THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Engineering (Information and Communication Technology for Embedded Systems)

DEGREE

Information and Communication Technology for Embedded Systems — **FIELD**

Electrical Engineering — **DEPARTMENT**

TITLE: A MRF-Based Approach for a Multisensor Land Cover Mapping of Mis- Registered Images

NAME: Mr. Ratchawit Sirisommai

THIS THESIS HAS BEEN ACCEPTED BY

..... **THESIS ADVISOR**

(Assistant Professor Teerasit Kasetkasem, Ph.D.)

..... **THESIS CO-ADVISOR**

(Associate Professor Apisit Eiumnoh, Ph.D.)

..... **THESIS CO-ADVISOR**

(Associate Professor Tsuyoshi Isshiki, Ph.D.)

..... **THESIS CO-ADVISOR**

(Mr. Preesan Rakwatin, Ph.D.)

..... **DEPARTMENT HEAD**

(Assistant Professor Teerasit Kasetkasem, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON .....

..... **DEAN**

(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

A MRF-BASED APPROACH FOR A MULTISENSOR LAND COVER MAPPING OF MIS-REGISTERED IMAGES

RATCHAWIT SIRISOMMAI

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Engineering (Information and Communication Technology for Embedded Systems)
Graduate School, Kasetsart University
2012

Ratchawit Sirisommai 2012: A MRF-Based Approach for a Multisensor Land Cover Mapping of Mis-Registered Images. Master of Engineering (Information and Communication Technology for Embedded Systems), Major Field: Information and Communication Technology for Embedded Systems, Department of Electrical Engineering. Thesis Advisor: Assistant Professor Teerasit Kasetkasem, Ph.D. 61 pages.

In this research, we proposed the joint registration and classification technique based on the Markov random field model (MRF) to obtain the land cover map of the scene. Instead of following the traditional approach to obtain the land cover map, where the registration and classification processes are performed separately, the proposed joint approach allows us to automatically evaluate the performance of both processes simultaneously. The expectation maximization (EM) algorithm is employed to solve our joint process problem by iteratively estimating the mapping parameters to obtain an optimal land cover map. In addition, spatial-context information of image pixels is exploited by using the Markov random field model to achieve better classification performance. We conducted experiments on two pairs of multisensor images to investigate the effectiveness of the proposed technique on practical use of satellite images. Results show that contextual information, derived from the MRF model, plays an important role in improving the classification performance. With our proposed algorithm, we achieve the correct mapping parameters and obtain the near perfect registration of multisensor images. However, as we continue to amplify the geometric difference of input images, we receive the poorer registration performance in the process. Our experiment on adjusting the initial mapping parameters suggests that the registration performance is most sensitive to errors in the scaling factors.

_____ _____ __ / __ / __

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

I wish to express my gratitude towards Geo-Informatics and Space Technology Development Agency (GISTDA) for providing PALSAR, THEOS images, and boundary data of land use of Sa Kaeo province. Assistance and expertise on processing SAR data from Mr. Amornchai Prakobya are greatly appreciated. Thanks are also due to Mr. Poonsak Miphoksap for kindly providing a PALSAR image and boundary data of sugarcane fields. The authors are grateful to Assoc. Prof. Mongkol Raksapatcharawong and researchers from the Satellite Management and Application Center (SMAC) for granting privileged access to SMMS data.

I would never be able to carry out this research without assistance, invaluable guidance, and kind suggestions from all of my advisors, especially Asst. Prof. Teerasit Kasetkasem, who promptly address the problems and oversee my progress without daunting. My deepest thanks are also due to my co-advisers including Assoc. Prof. Apisit Eiumnoh, Dr. Preesan Rakwatin, and Assoc. Prof. Tsuyoshi Isshiki.

This thesis would be a much rougher road without the assistance and helping hands provided by Ms. Pornpawee Kerdkor. I deeply appreciate her for being supportive since the first day of mine at ICTES Program. My special thanks go to Ms. Rattikar Punviset, Ms. Amornrat Khongma, Ms. Songjan Kasemsritanawat, and Ms. Taksa-on Putthaprasart for being kind and advisory when I was most desperate. I would be much slow in progress without the help from them. Finally, I would like to thank my parents and family for always being supportive throughout the course of this study and since I was born.

This research is financially supported by Thailand Advanced Institute of Science and Technology - Tokyo Institute of Technology (TAIST-Tokyo Tech), National Science and Technology Development Agency (NSTDA), Tokyo Institute of Technology (Tokyo Tech), National Research Council of Thailand (NRCT) and Kasetsart University (KU).

Ratchawit Sirisommai
October 2012

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Appendix Figure

LIST OF FIGURES (Continued)

A MRF-BASED APPROACH FOR A MULTISENSOR LAND COVER MAPPING OF MIS-REGISTERED IMAGES

INTRODUCTION

Remotely sensed images taken from satellites have been widely used for land cover mapping applications because of the capability to provide classification of different land cover types without having to physically assess the area of interest. In a situation where a single image does not provide sufficient classification performance, integrating multiple images of the same area is a common practice to increase the discrimination capability. In particular applications, especially agricultural field mapping, classification performance is benefited by using multitemporal sequences of satellite images. Moreover, multiple input images from different satellites can be used to further improve classification performance by providing better separation characteristics that a single sensor alone cannot do. A practical application is reported in (Lombardo *et al.,* 2003) where multitemporal sequences of synthetic aperture radar (SAR) images and a single optical image are used. Their results showed that overall discrimination performance was increased and agreed with other similar works where multisensor data are combined. (Skriver *et al.*, 2011) Emphasized the benefits of using multitemporal SAR images in short succession (weekly to monthly acquisitions) to crop classification, where the authors reported the improved classification accuracy by using multitemporal information. When it comes to multitemporal images, crop monitoring and classification problems have been recognized by many researchers in this field. Work presented in (Baghdadi *et al.,* 2009) is another example of several researches that exploits crop's phenology information to determine the growth stage of crops. They reported the notable accuracy achieved by using temporal profiles of Normalized Difference Vegetation Index (NDVI) extracted from different time series of optical images.

Although several studies have shown the potential of multitemporal and multimodal images to image classification, there exists a problem of image registration that is crucial to reliable result of land cover classification. The registration process is

required to align multiple satellite images into a common coordinate system. Only when all of input images are perfectly registered, classification algorithm can be applied to ensure the minimum error factored from the imperfect registration. However, the registration process is not always fully automated and usually requires a human operator to manually choose number of reference points known as control points (CPs). This process relies on individual's judgment on selecting number of appropriate CPs to properly register all input images. As a result, the land cover map (LCM) obtained in this way is not reliable and likely incorporates registration error that affects the final classification accuracy. Hence, the classification performance reflects registration performance. One might argue that there still exist a wide range of automatic registration algorithms that require little or no human supervision. However, these automated techniques do not always provide satisfactory results given diversity of applications and satellite data. Most importantly, the accuracy of the land cover map is not taken into account when applying the registration algorithms to available images.

In tradition, registration and classification are viewed as separate and sequential processes. However, based on the idea that the classification performance also depends on registration performance, we propose to consider the two processes as a single joint process. This allows us to evaluate registration and classification performances simultaneously and automatically. Another benefit of the joint registration and classification approach is that an algorithm can be designed in an attempt to maximize the accuracy of the LCM because, after all, the final product is the LCM not the co-registered images. Similar work was proposed in (Chen *et al.*, 2011) where joint image fusion and registration were considered. The expectation maximization (EM) algorithm was employed to solve the estimation problems of registration parameters and the true scene simultaneously. Different pairs of multisensor images were tested against the proposed joint process. Under the assumption that registration performance entails the quality of fusion result, the authors reported that better fusion performance can be achieved due to reduced registration errors. Our goal, however, is to obtain a LCM based on simultaneous estimation of registration and classification problems.

Under the assumption that pixels belonging to the same land cover class are likely close together in satellite images, we employ the Markov random field model (MRF) (Kindermann and Snell, 1980) to exploit this spatial information. Two satellite images from different sources are inputs of our proposed joint registration and classification algorithm. Our goal is to determine the mapping parameters for image registration and generate the LCM corresponding to the true scene simultaneously. The maximum likelihood (ML) estimate is used to estimate the mapping parameters. This allows us to determine the performance achieved by our proposed technique without concerning about adjusting parameters suitable for our problem, as in other sophisticated classifiers. However, the likelihood function is complex and cannot be practically computed. As a result, in this paper, the EM algorithm (Eliason, 2008) is used to approximate the likelihood function. We assume that affine transform exists between two images with imperfectly known mapping parameters including displacement and rotation. For each iteration of the EM algorithm, a new set of mapping parameters corresponding to the highest likelihood value is produced. Hence, the local optima point is obtained. In addition, the *a posteriori* probability is computed based on the current estimated mapping parameters. Finally, the optimum LCM under the maximum *a posteriori* can be directly obtained.

OBJECTIVES

According to the problem presented in section introduction, this thesis has the objective as follows.

1. To design an algorithm for registration and classification of multisensor satellite images.

2. To determine performance achieved by our proposed algorithm.

LITERATURE REVIEW

Image Registration

Image registration is a process of aligning two or more input images of the same scene. The input images might be taken at different times and/or by different types of sensor. The image registration is a crucial step and required prior to image interpretation and analysis in which final information is a result of integrating multiple data sources like image fusion, change detection, and land cover classification. The growing diversity of images due to development of image acquisition devices invokes the research on an automatic image registration process.

The image registration is widely adopted in various fields including remote sensing, medical imaging, satellite imagery, and computer vision. Its applications can be divided into four groups according to manner of image acquisition:

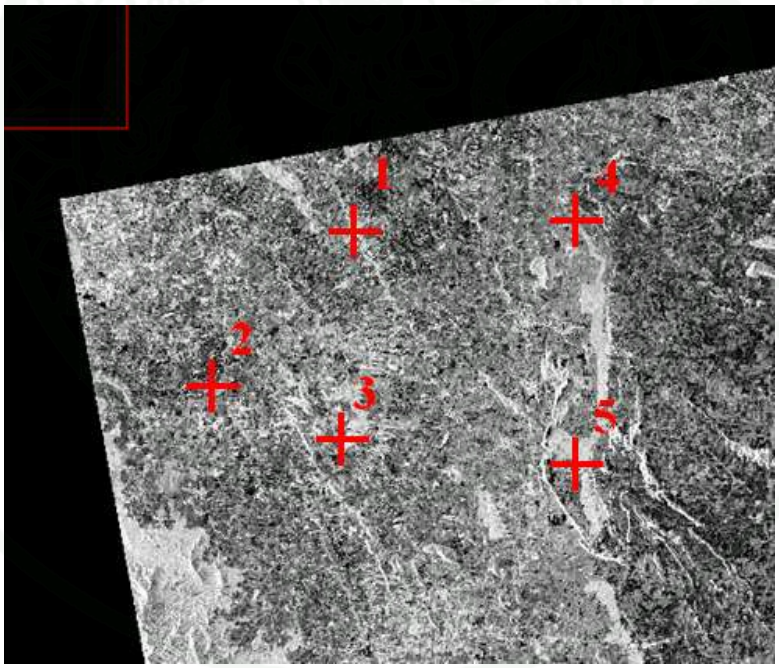
- Different viewpoints (multitier analysis): Images of the same scene are acquired at different viewpoints. Mosaicking of multiple scenes of satellite image is an example of this type of application.
- Different times (multitemporal analysis): Images of the same scene are acquired at different times, often on interval basis. The goal is to allow the analysis of changes in the scene as commonly found in crop monitoring applications in remote sensing.
- Different sensors (multimodal analysis): Images of the same scene are acquired by different sensors. The aim is to obtain more detailed information of the scene by integrating unique capability of information offered by different sensors. Examples of applications include image fusion and land cover classification that utilizes multisensor data in satellite images.
- Scene to model registration: Images and model of the same scene are registered, where a model can be computer representation of the scene such as digital elevation model (DEM) in GIS.

Due to the diversity of images and various types of noise and degradations, it is very difficult, suffice to say impossible so far, to design a universal method that is applicable to all registration tasks and scenarios (Fonseca and Costa, 1997). For instance, the type of geometric deformation has to be taken in to account when choosing a transformation model. According to (Zitova and Flusser, 2003), despite the difficulties of diversity mentioned above, registration methods can be generalized into following four steps:

- Feature detection: Distinctive objects (lines, corners, closed-boundaries regions, contours, edges, etc.) are manually or automatically detected. These detected features can be represented by their point representatives, which are commonly called control points (CPs), for processing in the next steps.
- Feature matching: The features detected in the sensed images and those detected in the reference image are matched and feature correspondence is established.
- Transform model estimation: The parameters of mapping functions used to align the sensed image with the reference image are estimated.
- Image resampling and transformation: The sensed image is transformed using the mapping functions. Appropriate interpolation technique is applied.

Figure 1 illustrates the feature detection and feature matching steps on remote sensing application where the symbols + and numerical labels denote the detected features and their correspondence pairs respectively.

(a)



(b)

Figure 1 Feature detection and matching on satellite images acquired by different sensors: (a) —a multispectral image, (b)—a SAR image.

Maximum Likelihood Estimation

The maximum likelihood estimate of the parameter(s) in the model is the value that maximizes the log-likelihood, given data being observed (Eliason, 2008).
Given the problem that we have the observed data $\mathcal{Y}$ but want to estimate the parameter(s) of model θ, we can write the conditional probability of the problem as $p(\theta|\mathcal{Y})$, and this is known as the inverse probability problem. By following Bayes' Theorem, the conditional probability of $p(\theta|\mathcal{Y})$ becomes

$$p(\theta|\mathcal{Y}) = \frac{p(\theta)p(\mathcal{Y}|\theta)}{p(\mathcal{Y})}. \quad (1)$$

Note that the denominator $p(\mathcal{Y})$ is not relevant to the parameters θ, thus it can be ignored. We can rewrite eq. (1) as

$$p(\theta|\mathcal{Y}) = k(\mathcal{Y})p(\mathcal{Y}|\theta) \quad (2)$$

, where $k(\mathcal{Y}) = \frac{p(\theta)}{p(\mathcal{Y})}$ is treated as a positive constant.

Since we cannot calculate the inverse probability $p(\theta|\mathcal{Y})$ directly, the likelihood notion introduced by R. A. Fisher (1912) (Aldrich, 1997) is used. The likelihood allows calculation from a traditional probability $p(\mathcal{Y}|\theta)$. We need to find the best estimator $\hat{\theta}$ that maximized the likelihood defined as

$$\mathcal{L}(\theta|\mathcal{Y}) \propto p(\mathcal{Y}|\theta). \quad (3)$$

Because of the proportional relationship, finding the best estimator that maximizes $p(\mathcal{Y}|\theta)$ will also maximize $\mathcal{L}(\theta|\mathcal{Y})$.

