Adaptive Multiple Region Segmentation Based On Outdoor Object Detection

Anju J.A, P. Jenopaul

Abstract— The main research objective of this paper is to detecting object boundaries in outdoor scenes of images solely based on some general properties of the real world objects. Here, segmentation and recognition should not be separated and treated as an interleaving procedure. In this project, an adaptive global clustering technique is developed that can capture the non-accidental structural relationships among the constituent parts of the structured objects which usually consist of multiple constituent parts. In this clustering technique, we deal with an image histogram and repeatedly find the number of local maxima of the histogram. This number specifies the number of different regions in the image. The adaptive global maximum clustering is simple and fast calculation compared to k-means clustering algorithm. The background objects such as sky, tree, ground etc. are recognized based on the color and texture information. This process groups them together accordingly without depending on a priori knowledge of the specific objects. The proposed method outperformed two state-of-the-art image segmentation approaches on two challenging outdoor databases and on various outdoor natural scene environments, this improves the segmentation quality. By using this clustering helps to overcome strong reflection and over segmentation. This proposed method shows better performance and improve background identification capability.

Keywords— Adaptive global maximum clustering, gestalt laws, image segmentation, outdoor scenes, perceptual organization.

I. INTRODUCTION

IMAGE segmentation is the process of partitioning a digital image into multiple segments. One of the fundamental problem in computer vision is considered as image segmentation. The primary goal of image segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyse [1]. In general, the outdoor scenes can be categorized into two namely, unstructured objects (e.g., sky, roads, trees, grass, etc.) and structured objects (e.g., cars, buildings, people, etc.). The unstructured objects

Anju J.A is a PG Student, currently doing ME in Applied Electronics, in PSN College of Engineering and Technology, Tirunelveli
(e-mail: anju1589@gmail.com).

Mr. P.Jenopaul is with PSN College of Engineering and Technology, Tirunelveli as the Assistant Professor in Electronics and Communication Department. (e-mail: jenopaul1@rediffmail.com).

mainly consist of backgrounds of images and structured objects consist of foreground of images. The background objects usually have nearly homogenous surfaces and are distinct from the structured objects in images. So many appearance based methods are used to achieve high accuracy in recognizing these background object classes [2],[3],[4].

The challenge for outdoor segmentation comes from the structured objects that are often composed of multiple parts, with each part having distinct surface characteristics. Without certain knowledge about an object, it is difficult to group these parts together [5],[6]. The research objective of this paper is to explore detecting object boundaries in outdoor scene images only 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.

Perceptual organization plays an important role in human visual perception. Perceptual organization, refers to the basic capability of the human visual system to derive relevant groupings and structures from an image without prior knowledge of its contents. The Gestalt psychologists summarized some underlying principles (e.g., proximity, similarity, continuity, symmetry, etc) that lead to human perceptual grouping. The classic Gestalt laws pointed out that convexity also plays an important role on perceptual organization because many real-world objects such as buildings, vehicles, and furniture tend to have convex shapes. These can be summarized by a single principle, i.e., the principle of nonacidentalness, which means that these structures are most likely produced by an object or process, and are unlikely to arise at random[7].

For applying Gestalt laws to real world applications there are several challenges. One of challenge is to find quantitative and objective measures of these grouping laws. The Gestalt laws are in descriptive forms. Therefore, one needs to quantify them for scientific use. Another challenge consists of finding a way to combine the various grouping factors since object parts can be attached in many different ways. Under different situations, different laws may be applied. Therefore, a perceptual organization system requires combining as many Gestalt laws as possible. The greater the number of Gestalt laws incorporated, the better chance the perceptual organization systems may apply appropriate Gestalt laws in practices. Ren [8] developed a

probabilistic model of continuity and closure built on a scale-invariant geometric structure to estimate object boundaries. Jacobs emphasized that convexity plays an

