

Article 3D-Gabor Inspired Multiview Active Learning for Spectral-Spatial Hyperspectral Image Classification

Jie Hu¹, Zhi He^{1,*}, Jun Li^{1,*}, Lin He² and Yiwen Wang¹

- ¹ Guangdong Province Key Laboratory of Urbanization and Geo-Simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China; hujie23@mail2.sysu.edu.cn (J.H.); yizhigiser@126.com (Y.W)
- ² College of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China; helin@scut.edu.cn
- * Correspondence: hezh8@mail.sysu.edu.cn (Z.H.); lijun48@mail.sysu.edu.cn (J.L.)

Received: 16 May 2018; Accepted: 3 July 2018; Published: 5 July 2018



Abstract: Active learning (AL) has been shown to be very effective in hyperspectral image (HSI) classification. It significantly improves the performance by selecting a small quantity of the most informative training samples to reduce the complexity of classification. Multiview AL (MVAL) can make the comprehensive analysis of both object characterization and sampling selection in AL by using various features of multiple views. However, the original MVAL cannot effectively exploit the spectral-spatial information by respecting the three-dimensional (3D) nature of the HSI and the query selection strategy in the MVAL is only based on the disagreement of multiple views. In this paper, we propose a 3D-Gabor inspired MVAL method for spectral-spatial HSI classification, which consists of two main steps. First, in the view generation step, we adopt a 3D-Gabor filter to generate multiple cubes with limited bands and utilize the feature assessment strategies to select cubes for constructing views. Second, in the sampling selection step, a novel method is proposed by using both internal and external uncertainty estimation (IEUE) of views. Specifically, we use the distributions of posterior probability to learn the "internal uncertainty" of each independent view, and adopt the inconsistencies between views to estimate the "external uncertainty". Classification accuracies of the proposed method for the four benchmark HSI datasets can be as high as 99.57%, 99.93%, 99.02%, 98.82%, respectively, demonstrating the improved performance as compared with other state-of-the-art methods.

Keywords: hyperspectral image classification; multiview active learning (MVAL); 3D-Gabor; feature assessment; sampling selection.

1. Introduction

Hyperspectral image (HSI) [1–4] contains hundreds of narrow bands, and has been extensively used in different application domains, such as forest monitoring and mapping [5,6], land-use classification [7,8], anomaly detection [9], endmember extraction [10] and environment monitoring [11]. Among those kinds of applications, supervised classification is a fundamental task and has been widely studied over the past decades [12,13]. Detailed spectral information is naturally beneficial for supervised land-cover classification in HSI. However, problems such as Hughes phenomenon can emerge due to the high dimension of HSI data [14]. To alleviate the above-mentioned problem, a lot of methods have been proposed in literature. For instance, band selection is studied to reduce the redundancy between contiguous bands [15] and spectral-spatial feature extraction takes advantages of the more distinguishable characteristics [16,17]. However, among those methods, sufficient labeled samples/pixels are crucial to get the reliable classification results [18]. Since it is difficult to obtain



a large number of labeled samples due to the time-consuming and expensive manual labeling process [19], defining a set of high informative training set is one of the solutions.

To select the ideal training samples, active learning (AL) has been widely studied [20,21]. As a resampling approach, it seeks interactively to construct a set of the most informative training samples from the unlabeled data pool in a biased way, thus substantially reducing the human labeling cost without sacrificing the classification accuracy [22]. It is a man-machine interacted learning procedure, which is able to significantly improve the performance of classifiers when the selected samples are of high information.

Selecting informative samples plays a critical role in AL [23] and therefore, several query strategies have been proposed in the literature. The sampling selection methods can be divided into three families [24]. (1) Posterior probability-based heuristic, which gives confidence of the class assignment to estimate classification uncertainty of each candidate, such as breaking ties (BT) [25], mutual information (MI) [26] and Kullback-Leibler (KL)-Max strategy [27,28]; (2) Large margin-based heuristic, which uses the distance to hyperplane to estimate the confidence of classifier, with straightforwardly utilizing classifiers like SVM [29]. This family includes several representative methods, such as margin sampling (MS) [30], multiclass level uncertainty (MCLU) [31] and significance space construction (SSC) [32]; (3) Committee-based heuristic, which qualifies the uncertainty of samples by using the inconsistent hypothesis between each committee. Typical methods include the normalized entropy query-by-bagging (nEQB) [33], maximum disagreement (MD)-based criteria [34] and adaptive maximum disagreement (AMD) [35]. The traditional single-view AL (SVAL) is usually based on the first two families, and a new branch of AL named multiview AL (MVAL), which adhered to the principles of the third family with the information of multiple views, has attracted considerable interest over the past few years [36].

The MV is first introduced into AL domain in [37], where the Co-Testing algorithm is proposed to learn the hypotheses of the model. Different from the traditional SVAL, MVAL describes the objects by learning from several views [38,39]. MVAL has been proven to be very successful for HSI classification and the advantages of the MVAL lie in three aspects [34]. First, it provides a direct way to estimate the value of candidates using disagreement between each classifier. Second, it significantly decreases the number of training samples by exploiting complementary information, and meanwhile the learning procedure converges quickly with fewer learning steps. Third, the result of MVAL is more credible than SVAL, since the classification output is the combination of different predictions. Based on the afore-mentioned advantages, we focus on the MVAL for HSI classification in this paper.

In MVAL, the method of constructing multiple views and the strategy of sampling selection are the two core issues, which are of vital importance. As to the first issue, [40] introduces a feature splitting method and [41] presents four view generation approaches by utilizing the original spectral information, while the spectral-spatial based view generation strategies proposed in [34,42] can incorporate spatial information to bring the original data into a more discriminative space. Moreover, a feature-driven AL is proposed by [43], in which Gabor filtering and morphological profiles are used for instantiation. However, the original data in [43] is increased to 52 times of the bands, leading to the extremely large amount of channels. To solve the problem, we propose a three-dimensional Gabor (3D-Gabor) and cube assessment based method to generate the multiple views without augmenting the dimensions. As to the second issue, various query selection methods have been proposed in the last decade [24,34,35]. They provide the simple and direct approaches to utilize the nature of disagreement between multiple views. However, those methods ignore the abundant information of the probability distributions within each independent views. What is worse, different classifiers are likely to gradually converge to a similar stage as the iteration number increases, which causes the decrease of the efficient inconsistency information. Therefore, we propose an internal and external uncertainty estimation (IEUE) strategy, which makes full use of both uncertainties within each independent classifier and between different classifiers.

Based on the above-mentioned analysis, we present a 3D-Gabor inspired MVAL method using spectral-spatial information for HSI classification in this paper. Compared to the existing literature, the contributions of this paper are twofold:

- 1. We propose a 3D-Gabor feature extraction and cube assessment based method for view generation. Representative views are generated with only one certain frequency and direction to ensure the low dimensional filtering outputs.
- 2. We present an IEUE strategy to rank the unlabeled samples for selection. Compared to most of the existing query strategies, the proposed method takes advantages of the integral of posterior distribution and committee-based disagreement.

The remainder of this paper is organized as follows. Section 2 introduces the proposed MVAL framework, in which 3D-Gabor feature extraction, cube assessment, IEUE query method and the output strategy are described in detail. Section 3 reports the experimental results on four benchmark HSI datasets. Section 4 gives some subsequent discussions of our proposed method and conclusions are drawn in Section 5.

2. Proposed Method

In this section, we describe the proposed spectral-spatial based MVAL in detail. The block diagram of our framework is shown in Figure 1, which is composed of two parts: view generation and query selection. Firstly, since the sufficiency, independency and accuracy are the vital issues in view generation [41], a novel approach obeying these principles is proposed to construct multiple views. Notice that each view is able to provide the sufficient information and is distinct from others, classifiers trained from those views can make reliable disagreement assumption. Sequentially, based on the prediction output of classifiers, the IEUE strategy ranks candidates by their internal-external uncertainties both independently and comprehensively. The samples with high controversy degree are regarded as informative. A certain batch of them are selected from the unlabeled pool in each iteration. The selected samples are then labeled by manual work and added into the training set to boost the generalization of the classifiers. The whole interaction repeats until *N* times or the proposed stopping criteria is met. Finally, we use the majority voting approach to combine the classification output of each view.



