Image Segmentation Method Selection for Vehicle Detection Using Unmanned Aerial Vehicle

Kirill Viktorovich Abramov[1], Pavel Vyacheclavovich Skribtsov[1] & Pavel Alexandrovich Kazantsev[1]

[1] PAWLIN Technologies Ltd, Dubna, Russia

Correspondence: Pavel Vyacheclavovich Skribtsov, PAWLIN Technologies Ltd, 141980, 13/11 st. Lenina, Dubna, Moscow Region, Russian Federation. E-mail: lvo@pawlin.ru

Abstract

This article discusses the possibility of applying the methods of allocating super pixels in the task for detecting moving and stationary vehicles in images obtained from the unmanned aerial vehicle (UAV) which flying over roads and parking lots. The paper will also consider the specificity of images obtained when shooting with the UAV, the specificity of the image processing, and formed the requirements for segmentation algorithm applicable to the task. Author of the article has developed the application required to measure the average image processing speed in the video stream and the application for evaluating vehicles partitioning quality. This application works with the test image, on which the location of the vehicle were determined by the human. A study was conducted of the several algorithms for image segmentation: LIC, Quick Shift, Felzenszwalb-Huttenlocher, and Model based clustering algorithm. The article presents data on the speed and accuracy of the evaluation of these algorithms in the task for UAV's images segmentation. In conclusion, author has chosen methods that suitable for their use in specific application task. For image segmentation, it was decided to use two of the most appropriate method: segmentation algorithm Felzenszwalb-Huttenlocher and developed by the author earlier algorithm based on the approach to clustering model based clustering. The article also discusses possible further ways of unification super pixels containing regions with vehicles. Further work will focus on the modification, parallelization and accelerate software implementation of FHS and MBC. The author will be also investigate the question of the possibility of Markov Chains to solving the task for super pixels association   to the regions and the question of the applicability of the binary classification of regions for the detection of vehicles.

Keywords: image segmentation, super pixel, UAV, vehicle detection, SLIC, quick shift, Felzenszwalb-Huttenlocher, model based clustering

1. Introduction

Due to the rapid technological development of unmanned aerial vehicles (UAVs), the tasks for automating road and parking lots monitoring using UAVs, detecting or recognising vehicles on the images obtained by aerial photography and aerial video recording, are getting more urgent. Now, there are ready solutions successfully using UAVs in emergencies (E) when searching for people, warning emergencies, exploration of fire seats in forests, assist in situation analysis, and coordinating the actions of rescue services at floods. UAVs are also used in monitoring power stations, electricity grids, agricultural lands, land resources, railways, and roads, security, e.g. border and facilities security.

Development of image recognition algorithms able to work with large volumes of video data raises the sharp question of the way to handle the data. This is connected to the fact that analysis of millions of pixels in each image frame can take considerable time and lead to a sharp increase in demands on computing equipment used to solve such tasks. In order to reduce the amount of data supplied to the recognition algorithm input, the most common are known image segmentation algorithms. These methods allow selecting closed connected areas based on similarity under any parameter and continue to work with them as with a single object.

The following discussion focuses on the segmentation algorithm choice applicable to specific practical tasks for recognizing vehicles in image data obtained by UAVs.

1.1 Statement of the Recognition Task

Flying over certain sections of the road or parking lots, UAVs will shoot and store the received video data. In

addition, UAVs have the ability to determine own position on the ground using such satellite navigation systems as GLONASS and GPS. On-board vision system (OVS) of a UAV is to detect different vehicles on the obtained images, identify their speed, direction, and identify the road traffic situation. Road situations mean the following: traffic jams, difficulty of movement in a particular road section, car accidents, vehicles violating the established speed limits, and improperly parked cars.

