



Perceptual Organization Based Technique for Segmenting the Outdoor Scene Images

N. Balaraman
M.E Student
Department of CSE
Annamalai University
Tamil Nadu, India

Dr. J. Sasikala
Assistant Professor
Department of CSE
Annamalai University
Tamil Nadu, India

Abstract— *A novel outdoor scene image segmentation algorithm based on background recognition and perceptual organization is used to recognize the background objects such as the sky, the ground, and vegetation based on the color and texture information. For the structurally challenging objects, which usually consist of multiple constituent parts, a developed 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. The experimental results shows that the proposed method outperformed two state-of-the-art image segmentation approaches on two challenging outdoor databases (Gould data set and Berkeley segmentation data set) and achieved accurate segmentation quality on various outdoor natural scene environments.*

Keywords— *Boundary energy, image segmentation, perceptual organization.*

I. INTRODUCTION

To recognize the background objects such as the sky, ground, and vegetation based on the color and texture information, we propose a novel outdoor scene image segmentation algorithm based on background recognition and perceptual organization.(i.e.) to explore detecting object boundaries in outdoor scene 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.

Images are considered as one of the most important medium of conveying information. Understanding images and extracting the information from them such that the information can be used for other tasks is an important aspect of Machine learning. An example of the same would be the use of images for navigation of robots. Other applications like extracting malign tissues from body scans etc form integral part of Medical diagnosis. One of the first steps in direction of understanding images is to segment them and find out different objects in them. To do this, features like the histogram plots and the frequency domain transform can be used. In this project, we look at three algorithms namely K Means clustering, Expectation Maximization and the Normalized cuts and compare them for image segmentation. The comparison is based on various error metrics and time complexity of the algorithms. It has been assumed that the number of segments in the image are known and hence can be passed to the algorithm. The report is organized as follows. Section 2 describes each segmentation algorithm in detail. Results generated from the algorithms are presented in section 3. Finally, section 4 draws some conclusions.

A. Image Detection

In the concept of feature detection refers to methods that aim at computing abstractions of image information and making local decisions at every image point whether there is an image feature of a given type at that point or not. The resulting features will be subsets of the image domain, often in the form of isolated points, continuous curves or connected regions.

B. Image Segmentation

The term image segmentation refers to the partition of an image into a set of regions that cover it. The goal in many tasks is for the regions to represent meaningful areas of the image, such as the crops, urban areas, and forests of a satellite image. In other analysis tasks, the regions might be sets of border pixels grouped into such structures as line segments and circular arc segments in images of 3D industrial objects. Regions may also be as groups of pixels having both a border and a particular shape such as a circle or ellipse or polygon.

When the interesting regions do not cover the whole image, we can still talk about segmentation, into foreground regions of interest and background regions to be ignored. Figure 1.3.1 Football image (left) and segmentation into regions (right). Each region is a set of connected pixels that are similar in colour. Segmentation has two objectives. The objective is to decompose the image into parts for further analysis. In simple cases, the environment might be well enough controlled so that the segmentation process reliably extracts only the parts that need to be analyzed further. For example, in the chapter on colour, an algorithm was presented for segmenting a human .



Fig.,1: Original Image



Fig.,2: Segmented Image

II. METHODS USING FOR IMAGE SEGMENTATION

A. K - Means Clustering

Images are considered as one of the most important medium of conveying information. Understanding images and extracting the information from them such that the information can be used for other tasks is an important aspect of Machine learning. An example of the same would be the use of images for navigation of robots. Other applications like extracting malign tissues from body scans etc form integral part of Medical diagnosis. One of the first steps in direction of understanding images is to segment them and find out different objects in them. To do this, features like the histogram plots and the frequency domain transform can be used. In this project, we look at three algorithms namely K Means clustering, Expectation Maximization and the Normalized cuts and compare them for image segmentation. The comparison is based on various error metrics and time complexity of the algorithms. It has been assumed that the number of segments in the image are known and hence can be passed to the algorithm. The report is organized as follows. Section 2 describes each segmentation algorithm in detail. Results generated from the algorithms are presented in section 3. Finally, section 4 draws some conclusions.

