

An improved Bayesian nonparametric mixture model to fusing both panchromatic and multispectral images for classification

T. Mao^{a, b}, H. Tang^{a, b}, Y. Shu^c, N. Yang^{a, b}

^a State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, 100875, China – (mting, yns@mail.bnu.edu.cn)

^b Key Laboratory of Environment Change and Natural Disaster, Beijing Normal University, 100875, China - tanghong@bnu.edu.cn

^c The national disaster reduction center of China, 100875, China – sysun@mail.bnu.edu.cn

KEY WORDS: Nonparametric Bayesian, Fusion, Classification, Object-Based Classification

ABSTRACT:

In this paper, we present an improved nonparametric Bayesian model based on a generalized metaphor of Chinese restaurant franchise (gCRF), which can take advantage of both panchromatic and multispectral images to obtain a classification map. There are two drawbacks in the gCRF when it is used to fuse panchromatic and multispectral image for classification, first, since superpixels which are obtained using other segmentation algorithm are considered as basic analysis units instead of pixels in the gCRF, the quality of final classification result depends on the calibre of over-segmentation map. Second, when classify PAN and MS image using the gCRF, semantic segments extracted from PAN image are sharing with MS image and then they are allocated clustering labels using MS image which is richer in spectral information. All the local semantic segments extracted from panchromatic image are supposed to be suitable for representing the local spatial structures in multispectral image, which is not objective in practice. In this paper we propose an improved gCRF, focusing on overcoming the two shortcomings of the gCRF. First of all, the formation of superpixels are integrated into the nonparametric Bayesian framework of the improved gCRF to obtain qualified superpixels. Second, the quality of the semantic segments is checked before sharing with MS image and corresponding measure is taken to dealing with the situation that the semantic segments are not suitable to represent the local structure of MS image. We evaluate the efficiency of the improved model and show it obtains the state-of-art results.

1. INTRODUCTION

With the development of improved sensors, many earth observation platforms can provide high-spatial-resolution remote sensing data, which often contain two types of images, i.e., panchromatic (PAN) and multispectral (MS) images, for example, QuickBird can provides us a PAN image with a resolution of 0.6 m and four-bands MS image with a resolution of 2.4 m. Several methods are proposed previously in order to make full use of both images to obtain a better classification results. Image fusion is the most common method to integrate the two images which merges the high resolution PAN and low-resolution MS imagery to create and enhanced high-resolution MS image which can be used in subsequent classification (Wang,2013). These methods can be termed as *sharpen-then-classification*. The shortcomings of these methods are twofold: firstly, classification results of these methods depend on the fusing algorithms which may have a negative effect on the classification due to the spectral distortion. Secondly, as pixel-based methods, the classification results may be unsatisfactory when dealing with high-spatial-resolution images due to the high degree of within-class spectral variability. *Segmentation-then-classification* approaches are other ways widely used to classify high-resolution PAN and MS imageries, which uses a segmentation of an image with higher spatial resolution as an input for the classification of MS image. Mao (2016) propose a unsupervised *segmentation-then-classification* method, named a generalized metaphor of Chinese restaurant franchise (gCRF), to fuse both PAN and MS image for classification under a nonparametric framework. In the gCRF, two interactive random processes, i.e., *table selection* and *dish selection* are used to interpret the generalization of PAN and MS images respectively. In the process of table selection, the local spatial structures of PAN image are extracted, which use superpixels from an over-segmentation map obtained using eCognition as basic analysis units rather than pixels. In the following process of *dish selection*, MS image with richer spectral information is used to allocate the

clustering label for the local spatial segments from PAN image. There are two deficiencies of the gCRF. Firstly, the quality of final classification result depends on the quality of over-segmentation map. Secondly, there is an assumption that all the local spatial segments of PAN image are supposed to be suitable for representing the spatial structure in MS image, which is not always in accordance with the fact. In this paper we propose an improved gCRF, focus on handling the two problems of the gCRF. The contribution of this paper are twofold: firstly, the formation of the superpixels is incorporated into the Bayesian framework; secondly, the examination that whether the local spatial segments of PAN image are suitable to represent the local spatial structures of MS images is taken into consideration and corresponding measures are taken to deal with the situation that some of the local spatial segments are not suitable to represent those in MS image. Experimental results show that the proposed method does improve the classification accuracy of the gCRF and has a state-of-art result than other comparative approaches, including *sharpen-then-classification* and *segmentation-then-classification* methods.

