A New Hybrid PCNN for Multi-Objects Image Segmentation

Zhenbo Li[1], Yu Jiang[1#], Jun Yue[2], Jingjing Fang[1], Zetian Fu[1], Daoliang Li [1*]

[1]College of Information and Electrical Engineering, China Agricultural University, Beijing, 100080
*Corresponding author, E-mail: zhenboli@gmail.com Tel: +86-10-62738751
[#]Yu Jiang is co-first author of the paper
[2]College of Information and Electrical Engineering, Ludong University, Yantai, 264025

***Abstract*— Many image based applications such as multi-object tracking were nagged by the problem of robust multi-objects image segmentation. In this paper, we propose a new hybrid Pulse Coupled Neural Network (PCNN) method for multi-object segmentation. Firstly, we use saliency detection methods, Graph-based visual saliency (GBVS) and Spectrum Residual (SR) to find more accurate object region (R1) and more number of object regions (R2) separately. Then an improved PCNN is used to work out the multi-objects with R1 and R2. The statistical result of R1 is selected as an adaptive generator threshold of PCNN and a selection standard of segmentation result. R2 determines the correct object number in the image. Experiments of images selected from BSD and VOC and two full image datasets (MSRC v2 and Weizmann) prove that our method can get more right object quantity and more accurate object region than GBVS-PCNN[1] and adaptive PCNN[2].**

I. INTRODUCTION

Image segmentation is a fundamental step for advanced tasks of image processing such as object recognition, object classification, image fusion and tracking etc. In object recognition and classification, particularly in medical image processing, many researchers discovered and classified the object by using image segmentation. Also some researchers studied image segmentation to improve video tracking performance.

For many image and video applications, people propose different means to segment multiple objects in images. Threshold-based algorithms, due to simplicity and fast speed, are widely considered to extract different objects in the image by using multiple thresholds [11] [12]. Some researchers think of more low level information like edge, texture to improve the accuracy of multi-objects segmentation [13] [14][15] . So as Graph Theory developed in recent decades, some cuts methods, contain Normalized Cuts [10], Collect Cut [16], are employed by considering the images as a graph.

However, traditional segmentation methods, such as threshold-based, edge-based, graph-based or mixed approaches, to some extent all troubles of inaccurate segmentation, especially of multi-object segmentation. Recently, with developing of artificial neural networks, more and more new strategies were proposed to solve the multi-objects segmentation and to use for many applications preferably[1][2][3][14]. But, for multi-objects segmentation, there are still remain two problems to solve:

a. Correct number of objects in the image
b. Accurate extraction of every object

In this paper, we proposed a new hybrid PCNN algorithm (Fig.1) to challenge these problems. For the first contribution of the paper, we show a saliency detection strategy, combined with GBVS[5] and SR[6] method, which can carry out accurate object region and correct number simultaneously. For the second contribution, we proposed the method to extract every object we detected in the image accurately using an improved PCNN with a new generator mode.

The remainder of the paper is organized as follows. In Section 2, we discuss the related works on saliency detection and segmentation. Section 3 demonstrates the mixed saliency method and the improved PCNN respectively and how they work together. Detailed experimental results for the problem of natural images multi-object segmentation are provided in Section 4. The conclusions are given in section 5.

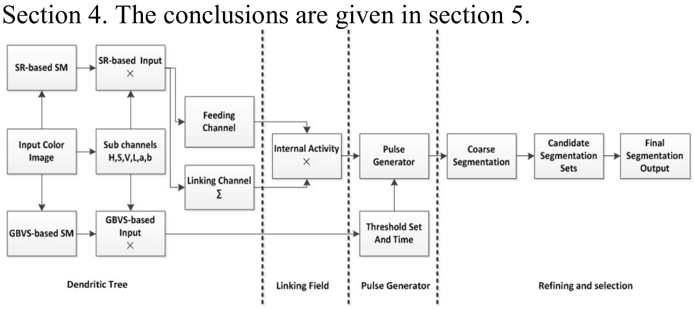

Fig. 1 structure of our approach

II. RELATED WORKS

We exploit PCNN approach to multi-objects segmentation. The PCNN approach offers a way to combine the object information and regions, obtained by saliency detection in the images to extract all the objects we are interested of. In this section, we review the related works on PCNN and saliency detection methods.