B. Image Segmentation Algorithm

Images can be segmented into regions by the following algorithms: K-means Clustering Algorithm K-Means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are clustered around centroids $\mu_i \forall i= 1 \dots k$ which are obtained by minimizing the objective.

$$v = \sum_{i=0}^k \sum_{x_j \in s_i} (x_j - \mu_i) \dots \dots \dots [1]$$

k are the clusters $S_i, i= 1, 2, \dots$

k and μ_i is the centroid or mean point of all the points $x_j \in s_i$

As a part of this project, an iterative version of the algorithm was implemented.

The algorithm takes a 2 dimensional image as input. Various steps in the algorithm are as follows:

1. Compute the intensity distribution(also called the histogram) of the intensities.
2. Initialize the centroids with k random intensities.
3. Repeat the following steps until the cluster labels of the image does not change anymore.
4. Cluster the points based on distance of their intensities from the centroid intensities.

Type equation here.

$$C^i: \arg \min_j ||x^{(i)} - \mu_j ||^2 \dots \dots \dots [2]$$

C^i - centroids of i random intensities

$x^{(i)}$ - mean point of all points

$$\mu_i := \frac{\sum_{i=1}^m 1\{c(i)=j\}x^i}{\sum_{i=1}^m 1\{c_i=j\}} \dots \dots \dots [3]$$

μ_i - centroid intensities

j - itearates over all the centroids.

where k is a parameter of the algorithm (the number of clusters to be found), iterates over the all the intensities.

III. EXISTING SYSTEM

Bottom-up image segmentation methods only utilize low-level features such as colors, textures, and edges to decompose an image into uniform regions Bottom-up methods can be divided into two categories, namely, region-based and contour-based approaches. A group of approaches treats image segmentation as a graph cut problem. In this work we are using two existing methods one for segmentation and another one for background identification, the existing systems are,

- A. Bottom-up segmentation.
- B. Background Identification.

A. BOTTOM UP SEGMENTATION

Bottom-up image segmentation methods only utilize low-level features such as colors, textures, and edges to decompose an image into uniform regions. Bottom-up method can be divided into two categories, namely, region-based and contour-based approaches. A group of approaches treats image segmentation as a graph cut problem. Shi and Malik [7] proposed the normalized cut criterion that removes the trivial solutions of cutting small sets of isolated nodes in the graph. Felzenszwalb and Huttenlocher [8] proposed an efficient graph-based generic image segmentation algorithm. As with the normalized cut method, this method also tries to capture nonlocal image characteristics. Comaniciu and Meer [9] treated image segmentation as a cluster problem in a spatial-range feature space. Their mean-shift segmentation algorithm has illustrated excellent performance on different image data sets and has been considered as one of the best bottom-up image segmentation methods. Some of these region-based methods have been widely used to generate coherent regions called super pixels for many applications [10], [11], [12], [13].

$$R_a = \text{arg}_{R_{min}} (E[\partial R]) \text{with} (a \in R) \wedge R \in R_s \dots \dots \dots [4]$$

R_a - One of the constituent parts of an unknown structured object.

∂R - The boundary of R.

$E[\partial R]$ - The boundary energy function.

Contour closure is one of the important grouping factors identified by Gestalt psychologists. Early contour-based studies such as active contour methods only utilize boundary properties such as intensity gradients. Zhu and Yuille [14] first used both boundary and region information within an energy optimization model. For their method to achieve good performance, a set of initial seeds needs to be placed inside each homogenous region. Jermyn and Ishikawa [15] proposed a new form of energy function. Their energy function is defined on the space of boundaries in the image domain by a ratio of two integrals around the boundary. The numerator of the energy function is a measure of the “flow” of some quantity into or out of the region. The denominator is a generalized measure of the length of the boundary.

