ISSN:2320-0790

An International Journal of Advanced Computer Technology

Survey On Segmentation And Recognition Of Categorized Objects

Nithina O, Prasanth Kumar P V

PG scholar, Asst.Professor.
Department of Computer Science
Vimal Jyothi Engineering College, Kannur University

Abstract: Object recognition is the task of finding and identifying objects in an image or video sequence. Object categorization is a typical task of computer vision which involves determining whether or not an image contains some specific category of object. The idea is related with recognition, identification, and detection. The object recognition problem can be defined as a labeling problem based on models of known objects. This is closely tied to the segmentation problem. Without at least a partial recognition of objects, segmentation cannot be done, and without segmentation, object recognition is not possible. Object recognition is generally posed as the problem of matching a representation of the target object with the available image features, while rejecting the background features. This paper compares various image segmentation methods and recognition of categorized objects from noisy web image collection using auto-context model.

Key Words: Image segmentation, object recognition, segmentation of categorized objects, auto-context model.

1. Introduction

Object recognition is the task of finding and identifying objects in an image or video sequence. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes and scales or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed from view. Algorithmic description of this task for implementation on machines has been very difficult. The object recognition problem can be defined as a labeling problem based on models of known objects. Formally, given an image containing one or more objects of interest (and background) and a set of labels corresponding to a set of models known to the system, the system should assign correct labels to regions, or a set of regions, in the image. The object recognition problem is closely tied to the segmentation problem. Without at least a partial recognition of objects, segmentation cannot be done, and without segmentation, object recognition is not possible.

Final task for image processing system will be to take an object region in an image and classify it thereby recognizing it. Generate a collection of classes, such as `bird', `boat', `bear', etc. Such a mechanism is called a classier. Many types of features are used for object recognition. Most features are based on either regions or boundaries in an image. It is assumed that a region or a closed boundary corresponds to an entity that is either an object or a part of an object.

Global features usually are some characteristics of regions in images such as area (size), perimeter, Fourier descriptors, and moments. Global features can be obtained either for a region by considering all points within a region, or only for those points on the boundary of a region. In each case, the intent is to find descriptors that are obtained by considering all points, their locations, intensity characteristics, and spatial relations.

Local features are usually on the boundary of an object or represent a distinguishable small area of a region. Curvature and related properties are commonly used as local features. The curvature may be the curvature on a boundary or may be computed on a surface. The surface may be an intensity surface or a surface in 2.5-dimensional space. High curvature points are commonly called corners and play an important role in object recognition. Local features can contain a specific shape of a small boundary segment or a surface patch. Some commonly used local features are curvature, boundary segments, and corners.

Relational features are based on the relative positions of different entities, regions, closed contours, or local features. These features usually include distance between features and relative orientation measurements. These features are very useful in defining composite objects using many regions or local features in images. In most cases, the relative position of entities is what defines objects. The exact same feature, in slightly different relationships, may represent entirely different objects.

2. Image Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). 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. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. 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 characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like marching cubes. Several general-purpose algorithms and techniques have been developed for image segmentation. To be useful, these techniques must typically be combined with a domain's specific knowledge in order to effectively solve the domain's segmentation problems.

A good initial segmentation is necessitated to obtain a good final segmentation result. To measure the quality of an initial segmentation [19] identified three useful measures:

(1) Connectivity, which requires the initial region to be a single region with closed contour.

2) Convexity, which requires the contour of the region to be convex.

3) Saliency, which indicates that the most salient region shall be more likely to be a good initial segmentation.

3. Categorization of Objects

Object categorization is a typical task of computer vision which involves determining whether or not an image contains some specific category of object. The idea is closely related with recognition, identification, and detection. Appearance based object categorization typically contains feature extraction, learning a classifier, and applying the classifier to new examples. There are many ways to represent a category of objects, e.g. from shape analysis, bag of words models, or local descriptors such as SIFT, etc. Examples of supervised classifiers are Naive Bayes classifier, SVM, mixtures of Gaussians, neural network, etc.

4. Context Based Object Recognition

The ability of humans to recognize thousands of object categories in cluttered scenes, despite variability in pose, changes in illumination and occlusions, is one of the most surprising capabilities of visual perception, still unmatched by computer vision algorithms. Object recognition is generally posed as the problem of matching a representation of the target object with the available image features, while rejecting the background features. In typical visual-search experiments, the context of a target is a random collection of distracters that serve only to make the detection process as hard as possible. However, in the real world, the other objects in a scene are a rich source of information that can serve to help rather than hinder the recognition and detection of objects.

