



รายงานวิจัยฉบับสมบูรณ์

โครงการ การปรับปรุงภาพถ่ายระยะไกล และการสร้างแผนที่จำแนกชนิดจากภาพที่มีเมฆปกคลุม
และความละเอียดต่ำด้วยการรวมภาพจากหลายรูปแบบและเวลา

โดย ผศ. ดร. อีรสีธิ เกษตรเกษม

กรกฎาคม 2556

รายงานวิจัยฉบับสมบูรณ์

โครงการ การปรับปรุงภาพถ่ายระยะไกล และการสร้างแผนที่จำแนกชนิดจากภาพที่มีเมฆปกคลุม
และความละเอียดต่ำด้วยการรวมภาพจากหลายรูปแบบและเวลา

ผศ. ดร. จีรสิทธิ์ เกษตรเกษม

มหาวิทยาลัยเกษตรศาสตร์

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

กิตติกรรมประกาศ

งานวิจัยชิ้นนี้ไม่อาจจะสำเร็จได้โดยปราศจากความช่วยเหลือจากผู้มีความรอบรู้อย่าง รองศาสตราจารย์ ดอกเตอร์ อภิสิทธิ์ เอี่ยมหน่อ จาก BIOTECH และ ขอขอบคุณ ดอกเตอร์ ปรีสาร รักวาทิน จาก สทอภ. ที่ให้ความเห็นในงานวิจัยและการนำงานวิจัยไปใช้งานอย่างมีประสิทธิภาพ

สุดท้ายผู้วิจัยขอขอบคุณสำนักงานคณะกรรมการอุดมศึกษา และสำนักงานกองทุนสนับสนุนการวิจัย ที่ให้ความโอกาสในการทำงานวิจัยชิ้นนี้

บทคัดย่อ

รหัสโครงการ: RSA5480031

ชื่อโครงการ: การปรับปรุงภาพถ่ายระยะไกล และการสร้างแผนที่จำแนกชนิด จากภาพที่มีเมฆปกคลุมและความละเอียดต่ำด้วยการรวมภาพจากหลายรูปแบบและเวลา

ชื่อนักวิจัย: ผศ. ดร. อธิสิทธิ์ เกษตรเกษม

Email Address: fengtsk@ku.ac.th

ระยะเวลาโครงการ: 15 กรกฎาคม 2554 – 14 กรกฎาคม 2556

รายงานฉบับนี้ประกอบด้วยเนื้อหา 3 ส่วน คือ 1) ขั้นตอนวิธีรวมภาพโดยตรงจากภาพถ่ายดาวเทียมที่ไม่วางประกบกัน 2) ขั้นตอนวิธีการสร้างแผนที่จำแนกชนิดโดยตรงจากภาพถ่ายดาวเทียมที่ไม่วางประกบกัน และ 3) ขั้นตอนวิธีในการระบุวันปลูกข้าวจากข้อมูลภาพ MODIS ที่มีเมฆปกคลุม ในส่วนของวิธีรวมภาพโดยตรงจากภาพถ่ายดาวเทียมที่ไม่วางประกบกัน ผู้วิจัยได้นำเสนอขั้นตอนวิธีที่ใหม่ในการรวมภาพความละเอียดเชิงพื้นที่ต่ำแต่มีความละเอียดเชิงสเปกตรัมสูงกับภาพความละเอียดเชิงพื้นที่สูงแต่มีความละเอียดเชิงสเปกตรัมต่ำเข้าด้วยกัน โดยภาพทั้ง 2 ไม่จำเป็นต้องวางประกบกัน จากปกติที่ต้องวางประกบภาพให้สมบูรณ์ก่อนที่จะทำการรวมภาพ ในผลงานวิจัยนี้ ขั้นตอนการวางประกบและการรวมภาพสามารถที่จะกำหนดให้อยู่ในรูปแบบของปัญหาการประมาณค่า ดังนั้นพารามิเตอร์ของการวางประกบภาพสามารถที่จะปรับค่าโดยอัตโนมัติ เพื่อให้ขั้นตอนการรวมภาพและการวางประกบภาพสามารถที่จะกระทำพร้อมกันได้ ขั้นตอนที่น่าเสนอให้รายงานฉบับนี้ อาศัยหลักเกณฑ์การประมาณค่าแบบ Maximum *a posteriori* (MAP) เพื่อใช้ในการรวมภาพ และใช้วิธีการหาค่าที่ดีที่สุดแบบ Metropolis เพื่อหาค่าที่เหมาะสมที่สุดในส่วนของการวางประกบและรวมภาพไปพร้อมๆกัน คณะผู้วิจัยได้ทำการคิดค้นสมการสำหรับการหาภาพรวมที่เหมาะสมที่สุดมาในรายงานฉบับนี้ด้วย ผลการทดสอบจากการทดลองแบบต่างๆ พบว่าผลการรวมภาพของขั้นตอนวิธีที่น่าเสนอสร้างภาพที่มีคุณภาพเชิงพื้นที่ และคุณภาพเชิงสเปกตรัมดี ถึงแม้ว่าภาพทั้ง จะวางประกบผิดพลาดในตอน 2 เริ่มต้นอยู่มากก็ตาม

ในการทำแผนที่จำแนกชนิดจากข้อมูลภาพที่ไม่วางประกบกันอย่างสมบูรณ์ รายงานฉบับนี้นำเสนอขั้นตอนวิธีการแบ่งแยกชนิดวัตถุร่วมกับการวางประกบภาพโดยอาศัยโมเดล Markov random field (MRF) เพื่อที่จะวางประกบภาพไปพร้อมๆกับการทำแผนที่ คณะผู้วิจัยได้อาศัยขั้นตอนวิธี Expectation-Maximization (EM) ในการแก้ปัญหการแบ่งแยกชนิดวัตถุร่วมกับการวางประกบภาพ โดยประมาณค่าการวางประกบสลับกับการประมาณค่าความน่าจะเป็นแบบ posterior ไปหลายๆรอบ หลักจากนั้นได้นำเกณฑ์แบบ Maximum *a posteriori* (MAP) มาใช้ในการทำแผนที่จำแนกชนิด ผู้วิจัยได้ทำการทดสอบขั้นตอนวิธีที่พัฒนาขึ้นกับชุดข้อมูลหลายๆชนิดทั้งแบบที่เป็นแบบจำลองและภาพจริง ผลที่ได้พบว่าเมื่อค่าพารามิเตอร์ของ MRF ได้รับการเลือกที่เหมาะสมแผนที่ที่ได้มีความถูกต้องเท่ากับกรณีที่มีข้อมูลภาพทั้งหมดถูกวางประกบโดยสมบูรณ์

ในส่วนของประมาณวันเพาะปลูกข้าว คณะผู้วิจัยได้พัฒนาวิธีการประมาณวันเพาะปลูกข้าวจากข้อมูล normalized difference vegetation index (NDVI) แบบทุกๆ วันจากภาพถ่าย 8 moderate resolution imaging spectroradiometer (MODIS) สัญญาณ NDVI จะถูกโมเดลด้วย triply modulated cosine function ที่ประกอบด้วยพารามิเตอร์หลัก ตัวคือ ค่าเฉลี่ย ขนาดสัญญาณ และ เฟสของสัญญาณ 3

และได้Extended Kalman filter (EKF) ในการประมาณพารามิเตอร์หลักทั้ง 3 จากโมเดลนี้ คณะผู้วิจัยสามารถลดผลกระทบของเมฆต่อสัญญาณ NDVI และพบว่าวันปลูกข้าวตรงกับวันที่ส่วนใหญ่เป็นฤดู มีค่าเพิ่มขึ้นจนเกินค่า threshold ที่กำหนดไว้ จากการทดลองพบว่าขั้นตอนวิธีที่พัฒนาขึ้นสามารถระบุวันเพาะปลูกได้ผิดพลาดน้อยกว่า 16 วันจากข้อมูลที่กรมการข้าวได้ถึง 73.3 และ 80 เปอร์เซ็นต์สำหรับนาปี และนาปรังที่ปลูก 2 ครั้งต่อปีตามลำดับ

คำหลัก: โมเดลมาคอฟฟ์แรนดอมฟิลด์ การรวมภาพพร้อมกับการวางประทับภาพ การทำแผนที่พร้อมกับการวางประทับภาพ การประมาณวันเพาะปลูกข้าว

ABSTRACT

Project Code: RSA5480031

Project Title: Remote Sensing Image Enhancement and Land Cover Mapping from Cloud-Covered and Low Resolution Images by Fusing with Images from Different Modalities and Times

Principle Investigator Asst. Dr. Teerasit Kasetkasem

Email Address: fengtsk@ku.ac.th

Project Period 15 July 2011 – 30 June 2013

The work in this report consists of three parts: 1) a joint image fusion algorithm from mis-registered image pairs, 2) a land cover mapping algorithms algorithm for mis-registered image datasets and 3) rice cultivation date estimation using cloudy multi-temporal Moderate Resolution Imaging Spectroradiometer (MODIS). In a joint image fusion algorithm from mis-registered image pairs, a new method for fusion and registration of low spatial and high spectral resolution image with high spatial and low spectral resolution image in a single step is developed. In the usual procedure, fusion is an independent process separated from registration. However, both image registration and fusion can be formulated as estimation problems. Hence, the registration parameters can be automatically tuned so that both fusion and registration can be optimized simultaneously. Here, we concentrate on the relationship between low-resolution multispectral and high-resolution panchromatic imagery. The proposed technique is based on a statistical model. It employs the maximum a posteriori (MAP) estimator to solve the fusion problem, and applies the Metropolis algorithm to solve the joint optimization problem. A closed-form solution to find the fused high spatial and spectral resolution image with correcting registration is also derived here. We examine our algorithm into datasets.

In a land cover mapping algorithms algorithm for mis-registered image datasets, we proposed a joint classification and registration technique based on a Markov random field (MRF) model to simultaneously align two or more images and obtain a land cover map (LCM) of the scene. The expectation maximization (EM) algorithm is employed to solve the joint image classification and registration problem by iteratively estimating the map parameters and approximate posterior probabilities. Then, the maximum *a posteriori* (MAP) criterion is used to produce an optimum land cover map. We conducted experiments on a set of four simulated images and one pair of remotely sensed images to investigate the effectiveness and robustness of the proposed algorithm. Our results show that, with proper selection of a critical MRF parameter, the resulting LCMs derived from an unregistered image pair can achieve an accuracy as high as when images are perfectly aligned. Furthermore, the registration error can be greatly reduced

In our last work, we developed a rice cultivation date estimation based on remote sensing data, since rice is critical information to evaluate the damages in rice fields from natural disasters. In this study, the 8-day composite normalized difference vegetation index (NDVI) derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) data was modeled as a triply modulated cosine function, and the extended Kalman filter (EKF) is

used to estimate the mean, amplitude and phase parameters of the cosine function. The cultivation dates are estimated as the date where the seasonal variation derived from the EKF is greater than a threshold after its minimum. From the experimental results, the estimated cultivation dates derived from the proposed algorithm agree with rice cultivation information from the National Rice Department. The 73.3 percentages of the estimated cultivation dates is within 16 days for the rain-fed rice areas, and more than 80 percentages of the estimated data is within 16 days for irrigated areas with two crop cycles per year

Keyword: Markov Random Field, joint image fusion and registration, joint image land cover mapping and registration, rice cultivation date estimation

Table of Contents

กิตติกรรมประกาศ	๒
บทคัดย่อ	๓
ABSTRACT	๕
Table of Contents	๗
บทคัดย่อสำหรับผู้บริหาร	- 1 -
Executive Summary	- 4 -
Research Detail	- 7 -
1. Image Fusion Algorithm for Mis-registered Images	- 7 -
1.1 Introduction	- 7 -
1.2 Observed Model	- 8 -
1.3 Optimum Image Fusion Rules	- 10 -
1.4 Experimental Results	- 14 -
2. Land Cover Mapping algorithm from Mis-Registered Image Datasets	- 29 -
2.1 Introduction	- 29 -
2.2 Problem Statement	- 31 -
2.3 Optimum Image Registration and Land Cover Mapping Criteria	- 35 -
2.4 Joint Image Registration and Land Cover Mapping Algorithm	- 37 -
2.5 Experimental Results	- 41 -
3. Rice Cultivation Date Estimation Using the Cloudy-Multi-Temporal Moderate Resolution Imaging Spectroradiometer	- 60 -
3.1 Introduction	- 60 -
3.2 Data Description	- 61 -
3.3 Proposed Algorithm	- 64 -
3.4 Experimental Results	- 65 -
3.5 Discussion	- 68 -
4. References	- 70 -
Outputs	- 76 -
Appendix	- 77 -

บทคัดย่อสำหรับผู้บริหาร

รหัสโครงการ: RSA5480031

ชื่อโครงการ: การปรับปรุงภาพถ่ายระยะไกล และการสร้างแผนที่จำแนกชนิด จากภาพที่มีเมฆปกคลุมและความละเอียดต่ำด้วยการรวมภาพจากหลายรูปแบบและเวลา

ชื่อนักวิจัย: ผศ. ดร. อีรสิทธิ์ เกษตรเกษม

Email Address: fengtsk@ku.ac.th

ระยะเวลาโครงการ: 15 กรกฎาคม 2554 – 14 กรกฎาคม 2556

ภาพถ่ายระยะไกล (Remote Sensing Images) ที่ถูกถ่ายจากมุมสูงนั้นมีประโยชน์อย่างมากต่อการจัดการพืชผลทางการเกษตร เนื่องจากภาพถ่ายระยะไกล 1 รูปสามารถครอบคลุมพื้นที่เป็นบริเวณกว้าง และภาพเหล่านี้สามารถทำการถ่ายภาพซ้ำๆบนพื้นที่ต่างในประเทศไทยในเวลาเพียงไม่กี่วัน ทำให้การนำภาพถ่ายระยะไกลมาช่วยในการประมาณการปริมาณพืชผลทางการเกษตร จะได้รับข้อมูลที่ใหม่และครบถ้วนอยู่เสมอหน่วยงานรัฐบาลได้นำภาพถ่ายระยะไกลมาใช้ในการแก้ไขและจัดการหลายๆอย่างในภาคการเกษตรและนอกภาคการเกษตรบ้างแล้ว เช่นการวางแผนการทรัพยากรน้ำ การจัดการป่าไม้ หรือแม้แต่ปัญหาภัยพิบัติต่างๆ ตัวอย่างเช่นภายหลังการเกิดภัยพิบัติสึนามิในปี 2547 ประเทศไทยได้มีการนำภาพถ่ายระยะไกลมาช่วยบ่งบอกและประมาณการความเสียหายที่เกิดจากภัยพิบัติ เช่นเดียวกันในฤดูฝนปี 2554 ที่เพิ่งผ่านมาได้มีการนำภาพถ่ายระยะไกลมาช่วยในการจัดการอุทกภัยของประเทศ

ถึงแม้ว่าเรามีแหล่งข้อมูลภาพถ่ายดาวเทียมจำนวนมาก การนำภาพถ่ายระยะไกลมาใช้งานต่างๆ ยังประสบปัญหาหลายด้าน เนื่องด้วยข้อจำกัดทางกายภาพของภาพถ่าย และสภาพแวดล้อมของประเทศไทย ซึ่งปัญหาที่พบบ่อยได้แก่ 1) เมฆที่ปกคลุมภาพถ่ายระยะไกลทำให้ไม่สามารถมองเห็นวัตถุที่ถูกบดบัง 2) เงาที่บดบังวัตถุที่ถูกปกคลุมโดยเงา และ 3) ความละเอียดเชิงพื้นที่และสเปกตรัมแสงของภาพที่ทำให้จำแนกชนิดวัตถุได้ไม่ถูกต้องตามที่ต้องการ

เพื่อแก้ปัญหาเหล่านี้ จำเป็นต้องใช้ภาพถ่ายระยะไกลจากหลายๆเวลา หรือจากหลายๆดาวเทียมเพื่อลดผลกระทบของเมฆและเงาในภาพ และยังเพื่อเพิ่มคุณภาพเชิงพื้นที่และสเปกตรัมของภาพ อย่างไรก็ตามเมื่อนำภาพหลายภาพมาใช้งานร่วมกัน มีความจำเป็นต้องวางประทับภาพเหล่านั้นให้ดี การวางประทับภาพคือการนำภาพถ่ายดาวเทียมอยู่ในแนวเดียวกัน เมื่อภาพเหล่านี้อยู่ในแนวเดียวกันแล้ว จึงจะนำภาพไปใช้ได้ ถ้าขั้นตอนนี้ไม่สมบูรณ์ประสิทธิภาพของการนำภาพไปใช้จะลดลง อย่างไรก็ตามไม่ขั้นตอนการวางประทับภาพใดที่สมบูรณ์ และการวางประทับภาพแต่ละอย่างมีวิธีการและขั้นตอนที่ต่างกัน ดังนั้นในงานวิจัยชิ้นนี้จึงเสนอให้รวมขั้นตอนการประทับภาพเข้ากับขั้นตอนการวิเคราะห์ภาพเป็นขั้นตอนเดียว เพื่อลดขั้นตอน และเพื่อให้ขั้นตอนการวิเคราะห์ภาพพิจารณาความผิดพลาดของการวางประทับภาพไว้ด้วย สุดท้ายผู้วิจัยจะนำภาพถ่ายดาวเทียมหลายๆภาพมารวมกัน เพื่อลดผลกระทบของเมฆในภาพและใช้ภาพเหล่านี้ในการระบุวันเพาะปลูกข้าวด้วย

การดำเนินโครงการ

โครงการนี้ ได้ดำเนินการศึกษาและพัฒนาวิธีการอยู่หลายแบบโดยมุ่งเน้นไปที่การนำภาพหลายๆชนิดมารวมเข้าด้วยกัน เพื่อสร้างเป็นแผนที่จำแนกชนิดที่มีความถูกต้องมากขึ้น คณะผู้วิจัยได้แบ่งการศึกษาออกเป็น 3 ส่วนหลักๆคือ

1. พัฒนาขั้นตอนวิธีรวมภาพโดยตรงจากภาพถ่ายดาวเทียมที่ไม่วางประทับกัน
2. พัฒนาขั้นตอนวิธีการสร้างแผนที่จำแนกชนิดโดยตรงจากชุดข้อมูลภาพถ่ายดาวเทียมที่ไม่วางประทับกัน
3. พัฒนาขั้นตอนวิธีในการระบุวันปลูกข้าวจากข้อมูลภาพ MODIS ที่มีเมฆปกคลุม

ขั้นตอนวิธีในการรวมภาพ

ในส่วนนี้ ผู้วิจัยได้พัฒนาวิธีการรวมภาพแบบใหม่ ที่รวมขั้นตอนการรวมภาพและวางประทับภาพเข้าด้วยกันให้เหลือเพียง 1 ขั้นตอน จากปกติที่มี 2 ขั้นตอน โดยเริ่มจากการวางประทับภาพ และหลังจากนั้นจะถูกนำไปรวมกันภายใต้ข้อสมมติที่ว่า การวางประทับภาพทำได้สมบูรณ์ และไม่มีข้อผิดพลาดใดๆเกิดขึ้น อย่างไรก็ตาม ข้อสมมตินี้อาจไม่เป็นจริงเสมอไป ดังนั้นจึงเสนอวิธีการรวมภาพที่คำนึงถึงว่าที่นำมารวมกันอาจจะไม่วางประทับกันอย่างสมบูรณ์ เพื่อให้รวมขั้นตอนทั้ง 2 เข้าด้วยกัน ขั้นตอนการวางประทับและการรวมภาพจะถูกดัดแปลงให้อยู่ในรูปแบบของปัญหาการประมาณค่า ดังนั้นพารามิเตอร์ของการวางประทับภาพสามารถที่จะปรับค่าโดยอัตโนมัติ เพื่อให้ขั้นตอนการรวมภาพและการวางประทับภาพสามารถที่จะกระทำพร้อมกันได้ เพื่อหาภาพรวมที่ดีที่สุด ผู้วิจัยได้นำหลักเกณฑ์การประมาณค่าแบบ Maximum *a posteriori* (MAP) เพื่อใช้หาภาพรวมที่เหมาะสมที่สุด และใช้วิธีการหาค่าที่ดีที่สุดแบบ Metropolis เพื่อหาค่าพารามิเตอร์ในการวางประทับภาพที่เหมาะสมที่สุด คณะผู้วิจัยได้ทำการคิดค้นสมการสำหรับการหาภาพรวมที่เหมาะสมที่สุดมาในรายงานฉบับนี้ด้วย ผลการทดสอบจากการทดลองแบบต่างๆ พบว่าผลการรวมภาพของขั้นตอนวิธีที่นำเสนอสร้างภาพที่มีคุณภาพเชิงพื้นที่ และคุณภาพเชิงสีที่ดี ถึงแม้ว่าภาพทั้ง 2 จะวางประทับผิดพลาดในตอนเริ่มต้นอยู่มากก็ตาม

ขั้นตอนวิธีการสร้างแผนที่จำแนกชนิดโดยตรงจากภาพถ่ายดาวเทียมที่ไม่วางประทับกัน

วิธีการทำแผนที่จากภาพจากชุดข้อมูลภาพที่วางประทับกันอย่างสมบูรณ์ ไม่ว่าข้อมูลภาพเหล่านี้จะมาจากภาพจากหลายๆเวลา หรือภาพจากหลายๆชนิด โดยปกติมักจะสมมติว่าภาพทั้งหมดที่นำมาใช้ได้รับการวางประทับกันอย่างสมบูรณ์ อย่างไรก็ตามเพราะภาพเหล่านี้ถ่ายกันคนละเวลา หรือถ่ายโดยดาวเทียมคนละดวง ข้อสมมติที่ว่า การวางประทับกันอย่างสมบูรณ์มักจะไม่เป็นจริงเสมอไป ดังนั้นขั้นตอนการทำแผนที่จำแนกชนิดที่ดีควรจะแก้ไขการวางประทับที่ผิดพลาดเหล่านี้ให้ดีขึ้นเท่าความสามารถในการแบ่งแยกวัตถุได้ถูกต้องแม่นยำ ในรายงานฉบับนี้นำเสนอขั้นตอนวิธีการแบ่งแยกชนิดวัตถุร่วมกับการวางประทับภาพโดยอาศัยโมเดล Markov random field (MRF) เพื่อที่จะวางประทับภาพไปพร้อมๆกับการทำแผนที่ คณะผู้วิจัยได้อาศัยขั้นตอนวิธี Expectation-Maximization (EM) ในการแก้ปัญหการแบ่งแยกชนิดวัตถุร่วมกับการวางประทับภาพ โดยประมาณค่าการวางประทับสลับกับการประมาณค่าความน่าจะเป็นแบบ posterior ไปหลายๆรอบ

หลักจากนั้นได้นำเกณฑ์แบบ Maximum *a posteriori* (MAP) มาใช้ในการทำแผนที่จำแนกชนิด ผู้วิจัยได้ทำการทดสอบขั้นตอนวิธีที่พัฒนาขึ้นกับชุดข้อมูลหลายๆชนิดทั้งแบบที่เป็นแบบจำลองและภาพจริง ผมที่ได้พบว่าเมื่อค่าพารามิเตอร์ของ MRF ได้รับการเลือกที่เหมาะสมแผนที่ที่ได้มีความถูกต้องเท่ากับกรณีที่มีข้อมูลภาพทั้งหมดถูกวางประทับโดยสมบูรณ์