In case that there are multiple observed data $\mathcal{Y}_i$ and they are all statically independent, the likelihood of the whole data becomes the product of individual likelihoods.

$$\mathcal{L} = \prod_{i=1}^{N} p(\mathcal{Y}_i|\hat{\theta}) \tag{4}$$

Note that N is the total number of observations. Because it is simpler to work with the equation in a summation form, the likelihood function is often derived to the log-likelihood.

$$\ln \mathcal{L} = \sum_{i=1}^{N} \ln p(\mathcal{Y}_i|\hat{\theta}) \tag{5}$$

Expectation Maximization Algorithm

The expectation maximization (EM) algorithm (Dellaert, 2002) is an iterative optimization technique that is used to estimate some unknown parameters in the Maximum Likelihood estimation (MLE) with the presence of hidden data. The aim is to estimate the most likely model parameters given data being observed in MLE.

In each iteration of the EM algorithm, two steps are performed: the expectation and the maximization steps. In the expectation step, or so-called E-step, the distribution of hidden data is estimated given the observed data and the current estimates of the model parameters. This is done by using the conditional expectation. In the maximization step, or so-called M-step, the estimates of the distribution of the hidden data in the E-step are used to maximize the likelihood function. By using the EM algorithm, the likelihood is guaranteed to increase with every progressing iteration (Borman, 2009).

Let us reuse the variables previously defined in eq. (1), where $\mathcal{Y}$ is our observed data and θ is a model parameter(s). In the MLE, our objective is to choose values of θ that maximize $p(\mathcal{Y}|\theta)$, and in turn maximizing $\mathcal{L}(\theta|\mathcal{Y})$ as defined in eq. (3). We use the EM algorithm to maximize $\mathcal{L}(\theta|\mathcal{Y})$ in an iterative basis. Each of EM's iteration produces a new set of estimated θ. Suppose that we are at the t'th iteration, we wish to obtain the updated parameters θ_t such that

$$\mathcal{L}(\theta_t|\mathcal{Y}) > \mathcal{L}(\theta_{t-1}|\mathcal{Y}) \tag{6}$$

In other words, we want to maximize the difference of the likelihood in the successive iterations, which can be expressed as

$$\mathcal{L}(\theta_t|\mathcal{Y}) - \mathcal{L}(\theta_{t-1}|\mathcal{Y}) = \ln p(\mathcal{Y}|\theta_t) - \ln p(\mathcal{Y}|\theta_{t-1}) \tag{7}$$

The expression ensures that the likelihood function is increased with the updated parameter θ_t. Remember that the EM algorithm allows inclusion of the hidden data. So, let us introduce the hidden variables to eq. (7) by denoting them as X and the realization of them as x. We may express $p(\mathcal{Y}|\theta_t)$ given the hidden variables as

$$p(\mathcal{Y}|\theta_t) = \sum_x p(\mathcal{Y}|x, \theta_t)\, p(x|\theta_t) \tag{8}$$

The difference of the likelihood eq. (7) can then be rewritten as

$$\mathcal{L}(\theta_t) - \mathcal{L}(\theta_{t-1}) = \ln \sum_x p(\mathcal{Y}|x, \theta_t)\, p(x|\theta_t) - \ln p(\mathcal{Y}|\theta_{t-1}) \tag{9}$$

Note that, to solve eq. (9), we need to compute the logarithm of sum, which is difficult. We can transform the equation into a sum of logarithms by using Jensen's inequality, i.e.

$$\begin{aligned}\mathcal{L}(\theta_t) - \mathcal{L}(\theta_{t-1}) &= \sum_x p(x|\mathcal{Y}, \theta_{t-1}) ln\left(\frac{p(x|\mathcal{Y}, \theta)p(\mathcal{Y}|\theta)}{p(x|\mathcal{Y}, \theta_{t-1})p(\mathcal{Y}|\theta_{t-1})}\right) \\ &\triangleq \Delta(\theta_t|\theta_{t-1})\end{aligned} \tag{10}$$

We can continue by writing

$$\mathcal{L}(\theta_t) \geq \mathcal{L}(\theta_{t-1}) + \Delta(\theta_t|\theta_{t-1}). \quad (11)$$

The EM algorithm can be explained in terms of lower bound maximization (Neal and Hinton, 1998). At each iteration, an optimal lower bound given the current θ_{t-1} is computed in the E-step, and then this bound is maximized in the M-step and we obtain the improved estimated of the model parameter θ_t. Thus, repeating the produre ensures that the value of the likelihood function $\mathcal{L}(\theta_t|\mathcal{Y})$ increases with each iteration. The optimal bound can be expressed as

$$l(\theta_t|\theta_{t-1}) \triangleq \mathcal{L}(\theta_{t-1}) + \Delta(\theta_t|\theta_{t-1}). \quad (12)$$

We can derive the relationship from eq. (11) and (12) as

$$\mathcal{L}(\theta_t) \geq l(\theta_t|\theta_{t-1}). \quad (13)$$

So, $l(\theta_t|\theta_{t-1})$ is bounded above by the likelihood function $\mathcal{L}(\theta_t)$. Our goal is to find θ that maximizes $\mathcal{L}(\theta)$. Because any increase in $l(\theta_t|\theta_{t-1})$ will also increase $\mathcal{L}(\theta_t)$, in the EM algorithm, we select new θ, denoted by θ_t, such that $l(\theta_t|\theta_{t-1})$ is maximized to obtain the greatest possible value of $\mathcal{L}(\theta_t)$. The process can be expressed as

$$\begin{aligned}\theta_t &= arg\max_{\theta}\{l(\theta_t|\theta_{t-1})\} \\ &= arg\max_{\theta}\left\{\mathcal{L}(\theta_{t-1}) + \sum_x p(x|\mathcal{Y},\theta_{t-1}) ln\left(\frac{p(x|\mathcal{Y},\theta)p(\mathcal{Y}|\theta)}{p(x|\mathcal{Y},\theta_{t-1})p(\mathcal{Y}|\theta_{t-1})}\right)\right\} \\ &= arg\max_{\theta}\left\{E_{x|\mathcal{Y},\theta_{t-1}}\{\ln p(\mathcal{Y},x|\theta)\}\right\} \end{aligned} \quad (14)$$

In the case that the maximization is difficult, the EM algorithm can be relaxed by simply selecting θ_t that increases $\Delta(\theta_t|\theta_{t-1})$ but not necessarily maximizing it. This approach is called the Generalized Expectation Maximization (GEM) algorithm.

Markov Random Fields Model

Markov random fields (MRF) model is another type of stochastic process that has been introduced in the probability theory. The concept of the MRF model is originated from attempts to put a very specific model, named the Ising model, into a general probabilistic setting. The first formulation of Ising model tries to measure the probability of possible configurations in a sequence of a small dipole called spin. At any moment, each of spin is either "up" or "down" signifying binary variables. And the probability measure is called a random field. There are two cases for neighboring interactions: the attractive case where spins are of the same orientation and the repulsive case where the spins are of opposite orientation (Kindermann and Snell, 1980).

Given the neighborhood system N, an MRF is a random field called X if for all sites $s \in \mathcal{S}$, the random variable $X(s)$ and $X(\mathcal{S}\backslash\widetilde{N}_s)$ are independent. We can express this system in terms of a mathematical equation as

$$\begin{aligned} p\{X(s) = x(s)|X(\mathcal{S}\backslash s) = x(\mathcal{S}\backslash s)\} \\ = P\{X(s) = x(s)|X(N_s) = x(N_s) = x(N_s)\} \end{aligned} \tag{15}$$

for all $s \in \mathcal{S}$. We go further by defining a clique, which is any singleton $\{s\}$. A clique of a graph is a subset $C \subset \mathcal{S}$ with more than one element, if and only if any two distinct sites of C are mutual neighbors. Next, we explore the definition of a Gibbs potential. A Gibbs potential on $\Lambda^{\mathcal{S}}$ relative to the neighborhood system N is a collection $\{V_C\}_{C\subset\mathcal{S}}$ of an energy function V_C: $\Lambda^{\mathcal{S}} \to \Re \cup \{+\infty\}$ such that

- $V_C \equiv 0$ if C is not a clique,
- For all $x, x' \in \Lambda^{\mathcal{S}}$ and all $C \subset \mathcal{S}$,
- $(x(C) = x'(C)) \Longrightarrow \big(V_C(x) = V_C(x')\big)$.

The energy function is derived from the potential $\{V_C\}_{C\subset\mathcal{S}}$ if

$$E(x) = \sum_{C} V_C(x). \tag{16}$$

A Gibbs field and an MRF share similar properties in the sense that a Gibbs field is defined over the neighborhood system N using the potential function and cliques. Similarly, an MRF is defined over the same neighborhood system by means of the local characteristics.

The MRF model can be used to model context-dependent entities such as image pixels with correlated features. MRF can also be used to describe the undirected graphical models, in which undirected edges represent the inter dependency among vertices.

Related Works

The registration and classification are traditionally treated as two separate processes, however, as the practice of using multiple input images to increase discrimination capability has become prominent to satellite imagery application, some researchers has proposed a joint approach to mitigate the effect of registration errors. This kind of error will result in reduced performance in a later step of application due to the nature of sequential execution presenting in the traditional practice. Lombardo, et al. (Lombardo *et al.,* 2003) has demonstrated the implementation of joint approach in satellite image segmentation problem where practical example of combining multitemporal synthetic aperture radar (SAR) images with a single multispectral optical image was the subject. In this work, the Markov random field (MRF) model was used to exploit the spatial context information of pixels and the joint segmentation is performed by identifying homogeneous regions. Classification performance was used to determine the performance of the proposed joint segmentation technique since the authors claimed that the quality of segmentation could determine the classification performance. The results suggested that the proposed segmentation technique increase the overall discrimination performance. The major advantage of this technique, when separating the joint segmentation process from the classification process, is that it is totally unsupervised. However, the authors reported the possible small fragments of

imperfect overlapping of the identified segments that resulted in degraded classification accuracy.

Another similar work of joint approach was proposed by (Chen *et al.*, 2011) where registration and fusion processes were integrated. The expectation maximization algorithm was employed to solve the estimation problems of registration parameters and the true scene simultaneously. Different pairs of multisensor images were tested against the proposed joint process. Under the assumption that registration performance entails the quality of fusion result, the authors reported that better fusion performance can be achieved due to reduced registration errors.

Among several supervised classifiers, the maximum likelihood (ML) estimation is common to the classification application of remotely sensed images due to its implementation simplicity. Although ML estimation yields acceptable results when training data provides sufficient discrimination characteristics, the classification is applied on pixel by pixel basis, thus negating the contextual information presented in remotely sensed images. One can derive context information by observing the pixels of the same class which are more likely close together. Incorporating this contextual information when estimating the probability of class labels would result in better class estimation as pixels are not treated separately. (Tso and Mather, 1999) Propose a work that exploits the contextual information by using the MRF model. The proposed technique proved to improve classification performance in terms of average producer's accuracy on multisource data comprising of optical and SAR images. However, the cost of computational time was not worth the accuracy improvement in some cases, leaving further investigation on how to efficiently estimate the model parameters to obtain higher accuracy.

Despite widely use of supervised classifiers in land cover mapping application of satellite images, unsupervised classification is still considered viable for the classification application due to its capability to deliver the valid results without the direct guidance from the expert. In (Korting *et al.*, 2008), expectation-maximization (EM) was applied to unsupervised classification of satellite images. The image pixels

were treated as the input vectors and mean and variance are the parameters to be estimated. In terms of class matching, the results showed better performance when compared to another unsupervised classifier called self-organizing maps (SOM). By visual inspection, EM also yielded better result in terms of smoothness and did not require post-processing to correct the spatial correlation. This can prove beneficial when incorporating EM with classification that exploits the spatial context information. However, in their work, EM still suffered from spectral similarities as pixel intensity was the only discriminating feature. Another drawback of EM was the computational time that was far higher than the opposing classifier. This is due to the computational cost of matrix inversion operations presented in each of EM iteration.

MATERIALS AND METHODS

This chapter is started by materials used in this study. Then, the methods of investigation and implementation of proposed approach are presented.

Materials

The proposed algorithm was implemented using Python as a development language. The experiments of the developed program was partially executed by a graphical processing unit (GPU) by Nvidia. The computation on GPU allowed for parallel processing, hence it increased the speed of execution of our program.

Methods

Problem Statement

Since our ultimate goal is to obtain the LCM, let $X(\mathcal{S})$ denote the LCM where $\mathcal{S}$is a set of pixels. We assume that there are L land cover classes in the area of interest and we let $\Lambda \in \{0, 1, \ldots, L-1\}$ be the class labels. Therefore, we can express the LCM as $X(\mathcal{S}) \in \Lambda^{\mathcal{S}}$.

It is worth nothing that we employ MRF to take advantage of contextual information of neighboring pixels, which likely belong to the same class. Assuming that the LCM satisfies the MRF properties with Gibbs potential $V_C(X)$, the marginal probability density function (PDF) of MRF can be written as

$$\Pr(X) = \frac{1}{Z_X} \exp\left(- \sum_{C \subset \mathcal{S}} V_C(X) \right) \tag{17}$$

where Z_X is a normalizing constant, C is a clique, and $E(X) = \sum_{C \subset \mathcal{S}} V_C(X)$ is called the Gibbs energy function (Geman & Geman, 1984). Cliques are groups of pixels that share similar characteristics with each other but show little similarity to other pixels outside the group, we can model our problem using cliques to identify subsets of

pixels belonging to the same class. Gibbs energy level corresponds to the clique configuration. Usually, low value of Gibbs energy indicates similarity while high value suggests dissimilarity of the configuration. Ising model (Kindermann and Snell, 1980), which is used to identify states of spins interacting with their neighbors, is applied to our work to describe the LCM. Interaction of neighboring pixels can be classified as follows, where N_s denotes a set of a neighboring s.

$$V_{\{s,r\}}(X) = \begin{cases} -\beta; if\ x_s = x_r\ and\ r \in N_s \\ \beta; if\ x_s \neq x_r\ and\ r\ \in N_s \\ 0; if\ r \notin N_s \end{cases} \quad (18)$$

We propose the joint process of registration and classification that requires multiple images as inputs. Hence, we assume to have N images acquired from different sensors and/or time denoted as $Y_n(\mathcal{S}) \in \Re^{\mathcal{S}\times B_n}$, where, B_n denotes the number of spectral bands of its corresponding Y_n. One of $Y_n$is chosen and assumed to be a perfect reference to the true scene, while the rests may or may not perfectly registered to the chosen Y_n, hence they are needed to be transformed in order to be perfectly aligned with the true scene. For demonstration purposes, assume we have two observed images namely Y_1 and Y_2. Given observation of both Y_1 and Y_2 are statistically independent, we can write PDF of the scenes being observed as

$$p\big(Y_1(\mathcal{S}), Y_2(\mathcal{S})|X(\mathcal{S})\big) = p\big(Y_1(\mathcal{S})|X(\mathcal{S})\big)p\big(Y_2(\mathcal{S})|X(\mathcal{S})\big). \quad (19)$$

Moreover, observation of each individual pixels in $Y_1(\mathcal{S})$ and $Y_2(\mathcal{S})$ is also assumed to be statically independent. Hence, the PDF of each pixel can be written as

$$p\big(Y_a(\mathcal{S})|X(\mathcal{S})\big) = \prod_{s\in\mathcal{S}} p\big(y_a(s)|x(s)\big) \quad (20)$$

where Y_a indicates the image being observed, given $a \in \{1, 2\}$, y_a denotes a vertical vector whose number of elements equal to the number of band, that is $y_a \in \Re^{Ba}$, and $x(s)$ is the class label of pixel s.

