Supervised Semantic Classification for Nuclear Proliferation Monitoring

Ranga Raju Vatsavai*, Anil Cheriyadat*, and Shaun Gleason†
*Computational Sciences and Engineering Division
†Measurement Science & Systems Engineering Division
Oak Ridge National Laboratory,
P.O. Box 2008, MS 6017, Oak Ridge, TN, USA
Email: vatsavairr@ornl.gov, cheriyadatam@ornl.gov, gleasonss@ornl.gov

***Abstract*—Existing feature extraction and classification approaches are not suitable for monitoring proliferation activity using high-resolution multi-temporal remote sensing imagery. In this paper we present a supervised semantic labeling framework based on the Latent Dirichlet Allocation method. This framework is used to analyze over 120 images collected under different spatial and temporal settings over the globe representing three major semantic categories: airports, nuclear, and coal power plants. Initial experimental results show a reasonable discrimination of these three categories even though coal and nuclear images share highly common and overlapping objects. This research also identified several research challenges associated with nuclear proliferation monitoring using high resolution remote sensing images.**

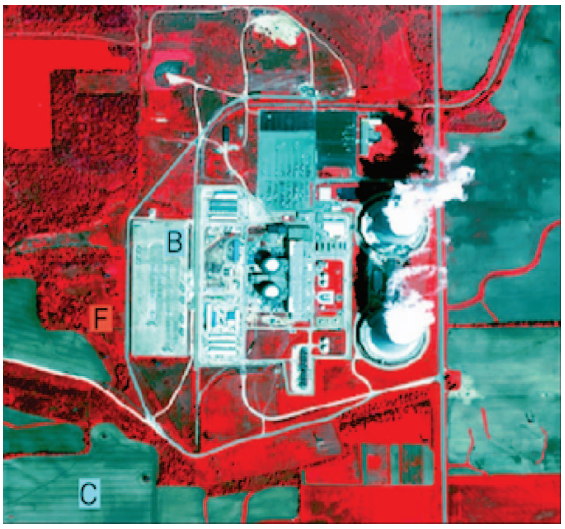
***Keywords*-GMM; LDA; Remote Sensing; Nuclear Nonproliferation;**

I. Introduction

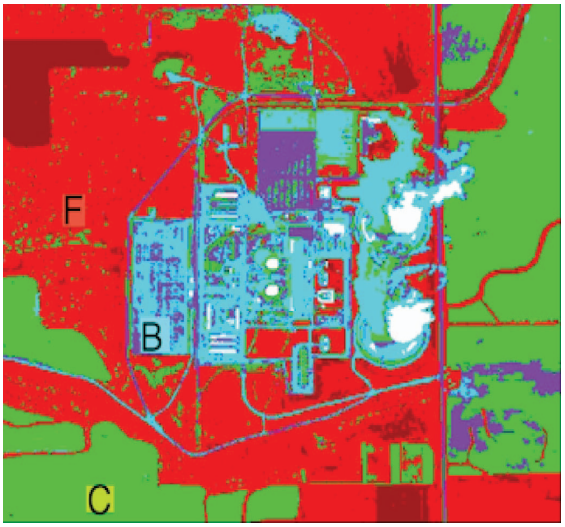
Nuclear proliferation is a major national security concern for many countries, especially to the United States. With more understanding and availability of nuclear technologies, and increasing persuasion of nuclear technologies by several new countries, it is increasingly becoming important to monitor the nuclear proliferation activities. Improvements in spatial and temporal resolutions, acquisition, and availability of remote sensing imagery made it possible to accurately identify key geospatial features and their changes over time. High-resolution remote sensing images can be highly useful in monitoring nuclear proliferation activities over any geographic region. Recent studies have shown the usefulness of remote sensing imagery for monitoring nuclear safeguards and proliferation activities [1]. However, there is a great need for developing technologies to automatically or semi-automatically detect nuclear proliferation activities using high-resolution remote sensing imagery.

Classification is one of the widely used technique for thematic information extraction from the remote sensing imagery. Classification is often performed on per-pixel basis; however proliferation detection requires identification of complex objects, patterns and their spatial relationships. One key distinguishing feature when identifying a complex geospatial object as compared to traditional thematic classification is that the objects and patterns that constitute a complex facility, such as a nuclear power plant, have interesting sub-objects with distinguishing shapes (e.g., circular shape of cooling towers) and spatial relationships (metric, topological, etc.) among those sub-objects. These complexities are clearly evident from Figure 1(c).

