Object Detection Based on RGB and Depth Information

Kangri Wang, Xiaolin Zhang

***Abstract*—**Object detection technology is an important technology for intelligent robot to understand the scene and conduct human-computer interaction. There are important applications in motion control. In this paper, the object detection technology based on RGB and depth information based on optimal dynamic programming algorithm is analyzed. Firstly, the related technology of robot object detection is introduced. On the basis of the discussion of object segmentation algorithm in RBG-D scene, the feasibility of optimizing dynamic programming algorithm for RGB and depth information object detection is studied.

***Keywords*—**optimized dynamic programming algorithm; depth information; object detection; feasibility study

I. INTRODUCTION

AT present, the more mature object detection method is to detect the object in the indoor scene by training on the basis of the annotated data[1]. In the practical application, the intelligent robot must first learn the object concept information and other related contents [2]. But in the process of learning and recognition, because of the difference of image semantic segmentation precision, the final detection result will also appear certain deviation [3-4]. In view of the complexity of indoor scene and layout, Using RGB and depth information processing method based on optimized dynamic programming algorithm can solve the problem of image semantic segmentation and improve image semantic segmentation [5].

II. OVERVIEW OF ROBOT OBJECT DETECTION TECHNOLOGY

A. Image semantic Segmentation Technology

With the increasing maturity of multimedia technology and the rapid development of machine vision technology, the ability of intelligent robot to understand image and video is gradually enhanced, and the research of image semantic segmentation has also made important progress [6]. Image technology mainly includes three levels of image understanding, image processing and image analysis, and each layer is progressive. Image segmentation is in the middle of image analysis and image processing in the framework of the whole image technology [7]. Image semantic segmentation is a very important link. Image semantic segmentation technology makes it become several sub-blocks by identifying the image semantics and accurately segmenting the image [8-9]. The three stages and main contents of image semantic segmentation are shown in Table 1.

Kangru Wang is with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China; University of Chinese Academy of Sciences, Beijing, 100049, China (corresponding Author Email: wangkangru@mail.sim.ac.cn)

Xiaolin Zhang is with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China.

Tab. 1 Three main stages of image semantic segmentation

Three main stages	Feature or method	Specific content
Image bottom feature extraction	Color feature	Global color feature index、Local color feature index
	textural feature	Spatial distribution of adjacent gray levels of pixels
	form feature	Wavelet descriptor, spline fitting curve, Fourier descriptor, angular radius exchange, orthogonal moment
image segmentation	Threshold segmentation method	Pixel comparison using Gray threshold
	Area-based segmentation method	Comparison of regional division by using seed region growth method and region division combination method
	Edge segmentation method	Detection and Segmentation method of Image grayscale violent change
Image semantic mapping	Mapping underlying features to high-level semantics	The semantic meaning of the note pixel, make the segmentation problem practical, attach the semantic label to the corresponding object

B. Depth information acquisition

With the development of science and technology, two-dimensional information extraction of objects has been gradually transferred to three-dimensional information extraction stage, which is necessary for the further human world. Depth information has three-dimensional characteristics, in different stages of image semantic segmentation, Therefore, object depth information acquisition method has an important impact on object detection. At present, the research on object depth information extraction method has made remarkable progress. For example, Kinect for Xbox360, a product developed by Microsoft, is used to obtain object depth information, as shown in figure 1.

Fig. 1. Microsoft Kinect for Xbox360

The product is used in the peripheral equipment of the Xbox 360 host computer. It can simplify the use of the device and can operate the system interface by voice or gesture without manual operation. The structure and technical parameters of the camera are shown in Table 2.

Tab. 2 Kinect Camera structure and main parameters

Composition structure	camera lens	parameter	function
Intermediate mirror	RGB color camera	Maximum resolution 1280*960	Collect object color information
Right mirror	infrared ray CMOS camera	Support 640*480 imaging resolution	Collection of scene depth data
Left mirror	Infrared emitter	——	Infrared emission device

In addition, Kinect also has the technology of chasing. The built in array microphone can be broadcast by four microphones at the same time, and the noise source and location can be identified after eliminating the noise. This technology product can provide support for the deep information acquisition of object detection.

