Automatic Rooftop Detection Using a Two-stage Classification

Bikash Joshi, Hayk Baluyan, Amer Al. Hinai, Wei Lee Woon
Department of Electrical Engineering and Computer Science (EECS), Abu Dhabi, UAE
{bjoshi, bhayk, wwoon, aalhinai}@masdar.ac.ae

Abstract—This paper presents a novel application of machine learning techniques to the automatic detection of building rooftops in satellite images. The image is first segmented into homogeneous regions using the k-means algorithm. These segments are then treated as candidate rooftop regions which are presented to a novel two-stage classification process; features are extracted from each segment and submitted to an ANN which serves as the first stage of the classification procedure. New features are then extracted from the outputs of the ANN and these are presented to an SVM which then performs the second classification pass. In this way, the first classification stage acts as a preprocessing step which, when processed by the SVM significantly reduces the number of false-positives. To establish the efficacy of the proposed method, its results are compared with those obtained using an alternative approach.

Index Terms—Machine Learning; Rooftop Detection; Artificial Neural Network; Support Vector Machine; Image Segmentation; Computer Vision

I. INTRODUCTION

Rooftop detection from satellite/aerial images is important in a variety of applications. Some examples of which are change detection in urban monitoring, the production of digital maps, verification and updating of GIS databases, land use analysis and route planning, military applications et.al [1][2]. For example accurate identification and localization of rooftops in urban images is a key step in territorial planning and city modeling. Similarly, identification of building rooftops helps to gain knowledge of the location, profile and density of buildings of a particular region. Such knowledge can be very useful in estimating the distribution of city’s population.

The detection of building rooftops is a tedious task. While manually detecting building rooftops from aerial images can be highly accurate, the effort required increases significantly with the size and complexity of the images. Hence, many techniques for automatic rooftop detection have been proposed. The main aim of these techniques is to perform the rooftop detection with minimal human interaction while maximizing detection accuracy. However, for a variety of reasons, the automatic rooftop detection is also a very challenging task [3]. The main reason is that the images used are different in terms of lighting conditions, quality and resolution. Another reason is that buildings can be of diverse and complicated shapes and structures, which can easily be confused with similar objects present in the image such as: cars and roads. In particular, there are currently no “universal” algorithms for detecting rooftops i.e. algorithms which can identify rooftops in all or most of the aerial or satellite images.

Computational solutions for rooftop detection are based on image processing operations such as edge detection, corner detection and image segmentation. One basic approach is to first generate rooftop candidates using image segmentation techniques and then to identify true rooftops using features like shape, area and the presence of shadows [4], [5]. Machine learning methods have also been gaining in popularity, where Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have been widely used. In [6] an ANN is used to facilitate effective rooftop identification in complex environments and it was able to distinguish small and bright buildings as well. Also, in [7] ANN is used for the detection of buildings, agricultural land, forest and water from satellite imagery. In [8] a review is presented of ANN based approaches in various related areas including medical image processing, biometric patterns and gestures extraction, letters and characters detection, edge and contour detection. For each of these applications ANNs performed reasonably well, thus demonstrating its potential for solving image processing tasks. Most of these approaches are based on pixel-level segmentation, which assign each pixel to a given segment based on features generated for the pixel in question. However this approach does not perform well for rooftop detection; in particular these methods fail to detect rooftops when applied to test images containing objects (e.g. cars, roads) which are of the same color as rooftops.

To address these apparent shortcomings, we propose here a novel method which is based on two consecutive classification stages. First, an ANN is used to classify candidates segments into rooftop and non-roof. New features are then extracted from the outputs of the ANN which are based on rooftop properties inferred using results gathered over the entire image. The second classification stage is then performed on these features using an SVM. While this is not unique, we also note that there appear to be relatively few studies in which SVMs are used for object detection (as in [9] and [10]). Also, in [11] SVM is used for rooftop detection from satellite images.

