

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**4,800**

Open access books available

**122,000**

International authors and editors

**135M**

Downloads

Our authors are among the

**154**

Countries delivered to

**TOP 1%**

most cited scientists

**12.2%**

Contributors from top 500 universities



**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



# Image Segmentation Integrating Generative and Discriminative Methods

Yuee Wu and Houqin Bian  
 Computer & Information Engineering Dept.  
 ShangHai University of Electric Power, ShangHai  
 China

## 1. Introduction

Image segmentation is a long standing problem in computer vision, and it is found to be difficult and challenging for two reasons.

The first challenge is the difficulty of modeling the vast amount of visual patterns that appear in generic images. The second challenge is the intrinsic ambiguities in image perception, especially when there is no specific task to guide the attention. Furthermore, an image often demonstrates details at multiple scales. Therefore, it must be wrong to think that a segmentation algorithm outputs only one result. It should output multiple distinct solutions dynamically so that solutions “best preserve” the intrinsic ambiguity. In our opinion, image segmentation should be considered a computing process not a vision task.

Motivated by the above two observations, we present a stochastic computing method for image segmentation. We define image partition to be the task of decomposing an image  $I$  into its constituent visual patterns. The output is represented by a hierarchical graph. Firstly, we formulate the problem as Bayesian inference, and the solution space is decomposed into union of many subspaces of varying dimensions. The goal is to optimize the Bayesian posterior probability. Secondly, top-down generative models are used to describe how objects and generic region models (e.g. texture and shading) generate the image intensities. The goal of image partition is to invert this process and represent an input image by the parameters of the generative models that best describe it together with the boundaries of the regions and objects. Thirdly, in order to estimate these parameters we use bottom-up proposals, based on low-level cues, to guide the search through the parameter space.

We test the algorithm on a wide variety of grey level and color images, and some results are shown in the paper.

## 2. The Bayesian formulation for segmentation

Let  $\Lambda = \{(i, j) : 1 \leq i \leq L, 1 \leq j \leq H\}$  be an image lattice, and  $I_\Lambda$  an image defined on  $\Lambda$ . For any point  $v \in \Lambda$ ,  $I_v \in \{0, \dots, G\}$  is the pixel intensity for a grey level image, or  $I_v = \{L_v, U_v, V_v\}$  for a color image. The problem of image segmentation refers to partitioning the lattice into an unknown number of  $K$  disjoint regions.

$$\Lambda = \bigcup_{i=1}^K R_i, \quad R_i \cap R_j = \emptyset, \quad i \neq j \quad (2.1)$$

Each image region  $I_R$  is assumed to be coherent in the sense that  $I_R$  is a realization from a probabilistic model  $p(I_R; \Theta)$ .  $\Theta$  represents a stochastic process whose model type is indexed by  $\ell_i$ .

Thus segmentation is denoted by a vector of hidden variables  $W$ , which describes the world state for generating the image  $I$ .

$$W = (K, \{(R_i, \ell_i, \Theta_i); i = 1, 2, \dots, K\}) \quad (2.2)$$

In a Bayesian framework, we make the inference about  $W$  from  $I$  over a solution space  $\Omega$ .

$$W \sim p(W|I) \propto p(I|W)p(W), W \in \Omega \quad (2.3)$$

The likelihood  $p(I|W)$  specifies the image generating processes from  $W$  to  $I$ , and the prior probability  $p(W)$  represents our prior knowledge of the world. The goal is to estimate the most probable interpretation of an input image  $I$ . This requires computing the  $W^*$  that maximized a posteriori probability over  $\Omega$ , the solution space of  $W$ ,

$$\begin{aligned} W^* &= \arg \max_{W \in \Omega} p(W|I) \\ &= \arg \max_{W \in \Omega} p(I|W)p(W) \end{aligned} \quad (2.4)$$

### 3. Stochastic grammar of image

One fundamental difficulty that we encounter in vision is to represent the enormous amount of visual knowledge needed for making robust inference from real world images. The origin of image grammar is that certain elements of the image tend to occur together more frequently than by chance. These elements are then composed recursively to form increasingly larger units which can share some "reusable" parts.