Figure 1. Flowchart of the proposed MVAL framework.

Figure 2 plots the proposed 3D-Gabor and cube assessment based view construction method. Firstly, the 3D-Gabor filters with various frequencies and directions are adopted for converting original HSI into multiple data cubes, which can provide different spectral-spatial information between each other. Sequentially, a cube evaluation criterion, which is motivated by the Fisher's ratio criteria (FR) [44] and conditional mutual information [45], is proposed to calculate the sufficiency and independency of the cubes obtained from the first step. Cubes which have qualified assessment are regarded as views in our proposed method.



Figure 2. The general view generation approach architectures.

2.1.1. 3D-Gabor

Gabor transform is closely related to human visual, and has been widely used in face recognition [46], texture classification [47] and information mining [48]. As the 3D extension of the traditional two-dimensional Gabor (2D-Gabor), 3D-Gabor is a powerful tool for feature extraction of the HSI data. It can capture the spatial and spectral information simultaneously. By respecting the 3D natural characteristic of the HSI data, 3D-Gabor has shown remarkable performance in HSI data analysis [49]. Specifically, a 3D-Gabor kernel can be designed as follows

$$\mathbf{G}_{\omega,\varphi,\theta}(x,y,\lambda) = g(x,y,\lambda) \exp\left\{j(x\omega_x + y\omega_y + \lambda\omega_\lambda)\right\}$$
(1)

where ω denotes the central frequency of the wave vector, and ω_x , ω_y , ω_λ are the projections of vector along *x* axis, *y* axis and *spectral* dimensions. φ denotes the angle between the vector and *spectral* dimensions and θ is the angle between the projection of the wave vector on *x*-*y* plane and *x* axis. The factor $g(x, y, \lambda)$ refers to a 3D Gaussian envelope in the (x, y, λ) domain, and other factors are exponential harmonic.

As shown in Equation (1), the 3D-Gabor filter captures spectral-spatial information in a comprehensive manner. It is noteworthy that in HSI, discriminative information for classification is tend to appear on low frequencies in the spatial dimension and high frequencies in the spectral dimension [50,51]. Therefore using spatial smoothing and differential spectral preservation can enhance the class separability [49]. We apply the low-pass filter using Gaussian-enveloped *cos* harmonic to extract the smooth spatial feature to preserve the integral structure of images and reduce the noise and high-pass filter with Gaussian-enveloped *sin* harmonic to obtain the distinct signal in spectral domain.

Moreover, several frequencies and orientations are adopted to generate several data cubes with different spectral-spatial features. Specifically, the parameters of frequencies and orientations used in this paper are set as follows

- $\omega \in \{\frac{1}{4}, \frac{1}{8}, \frac{1}{12}, \frac{1}{16}, \frac{1}{20}\};$
- $\{\varphi, \theta\} \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}.$

It is noteworthy that when $\varphi = 0$, the wave vector points to the same direction with different θ , thus leading to a total of 13 orientations. Furthermore, each cube is generated with a certain frequency ω and a pair of orientations { φ, θ }. Therefore, the obtained gabor cube $M = Gabor(data, \omega, \varphi, \theta)$ owns the same size as original HSI data. We can get 65 Gabor cubes varying in frequencies and directions, and the cube assessment criterion can be adopted to select the most suitable ones from the generated 65 Gabor cubes.

2.1.2. Cube Assessment

A. Fisher's ratio In this paper, the FR criteria is adopted to measure the class separability of each 3D-Gabor cube. Using the between-class and within-class scatter matrices, the FR of each pair of classes is modeled as

$$FR_{i,j} = \sum_{x \in D_s} \frac{(\mu_i - \mu_j)(\mu_i - \mu_j)^T}{(\mu_i - x)(\mu_i - x)^T + (\mu_j - x)(\mu_j - x)^T}$$
(2)

where the numerator represents between-class scatter matrix based on the means of class *i* and class *j*, and the denominator is within-class scatter matrix based on the variance of those two classes. $FR_{i,j}$ implies the class separability between classes, the larger the $FR_{i,j}$, i.e., the better discriminating capability. In this paper, the overall FR is calculated by the mean value of each two class $FR_{i,j}$, which can be written as

$$FR = \frac{2}{r(r-1)} \sum_{i,j \in \{1,2,\dots,r\}} FR_{i,j}$$
(3)

where *r* denotes the number of ground class, and $\frac{r(r-1)}{2}$ represents the number of $FR_{i,j}$ obtained by each pair of classes. It is noteworthy that we perform the FR on the small size of the initial labeled training samples D_s in this paper. To address the resulting dimensional issue, we calculate the FR on band level and then average the result. As such, the modified FR can be formed by

$$\widetilde{FR} = \frac{1}{B} \frac{2}{r(r-1)} \sum_{b}^{B} \sum_{i,j \in \{1,2,\dots,r\}} \sum_{x \in D_s} \frac{(\mu_i - \mu_j)(\mu_i - \mu_j)^T}{(\mu_i - x)(\mu_i - x)^T + (\mu_j - x)(\mu_j - x)^T}$$
(4)

As an example, Table 1 displays the FR (i.e., the FR obtained from Equation (4)) of each 3D-Gabor cube for the Indian Pines dataset. Detailed descriptions of this image is shown in Section 3.1. It is observed from Table 1 that cubes with lower frequency often achieve higher FR. Intuitively, the cubes with larger FR are more suitable for constructing views since they can separate the classes much better. However, it is noteworthy that similar cubes can yield close FR values. Notably that selecting similar cubes will reduce the effectiveness of the MVAL, we manually discard the cubes with repeated FR, and pick up the cubes whose FR are larger than a certain threshold τ . Subsequently, the conditional mutual information is adopted to select the cubes with both high discriminability and low redundancy.

Fisher's Ratio		Orientation											
Frequency	ori1	ori2	ori3	ori4	ori5	ori6	ori7	ori8	ori9	ori10	ori11	ori12	ori13
1/4	3.0616	0.1413	0.1095	0.1413	0.0892	0.0928	0.0892	0.1371	0.1054	0.1371	0.0892	0.0928	0.0892
1/8	3.1888	0.6837	0.1484	0.6837	1.6423	0.2357	1.6423	0.7148	0.1537	0.7148	1.6423	0.2357	1.6423
1/12	3.2212	2.5911	0.9349	2.5911	3.2340	2.0412	3.2340	2.3789	0.9393	2.3789	3.2340	2.0412	3.2340
1/16	3.2272	3.3114	2.3374	3.3114	3.3089	3.1888	3.3089	2.9055	2.2034	2.9055	3.3089	3.1888	3.3089
1/20	3.2308	3.3984	3.1311	3.3984	3.2926	3.3117	3.2926	3.0538	2.7531	3.0538	3.2926	3.3117	3.2926

Table 1. FR value of 3D-Gabor cubes with different frequencies and orientations of Indian Pines dataset. "ori1" to "ori13" denote the 13 orientations constructed by each pair of { φ , θ } varying in { $0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$ }.

B. Conditional mutual information Conditional mutual information measurement is utilized to generate multiple views that are dissimilar with each other. Let $c = \{c_1, c_2, ..., c_r\}$ be the ground truth labels, $M_i^{x \times y \times \lambda}$ be the *i*th 3D-Gabor cube with three dimensions *x*, *y* and λ , and B_{bi} be the *b*th band of *i*th cube, the mutual information of M_i to *c* can be defined as

$$I(M_i, c) = \sum_{b=1}^{B} H(B_{bi}) - H(B_{bi}|c)$$
(5)

where $H(B_{bi})$ denotes the entropy of B_{bi} and $H(B_{bi}|c)$ is the conditional entropy of B_{bi} given c. Subsequently, we can measure the information shared by M_j and c given M_i according to the following conditional mutual information

$$I(M_j, c|M_i) = \sum_{b=1}^{B} H(B_{bj}|B_{bi}) - H(B_{bj}|c, B_{bi})$$
(6)

where $H(B_{bj}|B_{bi})$ denotes the conditional entropy of B_{bj} given B_{bi} , and $H(B_{bj}|c, B_{bi})$ refers to the entropy of B_{bj} given B_{bi} and c. It is noteworthy that the larger $I(M_j, c|M_i)$, the more dissimilar between M_i and M_i .

