Detection of man-made-objects based on spatial aggregations

Hanns-F. Schuster
Institute for Photogrammetry
Bonn University
schuster@ipb.uni-bon.de

Abstract

This paper presents a method for detecting complex man-made-objects in images. The detection model is a bayesian net that aggregates cliques of image regions which may cover a complex object. Observable attributes of the regions are derived from a rich symbolic image description containing points, lines and regions as basic features including their relations. The model captures the dependency of the region aggregates on the features and their relations with respect to observability due to occlusions and to perspective deformations. Cliques are classified using MAP estimation. Up to now, the model captures cliques with one, two and three regions which is sufficient for detecting polyhedral objects. The model allows to detect and locate multiple appearances of object classes. The joint distribution of the Bayesian net is determined in a supervised learning step based on images with annotated regions. The method is realized and demonstrated for the detection of building roofs in aerial images.

1. Introduction

Object detection and categorization is an important challange in the area of computer vision. The solution requires methods of statistics, machine learning and basic computer vision algorithms. Graphical models are an increasingly accepted model in this area.

Many current approaches for object detection and categorization learn features which are defined in a global or local way. Some approaches learn fixed sets of features [14] or define cooccurance measures [15]. Other approaches model appearance probabilities of features [16, 2, 3, 4, 12], often equipped with scale invariant feature detectors.

In the domain of building detection and extraction from aerial images there are methods which address the problem often using specialized features of aerial images [5, 13, 11].

The presented approach is using features and their spatial relations as basic observations which are stored in a feature adjacency graph. According to the image model of [6] the extracted homogeneous image regions represent the piecewise smooth regions while points and lines are representing the discontinuities between those regions. Although the discontinuities carry the main information [1] they are spatially tied together via the regions. By doing this it is possible to insert spatial constraints in the semantic net which is used for the reasoning.

This helps to detect objects by limiting the search space and build a cover of interpretation over the image. The cover of interpretation is of special importance if a interpretation of the image is needed instead of categorizing the whole image by finding the most likely class. In this case every relevant Object in the image has to be detected, where relevance is bounded by the minimum size of the feature detection. The system is scale invariant for smaller changes of the scale because only the spatial relations instead of the size or length of the detected features are of importance.

2 Approach

The approach is motivated by the image model described in [6]. This model assumes objects to be homogeneous, piecewise smooth segments in the image. These segments are assumed to be bounded by discontinuities that are piecewise smooth boundary lines and points that are either boundary points of high curvature or junction points.

Most man-made-objects can be modeled in this way if they are mapped at their typical scale.

This approach is based on the homogeneous regions which connect the features spatially. Every region is a seed point for the classification, the points and lines are treated only as observations for these regions. Thus, image parts where are discontinuities but no homogeneous regions are not used for classification in this approach.

Due to occlusions in the image or failures of the feature detectors features may be not or wrong extracted. To overcome these deficiency, a probabilistic approach is chosen. Therefor a bayes net is used to represent the probability of appearance of objects and their features. The net is used to aggregate features that are introduced in the net as observations.

At the moment the approach is still focused on cliques of one, two and three adjacent regions in the feature graph. This is the most simple setup to provide a consistent cover of classification over polyhedral objects.

Inside the cliques the neighbor relations can be modeled as markov random field since the relations are symmetric.

2.1 Model

The probabilistic detection is modeled as a bayes net that consists of three levels. On the leaves on the bottom level of the net are nodes that are instantiated with the observations derived out of the feature adjacency graph. The middle level consists of nodes representing the type of cliques. The top level these are aggregated to the object-node that represents the class. Due to the three types of cliques there are three different bayes nets representing the structure of observations. A bayes net for two- and three-cliques does not only consist of adjacent regions. The neighborhood of adjacent regions delivers several new observations that make bigger cliques more reasonable.

There are observations only concerning single regions like symmetry of the boundary lines, corner points and textural parameters. The adjacency gives observations like the ration of the regions size, symmetry between the regions and their texture and in the ternary neighborhood again ratios and symmetries can be observed.

To deal with nodes of varying dimensionality of the Cliques-node we can treat this dimension as an additional object-node in a dynamical bayes net that represents the probability to represent a certain clique.

