Man-Made Structure Segmentation using Gaussian Processes and Wavelet Features

Hang Zhou and David Suter

Institute for Vision Systems Engineering, Dept Elect. & Comp. Syst. Eng.
PO Box 35, Monash University, Clayton, VIC 3800, Australia
{hang.zhou, d.suter}@eng.monash.edu.au

Abstract

We apply Gaussian process classification (GPC) to man-made structure segmentation, treated as a two class problem. GPC is a discriminative approach, and thus focuses on modelling the posterior directly. It relaxes the strong assumption of conditional independence of the observed data (generally used in a generative model). In addition, wavelet transform features, which are effective in describing directional textures, are incorporated in the feature vector. Satisfactory results have been obtained which show the effectiveness of our approach.

Keywords: Man-made structure segmentation, Gaussian process (GP), Gaussian process classification (GPC) Expectation Propagation (EP), wavelet transform.

1 Introduction

We aim to improve the performance of extracting man-made structure (specifically buildings) from 2D images. Several researchers have studied man-made structure extraction from 2D image data. Oliva and Torralba [1] used principal components of the power spectra to create a low-dimensional holistic representation of the scene. However, it is not suitable for images including both man-made regions and natural scenes. A technique is proposed in [2] for outdoor images classification using colour and texture features. But colour and texture features are not very effective for representing building structures. Hebert and Kumar [3] proposed a hybrid method which extracts generic features from the image blocks and labels image blocks based on the statistical distribution of the features learned from the training data. This Multi-Scale Random Field (MSRF) method yields better results compared with most other approaches.

The problem we observed with the current supervised learning based segmentation, is that it cannot build a model accurate enough. There always exists an overlap between man-made structures and other structure/scene classes, where the model cannot distinguish one class absolutely from another class. This leads to incomplete detection and false positives.

We try to make improvements on two aspects, i.e. 1. replacing the widely used generative model, such as in [3], with a discriminative Gaussian process (GP) and 2. adding some distinguishing wavelet transform features.

The probabilistic generative model used in [2, 3] models the joint probability of the observed data and the related labels where the posterior over the labels is the product of the prior and the likelihood under the Bayes's rule. Data is conditionally independent given the class labels [2] which is not true for man-made structures with obvious neighbouring dependencies. In [3], this dependency is modelled by a pseudo-likelihood approximation.

In contrast, a discriminative model is good at capturing the dependencies between the observations without any model approximations. It directly models what we want, the posterior over labels, without involving the complication of a generative model.

The GP classifier we used in our approach is discriminative, sophisticated, and yet tractable. By learning a function from examples, the man-made structure segmentation can be cast directly into the GP framework.

The remainder of the paper has the following structure. In Section 2, a description is given on Gaussian Process Classification (GPC) as well as an overview of the Expectation Propagation (EP) which is an approximate Bayesian inference technique for GPC. In Section 3, experiment details and results are presented. Section 4 summarizes the main conclusions of the work and discusses possible future extensions.

2 Gaussian Processes for Classification

2.1 Gaussian Process Classification

GP is a collection of random variables, any finite number of which has a joint Gaussian distribution [4]. It is fully specified by its mean function m(x) and covariance function k(x, x') , expressed as:

$$f \sim GP(m,k) \tag{2.1}$$

We use the GPC which as a Bayesian kernel classifier: specifically, as a binary classifier that discriminates between classes labeled as -1/+1.

Suppose we have a dataset D with n observations $D = \{(x_i, y_i) \mid i = 1,...,n\}$, let $X = [x_1,...,x_n]^{\mathrm{T}}$ and $y = [y_1,...,y_n]^{\mathrm{T}}$ be the input training image features and class labels respectively, $y_i \in \{-1,+1\}$.

Given this training data, we wish to make predictions on new class labels y_* for new inputs x_* by calculating $p(y_* \mid D, x_*)$ which is related to a latent function $f(x)$.

GPCs can be represented as graphical models [5] as shown in Figure 1. Class labels y are independent of input X and only depend on the latent function value f . The GP prior with hyperparameters θ is put on the latent function, making f_i and f_* jointly Gaussian.