ขั้นตอนวิธีในการระบุวันปลูกข้าวจากข้อมูลภาพ MODIS ที่มีเมฆปกคลุม

ในส่วนนี้ คณะผู้วิจัยได้พัฒนาวิธีการประมาณวันเพาะปลูกข้าวโดยอาศัยข้อมูล normalized difference vegetation index (NDVI) แบบทุกๆ 8 วันจากภาพถ่าย moderate resolution imaging spectroradiometer (MODIS) การรู้วันเพาะปลูกข้าวนี้สำคัญต่อการประเมินความเสียหายที่เกิดขึ้นจากภัยพิบัติไม่ว่าจะเกิดจากน้ำท่วมหรือภัยแล้ง และชดเชยความเสียหายให้กับเกษตรกรอย่างเหมาะสม คณะผู้วิจัยเลือกใช้ภาพถ่าย MODIS เพราะข้อมูลภาพเกือบทุกวัน และยังสามารถดาวน์โหลดได้โดยไม่มีค่าใช้จ่าย นอกจากนี้แล้วข้อมูล NDVI ของ MODIS ยังได้รับความนิยมในการศึกษาในเรื่องของชีพลักษณ์ของพืชทางการเกษตร (crop phenology)

ในรายงานฉบับนี้ผู้วิจัยได้โมเดล NDVI ด้วย triply modulated cosine function ที่ประกอบด้วยพารามิเตอร์หลัก 3 ตัวคือ ค่าเฉลี่ย ขนาดสัญญาณ และ เฟสของสัญญาณ และใช้ Extended Kalman filter (EKF) ในการประมาณพารามิเตอร์หลักทั้ง 3 จากโมเดลนี้ คณะผู้วิจัยพบว่าวันปลูกข้าวตรงกับวันที่ส่วนที่แปรผันเป็นฤดู มีค่าเพิ่มขึ้นจนเกินค่า threshold ที่กำหนดไว้ จากการทดลองพบว่าขั้นตอนวิธีที่พัฒนาขึ้นสามารถระบุวันเพาะปลูกได้ผิดพลาดน้อยกว่า 16 วันจากข้อมูลที่กรมการข้าวได้ถึง 73.3 และ 80 เปอร์เซ็นต์สำหรับนาปี และนาปรังที่ปลูก 2 ครั้งต่อปีตามลำดับ

ผลงานที่ได้:

1. T. Sritarapipat, **T. Kasetkasem**, and P. Rakwatin, "Fusion and Registration of THEOS Multispectral and Panchromatic Images," under review at the *International Journal of Remote Sensing*
2. **T. Kasetkasem**, P. Rakwating, R. Sirirsommai and A. Eiumnoh, "a Joing Land Cover Mapping and Image Registration Algorithm Based on a Markov Random Field Model," under review at *Remote Sensing*
3. **T. Kasetkasem**, P. Rakwatin, R. Sirisommai, A. Euimnoh, and T. Isshiki, "A MRF-Based Approach for a Multisensor Land Cover Mapping of Mis-Registered Images" *The 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS2012)*, July 22-27, Munich, Germany

บทสรุปโดยภาพรวม โครงการฯ ได้ดำเนินงานบรรลุวัตถุประสงค์ และโจทย์วิจัยที่ตั้งเป้าไว้ทุกประการ

Executive Summary

Project Code: RSA5480031

Project Name: Remote Sensing Image Enhancement and Land Cover Mapping from Cloud-Covered and Low Resolution Images by Fusing with Images from Different Modalities and Times

Principle Investigator: Assistant Professor Dr. Teerasit Kasetkasem

Email Address: fengtsk@ku.ac.th

Project Period: July 15th, 2011- July 14th, 2013

Remote Sensing Images due to their synoptic view and map like format are the crucial part of agriculture monitoring and management systems since one remote sensing image can capture a wide area. Furthermore, some remote sensing satellites can repeatedly capture images of the same scene for every 1 or 2 days. As a result, the remote sensing images provide timely and complete information on the ground cover. The various sectors in the Thai government, thus, have used remote sensing images in both agriculture and non-agriculture applications. For example, in 2004, the Thai government used remote sensing images in disaster assessment after the world biggest Tsunami, and in 2011, the remote sensing images was the vital parts in flood control and disaster relief effort in the recent massive flood.

Although remote sensing images have shown to be very promising, the use of remote sensing images in Thailand is still limited due to the physical and environmental limitations in Thailand. The common problems are 1) high cloud coverage, 2) high shadows and 3) the low spatial and spectral resolutions of available remote sensing images.

To tackle these problems, there is a need of combining multiple remotely sensed images together to reduce the effect of clouds and shadows and increase the spatial and spectral resolutions. These remotely sensed images can be from multiple times or multiple satellites. For multi-sensor image analysis, the registration process is a crucial initial step to the use of multiple images. Registration aligns multiple satellite images into a common coordinate system. Only when all of the input images are perfectly registered, the use of multiple images can be performed. Otherwise mis-registration will reduce the performance. In practice, perfect registration may not always be achievable since there are some unknown variations on satellite platforms and flight paths when capturing images. As a result, the performance is likely to suffer from mis-registration effects. As a result, in this research project, we attempt to combine image registration algorithm with image analysis techniques, namely, image fusion and land cover mapping to increase the image analysis. Since the proposed image analysis algorithms can handle mis-registered images, the number of steps to analyze images is reduced, and the image analysis algorithms are robust to image mis-alignment. Lastly, we attempt to use multi-temporal MODIS images to reduce the effect of cloud in the rice cultivation date estimation.

Operations Research

Since the goal of this research is aimed to fuse multiple images together to increase the accuracy in image analysis algorithm. We divide our research into three parts:

1. Remote sensing image fusion algorithm for mis-registered images

2. Land Cover Mapping algorithm from Mis-Registered Image Datasets
3. Rice cultivation date estimation using cloudy-multi-temporal Moderate Resolution Imaging Spectroradiometer (MODIS)

Remote sensing image fusion algorithm for mis-registered images

This work presents a new method for fusion and registration of low spatial and high spectral resolution image with high spatial and low spectral resolution image in a single step. In the usual procedure, fusion is an independent process separated from registration. However, both image registration and fusion can be formulated as estimation problems. Hence, the registration parameters can be automatically tuned so that both fusion and registration can be optimized simultaneously. Here, we concentrate on the relationship between low-resolution multispectral and high-resolution panchromatic imagery. The proposed technique is based on a statistical model. It employs the maximum a posteriori (MAP) estimator to solve the fusion problem, and applies the Metropolis algorithm to solve the joint optimization problem. A closed-form solution to find the fused high spatial and spectral resolution image with correcting registration is also derived here. We examine our algorithm into datasets. In the first dataset, a THEOS multispectral image with high spectral resolution and a THEOS panchromatic image with high spatial resolution are combined to produce a multispectral image with high spectral and spatial resolution. Similarly, in the second datasets, we fuse a high spectral resolution hyperspectral image with a high spatial resolution CCD image from SMMS satellite to produce a high-spatial resolution hyperspectral image. The results of our experiment show that the proposed fusion and registration algorithm can produce high quality high-resolution images from low spatial and high spectral resolution image and high spatial and low spectral resolution images even when they are severely mis-registered.

Land Cover Mapping algorithm from Mis-Registered Image Datasets

Traditional multi-modal and multi-temporal land cover mapping algorithms assume that all images are perfectly aligned. However, since multi-modal and multi-temporal images are likely to be obtained from different satellite platforms and/or acquired at different times, perfect alignment is very difficult to achieve. As a result, a proper land cover mapping algorithm must be able to correct registration errors as well as perform an accurate classification. In this paper, we proposed a joint classification and registration technique based on a Markov random field (MRF) model to simultaneously align two or more images and obtain a land cover map (LCM) of the scene. The expectation maximization (EM) algorithm is employed to solve the joint image classification and registration problem by iteratively estimating the map parameters and approximate posterior probabilities. Then, the maximum *a posteriori* (MAP) criterion is used to produce an optimum land cover map. We conducted

experiments on a set of four simulated images and one pair of remotely sensed images to investigate the effectiveness and robustness of the proposed algorithm. Our results show that, with proper selection of a critical MRF parameter, the resulting LCMs derived from an unregistered image pair can achieve an accuracy as high as when images are perfectly aligned. Furthermore, the registration error can be greatly reduced

Rice cultivation date estimation using cloudy-multi-temporal Moderate Resolution Imaging Spectroradiometer (MODIS)

Rice cultivation date estimation based on remote sensing data is critical information to evaluate the damages in rice fields from natural disasters. In this study, the 8-day composite normalized difference vegetation index (NDVI) derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) data was modeled as a triply modulated cosine function, and the extended Kalman filter (EKF) is used to estimate the mean, amplitude and phase parameters of the cosine function. The cultivation dates are estimated as the date where the seasonal variation derived from the EKF is greater than a threshold after its minimum. From the experimental results, the estimated cultivation dates derived from the proposed algorithm agree with rice cultivation information from the National Rice Department. The 73.3 percentages of the estimated cultivation dates is within 16 days for the rain-fed rice areas, and more than 80 percentages of the estimated data is within 16 days for irrigated areas with two crop cycles per year

Output:

1. T. Sritarapipat, **T. Kasetkasem**, and P. Rakwatin, "Fusion and Registration of THEOS Multispectral and Panchromatic Images," under review at the *International Journal of Remote Sensing*
2. **T. Kasetkasem**, P. Rakwating, R. Sirirsommai and A. Eiumnoh, "a Joing Land Cover Mapping and Image Registration Algorithm Based on a Markov Random Field Model," under review at *Remote Sensing*
3. **T. Kasetkasem**, P. Rakwatin, R. Sirisommai, A. Euimnoh, and T. Isshiki, "A MRF-Based Approach for a Multisensor Land Cover Mapping of Mis-Registered Images" *The 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS2012)*, July 22-27, Munich, Germany

Conclusion: This project has successful answer the project objectives and research problems

Research Detail

1. *Image Fusion Algorithm for Mis-registered Images*

1.1 Introduction

The fusion of multi-sensor image data is a widely used procedure for enhancing information extraction. Fusion uses already-existing data from different sources, acquired at different times, or captured by different sensors, to produce higher resolution data. When high spatial resolution data are merged with the high spectral resolution data, the fused image has both finer spatial and finer spectral resolution.

Image fusion algorithms can be divided into three categories, namely, feature-based, pixel-based and transform-based methods. The feature-based methods first segment input sources into homogeneous regions by using some segmentation techniques [1] and [2] according to texture characteristics. Next, a data combination algorithm combines texture information from different images together to produce a texture-enhanced image. The transformation-based methods convert the input images into a common transformed domain, such as intensity-hue-saturation (IHS) [3], principle components (PCA) [4], Brovey [5], 2D wavelet [6], [7] and [8], 3D wavelet [9], pyramids [10] or high pass filtering [11]. Image fusion is performed by combining the transform coefficients from all images. After that, the combined image is transformed by the inverse-transformation into the original image space. In the pixel-based methods [12] and [13], a pixel in the fused image is determined from a set of pixels from the input sources. For example, Hardie [14] and Li [15] applied statistical models to a pixel-based approach where intensity values from corresponding pixels from fusing images statistically depend on the intensity value of a fused image. The approach detailed in this paper is also pixel-based and use a statistical model.

Image registration is the process of transforming different sets of data into a single coordinate system. Data may come from different sensors, from different times, or from different viewpoints. Current registration methods can be classified into three categories, feature-based, transform-based, and intensity-based methods [16]. The feature-based approaches usually extract geometric features, also known as control points, such as intersections and landmarks, and use a least square criterion to estimate the registration parameters [17] and [18]. The transform-based methods work with images in the frequency domain [19] and [20] to compute the registration parameters by utilizing the properties of translation and rotation under the Fourier transform. In intensity-based methods, the registration parameters are estimated by maximizing some similarity measure between pixel values of the input images. Normalized cross-correlation [21], LS [22], and maximum likelihood (ML) [23] are some popular criteria used to measure similarity.

Although many algorithms have been proposed for image registration and image fusion, the operations are traditionally viewed as two independent processes. More precisely, image registration is normally performed first, followed by image fusion, which usually assumes that

perfect registration has been achieved [12], [13] and [24]. However, in practice, the registration process is not guaranteed to be error-free, and this registration error can negatively affect the fusion performance. The quality of fused images degrades significantly when the multispectral and panchromatic images are not perfectly registered.

The work by Chen [25] attempts to join the fusion and registration processes together by using the expectation maximization (EM) algorithm. However, they assume that the intensity value in each pixel of the high spatial resolution image can be selected from any one of the fused images. In the case of spatial enhancement, the intensity values of the fused image should follow the intensity values from the multispectral image only. As a result, the algorithm proposed in [25] is not suitable for spatial enhancement by combining multispectral and panchromatic images (so called “pan-sharpening”). In pan-sharpening, the fused image must have the same color distribution as the multispectral image but contain the spatial detail from the higher resolution panchromatic image.

In this paper, we develop the maximum a posteriori (MAP) estimator [26] to simultaneously fuse and register multispectral and panchromatic images to produce a high-resolution multispectral image. Here, we focus on the use of high-resolution panchromatic data to enhance multispectral imagery. However, the estimation framework developed here can be extended for any number of spectral bands in the primary and auxiliary sensors. The proposed technique is suitable for applications where some correlation, either localized or global, exists between the auxiliary image and the image being enhanced. A spatially varying statistical model is used to help exploit localized correlations between the primary and auxiliary image. Another important aspect of the proposed algorithm is that it uses an accurate observation model relating the “true” scene to the low-resolution observations. This means that a potentially wavelength-dependent spatially-varying system point spread function (PSF) can be incorporated into the estimator.

1.2 Observed Model

Let $X(\mathcal{S}) \in \mathfrak{R}^{M \times B}$ be a fine spatial and spectral resolution image (FSSRI) having M pixels and B bands where $\mathfrak{R}$ is a real number. Here, $\mathcal{S} = \{s_1, \dots, s_M\}$ denotes a set of pixels registered to some real world scene. The FSSRI is usually represented in a vector form so that, $\mathbf{x}(s_i) \in \mathfrak{R}^B$ is a fine spatial and spectral resolution vector (FSSRV) containing intensity values of a pixel s_i in the FRMI from all spectral bands. Here, we assume that the FSSRVs from different pixels are statistically independent, and have identical multivariate Gaussian distribution with a mean vector, $\boldsymbol{\mu}_X$, and a covariance matrix, C_X . The marginal probability density function (PDF) of the FRMI can therefore be written as

$$\Pr(X) = \prod_{s_i \in \mathcal{S}} \frac{\exp\left[-\frac{1}{2}(\mathbf{x}(s_i) - \boldsymbol{\mu}_X)^T C_X^{-1} (\mathbf{x}(s_i) - \boldsymbol{\mu}_X)\right]}{(2\pi)^{\frac{B}{2}} \sqrt{|C_X|}} \quad (1.1)$$

We omit $\mathcal{S}$ for the sake of abbreviation. In this paper, we assume that the direct observation of the FRMI is impossible, but we, instead, observe the coarse spatial and fine spectral resolution image (CSFSRI), and the fine spatial and coarse spectral resolution image (FRCRSRI) of the same scene.

Let $Y(\mathcal{T}) \in \mathbb{R}^{N \times B}$ be the CSFSRI with $N < M$ pixels and B bands where $\mathcal{T} = \{t_1, \dots, t_N\}$ is a set of pixels belonging to the CSFSRI. Since both CSFSRI and FSSRI are taken from the same scene, the relationship between $\mathcal{S}$ and $\mathcal{T}$ can be determined. Let us denote a coordinate of a pixel s in a FSSRI as (x, y) where x and y are the row and column of $\mathbf{x}(s)$. Similarly, we can write $t = (u, v)$ where u and v are the row and column of the pixel t in the CSFSRI. Hence, the relationship between s and t can be written as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \eta \cos \theta & \eta \sin \theta \\ -\eta \sin \theta & \eta \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d_x \\ d_y \end{bmatrix} \quad (1.2)$$

where η , θ , d_x , and d_y are scale, rotation angle, translation in column direction and translation in row direction between a pixel coordinates η and t , respectively. We denote $\omega = [\eta \ \theta \ d_x \ d_y]^T$ as the parameter vector. In practice, since the panchromatic and multispectral images are captured from the same satellite, they are already somewhat pre-registered to each other within the predefined accuracy range. As a result, the parameter vector can be assumed to be uniformly distributed within the defined set, i.e.,

$$\Pr(\omega) = \begin{cases} \frac{1}{|\Omega_\omega|}, & \omega \in \Omega_\omega \\ 0, & \text{otherwise} \end{cases} \quad (1.3)$$

where Ω_ω is the set of all possible values in the parameter space. If the parameter vector is known, the CSFSRI can be remapped and resampled into the coordinate system $\mathcal{S}$. As a result, let us denote $Y_m(\mathcal{S})$ as the remapped version of CSFSRI into $\mathcal{S}$. Clearly both $X(\mathcal{S})$ and $Y_m(\mathcal{S})$ have equal numbers of pixels and spectral bands. However, $X(\mathcal{S})$ has finer and clear detail than $Y_m(\mathcal{S})$ since all the remapping algorithms use only information from the coarser resolution image. We model this loss of finer and clearly detail as additive noise, i.e.,

$$\mathbf{y}_m(s) = \mathbf{x}(s) + \mathbf{n}_c(s) \quad (1.4)$$

where $\mathbf{y}_m(s) \in \mathbb{R}^B$ is the vector containing intensity values from all bands of the remapped CSFSRI, and $\mathbf{n}_c(s) \in \mathbb{R}^B$ is the additive noise vector. We assume further that noise vectors from all pixels are independent and Gaussian distributed with a zero mean vectors and a covariance matrix, C_c . Hence, the conditional PDF of $Y(\mathcal{S})$ give $X(\mathcal{S})$ and the map parameter is given by

$$\Pr(Y|X, \omega) = \prod_{s \in \mathcal{S}} \Pr(\mathbf{y}_m(s) | \mathbf{x}(s))$$

$$= \prod_{s \in S} \frac{\exp \left[-\frac{1}{2} (\mathbf{y}_m(s_i) - \mathbf{x}(s_i))^T \mathbf{C}_C^{-1} (\mathbf{y}_m(s_i) - \mathbf{x}(s_i)) \right]}{(2\pi)^{\frac{B}{2}} \sqrt{|\mathbf{C}_C|}} \quad (1.5)$$

Next, let $Z(S) \in \mathbb{R}^{M \times b}$ be the observed FSCSI having M pixels and $b < B$ bands. Here, we assume that the FSCSI has the same resolution as the FSSRI and is perfectly registered with the coordinate system of FSSRI. We assume further that the FSCSI is the weighted band summation of the FSSRI plus an additive noise, i.e.,

$$\mathbf{z}(s) = W\mathbf{x}(s) + \mathbf{n}_z(s) \quad (1.6)$$

where $\mathbf{z}(s)$ is the observation of the FSCSI at the pixel s , W is weight matrix, and $\mathbf{n}_z(s)$ is the additive noise to the FSCSI. Again, we also assume that noise vectors for all the pixels are independent and Gaussian distributed with a zero mean vector and a covariance matrix, $\mathbf{C}_z$, and that they are also independent of $\mathbf{n}_c(s)$. Hence, the conditional PDF of $Z(S)$ give $X(S)$ is given by

$$\begin{aligned} \Pr(Z|X) &= \prod_{s \in S} \Pr(\mathbf{z}(s)|\mathbf{x}(s)) \\ &= \prod_{s \in S} \frac{\exp \left[-\frac{1}{2} (\mathbf{z}(s) - W\mathbf{x}(s))^T \mathbf{C}_z^{-1} (\mathbf{z}(s) - W\mathbf{x}(s)) \right]}{(2\pi)^{b/2} |\mathbf{C}_z|^{1/2}} \end{aligned} \quad (1.7)$$

We formulate the image fusion problem as an M-ary hypothesis testing problem where each hypothesis corresponds to a different FSSRI. Furthermore, since we formulate our problem as an M-ary hypothesis testing problem, techniques developed to solve signal detection problems can be employed. We provide our methodology in the next section.

1.3 Optimum Image Fusion Rules

The maximum *a posteriori* (MAP) criterion [26] is used for solving the above problem in our work. This criterion is expressed as

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} [\Pr(X, \omega|Y, Z)] \right] \quad (1.8)$$

From Bayes' rule and assuming conditional independence of Y and Z given X and ω , Eq. (8) can be rewritten as

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[\frac{\Pr(Y|X, \omega) \Pr(Z|X) \Pr(X, \omega)}{\Pr(Y, Z)} \right] \right] \quad (1.9)$$

Since $\Pr(Y, Z)$ is independent of the choice of (X, ω) , it can be omitted and above equation reduces to

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} [\Pr(Y|X, \omega) \Pr(Z|X) \Pr(X, \omega)] \right] \quad (1.10)$$

By assuming further that X and ω are statistically independent, and substituting Eq. (1.1), Eq. (1.5), and Eq. (1.7) into Eq. (1.10), we have

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[\left(\prod_{s \in \mathcal{S}} \Pr(\mathbf{y}_m(s)|\mathbf{x}(s)) \Pr(\mathbf{z}(s)|\mathbf{x}(s)) \Pr(\mathbf{x}(s)) \right) \Pr(\omega) \right] \right] \quad (1.11)$$

Eq. (1.11) can be rewritten as

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[- \sum_s \{E_Y(\mathbf{y}_m, \mathbf{x}) + E_P(\mathbf{z}, \mathbf{x}) + E_X(\mathbf{x})\} - \ln|\Omega_\omega| \right] \right] \quad (1.12)$$

where

$$E_Y(\mathbf{y}_m, \mathbf{x}) = \frac{1}{2}(\mathbf{y}_m - \mathbf{x})^T C_C^{-1} (\mathbf{y}_m - \mathbf{x}) + \frac{B}{2} \ln 2\pi + \frac{1}{2} \ln |C_C| \quad (1.13)$$

$$E_Z(\mathbf{z}, \mathbf{x}) = \frac{1}{2}(\mathbf{z}(s) - W\mathbf{x}(s))^T \Sigma_Z^{-1} (\mathbf{z}(s) - W\mathbf{x}(s)) + \frac{b}{2} \ln 2\pi + \frac{1}{2} \ln |C_Z| \quad (1.14)$$

and

$$E_X(\mathbf{x}) = \frac{1}{2}(\mathbf{x}(s_i) - \boldsymbol{\mu}_X)^T C_X^{-1} (\mathbf{x}(s_i) - \boldsymbol{\mu}_X) + \frac{B}{2} \ln 2\pi + \frac{1}{2} \ln |C_X| \quad (1.15)$$

Again, we omit the term s for sake of abbreviation. Since the terms, $\frac{B}{2} \ln 2\pi$, $\frac{1}{2} \ln |C_C|$, $\frac{1}{2} \ln |C_Z|$, $\frac{B}{2} \ln 2\pi$, $\frac{b}{2} \ln 2\pi$, and $\frac{1}{2} \ln |C_X|$ are just constants, they can be ignored. Therefore, the optimum image fusion becomes

$$\begin{aligned} (\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[- \frac{1}{2} \sum_s \left\{ (\mathbf{y}_m - \mathbf{x})^T C_C^{-1} (\mathbf{y}_m - \mathbf{x}) \right. \right. \right. \\ \left. \left. + (\mathbf{z}(s) - W\mathbf{x}(s))^T \Sigma_Z^{-1} (\mathbf{z}(s) - W\mathbf{x}(s)) + (\mathbf{x} - \boldsymbol{\mu}_X)^T C_X^{-1} (\mathbf{x} - \boldsymbol{\mu}_X) \right\} \right. \\ \left. \left. - \ln|\Omega_\omega| \right] \right] \end{aligned} \quad (1.16)$$

The above equation can be modified to

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[- \frac{1}{2} \sum_s (\mathbf{x} - \mu_s(\mathbf{y}_m, \mathbf{z}, \boldsymbol{\mu}_X))^T C_s^{-1} (\mathbf{x} - \mu_s(\mathbf{y}_m, \mathbf{z}, \boldsymbol{\mu}_X)) - \ln|\Omega_\omega| \right] \right] \quad (1.17)$$

where

$$C_s = [C_X^{-1} + C_c^{-1} + WC_Z^{-1}W^T]^{-1}, \quad (1.18)$$

and

$$\mu_s(\mathbf{y}_m, \mathbf{z}, \mu_X) = C_s [C_X^{-1} \mu_X + C_c^{-1} \mathbf{y}_m + W^T C_z \mathbf{z}] \quad (1.19)$$

Next, we define the energy function $E(X, \omega)$ as the negative of the argument inside Eq. (19), i.e.,

$$E(X, \omega) = \frac{1}{2} \sum_s (\mathbf{x} - \mu_s(\mathbf{y}_m, \mathbf{z}, \mu_X))^T C_s^{-1} (\mathbf{x} - \mu_s(\mathbf{y}_m, \mathbf{z}, \mu_X)) + \ln |\Omega_\omega| \quad (1.20)$$

The optimum joint image fusion and registration problem becomes the energy minimization problem, i.e.,

$$(\hat{X}, \hat{\omega}) = \arg \left[\min_{(X, \omega)} E(X, \omega) \right] \quad (1.21)$$

For a given map parameter vector ω , the optimum FSSRI can be obtained by taking derivative of Eq. (20) with respect to $\mathbf{x}$, and the optimum FSSRV is given by

$$\mathbf{x} = \mu_s(\mathbf{y}_m, \mathbf{z}, \mu_X) \quad (1.22)$$

We note here that our result in Eq. (1.22) is similar to the work by Hardie (Hardie, 2004). However, in their work, images to be fused are assumed to be perfectly aligned. In order to obtain the map parameter, we need to find minimize Eq. (1.20) with respect to ω . However, the derivative of Eq. (1.21) with respect to the ω map parameter vector is very difficult to obtain analytically. Furthermore, the energy function $E(X, \omega)$ is an extremely non-convex function. Hence, a gradient based approach cannot be used since the resulting map parameter will very likely become stuck in one of many local optima. Instead, we employ the Metropolis algorithm [27] for determining the optimum map parameter vector.