A. *Saliency Detection*

Saliency region detection is inspired by human visual system (HVS) and rapidly becomes a crucial part of computer vision system. A number of relative works will be demonstrated in this lecture.

In 1998, Itti and his colleagues proposed a bottom-up selective visual attention models (SVAM) [8] which uses local feature contrast maps to carry out the saliency map (SM). Then Harel et al. [5] introduced a graph-based visual saliency model (GBVS or GBVSM) which uses graph theoretical normalization function to perform detection by using feature maps computed from Itti model in 2006. The GBVS model only adds a bit of computation complex and obtains preeminent performance, achieving 98% of the ROC area of a human-based control, whereas the classical algorithms of Itti achieve only 84%. Bruce et al. [4] applied information theory to improve SVAM and gained generic high quality performances.

In 2007, Hou et al. [6] proposed a new framework which completely different from Itti model – Spectrum Residual method (SR). Hou defined an image as a sum of two parts, Innovation and Prior Knowledge. The Innovation part denotes novelty and the Prior Knowledge part denotes redundancy. Then saliency map is computed from spectral residual, the Innovation part subtracted by Image and Prior Knowledge part. Experimental results show that this model is so simple but more efficient. In 2008, Hou et al. [7] introduced the Incremental Coding Length (ICL) to measure the perspective entropy gain of each feature. ICL model achieves superior accuracy in static saliency map generation and can capture less-reported dynamic visual search behaviors.

Recently, Liu et al. [1] presented an improved automatic hybrid model. A PCNN is exploited to extend traditional GBVS model. Meanwhile GBVS model and OTSU method determine iteration time and generation threshold of PCNN respectively. And other parameters of PCNN are set by new structure automatically. This improved hybrid model and showed more effective in saliency region detection accuracy than eight state-of-the-art SVAM approaches.

B. *Pulse Coupled Neural Network*

In the late 1980s, Eckhorn et al. [17] introduced the mammal neuron model during their study on the synchronous oscillation phenomenon in the visual cortex of cats. This model is further developed by Johnson [18], according to its ability to cause the adjacent neurons with similar inputs to pulse synchronously, finally evolved into PCNN. It is soon recognized as having significant potential in image processing problems. Because of a lot of parameters needed to determine, a number of works are applied to simplify the vital model and to set the parameters automatically.

Gu et al. [19] firstly brought forward a PCNN based on unit-linking that can use same model parameters to segment different kinds of images automatically and efficiently. This new strategy lets PCNN easier and more practical to use than original without choosing the enormous nagged parameters.

Chen et al. [3] presented a new automatic parameter setting method of a simplified PCNN. This method successfully determines all the adjustable parameters in SPCNN and does not need any training and trials as required by previous methods. They derived the general formula of dynamic properties of neurons and deduced the sub-intensity range expression in order to achieve their goal. The experimental segmentation results of the gray natural images from the Berkeley Segmentation Dataset, rather than synthetic images, prove the validity and efficiency of this automatic model.

Wei et al. [2] introduced PCNN model with adaptive threshold decay time constant. In his paper, authors utilized HVS theory to calculate a suitable threshold decay time constant of every image. They also designed a new mode of image entropy and selected the best result, as the final extraction output, from the entropy set of time series.

Recalling these related works on saliency region detection and PCNN model, we have come up with a train of thought on automatic multi-object segmentation. According to HVS theory, we combine the saliency region detection and PCNN to solve the problems mentioned in Section 1. Although Liu's work is similar to ours in structures, we will demonstrate the essential differences in section 3 and show our algorithm more effective in section 4.

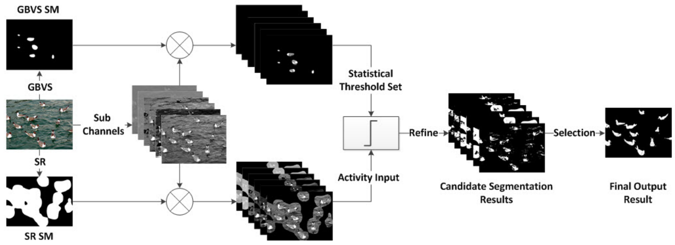

Fig. 2 the procedure of our method

III. MULTI-OBJECTS IMAGE SEGMENTATION MODEL

In this section, the two original improvements of our approach are introduced to detail the structure and workflow of proposed hybrid model (see Fig.1 and Fig.2). One is the correct decision of object numbers, the other is the accurate segmentation of every detected object. The former improvement reduces the region where handled by PCNN and the later makes it possible to segment automatically. When a person get the information from his eye, the focused objects can invoke more pulse than the background, namely we neglect the background signal. To simulate this phenomenon, we will use the saliency region detection methods to extend traditional simple PCNN. Some pre-experiment results, systemized in Table 1, help us to determine the strategy of saliency region detection.

