A GBVS Based Object Recognition Algorithm

Faxu Wang
School of Electronics and Information Engineering
Tongji University
Shang Hai, China
E-mail: wangfawd@126.com

Zhuping Wang
School of Electronics and Information Engineering
Tongji University
Shang Hai, China
E-mail: wangzp@126.com

Xiaonian Wang *
School of Electronics and Information Engineering
Tongji University
Shang Hai, China
E-mail: wangxiaonian@126.com
* Corresponding Author

Zhu Jin
School of Electronics and Information Engineering
Tongji University
Shang Hai, China
E-mail: zhujin@tongji.edu.cn

Abstract—Object recognition has always been regarded as hot research area in image field; especially there are many research difficulties of recognition in multiple objects scene. This paper puts forward an object recognition algorithm based on hierarchical decomposition GBVS. Firstly, get saliency map of current scene according to GBVS algorithm, and acquire multi focused areas according to its multiple peak features. Then, obtain object and scene described by tree via hierarchical focusing process. Thereafter, utilize edit distance to evaluate similarities among tress so as to realize object recognition. Finally, use real scene to verify method feasibility.

Keywords-component; object recognition; GBVS; tree matching; edit distance; Region of interest

I. INTRODUCTION

Object recognition is widely applied in the fields of biomedicine imaging, intelligent transportation system, and satellite imaging etc. This paper puts forward object recognition accomplishment by taking advantage of object recognition algorithm based on GBVS for multiple objects scene with premise condition in which object to be recognized is known. When observing a scene and seeking an object, there will always be some areas drawing our attention, and the recognition is believed to be the process from the course to the fine. In order to simulate human being visual model, significance analysis algorithm is introduced in object recognition process.

Object recognition is considered to be basic research, and there have been many recognition algorithms for two-dimension images: algorithms based on appearance model include sift/surf [1] [2], haar [3], and generalized though conversion [4] which all utilize grayscale image. Sift / surf and haar extract local features of images, whereas Hough conversion is for global features. Recognition algorithms based on context include [5] and [6], which combines object scene besides object features. Object recognition algorithms by using traditional algorithm to extract image features depend excessively on object appearance, and lack the representation of structure feature on object nature.

Visual Attention Theory supposes that human being visual system only processes a certain local part of the image in detail not the whole image. In early stage, Koch, Unman [7], and Itti [8] proposes that visual attention mechanism can be divided into two stages: quick, subconscious, bottom-up, and data driving significance extraction, slow, task dependent, up-bottom, and visibility driving significance extraction. Extracted significance image can be widely applied into many computer visual fields, which include interested object image segmentation [9] [10], object recognition [11], and image retrieval [12].

Current significance analysis algorithms can be generally divided into three categories: the first category is based on low level visual features, for example significance algorithm put forward by itti as mentioned above; the second category is not based on low level visual feature, for example, AC algorithm [13] and SR algorithm [14] make use of pure mathematics for calculation; the third category combines the preceding two algorithms, for example GBVS algorithm based on graph theory. Significance analysis algorithm extracts the most appealing part of the image, but during object recognition process, the object to be recognized only appeals attention compared with surrounding areas and the whole image may not be the most appealing part. Therefore, significance analysis algorithm needs to be improved. GBVS algorithm simulates visual theory during feature extraction process, introduces Markov chains during significance image generation process, calculates significance value, and maximum global value is significance area. This paper proposes all the maximum local values as appealed area, that is, all the areas appeals more attention compared with surrounding environment in the image. This is regarded as image description basis, and coordinates with tree matching algorithm [15] to accomplish object recognition.

This paper puts forward a new object recognition algorithm in image: the first step takes advantage of significance analysis algorithm to decompose the object, and then form object tree; the second step makes use of significance analysis algorithm to decompose scene image,

and then form scene tree; the third step analyzes via tree matching algorithm, and if scene tree includes part of sub-trees similar to object tree, object recognition is completed, otherwise, the scene does not include the object.

II. HIERARCHICAL DECOMPOSITION METHOD

This paper extends traditional GBVS algorithm into multiple objects environment. As illustrated in Figure 1(a), scene image includes three red boxes, and upper red box is the object to be recognized. Use GBVS algorithm to calculate image significance as illustrated in Figure 1(b): coordination in the bottom corresponds to pixel space position; vertical axis is significance value of corresponding pixel, and the larger the axis is, the stronger the significance in this point becomes. The highest peak is blue-green decoration pattern box, and the small peak on the left side is dark red box to be recognized. It can be concluded that although the object to be recognized is significant area in the local, it is not the most significant area in the global.