2.2. IEUE Query Selection

Before we propose our IEUE technology, we first introduce the multinomial logistic regression (MLR) classifier that we adopt to learn the prediction of classification. It is a discriminative classifier which can learn the useful features and has been proven to be very successful for HSI classification [20,52]. The MLR can directly models the posterior densities by using the following algorithm

$$P(y_i^{(k)} = 1 | x_i, \omega) \equiv \frac{exp(\omega^{(k)^T} h(x_i))}{\sum_{k=1}^{K} exp(\omega^{(k)^T} h(x_i))}$$
(7)

where $\omega \equiv [\omega^{(1)^T}, ..., \omega^{(K-1)^T}]^T$ denotes the regressors and $h(x_i) \equiv [h_1(x_i), ..., h_l(x_i)]^T$ is the spectral-spatial features [20]. The regressors are inferred by the logistic regression via variable splitting and augmented Lagrangian algorithm (LORSAL) [53]. In this paper, we use MLR to learn the posterior distributions and the prediction labels for query selection, and this classifier is also adopted to make the final outputs of the classification results.

Figure 3 provides a schematic illustration of how to calculate the *IEUE*. In order to fully exploit the discriminative information of both within-view and between-view, the *IEUE* of a certain candidate *x* can be defined as

$$IEUE(\mathbf{x}) = \frac{IU}{EU} \tag{8}$$

where *IU* and *EU* denote the internal and external uncertainty of views, respectively. On the one hand, we propose a multiview breaking ties criteria inspired by traditional breaking ties (BT) strategy [43] to estimate the *IU*. Let *c* be the ground truth labels, the *IU* is given as

$$IU(\mathbf{x}) = \max_{c \in \mathbf{c}} (P(c|\mathbf{x})) - \max_{c \in \mathbf{c} \setminus c^+} (P(c|\mathbf{x}))$$
(9)

where c^+ denotes the class label with the maximum probability and $P(\cdot)$ represents the comprehensive posterior probability of each class, which is the summation of the view-wise probability distributions

$$P(\mathbf{x}) = \sum_{i}^{k} p_{i}(\boldsymbol{c}|\boldsymbol{x})$$
(10)

where $p_i(c|x)$ denotes the independent distribution of the *i*th classifier. As shown in Equation (9), the IU focuses on the difference between the first and second maximum probabilities of classes. The candidate samples with smaller *IU* imply the more difficulty in labeling, therefore, those candidates are regarded as more informative, and vice versa. In this regard, the *IU* prefers the samples which lie on the boundary of two classes.



Figure 3. Schematic illustration of how to calculate the *IEUE*.

On the other hand, to evaluate the *EU*, a maximum disagreement criteria based on the disagreement of the prediction labels of different classifiers is adopted in this paper. Let $\{1, 2, ..., k\}$ be the classifier indexes and the *EU* is modeled as

$$EU(\mathbf{x}) = |unique(l_i(\mathbf{x}))|, i \in \{1, 2, ..., k\}$$
(11)

where $l_i(x)$ denotes the prediction label of the *i*th classifier for x and $|unique(\cdot)|$ is the number of unique prediction labels among classifiers. The samples which have higher EU are considered to be selected preferentially so as to guarantee that the chosen samples are the most controversial sets. Based on the above analysis, *IEUE* utilizes the contents of both posterior probability distributions and committee-based disagreement. We can rank the *IEUE* of candidate samples in an ascend manner, and the samples with minimum value are regarded as the most informative ones. Subsequently, the classifier selects the first few candidate samples with smaller *IEUE* and add them into the training set.

2.3. Output Strategy

To combine the classification results of individual classifiers, the *majority voting* is adopted in this paper. Since views are selected by cube assessment, each view is qualified to train the classifier independently. In *majority voting*, the classification labels of each view are treated equally. For a certain sample *x*, the algorithm adopts the label which emerges the most frequently among classifier hypothesis (i.e., owning the most votes) as its output label [37]. The final output of the prediction labels can be written as

$$Output(\mathbf{x}) = \underset{l_i(\mathbf{x}))}{\operatorname{arg\,max}} |count(l_i(\mathbf{x}))|, i \in \{1, 2, ..., k\}$$
(12)

where $|count(\cdot)|$ stands for counting the frequencies of output labels from each classifier l_i . As shown in Equation (12), *majority voting* is a very simple scheme to combine the results and it is appropriate when there are at least three views.

3. Experiments

In this section, we evaluate the performance of the proposed method on four benchmark hyperspectral datasets. For simplicity, we abbreviate our method as 3D-Gabor-IEUE. Extensive experiments are conducted to make the comprehensive comparisons with other state-of-the-art methods. Firstly, we make the comparison of the recently proposed methods, including multiview disagreement-intersection (MV3D-DisInt) based method [34], multiview disagreement-singularity (MV3D-DisSin) based method [34], Gabor-breaking ties (Gabor-BT) [43] and PCA-Gabor scheme [16]. To make more specifical comparison, we compare the 3D-Gabor-IEUE with other view generation and query selection methods as well. The comparison of view generation approaches is conducted using spectral information, 2D-Gabor [54] and the 3D-Gabor construction with no cube assessment named 3D-Gabor(no CA). The comparison of query selection is performed by maximum disagreement (MD) [34], entropy query-by-bagging (EQB) [24], adaptive maximum disagreement (AMD) [41], breaking ties (BT(SV)) [43] and randomly sampling (RS(SV)). It is notable that MD, EQB, AMD belong to the MVAL selection, while BT(SV) and RS(SV) are the query strategies of SVAL.

3.1. Data Description

Four publicly available HSI datasets are employed as benchmark sets, including the Indian Pines data, KSC data, University of Pavia data and University of Houston data. The details of four HSI datasets are described as follows.

1. Indian Pines data: this dataset was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over a mixed agricultural area in the Northwestern Indiana, USA, early in the growing season on 12 June 1992. Several bands affected by noise and water absorption phenomenon were removed from the original data, leaving a total of 200 channels of 10 nm width for experiments. This dataset contains 145×145 pixels with center wavelength from 0.4 to 2.5 µm. Figure 4 depicts the false color composition of the AVIRIS Indian Pines data as well as the map with 16 mutually exclusive ground truth classes. Since the number of samples is unbalanced and the spatial resolution is relatively low, the data poses a very challenging classification problem.



Figure 4. Indian Pines data. (**a**) Three band false color composite and (**b**) ground truth data with 16 classes.

2. **KSC data**: the KSC dataset was acquired by the AVIRIS sensor over the Kennedy Space Center, Florida, on 23 March 1996. The image contains a spatial coverage of 512×614 pixels, and has the spatial resolution of 18 m and 176 spectral bands in the wavelength range from 0.4 to 2.5 µm. Due to the noise and water-vapor absorption, 50 spectral bands were removed and 126 bands left. For classification purpose, 13 classes representing various land cover were manually defined. Figure 5 shows the false color composite and the corresponding ground truth map of the image.



Figure 5. KSC data. (a) Three band false color composite and (b) ground truth data with 13 classes.

3. **University of Pavia data**: this dataset was captured by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the urban area of the Pavia University, Italy, on 8 July 2002. The original data contains 115 spectral channels ranging from 0.43 to 0.86 μ m with spatial resolution of 1.3 m/pixel, and the pixels in this image is 610 × 340. The false color composite image together with the ground truth map are shown in Figure 6, where 9 classes of interest are provided for classification.



