Image Segmentation Based on RGBD Data

Elena Medvedeva
Department of Radio Electronics
Vyatka State University, VyatSU
Kirov, Russia
emedv@mail.ru

Elisaveta Varco
Department of Radio Electronics
Vyatka State University, VyatSU
Kirov, Russia
varkoelizaveta2011@hotmail.com

***Abstract*—The paper proposes a method of image segmentation based on the joint usage of color and depth data. The method consists of two stages. The first stage involves RGB image segmentation based on contour detection and the subsequent filling of closed regions. This procedure is followed by joint color and depth segmentation. Depth data make it possible to distinguish between pixels with similar brightness characteristics for different objects and improve the quality of image segmentation. To reduce computational resources, we suggest that contours should be detected in high order bit planes of a digital image using the mathematical model of two-dimensional Markov chain. The experimental results prove that the proposed method is effective.**

Keywords—RGBD segmentation, two-dimensional Markov chain, contour detection, depth map.

I. Introduction

Segmentation is used to solve a number of tasks related to detection and recognition of static and dynamic objects in video surveillance, autonomous driving, and others.

Traditional segmentation methods are mainly focused on the use of color or brightness features. According to these methods, the quality of image segmentation depends significantly on the pattern of the scene: smooth or sharp changes in lighting; shadows created by objects; complex backgrounds, and etc. Much work has been done in the field over the years; however, none of the existing segmentation techniques is able to obtain satisfactory results based on color data alone.

New RGBD sensors, for instance, the Microsoft Kinect, which provide synchronized depth and color video frames, have opened up new opportunities to solve the tasks related to object detection and recognition. Unlike RGB data, depth data are considered to be more resistant to changes in lighting and dynamic background objects and can be an effective additional feature for image segmentation.

Fusion of color and depth has become a new research topic in the field of computer vision recently. A number of papers offer various methods for segmenting RGBD data: methods based on combining background subtraction algorithms with depth data [1]; methods using convolution neural networks [2]; clustering [3]; contour, brightness and depth [4], and others.

However, almost all segmentation methods based on combining depth and color data are either insufficiently flexible or require significant computational resources. Therefore, research in this area is an urgent task.

The aim of this paper is to develop a method for image segmentation based on the joint usage of brightness and depth data which can improve the quality of segmentation with reduced computational resources.

II. Image segmentation based on RGBD data

In the RGB color space, each component is a digital halftone image. Its pixels are represented by g-bit binary numbers. The D component is also a multi-bit digital image (depth map) where each element corresponds to the information about the distance from the camera to each point of the observed scene.

There are two ways to perform RGBD data segmentation. The first stage involves color-based image segmentation, and the second stage – segmentation based on depth data or vice versa. It is more preferable to use color data at the first stage. This is due to a number of defects on the depth map – lost and distorted depth values, uneven and noisy object boundaries, incorrectly measured depth values for some materials with mirror or fine-grained surfaces, and so on. Therefore, using depth data at the first stage will significantly distort the object boundaries and break the object contours at the second one.

In this paper, firstly, the RGB image is segmented. To improve the accuracy of selected boundaries of objects of interest, we use the method based on detecting contours with subsequent pixel filling in closed image regions. The second stage involves joint segmentation of color and depth data. Depth data make it possible to distinguish pixels with similar brightness or color characteristics for different objects and thus to improve the quality of image segmentation.

Digital halftone images corresponding to color components can be represented by a set of bit binary images (BBI). The most informative (detailed) regions are highlighted on the high order BBI of the digital halftone image. The low order BBI are binary images in the form of two-dimensional noise. Therefore, we propose to detect the contours of objects of interest in the high order BBI of the digital halftone image. To detect the contours, it is possible to use the mathematical model based on two-dimensional Markov chains with two equally probable states $M_1^{(l)}$, $M_2^{(l)}$ and matrices of probability of horizontal ${}^1\mathbf{\Pi}^{(l)} = \begin{Vmatrix} {}^1\pi_{11}^{(l)} & {}^1\pi_{12}^{(l)} \\ {}^1\pi_{21}^{(l)} & {}^1\pi_{22}^{(l)} \end{Vmatrix}$, and vertical ${}^2\mathbf{\Pi}^{(l)} = \begin{Vmatrix} {}^2\pi_{11}^{(l)} & {}^2\pi_{12}^{(l)} \\ {}^2\pi_{21}^{(l)} & {}^2\pi_{22}^{(l)} \end{Vmatrix}$ ($l = \overline{1, g}$) transitions [5, 6].