The main contribution of this paper is a developed perceptual organization model (POM) for boundary detection. The POM quantitatively incorporates a list of Gestalt laws and therefore is able to capture the nonaccidental structural relationships among the constituent parts of a structured object. With this model, we are able to detect the boundaries of various salient structured objects under different outdoor environments.The proposed method outperformed two state-of-the-art studies [9],[10] on two challenging image databases consisting of a wide variety of outdoor scenes and object classes.


```
┌─────────────────┐
│   READ IMAGE    │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│  PREPROCESSING  │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│ ADAPTIVE GLOBAL │
│   CLUSTERING    │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│    ADABOOST     │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│   PERCEPTUAL    │
│  ORGANIZATION   │
│     MODEL       │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│   FOREGROUND    │
│   EXTRACTION    │
└─────────────────┘
```
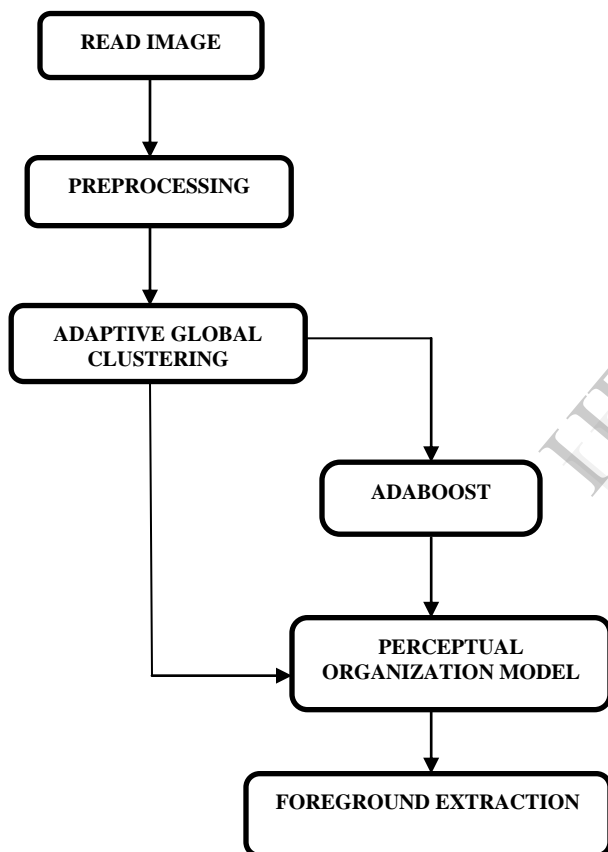
Fig 1: The proposed method

II. METHODOLOGY

The future method consists of three main steps for recognizing the common background and foreground objects.

A. Background Identification in Outdoor Natural Scenes

important role in perceptual organization and, in many cases, over rules other laws such as closure.

The objects seeming in natural scenes can be roughly divided into two categories namely, unstructured and structured objects. Unstructured objects typically have similar surfaces, whereas structured objects typically consist of several essential parts, with each part having distinct appearances in their color, texture, etc. The common backgrounds in outdoor natural scenes are those unstructured objects such as skies, roads, trees, and grasses and these objects have low visual variability in most cases and are distinct from other structured objects in an image. For instance, a sky commonly has a identical form with blue or white colours; a tree or a grass usually has a textured presence with green colours. Hence, these background objects can be precisely predictable only based on appearance data. Assume if we use a bottom-up segmentation method 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 a technique similar [2] . The key for this method is to use textons to represent object appearance information. The term texton is first presented for describing human textural perception. The whole textonization process proceeds as follows: First, the training images are converted to the perceptually uniform CIE color space. Then, the training images are convolved with a 17-D filter bank, which consists of Gaussians at scales 1, 2, and 4; the and derivatives of Gaussians at scales 2 and 4; and Laplacians of Gaussians at scales 1, 2, 4, and 8. These are applied to the uniform color space, whereas the other filters are applied only to the luminance channel. By doing so, the 17-D response is then augmented with the CIE channels to form a 20-D vector.

After augmenting the three color channels, we can achieve slightly higher classification accuracy [3]. Then, an adaptive global maximum clustering method is applied to automatically find the number of different regions and the meaningful clusters that have an important role in segmentation process. The adaptive global method is an unsupervised learning, trying to find the hidden structure of an image. Here we use an image histogram to get the number of different regions. In most images, there are too many local maxima of image histograms. But we need to find only the significant local maxima since those maxima are necessary to discriminate regions. To extract the significant local maxima, we first search for an interval, including the global maximum of an original histogram, and then fix the interval. Then such a fixed interval is called a cluster. Next, we eliminate the cluster gained from the former searching process, from the original histogram to find another new cluster. Then search for the new cluster of the reduced histogram and

repeat this process until we get the desired result. Fig. 2 explains this process. In this figure the black line shows the image histogram and red line indicates the local maxima. Here the x-axis denotes the gray level *l* and the y-axis denotes the number of occurrences of each gray level *h(l)*. Later the histogram changes every iteration, which is exactly the reduced version of the original histogram, the global maximum also adaptively changes every iteration. Therefore, we call such a maximum an adaptive global maximum that corresponds to one of the significant local maxima of the original histogram. This