Figure 6. University of Pavia data. (**a**) Three band false color composite and (**b**) ground truth data with 9 classes.

4. University of Houston data: this dataset was acquired using the ITRES-CASI 1500 over the University of Houston campus and the neighboring urban area, on 23 June 2012. The image data has a spatial coverage of 349×1905 pixels with spatial resolution of 2.5 m and 144 spectral bands ranging from 0.38 to 1.05 µm. 15 identified classes are used for classification. Figure 7 plots the false color composite of the image and corresponding ground truth map.





Figure 7. University of Houston data. (**a**) Three band false color composite and (**b**) ground truth data with 15 classes.

3.2. Experimental Setup

In order to validate the framework of our 3D-Gabor-IEUE, we first compare it with other state-of-the-art algorithms including MV3D-DisInt, MV3D-DisSin, Gabor-BT and PCA-Gabor. The first three schemes belong to AL families and the last one is a non-AL classification method. MV3D-DisInt and MV3D-DisSin are derived from MD scheme. MV3D-DisInt focuses on reducing the redundancy of the selected training sets, while MV3D-DisSin prefers candidates which have higher spatial singularity. We use 3D-Gabor to construct the multiple views and sigularity maps instead of 3D-RDWT filter in [34]. Gabor-BT is a feature-driven SVAL algorithm which constructs high-dimension discriminative features to fully exploit the potential of AL. PCA-Gabor is a spectral-spatial response classification method based on PCA transform and Gabor filter.

Meanwhile, to specifically evaluate the performance of our proposed 3D-Gabor-IEUE, we compare it with other state-of-the-art methods from two aspects. First, we assess the view generation methods by fixing the query strategy as IEUE. Second, query selection methods are evaluated by using multiple views generated by 3D-Gabor. In greater detail, to assess the efficiency of our proposed 3D-Gabor-based method, we compare it with other view generation algorithms, i.e., Spec, 2D-Gabor and 3D-Gabor(no CA). Among those approaches, Spec denotes using the original spectral information without constructing multiple views, and it is equal to the SV with BT strategy. 2D-Gabor generates 20 cubes by the 2D-Gabor filter, and then the first several cubes with the largest FR value are selected as multiple views. 3D-Gabor(no CA) generates multiple views by randomly selecting a few cubes from the 65 Gabor cubes obtained by 3D-Gabor without cube assessment. For the sake of fairness, the frequencies, directions and the number of views are set to the same in the 2D-Gabor, 3D-Gabor(no CA) and 3D-Gabor. That means, the frequencies ω of the those methods are set to $\left\{\frac{1}{4}, \frac{1}{8}, \frac{1}{12}, \frac{1}{16}, \frac{1}{20}\right\}$, the orientations θ of the 2D-Gabor are set to $\left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$, the orientations $\{\varphi, \theta\}$ of the 3D-Gabor(no CA) and 3D-Gabor are identically set to $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$, and number of views is set to 8. Moreover, cube assessment is adopted in the 3D-Gabor method, whose threshold τ is set to 1 and the number of qualified cubes are 18, 21, 21 and 19 for the four datasets, respectively.

Furthermore, in order to evaluate the performance of our IEUE query strategy, we compare it with other widely-used query methods, i.e., MD, EQB, AMD and two SV based strategies (BT and RS). MD focuses mainly on the number of distinct prediction labels among classifiers, while EQB is based on the entropy of the distribution of classification predictions. AMD selects the samples which have the maximum disagreement degree between each classifier. BT estimates the information of samples by calculating the difference between two highest class-wise probabilities and RS randomly selects the new training samples. In both BT(SV) and RS(SV), we use the principle component analysis (PCA) for dimension reduction, to keep the same size of combining 8 views as the original dataset.

In all the above-mentioned methods, the MLR classifier is adopted to learn the posterior probability distributions of each view. The MLR is implemented via the variable splitting and augmented Lagrangian (LORSAL) algorithm and the parameters are set following the work in [53]. For each dataset, we randomly select 5 samples per class from each view to initialize the classifiers and the remainder of labeled samples are set to be the initial testing/candidate samples. (For PCA-Gabor method, we randomly select the same number of training samples as other methods in the final step.) The batch size of candidate samples acquired in each iteration is 3 and the iteration number is empirically defined. Detailed settings of the four datasets are displayed in Table 2.

Moreover, the aforementioned methods are compared quantitatively by four indexes, including overall accuracy (OA), average accuracy (AA), kappa coefficient (Kappa) and the accuracy of each class. Each experiment is conducted with ten independent Monte Carlo runs using random selection of initial training and testing/candidate samples.

Data	Size	No.Classes	Initial Training	Initial Candidates	Batch Size	No. Iteration
Indian Pines	145×145	16	5 (per class)	10286	3	80
KSC	512×614	13	5 (per class)	5146	3	25
University of Pavia	610×340	9	5 (per class)	42731	3	80
University of Houston	349×1905	15	5 (per class)	14954	3	60

Table 2. Experimental settings of the four hyperspectral datasets.

3.3. Experimental Results

3.3.1. Comparison with Spectral-Spatial Classification Methods

Four recently proposed spectral-spatial classification methods are compared with the proposed 3D-Gabor-IEUE. The experimental results of different methods are displayed in Tables 3–6. Two conclusions can be drawn from the results.

- First, 3D-Gabor-IEUE outperforms the competing algorithms all over the four datasets, with both higher classification accuracy and lower standard deviations. For instance, in Table 3 of Indian Pines data, the OA of 3D-Gabor-IEUE achieves 99.57%, which is 2.72, 3.1%, 1.13% and 17.42% higher than another four state-of-the-art algorithms, respectively. In Tables 4 and 5, the proposed method obtains significantly higher value of OA, AA and Kappa than the MV3D-DisInt, MV3D-DisSin and PCA-Gabor, and slightly better than Gabor-BT. This observation can also be revealed from Figures 8 and 9, where 3D-Gabor-IEUE performs the best with the least misclassified samples. The impressive performance of 3D-Gabor-IEUE indicates its superiority in spectral-spatial HSI classifiation.
- Second, using the same amount of training datasets, AL based methods (i.e., MV3D-DisInt, MV3D-DisSin and Gabor-BT) achieves better results than non-AL scheme (i.e., PCA-Gabor). Specifically, the OAs of AL-methods are at least 14.32%, 6.21%, 3.27% and 17.84% higher than PCA-Gabor in Indian Pines data, KSC data, the University of Pavia data and the University of Houston data, respectively. It is observed that AL technology can perform well with very limited training samples. For instance, in Table 6, the 3D-Gabor-IEUE achieves the OA of 98.82% with only 255 training samples (5 for initialization and 180 are iteratively selected). Those results demonstrate the effectiveness of AL scheme.

3.3.2. Comparison with View Generation Methods

Three view generation methods are compared with the proposed 3D-Gabor by fixing the query strategy as IEUE in this paper. The classification accuracies of different methods are shown in Tables 3–6 and the classification maps are displayed in Figures 8 and 9. Two observations can be revealed from the experimental results.

• First, the 3D-Gabor-based feature extraction method outperforms the other applied algorithms on the four datasets. The encouraging improvements of classification accuracies confirms that the 3D-Gabor is powerful to extract discriminative information by obeying the 3D nature of the HSI data. On the contrary, the Spec and 2D-Gabor provide higher classification errors than the 3D-Gabor-based methods. It is observed from Table 5 that the OA of Spec is 8.47% and 9.62% lower than 3D-Gabor(no CA) and 3D-Gabor for the University of Pavia dataset, respectively. Similar properties of OA, AA and Kappa are also shown from the other results. The undesirable performance of original spectral data is plausible since the spectral data disregards the critical spatial information. Moreover, 2D-Gabor performs worse than the two 3D-Gabor-based methods in all of the datasets, suggesting that the spatial information can not be readily extracted with 2D-Gabor. When compared to original spectral data, it is even worse as unexpected in the University of Pavia data (see Table 5) and University of Houston data (see Table 6). This surprising finding indicates that the spatial information extracted by 2D-Gabor might contain many unwanted signals.