The normalized data obtained by all training samples are adopted as input information. The SOM network model is used to conduct training and simulation of these data. The schematic diagram of the competition neural network model structure is shown in Fig. 2 below:

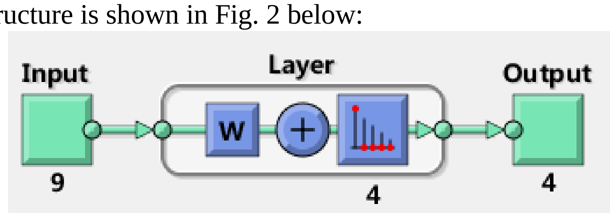


Fig. 2. Schematic diagram of the competition neural network model structure

III. Object Segmentation Algorithm in RGB-D Scene

A. Segmentation algorithm framework

RGB-D indoor scene segmentation framework is shown in Figure 1 based on, based on the framework, using gPb-ucm method for image segmentation, the main use of the image color information, and then change in the initial perception group. The segmented super pixels, to further determine the super pixel attribution, using intuitive variable group method of adjacent ultra the similarity of the pixels were compared. Calculation of super pixels between the normal angle, color distance has the average distance, thus confirming its similarity. If the image of the two super pixels meet the above conditions, is determined to be attributable to the same object, and then carries on the merger. The image no longer exist similar guidance After the super pixel block , the image segmentation can be completed, shown as Figure 3.

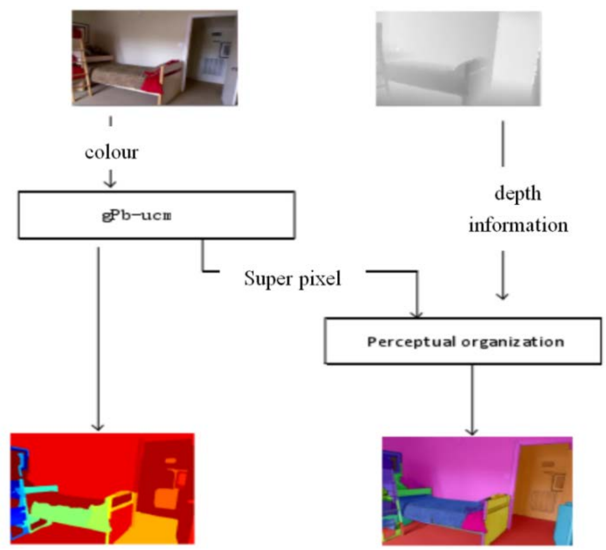


Fig. 3. Segmentation algorithm framework based on RGB-D indoor scene

B. Object Segmentation algorithm in RFB-D scene

In order to improve the accuracy of object image segmentation and avoid the influence of external moving speed, it is necessary to segment the object by combining the depth information of the image, and on the basis of the initial segmentation, the perceptual variable group is executed. The object is divided into two stages, one is the initial segmentation stage, and the other is the initial segmentation stage. Second, the stage of re-segmentation.

In the initial stage of image segmentation, it is necessary to segment a given image from a picture display object. Whether the global information or the local information judgment method is used alone, the result is one-sidedness. GPb-cum method, which contains global and local information of objects, can be used for image contour segmentation. The composition and function of the method are shown in Table 3.

Tab. 3 GPb composition and specific role

constituent part	full name	function
mPb	Multiscale probability of a boundary	Image processing using Image Local Information
sPb	Spectral probability of a boundary	Image processing using Image Global Information

When the gPb detector completes its work, it outputs a probabilistic value E, which represents the probability value at the pixel point in theta direction that belongs to the boundary of the image. A contour with the greatest probability can be obtained by using the watershed method, which is based on the theory of simulated geodesy and belongs to the mathematical morphological segmentation method, which simulates the target table segmentation image as topological geomorphology. The grayscale value of pixel points is used to express the corresponding altitude, and the local minimum value machine influence area is the catchment basin, and its marginal is the watershed. Usually, the process of watershed formation is demonstrated by simulating the entering process, and the surface of the local minimum value is found first. The model will enter the water slowly, and then build the dam at the confluence of the two basins and get the watershed. The main process is divided into two steps, one is sequencing. The second is inundation. By using this method, the maximum points that have not been submerged can be outputted. By using the pixel points with the greatest probability value, the object wheel can be segmented optimally and corrected by the contour arc. Ensure the accuracy of the contour. The initial segmentation effect is shown in figure 4.

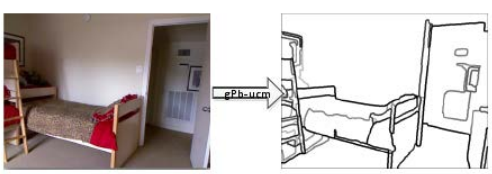


Fig. 4. Preliminary segmentation effect diagram

On this basis, combined with RGB and depth information perception variable group algorithm, the image is re-segmented to improve the problem of difficult segmentation in the initial segmentation process, and to introduce depth information which is not affected by external factors. To ensure the accuracy of the final segmentation results. Further segmentation based on the initial segmentation of the super-pixel, using the three criteria defined in the Gestalt criterion mining superpixel blocks, namely, the angle, the minimum distance and the color distance.