This approach to detecting contours will reduce computational resources by using 2×2 transition probability matrices.

Fig. 1 shows an element $\nu_3^{(l)}$ of a two-dimensional binary image with a neighborhood of neighboring elements $\Lambda_{i,j,k} = \{\nu_1^{(l)}, \nu_2^{(l)}\}$.

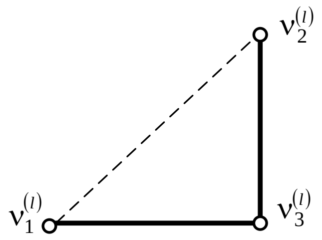


Fig. 1. Fragment of the bit plane of the digital halftone image.

In accordance with the mathematical model of a two-dimensional random Markov process, the amount of information in the $\nu_3^{(l)}$ element for various combinations of neighboring $\Lambda_{i,j,k} = \{\nu_1^{(l)}, \nu_2^{(l)}\}$ elements is determined using the formulas [5,6]:

$$I\left(\nu_3^{(l)} = M_i^{(l)} \middle| \nu_1^{(l)} = M_i^{(l)}, \nu_2^{(l)} = M_i^{(l)}\right) = -\log \frac{{}^1\pi_{ii}^{(l)}\,{}^2\pi_{ii}^{(l)}}{{}^3\pi_{ii}^{(l)}}; \quad (1)$$

$$I\left(\nu_3^{(l)} = M_i^{(l)} \middle| \nu_1^{(l)} = M_i^{(l)}, \nu_2^{(l)} = M_j^{(l)}\right) = -\log \frac{{}^1\pi_{ii}^{(l)}\,{}^2\pi_{ij}^{(l)}}{{}^3\pi_{ij}^{(l)}};$$

$$I\left(\nu_3^{(l)} = M_i^{(l)} \middle| \nu_1^{(l)} = M_j^{(l)}, \nu_2^{(l)} = M_i^{(l)}\right) = -\log \frac{{}^1\pi_{ij}^{(l)}\,{}^2\pi_{ii}^{(l)}}{{}^3\pi_{ij}^{(l)}};$$

$$I\left(\nu_3^{(l)} = M_i^{(l)} \middle| \nu_1^{(l)} = M_j^{(l)}, \nu_2^{(l)} = M_j^{(l)}\right) = -\log \frac{{}^1\pi_{ij}^{(l)}\,{}^2\pi_{ij}^{(l)}}{{}^3\pi_{ii}^{(l)}},$$

where ${}^r\pi_{ij}^{(l)}$ $\left(i,\, j = \overline{1,2};\, r = \overline{1,3}\right)$ are elements of transition probability matrices in one-dimensional Markov chains with two states – ${}^1\mathbf{\Pi}^{(l)}$ (horizontally), ${}^2\mathbf{\Pi}^{(l)}$ (vertically), and ${}^3\mathbf{\Pi}^{(l)} = {}^1\mathbf{\Pi}^{(l)} \times {}^2\mathbf{\Pi}^{(l)}$.

The elements of the transition probability matrices are supposed to be known a priori and obtained from a large number of samples of real images.

After comparing the calculated amount of information with the threshold, the decision on whether the analyzed element belongs to the contour point is made. The threshold value is calculated as the average value between the minimum amount of information and the amount of information when at least one of the neighboring elements assumes a different state.

For an 8-bit digital halftone image represented by 256 brightness values, it is possible to select all light regions with brightness ranging from 128 to 255 in a dark background using the high order (8th) bit plane, or, conversely, all dark objects in the background with brightness above 128. To highlight regions in less contrasting images with indistinct boundaries, it is necessary to detect the contours in the following binary images of the 7th or 6th bit of the digital halftone image. In this case, the contour image will represent the sum of contour images of several bits.

The proposed method of contour detection requires insignificant computational resources which are determined by comparison operations with two neighboring elements. As a result, one-pixel closed contour is obtained. This property is important when performing the following procedure – filling closed regions with color.

To fill closed regions with color, the range of brightness values $[Y_{\min}; Y_{\max}]$ for the object is specified. All the elements within the object area are assigned an average brightness value Y_{cp} (or a label with a specified value). To fill the regions with color, the line seed fill algorithm was chosen [7]. It provides a significant gain in memory and processing time by storing only one seed element for each filled regions. As a result of such image processing, the object can be divided into several parts or have inaccurate borders due to uneven illumination, the presence of shadows or glare. In addition, extraneous objects in the background of the scene can be seen in the image along with the objects of interest. All these factors will influence the quality of solution of the subsequent tasks of image detection, classification and recognition.