The joint distribution represented by the bayes net can be written as

$$p(O, C, R|F) = p(O|C, R, F)p(C, R|F)p(R|F)$$

Where O stands for Objects, C for Cliques , R for Regions and F for Features.

2.2 Learning

The task of learning is to estimate the parameters of the probability densities in the nodes of the bayes net. The goal is to find the best distribution to explain the observations.

To determine the distribution of the bayes net nodes there is a supervised learning step. As input for the learning step images are provided, in which the regions of the extracted feature adjacency graph are annotated. This annotation determines the class of the object represented by the features and is done in this case by a supervisor.

The so annotated images are learnt in a simple bayesian learning step [9] that updates the distribution in every involved node. where the maximum likelihood over the parameters is searched $r_{ml} = arg\ max_r p(O, C, R|F)$.

2.3 Classification

The classification starts with a random region in the feature graph. The derived observations are introduced in the leave nodes of the bayes net and the information is propagated through the net. This produces a hypothesis about the evidence of the region at the top level node of the net. Because there are three types of nets for the different cliques and several possibilities in the feature graph to instantiate these graphs, multiple hypotheses are generated in the top level node. These are rated in a MAP estimation to search for the most probable explanation. The so classified regions are labeled and the algorithm continues until all regions are labeled.

Thus there must be at least two categories to decide between, there is always a background class that models all objects that are not covered by the other classes. The problem here is like in other probabilistic approaches to parametrize the classs distribution.

3 Implementation

The Feature extraction is based on three disjunct programs. First there is a point detector that uses interest points [6]. The line detector fits lines through edge pixels that come out of a thresholded edge filter [7, 10]. As a third step homogeneous regions are detected where the variance in a texture measure computed with color histograms is below a threshold calculated out of the image noise. Thus the three detectors have to work together, they are run in aboves order, where in the following results is made a cut when detecting a feature at a position where a more important feature has been detected.

With the result of the feature detectors the feature adjacency graph is built. The feature adjacency graph is a planar graph with the features as nodes. Vertices in the graph mean that the features are adjacent. This adjacency is computed by the exoskeleton of the features [8]. The vertices are attributed with the distance and variance of the exoskeleton.

The observations which are derived from the feature adjacency graph are e.g. the symmetry of boundary lines, texture parameters, number of boundary lines and holes, number if corner points and number of neighbors for a single region. For a two region neighborhood there are the ratios between size, texture and moments of the regions and the shape and length of the shared boundary. The ternary clique has additional observations like symmetries of the shared boundary lines and ternary ratios of the above features.

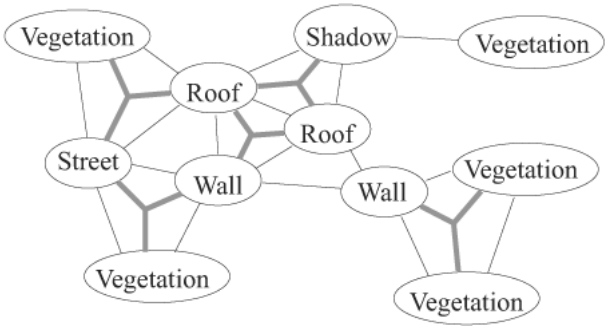

Figure 1: A cut out of the feature adjacency graph. The lines mark the observed two- and three-cliques.

