Recognition of the Target Image Based on POM

R.Haritha
Department of ECE
SJCET
Yemmiganur

K.Sudhakar
Associate Professor
HOD, Department of ECE
SJCET
Yemmiganur

T.Chakrapani
Associate Professor
Department of ECE
SJCET
Yemmiganur

Abstract—Image procession is the hot topic for the recent research area. Images are being used which gives more information than words in the form of test. The work of this paper is to develop a method based on perceptual organization model (POM) for boundary detection of the image. The recognition of the background objects such as the sky, the ground, and vegetation based on the color and texture information and development a perceptual organization model that can capture the non-accidental structural relationships among the constituent parts of the structured objects and, hence, group them together accordingly without depending on a priori knowledge of the specific objects. Proposed method integrates both bottoms-up and top-down information in a single grouping process. Image segmentation is used for identification of the boundary energy and the proposed method is encoded with five Gestalt laws into a boundary energy model to form a Perceptual Organization model.

Keywords: Boundary energy, image segmentation, perceptual organization model (POM), recognition.

I.INTRODUCTION

In the area of Image Processing the ongoing research work is image segmentation, which is problem to be considered in the internet. To get to know about image segmentation is that to partition an image into number of regions with coherent properties, so that each region corresponds to an object or area of interest. Here there are two categories objects, (a) structured objects (e.g., cars, buildings, people, etc.) and (b) unstructured objects (e.g., sky, roads, trees, grass, etc.). The structured objects are composed of multiple parts which has much segmentation, with each part having distinct surface characteristics (e.g., colors, textures, etc.). Unstructured objects are nothing but the backgrounds of images. The background objects have nearly homogenous surfaces and are distinct from the structured objects in images. Many recent appearance-based methods have achieved high accuracy in recognizing these background object classes. The objects must be organized and grouped together to form much better image. However, for working in this direction the methods may not be possible which are not organized as per the perception of the user. Many research activities are being done to explore detecting object boundaries in the images solely based on some general properties of the real-world objects, such as perceptual organization laws, without depending on a priori knowledge of the specific objects.
With respect to the user's point of view the images creation and manipulation is very important. Perceptual organization is nothing but how the image can be presented with set of attributes and properties. The Gestalt psychologists summarized some underlying principles (e.g. proximity, similarity, continuity, symmetry, etc.) that lead to human perceptual grouping. The laws are the user's view and constraints for the formation of the image of the real world. These Gestalt laws can be summarized by a single principle, i.e., the principle of non-accidentalness.

II. IMAGE SEGMENTATION ALGORITHM

The image segmentation method is used to find the boundary areas. The main intention of this paper is to explore detecting object boundaries solely based on some general properties of the real-world objects, such as perceptual organization laws, without depending on object-specific knowledge. The image segmentation algorithm is inspired by a POM, which is the main contribution of this paper. The POM quantitatively incorporates a list of Gestalt cues. Most studies to date apply Gestalt laws on zero- or 1-D image features (e.g., points, lines, curves, etc.). Different to these studies, our method applies Gestalt laws on 2-D image features, i.e., object parts.

A. Unstructured objects.

Bottom-up segmentation method can be used to segment an outdoor image into uniform regions. Then, some of the regions must belong to the background objects. To recognize these background regions, we use textons to represent object appearance information. The term texton is for describing human textural perception. After textonization process, each image region of the training images is represented by a histogram of textons. We then use these training data to train a set of binary Adaboost classifiers to classify the unstructured objects (e.g., skies, roads, trees, grasses, etc.). By these classifiers high accuracy will be achieved on classifying these background objects in outdoor images.

B. Perceptual Organization Model

Most images consist of background and foreground objects. Most foreground objects are structured objects that are often composed of multiple parts, with each part having distinct surface characteristics (e.g., color, texture, etc.). Assume that we can use a bottom-up method to segment an image into uniform patches, then most structured objects should be over segmented to multiple patches (parts). After the background patches are identified in the image, the majority of the remaining image patches correspond to the constituent parts of structured objects. The main thing here is to see how to piece the set of constituted parts of a structured object together to form a region that corresponds to the structured object without any object-specific knowledge of the object. To tackle this problem, we develop a POM. Accordingly, our image segmentation algorithm can be divided into the following three steps.