UAV is shooting from a height from one hundred to five hundred meters. A camera mounted on the aircraft, is usually directed vertically downward, and the images deliver only the upper part of the vehicle. If a frame is 640 pixels wide and 480 pixels high, the vehicle will have an area of about from 50 to 700 pixels. An example of an image obtained from a UAV camera is shown in Figure 2. Thus, the vehicle will be represented by a small number of homogeneous regions formed by the roofs of cars and trailers, bonnets, trunks, windshields, and rear glasses, etc.

This task is specific with the fact that the input images often do not have enough good quality, since UAVs often contain quite cheap equipment, which gives the advantage of easy replacement of the camera after a "hard landing" (Sechin et al., 2011). Based on this, it can be concluded that known blurring algorithms required to implement the most popular image segmentation algorithms, can be applied to the obtained images.
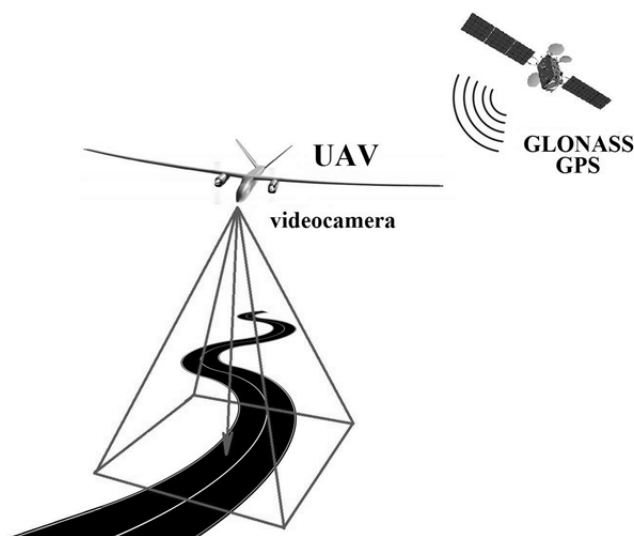

Figure 1. Schematic representation of the process of shooting sections of roads by UAVs

Figure 2. Example of an image from a UAV in highway monitor mode

1.2 Image Processing

Flying around certain parts of the roads, the UAV can travel long distances, making it difficult to transmit the video signal to the remote local devices, which could carry out data processing. Therefore, there is a need for processing at the aircraft. This is advantageous because time is not wasted on sending images and receiving the processed data. The aircraft with on-board software can quickly respond to road situations and can coordinate its path or trajectory. In this case, the input data processing shall be performed at a rate comparable to the rate of the video stream. The disadvantage of this solution can be considered that as a rule it is impossible to install on UAV's board the equipment with the computing power comparable with the possibilities of stationary solutions.

Based on the above-said, it can be concluded that a segmentation algorithm shall face sufficiently stringent requirements on speed, which will allow the vehicle speed detection comparable to the speed of the resulting video stream and provide a UAV's response to the existing road situation in real time.

It shall be noted that if the image segmentation methods are applied for the present task is also driven by a desire to reduce the time for single frame processing, because segmentation algorithms can simplify the task of recognition, reducing the amount of data to be processed. The need to develop an original recognition algorithm for vehicles is because there is a need for a reliable detection of not only moving vehicles, but stationary as well. Thus, the known algorithms for detecting moving objects (the inter-frame difference method, the background subtraction method, the Gaussian distributions method, the moving mean method, the optical stream determination method, and others (Sjuj Ljej & Gavrilov, 2011)) are not applicable in this case.

1.3 Segmentation and Super Pixels

Now let us consider the notion of image segmentation and formulate a number of requirements for such an algorithm for solving the recognition task. Segmentation is the process of image partitioning into several segments, combining sets of pixels, also called super pixels. Because of segmentation, all the processed images are divided into a number of connected non-intersecting regions - segments:

$$X = \bigcup_i S_i,$$

Where X is a set of segmented image pixels, S_i is a segment, $i -$ is an integer index and $S_i \cap S_j = \emptyset$, at $i \neq j$. The purpose of segmentation is to simplify and/or change the representation of the image to make it simpler and easier to analyse. All pixels in a segment are similar in some characteristic or calculated property, such as colour, brightness, or texture. The adjacent segments differ significantly in this characteristic. The result of image segmentation can be many contours selected on the image (Linda et al., 2011).