The Metropolis algorithm is a stochastic search method where, at each iteration, a new value of ω_n is randomly proposed. If ω_n results in a better fit (lower energy function), it is accepted, and the Metropolis algorithm sets $\omega = \omega_n$. However, even if ω_n corresponds to a higher energy value, ω_n is still accepted with some probability. These random moves allow the Metropolis algorithm escape from local optima. By employing the Metropolis algorithm, the optimization algorithm can be written as

1. Set $(X, \omega) = (X_0, \omega_0)$, $h=1$, and $T = \frac{T_0}{\log(h+1)}$ where X_0 and ω_0 are the initial

FSSRI and map parameters, h is the iteration number, T_0 is the initial temperature.

2. Find a FSSRI by using Eq. (1.23), and Let X be the resulting FSSRI.
3. Use Eq. (1.21) to compute energy function and assign $E_{old} = E(X, \omega)$.

4. A new map parameter ω_n is drawn from Eq. (1.3) and the corresponding energy function $E_n(X, \omega_n)$ is computed using Eq. (1.22).
5. Assign $\omega = \omega_n$ with probability $P = \max\left(1, \exp\left[-\frac{E_n(X, \omega_n) - E(X, \omega)}{T}\right]\right)$.
6. Set $h = h + 1$ and $T = \frac{T_0}{\log(h+1)}$. Go to (2) if $h < h_{max}$.

We observe that as the number of iteration increases to infinity, the temperature T decreases to zero. This implies that, after a small number of iterations, the Metropolis algorithm is likely to accept almost any proposed map parameter. However, when the number of iterations becomes large, the Metropolis algorithm prefers to accept the map parameter that results in a better fit (lower energy function). The convergence of the Metropolis algorithm to the global optimum regarding the choice of the initial map parameter and FSSRI is guaranteed if T_0 is sufficiently large. However, this value is often too large to be practically implemented. Nevertheless, in our problem, the initial map parameters and FSSRI are not far from the global optimum since both multispectral and panchromatic sensors are mounted on the same satellite. Figure 1.1 summarizes the proposed optimization algorithm in this paper.

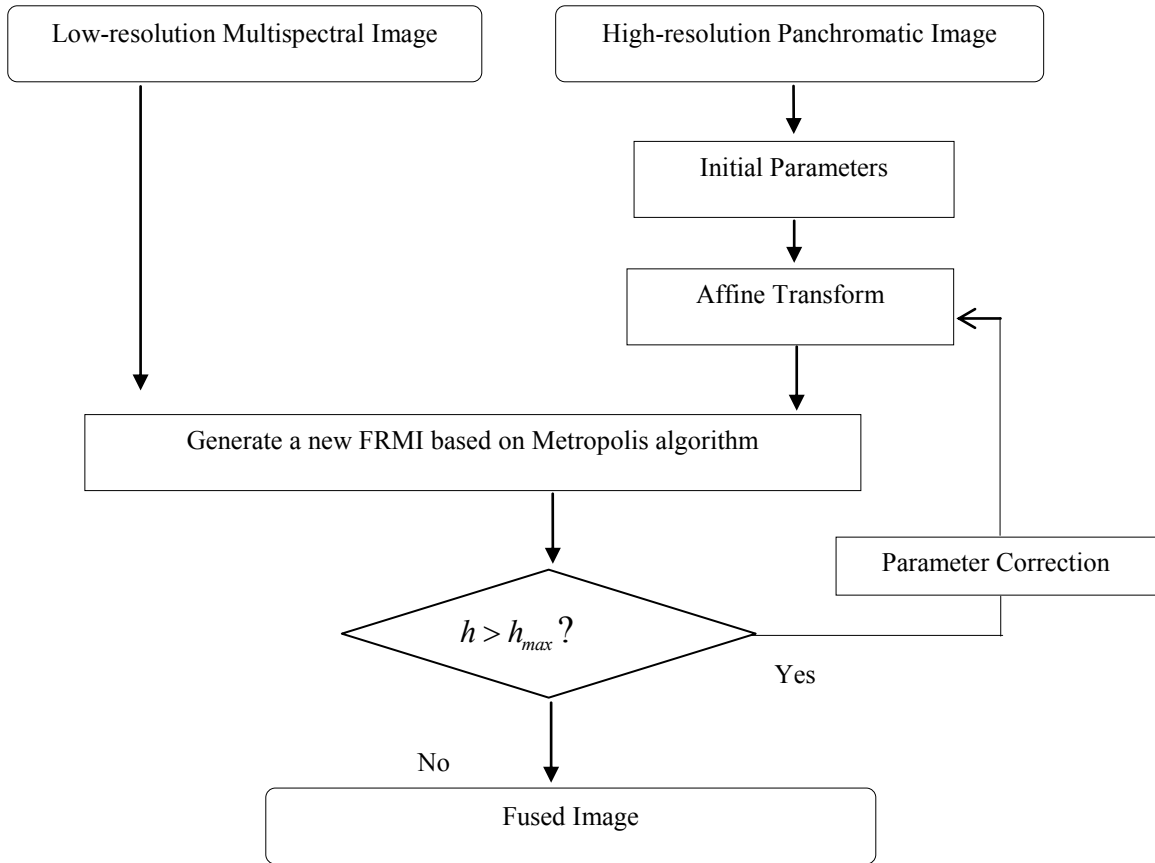


Figure 1.1: Flowchart of the proposed algorithm

1.4 Experimental Results

THEOS Data set

To examine the performance of the proposed algorithm, we used various image scenes from many parts of Thailand, namely, city (Figure 1.2), drought (Figure 1.3), agriculture (Figure 1.4), mountain (Figure 1.5) and seashore (Figure 1.6) areas. **Table 1.1** summarizes the detailed information for all the scenes used in this experiment. Multispectral and panchromatic images were acquired by THEOS, an Earth observation mission of Thailand, developed at EADS Astrium SAS, Toulouse, France. THEOS was launched from Dombrovskiy, Russia at 06:37 GMT on 1 October 2008, by a Dnepr rocket.



Figure 1.2: 750x750 THEOS multispectral image of city area

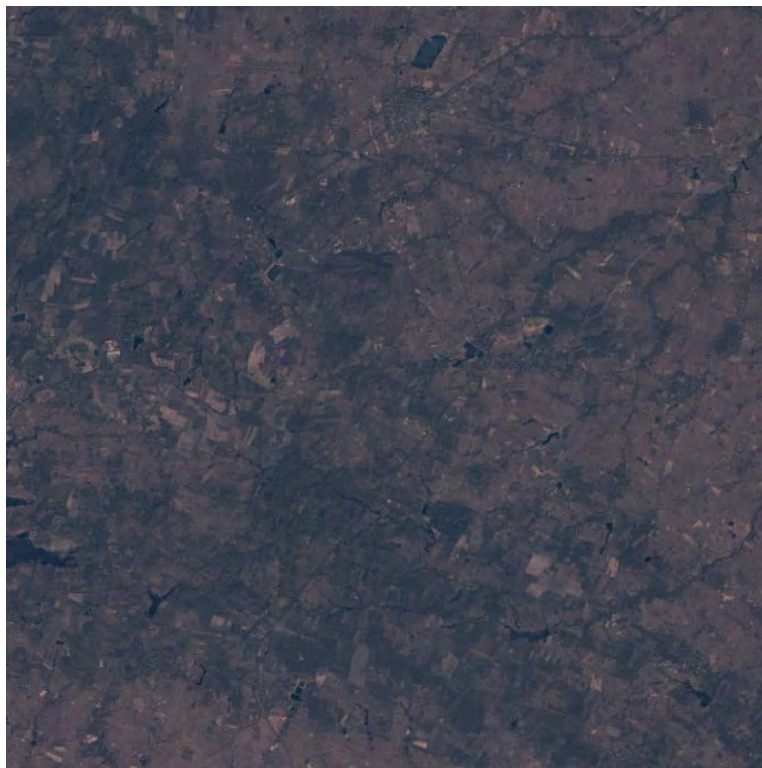


Figure 1.3: 750x750 THEOS multispectral image of drought area

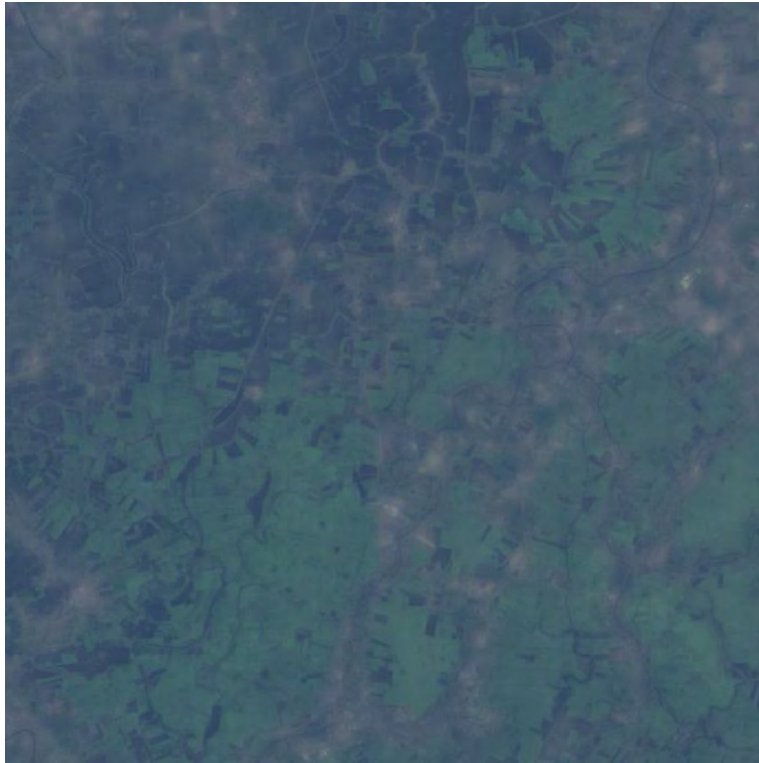


Figure 1.4: 750x750 THEOS multispectral image of agriculture area



Figure 1.5: 750x750 THEOS multispectral image of mountain area

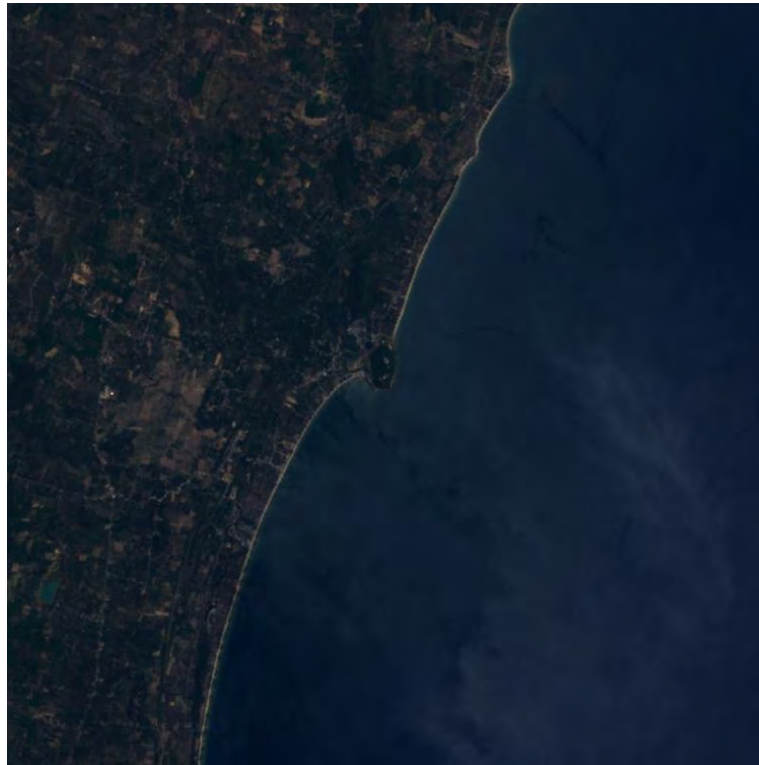


Figure 1.6: 750x750 THEOS reference multispectral image of seashore area

Table 1.1: Details of THEOS images in our experiment

Area	Longitude	Latitude	Date	Time
City	100.3845251902455N	13.82391143621505E	2009/12/13	10.32 AM
Drought	105.058671289676N	15.92080873025695E	2010/01/04	10.31 AM
Agriculture	100.5728597866535N	14.6219909729372E	2009/12/13	10.32 AM
Mountain	100.561439575278N	19.69589619804945E	2010/02/13	10.31 AM
Seashore	99.59914701590165N	11.3390399265943E	2010/03/06	10.33 AM

The THEOS imagery products include optical and near infrared spectroscopy as four multispectral bands, plus panchromatic images. The THEOS satellite orbit parameters are summarized in **Table 1.2**. Panchromatic products provide 2 meter resolution (at nadir) and 8 bits information depth. The output scene is a square scene of 22 km. x 22 km, and covers wavelengths ranging from 450 to 900 nm as shown in Figure 1.7. The multispectral products provide 15 meter resolution (at nadir) and 8 bits information depth. All four bands are delivered as one file. The output scene is a square scene of 90 km. x 90 km. Wavelength ranges are 450-520 nm for the blue band, 530-600 nm for green, 620-690 nm for red, and 770-

900 nm for near infrared as shown in Figure 1.7. **Table 1.3** summaries the information on multispectral and panchromatic images provided by the THEOS satellite.

Table 1.2: Orbital characteristics of THEOS

Satellite	THEOS
Orbit	Sun synchronous recurrent frozen orbit
Altitude	822 km
Inclination	97.95 degree
Repetition cycle	31 days
Descending node (local time)	10:30 AM
Orbital period	101.4 minutes
On-board capacity	16 Gbits

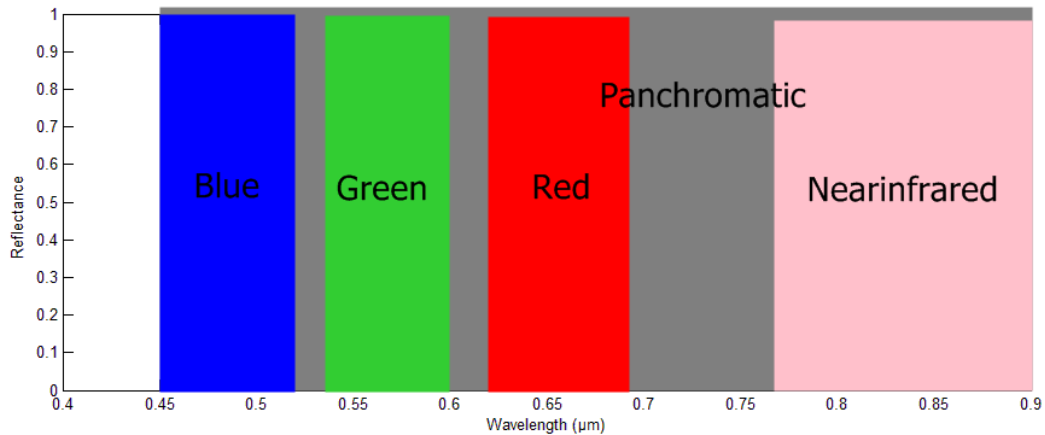


Figure 1.7: Ideal spectral response of THEOS

Table 1.3: Technical specifications of THEOS payload

	Panchromatic	Multispectral
Wavelength	450-900 nm	Blue: 450-520 nm
		Green: 530-600 nm
		Red: 620-690 nm
		Near Infrared: 770-900 nm
Resolution	2 m	15 m
Swath width	22 km	90 km
Pixels	12000	6000

Since our algorithm performs both image fusion and registration at the same time, the performance of our algorithm can be evaluated in terms of how far off the fused image is from the actual FRMI, that is, the estimation error between the actual parameters that map between the panchromatic and multispectral image pair and the estimated ones. If our algorithm performs perfect registration and fusion, the fused image will be exactly equal to FRMI and

the estimation error will be zero. In order to measure this key performance metric, we need to know both actual FSSRI and the map parameters. However, since we use observed THEOS multispectral and panchromatic images in this experiment, the actual FSSRI is unknown. As a result, we treat the observed multispectral image as the reference FSSRI, and perform the proposed image fusion algorithm on resampled multispectral and panchromatic image pairs. In this experiment, we resample multispectral and panchromatic images to 112.5 and 15 meters resolution, respectively. The resampled images have a resolution of 7.5 times lower than the original ones. After fusing a 112.5 meter multispectral image with a 15 meter panchromatic image, we obtain the fused multispectral image at 15 meter resolution which can be compared with the observed multispectral image.

Next, since our algorithm is designed to fuse unregistered image pairs, we would like to investigate the performance of our algorithm for different scenarios of initial registration errors. We would like to vary the relationship between the multispectral and panchromatic images in a pair by applying different values of displacements and rotations. However, since multispectral and panchromatic sensors are located at different locations on the THEOS satellite, their images are not perfectly aligned to start with. To be sure that we can precisely measure the amount of mis-registration we introduce, we manually register all multispectral and panchromatic image pairs first, using nine or more ground control points with the second-order polynomial transformation. The resulting root mean square errors were less than 0.0002 pixels for all images.

For the sake of brevity, we provide the details of our experiment only for the case of the city area image. The other cases followed the same steps. The multispectral (Figure 1.2) and the panchromatic (Figure 1.8) image pair were degraded by resampling to 112.5 (Figure 1.9) and 15 (Figure 1.10) meters resolution respectively. Then we fused the degraded images with our proposed algorithm to obtain a 15 meter-resolution multispectral image. We use the original 15-meter resolution multispectral image as the “ground truth” – that is, the (actually unobservable) FRMI. Hence, in our experiment, the resulting fused image is compared with the original multispectral image. If our algorithm performs perfect registration and fusion, the resulting mean square error (MSE) between the resulting FRMI and the original multispectral image will be zero.

In all of our experiments, the original multispectral and panchromatic registered images are of size 750×750 and 5625×5625 , respectively. Since the resolution ratio of the multispectral to the panchromatic is $15/2$ or 7.5 , we need 7.5 times as many panchromatic pixels as multispectral pixels to cover the same spatial extent.

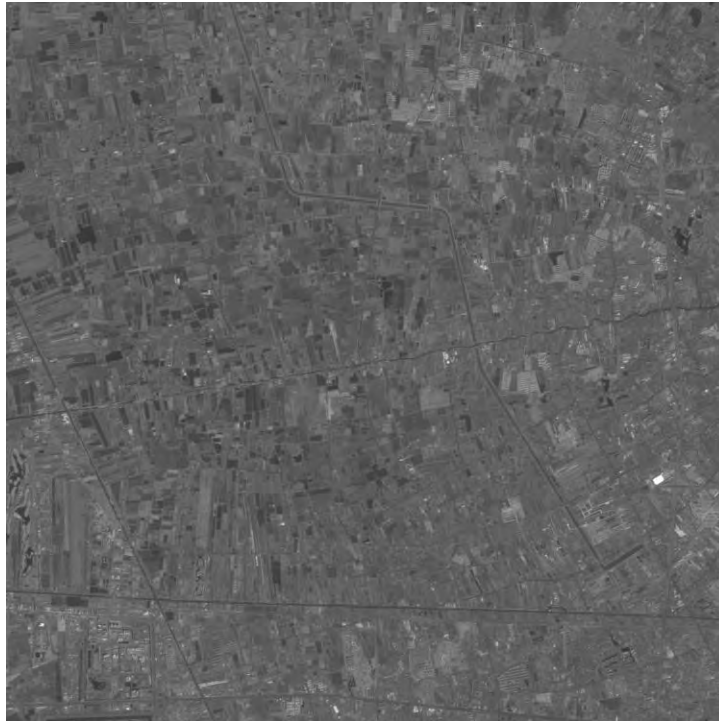


Figure 1.8: Panchromatic image of the city scene of size 5625×5625 pixels



Figure 1.9: Degraded multispectral image of city scene at 112.5×112.5 meters resolution of size 100×100 pixels

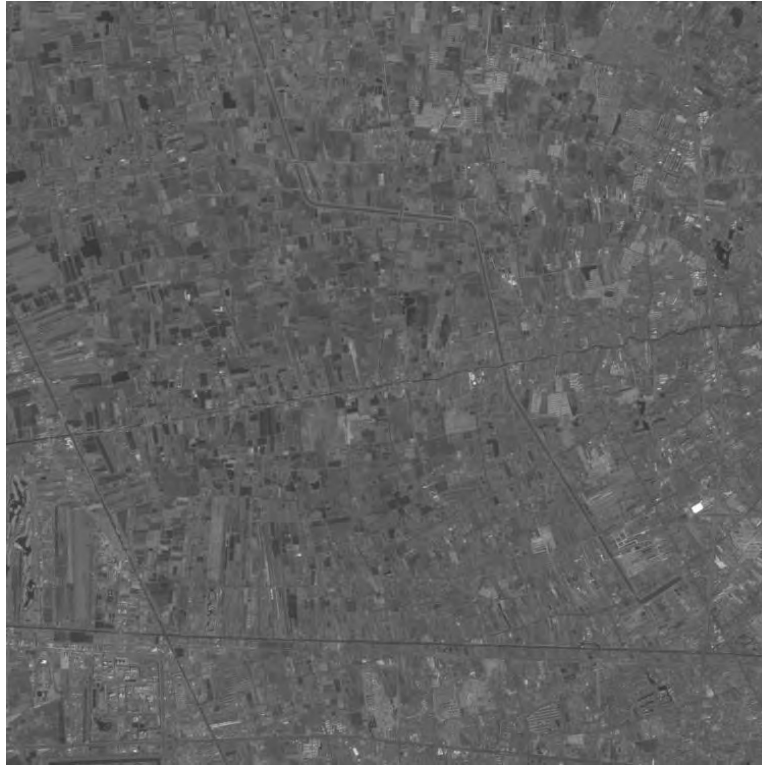


Figure 1.10: Degraded panchromatic image of the city scene at 15×15 meters resolution of size 750×750 pixels

Mis-registration errors were deliberately introduced into the degraded image pairs. Here, the affine transformation was used and there were four parameters, namely scale factor η , rotation angle θ in degree, displacement in column direction d_x in pixel, and displacement in row direction d_y in pixel. **Table 1.4** summarizes all mis-registration errors introduced in this experiment. Note here that mis-registration in the scale dimension was not investigated in this paper since images are taken from the same satellite at the same time. Figure 1.11 shows a panchromatic image distorted by 3 degrees of rotation, 3 pixels of displacement in the column and 3 pixels of displacement in the row.