1) Given an image, use a bottom-up method to Segment it into uniform patches.
2) Use background classifiers to identify background patches.
3) Use POM to group the remaining patches (parts). To larger regions that correspond to structured objects or semantically meaningful parts of structured objects.

The whole process works as follows: We first pick one part and then keep growing the region by trying to group its neighbors with the region. The process stops when none of the region's neighbors can be grouped with the region. To achieve this, we develop a measurement to measure how accurately a region is grouped. The region goodness directly depends on how well the structural relationships of parts contained in the region obey Gestalt laws. In other words, the region goodness is defined from perceptual organization perspective. With the region measurement, we can go find the best region that contains the initial part. In most cases, the best region corresponds to a single structured object or the semantically meaningful part of the structured object. This problem is formalized as follows.

Problem Definition: Let Ω represent a whole image that consists of the regions that belong to backgrounds R_B and the regions that belong to structured objects $R_S, \Omega = R_B \, U \, R_S$. After the background identification, we know that most of the structured objects in the image are contained in a sub-region R_S. Let P_0 be the initial partition of Ω from a bottom-up segmentation method. Let a denote a uniform patch from initial partition P_0. For $\forall$ ($a \in P_0$) $\wedge$ ($a \in R_S$), a is one of the constituent parts of an unknown structured object. Based on initial part a, we want to find the maximum region $R_a \subset R_S$ so that the initial part $a \in R_a$ and for any uniform patch i , where $(i \in P_0) \wedge (i \in R_a)$, i should have some special structural relationships that obey the *nonaccidentalness* principle with the remaining patches in R_a . This is formulated as follows:

$$R_a = arg_R \min(E[\partial R]) \quad with\ (a \in R) \wedge (R \in R) \qquad (1)$$

Where R is a region in RS, is the boundary of R , and E[∂R] is a boundary energy function. The boundary energy function provides a tool for measuring how good a region is. The goal here is to find the best region Ra in RS that contains initial part a. The boundary energy function is defined as follows

$$E[\partial R] = \frac{-\iint_R \int(x,y)dxdy}{L(\partial R)} \qquad (2)$$

Where L (∂R) is the boundary length of R.f(x,y) is the weight function in region R .The criterion of region goodness depends on how weight function f(x,y) is defined and also boundary length L(∂R). The convexity law, which is a global property, affects boundary energy E[∂R] based on the characteristic of boundary length L(∂R). First, we define the weight function f(x,y) in patch i as follows:

$$f(x,y) = e^{-\theta \bullet \eta (S_i - S_a)} \quad with\ (x,y) \in i, i \in R \quad (3)$$

where Θ is a weight vector. We empirically set Θ = [18, 3.5] in our implementation. Vector Si=[Bi,Ci], which is a point in the structural context space encoding the structure information of image patch i. Sa is a reference point in the structure's context spaces, which encodes the structural information of initial part a .. Weight function f(x,y) having a large value inside a newly included patch i means that current image patch i has a strong structural relationship with the constituent parts of the unknown structured object that contains initial part a. Ci is the cohesiveness strength, which we will define later. Bi is the boundary complexity of image patch i , which can be measured as

$$B_i = \frac{1}{N} \sum_{d=1}^{N} A(s,k) \bullet F(s,k) \qquad (4)$$

$$A(s,k) = 1 - \frac{\| p_{d+ks} - p_d \|}{\sum_{c=1}^{k} \| p_{d+cs} - p_{d+(c-1)s} \|} \qquad (5)$$