For binary classification, a sigmoid transformation is applied to the latent function so as to guarantee a valid probabilistic value within the range of [0,1]; We use the probit model $\sigma(z) = \Phi(z)$, where

$$\Phi(z) = \int_{-\infty}^{z} \mathrm{N}(x \mid 0,1)dx \tag{2.2}$$

denotes the cumulative density function of the standard Normal distribution.

2.2 Predictions

Following [6], inference is done by first computing the distribution of the latent variable corresponding to a test case

$$p(f_* \mid X, y, x_*) = \int p(f_* \mid X, x_*, f)\, p(f \mid X, y)df \tag{2.3}$$

where

$$p(f \mid X, y) = p(y \mid f)\, p(f \mid X) / p(y \mid X) \tag{2.4}$$

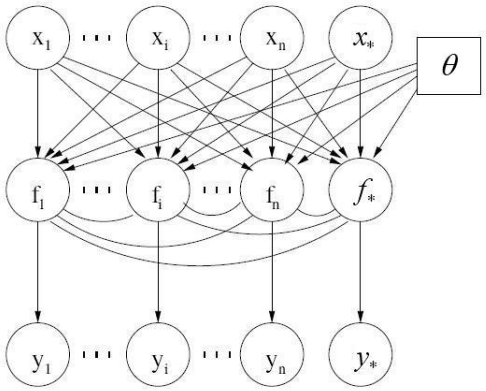

Figure 1. Graphical model for GPCs with n training data points and one test data point [5].

is the posterior over the latent variables,

As described before, the probit likelihood (2.2) is used for binary classification since $p(y_i \mid f_i)$ is not Gaussian.

$$p(y \mid f) = \prod_{i=1}^{n} p(y_i \mid f_i) = \prod_{i=1}^{n} \Phi(y_i f_i) \tag{2.5}$$

$p(f \mid X)$ which is the prior, is chosen to be a zero-mean Gaussian process with a covariance function $k(x, x' \mid \theta)$:

$$p(f \mid X) = N(0, k(x, x' \mid \theta)) \tag{2.6}$$

$p(y \mid X)$ is the marginal likelihood expressed as:

$$p(y \mid X) = \int p(f \mid X) \prod_{i=1}^{n} p(y_i \mid f_i)df \tag{2.7}$$

A probabilistic prediction of the class label is given by:

$$\begin{aligned} \overline{\pi}_* &= p(y_* = +1 \mid X, y, x_*) \\ &= \int \sigma(f_*)\, p(f_* \mid D, x_*)df_* \end{aligned} \tag{2.8}$$

Approximations are needed for integrals (approximate the non-Gaussian joint posterior with a Gaussian one).

2.3 Expectation Propagation Approximation

The EP is an approximate Bayesian inference technique which has been applied to GPC [6].

By comparing EP with other approximation methods such as Laplace method and the sophisticated Markov chain Monte Carlo (MCMC), it has been found that [7] EP is superior to Laplace and quite accurate compared with MCMC.

As stated in [6], the crucial probelm is the posterior distribution over the latent variables $p(f \mid X, y)$. It is given by Bayes's rule, as the product of a normalization term, the prior and the likelihood as shown in eq (2.4). The posterior $p(f \mid X, y)$ is calculated by approximations with a Gaussian joint posterior. The relevant parameters are updated sequentially. Implementation details of the EP algorithm can be found in Algorithm (3.5) and Algorithm (3.6) in [6].

3 Experiments and Results

3.1 Orientogram Features

A feature vector is computed at each 16×16 block to compute the sophisticated features catching the lines and edges patterns in man-made structure, as in [3].

More specifically, a 14 component feature vector [8] is generated at different scales: 1×1, 2×2, and 4×4 blocks. These features are derived from "orientograms" which are the histograms of gradient orientations in a region weighted by gradient magnitudes.

3.2 Wavelet Transform Features

LL	HL
LH	HH

Figure 2. Frequency bands of wavelet transform on the 16x16 image block.