As can be seen from Figure 1, thematic classification is designed to learn and predict thematic classes such as urban, forest, crops, etc., at pixel level. However, such thematic labels are not enough to capture the fact that the given image contains a nuclear power plant. What is missing is the fact that the objects, such as switch yard, containment building, turbine building, and cooling towers have distinguishing shapes, sizes, and spatial relationships (arrangements or configurations) as shown in Figure 1(c). These semantics are not captured in the traditional pixel based thematic classification. In addition, traditional image analysis approaches mainly exploit low-level image features (such as, color and texture and, to some extent, size and shape) and are oblivious to higher level descriptors and important spatial (topological) relationships without which we can not accurately discover these complex objects or higher level semantic concepts. A recent review paper [2] looked at the current state of art in image information mining and identified key research gaps with respect to nuclear proliferation monitoring using high-resolution images. To address one of the key gaps identified in that paper, that is, assigning a semantic label to a given image, we recently developed an unsupervised semantic classification framework [3]. Though the framework successfully demonstrated semantic clustering as a viable solution for identifying complex facilities such as nuclear and thermal power plants, it has several limitations. First, one has to manually assign the semantic labels to unsupervised topics found using the framework. Second, unsupervised approach do not scale well for large number of topics and image categories. In this paper, we address these limitations by adopting supervised LDA algorithm [4], [5] into the framework proposed in [3]. We further extended our evaluation to three broad categories: airports, nuclear, and coal plants. Initial results of supervised extension shows better performance than the unsupervised approach and has potential to scale to more semantic categories commonly found in remote sensing images. Though the ultimate goal of this research is to detect



(a) FCC Image with Thematic class labels (B-Buildings, C-Crop, F-Forest)



(b) Thematic Classified Image (B-Buildings, C-Crop, F-Forest)

(c) FCC Image with Semantic Labels (S-Switch Yard, C-Containment Building, T-Turbine Generator, CT-Cooling Tower

Figure 1. Thematic Classes vs. Semantic Classes

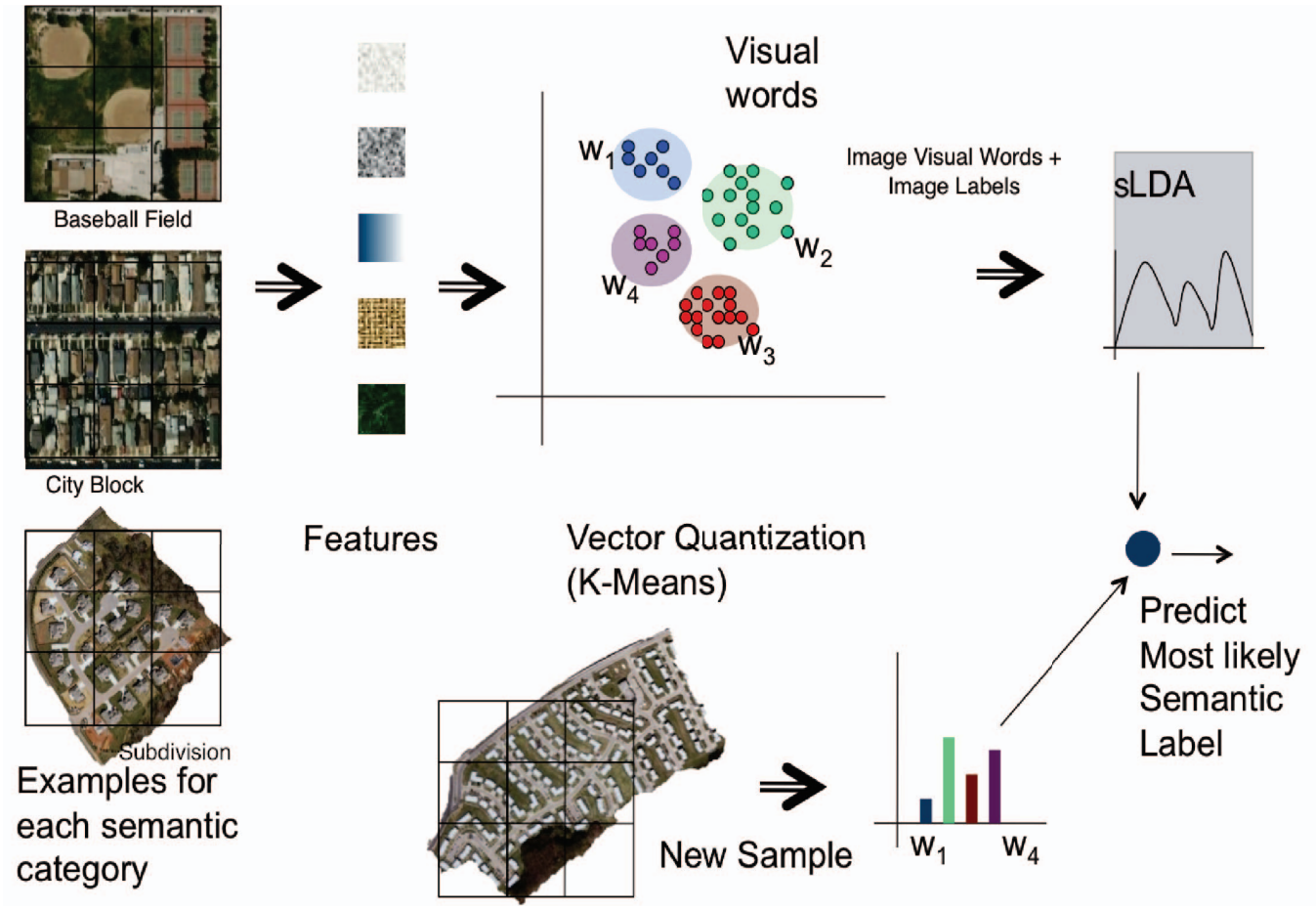
a variety of potential nuclear proliferation-related structures and activities, the current technology is still not mature enough to automate the end-to-end processing of high-resolution images to achieve this goal. We need advances in on all fronts: low-level and medium-level feature extraction, indexing, segmentation, spatial relationship modeling, and finally, semantic classification.