As mentioned in the previous section, in order to obtain a reliable LCM, each of the input images has to be well aligned. In this work, registration parameters denoted as θ_a are used to register one input image to another using affine transform, i.e.,

$$\theta_a = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ \gamma_4 & \gamma_5 & \gamma_6 \end{bmatrix} \tag{21}$$

The registration parameter θ_a composes of scale factors denoted by γ_1 and γ_5, skewness denoted by γ_2 and γ_4, and displacement denoted by γ_3 and γ_6. Taking mapping parameters θ_a into account, the conditional PDF of $Y_a(\mathcal{S})$ eq. (20) becomes

$$p\big(Y_a(\mathcal{S}, \theta_a)|X(\mathcal{S})\big) = \prod_{s \in \mathcal{S}} p\big(y_a(s, \theta_a)|x(s)\big). \tag{22}$$

The maximum *a posteriori* (MAP) is employed to estimate the most likely LCM given the observations (Y_a) and registration parameters. Hence, the optimized LCM can be expressed as

$$X^{opt} = arg\left\{\max_X\big[Pr\big(X|Y_1(\theta_1), Y_2(\theta_2)\big)\big]\right\} \tag{23}$$

where $Pr\big(X|Y_1(\theta_1), Y_2(\theta_2)\big)$ is the posterior probability of the LCM with known registration parameters and input images being observed. It is to be noted that $\mathcal{S}$ will be omitted for the rest of the section to reduce clutter in equations. According to the conditional probability, eq. (23) can be written as

$$X^{opt} = arg\left\{\max_X\left[\frac{p\big(Y_1(\theta_1), Y_2(\theta_2)|Pr(X)\big)}{p\big(Y_1(\theta_1), Y_2(\theta_2)\big)}\right]\right\}. \tag{24}$$

However, the denominator part $p\big(Y_1(\theta_1), Y_2(\theta_2)\big)$ can be omitted since it is independent of X. Hence, the eq. (24) can be reduced to

$$X^{opt} = arg\left\{\max_X\left[p\big(Y_1(\theta_1), Y_2(\theta_2)|Pr(X)\big)\right]\right\}. \tag{25}$$

The final goal of this work is to obtain the LCM that is closest to the true scene given input images and registration parameters. The energy function is used as a mean to determine the most likely LCM (X^{opt}). By minimizing this energy function, the LCM obtained is likely close to the optimal solution. Therefore, we substitute eq. (17) and eq. (22) into eq. (25), and we obtain

$$X^{opt} = arg\left\{\max_X\left[A \cdot exp\left(-E\big(X, Y_1(\theta_1), Y_2(\theta_2)\big)\right)\right]\right\} \tag{26}$$

where $A = \frac{1}{Z}$ and the energy function of a LCM (X) given observations with registration parameters $\big(Y_1(\theta_1)\ and\ Y_2(\theta_2)\big)$ is

$$E\big(X, Y_1(\theta_1), Y_2(\theta_2)\big) = -\sum_{a\in\{1,2\}}\sum_{s\in\mathcal{S}} log\left(p\big(y_a(s, \theta_a)|x(s)\big)\right) - \sum_{C\subset\mathcal{S}} V_C(X). \tag{27}$$

Since A is a constant and the exponential function is monotonic, eq. (26) can be further reduced to

$$X^{opt} = arg\left\{\min_X\big[E\big(X, Y_1(\theta), Y_2(\theta_2)\big)\big]\right\}. \tag{28}$$

We can conclude from eq. (28) that the LCM can be obtained by minimizing the energy function. However, the energy function as defined in eq. (27) is a non-convex function, and it cannot be solved efficiently by a general optimization technique given a vast possibility of LCM. Therefore, eq. (27) is approximated using the mean field theory (Crawford, 2007). The idea is to acquire one effective

interaction among all interactions to one pixel. To demonstrate the idea, the potential function defined in eq. (18) is replaced by its expected value as

$$V_{\{s,r\}}\big(x(s),x(r)\big) \approx \langle V_{\{s,r\}}\big(x(s),x(r)\big)\rangle_{x(r)|Y_1(\theta_1),Y_2(\theta_2)} \tag{29}$$

where $\langle V_{\{s,r\}}\big(x(s),x(r)\big)\rangle_{x(r)|Y_1(\theta_1),Y_2(\theta_2)}$ denotes the expected value of $V_{\{s,r\}}\big(x(s),x(r)\big)$ over the posterior probability $Pr\big(X|Y_1(\theta_1),Y_2(\theta_2)\big)$. The energy function in eq. (27) can be approximated as

$$E\big(X,Y_1(\theta_1),Y_2(\theta_2)\big) \approx \sum_{s\in S} W\big(x(s),y_1(s,\theta_1),y_2(s,\theta_2)\big) \tag{30}$$

where

$$\begin{aligned} &W\big(x(s),y_1(s,\theta_1),y_2(s,\theta_2)\big) \\ &= -\sum\nolimits_{a\in\{1,2\}} \log\Big(\rho\big(y_a(s,\theta_a)|x(s)\big)\Big) \\ &+ \sum\nolimits_{r\in N(s)} \langle V_{\{s,r\}}\big(x(s),x(r)\big)\rangle_{x(r)|y_1(s,\theta_1),y_2(s,\theta_2)} \end{aligned} \tag{31}$$

Since eq. (31) does not depend on $X(s)$ but $x(s)$, the optimization problem is reduced to searching for the individual $x(s)$. The approximation of energy function in eq. (30) implies that the posterior probability of X, given two observed input images, can also be approximated as

$$Pr\big(X|Y_1(\theta_1),Y_2(\theta_2)\big) \approx B\prod_{s\in S} e^{-W\left(x(s),y_1(s,\theta_1),y_2(s,\theta_2)\right)}. \tag{32}$$

From eq. (32), the posterior probability of a single pixel $x(s)$ can be approximated as

$$Pr\big(x(s)|Y_1(\theta_1),Y_2(\theta_2)\big) \approx Be^{-W\left(x(s),y_1(s,\theta_1),y_2(s,\theta_2)\right)} \tag{33}$$

where B is normalizing constant. The equation will be applied in the parameters estimation step where expectation and maximization algorithm (EM) is employed.

Optimum LCM and Mapping Parameters Estimated

In this study, the maximum likelihood estimation (MLE) is employed to estimate the registration parameters of mis-registered images and the LCM. The parameter estimation can be expressed as

$$(\theta_1, \theta_2)^{opt} = arg\left\{\max_{(\theta_1,\theta_2)}\left[\sum_{X\in\Lambda^S} p(Y_1(\theta_1), Y_2(\theta_2)|X)Pr(X)\right]\right\}. \quad (34)$$

The EM algorithm is employed in this study to find the solution of MLE in eq. (34). The algorithm is composed of two iterative steps. In the first step, called E-step, EM is used to find the expected value of the log-likelihood function of the observed input images and the LCM given the estimated registration parameters. This step is called E-step and it can be expressed as

$$Q(\theta_1, \theta_2 \parallel \theta_1^t, \theta_2^t) = \langle \log\big(p(Y_1(\theta_1), Y_2(\theta_2)|X)\big) + \log\big(p(X)\big)\rangle_{X|Y_1(\theta_1^t),Y_2(\theta_2^t)}. \quad (35)$$

It can be seen from eq. (35) that the expected values require the posterior probability $p(Y_1(\theta_1^t), Y_2(\theta_2^t)|X)$. Therefore, the expected values can be approximated as in eq. (32) and (33)

$$\begin{aligned} &Q(\theta_1, \theta_2 \parallel \theta_1^t, \theta_2^t) \\ &\approx D\sum_{s\in S}\sum_{l=0}^{L-1} \log\big(p(y_1(s,\theta_1), y_2(s,\theta_2)|x(s)=l)\big) \\ &\times e^{-W\left(x(s)=l, y_1\left(s,\theta_1^t\right), y_2\left(s,\theta_2^t\right)\right)} \end{aligned} \quad (36)$$

where

$$D = B\sum_{s\in S}\sum_{l=0}^{L-1}\log\big(p(X)\big)e^{-W\left(x(s)=l,y_1\left(s,\theta_1^t\right),y_2\left(s,\theta_2^t\right)\right)}.$$

After the E-step, a new set of registration parameters denoted by θ^{t+1} is estimated by maximizing eq. (35), i.e.

$$(\theta_1^{t+1},\theta_2^{t+1}) = \max_{(\theta_1,\theta_2)}[Q(\theta_1,\theta_2 \parallel \theta_1^t,\theta_2^t)]. \tag{37}$$

Equation (37) demonstrates the second step of EM algorithm called M-step and it can be approximated by using eq. (36) as

$$(\theta_1^{t+1},\theta_2^{t+1}) = \max_{(\theta_1,\theta_2)}\left[D\sum_{s\in S}\sum_{l=0}^{L-1}\log\big(p(y_1(s,\theta_1),y_2(s,\theta_2)|x(s)=l)\big) \times e^{-W\left(x(s)=l,y_1\left(s,\theta_1^t\right),y_2\left(s,\theta_2^t\right)\right)}\right] \tag{38}$$

Combining the results from eq. (28) and (38), the LCM can be obtained through our joint registration and classification technique. Assuming there are two input images with possible mis-registration, our algorithm can be written as follows.

1. Initialize the *a posteriori* probability of LCM, i.e. , $Pr\big(x(s)=l|Y_1(\theta_1),Y_2(\theta_2)\big)=\frac{1}{L}$ where L denotes number of class labels. Also, set registration parameters to some initial values for the first iteration.

2. Compute the expected value of potential function $\langle V_{\{s,r\}}\big(x(s),x(r)\big)\rangle$ using eq. (29).

3. Estimate a posteriori probability using eq. (33) and choose a normalizing constant B such that the total probability is one.

4. Estimate a new set of registration parameters using eq. (38)

5. Check if stopping criteria is met, i.e., the maximum iteration is reached or no significant changes in LCM for repeating iterations. Compute the final LCM using eq. (27) and (30) when the stopping criteria are satisfied. Otherwise, repeat step 2 with the updated parameters.

Obtaining Estimated Mapping Parameters of Likelihood

In each step of the EM algorithm, we need to find the estimations of mapping parameters that maximize the likelihood function. Doing exhaustive search, in order to obtain six parameters of θ that produce the largest likelihood, is not practical in our scenario. Therefore, we choose a numerical approach in which an optimization algorithm is employed to obtain the mapping parameters. In our work, the Newton's method (Hauser, 2012) is chosen.

Implementation of the Newton's method into our program involves guessing initial mapping parameters θ and approximating new sets of θ until $W\big(x(s), y_1(s,\theta_1), y_2(s,\theta_2)\big)$ is smallest. Sequential steps for obtaining the estimated set of θ by means of iteration are summarized as follows:

1. For the first run, the first set of mapping parameters is initialized, say $\hat{\theta}_0$.

2. The magnitude in which $\hat{\theta}_0$ can vary, say λ_0, is also initially set to some number. This magnitude of change or step size is initially large to allow for faster convergence rate.

3. The direction (increase/decrease), say Δ_0, is initialized to govern the adjustment direction of the parameters. The initialization of the direction doesn't matter since we are likely required to make an adjustment according to the likelihood obtained in the later step.

4. We then apply $\hat{\theta}_0$, Δ_0 and λ_0 to the optimization algorithm. In this way, we can obtain the likelihood given the current $\hat{\theta}_0$. According to the previous steps, new mapping parameters obtained, $\hat{\theta}_1$, can be generally expressed as

$$\hat{\theta}_1 = \hat{\theta}_0 + \Delta_0 \lambda_0. \quad (39)$$

We'll now treat the new mapping parameters as $\hat{\theta}_{t-1}$ should we continue onto the next iteration.

Newton's method and Line Search

- Here, adjustment to the mapping parameters is decided based on the expected value obtained, given the current estimated mapping parameters. Remember that we want a set of $\hat{\theta}$ that provides smallest $W\big(x(s), y_1(s, \theta_1), y_2(s, \theta_2)\big)$. We do this by adjusting the direction and magnitude of change (step side) to justify the most likely $\hat{\theta}$ in the next iteration.

- The direction adjustment is based on the observation of the slope of the likelihood function by means of calculating the gradient matrix. We can obtain the gradient matrix by taking the first derivative of the log-likelihood function respect to the mapping parameters:

$$\Delta_{t-1} = \frac{\partial \ln \mathcal{L}}{\partial \hat{\theta}_{t-1}} \quad (40)$$

We gradually reach the top of the likelihood function by adjusting $\hat{\theta}$ based on the gradient matrix Δ. For instance, when Δ is positive, this means the likelihood is increasing with the current $\hat{\theta}$. We then want to further increase $\hat{\theta}$ to ensure the highest possible likelihood. However, if the gradient vector Δ is negative, then this means the likelihood is decreasing with the current $\hat{\theta}$. So we need to decrease the values of $\hat{\theta}$. As we get closer to the highest likelihood, the gradient matrix becomes closer to zero. Note that the norms of the gradient matrix does not involve in the direction adjustment.

- Once we decide the suitable direction of the $\hat{\theta}$, we have to consider the amount by which the $\hat{\theta}$ should be changed. Apparently, we need a way to determine how fast the log-likelihood is changing. The second derivative of the log-likelihood function with respect to $\hat{\theta}$ can provide us such information. This simply means if the second derivative is high, then we will try to reduce the change to $\hat{\theta}$ so that the function is not changing too fast. Using Newton's method, the second derivative can be obtained by means of the Hessian matrix:

$$\mathrm{H} = \frac{\partial^2 \ln \mathcal{L}}{\partial \hat{\theta}^2} \tag{41}$$

The Hessian matrix, H, is a square matrix that consists of the second derivatives along its diagonal, whereas the off-diagonal elements consist of the partial derivative of $\ln \mathcal{L}$ in the cross $\hat{\theta}$ directions. Hence, with both direction and rate of change considered, the new parameter estimates can be obtained using the following equation.

$$\begin{aligned} \hat{\theta}_t &= \hat{\theta}_{t-1} - H_{t-1}^{-1} \cdot \Delta_{t-1} \\ &= \hat{\theta}_{t-1} - \left[\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \hat{\theta}_{t-1} \partial \hat{\theta}'_{t-1}} \right)^{-1} \cdot \frac{\partial \ln \mathcal{L}}{\partial \hat{\theta}_{t-1}} \right] \end{aligned} \tag{42}$$

From the equation, the new parameter estimates, $\hat{\theta}_t$, are adjusted in the direction by the first derivative, and in the change magnitude by the inverse of the second derivative.

