IMPLEMENTATION OF THE MARKOV RANDOM FIELD FOR URBAN LAND COVER CLASSIFICATION OF UAV VHIR DATA

J. Pratomo [a], T. Widiastomo [b]

[a] *Lokalaras Indonesia Institute, Indonesia*
[b] *Institut Penelitian Inovasi Bumi, INOBU, Indonesia*

Corresponding Author:
Jati Pratomo
Lokalaras Indonesia Institute, Indonesia
Email: jati@lokalaras.org

Abstract: *The usage of Unmanned Aerial Vehicle (UAV) has grown rapidly in various fields, such as urban planning, search and rescue, and surveillance. Capturing images from UAV has many advantages compared with satellite imagery. For instance, higher spatial resolution and less impact from atmospheric variations can be obtained. However, there are difficulties in classifying urban features, due to the complexity of the urban land covers. The usage of Maximum Likelihood Classification (MLC) has limitations since it is based on the assumption of the normal distribution of pixel values, where, in fact, urban features are not normally distributed. There are advantages in using the Markov Random Field (MRF) for urban land cover classification as it assumes that neighboring pixels have a higher probability to be classified in the same class rather than a different class. This research aimed to determine the impact of the smoothness (λ) and the updating temperature (T_{upd}) on the accuracy result (κ) in MRF. We used a UAV VHIR sized 587 square meters, with six-centimetre resolution, taken in Bogor Regency, Indonesia. The result showed that the kappa value (κ) increases proportionally with the smoothness (λ) until it reaches the maximum (κ), then the value drops. The usage of higher (T_{upd}) has resulted in better (κ) although it also led to a higher Standard Deviations (SD). Using the most optimal parameter, MRF resulted in slightly higher (κ) compared with MLC.*

1. INTRODUCTION

The land cover represents composition and the characteristics of the land surface in a various spatiotemporal scale, and it is used in a various domain, e.g., resources management, policy purposes (Cihlar, 2000). The land cover is a product of the remote sensing (RS), initially obtained from the aerial photography, and it has been the most common application of satellite imagery data (Colwell, 1960). Satellite imagery has been demonstrated its efficiency to acquire information regarding earth's topography in a large spatial and spectral extent (Yan et al., 2015). However, although the most recent optical satellite sensor can provide a Very High Image Resolution (VHIR) (Yan et al., 2015), imagery is often contaminated by aerosols, cloud and cloud shadows (Liang et al., 2001), causing difficulties in the land cover classification process.

Unmanned Aerial Vehicle (UAV) based RS has been developed rapidly due to some advantages. Firstly, it has better flexibility than different RS methods, e.g., airborne and satellite (Lin et al., 2015). Secondly, UAV-based photography can provide an image with extremely high spatial resolutions, able to observe previously-undetectable features (Getzin et al., 2012). Thirdly, since the UAV flying at a low altitude, UAV-based photography is rarely affected by cloud cover (Rango et al., 2009). Fourthly, UAV has become available at an affordable price (Nasrullah, 2016). Due to its advantages, UAV-based RS has been implemented in a various field. For instance, Getzin et al. (2012) discussed the usage of Very High Image

Resolutions (VHIR) data obtained from UAV to assess the biodiversity in a forest. Regarding the application in the urban area, Lin et al. (2015) studied the use of oblique imaging to detect trees in the residential area. Another study from Yan et al. (2015) used data from Light Detection and Ranging (LiDAR) obtained from UAV for land cover classification.

To understand the land cover in a particular area, extracting information in the image is a crucial stage. It has attracted many researchers from the remote-sensing community, particularly to develop an advanced classification approach and techniques to improve the classification accuracy (Lu & Weng, 2007). The image classification is a process of generating groups of identical pixels into classes that match the information categories from the user (Perumal & Bhaskaran, 2010). Among different image classification methods, supervised classification has been developed and broadly used to tackle the classification problems in multispectral data (Perumal & Bhaskaran, 2010; Richards, 2013). Earlier studies have carried out land cover classification using Maximum Likelihood Classification (MLC) (Perumal & Bhaskaran, 2010). In general, MLC assign pixels into certain class having the highest probability membership (Maselli et al., 1994). Furthermore, it considers the variance and covariance among class distributions (Otukei & Blaschke, 2010). When applied to normally distributed data, MLC may perform better than others classifiers. However, low accuracy may occur when data with a non-normal distribution were implemented (Otukei & Blaschke, 2010).