In addition, we experimented with adding 3 wavelet features: using the Haar wavelet transform to capture more directional texture information. Each 16x16 block is decomposed by a one-level Haar transform into four frequency bands as shown in
Figure 2. The three features are computed from HL, LH, HH respectively [9] as in eq (3.1).

$$f = \left(\frac{1}{64}\sum_{i=0}^{8}\sum_{j=0}^{8} C_{i,j}^2\right)^{\frac{1}{2}} \tag{3.1}$$

Where f is one of the three features, $C_{i,j}$ is the wavelet coefficients in each frequency band.

3.3 Results

The proposed approach was trained and tested using images that Kumar [3] used [1]. We used 93 images and the corresponding labels, with all images being cut to 256x256.

The training images are divided into non-overlapping 16x16 pixels blocks which are labeled as one of the two classes, i.e. building or non-building blocks.

The 93 training images originally contain 1813 building blocks and 20594 non-building blocks. There exists redundancy in the non-building blocks, such as sky, ground and vegetations, etc. In our test, a training set containing 348 building blocks and 788 non-building blocks were used by selecting the typical scene blocks.

We run Rasmussen and Williams's [6] program[2]. It employs the EP algorithm for binary GPC. The squared exponential covariance functions with isotropic distance measure is used.

$$k(x, x' \mid \theta) = \sigma^2 \exp(-\tfrac{1}{2}\|x - x'\|^2 / \ell^2 \tag{3.2}$$

where $\theta = [\sigma, \ell]$. We refer to σ^2 as the signal variance, and to ℓ as the characteristic length-scale. We set these as $\log(\ell) = 2.5, \quad \log(\sigma) = 0$ (which seem effective for the scale of data we experimented with).

Figure 3 shows some of the test images together with GPC predicative probabilities and segmentation results as well as Kumar's results. It can be seen that the predicative probability values of the building blocks are reasonably well separated from the rest of the scene block values. The GPC segmentation results tend to cover more building blocks and have less false detections compared with Kumar's results.

Detection rate and false positives are compared in Figure 4 with test results on 12 images. Each point in the chart represents one of the 12 images. (a) indicates that with much less training data, our solution has a similar detection rate as Kumar's which implies GP plus wavelet features converge well on sparse data. (b) shows a with/without wavelet features comparison on detection rate. Obviously, the detection rate is better by using wavelet features. (c) and (d) illustrate that the false positives are all similar either with or without wavelet features.

4 Conclusions and Future Work

In this paper, we have proposed an effective method for segmentation of man-made structure from natural scene. Satisfactory results are achieved on small training set by

[1] http://www.cs.cmu.edu/~skumar/manMadeData.tar
[2] http://www.gaussianprocess.org/gpml/code/matlab/doc/classification.html