At the second stage, a range of data values $[X_{\min}; X_{\max}]$ is set on the depth map that the object of interest can take, and a mask is formed. Next, the mask is superimposed on the result of segmentation of the RGB image and the final stage of selecting objects is performed.

This procedure allows you to distinguish between objects that have similar brightness or color characteristics, but varied range characteristics, as well as improve the segmentation of objects in uneven lighting, the presence of shadows, etc.

Fig. 2 shows a flowchart explaining the algorithm.

III. EXPERIMENTAL RESULTS

The RGBD Object Dataset was used to do research [8]. The RGBD dataset contains pairs of sequences of color images and depth maps, as well as segmentation results based on depth and color data, using the RANSAC algorithm and an adaptive Gaussian mixture (AGM) model [9]. Each video sequence consists of 199 of size frames. In each image, an object of interest is only one item.

Fig. 3 shows examples of segmentation algorithms: (a) – the original RGB image; (b) – reference marking; (c) – segmentation using the RANSAC algorithm and AGM; (d) – segmentation based on brightness data; (e) – the result of joint segmentation according to brightness and depth. The brightness segmentation of the image “Apple” is performed using the R component the 8th BBI; segmentation of images “Banana” and “Scissors” - according to the B-component the 8th BBI ; segmentation image “Coffee mug” - according to the G-component the 8th BBI .

The results given (Fig.3d) prove that the segmentation algorithm based on contour detection accurately localizes the boundaries of objects.

Additional use of depth data (Fig.3e) makes it possible to improve the quality of segmentation: to remove the selected fragments which are close in brightness to the object of interest, get rid of shadows, etc. In addition, when comparing the results in Fig. 3e and Fig.3c, it can be seen that the developed method allows more accurate selection of objects of interest than the method proposed in [9].

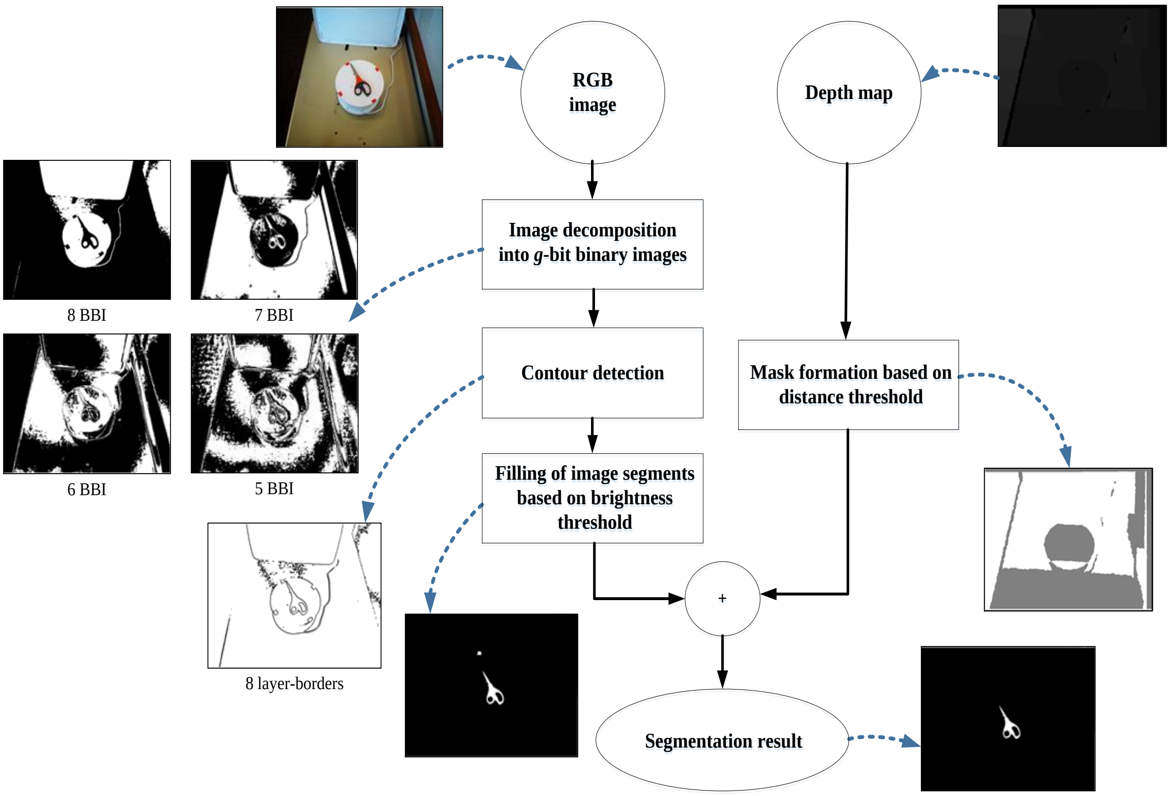


Fig. 2. Flowchart of RGBD image segmentation algorithm.

