SUPERPIXELS FOR HYPERSPECTRAL IMAGE ANALYSIS

A Thesis

Presented to

the Faculty of the Department of Electrical and Computer Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in Electrical Engineering

By

Tanu Priya

December 2014

SUPERPIXELS FOR HYPERSPECTRAL IMAGE ANALYSIS

Tanu Priya

Approved:

Committee Chair
Dr. Saurabh Prasad, Assistant Professor
Electrical and Computer Engineering

Committee Members:

Dr. Badri Roysam, Department Chair
Electrical and Computer Engineering

Dr. Shishir Shah, Associate Professor
Computer Science

Dr. Suresh K. Khator, Associate Dean
Cullen College of Engineering

Dr. Badri Roysam, Department Chair
Electrical and Computer Engineering

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor, Dr. Saurabh Prasad, for his continuous guidance, motivation and support, that helped me complete my Master's degree at University of Houston. His expert advice and immense knowledge have been invaluable throughout all stages of my thesis and research work.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Badri Roysam and Dr. Shishir Shah, for their encouragement and insightful comments. I would also like to thank my fellow lab-mates in Hyperspectral Image Analysis Group: Minshan Cui, Wu Hao, Yuhang Zhang, and Xiong Zhou for their help and support throughout my research.

Last but not the least, I would like to thank my husband, Abhishek, my parents, family and friends who have been my continual love and support.

SUPERPIXELS FOR HYPERSPECTRAL IMAGE ANALYSIS

An Abstract

of a

Thesis

Presented to

the Faculty of the Department of Electrical and Computer Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in Electrical Engineering

By

Tanu Priya

December 2014

Abstract

With rapid development of multi-channel optical imaging sensors, hyperpsectral data has become increasingly popular, necessitating development of algorithms for robust image analysis with such data. This thesis contributes methods that efficiently and robustly exploits superpixels for hyperspectral data. We study and quantify the efficacy of state-of-the-art superpixel generation algorithms for a variety of hyperspectral images. In this work, superpixel level analysis is proposed for two different hyperspectral image analysis problems — remote sensing image classification and person re-identification via forward looking hyperspectral imagery. Specifically, for remote sensing images, we propose a framework based on superpixels that provides spatial context for robust classification, and, for ground-based "natural" hyperspectral images, efficacy and utility of superpixels is demonstrated, in a multi-view setup, through a pilot study on a person re-identification task.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Hyperspectral images (HSI) record hundreds of narrowband measurements and provide fine spectral resolution over a wide range of the electromagnetic spectrum. HSI is able to capture material specific information which greatly improves classification performance. Multi- and hyperspectral images have been used extensively to improve the performance of image segmentation and classification in specific application areas such as remote sensing (see [1] for a survey of the field), medical diagnosis and bioinformatics [2, 3], and military surveillance [4]. A comprehensive overview of classification using HSI is offered in [5]. Although HSI data has the potential to provide useful information, several challenges still remain for effective exploitation, such as, addressing high dimensionality, complex statistical distributions, and including contextual information for image analysis.

Pixel level analysis is traditionally employed for material classification and recognition. However, it has been observed that pixels when taken independently become sensitive to noise and intra-class variability. In order to overcome this problem, the entire image was represented as multiple local regions [6, 7], defined by a window or grid. Nevertheless, such approaches are sensitive to the choice of grid dimensions and could cause drastic decrease in recognition performance, specially — when the image has a complex background, very common in natural scenes, or, when analyzing spectra, where any unwanted inclusion could cause spectral mixing. Hence, image segmentation, having well-defined boundaries between meaningful regions, is often preferred over a fixed window analysis. The challenge comes when we strive to obtain a "perfect" segmentation — wherein a unique object is associated with a unique segment. Obtaining such segmentation accurately is difficult to automate, and is not required from the perspective of simply providing spatial context. An alternate approach is an intentional oversegmentation of the image resulting in homogeneous regions called *superpixels* [8], comprising of several contiguous pixels.

Superpixel analysis segments the image into a set of spectrally similar regions, exploiting the fact that physical features are spatially contiguous. Superpixels have shown promise as a preprocessing step for object level classification in RGB and single-channel images [8, 9, 10] in the computer vision community. Features are computed at the superpixel level, with an aim of representing the image with "optimal" feature vectors. Regions extracted by oversegmentation, have their true boundaries preserved, and form a representation of an image that is much more compact than the original pixel grid. Superpixels have been successfully applied in image segmentation [11], object localization [12] and tracking [13]. Unlike traditional color and gray-scale imagery, hyperspectral imagery can suffer from additive noise. Traditionally, such noise can be reduced by enforcing local smoothness which is employed by grouping contiguous image areas where pixels are likely to be merged together. Hence, superpixels can be exploited to provide noise-robust hyperspectral image analysis [14, 15]. Using superpixels in analyzing high-resolution data not only represents an obvious advantage of noise reduction, as image pixels are meaningfully grouped into uniform partitions, but also the homogeneity and distributions within a superpixel makes the statistics stable and reliable. Hence, several statistical approaches, which have been

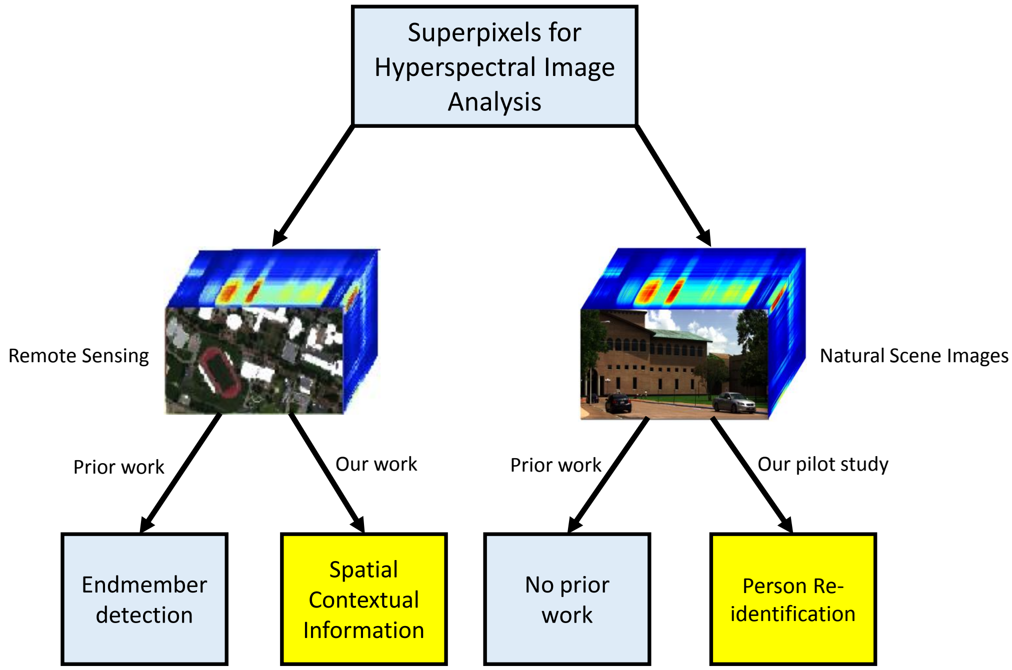


Figure 1.1: Block diagram representing applications of superpixels for hyperspectral data

widely used in pixel based methods, can be utilized for such analysis.

Fig. 1.1 provides an overview of the applications of superpixels for hyperspectral image analysis. Hyperspectral data can be broadly categorized into remote sensing imagery and natural scene imagery. There has been recent work using remote sensing imagery, where superpixel level analysis was employed for segmentation and endmember detection [14]. But, the strong contextual correspondence between a superpixel and its neighborhood superpixels was never systematically exploited. In this thesis we utilize spatial contextual information via superpixels for robust image analysis. Also, in this thesis, a pilot study is conducted to investigate the utility of superpixels for analysis of forward looking ground-based natural hyperspectral imagery. We acknowledge that in principle, "optimal" utilization of such super-pixels for analysis of natural scenes would be to use these as a tool to characterize contextual information (for example, by creating a vocabulary tree) [16, 17, 18, 19], our scope of this study is limited to a simple analysis for a practical problem (person re-identification), owing in part to the limited amount of data available. We use this case study to demonstrate that hyperspectral superpixels from natural scenes (when using metric appropriate for hyperspectral data — the spectral angle distance) indeed characterize spectrally coherent regions that can be utilized for various applications. Although, superpixels have been used extensively for re-identification in natural RGB images [16, 20], to the best of our knowledge, they have not been utilized for analysis of natural hyperspectral scenes.

The remainder of the thesis is organized as follows. In Chapter 2, we describe graph based superpixel segmentation approaches and quantify superpixel quality with hyperspectral images based on standard metrics. In Chapter 3, we present a novel superpixel based approach for hyperspectral image analysis which exploits spatial context within spectrally similar contiguous pixels for robust hyperspectral classification. In Chapter 4 we present a potential future application of superpixels, for natural forward looking hyperspectral images, through a pilot study on a person re-identification problem. We conclude by summarizing results in Chapter 5.

Chapter 2
Graph based approaches for superpixel segmentation

Image segmentation can be interpreted as partitioning of an image into meaningful regions, where each region defines a unique object or scene feature in the data. Image pixels, within a segment, are similar to each other with respect to certain characteristics, such as, intensity, spectral response or texture. In the computer vision community, a perceptually uniform region in an image is typically defined as a superpixel that contains one or more contiguous image pixels. Superpixels have traditionally been used to reduce the computational complexity of image analysis tasks, as they can assist in reducing the complexity of images from hundreds of thousands of pixels to only a few hundred superpixels. Hyperspectral images are associated with large number of spectral bands which makes segmentation a challenging task. Hence, interpretation of an image as a graph leads to a fast and efficient method of generating image segments.

Let $\{\mathbf{x}_i\}_{i=1}^{M}$ be the M pixels in our d-dimensional hyperspectral image, such that $\mathbf{x}_i \in \mathbb{R}^d$. An image can be represented as a graph $G = (V, E)$, where each pixel is the vertex of the graph denoted by set V and pair of pixels are connected by an edge e_{ij} that belongs to the set E. Each edge is associated with a weight w_{ij} which is a measure of similarity between the vertices/pixels. Graph representation of an image is shown in Fig. 2.1.

Current state-of-the-art graph-based superpixel segmentation approaches includes Felzenszwalb and Huttenlocher (FH) superpixel generation [21], Ren and Malik's Normalized Cut

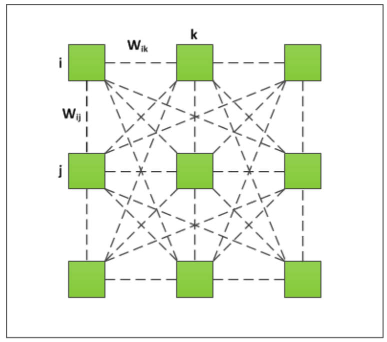


Figure 2.1: Graph Representation

(NCut) [22], Moore *et al.* superpixel lattices [9], Veksler *et al.* supervoxels [23], and Entropy Rate (ER) superpixels by Liu et al. [24]. Although, these methods have been extensively applied to traditional RGB or gray-scale imagery, they were never studied for multi-channel images (except FH, which was previously studied for spectral unmixing tasks). However, this thesis focuses on two superpixel generation algorithms — FH (which has been applied to hyperspectral imagery previously), and ER, which has been shown to be very effective for RGB images [24].

Hypothesis: Traditional graph-based superpixel segmentation approaches, commonly used for three channel images, can be naturally extended to multi-channel images with the goal of grouping spectrally coherent pixels.

2.1 Felzenszwalb-Huttenlocher superpixel segmentation

Felzenszwalb and Huttenlocher developed [21] an efficient segmentation algorithm based on a predicate that measures the evidence for a boundary between two segments using a graph-based representation of the image. A segmentation S is a partition of V into regions that corresponds to a connected component in the graph G. Ideally, all the elements in a component should be similar, and elements in different components should be dissimilar. This suggests that edges connecting two vertices within a component should have relatively low weights, and edges connecting vertices in different components should have higher weights. In order to obtain segmentation, a similarity predicate D is defined, such that, it provides a comparison between two regions.

Each vertex starts as a separate segment and neighboring sub graphs are merged when there is no evidence of boundary. The algorithm [21] determines the boundary evidence by comparing the inter component differences to the within component differences. Internal difference of a component S_a is the largest weight in the minimum spanning tree (MST) of the component, $MST(S_a,E)$. It is defined as

$$Int(S_a) = \max_{e \in MST(S_a,E)} w(e). \tag{2.1}$$

The difference between components is defined as the minimum weight edge connecting the two components (S_a,S_b), or

$$Dif(S_a,S_b) = \min_{x_i \in S_a, x_j \in S_b, e_{ij} \in E} w_{ij}. \tag{2.2}$$

If there is no edge between S_a and S_b we let $Dif(S_a,S_b) = \infty$. Evidence of boundary is true if difference between the components $Dif(S_a,S_b)$ is large relative to the internal difference within at least one of the components $Int(S_a)$ and $Int(S_b)$. Fig. 2.2 depicts the formation of components based on the predicate and the pairwise comparison predicate is defined as

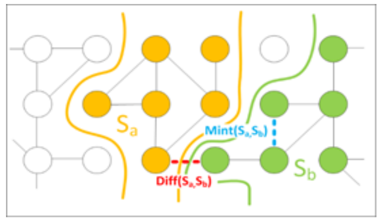


Figure 2.2: Segmentation

$$D(S_a,S_b) = \begin{cases} true & if \quad Dif(S_a,S_b) > MInt(S_a,S_b) \\ false & otherwise. \end{cases} \tag{2.3}$$

A threshold function is used to control the degree to which the difference between components must be larger than minimum internal difference. It is biased by a constant t and is inversely proportional to a superpixel's area $|S|$. The bias increases the internal variability of the smallest regions, thus, controlling the superpixel size. The minimum internal difference $MInt(S_a,S_b)$ is defined as

$$MInt(S_a,S_b) = min(Int(S_a) + \frac{t}{|S_a|}, Int(|S_b|) + \frac{t}{|S_b|}). \tag{2.4}$$

2.2 Entropy rate superpixel segmentation

In this method, superpixel segmentation is considered as a clustering problem presenting a new clustering objective function. Let A be a subset of the edge set E such that the resulting graph $G_A = (V,A)$ has fixed number (K_s) of connected sub-graphs. The objective function integrates entropy rate $H(A)$ and balancing term $B(A)$ — clustering is carried out by optimizing the function with respect to the edge set A : $\max_A\{H(A) + \lambda B(A)\}$, where

Algorithm 1 Segmentation algorithm

The input is a graph $G = (V, E)$, with n vertices and m edges.

The output is a segmentation of V into components S.

1. Sort E into $\pi = (o_1, ..., o_m)$, by non-decreasing edge weight.

2. Start with a segmentation S_a, where each vertex x_i is in its own component.

3. Repeat step 4 for $q = 1, ..., m$.

4. Construct S_q given S_{q-1} as follows. Let x_i and x_j denote the vertices connected by the q-th edge, i.e., $o_q = (x_i, x_j)$. If x_i and x_j are in disjoint components of S_{q-1} and $w(o_q)$ is small compared to the internal difference of both those components, then merge the two components otherwise do nothing.