$$F(s,k) = 1 - 2 * \left| 0.5 - \frac{n}{N-3} \right| \qquad (6)$$

where N is the number of pixels of the boundary of image patch i, k is the length of a sliding window over the entire boundary of patch i . A(s,k) and F(s,k) are the respective strength and frequency of the singularity at scale (step) s. pd and pd+ks are the two end pixels of a segment of the boundary in the window. pd+cs and pd+(c-1)s are the pixels between pd and pd+ks . n is the number of notches in the window. A notch means a nonconvex portion of a polygon, which is defined as a vertex with an inner angle larger than 1800 . Based on the similarity of the boundary complexity, we can distinguish man-made object parts, which usually have regular shapes, from vegetations, which usually have irregular shapes. This is especially useful for distinguishing the vegetation that may not be recognized solely based on appearance. Therefore, the first Gestalt law we encode in the POM is the similarity law.

After obtaining the boundary complexity of image patch i, we then measure how tightly image patch i is attached to the parts of the unknown structured object that contains initial part a, Ci is the cohesiveness strength and is calculated as

$$C(i) = \begin{cases} 1 & for\ i = a \\ \max_j (e^{\phi_{ij}\varphi_{ij}\lambda_{ij}} C_j) & for\ i \neq a \wedge j \in neighbors(i) \end{cases} \quad (7)$$

where j is a neighboring patch of patch i . The maximum value of cohesiveness strength is one. For patch a , the cohesiveness is always set as the maximum value since we know for sure that the patch belongs to the unknown structured object. Assume the cohesiveness strength of patch j is known. □ij measures the symmetry of i and j along a vertical axis and is defined as

$$\phi_{ij} = 1 - \delta(y_i, y_j) \qquad (8)$$

where δ is the Kronecker delta function; yi and yj are the column coordinates of the centroids of i and j. If yi and yj are very close, this means that patches i and j are approximately symmetric along a vertical axis. If patch j has a strong cohesiveness strength to the unknown object containing patch a, then patch i also has a strong cohesiveness strength to the unknown object containing patch a. This means that patch i is tightly attached to some parts of the unknown object. This is because parts that are approximately symmetric along a vertical axis are very likely belonging to the same object. Thus, this test encodes the symmetry law. φ_{ij} measures the alignment of patches and

$$\varphi_{ij} = \begin{cases} 0 & if\ e(\partial ij) \cap \partial i = \Phi \wedge e(\partial ij) \cap \partial j = \Phi \\ 1 & if\ e(\partial ij) \cap \partial i \neq \Phi \wedge e(\partial ij) \cap \partial j \neq \Phi \end{cases} \quad (9)$$

where ∂i and ∂j are the boundaries of i and j , respectively, and e(∂ij) is the extension of the common boundary between i and j. Φ denotes the empty set. This alignment test encodes the continuity law. If two object parts are strictly aligned along a direction, then the boundary of the union of the two components will have a good continuation. Accordingly, alignment is a strong indication that these two parts may belong to the same object. If i and j are neither symmetric nor aligned, then the cohesiveness strength of image patch i depends on how it is attached to j. λij measures the attachment strength between i and j . It is defined as

$$\lambda_{ij} = \beta * exp\left(-\alpha \frac{(\cos\omega) * L(\partial ij)}{L(\partial i) + L(\partial j)}\right) \qquad (10)$$

where α and β are constants (we empirically set α = 20 and β=3 in our implementation). The attachment strength depends on the ratio of the common boundary length between and to the sum of the boundary lengths of i and j. We have explicitly encoded four Gestalt laws (i.e., similarity, symmetry, continuity, and proximity) into our POM. The four Gestalt laws affect weight function f(x,y) assigned to different parts and hence affect the boundary energy E[∂R] for different regions R. Convexity is a global property, it affects the boundary energy E[∂R] in a different way. As shown in Figure 1,

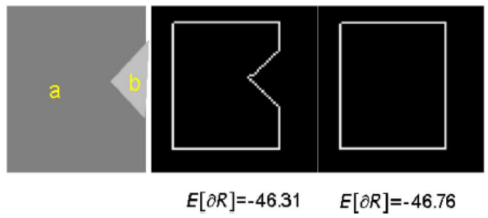

Figure1. Grouping a and b together with POM.