Based on the specifics of the application task to the segmentation algorithm, applicable in this case, one shall require the following from it:

- Super pixels shall have no common points and shall cover the entire image.

- Connectivity of the areas (in the spatial sense) of separate super pixels. Super pixels borders shall be close to the borders of recognised objects or pass inside of these objects.

- The high speed of the algorithm (or at least, the possibility of a significant acceleration of the algorithm). This requirement is necessary for the on-board image processing.

- Simplicity of forming an image for an object as a super pixels association. Detected vehicles shall be divided into a small number of super pixels. This will enable faster and more accurate solution of the task for searching vehicles in the image.

1.4 Literature Review

In their article, H. Meuel and M. Reso (Meuel et al., 2013) report on the application of the super pixels allocation methods at detecting vehicles. The main difference of the task considered in this paper from the above-mentioned paper is that the authors aim to allocate only moving vehicles, and that the super pixels segmentation method helps to organize compression of the resulting images for quick sending via the communication channels. Our task involves processing all the information by means of the equipment located on UAVs and detecting vehicles, including stationary. An important observation made in the mentioned article is that the images obtained from the UAVs, if applying image segmentation by super pixels, allow achieving an acceptable quality of vehicles edges selection and possibility of vehicles division into a small number of super pixels. Based on this experience, and it was decided to apply the algorithm for segmentation by super pixels. Our task implies that image partitioning into large segments, including the desired vehicles, is intended to speed up

and simplify the detection process.

Now, there are numerous super pixels allocation algorithms. Peer Neubert and Peter Protzel (Neubert and Protzel, 2007) tell in their article about the most popular modern open source algorithms. The article discusses various metrics for assessing the segmentation quality, based on comparison with the standard segmentation at the test images. In addition, this material provides comparison of different characteristics of known segmentation algorithms. The data on the images partitioning by different algorithms, their speed, and resistance to shearing, scaling, and rotation is provided.

2. Consideration of Various Algorithms for Super Pixels Allocation

Based on the data obtained in the paper by Peer Neubert and Peter Protzel (Neubert & Protzel, 2007), namely the quality of the partitioning (Boundary Recall, Under segmentation Error) and data on the speed of the different algorithms, the following algorithms were selected for further consideration: Quick Shift, oriSLIC (hereinafter SLIC) and Felzenszwalb-Huttenlocher Segmentation (hereinafter FHS). The article also discusses the results of the segmentation method based on model based clustering (Zhong & Ghosh, 2003). This segmentation method has already been implemented and applied by the author in other applications. It has shown good results of segmentation (Abramov & Dolgopolov, 2014).

It shall be noted that all the above algorithms involve image pre-processing. Various tasks involve different methods of pre-treatment, such as images contrasting, resolution reduction, and Gaussian blur is used for all methods.

Next, a brief look will be made at the principles of publicly available algorithms.

The Quick Shift algorithm performs image segmentation as clustering of points in the 5-dimensional space: 3 components of the pixel colour in the L_{uv} space and 2 coordinates of the pixel in the image. Clustering is based on an assessment of the points density: where the density is higher, there are clusters centres, and points are "draining" into the clusters centres on way to increase the density.

The Felzenszwalb-Huttenlocher algorithm uses a graph with weights for image segmentation. The edges of the graph represent the joining of two adjacent pixels in the image, and the weight of this edge is the dissimilarity measure of the two pixels united by this edge. Dissimilarity of the two pixels may be determined in different ways, but usually the distance is calculated from the three image channels. Having set a similar graph for the input image, the algorithm sorts all the edges according to the weight parameter. The algorithm then takes a sorted list of edges between a pair of combining the most similar pixels into one super pixel represented by a tree structure, where each pixel has a pointer to the parent pixel, with which it is combined. Association may not happen in the case if the pixels dissimilarity exceeds a predetermined threshold. Thus, all the pixels are nodes of a set of non-intersecting trees.

