MULTI-LEVEL FEATURE ANALYSIS FOR SEMANTIC CATEGORY RECOGNITION

Harini Sridharan and Anil Cheriyadat

Oak Ridge National Laboratory
Oak Ridge, TN USA 37832

ABSTRACT

At half-meter resolution the earth's surface has roughly 600 Trillion pixels. The need to process satellite imagery at such enormous scales for automated semantic categorization and the requirement to repeat this process at time-stipulated intervals demand optimal strategies to scan, extract, and, represent image features for accurate land-cover detection. In this paper we focus on developing optimal strategies for semantic categorization of image data which often involves computationally intensive feature extraction and mapping processes. Our proposed semantic categorization framework involves feature extraction and mapping at multiple levels. Initially, we examine low-level pixel features such as edge gradients, orientations, and intensity values to compute feature vector based on aggregate statistics. At the second level we generate line based representation by connecting edge gradients to extract higher-order spatial features on image scenes that are screened by the first level. By employing a multi-level feature analysis strategy we develop a semantic categorization framework that is computationally efficient and accurate. We tested our approach for the automated detection of *mobile home park* scenes, a challenging land-cover class, using one-meter aerial image data. We report the detection performance of our system. We envision that such changes to traditional feature analysis are necessary for the massive image analysis challenges.

Index Terms— mutli level analysis, semantic classification, mobile home parks

1. INTRODUCTION

With the availability of high spatial resolution data, mapping requirements are subsequently shifting from general land cover information such as grass, buildings, roads etc., to detailed semantic scene categories such as golf courses, school campuses etc. Semantic categorization of aerial imagery often involves deriving an intermediate representation of the imagery to quantify features that characterize the geometrical, spatial, and appearance attributes of physical structures in the scene. However, transforming image data to such intermediate representation and extracting high-order features from the representation is a computationally intensive process, but often necessary to generate accurate categorization for challenging semantic classes.

Fig. 1. Line based representation of aerial scenes. We compute local line pattern statistics based on difference in line orientations, length, position within a local neighborhood and line contrast statistics from the image patches to generate higher-order scene features that can yield accurate semantic categorization of the scene. The image patches shown here represent *mobile home parks*.

Previously, researchers have explored several interesting approaches based on pixel and object (homogeneous and contiguous pixels) attributes for high-resolution image classification. Most of the previous work [1] focused on improving the classification accuracy of the process. Often the challenge of scaling the solutions for large-scale operations is overlooked. Although pixel and object-based classification approaches, such as [2, 3], produce accurate thematic mapping, they rarely capture the complex relationship between physical objects that explain the semantics of the scene category. In this work we employ a line based feature extraction and representation strategy to capture the structural and spatial patterns in the scene. To address the computational challenges involved in generating such a representation for large-scale image analysis, we employ a tiered analysis scheme. We first use a local pixel-level attributes such as intensity and edge orientation statistics at the initial level for identifying image patches (scenes) with built-up presence. The image patches that have built-up presence are further processed to extract line representation at the second level. Figure 1