Figure 8. Classification maps of the Indian Pines data with a final of 320 training samples.



Figure 9. Classification maps of the University of Pavia data with a final of 285 training samples.

• Second, 3D-Gabor with the cube assessment (i.e., 3D-Gabor-IEUE) provides better or comparable performance than 3D-Gabor(no CA). For instance, as shown from Tables 3 and 5, the OA of 3D-Gabor is substantially higher than that of the 3D-Gabor(no CA). The reasons for better performance of 3D-Gabor lies in two important aspects, i.e., the cube assessment can successfully reject cascade of underqualified cubes by utilizing FR measurement, and can generate distinct views by conditional mutual information restriction. Both of two aspects lead to more reliable classification results.

3.3.3. Comparison with Query Selection Methods

Five query selection methods are compared by fixing the view generation as 3D-Gabor, to demonstrate the effectiveness of the proposed IEUE strategy. The quantitative results are displayed in Tables 3–6. Two conclusions can be obtained from the experimental results.

- First, the proposed IEUE performs best as comparing to other methods (including both MV-based and SV-based families) for the four datasets in terms of the highest accuracies (see 3D-Gabor-IEUE). It can further be observed from Table 3 that the OA, AA and Kappa of IEUE are much higher than those of the MD, AMD, BT(SV) and RS(SV), and slightly higher than the EQB. The experimental results displayed in Table 3 also demonstrate that our method is more stable, as evidenced by the smaller standard deviation. Classification performance of the KSC (see Table 4), the University of Pavia (see Table 5) and the University of Houston (see Table 6) also yield the similar properties. Specifically, the separation of "graminoid marsh" (i.e., class 8) and "spartina marsh" (i.e., class 9) from other classes is difficult in Table 4 of the KSC dataset, and the classification accuracies of those two classes by IEUE outperform other methods significantly. Unlike the most of the existing MVAL query strategies which only focus on the between-view inconsistencies, the main reason for the best performance of IEUE is that it takes both internal and external uncertainty into consideration. Therefore, the proposed IEUE method can provide more discriminative information for selecting the most informative samples.
- Second, the MVAL methods (i.e., MD, EQB, AMD and IEUE) can almost achieve better results than SVAL families (BT(SV), RS(SV)) for the four HSI datasets. This observation can be quantitatively concluded according to Tables 3 and 4. The main reason for the superior performance of MVAL is that this kind of method combines the results of each classifier, thus reducing the number of misclassified samples and leading to better classification performance.

3.3.4. Assessment of Selected Samples

It is known that in AL classification, the discriminative information carried by uncertain samples is significantly larger than others. In order to explore the sampling effectiveness of the proposed IEUE method, we display the selected samples and the initial class-wise accuracies for the four HSI datasets in Figure 10. However, the number of samples for different classes is different. To solve the problem, we adopt the ratios of the numbers to the corresponding classes for illustration. It can be observed that the IEUE prefers the samples from the classes which are difficult to classify. For instance, class "corn-min till" (class 3) is identified to be the difficult class in Indian Pines dataset (see Figure 10a), since the initial overall accuracy of this class is low as compared to other classes and the IEUE is inclined to choose more samples which belong to this class. On the contrary, as can be seen in Figure 10b, the IEUE selects the least samples in class "cattail marsh" (class 6) and class "water" (class 10), since these classes achieve the highest two accuracies and they are easy to classify. Similar trends for the other two datasets can be found in Figure 10c,d. In a nutshell, the preference to the informative samples which belong to difficult classes the sampling effectiveness of our IEUE query strategy.

Table 3. Classification accuracies [%] of different view generation and sampling selection methods for the Indian Pines dataset, using 80 initial training samples (5 per class) and 240 selected samples. For PCA-Gabor, 320 samples (20 per class) are used for training. The numbers in parenthesis are standard deviations. Bold values indicate the best result for a row.

Class	MU2D Distant	MV2D Disfin	Cabor PT	PCA Cabor	2D Cabor IELIE	IEUE					3D-Gabor			
Class	WIV 3D-DISIN	WV3D-DISSIN	Gabor-DI	rCA-Gabor	5D-Gabor-IEUE	Spec	2D-Gabor	3D-Gabor (No CA)	MD	EQB	AMD	BT (SV)	RS (SV)	
C1	99.79	99.58	95.85	98.82	98.75	76.83	93.47	93.62	99.59	97.11	98.95	94.39	87.93	
C2	99.46	99.54	98.27	72.83	99.66	77.56	96.80	96.85	98.87	99.09	99.19	94.16	81.88	
C3	99.09	99.10	98.67	72.15	99.69	64.11	97.68	94.65	96.46	97.36	98.86	94.51	75.19	
C4	99.04	98.26	99.01	95.61	99.16	59.63	96.61	96.07	97.97	99.55	99.46	92.42	79.76	
C5	98.44	98.07	92.93	93.60	99.67	91.02	96.56	96.82	98.24	98.37	98.84	95.63	91.47	
C6	99.35	99.63	99.04	98.35	99.55	92.99	96.97	98.95	99.15	99.48	99.75	99.22	98.01	
C7	100.00	99.52	97.62	100.00	96.64	93.02	98.57	97.62	99.52	97.62	98.10	99.23	98.83	
C8	100.00	99.98	99.96	97.96	99.75	97.50	98.83	99.75	99.71	99.25	99.94	99.59	98.42	
C9	99.33	98.67	100.00	100.00	99.33	93.00	94.57	99.33	100.00	99.33	98.62	99.50	99.50	
C10	94.93	93.89	98.81	77.84	99.37	69.77	92.43	92.96	95.40	98.65	99.07	90.12	79.10	
C11	94.75	93.79	98.89	70.61	99.70	81.11	96.52	96.77	96.64	98.94	97.13	95.35	85.24	
C12	97.74	96.07	98.68	81.30	98.98	74.69	95.70	96.10	97.20	96.82	96.84	92.56	78.35	
C13	99.46	98.39	99.81	99.67	99.81	98.77	91.18	99.47	99.22	97.78	99.71	99.62	98.94	
C14	95.86	96.59	99.14	94.10	99.91	93.28	96.90	98.42	95.12	99.73	96.60	94.83	96.35	
C15	90.05	89.16	95.31	92.63	99.29	61.10	97.60	91.87	90.41	99.40	89.80	95.63	84.55	
C16	90.56	91.92	97.51	98.53	96.03	71.91	90.93	95.28	91.22	96.64	93.14	93.04	86.78	
OA	96.85	96.47	98.44	82.15	99.57	80.63	96.28	96.57	96.89	98.78	97.86	94.93	86.43	
	(0.33)	(0.83)	(1.08)	(1.08)	(0.14)	(0.92)	(0.54)	(0.41)	(0.42)	(0.36)	(0.31)	(0.43)	(0.89)	
AA	97.37	97.01	98.09	90.25	99.08	81.02	95.71	96.53	97.17	98.45	97.75	95.61	88.77	
	(0.36)	(0.71)	(1.40)	(0.74)	(0.43)	(1.25)	(0.90)	(0.43)	(0.57)	(0.91)	(0.82)	(0.55)	(1.19)	
Kappa	96.42	95.98	98.22	79.86	99.51	77.85	95.76	96.09	96.46	98.61	97.57	94.22	84.55	
	(0.38)	(0.95)	(1.23)	(1.22)	(0.16)	(1.04)	(0.62)	(0.47)	(0.48)	(0.41)	(0.36)	(0.49)	(1.03)	

Table 4. Classification accuracies [%] of different view generation and sampling selection methods for the KSC dataset, using 65 initial training samples (5 per class) and 75 selected samples. For PCA-Gabor, 143 samples (11 per class) are used for training. The numbers in parenthesis are standard deviations. Bold values indicate the best result for a row.

