Global Image Analysis: Detection and Recognition of Basic Informative Elements of Road Scenes

Konstantin Kiy
Keldysh Institute of Applied Mathematics of Russian Academy of Sciences
Moscow, Russia
konst.i.kiy@gmail.com

Roman Dosaev
Keldysh Institute of Applied Mathematics of Russian Academy of Sciences
Moscow, Russia
romandosaev@gmail.com

***Abstract*—In this paper, an application of tools for global image analysis, provided by the geometrized histogram method for detecting and recognizing basic elements of road scenes, is described. The application to finding permanent and temporary road markings, detecting stop lines and features showing the fact of approaching to crossings are addressed. The possibility of using the technique in detecting and recognizing traffic signs and traffic lights is discussed. The approach is compared with other known methods for solving the described tasks. The algorithms for solving the considered tasks and their program implementation are outlined. The results of operation of the developed software system for particular images and video sequences are presented. The problems of designing a software system for integrated understanding road scenes, involving analyzing the road, roadsides, and other objects on the road, the sky, road markings, and traffic signs, are discussed.**

Keywords—computer vision, image understanding, global image analysis, mobile robots

I. Introduction

In this paper, a new approach is proposed for detecting and recognizing basic informative elements of road scenes, such as road marking and traffic signs, signals of traffic lights, etc. This topic is very urgent, and the application of the obtained results can be important in designing Advanced Driver-Assistance Systems (ADAS), widely implemented by the world leading automobile manufactures. This topic is also important for designing control systems for driverless vehicles. Unfortunately, the existence of unsolved problems in this field is proved by accidents with driverless vehicles (frequently, fatal). The solution of these problems is especially topical for the countries with the road and climate conditions similar to those in the Russian Federation. One of the most recent surveys of papers and results obtained in the field of detecting and recognizing road marking can be found in [1]. It is worth noting that the overwhelming majority of papers are devoted to lane detection. This means that they are mainly interested in finding the boundaries of the lane in which the vehicle locates. This is sufficient for designing systems for warning the driver that the vehicle departs the lane (the so-called LDDWS – lane departure detection warning system). For autonomous driving, it is necessary to solve a more complex problem – to find also the markings of the adjacent lanes and to recognize the nearest solid line road markings. In addition, it is necessary to analyze road markings when the vehicle passes several lanes entering the road. In this case, the road marking looks completely different, and occlusions caused by other vehicles may be substantial. All these circumstances make wrong the simplifying geometric assumptions that are adopted in conventional lane detection and make it much easier. Note also that the motion on curvilinear parts of the road makes it impossible to use methods based on the analysis of straight segments in the frame.

The first publication of the authors on this topic can be found in [2]. The method proposed in [2] allows us to solve the problem of finding road marking in the general statement. In this paper, methods proposed in [1] are developed, and such problems as detection and recognition of solid white marking and temporary colored marking (painted from yellow to red-orange) are considered. We also propose a method for separating temporary road marking under the presence of the existing solid road marking that has not been removed. We were not able to find any publications on this topic. It seems to us that it is connected with problems in dealing with color images in real time [1].

The problems of detecting stop lines are also discussed. Algorithms for finding traffic signs (mainly connected with regulation of motion on crossings) are presented. The technique for finding traffic lights is also described. In Section 2, the geometrized histograms method for concise image description and segmentation is briefly described. Its tools for global image analysis are presented. In Section 3, these tools are applied to the problem of detecting and recognizing road marking. Algorithms for finding stop lines, traffic signs, and traffic lights are outlined in Section 4.

II. Geometrized Histograms method

In contrast to the main methods for image segmentation, the geometrized histograms method is designed so that the main processing of video data can be made in parallel. Using the technique described in [3–5], the structural graph of color bunches *STG* is attached to each color image. To construct *STG*, the image is divided into strips of the same width St_n, with sides parallel to the horizontal or vertical axis of the image plane *Os*. The notion of the image geometrized histogram was introduced in [3]. This notion is a far generalization of the ordinary image histogram employed in many papers [11]. Note that the ordinary image histogram is a weak invariant of the image, since it is invariant under any one-to-one transformation of the image rectangle. It does not take into account the geometry of image objects at all. At the same time the geometrized histogram is only invariant relative to transformations inside strips for which points move perpendicular to the axis *Os*. Since we deal with narrow strips, these transformations only slightly change the geometry of the objects belonging to the image.