A. Related Work

The LDA model, originally proposed by Blei et al., [6] is an unsupervised statistical generative model developed for finding latent semantic topics in large collections of text documents. Since then, LDA technique has been widely applied and extended to several other domains. Previously, Lienou et al. [7] have shown that the 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 [8] presented a similar approach using LDA on visual words comprised of scale invariant feature transformation (SIFT) features to categorize diverse set of scene types including bed room, kitchen, living room, office, streets etc. Taking key insights from previous works, our initial work has adopted LDA for unsupervised semantic classification [3] for analyzing large volumes of satellite imagery for identification of complex facilities, especially nuclear and coal power plants. A unique feature of this framework is that it incorporated a richer set of features to generate a visual vocabulary that is more appropriate for recognizing these complex structures under a variety of scene acquisition conditions. Though initial evaluation showed promising results, one of the key limitations of this approach was that the topics found by the unsupervised LDA needs to be manually mapped into real semantic categories such as nuclear and coal plants. Thus the approach is not suitable for dealing with large number of topics which is often the case with high-resolution remote sensing imagery. In this paper, we extended our semantic classification framework by adopting the supervised LDA method proposed in [5]. We now describe this supervised semantic classification framework in detail.

B. Supervised Semantic Classification Framework

Figure 2 shows the overall semantic classification framework that we are developing as a first step towards semi-automatic monitoring of nuclear proliferation activities using high-resolution remote sensing imagery. This system consists of three core components: i) feature extraction, ii) visual vocabulary generation, and iii) semantic labeling using supervised LDA. As compared to the unsupervised approach, in the new framework, each image is associated with a class label. Likewise, LDA is replaced by sLDA. This system, once trained, can be used to predict semantic label



(a)

Figure 2. Supervised Semantic Labeling Framework

given a new image. We now briefly describe each of these key components.

II. Feature Extraction

The main objective behind the feature extraction process is to map each image to a set of visual words that correlates with the image content. Although various advance segmentation strategies are being developed to effectively segment the satellite image into the constituent parts, for this work we employ a straightforward tiling strategy. The original image is divided into 128x128 pixel non-overlapping tiles. The size of the tiles (128 square meters) is 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.

Feature extraction consists of three distinct steps: i) fixed or variable tile generation, ii) feature extraction, and iii) feature encoding. 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 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 computed 64-bin histograms for each channel and for panchromatic images, we computed 64-bin histograms over the pixel intensity values. To characterize the textural details of an image segment we used histograms computed over local binary patterns (LBP) [9]. 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 captured the structural information of the image tile

based on local edge patterns (LEP), edge orientation and line statistics. The LEPs [10] characterizing the structural details of the image segment are computed similarly to LBP except that in this case local binary patterns are computed based on the 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 computed a 72-bin histogram over the LEPs to capture the structural information. For additional implementation details on LBP and LEP, reader is referred to [11]. Edge orientation is a promising feature to discriminate man-made and natural structures present in the image. We computed edge orientations at each pixel using steerable filters [12]. We computed the 64-bin histogram of the edge orientation over angles from -90 to +90. To make the edge orientation histogram rotationally invariant, we computed a 64-point fast Fourier transform and kept the magnitude of the first 32 points as features. Previous work by [13] have shown that the line statistics derived from the imagery provides a promising feature set to discriminate various man-made structures. Line statistics are computed from the line support regions, which are contiguous groups of pixels having consistent gradient orientation [14]. We have computed the histograms based on line length (21 bins) and line contrast (24 bins) from line support regions.

Finally, the spectral, textural and structural features extracted from the image segment are stacked to form the full 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 method to produce a reduced 9-dimensional feature vector. Next, we applied feature vector quantization using Gaussian Mixture Model (GMM) clustering techniques on the reduced feature vectors to form the visual word vocabulary for the LDA method.

III. Visual Word Vocabulary

As compared to pixel based thematic classification, semantic classification works with words. As described in the previous section, a word could be a fixed tile, variable tile, or an image segment. As compared to text based semantic annotation, words in the same object category (e.g., building) in an images vary. For example, consider the baseball field in Figure 2, where tile (1,2) and (1,2) are very similar, they represent (predominantly) grass, while tile (1,1) and (2,2) are very similar, they represent bases. Therefore, the words (tiles) that are very similar (represent same object) in the image need to be grouped and assigned a single object label. These new words are called visual words. K-means clustering has been widely used in the past for visual word generation. In this work, in addition to K-means clustering we experimented with GMM clustering. GMM clustering offers better visual word generation especially if the samples follow a Gaussian distribution. In our experiments, we found that the visual word set generated by GMM is slightly better than the visual word set generated though K-means clustering. We now briefly describe the GMM clustering approach used in this work.