2. METHODOLOGY

In this section, we first introduce the gCRF and its deficiencies. Then we present the proposed improved model concentrating on handling the problems of the gCRF.

2.1 Generalized Chinese Restaurant Franchise

To fuse PAN and MS images for classification, Mao (2016) proposed a unified framework to iteratively discover semantic segments from PAN image and allocate clustering label using MS image, where the probabilistic generalization process of PAN and MS images can be explained by the gCRF. The illustration of satellite classification using gCRF the is shown in Fig.1 As shown in Fig.1(a), during the pre-processing, PAN image X^P is over-segmented into a set of superpixels which are used as

basic analysis units rather than pixels in the following processes and MS image \mathbf{X}^M is up-sampled to the size of PAN image. Besides, both images are partitioned into L groups in an identical way, i.e., $\{\tilde{x}_1^P, \tilde{x}_2^P, \dots, \tilde{x}_L^P\}$ and $\{\tilde{x}_1^M, \tilde{x}_2^M, \dots, \tilde{x}_L^M\}$.

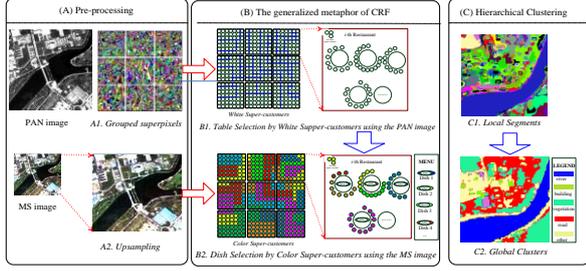


Fig.1 An illustration of satellite image classification using the gCRF [Mao, et al. 2016]

Then, superpixels from PAN and MS images can be explained by two types of people in L restaurants, i.e., white supercustomers $\theta^P = \{\theta_1^P, \theta_2^P, \dots, \theta_L^P\}$ and colour supercustomers $\theta^M = \{\theta_1^M, \theta_2^M, \dots, \theta_L^M\}$ in the generalized metaphor of CRF shown in Fig.1(b). There are two random processes in the gCRF that can explain the formation of semantic segments from PAN image and allocating the clustering label using MS image, i.e., *table selection* and *dish selection*.

(1) Table selection

As shown in the first row of Fig.1(b), assuming that all the white supercustomers with exception of θ_{ij}^P have selected tables \mathbf{T}_{-ij} and all of the tables with the exception of table t in the i -th restaurant have been served with dishes \mathbf{K}_{-it}^M that were ordered by colour supercustomers. Supercustomer θ_{ij}^P choose a table t according to (1).

$$p(t_{ij} = t | \mathbf{T}_{-ij}, \mathbf{K}_{-ij}^P)$$

$$\propto \begin{cases} n_{it}^{-ij} f_{k_{it}^P}(x_{ij}^P), & \text{if table } t \text{ has exist} \\ \gamma_0 f_{k_{it}^P}(x_{ij}^P), & \text{if } t \text{ is a new table} \end{cases} \quad (1)$$

Where n_{it}^{-ij} is the number of customers with exception of θ_{ij} already chosen table t , $f_{k_{it}^P}(x_{ij})$ is the likelihood of observation x_{ij} and γ_0 is the prior of a new table. If t is a new table, a dish k_{it}^P would be selected with the probability

$$p(k_{it}^{new} = k | \mathbf{T}_{-ij}, \mathbf{K}_{-it}^P) \propto \gamma f_k^P(x_{ij}^P) \quad (2)$$

Where \mathbf{K}_{-it}^P indicates the dish of all the tables selected by white supercustomers with exception of table t in i -th restaurant and γ is the prior for a new dish

2) Dish selection

After all the white supercustomers have a meal, the table selected by them are shared with colour supercustomers from MS image which take in charge to select a dish for each table according to (3), which is shown in the second row of Fig.1(b).

$$p(k_{it}^M = k | \mathbf{T}_{-it}, \mathbf{K}_{-it}^M) \propto \begin{cases} m_k f_{k_{it}^M}(x_{it}^M), & \text{if } k \text{ is previously used} \\ \gamma f_k^M(x_{it}^M), & \text{if } k \text{ is a new dish} \end{cases} \quad (3)$$

where m_k indicates the number of tables already served dish k chosen by colour supercustomers, and γ is the prior for a new dish. For more detail please refer to (Mao, 2016).