The geometrized histogram describes rather exactly the value distribution of the function specifying the image in its rectangle. The geometrized histogram is obtained by projecting pixels of the strip on its lower side. For a grayscale image [3], because of a discrete nature of the image, the projection of the set L_z of the level z of the function, specifying image on the lower side of the strip is a union of intervals $Pr(L_z) = \cup_k I_{kz}$ on it. In order to compare level sets in different strips, we can adopt that for all strips the intervals lie in the axis *Os*. The union of all systems of intervals for all strips describe well the value distribution of the monochrome function specifying the image. This construction is generalized to the case of a vector function specifying a color image [3]. In the projection, we manage to separate the whole set of the strip into subsets for which saturation, hue, and intensity vary in some ranges. We obtain systems of intervals, each of which *Sg* is characterized by the following parameters:

- The position of the interval [beg_{Sg}, end_{Sg}] *Sg* in the axis *Os*.
- The range $\Delta_H^{Sg} = [H_{min}^{Sg}, H_{max}^{Sg}]$ and the mean value of the hue H_{mean}^{Sg}.
- The range $\Delta_S^{Sg} = [S_{min}^{Sg}, S_{max}^{Sg}]$ and the mean value of the saturation S_{mean}^{Sg}.
- The range $\Delta_I^{Sg} = [I_{min}^{Sg}, I_{max}^{Sg}]$ and the mean value of the intensity I_{mean}^{Sg}.
- The cardinality of the interval $Card^{Sg}$ (it is equal approximately to the number of points of the strip located in the strip over *Sg* and having the color characteristics within the boundaries specified above for it).

Denote by $dens(Sg) = Card^{Sg}/(\ end_{Sg} - end_{Sg} + 1\ )$ the density of *Sg*. It is the substantial property of the algorithm of obtaining the geometrized histogram in a strip that it can be obtained for one pass of the array of the image pixels in the strip. It makes it possible to obtain it in real time. The union of the geometrized histograms of all strips give the geometrized histogram of the image. Using the original clustering operations [3], intervals of the geometrized histogram are joined into color bunches *g,* which are characterized by the same intensity color parameters as for intervals *Sg*, generating this bunch, as well as by total cardinality and density. The color bunches are united in a graph. In each strip, the adjacent color bunches (with the adjacent localization intervals) are joined by an edge of this graph. While in the adjacent strips the bunches whose localization intervals intersect each other are joined by edges as well. Informally, each color bunch gives a description of a certain part of a real object in the strip, its projection on the axis *Os*, and a description of the values of the numerical color characteristics of this part of the object. This graph is called STructural Graph (*STG)*. It can be interpreted geometrically by superimposing localization intervals of its bunches *b* ([beg_b, end_b]) on the central line of the corresponding strip and coloring these intervals using H_{mean}^b, S_{mean}^b, I_{mean}^b. Examples of color bunches, superimposed on the grayscale components of color images can be seen in open success in [2, 6]. The presented images show well that *STG* describes adequatelyl the intensity-color characteristics of images and the geometry of real objects in strips and in the whole image. Color bunches provide an analog of superpixels employed in classical segmentation methods.

A. Construction of the Search Lattice in STG

On the set of color bunches *STG* a “search lattice” SearchLat is constructed [5], which makes it possible to perform global image analysis. If we put on each middle line of each strip of the image the localization intervals [beg_b, end_b] of all bunches of the strip, then we obtain a certain covering of it. Recall that the density of the color bunch is $dens(b) = Card^b\ /L([beg_b,\ end_b])$ (the cardinality, divided by the length of the interval). The color bunches that have a maximal density in some of its points are called dominating color bunches. It is clear that the dominating color bunches generate a covering of the middle line. It is true [5] that we can always choose a linearly ordered sequence of dominating color bunches which generate a covering of the middle line. The dominating color bunches included in the linearly ordered covering are called basic color bunches. The basic color bunches of all strips generate the image search lattice SearchLat (*STG*) [5]. Many tasks of searching landmarks and objects in the frame can be reformulated as tasks of finding certain abstract objects in *STG*.