A. Estimating GMM Parameters

Let us assume that the training dataset D is generated by a finite Gaussian mixture model consisting of M components. If the labels for each of these components were known, then problem simply reduces to the usual parameter estimation problem and we could have used the maximum likelihood estimation (MLE) technique. Since labels for words are not known, we used the well-known expectation maximization algorithm to estimate the GMM parameters. Let us assume that each sample x_j comes from a super-population D, which is a mixture of a finite number (M) of clusters, $D_1, \ldots, D_M$, in some proportions $\alpha_1, \ldots, \alpha_M$, respectively, where $\sum_{i=1}^{M} \alpha_i = 1$ and $\alpha_i \geq 0 (i = 1, \ldots, M)$. Now we can model the data $D = \{x_i\}_{i=1}^{n}$ as being generated independently from the following mixture density.

$$p(x_i|\Theta) = \sum_{j=1}^{M} \alpha_j p_j(x_i|\theta_j) \tag{1}$$

$$L(\Theta) = \sum_{i=1}^{n} \ln \left[\sum_{j=1}^{M} \alpha_j p_j(x_i|\theta_j) \right] \tag{2}$$

Here $p_j(x_i|\theta_j)$ is the probability density function (pdf) corresponding to the mixture j and parameterized by θ_j, and $\Theta = (\alpha_1, \ldots, \alpha_M, \theta_1, \ldots, \theta_M)$ denotes all unknown parameters associated with the M-component mixture density. The *log-likelihood* function for this mixture density is given in 2. In general, Equation 2 is difficult to optimize because it contains the $\ln$ of a sum term. However, this equation greatly simplifies in the presence of unobserved (or incomplete) samples. We now simply proceed to the expectation maximization algorithm, and the interested reader can find detailed derivation of parameters for GMM in [15]. The expectation maximization (EM) algorithm at the first step maximizes the expectation of the *log-likelihood* function, using the current estimate of the parameters and conditioned upon the observed samples. In the second step of the EM algorithm, called maximization, the new estimates of the parameters are computed. The EM algorithm iterates over these two steps until the convergence is reached. For a multivariate normal distribution, the expectation $E[.]$, which is denoted by p_{ij}, is the probability that Gaussian mixture j generated the data point i, and is given by:

$$p_{ij} = \frac{\left|\hat{\Sigma}_j\right|^{-1/2} e^{\left\{-\frac{1}{2}(x_i-\hat{\mu}_j)^t \hat{\Sigma}_j^{-1}(x_i-\hat{\mu}_j)\right\}}}{\sum_{l=1}^{M} \left|\hat{\Sigma}_l\right|^{-1/2} e^{\left\{-\frac{1}{2}(x_i-\hat{\mu}_l)^t \hat{\Sigma}_l^{-1}(x_i-\hat{\mu}_l)\right\}}} \tag{3}$$

The new estimates (at the kth iteration) of parameters in terms of the old parameters at the M-step are given by the following equations:

$$\hat{\alpha}_j^k = \frac{1}{n}\sum_{i=1}^{n} p_{ij} \tag{4}$$

$$\hat{\mu}_j^k = \frac{\sum_{i=1}^{n} x_i p_{ij}}{\sum_{i=1}^{n} p_{ij}} \tag{5}$$

$$\hat{\Sigma}_j^k = \frac{\sum_{i=1}^{n} p_{ij}(x_i - \hat{\mu}_j^k)(x_i - \hat{\mu}_j^k)^t}{\sum_{i=1}^{n} p_{ij}} \tag{6}$$

Once the parameters are estimated using the EM algorithm described above, the resulting GMM can be used to assign cluster labels to the new samples. Visual words generated from this clustering process forms the vocabulary for LDA algorithm described in the next section.

IV. Latent Dirichlet Allocation (LDA)

In this section we briefly describe the LDA model originally proposed by Blei et al., [6]. 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. Let us first introduce the terminology and notations used before describing LDA model and parameter estimation technique.

- 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 the Figure 2, many words (windows) 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.
- A document d is a sequence of N words denoted by $w = (w_1, w_2, \ldots, w_N)$, where w_n is the nth word in the sequence. With respect to the image domain, the document corresponds to an image.
- A corpus is a collection of M documents (images) denoted by $D = w_1, w_2, \ldots, w_M$.
- A topic $z \in 1, , K$ is a probability distribution over the vocabulary of V words (visual words).

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

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:

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) \tag{7}$$

Now, by employing Bayes rule:

$$p(\theta, z|w, \alpha, \beta) = \frac{p(\theta, z, w|\alpha, \beta)}{p(w|\alpha, \beta)} \tag{8}$$