whole process is a series of clustering a gray level interval into several subintervals so that the original histogram has the adaptive global maximum over each subinterval. Thus, we call this process the AGMC process. After this process, each image region of the training images is represented by a histogram of textons. 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.). to achieve high accuracy on classifying these background objects in outdoor images.
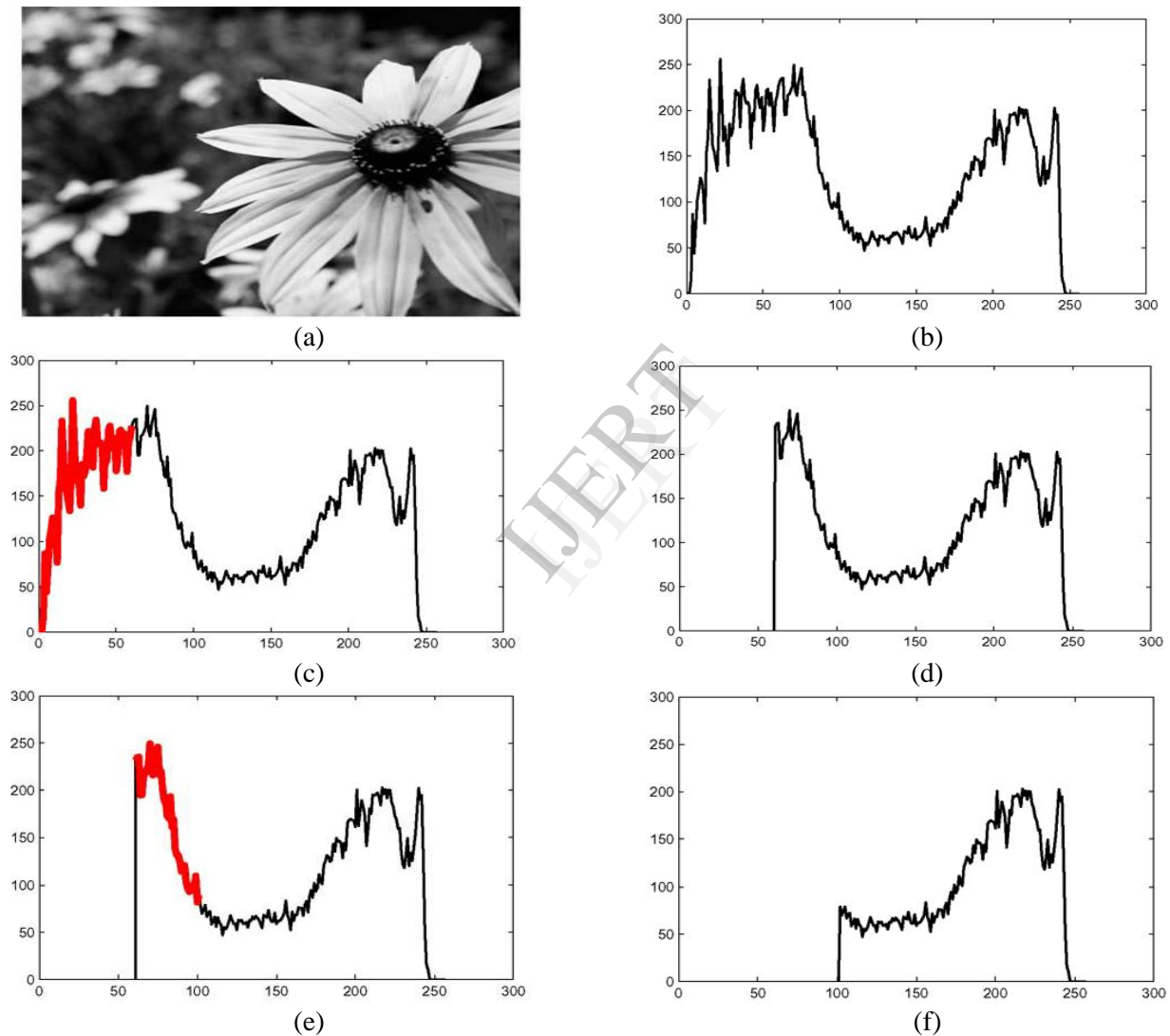

Fig 2: (a) Original image. (b) Its image histogram. (c) Domain of the thick red line is the first cluster. (d) Reduced histogram, which is the histogram to be used at the second iteration. (e) Domain of the thick red line is the second cluster. (f) Histogram at the third step. Note that the histogram changes every iteration of the AGMC process as (d) and (f).