Dividing the classification process into two separate stages appears to reduce the number of false positives, resulting in a significant improvement in the performance of the model in comparison with traditional approaches.

This approach of taking information from the classification result and using it for final prediction worked successfully in our previous work [12] as well. In this new approach we extend our previous work and perform the final prediction based on another pass of classification. When compared to our previous result in [12], addition of another pass of classification helped...
to detect many of the missed rooftops while keeping the
number of misclassified non-rooftop regions low. This resulted
in a high recall and low false-positive in the result as compared
to our previous result. From the obtained result we can observe
that second pass of classification significantly reduces the false
positives in the result. We believe that this unique combination
of two separate classification stages significantly improves
existing rooftop detection capabilities. Furthermore, the main
motivation of our research on rooftop detection is to assess
the overall rooftop area available in Abu Dhabi, UAE for
deployment of photovoltaics. Hence, we have used images
from this specific area to evaluate our model.

II. THE METHODOLOGY

The proposed methodology consists of three main steps:

1) Image Segmentation: First of all, we divide the images
into a set of candidate regions. K-means clustering
algorithm is used to perform this segmentation task,
which divides the pixels of the image into a number
of clusters based on the intensity values. Then, a flood-
fill algorithm is applied to group the segmented pixels
into a set of connected regions. Each of these regions
are used as rooftop candidates.

2) First classification stage: From each candidate region, 10
features are extracted (feature extraction is discussed in
more detail in next subsection). We formed one dataset
of the extracted features, where each row corresponds to
one candidate region. The ANN classifier is then trained
to distinguish between rooftop regions and non-rooftop
regions.

3) Second Classification Stage: The ANN in the first clas-
sification stage is able to detect many of the rooftops.
However, in practice there were many rooftop regions
which were either not detected or were misclassified.
A second classification stage using SVM helps to de-
tect these initially missed rooftops and reduces false-
positives in the first classification result. This second-
pass of classification uses information from the result of
first-pass of classification.

The overall process is shown in Figure 1. We will discuss each
of these steps in greater detail in the following subsections.

A. Image Segmentation

The goal of image segmentation is to divide the image into
various segments based on its color properties, so that each of
these segments will be used as a candidate region which can be
classified as either rooftop or non-rooftop region. In order to
perform this segmentation, we first enhanced the image using
bilateral filtering, then we performed K-means clustering of
the enhanced image and finally we extracted the connected
components in the clustered image such that each of these
connected regions can be considered as candidate regions.

1) Bilateral Filtering: Bilateral filter [13] is a non-linear
noise reducing smoothing filter which preserves edges in the
image. It combines two filtering approaches: domain filtering,
which helps to enforce closeness by weighing pixel values
with coefficients that fall off with distance and range filtering,
which averages pixel values with weights that decay with
dissimilarity. The result of applying bilateral filtering on image
and its influence on result of k-means clustering is shown in
Figure 2. From this figure, it is evident that bilateral filtering
resulted in smoother and visually cleaner segments.

![Fig. 2: (a) The original image (b) Original Image after applying bilateral
filtering (c) k-means clustering on the original image (d) k-means clustering
on the filtered image](image_url)
2) K-means clustering: After filtering the image, we perform clustering of image using K-means clustering algorithm. One important consideration, while using K-means clustering, is the choice of appropriate value of number of clusters i.e. k. In our previous work [12] of rooftop detection using histogram method, we found that k=4 worked best for segmenting images using k-means clustering. Since, in this study also we are using images from the same geographical area, we have chosen the value of k as 4.

3) Extracting connected components: In this step, the clustered pixels are grouped into connected regions to form candidate rooftop regions. There are two choices for performing this task: 4-connected flood-fill algorithm and 8-connected flood-fill algorithm. Practically, there was no significant difference between the result of both methods and since 8-connected flood-fill algorithm is computationally more demanding[14], we used 4-connected flood-fill algorithm in our experiments. The result of this step is a set of regions where each pixel in a region is connected to at least one other pixel in the same region in one of the four principle directions.