The main contribution of this energy function is to incorporate the general types of region information. Our method is based on this form of energy function, which is addressed in detail in Section III. In addition to the above energy function, some studies [16], [17] are built on different energy functions such as the Mumford–Shah segmentation model [18]. Recently, various various boundary detection methods based on statistical learning have been proposed in technical literature. Martin et al. [19] treated boundary detection as a supervised learning problem. They used a large data set of human-labeled boundaries in natural images to train a boundary model. Their model can then predict the possibility of a pixel being a boundary pixel based on a set of low-level cues such as brightness, color, and texture extracted from local image patches.

$$E[\partial R] = \frac{-\iint R^f(x,y) dx dy}{L(\partial R)} \dots \dots \dots [5]$$

$E[\partial R]$ - The boundary energy function.

$L(\partial R)$ - Boundary length of R

R^f - weight function in region R

Dollar et al. [20] and Hoiem et al. [21] followed a similar idea. Noticing the importance of context information, Dollar et al. [20] designed their boundary detection algorithm based on a large number of generic features calculated over a large image patch. This algorithm expects the context information to be provided by a large aperture. Hoiem et al. [21] estimated occlusion boundaries based on both 2-D perceptual cues and 3-D cues such as surface orientation and depth estimates. Multiclass image segmentation (or semantic segmentation) has become an active research area in recent years. The goal here is to label each pixel in the image with one of a set of predefined object class labels. Many studies operate on pixel level. Shotton et al. [22] assigned a class label to a pixel based on a joint appearance shape, and context model. In [23], Shotton et al. proposed the use of semantic texton forests for fast classification. A number of studies utilize superpixels as a starting point for their task. Gould et al. [24] proposed a superpixel-based conditional random field to learn the relative location offsets of categories. In their recent work [25], they developed a classification model that is defined in terms of a unified energy function over scene appearance and scene geometry structure. Other notable studies in this area include Micusik and Kosecka [11], Yang et al. [13], and He et al. [26]. Finally, we review some previous efforts attempting to apply Gestalt laws to guide image segmentation. A number of studies [1], [2], [27], [28] only applied one or two Gestalt laws (e.g., proximity, curvilinear, continuity, closure, or convexity, etc.) on 1-D image features (e.g., lines, curves, and edges) to find closed contours in images. Lowe [1] and Mahamud et al. [27] integrated proximity and continuity laws to detect smooth closed contour bounding unknown objects in real images. Ren et al. [28] developed a probabilistic model of continuity and closure built on a scale-invariant geometric structure to estimate object boundaries. Jacobs [2] emphasized that convexity plays an important role in perceptual organization and, in many cases, overrules other laws such as closure. Mohan and Nevatia [29] incorporated several Gestalt laws to detect a group of collated features describing objects. Their segmentation algorithm is based on a set of ad hoc geometric relationships among these collated features and is not based on the optimization of a measure of the value of a group. McCafferty [30] formulated the grouping problem in perceptual organization as an energy minimization problem where the energy of a grouping is defined as a function of how well it obeys the Gestalt laws. The total energy of a grouping is treated as the linear combination of the individual grouping energy values corresponding to the Gestalt laws.

Desolneux et al. [31] studied four Gestalt laws, namely, similarity in colors, similarity in sizes, alignments and proximity in point, and line and curve image features. They proposed the corresponding quantitative measurements for the significance of the four Gestalt laws and also showed the importance of the collaboration of Gestalt laws in the perceptual organization process.

B. BACKGROUND IDENTIFICATION

Objects appearing in natural scenes can be roughly divided into two categories, namely, unstructured and structured objects. Unstructured objects typically have nearly homogenous surfaces, whereas structured objects typically consist of multiple constituent parts, with each part having distinct appearances (e.g., color, texture, etc.). The common backgrounds in outdoor natural scenes are those unstructured objects such as skies, roads, trees, and grasses. These background objects have low visual variable and, in most cases, are distinguishable from other structured objects in an image. For instance, a sky usually has a uniform appearance with blue or white colors; a tree or a grass usually has a textured appearance with green colours.