Although UAV-based RS produces a higher spatial resolution compared with preceding sensors, e.g. Landsat TM, it may lead to complication in extracting urban features (Myint et al., 2006). When spatial resolutions increase, numerous small objects become further visible, which potentially lead to lower accuracy when it applied in an urban area (Myint et al., 2011). Among various classifiers that have been developed, Markov Random Field (MRF) were conceived as a stochastic approach to model a contextual information by including prior and posterior distribution of an image (Geman & Geman, 1984). This research aimed to implement an MRF approach in classifying urban features from UAV VHIR data. We explored different parameters to acquire the most appropriate parameter that produces the highest classification accuracy.

The MRF-based algorithm was proposed by considering the spatial correlation. It was developed according to the Toblers' First Law: "*Everything is related to everything else, but near things are more related than distant things*" (Tobler, 1970), which is the fundamental of the spatial dependence. Spatial dependence can be determined by the neighborhood systems, where neighboring pixels are more likely to fall under the same class than distant pixels. In the MRF model, each image has Markov properties according to spatial dependence, where isolated pixels are likely to be disappeared allowing a homogeneous region. Hence, the application of MRF is suitable for an image with a complex distribution of land cover class, such as that of a densely urban area.

2. DATA AND METHODS

2.1. Data Acquisition

Data used in this research were obtained from UAV RS, and the image was taken in Bogor Regency, Indonesia. For the UAV Platform, we used Ai450 version 2. This UAV is a fixed-wing platform and equipped with 24 Megapixels camera. This platform was designated to cruise in 60 to 90 kph, and it covers 70 km in distance for 70 minutes. Data acquisition process was provided in **Figure 1**.

This image was taken within two days under relatively similar conditions regarding wind speed, weathers and acquisition time. It takes between 2.00 PM to 5.00 PM. The flight speed was 36 to 54 kph at an altitude of 400 meters. Five Ground Control Points (GCP) was applied. For the orthophoto, we obtained Root Mean Square Meter (RMSE) at 0.0045 m. The spatial resolution of the image was 0.06 m.

Figure 1. Data Acquisition: (a) Platform preparation, (b) Before deployment, (c) Deployment, (d) landing (Authors, 2016)

2.2. Data Preparations

For training set, we developed samples which were limited into five class according to visual image interpretations (red roof, vegetation, concrete, dark roof, and bright roof) (**Figure 2b**). We only employed limited class due to two reasons. Firstly, we used a small subset (587 square meters). Secondly, we expected high difference among training data, and employing different classes within almost similar color will lead into lower class separability.

Figure 2. Subset Image and Training Set (Authors, 2016)

