IN DEFENSE OF ITERATED CONDITIONAL MODE FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Jianzhe Lin[1,3], *Qi Wang*[2], *Yuan Yuan*[1]

[1]Center for OPTical IMagery Analysis and Learning (OPTIMAL),
State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics,
Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P.R. China.
[2]Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P.R. China.
[3] Graduate University of the Chinese Academy of Science, 19A Yuquanlu, Beijing 100049, P.R. China.
linjianzhe@opt.cn; crabwq@gmail.com; yuany@opt.ac.cn.

ABSTRACT

Hyperspectral image classification is one of the most significant topics in remote sensing. A large number of methods have been proposed to improve the classification accuracy. However, the improvement often comes at the cost of higher complexity. In this work, we mainly focus on the Markov Random Fields related paradigm, which involves a demanding energy minimization procedure. Traditional methods are prone to employ the advanced optimization techniques. On the contrary, this paper is in defense of a simple yet efficient method for hyperspectral image classification, Iterated Conditional Mode, which has been generally considered inferior to other state-of-the-art methods. Our purpose is successfully achieved by tackling two inherent drawbacks of ICM, sensitive label initialization and local minimum. We apply our method to three real-world hyperspectral images, and compare the results with those of state-of-the-art methods. The comparisons show that the proposed method outperforms its competitors.

Index Terms— Iterated conditional mode, hyperspectral image classification, support vector machine

1. INTRODUCTION

Hyperspectral image is collected by the hyperspectral imaging sensor and often consists of hundreds to thousands of electromagnetic spectrums, each of which represents a characteristic of a particular physical property. The rich information is beneficial to a wide range of applications, such as object detection, recognition and classification[1], [2], [3]. Among these applications, hyperspectral image classification is possibly one of the most frequently researched topics. The classification results can give the ability to pinpoint ground objects laying on the observed surface, and to distinguish between spatially close ground classes. Therefore, constructing a proper method to support high quality classification is of great importance in the remote sensing field.

Existing hyperspectral classification methods can be roughly divided into two classes: supervised and unsupervised. The most distinctive difference between supervised and unsupervised classifiers is the existence of learning phase on training samples, which may bring some advantage to supervised classifiers. In this work, we focus on supervised classification. As for supervised hyperspectral classification, most of the traditional methods are based on the Bayesian framework [4] and sometimes, feature reduction technique is incorporated. Recently, the SVM based methods are considered as a most efficient [5], yet robust way to label hyperspectral pixels. Nevertheless, existing SVM based methods merely use spectral information of hyperspectral image, while is powerless to include the spatial relationships among pixels. But in fact, the usage of spatial information has been attached great significance in recent work.

Many problems of labeling pixels of a hyperspectral image can be addressed by making full use of both spectral and spatial information. One of the most common methods which take advantage of both information is Markov Random Fields (MRF) framework. A study of the related energy minimization methods for MRF has been presented in [6], which compares the state-of-arts. However, these works are usually with a high complexity. The work of Bruzzone [7] has attempted to integrate SVM technique with MRF framework (termed SVM-MRF for short) to get accurate classification for hyperspectral images. But it is not of very high speed. Later, Benediktsson [8] introduced the "fuzzy no-edge/edge" function into the spatial function to preserving edges while the accuracy is still need to be improved.

In this paper, we also focus on the SVM-MRF based hyperspectral image classification. A novel classifier is proposed contributing the following aspects: 1) SVM is intro-

This work is supported by the National Basic Research Program of China (973 Program) (Grant No. 2011CB707104), and the National Natural Science Foundation of China (Grant No. 61172143, 61379094 and 61105012).

duced to improve the traditional Iterated Conditional Modes (ICM) [9] to get better initialization; 2) A new energy function is proposed to constrain the result, which can make use of both the global and local information of the image. 3) A stepwise optimization is adopted to alternatively get the optimal result. Our new energy function is based on one of the oldest energy minimization method called ICM, which has been widely used in many fields such as character recognition and medical image segmentation [4]. But our experiments demonstrate that if the limitations of ICM are properly reduced, the obtained results are surprisingly promising.