Our production rules are graph operators and thus the image grammar is an attributed graph grammar. The graph grammar can be embedded in an And-Or graph representation, where each Or-mode points to alternative choices of sub-configuration, and an And-node is decomposed into a number of parts. Each non-terminal node generates child nodes starting with the scene label at the root and proceeds to objects, object parts, and ends with pixels as the leaves (terminal nodes). This hierarchical representation includes a dictionary  $\Delta_{gen}$  of generative image features used in the generative models. A special choice of the Or-nodes produces a configuration. The virtue of the grammar lies in its expressive power of generating a very large set of configurations through a relatively much smaller vocabulary. Figure 1 shows the grammar graph for an input image. Each node of the graph has an attribute variable for the labels and model parameters. The top node 0 is the scene label, and the nodes at the bottom are the image pixels. Three types of objects with different entropies are shown in nodes 1, 2, and 3.

To formulate this representation, we denote the graph by  $G = \langle V_N \cup V_T, E \rangle$ . The scene descriptions (attributes)  $W$  are defined on the non-terminal nodes  $V_N$ , and the image  $I$  is defined on the terminal nodes  $V_T$  (pixels).

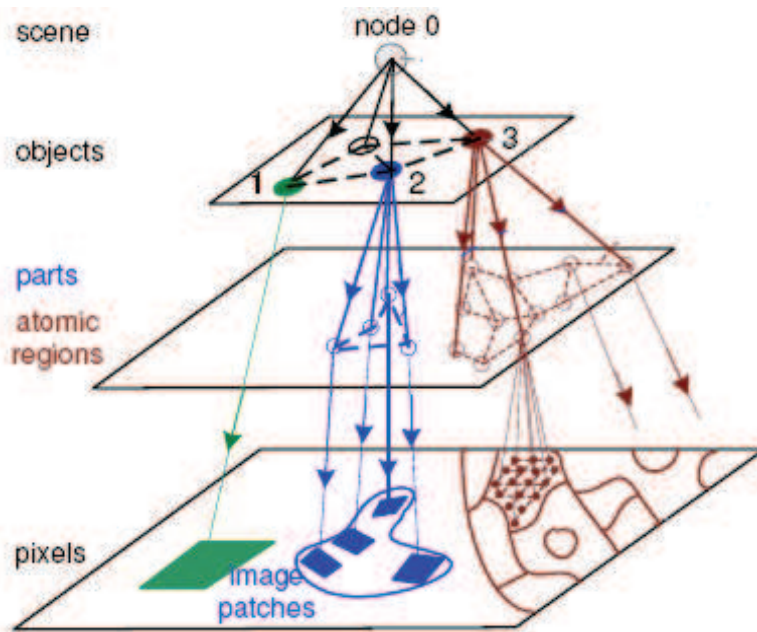


Fig. 1. The stochastic grammar for an input image

On each non-terminal node  $v \in V_N$ , we define attribute variables  $w(v) \in W$  for the labels and model parameters. Each non-terminal node  $v$  contains a list of children nodes  $child(v) = (u_1^v, \dots, u_{k^v}^v)$  which can be expanded for more specific representations of the object. These children nodes form a subgraph  $G_v$ . The leaf nodes at the bottom are the image pixels which form a lattice  $\Lambda$ . The attributes defined on the leaf nodes (lattice) are the image intensity values  $J$ . Depending on the visual patterns,  $J$  can be determined by  $W$  either deterministically  $J = g(W)$  or probabilistically  $J \sim p(J|W)$ . The correctness of a scene description  $W$  will depend on how well the synthesized image  $J$  matches the input image  $I$ . The full generative model allows the 30 different visual patterns to compete to explain the input intensity. The generative model is expressed as:

$$p(W; \Delta_{gen}) = p(w) \prod_{v \in V_N} p(w(u_1^v), \dots, w(u_{k^v}^v) | w(v); \Delta(v)), \quad (2.5)$$

$$J = g(W) \text{ or } J \sim p(J|W)$$

Where the conditional probability at  $v$  is a MRF defined on the graph  $G_v$  and captures the spatial relationship between the parts.  $\Delta_v$  is a generative image dictionary for  $w(v)$ , such as PCA, image patches, or textons. We will have a collection of generative image vocabularies for the 30 types of objects  $\ell$  over scale  $\sigma$

$$\Delta_{gen} = \{\Delta_{\ell, \sigma} : \ell = 1, 2, \dots, 30, \sigma = 1, 2, 3, \dots\} \quad (2.6)$$