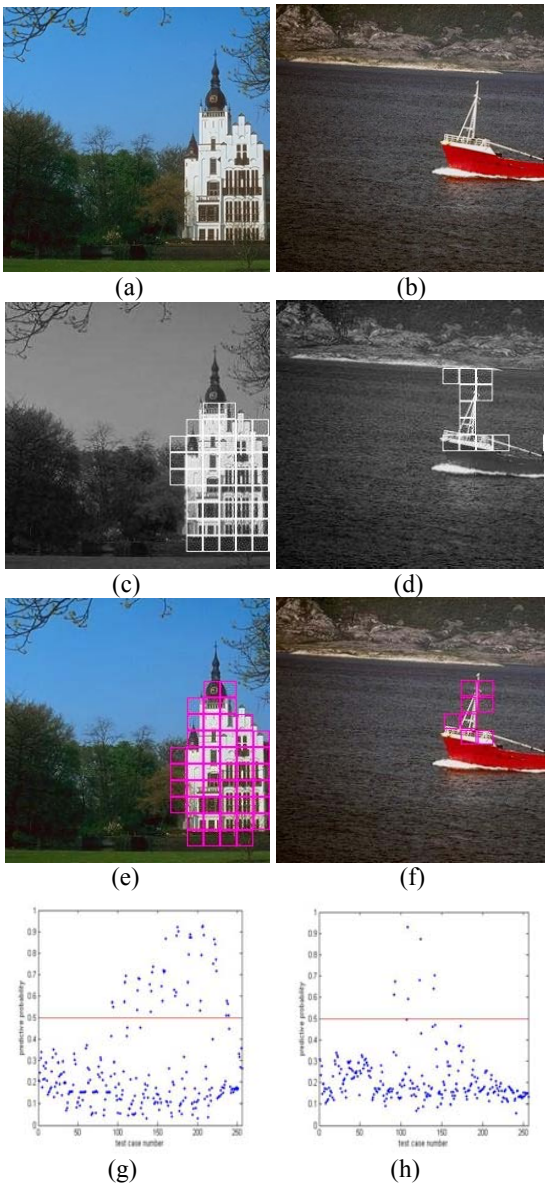

Figure 3. Segmentation results. (a)(b) original images (c)(d) Kumar's results (e)(f) our GP segmentation results (g)(h) predicative probabilities of 256 blocks in each image.

using the discriminative GP model as well as wavelet features. Our experiments have positively indicated that GPC has the potential to separate man-made structures from natural scenes. Wavelet features can well reflect the directional texture properties and are effective in enhancing the detection rate.

The performance of our solution can be improved in several aspects. More work should be done to better organize the training data coverage. Features of the model can be further optimized so as to be more discriminative between buildings and non-buildings. In the meantime, investigations can also be done on the use of different covariance functions and efficient calculation approximations for large datasets.

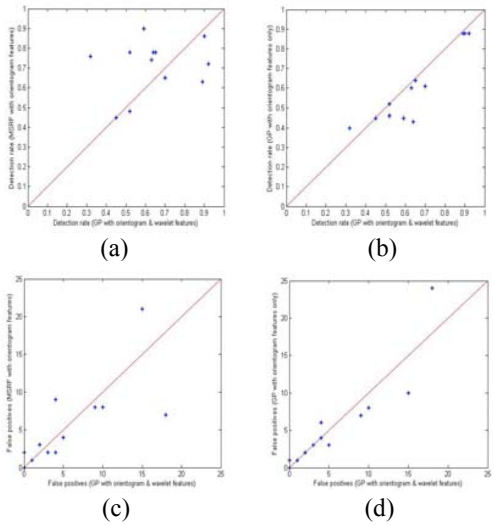

Figure 4. Detection rate (DR) and false positives (FP) comparison. (a) DR of GP+ wavelet vs Kumar's MSRF (b) DR of GP+wavelet vs GP w/o wavelet (c) FP of GP+ wavelet vs Kumar's (d) FP of GP+wavelet vs GP w/o wavelet.

5 References

[1] A. Oliva and A. Torralba, "The Shape of the Scene: a Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision,* vol. 42, pp. 145-175, 2001.

[2] X. Feng, C. K. I. Williams, and S. N. Felderhof, "Combining Belief Networks and Neural Networks for Scene Segmentation," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 467-483, 2002.

[3] S. Kumar and M. Hebert, "Man-Made Structure Detection in Natural Images using a Causal Multiscale Random Field," in *CVPR2003*, 2003, p. 119.

[4] C. E. Rasmussen, "Gaussian Processes in Machine Learning," *Advanced Lectures on Machine Learning,* 2003.

[5] H.-C. Kim and Z. Ghahramani, "The EM-EP algorithm for Gaussian process classification," in *Workshop on Probabilistic Graphical Models for Classification*, 2003.

[6] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*: The MIT Press, 2006.

[7] M. Kuss and C. E. Rasmussen, "Assessing Approximations for Gaussian Process Classification," *Advances in Neural Information Processing Systems,* pp. 699-706, 2006.

[8] C. Pantofaru, R. Unnikrishnan, and M. Hebert, "Toward Generating Labeled Maps from Color and Range Data for Robot Navigation," in *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2003.

[9] J. Z. Wang, *Integrated Region-Based Image Retrieval*: Kluwer Academic Publishers, 2001.