- In our work, the choice of step side is decided by using a line search method (Nocedal and Wright). The Newton's method combined with the step side introduced by the line search method can be expressed as:

$$\hat{\theta}_t = \hat{\theta}_{t-1} - \alpha_{t-1} H_{t-1}^{-1} \cdot \Delta_{t-1} \tag{43}$$

Notice the step side, α_{t-1}, introduced to the parameter estimation equation.

- In the ideal line search procedure, finding the global minimum of the problem is the goal. However, finding the global minimum is difficult and the computational cost could be very expensive. In practice, obtaining the step side that provides less than local minimum, called *inexact line search*, is sufficient. This is done by imposing some conditions on the obtained α to ensure the length of step size is favorable and each iterative execution does not take too long. We chose to follow the Wolfe conditions (Surhone, 2010) on adjusting the length of step side. The Wolfe conditions consist of two conditions called *sufficient decrease condition* and *curvature condition*. The sufficient decrease condition is defined in an inequality as

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0) \text{ with } c_1 \in (0,1) \tag{44}$$

, where $\phi(\alpha) = f(\hat{\theta}_{t-1} + \Delta\alpha)$, $\phi(0) = f(\hat{\theta}_{t-1})$, and $\phi'(0) = \Delta_{t-1}$ is the directional derivative. Basically, this condition allows us to obtain the length of step that is not too small. In our implementation, α is initially set to 1 to explore the largest step side possible. Then, the condition is checked; if it does not hold, α is reduced by half. The second condition is also intended to eliminate the short steps by means of the slope, and it can be expressed as

$$\phi'(\alpha) \geq c_2 \phi'(0) \text{ with } c_2 \in (c_1, 1). \tag{45}$$

The second condition of inequality is quite self-explanatory. The obtained α satisfies the condition if the slope of ϕ is greater than the initial slope $\phi(0)$ times the constant c_2. Observe that, in our implementation, determination of the length of step side is more influenced by the curvature condition than the sufficient decrease condition (c_2 is significantly greater than c_1.

- Incorporating the current mapping parameters $\hat{\theta}_{t-1}$, the gradient matrix Δ_{t-1} , the inverse of Hessian matrix, H_{t-1}^{-1}, and the step side α_{t-1} into the Newton's method optimization algorithm, the new mapping parameters $\hat{\theta}_t$ can be obtained in series of iterative execution. The estimation simply stops when changes in $\hat{\theta}_t$ are less than 5%

in the consecutive iterations or the maximum iteration threshold is reached (30 iterations for our program).

In the previous explanation of the optimization algorithm we employed, the Newton's methods, we are required to compute the second derivative of the Hessian matrix with respect to the mapping parameters. The derivation of obtaining the estimates of the mapping parameters can be found in the Appendix section of this paper.

Experimental Design

We have proposed a joint registration and classification technique for satellite images based on MRF. In this section, using the proposed technique, choice of satellite images and experiment methods will be explained.

Satellite Data

To evaluate the generality of our proposed technique, input images have to represent a variety of sensor types and other tailored properties contributing to registration performance (whether in a constructive or destructive way), e.g. pixel size and acquisition date/time. Based on this idea, we experimented on different satellite images acquired from different sensor types and set up two sets of experiment using different combination of these satellite images. The acquired satellite images and their respective properties are summarized in Table 1.

SMMS and PALSAR images

A multispectral image acquired from SMMS (Small Multi-Mission Satellite) is used in conjunction with an L- band SAR image from PALSAR. Integration of multispectral and SAR images is a practical example of classification using multisensor images. The aim of this setup is to evaluate the performance of our technique on this practical use of multiple satellite images.

QuickBird images

Multispectral and panchromatic images acquired from QuickBird are used as our input images. While the two images were captured from the same satellite, they have a significant difference in pixel size and spectrum number. The thought behind this setup is to evaluate the capability of our technique on high-resolution satellite images.

Table 1 Acquired Satellite Images with Their Respective Properties

Acquisition Date (dd/mm/yy)	Satellite/Sensor	No. of Bands	Pixel Size (Meters)
27/11/10	SMMS/multispectral	4 (B, G, R, NIR)	30
28/08/10	ALOS PALSAR/SAR	1 (HH Polarization)	15
26/02/04	QuickBird/multispectral	4 (B, G, R, NIR)	2.4
26/02/04	QuickBird/panchromatic	1	0.6

Study Areas

Study area I – sugarcane fields

The first study area is located in Sa Kaeo province of Thailand where several types of agriculture are present (central coordinates: latitude: 13.4385; longitude: 102.16; Figure 2). Major land uses in the area include sugarcane, rice paddy, corn, and cassava. However, forest and mountainous areas can also be found in some parts of the site.

Study area II – Kasetsart University

The study area is located in Bangkok, Thailand, on the premises of Kasetsart University covering around 0.2592 in km^2 (central coordinates: latitude: 13.8475;

longitude: 100.5696; Figure 3). Land covers in the study area are mostly composed of trees and buildings causing visible shadow in several parts of the images.

Ground Data

Ground data is used as a reference to evaluate the classification performance. While we obtained these ground data from different organizations and sources, ground data of study area 2 were created by a human operator familiar with the study area.

Boundaries data

For study area I, boundaries data of sugarcane fields were obtained via BIOTEC (National Center for Genetic Engineering and Biotechnology) with validity of year 2010, while other land use boundaries are from Land Development Department of Thailand. The reference map is created by using these boundaries data and composed of two classes: sugarcane and other land covers (Figure 4).

Human-generated map

Without any formal ground data of study area II, the reference map is generated by a human operator familiar with the study area (Figure 5). Land covers in the area are generalized into four groups: plant, water body, impervious surface, and shadow. The impervious surface class includes buildings, rooftops, and roads.

Experiments

While our primary goal is to test the performance of our algorithm, we've conducted two experiments using different sets of satellite data to evaluate the generality and discover any limitations of our algorithm. Also, we can derive suitable parameters configuration from different experiment results.

Experiment I

SMMS and PALSAR are used as input images in this experiment. We test the performance of our algorithm based on the classification of sugarcane fields among other land covers on study area I. First, the PALSAR image is assumed to be perfectly registered with the LCM while the SMMS image is not. In other words, we need to register the SMMS image with the PALSAR image to acquire the final LCM. Hence, the correct mapping parameters expressed in eq. (21) for SMMS image relative to PALSAR image are

$$\theta_a = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \end{bmatrix}.$$

It's to be noted that the values 0.5 denote the relative scale of the SMMS image to the PALSAR image, i.e., the original SMMS image (before undergoing registration process) is two times smaller than the PALSAR image and LCM in both sample and line directions. We thus initially transform the SMMS image so that its size is 92% of the PALSAR image's with the incorrect mapping parameters

$$\theta_a = \begin{bmatrix} 0.52 & 0 & 0 \\ 0 & 0.52 & 0 \end{bmatrix}.$$

After preparing the input images and initializing the mapping parameters, the LCM is obtained by applying our algorithm to the input images. We also experiment on different values of β (see eq. (18)): 0, 0.25, 0.5, and 0.75 to observe any improvement in classification performance. Based on the idea of MRF, the value of β regulates the probability of neighboring pixels, i.e. as β increases, neighboring pixels are more likely to belong to the same class.

Experiment II

Two QuickBird images with different number of bands and pixel size are used in this experiment setup. The land covers in study area II are our test case. Despite

the differences in input images and study area, the experiment is carried out in a similar fashion of that in the experiment I with a few exceptions:

- Incorrect scale and displacement in mapping parameters are tested separately.
- Greater number of β, ranging from 0.0 to 1.0, are experimented.
- The process is allowed to finish early if the resulting LCMs in consecutive 50 iterations have changed less than 0.1%. Otherwise the process will terminate at the maximum allowed iteration of 1000.

Figure 2 A subset of SMMS image over the study area I.

Figure 3 A subset of QuickBird image over the study area II.

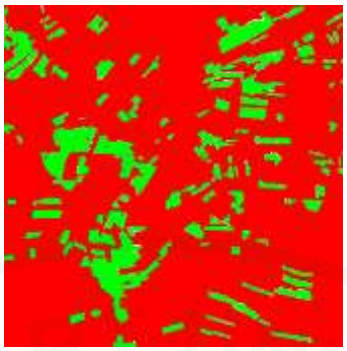


Figure 4 A ground data of study area I (green: sugarcane and red: others)

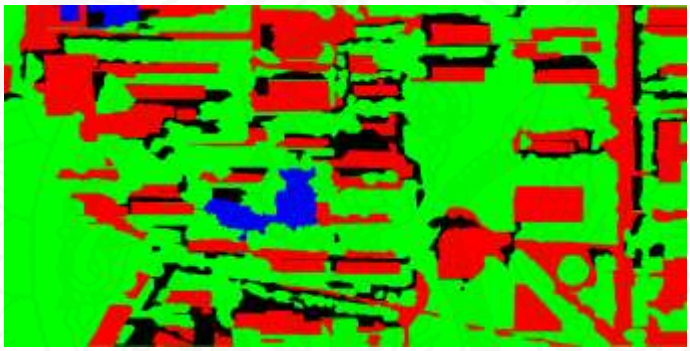


Figure 5 A ground data of study area II (green: plant, blue: water body, red: impervious surface, and black: shadow).

RESULTS AND DISCUSSION

Results

The resulting LCMs and classification performance from different experiments are summarized and discussed in this section. We address both the accuracy of classification and registration.

Result I

From the experiment I, with visual inspection, the pixels in resulting LCMs are more connected when the value of β is higher (

Figure **6**). Comparing the results of $\beta = 0$ and $\beta = 0.75$, the sugarcane-labeled pixels in LCM obtained with $\beta = 0.75$ are much more connected, whereas a large number of pixels are isolated with $\beta = 0$. The percentage of correct classification are equal to 71.91%, 78.53%, 79.09%, 79.23% for $\beta = 0, 0.25, 0.5, 0.75$, respectively. This further emphasizes that the MRF model increases the classification performance by exploiting the nature of neighboring pixels as suggested by both quantitative and qualitative analyses.

To evaluate the registration performance resulting from our algorithm, the estimation of the final mapping parameters are compared with the actual values. The results of mapping parameters with different values of β are summarized in Table 2 along with the corresponding errors. It can be observed that higher values of β tend to result in smaller errors in estimated mapping parameters with the exception of $\beta = 0.75$. This might suggest, for further implementation, we need to choose a suitable value of β to achieve the optimum registration performance. We can observe from Figure **7** that registration errors gradually decrease for every progressing iteration, and lower β values result in faster diminishing error rate. At this point, by inferring from the results, we can conclude that higher β values allow for better land cover classification while hindering the convergence rate of the mapping parameters estimated

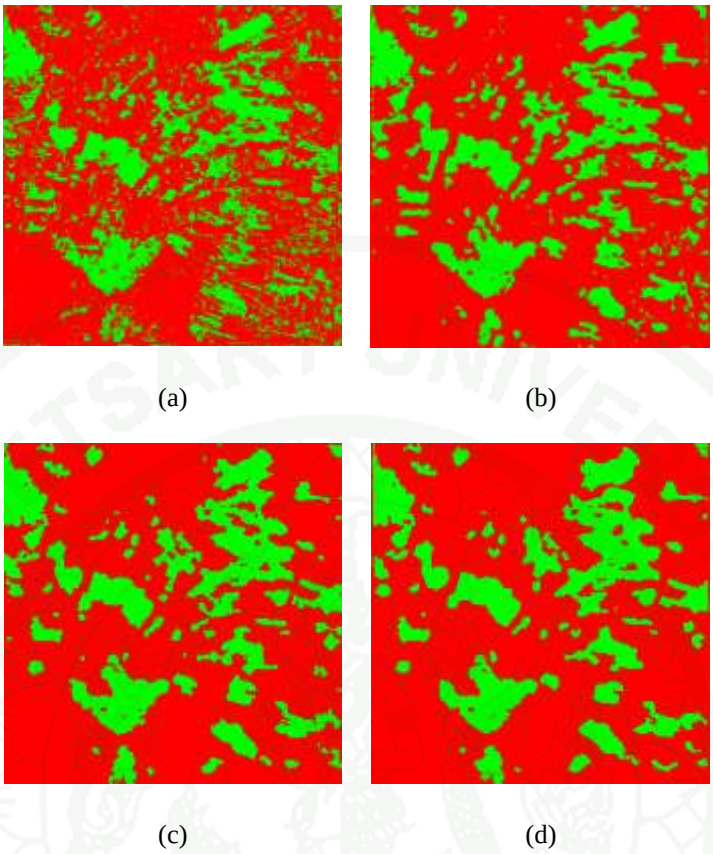


Figure 6 Resulting LCMs: $\beta = 0,\ 0.25,\ 0.5, 0.7$ for a, b, c, d, respectively.

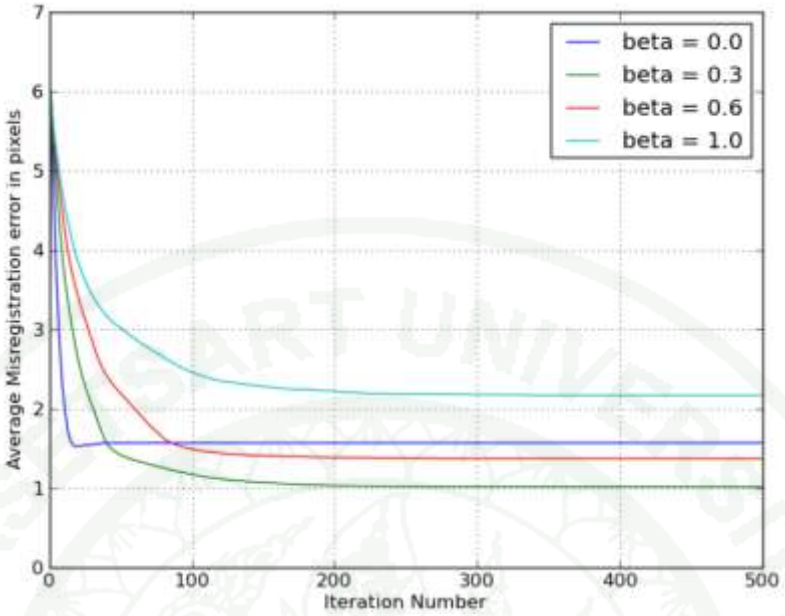


Figure 7 The mis-registration errors as a function of the number of iterations for $\beta = 0.0, 0.3, 0.6, 1.0$.

