Chapter 5

Semantic-aware image filtering: Applications to classification of hyperspectral images and semantic segmentation of natural images

5.1 Introduction

Although the filtered images generated by the filtering techniques provided spatial information, they were seldom exploited for image classification. In this thesis, we proposed several semantic-aware filtering techniques to protect the significant structural elements in the image while removing noises or smoothing textural details. As a result, the filtered images generated by these techniques provided useful spatial information of the input image. In this chapter, we explore spatial information of hyperspectral images (HSIs) for spectral-spatial classification using the filtered images generated by our developed filtering techniques.

The recent development of hyperspectral imaging has enabled numerous applications across various sectors, including agriculture, mining, forestry, and ur-

ban area monitoring. Hyperspectral images (HSIs) contain spectral data spanning hundreds of tiny wavelengths. The curse of dimensionality may result from such high-dimensional data's potential for redundant information [12, 51, 102]. But when carefully investigated using the right methods, the rich spectral information is discovered to be significantly aiding in the precise classification of different land cover types.

It is already established that along with spectral knowledge, incorporating the spatial information of the pixels greatly enhances the classification results of HSIs. The HSI literature includes numerous spectral-spatial classification techniques. One of the fundamental goal of all such techniques is to build a robust model that can capable of considering better spatial information of the pixels. Markov random fields (MRFs) is a tool serves as the foundation for a variety of spectral-spatial HSI classification techniques. This offers a versatile framework for considering spatial information of the pixels on the image that has been widely used with HSI data [57]. Spectral-spatial classification techniques based on sparse representation (SR) constitute another important category. In [52], a shape adaptive fixed region based SR model is proposed to consider spatial information of the pixels.In [44], a multiscale adaptive sparse representation (MASR) model is proposed to incorporate improved spatial information. In [66], unmixing and SR are combined together to extract appropriate spatial information associated to the pixels on the HSIs. In [46], a multiple-feature-based adaptive sparse representation (MFASR) is introduced. In [45], superpixels are employed to incorporate spatial information. Mathematical morphology (MM) is another popular tool widely used for considering spatial information of the image [14]. A large number of morphological operations are available in the MM to include spatial information which are often designed with fixed shaped structuring element (SE) [123]. An extended morphological profile (EMP) for the spectral-spatial classification of (HSIs) is established in [105]. Through variations in the shape and size of the structuring element (SE), it generates multiple filtered images from the HSI. Then, EMP is

constructed by concatenating these filtered images along with the HSI. Attribute filter (AP) is another MM tool that capable of considering geometrical shape of the objects. Instead of EMP, subsequently, an extended attribute profile (EAP) [36] is developed by using attribute profiles (AP) with different thresholds for the HSIs classification. In [16] threshold free EAP is suggested for considering spatial information. In recent times, several models based on deep neural networks have been introduced for HSI classification [41, 57, 148]. Also, sparse representation, MM, and AP based feature sets are now combined in deep ANN and CNN models for HSI classification [31, 78, 144].

Structure preserving image filtering is a well known approach used to prevent integrity of the structures of the objects of images while removing the noises or smoothing the textures. The effectiveness of such filtering techniques is assessed by how efficiently they are capable of removing noise or smoothing the textures with the minimum distortion of the structures of objects present on the images. The recent research focuses on the development of semantic-aware filtering techniques that incorporate the semantic information of the images to better differentiate structures and textures. Although the filtered images of such advanced techniques provided better spatial information, they are seldom exploited for image classification. In this chapter, to show the impact of our developed filtering approachs for incorporating spatial information, an extended semantic filtered profile (ESFP) is created by combining the filtered images generated by our technique. Then the constructed profile ESEP is utilized for the spectral-spatial classification of HSI. To assess its effectiveness, we compared our proposed profile against several recent state-of-the-art techniques using three real hyperspectral datasets.

In addition to this, the proposed filtering techniques are also tested by applying them to the semantic segmentation of natural images. Semantic segmentation, a fundamental task in computer vision, assigns class labels to every pixel in an image, enabling critical applications in domains such as autonomous driving,

medical imaging, and satellite analysis [6, 11, 26, 68, 73, 77, 100, 151]. Achieving precise boundary delineation and minimizing noise in segmentation maps remain key challenges in this field. Image filtering techniques, particularly edge-aware filters like bilateral and guided filters, have proven essential in overcoming these obstacles. These filters are widely employed in preprocessing to smooth noisy inputs while preserving crucial edges and in postprocessing to refine segmentation outputs, ensuring that predicted boundaries align closely with actual object edges. Moreover, advanced approaches such as Conditional Random Fields (CRFs) and edge-preserving neural architectures have incorporated filtering principles to further enhance segmentation accuracy. By addressing noise suppression, boundary precision, and feature enhancement, image filtering significantly improves the robustness and quality of semantic segmentation in complex real-world scenarios. We compared our proposed techniques with three other different filtering techniques by exploiting two different recently developed semantic segmentation techniques: one is an unsupervised technique called MeanShift++ [68], and the other is a supervised technique named BiSeNet V2 [151]. As the test images, we have taken two simple RGB images that are already used in the previous chapters to show the filtering results, and the other two test images are taken from a well known data set, BSD500 [100] using which the BiSeNet V2 model is also trained.

The next part of the chapter is outlined as follows: Section 5.2 provides details on the construction of the extended semantic filtered profile (ESFP) for each of the filtering techniques presented in this thesis. Section 5.3 describes the HSI datasets used for the experiments. Section 5.4 presents the experimental results of it. Section 5.5 presents an application in semantic image segmentation, while Section 5.6 concludes the chapter.

5.2 Construction of Extended Semantic Filtered Profile (ESFP)

Chapters 2, 3, and 4 of this thesis present four different semantic-aware image filtering techniques. The filtered images generated by these techniques may provide better spatial information of the input image as they considered semantic information. To take into account spatial information for spectral-spatial classification of HSI, first, several filtered images of the considered HSI are generated by applying one of our proposed filtering technique. Then, an extended semantic filtered profile is constructed by concatenating these filtered images. Finally, the spectra-spatial features of the pixels on the profile are used for classification of HSI.

The structure preserving filtering techniques proposed in this thesis excel at filtering out insignificant details and noises while minimizing distortion of the underlying structures in the image. Consequently, the filtered images produced by these techniques offer improved spatial information. Note that to incorporate appropriate semantic information all the four filtering techniques proposed in this thesis used window. Since the optimal value of the window is unknown, to consider a maximum amount of spatial information, one can generate multiple filter images by varying the size of the window. However, when dealing with HSI with hundreds of spectral bands, generating multiple filtered images for each band is not feasible. To circumvent this problem, a reduced set of principal components (PCs) extracted from the HSI, which preserves the most crucial information, is considered. For each of the considered PC, multiple filtered images are produced by varying the size of window. Finally, all the filtered images derived from the considered PCs are combined together to construct an extended semantic filtering profile (ESFP) for the HSI classification.

In more detail, let I represent an HSI with dimensions $M \times N \times P$. To reduce the dimension of I with minimal information loss, Principal Component

Analysis (PCA) is applied to I. Subsequently, the first m principal components (PCs) are chosen to represent the HSI. Depending on the semantic-aware filtering technique applied to these PCs, the extended semantic filtering profiles ESFP I, ESFP II, ESFP III, and ESFP IV are generated as follows:

ESFP I: This profile is constructed by applying the semantic-aware filtering technique proposed in Chapter 2. In this technique a window w is used to consider semantic information from the morphological gradient image. Since the best value of w is unknown, to account for a maximum amount of spatial information, we generate multiple filtered images by varying the size of w. Let I^{pc_i} be the ith PC of HSI I. The proposed filtering technique is applied to I^{pc_i} by considering increasing size window $w = w_1, w_2, \ldots, w_n$. As a result n filtered images denoted as $I^{pc_i}_{w_1}, I^{pc_i}_{w_2}, \ldots, I^{pc_i}_{w_n}$ are generated. The steps is repeated for other PCs. Finally all the generated filtered images are concatenated together to form an ESFP I as follows:

$$I_{ESFP\ I} = [\{I^{pc_1}, I^{pc_1}_{w_1}, \ldots, I^{pc_1}_{w_n}\}, \{I^{pc_2}, I^{pc_2}_{w_1}, \ldots, I^{pc_2}_{w_n}\}, \ldots, \{I^{pc_m}, I^{pc_m}_{w_1}, \ldots, I^{pc_m}_{w_n}\}]$$

(5.1)

ESFP II: This profile is constructed by applying Approach I of the semantic-aware filtering technique proposed in Chapter 3. In this technique a window w is used to consider semantic information by exploiting JS divergence. Since the best value of w is unknown, to account for a maximum amount of spatial information, we generate multiple filtered images by varying the size of w. Let I^{pc_i} be the ith PC of HSI I. The proposed filtering technique is applied to I^{pc_i} by considering increasing size window $w = w_1, w_2, \ldots, w_n$. As a result n filtered images denoted as $I^{pc_i}_{w_1}, I^{pc_i}_{w_2}, \ldots, I^{pc_i}_{w_n}$ are generated. The steps is repeated for other PCs. Finally all the generated filtered images are concatenated together to form an ESFP II as follows:

5.2. Construction of Extended Semantic Filtered Profile (ESFP)

$$I_{ESFP\ II} = [\{I^{pc_1}, I^{pc_1}_{w_1}, \ldots, I^{pc_1}_{w_n}\}, \{I^{pc_2}, I^{pc_2}_{w_1}, \ldots, I^{pc_2}_{w_n}\}, \ldots, \{I^{pc_m}, I^{pc_m}_{w_1}, \ldots, I^{pc_m}_{w_n}\}]$$

$$(5.2)$$

ESFP III: This profile is constructed by considering Approach II of the semantic-aware filtering developed in Chapter 3. In this technique four windows W_1, W_2, W_3 and W_4 are used to consider semantic information. The proposed filtering technique is applied to I^{pc_i} by considering the four windows. As a result 4 filtered images denoted as $I^{pc_i}_{W_1}, I^{pc_i}_{W_2}, I^{pc_i}_{W_3}$, and $I^{pc_i}_{W_4}$ are generated. The steps is repeated for other PCs. Finally all the generated filtered images are concatenated together to form an ESFP III as follows:

$$I_{ESFP\ III} = [\{I^{pc_1}, I^{pc_1}_{W_1}, I^{pc_1}_{W_2}, I^{pc_1}_{W_3}, I^{pc_1}_{W_4}\}, \ldots, \{I^{pc_m}, I^{pc_m}_{W_1}, I^{pc_m}_{W_2}, I^{pc_m}_{W_3}, I^{pc_m}_{W_4}\}] \quad (5.3)$$

ESFP IV: This profile is constructed by considering the semantic-aware filtering developed in Chapter 4. In this technique four options F_1, F_2, F_3 and F_4 are used to consider semantic information. The proposed filtering technique is applied to I^{pc_i} by considering the four options. As a result 4 filtered images denoted as $I^{pc_i}_{F_1}, I^{pc_i}_{F_2}, I^{pc_i}_{F_3}$, and $I^{pc_i}_{F_4}$ are generated. The steps is repeated for other PCs. Finally all the generated filtered images are concatenated together to form an ESFP IV as follows:

$$I_{ESFP\ IV} = [\{I^{pc_1}, I^{pc_1}_{F_1}, I^{pc_1}_{F_2}, I^{pc_1}_{F_3}, I^{pc_1}_{W_4}\}, \ldots, \{I^{pc_m}, I^{pc_m}_{F_1}, I^{pc_m}_{F_2}, I^{pc_m}_{F_3}, I^{pc_m}_{F_4}\}] \quad (5.4)$$

The profiles ESFP I, ESFP II, ESFP III or ESFP IV constructed from HSI I contains rich spatial information. Figure 5-1 show the first PC of the

Pavia University data set and the five filtered images generated by the filtering techniques presented in Chapter 2 and the Approach I presented in Chapter 3 for the window of sizes $7 \times 7, 11 \times 11, 17 \times 17, 27 \times 27$, and 43×43. Figure 5-2 show four filtered images generated from the first PC of Pavia University data set by applying the Approach II presented in Chapter 3 and the filtering techniques presented in Chapter 4. These figures, shows that the filtered images generated by the two of our developed technique well preserve the significant objects structures during the smoothing of textures and removing of noises. After constructing the profiles ESFP I, ESFP II, ESFP III, or ESFP IV, the derived spectral-spatial features of the pixels from these profiles are provided as input to a classifier for HSI classification.. The steps for the construction of profiles are summarized in Algorithm 5.
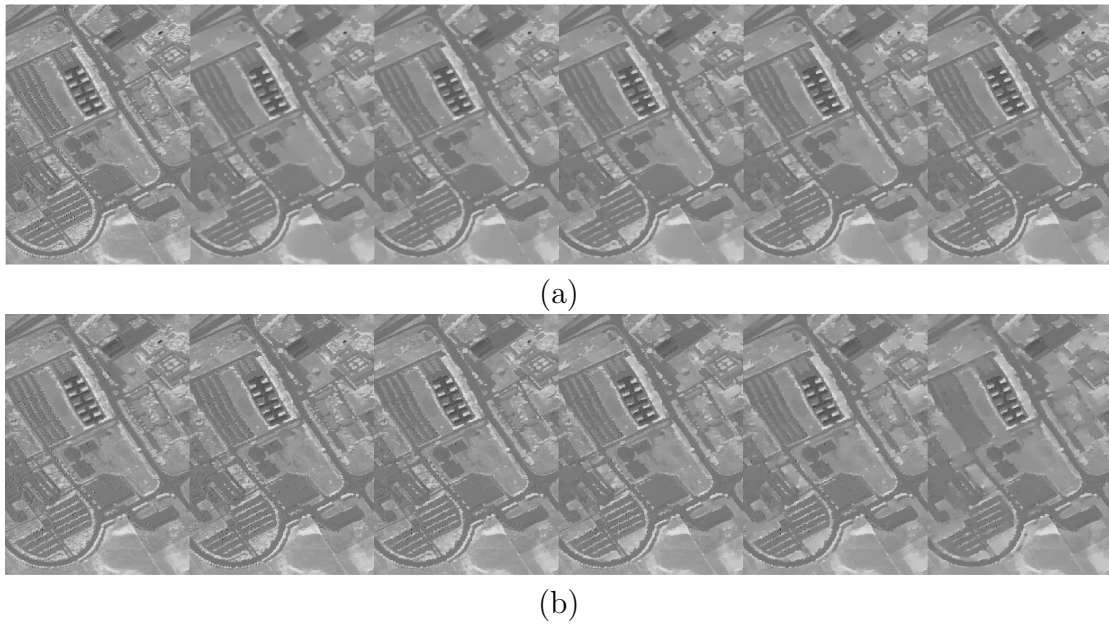

(a)

(b)

Figure 5-1: First PC of the Pavia University data set and the five filtered images generated from it by applying (a) the filtering technique presented in Chapter 2 and (b) the Approach I presented in Chapter 3, considering the size of windows 7×7, 11×11, 17×17, 27×27, and 43×43.
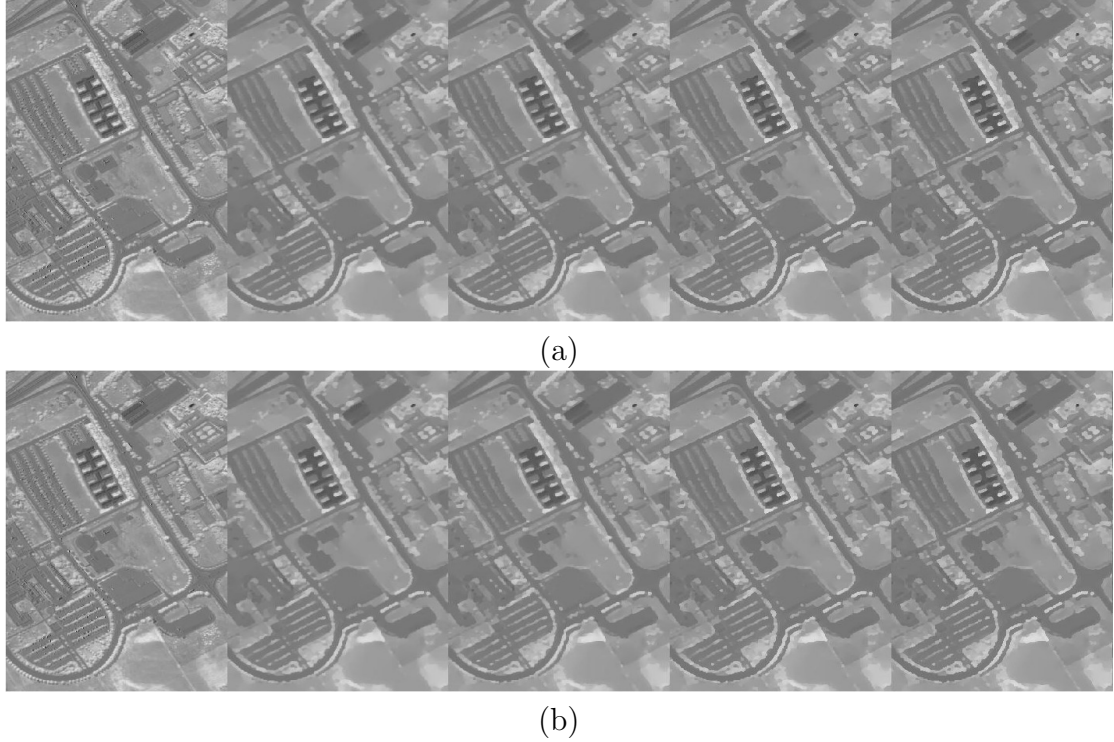
(a)

(b)

Figure 5-2: First PC of the Pavia University data set and the four filtered images generated from it by applying (a) the Approach II presented in Chapter 3 and (b) the filtering technique presented in Chapter 4 by using four discrete options.