B. Construction of global objects using the geometrized histograms method

We mean by global objects subsets of color bunches containing color bunches located in several adjacent strips. Global objects are constructed from left and right germs of global objects (left and right contrast curves) [4]. Germs of global objects contain dominating bunches having contrast intensity-color characteristics with their neighbors. To segment the image, a bipartite graph of left and right germs of global objects or contrast curves *LRG* is determined. If the image is divided into horizontal strips, then informally a left or right germ of a global object is a chain of color bunches in adjacent strips with similar intensity-color characteristics. Note that for left (right) germs left (right) ends of the localization intervals of their color bunches vary continuously from strip to strip, and the left (right) adjacent bunches of members of the left (right) germ have contrast color characteristics with it. These chains are constructed bottom-up, passing from strip to strip [4].

When we deal with chains of color bunches representing road markings, we should keep in mind that because of inhomogeneous illumination and shadows, the intensity and color characteristics of the corresponding color bunches may of bunches adjacent in the strip may vary significantly. In the next subsection, the principles for constructing chains of color bunches that can represent a road marking will be described.

C. Principles of construction of objects in STG that are candidates for the image in STG of a road marking

Methods for constructing candidates for an image of a road marking in *STG* have been briefly outlined in [2]. Let us present their more detailed description. The important specific feature of the approach is the fact that the problem of finding road marking is solved without knowing any information about external and internal parameters of the camera shooting the image. The conclusions about the metric characteristics of the frame are made when the image understanding system has detected the marking in the image. These conclusions will be taken into account when dealing with the subsequent frames of the video sequence.

If we deal with the problem of finding rather “narrow” objects, such as road markings, chains of color bunches, which belong to adjacent strips, contain both dominating and dominated color bunches [2]. It is clear that, for parts of the road marking belonging to distant strips, the corresponding color bunches have a smaller density than the color bunches generated by the road parts. Candidates for the image of the road marking in *STG* are constructed from color bunches that are adopted as candidates for parts of the road marking in the corresponding strip. The color bunches that may be images in *STG* of parts of road marking (white – permanent or colored – permanent or temporary) are distinguished from color bunches that are images of road parts by intensity or by intensity and color, as well as by the density and length of the corresponding localization intervals [beg_b, end_b]. Since there may be additional objects in frame, such as other participants of the traffic, puddles, dirt sports, wet parts of the road, bright spots from street lamps and vehicle lights, the differences in parameters between adjacent bunches can be two-sided and one-sided. Since the images of parts of the road generate basic color bunches from the search lattice SearchLat, a reasoning system is designed, which compares the parameters of a certain color bunch with those of the adjacent basic bunches from SearchLat. This system makes a conclusion on whether this color bunch belongs to candidates for being an image of a part of road marking in the part of *STG* connected with the given strip. To determine the difference in intensity, only small thresholds are employed at the level of distinguishability of human vision system (of order of several intensity grades), which can take place between the road marking and road when the illuminations varies in a wide range.

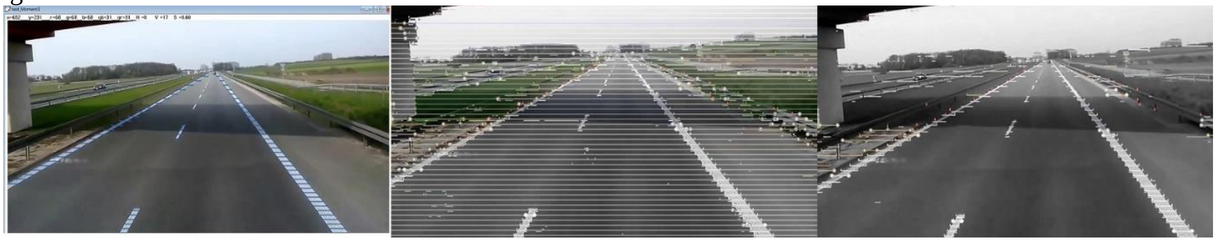

Fig. 1. The stages of constructing a white marking.

Figure 1 presents an example of an image with the detected road marking (the right part), the grayscale image corresponding to the color source image with superimposed color bunches (the middle part), and the candidates for being the images of parts of road marking in strips (the right part). Candidates for the image of parts of road marking in *STG* in strips were obtained using the search lattice SearchLat and the method (reasoning system) described above. Since there is no assumption that the vehicle is located on the road in the system, it is possible to find the marking on a square or in parking places. Since there are no data on the camera parameters and the position of the vehicle on the road, some candidates in the presented image look strange, but in constructing chains and recognizing true candidates, false candidates are eliminated (the right part of Fig. 1).

