An image understanding system based on the geometrised histograms method: finding the sky in road scenes

K.I. Kiy[1]

[1]Keldysh Institute of Applied Mathematicsof RAS, Miusskayasquare 4, Moscow, Russia, 145047

Abstract. In this paper, the technique provided by the geometrized histogram method for segmentation and description of color images is developed and improved in order to analyze the adjacency relation of left and right germs of contrast objects (left and right contrast curves) on the*STG*. This adjacency relation involves and generalizes the adjacency relation for regions in classical segmentation methods (the so called RAG). The adjacency relation of left and right germs of contrast objects is based on selecting in each layer of *STG* a "basic" set of linearly ordered bunches that in some sense covers this layer. Using this order relation, the adjacency relation for left and right germs of contrast objects is established. This order relation is also employed for finding the correspondence between left and right germs with prescribed geometric and color-intensity characteristic that are not adjacent and lay apart at a distance. Based on the adjacency relation proposed, a technique for constructing complex contrast objects with a prescribed geometric shape and color-intensity description is proposed. The developed technique is applied to analyzing road scenes in order to find the sky in video sequences. The results of finding the object in video sequences by a programming complex implementing these ideas are presented and discussed.

Keywords: Image segmentation, image understanding systems, object detection and tracking.

1. Introduction

In spite of serious progress in image segmentation and analysis [1--4] and many new ideas arising in machine learning and deep learning of convolutional networks, there are still serious difficulties in implementing global image analysis in real time.These difficulties are mainly connected with many different objects occurred in the scene, occlusion, and difficult and diverse illumination conditions in the real world. This makes it difficult to analyze the joint behavior of several real objects or even parts of them by one or another method (classical segmentation, sliding windows with machine learning or deep learning of convolutional networks). Moreover, as the mortal accident with the Tesla pilotless vehicle has shown, it is crucially important not only to classify any frame, but also to have clear understanding of the state of important objects in the image and to have their conceptual description in order to recognize the case of their complete change due to possible occlusion occurred. In this case it can prevent us from recognizing the body of a blue van as a part of the sky. If the image understanding system of a robot is able to select the sky region and determine its type (the shape and location of the

its boundary) and color and intensity characteristics in the image, as well as to interpret the regions over which it lies, the complete change of this information within a small number of frames will give the idea of the dangerous occlusion occurred to the system in order to prevent a possible accident.
In this paper we propose an image understanding system that can solve such problems in real time using only standard computational facilities for suburban and country roads. The approach to designing image understanding systems of such type is based on the geometrized histograms method proposed by the author [5-7]. This method not only segment color images in real time, but also makes it possible to construct adjacency relations between detected objects and to introduce higher-orderadjacency relations for objects that are rather distant in the image. This technique is applied to designing an image understanding system for finding and analyzing sky regions in images and video sequences of road scenes. The designed image understanding system finds the sky in video sequences of suburban and country roads very efficiently. The results can be found on the site [9].

2. A brief description of the output of the geometrised histograms method for an arbitrary color image

A detailed description of the geometrised histograms method can be found in [5—7].Usingthe methodsdevelopedin [5—7], wecanattachtoeach color image the graph of color bunches *STG*(STructural Graph). Toconstruct*STG*, theimageisdividedintostrips St_i, $i = 1,\ldots n$, of the same widthwith boundaries parallel to the horizontal or vertical axis of the image plane*Os*. Suppose that we deal with horizontal strips. Each strip $St_i$is described by the set of color bunchesB_i. $B = \cup B_i$ is the set on nodes of *STG*. Each color bunch $b \in B_i$is characterized by the following parameters: 1. a localization intervalint_b=[beg_b, end_b], belonging to *Os*; 2. $\Delta_H^b = [H_{min}^b, H_{max}^b]$ andH_{mean}^b– the range and the mean value of the hue of*b*; 3. $\Delta_S^b = [S_{min}^b, S_{max}^b]$ andS_{mean}^b –the range and mean value of saturation; 4. $\Delta_I^b = [I_{min}^b, I_{max}^b]$ andI_{mean}^b – the range and the mean value of the grayscale intensity, and 5. the cardinality$Card^b$(approximately, the number of points in the strip St_i whose coordinate *x*belongs to the localization interval[beg_b, end_b] that have the color characteristics belonging to the ranges Δ_H^b,Δ_S^b, and $\Delta_I^b$of the color bunch).

Color bunchesb_1 and b_2 lying in the same strip are called adjacent if their localization intervals int_{b1} and int_{b2} are adjacent. Color bunches lying in the adjacent strips are called adjacent if their localization intervals have nonempty intersection. Edges of *STG*join all adjacent color bunches.