<u></u>	MU2D District	MU2D Di-Sir	Calar PT	DCA Caltar	3D-Gabor-IEUE –		IE	UE			3D-Gabor			
Class	WV3D-DISINT	MV3D-D1551n	Gabor-DI	I CA-Gaboi		Spec	2D-Gabor	3D-Gabor (No CA)	MD	EQB	AMD	BT (SV)	RS (SV)	
C1	100.00	100.00	100.00	86.67	100.00	71.15	99.95	99.95	100.00	99.77	100.00	98.85	72.25	
C2	99.83	100.00	100.00	71.98	100.00	71.56	100.00	100.00	99.41	99.83	98.29	95.75	63.54	
C3	95.35	95.73	96.17	86.82	98.12	72.22	99.87	99.87	92.91	97.84	97.50	93.93	82.90	
C4	99.96	99.68	100.00	97.80	99.92	90.10	99.76	100.00	99.88	99.80	99.92	98.91	97.01	
C5	97.14	97.75	100.00	87.07	99.99	92.81	99.92	99.85	98.09	99.91	98.33	99.23	98.20	
C6	100.00	100.00	100.00	97.26	100.00	87.88	99.92	100.00	100.00	100.00	100.00	99.77	97.59	
C7	99.98	99.84	100.00	93.19	99.96	93.23	97.59	99.98	99.98	99.92	99.98	99.75	95.76	
C8	99.45	99.83	100.00	78.49	99.95	90.53	99.01	99.95	99.14	99.98	99.98	97.25	78.98	
C9	93.21	93.54	100.00	96.48	99.98	93.68	99.33	99.71	96.55	95.09	95.23	95.62	93.47	
C10	100.00	100.00	100.00	99.96	100.00	99.87	99.97	100.00	100.00	100.00	100.00	99.95	100.00	
C11	95.36	96.17	100.00	95.76	100.00	83.38	99.61	100.00	96.80	99.96	96.17	94.07	89.58	
C12	96.06	95.11	99.98	93.77	99.98	92.42	99.66	99.95	97.10	99.83	96.68	99.71	96.28	
C13	96.00	94.70	100.00	98.10	99.99	88.75	99.70	100.00	95.49	100.00	96.28	99.49	97.16	
OA	98.98	98.16	99.88	91.95	99.93	90.46	99.52	99.94	98.57	99.37	98.58	98.37	92.27	
	(0.84)	(0.75)	(0.24)	(1.60)	(0.18)	(1.24)	(0.42)	(0.06)	(0.53)	(0.81)	(0.77)	(0.46)	(1.34)	
AA	97.87	97.87	99.70	91.03	99.83	86.74	99.56	99.94	98.10	99.38	98.34	97.87	89.44	
	(0.72)	(0.83)	(0.62)	(1.78)	(0.45)	(1.12)	(0.32)	(0.05)	(0.58)	(0.76)	(0.72)	(0.91)	(1.94)	
Kappa	97.86	97.95	99.87	91.06	99.92	89.55	99.47	99.93	98.41	99.30	98.43	98.19	91.39	
	(0.93)	(0.83)	(0.27)	(1.78)	(0.20)	(1.39)	(0.47)	(0.07)	(0.58)	(0.74)	(0.86)	(0.52)	(1.49)	

Table 5. Classification accuracies [%] of different view generation and sampling selection methods for the University of Pavia dataset, using 45 initial training samples (5 per class) and 240 selected samples. For PCA-Gabor, 288 samples (32 per class) are used for training. The numbers in parenthesis are standard deviations. Bold values indicate the best result for a row.

Class		MV3D-DisSin	Calar PT	PCA-Gabor	3D-Gabor-IEUE -		IE	UE		3D-Gabor				
Class	MV3D-Disint		Gador-d I			Spec	2D-Gabor	3D-Gabor (No CA)	MD	EQB	AMD	BT (SV)	RS (SV)	
C1	99.77	99.73	98.83	75.03	99.07	85.27	79.94	98.32	98.89	98.15	98.72	94.98	81.05	
C2	94.52	93.89	99.91	92.09	99.85	96.70	89.41	99.52	91.20	99.80	90.75	99.79	98.97	
C3	95.46	93.20	95.35	86.70	96.36	72.65	74.79	93.20	89.78	95.28	93.39	89.47	72.25	
C4	95.30	95.21	98.70	93.50	97.46	83.79	42.31	96.68	96.25	98.01	97.22	97.82	90.72	
C5	99.86	99.85	99.46	99.58	99.92	92.13	94.92	99.72	99.93	99.80	99.96	99.16	98.06	
C6	82.82	81.73	99.82	96.95	99.71	80.13	78.33	97.96	86.98	99.81	82.83	99.65	96.51	
C7	94.08	94.68	97.45	92.89	98.46	78.49	77.36	92.93	96.77	96.25	96.10	92.37	79.23	
C8	85.56	87.47	97.65	89.23	97.46	86.89	72.36	94.03	96.14	95.44	94.22	96.34	91.01	
C9	89.82	91.69	91.03	86.63	94.97	99.17	30.37	94.97	94.10	89.81	93.05	67.91	65.83	
OA	93.32	93.02	98.94	89.75	99.02	89.40	79.60	97.87	93.11	98.50	92.43	97.15	91.94	
	(1.15)	(1.44)	(0.19)	(1.53)	(0.09)	(0.65)	(0.95)	(0.19)	(2.37)	(0.13)	(2.23)	(0.36)	(1.03)	
AA	93.02	93.05	97.58	90.29	98.14	86.14	71.09	96.37	94.45	96.93	94.02	93.05	85.96	
	(0.95)	(1.12)	(0.83)	(0.77)	(0.23)	(1.25)	(1.45)	(0.41)	(1.10)	(0.46)	(0.99)	(1.55)	(2.49)	
Kappa	91.15	90.75	98.60	86.62	98.70	85.83	72.60	97.17	90.96	98.00	90.06	96.20	89.21	
	(1.51)	(1.88)	(0.25)	(1.90)	(0.13)	(0.90)	(1.23)	(0.25)	(3.00)	(0.17)	(2.85)	(0.49)	(1.42)	

Table 6. Classification accuracies [%] of different view generation and sampling selection methods for the University of Houston dataset, using 75 initial training samples (5 per class) and 180 selected samples. For PCA-Gabor, 255 samples (17 per class) are used for training. The numbers in parenthesis are standard deviations. Bold values indicate the best result for a row.