a. Image

b. Training set
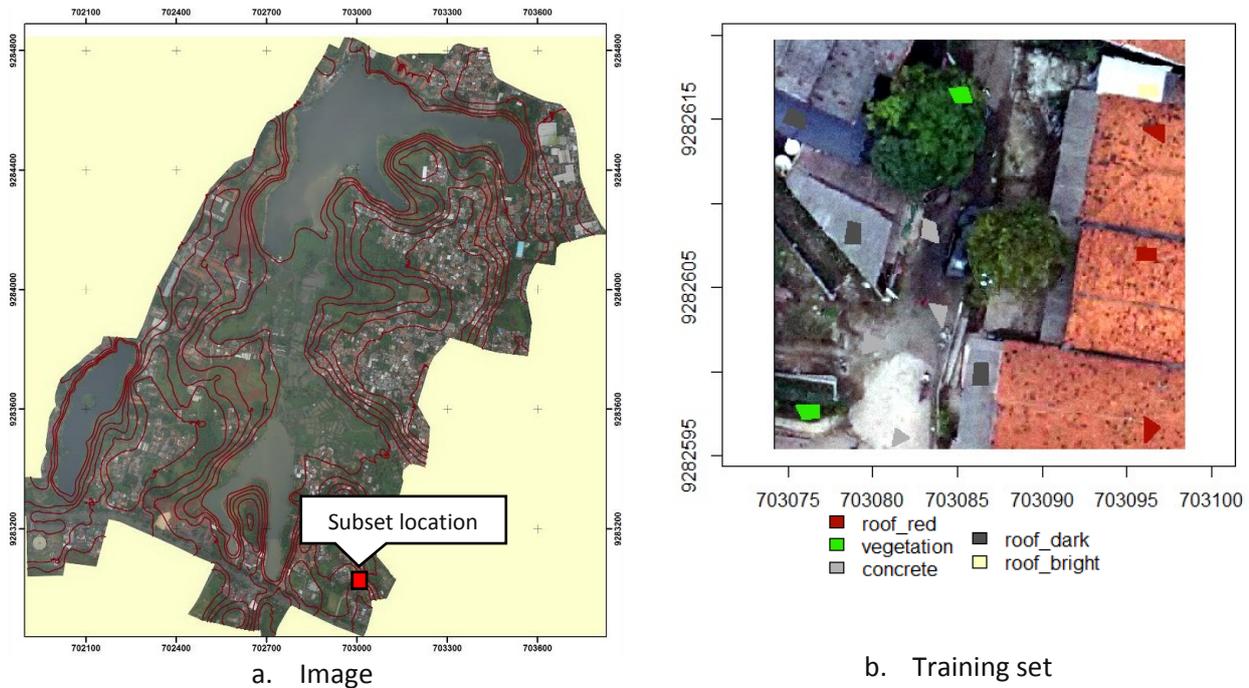
To ensure the quality of the training data set, we analyzed the class separability (**Table 1**). In this research, Jeffries-Matusita (JM) distance (see **equation 2**) was implemented to measure the separability between classes. It was widely used thanks to its advantage in suppressing the high separability values (Gunal & Edizkan, 2008). The JM distance transforms Bhattacharya (B) distance (see **equation 1**), to a value ranges from 0 to 2 because increasing B values do not imply that two classes have successfully separated. A higher value indicates higher separability between two classes.

$$B_{\alpha\beta} = \frac{1}{8}(\mu_\alpha - \mu_\beta)^T \left(\frac{C_\alpha + C_\beta}{2}\right)^{-1}(\mu_\alpha - \mu_\beta) + \frac{1}{2}\ln\left(\frac{\left|\frac{C_\alpha + C_\beta}{2}\right|}{\sqrt{|C_\alpha| \cdot |C_\beta|}}\right) \qquad [1]$$

$$JM_{\alpha\beta} = 2\left(1 - e^{-B_{\alpha\beta}}\right) \qquad [2]$$

According to **Table 1**, we notice that the sample is appropriate since the value of the JM distance is or near to 2, which indicate clear distinct among the class. However, slightly lower separability occurred between concrete and dark roof. Low-class separability will make the confusion between classes in the classification (Tolpekin & Stein, 2009).

Table 1. Class Separability of the Training Set (Analysis, 2016)

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	0	2	2.000	2.000	2
[2,]	2	0	2.000	2.000	2
[3,]	2	2	0.000	1.989	2
[4,]	2	2	1.989	0.000	2
[5,]	2	2	2.000	2.000	0

The scatterplot indicated that the class of concrete has spread from the lowest to the highest Digital Number (DN) (**Figure 3**). Concrete equally overlapped with the dark roof, which answered the lower class separability between them.