Table 1.4: The mis-registration errors examined in this experiment

	Scale	Ratio angle	Displacement in column direction	Displacement in row direction
Case I	1	1	1	1
Case II	1	-2	-2	-2
Case III	1	3	3	3
Case IV	1	-4	-4	-4
Case V	1	5	5	5

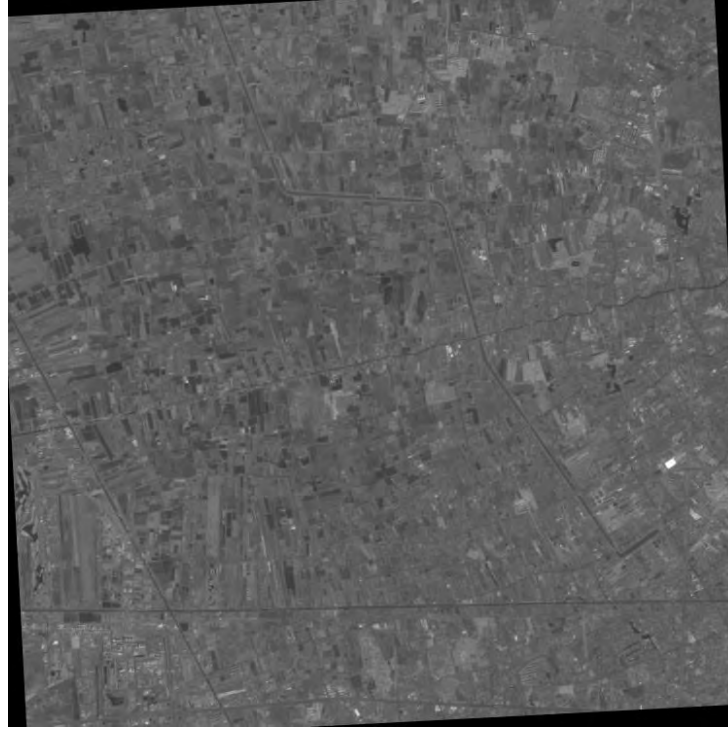


Figure 1.11: Degraded panchromatic image with mis-registration error of 3 degrees of rotation, 3 pixels of displacement in column and 3 pixels in row

To find the appropriate benchmark for our algorithm, we first fuse the registered image pair by using Eq. (1.23) and compare it to the FRMI from manual registration as shown in Figure 1.12. We note here that our algorithm and the work by Hardie are equivalent if no registration error is considered. In other words, the benchmark scenario is when we apply Hardie algorithm directly to multispectral and panchromatic image pairs. The corresponding MSE compared to the original multispectral image was equal to 7.2465.

To generate the initial FRMI for our algorithm for the non-benchmark case, we fused a mis-registered image pair using Eq. (1.23) and used it as the initial FSSRI. An example of the initial FSSRI for the case of 3 degrees of rotation, 3 pixels of displacement in column and 3 pixels in row is illustrated in Figure 1.13. It is obvious that the initial FSSRI appears to be blurred and does not contain any sharp edges. Next, we submit the initial FSSRI to our algorithm and iteratively refine it by applying different sets of transformation parameters as selected by the algorithm. Here, we set $T_0 = 0.1$, $h_{max} = 500$ iterations. The covariance matrices of the noise introduced by resampling to create the CRMI from different spectral bands are assumed to be statistically independent with variance of 25, i.e., $C_C = 25I$ where I is the identity matrix. Figure 1.14 displays the resulting FSSRI for the case of 3 degrees of rotation, 3 pixels of displacement in column and 3 pixels in row. By visual inspection, we

can observe that the resulting FRMI looks very similar to FRMI when perfect registration is obtained.

The results summarized in **Table 1.5** clearly show that our algorithm can move the map parameters closer to the correct values. We note here that, for all cases, our algorithm achieved lower RMSE values than those of the manual registration. These results imply that the manual registration that we performed may not have been perfect after all. Some tiny registration errors may still be present in the registered image pair.



Figure 1.12: THEOS fused multispectral image with city area and 750x750 pixels



Figure 1.13: THEOS initial fused multispectral image obtained by applying Hardie algorithm to the unregistered image pair with 3 pixels displacement and 3 degrees rotation with city scene



Figure 1.14: The resulting fused image for city area

Table 1.5: Comparison of fusion performance for images of city area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	7.2465	1	1	1	10.2043	0.4315	0.4195	0.1826	6.9368
0	0	0	7.2465	-2	-2	-2	9.7011	0.4426	0.4225	-0.0597	6.3584
0	0	0	7.2465	3	3	3	9.2295	1.5293	0.3358	-0.0037	6.9519
0	0	0	7.2465	-4	-4	-4	9.0814	1.2999	0.6848	-0.1954	7.4333
0	0	0	7.2465	5	5	5	8.8802	1.4716	0.6873	-0.0114	7.1084

Next, we investigated the other scenes. The resulting MSE and map parameters are summarized in Tables 1.6-1.9, for drought, agriculture, mountain, and seashore scenes. It is clear that the results from all cases are similar. Our algorithm can successfully move the map parameters closer to correct values regardless of the size of mis-registrations. Furthermore, in most cases, the resulting FSSRIs from the proposed algorithm yield lower RMSE values compared to the perfect registration case. This suggests that our algorithm finds errors remaining after manual registration.

Table 1.6: Comparison of fusion performance for images of drought area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	4.4493	1	1	1	6.7294	0.4178	1.6116	-0.0425	4.4184
0	0	0	4.4493	-2	-2	-2	6.3793	0.4561	0.5670	-0.0832	3.5014
0	0	0	4.4493	3	3	3	5.9539	1.4802	0.6065	0.0960	4.0669
0	0	0	4.4493	-4	-4	-4	5.7613	1.5929	0.4242	-0.1787	4.3226
0	0	0	4.4493	5	5	5	5.5656	0.6040	0.4540	0.0551	3.7374

Table 1.7: Comparison of fusion performance for images of agriculture area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	8.0216	1	1	1	11.3885	0.5491	1.3608	0.0223	5.6583
0	0	0	8.0216	-2	-2	-2	11.3888	0.4865	0.5077	0.0182	5.7027
0	0	0	8.0216	3	3	3	10.9708	0.3072	0.5173	0.0277	5.8182
0	0	0	8.0216	-4	-4	-4	10.6829	0.5111	0.7775	-0.0304	5.6377
0	0	0	8.0216	5	5	5	10.4727	0.4003	0.6520	0.0465	5.9665

Table 1.8: Comparison of fusion performance for images of mountain area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	3.6175	1	1	1	4.4446	0.3589	0.6566	-0.2068	3.2774
0	0	0	3.6175	-2	-2	-2	4.5898	0.4614	0.4915	-0.2390	3.3056
0	0	0	3.6175	3	3	3	4.6388	0.6505	-0.5470	-0.0401	3.2220
0	0	0	3.6175	-4	-4	-4	4.5726	0.8501	-0.4960	0.0323	3.1460
0	0	0	3.6175	5	5	5	4.6244	1.5957	0.4378	-0.0534	3.3172

Table 1.9: Comparison of fusion performance for images of sea shore area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	5.1446	1	1	1	7.4105	0.6091	0.6279	0.0773	3.9202
0	0	0	5.1446	-2	-2	-2	7.2712	0.4977	0.5948	0.0856	3.9134
0	0	0	5.1446	3	3	3	7.0384	0.3827	1.3204	0.0796	4.1890
0	0	0	5.1446	-4	-4	-4	6.7605	0.6050	0.6162	0.0149	3.8124
0	0	0	5.1446	5	5	5	6.6767	0.5031	0.6723	-0.0195	4.0117

Hyperspectral Data set

The small multi-mission satellite (SMMS) (HJ1-A) acquired an image data for our experiment. There are HI of size 300×300 pixels (Figure 1.15(a)) with 115 bands and 100 meter resolution, and MI of size 900×900 pixels (Figure 1.15(b)) with 4 bands and 30 meters resolution. These images cover the area of Suwannaphoom Airport in Thailand.



Figure 1.15: Hyperspectral and Multispectral image

To evaluate the performance of our proposed algorithm, the original hyperspectral image and multispectral image is resampled to simulate the CSFSRI and FSCSRI at 300 meter and 100 meter resolution resolutions, respectively. When these images are fused together, the fused hyperspectral image at 100 meter resolution is obtained. We can then compare to the original hyperspectral image to evaluate the quality of the fused image. First, the resampled hyperspectral image and multispectral image are fused together to evaluate the performance of the fusion algorithm when where two fusing images are perfectly registered. Here, the spatially enhanced hyperspectral image with perfect registration is shown in Figure 1.16. For more quantitative performance evaluation, the root mean square error (RMSE) between the original hyperspectral image and spatially enhanced hyperspectral image with perfect registration is computed and is equal to 3.3976.

To simulate the effect of the registration error, the resampled hyperspectral image is shifted 2 and 3 pixels in x and y directions, respectively. The resulting shifted and resampled HI (SRHI) is the simulated CSFSRI with some registration errors in this work. Then, the SRHI is fused with resampled FSSRI without registration error correction, and the resulting spatial enhancement HI without misregistration correction is shown in Figure 1.17. Since the misregistration is not corrected, the effect of misregistration is clearly visible, and the RME, comparing to the original HI, is equal to 6.2308 (about 82% more error).

Next, the SRHI together with the simulated CSFSRI is submitted to our proposed algorithm to obtain the spatially enhanced and spatially corrected FSSRI (Figure 1.18). With visual inspection, the effect of misregistration is no longer visible and the RME comparing to the original HI becomes 2.617. The RMSE is 60% better than the spatially enhanced HI without consider the registration error, and is 23% better than the spatially enhanced HI in the idea case. The performance improvement for the latter case may be the result from the registration error in the production of SMMS data in the first place. Our proposed algorithm estimates the registration errors to be 1.8 and 3.5 pixels in x - and y -directions. The differences of 0.2 and 0.5 in x - and y -direction from the experimental setup may be the original registration error in the production of the SMMS data discussed earlier.



Figure 1.16: The fused hyperspectral image without registraion error



Figure 1.17: The fused hyperspectral image with registration error



Figure 1.18: The fused hyperspectral image with registration error using our algorithm

2. *Land Cover Mapping algorithm from Mis-Registered Image Datasets*

2.1 Introduction

For multi-sensor image classification, the registration process is a crucial initial step. Registration aligns multiple satellite images into a common coordinate system. Only when all of the input images are perfectly registered can a classification algorithm be applied. Otherwise mis-registration will produce classification errors. In practice, perfect registration

may not always be achievable since there are some unknown variations on satellite platforms and flight paths when capturing images. As a result, the overall classification accuracy is likely to suffer from mis-registration effects.

Mahapatra and Sun [28] proposed an idea to incorporate the reduction of image registration error into an image classification tool. They attempted to integrate the segmentation information into an elastic image registration by using a Markov random field model. In their work, the configuration of a pixel contains both displacement of a pixel and a segmentation label. The multi-resolution graph-cut approach was employed to achieve sub-pixel registration accuracy. Although their results produced remarkable performance for non-rigid body image registration, this algorithm cannot be applied for rigid body image registration problems such as in the remote sensing since their model does not allow one set of the registration parameters to govern the remapping process of entire image. Furthermore, since they only consider the segmentation problem, their algorithm does not cover the multi-class scenarios that are often considered in the land cover mapping of remotely sensed images.

Another work by Chen *et al.* in [25] investigated the problem of joint image fusion and registration. In their paper, the observed images were remapped versions of the original images with possibilities of polarity reverse and/or DC offset. Chen *et al.* used an expectation maximization algorithm to solve the estimation problems of registration parameters and the true scene simultaneously. Different pairs of multi-sensor images were tested against the proposed joint process. Under the assumption that registration performance affects the quality of fusion result, the authors reported that better fusion performance can be achieved due to reduced registration errors. However, their work did not cover the problem of image classification in the presence of image registration errors.

In this paper, we employ an approach similar to [25] to incorporate correction of mis-registration effects into the land cover mapping process. To do this, we assume that remotely sensed images are derived from a common unobservable land cover map (LCM), and then distorted, with unknown remapping parameters, into the observed remote sensing images. (Note that if these map parameters are known, the observed remote sensing images can be directly aligned with the land cover map.) Next, we assume that a land cover class of interest is more likely to occupy several connected patches than a number of isolated pixels. As a result, the Markov random field (MRF) is employed as the model of the LCM. MRF models have been used in various fields ranging from statistical physics [29] and [30] to remote sensing. The original work by Geman and Geman [31] on MRF-based statistical methodology in 1984 has inspired a continuous stream of remote sensing researchers to employ the MRF model for a variety of image analysis tasks (e.g., [32], [33], [34], [35], [36], [37], [38]). Solberg *et al.* [32] developed MRF-based algorithms for image classification and change detection using multi-source data. A significant increase in classification and change detection accuracy was obtained using an MRF based classification algorithm compared to other approaches. Kasetkasem and Varshney [33] and Bruzzone and Prieto [34] also applied MRF models for an image change detection problem. Similarly, Xie *et al.* [35] applied the MRF

model to the recurring problem of speckle reduction in synthetic aperture radar (SAR) images. These promising results on image analysis problems have encouraged us to employ the concept of MRF models to the problem of generating an LCM.

Based on our image model, the registration and classification process can be performed in the following fashion. First, we estimate the unknown map transformation parameters based on the maximum likelihood (ML) criteria, and, then use these parameters to computer posterior probability for different arrangements of the land cover maps, where the MAP classifier selects the most likely LCM. However, in order to find the map parameters, the conditional probability of observed images given the map parameters is needed. This conditional probability can only be obtained by summing the joint probabilities of observed images and LCM associated with the map parameters, over all possible LCMs. This is impossible to obtain in most practical scenarios. As a result, the expectation-maximization (EM) algorithm [39] is also employed here. The EM algorithm iteratively searches for the most likely map parameters. The resulting parameters converge to one of the local optimum points of the likelihood function.

For a given iteration of the EM algorithm, our method computes the expected value of the logarithm of the probability of the observed images and land cover map given the map parameters, based on the *a posterior* probability of the LCM given observed remote sensing images and the current estimated map parameters. Then, new map parameters are obtained by maximizing the expected values. It has been shown in literature [39] that the new map parameters always correspond to a higher value of the likelihood function. Since each iteration of the EM algorithm calculates *a posterior* probability given the current estimated map parameter, an optimum LCM under MAP criteria can be easily obtained by choosing the LCM that maximizes *a posterior* probability. In other words, an optimum LCM for the most recent estimate of the map parameters under the maximum *a posteriori* (MAP) criterion is obtained on every iteration of the EM algorithm.

2.2 Problem Statement

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 $\Lambda \in \{0, 1, \dots, L-1\}$ be the class labels. Therefore, we can express the LCM as $X(\mathcal{S}) \in \Lambda^{\mathcal{S}}$. The label of LCM at pixel s is denoted by x_s which can also be called the configuration of $X(\mathcal{S})$ at the site s . Since land cover classes are more likely to occur in connected patches in the LCM than isolated pixels, the LCM is assumed to satisfy the MRF properties with Gibbs potential $V_C(X)$. Hence, the marginal probability density function (PDF) of a LCM can be written as

$$\Pr(X) = \frac{1}{Z_X} \exp \left(- \sum_{C \subset \mathcal{S}} V_C(X) \right) \quad (2.1)$$

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 [31]. Cliques are singleton or groups of pixels such that any two pixels are

mutually neighbors. Figure 2.1 shows all possible clique types for 4- and 8-neighborhood systems. The value of the Gibbs potential function depends on the configurations of the entire LCM and the clique. Usually, low values of the potential function correspond to similar configurations whereas high values correspond to dissimilar configurations of a clique. For instance, the Ising model [29] and [30], given by,

$$V_{\{s,r\}}(X) = \begin{cases} -\beta; & \text{if } x_s = x_r \text{ and } r \in NG_s \\ \beta; & \text{if } x_s \neq x_r \text{ and } r \in NG_s \\ 0; & \text{if } r \notin NG_s \end{cases} \quad (2.2)$$

for any two sites r and s , has been used extensively by statistical physicists to explain why neighboring particles are more likely to rotate in the same direction (i.e., either clockwise or counterclockwise). Here the notation NG_s is a set of neighboring pixels of s . We can extend the above model to our problem by letting x_s and x_r be the class labels of pixels s and r in $\mathcal{S}$, respectively. With this modification, the Ising model can be applied to describe the LCM because land cover class distributions are similar to the phenomenon described above (i.e., classes occupying neighboring pixels are likely to be the same).

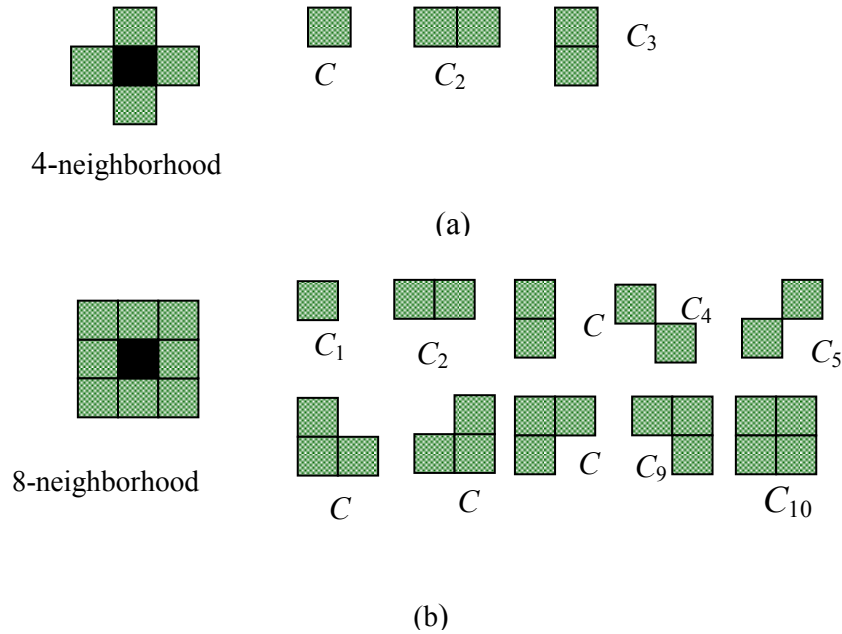


Figure 2.1: clique types for (a) 4-neighborhood; (b) 8-neighborhood

Furthermore, we assume that there are N remotely sensed images of the same scene acquired from different sensors and/or at different times. Here, $Y_n(\mathcal{T}_n) \in \mathfrak{R}^{\mathcal{T}_n \times B_n}; n = 1, 2, \dots, N$ denotes the n -th remotely sensed image where B_n denotes the number of spectral bands, and $\mathcal{T}_n$ is a map coordinate system to which the n -th remote sensing image is registered. Since all remotely sensed images and the LCM are from the same scene, the

relationship between $\mathcal{S}$ and $\mathcal{T}_n$ can be determined. Let us denote a coordinate of a pixel s in the LCM as (x, y) where x and y are the column and row of x_s . Similarly, we can write $t_n = (u_n, v_n) \in \mathcal{T}_n$ where u_n and v_n are the column and row of the pixel t_n in Y_n . If the affine transformation is used, the relationship between s and t_n can be written as

$$\begin{bmatrix} u_n \\ v_n \end{bmatrix} = \begin{bmatrix} m_{1,n} & m_{2,n} \\ m_{3,n} & m_{4,n} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_{5,n} \\ m_{6,n} \end{bmatrix} \quad (2.3)$$

where $m_{1,n}$ and $m_{4,n}$ are scale parameters, $m_{2,n}$ and $m_{3,n}$ are skew parameters, and $m_{5,n}$ and $m_{6,n}$ are displacement parameters in column and row-direction, respectively. We refer to $M_n = [m_{1,n}, m_{2,n}, m_{3,n}, m_{4,n}, m_{5,n}, m_{6,n}]$ as the map parameter vector between coordinate systems $\mathcal{S}$ and $\mathcal{T}_n$.

When all the map parameter vectors are given, one can remap all remotely sensed images to perfectly align with the LCM. Let us denote

$$Z_n(\mathcal{S}) = f(Y_n(\mathcal{T}_n), M_n) \quad (2.4)$$

as the remapped and resampled version of the n -th remote sensing image. Here, we assume further that the remapped and resampled images are statistically independent for a given LCM, i.e.,

$$\Pr(\mathbf{Y}|\mathcal{X}(\mathcal{S}), \mathbf{M}) = \Pr(Z_1(\mathcal{S}), \dots, Z_N(\mathcal{S})|\mathcal{X}(\mathcal{S})) = \prod_{n=1}^N \Pr(Z_n(\mathcal{S})|\mathcal{X}(\mathcal{S})). \quad (2.5)$$

where $\mathbf{M} = \{M_1, \dots, M_N\}$ and $Y = \{Y_1(\mathcal{T}_1), \dots, Y_N(\mathcal{T}_N)\}$ be collections of the map parameters and the observed multispectral images. Moreover, the intensity vectors from different pixels in $Z_n(\mathcal{S})$ are also assumed to be statistically independent when the LCM is given. Hence, the joint conditional PDF can be written as

$$\Pr(\mathbf{Y}|\mathcal{X}(\mathcal{S}), \mathbf{M}) = \prod_{n=1}^N \prod_{s \in \mathcal{S}} \Pr(\mathbf{z}_{n,s} | x_s) \quad (2.6)$$

where $\mathbf{z}_{n,s} \in \mathbb{R}^{B_n}$ denotes the intensity vector of the remapped image $Z_n(\mathcal{S})$ at a pixel s . We acknowledge that the assumption given in Eq.(2.6) may not always be true for all cases since some land cover classes have textural structure. One can incorporate texture information into our image model appropriately, which may further result in an increase in accuracy. This will, however, result in very complex problems which may not be desirable in practice.