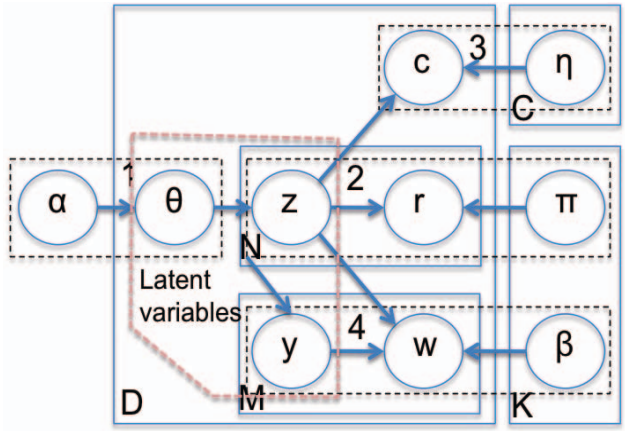
the likelihood of a document can be derived as follows:

$$\begin{aligned} & p(w|\alpha, \beta) \qquad (9) \\ = & \int p(\theta|\alpha)\left(\prod_{n=1}^{N}\sum_{z_n \in Z} p(z_n|\theta)p(w_n|z_n, \beta)\right)\mathrm{d}\theta \\ = & \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)}\int\left(\prod_{i=1}^{K}\theta_i^{\alpha_i - 1}\right)\left(\prod_{n=1}^{N}\sum_{i=1}^{K}\prod_{j=1}^{V}(\theta_i\beta_{ij})^{w_n^j}\right)\mathrm{d}\theta \end{aligned}$$

Then the objective is to find the corpus level parameters α and β such that log-likelihood of the entire image collection is maximized, that is,

$$L(\alpha, \beta) = \sum_{m=1} M \log p(w|\alpha, \beta) \tag{10}$$

Unfortunately, learning the parameters of LDA model is intractable. Well-known maximum likelihood estimation (MLE) can not be directly applied because of the presence of unobserved variables z, and θ. However, two approximations, namely mean-field variational expectation maximization (EM) [6], and the stochastic EM Gibbs sampling [16], are widely used in the literature. We have implemented the mean-field variational EM approach, which is briefly described in [3].



(a)

Figure 3. Graphical model representation of multi-class sLDA with annotation (source [5])

B. Supervised LDA

Supervised LDA, originally proposed in [4], is not suitable for multi-class classification problem as the algorithm was designed for continuous response variable. In the nuclear proliferation monitoring, our goal is to predict a class label for a given image, for example, nuclear or coal or airport, so that human analyst can select the appropriate image for further analysis. Therefore incorporating the response variable (class labels) into the learning process is critical. The supervised LDA was recently extended to multi-class classification problem [5], where the response variable admits discrete class labels. The discrete variable is assumed to be drawn from a softmax regression. The original approach [5] is designed to simultaneously model both classification and annotation tasks. Basic idea behind the approach is that annotation and classification are tightly coupled and by jointly modeling them will lead to better performance of both annotation and classification tasks. In this paper, we only implemented the classification task and we now briefly describe this procedure. We use the same concepts introduced in section IV for modeling image as bag of words. The graphical model of multi-class sLDA with annotation is given in Figure 3. Also known as plate model, in this graphical model, nodes represent variables, edges represent possible dependencies between random variables, plates (solid rectangles) denote replicated structure. Dotted rectangles are not part of the original plate model, they were drawn and numbered 1 through 4 to point to the corresponding numbered sections in the following description [5] of the generative process.

1) Draw topic proportions $\theta \sim Dir(\alpha)$
2) For each image region (fixed window) $r_n, n \in \{1, 2, \ldots, N\}$
   - Choose a topic $z_n \in \{1 \ldots K\}$ from the multinomial distribution $p(\theta|\alpha) = Dir(.|\alpha)$
   - Choose a word $r_n|z_n \sim Mult(\pi_{z_n})$
3) Draw class label $c|z_{1:N} \sim softmax(\tilde{z}, \eta)$, where $\tilde{z} = \frac{1}{N}\sum_{n=1}^{N} z_n$ is the empirical topic frequencies and the softmax function is given by the following distribution:
$$p(c|\tilde{z}, \eta) = \exp(\eta_c^T \tilde{z}) / \sum_{l=1}^{C} \exp(\eta_l^T \tilde{z})$$
4) Fore each annotation term $w_m, m \in \{1, 2, , \ldots, M\}$:
   - Draw region identifier $y_m \sim Unif\{1, 2, \ldots, N\}$
   - Draw annotation term $w_m \sim Mult(\beta_{z_n})$

As compared to the unsupervised LDA model described in the previous section, the annotated multi-class sLDA models both the image class (item 3) and image annotation (item 4) with same latent space. Notation are given in Table I.

Symbol	Meaning
K	Number of topics
C	Number of class labels
r	image visual word
w	annotation term
M	size of (annotation) vocabulary
$\eta_{1:C}$	set of C class coefficients
d	document (image)
θ_d	per document topic proportions
α	uniform Dirichlet prior on the per-document topic distribution
β	uniform Dirichlet prior on per-topic word distribution