Algorithm 5 Algorithm to generate the proposed profiles $I_{ESFP\ Ior\ IIor\ IIIor\ IV}$

INPUT- First m PCs $(I^{pc_1}, I^{pc_2}, \ldots, I^{pc_m})$ of the HSI I and n windows $(w_1, w_2, \ldots, w_n)$ of increasing size or n parameter options

OUTPUT- Extended semantic filtering profile: $I_{ESFP\ Ior\ IIor\ IIIor\ IV}$

$I_{ESFP\ Ior\ IIor\ IIIor\ IV} = \Phi,$

```
For i=1 to m do
For j=1 to n do
```

Generate the filtered image $I_{w_j}^{pc_i}$ or $I_{W_j}^{pc_i}$ or $I_{F_j}^{pc_i}$ by applying presented filtering respectively

$I_{ESFP\ Ior\ IIor\ IIIor\ IV} = \bigsqcup I_{w_j}^{pc_i}$ or $I_{W_j}^{pc_i}$ or $I_{F_j}^{pc_i}$

```
End For
End For
```

/* $\bigsqcup$ is the vector concatenation operator */

5.3 Data sets description

Three widely recognized hyperspectral data sets: the Indian Pines data set, the University of Pavia data set, and the University of Houston data set are used to determine the efficacy of our proposed technique. Here are brief descriptions of each of these datasets:

5.3.1 Indian Pines data set

The benchmark data set utilized in this study comprises hyperspectral imagery acquired over the rural landscape in Indian Pines, in northwest Indiana, USA. The data collection employed the Infrared Imaging Spectrometer/Airborne Visible (AVIRIS) sensor, equipped with 220 channels covering a spectral range from 0.2 to 2.4 micrometers (µm). Through preprocessing, the dataset underwent a reduction to 200 spectral bands where 20 water absorption bands are excluded within the ranges of [104-108], [150-163], and band 220. The resultant image has a 20 meters spatial resolution of dimensions of 145 by 145 pixels. The classes, training samples, test samples, ground truth map all the details required for experimental validation are shown in Figure 5-3 and Table 5.1.
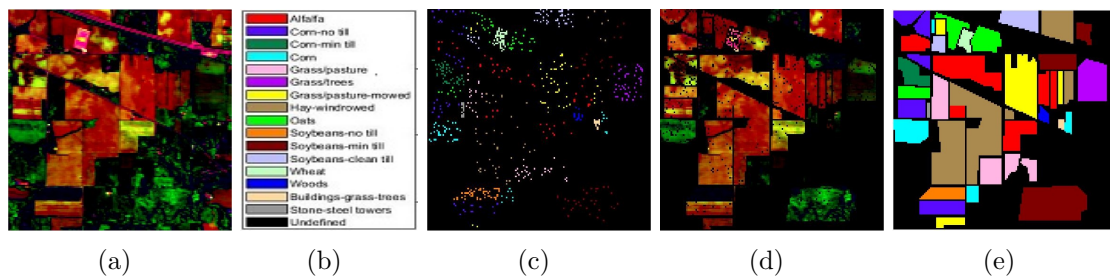

(a)   (b)   (c)   (d)   (e)

Figure 5-3: Indian Pines data set: (a) False color composition; (b) Classes with color code; (c) Training samples; (d) Test samples; and (e) Ground truth map.

5.3. Data sets description

Table 5.1: Class names, labels, as well as training and test samples selected by the IEEE GRSS Data Fusion Committee in 2013 [57], and available ground truths for Indian Pines dataset.

Land Classes	Class labels	Training samples	Test samples	Labeled Samples
Alfalfa	1	50	1384	1434
Corn (No till)	2	50	784	834
Corn (Min till)	3	50	184	234
Corn	4	50	447	497
Pasture/Grass	5	50	697	747
Trees/Grass	6	50	439	489
Pasture (Mowed)/Grass	7	50	918	968
Hay (Windrowed)	8	50	2418	2468
Oats	9	50	564	614
Soyabeans(No till)	10	50	162	212
Soyabeans (Min till)	11	50	1244	1294
Soyabeans (Clean till)	12	50	330	380
Wheat	13	50	45	95
Woods	14	15	39	54
Grass/Trees/Building	15	15	11	26
Steel/Towers/Stone	16	15	5	20

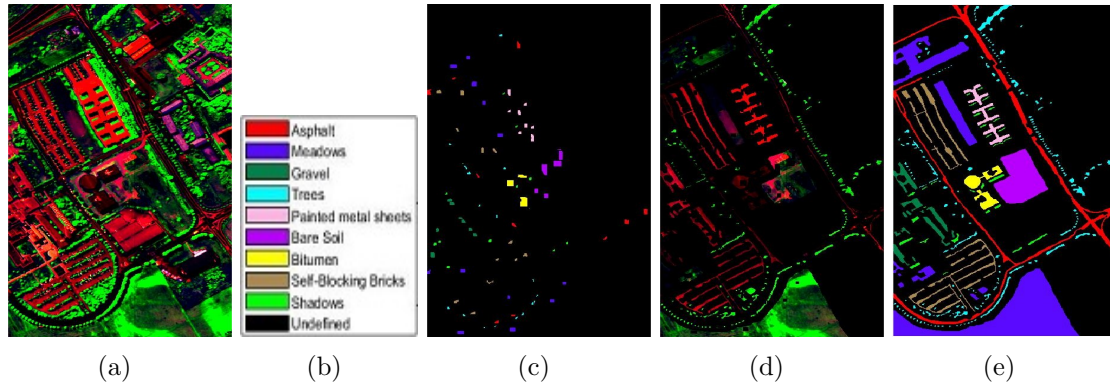## 5.3.2 Pavia University data set

(a)   (b)   (c)   (d)   (e)

Figure 5-4: Pavia University data set: (a) False color composition; (b) Classes with color code; (c) Training samples; (d) Test samples; and (e) Ground truth map.

This hyperspectral dataset was obtained over the urban area of the University of Pavia in Pavia, Italy. It was taken by the Reflective Optics System Imaging Spectrometer (ROSIS-03) airborne optical sensor. This sensor, equipped with 115 channels, 0.43 to 0.86 micrometers (µm) spanned the spectral range.

Following preprocessing, which entailed the removal of the 12 noisiest bands, the dataset was refined to consist spectral bands of 103. The resulting image possesses dimensions of 610 by 340 pixels showcases a spatial resolution of 1.3 meters. The details of the classes, training samples, test samples, ground truth map required for experimental validation are shown in Figure 5-4 and Table 5.2.

Table 5.2: Class names, labels, as well as training and test samples selected by the IEEE GRSS Data Fusion Committee in 2013 [57], and available ground truths for Pavia University dataset.

Land Classes	Class Labels	Training Samples	Test Samples	Labeled Samples
Asphalt	1	548	6304	6852
Meadow	2	540	18146	18686
Gravel	3	392	1815	2207
Tree	4	524	2912	3436
Metal sheet	5	265	1113	1378
Bare soil	6	532	4572	5104
Bitumen	7	375	981	1356
Brick	8	514	3364	3878
Shadow	9	231	795	1026

5.3.3   Houston University data set

The hyperspectral image of the University of Houston was captured through the Compact Airborne Spectrographic Imager (CASI) sensor, encompassing both the university grounds and the adjacent urban areas. The CASI sensor, with a 380 to 1050 nanometers (nm) ranged wavelength, equipped with 144 channels. This image features a spatial resolution of 2.5 meters and exhibits dimensions of 349 by 1905 pixels. The details of the classes, training samples, test samples, ground truth map required for experimental validation are shown in Figure 5-5 and Table 5.3.
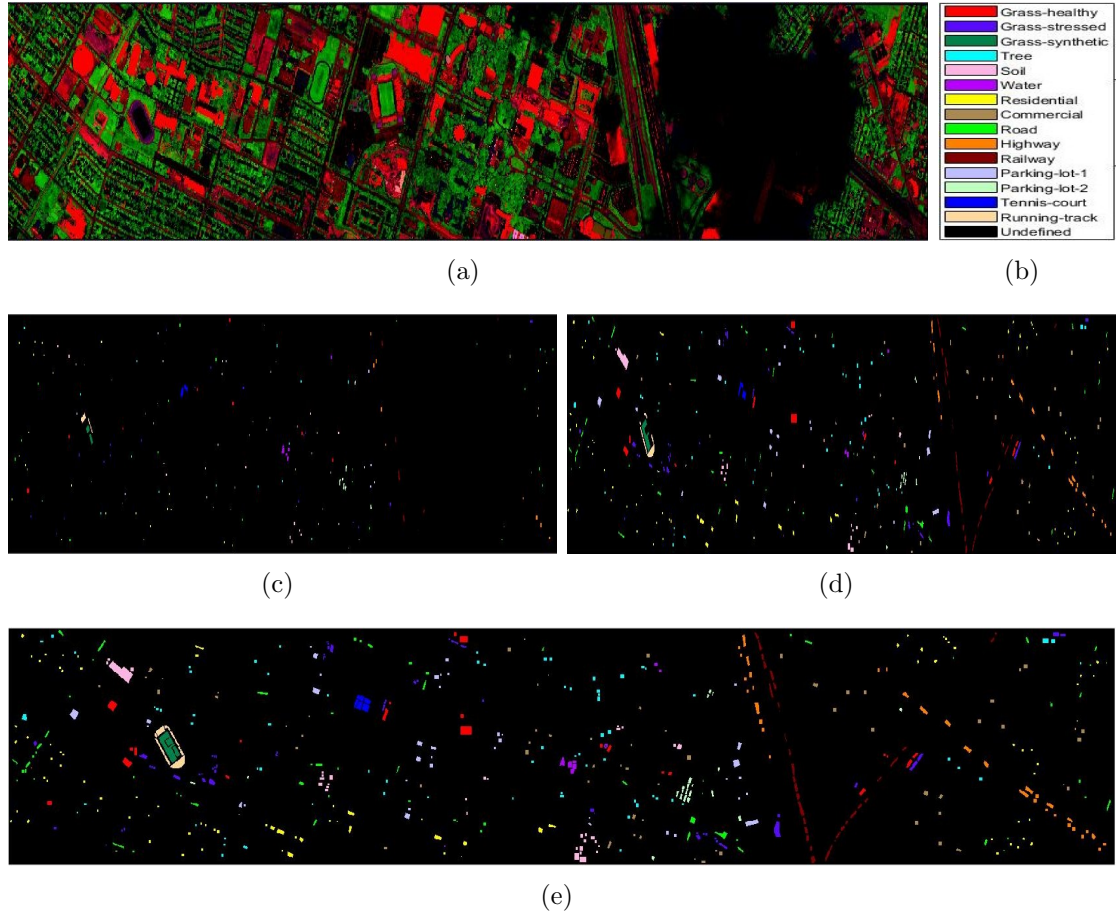
Figure 5-5: University of Houston data set: (a) False color composition; (b) Classes with color code; (c) Training samples; (d) Test samples; and (e) Ground truth map.