If we assume further that the intensity vector at a pixel s of the remapped image $Z_n(\mathcal{S})$ given the class label x_s is a multivariate normal random vector with mean vector $\boldsymbol{\mu}_{x_s,n}$ and covariance matrix $\Sigma_{x_s,n}$, Eq. (2.6) can be rewritten as

$$\begin{aligned} \Pr(\mathbf{Y}|\mathcal{X}(\mathcal{S}), \mathbf{M}) &= \prod_{n=1}^N \frac{1}{(2\pi)^{\frac{B_n}{2}} |\Sigma_{x_s,n}|^{1/2}} \exp \left[- \sum_{s \in \mathcal{S}} \frac{1}{2} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \Sigma_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) \right] \end{aligned} \quad (2.7)$$

where T denotes the matrix transpose operation.

By using chain rule, the posterior probability of the LCM given the observed multispectral images and the map parameters can be written as

$$\Pr(X|Y, \mathbf{M}) = \frac{\Pr(Y|X, \mathbf{M}) \Pr(X)}{\Pr(Y, \mathbf{M})}. \quad (2.8)$$

Since $\Pr(Y, \mathbf{M})$ is independent of the choice of X , it can be treated as a constant. Hence, we have

$$\Pr(X|Y, \mathbf{M}) = C \Pr(Y|X, \mathbf{M}) \Pr(X). \quad (2.9)$$

By substituting Eq. (1), and Eq. (7) into Eq. (9), we obtain

$$\Pr(X|Y, \mathbf{M}) = \frac{1}{Z'} e^{-E(X|Y, \mathbf{M})}. \quad (2.10)$$

where $Z' = \sum_{X \in \mathcal{S}} e^{-E(X|Y, \mathbf{M})}$ is a normalizing constant and independent of the choice of X , and

$$\begin{aligned} E(X|Y, \mathbf{M}) = & \frac{1}{2} \sum_{n=1}^N \left(\sum_{s \in \mathcal{S}} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \boldsymbol{\Sigma}_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\boldsymbol{\Sigma}_{x_s,n}| \right) \\ & + \sum_{C \subset \mathcal{S}} V_C(X) \end{aligned} \quad (2.11)$$

is called a conditional Gibbs energy function. Since, in this paper, we consider cliques comprised of pixel pairs only, hence, the conditional Gibbs energy function can be written as

$$\begin{aligned} E(X|Y, \mathbf{M}) = & \frac{1}{2} \sum_{n=1}^N \left(\sum_{s \in \mathcal{S}} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \boldsymbol{\Sigma}_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\boldsymbol{\Sigma}_{x_s,n}| \right) \\ & + \frac{1}{2} \sum_{s \in \mathcal{S}} \sum_{r \in NG_s} V_{\{s,r\}}(x_s, x_r) \end{aligned} \quad (2.12)$$

where NG_s denotes the set of neighboring pixels of s . The normalizing constant Z' cannot be computed in most practical scenarios due to the large number possible configurations (e.g., there are more than 2^{4096} possible configurations for binary LCM of size 64×64 .) As a result, we propose the use the mean field theorem [40] - [41] to remove the interaction between neighboring pixels defined in $V_C(X)$. The mean field theorem approximates the conditional Gibbs energy function as

$$E(X|Y, \mathbf{M}) \approx \frac{1}{2} \sum_{s \in \mathcal{S}} h_s(x_s|Y, \mathbf{M}) \quad (2.13)$$

where

$$\begin{aligned} h_s(x_s|Y, \mathbf{M}) = & \sum_{n=1}^N \left\{ (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \boldsymbol{\Sigma}_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\boldsymbol{\Sigma}_{x_s,n}| \right\} \\ & + \sum_{r \in NG_s} E_{x_r}[V_{\{s,r\}}(x_s, x_r)]. \end{aligned} \quad (2.14)$$

Here, $E_{x_r}[V_{\{s,r\}}(X)]$ is the expected value of the potential function with respect to the configuration of x_r . The expected value $E_{x_r}[V_{\{s,r\}}(x_s, x_r)]$ does not depends on x_r , and is equal to

$$E_{x_r}[V_{\{s,r\}}(X)] = \sum_{x_r \in \mathcal{X}} V_{\{s,r\}}(x_s, x_r) p^{MF}(x_r | \mathbf{Y}, \mathbf{M}) \quad (2.15)$$

where $p^{MF}(x_r | \mathbf{Y}, \mathbf{M}) = \frac{1}{Z_r} e^{-\frac{1}{2} h_r(x_r | \mathbf{Y}, \mathbf{M})}$. Note here that $Z_r = \sum_{x_r \in \mathcal{X}} e^{-\frac{1}{2} h_r(x_r | \mathbf{Y}, \mathbf{M})}$ is the normalizing constant for a pixel r . By using the approximation given in Eq. (2.13), the posterior probability can be written as

$$\Pr(X | \mathbf{Y}, \mathbf{M}) \approx \Pr^{MF}(X | \mathbf{Y}, \mathbf{M}) = \prod_{s \in \mathcal{S}} p^{MF}(x_s | \mathbf{Y}, \mathbf{M}) \quad (2.16)$$

The approximation in Eq. (2.16) is closest to $\Pr(X | \mathbf{Y}, \mathbf{M})$ from all approximations of the form $\prod_{s \in \mathcal{S}} p(x_s | \mathbf{Y}, \mathbf{M})$ when the Kullback-Leibler (KL) divergence [42]- [43] is used as a distance measure.

2.3 Optimum Image Registration and Land Cover Mapping Criteria

The standard approaches to multi-temporal and/or multi-modal image classification involve two steps. First images from different sources and/or times are registered to produce a set of images in a common coordinate system. Then, a land cover map is derived from this set of registered images. In this work, even though we propose an algorithm to simultaneously register and classify images, we still treat image registration and classification as two separate problems to follow standard approaches. As a result, we propose different optimization criteria for image registration and land cover mapping. However, we will show in Section IV that both image registration and land cover mapping can be combined into one algorithm so that the registration and land cover mapping can be performed simultaneously.

2.3.1 Optimum Image Registration

The maximum likelihood estimate (MLE) can be employed as the optimum map parameter estimator since the MLE is known to a consistent estimator [44]. The goal of the MLE is to determine the map parameters that maximize the joint probability density function (PDF) of all the observed images given the map parameters, i.e.,

$$(M_1, \dots, M_N)^{opt} = \arg \max_{M_1, \dots, M_N} \Pr(Y_1, \dots, Y_N | M_1, \dots, M_N) \quad (2.17)$$

In order to solve Eq. (2.17), the conditional PDF $\Pr(Y_1, \dots, Y_N | M_1, \dots, M_N)$ must be calculated and it is equal to

$$\begin{aligned} \Pr(Y_1, \dots, Y_N | M_1, \dots, M_N) &= \sum_{X \in \Lambda^{\mathcal{S}}} \Pr(Y_1(\mathcal{T}_1), \dots, Y_N(\mathcal{T}_N), X(\mathcal{S}) | M_1, \dots, M_N) \\ &= \prod_{n=1}^N \sum_{X \in \Lambda^{\mathcal{S}}} \Pr(Z_n | X(\mathcal{S})) \Pr(X(\mathcal{S})) \end{aligned} \quad (2.18)$$

Note here again that Z_n is the remapped and resampled version of Y_n . Since Eq. (2.18) is written as a multiplication of $\sum_{X \in \Lambda^S} \Pr(Z_n | X(S)) \Pr(X(S))$, the solution of Eq. (2.17) can be individually obtain, i.e.,

$$M_n^{opt} = \arg \max_{M_n} \sum_{X \in \Lambda^S} \Pr(Z_n | X(S)) \Pr(X(S)), \quad (2.19)$$

for $n = 1, \dots, N$. Since S is also unknown, there are many possible sets of M_n that maximize Eq. (2.19). For instance, if $M_n = [1, 0, 0, 1, 0, 0]$ is the solution of Eq. (2.19) for $S = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, we have that $M'_1 = [1, 0, 0, 1, 1, 0]$ is also the solution of Eq. (19) for $S' = \{(0, -1), (0, 0), (1, -1), (1, 0)\}$. As a result, it is imperative to limit the search space and number of possible solutions. Furthermore, in most practical situations, we may wish to produce the LCM registered to one of the input remote sensing images. Without loss of generality, we assume that the LCM is registered to Y_1 , i.e., we have $M_1 = M_1^* = [1, 0, 0, 1, 0, 0]$.

Next, let us consider a small LCM of size 100×100 pixels. In this case, there are $2^{10,000} \approx 2 \times 10^{3,010}$ possible binary LCMs. Therefore, the direct calculation of Eq. (2.19) is an impossible task, and hence, the solution of the MLE cannot be obtained in reasonable time. As a result, the expectation-maximization (EM) algorithm [39] is employed instead. The EM algorithm is an iterative parameter estimator which produces a new estimate for every iteration. It has been shown in [39] that this new estimate always results in higher or at least the same value of the likelihood function. In other words, if we let $\mathbf{M}^t = \{M_1^t, M_2^t, \dots, M_N^t\}$ be the collection of all estimated parameters at the t -th iteration from the EM algorithm, we will have $\Pr(Y_1, \dots, Y_N | \mathbf{M}^t) \geq \Pr(Y_1, \dots, Y_N | \mathbf{M}^{t-1})$ where $\mathbf{M}^{t-1}$ is the collection of estimated parameters at $(t - 1)$ -th iteration. Here, and throughout the rest of the paper, we omit S and T_n for the sake of abbreviation. In Section 4, we will discuss the details of the EM algorithm employed in this work and how it can be combined with the land cover mapping process. However, before going into the detail of the proposed algorithm, let us state the optimization criterion for the land cover mapping considered in this paper.

2.3.2 Optimum Land Cover Map

The classifier based on the maximum *a posteriori* (MAP) criteria selects the most likely LCM given the observed data and the map parameters since the resulting probability of error is minimum among all other classifiers [44]- [45]. The optimum solution under the MAP criterion is expressed as

$$X^{opt} = \arg \max_X [\Pr(X | \mathbf{Y}, \mathbf{M})]. \quad (2.20)$$

In general, $\Pr(X | \mathbf{Y}, \mathbf{M})$ is non-concave function and, therefore, conventional gradient-based optimization algorithms are not applicable for the solution of Eq. (2.20). Furthermore, the number of possible solutions is also very large. A direct search for the solution of Eq. (2.20) is too expensive to be practically implemented. As a result, we propose the use the mean field

theorem [40]- [41] to remove the interaction between neighboring pixels defined in $V_c(X)$. Hence, by substituting Eq. (2.16) into Eq. (2.20), the optimization problem becomes

$$X^{opt} = \arg \max_X \left[\prod_{s \in \mathcal{S}} p^{MF}(x_s | \mathbf{Y}, \mathbf{M}) \right]. \quad (2.21)$$

Since the optimizing function in Eq. (2.21) is written in the form of the multiplication of functions of an individual pixels, and $p^{MF}(x_s | \mathbf{Y}, \mathbf{M})$ is a non-negative function, the optimum solution can be solved from an individual function, i.e., for $s \in \mathcal{S}$,

$$x_s^{opt} = \arg \max_X [p^{MF}(x_s | \mathbf{Y}, \mathbf{M})]. \quad (2.22)$$

which is equivalent to

$$x_s^{opt} = \arg \min_X [h_s(x_s | \mathbf{Y}, \mathbf{M})]. \quad (2.23)$$

2.4 Joint Image Registration and Land Cover Mapping Algorithm

Since the EM algorithm is employed in this literature as the parameter estimator, we begin our discussion with the details of the EM algorithm. The EM algorithm [39] consists of two steps, namely the *expectation* (or *E*) and *maximization* (or *M*) steps. In the E-step, the EM algorithm finds the lower bound of the likelihood function given in the right hand side of Eq. (20) by calculating the expected value of the joint log-likelihood function of the observed images and the LCM. Here, the expected value is computed over the LCMs given the most recent estimate of the map parameter vectors and observed data, i.e.,

$$\begin{aligned} Q(\mathbf{M} || \mathbf{M}^{t-1}) &= E[\log \Pr(\mathbf{Y}, X | \mathbf{M}) | \mathbf{Y}, \mathbf{M}^{t-1}] \\ &= E[\log \Pr(\mathbf{Y} | X, \mathbf{M}) + \log \Pr(X) | \mathbf{Y}, \mathbf{M}^{t-1}] \end{aligned} \quad (2.24)$$

where $\mathbf{Y} = \{Y_1, \dots, Y_N\}$ is the set of all observed remotely sensed images, $\mathbf{M} = \{M_1, \dots, M_N\}$ is the set of all unknown map parameters, and $\mathbf{M}^t = \{M_1^t, \dots, M_N^t\}$ is the set of all estimated parameters from the t -th iteration of the EM algorithm. Note here that $M_1^t = M_1^*$. By substituting Eq. (2.1) and Eq. (2.7) into Eq. (2.24), the expected value becomes

$$\begin{aligned} Q(\mathbf{M} || \mathbf{M}^{t-1}) &= E \left[-\frac{1}{2} \sum_{n=1}^N \left(\sum_{s \in \mathcal{S}} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \quad \quad \quad \begin{matrix} -1 \\ x_{s,n} \end{matrix} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) \right) - \sum_{c \in \mathcal{S}} V_c(X) - Z_X \right] | \mathbf{Y}, \mathbf{M}^{t-1} \end{aligned} \quad (2.25)$$

In the M-step, the expected value given in Eq. (2.25) is maximized and a new set of map parameter vectors is obtained, i.e.,

$$\mathbf{M}^t = \arg \max_{\mathbf{M}, M_1=M_1^*} Q(\mathbf{M} || \mathbf{M}^{t-1}) \quad (2.26)$$

Clearly, the terms $\log |x_{s,n}|$, $\log(2\pi)^{B_n}$, $\sum_{c \in \mathcal{S}} V_c(X)$, and Z_X in Eq. (2.25) do not depend on $\mathbf{M}$. Hence, Eq. (2.25) can be modified to

$$\mathbf{M}^t = \arg \max_{\mathbf{M}, M_1=M_1^*} Q(\mathbf{M} || \mathbf{M}^{t-1}) \quad (2.27)$$

where

$$Q(\mathbf{M}||\mathbf{M}^{t-1}) = \arg \max_{\mathbf{M}, M_1=M_1^*} E \left[-\frac{1}{2} \sum_{n=1}^N \sum_{s \in S} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \mathbf{x}_{s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) \middle| \mathbf{Y}, \mathbf{M}^{t-1} \right]. \quad (2.28)$$

To find the solution of Eq. (2.28), the *a posteriori* probability of the LCM given the observed images and the map parameters from the $(t-1)$ -th iteration must be calculated in order to find the expected value. From the same reason as discussed in Section 2, the posterior probability cannot be practically calculated due to the huge number of possible LCMs. As a result, we employ the approximation given in Eq. (2.16), and hence, we have

$$\Pr(X|\mathbf{Y}, \mathbf{M}^{t-1}) \approx \prod_{s \in S} p_s^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1}) = \prod_{s \in S} \frac{1}{Z_s} e^{-\frac{1}{2}h_s(x_s|\mathbf{Y}, \mathbf{M}^{t-1})}. \quad (2.29)$$

By substituting d Eq. (2.26) into Eq. (2.29), we have

$$\begin{aligned} Q(\mathbf{M}||\mathbf{M}^{t-1}) &\approx Q^{MF}(\mathbf{M}||\mathbf{M}^{t-1}) \\ &= -\frac{1}{2} \sum_{n=1}^N \sum_{s \in S} \sum_{x_s \in} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \mathbf{x}_{s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) p_s^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1}) \end{aligned} \quad (2.30)$$

Hence, in the M-step, the new map parameters can be obtained by maximizing the approximation given Eq. (1.30), i.e.,

$$\mathbf{M}^t = \arg \max_{\mathbf{M}, M_1=M_1^*} Q^{MF}(\mathbf{M}||\mathbf{M}^{t-1}) \quad (2.31)$$

Since $\mathbf{z}_{n,s}$ depends only M_n and the right hand side of Eq. (1.30) is written as the summation of $\mathbf{z}_{n,s}$ from different images, the above optimization problem can be rearranged into the optimization of each individual mapping parameters, i.e.,

$$M_n^t = \arg \max_M Q_n^{MF}(M_n||\mathbf{M}^{t-1}); n = 2, \dots, N \quad (2.32)$$

where

$$\begin{aligned} Q_n^{MF}(M_n||\mathbf{M}^{t-1}) &= -\frac{1}{2} \sum_{s \in S} \sum_{x_s \in} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \mathbf{x}_{s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) p_s^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1}). \end{aligned} \quad (2.33)$$

Using the approximations given above, the modified EM algorithm is displayed in Figure 2.2. For each iteration, the posterior probability $\Pr(X|\mathbf{Y}, \mathbf{M}^t)$ is approximated by recalculating $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$. We follow the work by Zhang [9] which suggested that $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$ can be obtained from

$$\begin{aligned}
 h_s(x_s|\mathbf{Y}, \mathbf{M}^t) &= h_{obv}(x_s|\mathbf{Z}) + h_{ng}(x_s|X_{NG}) \\
 &= \sum_{n=1}^N \left\{ (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \mathbf{x}_{s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log | \mathbf{x}_{s,n} | \right\} \\
 &\quad + \sum_{x_r \in} V_{\{s,r\}}(x_s, x_r) p^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1}).
 \end{aligned} \tag{2.34}$$

where $h_{obv}(x_s|\mathbf{Z})$ and $h_{ng}(x_s|X_{NG})$ are the potential functions depending upon the observation and neighboring pixels, respectively.

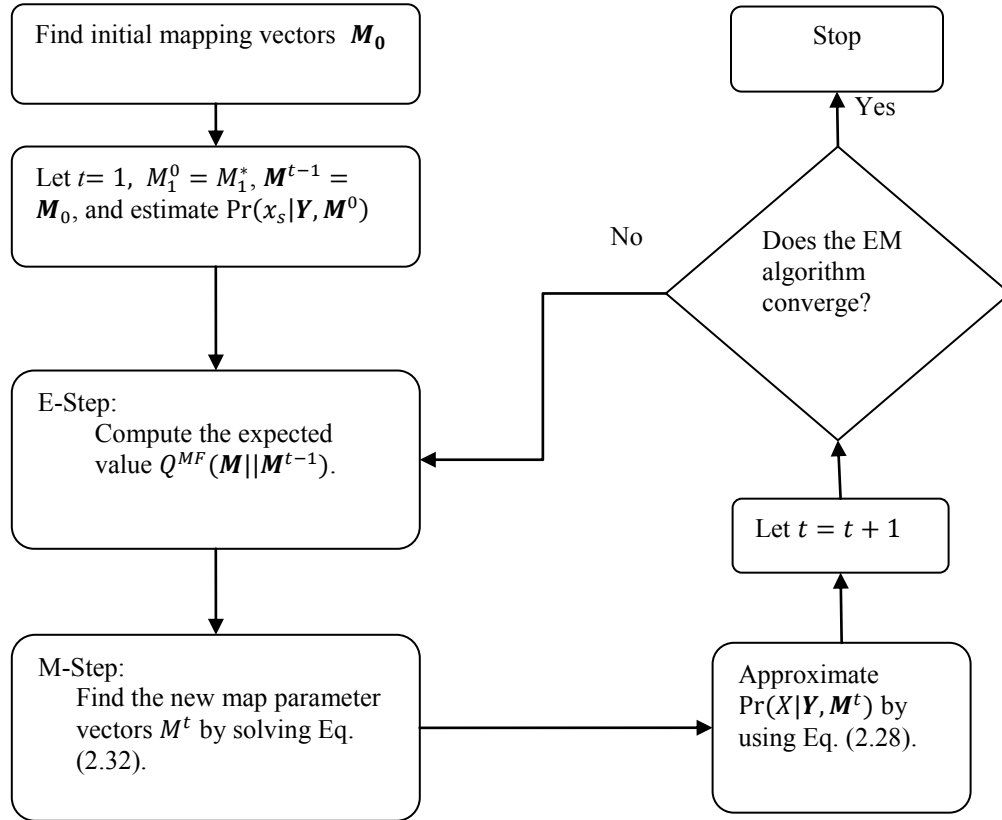


Figure 2.2: Block diagram of the modified EM algorithm

Since $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$ is recalculated for every iteration of the EM algorithm, we can choose a land cover class that minimizes $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$, and obtain the optimum LCM based on criterion given in Eq. (2.23) By combining the EM algorithm given in Figure 2.2 and the land cover mapping process by minimizing Eq. (2.23), the joint image registration and land cover mapping algorithm is given as

1. Initialize map parameters, i.e., $\mathbf{M}_1^0 = \mathbf{M}_1^*$ and $\mathbf{M}^0 = \{\mathbf{M}_1^0, \dots, \mathbf{M}_n^0\}$, let $t = 1$, and assign

$p^{MF}(x_s|Y, \mathbf{M}^0)$ based on some prior knowledge.

2. Compute $Q_n^{MF}(M_n||\mathbf{M}^{t-1})$ for $n = 2, \dots, N$.
3. Obtain M_n^t by solving Eq. (2.32) for $n = 2, \dots, N$, and assign $M_1^t = M_1^*$ and $\mathbf{M}^t = [M_1^t, \dots, M_N^t]$
4. Compute $h_s(x_s|Y, \mathbf{M}^t)$ by using Eq. (2.34).
5. Find the new LCM that minimizes $h_s(x_s|Y, \mathbf{M}^t)$ for all $s \in \mathcal{S}$.
6. Let $t = t + 1$, and go to Step 2 if a convergence criterion is not satisfied.

The critical step in the successful implementation of the joint image registration and land cover mapping algorithm proposed above is how to solve Eq. (2.32) efficiently. Here to find the maxima, we employ the particle swarm optimization (PSO) algorithm [46] since the traditional gradient search approaches are likely to fall into one of the local optimum points of $Q^{MF}(\mathbf{M}||\mathbf{M}^{t-1})$ due to its non-convexity. The PSO exploit the cooperative behavior for a group of animals such as birds and insects. In the PSO, an individual animal is called a particle and a group of animals is called swarm. These particles are initially distributed throughout the search space, and move around the search space. Based on some social and cooperative criteria, these particles will eventually cluster in the regions where the global optima can be found.