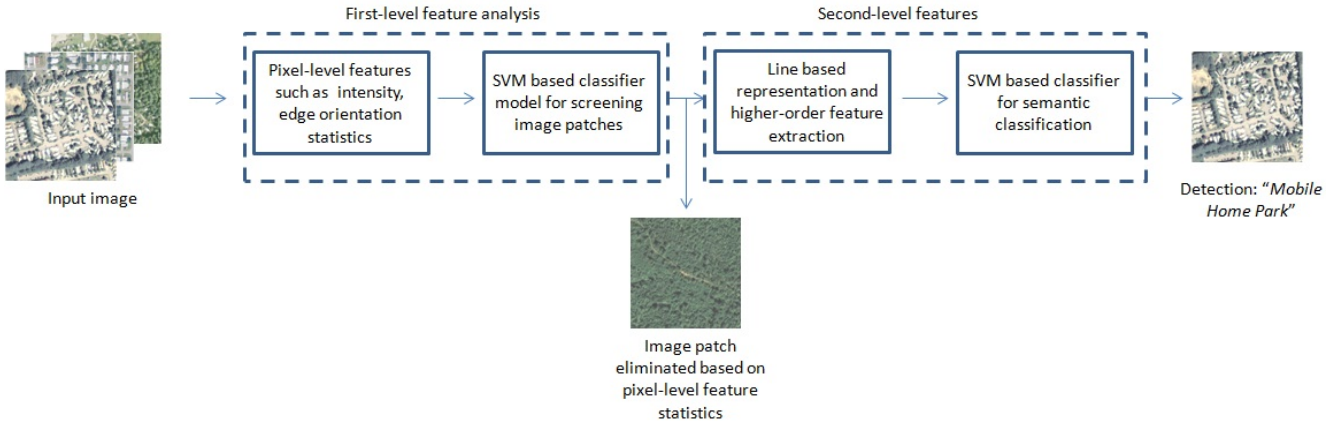

Fig. 2. Overview of the proposed framework. At the first level we examine pixel-level features for screening image patches that needs to be further processed for semantic categorization. The feature extraction process is based on first-order gradient and intensity statistics. We employ a SVM based classifier to identify image patches that need further processing. The second level of the process involves generating line based representation by connecting pixel gradients to generate higher-order line features for semantic category detection. The line generation process and feature extraction process is computationally intensive and by reducing the image volume that needs to be processed by the computationally intensive stage, we improve the overall efficiency of the process without compromising the detection performance.

.

shows line based representation generated from a few sample high-resolution overhead scenes. We compute local line pattern statistics such as difference in line orientations, length, position within a local neighborhood and line contrast statistics from the image patch to generate higher-order scene features that can yield accurate semantic categorization of the scene.We show that such a strategy can yield high processing efficiency without compromising the classification performance. We apply our technique to detect *mobile home park* scenes in high-resolution aerial imagery.

2. MULTI-LEVEL FEATURE ANALYSIS FRAMEWORK

Here we briefly describe our multi-level analysis framework. Figure 2 shows an overview of the proposed framework. First, we normalize the image by subtracting the mean out and dividing by the standard deviation. We operate on the gray scale image. Next, at each pixel we compute the intensity gradient magnitude and orientation.

2.1. First-Level Features

For the first level feature analysis we use the gradient features and raw intensity data for feature computation. We divide the image into non-overlapping square pixel-blocks consisting of 16×16 pixels. At each pixel block we compute statistical parameters from edge orientation histogram and first- and second-order local pixel intensity statistics. For the edge orientation statistics we compute three heaved central shift moments (order = 0,1, and 2) and two orientation measurements. The two orientation measurements are the magnitude of the histogram peak and the absolute sine difference of the orientations corresponding to the two highest peaks. For the first-order intensity statistics we simply compute the mean and variance measurement of the pixel intensity. For the second-order intensity statistics we compute intensity co-occurrence matrix and compute contrast feature from the co-occurrence matrix. For each pixel-block we perform the feature extraction at five scales where each scale is characterized by the spatial window from which the edge orientation and pixel intensity statistics are generated. We employ a square window and the window size at the five scales are 16, 32, 48, 64, and 80 pixels. Except for the first scale, we extract 7 features (5 edge orientation feature and 2 raw intensity statistics) from each scale. For the first scale along with the 7 features we also compute the gray-level co-occurrence feature. We refer interested readers to our previous work [4]. Next we learn a SVM based detection model to classify pixel-blocks with built-up presence. We use manually labelled data to perform the model training. For each image patch that has pixel-blocks with built-up labels less than a pre-defined threshold is eliminated from the second level computationally intensive feature extraction.

2.2. Second-Level Features

For the second level feature generation we use the pixel intensity gradient magnitude and orientation values to generate lines representing linear features in the scenes. Our approach

begins by extracting straight line segments from the image by grouping spatially contiguous pixels with consistent orientations as reported in [5]. The orientations are quantized into 8 bins ranging from 0° to 360° in 45° intervals. To avoid line fragmentation attributed to the quantization of orientations, we quantize the orientation into another 8 bins starting from 22.5° to $(360 + 22.5)^\circ$ in 45° intervals. Spatially contiguous pixels falling in the same orientation bin form the line supporting regions. Regions are generated based on the different quantization schemes separately and the results are integrated by selecting line regions based on a pixel voting scheme. We ignore pixels having gradient below a threshold (.5 for image intensity ranging between 0 and 1) to reduce noisy line regions. We compute the line centroid, length, and orientation from the Fourier series approximation of the line region boundary. Figure 1 shows lines generated from the input images. We refer interested readers to [6, 7].