Traditional approaches to object categorization use appearance features as the main source of information for recognizing object classes in real world images. Appearance features, such as color, edge responses, texture and shape cues, can capture variability in objects classes up to certain extent. In face of clutter, noise and variation in pose and illumination, object appearance can be disambiguated by the coherent composition of objects that real world scenes often exhibit.

4.1. Types Of Context

Contextual features can be grouped into three categories: semantic context (probability), spatial context (position) and scale context (size). Contextual knowledge can be any information that is not directly produced by the appearance of an object. It can be obtained from the nearby image data, image tags or annotations and the presence and location of other objects.

Semantic context corresponds to the likelihood of an object to be found in some scenes but not others. Hence we can define semantic context of an object in terms of its co-occurrence with other objects and in terms of its occurrence in scenes.

Spatial context can be defined by the likelihood of finding an object in some position and not others with respect to other objects in the scene.

Common approaches to object recognition require exhaustive exploration of a large search space corresponding to different object models, locations and scales. Prior information about the sizes in which objects are found in the scene can facilitate object detection. It reduces the need for multiscale search and focuses computational resources into the more likely scales. This contextual cue establishes that objects have a limited set of size relations with other objects in the scene. Scale context requires not only the identification of at least one other object in the setting, but also the processing of the specific spatial and depth relations between the target and this other object.

5. Auto-context Model

Contextual dependency is one major visual cue in segmentation. To better determine how fit a pixel belongs to foreground or background by including a large amount of contextual information, an auto-context model has been developed for automatic object extraction from images [13]. The auto-context model builds a multi-layer Boosting classifier on image features and context features surrounding a pixel to predict if this pixel is associated with the target concept, where subsequent layer is working on the probability maps from the previous layer. The auto-context model is embedded in the energy minimization formulation, and trained on all images of the categorized image collection in the automatic process of segmentation of categorized objects in [18].

6. Survey On Segmentation of Objects

Segmentation and recognition have long been treated as two separate processes. Stella X. Yuyz, Ralph Grossy and Jianbo Shiy [1] proposed a mechanism based on spectral graph partitioning. G.Saranya, L.M.Varalakshmi and R.Deepa [2] explores a semi supervised optimization model for determining an efficient segmentation of many input images. It can be used to automatically segment a large collection of images that are distinct but share similar features. It is proposed to conduct extensive experiments on various collections of biological images, it will be established that the model proposed is quite computationally efficient and effective for segmentation.

Neill D.F. Campbell et al [3] addresses the problem of automatically obtaining the object/background segmentation of a rigid 3D object observed in a set of images that have been calibrated for camera pose and intrinsic. Such segmentations can be used to obtain a shape representation of a potentially texture-less object by computing a visual hull. It confers improved performance in images where the object is not readily separable from the background in color space, an area that previous segmentation approaches have found challenging.

J. Winn and N. Jojic [4] address the problem of learning object class models and object segmentations from unannotated images. They introduce LOCUS (learning object classes with unsupervised segmentation) which uses a generative probabilistic model to combine bottom-up cues of color and edge with top-down cues of shape and pose. A key aspect of this model is that the object appearance is allowed to vary from image to image, allowing for significant within class variation.

H. Arora, N. Loe, D. A. Forsyth, and N. Ahuja [5] describe an unsupervised method to segment objects detected in images using a novel variant of an interest point template, which is very efficient to train and evaluate. Once an object has been detected, the method segments an image using a conditional random field (CRF) model. This model integrates image gradients, the location and scale of the object, the presence of object parts, and the tendency of these parts to have characteristic patterns of edges nearby. They enhance the method using multiple unsegmented images of objects to learn the parameters of the CRF, in an iterative conditional maximization framework.