Informally, each bunch describes a certain part of a real object in the strip, its projection on *Os* and the description of numerical characteristics of this part of the object. Thegraph*STG*can be interpreted geometrically by overlaying localization intervals of its bunches ([beg_b, end_b]) on the middle lines of the corresponding strips. Figure 1 demonstrates the representation of an image by the *STG* graph. Color bunches of each strip are superimposed on the middle lines of the corresponding strips. There are two types of color bunches.

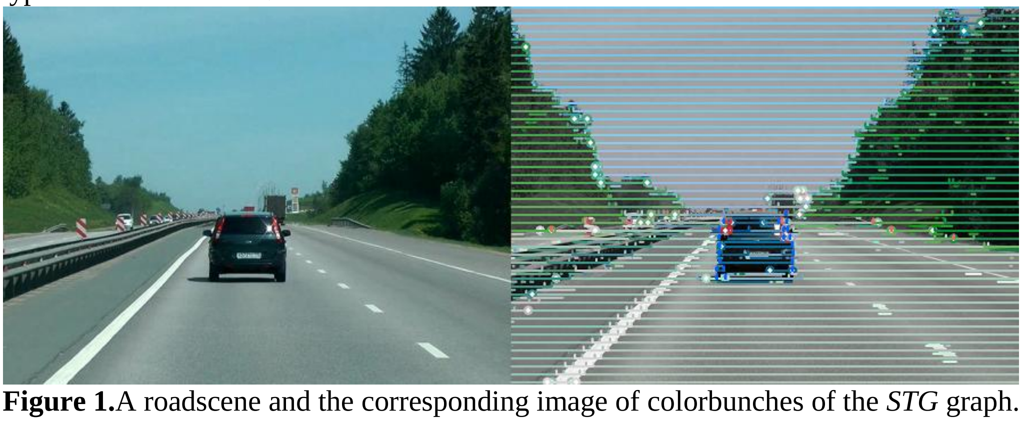

Figure 1.A roadscene and the corresponding image of colorbunches of the *STG* graph.

Color bunches of the first type are called dominating. Dominating bunches in some places of their localization interval int_b hasthe maximum density $dens_b = Card^b/l(int_b)$, where $l(int_b)$ isthe length of the interval int_b. It is clear that the localization intervals of the dominating color bunches generate a covering of the middle line of the corresponding strip. In the visualization, the localization intervals of dominating bunches are put on the entire middle line. In addition we have some kinds of color bunches that are not dominating. These bunches may also be very important. For example as a rule, the signal zones of a distant vehicle may have densitiesless than the densities of the bunches corresponding to the body of the vehicle. However, these bunches are very important in order to recognize the next actions of the driver of a vehicle going in front of our car. In the visualization, the color bunches of the second type are located slightly below the middle line. The construction and numerous experiments with images have shown that color bunches represent any connected color object in the real image with the size greater than three pixels. The description of a color image by color bunches compresses the information on images from millions of pixels to several hundreds of bunches. However, this image description contains all important features of the image, including a description of the geometry of objects belonging to it.

2.1. Continuous objects on STG

In [6, 7] the concepts of left and right contrast curves (left and right germs of global contrast objects) in *STG*were introduced, and a bipartite graph of left and right contrast curves*LRG*was constructed.Let theimagedivided into horizontal strips.A left (right) contrast curve is a chain of color bunches b_i with contrast right (left) neighbors located in adjacent strips St_i $i = k, k+1, \ldots k + d$, such that the intensity-color characteristics of these bunches vary continuously from strip to strip, as well as the coordinate x of their left (right) ends [6, 7]. Suchachainisconstructedupward, beginning from its lowest strip, finding the continuous extension of the previous bunch on the next strip [6, 7]. Figure 2 presents two examples of left (right) contrast curves in images.

Figure 2. Two germsof the sky represented by germs of global contrast objects in *STG*.