Then we formulate image partition as Bayesian inference where the goal is to compute the partition graph  $W$  from the input image  $I$  and, or verification, to synthesize  $J$  sampled from the likelihood,

$$W \sim p(W|I; \Delta_{gen}) \propto p(I|W)p(W; \Delta_{gen}), \quad W \in \Omega, J \sim p(J|W) \quad (2.7)$$

Our algorithm must achieve the difficult task of (1). constructing the stochastic grammar, whose structure, like parse trees in natural language, is not pre-determined but depend on the input image; (2). estimating the attributes of graph nodes (labels and model parameters); and (3). computing the spatial relations between sibling nodes.

The framework in the paper integrates two computing paradigms in vision –discriminative methods for fast speed and generative methods for generality and robustness.

#### 4. Generative and discriminative methods

Generative methods specify how the image  $I$  is generated from the scene representation  $W$ . it combines a prior  $p(W; \Delta_{gen})$  and a likelihood function  $p(I|W)$  to give a joint posterior probability  $p(W|I; \Delta_{gen})$ . To perform inference using generative methods requires estimating  $W^* = \arg \max p(W|I; \Delta_{gen})$ . This is often computationally demanding because there are usually no know efficient inference algorithm.

By contrast, discriminative methods are very fast to compute. They do not specify models for how the image is generated. Instead they give discriminative probabilities  $q(w_j|Tst_j(I))$  for components  $(w_j)$  of  $W$  based on a sequence of bottom-up tests  $Tst_j(I)$  performed on the image. The tests are based on local image features  $\{F_{j,n}(I)\}$  which can be computed from the image in a cascade manner,

$$Tst_j(I) = (F_{j,1}(I), F_{j,2}(I), \dots, F_{j,n}(I)), \quad j = 1, 2, \dots, K \quad (2.8)$$

These tests are selected from a dictionary of discriminative features  $\Psi_{dis}$ . In correspondence to the generative dictionary  $\Delta_{gen}$  in eqn.(2.7), we denote it by

$$\Psi_{dis} = \{\Psi_{\ell, \sigma} : \ell = 1, 2, \dots, 30, \sigma = 1, 2, \dots\} \quad (2.9)$$

The bottom-up tests generate two types of hypotheses.

- i. The *what-is-what* hypothesis for some node  $v$  in the partition graph  $W$ , which are marginal posterior probabilities  $q_v = q(w(v)|F_v(I))$  for the attributes  $w(v)$  (object label and model parameters) of  $v$ .  $F_v(I)$  denote the features used (in this paper we use Adaboost).
- ii. The *what-go-with-what* hypotheses for some horizontal edge  $e = \langle s, t \rangle$  in the partition graph, which are posterior probabilities  $q_e = q(e = \text{off} | f_e(I))$  for whether the two elements  $s, t$  belong to the same pattern. Equivalently it is the probability ratio  $\frac{q(e = \text{on} | I)}{q(e = \text{off} | I)} \cdot f_e(I)$  denotes the feature for measuring the dis-similarity between  $s, t$ . It has

been proved that with sufficient number of tests  $q(w(v)|F_v(I))$  can approach  $p(w(v)|I)$

asymptotically, and so  $\frac{q(e = \text{on} | f_e(I))}{q(e = \text{off} | f_e(I))}$  will approach  $\frac{p(e = \text{on} | I)}{p(e = \text{off} | I)}$ .

These two set of discriminative probabilities are then composed on-the-fly to generate hypotheses which are represented by importance proposal kernel,

$$Q_a(W, W' | Tst_t(I)) = q(W' | W, Tst_t(I)) \quad \text{Generative methods optimizes the joint Bayesian}$$

posterior  $p(W | I; \Delta_{gen})$  by a set  $A$  of reversible jumps, such as death-birth, spit-merge, model switching, etc. these jumps construct the partition graph and in combination they simulate an ergodic Markov chain search in the solution space of  $W$ . each type of jump  $a \in A$  is represented by a Markov kernel  $\kappa_a(W, W')$ , which is a conditional probability for moving from partition graph  $W$  to a new partition graph  $W'$ .

