This is the Author's version of the paper published as:

**Title:** Region Based Maximum Likelihood Estimation for Small Geospatial Object  
**Author:** D. Tien and X. Yi  
**Author Address:** dtien@csu.edu.au  
**Conference Name:** Digital Image Computing Techniques and Applications  
**Year of Conference:** 2007  
**Conference Location:** Adelaide  
**Editor:** M. Bottema  
**Pages:** e.g. pp73-78  
**URL:** http://dicta2007.infoeng.flinders.edu.au/  
**Keywords:** Remote Sensing, SVM, Image extraction  

**Abstract:** The availability of high resolution aerial imagery makes it possible to identify small geospatial objects in dense urban areas. The challenge lies in the overlapping problem in the feature space of the object classes. In this paper, we propose a region-based maximum likelihood (RML) method for geospatial object extraction in urban areas. For accurate extraction of the object region, the target objects are segmented by SVM classification. And the colour distribution is subsequently compared against the SVM training data via a Mahalanobis colour distance and a ML approach is developed to distinguish between regions that might be overlapped in their feature space. A quantitative measure which evaluates the resulting extractions is presented. The experimental results show that the proposed approach yields intuitively correct as well as accurate extraction of objects in urban aerial images.
Region Based Maximum Likelihood Estimation for Small Geospatial Object Extraction

David Tien and Yi Xiao
School of Information Technology, Charles Sturt University
Panorama Avenue Bathurst NSW 2795 Australia
yix@ee.usyd.edu.au

Abstract

The availability of high resolution aerial imagery makes it possible to identify small geospatial objects in dense urban areas. The challenge lies in the overlapping problem in the feature space of the object classes. In this paper, we propose a region-based maximum likelihood (RML) method for geospatial object extraction in urban areas. For accurate extraction of the object region, the target objects are segmented by SVM classification. And the colour distribution is subsequently compared against the SVM training data via a Mahalanobis colour distance and a ML approach is developed to distinguish between regions that might be overlapped in their feature space. A quantitative measure which evaluates the resulting extractions is presented. The experimental results show that the proposed approach yields intuitively correct as well as accurate extraction of objects in urban aerial images.

1. Introduction

Extracting geospatial objects from aerial imagery is much desired in urban development planning, emergency response, and Earth survey. The development of very high resolution (VHR) aerial imagery make it possible to identify small structures such as small houses, individual trees and swimming pools in dense urban areas.

For aerial imagery analysis, many classification techniques based on the spectral analysis of individual pixels have been proposed [1]. Classification algorithms, based on a statistical approach such as maximum likelihood [1] or a neural network have been frequently used [2-4], as well as possibility models or fuzzy logic for a fuzzy classifier [5, 6], and Nearest Neighbour method [7]. With the aim of achieving higher accuracy, integration of additional data has been investigated such as multi-source data fusion approach and contextual classification models [8-12].

However, even with SVM classification, overlapping samples still cause a serious problem to the classification task. In classical classification, each sample is treated equally in its appearance in an object region. However, in our application, some training samples, such as the outliers, do not appear more frequently in the target region than samples with values near the mean of the object class. We name these samples the core pixels. In this paper, we apply SVM classification for target object region extraction. To overcome the overlapping problem mentioned above, we propose a novel region based maximum likelihood approach (RML) to identify if a region classified by SVM is a target object.
region. For this, the colour distribution of the region is analysed and the colour pixels contributing to object identification are counted. To distinguish the colour samples (core pixels) contributing to object identifying and other samples, their colour mahalanobis distance is calculated and a maximum likelihood rule is applied to judge if the region is the target object region.

2. Object region segmentation by SVM classification

Given samples \((y_i, x_i), x_i \in \mathbb{R}^n, y_i \in \{-1, +1\}\) \(i = 1, ..., l\) (where \(l\) is the number of samples, \(x_i\) is the sample, and \(y_i\) is the class label of \(x_i\)), and a kernel function \(k(x_i, x_j)\), SVM is formed as solving the quadratic programming problem [15, 16]:

\[
\begin{align*}
\alpha = \arg \min & \quad \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{l} \alpha_i \\
\text{subject to} & \quad 0 \leq \alpha_i \leq c, \quad i = 1, ..., l \\
& \quad \sum_{i=1}^{l} \alpha_i y_i = 0
\end{align*}
\]

(1)

c is a parameter which trades off wide margin with a small number of margin failures.

All the \(x_i\) corresponded to non-zero \(\alpha_i\) are the Support Vectors (SVs). The classifier is

\[
f(x) = \text{sign} \left( \sum_{i \in SV} \alpha_i y_i K(x_i, x) + b \right)
\]

(2)

Where \(b = -\frac{1}{2} \sum_{i \in SV} \alpha_i y_i [K(x_i, x_i) + K(x_i, x_i)]\), \(x_i, x_i\) are different types of SVs.