B. First Classification Stage

This step can be further sub-divided into the following sections:

1) Data Preparation: After performing image segmentation of the training images into candidate regions as in the previous section, we manually labeled the rooftop regions in the image. Then, 10 features are extracted from the candidate regions (feature extraction process is discussed in more detail in the next section) which describe the segments. Thus, we prepared the training dataset, in which each row represents one of the segments and different columns show the corresponding feature values for that segment. We also include the manual label such that it is either 1 (if it is a rooftop) or 0 (if it is a non-rooftop).

2) Feature Extraction: Features are numerical attributes which allow rooftop and non-rooftop regions to be distinguished from each other. In this study, 10 features which are relevant to the classification task at hand are selected.

- Area: This denotes the area of a particular segment in terms of pixels. This feature can distinguish building rooftops from non-rooftop regions, since most of the rooftops have similar area which can be easily distinguished from non-rooftop objects such as cars, trees etc.
- Ratio of Minor to Major Axis Lengths: This is basically the width to length ratio of the segmented regions. Figure 3 shows the minor to major axes of building rooftops represented by thick and thin lines respectively. The axis ratio for buildings are comparable to each other, whereas this is drastically different for elongated non-rooftop objects such as roads, rivers etc.
- Mean Intensity: Mean intensity represents the mean of all grayscale intensity values of all pixels present inside one region. As can be seen in Figure 3, mean intensity value of rooftops in one region are similar to each other, as the color of roofs are mostly similar in one region.

\[ \text{Roundness} = \frac{4\pi \cdot \text{Area}}{\text{Perimeter}^2} \]  
\[(1)\]

The value of roundness varies from 0 to 1; objects such as rooftops have value of roundness close to 1, whereas elongated objects have roundness close to 0.

- Rectangularity: Rectangularity of a segmented region is a shape-based feature [15]. It is calculated as follows:

\[ R = \frac{\text{Area}}{\text{MajorAxisLength} \cdot \text{MinorAxisLength}} \]  
\[(2)\]

- Contrast: Contrast can be defined as the measure of the intensity difference between a pixel and its neighbor pixel over the whole image. Usually rooftop regions in the image are homogeneous, so they tend to have lower contrast.

- Energy: Energy represents the textual uniformity which is given by pixel pair repetitions. In other words, it detects disorderliness in textures. High energy value corresponds to a constant or homogeneous image.

- Homogeneity: Homogeneity measures the closeness of distribution of pixels in an image. Values of homogeneity ranges from 0 to 1. For a perfectly homogeneous image the value of homogeneity is 1. Since, rooftop regions are more homogeneous, so they tend to have a high value of homogeneity.

3) Normalization: After the features have been extracted as described above, each feature is normalized so as to have zero mean and unit standard deviation.

\[ \text{see: http://www.mathworks.com/help/images/ref/regionprops.html#brjijkp-1} \]

\[ \text{see: http://www.mathworks.com/help/images/ref/entropy.html} \]
4) ANN Classification: For this stage, a Multi-Layer Perceptron (MLP) with one hidden layer and ten hidden units was used. Preliminary experimentation with different numbers of hidden units indicated that this was a good choice although the performance did not seem to be very sensitive to this parameter. The output layer contains a single unit. The value of this output unit determines whether a segment is a rooftop or non-rooftop. A sigmoid function is chosen as the activation function for all neurons in all layers.

The numeric features calculated in the previous section are supplied to the neural network for training. During training, network’s output is compared to the actual output and eventually weight coefficients are modified using backpropagation algorithm. This training process repeats until network’s output is able to classify rooftop or non-rooftop regions with high accuracy. This trained neural network is then presented with each test image. The output will then be the predicted labels and class probabilities for each candidate region.