5. Return $S = S_m$

λ is the weight assigned to the balancing term. The entropy rate quantifies the uncertainty of a stochastic process — a random walk on the graph. Using a random walk model, the algorithm [24] estimates entropy rate of a random walk on G_A as

$$H(A) = -\sum_{i=1} \mu_i \sum_{j=1} p_{ij}(A) log(p_{ij}(A)), \tag{2.5}$$

where $\mu_i = w_i / \Sigma_{i=1}^{|V|} w_i$ is a stationary distribution with respect to a random walk on the graph [24] and p_{ij} is the associated transition probability, defined as

$$p_{ij}(A) = \begin{cases} \frac{w_{ij}}{w_i}, & if \quad i \neq j, e_{ij} \in A, \\ 0, & if \quad i \neq j, e_{ij} \notin A, \\ 1 - \frac{\sum_{j:e_{ij}\in A} w_{ij}}{w_i}, & if \quad i = j, \end{cases} \tag{2.6}$$

where w_i is the total incident weight falling on vertex v_i. Although including any edge in set A increases the entropy rate, the increase is larger when selecting edges that form uniform clusters. The balancing term is expressed as $B(A) = H(Z_A) - N_A$, where N_A is the number of connected components in the graph, and Z_A is the distribution of the cluster membership. If S_A is the set of graph partitions for the edge set A, $S_A = \{S_1, S_2, S_3, ..., S_{N_A}\}$, then the distribution Z_A is given by $p_{Z_A}(i) = \frac{|S_i|}{|V|}, \forall i \in \{1, 2, ..., N_A\}$. The entropy $H(Z_A)$ favors clusters with similar sizes; whereas N_A favors fewer number of clusters.

The entropy rate of the random walk on the graph and the balancing function are monotonically increasing submodular functions (shown in [24]) under the proposed graph construction. These properties can be defined as follows:

Submodularity — Let E be a finite set. A set function F is submodular if

$$F(A \cup \{a_1\}) - F(A) \geq F(A \cup \{a_1, a_2\}) - F(A \cup \{a_2\}) \tag{2.7}$$

for all $A \subseteq E$, a_1, $a_2 \in E$ and a_1, $a_2 \notin A$. This property is also known as the diminishing return property, which says that the impact of a module is less if used in a later stage.

Monotonically increasing set function — A set function F is monotonically increasing if $F(A_1) \leq F(A_2)$ for all $A_1 \subseteq A_2$

Matroid — A matroid is an ordered pair $M = (E, I)$ consisting of a finite set E and a collection I of subsets of E.

One standard approach to maximize a submodular function is through a greedy algorithm. Hence, we adopt the greedy heuristic approach proposed in [24] for the ER partitioning. The algorithm starts with set A being an empty set and iteratively adds those edges to A that provide the largest increase in the objective function subject to the matroid constraint (i.e, (1) A cannot include cycles or self-loop and (2) A forms a graph partition with K_s connected components). Initially, a max heap structure [25, 24] is constructed to compute the gain of adding each edge to A. The edge with the maximum gain is popped from the heap, at each iteration, and included to A. The addition of an edge affects the gains of some of the remaining edges present in the heap. Therefore, the heap needs to be updated after every inclusion. However, the submodular property enables an efficient update of the heap structure. Due to this diminishing return property, gain for each edge can never increase and therefore, it is adequate to update the gain of the top element and not necessarily the others. Since only the top element of the heap is updated everytime and the values for the other elements can only decrease, the top element is the maximum value. The iterations terminate when the number of connected subgraphs reaches a preset number, $N_A = K_s$, where K_s is the number of superpixels specified by the user (determined empirically in our work). We have observed that with hyperspectral imagery, this approach provides a very reliable superpixel partitioning of the image (results quantifying efficacy as a function of K_s with hyperspectral imagery are provided in the following section).

2.3 Quantifying Superpixels

2.3.1 Quantification metrics

For evaluating the quality of superpixels, we focus on two standard metrics : boundary recall [26, 27, 24] and undersegmentation error [26, 24].

Boundary Recall

It measures the fraction of ground truth boundaries that fall within a certain distance d of a superpixel boundary. In this work, the distance d is kept as 2 pixels. Given a ground truth boundary image G and the code generated boundary image S, we compute the boundary recall as

$$BR = \frac{TP}{TP + FN}, \tag{2.8}$$

where True Positives (TP) is the number of boundary pixels in G for which there exists a boundary pixel in S in range d and False Negatives (FN) is the number of boundary pixels in G for which there is no boundary pixel in S in range d.

Undersegmentation Error

It measures the fraction of pixel leak, i.e. inclusion of unwanted regions in the segmentation result, across ground truth boundaries. For each ground truth segment G_i, overlapping superpixels S_z's are considered to compute the size of the pixel leaks $|S_z - G_i|$'s. Pixel leaks over all segments are added together and normalized by image size to compute undersegmentation error metric, or

Figure 2.3: Hyperspectral Scene Image

$$UE = \frac{\sum_i \sum_{z:S_z \cap G_i \neq \emptyset} |S_z - G_i|}{\sum_i |G_i|}, \tag{2.9}$$

where i is the number of ground truth segments and z is the number of superpixel segments.

2.3.2 Validation:Experimental setup and Results

In this section, we introduced a natural scene image, captured by the Hyperspec Imaging Spectrometer, taken at University of Houston campus. The scene image contains 325 spectral bands over the $400 \sim 1000nm$ wavelength range with the spatial size of 1004 x 2500 pixels. The HSI image is shown in Fig. 2.3. For this work, we manually outlined objects with pixel boundaries — this annotation, shown in Fig. 2.4 (a), is used to quantify efficacy of different segmentation approaches with hyperspectral images. This scene is reasonably complex, with a variety of illumination conditions, object geometries and material composition.

The algorithms described above have been used for superpixels in three-channel images

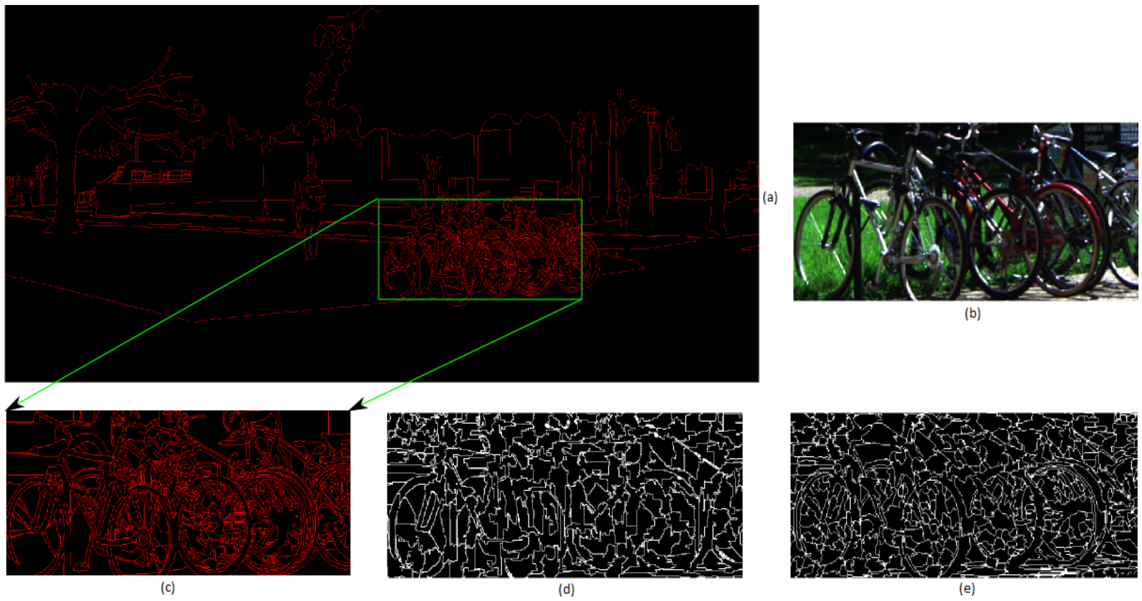
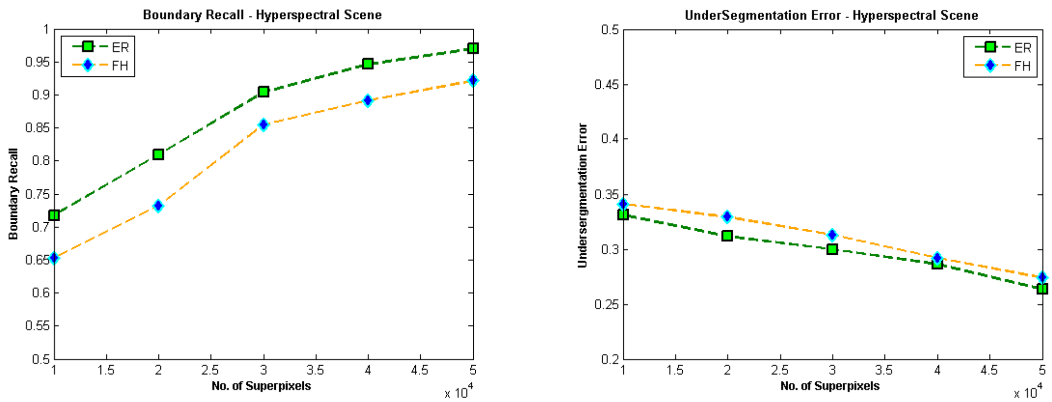


Figure 2.4: (a) Manual Annotations, (b) Cropped region of original image, (c) Annotation over cropped region, (d) FH generated superpixels, and (e) ER generated superpixels

and can be naturally translated to n-channel images. Therefore, we implemented a straightforward extension of these methods to the hyperspectral domain. In this work, we have used a Gaussian kernel to convert Euclidean distances, in the hyperspectral feature space, into similarities. Fig. 2.4 (d) and (e) show superpixel segmentation results, over a cropped region, obtained using FH and ER algorithms respectively. A visual interpretation of superpixels generated using the two approaches indicates that FH produces superpixels with irregular sizes and shapes while ER favors compact and homogeneous clusters. This was expected as the balancing term used in ER encourages clusters with similar sizes.

The results were quantified using two standard metrics which are commonly used for evaluating the quality of superpixels: UE [28] and BR [8]. These metrics are discussed in the previous section. For this work, we manually outlined objects with pixel boundaries —



(a) BR (b) UE

Figure 2.5: Comparison between ER and FH algorithms using

this annotation, shown in Fig. 2.4 (a), is used to quantify efficacy of superpixels generated using ER and FH. Ideally, it would be better to have more manual annotations per image for validation, however, the data we used in this work is unique and involves complex details with cluttered background, which makes it challenging to interpret visually. Nonetheless, with one manual annotation, our validation should still hold good, as BR has tolerance to error in the annotation,i.e., it measures the fraction of ground truth boundaries that fall within a certain distance d of a superpixel boundary. This distance is kept as 2 pixels in our work. Fig. 2.5 shows the UE and BR plots using both ER and FH algorithms on the hyperspectral scene. These performance metrics are plotted against the number of superpixels in the image. Lower values of UE (range - [0 1]) and higher values of BR (range - [0 1]) are preferable.

In comparing these methods, with this complex natural hyperspectral scene, we have observed ER to provide a superior quality of superpixels, and hence in this thesis, we have used ER as the base superpixel algorithm. Also, in subsequent chapters, we will be providing additional validation by comparing quantification results using different datasets.

Chapter 3
Superpixels for Remote Sensing Image Analysis

3.1 Introduction and Related Work

Hyperspectral imagery (HSI) provides rich information, captured over a wide range of the electromagnetic spectrum, for each pixel and typically have hundreds of narrow contiguous bands. This abundant spectral information allows a very accurate characterization of the materials present in the image. Recent work in sensor design has substantially improved the spatial resolution of hyperspectral images. Remote sensing imagery has high spatial resolution and can provide detailed information both spectrally and spatially. Such data is hence naturally suitable for image classification tasks. Although HSI data provide beneficial information, several challenges are yet to be exploited, including problems related to curse of high dimensionality, complex statistical distributions, and incorporation of spatial neighborhood information for image analysis.

It is often observed that pixels taken separately are sensitive to noise and intra-class variability. A higher spatial resolution may lead to lower signal-to-noise ratio (SNR) and higher intra-class variation. In previous work [6], a window or grid analysis was implemented for representing local regions. However, it is observed in such approaches that the classification accuracy, being sensitive to window size, decreases drastically with an inappropriate choice of grid dimensions. Further, "optimal" grid dimensions may vary within the scene.

In order to overcome this problem, image segmentation is often preferred over a fixed window analysis. Segmentation can reduce the variability by delineating boundaries between meaningful regions. Incorporating spatial neighborhood information in image analysis, as shown in [29, 30], has the potential to enhance recognition performance. Striving to obtain a perfect segmentation is a challenge, hence, an alternate approach is an intentional oversegmentation of the image resulting in homogeneous regions called *superpixels* [8], that comprises of several contiguous pixels.

Superpixels have shown promise as a preprocessing step for object level classification [8, 9, 10] in the computer vision community. The image is represented with feature vectors that are computed at the superpixel level. Unlike traditional color and gray-scale imagery, hyperspectral imagery can suffer from additive noise. Such noise can be reduced by enforcing local smoothness which is employed by grouping contiguous image areas. Hence, superpixels can be exploited to provide noise-robust hyperspectral image analysis [14, 15]. For example, computing the mean reflectance spectra within a superpixel will result in a significant reduction in additive noise. Additionally, for HSI analysis, superpixels have the potential to identify groupings of adjacent spectra that that can therein be assumed to have been generated from a distribution particular to the material within the superpixel.

In previous work (c.f. [31, 32]), spatial dependencies between pixels have been exploited by performing spectral-spatial classification. However, as acknowledged in their works, a key limitation is that the approaches consider outcomes from pixelwise classification as the "ground truth" for selecting markers that then inform the spectral-spatial classification, leading to a potential propagation of errors. Although superpixels have been successfully utilized for data mining and related applications in the computer vision community, and for tasks such as endmember extraction in the remote sensing community, in this thesis, we propose a framework that leverages from the deliberate over-segmentation provided by superpixels for effective and efficient spatial-spectral classification. We assert that the proposed framework [33] is very generic and is conducive to any feature reduction and classification method. Our approach provides a paradigm where superpixels can be used in conjunction with ensemble classification techniques, for effective exploitation of spatial context. Specifically, a multi-classifier, decision fusion approach is utilized within each superpixel, to derive a robust classification decision at the superpixel level.

Hypothesis : Superpixel based analysis is an effective way to integrate spectral and spatial neighborhood information into remote sensing image classification. Unlike window-based methods that are sensitive to grid size, superpixels form well-defined boundaries between regions and hence reduces spectral mixing — an issue quite problematic in hyperspectral image analyis.

3.2 Proposed Approach

Fig. 3.1 provides an overview of the proposed method. Superpixels were generated by over-segmenting the hyperspectral image using a graph-theoretic approach, i.e., Entropy rate segmentation algorithm, discussed in Chapter 2, due to it's preference for close-packed and homogenous regions. Next, feature extraction is employed to address the issue of high dimensionality. Following this, pixel-level classification within a superpixel was employed

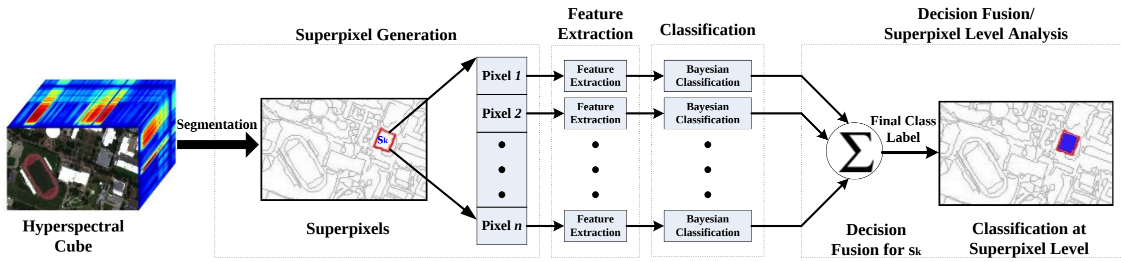


Figure 3.1: Block-level functionality of decision fusion approach for superpixels

to determine class labels/class-conditional posterior probabilities. A decision fusion approach is then invoked to fuse classification outcomes from each pixel in a superpixel to obtain final class label at the level of the superpixel.