This paper names local maximum value and its surrounding area in significance image as focused area. As illustrated in Figure 2 0 generation (b), extract three peak values and three focused areas represented by surrounding areas in scene, and then form 3 subareas, thereafter, extract respectively focused area from each subarea. The first generation (a) is decomposed into four focused areas, and the first generation (c) is decomposed into two focused areas, and the first generation (e) is not decomposed. The color of each decomposed subarea is simple without continuous decomposition capability. This forms tree structure regarding the whole image as root node, according to focus area decomposition, and representing leaf property by 256-dimension space color histogram.

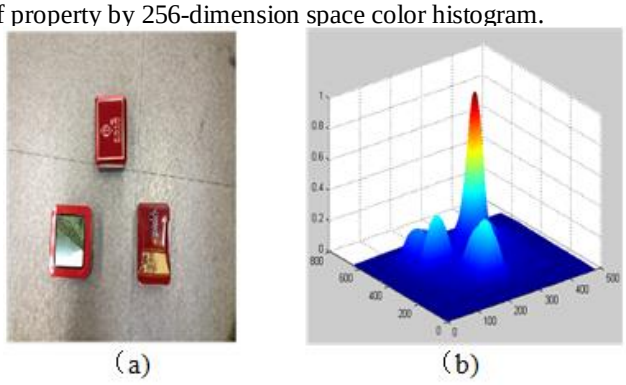

Figure 1. Scene and scene multiple peaks.

Hierarchal focus area accomplishes tree description on scene image. The object to be recognized also adopts the same method to form tree description. Compare the two trees, if there exists the sub-tree similar to object tree in scene tree, a block of focus area constructed by similar subtree is the position of object to be recognized and realize object recognition.

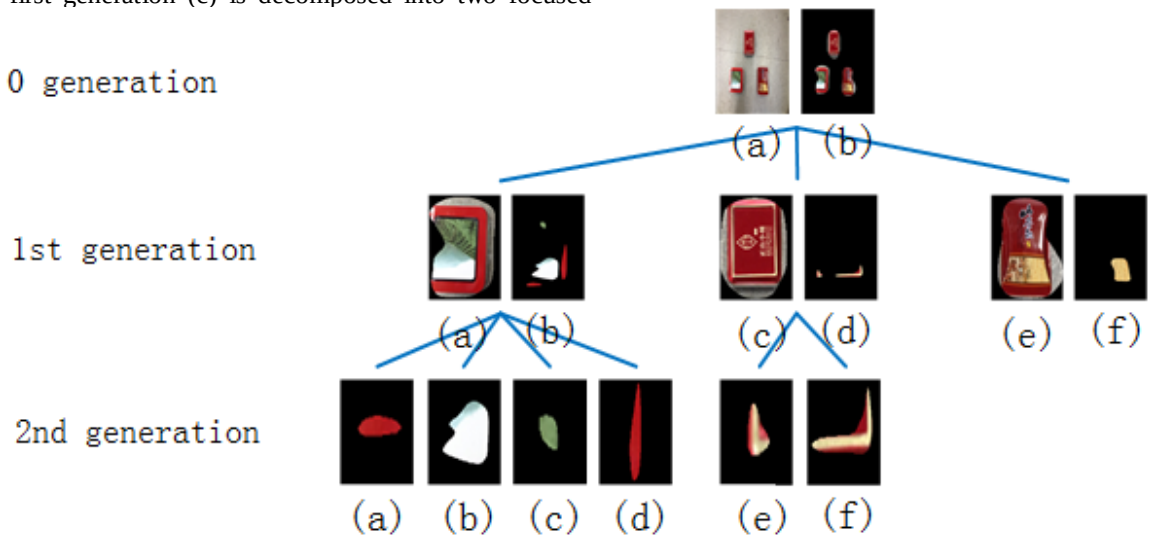

Figure 2. Scene tree.

III. TREE MATCHING ALGORITHM

Adopt GBVS algorithm to get hieratically get appealed area of image to form image tree description. The tree formed by the object to be recognized is named as object tree, and the tree formed by the scene is named as scene tree. Convert object recognition in multiple objects scene into the problem of similarities among trees.