Class	MV2D District	MW2D DicSin	Cabor PT	PCA Cabor	3D-Gabor-IEUE —		IE	UE	3D-Gabor				
Class	WV3D-DISIM	WV3D-DISSIN	DISSIII Gabor-DI I CA-Gabo		3D-Gabor-IEUE	Spec	2D-Gabor	3D-Gabor (No CA)	MD	EQB	AMD	BT (SV)	RS (SV)
C1	95.49	98.49	98.79	87.50	98.59	93.23	67.07	98.59	97.01	98.38	96.53	96.21	96.33
C2	97.07	96.61	98.07	84.96	98.44	90.85	54.62	99.57	97.19	97.47	97.04	98.83	93.15
C3	99.84	99.87	99.96	92.56	99.91	99.23	76.21	99.91	99.88	99.61	99.83	98.17	91.76
C4	95.28	92.53	97.01	82.44	98.81	93.32	51.17	98.95	93.59	99.69	96.41	97.80	94.11
C5	99.97	99.85	99.90	99.42	99.97	96.91	69.54	99.99	99.85	99.94	99.91	99.76	99.76
C6	100.00	99.74	95.97	75.23	98.44	85.33	75.92	98.09	99.64	94.89	98.65	72.62	67.16
C7	99.50	99.22	98.29	56.62	99.64	81.63	56.01	99.28	98.51	95.85	98.63	84.49	66.27
C8	97.78	97.73	95.18	56.71	97.99	56.66	63.97	94.15	95.54	92.75	96.72	84.29	71.13
C9	96.78	96.49	98.82	66.36	98.74	82.08	47.72	98.14	95.39	96.99	96.65	93.99	74.69
C10	91.25	90.20	99.73	70.27	99.78	84.45	64.21	99.90	93.49	99.98	92.71	92.10	86.64
C11	98.20	95.25	98.85	67.53	99.16	86.52	56.06	99.37	97.60	99.35	98.37	86.40	80.08
C12	83.47	80.76	93.79	73.91	97.63	76.85	50.84	95.56	84.24	96.09	90.56	90.24	79.03
C13	89.60	88.48	95.12	76.14	93.74	32.83	51.25	94.41	94.63	92.72	90.58	82.87	70.98
C14	96.85	96.75	99.93	98.43	99.93	95.14	84.32	99.95	99.29	99.81	97.56	98.55	96.18
C15	93.68	95.20	99.97	98.32	99.74	97.54	50.87	99.91	95.48	99.88	98.10	99.25	99.63
OA	95.57	94.96	97.99	77.12	98.82	84.33	59.55	98.43	95.67	97.69	96.51	92.47	84.69
	(0.80)	(1.27)	(0.64)	(1.20)	(0.47)	(2.21)	(1.18)	(0.75)	(0.88)	(0.30)	(1.08)	(1.16)	(1.43)
AA	95.65	95.15	97.97	79.09	98.70	83.50	61.32	98.39	96.09	97.56	96.55	91.70	84.46
	(0.79)	(1.30)	(0.55)	(1.08)	(0.43)	(1.83)	(1.67)	(0.66)	(0.78)	(0.33)	(0.86)	(1.16)	(1.62)
Kappa	95.21	94.55	97.83	75.27	98.72	83.05	56.17	98.29	95.32	97.50	96.22	91.85	83.44
	(0.86)	(1.38)	(0.77)	(1.30)	(0.50)	(2.38)	(1.28)	(0.81)	(0.95)	(0.33)	(1.16)	(1.26)	(1.55)



Figure 10. The ratios of the selected samples to the corresponding classes of the IEUE method (see the histograms) and the initial class-wise accuracies (see the accuracy curves) for the four datasets. (a) Indian Pines, (b) KSC, (c) University of Pavia, (d) University of Houston.

3.3.5. Analysis of Computational Complexity and Learning Rate

Taking the the learning speed into consideration, we show the required elapsed time of three query algorithms for the Indian Pines and the KSC data in Table 7. The hardware that we utilize for the experiments is Intel CPU, 3.40 GHZ. From the result table, the MD and IEUE is faster than EQB, while the time of MD is slightly less than the proposed IEUE. It is due to the fact that IEUE operates the probability subtraction based on MD algorithm. Moreover, the time of EQB is relatively higher. It is because that EQB operates the calculation of the entropy for each candidate. Moreover, we display the learning curves of the above-mentioned three methods in Figure 11 to compare the learning rates. It can be observed that the proposed IEUE converges significantly faster with higher accuracies. For KSC data, although both the IEUE and EQB can achieve 100% accuracies finally, the training steps are fewer for IEUE. Those observations demonstrate the learning efficiency of the proposed method.

Table 7. CPU processing time (in seconds) of different AL selection methods for the Indian Pines and KSC data.

Methods	MD	EQB	IEUE
Indian Pines	272.61	287.47	274.61
KSC	651.79	684.55	660.11



Figure 11. Learning curves (i.e., OA and standard deviation versus the query steps) of different query selection methods for (**a**) Indian Pines and (**b**) KSC.

4. Discussion

According to the above-mentioned experiments, the proposed 3D-Gabor-IEUE outperforms almost all the applied methods, especially the experiments using only the spectral information. From the Spec to 3D-Gabor-IEUE, the improvements of OA of the four datasets are 19.21%, 9.47%, 9.62% and 14.49%, respectively. This illustrates that the 3D-Gabor-IEUE method is a reliable spectral-spatial classification method. Furthermore, it seems the proposed method has potential to deal with the challenging data like Indian Pines (with the spatial resolution of 30 m and 200 channels). The main reason of the significant improvements of Indian pines are two folds. First, the original HSI with only spectral characteristics are often mixed. Samples from different classes tend to be overlapped, especially when then spatial resolution is limited. In addition, since the contiguous bands in HSI data are highly correlated, the differential information between each band can not be fully exploited. The proposed IEUE technology is based on the posterior probability distributions of each classifier. Therefore, it is necessary to construct the feature space where the class boundaries are clearer. In this paper, the 3D-Gabor with smoothness in spatial domain and discrimination in spectral domain is used, leading to a more discriminative feature space as comparing to the original data. The 3D-Gabor features significantly boost the performance of IEUE. Second, the samples queried by 3D-Gabor-IEUE are very informative. Adding those samples to the training sets can improve the performance quickly. Therefore the accuracies are significantly improved with very limited training data.

5. Conclusions

In this paper, we have presented a 3D-Gabor inspired MVAL framework (i.e., 3D-Gabor-IEUE) for spectral-spatial HSI classification. The proposed method consists of two major steps. The first one is the view generation approach based on 3D-Gabor feature extraction and cube assessment criterion. The main advantage of our view generation is that it not only provides sufficient information for independent classification, but also satisfies the requirement of diversity between views. Compared to the 2D-Gabor and original HSI data, the class separating capability is significantly increased with 3D-Gabor transformation. The second one is the proposed IEUE method for sampling selection. Using the posterior probability distribution and the disagreement between views, this strategy comprehensively estimates the uncertainty on both internal and external aspects. Compared to the traditional SVAL and disagreement-based MVAL scheme, samples are evaluated in an more discriminative manner by using the IEUE and the highly informative candidates can be accurately selected. The effectiveness of the proposed method is evaluated on four AVIRIS and ROSIS datasets. Quantitatively, the OA of the 3D-Gabor-IEUE improves at most 22% compared to other state-of-the-art methods. In the future, we will fully exploit the internal information within each view to make more

convincing estimation for sampling selection. Furthermore, combining other data sources in generating views and using other HSI data for experiments is also a probable future research direction.

Author Contributions: All coauthors made significant contributions to the manuscript. J.H., Z.H. and J.L. designed the research framework, analyzed the results and wrote the paper. L.H. programmed the experiments. Y.W. assisted in the prepared work and validation work. Moreover, all coauthors contributed to the editing and review of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 41501368, 61571195, and the Fundamental Research Funds for the Central Universities under Grant 161gpy04.