Therefore, these background objects can be accurately recognized solely based on appearance information. Suppose if a bottom-up segmentation method is 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, a method is to use textons to represent object appearance information. The term texton is first presented in [32] for describing human textural perception.

The whole textonization process proceeds as follows:

- (i). The training images are converted to the perceptually uniform CIE color space.
- (ii). The training images are convolved with a 17-D filter bank which consists of Gaussians at scales 1, 2, and 4;
- (iii). The derivatives of Gaussians at scales 2 and 4; and Laplacians of Gaussians at scales 1, 2, 4, and 8.
The Gaussians are applied to all three color channels, whereas the other filters are applied only to the luminance channel.
- (iv). The process is repeated to obtain a 17-D response for each training pixel.
- (v). The 17-D response is then augmented with the CIE L^* , a^* , b^* channels to form a 20-D vector.
- (vi). Augmenting the three color channels, we can achieve slightly higher classification accuracy.
- (vii). The Euclidean-distance – means clustering algorithm is performed on the 20-D vectors collected from the training images to generate cluster centers. These cluster centers are called textons.
- (viii). Each pixel in each image is assigned to the nearest cluster center, producing the texton map. After this 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.). Our classifiers also achieve high accuracy on classifying these background objects in outdoor images.

Disadvantages:

- Using a bottom-up method to segment an image into uniform patches then most structured objects should be over segmented to multiple patches.
- After the background patches are identified the majorities of the remaining image patches corresponds to the constituent parts of the structured objects.
- The problem is how to piece the set of constituted parts of a structured objects together to form a region that corresponds to the structured object without any object-specific knowledge of the object

IV. PROPOSED SYSTEM

A. Perceptual Organization Model [Pom]

To overcome the problem of existing system the proposed one, the Perceptual Organization Method [POM] developed for boundary detection. POM can detect many structured object boundaries without having any object-specific knowledge of these objects. A salient structured object refers to a structured object with an independent and detectable physical boundary.

ex: window of a building.

An object parts refers to a homogeneous portion of the salient structured objects surface in an image. 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.

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

Similar to the human visual system, our POM can “perceive” a list of special structural relationships that obey the principle of non-accidentalness such as similarity in shape regularity, symmetry, alignment, adjacency, and embedment. The “perception” is quantified by boundary energy whenever a new member is added to a group. If the new

member has some structural relationships obeying the principle of non-accidentalness with other members in the group, then the boundary energy of the new formed group is smaller than the one of the old group. Otherwise, the boundary energy of the new-formed group is larger than the one of the old group.

The remaining task is to find the region R_a in (1) that has the minimum boundary energy among all the regions that contain image patch . In other words, we want to find the best region that contains image patch a and all image patches contained in the region that have some special structural relationships obeying the principle of non-accidentalness with each other. This region often corresponds to the whole structured object or the semantically meaningful portion of the structured object.

The challenge is that there may exist a large number of possible regions that contain image patch a in R_s , and it is computationally expensive to search all the possible regions to find the one with the global minimum boundary energy. Therefore, we develop an efficient boundary detection algorithm based on a breadth-first search strategy. Instead of finding the region with the global minimum boundary energy, the algorithm tries to find a region with the local minimum boundary energy.

Although it is not guaranteed that the algorithm is always able to find the region with the optimal boundary energy, we have found that it works quite well in practice.