C. Second Classification Stage

In practice, the ANN was not able to detect all of the rooftop regions in the image, and also produced a significant number of false positives. To address this problem, a second classification stage was performed, where the outputs of the first classification stage were treated as an additional preprocessing step. It was found that this additional step helped to greatly improve the overall performance of the system.

1) Data Preparation: Only candidate regions with a class probability exceeding 0.2 were considered (the class probability is basically the output of the first classification stage, and denotes the estimated probability that a candidate region is a rooftop). To prepare the training dataset for the second-pass of classification, same training images were prepared as in the first pass. All those training images are processed with first pass of classification and then used to prepare the training dataset in the second stage. Then, we extracted three features from these candidate regions.

2) Feature Re-extraction: New features are extracted based on the results of first stage. As stated above, features are only extracted for those regions with class probabilities exceeding 0.2. Each of these features are described below:

- Class probability: class probability is the measure of each candidate region actually being a rooftop. Class probability values are obtained by the result of ANN in first-classification stage. Its values range from 0 to 1. rooftops candidates have high class probability value (close to 1) whereas non-roof objects have considerably low class probability value.

- Confidence Value: rooftops in one region are similar to each other. So, the result of first-stage classification shows the most prevalent intensity of rooftop. So, we take a histogram of the rooftops identified by the first-stage classification by dividing it into 8 bins. Then we provide each bin a confidence value based on the number of pixels belonging to that bin. So, for second-stage classification we calculate the confidence value of each segment by checking to which bin it belongs. Ideally, the candidate regions corresponding to rooftop should have a higher confidence value, whereas the non-rooftop regions have a lower confidence value.

- Shadow: Shadow information mostly lies on the blue band of the RGB image [16]. So, shadow regions can be enhanced using the below formula:

\[ I = 4 \pi \arctan \left( \frac{B - G}{B + G} \right) \]

(3)

Where, B and G denote the blue and green bands in the color image respectively. In the image I, shadow regions have high intensity, so they can be segmented easily using thresholding methods such as Otsu’s thresholding. Rooftop regions are accompanied by their shadows. Mostly, they overlap with the roof edges. But, sometimes they may not overlap if the edges are broken or are not smooth. So, to ensure that the rooftop region overlap with its shadow we dilated the candidate regions[17]. Then, we counted the number of boundary pixels of the regions overlapping with the shadow pixels. We have used this count of shadow pixels are our feature.

3) Classification using Support Vector Machine (SVM): SVM is a supervised learning technique which finds an optimal decision boundary to separate the feature space into the corresponding classes. Like many other kernel methods, the performance of SVM is highly influenced by the choice of kernel function. Preliminary investigations found that the polynomial kernel function gave the best results for our data and so this was used for all subsequent experiments.

Polynomial kernel function of order d can be represented as below:

\[ K(x, y) = (x^T \cdot y + 1)^d \]

(4)

First, we trained a SVM using the training images. Then we test each of our test images using the trained SVM model. Thus, we get our final rooftop and non-rooftop labels.

III. RESULTS AND DISCUSSION

A. Data

For the experiments described here, images gathered from Google Maps were used. The research presented here was part of a broader effort to determine the total rooftop area in Abu Dhabi which was available for solar PV installation; as such, images depicting residential areas in Abu Dhabi were used. To ensure the generality of our model it was tested with two image datasets with some differing characteristics; these are referred to as Khalifa and Raha, which are names of the corresponding regions in Abu Dhabi.

To keep the number of clusters manageable and thus facilitate proper segmentation, the satellite images of these areas were divided into 512 x 512 pixel sized tiles, which corresponds to an area of approximately 70m x 70m. The Khalifa dataset consists of 25 such images of which 10 are used for training and 15 for testing; the Raha dataset consists of 23 such images of which 10 are used for training and 13
for testing. Rooftops in these images are labeled manually and these labels are subsequently used to label each segment as either a rooftop (1) or a non-rooftop (0).