The framework is built in a way, such that, any state-of-the-art feature extraction and classification method could be employed to get a final decision at the superpixel level. We validated our approach with commonly used feature reduction methods, such as, Linear Discriminant Analysis (LDA) and Local Fisher's Discriminant Analysis (LFDA) [34, 35], and frequently used classifiers such as the Gaussian Maximum likelihood and Infinite Gaussian mixture model (IGMM) classifiers. Decision fusion was accomplished using majority voting (MV), linear opinion pool (LOP) and logarithmic opinion pool (LOGP) per superpixel [36]. However, much of the remote sensing data is observed to be non-Gaussian and multi-modal. Therefore, feature reduction methods that aims at preserving the multimodal structure by maximizing between-class separability and simultaneously maintaining the within-class local structure are likely to perform better, particularly when coupled with statistical classifiers such as GMMs and IGMMs [36, 33].

3.2.1 Feature Extraction

Locality Preserving Fisher's Discriminant Analysis (LFDA), a recently proposed dimensionality reduction method, is an extension to traditional Linear Discriminant Analysis (LDA) [35] that does not make the assumption for the data to have a uni-modal Gaussian distribution. It has been shown to retain the local structure under the embedding, and is therefore suitable for non-Gaussian, even multi-modal data. LFDA maintains a good between-class separation in the projected subspace while preserving the within-class local structure. This is achieved by modifying the traditional LDA construct and weighing the within-class and between-class matrices used in the Fisher's ratio via an affinity matrix that represents *affinities* (e.g. through a nonlinear, radial-basis function kernel — which is the kernel function used in this work) between all pairs of points. The between class, $S^{(lb)}$, and the within class, $S^{(lw)}$, scatter matrices are defined as

$$\begin{aligned} S^{(lb)} &= \frac{1}{2}\sum_{i,j}^{n} W_{i,j}{}^{(lb)}(\mathbf{x_i}-\mathbf{x_j})(\mathbf{x_i}-\mathbf{x_j})^T, \\ S^{(lw)} &= \frac{1}{2}\sum_{i,j}^{n} W_{i,j}{}^{(lw)}(\mathbf{x_i}-\mathbf{x_j})(\mathbf{x_i}-\mathbf{x_j})^T, \end{aligned} \tag{3.1}$$

where the $n \times n$ matrices $W_{i,j}{}^{(lb)}$ and $W_{i,j}{}^{(lw)}$ are defined as

$$W_{i,j}{}^{(lb)} = \begin{cases} A_{i,j}(1/n - 1/n_l), & if \quad y_i = y_j = l \\ 1/n, & if \quad y_i \neq y_j \end{cases}$$

$$W_{i,j}{}^{(lw)} = \begin{cases} A_{i,j}/n_l, & if \quad y_i = y_j = l \\ 0, & if \quad y_i \neq y_j \end{cases} \tag{3.2}$$

and affinity matrix A is defined as

$$A_{i,j} = \exp\left(-\frac{||\mathbf{x_i}-\mathbf{x_j}||^2}{||\mathbf{x_i}-\mathbf{x_i}^{(knn)}||\cdot||\mathbf{x_j}-\mathbf{x_j}^{(knn)}||}\right). \tag{3.3}$$

LFDA exploits the manifold structure of data, and can be easily computed by solving a generalized eigenvalue problem. The neighborhood relationships are retained in LFDA by employing an affinity matrix that is described above. Since, LFDA does not force pair of points that are far-apart but belonging to the same class, to be close together in projection, it can be perceived as a localized variant of LDA. It can be easily verified that when $A_{i,j}$ = 1, for all i and j, LFDA degenerates to LDA. Due to the weaker limitation, LFDA outperforms LDA in maximizing separability between different classes. The reader is referred to [35] for a description of LFDA, and to [34] for our previous work describing its use with mixture of Gaussian classifiers for high dimensional hyperspectral imagery data.

3.2.2 Bayesian Classification

Gaussian mixture models (GMMs) is a finite parametric model that estimates the data to be distributed according to a finite number of Gaussian mixture densities. GMM based classifier, as shown in [34], can effectively capture statistics of hyperspectral data. The main drawback with GMM based approaches is that they rely on the assumption that the number of modes (components) in the model is fixed and known a-priori. Hence, a model selection scheme, such as the Bayesian Information Criteria (BIC) or Akaike Information Criteria (AIC) [37] is traditionally used to estimate the "optimal" number of modes. More recently, an extension to GMM, known as Infinte Gaussian Mixture Model (IGMM) [38] has been developed. It automatically infers the number of modes from training data and hence, does not restrict itself to the assumption. In recent preliminary work, we demonstrated that this is a very effective Bayesian classification approach for classification of hyperspectral and LiDAR data [39] at the pixel level (i.e., the study did not consider spatial context).

Data points of each class can be considered to be generated from a Gaussian mixture model with an unknown number of modes. Let $\{\mathbf{x}_i\}_{i=1}^N$ where $\mathbf{x}_i \in \mathbb{R}^d$ be the N training pixels of one specific class in the image and c_i the corresponding mode label indicating which Gaussian component $\mathbf{x}_i$ belongs to. The data $\{\mathbf{x}_i\}_{i=1}^N$ could be in the raw (input) feature space or a projected subspace (e.g. an LFDA projected subspace), although in this work, we consider the latter case, since a projection alleviates statistical ill-conditioning when employing IGMMs. A traditional Gaussian mixture model with K components is described as

$$p(\mathbf{x}_i|\pi,\theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_i|\theta_k), \tag{3.4}$$

where $\pi = (\pi_1, \pi_2, ..., \pi_K)$, $\Theta = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_K\}$ and $\sum\limits_{k=1}^{K} \pi_k = 1$. $\theta_k = (\mu_k, \sum_k)$ represents the mean vector and covariance matrix of the k^{th} component.

In the absence of prior knowledge of the "true" number of modes, K, IGMM assumes K to be ∞, while an approach that uses finite GMMs must infer the optimal number of mixtures by an information theoretic criteria.

Although we assert that the framework can utilize any supervised classification approach, we also believe that, for such non-Gaussian and multi-modal data, a Gaussian mixture model or an Infinite Gaussian mixture model (IGMM) classifier, which learns and employs a statistical model for each class within the image, would yield better results. Conceptually, the proposed idea of utilizing superpixels to extract spatial neighborhood information via decision fusion holds irrespective of the nature of the backend classifiers utilized. GMMs or IGMMs are attractive choices in this framework as the membership functions are posterior probabilities that can handle complex distributions of point clouds in the feature space arising from various sources of variability in remote sensing environments.

3.2.3 Superpixel Level Analysis via Decision Fusion

Decision fusion allows for fusion of posterior probabilities (and class-labels) in an ensemble-classifier framework. The ensemble here denotes the collection of pixel-level Bayesian classifiers within each superpixel. Our choice of a Bayesian classifier is motivated in part by the fact that the resulting approach would provide reliable estimates at the hyperspectral pixel-level. In this work, we employ and study the efficacy of three decision fusion schemes to provide superpixel level classification by fusing pixel-level classification outcomes — majority voting (MV), linear opinion pools (LOP) and logarithmic opinion pool (LOGP) [36]. However, MV is sub-optimal by design since it only "fuses" class labels based on a vote over individual class labels from each classifier in the ensemble, and hence LOP/LOGP are desirable when using Bayesian classifiers as they utilize class-conditional posterior probabilities [40]. For the k'th superpixel $\mathbf{s}_k$, (which is a matrix, $\mathbf{s}_k \in \mathbb{R}^{d\times|S_k|}$, containing hyperspectral pixels in that superpixel), a LOP formulation has the following form

$$C(\omega_i|\mathbf{s}_k) = \sum_{j=1}^{N_s} \lambda_j P_j(\omega_i|\mathbf{s}_k), \omega = \arg\max_i C(\omega_i|\mathbf{s}_k), \tag{3.5}$$

where $P_j(\omega_i|\mathbf{s}_k)$ is the individual pixel-level class-conditional posterior probability (or likelihood) corresponding to the j'th pixel, N_s is the number of pixels in the k'th superpixel, and $\{\lambda_j\}_{j=1}^{N_s}$ are weights (assumed uniform in this work). The global class membership function $C(\omega_i|\mathbf{s}_k)$ enables classification at the superpixel level, and is a weighted sum of the posterior probabilities of all classifiers.

Likewise, LOGP is a weighted product of the posterior probabilities (or likelihoods, assuming uniform class priors) of all classifiers, and, in this formulation, can be expressed as

$$C(\omega_i|\mathbf{s}_k) = \prod_{j=1}^{N_s} P_j(\omega_i|\mathbf{s}_k)^{\lambda_j}, \omega = \arg\max_i C(\omega_i|\mathbf{s}_k), \tag{3.6}$$

or

$$\log C(\omega_i|\mathbf{s}_k) = \sum_{j=1}^{N_s} \lambda_j \log P_j(\omega_i|\mathbf{s}_k), \omega = \arg\max_i \log C(\omega_i|\mathbf{s}_k). \tag{3.7}$$

Although, LOP is a simple fusion approach, it has some limitations [40], e.g., decisions from different classifiers are not treated individually in the fusion process. For IGMM classifiers, we expect LOGP (a weighted product) to fare better, owing to its ability to result in uni-modal (less-dispersed) membership functions [36]. Another benefit of LOGP is that it treats outputs from various classifiers independently. In previous work [39], it was experimentally observed that LOP provides reliable decision level fusion with Gaussian maximum-likelihood (ML) classifiers, while LOGP consistently outperforms other fusion

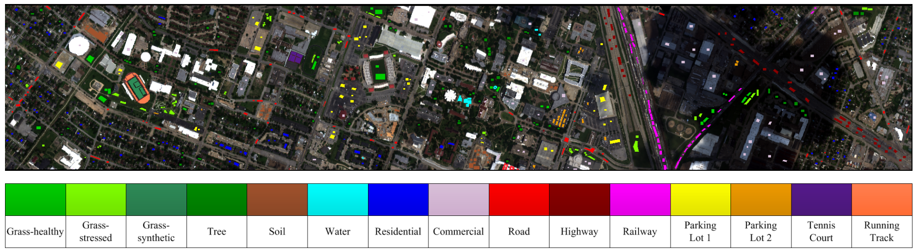
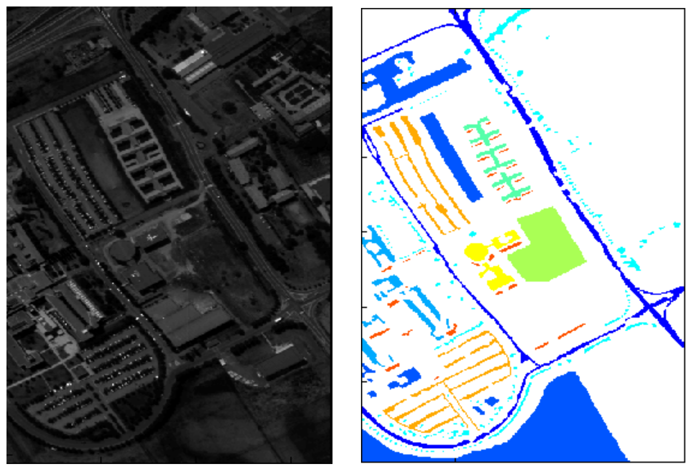


Figure 3.2: University of Houston HSI data



(a) (b)

Figure 3.3: (a) University of Pavia HSI data and (b) University of Pavia Ground truth

approaches when the base classifiers in the ensemble are mixture of Gaussians (e.g. GMMs or IGMMs).

3.3 Validation:Experimental setup and Results

3.3.1 Dataset description

In this work, we employed two datasets to validate the proposed framework — the "University of Houston" (UH) hyperspectral dataset, and the University of Pavia (PaviaU) hyperspectral dataset. The UH data was acquired by the NSF-funded National Center for

Airborne Laser Mapping (NCALM) over the University of Houston campus and the neighboring urban area and contains 15 identified urban ground cover classes and was hosted by the Hyperspectral Image Analysis group at the University of Houston for the 2013 IEEE Geoscience and Remote Sensing Data Fusion Contest, as a benchmarking dataset. This dataset contains a hypersectral image with a spatial resolution of $2.5m$. The hyperspectral image contains 144 spectral bands over the $364 \sim 1046nm$ wavelength range with a spatial size of 349 by 1905. The UH HSI image with class legends (15 classes) is shown in Fig. 3.2. The PaviaU hyperspectral data (another commonly used benchmarking hyperspectral image) was acquired by the ROSIS sensor over an urban area surrounding the University of Pavia (PaviaU) in Italy, containing 9 ground cover classes. The image has a spatial size of 610 by 340 pixels with the spatial resolution of $1.3m$ per pixel and contains 103 spectral bands and is shown in Fig. 3.3

3.3.2 Superpixel Quantification

As discussed in the earlier section, to measure the quality of superpixels generated with the hyperspectral image, outcomes were quantified using two standard metrics which are commonly used for evaluating the quality of superpixels : undersegmentation error (UE) [28] which measures the fraction of pixel leak across ground truth (manual annotation) segments, and boundary recall (BR) [8] which measures the percentage of natural boundaries recovered by the superpixel boundaries.

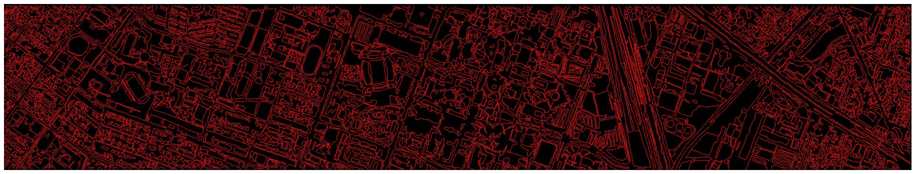


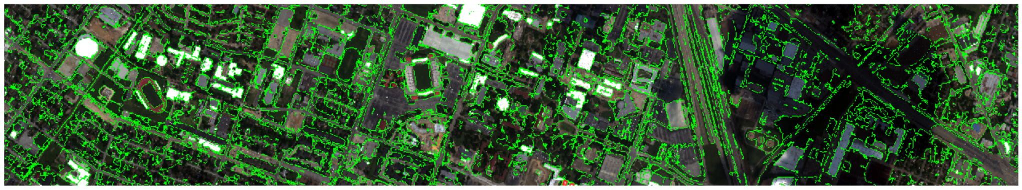
Figure 3.4: Manual Annotation — UH campus data

3.3.2.1 Quantification results with superpixel generating algorithms — ER and FH

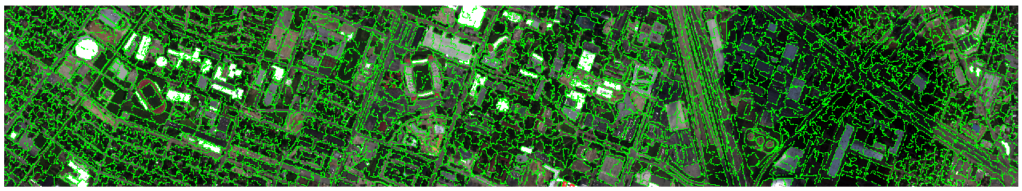
For this work, we created an annotation (by manually outlining objects with pixel boundaries) — this annotation, shown in Fig. 3.4, is used to quantify efficacy of ER with hyperspectral images. We note that unlike the computer vision community wherein several benchmarking datasets exist for quantifying efficacy of segmentation and oversegmentation algorithms, this is the first systematic manual annotation of a remotely sensed hyperspectral image for such purposes. In future, we will expand this with multiple annotations (from different annotators) for increased robustness to any biases caused by visual interpretation. Superpixels are generated using ER and FH applied to the hyperspectral image — the results over the entire UH campus data are shown in Fig. 3.5.