E. Borenstein and S. Ullman [6] describe a new approach for learning to perform class based segmentation using only unsegmented training examples. As in previous methods, first use training images to extract fragments that contain common object parts. They show how these parts can be segmented into their figure and ground regions in an automatic learning process. This is in contrast with previous approaches, which required complete manual segmentation of the objects in the training examples

Z. L. Cao and L. Fei-Fei [7] present a novel generative model for simultaneously recognizing and segmenting object and scene classes. This model is inspired by the traditional bag of words representation of texts and images as well as a number of related generative models, including probabilistic latent semantic analysis (pLSA) and latent Dirichlet allocation (LDA). They propose a spatially coherent latent topic model (spatial-LTM). Spatial-LTM represents an image containing objects in a hierarchical way by over-segmented image regions of homogeneous appearances and the salient image patches within the regions.

J. Cui et al. [8] identifies the issue of transducing the object cutout model from an example image to novel image instances. Although object and background are very likely to contain similar colors in natural images, it is much less probable that they share similar color configurations. They propose a local color pattern model to characterize the color configuration in a robust way and an edge profile model to modulate the contrast of the image, which enhances edges along object boundaries and attenuates edges inside object or background. The local color pattern model and edge model are integrated in a graph-cut framework.

B. Alexe, T. Deselaers, and V. Ferrari [9], propose a novel method for unsupervised class segmentation on a set of images. It alternates between segmenting object instances and learning a class model.

L. Mukherjee, V. Singh, J. Xu, and M. D. Collins, [10] develop new algorithms to analyze and exploit the joint subspace structure of a set of related images to facilitate the process of concurrent segmentation of a large set of images. They introduce efficient iterative algorithm (with small computational requirements) whose key steps reduce to objective functions solvable by max-flow and/or nearly closed form identities.

L. Wang, J. Xue, N. Zheng, and G. Hua [11] present a method for automatically extracting salient object from a single image, which is cast in an energy minimization framework. Unlike most previous methods that only leverage appearance cues, they employ an auto-context cue as a complementary data term.

J. Xue, L. Wang, N. Zheng, and G. Hua [12] present a method for automatically extracting salient object from a single image. The method is cast in an energy minimization framework. Unlike that only appearance cues are leveraged in most previous methods, an auto-context cue is used as a complementary data term. They introduce some initial recognition results from the induced auto-context model.

Table I
Segmentation Accuracies Different Methods Tested On Weizmann Horse Dataset

Algorithms	Technique used	Features Selected	Accuracy
[18]	Graph cut (mincut/ maxflow)	Edge	92.9 %
[19]	Region's saliency	Shape, edge	89.3 %
[20]	Cosegmentation	Shape	74.9 %
[21]	Bicos Algorithm for Cosegmentation	Pixels & color distribution	90.0 %
[9]	Inspired by garb cut	Appearance, Jshape & location	86.2 %
[22]	Graph cut	Texture & edges	86.3 %
[23]	A hybrid graph model	Local shape, color/texture	95.9 %

In Table I, the average segmentation accuracy of [] object segmentation method outperforms [4], [7], [13], [27], [37], [39], [40], and compares competitively with LOCUS [1]. The accuracy of the method is lower than [60], and this may be explained by the fact that [60] has strong ability of encouraging segmentation of images along boundaries of homogeneous color/texture.

7. Survey On Auto-Context Models

Auto-context model originally proposed by Tu [13] and later extended by Wang et al. [19] for automatic object extraction from images. The auto-context model builds a multi-layer Boosting classifier on image features and context features surrounding a pixel to predict if this pixel is associated with the target concept, where subsequent layer is working on the probability maps from the previous layer.

To evaluate the importance and usefulness of the auto-context model in the segmentation, [au salient obj ex] implemented two versions of the method. Their only difference is that one includes the auto-context model in the energy function, and the other does not include it.

8. Category Recognition

The goal of image category recognition is to predict whether an image belongs to a certain category. There are a number of recent works on category recognition using various models, such as part-based models and bag-of-words models.

A lot of methods based on bag-of-words model have shown impressive results on image recognition in many settings [24], [25].

9. Conclusion

Object recognition has been one of the most important topics in image processing. In one form or another, it has attracted significant attention. Many approaches have been developed for object segmentation and recognition of categorized images. Automatically extracting and recognizing categorized objects from noisy Web image collections is a difficult task. Auto-Context model is able to automatically extract the object of interest from its background without any user intervention.

12. References

[1] S. Yu, R. Gross, and J. Shi: Concurrent object recognition and segmentation by graph partitioning. in Advances in Neural Information Processing Systems. Cambridge, MA, USA:MIT Press, 2002.