Table 2 The Resulting Estimated Mapping Parameters and Errors.

Estimated mapping parameters	***β***				**Actual value of mapping parameters**
	0	**0.25**	**0.5**	**0.75**	
γ_1	0.496	0.503	0.499	0.501	0.5
γ_5	0.504	0.499	0.499	0.500	0.5
γ_4	-0.001	-0.004	0.0	0.0	0.0
γ_2	0.018	-0.001	-0.010	0.0	0.0
γ_3	-0.174	0.220	0.023	0.197	0.0
γ_6	0.631	0.475	0.497	0.151	0.0
RMSE	1.566	1.013	1.370	2.166	-

Result II

In the experiment II, we have observed more displacement and scale in mapping parameters with the values of β ranging from 0.0 to 1.0. However, we also experimented on MRF model to further investigate the robustness of it on perfectly registered images given there are more number of classes than that in the previous experiment. The results showed, as we progress to greater values of β, more consolidated pixels of the same classes as seen in Figure 8. The quality assessment does agree with the visual inspection where percentage of correct classification is increased along with the increment of β.

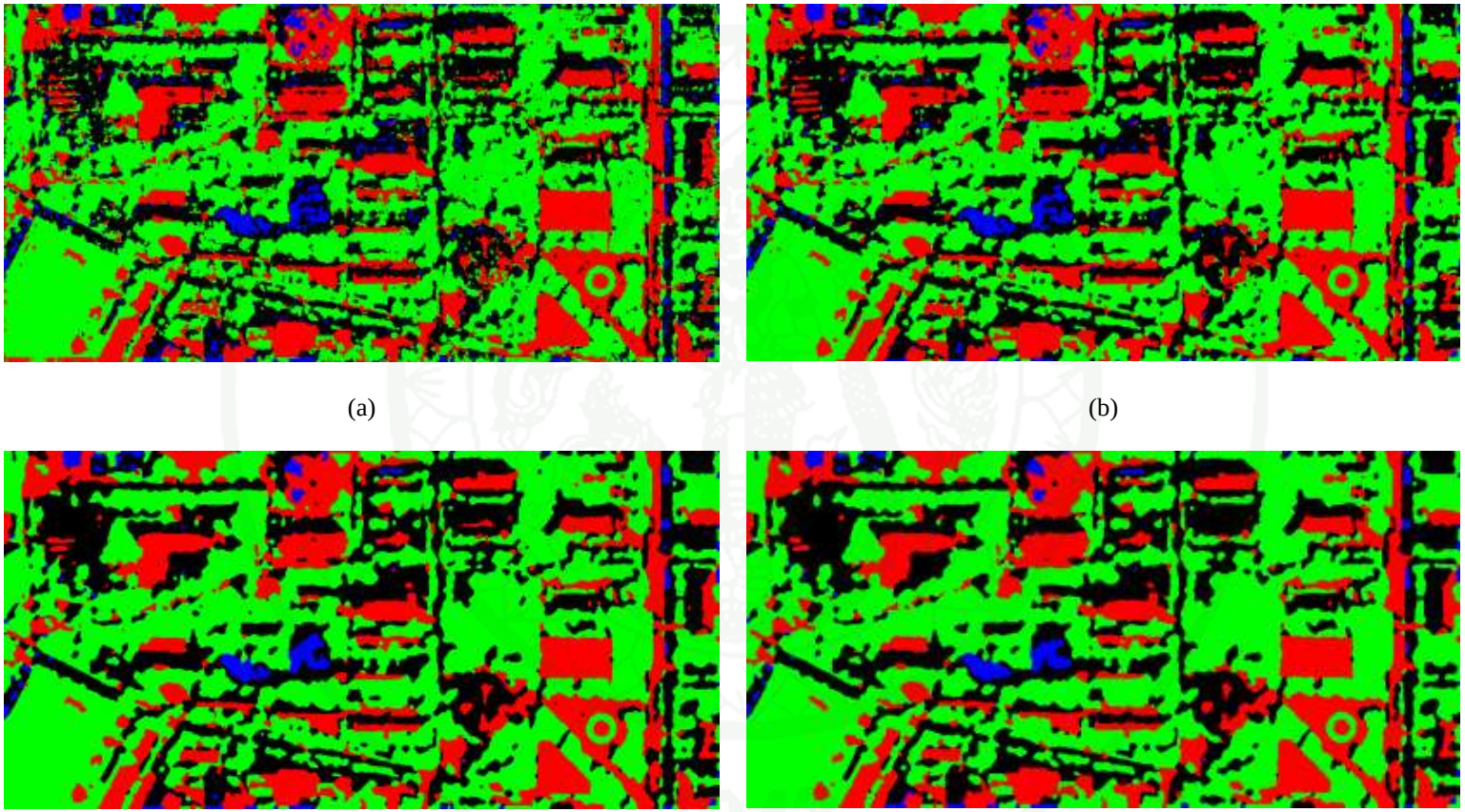
The results of classification performance when mapping parameters are incorrect are shown in Figure 11. It's worth noting that the incorrect mapping parameters imposed when implementing our proposed algorithm were applied to both sample and line directions of input images. That means, for an instance, 60 pixels in mis-displacement is resulted from 30 pixels mis-displacement each in sample and line directions. Therefore, the corresponding mapping parameters for the given example are

$$\theta_a = \begin{bmatrix} 0.25 & 0 & 30 \\ 0 & 0.25 & 30 \end{bmatrix}$$

when two input images are four times different in pixel size depicted by 0.25. This implementation is also applied to the case of mis-registration in scale. It's also worth mentioning that the abrupt drop in classification performance is caused by the sudden increase in degree of mis-registration, for example, the increase from 8 pixels of displacement to 10 that suddenly becomes 20. This is, however, a by-design experiment since we wanted to evaluate the limitation of our algorithm.

Discussion

As for the results, it suffices to say that classification performance continued to drop as the degree of mis-registration in the mapping parameters increased. This shows the importance of registration that contributes to the classification performance. Apart from the trend of classification performance in general, the results suggest that smaller β values are more tolerant to degree of mis-registration. As can be observed from Figure 11, classification accuracies whose β values are below 0.4 remain almost constant for some mis-registration degrees, whereas accuracies drop at noticeably magnitude when β values are higher. The explanation behind this is best discussed in convergence-speed context. Figure 10 shows total number of iteration the algorithm requires to obtain the final LCM. It's to be noted that the maximum iteration number for our experiment setup is one thousand so the algorithm is forced to conclude however the current LCM is. Since the LCM obtained in this way is not likely the most desired result, we can expect low performance in both registration and classification performance. The 8.16%-mis-registration in Figure 10 illustrates this kind of scenario wherein the algorithm terminated at 1000^{th} iteration for higher β values. This, however, does not apply to lesser mis-registration of 4.04% where the algorithm converged before the maximum-allowed iterations. The increase in β value still dictates the increasing number of iterations in both cases though.



(a) (b)

(c) (d)

Figure 8 Resulting LCMs: $\beta = 0,\ 0.2,\ 0.6,\ 1.0$ for a, b, c, d, respectively

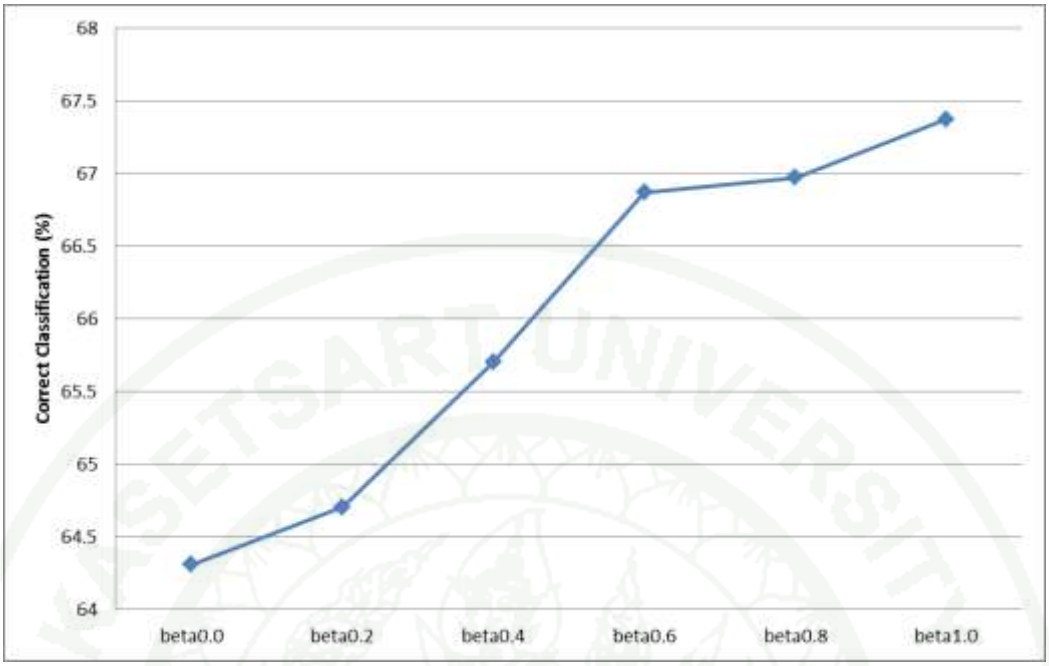


Figure 9 The trend of classification performance as β is increased.

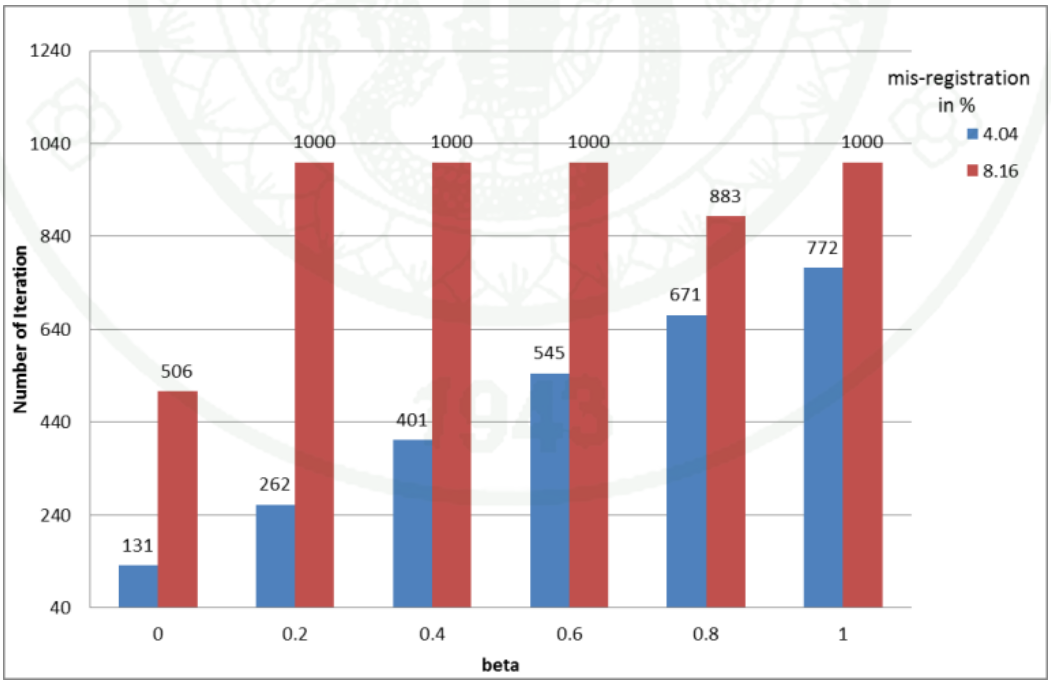
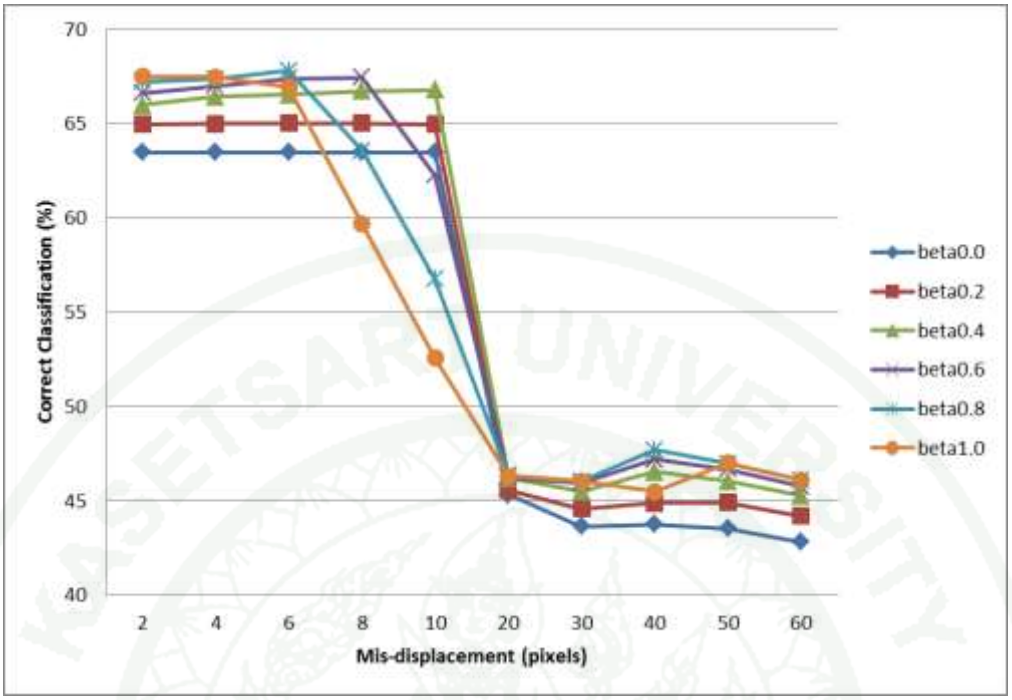
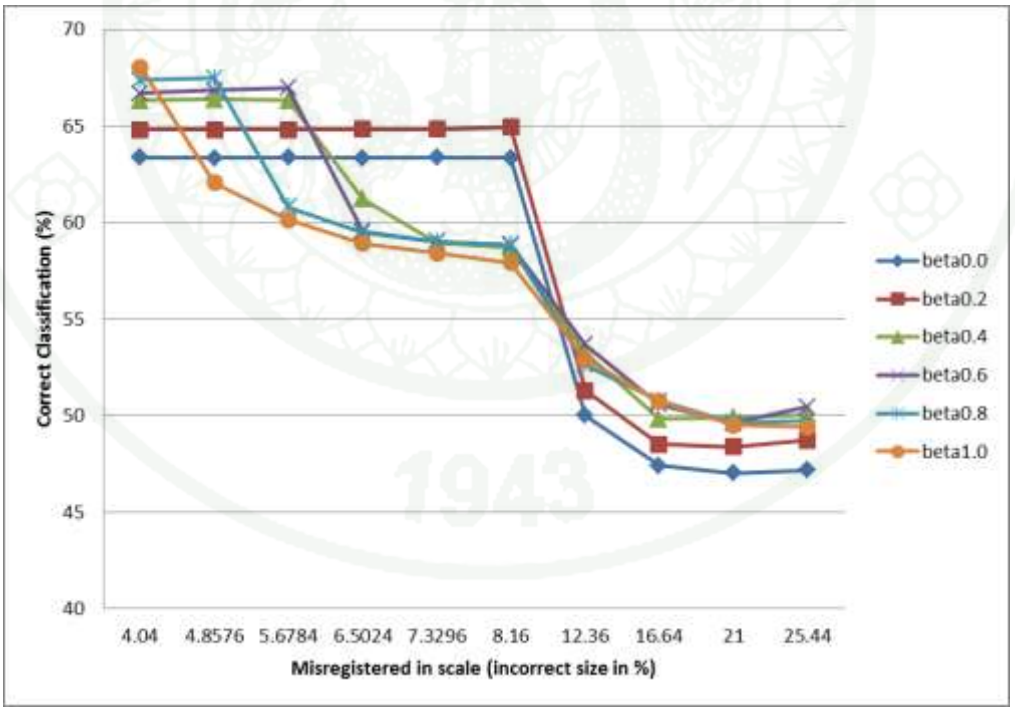


Figure 10 Comparison of iteration number as β is increased.