The rest of this paper is organized as follows. In section 2, we discuss the main challenges for the improvement of ICM. In section 3, we show the procedure of the stepwise ICM (SICM). The experiments and comparisons are demonstrated in section 4, and the conclusion is made in section 5.

2. LIMITATIONS OF ICM

ICM is a straightforward method for energy optimization and its popularity is mainly due to the simplicity. However, its accuracy can still be improved if some specific limitations of ICM are overcome. In this section, we will first give a brief introduction to ICM, as well as MRF. Then the limitations of ICM are derived, which motivate the work in this paper.

2.1. Overview of ICM

ICM was firstly introduced as an energy optimization method for MRF. Though there are various applications in computer vision and image processing, many of them finally come down to the energy minimization problem, which can be effectively modeled as MRF. Actually, MRF is a labeling problem that tries to give each pixel its best label. The criterion for the "best" is closely related to the problem to be tackled and it directly influences the definition of energy function.

For the accurate description of ICM, several notations are presented first. Assume the image I with n pixels is denoted as $\mathcal{X} = \{x_i\}_{i=1}^n$ and the task is to get a label set $\mathcal{L} = \{l_i\}_{i=1}^n$ for each pixel, where $l_i \in \{1, 2, ..., L-1\}$ and L is the total number of labels. The energy function is generally defined as

$$E = E_d + \lambda E_s, \tag{1}$$

where E_d is the data term reflecting the likelihood of the observed data, and E_s is the smoothness term representing the joint Gibbs distribution on the label field and satisfying the Markov property. By finding the minimum solution of the energy function E, the label set can be obtained by

$$\mathcal{L}^* = \arg\min_{\mathcal{L}} E. \tag{2}$$

For the minimization of this energy function, ICM is one of the most popular approaches. The main steps of ICM can be summarized as follows:

Step 1. Assume an initial labeling $\mathcal{L}_0$ and specify the convergent conditions: maximum iteration number N and the minimum energy $Delta$.

Step 2. For the kth iteration, check all the possible label choice of each pixel in I. If no reduction to the energy function can be obtained, the label remains unchanged. Otherwise, update the pixel's label to the one that can reduces the energy value most.

Step 3. If $k > N$ or $E_k - E_{k-1} \leq Delta$, the method stops. Otherwise, update $k = k + 1$ and return to step 2.

2.2. What result in the bad performance of ICM?

ICM is an efficient method that can be completed quickly after a few iterations. However, this rapidly converging method cannot guarantee a globally optimal solution. The drawbacks of ICM are primarily from two aspects.

1) Label initialization. Compared with other method, ICM is heavily influenced by the label initialization [6], which is a double-edged sword. If a bad initialization is obtained, the error area will spread out in the image after iterations; or if a good initialization, the result will might be surprisingly satisfying. When using other energy minimization methods such as graph cut, LBP and so on, this problem doesn't exist. Therefore, getting an appropriate initialization is of great importance.

2) Local convergence. In ICM, every pixel gets its new label after each iteration, according to the pixel data itself and its neighborhood. However, compared to many other energy minimization methods, ICM has an obvious disadvantage that it provides good result only when the number of local minima is small. A huge number of local minima may result in a solution that is far from global optimum, which is not acceptable. But unfortunately, this is often the case.

3. SICM: STEPWISE ITERATED CONDITIONAL MODES

To address the problem discussed in the previous section and make good use of both spectral and spatial information, we propose a stepwise ICM approach. In SICM, we introduce a SVM based strategy to get a prominent initialization. Then the detailed optimization to the energy function of ICM is followed next. At last, a new local-preserving descriptor is proposed to overcome some local problem for the hyperspectral classification result of ICM.