Table I
NOTATIONS

C. Inferencing and Parameter Estimation

As noted in previous section, for LDA model inferencing posterior over hidden variables is intractable. Approximate solution was obtained in [5] using mean-field variational methods. There are three latent variables in the annotated multi-class sLDA model: per-image (d) topic proportions θ, per-visual word (r) topic assignment z_n, and the per-annotation (w) region identifier y_m. The mean-field variational distribution over latent variables is given by:

$$q(\theta, z, y) = q(\theta|\gamma) \prod_{n=1}^{N} q(z_n|\phi_n) \prod_{m=1}^{M} q(y_m|\lambda_m) \qquad (11)$$

where γ is a variational Dirichlet, ϕ_n is a variational multinomial over the K topics, and λ_m is a variational multinomial over the image regions. These parameters are found using coordinate ascent by minimizing the KL divergence between q and the true posterior. Once the parameters are estimating using the training data, classification is performed (predict class label for a new image) by estimating the probability of class label using the following approximation:

$$
\begin{aligned}
& p(c|r,w) \qquad (12) \\
\approx \quad & \int \exp\left(\eta_c^T \tilde{z} - log\left(\sum_{l=1}^{C} \exp(\eta_l^T \tilde{z})\right)\right) q(z)dz \\
\geq \quad & \exp\left(E_q[\eta_c^T \tilde{z}] - E_q\left[\log\left(\sum_{l=1}^{C} \exp(\eta_l^T \tilde{z})\right)\right]\right)
\end{aligned}
$$

The log of sum term in Eq. 13 poses problem, however can be approximated using Jensen's inequality. The prediction rule can be written as:

$$
\begin{aligned}
c^* = \quad & \arg \max_{c \in \{1,\ldots,C\}} E_q[\eta_c^T \tilde{z}] \qquad (13) \\
= \quad & \arg \max_{c \in \{1,\ldots,C\}} \eta_c^T \tilde{\phi}
\end{aligned}
$$

In this paper, we utilized only classification part of the annotated multi-class supervised LDA model. We now present the experimental results.

V. Experimental Results

Over 200 multi-spectral 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 (1m) commercial satellites, primarily Quickbird and Ikonos. These images are preprocessed, stored and cataloged for each acquisition time: the high resolution panchromatic image, the lower resolution multi-spectral image data, and a pan-sharpened version integrating multi-spectral data with the higher resolution panchromatic image. Analysis has primarily been performed on the 11 bit grayscale panchromatic images. Pixel data from the images and results from image segmentation and feature extraction methods have been organized and stored in a relational database. 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.

For supervised semantic labeling using LDA, we have selected and preprocessed 123 images of which 18 images contained airports, 75 images contained nuclear power plants, and 30 contained coal power plants. These images are divided into two independent training and test datasets consisting of 41 and 82 images respectively (see Table II). Each of these images has gone through the fixed tiling and feature extraction processes described in Section II. We have applied both k-means and GMM clustering (Section III) 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 Mahalanobis distance. K-Means is a special case of GMM clustering under certain assumptions. On of the challenges in applying these clustering techniques is to specify optimal number of clusters. We did two kinds of experiments to find optimal number of clusters. In the first approach, we used 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 clusterings (15 to 30, in increments of 5) as we know that the optimal value lies in this interval (from BIC experiment). The size of visual vocabulary (V) for LDA is equal to the number of clusters.

Figure 5 shows example airport, nuclear and coal plant images. We fitted both LDA and multi-class sLDA models to the training data by using the visual vocabulary generated from the GMM clustering. We evaluated both training and

Type	*Training*	*Test*	Total.
Airport	8	10	18
Coal	13	17	30
Nuclear	20	55	75
Total	41	82	123

Table II
Total Image Collection

G.Truth	*Airport*	*Coal*	Nuclear	Prod. Acc. (%)
Airport	5	3	0	62.5
Coal	1	9	3	69.2
Nuclear	1	9	10	50.0
Users Acc (%)	71.4	42.85	76.9	(OA) 58.5

Table III
Training Accuracy (Unsupervised LDA)

G.Truth	*Airport*	*Coal*	Nuclear	Prod. Acc. (%)
Airport	7	3	0	70.0
Coal	2	7	8	41.2
Nuclear	3	25	27	49.1
Users Acc (%)	58.3	20.0	77.2	(OA) 50.0