patch *b* is embedded into patch *a*, which causes a concavity on patch *a*. The boundary length of the region that contains *a* and *b* is shorter than that of the region that contains only patch *a* due to the concavity on patch *a*. As a result, the boundary energy of the region that contains patches *a* and *b* is smaller than that of the region that only contains patch *a*. Therefore, patches *a* and *b* are treated as one entity by our POM. In summary, any parts that are embedded into a big entity will be grouped together with the big entity due to their contribution of decreasing the boundary length of the new-formed big entity. This is because the embedded components increase the degree of convexity of the big entity they are embedded in.

Algorithm 1: Boundary detection based on perceptual organization

INPUT: P_0, R_S, and reference region *a*

OUTPUT: region R_a that contains *a* with the minimal boundary energy in a local area of $R_a \cup$ neighbors (R_a)

1. Let $R_a = a$.
2. Let $N\,R_a = \{\square_n \,|(\square_n \in P_0) \wedge (\square_n \in R_S) \wedge (\square_n \in$ neighbors (R_a)))} .
3. Repeat steps 4–7 for $q=1,\ldots,n$
4. Select a subset of $N\,R_a$ with q regions: $\mu = (u_1,\ldots,u_q)$, so that $\forall x, y <= q$, there exists a path in μ connecting u_x to u_y .
5. Measure the boundary energy of $R_a \cup \mu$ with (2).
6. If $E[\partial(R_a \cup \mu)] < E[\partial(R_a)]$, set $R_a = R_a \cup \mu$, GOTO step 2.
7. Otherwise select the next set of μ from $N\,R_a$ and repeat steps 4–7 until all possible μ have been tested.
8. Return R_a.

C. Image Segmentation Algorithm

The POM introduced in Section II.B can capture the special structural relationships that obey the principle of nonaccidentalness among the constituent parts of a structured object. We now turn to the image segmentation algorithm. Given an outdoor scene image, we first apply the segment-merge technique described above to generate a set of improved superpixels. Most of the superpixels approximately correspond to object parts in that scene. We build a graph to represent these superpixels: Let G = (V,E) be an undirected graph. Each vertex □e □ V corresponds to a superpixel, and each edge (□e, □f) □ E corresponds to a pair of neighboring vertices. We then use our background classifiers to divide V into two parts: backgrounds such as sky, roads, grasses, and trees RB and structured parts RS . Most of the structured objects in the scene are therefore contained in RS. We then apply our perceptual organization algorithm on RS. At the beginning, all the components in RS are marked as unprocessed. Then, for each unprocessed component □u in RS, we use the boundary detection algorithm described in Section III to detect the best region Ou that contains vertex □u . Region Ou may correspond to a single structured object or the semantically meaningful part of a structured object. We mark all the components comprising Ou as processed. The algorithm gradually moves from the ground plane up to the sky until all the components in RS are processed. Then, we finish one round of perceptual organization procedure and use the grouped regions in this round as inputs for the next round of perceptual organization on RS. At the beginning of a new round of perceptual organization, we merge the adjacent components if they have similar colors and build a new graph for the new components in RS. This perceptual organization procedure is repeated for multiple rounds until no components in RS can be grouped with other components. In practice, we find that the result of two rounds of grouping is good enough in most cases. At last, in a post-process procedure, we merge all the adjacent sky and ground objects together to generate final segmentation.

III .EXPERIMENTAL RESULTS

The segmentation of our POM is mainly based on the geometric relationships between different object parts. A segmentation accuracy score is defined as

$$K = \frac{GT \cap RM}{GT \cup RM} \qquad (11)$$

where and represent the set of pixels in the ground truth segment of an object and the machine generated object segment, respectively. The Segmentation results are as shown below figure a & b. The Segmentation Accuracy measurement is based on equation (11) is shown in the TABLE I for each class the score is averaged over the entire salient object segment. Table I gives the comparison of the performance of POM with that of the baseline method on the GDS. To future enhancement is to provide the background identification capability for more classifiers to identify mountains, buildings, walls, etc.