5.4   Experimental results

5.4.1   Experimental setting

To reduce the curse of dimensionality problem, the first m PCs of the considered HSI which preserve more than 99% information are used to construct the ESFP. For each PC, If n filtered images are generated, then the size of the constructed ESFP will be $m \times n$. In the present experiment regardless of the considered data set first 10 PCs are used to construct the ESFP. Once the ESFP is obtained, for experimental validation random forest (RF) classifier is used. The $m \times n$ spectral-

Table 5.3: Class names, labels, as well as training and test samples selected by the IEEE GRSS Data Fusion Committee in 2013 [57], and available ground truths for the University of Houston dataset.

Land Classes	Class Labels	Training Samples	Test Samples	Labeled Samples
Grass (Healthy)	1	198	1053	1251
Grass (Stresses)	2	190	1064	1254
Grass (Synthetic)	3	192	505	697
Tree	4	188	1056	1244
Soil	5	186	1056	1242
Water	6	182	143	325
Residential	7	196	1072	1268
Commercial	8	191	1053	1244
Road	9	193	1059	1252
Highway	10	191	1036	1227
Railway	11	181	1054	1235
Parking (Lot 1)	12	192	1041	1233
Parking (Lot 2)	13	184	285	469
Court (Tennis)	14	181	247	428
Track (Running)	15	187	473	660

spatial features of the profile are fed into the classifier for computing classification results. In our experiment, 200 trees are considered to construct the random forest classifier.

In the experiment the ESFP I and ESFP II are constructed by taking five windows of increasing size. The size of these windows is taken from the Fibonacci numbers [1 2 3 5 8 13 . . .], which grow exponentially with a non-uniform spacing of numbers, but their ratios converge to the golden ratio (approximately 1.618). The formula used to determine the size of the windows is $2 \times f + 1$, where the value of f is taken from the Fibonacci series [1 2 3 5 8 13 21 . . .]. Both the ESFP I and ESFP II are constructed by considering the window of sizes 5×5, 7×7, 11×11, 17×17, and 27×27. So the size of both the profiles is $10 \times 5 = 50$. The ESFP III and ESFP IV is constructed by generating 4 filtered images from each PC using the options $W_1|W_2|W_3|W_4$ and $F_1|F_2|F_3|F_4$, respectively. As a results the size of these profiles is $10 \times 4 = 40$.
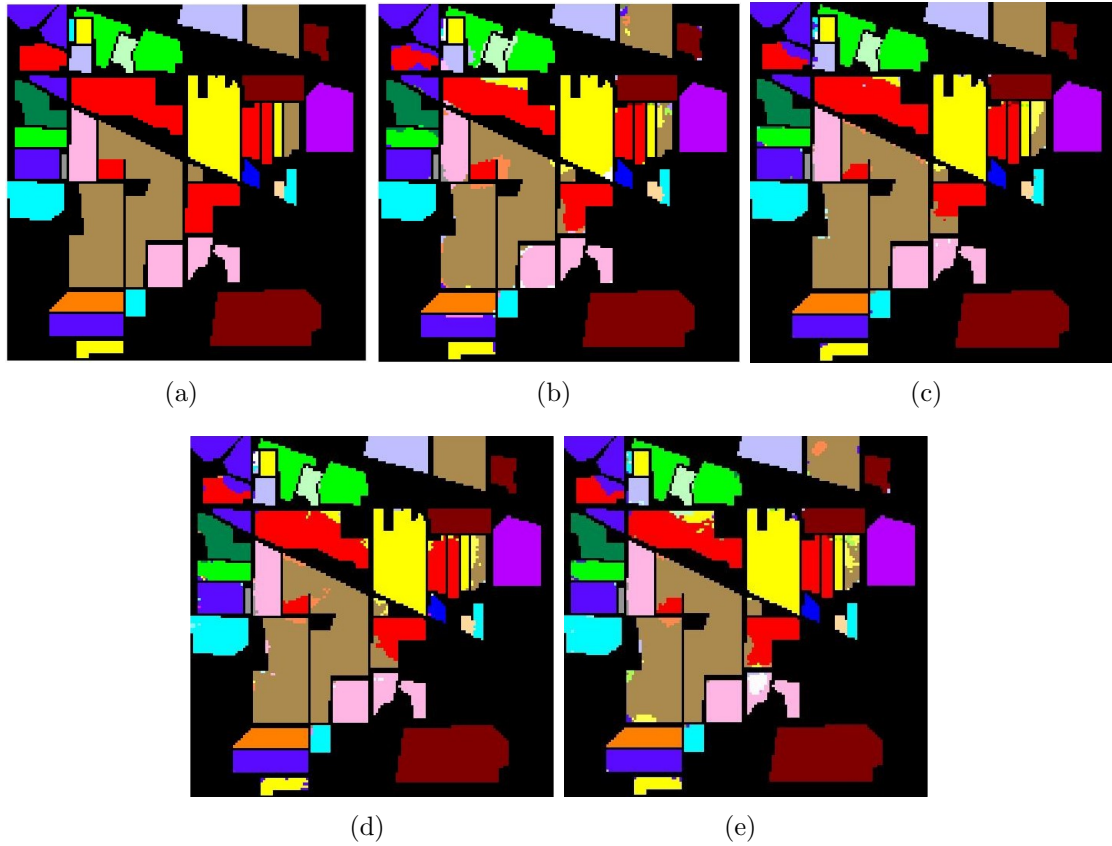
Figure 5-6: (a) Ground truth map and the classification maps provided by (b) ESFP I, (c) ESFP II (d) ESFP III and (e) ESFP IV for the Indian Pines data set.

5.4.2 Results analysis

In [57], a more challenging training and test samples for the above three described data sets are exercised to compare the overall performance of many state-of-the-art spectral-spatial HSI classification techniques that incorporate spatial information by using MRF, MM, segmentation, sparse representation, and deep neural networks. To show the potential of the proposed approaches, the ESFPs constructed by our four different method are also classified using the same training and test samples as used in [57]. This allows us to measure our developed techniques against the large number of state-of-the-art techniques reported in [57], which include other existing mathematical tools such as MRF, MM, segmentation, sparse representation, and deep neural networks for incorporating spatial information in

HSI.

Figure 5-7: (a) Ground truth map and the classification maps provided by (b) ESFP I, (c) ESFP II (d) ESFP III and (e) ESFP IV for the Pavia University data set.