There are two major deficiencies of the gCRF. Firstly, in the process of pre-treatment, an over-segmentation algorithm is applied to obtain the superpixels, for example, the multi-resolution segmentation in eCognition. Thus the quality of classification result depends on the quality of the over-segmentation map. The inferior over-segmentation algorithm may result in poor classification result. Secondly, there is an assumption in the gCRF for sharing the tables between white supercustomers and colour supercustomers that all the semantic segments, i.e., tables, discovered from PAN image can represent the local spatial structure of MS image in the same area, which sometimes may not be accordance with the fact. For example, in the area of Miyun, Beijing, China, shown in Fig.3, the gray values of some buildings are quite similar with those of roads in PAN image. By contrast, the spectral values of the two in MS

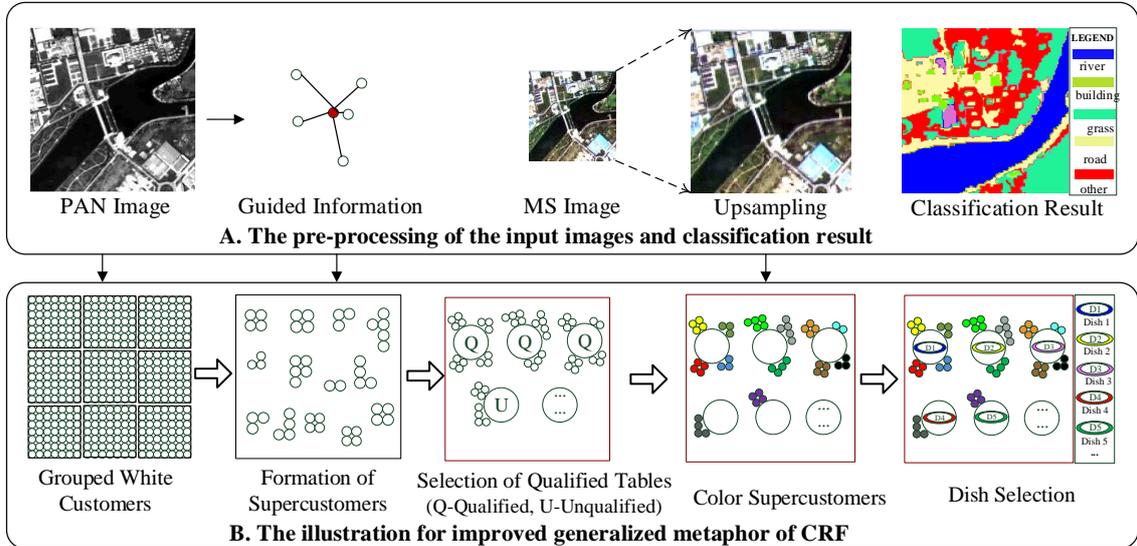


Fig. 2 Illustration of improved generalized metaphor of CRF

image may be distinguishable. Thus those buildings and roads may be wrongly discovered in the same semantic segments in PAN image which is not suitable for sharing with MS image.

2.2 The Improved Generalized Chinese Restaurant Franchise

In this paper an improved gCRF are proposed concentrating on two aspects as shown in Fig.2: (1) the formation of supercustomers (i.e., superpixels) from raw PAN image; (2) the selection of qualified tables from PAN image which are suitable for sharing with MS image to allocate clustering labels.

2.2.1 Formation of supercustomers: The formation of superpixels from pixels of PAN image can be analogous to converging customers with similar characteristic to a supercustomer. To group customers to supercustomers, guided information which is the descriptor of combining spatial and spectral information among pixels is used (Shu et al., 2015). The guided information is defined as follows

$$r_{m,n} = \exp\left(-\frac{\|x_m - x_n\|^2}{\sigma_{spa}^2} - \frac{\|G_m - G_n\|^2}{\sigma_{spe}^2}\right) \quad (4)$$

where x_m indicates the location of the m -th pixel, G_m indicates the m -th pixel value in guided image, σ_{spa}^2 and σ_{spe}^2 are the spatial and spectral variance, respectively. w -th customer in j -th restaurant, indicated as c_{jw} , chooses a supercustomer θ_{jw} to join with proportional to

$$p(\theta_{jw} = \theta | \theta_{-wj}, \mathbf{K}_{-wj}) \propto \begin{cases} \left(\sum_{\theta_{jl}=\theta} r_{j,l}\right) f_k(x_{ji}), & \text{if supercustomer } s \text{ has exist} \\ \alpha_0(x_{ji}) f_k(x_{ji}), & \text{if } s \text{ is a new supercustomer} \end{cases} \quad (5)$$

where $\sum_{\theta_{jl}=\theta} r_{j,l}$ is the sum of guided information of all the customers in j -th restaurant who belong to supercustomer θ and α_0 is the prior for a new supercustomer.