Table 1 the effectiveness with respect to object number correction and object region accuracy

	SR	75% GBVS	95% GBVS	98% GBVS	99% GBVS
ONCR	88%	70%	56%	–	–
ORAR	65%	75%	80%	87%	78%

In table 1, ONCR means the rate of object number correction and ORAR means the rate of object region accuracy, as described in Eq. 1 and Eq. 2 respectively. The statistical result of experiments shows that 98% GBVS is more effective for ORAR and SRSM is used for ONCR better.

$$ONCR = correct\ image\ number / total\ image\ number \tag{1}$$

$$ORAR = contain\ accurate\ \ feature\ image\ number \\ /total\ image\ number \tag{2}$$

Where correct image refers to that has same detected object number with ground truth; contain accurate feature indicates that the saliency region associated with input image contains main feature of object.

A. Dendritic Tree

For a given input color image $I(x, y)$, firstly, it is processed by GBVS model and SR model, and GBVS-based saliency map (GBVSSM) and SR-based saliency map (SRSM) are gained by respectively. According to our experiments, GBVS-based mask and SR-based mask are work out by 98% and 50% of SM separately (see Table 1 and [1]). GBVS-based mask establishes the accurate object region and the SR-based establishes the correct object number. Then SR input (SRI) and GBVS input (GBVSI) are generated by using Eq. 3 and divided into six sub-channels (H, S, V, L, a, b) subsequently. Each of pixels in every channel (H, S, V, L, a, b) is considered as external stimulus of the PCNN, namely the feeding input (F). The linking input (L) is computed from the external stimulus neighborhoods $(SRI_{k,l})$ via linking synaptic weight W. Eq. 4 and Eq. 5 describe these separately:

$$SMI_c = I_c \times SM \tag{3}$$

$$F_c[1] = SMI_c \tag{4}$$

$$L_c[1] = \sum_{k,l \in N(i,j)} SMI_{c,k,l} \times W \tag{5}$$

Where SM is the GBVS-based or SR-based saliency map; c is the channel index; $[1]$ denotes that this channel has no change in each of iterations; $N(i, j)$ refers to the eight neighborhoods of $SMI(i, j)$; W is the linking weight matrix, which decided by the [9]'s method, i.e.,

$$W = [0.5\ 1\ 0.5;\ 1\ 0\ 1;\ 0.5\ 1\ 0.5] \tag{6}$$

Then, in the linking field of PCNN, the feeding channel and linking channel are united like traditional SPCNN, as described in Eq. 7:

$$U_c = F_c \times (1 + \beta \times L_c) \tag{7}$$

Where U is the internal active neurons. β is the internal linking factor, indicates how strongly a neuron is affected by its neighborhood. [9]'s method computes it as Eq. 8:

$$\beta = 1 / \sqrt{1 + STD_{i,j}} \tag{8}$$

Where $STD_{i,j}$ is the standard deviation of F value of the eight neighborhoods.

Consequently, we get six U sets of each sub-channel of input image I.

B. Pulse Generator

In our method, we apply a new generator of PCNN to improve its performance. Firstly, we count the number of each pixel value in the region of per sub channel detected by GBVS. Consequently, the top k percent of the number of pixels are selected as pulse threshold set (TS) and the quantity of types of them is the iteration time of PCNN model. In each iteration, the pulse threshold is determined by a relative pixel value, as described in Eq. 9:

$$\theta_i = TS_i \tag{9}$$

$$Y_c = \begin{cases} 1, & U_c(x, y) = \theta_i \\ 0, & otherwise \end{cases} \tag{10}$$

Where i indicates the i th iteration of PCNN. θ is the pulse threshold. Y is the coarse segmentation result of input image.

Hitherto, most of important parameters of the new hybrid PCNN, such as iteration time and pulse threshold in each iteration, are set automatically. The further development of PCNN is constrained by the nuisance of parameter setting no matter manually or from a prior training [1][3]. The self-adaptive statistical thresholds and automatic region shrinkage strategy are more suitable and efficient than other self-adaptive methods (see Section 4).