Fig. 4: (a) Sample image from Khalifa City A (b) Sample image from Raha Gardens (c) and (d) An example of manual labeling of image

Figure 4 presents examples of image from both datasets and also a manually labeled image. In these images we can see that, many non-roof objects (for example cars and roads) look very similar to the rooftops. For example for "Raha" images, the color of rooftops is very similar to that of the road. So, our model needs to be able to distinguish these non-roof objects from true rooftops.

B. Experimental Results

We used commonly adopted performance metrics for evaluation. The used metrics are Precision and Recall, which can be defined as:

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{5}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{6}
\]

Where, True positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) represent number of correctly identified rooftops, number of incorrectly identified rooftops, number of correctly rejected rooftops and number of incorrectly rejected rooftops respectively.

Precision represents the percentage of correctly identified rooftops (TP) in the total detected rooftops and Recall gives the percentage of correctly identified rooftops (TP) detected by our approach to the actually present rooftops. So, recall is the most important measure to assess performance of the model.

We evaluate the performance of our model based on the overall area covered by detected rooftops. We also compared the results obtained from the first and second pass of classification. Sample result of our algorithm is shown in Figure 5. It can be seen that the result of First-stage classification contains many false positive regions. This is what our second-stage classification helps to improve. Second-stage classification reduces false positives largely, which in turn significantly improves the precision. Even though there is slight decrease in recall values as some of the rooftops identified in the first-stage may be lost, the improvement in precision and false positives is significant.

The results for Khalifa and Raha dataset are presented in Table I. This table shows overall precision, recall and FP values for the test datasets. As can be seen in the result, the addition of second-pass of classification significantly reduces the false positive in the result as compared to the first-pass result (especially for Raha dataset, there is a drastic reduction in false positives).

As we already mentioned Recall value is an important measure of performance of any rooftop detection model. So, our proposed two-stage classification model works excellently by detecting many of the rooftops present in the image. So, the recall value of our approach is high.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Result of First-Stage</th>
<th>Result of Second-Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khalifa</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td></td>
<td>89.97%</td>
<td>82.41%</td>
</tr>
<tr>
<td>Raha</td>
<td>86.58%</td>
<td>87.54%</td>
</tr>
</tbody>
</table>

C. Comparison with other similar work

Finally, we compare our current results with results obtained using an earlier method which we had presented in [12]. Both of these models are using images from same geographical area. The comparison is shown in Table II. From this comparison, it can be observed that our approach works better than the previous approach. We can see that the area of detected rooftop (Recall) is significantly higher than the previous approach.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Histogram Method</th>
<th>Our Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>Khalifa</td>
<td>93.16%</td>
<td>70.01%</td>
</tr>
<tr>
<td>Raha</td>
<td>97.6%</td>
<td>79.7%</td>
</tr>
</tbody>
</table>

IV. CONCLUSION AND FUTURE WORK

In this paper, we presented a new approach for rooftop detection using a two-stage classification procedure using a novel combination of Artificial Neural Networks and Support Vector Machine. We showed that this methodology was able to detect a high percentage of rooftops and the second-stage classification really helped to reduce the false positives returned by the first classification stage. One unique feature of our approach is that we extract information from the first-pass of classification and use this information in the second pass to improve the result. So, addition of this second classification stage helps to detect missed rooftops from first stage and significantly reduces the false positives in the result.
However one weakness of this approach was that a small number of cases where the first-stage classification was very accurate, the second classification stage actually removed some of the true rooftops. Also the performance of this model deteriorates when there are rooftops with different colors. The probable reason for this drawback is that, such images with multicolor rooftops make those features insignificant which use color information of rooftops.

For future work we plan to work on two main areas:

1) To fulfill the main motive of our research, we will use the rooftop area detected by our rooftop detection model to estimate the solar energy produced by the installation of solar Photovoltaic cells on the rooftops of Abu Dhabi City. Also, we plan to assess the cost involved in such installation.

2) We also plan to test this method on images from different geographical locations and also test the method with more complex images and over a larger geographical area.

REFERENCES


