SEMANTIC INFORMATION EXTRACTION FROM MULTISPECTRAL GEOSPATIAL IMAGERY VIA A FLEXIBLE FRAMEWORK

Shaun Gleason[1], Regina Ferrell[1], Anil Cheriyadat[1], Raju Vatsavai[1], Soumya De[2]

[1]Oak Ridge National Laboratory
[2]Missouri University of Science and Technology

ABSTRACT

Identification and automatic labeling of facilities in high-resolution satellite images is a challenging task as the current thematic classification schemes and the low-level image features are not good enough to capture complex objects and their spatial relationships. In this paper we present a novel algorithm framework for automated semantic labeling of large image collections. The framework consists of various segmentation, feature extraction, vector quantization, and Latent Dirichlet Allocation modules. Initial experimental results show promise as well as the challenges in semantic classification technology development for nuclear proliferation monitoring.

Index Terms— Satellite image analysis, semantic classification, invariant features, Latent Dirichlet Allocation

1. INTRODUCTION

The Oak Ridge National Laboratory is developing an algorithmic framework to assist the geospatial image analyst with the tedious process of searching through geospatial image libraries for potential nuclear proliferation activities. An overview of this research framework is shown in Figure 1. Geospatial libraries are continuously collected today in higher spatial and spectral resolution than ever before. The ability to process and comprehend this data is limited by the number of analysts and available software tools. The ultimate goal of this research is to detect a variety of potential nuclear proliferation-related structures and activities. As a first step, development and testing are focused on detecting and identifying nuclear power facilities. First, images are segmented using a fixed tiling approach. Approximately radiometric-invariant features are extracted and used to quantize the segmented regions, or tiles. These tile features are fed into a probabilistic model, which predicts the most likely image "topic," or semantic description of the facility of interest (e.g. nuclear plant, coal plant, airport) based on the distribution of identified objects within the image. Initial results on semantic identification of complex structures within imagery show the strong potential of the approach.

Previously, researchers have explored similar paths for recognizing the semantic category of an image based on the co-occurrence of visual features. Lienou et al. [1] have shown that Latent Dirichlet Allocation (LDA) based semantic classification of satellite image content using simple visual features such as the mean and standard deviation of pixel intensity values in a local neighborhood yielded promising results. In the context of terrestrial image categorization, Li and Perona [2] presented a similar approach using LDA on visual words comprised of SIFT features to categorize a diverse set of scene types including bedroom, kitchen, living room, office, streets etc. Taking key insights from previous works, we advance the semantic classification of satellite imagery based on LDA framework by incorporating unique and richer feature sets to form our visual word vocabulary.

It is also important to note that this work is unique in that the facilities of interest being studied here are highly overlapping in terms of their overall visual features (buildings, parking lots, containment vessels, etc.). It is the presence of unique and perhaps infrequently occurring features (cooling towers, coal piles, nuclear reactors, etc.) and their spatial relationship to one another that separates one class of facility from another, which makes this a challenging classification problem.

2. SATELLITE DATA

Over 130 multispectral satellite images have been collected from commercial satellites of 4 basic categories of facilities: U.S and international nuclear plants, coal power plants, refineries, and airports. These images cover over 80 distinct geographical sites and when possible, 2 images taken at different times have been collected for each site. These images were from high resolution (~1 m) commercial satellites, primarily Quickbird and Ikonos, and 3 images are stored and catalogued from each acquisition time: the high resolution panchromatic image, the lower resolution multispectral image data, and a pan-sharpened version integrating multispectral data with the higher resolution panchromatic image. Analysis has primarily been performed on the 11 bit grayscale panchromatic images.