The kernels are "informed" by proposal kernels computed by discriminative method, and are realized by the Metropolis-Hastings method,

$$\kappa_a(W, W' | Tst_t(I)) = Q_a(W, W' | Tst_t(I)) \min\left(1, \frac{Q_a(W', W | Tst_t(I)) \cdot p(W' | I)}{Q_a(W, W' | Tst_t(I)) \cdot p(W | I)}\right) \quad \text{for } W \neq W', a \in A \quad (2.10)$$

The metropolis-Hastings step compares the discriminative probability ratio with the true Bayesian posterior probability ratio, and can be considered as a probabilistic version of hypothesis-and-test.

## 5. Experiments

The image segmentation algorithm is applied on a number of outdoor/indoor images. The speed in PCs is comparable to segmentation methods such as normalized cuts. It typically runs around 10-20 min. the main portion of the computing time is spent in segmenting the generic patterns and by boundary diffusion. Figure 3 and 4 show some example. We present the results in two parts. One shows the segmentation boundaries for generic region and objects, and the other shows the labelmap for generic region and objects to indicate objects recognition. From the segmentation results we can see high-level knowledge helps segmentation to overcome problem of oversegmentation.

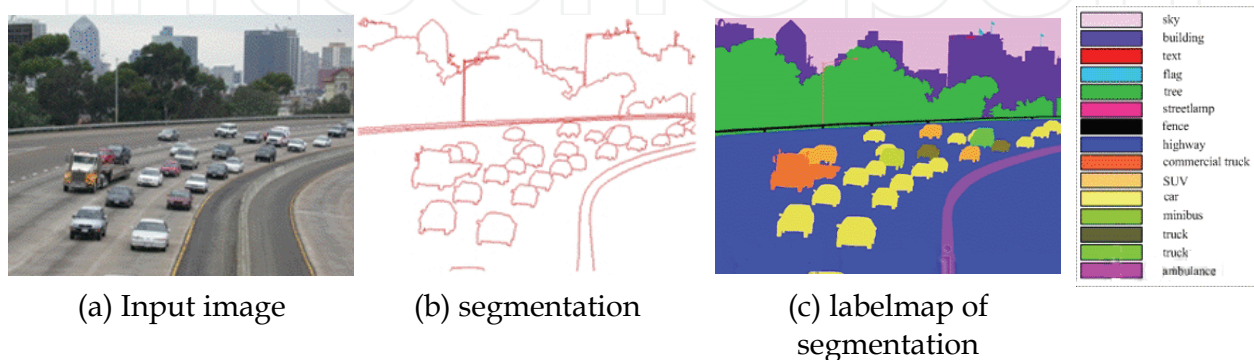


Fig. 2. Results of segmentation on an outdoor image



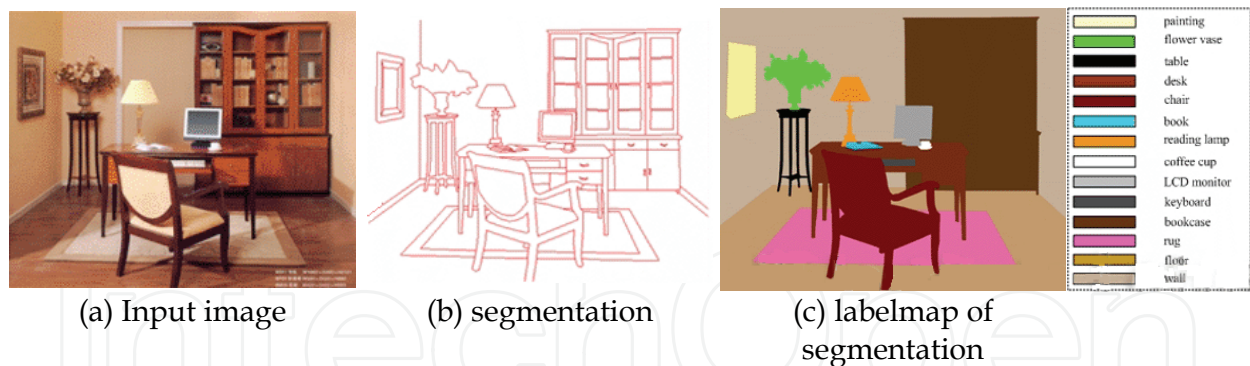


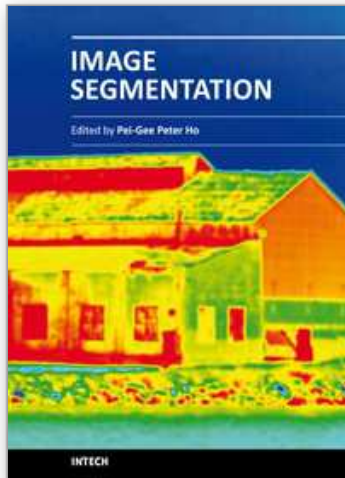
Fig. 3. Results of segmentation on an indoor image

## 6. Acknowledgments

This work is supported by the Project of ShangHai Key Disciplines Construction (p1303) and Special Foundation for Outstanding Young Teachers in Colleges and Universities of ShangHai (z-2009-03).

## 7. References

- [1] S.C. Zhu and D. Mumford, "A Stochastic Grammar of Images", *Foundations and Trends in Computer Graphics and Vision*, 2007.
- [2] T.F. Wu, G.S. Xia, and S.C. Zhu, "Compositional Boosting for Computing Hierarchical Image Structures" *CVPR*, 2007.
- [3] S. Agarwal and D. Roth, "Learning a sparse representation for object detection", *Proc. ECCV*, 2002.
- [4] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting", *The Annals of Statistics*, 38(2):337-374, Oct. 2001.
- [5] H. Chen and S.C. Zhu, "A generative model of human hair for hair sketching", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, 2005.
- [6] Shiliang Sun, Changshui Zhang and Guoqiang Yu. "A Bayesian Network Approach to Traffic Flow Forecasting." *IEEE Transactions on Intelligent Transportation Systems*. March 2006. Volume 7, issue 1. pp. 124- 132.