The segmentation process can be performed automatically for typical images (or sequences of video frames) in which objects of interest have similar characteristics in brightness and depth.

Precision (P) and recall (R) criteria were used to assess the quality of segmentation, and the error coefficient was calculated (E) [10]:

$$Precision = \frac{TP}{TP + FP}, \tag{2}$$

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

$$E = \frac{FP + FN}{TP + TN + FP + FN}, \tag{4}$$

where TP – true positives; TN – true negatives; FP – false positives; FN – false negatives.

The precision within the segmented region is the percentage of pixels which actually belong to the given region in relation to all the pixels that are assigned to this region. The recall criterion measures the percentage of all truly defined pixels which belong to the segmented region in relation to all the pixels. The error coefficient E takes all the error pixels into account in relation to the total number of pixels.

Reference segmentation images were used to calculate precision, recall and error coefficient.

Table 1 contains the results of assessments of the quality of image segmentation using the developed method and the known method [9]. The assessments were calculated using individual images and averaged over the entire video sequence.

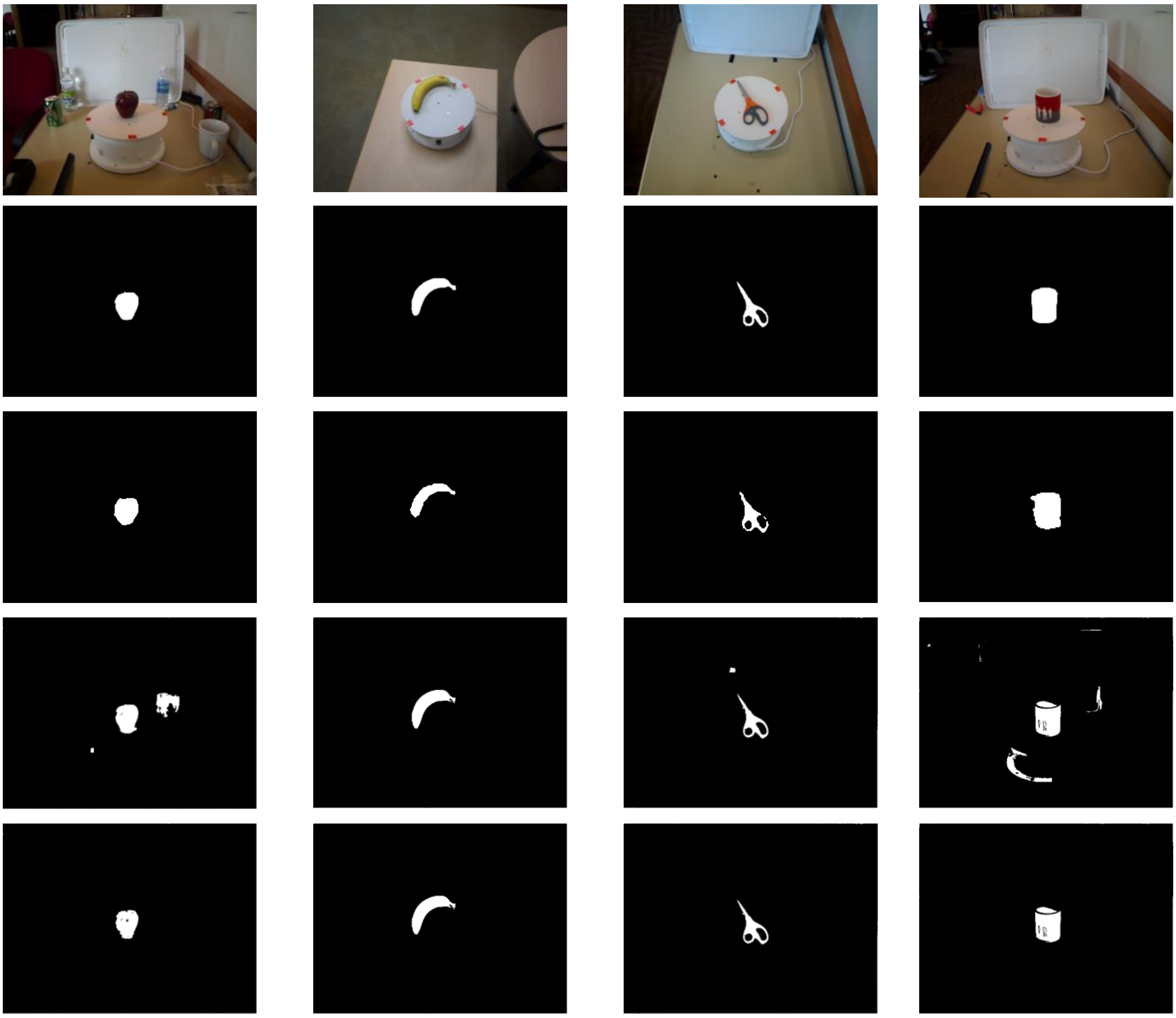
TABLE I. ESTIMATION THE RESULTS OF SEGMENTATION ALGORITHMS.