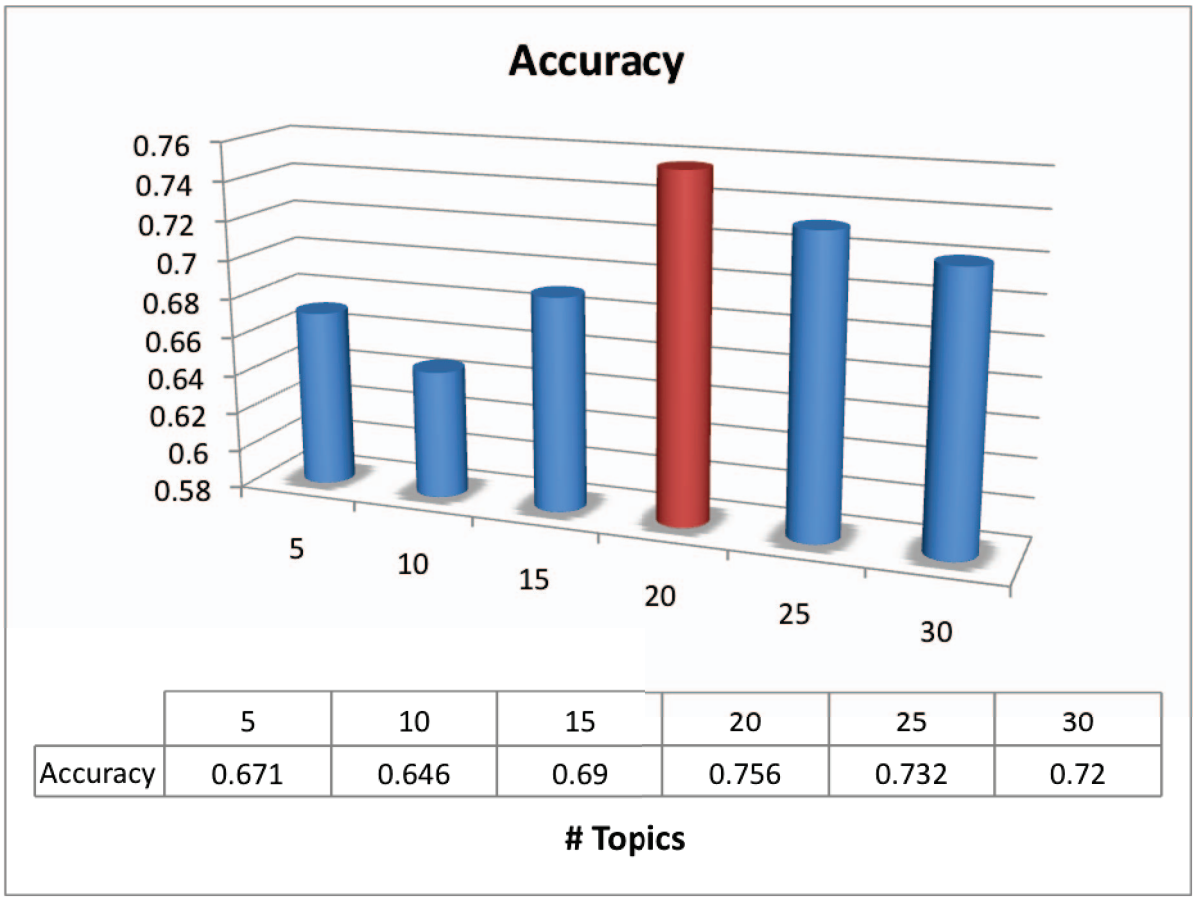
Table IV
Test Accuracy (Unsupervised LDA with manual class assignment)

G.Truth	*Airport*	*Coal*	Nuclear	Prod. Acc. (%)
Airport	6	2	0	75
Coal	0	11	2	84.6
Nuclear	0	0	20	100.0
Users Acc.	100.0	84.6	90.9	(OA) 90.2

Table V
Training Accuracy (Multi-class sLDA)

G.Truth	*Airport*	*Coal*	Nuclear	Prod. Acc. (%)
Airport	6	3	1	60.0
Coal	1	10	6	58.8
Nuclear	0	9	46	85.18
Users Acc.	85.71	45.45	86.79	(OA) 75.6

Table VI
Test Accuracy (Multi-class sLDA)



	5	10	15	20	25	30
Accuracy	0.671	0.646	0.69	0.756	0.732	0.72

Figure 4. Overall Classification Accuracy As a Function of Number of Topics

test accuracy of these two models for different topic and vocabulary sizes, and found the best accuracy for the following combination: number of visual words = 25 and number of topics = 20. For unsupervised LDA we choose only top 5 topics to manually assign labels to test images, beyond that finding the right number and combination of topics would be cumbersome. Training and test accuracies were summarized in Tables III to VI. Accuracy against number of categories is summarized in Figure 4.

VI. Conclusions and Future Directions

In this paper, we presented a supervised semantic labeling framework for identification of complex facilities in high-resolution satellite images. The framework consists of three key components. We developed several feature extraction techniques including intensity histograms, local binary patterns (LBP), local edge patterns (LEP), and edge orientation. The features extracted from fixed tiles are then quantized using GMM clustering technique. LDA and sLDA models are trained on 51 images collected under different spatial and temporal settings spread across the globe. The models learned are then applied on independent test dataset consisting of 82 images. The overall test accuracy for unsupervised LDA is 50% and multi-class supervised LDA is 75.6%, which shows good improvement in performance. These experimental results show good promise of the proposed framework. It is important to note that the image corpus contains complex objects with highly overlapping visual words, especially buildings. Our initial experiments also show several limitations and challenges in semantic labeling of complex facilities in high resolution images. First, the existing feature sets do not account for object geometry (e.g., large building vs. small buildings vs. circular buildings) which is highly useful in distinguishing the nuclear plants from thermal plants. One of the important challenges that needs to be addressed is the tile size. In this study we chose the tile size (128 m square) empirically, large enough to capture the salient features of the underlying structures (buildings, reactors, cooling towers, etc.). We are experimenting with several tile sizes to find if there is a relationship between the tile size and the quality of label prediction. In future we will also compare the performance of new features against most commonly used SIFT features. Another limitation of LDA model is that spatial relationships among the objects are ignored due to the 'bag of words'

assumption. As can be seen from the example images, spatial relationships are important as the objects (e.g., cooling towers, turbine building, switchyard) are arranged in a specific spatial configuration, therefore incorporating neighborhood relationships should improve the prediction performance. Another limitation that we observed is the equal weighting of visual words by the LDA method. There are a few distinguishing visual words between these semantic categories, however the frequency of these words (e.g., many coal plants have open coal dumps within the plant vicinity) is extremely low as compared to dominant visual words such as buildings. Our future research will focus on: (i) supervised approaches for visual word/vocabulary generation, (ii) modeling spatial relationships which becomes even more critical with the addition of more complex facilities into the mix, and (iii) visual word weighting. Since obtaining ground truth for visual words and as well as semantic labels for large number of images is difficult, we will also look at employing semi-supervised [17] approaches in the context of semantic labeling.