In the case of linear SVM, kernel function in the equation (2) is replaced by:

\[
K(x_i, x_j) = (x_i \cdot x_j)
\]

(3)

For non-linear SVM with Gaussian kernel, the kernel function is as follows:

\[
K(x_i, x_j) = e^{-\frac{1}{2\sigma^2} \|x_i - x_j\|^2}
\]

(4)

As we can see from formula (3) (4), the speed of nonlinear SVM is directly related to the number of SVs. SVM is fundamentally developed for such binary classification case and is extendable for multi-class situation.

We collected a variety of colour samples of a number of classes from a dataset for SVM training. Pixels in the test images of the dataset are classified by SVM classification and regions are extracted from the connected components of the pixels in the same classes classified by SVM.

3. Core pixel calculation

Core pixels of a target object are distinguished from the remaining pixels in a region by Mahalanobis colour distance measurement. That is, core pixels are defined as pixels with values within a defined distance (Mahalanobis) of the mean of the object class.

\[
d_i = ((x_i - t)^T C^{-1}(x_i - t))^{1/2} \quad \text{for} \quad i = 1, ..., n
\]

(5)

Core pixels are obtained as

\[
\{ | \quad ore \quad i \quad \alpha \quad y \quad d \quad d \quad c \quad r \quad e \quad s \quad y \quad y \quad C \quad y \quad = \quad \frac{1}{\sum_{j \in \text{ore}} y_j} \quad \text{for} \quad x_i \in r .
\]

(6)

\(n\) is the number of training samples of the target object. \(r\) is a candidate region classified by SVM.

4. Maximum likelihood recognition

In our approach, maximum likelihood (ML) estimation is applied to each candidate region in the segmentation.

Let \(\omega_i, i = 1, ..., M\) represent the object classes. The class likelihood functions are denoted by \(p(r | \omega_i)\), and modelled by the ratio of the pixel number that belong to the core pixels for \(\omega_i\) in the region \(r\) over the total number of the pixels in the region, i.e.,

\[
p(r | \omega_i) = \frac{1}{N} \sum_{j=1}^{N} y_j \quad \text{where} \quad N \quad \text{is \ the \ total \ number \ of \ the \ pixels \ in \ the \ region, \ i.e.,} \\
p(r | \omega_i) = \frac{1}{N} \sum_{j=1}^{N} y_j \quad \text{where} \quad N \quad \text{is \ the \ total \ number \ of \ the \ pixels \ in \ the \ region, \ i.e.,}
\]

\[
\omega_k \iff p(r | \omega_i) \geq p(r | \omega_k) \forall k \neq i, k = 1, ..., M
\]

(7)

5. Experiments
For performing experiments on aerial imagery, a dataset provided by the department of Land, Australia, has been used. The dataset covers the areas of Mudgee in New South Wales, Australia. The configuration of the dataset is shown in Table I.

**Table I Image configuration**

<table>
<thead>
<tr>
<th>Image</th>
<th>Resolution</th>
<th>Covering area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mudgee</td>
<td>0.2m</td>
<td>5.2 × 4.7 (km × km)</td>
</tr>
</tbody>
</table>

In experiments using our method, a sub-image for training samples has been chosen from the original image. It contains five major classes in a number of training samples employed for SVM classification. The training samples are collected from a variety of regions of the target objects. The number of samples corresponding to these classes is given in Table II.

**Table II Training samples**

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of Pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swimming pool</td>
<td>420</td>
</tr>
<tr>
<td>Light grey and green roof</td>
<td>186</td>
</tr>
<tr>
<td>Light blue and brown roof</td>
<td>172</td>
</tr>
<tr>
<td>Red and brown roof and soil</td>
<td>315</td>
</tr>
<tr>
<td>Tree and grass</td>
<td>784</td>
</tr>
</tbody>
</table>

The overlap between classes for training samples is estimated using the transformed divergence measure [17]. The transformed divergence distance measures are real values between 0 and 2, where ’0’ indicates complete overlap between the two classes and ’2’ indicates a complete separation between the two classes. The results are given in Table III. It can be seen that swimming pool and the light grey and green roof or light blue and brown class pair is overlapped.

**Table III The measurement of colour overlap between the classes – transformed divergence**

<table>
<thead>
<tr>
<th>Class</th>
<th>Swimming pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light grey and green roof</td>
<td>1.6</td>
</tr>
<tr>
<td>Light blue and brown roof</td>
<td>1.5</td>
</tr>
<tr>
<td>Red and brown roof and soil</td>
<td>2</td>
</tr>
<tr>
<td>Tree and grass</td>
<td>2</td>
</tr>
</tbody>
</table>

The overlap between classes for training samples is estimated using the transformed divergence measure [17]. The transformed divergence distance measures are real values between 0 and 2, where ’0’ indicates complete overlap between the two classes and ’2’ indicates a complete separation between the two classes. The results are given in Table III. It can be seen that swimming pool and the light grey and green roof or light blue and brown class pair is overlapped.