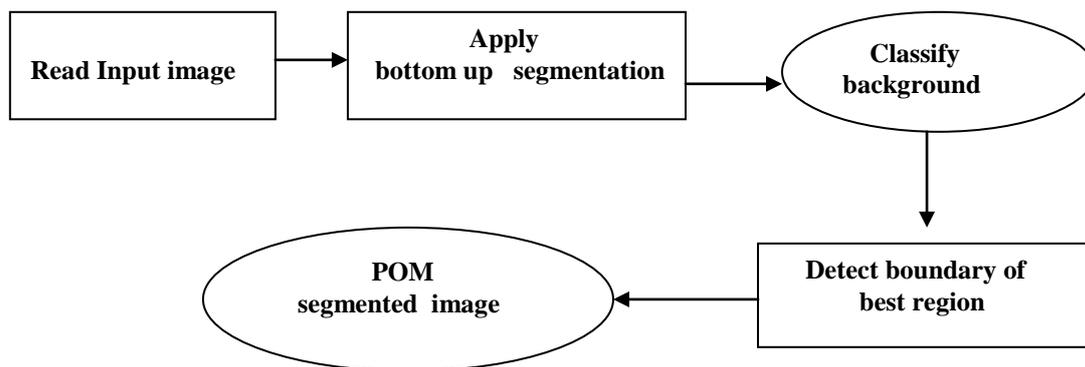


Fig.,3. Block Diagram Of Proposed Work [POM]

V. EXPERIMENTAL RESULTS



Fig.4. Modules Results

VI. PERFORMANCE ANALYSIS AND CONCLUSION

In this work, the bottom – up method is used in order to compare of the correlation between the original image and POM segmented images. To evaluate the performance of the proposed method outdoor images are used.

Table Accuracy Values of Three Modules

S.No	Images	Accuracy for Bottom up Method	Accuracy for Background Identification	Accuracy for POM Segmentation
1	HORSE	62.00	60.3	64.07
2	LION	63.00	68.8	69.02
3	HUMAN	71.00	65.3	74.06
4	LIZARD	68.00	64.7	70.08
5	ZEBRA	67.00	60.3	68.35

From the above tabulation, it is observed that the performance is consistent when going from the Berkeley Dataset.

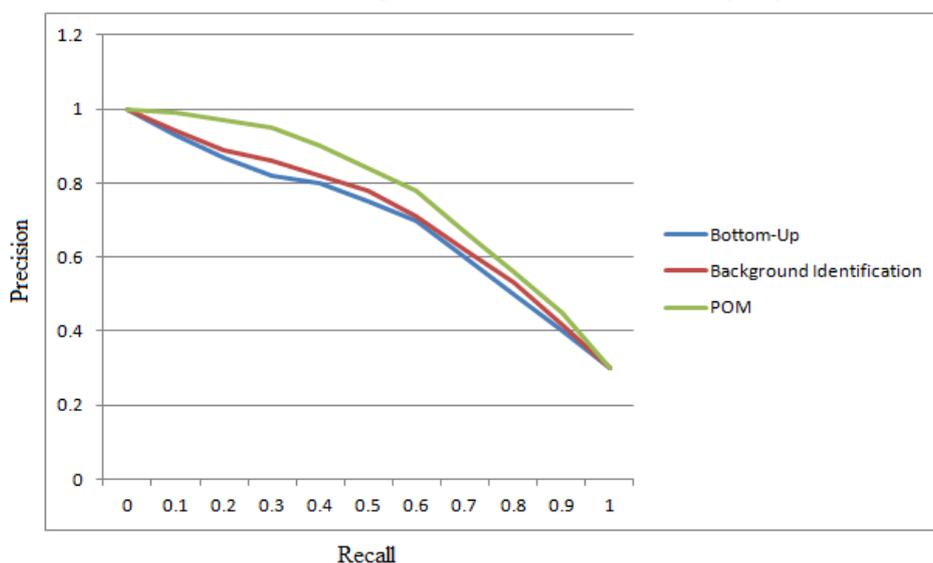


Fig .6.Comparison between existing and proposed systems

X axis 1 cm = 0.1 pixel
Y axis 1 cm = 0.2 pixel

In this work, to identify the outdoor natural scenes, a novel segmentation natural scenes, a novel segmentation algorithm, that is a proposed method POM is developed. The experimental results of POM method shows that the proposed one outperformed two competing state-of-the art-image segmentation approaches (BSDS [60]). It is well accepted that segmentation and recognition should not be separated and should be treated as an interleaving procedure. The method basically follows this scheme and requires identifying some background objects as a starting point. Compared to the large number of structured object classes, there are only a few common background objects in outdoor scenes. These background objects have low visual variety and hence can be reliably recognized. After background objects are identified, where the structured objects are and delimit perceptual organization in certain areas of an image. For many objects with polygonal shapes, such as the major object classes appearing in the streets (e.g., buildings, vehicles, signs, people, etc.) and many other objects, 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, the POM method provides a way to separate segmentation and recognition.