In our work, for a given image Y_n , each particle represents a mapping parameter and we denote the i -th particle as $M_{n,i}$. At each iteration, the i -th particle moves by a velocity V_i which is a function of the best-known positions (mapping parameter) discovered by the i -th particle (P_i) itself, and from all particles (G), i.e.,

$$V_i^r = \omega V_i^{r-1} + \varphi_1 u_1 (P_i - M_{n,i}^{r-1}) + \varphi_2 u_2 (G - M_{n,i}^{r-1}) \quad (2.35)$$

and

$$M_{n,i}^r = M_{n,i}^{r-1} + V_i^r \quad (2.36)$$

for $n = 2, \dots, N$. where ω is the inertial weight, φ_1 and φ_2 are acceleration constants, and u_1 and u_2 are uniform random numbers between zero and one. The velocity is usually kept in the range of $[V_{min}, V_{max}]$ to make sure that $M_{n,i}^r$ is in the valid regions. Note here that the performance of the PSO depends on the selection of, ω , φ_1 and φ_2 , and the number of iterations. In this paper, we set the number of particles to 80 and the maximum number of iterations to be 200 as a suitable setup for our experiment. We acknowledge that different setups of these parameters may result in different convergence rate. However, the investigation of the optimum parameter selection of the PSO in term of convergence rate is out of scope of this paper. We refer to the work by [47] for more details.

2.5 Experimental Results

In this section, we provide the results of two experiments based on the methodology derived in Section 4 to jointly register and classify a set of remotely sensed images. The first experiment is conducted over a simulated dataset in order for us to investigate many aspects of our proposed algorithm. Next, we will examine the performance of our algorithm in the actual remote sensing image. For both examples, the goal is to examine the performance of algorithm to different degrees of initial registration errors. If our algorithm performs perfectly, it should be able to align images together and produce a LCM from unregistered images as accurate as when images are registered.

Experiment 1:

In the experiment, we examine the performance of the proposed algorithm in term of classification performance and registration accuracy by attempting to produce a land cover map from a set of four simulated images. All the simulated images have an equal size of 512×512 pixels (Figure 2.3) and contain four land cover classes (Classes 1-4) with intensity values of zero, one, two and three for black, dark gray, light gray and white areas, respectively. Based on the noiseless image, the ground truth image in this example is given in Figure 2.4 where the blue, black, green and red colors correspond to Classes 1-4, respectively. Next, all of the input images are added with the independent and identical Gaussian noise with zero mean and standard deviation of $\sigma = 1$ to examine the performance of our proposed algorithm to the image noise. Figure 2.5 shows an example of the input image for $\sigma = 1$. We observe that the observed image appears to be very noisy.



Figure 2.3: Noiseless Simulated Image in Example 1

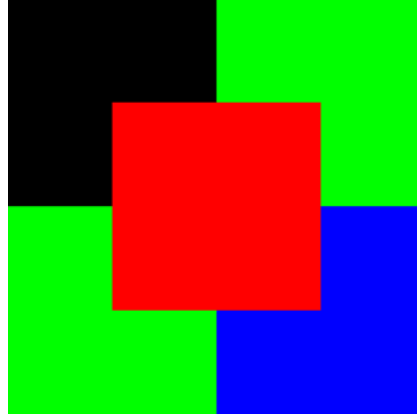


Figure 2.4: The ground data of Example 1

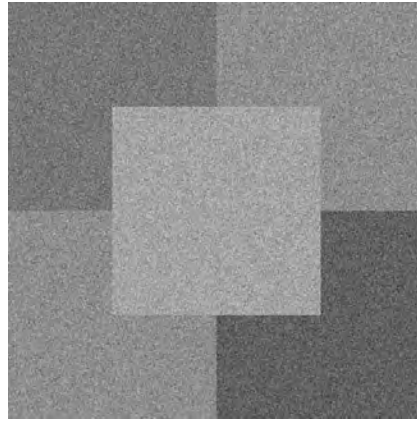


Figure 2.5: An example of the noisy input image at $\sigma=1$ in Example 1

Since our algorithm performs both image registration and land cover mapping at the same time, the performance of our algorithm can be evaluated in terms of how much the resulting LCM deviates from the reference LCM, and the estimation error between our calculated map parameters and the actual parameters that relate the LCM to the simulated images. If our algorithm performs perfect registration and land cover mapping, the resulting percentages of mis-classified pixels will be zero, and the registration error between images and LCM will be zero. In this example, the correct mapping parameters for all observed images are the same and equal to $M_{perfect} = [1,0,0,1,0,0]$ which correspond to unit scale, zero skew, and zero displacement. Next, since we want to examine the effect of the initial registration errors to the performance of our algorithm, we investigate different scenarios of initial registration errors by varying the initial mapping parameters between the observed images and LCM at different values of displacement, scale and skew parameters. In particular, we investigate three scenarios for the only displacement, only scale and only skew errors, respectively. Table 2.1 shows the initial mapping parameters for all three scenarios. Here, δ , ρ and η are the initial

displacement, scale, and skew parameter errors. Note that the initial mapping parameter errors for Image 1 for all scenarios are zero since we assume that the first image is registered to the LCM as mentioned in Section 2.31.

Before examining the performance of our proposed algorithm, we examine the effect of registration errors to the performance of image classification. This value can be viewed as the worst case scenario where the LCM is derived directly from the set of mis-registered images. Here, we employ the maximum likelihood classifier (MLC) [44] to the set of four remapped images, and the LCM is obtained from

$$x_s^{MLC} = \arg \min_x \left[\sum_{n=1}^4 \left\{ (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \boldsymbol{\Sigma}_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\boldsymbol{\Sigma}_{x_s,n}| \right\} \right] \quad (2.37)$$

where the subscript n denotes the n -th remapped image. We note here that Eq. (2.37) is the special of the optimum LCM obtained from Eq. (2.22) when $\beta = 0$. Figure 2.6 (a), (b) and (c) display the resulting LCM for $\delta = 12$ and $\sigma = 1$ for Scenario I, $\rho = 0.05$ and $\sigma = 1$ for Scenario II, and $\eta = 0.05$ and $\sigma = 1$ for Scenario III. The averaged percentages of misclassified pixels after a hundred independent runs are equal to 28.66%, 31.93 and 27.03, for Scenarios I, II and III given above, respectively.

Table 2.1: Mapping parameter errors are three scenarios in Example 1

	Image	Mapping parameters					
		m_1	m_2	m_3	m_4	m_5	m_6
Scenario I: Displacement error (δ)	1	1	0	0	1	0	0
	2	1	0	0	1	δ	0
	3	1	0	0	1	0	$-\delta$
	4	1	0	0	1	$-\delta$	δ
Scenario II: Scale error (ρ)	1	1	0	0	1	0	0
	2	$1 + \rho$	0	0	1	0	0
	3	1	0	0	$1 + \rho$	0	0
	4	$1 - \rho$	0	0	$1 - \rho$	0	0
Scenario III: Sheer error (η)	1	1	0	0	1	0	0
	2	1	η	0	1	0	0
	3	1	0	η	1	0	0
	4	1	$-\eta$	$-\eta$	1	0	0

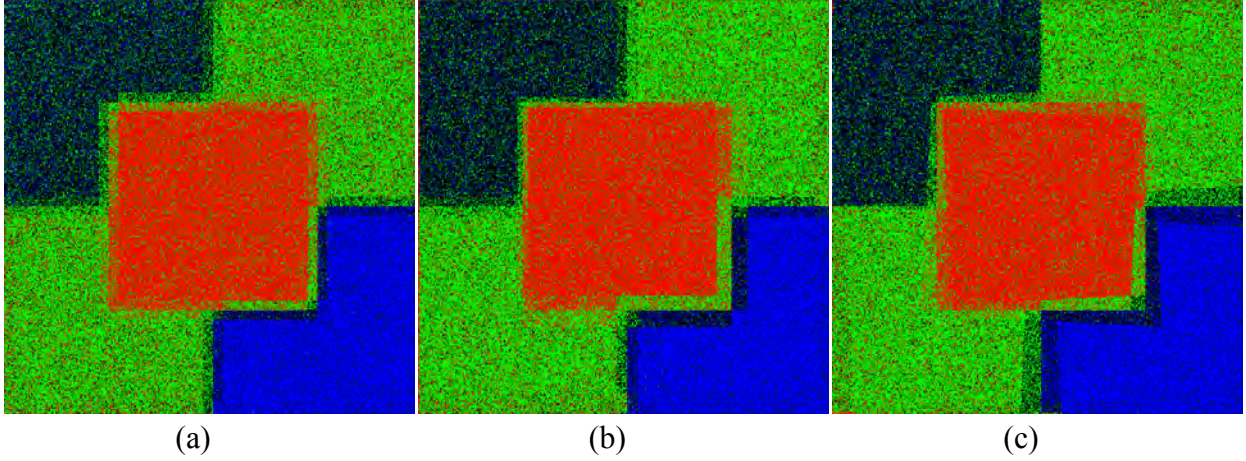


Figure 2.6: Examples of the MLC-Based LCMs for (a) Scenario I with $\delta=12$ and $\sigma=1$; (b) Scenario II with $\rho=0.05$ and $\sigma=1$; and (c) Scenario III with $\eta=0.05$ and $\sigma=1$

Next, the proposed algorithm is applied to the above datasets. The whole process was implemented using CUDA on NVIDIA Tesla M2090 with 1 GB memory. Here, we assign $p^{MF}(x_s|\mathbf{Y}, \mathbf{M}^0) = \frac{1}{4}$ as the most extreme case where no prior information is given. In different trials, the value of β is set to be 0.00, 0.25, 0.50, and 0.75 (see Eq. (2)). Since our algorithm performs both image classification and registration, the termination criteria must ensure the convergences in both the estimated posterior probability and mapping parameters. As a result, we define

$$p_{changes} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \sum_{x_s \in} |p_s^{MF}(x_s|\mathbf{Y}, \mathbf{M}^t) - p_s^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1})|, \quad (2.38)$$

to measure changes in the posterior probabilities from two consecutive iterations. We also define

$$d_{movement,n} = \frac{1}{|\mathcal{T}_n|} \sum_{(u_n, v_n) \in \mathcal{T}_n} \sqrt{(x_n^t - x_n^{t-1})^2 + (y_n^t - y_n^{t-1})^2} \quad (2.39)$$

to characterize the movement of coordinates of the remapped image Z_n from two consecutive iterations where

$$\begin{bmatrix} x_n^t \\ y_n^t \end{bmatrix} = \begin{bmatrix} m_{1,n}^t & m_{2,n}^t \\ m_{3,n}^t & m_{4,n}^t \end{bmatrix}^{-1} \begin{bmatrix} u_n - m_{5,n}^t \\ v_n - m_{6,n}^t \end{bmatrix}. \quad (40)$$

Here, $m_{i,n}^t$ denotes the mapping parameter m_i from the n^{th} at the t^{th} iteration. In this example, the algorithm terminates when $p_{changes}$ is less than $p_{min} = 10^{-5}$, and $d_{movement,n}$ is less than 0.1 pixels for five consecutive iterations for $n = 2, 3, 4$. To create a benchmark for our

proposed algorithm, we examined two extreme cases where LCMs are derived directly from the unregistered image pairs and from perfect registered image pair. The LCMs from these extreme cases are classified using our proposed algorithm by fixing $M^t = M^*$. For perfect registration, we have $M^* = M_{perfect}$ whereas, for unregistered image pairs, we set M^* equal the values given in Table 2.1 for the respective scenarios. The first extreme case can be considered as the lower limit on the classification accuracy if we perform the land cover mapping without alignment of images first. The second case is an upper bound on the classification accuracy when we produce a map from a registered image pair. By setting up our experiment in this fashion, we can investigate how much improvement our algorithm can gain by integrating the registration and classification together, and how far the performance of our algorithm is from the upper limit where all uncertainties in registration are removed. To ensure the statistical significance of our experiment, all experiments are repeated ten times.

Table 2.2 displays the averaged percentages of misclassified pixels (PMP) of the LCMs for different values of β and for Scenario I with $\delta = 12$, Scenario II with $\rho = 0.05$ and Scenario III with $\eta = 0.05$ when $\sigma = 1$. Note here that, in this example, we employ the percentages of mis-classified pixels as the performance metric to evaluate the classification performance rather than the overall accuracy to highlighted small differences in the classification performance between LCMs derived from image datasets without registration error and LCMs obtained from our proposed algorithm. From Table 2.2, it is clear that, from all scenarios, the PMPs derived from image datasets without registration errors corrections are always significance poorer than those derived from registered image datasets. These results support our claims that it is important to consider lack of alignments in performing image classification. We also observe that, for $\beta = 0.25, 0.5$ and 0.75 , our proposed algorithm produced the LCM with the accuracy similar to those obtained from image dataset without any registration error. These results imply that our proposed algorithm attain the upper-bound accuracy with proper selection of MRF parameter. To ensure the statistical significance, we compute the pairwise t -statistics for unequal variance populations [43] of the PMPs obtained from LCMs derived from the proposed algorithm for various initial registration errors against those obtained from image dataset with no registration error, and the resulting p -values [43] of the t -statistics are given in Table 2.3. The p -value represents the probability that there is no difference in PMPs. Hence, a smaller p -value implies that PMPs from two experiments are different. We also compute the t -statistics comparing LCMs obtained from image dataset with and without registration errors. The resulting p -values of these t -statistics are also summarized in Table 2.3. It is clear from Table 2.3 that there is significant differences in term of PMPs from LCMs obtained from image dataset with and without registration errors. Furthermore, the p -values also support our claim that or $\beta = 0.25, 0.5$ and 0.75 , our proposed algorithm produced the LCM with the accuracy similar to those obtained from image dataset without any registration error. However at $\beta = 0$, our proposed algorithm perform significantly poorer than those of perfect registration. In fact, at $\beta = 0$, our proposed algorithm achieves roughly the same performance as in situation where there is no registration error correction since at

$\beta = 0$, our proposed algorithm cannot correctly estimate the map vectors. Figure 2.7 shows examples of the resulting LCMs at $\beta = 0.75$ for all scenarios. We observe that all the LCMs appeared to be more connected than the MLC-based LCMs given in Figure 2.6: Examples of the MLC-Based LCMs for (a) Scenario I with $\delta=12$ and $\sigma=1$; (b) Scenario II with $\rho=0.05$ and $\sigma=1$; and (c) Scenario III with $\eta=0.05$ and $\sigma=$

Table 2.2: Comparison of the averaged percentages of misclassified pixels (PMP) between two extreme cases and our proposed algorithm

β	No registration Error	No registration error correction			Proposed Algorithm with initial registration errors		
		Scenario I with $\delta = 12$	Scenario II with $\rho = 0.05$	Scenario III with $\eta = 0.05$	Scenario I with $\delta = 12$	Scenario II with $\rho = 0.05$	Scenario III with $\eta = 0.05$
0.0	25.65%	28.66%	26.87%	27.05%	28.65%	26.07%	27.12%
0.25	0.43%	4.81%	5.96%	6.45%	0.45%	0.43%	0.43%
0.5	0.039%	4.24%	5.65%	6.21%	0.039%	0.041%	0.043%
0.75	0.021%	4.19%	5.56%	6.13%	0.024%	0.032%	0.026%

Table 2.3: the p-values of the pairwise t-test with unequal variances of our proposed algorithm to the perfect registration cases, and no registration error correction to the perfect registration cases

β	No registration Error	No registration error correction			Proposed Algorithm with initial registration errors		
		Scenario I with $\delta = 12$	Scenario II with $\rho = 0.05$	Scenario III with $\eta = 0.05$	Scenario I with $\delta = 12$	Scenario II with $\rho = 0.05$	Scenario III with $\eta = 0.05$
0.0	1	1.5×10^{-22}	1.6×10^{-14}	4.0×10^{-18}	1.9×10^{-23}	4.0×10^{-15}	3.9×10^{-15}
0.25	1	2.0×10^{-17}	3.5×10^{-19}	3.6×10^{-18}	0.457	0.717	0.500
0.5	1	1.5×10^{-15}	2.8×10^{-17}	1.8×10^{-16}	0.712	0.167	0.401
0.75	1	1.5×10^{-14}	1.4×10^{-15}	6.2×10^{-17}	0.060	0.033	0.079

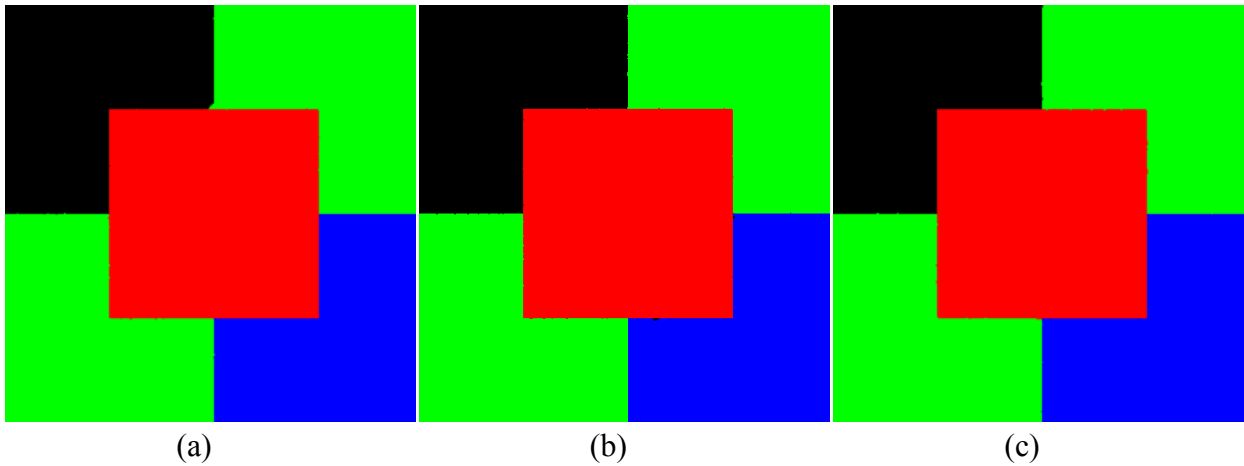


Figure 2.7: Examples of the resulting LCMs from our proposed algorithm (a) Scenario I with $\delta=12$ and $\sigma=1$; (b) Scenario II with $\rho=0.05$ and $\sigma=1$; and (c) Scenario III with $\eta=0.05$ and $\sigma=1$

Since at $\beta = 0.75$, our proposed algorithm achieves highest performance, we examine the effect of the initial registration errors to the performance of our algorithm by varying values of δ , ρ , and η for Scenarios I, II and III, respectively for $\beta = 0.75$. Again, ten independent runs are performed to ensure the statistical significance and the results are given in Table 2.4. We observe that, for all scenarios, the PMPs are roughly the same. In other words, the initial registration errors have little effect on the performance of our algorithm. These results imply the robustness of our proposed algorithm to the initial mis-registration errors if the proper value of β is chosen.

Table 2.4: The averaged percentages of mis-classified pixels as the function of the initial registration error for all Scenarios

Scenario I		Scenario II		Scenario III	
δ	PMP	ρ	PMP	η	PMP
0	0.019%	-0.05	0.035%	-0.05	0.036%
4	0.032%	-0.03	0.035%	-0.03	0.029%
8	0.029%	-0.01	0.022%	-0.01	0.043%
12	0.026%	0.01	0.030%	0.01	0.040%
		0.03	0.024%	0.03	0.036%
		0.05	0.032%	0.05	0.026%

Another key performance metric in this example is the residual registration errors after processing. Table 2.5 displays the means and standard deviations of the root mean square errors (RMSEs) from ten independent runs between each simulated images and the reference LCM. The RMSE of the n -th image is computed from

$$RMSE_n = \frac{1}{|\mathcal{T}_n|} \sum_{(u_n, v_n) \in \mathcal{T}_n} \sqrt{(x_n^{gt} - x_n^{est})^2 + (y_n^{gt} - y_n^{est})^2} \quad (2.41)$$

where (x_n^{gt}, y_n^{gt}) and (x_n^{est}, y_n^{est}) are the ground truth and estimated coordinates. Here, the ground truth coordinates obtained by letting $M_n = M_{perfect}$. Clear, for $\beta = 0.25, 0.5$, and 0.75 , our algorithm can successfully registered all images with the LCMs. However, at $\beta = 0$ our algorithm cannot align these images with the LCM. The results in Table 2.5 emphasize the importance of parameter selection. Note here that the RMSE of Image 1 is not shown in the Table 2.5 since it is assumed to be perfectly aligned (registration error is zero.) with the LCM. Next, we examine the effect of image noise to the registration accuracy by varying the noise variance σ^2 from -30dB to 0 dB and the resulting averaged RMSEs for $\beta = 0.0$ and 0.75 are given in Table 2.6 and Table 2.7, respectively. We observe here that there are slight performance differences in term of the RMSEs for σ^2 of -30, -20 and -10 dB for both $\beta = 0.00$ and 0.75 . However, for the noise variance equal to 0 dB, our algorithm can only correctly aligned Images 2-4 to the LCM at $\beta = 0.00$. This result emphasizes the importance of a parameter selection to the convergence of our algorithm. For the performance comparison, we compare the registration accuracy of our proposed algorithm for various scenarios and $\beta = 0.75$ with a traditional image-to-image registration technique. Here we employ the mean square error criteria (MSEC) [16] since the MSEC is suitable for register images with the same modality and suffered from additive Gaussian noise. For the traditional image-to-image registration, we register Images 2-4 with Image 1 since Image 1 is assumed to be aligned with the LCM. The averaged RMSEs from ten independence runs for various noise variances are given in Table 2.8. Again the particle swarm optimization algorithm with eighty particles is employed to ensure global optimality. As expected, the registration accuracy decreases as the noise variance increase. By comparing Table 2.6 and Table 2.8, the RMSEs from our proposed algorithm seem to be lower (better) than those obtained from the MSEC for noise variances equal to -20, -10 and 0 dBs. Next, we again perform the pairwise t -test to determine whether there are significant differences in RMSEs obtained from our proposed algorithm and the MSEC, and the resulting p -values [43] are shown in Table 2.9. From the p -values, we can conclude that our proposed algorithm achieves significantly better registration accuracies than those obtained from the MSEC for the noise variances of -20, -10 and 0 dBs. Note here that, for a noise variance equal to -30 dB, the registration errors from our proposed algorithm and the MSEC are roughly zero and, therefore there is no different in term of registration accuracy.

Table 2.5: The residue registration errors of our proposed algorithm for various scenarios and values of β .

Scenario			No registration Error Correction	$\beta = 0.0$	$\beta = 0.25$	$\beta = 0.50$	$\beta = 0.75$
I ($\delta = 12$)	Image 2	Mean	12	11.99	0.111	0.295	0.280
		STD	-	0.0015	0.259	0.139	0.100
	Image 3	Mean	12	11.99	0.031	0.192	0.312
		STD	-	0.0018	0.020	0.120	0.156
	Image 4	Mean	16.97	16.96	0.213	0.338	0.212
		STD	-	0.0017	0.566	0.088	0.136
II ($\rho = 0.05$)	Image 2	Mean	14.06	13.56	0.028	0.281	0.327
		STD	-	0.072	0.010	0.130	0.113
	Image 3	Mean	14.06	13.49	0.020	0.353	0.312
		STD	-	0.032	0.080	0.102	0.106
	Image 4	Mean	21.97	20.97	0.253	0.245	0.315
		STD	-	0.095	0.636	0.120	0.082
III ($\eta = 0.05$)	Image 2	Mean	14.76	14.71	0.025	0.295	0.296
		STD	-	0.204	0.020	0.149	0.098
	Image 3	Mean	14.76	14.73	0.017	0.415	0.350
		STD	-	0.182	0.006	0.090	0.136
	Image 4	Mean	21.72	22.04	0.350	0.312	0.371
		STD	-	0.0325	0.983	0.155	0.088

Table 2.6: The residue registration errors for various noise variances and $\beta=0.75$.