Let us describe the rules for generating chains of candidates for parts of marking in strips. These chains are called candidates for the image of road marking in *STG*. Chains are constructed from the zero (bottom) strip. All bunches involved in a certain candidate are eliminated from the finding next candidates. Assume that a chain of bunches b_i $i = 0, \ldots, k$, located in a chain of adjacent strips, has been already constructed. Denote by $\mathrm{Int}_i = [beg_i, end_i]$ the localization intervals of the bunches b_i, by e_i the vectors that join the ends of the interval Int_{i-1} and Int_i, and by D_i the vectors of directions joining the ends of Int_0 and Int_i. In constructing continuous left chains of intervals, we deal with the left ends of intervals, while dealing with the right chains– with the right ones. Since there is no assumption on the position of the vehicle on the road, at the first step of the algorithm, we require only that the intervals Int_0 and Int_1 have a strong intersection. Let us introduce in the consideration the reduced direction vectors $d_i = D_i/(i - 1)$. The vectors e_i and d_i have the *x* coordinate equal to the strip width, their direction vector is parameterized by the *y* coordinate. The continuity conditions are formulated in terms of relations between e_i and d_i and restrictions on the jumps of the corresponding angles. Applying these rules, we obtain a set of candidates for the image of the road marking in *STG*. The matrices of traces of marking candidates on the sets of candidates for parts of the marking in strips are constructed. Matrices of traces show the number of the candidate for global marking (left and right separately) that passes in particular strips through each local candidate in each strip. Using the trace matrices, we can establish relations between left and right candidates for the global road marking. The related left and right candidates pass through the same candidates in strips. In the matter of fact, each real global candidate (left or right) has a small number of connected (right or left) candidates. The real road marking has to generate both left and right candidates, which pass through the same color bunches, i.e. they contain the same bunches in *STG*. In the ideal case, one left global candidate corresponds to one right global candidate. In the case of segmentation faults, the connected candidate can be divided into several connected candidates. For example, it may occur that in a number of strips, the marking disappears (occlusion) or because of noise, some false bunches connected with another object (a part of another vehicle or a puddle) may be added.

III. Road Marking Recognition Based on Candidates Constructed in *STG* and Comparison with the Other Methods

Among the constructed continuous chains of candidates for road marking in strips, i.e. images of the marking in strips, false chains may be found. These chains may be determined by road railings, other participants of the traffic, wet places, etc. The presence of such false objects is noted in all main papers on this subject [1, 8]. To eliminate them, the shape of candidates for marking is analyzed. For this purpose, methods developed in [5] are employed. For each candidate, the straightforwardness hypothesis is tested and a certain inclination is attached to it. A curvature index is determined for curvilinear parts. The method for determining inclination is stable to segmentation errors and is based on the analysis of histograms of inclinations for segments joining points of the ends of localization intervals of the chain [5]. To eliminate false candidates for road marking, the relation between connected left and right candidates, described in the previous subsection are studied. If all connected right candidates have attached inclinations, then their proximity to that of the selected left candidate is tested. Based on these evaluations a certain validity index is assigned to the obtained chains. Using such parameters as the length of the chain, the number of the initial strip, the coordinates of the point of intersection of the extension of the chain with the lower image boundary, the position of the chain relative to the vertical central line of the image, the behavior of the chain at “infinity”, relative distances between chains, the coordinates of their vanishing points, a group of chains with the coordinated behavior are selected. This specifies basic elements of the road marking. The closest to the center chains are distinguished and certain chains are interpreted as continuations of each other (interrupted road marking, partially occluded road marking). The closest candidates for left and right solid lines are found. When the basic candidates are selected and their relative position is determined, other candidates are checked for fitting. After this testing, the false candidates are eliminated. The chosen candidates give a ground for dealing with the next frame. These actions and the generation of trace matrices are performed for candidates for white and yellow marking separately. It is explained by the fact that in the winter period white road marking may have a slightly yellow color, and it is a particular intelligent problem to distinguish both types of road marking during this period. Candidates for yellow road marking start from color bunches with intensity-color characteristics in the yellow-orange domain. As the initial bunches, the bunches with a rather big saturation are chosen. Then a fall of saturation is possible (which is really occurred in images shot by real cameras), and the construction is continued based mainly on the geometric fitting of localization intervals.

Figure 2 presents examples of the operation of the proposed reasoning system in the form of several frames from records of the results of processing a particular video sequence. The complete records of processing this video sequence can be found in [9].

Fig. 2. Examles from records of processing a particular video sequence.