The results for all three data sets, provided by the numerous spectral-spatial classification techniques are reported in [57]. By looking into these results, we found that three techniques, namely EMEP[56], MFASRC[46], and Gabor-CNN [30], provided significantly better results than the others. To gauge the adequacy of the proposed ESFPs, their classification results are contrasted with those of the three methods. Moreover, the threshold free attribute profile (TEAP) technique
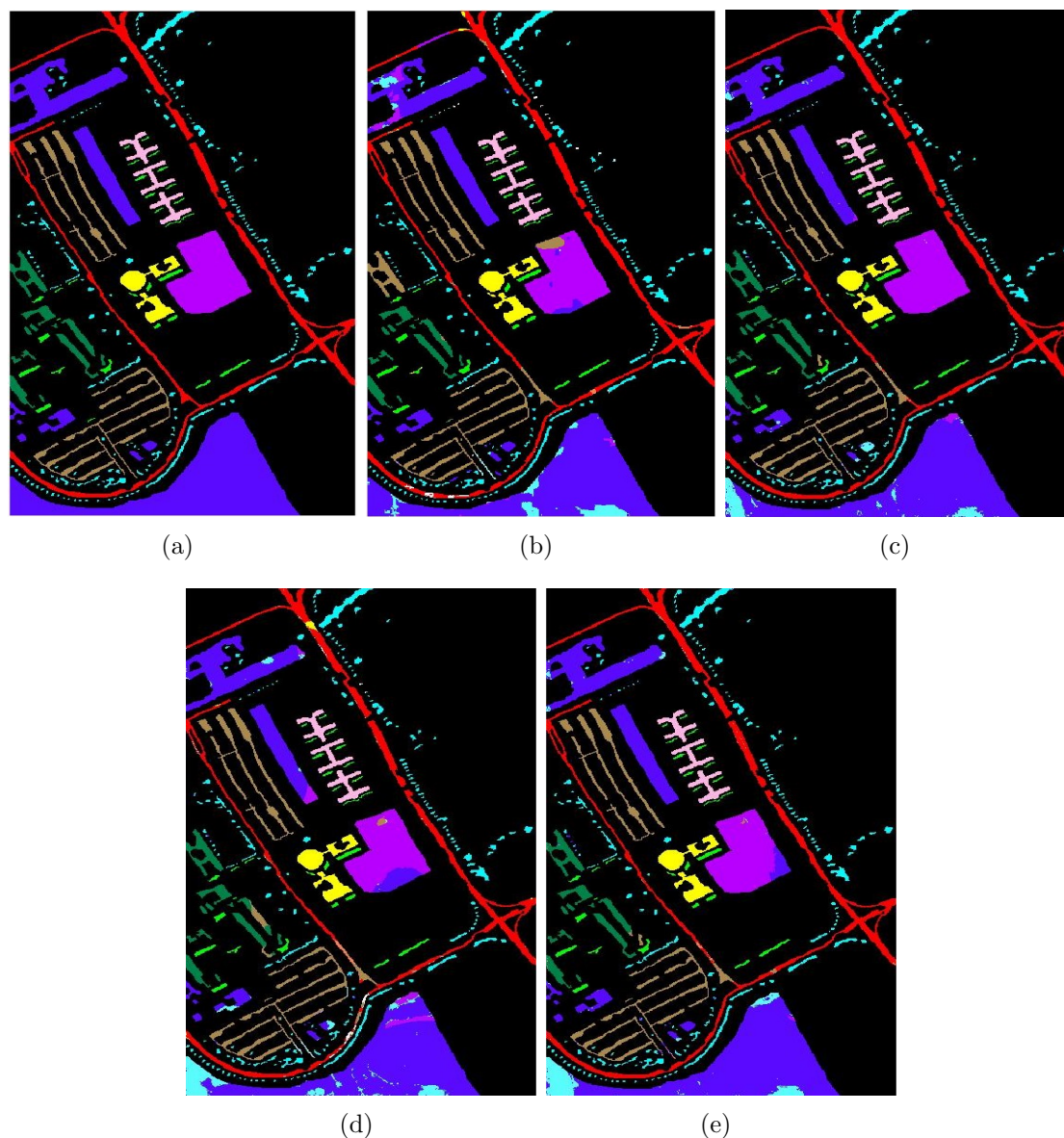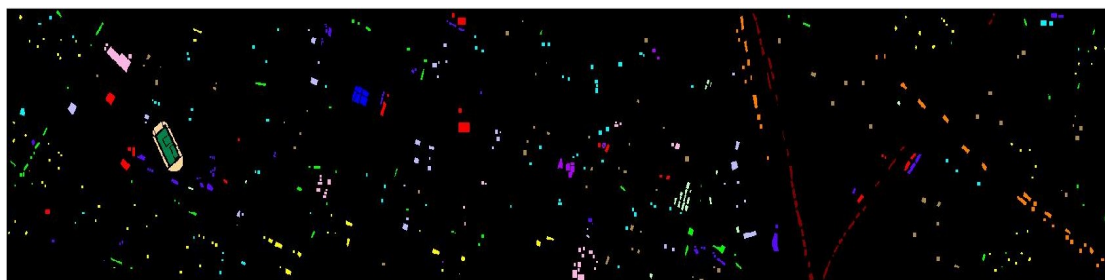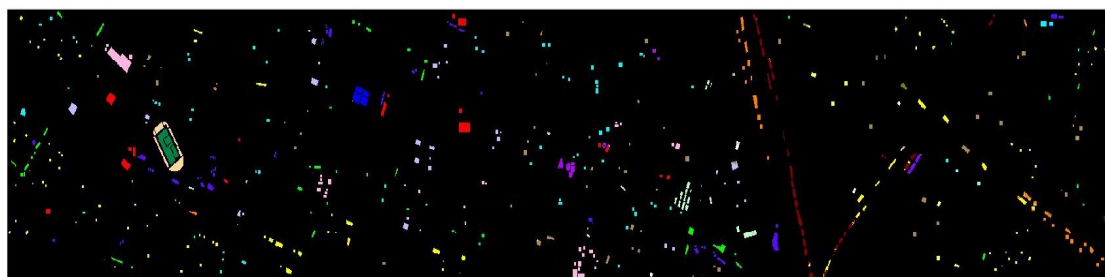
(a)

(b)

(c)

(d)

(e)

Figure 5-8: (a) Ground truth map and the classification maps provided by (b) ESFP I, (c) ESFP II (d) ESFP III and (e) ESFP IV for the University of Houston data set.
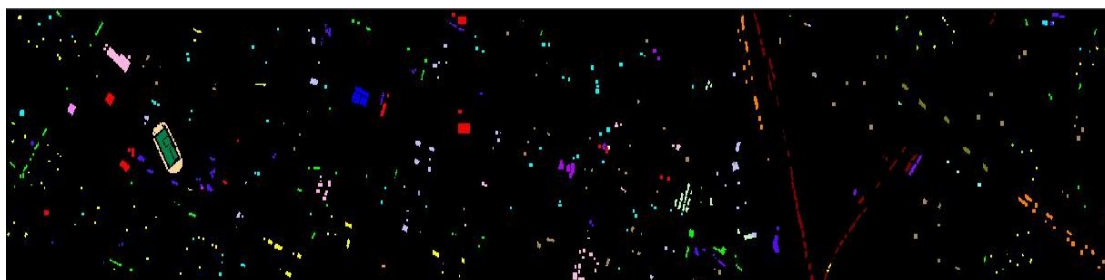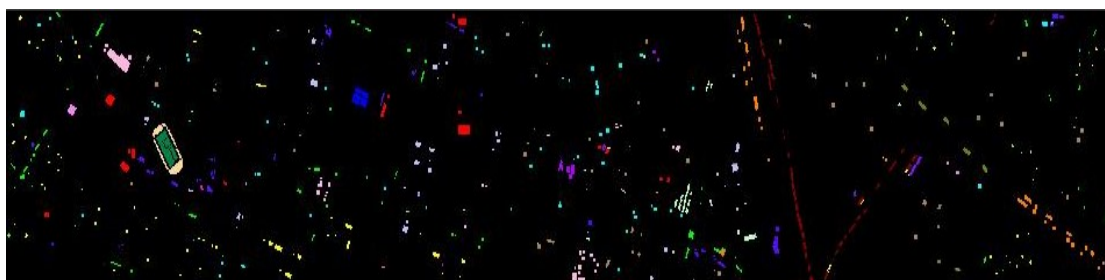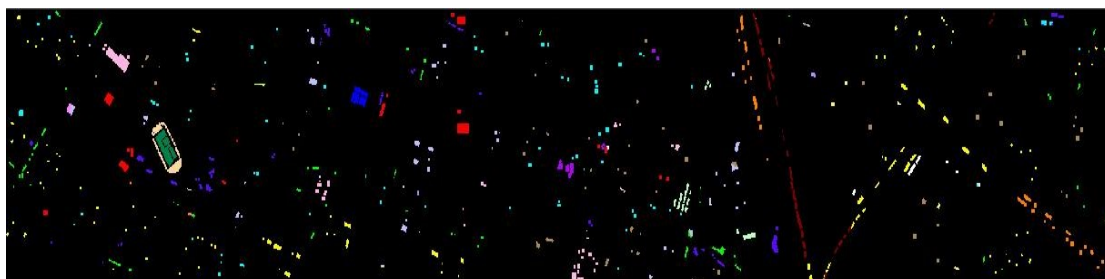
presented in [16] and the PCA-based edge-preserving filtering (EPF) technique presented in [71] are used for comparison. The average class-wise accuracy (AA), the overall accuracy (OA), the kappa accuracy (κ) and the class-by-class accuracy metrics are used for comparison. where both OA and AA have a range of [0, 100], while κ ranges from [0, 1], with higher values in each metric indicating better performance. Table 5.4, Table 5.5, and Table 5.7 reports the classification results provided by the different techniques for Indian Pines, Pavia University and University of Houston data sets, respectively. Where for the proposed techniques, four different classification models are applied for classifications, *viz* Random forest (RF), Support Vector Machine (SVM), Sparse Representation Classifier (SRC) and Convolution Neural Network (CNN). Among all these classifiers, RF provides the best results for the proposed approaches. Random Forest outperforms SVM, CNN, and SRC in HSI classification due to its robustness to high dimensionality, built-in feature selection, resilience to noise, and ability to handle limited training data (as the standard train test used here has limited and nonuniform training samples throughout the classes). It models complex nonlinear relationships effectively without extensive hyperparameter tuning. While CNNs excel with large datasets and spectral-spatial patterns, RF's simplicity and ensemble learning make it ideal for HSI data with limited training samples. From these tables, it can be observed that for the Indian Pines data set, our proposed ESFP I, ESFP II, ESFP III and ESFP IV provide higher AA than the existing techniques, while ESFP II produces the highest AA of 98.23% and OA of 97.26%. Whereas the best literature method (i.e., MFASRC [46]) yields an AA of 97.25% and an OA of 95.22%. For Pavia University data set, the proposed ESFP II provides the highest AA (97.84%) and OA (95.73%) while the best literature method (i.e., EMEP [56]) results in an AA of 96.57% and an OA of 95.46%. Similar results also observed on University of Houston data set. For this data set, the proposed technique provides classification accuracy higher of 2% and 3% than those of the best TEAP [16] and second best MFASRC [46] literature methods, respectively.

Table 5.4: The classification results obtained from various techniques using the same training and test samples as those in [57] for the Indian Pines dataset.