2.2.2 Selection of Qualified Tables: After we obtain the supercustomers, i.e., superpixels, each supercustomer selects a table (semantic segments) in a restaurant according to (1) and (2). In other words, superpixels in PAN image are grouped into semantic segments in the same way as that of gCRF. In the gCRF, the tables selected by white supercustomer are all shared with colour supercustomers, while in the improved model the quality of tables is checked before sharing with MS image. In other words, the validity of the semantic segments from PAN image are examined before transferred to MS image. During the examination, a semantic segment can be regarded as qualified to represent the local spatial structure of MS image if the spectral variability of the area in MS image with the same location of the semantic segments is lower than an experimental threshold, and it is regarded as unqualified if not. The qualified semantic segments of PAN image are shared with MS image and they are allocated a clustering label using MS image according to (3). While the unqualified semantic segments are split into superpixels contained in it. Then each superpixel in an unqualified semantic segments is regarded as a qualified table which is considered suitable to represent the local spatial structure of MS image and is allocated a clustering label of MS image according to (3), too.

The Gibbs-like sampling algorithm for the improved gCRF includes six steps.

- 1) Step 1: Initializing the model parameters and calculating the guided information.
- 2) Step 2: sampling a superpixel label for each pixel. Given the rest of the variables, sample a superpixel label θ_{wj} for w -th pixel in j -th sub-image according to (5).

- 3) Step 3: sampling a table label, i.e., semantic segments label for each superpixel using PAN image. Sampling table label t_{ij} for i -th superpixel in j -th sub-image according to (1) and sampling a dish for new table according to (2).
- 4) Step 4: selecting qualified tables. Examine the quality of each semantic segments, semantic segments meet the requirement mentioned before are considered as qualified tables. Each unqualified semantic segment is split to superpixels contained in it and each superpixel is treated as a qualified table.
- 5) Step 5: allocating clustering labels using MS image. Allocate each qualified table a clustering label according to (3).
- 6) Step 6: checking the convergence of the clustering result. If not, repeat Step 2 to Step5.

3. EXPERIMENT RESULT AND EVALUATION

The experimental data we choose to evaluate the performance of the proposed method is shown in Fig.3 including a PAN (2-m resolution) and MS image (6-m resolution; four bands: red, green, blue and near-infrared) acquired by Mapping satellite-1 images with a size of 1200*1200 pixels (PAN image), covering a suburban area (Miyun, Beijing, China) with five major geo-objects (water, grass/tree, building, road and bare soil), as shown in Fig.3(c).

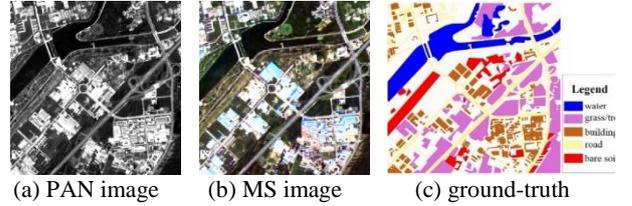


Fig. 3 Panchromatic, multispectral satellite imagery and ground-truth image

We compare the proposed method with gCRF and two types of approaches using both PAN and MS image to obtain classification results, (1) pixel-based methods: firstly, we use GS in ENVI to obtain the pan-sharpened image, then we use Kmeans and SVM to classify the pan-sharpened image, termed as Kmeans+GS+pix and SVM+GS+pix. (2) object-based methods: firstly, MS image is up-sampled to the size of PAN image and averaged according to a segment map obtained from PAN image. Then the pixels of averaged MS image in a segment is considered as basic analysis units instead of pixels and we perform Kmeans and SVM to obtain the classification results. Methods of this type are termed as Kmeans+MS+seg and SVM+MS+seg. The classification results and quantitative evaluation is shown in Fig.4 and TABLE I.

As shown in Fig.4, the proposed method performs similar with gCRF from visual inspection and better than Kmeans+GS+pix and Kmeans+MS+seg (Fig.4(c) and Fig.4 (e)). It also outperforms SVM+MS+seg, since the performance of the latter depends on the quality of segmentation map from PAN image. The proposed methods is comparable to SVM+GS+pix in Fig. 4 (b) by visual examination.