In this way up to 256 different left and right contrast curves are found. By the construction, no more than one left and one right germ of global contrast objects can pass through any color bunch. On the set of all color bunches the functions $\mathrm{Germ}_{\mathrm{left}}(STG)$ and$\mathrm{Germ}_{\mathrm{right}}(STG)$ are determined. These functions take at each color bunchas the value the number of the germs passing through this bunch or -1 if there is no such germ. Each color bunch of a left (right) contrast curve has contrast contact with its right (left) neighbor. It is supposed that each left (right) contrast curve is a left (right) part of a certain hypothetical global object. Any left (right) contrast curve has its own geometric pattern determined by the discrete set of left (right) ends of the localization intervals of its color bunches. In the right image of Fig. 2, the presented left contrast curve (painted for visibility by dark intervals) is simultaneously a right contrast curve, since each of its bunches has both left and right contrast neighbors (parts of the forest or the boundary of the frame). Left and rights ends of color bunches of the contrast curve of the right side of Fig. 2 specify standard boundaries of a sky region in a road scene. The right contrast curve in the left image of Fig. 2 has to be completed to generate the whole sky region. Of course it is the most typical situation. In what follows we present a reasoning system that performs the operation of extension sky regions. In addition to the geometric pattern, each contrast curve has intensity-color characteristics determined by its color bunches. For the image of the right side of Fig. 2, together with the geometry of boundaries, we can produce the label “bright blue sky without clouds”. To be able to assemble the sky region from available left and right germs of global contrast objects, we have to introduce a new technique applied for this purpose [6 –8].

2.2. Adjacency relations graphs

In each strip among all the color bunches, we are able to select a basic set of bunches dominating in some part of the strip (having the greatest density in it). It is obvious that localization intervals of the dominating bunches give a covering of the middle line of the strip. For each dominating, bunch it is possible to find its closest left and right dominating neighbors. Using this construction we can select a completely ordered basic subset of dominating color bunches that provides a covering of the middle line.On thisbasic subset, it is possible to introduce a complete ordering (to number dominating color bunches of this subset from 0 to a certain *k*). Figure 3 demonstrates basic subsets for two strips of the image of Fig. 1. In addition, Figure 3 shows that all important parts of objects are included in the descriptions of strips by color bunches.

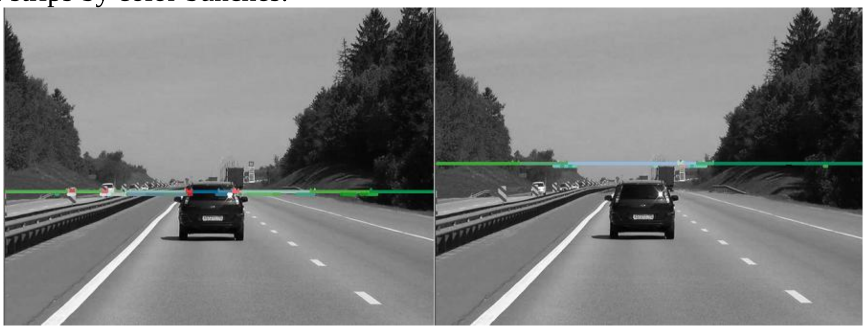

Figure 3. Two basic sets of colorbunches in two different strips.

All linear ordered basic subsets of bunches joined for all strips generate on the image a "search lattice"SeachLat (*STG*).The constructed SeachLat (*STG*)(bunches are numbered with preservation of adjacency relation) allows one to construct the adjacency graph *ADG*, which determines adjacency relation for constructed contrast objects in *STG*. Each left (right) contrast curve (germ of global contrast object) is a continuous sequence of color bunches in a chain of adjacent strips (see Fig. 2). The sequence of left (right) ends of the localization intervals vary continuously, as well as their intensity-color characteristics [6]. By construction only one left or right contrast curve can pass through any color bunch. If this curve exists, then it is uniquely determined by the functions $\mathrm{Germ}_{\mathrm{left}}$(*STG*) and $\mathrm{Germ}_{\mathrm{right}}$(*STG*). Suppose that we have a germ of a contrast global object *G*.Starting from the first bunch b_1 of this germ in the first its strip and moving to the left and right of it, we find all adjacent germsof *G* in the considered strip. In this way considering the germs passing through the direct neighbors of b_1 we can construct the direct adjacent germs of *G* in the strip. Moving from strip to strip we are able to construct the partadjacency graph *ADG*connected with the left and right adjacent germs.Consider the extension of adjacency relations in the down and up directions. In the construction of left and right germs, we introduced the structure that describe for each color bunch the set of color bunches that locate in the next strips above and under the considered strip. Using this structure, we are able to extend the adjacency relations downward and upward. Considering in each strip adjacent germs, passing through the next bunches of the search lattice, we are able to introduce a multiple adjacency graph *MADG*or adjacency graphs of higher orders Adg_i(*STG*). *MADG* makes it possible to perform global image analysis, e.g., to analyze component of the same global object even in case of occlusion [8]. For example we are able to investigate two road sides (left and right) simultaneously. The graph *MADG* makes it possible to assemble complex real objects which contain heterogeneous parts. In this graph not only relations between objects that have common boundaries are established, but between objects separated by occlusion as well. New results connected with a detailed construction and application of *ADG* and *MADG*can be found in [8].