B. Perceptual Organization Model (POM)

Most images consist of background and foreground objects and these foreground objects are structured objects that are often composed of multiple parts, with

each part having distinct surface characteristics. Assume that we can use a bottom-up method to segment an image into uniform patches, then most structured objects should be oversegmented to multiple 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 challenge here is 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.

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

We now go through the details of our POM. Even after background identification, there is still a large number of parts remaining. Different combinations of the parts form different

regions. We want to use the Gestalt laws to guide us to find and group these kinds of regions. Our strategy is that, since there always exist some special structural relationships that obey the principle of non accidentalness among the constituent parts of a structured object, we may be able to piece the set of parts together by capturing these special structural relationships. 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.

C. Image Segmentation Algorithm

The POM can capture the special structural relationships that obey the principle of nonaccidentalness among the constituent parts of a structured object. To apply the proposed POM to real-world natural scene images, we need to first segment an image into regions so that each region approximately corresponds to an object

part. In this implementation, Felzenszwalb and Huttenlocher's approach [11] are used to generate initial superpixels for an outdoor scene image. We select this method because it is very efficient and the result of the method is comparable to the mean-shift algorithm [12]. To further improve the segmentation quality, we apply a segment-merge method on the initial superpixels to merge the small size regions with their neighbors. These small size regions are often caused by the texture of surfaces or by the inhomogeneous portions of some part surfaces. Since these small size image regions contribute little to the structure information of object parts, we merge them together with their larger neighbors to improve the performance of our POM. In addition, if two adjacent regions have similar colors, we also merge them together. By doing so, we obtain a set of improved superpixels. Most of these improved superpixels approximately correspond to object parts. 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 be an undirected graph. Each vertex corresponds to a superpixel, and each edge corresponds to a pair of neighboring vertices. We then use our background classifiers are divide into two parts: backgrounds such as sky, roads, grasses, and trees and structured parts. We then apply our perceptual organization algorithm at the beginning, all the components in are marked as unprocessed. Then, for each unprocessed component to detect the best region that contains vertex . Region may correspond to a single structured object or the semantically meaningful part of a structured object. We mark all the components comprising as processed. The algorithm gradually moves from the ground plane up to the sky until all the components in 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. 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. This perceptual organization procedure is repeated for multiple rounds until no components in 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

A. Gould Database

First test the image segmentation algorithm using Gould image data set (GDS). This data set contains 10 images of urban and rural scenes assembled from a collection of public image data sets. The images on this data set are downsampled to approximately 320 pixels × 240 pixels. The images contain a wide variety of man-made and biological objects such as buildings, signs, cars, people, cows, and sheep. This data set provides ground truth object class segmentations that associate each region with one of eight semantic classes (sky, tree, road, grass, water, building, mountain, or foreground). For reference of this data set a Gould09 method is used here. The Gould09 method is a slight variant of the baseline method and achieved comparable result against the relative location prior method in Shotton's method and Yang's method on the MSRC-21 data set. Gould09 is trained on the training set and tested on the testing set. We first use the training images to train five background classifiers for background identification. Then, we test our POM method on both the testing set and the full GDS data set.

For the structured objects, POM does not gain any prior knowledge from training images. Here POM achieves very stable performance on segmenting the difficultly structured objects on the full data set. This shows that our POM can successfully handle various structured objects appearing in outdoor scenesl. POM is not a multiclass segmentation method because it does not label each pixel of an image with one of eight semantic classes as Gould09. Gould09 seems to be adaptable to the variation of the number of semantic classes. But, the foreground class in GDS includes a varied selection of structured object classes such as cars, buses, people, signs, sheep, cows, bicycles, and motorcycles, which have totally different appearance and shape features. This makes training an accurate classifier for classifying the foreground classes difficult. As a result, the Gould09 method cannot handle complicated environments where multiple foreground objects may appear close to each other. In such cases, the Gould09 method often labeled the whole group of physically different object instances such as people, car, and sign as one continuous foreground class region. In this method only requires identifying five background object classes. The remaining object classes are treated as structured objects.