- [7] Z.Y. Yao, X. Yang, and S.C. Zhu, "Introduction to a large scale general purpose groundtruth dataset: methodology, annotation tool, and benchmarks", *6th Int'l Conf on EMMCVPR*, August 2007.



## **Image Segmentation**

Edited by Dr. Pei-Gee Ho

ISBN 978-953-307-228-9

Hard cover, 538 pages

**Publisher** InTech

**Published online** 19, April, 2011

**Published in print edition** April, 2011

It was estimated that 80% of the information received by human is visual. Image processing is evolving fast and continually. During the past 10 years, there has been a significant research increase in image segmentation. To study a specific object in an image, its boundary can be highlighted by an image segmentation procedure. The objective of the image segmentation is to simplify the representation of pictures into meaningful information by partitioning into image regions. Image segmentation is a technique to locate certain objects or boundaries within an image. There are many algorithms and techniques have been developed to solve image segmentation problems, the research topics in this book such as level set, active contour, AR time series image modeling, Support Vector Machines, Pixon based image segmentations, region similarity metric based technique, statistical ANN and JSEG algorithm were written in details. This book brings together many different aspects of the current research on several fields associated to digital image segmentation. Four parts allowed gathering the 27 chapters around the following topics: Survey of Image Segmentation Algorithms, Image Segmentation methods, Image Segmentation Applications and Hardware Implementation. The readers will find the contents in this book enjoyable and get many helpful ideas and overviews on their own study.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yuee Wu and Houqin Bian (2011). Image Segmentation Integrating Generative and Discriminative Methods, Image Segmentation, Dr. Pei-Gee Ho (Ed.), ISBN: 978-953-307-228-9, InTech, Available from: <http://www.intechopen.com/books/image-segmentation/image-segmentation-integrating-generative-and-discriminative-methods>

**INTECH**  
open science | open minds

#### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

#### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821



© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen