Volume : 1 | Issue : 3 | December 2011

ISSN - 2249-555X

Research Paper

Image Parsing

Image Parsing : Models And Algorithms

* Pankaj Bhambri ** Amandeep Kaur

* Assistant Professor, Department of Information Technology, Guru Nanak Dev Engg. College, Ludhiana

** Research Scholar, Guru Nanak Dev Engg. College, Ludhiana

ABSTRACT

In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels, also known as superpixels). Image segmentation is a long standing problem in computer vision. Real world images consist of multiple layers of stochastic processes, such as texture, texton, stochastic point, line, curve, graph, region, and object processes, which generate images through spatial organizations. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Image parsing is basically done to properly understand an image.

Keywords : Image parsing, Image segmentation, Object detection.

Introduction

Image segmentation is a long standing problem in computer vision. Often it is viewed as an ill-defined problem in comparison to other vision tasks which have apparently well defined objectives, such as detection, recognition, and tracking. Unfortunately, without addressing segmentation problems, those special purpose vision tasks are fundamentally ill-defined. It is fair to say that computer vision or image understanding is all about parsing images. Image parsing attempts to find a semantically meaningful label for every pixel in an image. Many vision problems that use natural images involve image parsing and a wealth of possible applications include: autonomous navigation, augmented realities and image database retrieval. Natural images consist of an overwhelming number of visual patterns generated by very diverse stochastic processes in nature. The objective of image understanding is to parse an input image into its constituent patterns.

Depending on the type of patterns that a task is interested in, the parsing problem is called respectively

1). Image segmentation --- for homogeneous grey/color/texture region processes.

2). Perceptual grouping --- for point, curve, and general graph processes

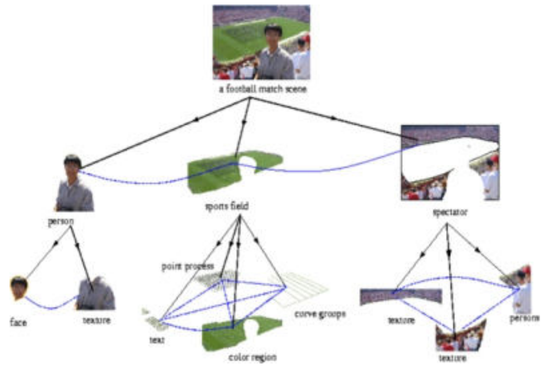
3). Object recognition --- for text and objects.

The figure below is an example of parsing a stadium scene hierarchically: human (face and clothes), sports field (a point process, a curve process, homogeneous colorregions, text) and spectators (textures, persons).

Objectives Of Image Parsing

Image parsing is basically the task of decomposing an image I into its constituent visual patterns. The output is represented by a hierarchical graph W called the " parsing graph".

Figure1 : Image Parsing Example



The goal is to optimize the posterior probability p (W|I). Figure 1 illustrates a typical example where a football scene is first divided into three parts at a coarse level : a person in the foreground, a sports field, and the spectators. These three parts are further decomposed into nine visual patterns in the second level : a face, three texture regions, some text, a point process (the band on the field), a curve process (the markings on the field), a color region, and a region for nearby people. In principle, an image can be decomposed until reaching a resolution criterion. The confusion in defining image segmentation (and also perceptual grouping) simply reflect the following two facts.

- It is difficult to model all types of stochastic patterns in generic vision.

Real world images consist of multiple layers of stochastic processes, such as texture, texton, stochastic point, line, curve, graph, region, and object processes, which generate images through spatial organizations.

This subsumes image segmentation as region process, and naturally integrates object recognition and perceptual organization. The latter deal with point, line, curve, and object processes. Implicit in this formulation is the notion of generative models for image interpretation in contrast to classification and discrimination methods. This observation motivated our work and many others in modeling and learning various stochastic patterns.

- Image segmentation is a computational process and should NOT be treated as a task.

Real images are intrinsically ambiguous, and our perception changes dynamically, even in a very short time duration, depending on our attention. Generally speaking, the more onelooks at an image, the more one sees. It seems narrow-minded to think that a segmentation algorithm just outputs one final result. Instead a segmentation algorithm should realize the intrinsic ambiguities characterized, say, in a Bayesian posterior probability, and outputs multiple distinct solutions dynamically and endlessly so that these solutions, as samples, ``best preserve'' the posterior density. Then we need a mathematical principle for pruning and selecting the outputs.

Models Of Image Parsing

Region Competition

A main contribution of this work is the derivation of a diffusion equation from a general MDL/Bayesian energy function (posterior probability). This equation shows that region growing and SNAKE/Balloon type algorithms are different ways for minimizing an energy. A remaining problem is that the split-merge steps are not reversible, therefore, the whole algorithm is greedy, and is not guaranteed to find global optima. This problem is then resolved by the DDMCMC framework and the Swendson-Wang Cut algorithm.

Energy formulation for image segmentation: An image contains n regions and each region is fitted to a model. The contours of the regions are assumed to be smooth (short length).

$$E[\Gamma, \alpha] = \sum_{i=1}^{M} \{ \ \frac{\mu}{2} \oint_{\partial R_i} ds - \iint_{R_i} \log p(I(x, y); \alpha_i)\, dxdy \ + \lambda \}$$

Image Segmentation By Data Driven Markov Chain Monte Carlo

DMCMC is a follow-up work of region competition. DDMCMC contributes in the following aspects:

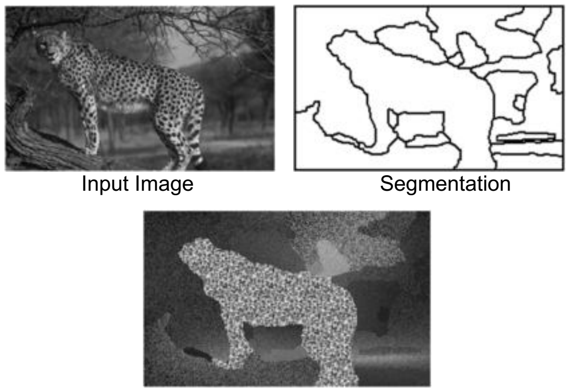
1. It makes the split-merge process reversible, thus the algorithm form ergodic Markov chain searching in the heterogeneously structured solution space, and it achieves global optimization independent of initial segmentation conditions.

2. We regard image segmentation as a computing process Not a vision task. Thus the algorithm should run endlessly and output many distinct solutions, i.e. the more it looks, the more it sees. This is consistent with human attention at a single image. We studied a k-adventurers algorithm which prune trivial solutions and preserve K-most important and distinct segmentations. This generalizes the conventional maximum a posteriori (MAP) estimation.

3. The DDMCMC framework engages various image models: multinomial, texture, color, global spline et al. These are generative models, and are compatible with each other.

4. The DDMCMC framework provides a unified view for the role of conventional algorithms, region growing, SNAKE/Balloon, split-merge, model switching and adaptation, PDEs and diffusions, region competition and subsume these algorithms. It provides a compatible platform to embed perceptual organization and object recognition.

Figure2 Image Segmentation By DDMCMC



Input Image Segmentation

Synthesis

3.3 3D Scene Segmentation And Reconstruction By DDMCMC

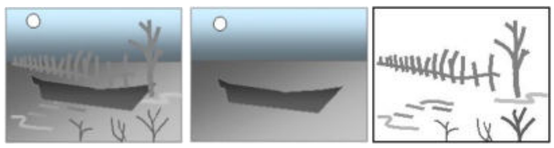
Recently there are renewed and growing interest in computer vision research for parsing and reconstructing 3D scenes from range images, driven by new developments in sensor technologies and new demands in applications. Firstly, high precision laser range cameras are becoming accessible to many users, which makes it possible to acquire complex real world scenes like the following ones:

Figure3 3D Scene Segmentation And Reconstruction By DDMCMC

3.4 Parsing Images Into Region And Curve Processes

Curve is a typical pattern appearing in natural images and has been studied extensively in the literature. Also, curves often fall in groups corresponding to high level structures such as railings, trees, and textures, which are key components to be recognized in understanding scenes for visual systems. Without analyzing curves explicitly, image segmentation algorithms assuming 2D region patterns often have problems in dealing with images with rich 1D structures, for example, parallel railings, weeds, tree branches, etc.

Figure4 Parsing Images Into Region And Curve Processes



Input Image Region Layer Curve Layer

Algorithms Of Image Parsing

4.1 Shortest Paths

path(s) = 0

For each edge (u, v):

path(u) = w

path(v) = w +w(u, v)

Lightest derivation of path(v) is a shortest path from s to v.

4.2 Vision Pipeline

Edges → Contours → Objects

Vision system are often defined in terms of processing stages.

Can define each stage using rules for building structures by composing structures from previous stage.

Solving a single lightest derivation problem allows high-level knowledge to influence low-level interpretations.

But leads to a difficult computational problem.

4.3 Prioritized Rules

Rules A1, . . . ,An →g C with priority function p(w1, . . . ,wn).

Execution generates weight assignments (B = w).

S: set of assignments

Q: priority queue of assignments

Procedure Run(P)

1. S ←
2. Initialize Q using rules with no antecedents
3. while Q is not empty
4. Pop assignment (B = w) from Q
5. if B has no assigned weight in S
6. S ← S U {(B = w)}
7. Push assignments derivable in one step using S into Q
8. return S

4.4 KNUTH'S LIGHTEST DERIVATION

Define prioritized rules K(R) by setting the priority of each rule in R to be g(w1, . . . ,wn).

Generalizes Dijkstra's shortest paths algorithm.

Computes lightest derivations in increasing weight order.

We can stop without deriving every statement.

Requires non-decreasing and superior weight functions.

Example: g(w1, . . . ,wn) = w1 + . . . +wn +v

Conclusion

Image segmentation is a long standing problem in computer vision. Real world images consist of multiple layers of stochastic processes, such as texture, texton, stochastic point, line, curve, graph, region, and object processes, which generate images through spatial organizations. Thus an appropriate formulation should be image decomposition or parsing, which decomposes an image into its natural constituents as various stochastic processes. With the help of image parsing the complex image structures are defined in a coarse-to-fine manner.

REFERENCES

F. Jelinek and J. D. Lafferty, “Computation of the probability of initial substring generation by stochastic context-free gram-mars,” Computational Linguistics, vol. 17, no. 3, pp. 315323,1991. | M. Collins, “Head-driven statistical models for natural languageparsing,” Ph.D. Thesis, University of Pennsylvania, 1999. | K. Lari and S. J. Young, “The estimation of stochastic context-free grammars using the inside-outside algorithm,” in ComputerSpeech and Languag, 1990. | M. Shilman, P. Liang, and P. A. Viola, “Learning non-generative grammatical models for document analysis,” in Proceedings of IEEE International Conference on Computer Vision, 2005, pp.962969. | Z. Tu and S. C. Zhu, “Image segmentation by data-driven markov chain monte carlo,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 657673,2002. | Z. Tu, X. Chen, A. L. Yuille, and S. C. Zhu, “Image parsing: Unifying segmentation, detection, and recognition,” in Proceedings of IEEE International Conference on Computer Vision, 2003, pp. 1825. | S. Zhu and D. Mumford, “A stochastic grammar of images,”Foundations and Trends in Computer Graphics and Vision,vol. 2, no. 4, pp. 259362, 2006. | J. D. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditionalrandom fields: Probabilistic models for segmenting and labelingsequence data,” in Proceedings of International Conference onMachine Learning, 2001, pp. 282289. | X. He, R. S. Zemel, and M. A. Carreira-Perpi˜nιιn, “Multiscaleconditional random fields for image labeling,” in Proceedingsof IEEE Computer Society Conference on Computer Vision andPattern Recognition, 2004, pp. 695702. | J. Shotton, J. M. Winn, C. Rother, and A. Criminisi, “Texton-Boost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation,” in Proceedings ofEuropean Conference on Computer Vision, 2006, pp. 115