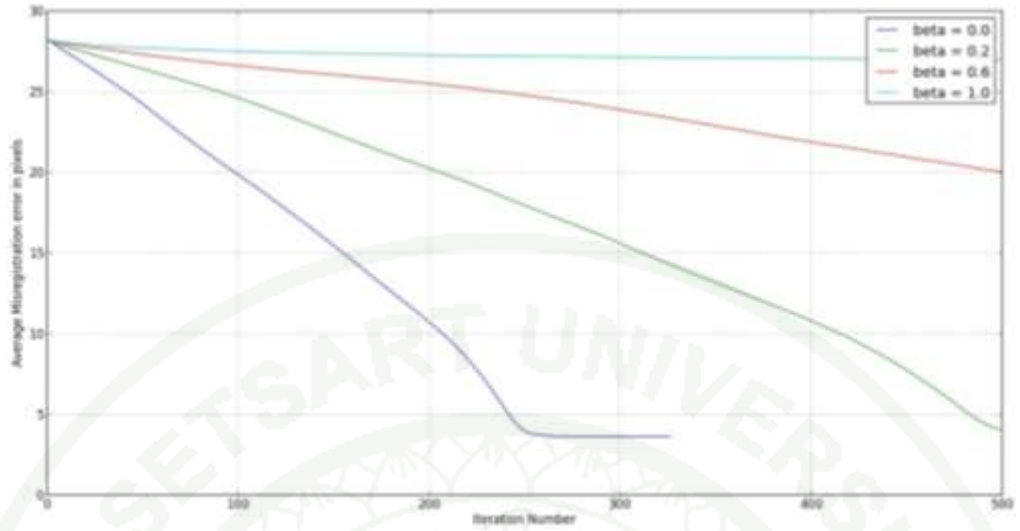


(a)

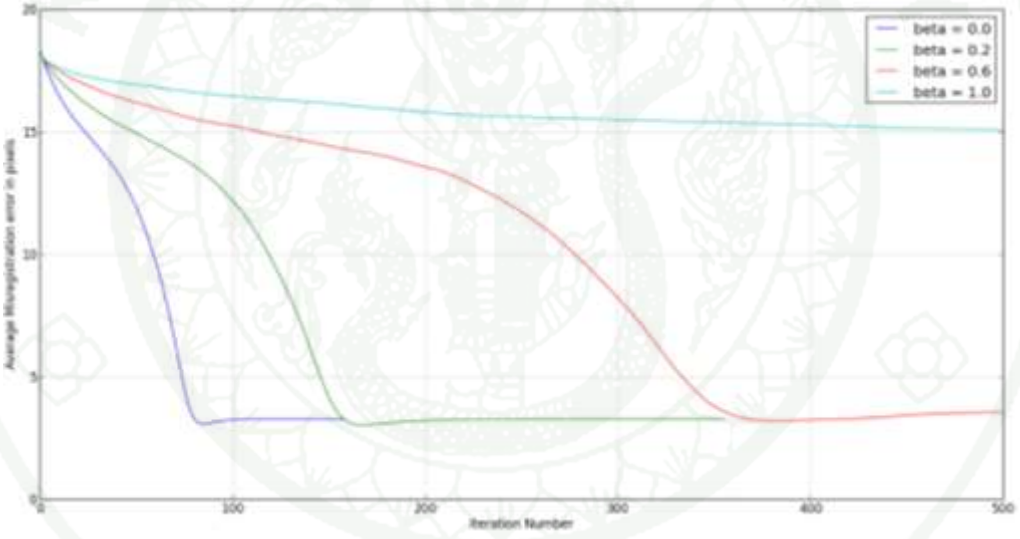


(b)

Figure 11 The trends of classification performance when mis-registration in: a) displacement increased; b) scale increased.



(a)



(b)

Figure 12 The mis-registration errors as a function of the number of iterations in: a) 10 pixels displacement; b) 4.86% scale error.

CONCLUSION AND RECOMMENDATIONS

Conclusion

In this paper, we have proposed the joint registration and classification technique based on the Markov random field model to obtain the land cover map of the scene.

Instead of following the traditional approach to obtain the land cover map, where the registration and classification processes are performed separately, the proposed joint approach allows us to automatically evaluate the performance of both processes simultaneously. The EM algorithm is employed to solve our joint process problem by iteratively estimating the mapping parameters to obtain an optimal land cover map. In addition, spatial-context information of image pixels is exploited by using the Markov random field model to achieve better classification performance.

We have applied our proposed joint method to a practical application of satellite imagery where multisensor pairs of satellite images are used. The results show that contextual information, derived from the Markov random field model, plays an important role in improving the classification performance. Both quantitative and qualitative assessments suggest that amplifying the contextual-interaction parameter results in the better land cover maps obtained. Furthermore, we have evaluated the final estimated of the mapping parameters after using our joint method. The results suggest that the registration performance also significantly contributes to the quality of the land cover map. With our proposed algorithm, we have managed to achieve the correct mapping parameters and obtain the near perfect registration of multisensor images. However, as we artificially continue to increase the volume of error to the initial mapping parameters, which is equivalent to amplifying the geometric difference of input images, we receive the poorer registration performance in the process. Our experiment on adjusting the initial mapping parameters suggests that the registration performance is most sensitive to errors in the scaling factors.

Recommendations

The proposed joint registration and classification technique shows the potential solution to the satellite image classification based on multiple sources of data, which is a prominent practice in the applications of satellite imagery in the present. The ability to automatically produce a classified map based on multisensor images is a major advantage of this proposed algorithm. The increasing availability of satellite images inevitably introduces greater diversity and complexity to satellite image applications that make use of multiple sources of data; hence the increased demand for an automated practice to extract valuable information from those diverse data. Our proposed joint method directly responds to the call. Still, the performance of the algorithm in terms of execution speed is still left to be explored and improved as well as the capability to handle multiple images with various differences in geometry and noise.

LITERATURE CITED

Aldrich, J. 1997. R. A. Fisher and the Making of Maximum Likelihood 1912 - 1922. **Statistical Science**, issue 12, 162-176.

Baghdadi, N., P. Todoroff, M. E. Hjj and A. Begue. 2009. Potential of SAR Sensors TerraSAR-X, ASA/ENVISAT and PALSAR/ALOS for Monitoring Sugarcane Crops on Reunion Island. **Remote Sensing of Environment**, issue 113, 1724-1738.

Borman, S. 2009. The Expectation Maximization Algorithm.

Chen, S., Q. Guo, H. Leung and E. Bosse. May 2011. A Maximum Likelihood Approach to Joint Image Registration and Fusion. **IEEE Transactions on Image Processing**, 1363-1372.

Crawford, N. J. 2007. Mean Field Theories and Models of Statistical Physics. University of California.

Dellaert, F. 2002. The Expectation Maximization Algorithm. College of Computing, Georgia Institute of Technology.

Eliason, S. R. 2008. Maximum Likelihood Estimation: Logic and Practice (2 ed.). (S. R. Eliason, Ed.) Sage.

Fonseca, L. M. and M. H. Costa. 1997. Automatic Registration of Satellite Images. **Brazillian Symposium on Computer Graphics and Image Processing**, issue 10, 219-226.

Geman, S., and D. Geman. 1984. Stochastic Relaxation, Gibbs Distribution, and the Bayesian Retoration of Images. **IEEE Trans. Pattern Anal. Machine Intell**, 721-741.

Hauser, K. 2012. Multivariate Newton's Method and Quasi-Newton Methods.

Kindermann, R. and J. L. Snell. 1980. Markov Random Fields and Their Applications. American Mathematical Society.

Korting, T. S., L. Maria and G. Fonseca. 2008. Expectation-Maximization x Self-Organizing Map for image classification. **IEEE International Conference on Signal Image Technology and Internet Based Systems**.

Lombardo, P., C. J. Oliver, T. M. Pellizzeri and M. Meloni. 2003. A New Maximum-Likelihood Joint Segmentation Technique For Multitemporal SAR and Multiband Optical Images. **IEEE Transactions on Geoscience and Remote Sensing**, 2500-2518.

Neal, R. and G. Hinton. 1998. A View of the EM Algorithm That Justifies Incremental, Sparse, and Other Variants. In M. Jordan (Ed.), **Learning in Graphical Models**. Kluwer Academic Press.

Nocedal, J. and S. J. Wright. n.d. **Numerical Optimization, 2nd edition** (2 ed.).

Skriver, H., F. Mattia, G. Satalino, A. Balenzano, V. R. Pauwels, N. E. Verhoest and M. Davidson. 2011. Crop Classification Using Short-Revisit Multitemporal SAR Data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, vol. 4, 423-431.

Surhone, L. M. 2010. Wolfe Conditions. (L. M. Surhone, M. T. Timpledon, & S. F. Marseken, Eds.) VDM Verlag Dr. Mueller e.K.

Tso, B. C., and P. M. Mather. May 1999. Classification of Multisource Remote Sensing Imagery Using a Genetic Algorithm and Markov Random Field. **IEEE Transactions on Geoscience and Remote Sensing**, issue 37, 1255-1260.

Zitova, B. and J. Flusser. 2003. Image Registration Methods: A Survey. **Image and Vision Computing *21***, 977-1000.

APPENDIX

Derivation to Obtain the Estimates of the Mapping Parameters

In the Problem Statement section, we have assumed a scenario where two observed images are needed to be co-registered. Given one of the two is assumed to be a perfect reference to the true scene X, we can register the other observed image, Y, to the true scene by means of affine transformation. The mapping parameters, θ, are required to perform the transformation and they can be expressed in a matrix form as in eq. (21), rewritten here for convenience.

$$\theta = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ \gamma_4 & \gamma_5 & \gamma_6 \end{bmatrix}$$

where γ_1 and γ_5 represent scaling, γ_2 and γ_4 represent skewing, γ_3 and γ_6 represent displacement in x and y axis repectively. Then, consider letting (U, V) as a reprentative coordinate of the true scen. We can express the transformation of a coordinate (u, v) in Y to the true scene's coordinate as follows.

$$\begin{aligned} X &= \theta \cdot Y \\ \begin{bmatrix} U \\ V \end{bmatrix} &= \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_4 & \gamma_5 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} \gamma_3 \\ \gamma_6 \end{bmatrix} \\ U &= \gamma_1 u + \gamma_2 v + \gamma_3 \\ V &= \gamma_4 u + \gamma_5 v + \gamma_6 \end{aligned} \tag{46}$$

We obtain the estimates of the mapping parameters by solving the following equation.

$$E\big(x, y(\theta)\big) = \sum_{s\in S}\sum_{l=0}^{L-1}[(y(\theta)-\mu_x)^T\Sigma_x^{-1}(y(\theta)-\mu_x) + \log|\Sigma_x|]\,\hat{P}_x$$

$$\frac{\partial E}{\partial \gamma_k} = \sum_{s\in S}\sum_{l=0}^{L-1}\left[2(y(\theta)-\mu_x)^T\Sigma_x^{-1}\frac{\partial}{\partial \gamma_k}y(\theta)\right]\hat{P}_x$$

$$\frac{\partial^2 E}{\partial \gamma_k^2} = \sum_{s\in S}\sum_{l=0}^{L-1}\left[2(y(\theta)-\mu_x)^T\Sigma_x^{-1}\frac{\partial^2}{\partial \gamma_k^2}y(\theta) + \frac{\partial}{\partial \gamma_k}y(\theta)^T\Sigma_x^{-1}\frac{\partial}{\partial \gamma_k}2(y(\theta)-\mu_x)\right]\hat{P}_x$$

$$= \sum_{s\in S}\sum_{l=0}^{L-1}\left[2(y(\theta)-\mu_x)^T\Sigma_x^{-1}\frac{\partial^2}{\partial \gamma_k^2}y(\theta) + 2\left(\frac{\partial}{\partial \gamma_k}y(\theta)^T\Sigma_x^{-1}\frac{\partial}{\partial \gamma_k}y(\theta)\right)\right]\hat{P}_x \quad (47)$$

The first derivatives of $y(\theta)$ with repect to γ_k, which can be obtained via eq. (46), are summarized as follows:

$$\begin{aligned}
\frac{\partial}{\partial \gamma_1}y(\theta) &= \frac{\partial}{\partial U}y(\theta)\cdot\frac{\partial U}{\partial \gamma_1}\\
\frac{\partial}{\partial \gamma_2}y(\theta) &= \frac{\partial}{\partial U}y(\theta)\cdot\frac{\partial U}{\partial \gamma_2}\\
\frac{\partial}{\partial \gamma_3}y(\theta) &= \frac{\partial}{\partial U}y(\theta)\cdot\frac{\partial U}{\partial \gamma_3}\\
\frac{\partial}{\partial \gamma_4}y(\theta) &= \frac{\partial}{\partial V}y(\theta)\cdot\frac{\partial V}{\partial \gamma_4}\\
\frac{\partial}{\partial \gamma_5}y(\theta) &= \frac{\partial}{\partial V}y(\theta)\cdot\frac{\partial V}{\partial \gamma_5}\\
\frac{\partial}{\partial \gamma_6}y(\theta) &= \frac{\partial}{\partial V}y(\theta)\cdot\frac{\partial V}{\partial \gamma_6}
\end{aligned} \quad (48)$$