The overlap between classes for training samples is estimated using the transformed divergence measure [17]. The transformed divergence distance measures are real values between 0 and 2, where ’0’ indicates complete overlap between the two classes and ’2’ indicates a complete separation between the two classes. The results are given in Table III. It can be seen that swimming pool and the light grey and green roof or light blue and brown class pair is overlapped.

The SVM used in these experiments are based on OSU Support Vector Machines Toolbox [18]. The linear kernel provided efficient performance. Since the classification accuracy in our experiments were not significantly different in terms of different values for c, we adopted the regularization parameter c=1.

We focus on the results and the discussion of swimming pool extraction. The swimming pool regions, together with some light grey and green roof regions were separated from others by SVM classification. The Maximum likelihood estimation on these regions further excluded the light grey and green roofs.

In this paper, we utilise the measure and the criteria of completeness and correctness [19] to assess the accuracy of the object extraction. The extraction results are classified as true positive (TP), true negative (TN) or false positive (FP). TP is the number of objects detected manually and by the proposed method, FP is the number of detected by the proposed method but not manually, and TN is the number of objects detected manually but not by the proposed method. completeness and correctness are defined as follows:

$$\text{completeness} = \frac{TP}{TP + TN}, \quad \text{correctness} = \frac{TP}{TP + FP}$$

completeness measures the ability to find the object regions and correctness measures the accuracy of object detection.

![Fig. 1](image_url)  
Fig. 1 completeness and correctness of our method in terms of $d_{th}$

Fig. 1 shows the completeness and correctness of our method in terms of $d_{th}$ in equation (6) for swimming pool extraction. As the small value of $d_{th}$ (<4) can not include all the pool pixels, completeness is less than 1. Some pool regions are missed. When $d_{th}$ increases, completeness increases, all the pools are extracted but some light grey roofs are also included, which leads the decreasing of correctness. The threshold value for core pixel calculation is set to $d_{th} = 5$ for swimming pool extraction.
Fig. 2 illustrates the extraction procedure.

**Fig. 2 (a)** SVM classification. (swimming pool; light grey and green roof; red roof and brown soil; tree and grass; light blue and brown roof)

**Fig. 2 (b)** Candidate pool regions and the core pixels for swimming pools in the regions.

**Fig. 2 (c)** Extracted swimming pools (Swimming pool boundaries are highlighted.).

In the experiment, 294 out of 316 pools were detected, 74 false alarms occurred. False alarms were caused by some blue roofs that shared colours with swimming pools.

The performance of the proposed RML algorithm has been compared with the conventional algorithm of Mahalanobis colour distance (MCD) [20] and nearest neighbour (NN) algorithm. The three methods are trained and tested on the same datasets. Fig. 3 shows an example of the results with the three different methods. The accuracy of extraction results produced from the best performances of our RML, the MCD and the NN algorithm are presented in Table IV, where the threshold value for MCD is set to $d_m = 10$.

<table>
<thead>
<tr>
<th>Method</th>
<th>completeness</th>
<th>correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mahalanobis distance</td>
<td>0.93</td>
<td>0.47</td>
</tr>
<tr>
<td>Nearest neighbour</td>
<td>0.90</td>
<td>0.5</td>
</tr>
<tr>
<td>Our method</td>
<td>0.93</td>
<td>0.98</td>
</tr>
</tbody>
</table>

It can be seen, that the performed on the swimming pool extraction from RML and MCD leads to the highest completeness (i.e., 93%) as compared to those achieved by NN. NN covered less object appearance variation than RML and MCD did. The correctness obtained by RML is significantly higher than the others and the correctness achieved by MCD is the lowest when compared to the correctness achieved by RML and NN. The high correctness of RML in extracting the object indicates its ability to handle the overlap. Thus, the results clearly show that the RML algorithm outperforms the conventional MCD and NN algorithms for the extraction of swimming pool in this experiment.

**Fig. 3 (a)** Extracted swimming pools with RML method.

**Fig. 3 (b)** Extracted swimming pools with MCD method.
6. Conclusions

The primary issue that motivated this research is the limitation of traditional supervised algorithms, for classifying (identifying) high resolution aerial images. The challenge is the overlap problem caused by the increasing of intra-class distance and decreasing of inter-class distance. In this paper, we have proposed a region-based maximum likelihood (RML) method to improve the misclassification of SVM due to the overlap. Through our experiments, we have demonstrated the effective application of the proposed algorithm in identifying target objects. The proposed region based maximum likelihood algorithm produced significantly higher accuracy as compared to the Mahalanobis colour distance (MCD) and nearest neighbour (NN) algorithm.

ACKNOWLEDGEMENT

The work has been supported by the Cooperative Research Centre for Spatial Information, whose activities are funded by the Australian Commonwealth's Cooperative Research Centre Program.

References


[16] Haizhou Ai, Lihang Ying, Guangyou Xu, A Subspace Approach to Face Detection with Support Vector Machines,