Also, for clarity, we are showing a cropped region of UH in Fig. 3.6, where ground truth segments are defined. The results obtained from ER and FH, over this cropped region of UH, are shown in Fig. 3.7. Looking at the figures, it can be observed that ER superpixels retains most of the object boundaries, over FH superpixels, specially regions largely covered by shadows.

Fig. 3.8 shows UE and BR quantification curves with both ER and FH, for this hyperspectral image, plotted against the number of superpixels in the image. Lower values of UE



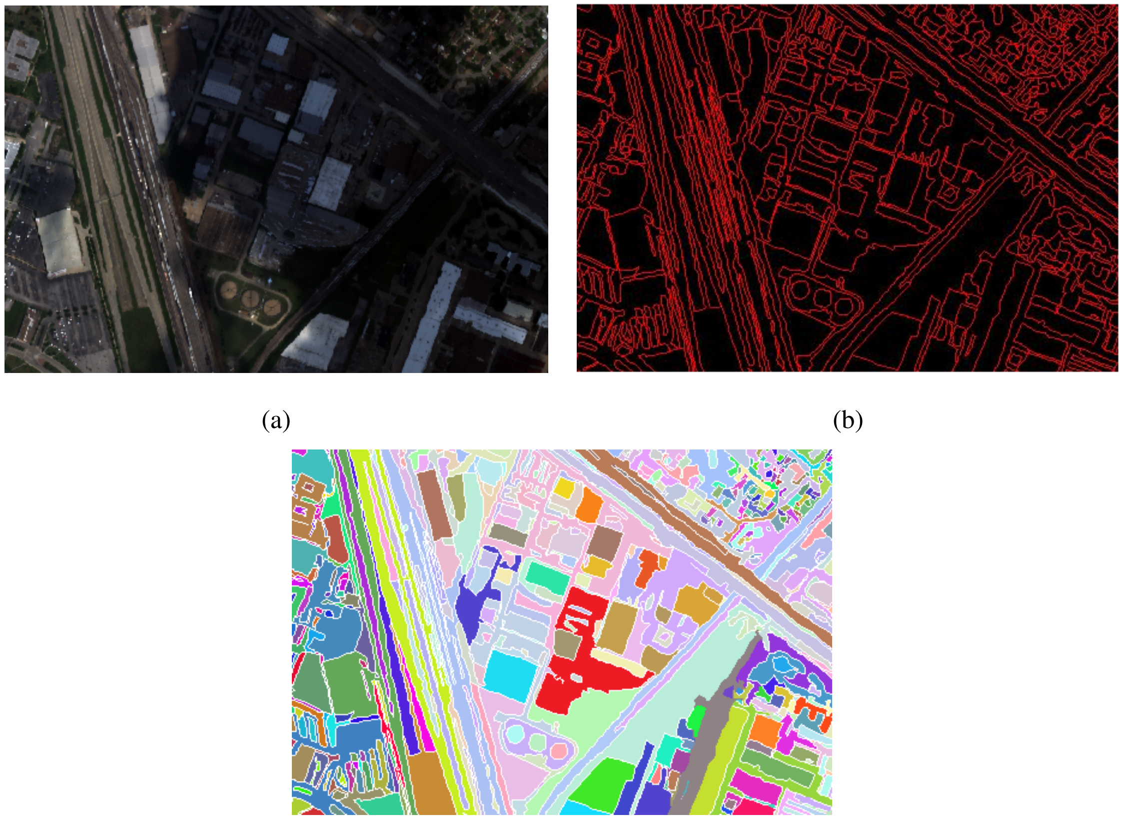
(a)



(b)

Figure 3.5: Superpixels(2000) over UH data using (a) FH (b) ER

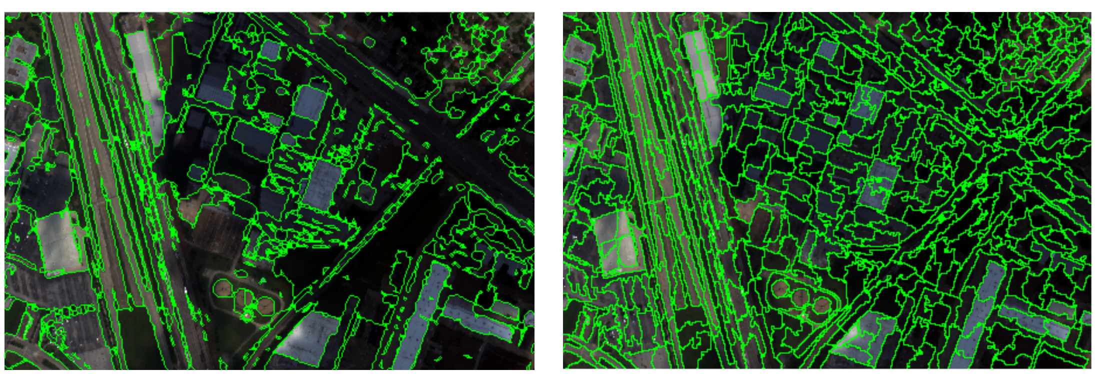
(range — [0 1]) and larger values of BR (range — [0 1]) are preferable. It can be seen that (a) larger number of superpixels give better boundary recall rates and under-segmentation error, and, (b) ER outperforms FH. Parameters for each algorithm can be tuned to obtain desired number of superpixels. For ER, number of superpixels K_s is specified by the user, while for FH, a smaller value of minimum component size and segmentation co-efficient [21] generates a larger number of superpixels. BR and UE curves help in determining the optimal values for these parameters — favorable values for these system parameters will be those which generate appropriate number of superpixels such that the BR and UE curves (Fig. 3.8) saturate to their "optimal" values. In this case, these values saturate after 50,000 — based on this observation, in what follows, we use 50,000 superpixels, although it can be expected that for similar types of scenes at similar resolutions, the ratio of total number



(a) (b)

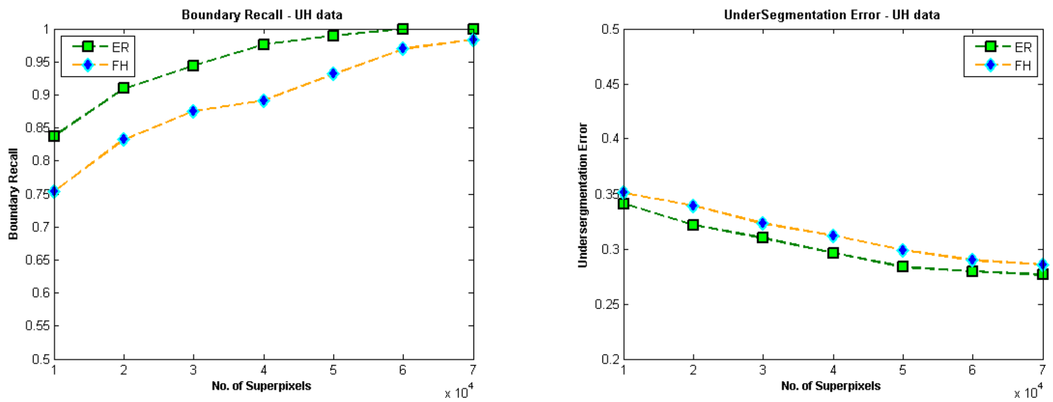
(c)

Figure 3.6: UH (a) Cropped region (b) Manual Annotations and (c) Ground truth Segments



(a) (b)

Figure 3.7: Superpixels(500) over a cropped region of UH using (a) FH (b) ER



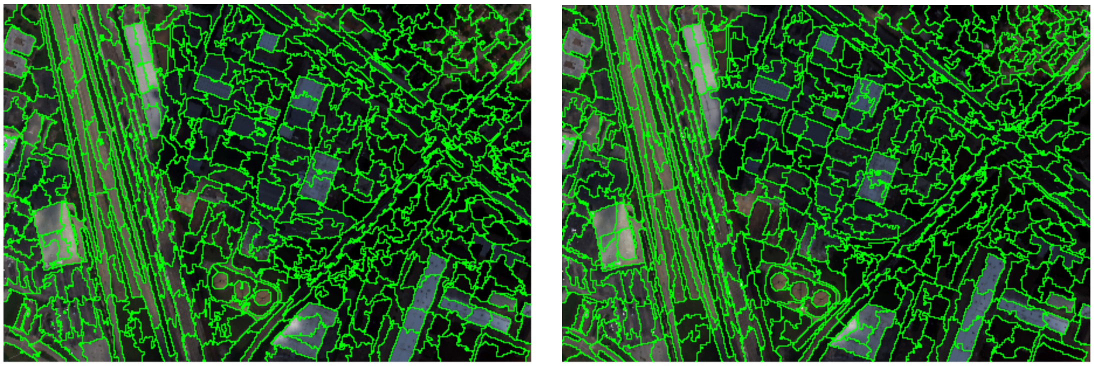
(a) (b)

Figure 3.8: Quantification Results (a) Boundary recall and (b) Undersegmentation error

of pixels to number of superpixels “required” would be similar.

3.3.2.2 Quantification results by varying the number of channels

Superpixels are widely used by the vision community for three channel (RGB) images, but they were never utilized for remote sensing hyprspectral imagery. Since, multi-channel data provides abundant spectral information across hundreds of contiguous bands, it is therefore expected that segmentation using HSI data would yield better quality superpixels. Hence, in this work, we have compared superpixels, generated using RGB data, with those generated using HSI data. As it has been established, from previous quantification results, that ER outperforms FH for this dataset, hence, we have used ER as the base superpixel generation algorithm for rest of the experiments. HSI and RGB superpixels using ER algorithm are shown in Fig. 3.9. Also, HSI data is projected to fewer bands 3, 5, 10, 15, 20 and 25, using Principal Component Analysis (PCA), following which, HSI superpixels were generated over that reduced space. UE and BR values were computed over the



(a) (b)

Figure 3.9: ER Superpixels(500) over a cropped region of UH using (a) RGB (b) HSI data

reduced HSI data and RGB data as shown in Fig. 3.10. Quantification results show that (a) HSI superpixels reduces missed boundaries and pixel leaks more than RGB Superpixels and (b) dimensionality reduction helps in improving the quality of superpixels, however, the improvement is marginal if the number of superpixels are large.

It is also observed that HSI superpixels, generated by projecting multi-channel HSI data to three channels, yields superior quality in terms of the quantification metrics.

3.3.3 Classification results

We collected pixel-level training samples randomly from the labeled data (to form a training reference library of spectra corresponding to different classes), and superpixel level test samples. Only those superpixels that overlap with the ground truth were considered. Hence, in our experiments, for the UH dataset, out of available 50000 superpixels, we used 2374 and 2216 superpixels for testing with ER and FH methods respectively, while, for PaviaU dataset, out of the 30000 superpixels, we used 7816 and 7670 superpixels for testing with ER and FH methods respectively. Further, to avoid any bias, if any test superpixel

Table 3.1: Average overall accuracy (with standard deviation) (in %) for UH

Method	No. of Training Samples per class			
	20	50	80	110
Pixel Level				
LDA-ML-RGB	50.0(1.7)	68.9(1.1)	77.0(1.2)	78.6(1.9)
LDA-ML-HSI	50.1(1.1)	71.6(1.2)	78.3(1.5)	81.1(0.8)
LFDA-IGMM-RGB	71.3(1.8)	78.5(1.2)	79.7(1.5)	80.0(1.6)
LFDA-IGMM-HSI	74.0(1.9)	80.5(2.0)	82.4(1.8)	83.5(1.5)
Spectral Averaging				
LDA-ML-Window	50.9(1.2)	72.2(1.5)	79.1(1.8)	81.8(1.4)
LDA-ML-FH-SP	55.0(1.2)	73.2(1.4)	79.5(1.5)	82.3(1.8)
LDA-ML-ER-SP-RGB	50.2(1.3)	51.9(1.8)	60.7(1.1)	69.4(2.0)
LDA-ML-ER-SP-HSI	55.7(1.3)	74.1(1.1)	80.2(1.4)	83.1(2.0)
LFDA-IGMM-Window	71.8(1.3)	77.1(1.1)	80.2(2.1)	82.9(1.3)
LFDA-IGMM-FH-SP	75.7(1.0)	81.4(1.1)	82.0(1.8)	83.5(1.4)
LFDA-IGMM-ER-SP-RGB	71.8(1.2)	72.5(1.1)	74.6(1.4)	76.9(1.9)
LFDA-IGMM-ER-SP-HSI	75.5(1.3)	81.6(1.5)	82.8(1.1)	84.2(1.2)