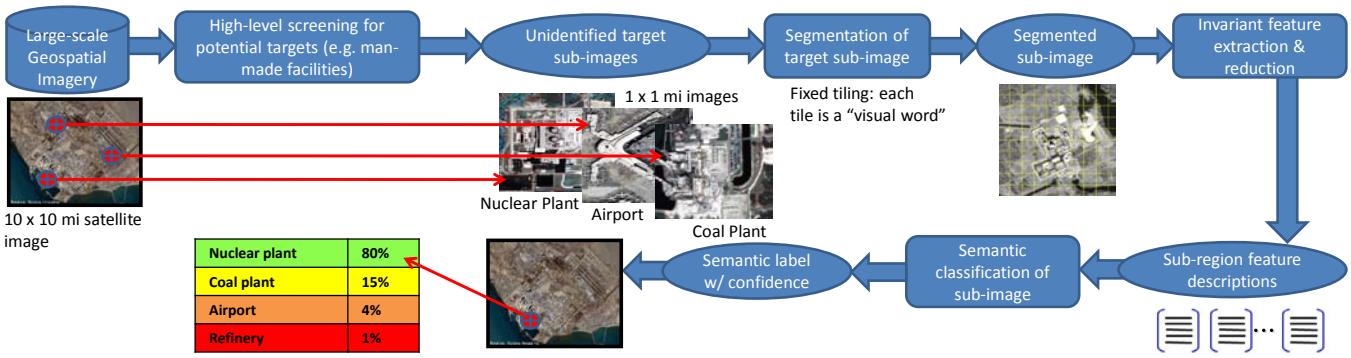

Figure 1. Schematic of developed framework for semantic image classification.

Pixel data from the images and results from image segmentation and feature extraction methods have been organized and stored in a relational database. In addition to detailed information about the source and location of the images, the database contains information about low level processing results and manually labeled ground truth information data that are available for selective retrieval. This organization of data allows the automated retrieval of images, segmentation results, and feature data that can be useful for efficiently generating consistent results across multiple experiments.

3. SEGMENTATION & FEATURE EXTRACTION

Various segmentation strategies are being developed to effectively segment the satellite image into the constituent parts, including variations of the J-value segmentation approach [3], [4]. Results using these more advanced segmentation strategies will be reported in future work, however, for the results reported here, a straightforward tiling strategy was implemented where the original image was divided into 128x128 pixel non-overlapping tiles. The size of the tiles (128 m square) was empirically chosen to be large enough to capture the salient features of the underlying structures (buildings, reactors, cooling towers, etc.), but not so large that a single tile would consistently contain structural features from multiple objects. This is important because the feature vectors representing the "visual words" used later in the semantic classification process are extracted from these individual tiles.

The objective of the feature extraction step is to represent each segment of a tessellated image by a unique feature vector characterizing the spectral, textural and structural details. The spectral, textural and structural details are represented through a statistical distribution of low-level features. The spectral attributes of an image segment provide valuable cues in distinguishing certain land-cover classes, and these are represented through intensity histograms. For multi-spectral images, we compute 64-bin histograms for each channel and for panchromatic images, we compute 64-bin histograms over the pixel intensity values. To characterize the textural details of an image segment we use histograms computed over local binary patterns (LBP) [5]. For generating LBPs, a 3x3 pixel neighborhood around each pixel is thresholded based on the intensity value of the center pixel to form a binary pattern from eight neighboring pixels. To make the LBPs rotationally invariant, we only consider the 36 binary patterns from the total of 256 patterns based on rotation invariance.

We capture the structural information of the image segment based on local edge patterns (LEP), edge orientation and line statistics. The LEPs [6] characterizing the structural details of the image segment is computed similarly to LBP except that in this case local binary patterns are computed based on binary edge map rather than intensity values. In the case of LEPs there are 36 rotationally invariant binary patterns, but based on the state of the center pixel (edge=1, no-edge= 0), possible patterns are mapped to 72 unique patterns. We compute a 72-bin histogram over the LEPs to capture the structural information. For additional implementation details on LBP and LEP, reader is referred to [7].

Edge orientation is a promising feature to discriminate man-made and natural structures present in the image. We compute edge orientations at each pixel using steerable filters [8]. We compute the 64-bin histogram of the edge orientation over angles from -90 to +90. To make the edge orientation histogram rotationally invariant, we compute a 64-point fast Fourier transform and keep the magnitude of the first 32 points as features.

Previous work by [9] have shown that line statistics derived from imagery provide a promising feature to discriminate different man-made structure categories. Line statistics are computed from line support regions-contiguous groups of pixels having consistent gradient orientation [10]. We computed histograms based on line length (21 bins) and line contrast (24 bins) computed from line support regions.

The spectral, textural and structural features extracted from the image segment are stacked to form the original feature vector. The original 249-dimensional

(64+36+72+32+21+24) feature vector is subjected to standard dimensionality reduction technique based on PCA followed by linear discriminant analysis to produce a reduced 9-dimensional feature vector. Next, we applied feature vector quantization using K-means and Gaussian Mixture Model (GMM) clustering techniques on the reduced feature vectors to form our visual word vocabulary.