Noise variance (dB)	Average root mean square errors								
	Scenario I, $\delta = 12$			Scenario II, $\rho = 0.05$			Scenario III, $\eta = 0.05$		
	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4
-30	0.007	0.011	0.009	0.006	0.010	0.019	0.012	0.019	0.013
-20	0.010	0.012	0.009	0.023	0.016	0.012	0.017	0.016	0.011
-10	0.036	0.035	0.037	0.028	0.018	0.029	0.028	0.030	0.022
0	0.244	0.280	0.185	0.119	0.138	0.071	0.078	0.053	0.200

Table 2.7: The residue registration errors for various noise variances and $\beta=0$.

Noise variance (dB)	Average root mean square errors								
	Scenario I, $\delta = 12$			Scenario II, $\rho = 0.05$			Scenario III, $\eta = 0.05$		
	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4
-30	0.016	0.08	0.010	0.015	0.007	0.019	0.009	0.011	0.019
-20	0.017	0.012	0.014	0.015	0.018	0.015	0.010	0.015	0.017
-10	0.014	0.018	0.015	0.018	0.018	0.023	0.019	0.016	0.014
0	11.99	11.99	16.97	11.91	11.89	20.28	12.75	12.79	20.61

Table 2.8: The residue registration errors using the minimum mean square error criteria for various noise variances.

Noise variance (dB)	Image 2		Image 3		Image 4	
	Mean	STD	Mean	STD	Mean	STD
-30	0.008	0.0029	0.007	0.0041	0.010	0.0054
-20	0.422	0.0040	0.425	0.0033	0.423	0.0049
-10	0.663	0.0037	0.665	0.0014	0.664	0.0017
0	0.875	0.516	1.637	1.441	1.352	0.9744

Table 2.9: The p-value from the pairwise t-test between the traditional registration method and our proposed algorithm for various Scenario at $\beta=0.75$.

Noise variance (dB)	Average root mean square errors								
	Scenario I, $\delta = 12$			Scenario II, $\rho = 0.05$			Scenario III, $\eta = 0.05$		
	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4
-30	0.829	0.402	0.883	0.413	0.413	0.201	0.507	0.092	0.407
-20	1×10^{-18}	4×10^{-14}	2×10^{-21}	1×10^{-13}	1×10^{-13}	3×10^{-15}	2×10^{-13}	2×10^{-13}	5×10^{-17}
-10	3×10^{-14}	2×10^{-14}	3×10^{-14}	3×10^{-15}	3×10^{-15}	5×10^{-16}	2×10^{-23}	1×10^{-14}	7×10^{-17}
0	0.004	0.016	0.004	0.001	0.001	0.003	0.0010	0.007	0.004

Figure 2.8 shows the averaged numbers of iterations that the algorithm requires before the convergence criterion is satisfied for different scenarios and β . For $\beta = 0.25, 0.5$ and 0.75 , more iterations are needed as the value of β increases. However, at $\beta = 0$, our algorithm terminates at the higher numbers of iterations for Scenarios II and III. The main reason to the slow convergences is due to the small changes in the mapping parameters from one iteration to another and since $\beta = 0$, this small changes in the mapping parameters have significant influence on the posterior probability.

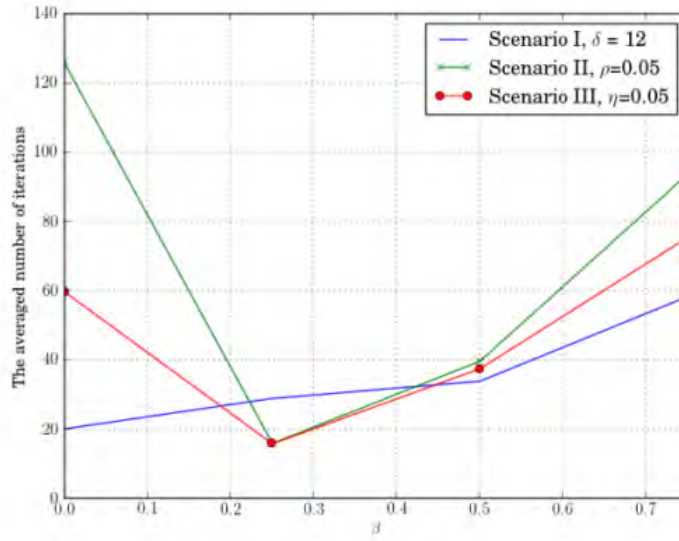


Figure 2.8: The averaged number of iterations required before the termination criteria are satisfied for different scenarios in Example 1

Experiment 2

A QuickBird dataset consisting of one multispectral image (MI) of size 150×300 pixels and one panchromatic image (PAN) of size $600 \times 1,200$ pixels was used in this experiment (Figure 2.9). The MI and PAN have resolutions of 2.4 and 0.6 meters, respectively. Both images captured a part of Kasetsart University in Bangkok, Thailand, covering around 0.2592 in km^2 on July 10th, 2008. By visual interpretation, we classified the area into five classes, namely, water, shadows, vegetations and impervious type 1 and impervious type 2, and the ground truth image is shown in Figure 2.10 where blue, black, green, red and white colors correspond to water, shadow, vegetation, impervious type 1 and impervious type 2, respectively. Here, the impervious is divided into two types due to different roof and pavement colors in the scene. By using both PAN and MI images, we randomly select 1000 samples for each land cover classes.



Figure 2.9: QUICKBIRD dataset of a part of Kasetsart University (a) False color composite MI; and (b) PAN

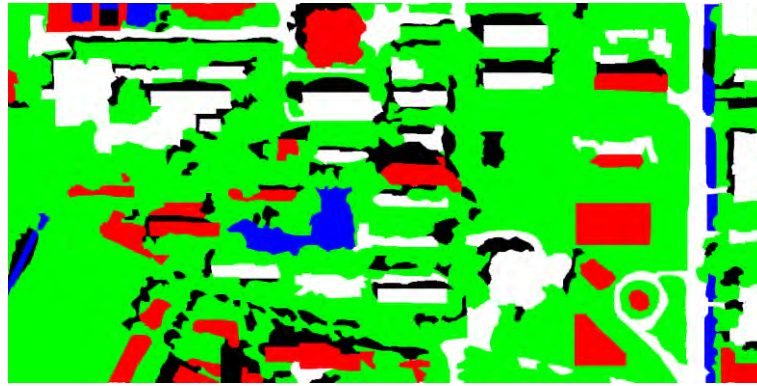


Figure 2.10: Ground truth image for Example 2 (green, blue, black, red and white colors for vegetation, water, shadow, impervious type 1 and impervious type 2, respectively)

In Experiment 2, we focused on the robustness of the proposed algorithm with different degrees of the initial displacement, scale and rotation errors. In fact, there are six displacement errors in x -direction and y - directions, four scale errors and six rotational errors used in this experiment. The termination criteria used in this example is similar to those in Example 1, i.e., our algorithm is terminated if $p_{changes}$ (see Eq. (2.38)) is less than 10^{-5} and $d_{movement,MI}$ (see Eq. (2.39)) is less than 0.1 pixels for five consecutive iterations. Before examining the robustness of our algorithm, we determined the benchmark performance of the MRF-based land cover mapping when MI and PAN are perfectly registered. The resulting LCMs are shown in Figure 2.11. Again, as we progress to greater values of β , more connected LCMs are obtained. The overall accuracy graph shown in Figure 2.12 agrees with the visual inspection that the classification performance increases as the values of β increases. In this example, we employ the overall accuracy rather than the percentages of mis-classified pixels used in Example 1 since overall accuracy is more widely used performance metrics in remote sensing image classification.

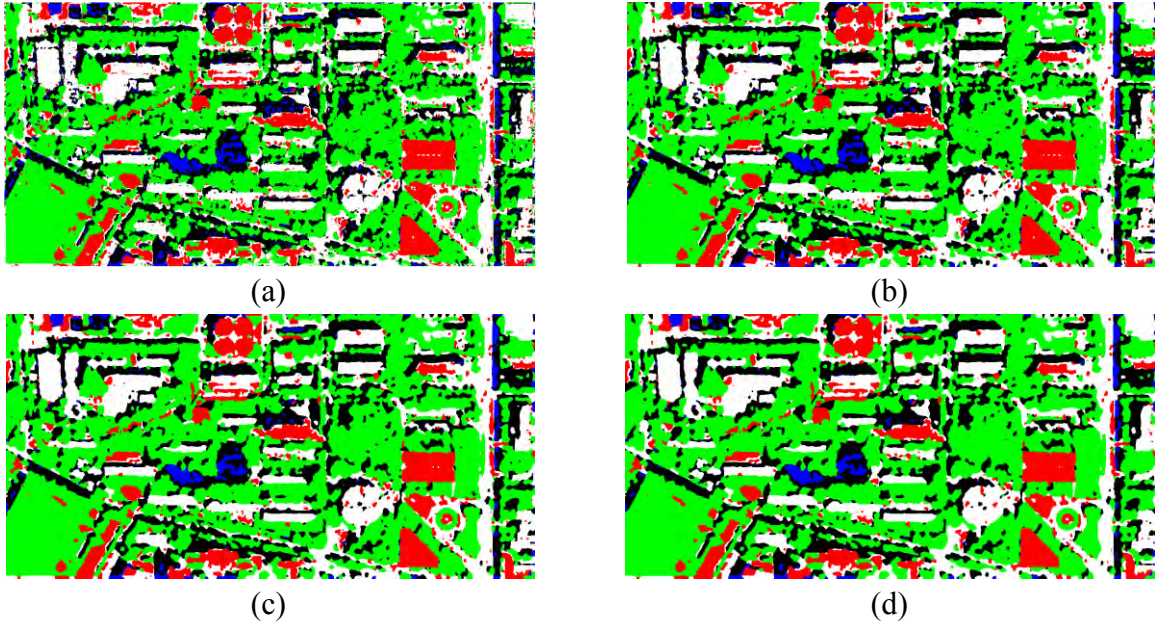


Figure 2.11: LCMs for the perfect registration case for (a) $\beta=0$; (b) $\beta=0.25$; (c) $\beta=0.50$; and (d) $\beta=0.75$

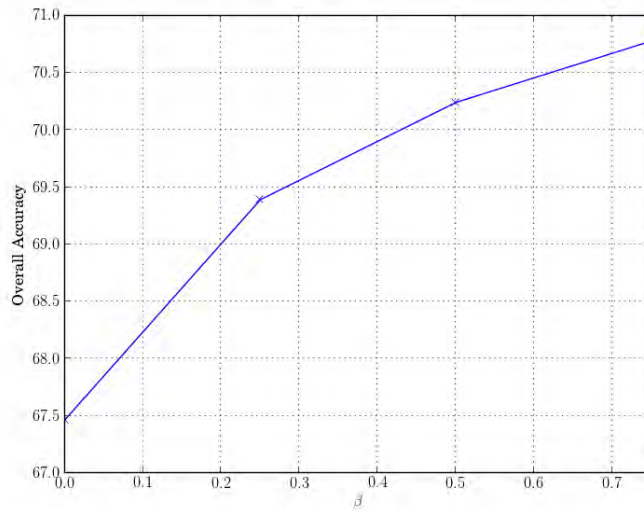


Figure 2.12: Overall accuracies for different values of β when MI and PAN are perfectly aligned

Since the PAN has a higher resolution, we assume that it is aligned with the LCM, and we only need to find map parameters of the MI. Here, the PAN has a higher resolution than the multispectral image by the factor of four, and both MI and PAN are obtained from the same satellite, the optimal map parameter vector relating the two images should be equal to

$$M^{opt} = [0.25, 0, 0, 0.25, 0, 0].$$

To investigate the robustness of our algorithm to displacement, scale and rotation errors, we introduce the registration errors in displacement in x -direction, displacement in y -direction, scale and rotation into the MI and PAN pair. If we introduce the displacements into the image pair, the initial map parameter is set to be equal to

$$M^0 = [0.25, 0, 0, 0.25, x, 0]$$

and

$$M^0 = [0.25, 0, 0, 0.25, 0, y],$$

for displacement errors in x - and y -direction, respectively. Here, the values of x and y are set to be $-5, -3, -1, 1, 3$, and 5 . For the scale and rotation errors, we assume that the both images are aligned at mid points, i.e., the pixel (75,150) of MI is at a pixel (300,600) of PAN. Next, we apply the initial scale errors, s , of $-5\%, -2.5\%, 0\%, 2.5\%$ and 5% when comparing with the PAN image to the multispectral image. Here, the scale errors of $-5\%, -2.5\%, 0\%, 2.5\%$ and 5% correspond to the initial scales of PAN to MI of 3.8, 3.9, 4.0, 4.1 and 4.2, respectively. For rotation errors, we rotate the MI by θ degrees in the counter clockwise direction. Here, the initial rotation errors θ are set to be $-3, -2, -1, 1, 2$, and 3 . The initial RMSE_{MI} (see Eq. (41)) for all cases are given in Table 2.10. Again, if our algorithm performs perfectly, the estimated map parameter will converge back to M^{opt} . In other words, we will eventually have $M^t = M^{opt}$. Once the correct map parameter vector is obtained, the classification accuracies of the LCMs should be equal to that in the perfect registration cases (Figure 2.11 (a)-(d)). In this example, we again assign $p^{MF}(x_s|Y, M^0) = \frac{1}{5}$, the most extreme case where no prior information is given.

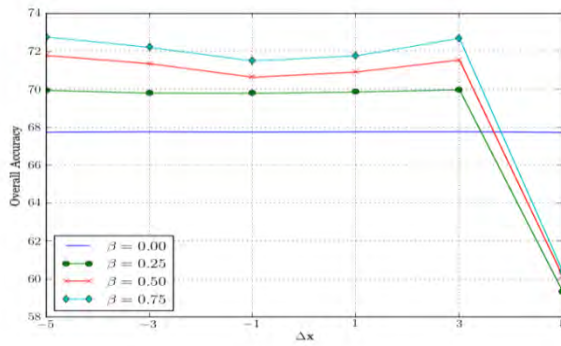
Table 2.10: The initial RMSEMI in meters (pixels in LCM) for various cases in Example 2

Error in x - direction		Error in y - direction		Error in scale		Error in Rotation	
x	RMSE _M	y	RMSE _M	s	RMSE _M	θ	RMSE _M
-5	12 (20)	-5	12 (20)	-5%	21.3 (36)	-3	11.12 (19)
-3	7.2 (12)	-3	7.2 (12)	-2.5%	10.7 (18)	-2	7.45 (12)
-1	2.4 (4)	-1	2.4 (4)	0%	0.0 (0)	-1	3.72 (6.2)
1	2.4 (4)	1	2.4 (4)	2.5%	10.7 (18)	1	3.72 (6.2)
3	7.2 (12)	3	7.2 (12)	5%	21.3 (36)	2	7.45 (12)
5	12 (20)	5	12 (20)			3	11.12 (19)

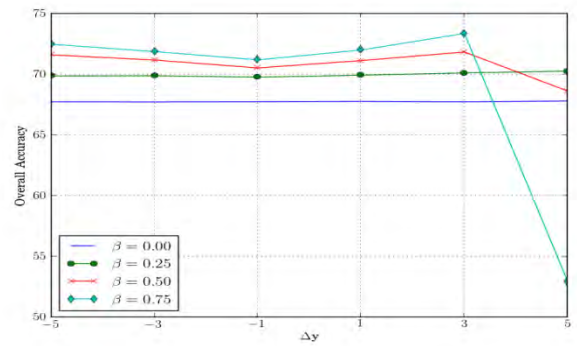
The overall accuracies as the function of x , y , s , and θ are shown in Figure 2.13 (a)-(d), respectively. From all most all scenarios, the overall accuracies increase as the value of β increase since the MRF model promote more connected land cover maps, and, therefore, remove the isolated misclassified pixels. However, for $x = 5$ and $y = 5$, the overall accuracies of our algorithms decrease as β increases. The main reason to these performance degradations are due to the fact that our algorithm terminate on one of the local optima since

the EM algorithm employed in our work cannot guarantee the global optimum solution. The evidence can be seen in Figure 2.14 (a) and (b). In Figure 2.13: The effect of initial registration errors to the overall accuracies

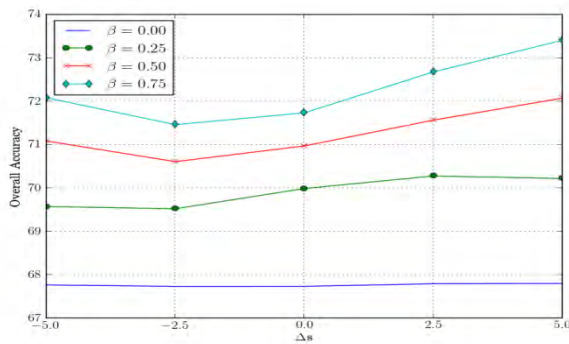
(a), we observe that, in almost all of the initial values of x , the number of iterations increases as the value β increases. However, for $x = 5.0$, our algorithm terminates at only 75, 109 and 129 iterations for $\beta = 0.25, 0.5$ and 0.75 whereas, for $\beta = 0.0$, our algorithm terminates after 180 iterations. Similarly, we observe the same phenomenal in Figure 2.13 (b) for $\Delta x = 5.0$ where the algorithm terminates at the lower number of iterations for $\beta = 0.75$ than $\beta = 0.0, 0.25$ and 0.50 . This result shows the effect of the initial registration errors to the convergence of our algorithm. In most practical situation, such an large initial registration errors is unlikely to occur since most remote sensing images are embedded with coordinate information from a producer.



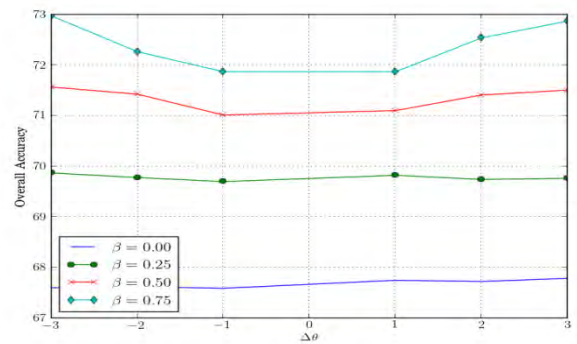
(a)



(b)

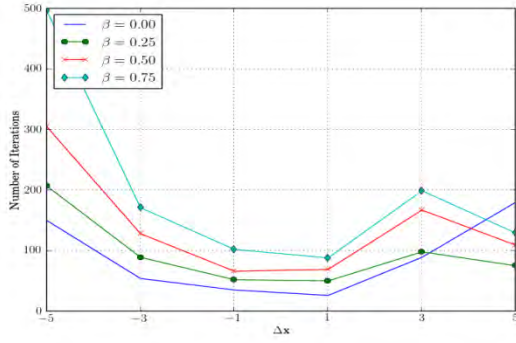


(c)

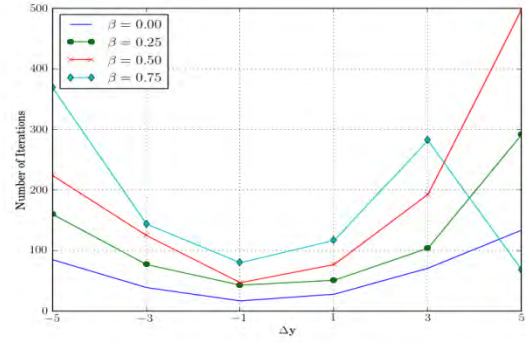


(d)

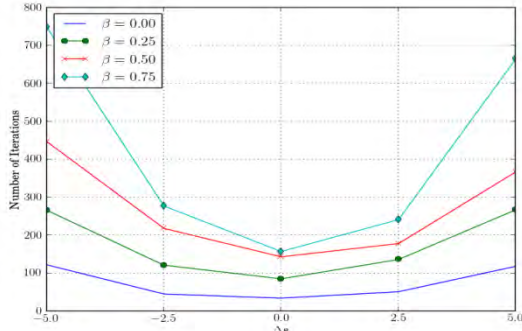
Figure 2.13: The effect of initial registration errors to the overall accuracies



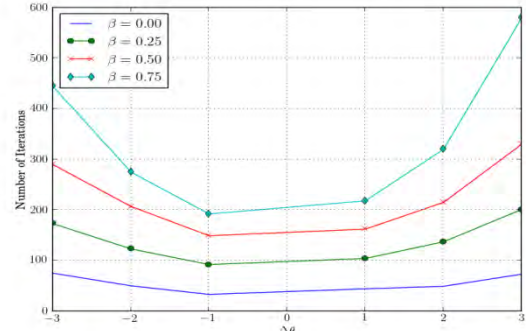
(a)



(b)



(c)



(d)

Figure 2.14: The effect of the initial registration errors to the number of iterations

Similar to previous example, we also compare the performance of our proposed algorithm (PA) with two extreme cases where images are perfectly registered (PR) and there is no registration error correction (NC) and the results are given in Tables 2.11-2.14. From this comparison, we observe that if our algorithm converges to the global optimum solutions, the resulting overall accuracies from our proposed algorithm are similar to those of the perfect registration cases, and the significant improvements are obtained from the cases where there is no registration error correction. The maximum performance improvements from no registration error correction for each cases are 12.6% for $\Delta x = -5$ and $\beta = 0.75$, 12.4% for $\Delta y = -5$ and $\beta = 0.75$, 17.4% for $\Delta s = -5\%$ and $\beta = 0.75$, and 14.9% for $\Delta \theta = 3^\circ$ and $\beta = 0.75$. We observe that the maximum improvements are achieve at $\beta = 0.75$. This observation suggests that a higher performance gain can be obtain by increasing the value of β . Next, we also notice that our proposed algorithm can sometimes achieve even higher accuracies than those of the prefect registration cases. The reason is due to the fact that our algorithm requires more iterations than the scenarios where image pair is perfectly registered since our algorithm terminates if both the estimated map parameters and the resulting LCM converge whereas, in the perfect registration case, the process terminates if only the resulting

LCM converges. Hence, our algorithm may terminate at lower percentages of changes in the LCM, and result in more accurate LCM which results in higher precision.

Table 2.11: Overall accuracies for different values of β on two extreme cases and our proposed algorithm for different initial displacement error in x-direction Δx where PA and NC denote the cases of the proposed algorithm and no registration error correction, res

β	PR	$\Delta x = -5.0$		$\Delta x = -3.0$		$\Delta x = -1.0$		$\Delta x = 1.0$		$\Delta x = 3.0$		$\Delta x = 5.0$	
		PA	NC	PA	NC	PA	NC	PA	NC	PA	NC	PA	NC
0.0	67.5	67.7	57.6	67.8	62.2	67.7	66.9	67.8	66.7	67.7	61.8	67.8	57.0
0.25	69.4	70.0	58.8	69.8	63.7	69.8	68.6	69.9	68.3	70.0	63.4	59.3	58.4
0.5	70.3	71.8	59.7	71.4	64.6	70.6	69.6	70.9	69.2	71.5	64.4	60.2	59.2
0.75	71.1	72.8	60.2	72.2	65.2	71.5	70.3	71.8	70.0	72.7	65.0	60.4	59.9

Table 2.12: Overall accuracies for different values of β on two extreme cases and our proposed algorithm for different initial displacement error in y-direction Δy where PA and NC denote the cases of the proposed algorithm and no registration error correction, res

β	PR	$\Delta y = -5.0$		$\Delta y = -3.0$		$\Delta y = -1.0$		$\Delta y = 1.0$		$\Delta y = 3.0$		$\Delta y = 5.0$	
		PA	NC	PA	NC	PA	NC	PA	NC	PA	NC	PA	NC
0.0	67.5	67.7	57.6	67.7	62.2	67.7	66.9	67.7	66.7	67.7	61.8	67.8	57.0
0.25	69.4	69.9	58.8	69.9	63.7	69.8	68.6	70.1	68.3	70.1	63.4	70.3	58.4
0.5	70.3	71.6	59.7	71.2	64.6	70.5	69.6	71.8	69.2	71.8	64.4	68.6	59.2
0.75	71.1	72.5	60.1	71.9	65.2	71.2	70.3	73.4	70.0	73.4	64.9	62.9	59.9

Table 2.13: Overall accuracies for different values of β on two extreme cases and our proposed algorithm for different initial scale error Δs where PA and NC denote the cases of the proposed algorithm and no registration error correction, respectively.