Cls./ Acc.	EMEP [56]	MFASRC [46]	Gabor-CNN [30]	TEAP [16]	EPF [71]	Proposed ESFP I	Proposed ESFP II	Proposed ESFP III	Proposed ESFP IV
	Indian Pines- Random Forest (RF) (Trees: 200)								
1	87.43	86.78	84.44	79.77	79.26	87.34	85.98	**91.25**	85.69
2	95.79	98.98	91.53	90.18	90.31	98.05	**99.62**	99.16	99.74
3	98.91	98.91	98.77	95.11	95.11	98.91	100	100	100
4	95.53	97.54	94.70	88.81	89.49	94.05	96.20	**98.14**	94.85
5	95.98	97.70	**99.28**	96.56	96.27	93.87	98.13	98.70	90.24
6	99.09	99.54	100	99.32	99.32	100	100	100	98.86
7	92.81	94.99	95.84	89.98	89.98	96.13	**97.39**	95.61	95.86
8	93.13	94.17	90.94	88.88	89.29	93.21	96.69	**97.39**	93.59
9	87.06	92.55	88.59	86.88	87.23	87.78	96.81	**98.09**	97.34
10	99.38	98.38	100	98.77	98.15	99.38	100	100	100
11	97.03	99.76	99.34	99.76	99.68	99.43	99.92	100	99.75
12	**99.09**	98.18	89.66	98.79	98.79	96.48	98.18	**99.09**	97.27
13	100	100	100	100	100	95.56	100	95.56	97.78
14	94.87	97.44	97.37	97.44	97.44	100	100	98.72	100
15	100	100	100	100	90.91	100	100	100	100
16	100	100	100	100	100	100	100	100	100
AA	96.00	97.25	95.24	93.86	93.84	98.01	98.01	**98.23**	97.09
OA	93.70	95.22	92.84	91.04	90.99	96.41	96.84	**97.26**	94.59
κ	0.9279	0.9452	0.9161	0.8975	0.8969	0.9588	0.9638	**0.9686**	0.9380
	SVM (C: 1000, gamma: 1, kernel: *rbf*)								
AA						93.98	93.98	94.19	93.10
OA						87.76	88.15	88.53	86.10
κ						0.8607	0.8652	0.8695	0.8420
	SRC (Alpha: 0.01, Cross-Validation Accuracy: 0.9237)								
AA						93.25	93.25	93.46	92.38
OA						88.15	88.55	88.93	86.49
κ						0.8651	0.8696	0.8739	0.8463
	CNN (Filters Layer 1: 64, Layer 2: 128, Dense Units: 256, Dropout Rate 1: 0.3, Learning Rate: 0.0001)								
AA						88.32	88.32	88.52	87.49
OA						85.51	85.89	86.26	83.89
κ						0.8400	0.8444	0.8486	0.8218

Table 5.5: The classification results obtained from various techniques using the same training and test samples as those in [57] for the Pavia University dataset.

Cls./ Acc.	EMEP [56]	MFASRC [46]	Gabor-CNN [30]	TEAP [16]	EPF [71]	Proposed ESFP I	Proposed ESFP II	Proposed ESFP III	Proposed ESFP IV
Pavia University- Random Forest (RF) (Trees: 200)									
1	96.05	82.95	87.75	93.73	86.29	90.91	**96.27**	93.58	91.62
2	93.45	56.79	97.25	87.14	80.98	78.62	92.60	96.12	**98.89**
3	81.37	91.07	70.92	79.12	80.94	58.55	**98.02**	66.33	89.03
4	**99.87**	96.36	97.09	98.83	94.88	95.80	98.59	98.12	96.98
5	99.93	97.75	98.83	99.91	**100**	99.43	**100**	**100**	99.64
6	99.26	81.87	64.62	75.15	80.93	97.15	**99.91**	98.58	86.43
7	**99.85**	99.39	76.66	99.49	90.42	89.25	97.35	95.31	93.86
8	**99.43**	95.87	99.05	99.14	90.69	94.99	99.29	93.11	94.96
9	**100**	92.33	98.36	99.62	94.59	97.86	98.62	98.65	99.12
AA	96.57	88.26	87.83	92.47	88.59	90.25	**97.84**	93.31	94.68
OA	95.46	74.39	91.62	89.17	85.30	88.88	**95.73**	94.68	93.33
κ	0.9407	0.6828	0.8914	0.8562	0.8072	0.8568	**0.9429**	0.9287	0.9300
SVM (C: 1000, gamma: 1, kernel: *rbf*)									
AA						86.54	93.82	89.47	90.79
OA						80.90	87.14	86.18	84.95
κ						0.7691	0.8464	0.8337	0.8348
SRC (Alpha: 0.01, Cross-Validation Accuracy: 0.9237)									
AA						85.87	93.09	88.78	90.09
OA						81.27	87.53	86.57	85.33
κ						0.7731	0.8508	0.8380	0.8391
CNN (Filters Layer 1: 64, Layer 2: 128, Dense Units: 256, Dropout Rate 1: 0.3, Learning Rate: 0.0001)									
AA						81.33	88.17	84.09	85.32
OA						78.83	84.90	83.97	82.78
κ						0.751	0.826	0.814	0.815

Table 5.6: The classification results obtained from various techniques using the same training and test samples as those used in [57] for the University of Houston dataset

Cls./ Acc.	EMEP [56]	MFASRC [46]	Gabor-CNN [30]	TEAP [16]	EPF [71]	Proposed ESFP I	Proposed ESFP II	Proposed ESFP III	Proposed ESFP IV
University of Houston- Random Forest (RF) (Trees: 200)									
1	77.78	80.82	**87.47**	82.03	54.04	82.03	81.67	81.20	81.67
2	76.88	82.52	**86.01**	85.10	82.24	85.15	83.46	85.06	85.15
3	**100**	**100**	78.22	99.90	**100**	99.98	**100**	**100**	**100**
4	82.77	82.77	85.02	92.10	44.41	**92.17**	90.06	86.27	91.00
5	96.02	**100**	99.89	99.80	**100**	99.89	99.81	99.91	99.91
6	95.80	**99.30**	89.44	94.60	95.80	94.62	**99.30**	95.80	95.80
7	72.95	86.29	**90.19**	85.70	63.53	85.70	73.60	71.55	84.61
8	**82.05**	68.66	74.44	77.19	45.68	77.19	78.06	79.49	79.11
9	63.17	78.75	84.42	82.40	51.37	82.47	**88.86**	78.47	82.81
10	67.57	66.60	63.61	57.25	48.55	57.29	**68.53**	68.24	**68.53**
11	82.83	81.50	80.06	77.84	90.89	77.84	**97.72**	95.92	73.34
12	84.53	74.35	**87.30**	80.90	78.77	80.99	79.06	86.26	76.85
13	73.68	63.86	**85.06**	70.05	71.57	70.07	69.47	64.91	65.26
14	99.60	**100**	**100**	99.50	**100**	99.59	99.60	**100**	99.59
15	98.94	**100**	56.95	99.65	98.94	99.70	99.79	**100**	99.79
AA	83.64	84.36	82.94	85.33	75.32	85.65	**87.24**	85.95	85.90
OA	80.83	82.09	84.12	82.82	70.21	83.76	**85.47**	83.36	83.16
κ	0.7920	0.8058	0.8251	0.8138	0.6773	0.8238	**0.8423**	0.8183	0.8173

SVM (C: 1000, gamma: 1, kernel: *rbf*)

Acc.	Proposed ESFP I	Proposed ESFP II	Proposed ESFP III	Proposed ESFP IV
AA	82.13	83.65	82.41	82.37
OA	76.24	77.80	75.88	75.70
κ	0.739	0.756	0.735	0.734

SRC (Alpha: 0.01, Cross-Validation Accuracy: 0.9237)

Acc.	Proposed ESFP I	Proposed ESFP II	Proposed ESFP III	Proposed ESFP IV
AA	81.49	83.01	81.78	81.73
OA	76.58	78.15	76.22	76.04
κ	0.743	0.760	0.738	0.737

CNN (Filters Layer 1: 64, Layer 2: 128, Dense Units: 256, Dropout Rate 1: 0.3, Learning Rate: 0.0001)

Acc.	Proposed ESFP I	Proposed ESFP II	Proposed ESFP III	Proposed ESFP IV
AA	77.18	78.62	77.45	77.41
OA	74.29	75.80	73.93	73.76
κ	0.722	0.738	0.717	0.716

Thus, for all the three data sets the constructed ESFPs are not only provided best classification results, but it also showed better consistency of the results obtained through all the classes and the datasets than the literature methods. For visual analysis, the classification maps generated by our ESFPs are shown in Figures 5-6, 5-7, and 5-8. These results confirmed the superiority of our proposed filtering techniques for incorporating spatial information in HSI classification.

5.5 Semantic segmentation of natural images

Semantic segmentation is a valuable approach for testing the effectiveness of edge-aware filtering in real-time image processing. Edge-aware filters aim to suppress noise while preserving critical boundaries, which is essential for maintaining the visual and structural integrity of real images. By applying semantic segmentation algorithms to both the raw and edge-aware filtered images, the effectiveness of boundary preservation can be evaluated. Metrics like boundary accuracy and edge alignment, combined with visual inspection, help determine whether the filtering retains fine details and sharp edges without introducing over-smoothing or distortions. This ensures the filtered images are optimized for real-time applications where edge fidelity is crucial.

In this section, firstly we have evaluated the proposed filtering techniques on two simple RGB images (without ground truth edge maps), previously used to show the filtering results in the previous chapters, to demonstrate the basic functionality and effectiveness of the methods. Additionally, we tested the techniques on two real test images (with ground truth edge maps) selected from the BSD500 dataset [100], a widely used benchmark image set for image processing tasks like edge detection and segmentation. These real-world images provide a robust basis to analyze the performance of the filtering techniques under more complex and natural conditions, emphasizing their capability to preserve meaningful edges while reducing noise.