Acknowledgments: The authors would like to thank David A. Landgrebe from Purdue University for providing the AVIRIS image of Indian Pines dataset; Paolo Gamba from University of Pavia for providing the ROSIS University of Pavia dataset and the Hyperspectral Image Analysis group and the NSF-NCALM for providing the Houston University dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Chen, C.; Li, W.; Su, H.; Liu, K. Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine. *Remote Sens.* **2014**, *6*, 5795–5814. [CrossRef]
- 2. Sun, W.; Jiang, M.; Li, W.; Liu, Y. A symmetric sparse representation based band selection method for hyperspectral imagery classification. *Remote Sens.* **2016**, *8*, 238. [CrossRef]
- 3. Feng, F.; Li, W.; Du, Q.; Zhang, B. Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity. *Remote Sens.* **2017**, *9*, 323. [CrossRef]
- 4. Su, H.; Zhao, B.; Du, Q.; Du, P.; Xue, Z. Multifeature dictionary learning for collaborative representation classification of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2467–2484. [CrossRef]
- 5. Awad, M. Forest mapping: A comparison between hyperspectral and multispectral images and technologies. *J. For. Res.* **2017**. [CrossRef]
- Awad, M.; Jomaa, I.; Arab, F. Improved capability in stone pine forest mapping and management in lebanon using hyperspectral CHRIS PROBA data relative to Landsat ETM+. *Photogramm. Eng. Remote Sens.* 2014, 80, 724–731. [CrossRef]
- 7. Liang, H.; Li, Q. Hyperspectral imagery classification using sparse representations of convolutional neural network features. *Remote Sens.* **2016**, *8*, 99. [CrossRef]
- Sun, W.; Yang, G.; Du, B.; Zhang, L.; Zhang, L. A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* 2017, 55, 4032–4046. [CrossRef]
- 9. Zhao, C.; Wang, Y.; Qi, B.; Wang, J. Global and local real-time anomaly detectors for hyperspectral remote sensing imagery. *Remote Sens.* 2015, *7*, 3966–3985. [CrossRef]
- 10. Sun, W.; Ma, J.; Yang, G.; Du, B.; Zhang, L. A Poisson nonnegative matrix factorization method with parameter subspace clustering constraint for endmember extraction in hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *128*, 27–39. [CrossRef]
- 11. Awad, M. Sea water chlorophyll—A estimation using hyperspectral images and supervised Artificial Neural Network. *Ecol. Inform.* **2014**, *24*, 60–68. [CrossRef]
- 12. Chen, J.; Xia, J.; Du, P.; Chanussot, J.; Xue, Z.; Xie, X. Kernel supervised ensemble classifier for the classification of hyperspectral data using few labeled samples. *Remote Sens.* **2016**, *8*, 601. [CrossRef]
- Yu, H.; Gao, L.; Li, J.; Li, S.S.; Zhang, B.; Benediktsson, J.A. Spectral-spatial hyperspectral image classification using subspace-sased support sector machines and adaptive markov random fields. *Remote Sens.* 2016, *8*, 355. [CrossRef]
- Sun, W.; Halevy, A.; Benedetto, J.J.; Czaja, W.; Liu, C.; Wu, H.; Shi, B.; Li, W. UL-Isomap based nonlinear dimensionality reduction for hyperspectral imagery classification. *ISPRS J. Photogramm. Remote Sens.* 2014, 89, 25–36. [CrossRef]
- 15. Yuan, Y.; Zhu, G.; Wang, Q. Hyperspectral band selection by multitask sparsity pursuit. *IEEE Trans. Geosci. Remote Sens.* **2015**, 53, 631–644. [CrossRef]
- 16. Wei, Y.; Zhou, Y.; Li, H. Spectral-spatial ersponse for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 203. [CrossRef]

- 17. He, L.; Li, J.; Liu, C.; Li, S. Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1579–1597. [CrossRef]
- 18. Li, C.; Wang, J.; Wang, L.; Hu, L.; Gong, P. Comparison of classification algorithms and training sample sizes in urban land cassification with landsat thematic mapper imagery. *Remote Sens.* **2014**, *6*, 964–983. [CrossRef]
- 19. He, Z.; Liu, H.; Wang, Y.; Hu, J. Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 1042. [CrossRef]
- 20. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 844–856. [CrossRef]
- 21. Tan, K.; Zhu, J.; Du, Q.; Wu, L.; Du, P. A novel tri-training technique for semi-supervised classification of hyperspectral images based on diversity measurement. *Remote Sens.* **2016**, *8*, 749. [CrossRef]
- 22. Sun, S.; Zhong, P.; Xiao, H.; Wang, R. Active learning with gaussian process classifier for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1746–1760. [CrossRef]
- Huang, S.J.; Jin, R.; Zhou, Z.H. Active learning by querying informative and representative examples. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 6–11 December 2010; pp. 892–900.
- 24. Tuia, D.; Volpi, M.; Copa, L.; Kanevski, M.; Munoz-Mari, J. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 606–617. [CrossRef]
- 25. Luo, T.; Kramer, K.; Goldgof, D.B.; Hall, L.O.; Samson, S.; Remsen, A.; Hopkins, T. Active learning to recognize multiple types of plankton. *J. Mach. Learn. Res.* **2005**, *6*, 589–613.
- 26. MacKay, D.J.C. Information-based objective functions for active data selection. *Neural Comput.* **1992**, *4*, 590–604. [CrossRef]
- Jun, G.; Ghosh, J. An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis. In Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Boston, MA, USA, 7–11 July 2008; pp. 52–55.
- 28. Rajan, S.; Ghosh, J.; Crawford, M.M. An active learning approach to hyperspectral data classification. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1231–1242. [CrossRef]
- Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the 15th Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
- Scheffer, T.; Decomain, C.; Wrobel, S. Active hidden markov models for information extraction. In Proceedings of the 4th International Symposium on Intelligent Data Analysis (IDA), Cascais, Portugal, 13–15 September 2001; pp. 309–318.
- 31. Demir, B.; Persello, C.; Bruzzone, L. Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1014–1031. [CrossRef]
- 32. Pasolli, E.; Melgani, F.; Bazi, Y. Support vector machine active learning through significance space construction. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 431–435. [CrossRef]
- 33. Tuia, D.; Ratle, F.; Pacifici, F.; Kanevski, M.F.; Emery, W.J. Active learning methods for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* 2009, 47, 2218–2232. [CrossRef]
- 34. Zhou, X.; Prasad, S.; Crawford, M.M. Wavelet-domain multiview active learning for spatial-spectral hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4047–4059. [CrossRef]
- 35. Di, W.; Crawford, M.M. Active learning via multi-view and local proximity co-regularization for hyperspectral image classification. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 618–628. [CrossRef]
- Crawford, M.M.; Tuia, D.; Yang, H.L. Active learning: any value for classification of remotely sensed data? *Proc. IEEE* 2013, 101, 593–608. [CrossRef]
- 37. Muslea, I.; Minton, S.; Knoblock, C.A. Active learning with multiple views. J. Artif. Intell. Res. 2006, 27, 203–233.
- Zhao, J.; Xie, X.; Xu, X.; Sun, S. Multi-view learning overview: Recent progress and new challenges. *Inf. Fusion* 2017, 38, 43–54. [CrossRef]
- 39. Sun, S. A survey of multi-view machine learning. Neural Comput. Appl. 2013, 23, 2031–2038. [CrossRef]
- 40. Sun, S.; Jin, F.; Tu, W. View construction for multi-view semi-supervised learning. In Proceedings of the International Symposium on Neural Networks (ISSN), Guilin, China, 29 May–1 June 2011; pp. 595–601.

- 41. Di, W.; Crawford, M.M. View generation for multiview maximum disagreement based active learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1942–1954. [CrossRef]
- 42. Xu, X.; Li, J.; Li, S. Multiview intensity-based active learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 669–680. [CrossRef]
- 43. Liu, C.; He, L.; Li, Z.; Li, J. Feature-driven active learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 341–354. [CrossRef]
- 44. Li, W.; Prasad, S.; Fowler, J.E.; Bruce, L.M. Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification. *IEEE Signal Process. Lett.* **2011**, *8*, 894–898. [CrossRef]
- 45. Sotoca, J.M.; Pla, F. Supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recognit.* **2010**, *43*, 2068–2081. [CrossRef]
- 46. Liu, C.; Wechsler, H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Process.* **2002**, *11*, 467–476. [PubMed]
- 47. Riaz, F.; Hassan, A.; Rehman, S.; Qamar, U. Texture classification using rotation- and scale-invariant gabor texture features. *IEEE Signal Process. Lett.* **2013**, *20*, 607–610. [CrossRef]
- 48. Li, J.; Narayanan, R.M. Integrated spectral and spatial information mining in remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 673–685. [CrossRef]
- 49. He, L.; Li, J.; Plaza, A.; Li, Y. Discriminative low-Rank gabor Filtering for spectral-spatial hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1381–1395. [CrossRef]
- 50. He, L.; Li, Y.; Li, X.; Wu, W. Spectral-spatial classification of hyperspectral images via spatial translation-invariant wavelet-based sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2696–2712. [CrossRef]
- 51. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE* **2013**, *101*, 652–675. [CrossRef]
- 52. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 318–322.
- 53. Bioucas-Dias, J.; Figueiredo, M. *Logistic Regression via Variable Splitting and Augmented Lagrangian Tools;* Instituto Superior Técnico: Lisboa, Portugal, 2009.
- 54. Haghighat, M.; Zonouz, S.; Abdel-Mottaleb, M. CloudID: Trustworthy cloud-based and cross-enterprise biometric identification. *Expert Syst. Appl.* **2015**, *42*, 7905–7916. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).