REFERENCES

[1] D. Lowe, Perceptual Organization and Visual Recognition. Dordrecht, The Netherlands: Kluwer, 1985.
 [2] D. W. Jacobs, "Robust and efficient detection of salient convex groups," IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 1, pp. 23–37, Jan. 1996.
 [3] Z. L. Liu, D. W. Jacobs, and R. Basri, "The role of convexity in perceptual completion: Beyond good continuation," Vis. Res., vol. 39, no. 25, pp. 4244–4257, Dec. 1999.
 [4] A. Witkin and J. Tenenbaum, "On the role of structure in vision," in Human and Machine Vision, J. Beck, B. Hope, and A. Rosenfeld, Eds. New York: Academic, 1983.

- [5] 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.
- [6] V. Bruce and P. Green, *Visual Perception: Physiology, Psychology and Ecology*. Hillsdale, NJ: Lawrence Erlbaum Associates Ltd., 1990.
- [7] J. B. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [8] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
- [9] 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.
- [10] T. Malisiewicz and A. A. Efros, "Improving spatial support for objects via multiple segmentations," in *Proc. BMVC, 2007*.
- [11] B. Micusik and J. Kosecka, "Semantic segmentation of street scenes by superpixel co-occurrence and 3-D geometry," in *Proc. IEEE Workshop VOEC, 2009*.
- [12] C. Pantofaru, C. Schmid, and M. Hebert, "Object recognition by integrating multiple image segmentations," in *Proc. ECCV, 2008*, pp. 481–494.
- [13] 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.
- [14] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing and Bayes/MDL for multi-band image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.
- [15] I. H. Jermyn and H. Ishikawa, "Globally optimal regions and boundaries as minimum ratio weight cycles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1075–1088, Oct. 2001.
- [16] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [17] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 271–293, Dec. 2002.
- [18] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, Jul. 1989.
- [19] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.
- [20] P. Dollar, Z. W. Tu, and S. Belongie, "Supervised learning of edges and object boundaries," in *Proc. IEEE CVPR, 2006*, vol. 2, pp. 1964–1971.
- [21] D. Hoiem, A. N. Stein, A. A. Efros, and M. Hebert, "Recovering occlusion boundaries from a single image," in *Proc. IEEE ICCV, 2007*, pp. 1–8.
- [22] 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.
- [23] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in *Proc. IEEE CVPR, 2008*, pp. 1–8.
- [24] 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.
- [25] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in *Proc. IEEE ICCV, 2009*, pp. 1–8.
- [26] X. He, R. Zemel, and M. Carreira-Perpinan, "Multiscale CRFs for image labeling," in *Proc. IEEE CVPR, 2004*, pp. 695–702.
- [27] S. Mahamud, L. R. Williams, K. K. Thornber, and K. Xu, "Segmentation of multiple salient closed contours from real images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 4, pp. 433–444, Apr. 2003.
- [28] 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.
- [29] R. Mohan and R. Nevatia, "Perceptual organization for scene segmentation and description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 6, pp. 616–635, Jun. 1992.
- [30] J. D. McCafferty, *Human and Machine Vision: Computing Perceptual Organization*. Chichester, U.K.: Ellis Horwood, 1990.
- [31] A. Desolneux, L. Moisan, and J. M. Morel, "A grouping principle and four applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 4, pp. 508–513, Apr. 2003.
- [32] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 7–27, Jun. 2001.