4. SEMANTIC LABELLING

The next step is to apply a semantic label to the image of interest based on the distribution of visual words contained within the image. The algorithm selected to determine the best semantic label is based on the LDA model and is described next.

4.1. Latent Dirichlet Allocation (LDA)

The LDA model, originally proposed by Blei et al., [11], is an unsupervised statistical generative model developed for finding latent semantic topics in large collections of text documents. In the LDA model, each document d is assumed to be generated by a K-component mixture model, where the mixing probabilities θ_d for each document are governed by a global Dirichlet distribution. We now briefly introduce the terminology and notations:

1. A *word* w $\in$ {1, …, V} is the most basic unit of data. Here V denotes the vocabulary. As we are applying LDA to the remotely sensed image domain, a word w corresponds to a region (each cell or window in a grid as shown in Figure 1, or any arbitrary segment in the image). As can be seen in Figure 1, lot of words (tiles) are similar (for example, building tiles, water tiles), therefore these words need to be grouped together first into visual words. Thus, in the image domain, the basic unit of discrete data is the visual word.
2. A *document* d is a sequence of N words denoted by **w** = (w_1, w_2, …, w_N), where w_n is the nth word in the sequence. With respect to the image domain, a document corresponds to an *image*.
3. A *corpus* is a collection of M documents (images) denoted by D = {$\mathbf{w}_1$, $\mathbf{w}_2$, …, $\mathbf{w}_M$}.
4. A *topic* z $\in$ {1, …, K} is a probability distribution over the vocabulary of V words (visual words).

4.1. LDA as a Generative Process

We now briefly describe the generative process modeled by LDA. Given a corpus of unlabeled images, the LDA model discovers hidden topics as distributions over visual words in the vocabulary. In this process, words are modeled as observed random variables and topics are latent random variables. LDA assumes the following generative process.

- For each image indexed by d $\in$ {1…M} in a corpus:
  - Sample a K-dimensional topic weight vector (mixing proportions) θ_d from the distribution $p(\theta \mid \alpha) = Dir(. \mid \alpha)$.
- For each word, w_n $\in$ {1…N} in a document d:
  - Choose a topic z_n $\in$ {1…K} from the multinomial distribution $p(z_n = k \mid \theta_d) \sim Mult(. \mid \theta_d) = \theta_d^k$
  - For a chosen topic z_n, draw a word, w_n, from the probability distribution: $p(w_n = i \mid z_n = j, \beta) \sim Mult(. \mid \beta) = \beta_{ij}$

As can be seen from the above-described generative process, LDA is a hierarchical model. Each of K multinomial distributions β_k assigns a high probability to a specific set of words that are frequently occurring or semantically coherent in a topic. Since the generative process assumes that each word in a document is generated by a different topic, the LDA model allows multiple topic assignments to a single image. This generative process defines a joint distribution for each document w_m. For given α and β, the joint distribution over the topic mixtures θ is given by:

$$p(\theta, z, w \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta) \quad (1)$$

Then the objective is to find the corpus level parameters α and β such that the loglikelihood of the entire image collection is maximized. Unfortunately, learning the parameters of LDA model is intractable. However, two approximations, namely mean-field variational expectation maximization (EM) [11], and the stochastic EM Gibbs sampling [12], are widely used in the literature. We have implemented the mean-field variational EM approach. Mathematical details are omitted for brevity, but the interested reader can refer to [11] for full details. First, various feature are extracted from each fixed grid cell or word. These words are then quantized into visual words, using K-Means and GMM clustering techniques. The LDA model is then applied to these visual words. The following section gives more details on various parameters used.

