# A COMPARATIVE STUDY BETWEEN PAIR-POINT CLIQUE AND MULTI-POINT CLIQUE MARKOV RANDOM FIELD MODELS FOR LAND COVER CLASSIFICATION

Bo Hu [a, *], Peijun Li [a]

[a] Institute of Remote Sensing and GIS, Peking University, Beijing 100871, PR China – alex1949c@163.com, pjli@pku.edu.cn

**KEY WORDS:** Image Classification, Spatial Dependence, Markov Random Field, Pixel Interaction, Parameter Estimation

**ABSTRACT:**

Markov random field (MRF) is an effective method for description of local spatial-temporal dependence of image and has been widely used in land cover classification and change detection. However, existing studies only use pair-point clique (PPC) to describe spatial dependence of neighbouring pixels, which may not fully quantify complex spatial relations, particularly in high spatial resolution images. In this study, multi-point clique (MPC) is adopted in MRF model to quantitatively express spatial dependence among pixels. A modified least squares fit (LSF) method based on robust estimation is proposed to calculate potential parameters for MRF models with different types. The proposed MPC-MRF method is evaluated and quantitatively compared with traditional PPC-MRF in urban land cover classification using high resolution hyperspectral HYDICE data of Washington DC. The experimental results revealed that the proposed MPC-MRF method outperformed the traditional PPC-MRF method in terms of classification details. The MPC-MRF provides a sophisticated way of describing complex spatial dependence for relevant applications.

## 1. INTRODUCTION

Markov random field (MRF) provides a tractable way of taking advantage of spatial correlation. It has been used in many fields, such as image segmentation (Besag,1986; Derin and Elliott, 1987), texture modelling (Cross and Jain 1983), denoising and restoration(Lu and Jiang, 2001; Cui and Wang, 2005), spatial-temporal classification (Solberg et al., 1996; Melgani and Serpico, 2003), data fusion(Nishii ,2003) and so on.
Spatial information is especially important in classification of remotely sensed image because the use of spectral information only usually achieves limited classification accuracy. By using contextual information quantified by MRF, classification results may have a notable improvement.
Most existing applications of MRF measure spatial dependence only with pair-point clique (PPC). MRF with PPC has an effect of spatial smoothing that will reduce 'salt-and-pepper' appearance in classification results. However, some structural information may also be lost. The multi-point clique (MPC) which is defined with more complicated structures, hence, could possess a better potential in classification. In addition, with development of computer hardware, expense in calculation of MPC is gradually acceptable.
In this paper, we aim to compare performances of MRFs with PPC and MPC in classification of remotely sensed image. In order to make objective comparison, clique parameters involved in calculation are acquired from parameter estimation and a novel robust LSF method is proposed.
This paper is organized as follows. In Section II, we present some background theory on MRF and introduce a new method for clique parameter estimation. In Section III the results of experiment are presented, and discussion is given. Finally, some concluding remarks are given in Section IV.

## 2. METHODOLOGY

### 2.1 Markov random field

MRF is a technique for defining local dependence. Here basic notions of MRF involved in this paper are presented and specific details can refer to relevant papers (Geman ,1984; Derin and Elliott,1987).

**Pair-point clique and multi-point clique**
In this paper MRF is defined on a 3 by 3 neighborhood and the clique types involved in this paper are presented in Fig.1. The left specifies the site $(i, j)$ and its neighbourhood. The first row on the right are cliques of PPC and all 10 cliques constitute the MPC. Letters bellow the cliques are potential parameters of each clique. Most existing studies use PPC MRF to quantify spatial information. However MPC has a better ability in describing the local structure and we take this research to compare the capabilities in classification of two clique types.
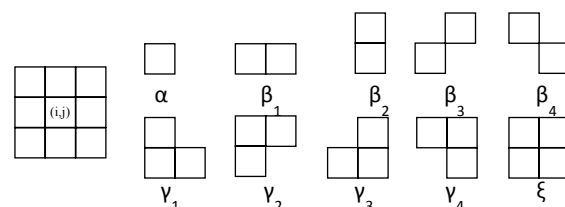


Figure 1. Neighborhood and clique types

$V_c(x)$ is the potential function of the clique which depends only on the class labels in clique c. $V_c(x)$ has the form as bellow

$$V_c(x) = \begin{cases} - \zeta & \text{if all x in c are equal} \\ \zeta & \text{otherwise} \end{cases} \quad (1)$$

Where $\zeta$ is potential specified for clique c. The clique potentials parameters are defined as $\theta = [\beta_1, \beta_2, \beta_3, \beta_4]^T$ for PPC and $\theta = [\beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \xi]^T$ for MPC.