The SLIC algorithm segmentation is based on K-means data clustering. The desired super pixel size shall be set for its work. Further, the entire image is divided into a grid of squares of a given size where the K-means algorithm starts the clustering process in order to minimize the deviation from the pixels centres of the forming clusters. Initially, the clusters centres are selected in the centres of four adjacent grid squares.

3. Methods

In order to compare the segmentation results by different algorithms, a computer application was developed capable of running the segmentation process in turn by four different methods of allocating super pixels, measure the operating time of each algorithm and visualizes the received super pixels. The application input received a series of frames from the test video stream. Each frame in the video stream was 204 pixels high and 357 pixels wide. Such resolution ensured maximum speed of the test method, in which a stable allocation of vehicle is possible. Measurements of the operation time were carried out on a computer with average technical specifications: Intel Core i5 2.4 GHz and 8 GB RAM.

In order to assess the accuracy of the vehicles partitioning, an application was developed to carry out the layout of frames from the video stream. The layout was performed by the application user, who focused thee rotated rectangle on the image according to the vehicle position. With this layout, the application may determine whether a particular super pixel intersects a part of the image representing a vehicle, and how many pixels of the super pixel intersect the area marked as the vehicle and how many of them do not intersect.

4. Results

Figures 3 and 4 show the results of segmentation for SLIC (Achanta et al., 2012), Quick Shift (Vedaldi & Soatto, 2008), Felzenszwalb-Huttenlocher (Felzenszwalb & Huttenlocher, 2004) algorithm and a model based clustering

- MBC - method. All algorithms, except for the last, are publicly available.

The requirement of super pixels proximity to the vehicles boundaries in the images is observed by all the compared algorithms. The conclusion about this can be done not only based on the observations, but also at a low enough error level of insufficient segmentation calculated from the test image from the segmentation standards base Berkeley Segmentation Data Set 500 (Note 1) in the article by Peer Neubert and Peter Protzel (Neubert & Protzel, 2007). It shall be noted that the SLIC algorithm is often unable to identify road marking, as its work requires a desired size of super pixels, and in case of setting an adequate parameter value for identifying road marking, the algorithm divides the vehicles on a large number of small super pixels, making them difficult to find. Figures 3 and 4 show examples of image processing obtained from the UAVs (see. Figure 2) using a series of algorithms for image separation into super pixels.
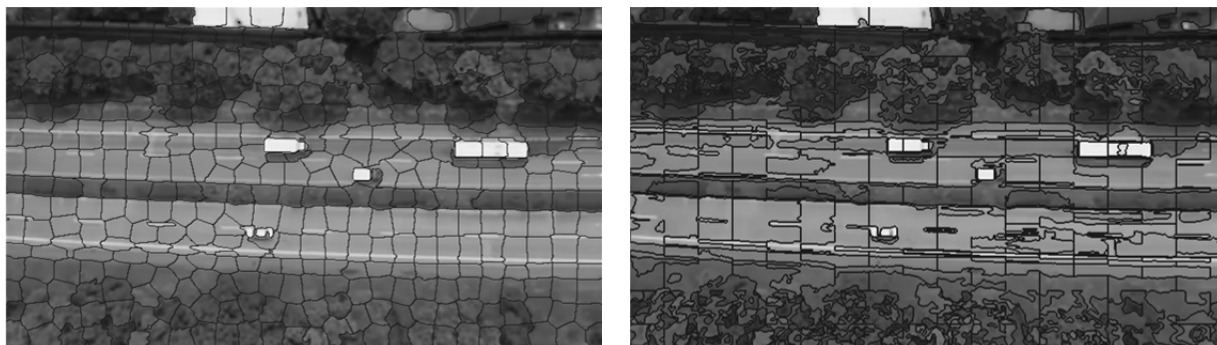