B. Berkeley Segmentation Data set

POM image segmentation method can be evaluated by using Berkeley segmentation data set (BSDS) also. BSDS contains a training set of 10 images and a test set of 5 images. For each image, BSDS provides a collection of hand-labeled segmentations from multiple human subjects as ground truth. BSDS has been widely used as a benchmark for many boundary detection and segmentation algorithms in technical literature.

Figure2:Illustration of segmentation of an outdoor scene (a)input image,(b)preprocessing the input image, (c)the image is labeled

(a)

(b)

(c)

(d)

(e)

(f)

(g)
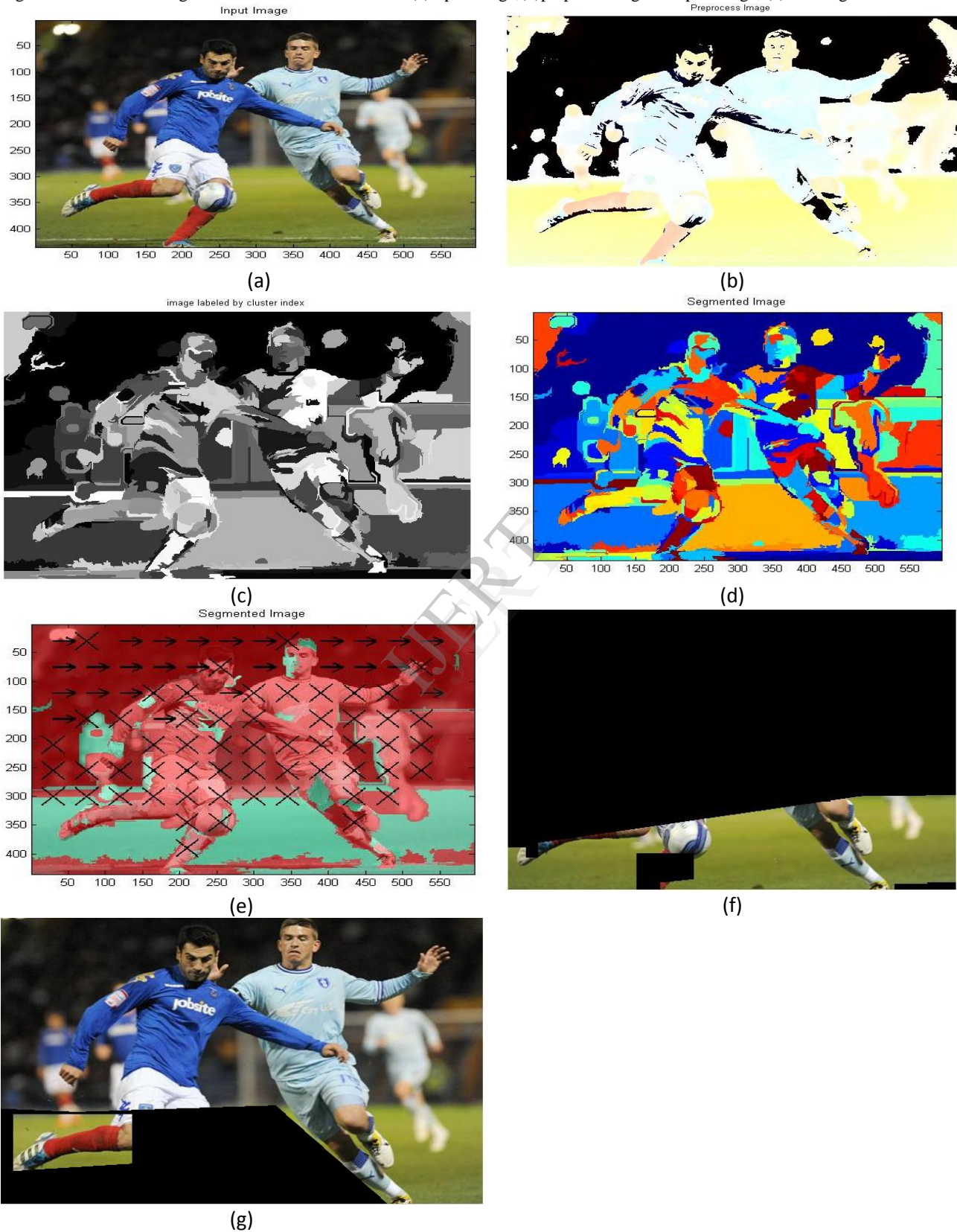
by cluster analysis using adaptive global maximum clustering technique, (d)segmentation of an input image: for background objects identification. Sky is labeled as blue, ground is labeled as yellow, and vegetations (tree or grass) are labeled as green, (e) label the input images, (f) and (g) shows the extraction of background and foreground objects of an image.

