Brightness[23]). Obtained image J is analyzed for each region. Advantage of HSV over RGB is what color channel is explicitly presented in it. Color model LAB

(CIELAB [24]) was also studied, but the result appeared to be worse.

Calculation of the histogram was performed for each channel of the region $R = \bigcup_{k=0}^{n} SP_{i_k}$, where SP_{i_k} is a set of superpixels composing the region.

$$\mathrm{H}_{ij} = Z^{-1} \sum_{SP \subseteq R} W^2(SP) \sum_{p \in SP} h_i(J_i(p), j)$$

$$Z = \sum_{SP \subseteq R} W^2(SP)\, S(SP)$$

Where: I = channel number, j = bin number, H_{ij} = histogram value, SP = superpixel of region, $W(SP)$ = weight of superpixel, $S(SP)$ = square of superpixel, Z^{-1} = normalizing factor, p = pixel of superpixel, $J_i(p)$ = value of channel intensity I for pixel p, $h_i(\cdot, j)$ = characteristic function of intensity value belonging to bin j:

$$h_i(x, j) = \begin{cases} 1, r_{i,j} \le x < r_{i,j+1} \\ 0, otherwise \end{cases}$$

Where $r_{i,j}$ defines lower bound of bin j for channel i. For channel H (Hue) proportional division by 6 bins was applied. For channels S (Saturation), V (Value, Brightness) non-proportional division by three bins was applied. As before classification was provided by use of the cascade of linear classifiers. Additionally "Unknown" color class was introduced. Mean accuracy of detection 0.65 was achieved under the test set.

4. Discussion

The paper presents the following results.

1) The largest DB of vehicle images was created which allow training universal algorithm of vehicle detection and classification.
2) Distinction of the proposed approach to image segmentation from the work[4] is in application of automatic adjustment of parameters of segmentation. Parameters were found at which SEEDS method is applied with combination of not more than two segments. At the same time coverage of vehicles appeared not worse than for FHS with combination of not more than three segments, but segmentation speed has been increased essentially and the number of regions has been decreased a little.
3) The use of LBP-descriptors (with HoG-descriptors) and other improved detection scheme allowed an increase to classification accuracy.
5) Detected objects are classified by type.
6) Detected objects are classified by color.

5. Conclusion

Method of detection and classification for both moving and stationary vehicles in images, received from UAV, has been proposed.

The largest DB of the vehicle images was created. An approach has been developed which allows automatic selection of image segmentation parameters and sufficiently quick method SEEDS has been adopted for image segmentation. It became possible to limit regions describing vehicles by two superpixels. Combination of HoG and LBP descriptors was successfully applied as well as autoencoder for building of multilevel cascade detector. Classification of vehicles by color and size was realized.

Obtained results confirm the possibility of use of the proposed method during development of UAV on-board systems, designed for automatic monitoring of traffic situation or parking.

Further development of the proposed method:
1) Increase the number of types of cars.
2) Use new features, such as local autoencoder.
3) Use convolution network at the final stage of the detector.
4) Use new segmentation methods, including methods of segmentation optimized for work with videos (for example gSLIC algorithm[31]).
5) The method can be used for detection of other objects which are not vehicles, for example, water surface object (such as board) or animals.

6. Acknowledgments

The research was financed by the Ministry of Education and Science of the Russian Federation under the Grant Contract from 8 September, 2014, (grant number № 14.576.21.0051, contract unique identifier RFMEFI57614X0051), grant assigned for the research into "Development of intelligent algorithms of traffic situations detection and identification for the on-board systems of the unmanned aerial vehicles performing automatic traffic patrolling by means of GLONASS".

7. References

1. Meuel H, Reso M, Jachalsky J, Ostermann J. Superpixel-based segmentation of moving objects for low bitrate ROI

coding systems. 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). . 2013. 27-30: 395, 400. DOI: 10.1109/AVSS.2013.6636672.