VII. Acknowledgments

This research is sponsored by the NA-22 office of the National Nuclear Security Administration within the Department of Energy, USA. We would like to thank Regina Ferrell, Soumya De, and Mesfin Dema for the help and inputs to this research.

Copyright: This manuscript has been authored by employees of UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. Accordingly, the United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

[1] B. Jasani, S. Nussbaum, and I. Niemeyer, *International Safeguards and Satellite Imagery: Key Features of the Nuclear Fuel Cycle and Computer-Based Analysis*. Berlin: Springer, 2009.

[2] R. R. Vatsavai, B. Bhaduri, A. Cheriyadat, L. Arrowood, E. Bright, S. Gleason, C. Diegert, A. Katsaggelos, T. Pappas, R. Porter, J. Bollinger, B. Chen, and R. Hohimer, "Geospatial image mining for nuclear proliferation detection: Challenges and new opportunities," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, 2010, pp. 48 –51.

[3] R. R. Vatsavai, A. Cheriyadat, and S. Gleason, "Unsupervised semantic labeling framework for identification of complex facilities in high-resolution remote sensing images," in *IEEE ICDM International Workshop on Spatial and Spatiotemporal Data Mining (SSTDM-10)*. IEEE, 2010.

[4] D. M. Blei and J. D. McAuliffe, "Supervised topic models," in *NIPS*, 2007.

[5] C. Wang, D. M. Blei, and F.-F. Li, "Simultaneous image classification and annotation," in *CVPR*, 2009, pp. 1903–1910.

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

[7] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent dirichlet allocation," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, no. 1, pp. 28 –32, jan. 2010.

[8] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, 20-25 2005, pp. 524 – 531 vol. 2.

[9] J.-L. Chen and A. Kundu, "Rotation and gray scale transform invariant texture identification using wavelet decomposition and hidden markov model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 2, pp. 208 –214, feb 1994.

[10] C.-H. Yao and S.-Y. Chen, "Retrieval of translated, rotated and scaled color textures," *Pattern Recognition*, vol. 36, no. 4, pp. 913 – 929, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/B6V14-47F1HVN-4/2/2849e74eafad851a14748f7add030e98

[11] K. W. Tobin, B. L. Bhaduri, E. A. Bright, A. Cheriyadat, T. P. Karnowski, P. J. Palathingal, T. E. Potok, and J. R. Price, "Large-scale geospatial indexing for image-based retrieval and analysis," in *Proc. International Symposium on Visual Computing*. 69121 Heidelberg, Germany: LNCS 3804, Springer Verlag, 2005.

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

[13] C. Unsalan and K. Boyer, "Classifying land development in high-resolution panchromatic satellite images using straight-line statistics," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 4, pp. 907 – 919, april 2004.

[14] J. B. Burns, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 4, pp. 425 –455, july 1986.

[15] J. Bilmes, "A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," Technical Report, University of Berkeley, ICSI-TR-97-021, 1997., 1997.

[16] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. Suppl 1, pp. 5228–5235, April 2004.

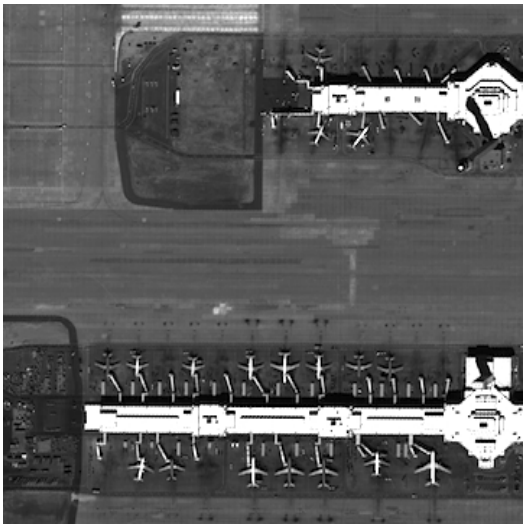
[17] R. R. Vatsavai, S. Shekhar, and T. E. Burk, "An efficient spatial semi-supervised learning algorithm," *IJPEDS*, vol. 22, no. 6, pp. 427–437, 2007.

(a) Coal Plant 1

(b) Coal Plant 2

(c) Nuclear Plant 1

(d) Nuclear Plant 2



(e) Airport 1

(f) Airport 2

Figure 5. Examples From Three Image Categories