**Clique parameter estimation**

Potential parameter $\theta$ involved in energy function is essential to the performance of MRF. In some applications, $\theta$ is set empirically while in this research, we aim to make an objective comparison between PPC and MPC to avoid subjective judgement. For this reason both $\theta$ of two MRFs are acquired by an estimation method.

Least squares fit (LSF) for estimation of MRF parameters is designed by Derin and Elliott(1987). The procedure consists of the following steps.

1) Equation (2) is built between class label $k$ and $j$ with the same neighbourhood $\eta$ in which $\theta$ is considered as a variable

$$(\phi(k, \eta) - \phi(j, \eta))^T \theta = \ln \frac{P(j \mid \eta)}{P(k \mid \eta)} \quad (2)$$

Where $\phi(k, \eta)$ is a vector with the number of each clique on class label k and its neighbourhood $\eta$   $P(j \mid \eta)$ is obtained from histogram statistics which denotes the probability of class label j with its neighbourhood $\eta$ .

2) An overdetermined system of equations is built in terms of the probabilities and the parameters.

3) Solve it using the LSF method.

However traditional LSF method could not give an ideal regression when some observations do not obey the same distribution and then a biased solution will be got. For remotely sensed image, misclassification is unavoidable and if the fallacious classified labels are used then the effect of LSF may discount. To solve this, we take a robust estimation method (Yang, 2002) which gives every observation a weight and adjusts the weight according to its residual, thus the misclassified pixels would have low weight and could hardly impact the estimation.

An observation is consist of each pair of k and j, and equation is built with the form

$$L = A\theta + e \quad (3)$$

Where $A = \phi(k, \eta) - \phi(j, \eta)$, $L = \ln \frac{P(j \mid \eta)}{P(k \mid \eta)}$ and e is an random error obeying the standard Gaussian distribution. An equation system is consist of all observations. Suppose $\hat{\theta}$ is the estimate of $\theta$ and the residual for i th observation is

$$v_i = a_i\hat{\theta} - L_i \quad (4)$$

As illustrated above, it is not proper to treat every observation equally that a weight vector $W$ is introduced as

$$W_i = W_i' \frac{\psi(v_i)}{v_i}$$

$$\psi(v_i) = \begin{cases} 0 & |\bar{v}_i| \leq k_0 \\ W_i' \frac{k_0}{|\bar{v}_i|} (\frac{k_1 - |\bar{v}_i|}{k_1 - k_0})^2 & k_0 \leq |\bar{v}_i| \leq k_1 \\ W_i' & |\bar{v}_i| \geq k_1 \end{cases} \quad (5)$$

Where $k_0$ and $k_1$ is chosen 2.0-3.0 and 4.5-8.5. $\bar{v}_i$ is standardized residual. Hence $\hat{\theta}$ is acquired by

$$\hat{\theta} = (A^T WA)^{-1} A^T WL \quad (6)$$

**2.2 Contextual classification by MRF**

The main goal of this experiment is to compare the capabilities of PPC and MPC based MRF in classification thus we design the experiment. The flow chart of this experiment is as bellow
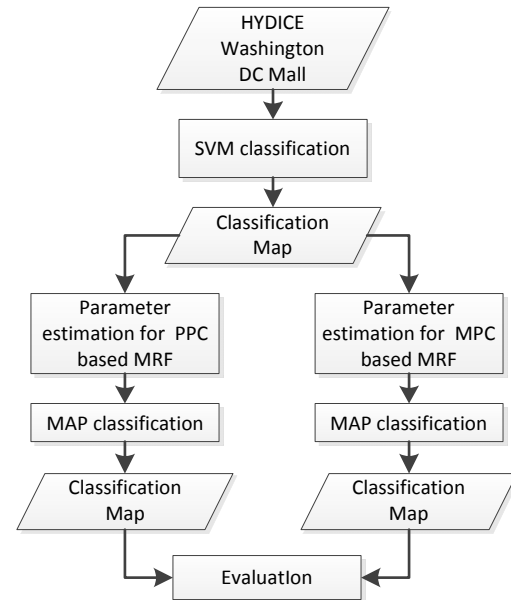