2. Kim N, Chervonenkis M. Situation Control of Unmanned Aerial Vehicles for Road Traffic Monitoring. Canadian Center of Science and Education. Modern Applied Science. 2015; 9 (5): 295–303. DOI:10.5539/mas.v9n5p1.
3. Abramov K, Skribtsov P, Kazantsev P. Image Segmentation Method Selection for Vehicle Detection Using Unmanned Aerial Vehicle. Canadian Center of Science and Education. Modern Applied Science. 2015; 9 (5): 295–303. DOI:10.5539/mas.v9n5p295.
4. Sincha D, Chervonenkis M, Skribtsov P. Vehicle Detection in Aerial Traffic Monitoring. American Journal of Applied Sciences. 2016; 13 (1): 46–54. DOI:10.3844/ajassp.2016.46.54.
5. Forssyth D., Ponce J. Computer Vision: A Modern Approach. Prentice Hall, Pearson Education: 2003.
6. Jähne B. Digital Image Processing. 6th edn. Springer Berlin Heidelberg: 2005. 607 p. DOI: 10.1007/3-540-27563-0. http://link.springer.com/book/10.1007%2F3-540-27563-0
7. Kim N, Bodunkov N. Adaptive Surveillance Algorithms Based on the Situation Analysis. In: Favorskaya MN, Jain LC. (eds.) Computer Vision in Control Systems-2: Innovations in Practice. Springer International Publishing. 2015. p. 169–200. DOI: 10.1007/978-3-319-11430-9_7.
8. Yilmaz A, Javed O, Shah M. Object tracking: A survey. ACM Computer Survey. 2006 December; 38 (4): 260–7. DOI: 10.1145/1177352.1177355.
9. Feng Lin, Kai-Yew Lum, Ben M. Chen, Tong H. Lee. Development of a vision-based ground target detection and tracking system for a small unmanned helicopter. Science in Chine. Series F: Information Sciences, Springer: 2009; 41: 3071–7.
10. Lin Y, Yu O, Medioni G. Efficient detection and tracking of moving objects in geo-coordinates. Machine Vision and Applications. Springer: 2010. DOI: 10.1007/s00138-010-0264-
11. Kim N. Automated Decision Making in Road Traffic Monitoring by on-Board Unmanned Aerial Vehicle System. Indian Journal of Science and Technology. 2015 December; 8 (10).
12. Turmer S. Car detection in low frame-rate aerial imagery of dense urban areas [PhD thesis]., Munchen.
13. Dalal N, Triggs B. Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005. 2005. p. 886–93.
14. Dalal N.Finding People in Images and Videos[PhD thesis]. Institut National Polytechnique de Grenoble.
15. Felzenszwalb P, Huttenlocher D. Efficient Graph-Based Image Segmentation. International Journal of Computer Vision. 2004 September; 59 (2).
16. Van den Bergh M, Boix X, Roig G. SEEDS: Superpixels Extracted via Energy-Driven Sampling. European Conference on Computer Vision. 2012; 13–26.
17. Efficient High-Quality Superpixels: SEEDS Revised. 2016 September 30. Available from: http://davidstutz.de/efficient-high-quality-superpixels-seeds-revised
18. Implementation of the superpixel algorithm called SEEDS. 2016 September 30. Available from: https://github.com/davidstutz/seeds-revised
19. Achanta, R. et al. SLIC Superpixels Compared to State-of-the-art Superpixel Methods, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012 May; 34 (11):2274–82.
20. Zhong S, Ghosh J. A Unified Framework for Model-based Clustering. Journal of Machine Learning. 2003; 1001–37.
21. Vedaldi A, Soatto S. Quick shift and kernel methods for mode seeking. Proceedings of ECCV. 2008; 705–18.
22. Ng A. Sparse autoencoder. CS294A Lecture Notes, Stanford University, Stanford, CA, USA, Technical Report, 2011.
23. Agoston M. Computer Graphics and Geometric Modeling: Implementation and Algorithms. London, Springer: 2005; 300–306.
24. Schanda J. Colorimetry. Understanding the CIE System. Wiley Interscience. Wiley: 2007.
25. Aeroterrascan Indonesia Highway Monitoring. 2016 October 21. Available from: https://www.youtube.com/watch?v=0g5dAuhMYjs
26. KIT AIS Data Set. 2016 October 21. Available from: http://www.ipf.kit.edu/downloads_People_Tracking.php
27. Razakarivony S, Jurie F. Vehicle Detection in Aerial Imagery: A small target detection benchmark. Journal of Visual Communication and Image Representation. 2015 January; 34: 187–203.
28. Chen Q, Koltun V. Fast MRF optimization with application to depth reconstruction. CVPR. IEEE Conferenc. 2014; 3914–21.
29. Vehicle Detection in Aerial Images. 2016 October 21. Available from: http://www.dlr.de/caf/en/desktopdefault.aspx/tabid-5431/9230_read-42467
30. Ojala T, Pietikäinen M, Harwood D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions. Pattern Recognition. 1996; 29, 51–9.
31. Birkus R. Accelerated gSLIC for Superpixel Generation used in Object Segmentation. Proceedings of CESCG 2015: The 19th Central European Seminar on Computer Graphics. 2015 April.