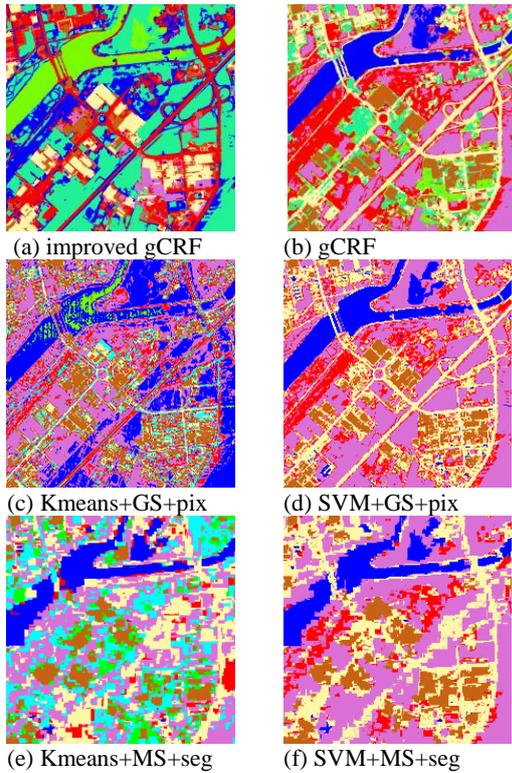


Fig.4 The classification results of all methods

As we can see from Table 1, the improved gCRF has the highest Kappa coefficient and lowest overall entropy except those of SVM+GS+pix. Since we know that the higher Kappa coefficient indicates higher classification accuracy, and the lower overall entropy indicates better homogeneity, the proposed method performs better than the gCRF, all the unsupervised methods which use Kmeans as the classifier, besides, it performs better than SVM+MS+seg which is a supervised method while it performs worse than SVM+MS+seg according to Kappa and overall entropy.

| | Impr oved gCRF | gCR F | pixel-based methods | | object-based methods | |
|-------|----------------------|----------|---------------------------|-----------------------------|--------------------------------|-----------------------------|
| | | | <i>Kmeans+G S+pix</i> | <i>SVM +GS +pix</i> | <i>Kmeans +MS+ seg</i> | <i>SVM +MS +seg</i> |
| Kappa | 0.8 | 0.79 | 0.39 | 0.80 | 0.35 | 0.72 |
| OE | 0.51 | 0.51 | 0.91 | 0.50 | 1.01 | 0.70 |

Table 1 Evaluation of the kappa coefficient and overall entropies of different models

4. CONCLUSION

In this paper we proposed an improved nonparametric Bayesian model based on the gCRF, which focuses on overcoming the two defects of the gCRF when used to fuse PAN and MS image for classification, i.e. the dependence of other over-segmentation algorithm and lacking of examination of tables when sharing them between PAN image and MS image. In the future, more information such as edge, shape of the objects can be taken into consideration to extract the spatial information of images for a better classification.

ACKNOWLEDGEMENTS

This work is partly supported by the National Natural Science Foundation of China (no. 41571334) and the Fundamental Research Funds for the Central Universities.

REFERENCES

- Shackelford, Aaron K., and Curt H. Davis, 2003. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE transactions on geoscience and remote sensing* 41(9). pp. 1920-1932.
- Bruzzone, L., & Carlin, L., 2006. A Multilevel context-based system for classification of very high spatial resolution images. *IEEE Transaction on Geosciences and Remote Sensing*, 44(9), 2587–2600.
- Johnson, B., Xie, Z., 2013. Classifying a high resolution image of an urban area using super-object information[J]. *ISPRS journal of photogrammetry and remote sensing*, 83, pp. 40–49.
- Mao, T., Tang, H., He, S., Shu, Y. *et al.*, 2015. Fusion of panchromatic and multispectral images for classification using the Chinese restaurant franchise with shaped tables. In: *Joint Urban Remote Sensing Event (JURSE)*, Lausanne, Switzerland, pp. 1-4.
- Shu, Y., Mao, T., Tang, H. *et al.*, 2015. Unsupervised Classification of VHR Panchromatic Images Using Guided Chinese Restaurant Franchise Mixture Model. In: *IEEE International Geoscience and Remote Sensing Symposium*, Milan, Italy, pp. 2413-2416.
- Teh, J. W., Jordan, M. I., Beal, M. J., *et al.* ,2006. Hierarchical dirichlet processes. *Journal of the American statistical association*, 101(476).
- Wang, G., Liu, J. and He, G., 2013. A Method of Spatial Mapping and Reclassification for High-Spatial-Resolution Remote Sensing Image Classification. *The Scientific World Journal*, 2013.