C. Performance test analysis

The following is based on the NYUD2 data set to test the above algorithm, the data set has semantic label RGB-D dataset, which contains RGB data, manually annotated semantic picture and depth data. The data needed for the experiment is captured by Kinect sensor. The composition of the data set and the number of specific parameters are shown in Table 4.

Tab. 4 NYUD2 dataset characteristics

data item	number
RGB	1449 groups
Depth image	795 groups training and 654 groups test
Scene information	464
Unlabeled frame	407024

This experiment is carried out in the environment of Ununtu14.04 workstation. The main frequency is 3.0 GHz. Intel is used to deal with it and 32GRAM. It is a standard HPZ820 workstation. From the result of the experiment, we use RGB information separately. The performance comparison of the segmentation algorithm using depth information alone and using RGB and depth information is shown in Table 5.

Tab. 5 Performance comparison of different object segmentation algorithms

algorithm	error rate	Regional quality	computation time
Using RGB information alone	0.31	0.55	10.9ms
Separate use of depth information	0.25	0.62	11.3ms
Combined use of RGB and depth information	0.21	0.64	14.7ms

It can be seen from Table 5 that the image segmentation method based on gPb-ucm algorithm after adding perceptual change groups (using RGB and depth information) can make the support relationship between objects in the scene clearer and have a significant effect on improving the segmentation effect. Moreover, the accuracy of object segmentation can be improved, but the computational time is less. The specific segmentation effect of the three algorithms is shown in figure 4. The three columns from left to right correspond to the three algorithms from top to bottom in Table 5 and figure 5.

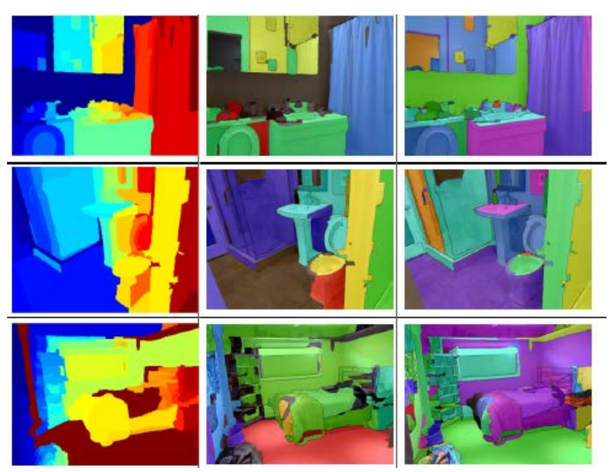


Fig. 5. Effect diagram of three segmentation algorithms

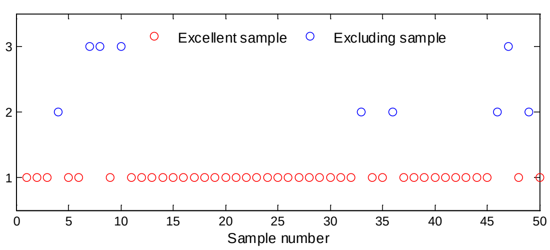
IV. Object Monitoring Based on RGB and Depth Information Based on Optimal Dynamic Programming Algorithm

A. Multi-feature semantic annotation algorithm

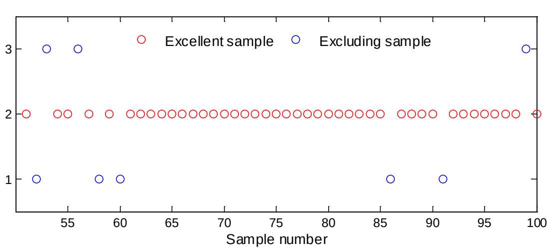
After the image segmentation is processed by the above image segmentation method, the image semantic tagging and image understanding should be carried out. The key problems are region feature selection and semantic classification. It is necessary to use multi-feature semantic annotation algorithm to annotate the segmented image region, extract regional features, construct super-pixel semantic annotation pool, and then introduce the features into the classifier for training. Finally, the image segmentation semantics is obtained. The selected NYUD2 data set in the above experiments can be used to obtain the color and depth information features, and then normalize the region feature units of the hyperpixel block. Using the super-pixel semantic tag pool constructed by the previous steps, the training sample is trained by classifier, and each super-pixel is assigned semantic label.