Table 3.2: Average overall accuracy (with standard deviation) (in %) for UH

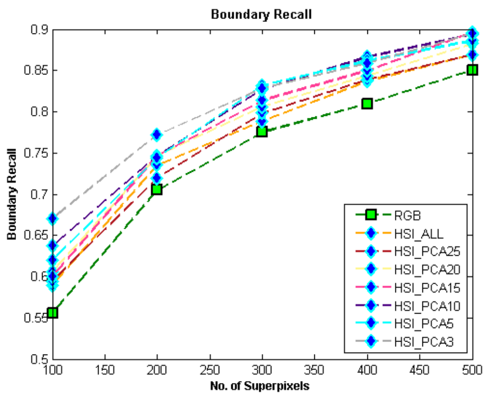
Method	No. of Training Samples per class			
	20	50	80	110
Decision Fusion				
LDA-ML-MV-Window	55.0(1.1)	74.5(1.0)	77.9(1.5)	82.1(1.2)
LDA-ML-MV-FH-SP	57.0(1.1)	76.5(1.5)	81.2(1.6)	83.5(1.1)
LDA-ML-MV-ER-SP-RGB	52.7(1.1)	73.9(1.2)	77.4(1.8)	79.0(1.1)
LDA-ML-MV-ER-SP-HSI	57.4(1.3)	78.4(1.2)	82.3(1.1)	84.1(1.7)
LDA-ML-LOP-Window	55.7(1.9)	75.1(1.3)	78.8(1.7)	82.9(1.8)
LDA-ML-LOP-FH-SP	58.9(1.7)	78.6(1.2)	82.7(1.1)	84.9(18)
LDA-ML-LOP-ER-SP-RGB	53.0(1.7)	74.5(1.1)	78.0(1.4)	79.6(1.1)
LDA-ML-LOP-ER-SP-HSI	59.1(1.2)	79.2(0.9)	83.7(1.5)	85.5(1.0)
LDA-ML-LOGP-Window	56.3(1.2)	76.9(1.9)	79.7(2.1)	83.3(1.5)
LDA-ML-LOGP-FH-SP	59.1(1.2)	79.5(1.5)	83.8(1.3)	86.0(1.9)
LDA-ML-LOGP-ER-SP-RGB	53.9(1.4)	75.2(1.2)	79.1(1.8)	80.2(1.3)
LDA-ML-LOGP-ER-SP-HSI	59.9(1.1)	80.1(1.1)	84.6(1.4)	86.7(1.4)
LFDA-IGMM-MV-Window	71.5(1.2)	77.8(1.5)	80.4(1.1)	83.9(1.4)
LFDA-IGMM-MV-FH-SP	76.2(1.0)	82.7(1.1)	84.4(1.8)	86.9(1.1)
LFDA-IGMM-MV-ER-SP-RGB	72.4(1.0)	78.5(1.9)	79.3(1.5)	80.8(1.1)
LFDA-IGMM-MV-ER-SP-HSI	77.8(1.5)	83.4(1.7)	85.9(1.4)	87.8(1.3)
LFDA-IGMM-LOP-Window	72.2(1.8)	78.7(1.6)	81.2(2.0)	84.8(1.3)
LFDA-IGMM-LOP-FH-SP	77.9(1.9)	83.8(1.2)	85.9(1.1)	87.7(1.1)
LFDA-IGMM-LOP-ER-SP-RGB	73.1(1.3)	79.2(1.5)	80.1(1.8)	81.6(1.2)
LFDA-IGMM-LOP-ER-SP-HSI	76.8(2.1)	83.4(1.8)	85.2(1.7)	88.5(1.1)
LFDA-IGMM-LOGP-Window	74.4(2.1)	80.3(1.1)	83.9(1.5)	87.3(1.8)
LFDA-IGMM-LOGP-FH-SP	78.7(1.0)	85.5(1.3)	87.0(1.4)	89.2(1.5)
LFDA-IGMM-LOGP-ER-SP-RGB	74.9(1.1)	80.8(1.6)	81.5(1.5)	82.7(1.6)
LFDA-IGMM-LOGP-ER-SP-HSI	78.4(1.4)	85.7(1.7)	87.6(1.6)	91.1(1.1)

Table 3.3: Average overall accuracy (with standard deviation) (in %) for Pavia

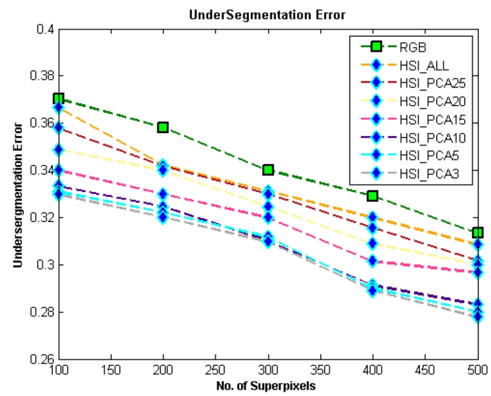
Method	**No. of Training Samples per class**			
	20	**50**	**80**	**110**
Pixel Level				
LDA-ML-RGB	54.6(1.3)	77.2(1.1)	79.3(1.7)	81.4(1.6)
LDA-ML-HSI	55.1(1.1)	78.1(1.2)	80.0(1.5)	82.5(1.4)
LFDA-IGMM-RGB	74.0(1.5)	79.5(1.1)	81.8(1.3)	83.1(1.0)
LFDA-IGMM-HSI	76.8(1.3)	81.1(1.5)	83.7(1.7)	84.6(1.1)
Spectral Averaging				
LDA-ML-Window	56.2(1.2)	77.9(1.1)	80.7(1.9)	81.4(1.3)
LDA-ML-FH-SP	59.1(1.1)	79.5(1.0)	81.9(1.2)	83.0(1.4)
LDA-ML-ER-SP-RGB	55.6(1.7)	77.9(1.1)	80.1(1.0)	81.0(1.4)
LDA-ML-ER-SP-HSI	59.8(1.9)	80.8(1.2)	82.6(1.3)	83.9(1.5)
LFDA-IGMM-Window	75.1(1.4)	79.8(1.2)	82.4(1.1)	83.9(1.8)
LFDA-IGMM-FH-SP	77.2(1.7)	81.1(1.1)	84.6(1.8)	85.3(1.5)
LFDA-IGMM-ER-SP-RGB	74.8(1.1)	80.2(1.4)	82.6(1.3)	84.0(1.1)
LFDA-IGMM-ER-SP-HSI	77.9(1.2)	82.4(1.4)	85.0(1.1)	86.1(1.2)

Table 3.4: Average overall accuracy (with standard deviation) (in %) for Pavia

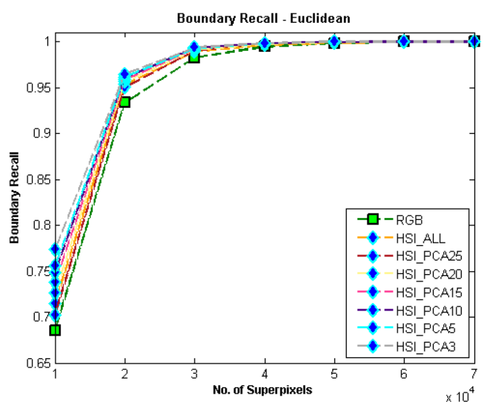
Method	No. of Training Samples per class			
	20	**50**	**80**	**110**
Decision Fusion				
LDA-ML-MV-Window	60.2(1.2)	80.9(1.1)	82.3(1.9)	83.5(1.4)
LDA-ML-MV-FH-SP	61.9(1.0)	81.5(1.1)	84.0(1.6)	85.2(1.3)
LDA-ML-MV-ER-SP-RGB	58.6(1.0)	77.7(1.2)	81.5(1.4)	82.2(1.8)
LDA-ML-MV-ER-SP-HSI	62.7(1.3)	82.4(1.5)	84.8(1.4)	85.7(1.6)
LDA-ML-LOP-Window	63.1(1.8)	81.8(1.7)	83.4(1.6)	85.6(1.1)
LDA-ML-LOP-FH-SP	62.8(1.4)	82.6(1.2)	84.4(1.0)	86.0(1.4)
LDA-ML-LOP-ER-SP-RGB	60.1(1.8)	79.5(1.3)	82.3(1.0)	84.1(1.5)
LDA-ML-LOP-ER-SP-HSI	64.6(1.7)	83.9(1.2)	85.7(1.5)	87.0(1.4)
LDA-ML-LOGP-Window	64.4(1.6)	82.6(1.1)	84.5(1.7)	86.7(1.9)
LDA-ML-LOGP-FH-SP	63.5(1.1)	82.0(1.0)	85.3(1.2)	86.8(1.6)
LDA-ML-LOGP-ER-SP-RGB	62.1(1.3)	81.0(1.6)	83.5(1.4)	85.1(1.9)
LDA-ML-LOGP-ER-SP-HSI	65.8(1.0)	85.1(1.3)	86.9(1.7)	88.2(1.9)
LFDA-IGMM-MV-Window	77.3(1.1)	82.2(1.3)	84.6(1.8)	85.9(1.1)
LFDA-IGMM-MV-FH-SP	78.6(1.1)	83.3(1.5)	85.7(1.2)	87.3(1.2)
LFDA-IGMM-MV-ER-SP-RGB	75.7(1.3)	81.6(1.2)	82.8(1.1)	84.5(1.8)
LFDA-IGMM-MV-ER-SP-HSI	79.1(1.2)	84.0(1.1)	86.5(1.3)	87.9(1.9)
LFDA-IGMM-LOP-Window	79.3(1.8)	84.4(1.3)	86.3(1.7)	87.6(1.5)
LFDA-IGMM-LOP-FH-SP	80.4(1.3)	85.9(1.1)	88.0(1.9)	88.7(1.0)
LFDA-IGMM-LOP-ER-SP-RGB	77.4(1.1)	83.2(1.5)	84.3(1.6)	85.7(1.2)
LFDA-IGMM-LOP-ER-SP-HSI	81.3(1.8)	86.7(1.2)	88.2(1.6)	89.3(1.4)
LFDA-IGMM-LOGP-Window	80.5(1.1)	85.6(1.4)	87.8(1.5)	88.7(1.7)
LFDA-IGMM-LOGP-FH-SP	81.8(1.7)	87.0(1.4)	88.7(1.6)	89.8(1.2)
LFDA-IGMM-LOGP-ER-SP-RGB	79.1(1.1)	84.5(1.3)	86.3(1.8)	87.2(1.1)
LFDA-IGMM-LOGP-ER-SP-HSI	82.5(1.1)	87.8(1.8)	89.5(1.4)	91.4(1.5)



(a) BR-cropped image



(b) UE-cropped image



(c) BR-entire image

Figure 3.10: Superpixel Quantification Results

contained even a single training pixel, we discarded the superpixel from our validation dataset. As a result of this sequestration, as the number of training pixels increases, the number of test superpixels decreases (albeit slightly). By choosing non-overlapping regions of interest for training and testing, this reduction is minimized — for instance, when using ER approach, the range of total testing superpixels (over all classes) drops from 2343 to 2192, for UH, and, from 7636 to 6826, for PaviaU, as the number of training pixels per class are increased from 20 through 110. The results as quantified by overall accuracy were evaluated as a function of different number of training samples per class, and summarized in Tab. 3.1, Tab. 3.2, Tab. 3.3 and Tab. 3.4.

Each entry in the table corresponds to the mean accuracy from 10 random runs (i.e., randomly selecting different training points and test superpixels from the image), along with the standard deviation around the mean. We present results with — (1) *Pixel level* approach, where training and test pixels are drawn from the labeled pool of hyperspectral data, (2) *Spectral averaging* followed by per-pixel classification, i.e., computing mean spectral content overall all pixels within each superpixel, followed by pixel-level classification using LDA and a quadratic Gaussian maximum-likelihood (ML) classifier (LDA-ML), or LFDA-IGMM on the *smoothed* hyperspectral signatures; (3) Superpixel based *decision fusion* approach (incorporating spatial context): Different variants of base-classifiers (LDA-ML/LFDA-IGMM) and decision fusion approaches (MV/LOP/LOGP). Additionally, we compared classification results, using corresponding RGB data (we extracted red, blue, and green channels to form a natural color RGB image — bands 21 (460*nm*), 40 (550*nm*) and 59 (640*nm*) were selected for UH data and bands 18 , 38 and 55 for Pavia data). For superpixel based approach RGB superpixels were generated using ER and corresponding RGB data. Also, the superpixel based approaches — Spectral averaging and Decision fusion, were compared with a window-based-approach, where a fixed size sliding window, 5 x 5 for UH data and 4 x 4 for PaviaU data (chosen to match the average number of pixels in a superpixels for the datasets, for fair comparison), is defined across every test pixel, followed by pixel-level classification within each window.

It is clear from these results that for hyperspectral remote sensing data (a) a superpixel

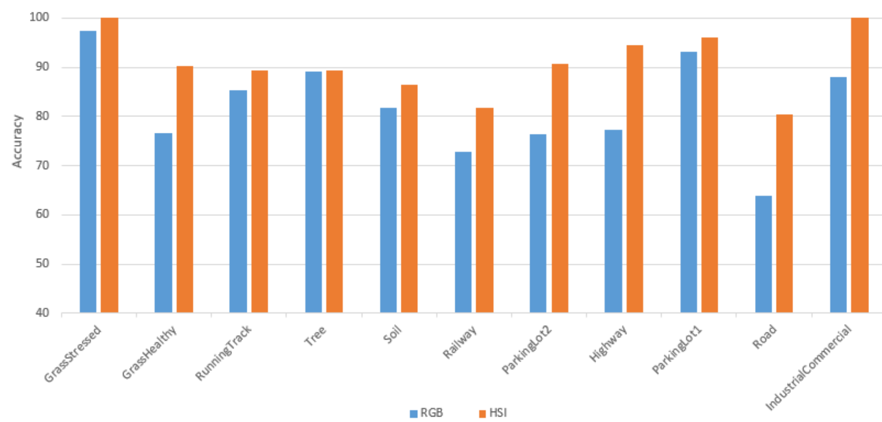


Figure 3.11: Confusion Matrix Plot — UH data (using LFDA-IGMM-LOGP)

based decision fusion approach outperforms per-pixel and window-based-approaches, even when using very limited training data, (b) an over-segmentation using ER and FH methods reveal similar classification performance, even though ER provides superior quality of superpixels, and (c) additional spectral information, captured over a wide range of the electromagnetic spectrum, leads to significantly improved recognition performance. A bar plot depicting class specific accuracies, with UH HSI and RGB data, using LFDA-IGMM-LOGP, is shown in Fig. 3.11. Note that HSI outperforms RGB significantly for "hard classes" such as — Grass healthy, Parking lot 2, Highway, Road and Industrial Commercial. Also, our approach would yield better performance when using higher spatial resolution imagery due to the improved quality of over-segmentation, and appropriateness of our working assumption that each superpixel helps in creating a pool of data assumed to be drawn from identical distributions.

A graph of accuracy plotted against the number of superpixels is shown in Fig. 3.12. We also employed two segmentation algorithms commonly utilized in prior spectral-spatial

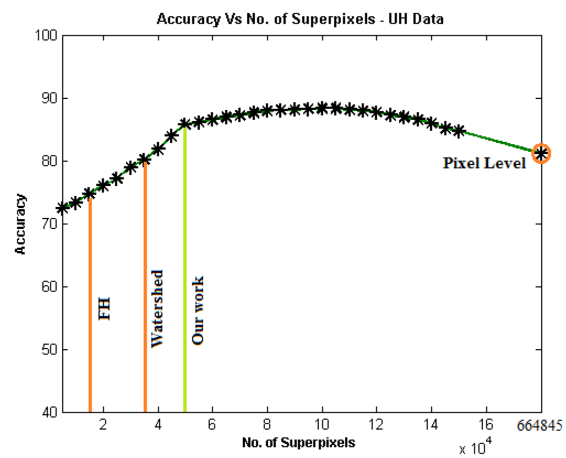


Figure 3.12: Accuracy vs No. of Superpixels

work, including watershed [31] and FH, to segment the UH dataset. It was observed that watershed gives 37720 segments, while FH gives 13843 segments. Using the proposed approach, the accuracy that can be achieved with segments ranging from 10000 - 40000 is between 73.2% - 81.8%, while it is approximately 80.5% when operating at the pixel level — both of which are far less than that obtained with 50000 segments (used in our experiments, motivated by Fig. 3.12). This verifies our assertion that over-segmentation as provided by superpixels provides a natural framework for spatial-spectral classification via the proposed Bayesian ensemble classification framework.