A. A comparison with the other methods for detecting road marking and the advantages of the developed method

As the main source of information about the previously developed methods, we use the survey [1]. The bibliography of [8] was also very useful. In addition, we also analyzed the papers from the bibliography of [2]. In [2], we have already conducted a certain comparative analysis of the abilities and potentialities of the proposed method compared with the methods developed earlier. Let us present some results of the comparison.

- The developed method does not require the data on the internal and external camera parameters and operates well when the inclination of the road changes sharply, in contrast to the methods using the flat model of the road vicinity.
- The method creates a big list of candidates for road marking parts, including possible false candidates such as, railings, parts of other participants of the traffic, posts, wet places, etc. Then there is a reasoning system that based on comparative logical analysis selects true parts of the road marking.
- In each processed frame, program gives a convenient description of the list of details of parts of marking, such as inclination, curvature, the distances between parts of marking, their intensity-color characteristics. These data can be quickly and efficiently used for estimating the next frame.
- The method does not impose the restrictions on the shape of road marking and is efficient under sharp changes of the shape of temporary marking, in contrast to methods based on the Hough transform. It is also possible to find the marking of adjacent lanes and construct the marking on distant parts of the road.
- The method subtly takes into account the color characteristics of the road marking and allows one to separate temporary road marking from the permanent one in case of winter dirty road marking.
- The method can deal with partially rubbed out and dirty road marking, typical for the winter-spring season in countries with the climate similar to that of the Russian Federation.
- The method has very high performance for HD video (1280x720, 1920x1080), which is reached on standard personal computers.
- Simultaneously, complete segmentation of the frame is performed, and such important objects as the sky, roadsides, the road, other participants of the traffic, dangerous objects, like posts near the road are detected [5, 6]. This makes it possible to semantically analyze the road scene, even when some of modules give incorrect results.

The performance of developed method was studied using video sequences of road scenes shot by different cameras, even by cheap car video registrators. The detailed description of the developed software system implementing the geometrized histograms method and the image understanding systems for the whole list of tasks was presented in [7]. In [7], the results of operation of the developed method was conducted in two modes [7]: 1. With multithreading performed based on parallel processing of strips on which the image was divided (“on” in the table), 2. Without multi-threading (“off” in the table). Note that not only the road marking was detected, but the whole complex of tasks mentioned above were solved. Figure 3 shows the results obtained in [7]. The performance was estimated for a video sequence with 1400 frames, which can be found in [9] with superimposed results of processing. In particular, for HD video of resolution 1920x1080 was obtained performance of

25 fps. In [10], the results are applied to controlling a vehicle.

TABLE I. PERFORMANCE EVALUATION

Multithreading state	Num of strips	Resolution of images		
		640x480	*1280x 720*	*1920x 1080*
Off	48	25.381	59.1323	118.512
On	48	10.2064	18.8395	36.169
Off	72	29.9551	66.3633	124.652
On	72	16.3946	22.6448	38.6855
Off	108	35.3268	76.5239	136.415
On	108	24.1019	29.0153	41.0169

Fig. 3. Estimates of system performance in ms for different frame resolution and different numbers of strips in the image division for a video sequence of road scenes.

The development of the reasoning system is under progress and some disadvantages are being eliminated.

IV. Detection and Analysis of Stop Lines, Traffic Lights, Traffic Signs

These elements should be evaluated in producing a control mode for an autonomous vehicle when going into a crossing, as well as in order to warn the driver in ADASs. Using even rough camera calibration of the camera close zone, knowing the approximate distance from the vehicle to upper boundaries of the closest strips of the image, it is easy to detect the solid marking near the vehicle. The solid lane marking is one of the features of the approaching stop line. As processing of a big number of images has shown, a rather distant stop line is represented in *STG* by one or several dominated color bunches in a certain strip in front of the vehicle. These bunches overlap the lane. The intensity-color characteristics of these bunches are close to those of the lane marking. These bunches are found by simple search in *STG*. While approaching the stop line, these bunches become dominating. A part of the image containing candidates for the stop line is cut from the image. Using the division of the cut image into vertical strips, the stop line is exactly localized. To detect a real stop line, it is necessary to analyze additional features (the red traffic light, stop traffic sign, give way sign).An important component of planning the motion on crossings is to detect traffic signs and to determine their state. The geometrized histograms method makes it possible to detect in images contrast color objects of size less than 3 pixels. To detect traffic lights in advance, it is necessary to work with HD video in real time. At present, using multithreading in constructing *STG* on standard processors with four 4 CPU cores, the performance is about 50 fps for resolution of 1280x720 (about 25 fps for 1920x1080) for 2.5 GHz CPUs [7]. Traffic lights have rather big saturation and intensity and a small size. At the first stage, all such candidates (as a rule, dominating) are found in *STG* in upper strips. A real traffic light gives a chain of bunches on a sequence of frames, moving in a certain direction. The search is simultaneously conducted in the all three ranges (green, yellow, and red) finding the moment of changing the signal.