B. Region feature extraction

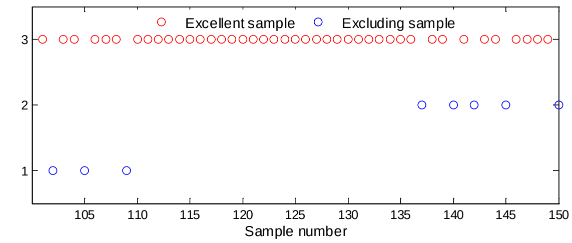
Region feature extraction mainly includes the following aspects: (1) Color feature extraction. In the existing algorithms, in order to extract high complexity features, It may take a huge amount of image processing to achieve the object detection purpose. Color features do not need to deal with a large number of calculations, as long as the image pixel value conversion can be achieved, the complexity is lower. By establishing a color histogram, To describe the color feature of super pixel and measure the difference of adjacent super pixel,shown as figure 6;



(a)



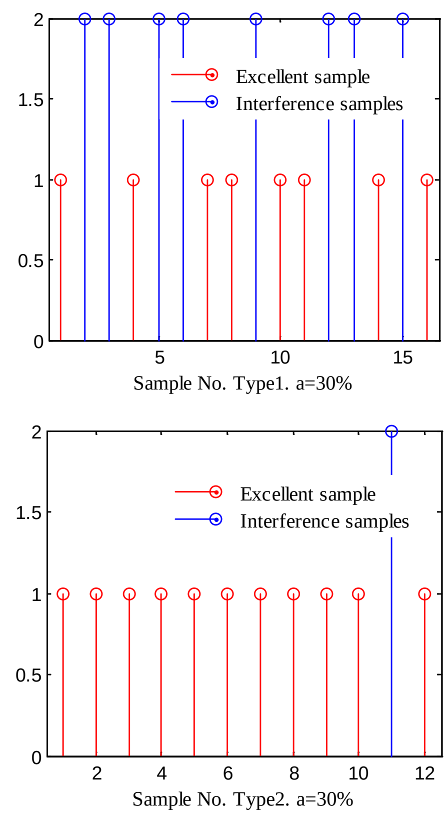
(b)



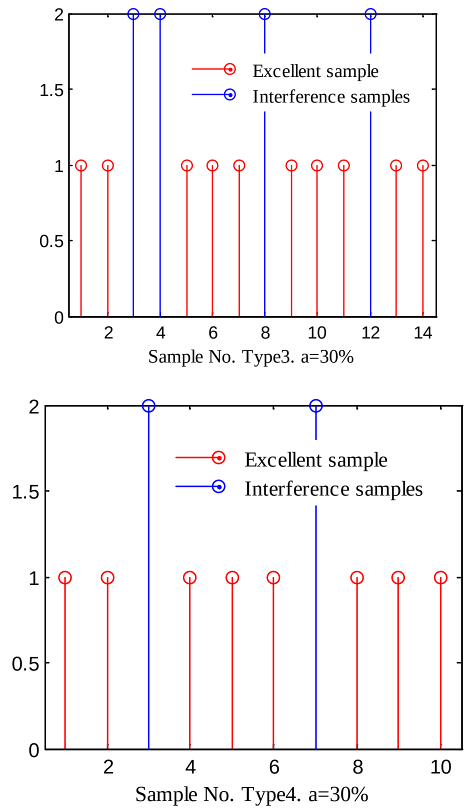
(c)

Fig. 6. Elimination of disturbing samples from samples of different sources

(2)Shape feature extraction is usually based on contour or region, the area of super pixel is represented by the number of pixels in the region, and the shape of super pixel is described by rectangular degree. (3)Position feature extraction, including direction feature and height feature extraction, takes the result of less direction error after five iterations as directional feature. The maximum pixel distance between the super-pixel block and the lowest position pixel and the minimum pixel distance average are used as the height of the actual object and the ground, where the direction diagram is shown in figure 7;



(a) (b)



(c) (d)

Fig. 7. Directional diagram of different samples

(4)The feature extraction of texture elements reflects the arrangement properties of the structure of the object surface structure, the extraction method and the four feature types, as shown in Table 6.