Figure 3. Visualization of the results of SLIC (left) and Quick Shift (right) algorithm
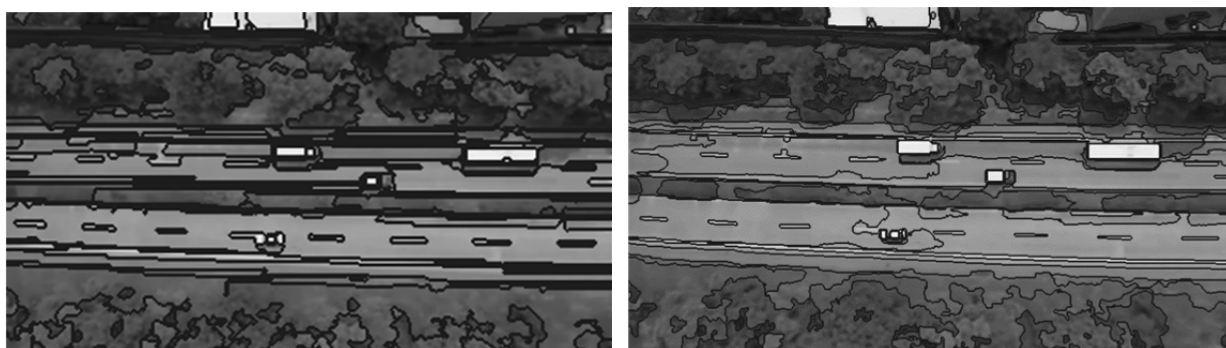

Figure 4. Visualization of the results for Felzenszwalb-Huttenlocher (left) and model based clustering (right) algorithm

In order to determine the operating time, the algorithms process data from the same video stream captured by UAV while passing over a motor road. The average processing time for one frame is the value used for comparison. Speed data are given in Table 1. The table shows that Quick Shift and Model based clustering (MBC) algorithm are not suitable for image segmentation in real time at a speed close to the video stream speed. Figure 5 shows graphs of algorithm operation time (per image pixel), depending on the frame size in the video stream. However, the ordinate axis contains the ratios of milliseconds spent by an algorithm on segmentation to the number of pixels in the image. Looking at the data, it can be said that while the MBC algorithm practically linearly depends on the image size (processing time per pixel is almost independent of the size of the image). The SLIC and FHS algorithm have time growth slightly higher than the linear algorithm, but for the FHS method, it cannot be considered as a disadvantage, since it is generally much faster than other solutions.

Table 1. Data on the algorithm speed (the average processing speed of the video stream test frame)

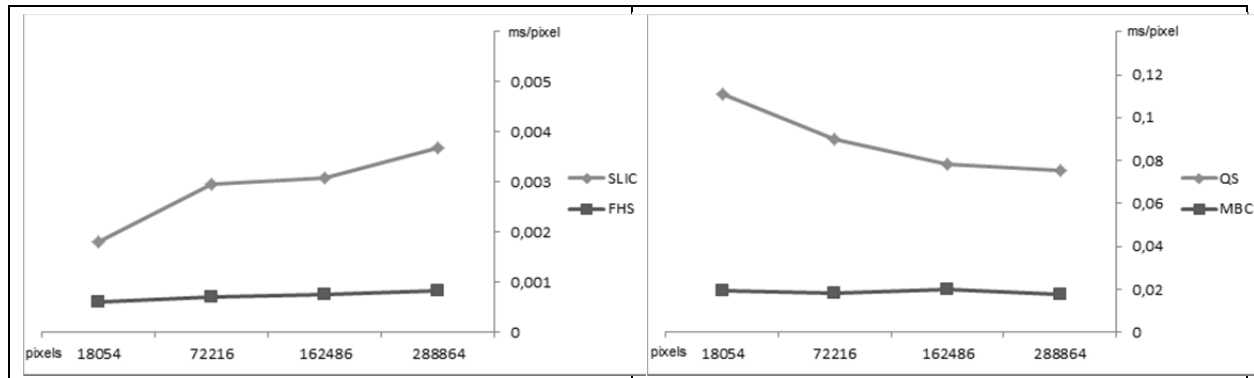| Method name	Average operating time (ms)
FHS	51.35
SLIC	214.17
MBC	1321.45
Quick Shift	6510.54