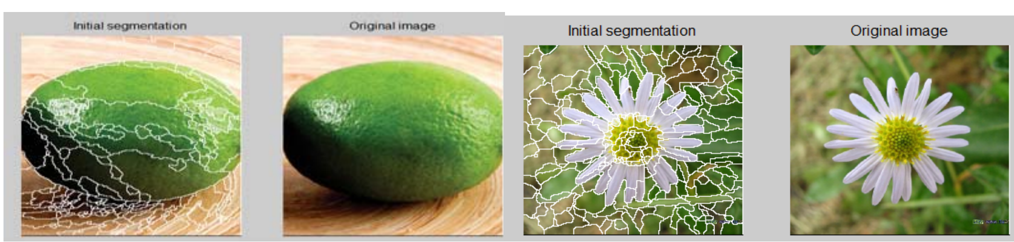

Figure 2. Segmentation process on the images showing original image and segmented image.

TABLE-I

Segmentation Accuracy score on GDS

	lemon	flower	Grass	water	Average
Gould09(on test set)	32.5	13.5	26.4	37.6	27.5
POM (on test set)	46.1	29.3	42.9	48.8	41.77
POM (on full data set)	48.9	34.4	46.5	44.9	43.67

IV. CONCLUSION

The segmentation and recognition are performed for identifying the foreground and background objects in the form of an interleaving procedure. The Segmentation Accuracy measurement is based on equation (11) is shown in the TABLE I for each class the score is averaged over the entire salient object segment. Background objects have low visual variety and hence can be reliably recognized. After background objects are identified, we roughly know where the structured objects are and delimit perceptual organization in certain areas of an image by using the perceptual organization. To future enhancement is to provide the background identification capability for more classifiers to identify mountains, buildings, walls, etc.

REFERENCES

[1] T. Malisiewicz and A. A. Efros, "Improving spatial support for objectsvia multiple segmentations," in *Proc. BMVC*, 2007.

[2] E. Borenstein and E. Sharon, "Combining top-down and bottom-up segmentation," in *Proc. IEEEWorkshop Perceptual Org. Comput. Vis.CVPR*, 2004, pp. 46–53.

[3] X. Ren, "Learning a classification model for segmentation," in *Proc.IEEE ICCV*, 2003, vol. 1.

[4] U. Rutishauser and D. Walther, "Is bottom-up attention useful for object recognition?," in *Proc. IEEE CVPR*, 2004, vol. 2, pp. 37–44

[5] C. Cheng, A. Koschan, D. L. Page, and M. A. Abidi, "Scene image segmentation based on perception organization," in *Proc. IEEE ICIP*, 2009, pp. 1801–1804.

[6] B. C. Russell, A. B. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation," *Int. J. Comput. Vis.*, vol. 77, no. 1–3, pp. 157–173, May 2008.

[7] H. Su, A. Bouridane, and D. Crookes, "Scale adaptivecomplexity measure of 2D shapes," In Proc. *ICPR* pp.134-137, 2006.

[8] D.W. Jacobs, "What makes viewpoint-invariant properties perceptually salient?" Journal of the optical society of America, A, Optics, image, science and vision, 20(7), pp 1304-1320, 2003.

[9] B.C. Russell, "Using Multiple Segmentations to discover objects and their extent in image collections," *In Proc. CVPR*, Vol. 2, pp. 1605-1614, 2006.

[10] I.H. Jermyn, and H. Ishikawa, "Globally optimal regions and boundaries as minimum ratio weight cycles," *IEEE Trans. PAMI,* 23(10), pp.1075-1088, 2001.

[11] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from RGBD images. ECCV, 2012.

[12] J. Tang, S. Miller, A. Singh, and P. Abbeel. A textured object recognition pipeline for color and depth image data. ICRA, 2012.

[13] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek.Evaluating color descriptors for object and scene recognition.TPAMI, 2010.