Figure 2. Flow chart of experiment

1. Image is classified by spectral and get conditional probability.
2. $\theta$ for MRF is estimated by preliminary classification map. For PPC-MRF $\theta$ is formed as $\theta = [\beta_1, \beta_2, \beta_3, \beta_4]$ and for MPC-MRF with $\theta = [\beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \xi]$. In this step a robust LSF method is taken.
3. The image is classified using two MRFs according MAP criterion.
4. The classification maps is evaluated.

**3.   EXPERIMENTAL RESULTS AND DISCUSSION**

**3.1  Study area**

The data set used in this research is HYDICE of Washington DC with resolution 2.5m. There are 191 bands ranging from 0.4μm to 2.4μm and bands in opaque region for atmosphere are omitted. The data set contains 1208 scan lines with 307 pixels in each scan line.

The main features in the image are roofs, roads, path (graveled paths down the mall center), grass, trees, water, and shadow. We use classification of support vector machine (SVM) as the input for MRF.

| Number | Class name |
|---|---|
| 1 | Roof |
| 2 | Paths |
| 3 | Roads |
| 4 | Shadow |
| 5 | Trees |

| 6 | Grass |
|---|-------|
| 7 | Water |

Table 1. Class names and number in experiment

## 3.2 Spectral classification

Image is classified by SVM with radial basis function kernel. The gamma and penalty involved are 0.005 and 500. There are 7 classes which are roof, paths, roads, shadow, trees, grass and water. The confusion matrix is as bellow

| Class number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | total |
|---|---|---|---|---|---|---|---|---|
| 1 | 104 | 7 | 28 | 0 | 0 | 5 | 0 | 144 |
| 2 | 21 | 33 | 0 | 0 | 0 | 0 | 0 | 54 |
| 3 | 17 | 1 | 162 | 11 | 16 | 4 | 0 | 211 |
| 4 | 1 | 0 | 4 | 37 | 1 | 0 | 3 | 46 |
| 5 | 0 | 0 | 0 | 0 | 149 | 9 | 0 | 158 |
| 6 | 4 | 1 | 3 | 0 | 26 | 282 | 0 | 316 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 80 | 80 |
| Total | 147 | 42 | 197 | 48 | 192 | 300 | 83 | 1009 |
| Overall accuracy = 83.94% | | | | | | | | |
| Kappa coefficient = 0.8004 | | | | | | | | |

Table 2. Classification confusion matrix for SVM

Classification achieves an overall accuracy 83.94% and Kappa coefficient 0.8004. It can be seen that the main error is coursed by confusion of roof, paths, and roads. The reason for that is the materials of the three are similar and they can hardly separate by spectral information.

## 3.3 Clique parameters estimation and classification

Classification map by SVM is applied as input for this step. As described before a robust LSF method is taken which help to acquire a more accurate estimate for MRFs. According to our experiment the estimation of robust LSF could result in a better classification. Observing $\theta$ in Table 3, we find that the parameters controlling the horizontal and vertical directions are obviously larger than others, which is consistent to the character of the image whose main topographic features, buildings, roads, paths are almost distributed horizontally or vertically.

| | Overall Accuracy | Kappa Coefficient | $\theta$ |
|---|---|---|---|
| PPC | 85.43% | 0.8188 | [0.70, 0.48, 0.02, 0.13] |
| MPC | 85.03% | 0.8137 | [0.75, 0.76, 0.28, -0.11, 0.04, -0.30, -0.32, 0.02, 0.13] |