Figure 5. Graph of algorithm operation time (per image pixel) depending on the image size (SLIC, FHS in the left and QS, MBC in the right)

The Quick Shift algorithm cannot be applied because of the low speed and the lack of apparent capacity to accelerate significantly the algorithm, as the Quick Shift method does not guarantee connectivity of the received segments. For these reasons, Quick Shift is not recommended for vehicle search tasks and is not considered further in the article.

To compare the quality of segmentation, it is proposed to analyse the following characteristics of true image segmentation for the object (ground trust) G:

1) N, the total number of super pixels (segments) crossing G;
2) S_G, area G (number of pixels);
3) S_1 set of superpixels having less than a half of the area falling into G;
4) N_1, number of superpixels included in the set S_1;
5) S_2, set of superpixels having a half and more of the area falling into G; N_2, number of superpixels included in the set S_2.
6) $Recall$, the ratio of the total number of pixels (area) G, included into S_1

$$Recall = 1 - \frac{\sum_{i=0}^{N_1} r_i}{S_G}$$

Where r_i is the number of pixels in the i-th superpixel belonging to the set S_1

7) $Precision$, the ratio of the total number of pixels (area) G, included into the set S_2 to G area

$$Precision = 1 - \frac{\sum_{i=0}^{N_2} p_i}{S_G}$$

Where p_i is the number of pixels in the i-th superpixel belonging to the set S_2.

Experiments were conducted with a number of images obtained from the UAVs. These images had manually allocated vehicles regions (in the form of rotated rectangles). Images were processed using three different methods (algorithms) - FHC, MBC, and SLIC. Average values of the characteristics are shown in Table 2.

Table 2. Average values of segmentation methods characteristics based on the experimental data

Method	N	N2	Recall	Precision
FHS	5.51	2.18	0.792	0.864
MBC	4.70	1.75	0.797	0.917
SLIC	5.21	0.54	0.341	0.911

The SLIC algorithm cannot be used to solve the task of recognition (detection) of vehicles. as it has too low recall. Approximately, in 48% of cases, N_2 equals to 0 (a super pixel is in G for less than a half). Another SLIC disadvantage is that it does not identify road marking.

The MBC algorithm (method) has the best recall and precision (especially). The method also identifies road marking well. However, this method leads to an excessive number of allocated super pixels on one object. In the future, it is planned to finalize the MBC implementation, so that the number of super pixels on the object is reduced, and the speed is increased (due to optimization and parallelization). If the resulting MBC modification maintains the quality of the segmentation, the MBC method can be used to solve this task.

Among the examined methods, the FHS algorithm has the highest average speed of image processing in the video stream, as well as good accuracy. This algorithm meets all other requirements imposed in section 2, so the task for detecting vehicles shall be probably solved by the FHS algorithm fully satisfying the image segmentation speed and quality speed.

5. Discussion

Among the examined methods, the FHS algorithm has the highest average speed of image processing in the video stream, as well as good accuracy. This algorithm meets all other requirements imposed in section 2, so the task for detecting vehicles shall be probably solve using the FHS algorithm fully satisfying the image segmentation speed and quality speed.