3.4 Conclusion

In this work, we presented a new approach to exploiting spatial context via superpixels for hyperspectral image analysis. We created a systematic ground-truth of boundaries via manual annotation, and thoroughly validated the efficacy of ER (a technique that has previously only been studied with natural color images — to our knowledge, this is the first time such an annotation has been created and superpixel efficacy validated with remotely sensed hyperspectral data) as a superpixel generation tool for high dimensional hyperspectral imagery. We then used the resulting superpixels with an ensemble of classifiers — incorporating spatial context via superpixels using the proposed approach results in robust classification, even with very little training data. Interestingly, although ER provides superior superpixels in terms of popular quantification metrics (UE and BR), the classification accuracies resulting from an over-segmentation using both methods reveal similar classification performance when using superpixels generated by these methods. Also, as expected, an improvement in classification performance is observed, when LFDA is used in conjunction with IGMM based classifier, owing to their ability to successfully preserve and capture complex multi-modal statistical structure of remote sensing HSI data, although, the proposed framework can be easily extended to any feature reduction and classification approach. However, this enhancement in performance comes with added computational complexity, that comes with the use of statistical Bayesian methods. We note that the present framework will specifically work well when images with higher spatial resolution are used. Superpixel generation itself is computationally very efficient (e.g., partitioning the entire UH hyperspectral image into $50,000$ contiguous superpixels using ER took approximately 19 seconds on a dual-core, $3GHz$ CPU). From the results, we conclude that superpixels provide a very effective, simple and unsupervised pre-processing for supervised analysis of remotely sensed images.

Chapter 4
Superpixels for natural hyperspectral image analysis - a pilot study

4.1 Introduction and Related work

Among the various applications of multi-camera surveillance systems, person reidentification is an active research area. Re-identification aims to recognize a person separated in location and time. Although, detection and extraction of the desired target decreases computational time and helps in the matching process, re-identification still remains a difficult problem owing to several challenges — (1) variation in illumination conditions, (2) variation in pose and (3) different viewpoints etc. There has been substantial amount of work [41, 42, 43], in the past years to address these challenges. Current state-of-the-art can be categorized into two categories: one category that focuses on extracting features that are pose, viewpoint and illumination invariant [41, 42], while another category emphasizes on utilizing similarity metrics [43]. By analyzing various approaches, we find that a majority of these methods take into account global appearance of a person, such as weighted color histogram [42]. Such features represent the dominant characteristics very well, but, minute details, such as a unique patch on the clothes, can not be successfully described. As a result, often mismatching occurs. Moreover, all the previous approaches perform re-identification by analyzing the information of a small cropped area, which is typically composed of one or more rectangles around eyes, nose and lips [44]. However, there are other important facial features, such as, features extracted from forehead, chin, and hair areas, which are not well studied since, they are generally hard to be detected and extracted owing to their non-rigid shapes.

A standard pre-processing step in many recognition tasks is to partition the input image into multiple segments, where each segment represents meaningful and consistent features. The goal of segmentation is to decompose an image into meaningful regions, resulting in a higher level representation of the image pixels. A coarser partition would result in regions that could extend across boundaries between perceptually distinct segments and are no longer compact. Hence, an oversegmentation of the image into a set of superpixels [8] : "semantically meaningful atomic regions", is preferred. Regions extracted by oversegmentation, have their true boundaries preserved, and form a representation of an image that is much more compact than the original pixel grid. In recent years, superpixels, have been successfully applied to natural images for image segmentation [11], object localization [12] and tracking [13]. Superpixels enable us to measure feature statistics on a naturally adaptive domain. Such representation captures redundancy in the image and is computationally efficient — reduces the complexity of images from hundreds of thousands of pixels to only a few hundred superpixels. Also, superpixels have the potential to avoid misalignment [16] caused by object variance on the Histogram of Gradient [45] and Haar-like [46] features.

Due to the inherent benefits of superpixels, there has been significant amount of work in the computer vision community that incorporates superpixel level analysis. Recently, matching and person re-identification are two such applications where superpixels have significantly improved the detection performance. In most of such approaches re-identification is implemented by forming a vocabulary tree [16, 17, 18, 19] based on local features extracted at the superpixel level. Superpixel segmentation is incorporated to generate visual patches as regions of interest for local features. Following this, a vocabulary tree is created that contains each local feature as a visual word [16]. It is observed that the boundaries and details of the object are well detected when superpixels are utilized to segment the image into visual patches. In such scenarios, different images of the same individual would have similar visual patches, making superpixel based local features a reasonable cue for matching and re-identification.

Although, superpixel based approaches seem to work well with traditional color imagery, they have never been studied for natural hyperspectral images. It has been observed that hyperspectral images can suffer from additive noise. Traditionally, such noise can be reduced by imposing local smoothness, which is achieved by grouping image areas into superpixels [14]. We acknowledge that a vocabulary tree based approach would be the optimal use of superpixels for such a problem. However, such an effort would be beyond the scope of this thesis, which is focused on studying the efficacy of hyperspectral superpixels. As a case study, we consider the person re-identification problem, for which a small pilot dataset was acquired using a hyperspectral camera, and demonstrate that hyperspectral superpixels for this task perform well when compared to the corresponding "RGB" image. Specifically, a spectral angle based distance metric (which has been shown to possess illumination invariance for spectral matching tasks [47, 48, 49]) is suitable when setting up the re-identification task at the superpixel level.

Hypothesis : Superpixel based approach is an effective pre-processing step for natural scene image analysis. Superpixels have the potential to facilitate a very simple and effective re-identification.

4.2 Dataset description

In this work, we have introduced a new dataset, captured outdoors on the University of Houston campus. The data was acquired using a Headwall Photonics hyperspectral imager which provided measurements in 325 spectral bands spanning the visible and near-infrared spectrum from $400nm - 1000nm$ uniformly. In order to evaluate the performance of a re-identification model, the dataset must represent commonly encountered confounding factors such as viewpoint and illumination variation. Hence, our data, consisting of images of 15 humans, was collected at different times of the day — before noon and afternoon. Also, each person was asked to pose arbitrarily for the camera in order to get arbitrary viewpoints. Due to the temporal difference of the captured data, variations in physiological manifestations were also accommodated. One of the hyperspectral images from the dataset is shown in Fig. 4.1. The spatial size of the image is 1004 by 400.

4.3 Validation study for person re-identification

Human detection plays an important role in several applications, such as, video surveillance, face recognition and human computer interface. It has been observed, in many recognition tasks, that the object usually covers an insignificant part of the image. Consequently, partitioning of images into perceptually relevant regions can greatly accelerate the process of identification. Hence, our approach starts with an initial oversegmentation, such that, the image is divided into homogenous partitions called superpixels. ER algorithm, discussed

(a) AM (b) PM (c) AM (d) PM

Figure 4.1: Hyperspectral images of Person 1 and Person 2

in Chapter 2, is utilized for generation of superpixels, as it favors compact and uniform regions and the results from this are shown in Fig. 4.2.

Superpixels were generated over the face dataset using HSI and corresponding RGB data (we selected red, green and blue channels, to form a natural color RGB image — bands 34 (460*nm*), 82 (549*nm*) and 131 (639*nm*) were chosen from the visible range of the dataset). It can be observed from Fig. 4.2 that while RGB superpixels failed to capture the variability in clothes of the person, HSI superpixels seem to identify different colors and patterns accurately with very little confusion. We will be utilizing this fact, in the following experiment, to study the benefits of a simple superpixel based re-identification task.

The principal advantage of hyperspectral over conventional color imagery is that materials can be identified by their unique spectral reflectance profiles. These spectral signatures

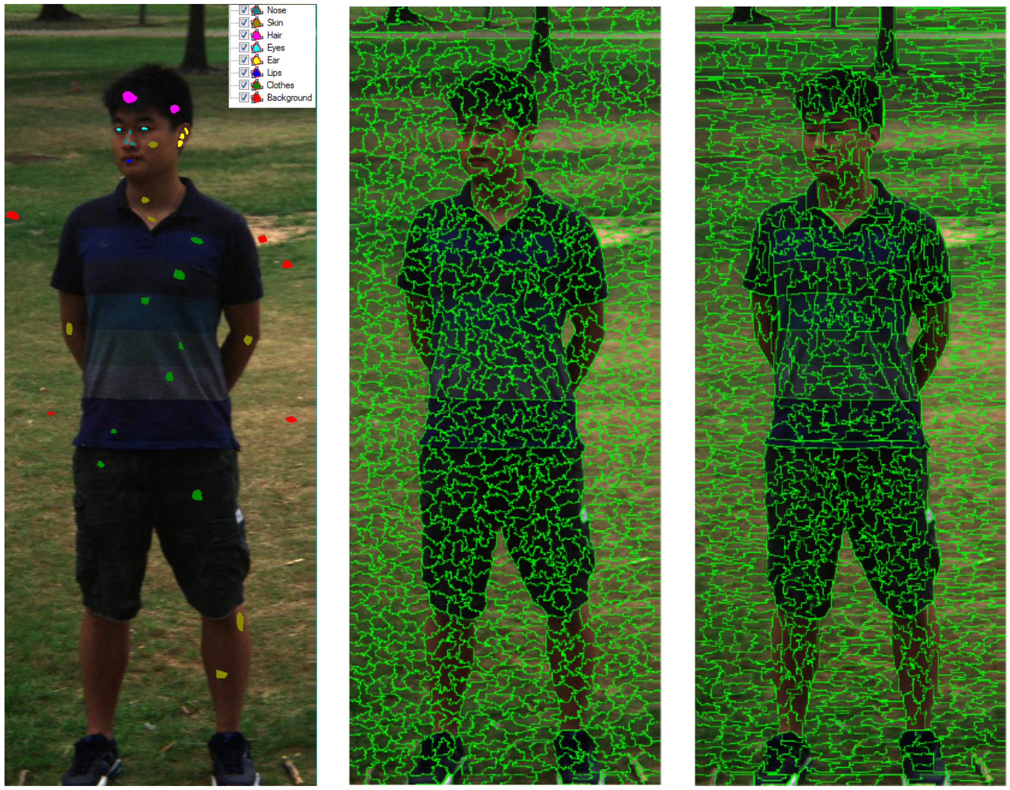


Figure 4.2: (a) Manual Annotation (Ground truth) and Superpixels (b) RGB (c) HSI

are distinctive of the material characteristics, hence, creating a spectral library of signatures will be an effective way to recognize unknown objects by comparing reflectance from any unidentified object with a library of known materials. Mean spectral signatures of all classes present in the image is shown in Fig. 4.3.

In [16, 17, 18, 19], superpixels were used to provide spatial context for the re-id problem. Hence, it would be appropriate to test the efficacy and quality of hyperspectral superpixels of natural images for matching and re-identification, where relative location and neighborhood information from superpixels can be very useful. Traditionally, for RGB images, color histogram [50] is more robust than other feature descriptors. This is because, as

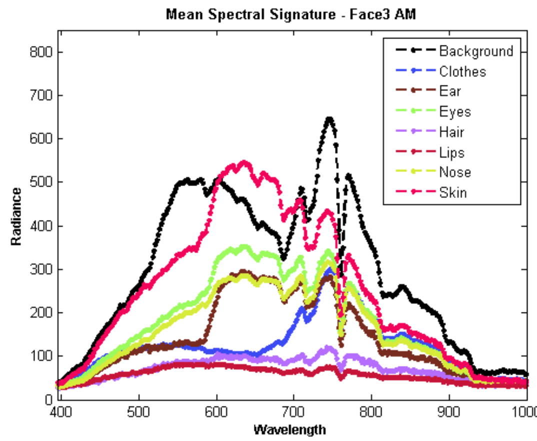


Figure 4.3: Mean Spectral Signature - All Classes : Face 3 AM

the image is oversegmented, the small regions often vary substantially in shape, while colors from the same object possess high similarity. However, for multi-channel images, a color histogram of all channels will result in a very large feature space, which, in turn, will make the computations exhaustive. Hence, we employed spectral angle distance as the similarity measure in our approach, due to it's potential to quantify the variations in hyperspectral data [47]. Another benefit of this metric is illumination invariance [47, 48, 49]. Let S_i and S_j represent any two regions of the initial oversegmentation given by $S_{os} = \{S_1, S_2, S_3, ..., S_{K_s}\}$, where K_s is the number of superpixels. The spectral angle distance $d_{sa}(S_i, S_j)$ between the two regions S_i and S_j can be defined as

$$d_s a(S_i, S_j) = \frac{\hat{\mathbf{x}}_{S_i} \hat{\mathbf{x}}_{S_j}}{\| \hat{\mathbf{x}}_{S_i} \| \| \hat{\mathbf{x}}_{S_j} \|}, \tag{4.1}$$

where, $\hat{\mathbf{x}}_{S_i}$ is the mean spectrum of region S_i across all spectral wavelenghts λ.

For the next set of experiments, the dataset is divided into two parts, each comprising of 15 images. The first set, used as the reference set, consists of images taken before noon. While, the second set, used as the test set, consists of images of the same set of people, taken in the afternoon. An initial oversegmentation is employed to get all superpixels that overlap the ground truth segments, and then, spectral averages are computed over all pixels within these superpixels. This is followed by determining the spectral angle distance between each pair of images, and, for every category. The distances, also interpreted as similarity scores, are reported in the tables below. Each entry in the table corresponds to the spectral angle distance between any pair of images. Maximum spectral angle distance for an image pair denotes maximum similarity between the two images. Matching or re-identification occurs when an image of a person taken before noon gets the highest score against the image of same person taken in the afternoon. Highest scores are marked in red.

Identification performance is evaluated using the Cumulative Matching Characteristic (CMC) curves. The CMC curve represents the expectation of finding the correct match in the top *r* matches. In other words, a rank-*r* recognition rate shows the percentage of the test images that are correctly recognized from the top *r* matches in the reference set. Comparison between HSI and RGB superpixel based approach, for every category, up to rank 15 are shown in the Fig. 4.4.