Leaf feature is represented by 256-dimension HSV space color histogram. Because focused area levels are different and pixel amount differences are large, the types including colors are focuses and pixel amount of each color is not cared, and set all the positive values in color histogram to be 1. Suppose that histogram vector of leaf X is (x1, x2 …x256) and leaf Y is (y1, y2…y256). Leaf distance d (X, Y) refers to Euclidean distance between two vectors, which is defined as Formula (1). When leaf distance is less than or equal to set threshold, it is considered that they are the same leaves; when leaf range is more than set threshold, it is considered that they are different leaves.

$$d(X,Y) = \sqrt{\sum_{i=1}^{256}(x_i - y_i)^2} \qquad (1)$$

Edit distance is often used to evaluate the similarities between two trees. Suppose two trees T, and T', edit distance refers to the sum of all the edit operations experienced for the T to be clipped so as to be the fully same as T'. Use T[i] to represent T leaf, and use T'[j] to represent T' leaf, and Λ represents blank leaf. The edits include: 1. Leaves on the T and T' with the same positions and different properties, change corresponding leaf property on T, which is represented as T[i]→T'[j]; 2. Leaf T[i] on the T and there is no corresponding leaf on the position of T', delete T leaf, which is represented as T[i]→ Λ ; 3. There is no such leaf on T as the leaf T'[j] on corresponding position on T', and insert the leaf on T, which is represented as Λ →T'[j]. After traverse on all the leaves on the two trees, conduct the preceding operations to make T the same as T'. If the clip from one tree to another tree is defined as M.

$$cost(M) = \sum_{(i,j)\in M} r(T[i] \to T'[j]) + \sum_{i\in I} r(T[i] \to \Lambda) + \sum_{j\in J} r(\Lambda \to T'[j]) \quad (2)$$

Tree edit distance can effectively represent the similarities between two trees: the less the value is, the more similarities the two trees have; when is 0, it means the two trees are fully the same. In object recognition of multiple objects scene, it cannot directly calculate edit distance of scene tree and object tree. This is because of the fact that the scene includes many objects and only one tree may be similar to the object. Only when edit distance between sub-tree and object tree is less than threshold and it is the minimum in all the sub-trees can sub-tree scope be confirmed as the object to be recognized.

In order to find possible sub-tree as quick as possible, this paper adds a step before calculating edit distance between trees: fast screen scene tree leaf. Because it may exist replacement lack when focused area is divided, all the leaves except root nodes should be traversed. Start from the first generation of leaves of scene tree, compare the leaves and root nodes of object tree one by one: if leaf property is similar, it is regarded as suspected sub-tree root node. After the selection of all the suspected sub-trees is completed, calculate tree edit distance on sub-tree and object tree one by one, and based on this judge whether there is the sub-tree similar to object tree and accomplish object recognition.

Tree matching algorithm steps:

Step 1: traverse scene tree leaves, and respectively calculate leaf distance of root node in object tree, which is shown in Formula (1); select all the leaves with less than 8 leaf range.

Step 2: regard selected leaf as root node, and include all its leaves to form suspected tree.

Step 3: traverse suspected tree leaves and object tree leaves, and calculate leaf distance, and the minimum value in all the leaves with less than 8 distance is considered as similar leaf.

Step 4: modify, add, or delete suspected tree based on the judgment whether the leaf is similar, and calculate its edit distance with object tree. The leaf with the value less than threshold and smallest distance is regarded as the object to be recognized.

IV. EXPERIMENTAL RESULT

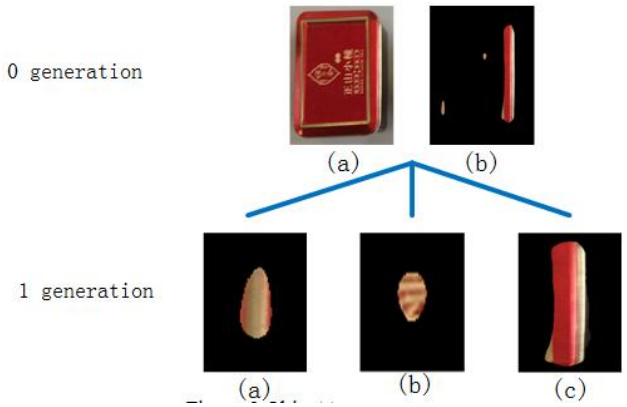

Figure 3. Object tree.

Respectively calculate its leaf distances with all the leaves in scene tree, and get the result shown in Table 1, 1(a) represents scene tree the first generation (a) in Figure 1. Seen from the result, the follows are get due to the fact that results of 1(c), 1(e), 2(e), and 2(f) are less than threshold: suspected "tree 1" regards 1(c) as root node; suspected "tree 2" regards 1(e) as root node; suspected "tree 3" regards 2(e) as root node; alternative "tree 4" regards 2(c) as root node.

TABLE I. LEAF DISTANCE BETWEEN OBJECT ROOT NODE AND SCENE TREE LEAF.

Scene tree	1(a)	1(c)	1(e)	2(a)	2(b)	2(c)	2(d)	2(e)	2(f)
Distance	9.54	**7.42**	**7.87**	8.94	11.0	10.7	8.94	**7.14**	**7.55**

Respectively calculate leaf distances of three leaves and all the suspected leaves.