Table 3 Classification accuracy obtained by LSF and robust

| Class number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | total |
|---|---|---|---|---|---|---|---|---|
| 1 | 106 | 6 | 24 | 0 | 0 | 2 | 0 | 138 |
| 2 | 20 | 34 | 0 | 0 | 0 | 0 | 0 | 54 |
| 3 | 17 | 0 | 168 | 9 | 13 | 4 | 0 | 211 |
| 4 | 1 | 0 | 3 | 37 | 0 | 0 | 1 | 42 |
| 5 | 0 | 0 | 0 | 2 | 153 | 12 | 0 | 167 |
| 6 | 3 | 2 | 2 | 0 | 26 | 282 | 0 | 315 |

| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 82 | 82 |
|---|---|---|---|---|---|---|---|---|
| Total | 147 | 42 | 197 | 48 | 192 | 300 | 83 | 1009 |
| Overall accuracy = 85.43% | | | | | | | | |
| Kappa coefficient = 0.8188 | | | | | | | | |

Table 4. Classification confusion matrix for SVM +PPC MRF

| Class number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | total |
|---|---|---|---|---|---|---|---|---|
| 1 | 106 | 6 | 25 | 0 | 0 | 2 | 0 | 139 |
| 2 | 20 | 34 | 0 | 0 | 0 | 0 | 0 | 54 |
| 3 | 17 | 0 | 164 | 8 | 13 | 4 | 0 | 206 |
| 4 | 1 | 0 | 2 | 38 | 0 | 0 | 1 | 42 |
| 5 | 0 | 0 | 2 | 2 | 151 | 10 | 0 | 165 |
| 6 | 3 | 2 | 4 | 0 | 28 | 284 | 0 | 321 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 82 | 82 |
| Total | 147 | 42 | 197 | 48 | 192 | 300 | 83 | 1009 |
| Overall accuracy = 85.03% | | | | | | | | |
| Kappa coefficient = 0.8137 | | | | | | | | |

Table 5. Classification confusion matrix for SVM+MPC MRF

From the confusion matrix we can see that the inclusion of contextual information by MRFs improved the classification accuracy. More roads and trees pixels were correctly classified. Overall accuracies of the two classifications are over 85% which is 1% greater than that of SVM. Kappa coefficients are over 0.81 and 0.01 greater than SVM's. However MPC does not notably differentiates itself in accuracy values with PPC.
From classification accuracies, there is no obvious difference of the two MRFs. Next, the classification maps are compared. Following three groups of subsets in the image are used to exam the capabilities of the MRFs.
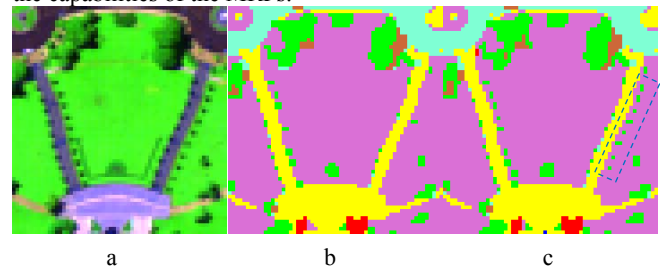


a        b        c
Figure 3. Classification details I
From a to c is HYDICE, PPC MRF, MPC MRF
Blue rectangle points the differences between PPC and MPC

As can be seen from Figure 3a, grass, trees roads and paths are the main feature in this plot. Most trees gather on the upside of the image and the others distributed alone the path. In Figure 2b and Figure 2c, something in common is that grass, roads and clustered trees are almost correctly classified while paths are all misclassified as roof. However the different is the trees along the paths which are smoothed a lot by PPC MRF while they are retained by MPC MRF (as the pointed in blue rectangle in Figure 3c).

a                           b                           c

Figure 4. Classification details II
From a to c is HYDICE, PPC MRF, MPC MRF
Blue rectangles point the differences between PPC and MPC