By observing at the CMC curves, it can be interpreted that (a) abundant spectral information from hyperspectral data helps in the task of re-identification and (b) skin seems to be a good identifier for a matching algorithm using spectral content, while, hair seems to be a poor identifier. Also, for re-identification based on clothes, hyperpspectral data shows less discriminatory power, due to a lot of variability in clothes of different persons. Our dataset

Table 4.1: Similarity Score for Clothes : RGB Superpixels

Person AM\PM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.9997	0.9884	0.9919	0.9949	0.998	0.9939	0.9989	0.9933	0.9906	0.9973	0.9992	0.9882	0.9991	0.9976	0.9896
2	0.9877	0.9791	0.9835	0.9881	0.9932	0.987	0.9948	0.9858	0.9821	0.9924	0.9955	0.9791	0.9958	0.9925	0.9809
3	0.9900	0.9823	0.9862	0.9905	0.9948	0.9895	0.9961	0.9884	0.9851	0.9943	0.9967	0.9823	0.9971	0.9943	0.9840
4	0.9918	0.9849	0.9883	0.9983	0.9961	0.9915	0.9971	0.9904	0.9874	0.9957	0.9976	0.9849	0.9981	0.9956	0.9864
5	0.9912	0.9841	0.9876	0.9917	0.9969	0.9908	0.9968	0.9897	0.9867	0.9952	0.9973	0.9841	0.9978	0.9951	0.9856
6	0.9896	0.9815	0.9857	0.9899	0.9945	0.9888	0.996	0.9878	0.9843	0.9937	0.9966	0.9814	0.9968	0.9939	0.9831
7	0.9891	0.9810	0.9851	0.9895	0.9942	0.9885	0.9956	0.9873	0.9839	0.9935	0.9962	0.9811	0.9966	0.9936	0.9827
8	0.9871	0.9783	0.9827	0.9875	0.9927	0.9864	0.9943	0.9961	0.9814	0.9919	0.9950	0.9783	0.9954	0.9920	0.9801
9	0.9981	0.9990	0.9991	0.9980	0.9949	0.9985	0.9930	0.9988	0.9996	0.9957	0.9924	0.9992	0.9922	0.9955	0.9992
10	0.9902	0.9824	0.9864	0.9906	0.9950	0.9896	0.9963	0.9885	0.9852	0.9974	0.9969	0.9824	0.9972	0.9944	0.9840
11	0.9875	0.9788	0.9832	0.9879	0.993	0.9867	0.9946	0.9856	0.9819	0.9922	0.9963	0.9788	0.9957	0.9923	0.9806
12	0.9896	0.9817	0.9857	0.9900	0.9945	0.9890	0.9959	0.9879	0.9845	0.9939	0.9965	0.9817	0.9969	0.9939	0.9833
13	0.9942	0.9877	0.9912	0.9944	0.9977	0.9935	0.9986	0.9928	0.9900	0.9971	0.9989	0.9876	0.9990	0.9973	0.9890
14	0.994	0.9876	0.9911	0.9943	0.9976	0.9934	0.9985	0.9927	0.9899	0.9970	0.9989	0.9875	0.9971	0.9972	0.9889
15	0.9933	0.9867	0.9900	0.9936	0.9971	0.9928	0.9980	0.9919	0.9891	0.9967	0.9984	0.9867	0.9987	0.9966	0.9881

Table 4.2: Similarity Score for Clothes : HSI Superpixels

Person AM\PM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.9997	0.9884	0.9919	0.9949	0.998	0.9939	0.9989	0.9933	0.9906	0.9973	0.9992	0.9882	0.9991	0.9976	0.9896
2	0.9877	0.9791	0.9835	0.9881	0.9932	0.987	0.9948	0.9858	0.9821	0.9924	0.9955	0.9791	0.9958	0.9925	0.9809
3	0.9900	0.9823	0.9862	0.9905	0.9948	0.9895	0.9961	0.9884	0.9851	0.9943	0.9967	0.9823	0.9971	0.9943	0.9840
4	0.9918	0.9849	0.9883	0.9983	0.9961	0.9915	0.9971	0.9904	0.9874	0.9957	0.9976	0.9849	0.9981	0.9956	0.9864
5	0.9912	0.9841	0.9876	0.9917	0.9969	0.9908	0.9968	0.9897	0.9867	0.9952	0.9973	0.9841	0.9978	0.9951	0.9856
6	0.9896	0.9815	0.9857	0.9899	0.9945	0.9888	0.996	0.9878	0.9843	0.9937	0.9966	0.9814	0.9968	0.9939	0.9831
7	0.9891	0.9810	0.9851	0.9895	0.9942	0.9885	0.9956	0.9873	0.9839	0.9935	0.9962	0.9811	0.9966	0.9936	0.9827
8	0.9871	0.9783	0.9827	0.9875	0.9927	0.9864	0.9943	0.9961	0.9814	0.9919	0.9950	0.9783	0.9954	0.9920	0.9801
9	0.9981	0.9990	0.9991	0.9980	0.9949	0.9985	0.9930	0.9988	0.9996	0.9957	0.9924	0.9992	0.9922	0.9955	0.9992
10	0.9902	0.9824	0.9864	0.9906	0.9950	0.9896	0.9963	0.9885	0.9852	0.9974	0.9969	0.9824	0.9972	0.9944	0.9840
11	0.9875	0.9788	0.9832	0.9879	0.993	0.9867	0.9946	0.9856	0.9819	0.9922	0.9963	0.9788	0.9957	0.9923	0.9806
12	0.9896	0.9817	0.9857	0.9900	0.9945	0.9890	0.9959	0.9879	0.9845	0.9939	0.9965	0.9817	0.9969	0.9939	0.9833
13	0.9942	0.9877	0.9912	0.9944	0.9977	0.9935	0.9986	0.9928	0.9900	0.9971	0.9989	0.9876	0.9990	0.9973	0.9890
14	0.994	0.9876	0.9911	0.9943	0.9976	0.9934	0.9985	0.9927	0.9899	0.9970	0.9989	0.9875	0.9971	0.9972	0.9889
15	0.9933	0.9867	0.9900	0.9936	0.9971	0.9928	0.9980	0.9919	0.9891	0.9967	0.9984	0.9867	0.9987	0.9966	0.9881

Table 4.3: Similarity Score for Hair : RGB Superpixels

Person AM\PM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.9979	0.9960	0.9973	0.9972	0.9973	0.9973	0.9972	0.9976	0.9976	0.9975	0.9975	0.9977	0.9974	0.9974	0.9972
2	0.9978	0.9966	0.9978	0.9978	0.9979	0.9978	0.9977	0.9981	0.9981	0.9980	0.9981	0.9982	0.9979	0.9979	0.9977
3	0.9921	0.9901	0.9922	0.9921	0.9923	0.9922	0.9920	0.9928	0.9927	0.9926	0.9926	0.9928	0.9924	0.9925	0.9921
4	0.9924	0.9903	0.9925	0.9924	0.9926	0.9925	0.9922	0.9931	0.9929	0.9928	0.9928	0.9931	0.9927	0.9927	0.9923
5	0.9941	0.9923	0.9942	0.9941	0.9943	0.9942	0.9939	0.9947	0.9946	0.9944	0.9944	0.9947	0.9943	0.9944	0.9940
6	0.9928	0.9908	0.9929	0.9928	0.9930	0.9929	0.9927	0.9935	0.9934	0.9932	0.9932	0.9935	0.9931	0.9931	0.9928
7	0.9932	0.9912	0.9933	0.9932	0.9934	0.9933	0.9930	0.9938	0.9937	0.9936	0.9936	0.9938	0.9935	0.9935	0.9931
8	0.9931	0.9911	0.9931	0.9931	0.9932	0.9931	0.9929	0.9937	0.9936	0.9934	0.9934	0.9931	0.9933	0.9934	0.9930
9	0.9967	0.9954	0.9968	0.9967	0.9968	0.9968	0.9967	0.9972	0.9961	0.9970	0.9971	0.9972	0.9969	0.9970	0.9967
10	0.9966	0.9952	0.9967	0.9966	0.9967	0.9967	0.9965	0.9970	0.9970	0.9969	0.9969	0.9971	0.9968	0.9968	0.9965
11	0.9965	0.9951	0.9966	0.9965	0.9966	0.9966	0.9964	0.9970	0.9969	0.9968	0.9968	0.9970	0.9967	0.9967	0.9964
12	0.9963	0.9948	0.9963	0.9963	0.9964	0.9963	0.9962	0.9967	0.9966	0.9965	0.9965	0.9968	0.9965	0.9965	0.9962
13	0.9972	0.9959	0.9973	0.9972	0.9973	0.9972	0.9971	0.9976	0.9975	0.9974	0.9974	0.9976	0.9974	0.9974	0.9971
14	0.9969	0.9956	0.9970	0.9969	0.9970	0.9970	0.9968	0.9973	0.9973	0.9972	0.9972	0.9974	0.9971	0.9971	0.9968
15	0.9990	0.9981	0.9990	0.9990	0.9990	0.9990	0.9989	0.9992	0.9992	0.9991	0.9991	0.9992	0.9991	0.9991	0.9989

Table 4.4: Similarity Score for Hair : HSI Superpixels

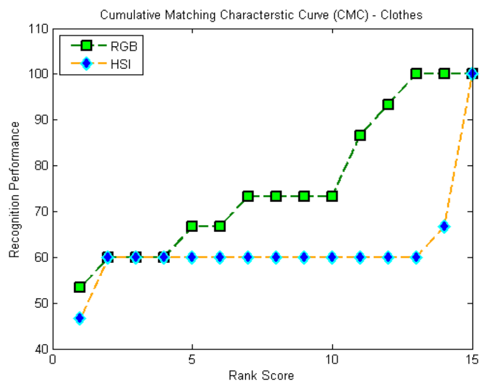
Person AM\PM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.9954	0.9986	0.9891	0.9906	0.9868	0.9891	0.9925	0.9877	0.9863	0.9870	0.9870	0.9839	0.9904	0.9919	0.9895
2	0.9945	0.9982	0.9879	0.9894	0.9853	0.9877	0.9914	0.9863	0.9849	0.9858	0.9857	0.9823	0.9893	0.9908	0.9891
3	0.9938	0.9939	0.9842	0.9868	0.9832	0.9862	0.9908	0.9840	0.9819	0.9806	0.9820	0.9806	0.9853	0.9885	0.9774
4	0.9940	0.9942	0.9845	0.9870	0.9835	0.9864	0.9910	0.9843	0.9822	0.9809	0.9823	0.9808	0.9856	0.9888	0.9780
5	0.9940	0.9942	0.9845	0.9871	0.9836	0.9865	0.9911	0.9843	0.9823	0.9810	0.9824	0.9809	0.9856	0.9888	0.9779
6	0.9940	0.9942	0.9846	0.9871	0.9836	0.9866	0.9911	0.9844	0.9823	0.9811	0.9824	0.9810	0.9857	0.9889	0.9780
7	0.9940	0.9941	0.9844	0.9870	0.9835	0.9864	0.9960	0.9843	0.9822	0.9809	0.9823	0.9808	0.9855	0.9888	0.9778
8	0.9940	0.9942	0.9845	0.9871	0.9836	0.9865	0.9911	0.9843	0.9822	0.9810	0.9823	0.9809	0.9856	0.9888	0.9779
9	0.9961	0.9963	0.9882	0.9903	0.9872	0.9897	0.9936	0.9879	0.9861	0.9851	0.9862	0.9847	0.9892	0.9919	0.9823
10	0.9964	0.9965	0.9886	0.9906	0.9876	0.9901	0.9939	0.9883	0.9865	0.9855	0.9866	0.9852	0.9895	0.9922	0.9826
11	0.9962	0.9963	0.9882	0.9903	0.9872	0.9897	0.9936	0.9879	0.9861	0.9851	0.9863	0.9847	0.9892	0.9919	0.9823
12	0.9962	0.9960	0.9882	0.9903	0.9873	0.9898	0.9937	0.9880	0.9862	0.9850	0.9862	0.9849	0.9891	0.9919	0.9816
13	0.9962	0.9960	0.9882	0.9904	0.9873	0.9898	0.9937	0.9880	0.9862	0.9850	0.9863	0.9849	0.9891	0.9919	0.9816
14	0.9963	0.9962	0.9883	0.9905	0.9874	0.9899	0.9938	0.9881	0.9863	0.9852	0.9864	0.9850	0.9893	0.9992	0.9820
15	0.9950	0.9977	0.9875	0.9892	0.9853	0.9879	0.9918	0.9863	0.9847	0.9850	0.9852	0.9824	0.9888	0.9907	0.9866

Table 4.5: Similarity Score for Skin : RGB Superpixels

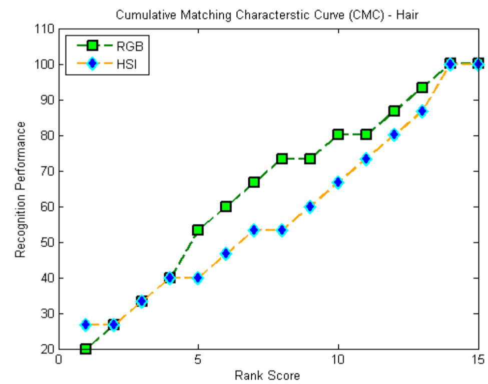
Person AM\PM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.9941	0.9950	0.9930	0.9910	0.9924	0.9938	0.9873	0.9903	0.9928	0.9906	0.9906	0.9925	0.9916	0.9892	0.9923
2	0.9950	0.9974	0.9952	0.9939	0.9949	0.9965	0.9901	0.9926	0.9950	0.9931	0.9930	0.9950	0.9941	0.9919	0.9948
3	0.9989	0.9991	0.9996	0.9976	0.9983	0.9987	0.9956	0.9973	0.9985	0.9975	0.9975	0.9984	0.9980	0.9967	0.9982
4	0.9989	0.9999	0.9994	0.9989	0.9993	0.9998	0.9971	0.9983	0.9993	0.9986	0.9985	0.9994	0.999	0.9980	0.9993
5	0.9993	0.9997	0.9995	0.9988	0.9998	0.9996	0.9973	0.9985	0.9994	0.9987	0.9987	0.9994	0.9991	0.9981	0.9993
6	0.9994	0.9991	0.9995	0.9993	0.9997	0.9998	0.9980	0.9990	0.9997	0.9992	0.9991	0.9997	0.9995	0.9987	0.9996
7	0.9992	0.9999	0.9997	0.9992	0.9996	0.9998	0.9977	0.9987	0.9996	0.9990	0.9990	0.9996	0.9993	0.9985	0.9996
8	0.9989	0.9990	0.9985	0.9974	0.9982	0.9986	0.9954	0.9972	0.9984	0.9973	0.9973	0.9982	0.9978	0.9965	0.9981
9	0.9992	0.9995	0.9992	0.9983	0.9989	0.9992	0.9966	0.9980	0.9998	0.9982	0.9982	0.999	0.9986	0.9976	0.9989
10	0.9999	0.9992	0.9998	0.9992	0.9996	0.9993	0.9985	0.9994	0.9998	0.9994	0.9994	0.9996	0.9996	0.9991	0.9996
11	0.9990	0.9986	0.9997	0.9994	0.9996	0.9989	0.9993	0.9918	0.9997	0.9997	0.9998	0.9996	0.9997	0.9996	0.9996
12	0.9997	0.9996	0.9998	0.9993	0.9997	0.9996	0.9982	0.9992	0.9998	0.9993	0.9993	0.9997	0.9995	0.9989	0.9996
13	0.999	0.9978	0.9993	0.9995	0.9994	0.9984	0.9999	0.9999	0.9995	0.9998	0.9999	0.9993	0.9997	0.9999	0.9994
14	0.9994	0.9979	0.9994	0.9994	0.9994	0.9984	0.9998	1.0000	0.9995	0.9998	0.9999	0.9994	0.9997	0.9998	0.9994
15	0.9993	0.9994	0.9990	0.9999	0.9994	0.9997	0.9993	0.9997	0.9993	0.9999	0.9999	0.9995	0.9998	0.9997	1.0000