3.1. SVM initialization

Using SVM to get the initialization is one of the key components in our method. Traditionally, the maximum likelihood method is applied to get the initialization for ICM. However, the accuracy of classification result is no more than 80%. To get an effective initialization, SVM, a pixelwise classification

method which has been proven to be effective for the classification of high-dimensional data, is adopted in our method. We use one-against-all SVM to complete the label initialization.

3.2. Stepwise optimization

Traditionally, the energy function of MRF is defined as Eq. 1, in which a coefficient λ is imposed to give a proper proportion between the data term and the smoothness term. We find that this coefficient is hard to be determined adaptively. In some part of an image where there is nearly only one class of pixels, the smoothness term counts more, while in other parts that different classes coexist and no single class takes the advantage, the data term is more important. To cope with this problem, a stepwise strategy is introduced to separate these two terms in which case the coefficient λ can be abandoned. The principle is that the data term and smoothness term should take their respective effect according to the actual configuration of the observed data. To be specific, we first use the smoothness term to make a local classification, and then use the data term to wipe off the wrong labels which can be seen as a global restriction.

With this strategy, an improved energy function is proposed. For the smoothness term, a revised ISING model is proposed, while for the data term, a GMM model is introduced.

3.2.1. Smoothness term

A typical definition of the smoothness term is $E_s = \sum_{p,q\in C} V_c(l_p, l_q)$, where $p, q \in \mathcal{X}$ and C is the set of cliques defined on a specific neighborhood system. Traditional ICM usually employs the 4 adjacent neighbors to smooth the labeling results. However, due to the local convergence, enlarging the scale of the neighborhood system might help get a more accurate optimization result. Therefore in this work, a 5-order neighborhood system is adopted to improve the performance. [1]

Besides, we think pixels with different distances to the central one should be treated separately. With respect to this point, the weight of pixel x_i is set as $w_i = 1/n$, where n is its order. This weighting strategy reflects the influences and confidences from different neighboring pixels. Based on this formulation, we calculate the smoothness term as the sum of potential functions $V(l_p, l_q)$, which takes the form

$$V_{\{p,q\}}(l_p, l_q) = \begin{cases} \frac{-\rho}{n} & if \quad l_p = l_q \\ \frac{+\rho}{n} & if \quad l_p \neq l_q \end{cases} . \tag{3}$$

[1] If the distance between then neighbor pixel x_i and the central pixel x_o is n, then x_i is the n-order neighbor of x_o. In our experiments, we find that with the increase of the neighbor order, the classification accuracy boosts. But when the order is larger than 5, the performance decreases.

In fact, this is a revised version of ISING model motivated by the different weights of neighboring pixels. The farther the pixel is from the center one, the less effect it will take.

3.2.2. Data term

In the energy function, the data term is used to penalize solutions that are not in consistence with the prior information. A general form of the data term is $E_d = \sum_{p\in\mathcal{X}} w_p(l_p)$, where $w_p(l_p)$ is the cost of assigning a label l_p to pixel p. For the hyperspectral image classification scenery, there are several categories to be identified. Therefore, the Gaussian Mixture Model (GMM) is introduced to define the cost function.

Assume there are K Gaussian distributions (classes in the hyperspectral image) in the GMM model, indicating the k classes contained in the hyperspectral image. Each of them can be specified by a set of parameters $\theta_k = \{\mu_k, \sigma_k, z_k\}, k \in \{1, ..., K\}$, where μ is the expectation, σ is the variance, and z_k is the mixing coefficient of the kth category. These parameters can be iteratively estimated and updated by the Expectation-Maximization (EM) method. Then we define

$$w_p(l_p) = f(p|\theta_{l_p}) \tag{4}$$

where $f(.)$ is the probability density function. This definition measures the probability of a pixel belonging to a specific category.

3.2.3. Optimization

With the above definitions, the stepwise optimization is introduced as below.

Step 1. Use the smoothness term to make a local classification. The labels obtained by the SVM initialization is not very robust. Some parts of the image may have isolated class labels. Therefore, a smoothed label set is acquired by $\mathcal{L}_s = \min_{\mathcal{L}} E_s$ through the ICM optimization.