[2] G.Saranya, L.M.Varalakshmi and R.Deepa: Computationally Ecient Segmentation Model for

Collection of Images International Journal of Computer Science Engineering Technology (IJCSET), April 2013.

[3] Neill D.F. Campbell et al: Automatic Object Segmentation from Calibrated Images. 21st ICPR, Nov. 2012, pp. 33093312.

[4] J. Winn and N. Jojic: LOCUS: Learning object classes with unsupervised segmentation. 10th IEEE ICCV, Oct. 2005.

[5] H. Arora, N. Loe, D. A. Forsyth, and N. Ahuja: Unsupervised segmentation of objects using e cient learning. IEEE CVPR,Jun. 2007, pp. 17.

[6] E. Borenstein and S. Ullman: Learning to segment. 8th ECCV,May 2004, pp. 315328.

[7] L. Cao and L. Fei-Fei: Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes. IEEE 11th ICCV, Oct. 2007, pp. 18.

[8] J. Cui et al.: Transductive object cutout. IEEE CVPR, Jun. 2008, pp. 18.

[9] B. Alexe, T. Deselaers, and V. Ferrari: ClassCut for Unsupervised Class Segmentation. 11th ECCV, 2010, pp. 380393.

[10] L. Mukherjee, V. Singh, J. Xu, and M. D. Collins: Analyzing the subspace structure of related images: Concurrent segmentation of image sets 12th ECCV, 2012, pp. 128142.

[11] L. Wang, J. Xue, N. Zheng, and G. Hua: Automatic salient object extraction with contextual cue. IEEE ICCV, Nov. 2011, pp. 105112.

[12] J. Xue, L. Wang, N. Zheng, and G. Hua: Automatic salient object extraction with contextual cue and its applications to recognition and alpha matting. Pattern Recognition., vol.46, no. 11, pp. 28742889, 2013.

[13] Z. Tu, “Auto-context and its application to high-level vision tasks,” in Proc. IEEE CVPR, Jun. 2008, pp. 1–8.

[14] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman: Learning object categories from Googles image search. 10th IEEE ICCV, Oct. 2005, pp. 18161823.

[15] F. Schro, A.Criminisi, and A. Zisserman: Harvesting image databases from the web. IEEE 11th ICCV, Oct. 2007, pp. 18.

[16] N. Sawant, J. Z. Wang, and J. Li: Enhancing training collections for image annotation: An instance-weighted mixture modeling approach. IEEE Transaction on Image Processing, vol. 22, no. 9, pp. 35623577, Sep. 2013.

[17] J. Wu: Efficient HIK SVM learning for image classification. IEEE Transaction on Image Processing, vol. 21, no. 10, pp. 44424453, Oct. 2012.

[18] Le Wang, Gang Hua: Joint Segmentation and Recognition of Categorized Objects From Noisy Web Image Collection. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 9, SEPTEMBER 2014.

[19] L. Wang, J. Xue, N. Zheng, and G. Hua: Automatic salient object extraction with contextual cue. IEEE ICCV, Nov. 2011,pp. 105112.

[20] J. C. Rubio, J. Serrat, A. López, and N. Paragios, “Unsupervised cosegmentation through region matching,” in Proc. IEEE CVPR, Jun. 2012, pp. 749–756.

[21] Y. Chai, V. S. Lempitsky, and A. Zisserman, “BiCoS: A bi-level cosegmentation method for image classification,” in Proc. IEEE ICCV, Nov. 2011, pp. 2579–2586.

[22] C. Rother, V. Kolmogorov, and A. Blake, “‘GrabCut’: Interactive foreground extraction using iterated graph cuts,” ACM Trans. Graph., vol. 23, no. 3, pp. 309–314, 2004.

[23] G. Liu, Z. Lin, X. Tang, and Y. Yu, “A hybrid graph model for unsupervised object segmentation,” in Proc. IEEE 11th ICCV, Oct. 2007, pp. 1–8.

[24] L.-J. Li and L. Fei-Fei, “OPTIMOL: Automatic online picture collection via incremental model learning,” Int. J. Comput. Vis., vol. 88, no. 2, pp. 147–168, 2010.

[25] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in Proc.
IEEE CVPR, 2006, pp. 2169–2178.