Table 4.6: Similarity Score for Skin : HSI Superpixels

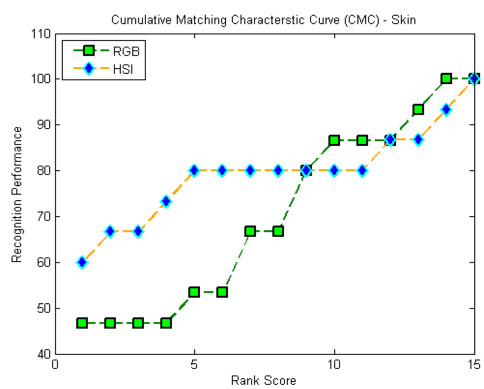
Person AM\PM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.9954	0.9993	0.9995	0.9991	0.9994	0.9991	0.9965	0.9965	0.9990	0.9981	0.9994	0.9999	0.9991	0.9993	0.9975
2	0.9888	0.9993	0.9965	0.9956	0.9961	0.9954	0.9903	0.9902	0.9950	0.9931	0.9958	0.9978	0.9952	0.9975	0.9921
3	0.9918	0.9990	0.9997	0.9968	0.9975	0.9967	0.9927	0.9927	0.9965	0.9951	0.9975	0.9987	0.9969	0.9988	0.9941
4	0.9951	0.9991	0.9994	0.9993	0.9993	0.9992	0.9966	0.9965	0.9990	0.9982	0.9992	0.9998	0.9990	0.9991	0.9977
5	0.9952	0.9994	0.9995	0.9993	0.9995	0.9993	0.9966	0.9965	0.9990	0.9982	0.9993	0.9999	0.9991	0.9991	0.9977
6	0.9970	0.9980	0.9996	0.9991	0.9998	0.9999	0.9986	0.9985	0.9991	0.9995	0.9996	0.9996	0.9996	0.9982	0.9994
7	0.9967	0.9983	0.9996	0.9999	0.9997	0.9999	0.9983	0.9982	0.9998	0.9994	0.9995	0.9997	0.9996	0.9983	0.9992
8	0.9955	0.9985	0.9994	0.9994	0.9994	0.9994	0.9972	0.9998	0.9995	0.9986	0.9992	0.9997	0.9991	0.9981	0.9986
9	0.9961	0.9984	0.9995	0.9998	0.9996	0.9997	0.9978	0.9977	0.9999	0.9991	0.9994	0.9997	0.9993	0.9981	0.9990
10	0.9982	0.9962	0.9992	0.9996	0.9993	0.9996	0.9995	0.9995	0.9998	0.9998	0.9992	0.9985	0.9995	0.9970	0.9999
11	0.9973	0.9984	0.9997	0.9993	0.9996	0.9993	0.9978	0.9978	0.9991	0.9989	0.9998	0.9995	0.9997	0.9997	0.9980
12	0.9964	0.9988	0.9997	0.9998	0.9997	0.9997	0.9977	0.9977	0.9996	0.999	0.9996	0.9999	0.9995	0.9988	0.9987
13	0.9978	0.9981	0.9992	0.9998	0.9991	0.9998	0.9987	0.9987	0.9997	0.9996	0.9990	0.9995	0.9994	0.9989	0.9991
14	0.9973	0.9967	0.9985	0.9976	0.9983	0.9976	0.9968	0.9968	0.9973	0.9978	0.9988	0.9978	0.9987	0.9996	0.9961
15	0.9966	0.9910	0.9958	0.9971	0.9963	0.9973	0.9986	0.9985	0.9976	0.998	0.9961	0.9947	0.9966	0.9919	0.9990



(a)



(b)



(c)

Figure 4.4: CMC curves

includes varying illumination conditions, pose variability of persons and camera viewpoint differences, aggravating the drift in location of interest points when a local descriptor is utilized. This would result in partial uniformity of captures of the same person and makes re-identification more challenging. However, incorporating superpixel segmentation to generate visual words as regions of interest for local features can potentially alleviate the influence of shift in interest points by utilizing the contextual information via superpixels. Different images of the same person would have many similar visual patches, and hence a superpixel based vocabulary tree would be a reasonable cue for matching individuals, that can be combined with matching of spectral content. Identification performance can also be further improved by exploiting relative positions (spatial neighborhood information from superpixels) of interest points.

4.4 Future work and Conclusion

In this chapter, we performed a limited study on superpixel based matching and person re-identification for natural hyperspectral images. To the best of our knowledge, this is the first time ground based hyperspectral images are used for such a validation. We start with an initial oversegmentation of the image resulting in superpixels. Re-identification is performed by comparing spectral angle distances of the test and reference subjects. We created superpixel labels via manual annotation, and validated the efficacy of superpixels for person re-identification via CMC curves. On testing the efficacy of HSI superpixels for the re-id problem, it can be observed that superpixels are a convenient way for spectral matching and identifying non-rigid shapes in the image — although our experimental setup in this pilot study was simplistic, we suggest a practical implementation wherein hyperspectral superpixels help characterize a vocabulary tree, that when utilized in conjunction with spectral matching (e.g. via the spectral angle distance), provides enhanced re-id performance. Also, it was observed that HSI superpixels have discriminative ability and thus, are expected to be effective in person re-identification task. Spectral angle distance between every superpixel, in an initial oversegmentation of the image, and the reference signatures, has the potential to rapidly localize and match objects in an image. We would like to note another potential benefit of such a framework in terms of simple spectral matching — shot-wave infrared imagery is known to have low Signal to Noise (SNR), a property that will likely be exaggerated in ground based imaging of natural scenes with complex illumination conditions. Spectral averaging over superpixels would provide a natural smoothing of the additive noise prior to spectral matching.

The results and conclusions of this work unavoidably come together to open questions and new ideas. We enumerate some of the research avenues for future work. We explored the idea that information from the entire spectra, captured by a hyperspectral sensor is useful in visual recognition. It was observed that HSI superpixels have discriminative ability and can potentially be beneficial in person re-identification tasks. Hence, for future studies, it would be optimal to formulate re-identification as an image search problem by incorporating a vocabulary tree approach, where local features from HSI superpixels can be regarded as visual words — additionally, information from a vocabulary tree can be "fused" with spectral reflectance (providing information about material properties) features for enhanced re-id performance. Illumination variation leads to unreliable representation of same individual, specifically, when using RGB images in low illumination conditions, color and texture information differs slightly between different persons and hence shows no discriminatory power. This suggests that illumination factors affect recognition performance. A spectral angle based distance metric has been shown to possess illumination invariance and would be suitable when setting up the re-identification task at the superpixel level. Hence, for future work, an improved spectral angle based similarity metric can be investigated.

Chapter 5
Conclusion

A tenet of object classification is that accuracy improves with an increasing number (and variety) of spectral channels available to the classifier. The input for many classification tasks are images taken using conventional cameras containing three broadband spectral measurements (the red, green, and blue channels of the image). RGB cameras are bountiful, cheap, and easy to use; however, the coarse sampling of the visible spectrum limits classification accuracy, especially in the presence of metameric scene elements. Hyperspectral imaging (HSI) systems, on the other hand, record hundreds of measurements and provide fine spectral resolution over a wide range of the electromagnetic spectrum that greatly improves classification performance. The improved performance comes at a cost. HSI camera systems require specialized processing units, are expensive and bulky, have long acquisition times.

In this thesis we presented a superpixel based analysis for hyperspectral data. In Chapter 2 we discussed two graph based segmentation approaches — ER and FH, for computing superpixels. Although, these methods are widely used for three-channel images, ER has never been utilized for hyperspectral imagery, FH has recently been studied for pixel unmixing, but not for classification tasks. We implemented a straightforward extension of these methods to multi-channel images. Segmentations obtained from these methods on hyperspectral data were quantified using standard metrics. In Chapter 3 we presented a new approach to exploiting spatial context via superpixels for hyperspectral remote sensing image analysis. A multi-classifier decision fusion approach is employed within every superpixel, to derive a robust classification decision at the superpixel level. From the experimental results shown in Chapter 3, we conclude that incorporating spatial context via superpixels using the proposed approach results in robust classification, even with very little training data. We compared our system with that of three spectral channels only and experiments indicate that hyperspectral data provides an improved recognition performance. In Chapter 4 we performed a pilot study on utilizing superpixels for the person re-identification problem. On testing the efficacy of HSI superpixels for re-id problem, it was observed that HSI superpixels have discriminative ability and are a convenient way for spectral matching and identifying non-rigid shapes in the image, and thus, are expected to be effective in person re-identification task — we acknowledge the limitations of this pilot study for the re-identification problem. The promising results for this problem are indicative of the potential of hyperspectral superpixels with natural scenes — in a rigourous person re-identification framework, they can be utilized to efficiently characterize a vocabulary tree based model. Additionally, by spatially averaging over superpixels, one can expect noise robustness — this can be particularly beneficial to low SNR hyperspectral data.

References

[1] D. Lu and Q. Weng, "A Survey of Image Classification Methods and Techniques for Improving Classification Performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.

[2] B. Khoobehi, J. M. Beach, and H. Kawano, "Hyperspectral Imaging for Measurement of Oxygen Saturation in the Optic Nerve Head," *Investigative ophthalmology & visual science*, vol. 45, no. 5, pp. 1464–1472, 2004.

[3] C. P. Huynh and A. Robles-Kelly, "Hyperspectral Imaging for Skin Recognition and Biometrics," in *IEEE International Conference on Image Processing (ICIP)*, pp. 2325–2328, IEEE, 2010.

[4] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral Image Processing for Automatic Target Detection Applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.

[5] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. John Wiley & Sons, 2007.

[6] J. Shotton, M. Johnson, and R. Cipolla, "Semantic Texton Forests for Image Categorizationand Segmentation," in *Proceedings of IEEE Conference on Computer vision and pattern recognition (CVPR)*, pp. 1–8, 2008.

[7] J. Aghajanian, J. Warrell, S. J. Prince, P. Li, J. L. Rohn, and B. Baum, "Patch-based within Object Classification," in *IEEE 12th International Conference on Computer Vision (ICCV)*, pp. 1125–1132, IEEE, 2009.

[8] X. Ren and J. Malik, "Learning a Classification Model for Segmentation," *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pp. 10–17, 2003.

[9] A. P. Moore, S. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel Lattices," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.

[10] G. Mori, X. Ren, A. A. Efros, and J. Malik, "Recovering Human Body Configurations: Combining Segmentation and Recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. II–326, 2004.

[11] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic Superpixels Compared to State-of-the-art Superpixel Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 34, no. 11, pp. 2274–2282, 2012.

[12] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class Segmentation and Object Localization with Superpixel Neighborhoods," in *IEEE 12th International Conference on Computer Vision (ICCV)*, pp. 670–677, IEEE, 2009.

[13] S. Wang, H. Lu, F. Yang, and M.-H. Yang, "Superpixel Tracking," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 1323–1330, IEEE, 2011.

[14] D. R. Thompson, L. Mandrake, M. S. Gilmore, and R. Castaño, "Superpixel Endmember Detection," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 48, no. 11, pp. 4023–4033, 2010.

[15] D. R. Thompson, R. Castano, and M. S. Gilmore, "Sparse Superpixel Unmixing for Exploratory Analysis of CRISM Hyperspectral Images," in *IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, 2009.

[16] G. Shu, A. Dehghan, and M. Shah, "Improving an Object Detector and Extracting Regions using Superpixels," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3721–3727, IEEE, 2013.

[17] B. Micusik and J. Kosecka, "Semantic Segmentation of Street Scenes by Superpixel Co-occurrence and 3d Geometry," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pp. 625–632, IEEE, 2009.

[18] H. Liu, Y. Qu, Y. Wu, and H. Wang, "Class-specified Segmentation with Multi-scale Superpixels," in *Computer Vision-ACCV 2012 Workshops*, pp. 158–169, Springer, 2013.

[19] J. Kaufhold, R. Collins, A. Hoogs, and P. Rondot, "Recognition and Segmentation of Scene Content using Region-based Classification," in *International Conference on Pattern Recognition (ICPR)*, vol. 1, pp. 755–760, IEEE, 2006.

[20] C. Liu and Z. Zhao, "Person Re-identification by Local Feature based on Super Pixel," in *Advances in Multimedia Modeling*, pp. 196–205, Springer, 2013.

[21] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[22] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 22, no. 8, pp. 888–905, 2000.

[23] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and Supervoxels in an Energy Optimization Framework," in *Computer Vision–ECCV 2010*, pp. 211–224, Springer, 2010.

[24] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy Rate Superpixel Segmentation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2097–2104, 2011.

[25] M. D. Atkinson, J.-R. Sack, N. Santoro, and T. Strothotte, "Min-max Heaps and Generalized Priority Queues," *Communications of the ACM*, vol. 29, no. 10, pp. 996–1000, 1986.

[26] A. Schick, M. Fischer, and R. Stiefelhagen, "Measuring and Evaluating the Compactness of Superpixels," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 930–934, Nov 2012.

[27] A. Hanbury and J. Stottinger, "On Segmentation Evaluation Metrics and Region Counts," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4, Dec 2008.

[28] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast Superpixels using Geometric Flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 31, no. 12, pp. 2290–2297, 2009.

[29] M. Mahdianpari, M. Motagh, and V. Akbari, "Speckle Reduction and Restoration of Synthetic Aperture Radar Data with an Adoptive Markov Random Field Model," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 276–279, 2012.

[30] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial Hyperspectral Image Segmentation using Subspace Multinomial Logistic Regression and Markov Random Fields," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 50, no. 3, pp. 809–823, 2012.

[31] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple Spectral–Spatial Classification Approach for Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 48, no. 11, pp. 4122–4132, 2010.

[32] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances

in Spectral-Spatial Classification of Hyperspectral Images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.

[33] T. Priya, S. Prasad, and H. Wu, "Superpixels for Spatially Reinforced Bayesian Classification of Hyperspectral Images - under revision," *IEEE Geoscience and Remote Sensing Letters (GRSL)*, 2014.

[34] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality Preserving Dimensionality Reduction and Classification for Hyperspectral Image Analysis," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 50, pp. 1185–1198, Apr. 2012.

[35] M. Sugiyama, "Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis," *The Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, May 2007.

[36] S. Prasad and L. M. Bruce, "Decision Fusion with Confidence-based Weight Assignment for Hyperspectral Target Recognition," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 46, no. 5, pp. 1448–1456, 2008.

[37] G. Schwarz, "Estimating the Dimension of a Model," *The annals of statistics*, vol. 6, pp. 461–464, Mar. 1978.

[38] C. E. Rasmussen, "The Infinite Gaussian Mixture Model.," in *Neural Information Processing Systems (NIPS)*, vol. 12, pp. 554–560, 1999.

[39] H. Wu and S. Prasad, "Infinite Gaussian Mixture Models for Robust Decision Fusion of Hyperspectral Imagery and Full Waveform LiDAR Data," in *Proceedings of IEEE*

Global Conference on Signal and Information Processing (GlobalSIP), pp. 1025–1028, Dec 2013.

[40] J. A. Benediktsson and J. R. Sveinsson, “Multisource Remote Sensing Data Classification based on Consensus and Pruning,” *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 41, no. 4, pp. 932–936, 2003.

[41] N. Gheissari, T. B. Sebastian, and R. Hartley, “Person Reidentification using Spatiotemporal Appearance,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 1528–1535, IEEE, 2006.

[42] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person Re-identification by Symmetry-driven Accumulation of Local Features,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2360–2367, IEEE, 2010.

[43] W.-S. Zheng, S. Gong, and T. Xiang, “Person Re-identification by Probabilistic Relative Distance Comparison,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 649–656, IEEE, 2011.

[44] A. Dantcheva and J.-L. Dugelay, “Frontal-to-side Face Re-identification based on Hair, Skin and Clothes Patches,” in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, pp. 309–313, IEEE, 2011.

[45] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,”

in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893, IEEE, 2005.

[46] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. I–511, IEEE, 2001.

[47] N. Keshava, "Distance Metrics and Band Selection in Hyperspectral Processing with Applications to Material Identification and Spectral Libraries," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 42, no. 7, pp. 1552–1565, 2004.

[48] S. M. Adler-Golden, R. Y. Levine, M. W. Matthew, S. C. Richtsmeier, L. S. Bernstein, J. H. Gruninger, G. W. Felde, M. L. Hoke, G. P. Anderson, *et al.*, "Shadow-insensitive Material Detection/Classification with Atmospherically Corrected Hyperspectral Imagery," in *Aerospace/Defense Sensing, Simulation, and Controls*, pp. 460–469, International Society for Optics and Photonics, 2001.

[49] S. Robila, "An Investigation of Spectral Metrics in Hyperspectral Image Preprocessing for Classification," in *Geospatial goes global: From Your Neighborhood to the Whole Planet. ASPRS Annual Conference, Baltimore, Maryland*, pp. 7–11, 2005.

[50] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Interactive Image Segmentation by Maximal Similarity based Region Merging," *Pattern Recognition*, vol. 43, no. 2, pp. 445–456, 2010.