5.5. Semantic segmentation of natural images

For this analysis, we have selected two recently developed semantic segmentation techniques to evaluate the edge-aware filtering results. The first is MeanShift++ [68], an unsupervised method that clusters pixels based on spatial and color similarities, making it suitable for evaluating edge preservation without reliance on labeled data. The second is BiSeNet V2 [151], a supervised model trained on the BSD500 [100] training set, which leverages a real-time architecture to produce precise segmentation maps. By using these complementary approaches, we aim to comprehensively assess the performance of the filtering techniques in terms of edge retention and noise reduction across both simple RGB and real-world BSD500 test images [100].

In this evaluation, we used two types of metrics, one is reference based and the other is non-reference based, to quantitatively assess the quality of segmentation for the edge-aware filtered images. This distinction was applied because the first two simple RGB images have no ground truth edge map for reference, while the two test images from the BSD500 dataset [100] include annotated ground truth data. Hence, to evaluate the quality of segmentation for edge-aware filtered images comparatively, we utilized both reference-based and non-reference-based metrics, ensuring comprehensive analysis across images with and without ground truth.

The non-reference based metrics, applicable to all images, include the Edge Preserving Index (EPI) [6] (ranges from 0 to 1, where higher is better) to measure edge retention. Spatial Entropy [26] (ranges from 0 to $\log n$, where n is the number of spatial zones, which also lower is better) to assess the uniformity of segmented regions, Color Homogeneity [11] (ranges from 0 to 1, where higher is better) for evaluating consistency within segments, and Normalized Gradient Deviation (NGD) [77] (lower is better) to determine the preservation of gradient information. Together, these metrics provide a balanced framework to assess edge preservation, semantic coherence, and overall segmentation quality, with optimal

values indicating better performance for each respective criterion. The metrics Spatial Entropy, Color Homogeneity, Edge Precision Index (EPI), and Normalized Gradient Difference (NGD) provide complementary yet often contradictory perspectives on segmentation quality. Spatial Entropy assesses the balance of segment sizes, favoring evenly distributed regions, while Color Homogeneity emphasizes the uniformity within segments, prioritizing smoothness over boundary precision. EPI focuses on accurate edge detection, rewarding sharp and well-defined boundaries but potentially penalizing over-smoothing. In contrast, NGD evaluates the preservation of smooth transitions, often conflicting with EPI by penalizing sharp gradients. These metrics highlight the inherent trade-offs between uniformity, edge accuracy, and smoothness in evaluating segmentation results.

The first two rows of Table 5.7 show the quantitative values for these metrics for two simple RGB image taken from the previous chapters. Where the segmentation is done by the unsupervised MeanShift++ [68] semantic segmentation technique. Figure 5-9 is showing the corresponding images and their comparative segmented images, respectively.

The evaluation reveals distinct strengths across techniques, with the original image of Pompeii Marine Life 5-9(a) excelling in edge preservation ($EPI = 0.4543$) and GISF [92] leading in spatial entropy (1.7366) and color homogeneity (0.4701), highlighting their suitability for structural detail and region uniformity, respectively. The proposed three techniques, particularly the proposed III from Chapter 4, offer a balanced trade-off, excelling in gradient consistency ($NGD = 0.4820$) and achieving the best edge retention ($EPI = 0.2071$) among them while maintaining reasonable spatial entropy and color homogeneity. Proposed III stands out as the most versatile for general-purpose segmentation tasks. Similarly, for the image of the flower in Figure 5-9(i) the proposed I technique produces the overall best balanced semantic segmented map using MeanShift++ [68].

5.5. Semantic segmentation of natural images

The reference-based metrics, applicable to the BSD500 dataset with ground truth, include the F1 Score [6] for Edge (ranges from 0 to 1, where higher is better) for measuring overlap between predicted and true edges, Fragmentation Index (FI) [126] (ranges from 0 to 1, where lower is better) for assessing edge continuity, Normalized Mutual Information (NMI) [9] (ranges from 0 to 1, where higher is better) for quantifying agreement with ground truth, and Hausdorff Distance [73] (ranges from 0 to the diagonal length of input image, where lower is better) for evaluating the proximity of predicted and true boundaries. The metrics F1-Score for Edges, Fragmentation Index (FI), Normalized Mutual Information (NMI), and Hausdorff Distance offer diverse and sometimes conflicting evaluations of segmentation quality. The F1-Score for Edges assesses the overlap between predicted and ground truth edges, prioritizing precision and recall for boundary accuracy. Fragmentation Index (FI) evaluates the structural coherence within segments, penalizing excessive splitting of regions and favoring compactness. Normalized Mutual Information (NMI) measures the consistency between predicted and ground truth labels, emphasizing alignment in segmentation structure without focusing on spatial precision. Hausdorff Distance quantifies the maximum spatial deviation between segment boundaries, focusing on extreme mismatches but ignoring overall accuracy. Together, these metrics reflect trade-offs between edge precision, segment cohesion, label alignment, and boundary deviations, providing a comprehensive yet contrasting view of segmentation performance.

Figure 5-10 shows the semantic segmentation results of two different test images from the well-known dataset BSD500 [100] applying the real-time supervised technique BiSeNet V2 [151]. From the segmented images, it can be seen that the proposed techniques produce an overall balanced segmentation map for both images. The last two rows of Table 5.7 show the quantitative values for these metrics for two test images, respectively.

Table 5.7: The quantitative measure for the segmented image quality using no reference (without ground truth) metrics like Spatial Entropy, Color homogeneity, Edge preserving index (EPI), Normalized gradient deviation (NGD) and reference based (with ground truth) metrics F_1 score, Fragmentation Index (FI), Normalized mutual information (NMI) , and Hausdorff distance (HD) of the original images and the filtered images produced by BTF [32], SATF [69], GISF [92] and three proposed techniques of the previous chapters. For semantic segmentation, two recently developed different methods: MeanShift++ [68] and BiSeNet V2 [151]. The bold fonts indicate the best scores of the respective metrices.