Next, to compute higher-order local line pattern statistics, we perform a neighborhood analysis around each line using the nearest 30 neighbors. For each line, we compute the histogram of the distance to the neighbors as one of the line features. For each line, we also compute the histograms of differences in line length and orientation between the line and it's neighbors as addiitional features. This results in a 154 length feature vector for each line. In order to model the line patterns, we develop a codebook representing the cluster centers generated from line feature vector by employing K-means clustering. Next, each line is mapped to the corresponding codebook entry based on the distance proximity. For the input image we produce a histogram of the cluster labels in addition to the histogram of line brightness values to form the final feature vector. We learn a SVM based detection model based using labeled feature vectors to detect mobile home parks.

3. EXPERIMENTS

The performance of the multi-level framework was tested at two separate stages: (i) detection performance at second-level, and (ii) overall detection performance when both the first and second level are combined. The performance of the first level analysis, which predominantly detects built-up areas, is close to 85%. We refer interested readers to our previous work [4] for the detailed performance assessment. Two separate datasets were used for the two stages as discussed in the following paragraphs.

3.1. Dataset

To assess the higer-order level features, a preliminary visual analysis of the mobile parks using USDAs National Agriculture Imagery Program (NAIP) imagery was performed across the US identity different *mobile home park* scenes. Based on this analysis, a database of one-meter aerial scenes (300 x 300 pixels) was compiled from 10 different cities that exhaustively covered the majority type of *mobile home parks* and other categories. For each scene a context size of 300 meters was carefully chosen to be large enough to capture the spatial context of the *mobile home parks* scene. A total of 610 positive samples and 7,813 negative samples representing *mobile home parks* scene category were collected. In order to assess the performance of these features in detecting mobile-home park scenes, a geographically stratified cross validation technique was designed using the aerial scene database. Under this scheme we iteratively hold samples from one geographic area for testing and trained a linear SVM classifier using the samples from all the other geographic area. To assess the overall accuracy, a 5-fold cross validation was also performed without geographical stratification.

To evaluate the performance of the entire framework, two large-scale aerial images ($25km^2$) from San Ysidro, CA and Rochester, NY were chosen. In the case of large-scale image analysis, we use a moving window of 300 meters context size and a step size of 100 meters. The moving windows were classified using a SVM model developed based on the samples collected previously for the higher-order level feature assessment.

3.2. Results

For the second-level analysis the overall *mobile home parks* scene detection accuracy for each chosen city is shown in Table 1. The average overall accuracy is 96.80% with a precision and recall of 72.13% and 81.48% respectively for mobile home samples . The very low false positive rate and high true positive rate suggests that the chosen higher-order features can be very useful for*mobile home parks* scene detection. For the overall assessment of our multi-level framework we conducted *mobile home park* detection on the Rochester and San Ysidro imagery. The results are shown in Figure 3. We processed a total of 4484 windows resulting in detection rate of 87.9 % and false positive rate of 0.25 %. Similarly for the San Ysidro imagery, a total of 2695 windows were processed with a detection rate of 70 % and a false positive rate of 0.55 %. Filtering of image patches using the first level features contributed to the reduction of false positives as well as for improving the computational efficiency.

4. CONCLUSIONS

In this paper, we have presented a multi-level framework for detecting *mobile home park* scenes from high-resolution imagery. We employ a tiered analysis approach to address the computational challenges in generating complex higher order features. The higher order features are designed for detecting dense *mobile home park* scenes. Currently we are extending our work to detect sparse *mobile home park* scenes by investigating adaptive window sizes and multi-scale approaches.