Figure 3. Scatter Plot of Training Class in Band 1 and 2 (Analysis, 2016)
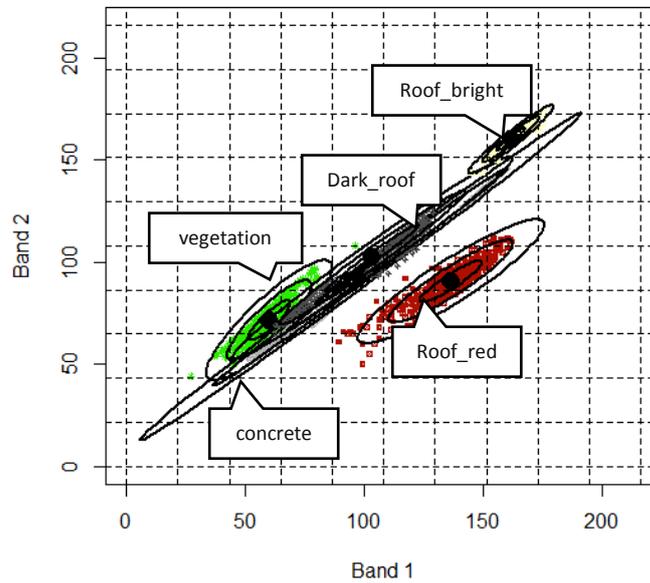
2.3. Markov Random Field

In our research, we employed the MRF model as previously shown by Ardila et al. (2011). They have successfully developed MRF model to identify urban trees by increasing the spatial resolution of the final classified image. We used the first order neighborhood system, which was considered as four adjacent pixels. It has a faster computation than the second order, which included eight adjacent pixels. The adjacent pixels included in our model were (1,0), (0,1), (1,1) and (1,-1) (**Figure 4**) covering the vertical, horizontal and diagonal adjacent pixels.

Figure 4. Neighborhood system implemented in this research (Ardila et al., 2011)
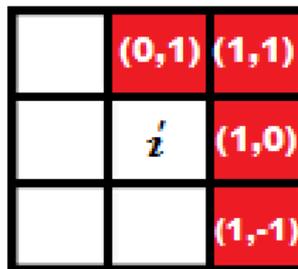
The weight of the contribution from the neighboring pixels ($w(a_l)$) controls the smoothness level of the prior model. A higher contribution leads to a smoother solution of the prior model, which increases the probability of same class label within a neighborhood system. The $w(a_l)$ is modelled as $w(a_l) = q \cdot \varphi(a_l)$, where the minimum value of q=0, and the maximum value is ∞. $\varphi(a_l)$ relies only on the distance $d(a_{j|i}, a_l)$. In this weighted system, a further distance gives a lower contribution. The prior energy function ($U(c)$) is aimed to model the contribution of adjacent pixels in a neighbourhood system, and defined as:

$$U(c) = \sum_{ij} U\left(c\left(a_{j|i}\right)\right)$$

[3]

$$= \sum_{ij} \sum_{l \in N(a_{j|i})} w(a_l) I\left(c(a_{j|i}), c(a_l)\right) \qquad [4]$$

Where

$U\left(c(a_{j|i})\right)$ = the local contribution to the prior energy from pixel $c(a_{j|i})$

$N(a_{j|i})$ = the neighbourhood system of pixels $a_{j|i}$

$w(a_l)$ = the weight of the contribution from the neighbouring pixel $a_l \in N(a_{j|i})$

$I\left(c(a_{j|i}), c(a_l)\right)$ takes the value 0 if $c(a_{j|i}) = c(a_l)$ and 1 otherwise

The likelihood energy function is used to model the probability of a pixel belong to which class in the classified image. In this research, the Maximum Likelihood Classifier was applied to implement the probability model. This likelihood model works by assuming each class in the image has a normal distribution. It computes the highest probability to label pixels into a particular class (Maselli et al., 1994). The likelihood model can be modelled by considers a normal distribution of a pixel b_i with value $y(b_i)$ with mean vector μ_i and covariance matrix C_i as:

$$U(y|c) = \sum_{i,j} U\left(y(b_i)\big|c(a_{i|j})\right) \qquad [5]$$

$$= \sum_{i,j} \left[\frac{1}{2}(y(b_i) - \mu_i)'C_i^{-1}(y(b_i) - \mu_i) + \frac{1}{2}\ln|C_i|\right] \qquad [6]$$