| Metrics	Original Images	BTF [32]	SATF [69]	GISF [92]	Proposed I [2]	II [3]	III [4]
For two simple RGB images (see Figure: 5-9) without ground truth (By applying MeanShift++ [68])							
Spatial Entropy [26]	1.0357	1.4571	1.0691	**1.7366**	1.2668	1.1671	1.0770
Color Homogeneity [11] 0.1430	0.2725	0.1284	**0.4701**	0.2670	0.1687	0.2153	
EPI [6]	**0.4543**	0.1907	0.0941	0.1175	0.2053	0.1909	0.2071
NGD [77]	0.6745	0.5030	**0.3711**	0.4043	0.4416	0.4989	0.4820
Spatial Entropy [26]	1.4506	1.5083	1.4701	1.6047	**1.6758**	1.5403	1.3776
Color Homogeneity [11] 0.2295	0.2483	0.2337	0.3390	**0.5011**	0.2451	0.1025	
EPI [6]	**0.202297**	0.1275	0.1384	0.0994	0.2269	0.1644	0.1604
NGD [77]	0.5132	0.4215	0.437	**0.3734**	0.5394	0.4705	0.468
For two test images (see Figure: 5-10) with ground truth from the BSD500 dataset [100] (By applying BiSeNet V2 [151])							
F1-Score for Edges [6]	0.4811	0.4811	0.4710	0.4810	0.4381	**0.4813**	0.4801
Fragmentation Index (FI) [126]	0.8292	0.8323	0.6920	0.6834	0.7507	**0.6833**	0.6841
NMI [9]	0.0112	0.0110	**0.0236**	0.0129	0.0137	0.0135	0.0139
Hausdorff Distance [73]	71.45	70.72	72.24	78.24	72.48	**70.48**	70.72
F1-Score for Edges [6]	0.2677	0.1783	0.1482	0.1479	0.2692	0.2738	**0.2739**
Fragmentation Index (FI) [126]	0.8577	0.8780	0.9065	0.9099	0.8444	0.8804	**0.8413**
NMI [9]	0.0204	0.0193	0.0190	**0.0226**	0.0198	0.0185	0.0186
Hausdorff Distance [73]	121.79	117.44	116.66	**111.83**	115.52	116.18	118.01
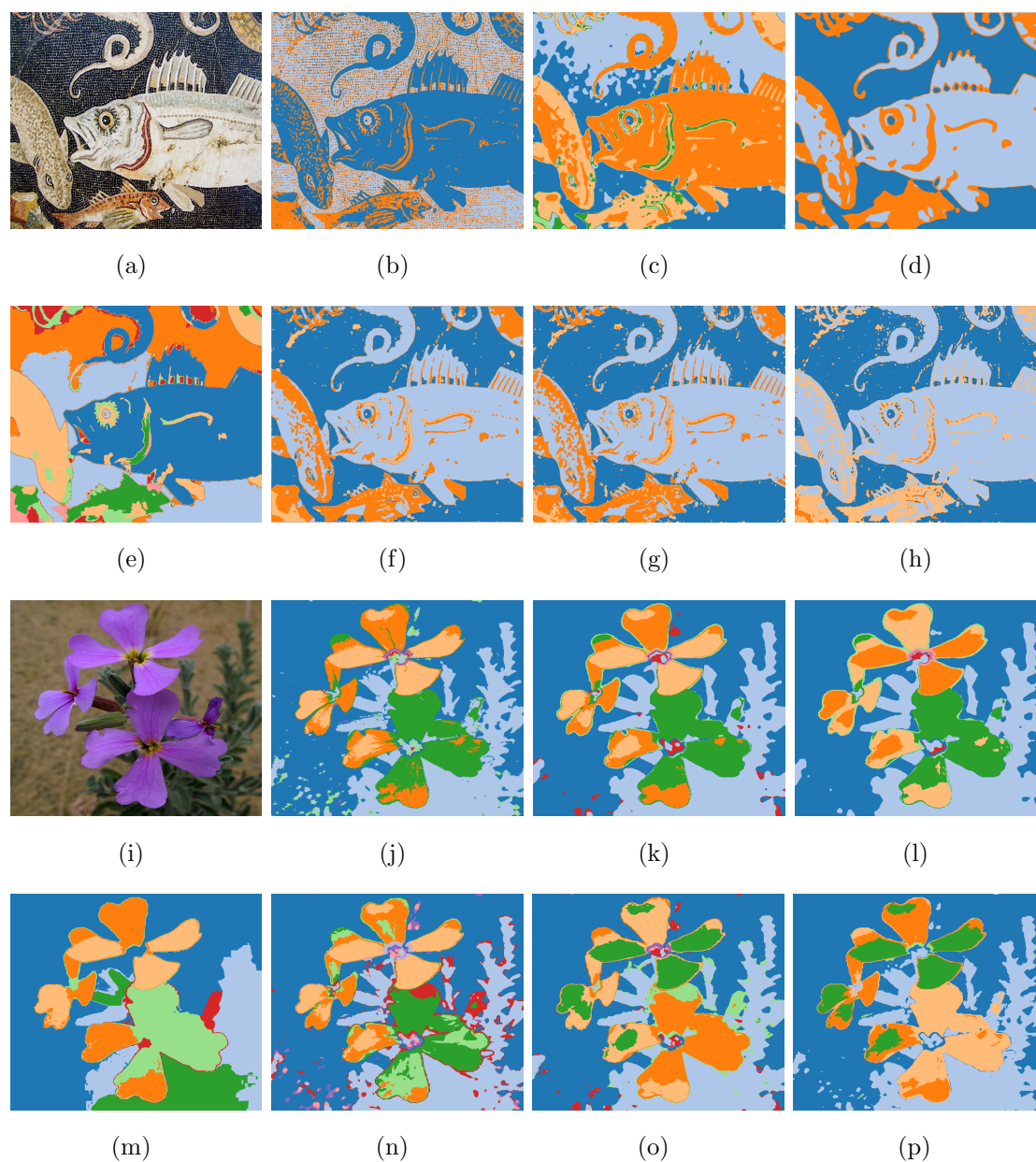
Figure 5-9: (a)(i) Original images (Two simple RGB images, those already used in the previous chapters to show the filtering results) and (b)(j) Segmented image from those original images, (c)(k) Segmented image from the filtered images of BTF [32], (d)(l) SATF [69], (e)(m) GISF [92], and (f-h)(n-p) proposed I, II, and III techniques from Chapters: 2, 3, and 4 respectively. Segmented using an unsupervised technique, MeanShift++ [68].
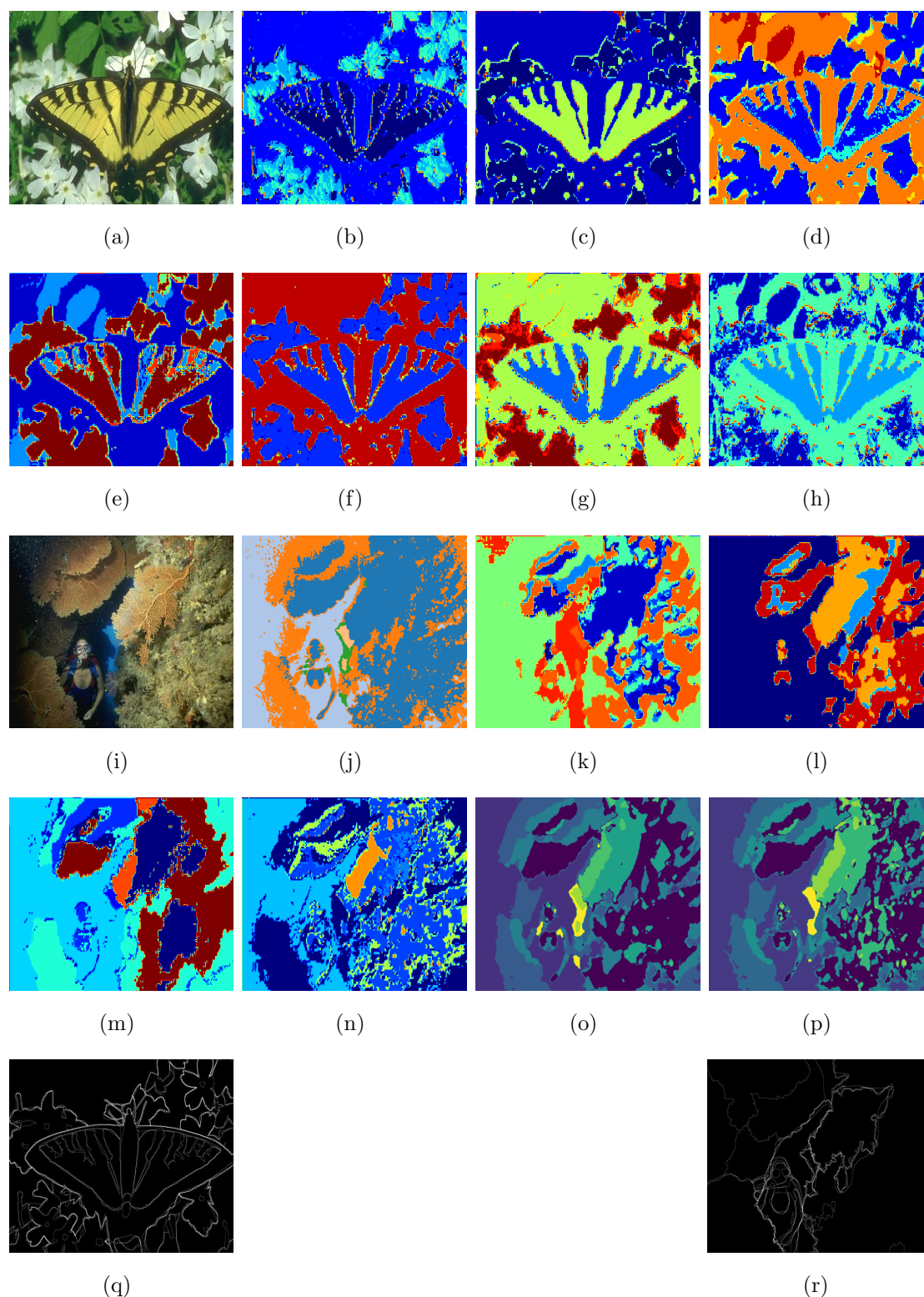
155

Figure 5-10: (a)(g) Two original test images from BSD500 dataset [100] and (b)(j) segmented images from those original images, (c)(k) segmented images from the filtered images of BTF [32], (d)(l) SATF [69], (e) GISF [92], and (f-h)(n-p) proposed I, II, and III techniques from the Chapters: 2, 3, and 4 respectively. Segmented by applying a supervised technique called BiSeNet V2 [151], (q)(r) Respective ground truth edges of the two test images.

The superiority of the last three segmentation maps is evident when analyzed across the four metrics: F1-Score, Fragmentation Index (FI), Normalized Mutual Information (NMI), and Hausdorff Distance (HD). While for the first butterfly image the proposed II technique excels in the three different metrics F_1 score (0.4813), FI (0.6833), and HD (70.48) indicating the highest agreement with ground truth, reflecting reduced fragmentation and minimal mismatch. Also, both techniques, the proposed I and III, demonstrate a balanced performance across all the metrics. These metrics collectively highlight the robustness of the last three techniques, with the proposed II standing out as the most consistent performer, demonstrating their overall superiority in handling the trade-offs between precision, fragmentation, agreement, and boundary alignment. Similarly for the second test image of the plunger, all three proposed techniques produce overall well-balanced results, where proposed III excels the most.

5.6 Conclusions

Semantic-aware structure preserving image filtering techniques play a pivotal role in preserving crucial structural information while reducing irrelevant textures and noises in the image. Even if the images produced by such an approach provide rich spatial information, they are seldom used for HSI classification. In this chapter, the semantic edge-aware structure preserving image filtering techniques proposed in this thesis are used to construct ESFPs for spectral-spatial classification of HSI. The performance of the proposed techniques have been validated by comparing it with that of many state-of-the-art spectral-spatial HSI classification techniques that incorporate spatial information by exploiting MRF, MM, segmentation, sparse representation, or deep neural networks. The results of the comparison pointed out the supremacy of the semantic filtering profile over the state-of-the-art techniques for incorporating spatial information. As a future development of this part of work, we plan to use the ESFP as an input to the deep

models for guiding the extraction of more effective features. To the end, in addition to this, the proposed filtering techniques are also validated through semantic segmentation of natural images by one unsupervised technique, MeanShift++, and another supervised technique, BiSeNet V2. The comparative results have shown the potential of the techniques in the overall balanced image segmentation. As a future task as an extension to this part, we plan to develop a new approach for semantic segmentation, incorporating our filtering techniques.

List of publications from this chapter:

Journals

1. Pradhan K., Patra S. and Bruzzone L.,"Extended semantic edge-aware filtering profile for hyperspectral image classification," in IEEE Geoscience and Remote Sensing Letters, https://doi: 10.1109/LGRS.2024.3387473. *(IF-4.0)*.

Conference

1. K. Pradhan and S. Patra, "Semantic-Aware Image Filtering for Classification of Hyperspectral Images," 3rd International Conference on Intelligent Vision and Computing (ICIVC 2023) (*Presented*).