The detection and recognition of traffic signs are other important elements of controlling the motion into crossings. The traffic signs (give way, stop, main road, etc.) contain informative color parts and even the combination of these components (e.g. green and blue ones in the traffic sign of pedestrian crossings). As a rule, these components have rather big saturation and can be effectively found in *STG*. After localizing a candidate, a small its neighborhood is cut and *STG* is constructed for it. Methods have been developed that allows one to find in *STG* typical geometrical figures (rectangles, rhombs, triangles, etc.). This specific technique will be published separately.

Complex scene analysis is a substantial element of scene analysis. The detection of road marking under the bridge (Fig. 2) makes it possible to save the correct motion direction, when the other systems (detection of the road, roadsides, and sky) produce wrong results. It is very important to have a reasoning system that takes into account the results for previous frames. At present this system is under development. A particular publication will be devoted to this topic.

V. Conclusion

In this paper, algorithms for detecting and recognizing road marking were briefly described. Examples of operation of a software system implementing these algorithms were presented. Estimates of the performance of this system were given for HD video. Algorithms for solving tasks for finding stop lines, traffic lights, and traffic signs based on the graph *STG* were also outlined.

Acknowledgment

This work was supported by the Russian Foundation for Basic Research, projects N. 18-07-00127 and 19-08-01159.

References

[1] S.P. Narote, P.N. Bhujbal, A.S. Narote and D.M. Dhane, “A review of recent advances in lane detection and departure warning system,” Pattern Recognition, vol. 73, no. 1, pp. 216-234, 2018.

[2] R.V. Dosaev and K.I. Kiy, “A new real-time method for finding temporary and permanent road marking,” CEUR Workshop Proceedings, vol. 2391, pp. 86-96, 2019.

[3] K.I. Kiy, “A new real-time method for description and generalized segmentation of color images,” Pattern Recognition and Image Analysis, vol. 20, no. 2, pp. 169-178, 2010. DOI: 10.1134/S1054661810020082.

[4] K.I. Kiy, “Segmentation and detection of contrast objects and their application in robot navigation,” Pattern Recognition and Image Analysis, vol. 25, no. 2, pp. 338-346, 2015. DOI: 10.1134/ S1054661815020145.

[5] K.I. Kiy, “A new method of global image analysis and its application in understanding road scenes,” Pattern Recognition and Image Analysis, vol. 28, no. 3, pp. 483-494, 2018. DOI: 10.1134/S1054661818030100.

[6] K.I. Kiy, “An image understanding system based on the geometrized histograms method: finding the sky in road scenes,” CEUR Workshop Proceedings, vol. 2210, pp. 291-299, 2018.

[7] K.I. Kiy, D.A. Anokhin and A.V. Podoprosvetov, “A software system for image processing with parallel computing,” Programming and Computer Software, vol. 46, no. 6, 2020, in press.

[8] P.Ch. Wu, Ch.U. Chang and Ch.H. Lin, “Lane-mark extraction for automobiles under complex conditions,” Pattern Recognition, vol. 47, no. 8, pp. 2756-2767, 2014.

[9] Project “AvtoNiva”. URL: http://project1054516.tilda.ws/.

[10] A. Podoprosvetov, K. Kiy, V. Pavlovsky and D. Anokhin, "Development of Complex Control System for the Autonomous Vehicle Niva," Proceedings of XXI Int. Conf. on Complex Systems: Control and Modeling Problems (CSCMP), Samara, Russia, pp. 311-315, 2019. DOI: 10.1109/CSCMP45713.2019.8976523.

[11] A.Yu. Denisova and V.V. Sergeev, “Algorithms for calculating multichannel image histograms using hierarchical data structures,” Computer Optics, vol. 40, no. 4, pp. 535-542, 2016. DOI: 10.18287/2412-6179-2016-40-4-535-542.