Table 1. *Mobile home parks* scene detection accuracy (in %) for different cities. Al - Albuquerque, Bu- Buffalo, He-Helena, Ja-Jackson, Lo-Louiseville, Ph-Phoenix, Ri-Ri-Riverside, Sac-Sacremento, Sfo-San Francisco, San-Santa Fe.

City	Al	Bu	He	Ja	Lo	Ph	Ri	Sac	Sfo	San
Accuracy(%)	93.5	82.5	61.1	73.5	96.3	93.2	66.7	82.1	99.6	85.3

We are also porting our work to GPU based cluster architecture to support national level critical infrastructure assessment and monitoring.

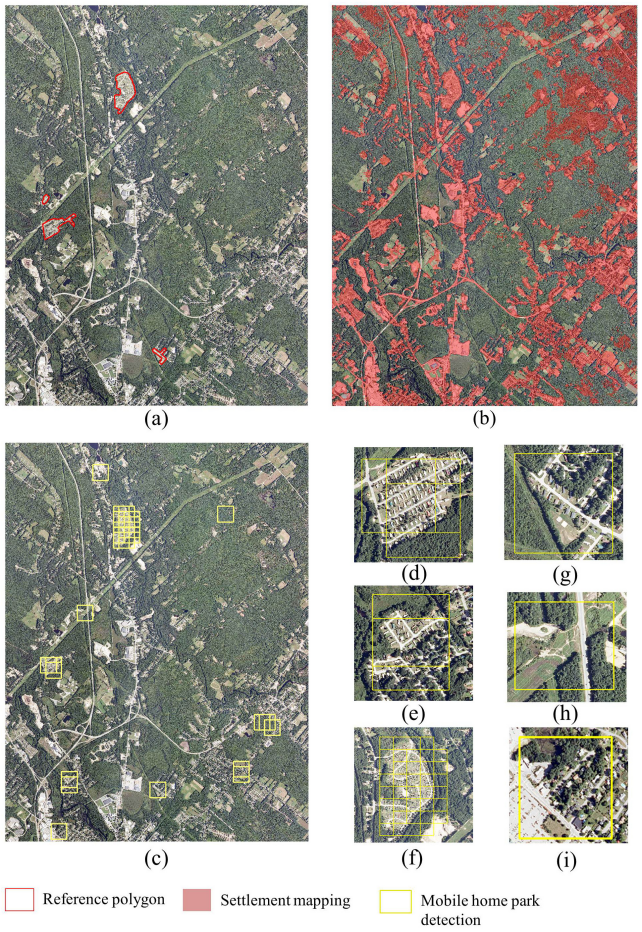

Fig. 3. (a) Rochester image marked with manually delineated reference polygons for *mobile home parks* (b) Red regions overlaid on the imagery represents output from first level (c) *mobile home park* detections produced second level marked in yellow. (d) - (g) show true positive and (h) - (i) show false positive detections.

5. COPYRIGHT FORMS

6. REFERENCES

[1] G. G. Wilkinson, "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE TGRS*, vol. 43, pp. 433–440, 2005.

[2] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE TGRS*, vol. 44, no. 9, pp. 2587–2600, sept. 2006.

[3] R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J. Cheung-Wai Chan, and F. Canters, "Improved classification of VHR images of urban areas using directional morphological profiles," *IEEE TGRS*, vol. 46, no. 10, pp. 2803–2813, 2008.

[4] E. Bright A. Cheriyadat D. Patlolla, J. Weaver, "Accelerating satellite image based large-scale settlement detection with the GPU," in *Proc. of ACM GIS SigSpatial Workshop*, 2012.

[5] J. B. Bums, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *IEEE Trans. on PAMI*, vol. 8, pp. 425–455, 1986.

[6] C. Unsalan and K. L. Boyer, "Classifying land development in high-resolution panchromatic satellite images using straight-line statistics," *IEEE Trans. on GeoScience and Remote Sensing (TGRS)*, vol. 42, pp. 907–919, 2004.

[7] A. Cheriyadat, "Learning scene categories from high resolution satellite image for aerial video analysis," in *Proc. of Computer Vision and Pattern Recognition Workshop*, 2011.