The second derivatives of $y(\theta)$ with repect to $\gamma_k$can then be obtained from eq. (48), and are summarized as follows.

$$\frac{\partial^2}{\partial\gamma_1^2}y(\theta) = \frac{\partial^2}{\partial U^2}y(\theta)\cdot u^2$$
$$\frac{\partial^2}{\partial\gamma_2^2}y(\theta) = \frac{\partial^2}{\partial U^2}y(\theta)\cdot v^2$$
$$\frac{\partial^2}{\partial\gamma_3^2}y(\theta) = \frac{\partial^2}{\partial U^2}y(\theta)$$
$$\frac{\partial^2}{\partial\gamma_4^2}y(\theta) = \frac{\partial^2}{\partial V^2}y(\theta)\cdot u^2$$
$$\frac{\partial^2}{\partial\gamma_5^2}y(\theta) = \frac{\partial^2}{\partial V^2}y(\theta)\cdot v^2$$
$$\frac{\partial^2}{\partial\gamma_6^2}y(\theta) = \frac{\partial^2}{\partial V^2}y(\theta) \quad (49)$$

For demonstration purposes, we have included the derivation in first and second derivatives with respect to γ_1 as follows:

$$\begin{aligned}\frac{\partial}{\partial\gamma_1}y(\theta) &= \frac{\partial}{\partial U}y(\theta)\cdot\frac{\partial U}{\partial\gamma_1} + \frac{\partial}{\partial V}y(\theta)\cdot\frac{\partial V}{\partial\gamma_1}\\ &= \frac{\partial}{\partial U}y(\theta)\cdot\frac{\partial U}{\partial\gamma_1}\\ \frac{\partial^2}{\partial\gamma_1^2}y(\theta) &= \frac{\partial}{\partial\gamma_1}\left(\frac{\partial}{\partial U}y(\theta)\cdot\frac{\partial U}{\partial\gamma_1}\right)\\ &= \frac{\partial}{\partial U}y(\theta)\frac{\partial^2}{\partial\gamma_1^2}U + \frac{\partial U}{\partial\gamma_1}\left(\frac{\partial}{\partial\gamma_1}\frac{\partial}{\partial U}y(\theta)\right)\\ &= \left[\frac{\partial^2}{\partial U^2}y(\theta)\frac{\partial}{\partial\gamma_1}U + \frac{\partial}{\partial V}\frac{\partial}{\partial U}y(\theta)\frac{\partial V}{\partial\gamma_1}\right]\frac{\partial}{\partial\gamma_1}U\\ &= \frac{\partial^2}{\partial U^2}y(\theta)\cdot u\end{aligned} \quad (50)$$

Other derivatives with respect to the rest of γ_k can be computed in a similar fashion. According to eq. (48) and (49), we still need to compute $\frac{\partial}{\partial U}y(\theta)$, $\frac{\partial}{\partial V}y(\theta)$, $\frac{\partial^2}{\partial U^2}y(\theta)$, and $\frac{\partial^2}{\partial V^2}y(\theta)$. This can be achieved by using a cubic convolution as an interpolation

technique [20]. The following kernel of cubic convolution corresponding to x-axis is used.

$$k(x) = \begin{cases} \frac{3}{2}x^3 - \frac{5}{2}x^2 + 1, & 0 \le x \le 1 \\ -\frac{3}{2}x^3 - \frac{5}{2}x^2 + 1, & -1 \le x \le 0 \\ -\frac{1}{2}x^3 + \frac{5}{2}x^2 - 4x + 2, & 1 < x \le 2 \\ \frac{1}{2}x^3 + \frac{5}{2}x^2 + 4x + 2, & -2 < x \le -1 \\ 0, & otherwise \end{cases} \quad (51)$$

Given the mapping parameters θ, let us assume that (u', v') is the interpolated pixel coordinate corresponding to the coordinate (u, v) in the observed image Y. If the coordinate (u', v') falls within the rectangular window $[\lfloor u' \rfloor, \lfloor u' \rfloor + 1] \times [\lfloor v' \rfloor, \lfloor v' \rfloor + 1]$, where $\lfloor u' \rfloor$ denotes the largest integer that is less than or equal to u', we can write the cubic interpolation function as

$$y(u, v) = \sum_{s_u=-1}^{2} \sum_{s_v=-1}^{2} y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v) \cdot k\big(u' - (\lfloor u' \rfloor + s_u)\big) \cdot k\big(v' - (\lfloor v' \rfloor + s_v)\big) \quad (52)$$

where s_u and s_v are integer distance along x and y axis respectively. We can finally compute $\frac{\partial}{\partial U} y(\theta)$ and $\frac{\partial}{\partial V} y(\theta)$ as follows.

$$\frac{\partial}{\partial U}y(\theta) = \frac{\partial}{\partial U}y(u,v) = \frac{\partial}{\partial k(x)}y(u',v')\frac{\partial k(x)}{\partial x}\frac{\partial x}{\partial U}$$

$$= \begin{cases} y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v)\left(\frac{9}{2}x^2 - 5x\right)k\big(v' - (\lfloor v' \rfloor + s_v)\big), & 0 \le x \le 1 \\ y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v)\left(-\frac{9}{2}x^2 - 5x\right)k\big(v' - (\lfloor v' \rfloor + s_v)\big), & -1 \le x \le 0 \\ y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v)\left(-\frac{3}{2}x + 5x - 4\right)k\big(v' - (\lfloor v' \rfloor + s_v)\big), & 1 < x \le 2 \\ y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v)\left(\frac{3}{2}x + 5x + 4\right)k\big(v' - (\lfloor v' \rfloor + s_v)\big), & -2 < x \le -1 \\ 0, & otherwise \end{cases} \quad (53)$$

The $\frac{\partial}{\partial V}y(\theta)$ can be computed in a similar way using the kenerl corresponding to y axis. We then can derive $\frac{\partial^2}{\partial U^2}y(\theta)$ based on eq. (53) as

$$\frac{\partial^2}{\partial U^2}y(\theta) = \frac{\partial^2}{\partial U^2}y(u,v) =$$

$$= \begin{cases} y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v)(9x - 5)k\big(v' - (\lfloor v' \rfloor + s_v)\big), & 0 \le x \le 1 \\ y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v)(-9x - 5)k\big(v' - (\lfloor v' \rfloor + s_v)\big), & -1 \le x \le 0 \\ y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v)(-3x + 5)k\big(v' - (\lfloor v' \rfloor + s_v)\big), & 1 < x \le 2 \\ y(\lfloor u' \rfloor + s_u, \lfloor v' \rfloor + s_v)(3x + 5)k\big(v' - (\lfloor v' \rfloor + s_v)\big), & -2 < x \le -1 \\ 0, & otherwise \end{cases} \quad (54)$$

We can now continue on what we left off in eq.(47). Combining the results from eq. (48), (49), (54), and (53), we can write eq.(47) with respect to γ_{k1}^2 as

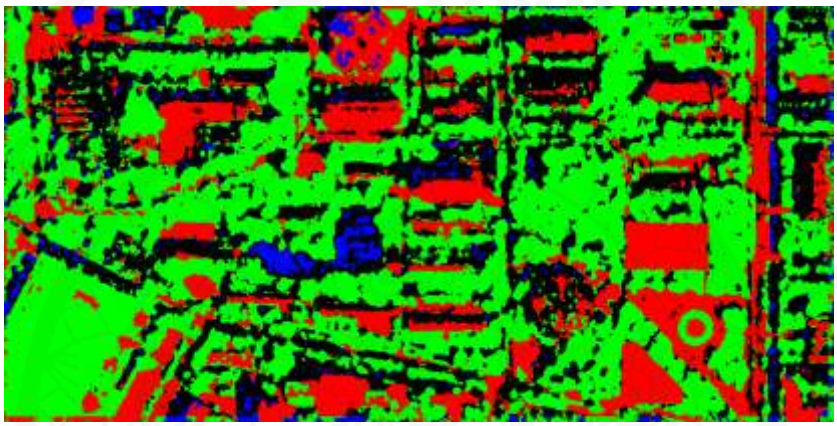
$$\begin{aligned} \frac{\partial^2 E}{\partial \gamma_1^2} &= \sum_{s \in \mathcal{S}} \sum_{l=0}^{L-1} \Bigg[2(y(\theta) - \mu_x)^T \Sigma_x^{-1} \frac{\partial^2}{\partial \gamma_1^2} y(\theta) \\ &\qquad + 2\left(\frac{\partial}{\partial \gamma_1} y(\theta)^T \Sigma_x^{-1} \frac{\partial}{\partial \gamma_1} y(\theta)\right)\Bigg] \hat{P}_x \\ &= \sum_{s \in \mathcal{S}} \sum_{l=0}^{L-1} \Bigg[2(y(\theta) - \mu_x)^T \Sigma_x^{-1} \left(\frac{\partial^2}{\partial U^2} y(\theta) \cdot u^2\right) \\ &\qquad + 2\left(\frac{\partial}{\partial U} y(\theta)^T \Sigma_x^{-1} \frac{\partial}{\partial U} y(\theta) \cdot u\right)\Bigg] \hat{P}_x \end{aligned} \quad (55)$$

The Resulting LCMs

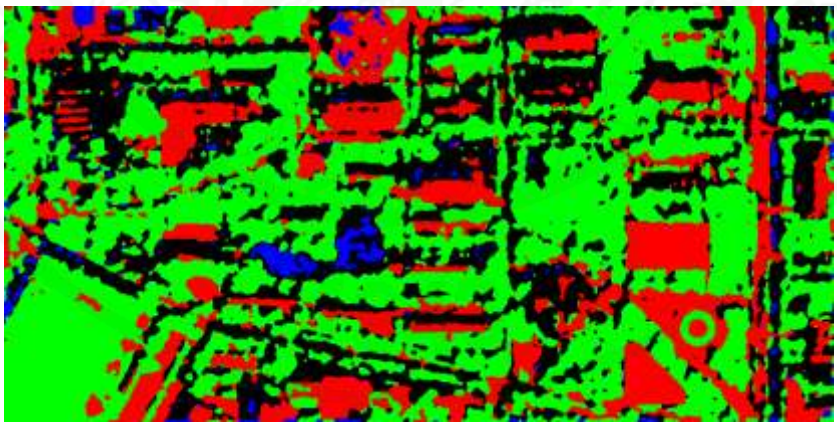
Perfect Registered LCMs

Initial mapping parameters:

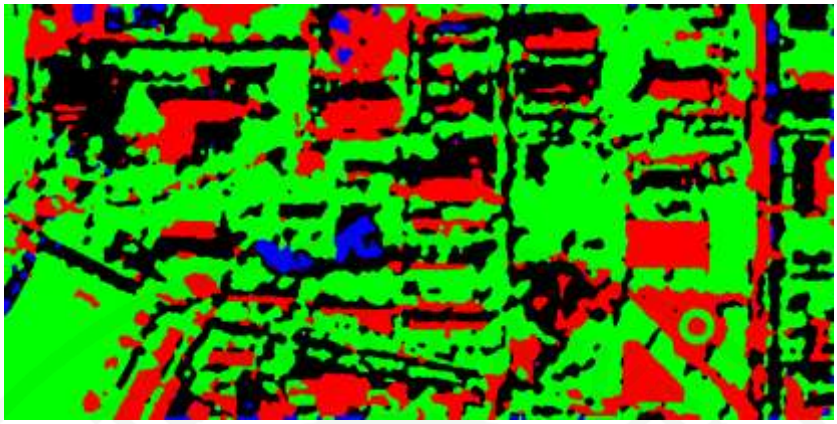
$$\theta_a = \begin{bmatrix} 0.25 & 0.0 & 0.0 \\ 0.0 & 0.25 & 0.0 \end{bmatrix}$$



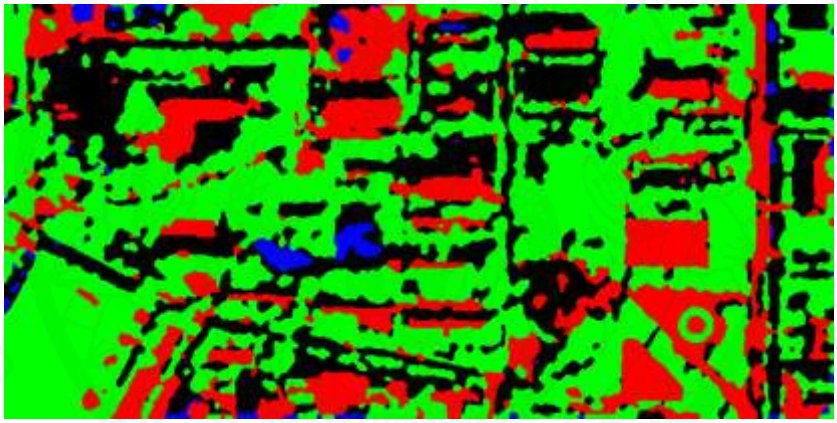
Appendix Figure 1 Resulting LCM with perfect registration, $\beta = 0.0$.



Appendix Figure 2 Resulting LCM with perfect registration, $\beta = 0.2$.



Appendix Figure 3 Resulting LCM with perfect registration, $\beta = 0.6$.

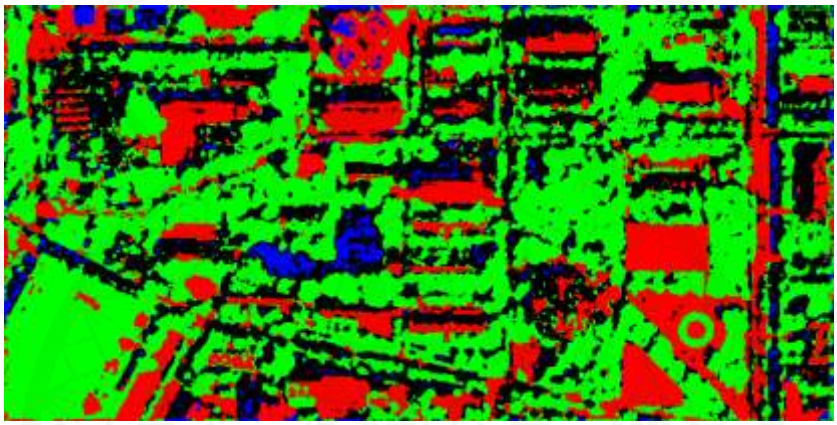


Appendix Figure 4 Resulting LCM with perfect registration, $\beta = 1.0$.