The possibility of applying the MBC algorithm shall be considered separately. In a specific implementation, the model based clustering algorithm, involved in the above-mentioned testing performance, showed quite bad timing. However, this algorithm is fully parallelizable and its parallel implementation, which has the possibility of GPU computing, can have ten times better performance. In this connection, it is not necessary to discard immediately the possibility of using the MBC method to solve the task. The following discussion focuses on the advantages of the MBC method.

The algorithm is based on the principle of data model based clustering. Models refer to data structures describing the laws according to which different objects adjacent to each other can be combined into one cluster, and decide that the objects refer to different clusters. The algorithm is iterative. At the initial iteration, each cluster is represented by one pixel of the image (for the image segmentation tasks). Further, there is the process of combining the clusters located as close to each other as possible. The feature of the algorithm is that the function assessing clusters proximity measure can be arbitrary, and is defined for a particular implementation. This gives a possibility to create more "intelligent" segmentation methods. Since the clusters (super pixels for the image segmentation task) similarity evaluation function can estimate the geometric properties of the cluster and combine objects of similar shape. The author has developed a software implementation allowing allocating portions of the textured surface areas at the image with normal pixel distribution and areas with linear dependence of the pixels intensity in the image depending on their coordinates at the image. Computationally simple function is developed and can be applied for calculation of clusters proximity measure, based on an assessment of the central intensity and spread of the pixels intensities within the same cluster.

6. Combining Super Pixels in Regions

When using the most appropriate methods for decomposition at super pixels of image obtained from the UAV with a suitable choice of parameters, as a rule, it is possible to decompose the image of a single vehicle to a small number of super pixels. This is explained by the fact that the mapping of the vehicle is small enough at such images. In most cases, the number of super pixels representing the vehicle does not exceed three for the FHS method. When using the MBC method, increasing the number of iterations, the neighbouring super pixels (clusters) can be combined into one if they have similar characteristics.

In the detection of the vehicle using the methods of super pixels isolation, it is desirable to bring together those super pixels, which presumably correspond, to the vehicle in one region.

The simplest approach for super pixels combining is bringing all the small super pixels (in which the number of

pixels is less than a certain threshold) together with neighbouring super pixels with the most similar characteristics. In particular, this approach is included the implementation of the FHS method. When combining super pixels (regardless of the method for obtaining them, including the MBC method), it is proposed to use one or more iterations of the MBC method. If the clusters (super pixels) are partitioned by the MBC method, then computationally simple clusters proximity measure function is used (e.g., based on the calculation of the central cluster intensity and measuring the spatial clusters proximity), then computationally much more complex function can be used for the final uniting of the super pixels. This is due to the fact that in the initial phases of the MBC method, the number of clusters is equal to or commensurate with the number super pixels in the image (usually, hundreds of thousands) and the number of iterations is measured in tens, and in the case of the final merging, the number of super pixels is measured in hundreds, and the number of the required iterations - in units. Consequently, the use of more complex clusters proximity measure functions is permissible when implemented on the UAV board. More advanced clusters proximity measure functions may take into account the colour and texture characteristics, such as colour and gradient correlation or Hu covariance matrix invariants (Hu, 1962), pixels intensities histograms and gradient orientation (HOG) (Dalal & Triggs, 2006), as well as other characteristics (Tuytelaars and Mikolajczyk, 2008).

Eventually, after images processing, the set of regions shall be obtained, part of which is quite close to the image of the vehicle. AT that, the vast majority of the vehicles in this set shall have the corresponding regions. In this case, the task for vehicles detecting is reduced to a binary classification of regions, and can be solved by the methods of retrieving a set of characteristics of the region and the automatic learning using examples (Leitloff et al., 2011). In particular, it is possible to use the Deep Learning method (Deng, Yu., 2006, Szegedy et al., 2013, Goh et al., 2011).