C. Refining and Selection Standard

As shown in Fig. 1, an automatic refine and selection algorithm are proposed in this part. With the coarse segmentation result, firstly, we should refine the object part, preliminary masked by SRSM, to gain intact object as possible as we can. If it has similar pixel value to its eight neighborhoods, the masked part will be outputted like the most state of neighborhoods, as described in Eq.11. The above step will be executed until the output has no changing, consequently, a candidate result is carried out by adding the two coarse results. Then another vital step is the selection of output from six sub-channel outputs. A containing scores (CS) and deviation (MSD), between mean pixel value of image and statistical threshold set, are calculated by Eq. 12 and Eq.13 respectively. Then the outputs of sub-channel which has lower CS and higher MSD will be chosen as the final segmentation result.

$$Y_c(i, j) = Y_c(k, l), \\ (k,l) = \begin{cases} (k,l), & \min abs(U_c(k,l) - \theta) < \theta_r \\ (i, j), & otherwise \end{cases} \tag{11}$$

$$CS_c = \sum_{i=1}^{it} \sum (U_c(x, y) - TS_i)^2 \tag{12}$$

$$MSD_c = \sum (I_c(x, y) - \overline{I}_c)^2 \tag{13}$$

Where $k, l \in N(i, j)$. θ_r is the threshold of refining. it denotes the iteration number. $\overline{I}_c$ is the average value of sub-channel c of input image I.

IV. EXPERIMENTS AND DISCUSSION

In order to evaluate our model quantitatively, the following experiments are processed. The full of Weizmann 2-obj Segmentation Database [20], MSRC v2, and some selected images of VOC 2011 and Berkeley BSD 500 [21], totally 880 images, are used to testify the effectiveness of ours. We also compared the hybrid model with other self-adaptive PCNN method, Wei's [2] and Liu's [1], in two measurements, correct number of object and overlap scores. The representative results are showed in Fig.3, Fig. 4 and Fig. 5. We set k as 30% and θ_r as 0.05.

Table 2 the statistical results of ONCR and OS

	ONCR on MSRC v2	ONCR on parts of VOC and BSD 500	ONCR on Weizmann	OS on MSRC v2	OS on Weizmann
Ours	95.65%	93%	88%	0.5939	0.6838
Liu's	43.48%	50%	53%	0.5013	0.3938
Wei's	43.48%	50%	45%	0.1414	0.3651

Fig. 3 the representative results come from MSRC v2, VOC 2011, BSD500. **(a)** is the original color image; **(b)**, **(c)** and **(d)** are the results of ours, Liu's and Wei's respectively.

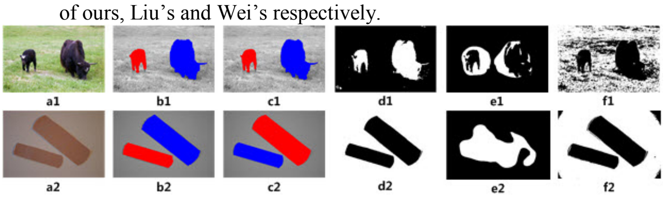

Fig. 4 the representative results with two GTs come from Weizmann Database. **(a)** is the original image; **(b)** and ***(c)*** are GTs labeled by different person; **(d)**, **(e)** and **(f)** are results of ours, Liu's and Wei's respectively.

A. The number of objects in image

For the multi-objects segmentation, the correct number of objects is the crucial aspect. We count the number of extracted objects in each results processed by three methods ours, Wei's and Liu's and consequently compare them with the ground truth. The statistical result shows in Table 2. As illustrated in Table 2, our method can not only mostly segment all objects in the images, but also captures accurate part of objects.

The number of interested objects, in most images, is carried out by our model correctly. Especially in a3-a6 of Fig. 3, a2, a6, a8 of Fig. 5, due to the new dendritic tree strategy, our model can calculate the right number while Liu's and Wei's cannot. For the eminent improvement, the crux is that the SR method is more useful than GBVS in the decision of object number. But if the saliency region of SR is wrong, the result of ours is invalid, just like a1, a3, a4 and a7 of Fig. 6. The reason is that SR is incorrect.