Tab. 6 Feature of texture primitives and classification of extraction methods

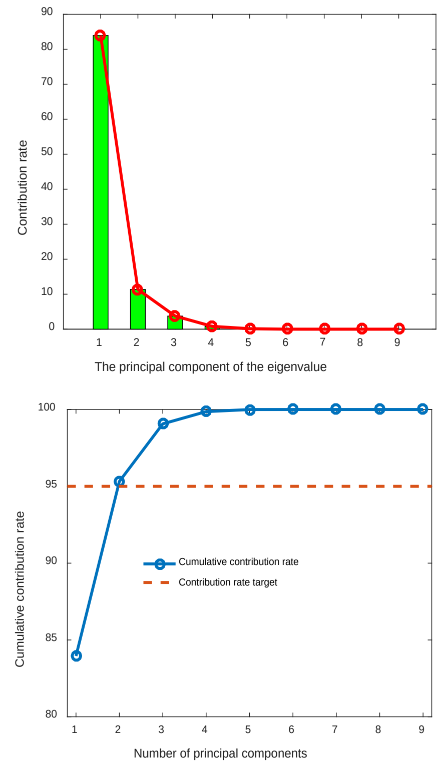
Four texture features	Extraction method
Statistical type Model type	Direct extraction from image domain
Signal processing type Structural type	Extraction from transform domain

C. Constructing hyperpixel semantic annotation pool and training classifier

The semantic annotation framework takes pixels or super-pixels as tagging primitives and uses region-level annotation methods for semantic segmentation. The so-called region-level is the super-pixel region. NYUD2 dataset formed by image segmentation and as the international authority RGB-D indoor scene data set. With dense benchmark RGB and depth annotation information, a model with good generalization performance can be obtained after training, and the annotation quality is shown in Table 7 and figure 8.

Tab. 7 NYUD2 dataset semantic annotation quality

Number of indoor scenes	Structural semantic category label	Core semantic tag	Dense label coverage
464	4	12	55.3%



(a) (b)

Fig. 8. Calculation results of the CNN model

By using super-pixel class labels and feature vectors, each item of super-pixel can be integrated into the whole image to form a super-pixel semantic annotation pool corresponding to the super-pixel set, and it is used as a training sample to train the random forest classifier. The class probability is assigned to each super pixel to determine which specific scene object the super pixel belongs to, so as to complete the object detection task.

V. CONCLUSION.

To sum up, through the experimental verification of image segmentation algorithm in object detection, we can conclude that the segmentation algorithm using RGB information and depth information has a high segmentation accuracy. Therefore, the reliability of object detection results can be improved. On the basis of this, by using multi-feature semantic labeling algorithm and training classifier, the object recognition and detection task can be completed, and the method is more feasible.

REFERENCES

[1] Wang J H. “Feasibility study of the detection of chlorpyrifos residuals on apple skin based on infrared micro-imaging”. *Optical Engineering*, 2012, 51(10), pp. 2002-2009.

[2] Kain J S, Coniglio M C, Correia J, et al. “A feasibility study for probabilistic convection initiation forecasts based on explicit numerical

guidance". *Bulletin of the American Meteorological Society*, 2013, 94(8), pp. 1213-1225.

[3] Mathews D M, Cirullo P M, Struys M M, et al. "Feasibility study for the administration of remifentanil based on the difference between response entropy and state entropy". *British Journal of Anaesthesia*, 2007, 98(6), pp. 785-791.

[4] Papaioannou A, Kennedy C C, Campbell G, et al. "A team-based approach to warfarin management in long term care: A feasibility study of the MEDeINR electronic decision support system". *Bmc Geriatrics*, 2010, 10(1), pp. 1-10.

[5] Dsouza R, Subhash H, Kai N, et al. "Feasibility study of phase-sensitive imaging based on multiple reference optical coherence tomography". *Chinese Optics Letters*, 2017, 15(9), pp. 090007.

[6] Petrinic I, Korenak J, Povodnik D, et al. "A feasibility study of ultrafiltration/reverse osmosis (UF/RO)-based wastewater treatment and reuse in the metal finishing industry". *Journal of Cleaner Production*, 2015, 101, pp. 292-300.

[7] Espinoza A, Beeksma B, Petasecca M, et al. "The feasibility study and characterization of a two-dimensional diode array in "magic phantom" for high dose rate brachytherapy quality assurance". *Medical Physics*, 2013, 40(11), pp. 111702.

[8] Zheng W L, Shen S C, Lu B L. "Online Depth Image-Based Object Tracking with Sparse Representation and Object Detection". *Neural Processing Letters*, 2016, 45, pp. 1-14.

[9] Agyenim F, Hewitt N, Eames P, et al. "A review of materials, heat transfer and phase change problem formulation for latent heat thermal energy storage systems (LHTESS)". *Renewable & Sustainable Energy Reviews*, 2010, 14(2), pp. 615-628.