7. Conclusion

The proposed methods of automatic images partitioning into super pixels can effectively reduce the task for finding and detecting vehicles in images to the binary classification task. This approach will allow realizing vehicles detection on the images or in the video stream received by UAVs on board of the aircraft. Among the considered methods for images partition on super pixels, the two methods were allocated: FHS, MBC. They are suitable for the task, and in the current implementation only the FHS method is suitable for primary super pixels isolation. The MBC algorithm can be used both as a means of allocating super pixels (provided its parallel implementation), and as a method of combining already found super pixels constituting one vehicle. Further papers on the topic of the article are supposed to be written on modification, parallelization, and acceleration of FHS and MBC software implementation; on research of the possibility of using the Markov Chains for solving the task of uniting super pixels into regions and researching the methods for binary classification of regions for vehicles detection. The proposed approach can be extended to the task of detecting other objects in the image.

Acknowledgments

References

Abramov, K. V., & Dolgopolov, A. V. (2014). Algoritm klasterizacii na osnove modelej, ego primenenie v zadachah segmentacii i analiz vozmozhnosti parallel'noj realizacii, XII Vserossijskaja nauchnaja konferencija "Nejrokomp'jutery i ih primenenie". Tezisy dokladov, otdel operativnoj poligrafii MGPPU, Moskva, 89.

Achanta, Radhakrishna et al. (2012). SLIC Super pixels Compared to State-of-the-art Super pixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(11), 2274 - 2282.

Dalal, N., & Triggs, B. (2006). Object Detection using Histograms of Oriented Gradients. In: Pascal VOC Workshop, ECCV.

Deng, L., & Dong, Y. (2014). Deep learning: Methods and applications. Now Publishers Inc.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient Graph-Based Image Segmentation. *Int. J. Comput. 59*(2), 167-181. http://dx.doi.org/10.1023/B:VISI.0000022288.19776.77

Goh, H. et al. (2011). Learning invariant color features with sparse topographic restricted Boltzmann machines.

Image Processing (ICIP), 2011 18th IEEE International Conference on. IEEE.

Hu, M. K. (1962). Visual Pattern Recognition by Moment Invariants. IRE Trans. Info. *Theory, IT-8*, 179–187,

Meuel, H., Reso, M., Jachalsky, J., & Ostermann, J. (2013). Super pixel-based segmentation of moving objects for low bitrate ROI coding systems. Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference, 395, 400, 27-30. http://dx.doi.org/10.1109/AVSS.2013.6636672

Neubert, P., & Protzel, P. (2007). Super pixels benchmark and comparison, Forum Bildverarbeitung, Scientific Publishing, Karlsruhe, 358.

Sechin, A., Drakin M., & Kiseleva, A. (2011). Bespilotnye letatel'nye apparaty: Primenenie v celjah ajerofotos#emki dlja kartografirovanija (chast' 2), Racurs, Moscow. Retrieved from http://www.racurs.ru/?page=699

Szegedy, C., Alexander, T., & Dumitru E. (2013). Deep Neural Networks for Object Detection. Advances in Neural Information Processing Systems.

Tinne, T., & Krystian, M. (2008). Local invariant feature detectors: A survey. Found. Trends. *Comput. Graph. 3*(3), 177-280. http://dx.doi.org/10.1561/0600000017

Türmer, S., Leitloff, J., Reinartz, P., & Stilla, U. (2011). Evaluation of selected features for car detection in aerial images. ISPRS Hannover Workshop, 14-17. Hannover.

User Aero Terrascan. Aeroterrascan Indonesia Highway Monitoring. Retrieved from www.youtube.com/watch?v=0g5dAuhMYjs

Van Rijsbergen, C. J. (1979). *Information Retrieval* (2nd ed.). Butterworth.

Vedaldi, A., & Soatto, S. (2008). Quick shift and kernel methods for mode seeking. In Proc. ECCV.

Zhong, S., & Ghosh, J. (2003). A Unified Framework for Model-based Clustering. *Journal of Machine Learning,* 1001-1037.

Note

Note 1. www.eecs.berkeley.edu/Research/Projects/ CS/vision/grouping/resources.html