3. Construction of a reasoning system for finding the sky in road scenes

The problem of finding the sky is one of the problems solved in the course of developing the control system of the autonomous robot AvtoNiva, produced by a research group in the Keldysh Institute of Applied Mathematics of RAS based on a stock Niva car. For this purpose, it is not necessary to obtain the detailed description of the sky region at the pixel level. We need only approximate, qualitative and semantic description of the sky region that can be used in the control system for qualitative estimation of the road neighborhood and for detecting possible occlusion caused by unpredictable actions of other participants of the traffic. The statement of the problem under these conditions is described in the next subsection.

3.1. Problem statement and quality estimation

It is supposed that the image of a road scene is divided into a number of strips of the same width with the boundaries parallel to the horizontal axis of the image plane. For example for an image of resolution 640x480 we used 48 strips. We have to specify the sky region by a system of intervals on the middle lines of some of the strips. It is not supposed that the sky region is simply connected. Due to occlusion it may contain several components. We have to find approximate color and intensity characteristics of each connected component and to describe its possible semantic type, e.g., “bright blue sky without clouds”. If thedetected regionof the sky in the form specified above takes into account about 90% of real pixel sky region (with minimum possible false positives) and the lower boundary of the sky region is found with the precision up to one strip, then the solution found is considered as quite successful.In this case, we can approximately deduce the qualitative character of the road (a straight road, a turn, etc.) even in the case of heavy occlusion of the road caused by other vehicles. We are also able to recognize the dangerous occlusion caused by the car in front of us taking into account among other features the complete change in the pattern of the sky region. It is also supposed that the problem has to be solved on a standard PC in real time.

3.2. Specific features of the problem and its solution

If we attentively study a sufficiently large number of images of scenes containing sky regions taken under different illumination conditions, at different times of the day, and during different seasons, then it will be clear that the sky region may be a very complex object.It may contain many different more or less homogeneous parts that are quite different in color and intensity. The sky region is especially diverse during sunset or sunrise. The appearance of clouds may change the pattern dramatically. The presence of sky-similar objects such as walls of buildings (especially without windows) makes the problem even more complex. Under these conditions, it may be impossible to solve the problem using only one frame. Sometimes even a human may fail to determine the boundary of the sky quite correctly using only one image. Only external knowledge and views from other positions may help. For instance in Fig. 4, a part of a circle of the white antenna left of the low clouds over the roof of a Sberbank building gives an example of such a situation.

Therefore to obtain an adequate solution, at least two stages of solving the problem are necessary. At the first stage, a single frame is analyzed and a preliminary solution is described. At the second stage, we compare and study a set of adjacent frames in order to provide the final solution.

Let us describe the first stage of solution. At the first step, we generate a conceptual and semantic description of all left and right contrast curves constructed by algorithms described in [6, 7]. To describe the geometry of any contrast curve, we use methods proposed in [8]. Let us briefly describe them. For this purpose, we divide the boundary points of any contrast of curves (a set of left (right) ends of localization intervals of thecolor bunches involved in this contrast curve) into branches on which the coordinate x(the horizontal coordinate in the image plane) ofitsnodesincreases or decreases. As the additional constraint, we suppose that on thesebranches the absolute valuesof the differences of abs($end_{b(k+1)} - end_{bk}$)(the right curves) or abs($beg_{b(k+1)} - beg_{bk}$)(the left curves) for the adjacent nodes is bounded by a constant connected with the width of the strip. Introducing these constraints, we eliminate the effect of sharp change of the shape of the boundary curve. Then we test whether these branches belong to certain straight line segmentsor they are convex or concave. To test linear hypothesis, we use histograms of inclines of the segments connecting adjacent nodes of the contrast curves. Details can be found in [8]. Since each contrast curve is a collection of color bunches,including their intervals of localization, we introduce for any contrast curve b its weight $W(b)$, which is equal to the sum of lengths of its localization intervals. It is clear that $W(b)$ characterizes the area of the figure determined by the set of localization intervals. We include in this figure all parts of the strip over and under the localization intervals. In addition, the distribution of lengths of localization intervals along the curve determines its behavior at infinity. This parameter separates contrast curves that are long and narrow with decreasing lengths (like parts of the road) and long and wide with increasing and non-decreasing lengths (like forests, fields, parts of the sky, bodies of cars). Then we select both the left and right contrast curves with the maximum weight $W(b)$ that havecolor-intensity characteristics possible for parts of the skyand are located in the top part of the frame. Based on the theleft and right curves of the maximum weightfound, using the search lattice on the image,weconstruct the whole region of the sky.When we move to the left or right on the search lattice, we stop the extension of the sky region when the regions classified as forests, fields, roads,etc., occur. In a similar way, when we move on the search lattice to the bottom of the image, we stop the extension when the regions mentioned above are met.