Step 2. Take the data term to justify the correctness of the pixels' label. After the first step, there will be pixels with a changed label. But this change is not necessarily appropriate. Besides, the other pixels that remains unchanged might be assigned a wrong label, too. Therefore, we employ the data term to re-identify every pixel to see if its label is consistent with the corresponding class property. This is achieved by checking the pixels' labels by ICM as $\mathcal{L}_d = \min_{\mathcal{L}} E_d$.

Step 3. Check if this iteration has cause enough changes. If so, the optimization stops and the final labeling results are obtained. Otherwise, go to step 1 and start a new iteration.

3.3. An alternative descriptor

The new stepwise strategy can get a good classification result while we find in a few sensitive areas of the hyperspectral images, the result is still not satisfying. This failure is not

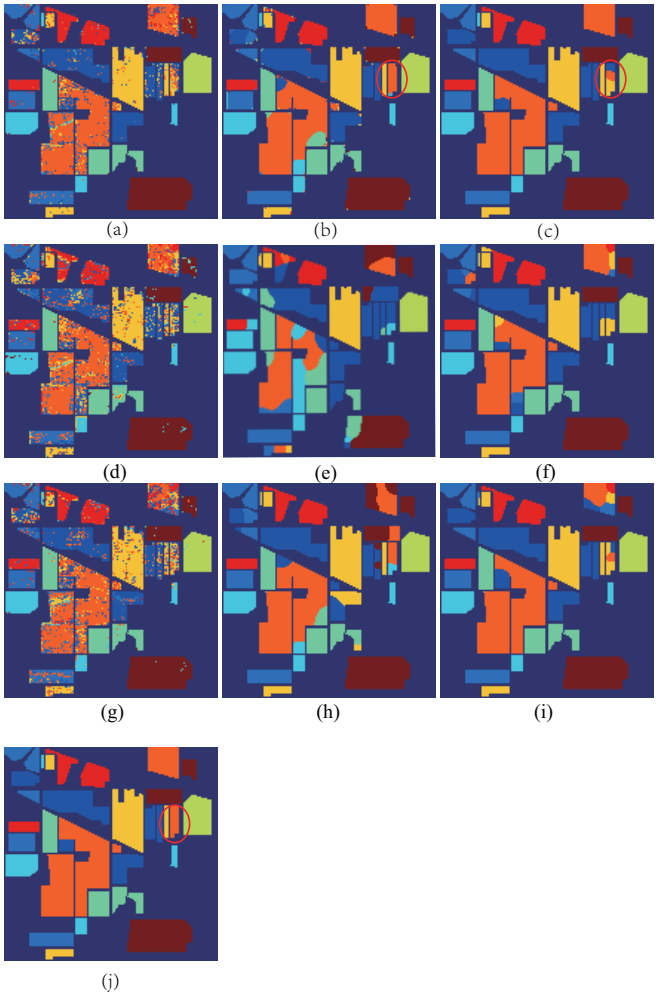

Fig. 1. Results on Indian Pine Image. From the first to the third row, the experiments are conducted based on 30%, 10% and 5% training samples. For each row, the three classification maps are respectively obtained by SVM, SICM with SSIFT descriptor and SICM with the raw image feature. The last row is the Ground truth.

particular for the proposed method ,but for many other state-of-the-art methods. To optimize these areas, a new feature descriptor based on spectral angle [10] is utilized. We describe the feature in a SIFT-like manner and it is thus termed as S-SIFT. The procedure is as follows: first, calculate the spectral angles [10] of every pixel with its neighbor pixels of different direction. Second, define the direction with maximum spectral angle as the main direction. Third, take 16 pixels as a seed and 16 seed as a window to describe every central pixel just like SIFT. The using of spectral angle is with a view to the consistence of neighboring pixels. Experiments demonstrate that this descriptor can accurately segment these areas.