Here, we directly evaluate POM method on the test set of BSDS. The sizes of images in this data set are larger than the sizes of images in GDS. We use larger parameters to generate the initial superpixels for an input image. The same background classifiers trained in the GDS data set to identify background objects in this data set are used here also.

For each image, BSDS provides a collection of multiple human-labeled segmentations. For simplicity, we only select the first human-labeled segmentation of the collection as ground truth for the image. If the size of a ground truth segment size is smaller than % 0.5 of the image size, it is not a salient object. For the boundary-based measurement, we use the precision–recall framework recommended by BSDS. A precision–recall curve is a parameterized curve that captures the trade off between accuracy and noise. Precision is the fraction of detections that are true boundaries, whereas recall is the fraction of true boundaries that are detected. Thus, precision is the probability that the segmentation algorithm's signal is valid, and recall is the probability that the ground truth data is detected. These two quantities can be combined in a single quality measure, i.e., F-measure, defined as the weighted harmonic mean of precision and recall. Boundary detection algorithms usually generate a soft boundary map for an image.

IV. CONCLUSION

The main contribution of this paper is to develop a perceptual organization model for extracting background and foreground images of an object. The experimental results show that the future method outpaced two competing state-of-the-art image segmentation approaches and achieved good segmentation quality on two challenging outdoor scene image data sets. It is well accepted that segmentation and recognition should not be separated and should be treated as an interleaving procedure. In this method mainly follows the scheme and requires identifying some background objects as a starting point and compared to the large number of structured object classes. There are only a few common background objects in outdoor scenes and these 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. Our method can piece the whole object or the main portions of the objects together without requiring

recognition of the individual object parts. In other words, for these object classes, our method provides a way to separate segmentation and recognition. This is the major difference between our method and other class segmentation methods that require recognizing an object in order to segment it. This paper shows that, for many fairly articulated objects, recognition may not be a requirement for segmentation. The geometric relationships of the constituent parts of the objects provide useful cues indicating the memberships of these parts.

REFERENCES

[1] S. K. Shah, "Performance modeling and algorithm characterization for robust image segmentation," *Int. J. Comput. Vis.*, vol. 80, no. 1, pp. 92–103, Oct. 2008.

[2] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," Int. J. Comput. Vis., vol. 81, no. 1, pp. 2–23, Jan. 2009.

[3] J. Winn, A. Criminisi, and T. Minka, "Categorization by learned universal visual dictionary," in *Proc. IEEE ICCV*, 2005, vol. 2, pp. 1800–1807.

[4] L. Yang, P. Meer, and D. J. Foran, "Multiple class segmentation using a unified framework over man-shift patches," in *Proc. IEEE CVPR*, 2007, pp. 1–8.

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

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

[7]D. W. Jacobs, "What makes viewpoint-invariant properties perceptually salient?," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 20, no. 7, pp. 1304–1320, Jul. 2003.

[8] X. F. Ren, C. C. Fowlkes, and J.Malik, "Learning probabilistic models for contour completion in natural images," Int. J. Comput. Vis., vol. 77, no. 1–3, pp. 47–63, May 2008.

[9] S. Gould, O. Russakovsky, I. Goodfellow, P. Baumstarck, A. Y. Ng, and D. Koller, The STAIRVision Library (v2.3) 2009 [Online].Available: http://ai.stanford.edu/~sgould/svl

[10] M.Maire, P. Arbelaez, C. C. Fowlkes, and J.Malik, "Using contours to detect and localize junctions in natural images," in *Proc. IEEE CVPR*, 2008, pp. 1–8.

[11] 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.

[12] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[13] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[14] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in Proc. IEEE ICCV, 2009, pp. 1–8.

[15] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in Proc. IEEE CVPR, 2008, pp. 1–8.

[16] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller, "Multi-class segmentation with relative location prior," *Int. J. Comput. Vis.*, vol. 80, no. 3, pp. 300–316, Dec. 2008.

[17] C. Pantofaru, C. Schmid, and M. Hebert, "Object recognition by integrating multiple image segmentations," in *Proc. ECCV*, 2008, pp. 481–494.

[18] B. Micusik and J. Kosecka, "Semantic segmentation of street scenes by superpixel co-occurrence and 3-D geometry," in Proc. IEEE Workshop VOEC, 2009.