It is especially difficult to eliminate regions generated by buildings, having intensity-color characteristics similar to those of sky regions. For this purpose, the reasoning system finds sky-similar regions with straight boundaries and tests whether these regions have subobjectsinside with vertical boundaries (windows, doors). Figure 4 demonstrates two complex examples of images of a city landscape.

Figure 4. Skyregions in images of a city landscape.

In spite of several small mistakes, the level of the sky is found quite correctly. The results of processing of video sequences by the presented system can be found in [9]. It is important to note that the results completely support the conclusion that the problem of finding the sky in images is not local and requires elaborate global analysis of the frame.At the end of the first stage, we find the boundary of the sky region in the form of an array specifying the number of the first pixel of the sky region in each column of the image*sk_boun*(k), k is the number of the corresponding column. We also have the set *sk_germs*that contains of all germs (contrast curves) included in the sky region. Analyzing the parameters of the germs of the set, we produce the semantic description of the sky image.

At the second stage of solution, we compare arrays *sk_boun*for the current and previous frames. We also compare the semantic descriptions of the sky regions found of adjacent frames In the case of their correspondence, we adopt a new solution. Otherwise, we analyze the differences and decide whether a dangerous occlusion occurs, taking into account other features such as possible signal zone of the vehicle in front of our car, etc.

4. Programming implementation, demonstration of the results and discussion

The image understanding system has been implemented by a program written in C++ and operating under Windows and Linux environments. This program processes video sequences in real time on standard computers with processors I3-I7 and records the results for each frame of the video sequence tested. For frames of resolution 640x480, the operation speed is 20fps. The program has been tested on dozens of video sequences taken from cars on different Russian roads under different seasons, times of the day, under different illumination conditions. Figure 5 presents several examples from records of the results for three videosequences. On country roads from the considered series the percentage of positive results variedfrom 98 to 100 %. Even on new video sequences of this type processed in the first time without modifying the program a very high percentage of positive results is obtained. Some problems may appear when processing video sequences taken in villages and towns. This is connected with buildings with walls that cannot be distinguished from the adjacent parts of the sky using only one frame. The data base of rules and features are being modified, as well as the work with adjacent frames, in order to eliminate these problems. The results of processing several video sequences can be found in [9]. In addition to the sky region, the video system of the control system of AvtoNiva finds other regions interesting for controlling the vehicle such as the boundaries of the vegetation regions, road regions, the other vehicles on the road. The solution to a part of these tasks was described in [8].

Figure 5. Examples of records of experiments with video sequences.

5. Acknowledgments
This work was supported by the Russian Foundation for Basic Research, projects no. 16-08-00880,16-07-01264a, and 18-07-00127.

6. References

[1] Forsyth, D.A.Computer Vision, a Modern Approach/D.A.Forsyth, J. Ponce. - London: Prentice Hall Ltd., 2003. – 244 p.

[2] Mishra, A.K. Active segmentation/ A.K. Mishra, Y. Aloimonos// Int. J. of Humanoid Robotics.– 2009. – Vol. 6(3). –P. 361-386.

[3] Chen,L.-Ch. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs/L.-Ch. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille[Electronic resource].- Access mode: https://arxiv.org/pdf/1606.00915.

[4] Divvala, S.K. Context and subcategories for sliding window object recognition/ S.K. Divvala // PhD Thesis. –Carnegie Mellon University, Pittsburgh, 2012.

[5] Kiy, K.I. A new real-time method for description and generalized segmentation of color images/K.I. Kiy//Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. – 2010. – Vol. 20(2). –P. 169-178.

[6] Kiy, K.I. Segmentation and detection of contrast objects and their application in robot navigation/ K.I. Kiy//Patt. Rec. Image Analysis: Advances in Mathematical Theory and Applications. – 2015. – Vol. 22(2). – P. 338-346.

[7] Kiy, K.I. A new real-time method of contextual image description and its application in robot navigation and intelligent control/ K.I. Kiy// Chapter 5 in Computer Vision in Control Systems-2, Innovations in Practice, Intelligent Systems Reference Library 75, eds. M. Favorskaya, L.C. Jain. - Springer, 2015. –P. 109-133.

[8] Kiy, K.I. Adjacency graph of contrast objects– globalframe understanding and their application in robotnavigation/ K.I. Kiy//Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, 2018 (in press).

[9] Demonstration site [electronic resource]. – Access mode:http://video.mail.ru/kikip_46/_myvideo.