The posterior energy function is required to assign the pixels to a particular class. The posterior model incorporates the prior and likelihood model and can be defined as:

$$U(c|y) = q \sum_{ij} \sum_{l \in N(a_{j|i})} \varphi(a_l) I\left(c(a_{j|i}), c(a_l)\right) + U(y|c) \qquad [7]$$

An MRF-based image classification requires a smoothness parameter (λ) to complete the computation. This parameter determines the smoothness level of the classified image obtained from MRF process. The smoothness parameter should be defined by considering its impact on the accuracy of the result. A higher (λ) leads to a smoother result. A smoother classified image does not imply a better result. Therefore, the optimal value should be estimated to obtain the highest accuracy. The value of (λ) ranges from 0 to 1. It works in the posterior model by controlling the prior and likelihood model. The smoothness parameter (λ) was introduced in this model by dividing the **equation 7** by $1 + q$, which will transform the parameter so that it has the value one as the maximum value. The posterior model can be modelled as:

$$U(c|y) \propto \lambda \sum_{ij} \sum_{l \in N(a_{j|i})} \varphi(a_l) I\left(c(a_{j|i}), c(a_l)\right) + (1 - \lambda) U(y|c) \qquad [8]$$

In this research, the Simulated Annealing (SA) approach is used for implementing the energy optimisation. It is adapted from the chemical process of metals cool and reaches its desired shape. In this study, it will adapt the initial maximum likelihood classification to consider the spatial correlation. The energy optimisation is required to achieve the global optima instead of the local optima. In SA is widely used and works well as an optimisation algorithm, and it has been applied in wide range of areas (Aarts et al., 2005). SA is controlled by two main parameters, initial temperature (T_0) and updating temperature (T_{upd}).

$$T = T_0 \times T_{upd} \qquad [9]$$

Initially, the SA process starts at high (T_0) value and then cools down based on defined (T_{upd}). After employing the (T_{upd}), the (T_0) will decrease until reaching the equilibrium. The (T_{upd}) should be defined to make sure that the SA has enough iteration steps for obtaining the highest accuracy result in acceptable

computation time. The (T$_{upd}$) controls the rate of temperature decrease. The process will stop if there are no more pixels are updated. The final result is expected to be achieved by applying the smoothness parameter (λ) that adjusts the smoothness level. The energy minimization is used for maximisation of pixels' class labeling.

## 2.4. Experiment Design

In this research, the estimation of (λ) and (T$_{upd}$) was carried out by trial and error experiments. The value of the parameter (λ) and (T$_{upd}$) that leads to the highest kappa (κ) is chosen as the optimum value. For the (λ), we tested different value from 0.95 to 0.05, with an interval of 0.05. Meanwhile, for the (T$_{upd}$), we use five different value, which are 0.9, 0.75, 0.5, 0.25 and 0.1.

3. RESULTS AND DISCUSSION

According to the used training data (**Figure 2b**), the classification resulted in two images (**Figure 5**). On **Figure 5a**, we noticed classification resulted from the MLC, while **Figure 5b** was from MRF. In the area highlighted by red circle (**Figure 5**), we noticed the impact of the (λ) into classification result. MLC is resulting some isolated pixels (**Figure 5a**), while this pixel is disappeared in MRF (**Figure 5b**).