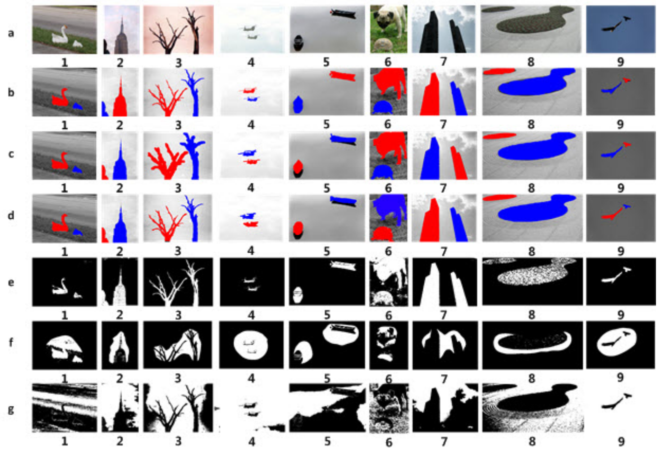

Fig. 5 the representative results with three GTs come from Weizmann Database. **(a)** Original image; **(b)**, (c) and **(d)** are the GTs labeled by different person; **(e)**, **(f)** and **(g)** are the results of ours, Liu's and Wei's separately.

Table 3 the OS score of images in Fig. 5

	a1	a2	a3	a4	a5	a6	a7	a8	a9
Ours on GT a	0.55	0.78	0.65	0.47	0.79	0.68	0.91	0.69	0.51
Ours on GT b	0.56	0.79	0.62	0.49	0.55	0.70	0.90	0.71	0.62
Ours on GT c	0.57	0.78	0.65	0.47	0.49	0.69	0.92	0.72	0.61
Liu's on GT a	0.43	0.41	0.35	0.02	0.69	0.39	0.32	0.48	0.62
Liu's on GT b	0.43	0.42	0.44	0.02	0.46	0.42	0.31	0.49	0.53
Liu's on GT c	0.43	0.42	0.35	0.02	0.36	0.42	0.32	0.49	0.63
Wei's on GT a	0.08	0.47	0.27	0.24	0.11	0.27	0.43	0.40	0.49
Wei's on GT b	0.08	0.49	0.35	0.28	0.08	0.29	0.42	0.42	0.60
Wei's on GT c	0.08	0.48	0.27	0.27	0.07	0.29	0.43	0.42	0.59

B. Overlap Scores

To quantify accuracy, we use the pixel-level segmentation overlap score, OS. The quality of segmented region with respect to the ground-truth object segmentation is measured as Eq. 14:

$$OS = \frac{|GT \cap R|}{|GT \cup R|} \tag{14}$$

Where we take as GT the full object region associated with region R's majority pixel; GT denotes the segmentation regions handled by human manually; R is the segmentation region of ours. The higher OS the output gains, the better effective and accurate it is. The quantity results with respect to the images are shown in Table 2.

As depicted in Table 2, the high average OS scores of ours, whether in MSRC v2 or Weizmann, have proven that the new pulse generator is efficient. Particularly, in a7 of Fig. 5, the OS score is more than 90%. But if the background is so simple or the GBVSSM is not accurate, our model is surpassed by others, i.e. a9 of Fig. 5.

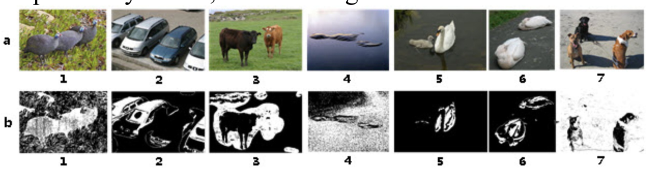

Fig. 6 some incorrect results carried out by our model. (**a**) original image; (**b**) result of ours.

V. CONCLUSIONS AND FUTURE WORK

In this paper, a new hybrid PCNN model is proposed, where the saliency region detection methods SR and GBVS assist PCNN with all the object regions and optimal self-adaptive selection of PCNN parameters. Experimental results have proven that this strategy, inspired by HVS theory, solves both of the two nettlesome but crucial problems, correct number of objects in the image and accurate extraction of every object.

However, some images in Fig. 6 are not figured perfectly due to several unsettled aspects. One is that objects in the image have different pixel values, the other one is the ineffectiveness of saliency region detection method. To solve these problems, in future, we will try to consider the following ways: i) feature selection, not only pixel value but edge, texture and so on are thought of; ii) the new statistical strategy of pulse threshold set; iii) the new structure of PCNN is more like humans'. Meanwhile, the time cost of our model constraint its application on real-time program. Surely, to extend this model, we will attempt to use it for object recognition, classification and real-time multi-objects tracking via optimizing the algorithm.