4.1. Experiments and Results

As described in Section 2, we have collected a variety (different spectral, spatial, and temporal resolutions) of images. For semantic classification using LDA, we have selected and preprocessed 52 images of which 31 images contained nuclear power plants and 21 contained coal power plants. An independent test data set consisting of 19 nuclear images and 5 coal images were also collected. Each of these images has gone through the fixed tiling and feature extraction processes described in Section 3. We have applied both k-means and GMM clustering techniques on these images. Unlike k-means, which identifies clusters by nearest centroids using Euclidean distance, GMM finds a set

of k Gaussians from the data by using the Mahalanobis distance. K-Means is a special case of GMM clustering under certain conditions. On of the challenges in applying these clustering techniques is to specify optimal number of clusters. We performed two kinds of experiments to find the optimal number of clusters. In the first approach, we used an information theoretic measure, Bayesian information criterion (BIC) [13], to find optimal number of clusters. However, BIC based approach is quite computationally expensive as it evaluates BIC optimization in an incremental fashion for different values of K. In the second approach, we tried only a fixed number of clustering's (15 to 30, in increments of 5) as we know that the optimal value lies in this interval. The size of vocabulary (V) for LDA is equal to the number of clusters. We fit the LDA model to the training data by using the vocabulary generated from the clustering. We evaluated training accuracy of LDA for different topic and vocabulary sizes, and found the best accuracy for the following combination: number of visual words = 25 and number of topics = 3. We used the GMM and LDA models learned from training data to evaluate the predictive performance of LDA in semantic labeling by applying them to an independent test data. Both the training and test accuracies are summarized in Table I. Since we used LDA in unsupervised mode, we manually mapped the topics predicted by LDA onto the nuclear and coal categories.

Table I. LDA performance on training and test images.

	Nuclear	Coal	Nuclear	Coal
Nuclear	26	5	13	6
Coal	6	5	2	3
Overall Accuracy	Training = 73%		Test = 67%	

5. CONLCUSIONS

Given the highly overlapping visual words from one facility type to another, the achieved results are quite promising. There are clearly several avenues of research needed to improve the classification accuracy. First, to improve the uniqueness of visual words across facility types, a supervised version of LDA is being implemented. Second, the bag-of-words model inherent to LDA does not take into account unique spatial relationships between objects that are specific to a given facility type, so we are investigating methods of including these spatial relationships into our modeling approach. Future work will focus on implementation of these new approaches and evaluation of new features and categories on a larger set of images.

6. ACKNOWLEDGEMENTS

Research sponsored by the NA-22 office of the National Nuclear Security Administration within the Department of Energy. This manuscript has been authored by employees of UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy.

7. REFERENCES

[1] M. Lienou, H. Maitre and M. Datcu, "Semantic Annotation of Satellite Images Using Latent Dirichlet Allocation", *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 1, January 2010.

[2] L. Fei-Fei and P. Perona, " A Bayesian Hierarchical Model for Learning Natural Scene Categories", *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[3] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, pp. 800–810, August 2001.

[4] Celebi ME, Aslandogan YA, Stoecker WV, Iyatomi H, Oka H, Chen X, "Unsupervised border detection in dermoscopy images," *Skin Res Tech*, vol. 13:1–9, 2007

[5] M. Pietikainen, T. Ojala and, Z. Xu, "Rotation-invariant texture classification using feature distributions", *Pattern Recognition*, vol. 33, no.1, pp. 43-52, 2000.

[6] C. H Yao and, S. Y Chen, "Retrieval of translated, rotated and scaled color textures", *Pattern Recognition*, vol. 36, pp. 913-929, no. 4, 2003.

[7] K. W. Tobin, B. L. Bhaduri, E. A. Bright, A. M. Cheriyadat, T. P. Karnowski, P. J. Palathingal, T. E. Potok, and J. R. Price, "Large-Scale Geospatial Indexing for Image-Based Retrieval and Analysis," *Proc. International Symposium on Visual Computing*, LNCS 3804, Springer-Verlag, Berlin, 2005, pp. 543-552, 2005.

[8] W. Freeman and, E. Adelson, "The design and use of steerable filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891-906, 1991.

[9] 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* ,vol. 42, no. 4, April 2004

[10] J. B. Bums, A. R. Hanson, and E. M. Riseman, "Extracting straight lines", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. PAMI-8, pp.425–455, July 1986.

[11] Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent dirichlet allocation. *J. Mach. Learn.* Res., vol. 3, 993-1022, 2003.

[12] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences*, (101) Suppl. 1: 5228-5235, 2004.

[13] Fraley, C., Raftery, A.E. and Wehrens, R., Incremental Model-Based Clustering for Large Datasets with Small Clusters. *Journal of Computational and Graphical Statistics*, vol. 14, 529-546, 2005.