Figure 5. Classification Result from MLC and MRF (Analysis, 2016)

a.   MLC   b.   MRF
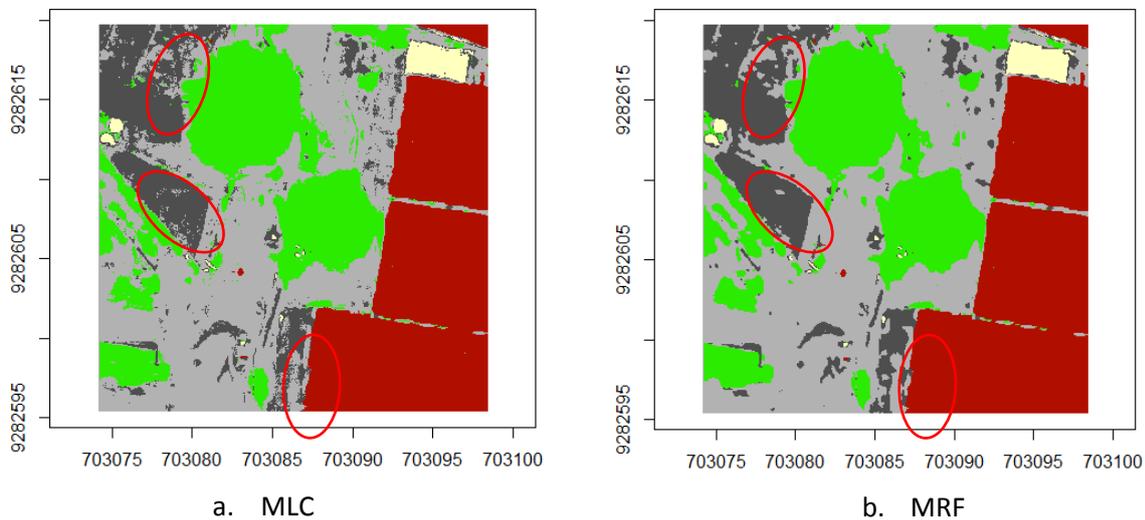
Regarding classification accuracy, among a different combination of (λ) and (T$_{upd}$), the highest (κ) obtained from MRF was 0.909679. This value was obtained from λ=0.9 and T$_{upd}$ = 0.9. Meanwhile, (κ) resulted from MLC was 0.908143, which means that MRF resulted slightly higher accuracy than MLC. This result is not unexpected, since Wang and Wang (2013) have proved that MRF generates a more reasonable result compared with the sub-pixel/pixel spatial attraction model, Hopfield neural networks (HNNs), HNN with SSRSI, image interpolation then hard classification.

Figure 6. (κ) value resulted from different (λ) and (T$_{upd}$) (Analysis, 2016)
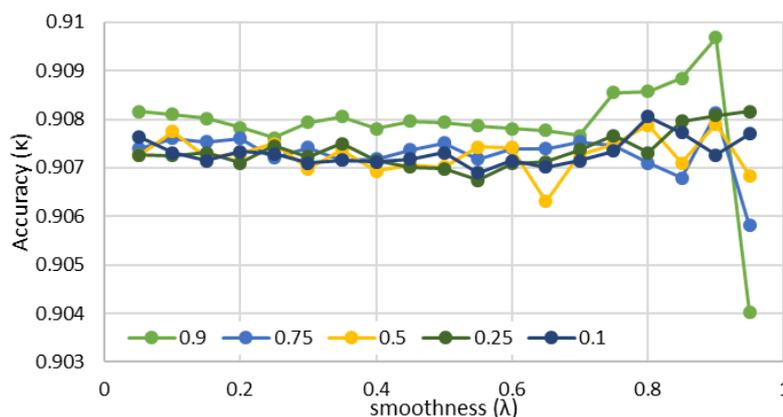
As we can see in **Figure 6**, the (κ) for T_{upd} = 0.9 are increased and reach its peak in λ=0.9 and suddenly the (κ) value is dropped. Since the (λ) controls the contribution of prior and likelihood models in the posterior energy (Tolpekin & Stein, 2009), the highest the (λ), the less considered the likelihood. However, if the (λ) is too high, it will lead to over smoothing (Tiwari et al. , 2015). The lowest (κ) is 0.904031, and this value was obtained in T_{upd} = 0.9 and λ=0.95. In one hand, the likelihood energy assumed that the land cover classes are evenly distributed. In another hand, the prior energy aimed to model the contribution of adjacent pixels in a neighborhood system. In the case of over-smoothing, the value of the particular class is only assigned according to their neighborhood class, while the mean and covariance of pixels from the training sample is ignored.