ACKNOWLEDGEMENT

This research is financially supported by the National Natural Science Foundation of China (61100115&30971693), the National Science and Technology Support Program (2011BAD21B01 & 2012BAD35B07), the Fundamental Research Funds for the Central Universities (2012RC022), the Natural Science Foundation of Shandong Province (ZR2012FM008), Program for New Century Excellent Talents in University (NCET-09-0731) and the project of agricultural key programs for science and technology development of Ningbo (2011c11006).

REFERENCES

[1] Shangwang Liu, Dongjian He, and Xinhong Liang, An Improved Hybrid Model for Automatic Salient Region Detection, IEEE Signal Processing Letters, vol. 19, no. 4, pp. 207-210, April 2012.

[2] Shuo Wei, Hong Qu, Mengshu Hou, Automatic image segmentation based on PCNN with adaptive threshold time constant, Neurocomputing (74), pp.1485-1491, 2011.

[3] Y. Chen, S.-K. PARK, Y.Ma, and R. Ala, A new automatic parameter setting method of a simplified PCNN for image segmentation, IEEE Trans. on Neural Networks, vol. 22, no. 6, pp. 880–892, Jun. 2011.

[4] N. D. B. Bruce and J. K. Tsotsos, An information theoretic model of saliency and visual search, in Proc. WAPCV, 2007, pp. 171–183.

[5] J. Harel, C. Koch, and P. Perona, Graph-based visual saliency, in Proc. NIPS, 2006, pp. 545–552.

[6] X. Hou and L. Zhang, Saliency detection: A spectral residual approach, in Proc. CVPR, 2007, pp. 1–8.

[7] X. Hou and L. Zhang, Dynamic visual attention: Searching for coding length increments, in Proc. NIPS, 2008, pp. 681–688.

[8] L. Itti, C. Koch, and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[9] H.Wang, C. Ji, B. Gu, and G. Tian, A simplified pulse-coupled neural network for cucumber image segmentation, in Proc. ICCIS, 2010, pp.1053–1057.

[10] J. Shi and J. Malik. Normalized Cuts and Image Segmentation., IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888–905, August 2000.

[11] Zhihui Liu, Xuecheng Zou, Zhenglin Liu, A contrast enhancement algorithm based on double-threshold segmentation for LCD controller, in Cong. ISP, 2010, pp. 2434-2437.

[12] Lijuan Wang, Sha Feng, Yunxia Qu, Threshold-Based Level Set Method of Image Segmentation, in Conf. INIS, 2010, pp. 703-706.

[13] Li, J., Pan, D., Zhu, Q., Mao, Z., Edge-Guided Multi-scale Segmentation of Satellite Multispectral Imagery, IEEE Trans. on Geoscience and Remote Sensing, 2012 (in press).

[14] Chacon M, M.I, A PCNN-FCM time series classifier for texture segmentation, Annual Meeting of FIPS, 2011, pp.1-6.

[15] Melendez, J. Girones, X., Puig, D., Supervised texture segmentation through a multi-level pixel-based classifier based on specifically designed filters, IEEE Int'l. Conf. on IP, 2011, pp. 2869-2872.

[16] Yong Jae Lee, Kristen Grauman, Collect-Cut: Segmentation with Top-Down Cues Discovered in Multi-Object Images, in Proc. CVPR, 2010, pp. 3185- 3192.

[17] R.Eckhorn, H.J.Reitboeck, M.Arndt, Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex, Neural Computation, vol.2, no.3, pp. 293–307, 1990.

[18] J.L.Johnson, M.L.Padgett, PCNN model sand applications, IEEE Trans. on Neural Networks, vol. 10, no. 3, pp. 480–498,1999.

[19] X.Gu, S.Guo, D.Yu, A new approach for automated image segmentation based on unit-linking PCNN, in Proc. MLC, 2002, pp.175–178.

[20] Sharon Alpert, Meirav Galun, Ronen Basri, Achi Brandt, Image Segmentation by Probabilistic Bottom-Up Aggregation and Cue Integration, in Proc. CVPR, 2007, pp. 1-8.

[21] D. Martin, C. Fowlkes, D. Tal, J. Malik, A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics, in Proc. CV, 2001, pp. 416-423.