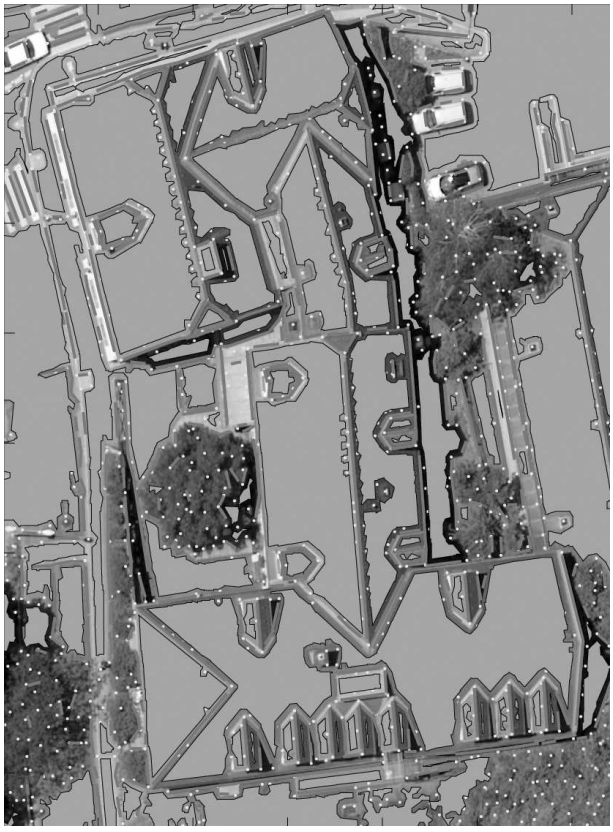

Figure 2: feature extraction on an aerial image with a larger building. The extracted points, lines and regions are shown.

4 Results

The experiments were done with aerial images of suburban and urban regions. The figure 1 shows the extracted points, lines and regions. Figure 2 shows the adjacency between these features. The foundation for learning from which the distribution of the net is estimated is in this case about a square kilometer of suburban houses. The houses extracted by a semiautomatic tool for building extraction and backprojected in the image as a annotated layer of the image. The figure 4 shows results of the classification of two parts of aerial images. The detected categories are house, street and background/vegetation.

5. Conclusions and Outlook

This paper presents an approach for the detection and classification of objects in images that can create a cover of interpretation over the image regions. This is from importance if the detection prepares a reconstruction...

The presented approach is not for the categorization of the image but for the detection of multiple objects in the image. This is for example needed if the following step is a reconstruction. The probabilistic model can handle errors of the feature extraction, of the supervising teacher in the learning step and the perspective deformations.

The extraction of the feature adjacency graph is a well understood and fast implemented algorithm. By inserting the spatial aggregation of features the number of generated hypotheses is kept low in contrast to other methods. This keeps the computational overhead of the bayes net small.

For the future work there will be some tests necessary regarding the speed and detection rate in image databases

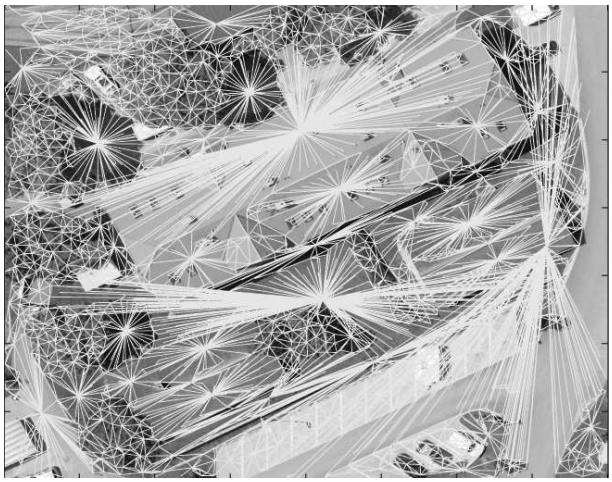

Figure 3: the same image showing the adjacency of the extracted features

to explore the limits of the approach. To make the detection more robust the interest points can be replaced by some scale and rotation invariant feature points.

The next step for the detection and classification will be to expand the bayes net with a dynamic part. With this the net will be able to model complex polyhedral objects even if they are adjacent to each other. It is possible to expand the net until the top level node represents the entire scene. Even if this is in most cases irrelevant due to the high number of possible states, it can be very helpful in the context of the building extraction to know whether the extraction takes place on the country side, on a suburban or urban scene. In the context of an AI-System it would be possible here to choose the algorithms to examine the hypotheses.

References

[1] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.

[2] L. Fei-Fei, R. Fergus, and P. Perona. A Bayesian approach to unsupervised one-shot learning of object categories. In *Proceedings of the 9th International Conference on Computer Vision, Nice, France*, pages 1134–1141, 2003.

[3] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, http://www.robots.ox.ac.uk/ vgg 2003.

[4] Robert Fergus, Pietro Perona, and Andrew Zisserman. A visual category filter for google images. In *ECCV (1)*, pages 242–256, 2004.