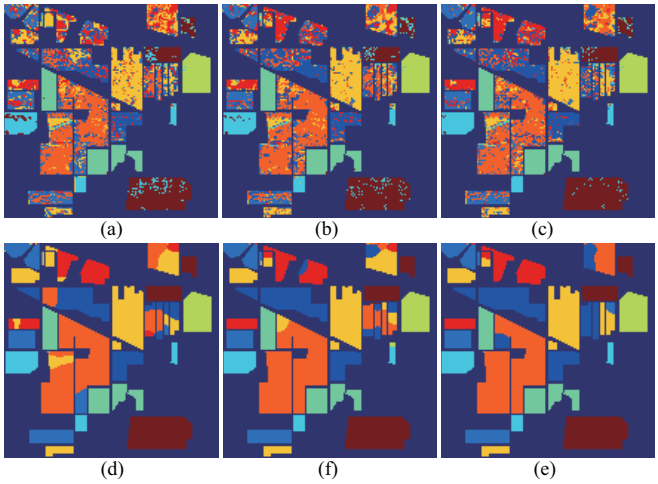

Fig. 2. Comparative results of kNN classification (first row) and kNN based SICM (second row). From the first column to the third column, the results are respectively obtained with 5%, 10%, and 30% training samples.

4. EXPERIMENTS

4.1. Data set

We test our method on three publicly available hyperspectral images, namely, Salinas, Pavia University, and Indian Pines. These three images are described as follows.

The *Indian Pines* image was collected by AVIRIS sensor over a vegetation area in North-western Indiana, and consists of 145×145 pixels, 224 spectral reflectance bands with spatial resolution of 20m/pixel. This image contains 16 classes of interest. We select the major nine categories of them to accomplish our experiment. We choose our training samples in the image randomly, and the percentage of training samples includes 5%, 10%, and 30%.

The *Salinas scene* was also gathered by AVIRIS sensor, on Salinas Valley, California. This image is characterized by spatial resolution of 3.7 m/pixels with 224 spectral bands, and comprises 512×217 samples. The image contains 16 classes of interest. We have our test on these 16 classes of interest, and the test is accomplished with 5% and 10% training samples.

The *Pavia University* is acquired by ROSIS sensor over Pavia, northern Italy. This image consists of 610×340 pixels with 103 spectral bands, and the resolution is 1.3m/pixel. For this image, we conduct our experiments with 5% and 10% training samples.

4.2. Experimental Results

For the Indian Pines image, the experiments start with the classification of 30% training samples. The comparison is illustrated in the first row of Fig. 1 with a ground truth in

Fig. 1(j). The three results are respectively obtained by the traditional SVM classifier, SICM method with SSIFT feature and SICM method with the raw image feature. The corresponding accuracy is 90.11%, 95.17% and 98.97%.

Obviously, Both the raw image feature and SSIFT feature based SICM obviously outperform the traditional pixel-wise SVM. But using SICM with the raw image information has produced the highest accuracy. This level of accuracy is satisfying and it is difficult to see a better one in literatures. Besides, it is found that the SICM iteration only cost three to four seconds, which means the new methods does not degrade the efficiency of ICM. Seen from the result of Fig. 1(a), we find that compared with ground truth, the errors are concentrated in the upper right soybeans-mintill area, which is highlighted with a red circle. This area is well known that it is difficult to be classified accurately [1]. For other area, there is nearly no wrong label when SICM is employed. However, in Fig. 1(b) we can find that highlighted area almost have no error. Benefited from the new SSIFT features we proposed, the soybean-mintill can almost be separated out perfectly. This advantage cannot be achieved by most of other features. As for its accuracy, although it is slightly lower than using gray feature, it is still obviously better than using the traditional SVM classifier. Therefore, it is evident that the new SSIFT feature is of great value for some error-prone area of the image owing to the existence of its spatial characteristic.