In Figure 4a, the main geographical features are building, roads, trees and grass. From Figure 4b and Figure 4c it can be seen that most part of the image is classified correctly and consistently except some trees and grass in the middle of the road are omitted in PPC MRF while they are not neglected by MPC MRF. In addition the same situation was encountered when there were isolated trees in the grass. The differences are displayed by blue rectangle in Figure 4c.
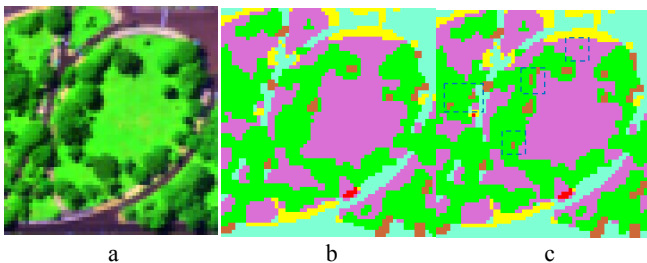


a                           b                           c

Figure 5. Classification details IIIs
From a to c is HYDICE, PPC MRF, MPC MRF
Blue rectangles point the differences between PPC and MPC

In Figure 5a there is a garden located in the middle of the road which is planted with grass and trees. From Figure 5b and Figure 5c it can be seen that roads, grass and trees are classified with no difficulty but the shadow beside the trees can't be recognized easily. Specifically, as pointed in the blue rectangle in Figure 5c, PPC MRF ignores the small ones but MPC MRF maintains the details.

In this experiment, MRF with PPC and MPC don't have significant difference in classification accuracy. However, MRF has an obvious advantage in preserving the isolated and small objects over the MRF with PPC which makes a relatively smooth classification and a lot of details are neglected. It indicates that the accuracy values do not reflect the differences in classification capabilities of the two MRFs. We speculate that the reason for this attributes to the test samples, 1009 altogether and 2.7% of all pixels, which are too few and some image details could not be test by their distribution. So we expect that the MRF with MPC could have an outstanding performance in classification accuracy for the case that test samples with adequate amount are available.

## 4. CONCLUSION

In this study we compared the classification performances by MRFs based on PPC and MPC. By estimating clique parameters, two MRFs entirely exhibit their capabilities in classification for HYDICE. The results reveal that, MRF with PPC and MPC do not exhibit an obvious difference in classification accuracy with limited test samples. MRF with PPC inclines to neglect the isolated and small pixels and

courses an information loss. However, MRF with MPC has an advantage in preserving the classification details.
A new clique parameter estimation method, a robust LSF, is proposed in the study. The new method results in a better classification in the experiment which means a better estimation is acquired. Hence, the proposed estimation method is worthy of being taken in application.

## 5. REFERENCES

Cui, Y., Wang, K., 2005. Markov random field modeling in the wavelet domain for image denoising. In: Proc. Internat. Conf. on Machine Learning and Cybernetics, 9, pp. 5382-5387.
G. R. Cross, A. K. Jain, 1983. Markov random field texture models. IEEE Trans. Pattern Anal. Machine Intell., PAMI-5, pp. 25-39

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell, 6 (6), pp.721-741.

H. Derin, H. Elliott, 1987. Modelling and segmentation of noisy and textured images using Gibbs random fields. IEEE Trans. Pattern Anal. Machine Intell., PAMI-9, no. 1, pp. 39-55.

J. Besag, 1986. On the statistical analysis of dirty pictures. J. R. Statist. Soc.B, 48, pp. 259-302

Lu, Q., Jiang, T., 2001. Pixon-based image denoising with Markov random field. Pattern Recognition, 34 (10), 2029-2039.

Melgani F., Serpico, S. B., 2003. A Markov random field approach to spatiotemporal contextual image classification. IEEE Transactions on Geoscience and Remote Sensing, 41(11), pp. 2478-2487

P. Andrey, P. Tarroux, 1998. Unsupervised segmentation of Markov random field modeled textured images using selectionist relaxation. IEEE Trans. Pattern Anal. Machine Intell., 20, pp. 252-262

Ryuei Nishii ,2003. A Markov random field-based approach to decision-level fusion for remote sensing image classification. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 41(10), pp.2316-2319

Solberg, A. H. S., Taxt, T., Jain, A. K., 1996. A Markov random field model for classification of multisource satellite imagery. IEEE Transactions on Geoscience and Remote Sensing, 34(1), pp.100-113.

Yang Cao, Yupin Luo, Shiyuan Yang, 2011.Image denoising based on hierarchical Markov random field. Pattern Recognition Letters, 32, pp.368-374

Yang Y, 2002. Robust estimator for correlated observations based on bifactor equivalent weights. Journal of Geodesy, 76, pp.353-358