Regarding the impact of (T_{upd}) to (κ), we noticed from **Figure 7** that higher (T_{upd}) resulting a higher (κ), and T_{upd}=9 resulted the highest (κ). In average, T_{upd}=0.9 generates highest (κ) compared with other (T_{upd}). However, it also provide the highest standard deviations (SD) (**Figure 7**).

Figure 7. Average and Standard Deviation (SD) of (T_{upd}) among different (λ) (Analysis, 2016)
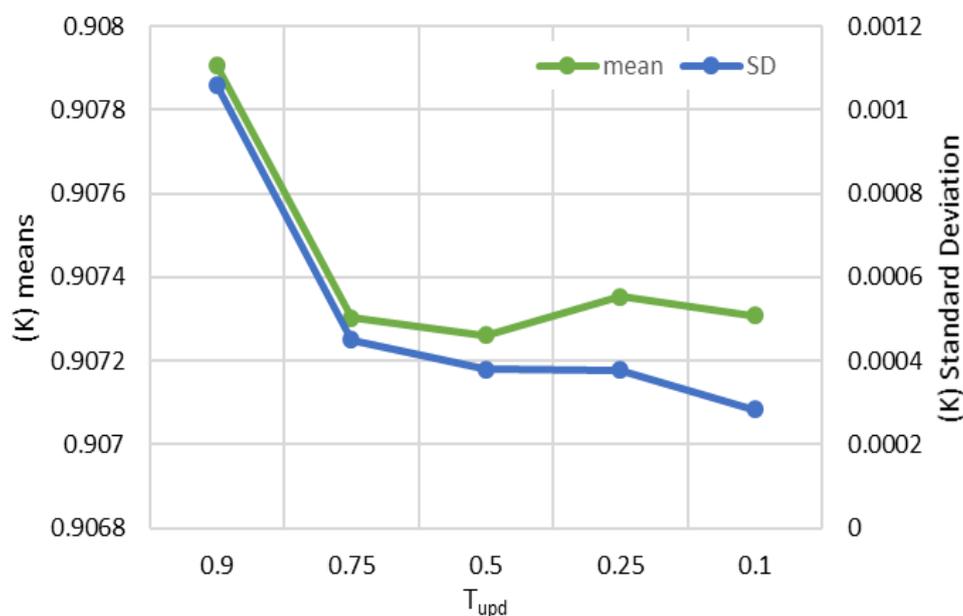
The (T_{upd}) has the trade-off between the accuracy of the result and the computation time (Lam, 1988). A higher (T_{upd}) generating a higher (κ) and requires more iteration steps for the process to reach its solution. The more iteration steps lead to an additional computation time. Moreover, the increasing of iteration steps will increase the probability of the process to reach a different equilibrium. **Figure 7** shows that the highest (T_{upd}) leads to the highest standard deviation of (κ), while the lowest (T_{upd}) creates the lowest standard deviation of (κ).

4. CONCLUSION

This research demonstrated the use of Markov Random Field for urban land cover classification from UAV VHIR data. We showed the impact of smoothness parameter (λ) and temperature update (T_{upd}) toward accuracy (κ). First, the usage of higher (λ) has resulted in better (κ), until it reaches its peak and suddenly the (κ) is dropped due to over smoothing. However, the (κ) resulted from MRF with the most optimal parameter is only slightly better than MLC. This result is not as we expected since we assume the usage of MRF might significantly increase the (κ). Second, the usage of higher (T_{upd}) has resulted in a better (κ) although it correspondingly leads to higher Standard Deviation (SD), since higher (T_{upd}) results in a higher probability of process to end in a different equilibrium. For further research, we recommend using a larger subset supported with better computation resources. Also, it is possible to use the estimation of the smoothness parameter, T_0 and T_{upd} as demonstrated by Tiwari et al. (2015).