[5] André Fischer, Thomas H. Kolbe, Felicitas Lang, Armin B. Cremers, Wolfgang Förstner, Lutz Plümer, and Volker Steinhage. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding: CVIU*, 72(2):185–203, 1998.

[6] W. Förstner. A Framework for Low Level Feature Extraction. In J. O. Eklundh, editor, *Computer Vision - ECCV '94, Vol. II*, volume 802 of *LNCS*, pages 383–394. Springer, 1994.

[7] Wolfgang F"orstner and Bernhard Wrobel. *Mathematical Concepts in Photogrammetry*, pages 15–180. ASPRS, 2004.

[8] C. Fuchs. *Parameterarme Verfahren zur Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung für die Bildsegmentierung*. PhD thesis, Deutsche Geodätische Kommission, München, 1998.

[9] David Heckerman, Dan Geiger, and David Maxwell Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.

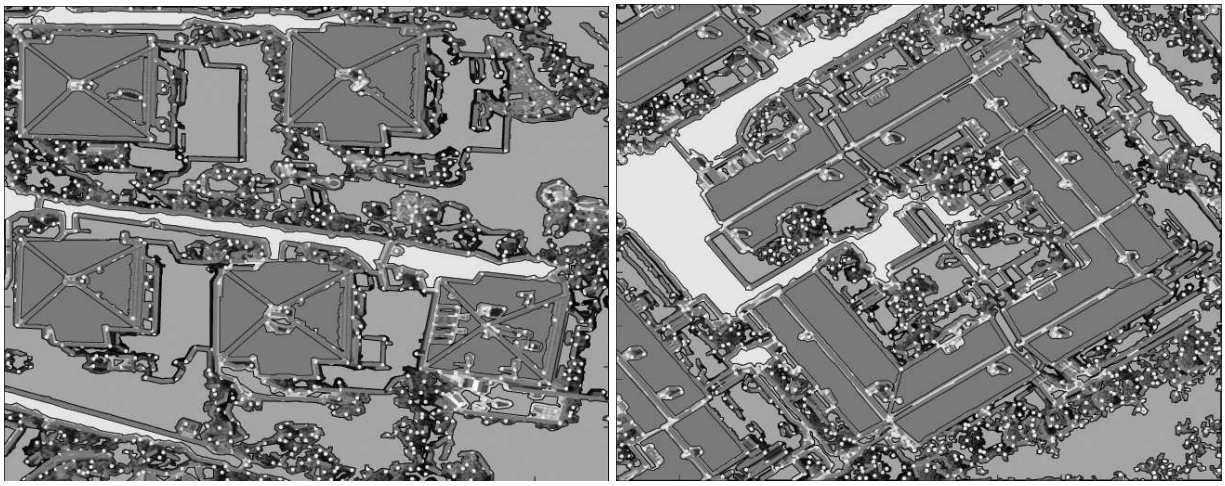

Figure 4: Two parts of aerial images overlayed with the categorization results. The bright regions are classified as street, the dark ones as house. The middle gray is the background class.

[10] Stephan Heuel. *Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction*. Lecture Notes in Computer Science. Springer, 2004.

[11] Zu Whan Kim and Ramakant Nevatia. Learning bayesian networks for diverse and varying numbers of evidence sets. In *Proc. 17th International Conf. on Machine Learning*, pages 479–486. Morgan Kaufmann, San Francisco, CA, 2000.

[12] Bastian Leibe and Bernt Schiele. Scale-invariant object categorization using a scale-adaptive mean-shift search. In *DAGM-Symposium*, pages 145–153, 2004.

[13] H. Mayer. Automatic Object Extraction from Aerial Imagery – A Survey Focussing on Buildings. *Computer Vision and Image Understanding*, 74(2):138–149, 1999.

[14] Paul A. Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR (1)*, pages 511–518, 2001.

[15] Julia Vogel and Bernt Schiele. A semantic typicality measure for natural scene categorization. In *Pattern Recognition Symposium DAGM'04*, Tübingen, Germany, September 2004.

[16] Markus Weber, Max Welling, and Pietro Perona. Unsupervised learning of models for recognition. In *ECCV (1)*, pages 18–32, 2000.