Next, we conduct our experiments with the training samples of 10% and 5%, respectively. The comparisons are also presented in the second and third row of Fig. 1. By introducing SICM, the accuracy generally has made a nearly 20% promotion than traditional pixel-wise SVM. Especially for 5% samples which is often used in other literatures, the accuracy of the traditional SVM classifier is only 77.36%, while for SICM, the accuracy reached 94.31%. As for 10% training samples, the original SVM achieves the accuracy of 79.69%, and SICM can achieve the accuracy 96.08%. These results illustrate an obviously better performance of our methods.

4.2.1. Comparison with kNN and other classification method

We also applied a mostly used kNN method to get the initial classification results. The comparisons conducted with 5%, 10%, 30% training examples are shown in Fig. 2. It is found that the accuracy of the traditional method kNN tends to be around 70%, which is not a nice result. However once the kNN method is improved by our SICM method, the averaged accuracy reaches 90%, which is a distinct improvement. As shown in Fig. 2, the classification accuracy is 65.43%, 69.46% and 76.08% for kNN and 87.06%, 87.77%, 87.77% for kNN based SICM. These results demonstrate that SICM can improve the accuracy of many other methods, not only the SVM based one.

We make a detailed comparison with another two methods: the Feature Extraction by Genetic Algorithm combined with SVM (FEGA-SVM) and Extended Morphological Profiles based SVM (EMP-SVM) methods, whose result is provided by [1]. We choose 10% training samples as a representation and the comparison results are shown in Tab. 1.

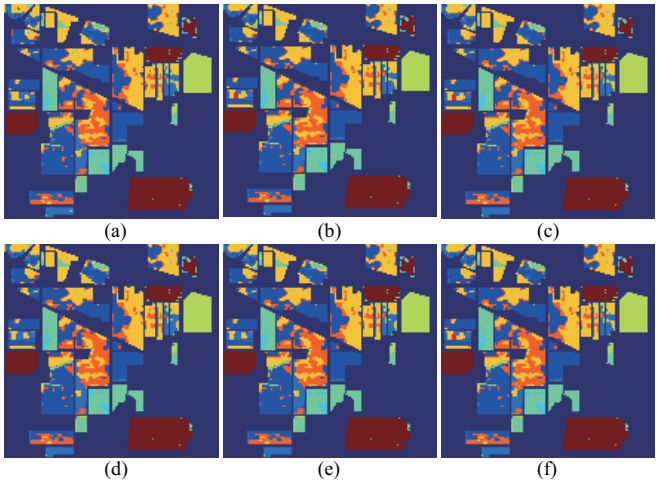

Fig. 3. Comparison result of energy minimization "family". (a)ICM. (b)LBP-s. (c)LBP-M. (d)Graph cut-swap. (e)Graph cut-expansion. (f)TRW-s.

We can see the precision of the former two methods is less than 90%, and for our method, except for kNNICM method (87.77%), the accuracy of every method is not less than 90%. The bold font means the best classification results in every rows. we can see that except for the classification result of soy which EMP-SVM has the best result, the other best ones are in our method. For many classes, our method has the perfect results which reach 100%.

4.2.2. Comparison with different energy minimization methods

We also compare the proposed stepwise optimization method with five other energy minimization methods. They are the original ICM, Graph Cuts, LBP, Tree Reweighted Passing(TRW) [6], and the proposed SICM. Although it is widely accepted that the global optimizing methods such as Graph Cuts, LBP, and TRW often outperform ICM, we find that the SICM we proposed can obviously surpass the other competitors. The accuracy of the competitors are all no more than 50%(the highest accuracy is 47.92%). While, for SICM the averaged accuracy is 98.97%. The result can be seen in Fig. 3.