TABLE II. LEAF DISTANCES OF OBJECT LEAF AND SUSPECTED LEAVES.

Scene tree / Object tree	2(e)	2(f)	1(f)
1(a)	6.4807	6.1644	5.4772
1(b)	6.5574	5.5677	5.3851
1(c)	5.6568	5.8309	7.8740

Calculate edit distance between suspected tree and object tree, and judge which tree is the object to be recognized.

TABLE III. EDIT DISTANCE BETWEEN OBJECT TREE AND SUSPECTED TREE.

	tree 1	tree 2	tree 3	tree 4
Modify leaf	0	0	0	0
Add leaf	1	2	3	3
Delete leaf	0	0	0	0
$cost(M)$	**1**	2	3	3

Where, edit distance between suspected "tree 1" and object tree is less than threshold and is the minimum value in all suspected trees, therefore, the area represented by the sub-tree regarding the first generation (c) as root node. The test verifies that object recognition algorithm proposed by this paper accomplishes the recognition on the object to be recognized and the removal on close objects in multiple objects scene.

In order to verify the method proposed by this paper design verification test as the follows: respectively take photos of multiple objects scene and the object to be recognized, and recognize the object in scene after the image is processed. The scene is illustrated as Figure 1, and utilizes the method of extracting focused area proposed in Chapter 2 to decompose, and then form scene tree shown as Figure 2; the object to be recognized is illustrated as Figure 3 0 generation (a), process object image, and extract focused area, and get three subareas illustrate as0 generation (b). Decompose three subareas and form the first generation, which is illustrated as in Figure 3 the first generation (a), (b), and (c), and three subareas cannot be further decomposed, and this accomplishes processing on the object to be recognized. Figure 3 is object tree diagram formed by the object to be recognized.

ACKNOWLEDGMENT

In this paper, the research is sponsored by National Natural Science Foundation of China 91420103, 61473209.

REFERENCES

[1] Liang Dong and Yan Pu, "Spectral matching algorithm based on nonsubsampled contourlet transform and scale-invariant feature transform." Systems Engineering and Electronics, Journal of Volume: 23, Issue: 3, pp. 453 – 459, 2012.

[2] Wang Kai, Cheng Bo and Ma Lu, "Multi-source Remote Sensing Image Registration Based on Normalized SURF Algorithm.", Computer Science and Electronics Engineering (ICCSEE), pp.373 – 377, 2012.

[3] Sevcenco, I. S.,Hampton, P. J. and Agathoklis "Seamless stitching of images based on a Haar wavelet 2D integration method" Digital Signal Processing (DSP), pp.1 – 6, 2011

[4] Chun-Pong Chau and Wan-Chi Siu, "Generalized dual-point Hough transform for object recognition Image Processing" ICIP 99. pp 560 - 564 vol.1,1999.

[5] Karlinsky Leonid and Dinerstein M. "The chains model for detecting parts by their context." 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 25 – 32, 2010

[6] Yao Bangpeng and Li Fei-Fei., "Modeling mutual context of object and human pose in human-object interaction activities." 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 17 – 24, 2010

[7] Koch, C. and Ullman, S., "Shifts in selective visual attention: towards the underlying neural circuitry", Human neurobiology, 4(4), pp.219, 1985

[8] Itti, L., Koch, C. and Niebur, E., "A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11), pp.1254-1259, 1998.

[9] Han, J., Ngan, K., Mingjing Li. and Hongjiang Zhang, Unsupervised extraction of visual attention objects in color images [J]. IEEE Transactions on Circuits and Systems for Video Technology, 16(1), pp.141-145, 2006.

[10] Ko, B. and Nam, J., Object-of-interest image segmentation based on human attention and semantic region clustering [J]. Journal of Optical Society of America, 23(10), pp.2462-2470, 2006.

[11] Rutishauser, U., Walther, D., Koch, C. and Perona, P., Is bottom-up attention useful for object recognition? [C]. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.II-37 - II-44 Vol.2, 2004.

[12] Kashif Iqbal, Michael O. Odetayo and Anne James., Content-based image retrieval approach for biometric security using colour texture and shape features controlled by fuzzy heuristics [J]. Journal of Computer and System Sciences, 78 (4), pp.1258-1277, 2012.

[13] Achanta, R., Estrada, F., Wils, P. and Susstrunk, S., Salient region detection and segmentation. International Conference on Computer Vision Systems, 5008, pp.66-75, 2008.

[14] Xiaodi Hou and Liqing Zhang, Saliency Detection: A spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition, pp.1-8, 2007.

[15] Harel, J., Koch, C. and Perona, P., Graph-based visual saliency. Journal of the ACM (JACM), pp.545-552, 2007..