5. ACKNOWLEDGMENTS

6. REFERENCES

Aarts, E., Korst, J., & Michiels, W. (2005). Simulated Annealing. In E. K. Burke & G. Kendall (Eds.), *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques* (pp. 187–210). Boston, MA: Springer US. http://doi.org/10.1007/0-387-28356-0_7

Ardila, J. P., et al. (2011). Markov-random-field-based super-resolution mapping for identification of urban trees in VHR images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6), 762–775.

Cihlar, J. (2000). Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, *21*(6–7), 1093–1114.

Colwell, R. N. (1960). *Manual for Photographic Interpretation*. Washington DC: The American Society of Photogrammetry.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741. http://doi.org/10.1109/TPAMI.1984.4767596

Getzin, S., Wiegand, K., & Schöning, I. (2012). Assessing biodiversity in forests using very high-resolution images and unmanned aerial vehicles. *Methods in Ecology and Evolution*, *3*(2), 397–404.

Gunal, S., & Edizkan, R. (2008). Subspace based feature selection for pattern recognition. *Information Sciences*, *178*(19), 3716–3726.

Lam, J. (1988). *An Efficient Simulated Schedule*. Yale University.

Liang, S., Fang, H., & Chen, M. (2001). Atmospheric correction of Landsat ETM+ land surface imagery. I. Methods. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(11), 2490–2498. http://doi.org/10.1109/36.964986

Lin, Y., et al. (2015). Use of UAV oblique imaging for the detection of individual trees in residential environments. *Urban Forestry & Urban Greening*, *14*(2), 404–412. http://dx.doi.org/10.1016/j.ufug.2015.03.003

Lu, D., & Weng, Q. (2007). Review article A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, *28*(5), 823–870. http://doi.org/10.1080/01431160600746456

Maselli, F., Conese, C., & Petkov, L. (1994). Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. *ISPRS Journal of Photogrammetry and Remote Sensing*, *49*(2), 13–20.

Myint, S. W., et al. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, *115*(5), 1145–1161.

Myint, S. W., Mesev, V., & Lam, N. (2006). Urban textural analysis from remote sensor data: Lacunarity measurements based on the differential box counting method. *Geographical Analysis*, *38*(4), 371–390.

Nasrullah, A. R. (2016). *Systematic Analysis of Unmanned Aerial Vehicle (UAV) Derived Product Quality*. Faculty of Geo-Information Science and Earth Observation of the University of Twente.

Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, *12*(SUPPL. 1), 27–31. http://doi.org/10.1016/j.jag.2009.11.002

Perumal, K., & Bhaskaran, R. (2010). Supervised classification performance of multispectral images. *arXiv Preprint arXiv:1002.4046*.

Rango, A., et al. (2009). Unmanned aerial vehicle-based remote sensing for rangeland assessment, monitoring, and management. *Journal of Applied Remote Sensing*, *3*(1), 33542.

Richards, J. A. (2013). *Remote Sensing Digital Image Analysis*. Springer Nature. http://doi.org/10.1007/978-3-642-30062-2

Tiwari, L. K., et al. (2015). Markov random field based method for super-resolution mapping of forest encroachment from remotely sensed ASTER image. *Geocarto International*, *31*(4), 428–445. http://doi.org/10.1080/10106049.2015.1054441

Tobler, W. R. (1970). A Computer Movie Simulation Urban Growth in Detroit Region. *Economic Geography*, *46*, 234–240. http://dx.doi.org/10.2307/143141

Tolpekin, V. A., & Stein, A. (2009). Quantification of the Effects of Land-Cover-Class Spectral Separability on the Accuracy of Markov-Random-Field-Based Superresolution Mapping. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(9), 3283–3297. http://doi.org/10.1109/TGRS.2009.2019126

Wang, L., & Wang, Q. (2013). Subpixel mapping using Markov random field with multiple spectral constraints from subpixel shifted remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, *10*(3), 598–602.

Yan, W. Y., Shaker, A., & El-Ashmawy, N. (2015). Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment*, *158*, 295–310. http://dx.doi.org/10.1016/j.rse.2014.11.001