4.3. Results on Pavia university and Salinas

In order to further demonstrate that SICM is appropriate for different hyperspectral images, experiments are also conducted on another two databasets, Pavia University and Salinas. But

Table 1. Comparison results

	Number of sample	FEGA-SVM	EMP-SVM	kNN	kNN-SICM	SICM	SSIFT-SICM
feature number		(95)	(18)	(206)	(206)	(206)	(128)
Overall accurary		76.36%	88.63%	69.46%	87.77%	**96.08%**	91.86%
Average accurary		78.56%	88.22%	74.78%	92.68%	**98.09%**	95.17%
Classes							
Corn-notill	1428	67.00%	84.00%	54.07%	72.82%	**100%**	79.74%
Corn-mintill	830	60.00%	80.00%	55.34%	93.75%	**100%**	98.77%
Grass-pasture	483	88.00%	90.00%	87.99%	96.34%	96.34%	**98.69%**
Grass/Trees	730	89.00%	95.00%	99.05%	**100%**	**100%**	96.83%
Hay-windrowed	478	93.00%	95.00%	99.20%	**100%**	**100%**	**100%**
soybeans-notill	972	67.00%	85.00%	77.55%	94.16%	**99.89%**	**99.89%**
Soybeans-min	2455	73.00%	**90.00%**	57.60%	77.08%	86.67%	88.35%
Soybean-clean	593	77.00%	76.00%	53.55%	**100%**	**100%**	**100%**
Woods	1265	93.00%	99.00%	88.67%	**100%**	**100%**	94.25%

for the limited paper length, the figural results are not illustrated. Here we only report the quantitative results. The precision of traditional pixel-wise SVM is only about 90% (90.65% and 90.79% for 50 and 100 training samples, respectively), while the accuracies of SICM are higher than 95% (96.59% and 97.76% for 5% and 10% training samples, respectively). As for the Pavia University image, the classification accuracy of the traditional SVM is around 85.00%. The SICM increase the accuracy by 13.39% and 10.87% for 5% and 10% training sample respectively.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we present a novel classifier named SICM for hyperspectral image classification, which can make full use of spatial and spectral information. Experiments prove the superiority of the proposed SICM. Moreover, the newly proposed local-preserving feature SSIFT is helpful for some traditionally error-prone area. In the future work, we will improve the local-preserving feature and apply it to address many other problems in hyperspectral classification.

6. REFERENCES

[1] J. Palmason J. Benediktsson and J. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.

[2] Q. Wang, Y. Yuan, and P. Yan, "Visual saliency by selective contrast," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 23, no. 7, pp. 1150–1155, 2013.

[3] Q. Wang, Y. Yuan, P. Yan, and X. Li, "Saliency detection by multiple-instance learning," *IEEE Trans. Cybernetics*, vol. 43, no. 2, pp. 660–672, 2013.

[4] Y. Chen J. Meng, J. Zhang and Y. Huang, "An iterated conditional mode solution for bayesian factor modeling of transcriptional regulatory networks," in *Proc. IEEE Int. Workshop on Genomic Signal Processing and Statistics*, 2010, pp. 1–4.

[5] R. Rajabi M. Khodadadzadeh and H. Ghassemian, "A novel approach for spectral-spatial classification of hyperspectral data based on svm-mrf method," in *Proc. IEEE Int. Geoscience and Remote Sensing Symposium*, 2011, pp. 1890–1893.

[6] D. Scharstein O. Veksler V. Kolmogorov A. Agarwala M. Tappen R. Szeliski, R. Zabih and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.

[7] F. Bovolo and L. Bruzzone, "A unified framework for map estimation in remote sensing image segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 1617–1634, 2005.

[8] J. Chanussot Y. Tarabalka, M. Fauvel and J. A. Benediktsson, "Svm- and mrf- based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 736–740, 2010.

[9] B. N. Subudhi A. Ghosh and L. Bruzzone, "Integration of gibbs markov random field and hopfield-type neural networks for unsupervised change detection in remotely sensed multitemporal images," *IEEE Trans. Image Processing*, vol. 22, no. 8, pp. 3087 – 3096, 2013.

[10] J. B. Boardman K. B. Heidebrecht A. T. Shapiro P. J. Barloon F. A. Kruse, A. B. Lefkoff and A. F. H. Goetz, "The spectral image processing system (sips)-interactive visualization and analysis of imaging spectrometer data," *Remote Sensing Environment*, vol. 44, no. 2/3, pp. 145–163, 1993.