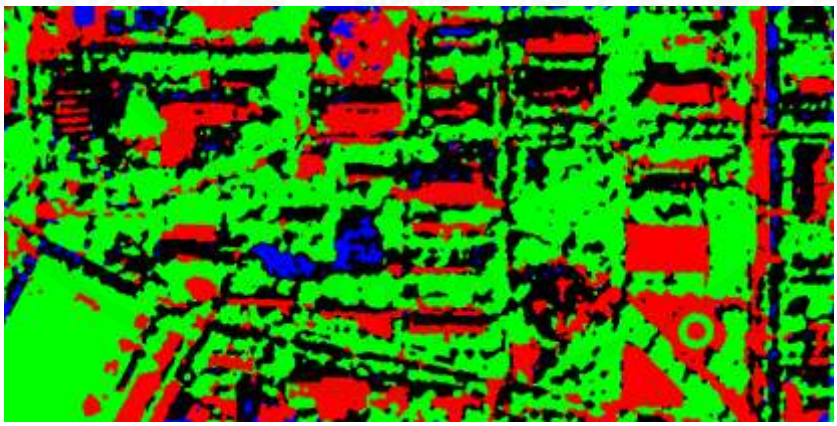
Imperfect Registered LCMs with Initial Errors in Displacement

Initial mapping parameters:

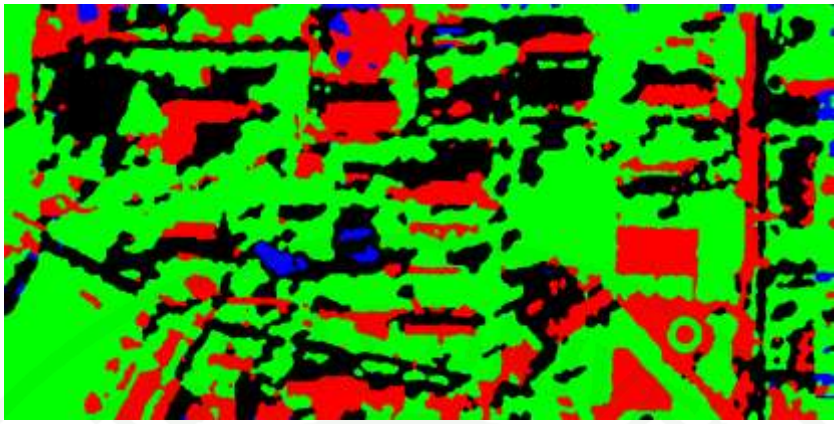
$$\theta_a = \begin{bmatrix} 0.25 & 0.0 & 5.0 \\ 0.0 & 0.25 & -5.0 \end{bmatrix}$$



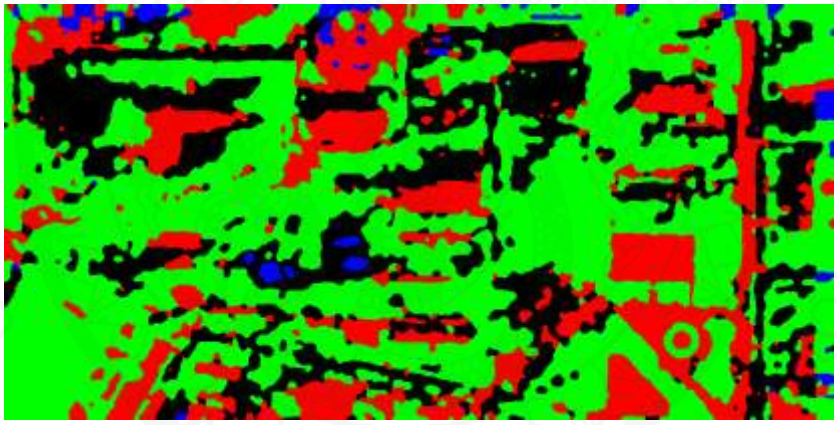
Appendix Figure 5 Resulting LCM with imperfect registration in displacement, $\beta = 0.0$.



Appendix Figure 6 Resulting LCM with imperfect registration in displacement, $\beta = 0.2$.



Appendix Figure 7 Resulting LCM with imperfect registration in displacement, $\beta = 0.6$

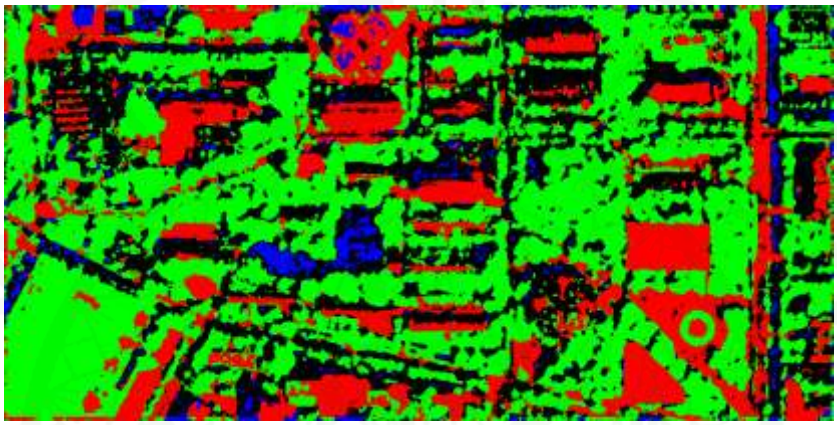


Appendix Figure 8 Resulting LCM with imperfect registration in displacement, $\beta = 1.0$.

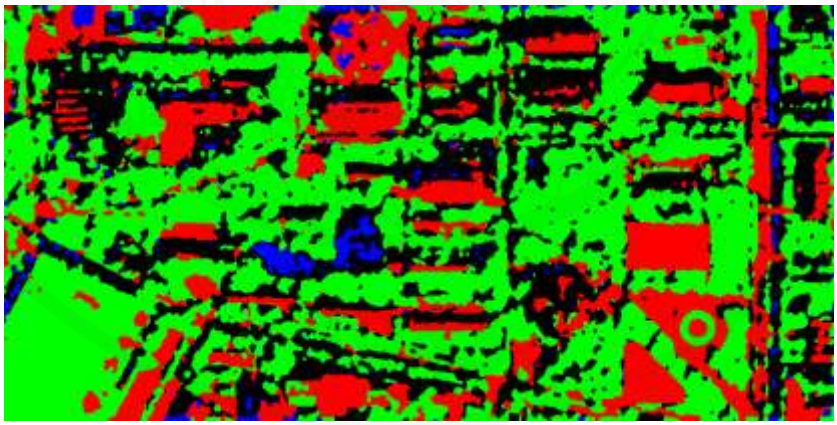
Imperfect Registered LCMs with Initial Errors in Scaling

Initial mapping parameters:

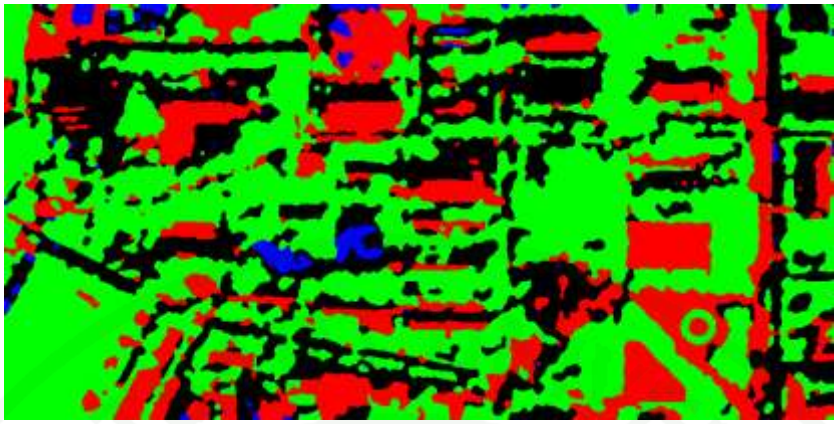
$$\theta_a = \begin{bmatrix} 0.256 & 0.0 & 0.0 \\ 0.0 & 0.244 & 0.0 \end{bmatrix}$$



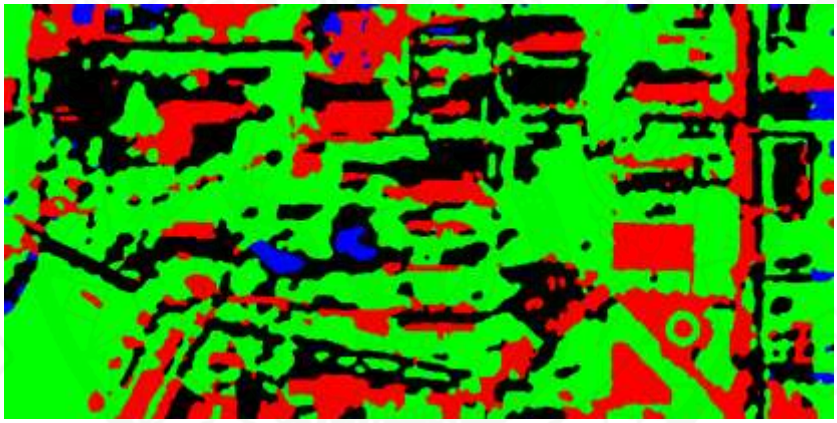
Appendix Figure 9 Resulting LCM with imperfect registration in scaling, $\beta = 0.0$



Appendix Figure 10 Resulting LCM with imperfect registration in scaling, $\beta = 0.2$.



Appendix Figure 11 Resulting LCM with imperfect registration in scaling, $\beta = 0.6$



Appendix Figure 12 Resulting LCM with imperfect registration in scaling, $\beta = 1.0$

Accuracy Assessment of Resulting LCMs

Perfect Registered LCMs

Initial mapping parameters:

$$\theta_a = \begin{bmatrix} 0.25 & 0.0 & 0.0 \\ 0.0 & 0.25 & 0.0 \end{bmatrix}$$

Appendix Table 1 Confusion Matrix of LCM with perfect registration, $\beta = 0.0$.

Classified Map	Reference Map				
	Plant	**Water**	**Impervious Surface**	**Shadow**	**User's Accuracy (%)**
Plant	296,470	1,140	33,905	8,956	87.07
Water	7,258	6,190	2,945	6,843	26.63
Impervious Surface	35,604	334	107,778	6,355	71.81
Shadow	87,604	6,594	59,472	52,552	25.48
Producer's Accuracy	69.44	43.41	52.80	70.34	
Overall Accuracy (%): 64.30			Kappa Coefficient: 0.4335		

Appendix Table 2 Confusion Matrix of LCM with perfect registration, $\beta = 1.0$.

Classified Map	Reference Map				
	Plant	**Water**	**Impervious Surface**	**Shadow**	**User's Accuracy (%)**
Plant	311,544	1,229	29,536	8,632	88.77
Water	2,304	5,717	1,439	1,929	50.20
Impervious Surface	30,768	298	110,162	6,508	74.57
Shadow	82,320	7,014	62,963	57,637	27.45
Producer's Accuracy	72.97	40.10	53.97	77.15	
Overall Accuracy (%): 67.37			Kappa Coefficient: 0.4756		

Imperfect Registered LCMs with Initial Errors in Displacement

Initial mapping parameters:

$$\theta_a = \begin{bmatrix} 0.25 & 0.0 & 5.0 \\ 0.0 & 0.25 & -5.0 \end{bmatrix}$$

Appendix Table 3 Confusion Matrix of LCM with imperfect registration in displacement, $\beta = 0.0$

Classified Map	Reference Map				
	Plant	**Water**	**Impervious Surface**	**Shadow**	**User's Accuracy (%)**
Plant	291,336	1,058	33,110	9,397	86.99
Water	7,773	6,410	2,820	7,131	26.56
Impervious Surface	32,780	321	107,301	6,245	73.17
Shadow	95,047	6,469	60,869	51,933	24.23
Producer's Accuracy	68.24	44.96	52.57	69.52	
Overall Accuracy (%): 63.47			Kappa Coefficient: 0.4246		

Appendix Table 4 Confusion Matrix of LCM with imperfect registration in displacement, $\beta = 1.0$.

Classified Map	Reference Map				
	Plant	**Water**	**Impervious Surface**	**Shadow**	**User's Accuracy (%)**
Plant	268791	3730	77459	28595	71.00
Water	4655	3319	1808	2354	27.35
Impervious Surface	68375	487	72121	9481	47.93
Shadow	85115	6722	52712	34276	19.17
Producer's Accuracy	62.9581	23.2782	35.3361	45.8812	
Overall Accuracy (%): 52.57			Kappa Coefficient: 0.2133		

Imperfect Registered LCMs with Initial Errors in Scaling

Initial mapping parameters:

$$\theta_a = \begin{bmatrix} 0.256 & 0.0 & 0.0 \\ 0.0 & 0.244 & 0.0 \end{bmatrix}$$

Appendix Table 5 Confusion Matrix of LCM with imperfect registration in scaling, $\beta = 0.0$.

Classified Map	Reference Map				
	Plant	**Water**	**Impervious Surface**	**Shadow**	**User's Accuracy (%)**
Plant	291057	1096	33658	9247	86.87
Water	8304	6417	2747	7279	25.93
Impervious Surface	33289	292	107088	6596	72.72
Shadow	94286	6453	60607	51584	24.23
Producer's Accuracy	68.1734	45.0063	52.4684	69.0493	
Overall Accuracy (%): 63.35			Kappa Coefficient: 0.4226		

Appendix Table 6 Confusion Matrix of LCM with imperfect registration in scaling, $\beta = 1.0$

Classified Map	Reference Map				
	Plant	**Water**	**Impervious Surface**	**Shadow**	**User's Accuracy (%)**
Plant	300181	1666	51285	16854	81.13
Water	3313	4905	1301	1037	46.47
Impervious Surface	46348	325	91912	7025	63.12
Shadow	77094	7362	59602	49790	25.69
Producer's Accuracy	70.3105	34.4017	45.0328	66.6479	
Overall Accuracy (%): 62.05			Kappa Coefficient: 0.3777		

CIRRICULUM VITAE

NAME : Mr. Ratchawit Sirisommai

BIRTH DATE : September 23, 1987

BIRTH PLACE : Chonburi, Thailand

EDUCATION :

YEAR	**INSTITUTE**	**DEGREE/DIPLOMA**
2010	Mahidol University	B.Sc. (Information and Communication Technology)
2012	Kasetsart University	M.Eng. (Information and Communication Technology for Embedded Systems)

POSITION/TITLE : -

WORK PLACE : -

SCHOLARSHIP/AWARDS : TAIST ICTES Master Degree Scholarship

PUBLICATION :

PUBLICATIONS :