Video sequences	*based on brightness*			*based on brightness and depth*			*based on RANSAC and AGS model [9]*		
	P	*R*	*E*	*P*	*R*	*E*	*P*	*R*	*E*
Apple	0.93	0.58	0.0086	0.89	0.98	0.0015	0.90	0.97	0.0013
Banana	0.93	0.93	0.0018	0.93	0.93	0.0018	0.80	0.97	0.0027
Scissors	0.80	0.91	0.0026	0.82	0.97	0.0021	0.78	0.93	0.0026
Coffee mug	0.83	0.63	0.0118	0.98	0.80	0.0030	0.95	0.94	0.0014
Comb	0.89	0.40	0.0237	0.91	0.90	0.0035	0.85	0.94	0.0035

Joint segmentation has similar values of precision with those for brightness segmentation but increases the recall score (up to 2.1 times) and reduces the segmentation error (up to 5.7 times).

IV. Conclusion

The proposed method of image segmentation based on the joint usage of color and depth data makes it possible to accurately select the boundaries of objects of interest and effectively distinguish the pixels with similar brightness characteristics for different objects. Due to the detected contours in high order bit planes of the digital image using the mathematical model of two-dimensional Markov chain, it is possible to reduce the computational resources when implementing the algorithm. The algorithm can be used to solve a number of tasks related to object detection and recognition in video surveillance systems, autonomous driving, etc.



a)
b)
c)
d)
e)

Apple | Banana | Scissors | Coffee mug

Fig. 3. Comparison of RGBD data segmentation methods.

REFERENCES

[1] R. Trabelsi, I. Jabri, F. Smach and A. Bouallegue, “Efficient and fast multi-modal foreground-background segmentation using RGBD data,” Pattern Recognition Letters, vol. 97, pp.13-20, 2017.

[2] X. Qi, R. Liao, J. Jia, S. Fidler and R. Urtasun, “3D Graph Neural Networks for RGBD Semantic Segmentation,” IEEE International Conference on Computer Vision, Venice, Italy, pp. 5199-5208, 2017.

[3] H. Yalic and A. Can, “Automatic Object Segmentation on RGB-D Data using Surface Normals and Region Similarity,” Proceedings of the 13th Intern. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol. 4, pp. 379-386, 2018.

[4] F. Xiang, C. Chen, C. Wang and C.-C. Jay Kuo, “Image segmentation using contour, surface, and depth cues,” IEEE Intern. Conf. on Image Processing, 2017.

[5] E.P. Petrov, I.S. Trubin, E.V. Medvedeva and S.M. Smolskiy, “Mathematical Models of Video-Sequences of Digital Half-Tone Images,” Integrated models for information communication systems and net-works: design and development, IGI Global, pp. 207-241, 2013.

[6] E.V. Medvedeva and E.E. Kurbatova, “Image Segmentation Based on Two-Dimensional Markov Chains,” Computer Vision in Control Systems-2. Innovations in Practice, Springer International Publishing Switzerland, pp. 277-295, 2015.

[7] T. Pavlidis, “Algorithms for Graphics and Image Processing,” M.: Radio and Communications, 1986.

[8] RGBD Object Dataset [Online]. URL: http://rgbd-dataset.cs.washington.edu/dataset.html (23.05.2019).

[9] K. Lai, L. Bo, X. Ren and D. Fox, “A Large-Scale Hierarchical Multi-View RGB-D Object Dataset,” IEEE Intern. Conf. on Robotics and Automation, 2011.

[10] S. Russell and P. Norving, “Artificial Intelligence: A Modern Approach,” M.: Williams Publishing House, 2006.