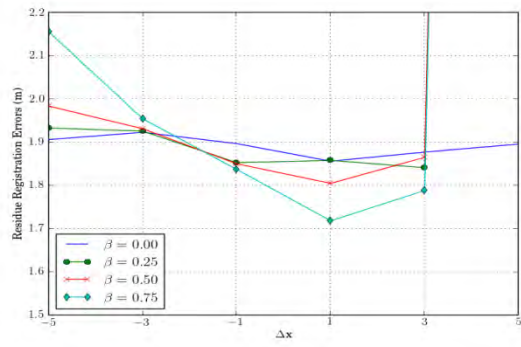
β	PR	$\Delta s = -5\%$		$\Delta s = -2.5\%$		$\Delta s = 0\%$		$\Delta s = 2.5\%$		$\Delta s = 5\%$	
		PA	NC	PA	NC	PA	NC	PA	NC	PA	NC
0.0	67.5	67.8	52.7	67.7	61.0	67.7	67.5	67.8	64.9	67.8	57.8
0.25	69.4	69.6	53.4	69.5	62.4	70.0	69.4	70.3	66.1	70.2	58.9
0.5	70.3	71.1	54.2	70.6	63.3	71.0	70.3	71.6	67.0	72.1	59.7
0.75	71.1	72.1	54.7	71.5	64.2	71.1	71.1	72.7	67.6	73.4	60.1

Table 2.14: Overall accuracies for different values of β on two extreme cases and our proposed algorithm for different rotation error θ where PA and NC denote the cases of the proposed algorithm and no registration error correction, respectively.

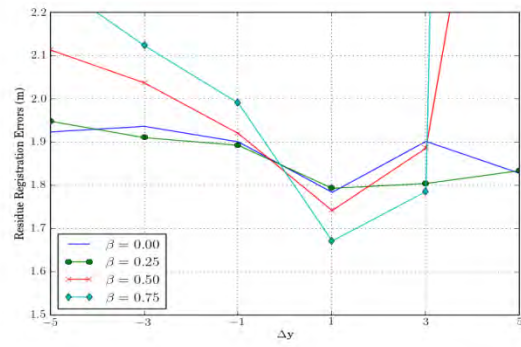
β	PR	$\Delta\theta = -3^\circ$		$\Delta\theta = -2^\circ$		$\Delta\theta = -1^\circ$		$\Delta\theta = 1^\circ$		$\Delta\theta = 2^\circ$		$\Delta\theta = 3^\circ$	
		PA	NC	PA	NC	PA	NC	PA	NC	PA	NC	PA	NC
0.0	67.5	67.6	57.3	67.6	60.8	67.6	65.3	67.7	64.8	67.7	59.8	67.8	55.5
0.25	69.4	69.9	58.5	69.8	62.2	69.7	66.9	69.9	66.5	69.7	61.1	69.8	56.6
0.5	70.3	71.6	59.3	71.4	63.0	71.0	67.8	71.1	67.4	71.4	62.0	71.5	57.4
0.75	71.1	73.0	59.7	72.3	63.6	71.9	68.4	71.9	68.1	72.5	62.6	72.9	58.0

Another key performance metric of our algorithm is the resulting registration errors. Figure 2.15 (a)-(d) show the residue registration errors in term of RMSE (in meters) between the MS image and the LCM for different initial registration errors. We observe that, if our algorithm converges to the global optimum solutions, it can successful reduce the registration error down to around 1.8 meters in the LCM (or equivalently 0.75 pixels on MS image and 3 pixels on PAN image and LCM). These results imply that our algorithm can align images together to the accuracy less than those of the lowest resolution (here is MS image). For each initial registration error cases, the minimum RMEs of 1.718 (2.86 pixels in the LCM) meters for $\Delta x = 1.0$, 1.672 (2.79 pixels in the LCM) meters for $\Delta y = 1.0$, 1.730 (2.88 pixels in the LCM) meters for $\Delta s = 0\%$ and 1.704 (2.84 pixels in the LCM) meters for $\Delta\theta = -1^\circ$ occurs at $\beta = 0.75$. These results suggest that, if our algorithm converges, the larger value of β increases the accuracy of registration as well as the classification. However, for the cases of $\Delta x = 5.0$ and $\Delta y = 5.0$, our algorithm cannot register the MS image to the LCM since our algorithm is stuck in one of the local optima. The residue registration errors for $\Delta x = 5.0$ are 1.896 (3.16), 10.96 (18.3), 11.14 (18.6), and 11.41 (19.0) meters (pixels in the LCM), and for $\Delta y = 5.0$ are 1.827 (3.05), 1.834 (3.06), 3.133 (5.22) and 11.57 (19.3) meters (pixels in the LCM) for $\beta = 0.0, 0.25, 0.50$, and 0.75, respectively. Here, the initial displacement error corresponds to the RME of 20 pixels in the LCM. Such a large initial RME are only found when remote sensing images have significant different in spatial resolutions. LCMs derived from remote sensing image dataset with such a large scale difference are often unreliable and not often found in practice.

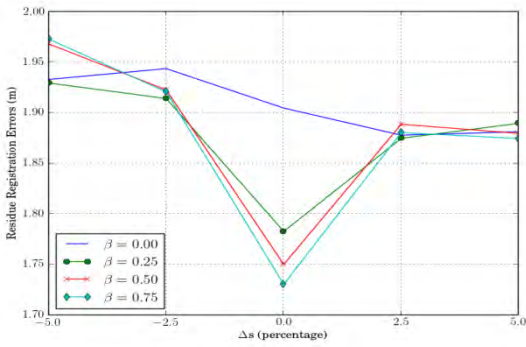
For performance comparison, we apply the normalized cross correlation method [16] to register PAN and MS images together and the resulting RME is equal to 1.836 meters or 3.06 pixels in the LCM. From Figures 2.15(a)-(d), we found that, with proper parameter selections and the initial registration errors, our proposed algorithm can achieved higher registration accuracy than those from the normalized cross correlation method. For example, our algorithm obtains the registration errors of 1.718 (2.86 pixels) meters for $\beta = 0.75$ and $\Delta x = 1$, 1.671 (2.79 pixels) meters for $\beta = 0.75$ and $\Delta y = 1$, or 1.702 (2.84 pixels)meters for $\beta = 0.75$ and $\Delta\theta = -1^\circ$.



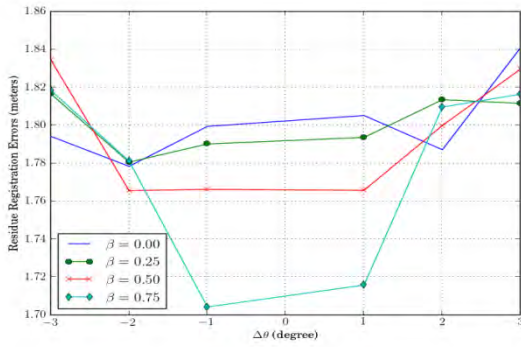
(a)



(b)



(c)



(d)

Figure 2.15: The effect of the initial registration errors to the residue registration error of our proposed algorithm in Example2

3. *Rice Cultivation Date Estimation Using the Cloudy-Multi-Temporal Moderate Resolution Imaging Spectroradiometer*

3.1 Introduction

As an agricultural country, Thailand has many vital industrial crops such as rice, oil palm, rubber tree and cassava. Among them, rice is the most important crop since it is a major exporting vegetation product of Thailand and it is also a staple food of many people around the world. Particularly, Thai HomMalee rice is also the federation of rice which is famous around the world. In Thailand, rice fields are located mostly in the central and north-east regions which are low-land and flat areas. Due to its low attitude and flatness, the rice cultivation in these areas can be easily damaged by natural disaster such as floods, draught and storm [48]. In order to evaluate the damage in rice fields, the cultivation date is a crucial factor that can be used to assess the compensation from those disasters, can be used in estimating crop progress as well as crop yield in an agricultural support system [49].

To achieve this goal, we need the monitoring system that can cheaply monitor the rice field area in the county-wide area. A good candidate for such the monitoring equipment is the remote sensing satellite due to its synoptic view and map-like format. Among the remote sensing satellites, the Moderate Resolution Imaging Spectroradiometer (MODIS) is most promising since its daily repeated cycle and free access to the raw and processed data. In literatures [50], [51], and [52] the time-series data of the normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI) are the most commonly used remote sensing data. They have been used extensively in vegetation-condition monitoring, land-cover classification and mapping, and environment-impact analysis. The EVI has a higher sensitive than that of the NDVI in high biomass areas, but the EVI have only 500-m resolution on MODIS imagery [53] which is too coarse to be used for most rice field in Thailand. The NDVI, on the other hand, can provide a higher resolution than EVI. However, it includes various noise components such as clouds, aerosols and bidirectional reflectance distribution factors. Therefore, noise reduction is important method before data will be used [54].

Time-series NDVI has been used for detecting phenology dates of corn and soybean. Zha *et al.* [49] proposed crop phenology date estimation method in the United States by comparing two techniques between Savitzky-Golay filter and the double logistic function to smooth MODIS NDVI data from daily surface reflectance data at 250 m resolution and then using first and second derivatives to extract date from smoothed data. They concluded that two techniques have their own advantages and their error result is less than 2 weeks.

The work by Kleynhans *et al.* in [55] used the extended Kalman filter and a triply modulated cosine function phase with the 8-days composited MODIS NDVI data at 500 m resolution to classify between settlement and natural vegetation in northern South Africa. The NDVI data are classified by its mean, amplitude and phase parameters. Then, the classifier used the mean and amplitude parameters to divide a remotely sensed image into the settlement and natural vegetation areas. The classification accuracy result from their approach is higher than those from the method using the Fourier transform. Kleynhas *et al.* also used the same

method and the same data to estimate the land cover changes in [56]. Here, the mean and amplitude parameters are used to characterize changes in term of a change index. Their result of change detection has an accuracy of 89% with a 13% false alarm rate.

In this study, the extended Kalman filter and a triply modulated cosine function with MODIS NDVI time-series data (8-days composited with 250 m resolution) are applied to estimate rice cultivation date. The triply modulate cosine function can be divided into 2 components, namely, the long-term trend and the periodic part [57]. Since rice cultivation changes from low NDVI to a high NDVI value from the growing period and return from high to low NVDI for harvestmen. Hence, the rice growing period for one cycle produces a NVDI signal that appears to be the inverted bell structure. Therefore, the cultivation time of rice should be appeared in the area where the NVDI signal changes from low to high values. In our work, we use the point where the season variation of the seasonal part from the triply modulate cosine function crosses a given threshold as the rice cultivate date.

3.2 Data Description

The study areas consists of Bangkok, Chachoengsao, Nonthaburi, Saraburi, Chainat, Chaiyaphum, Khonkaen and Ayutthaya provinces with the rice cultivation information obtained from the rice department of Thailand (Figure 3.1). Chainat and Khonkaen provinces are located in northeast part of Thailand and the others are located in the central area of Thailand. Northeast Thailand is arable for major rice. The rain-fed rice or major rice is the highest quality rice which can be cultivated only one time in a year. They are planted in rainy season and have crop cycle of about six months. Almost all irrigated rice or out of season rice are grown in the central of Thailand. They can be cultivated many times in a year with the cultivation period about four months. The irrigated rice had many characteristics such as two crop cycles in a year and three crop cycles in a year. Examples of the rice phenology are showed in Figure 3.2 for rain-fed and irrigated rice.

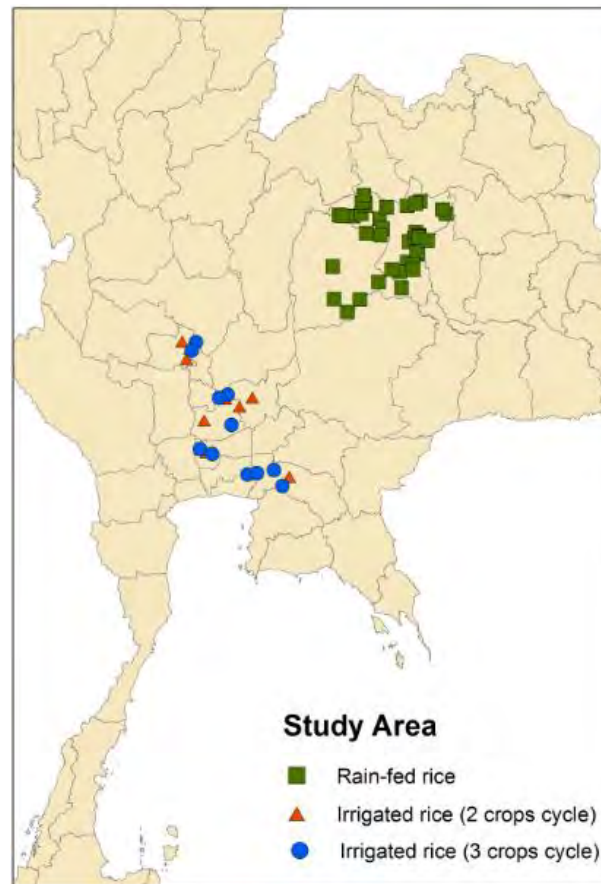


Figure 3.1: Study areas and rice coordinates relate to data from the National Rice Department.

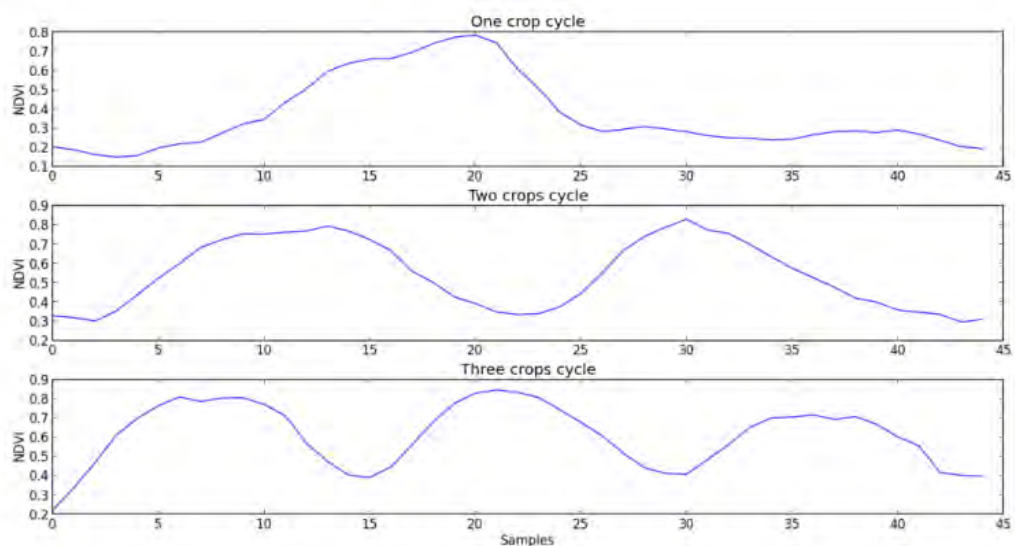


Figure 3.2: NDVI rice phenology (a) rain-fed rice; (b) two crops cycle in a year; and (c) three crops cycle in a year.

3.2.1 Ground-truth

The rice cultivation information provided by the rice department contains the UTM coordinates, province, planting date and harvest date of rice. This ground data are used to examine the accuracy of the proposed rice estimation date algorithm. The obtained data can be divided into 45 samples for rain-fed rice, 24 sample for two crops cycle and 15 samples for three crops cycle. In order to separate areas into rice and non-rice, we employ the K-mean algorithm to classify the MODIS NDVI time-series data into two classes since the NDVI pattern from rice and non-rice area is very different (Figure 3.3).

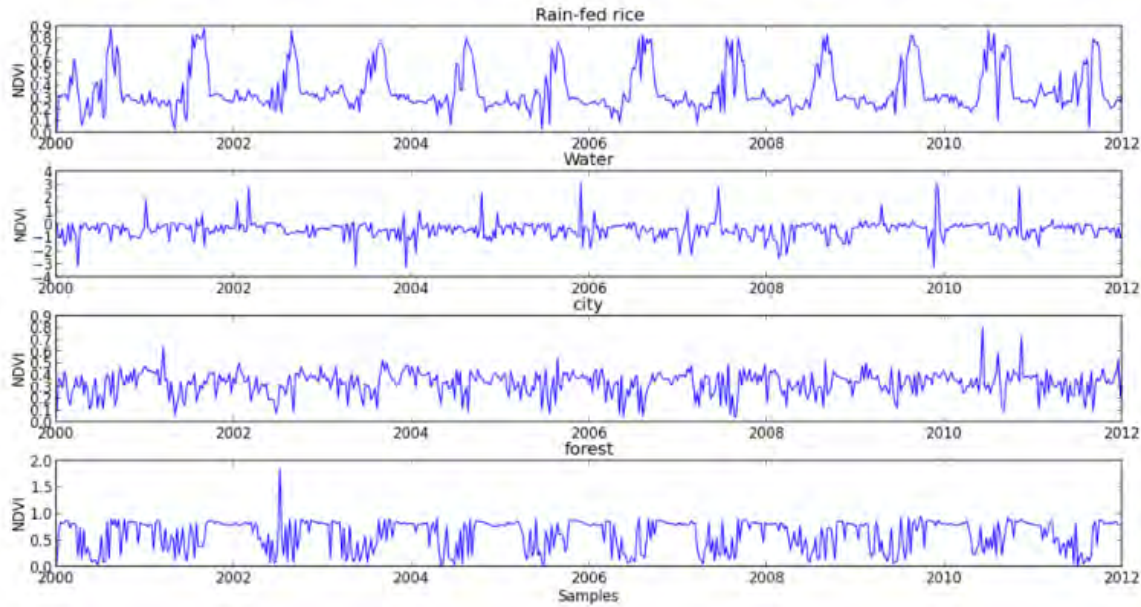


Figure 3.3: NDVI pattern; (a) Rain-fed rice; (b) Water;(c) City; and (d) Forrest.

3.2.2 NDVI Time-series Data

The 8-day MODIS/Terra data (MOD09Q1) acquired from NASA from 2000 to 2012 used in this analysis. It has two spectral bands for red and near-infrared spectral bands with a spatial resolution of 250 meters. The data were re-projected to the Geographic coordinate over Thailand. The NDVI is used to study the vegetation growth estimation is normalized transform of the near-infrared to red reflectance ratio follow as

$$NDVI = \frac{(NIR - RED)}{NIR + RED}$$

where *NIR* and *RED* are the digital numbers in near-infrared and red spectral bands, respectively. Since vegetation reflect the NIR color spectral better than the red color, the NDVI has a higher value if there is more vegetation, and NDVI is low when a scene does not have any vegetation. Usually, the NDVI data is between -1 and 1 where 1 and -1 indicates full vegetation and no vegetation, respectively. To create the time-series dataset for this 13 year

period; the NDVIs for every 8day MODIS scene were computed from the 8-day MODIS data. There are in total of 587 NDVI data for each pixel from Feb. 18, 2000 to Dec. 11, 2012.

3.3 Proposed Algorithm

3.3.1 A triply modulated cosine function

The NDVI time-series for a given pixel is modeled by a triply modulated cosine function [8][9] given as

$$y_k = \mu_k + \alpha_k \cos(\omega k + \varphi_k) + n_k \quad (3.1)$$

where y_k is the observed value of the NDVI time-series at time k and n_k is the noise sample at time k . Here, the cosine function at the time k is modeled to have the angular frequency of ω , mean of μ_k , amplitude of α_k , and phase of φ_k . The angular frequency can be explicitly computed as $\omega = 2\pi f$, where f is based on the annual vegetation growth cycle. Given the eight daily composite MODIS data, f is equal to 8/365 in rain-fed rice data, and 8/240 in irrigated rice data. The values of μ_k , α_k and φ_k are functions of time and must be estimated given y_k for $k \in 1, \dots, N$ where $N = 587$.

3.3.2 The Extended Kalman Filter

The estimation of triply modulated cosine function parameters is nontrivial and requires a nonlinear estimator. According to the EKF formulation, for every increment of k (the discrete time), a state vector x_k is defined containing the parameters to be estimated in the form $x_k = [\mu_k \ \alpha_k \ \varphi_k]^T$. The relation between x_k and x_{k-1} is given by a function $f(\cdot)$ which is a known and can be a nonlinear function. The state vector x_k is related to the observation vector y_k via a nonlinear measurement function $h(\cdot)$. Due to noisy nature of the measurement, both systems are corrupted with unknown noises, and they can be written as

$$x_k = f(x_{k-1}) + w_k, \quad (3.2)$$

and

$$y_k = h(x_k) + v_k \quad (3.3)$$

for the estimations of the state and observation vectors, respectively. The terms w_k and v_k are process and observation noises, respectively. The state vector parameters may be estimated over time k by recursive iteration based on the observation data y_k up to the time k . In the observation equation (3), y_k is the predicted measurement.

3.3.3 Cultivation Time Estimation Method

Interpretation of the term μ_k and $\alpha_k \cos(\omega k + \varphi_k)$ of a triply modulated cosine function can be given that μ_k can be considered as the trend variation and should be associated with long term vegetation changes such as the deforestation whereas the term $\alpha_k \cos(\omega k + \varphi_k)$ can be considered as the seasonal variation, and should be related to the growth of rice. As a result, we use only the term $\alpha_k \cos(\omega k + \varphi_k)$ in the estimation of cultivation time in this paper.

Before cultivation, the rice field is clear of vegetation as a result, the NDVI is low. As rice grow up, more vegetation cover the rice field and the NDVI increase. However, in certain area, farmer put down the some other crops such as green beans or grasses before cultivate rice to increase the Nitrogen in the soil. Hence, we identify the rice cultivation date as the first time instance that seasonal term $\alpha_k \cos(\omega k + \varphi_k)$ is greater than a predefined threshold τ after the lowest point.

3.4 Experimental Results

The initial state parameters of EKF method as well as the variances of observation and process noise are estimated based on known training data from the study areas. The training data were used all of time-series data at its pixel. The initial state parameters calculated using simple mathematic formula as

$$\mu_1 = \sum_{i=1}^N \frac{NDVI_i}{N} \quad (3.4)$$

and

$$\alpha_1 = \frac{\max(NDVI) - \min(NDVI)}{2} \quad (3.5)$$

where N is the number of train time-series and $NDVI_i$ is the NDVI component of the time-series i . The NDVI time-series data starts at Feb. 18, 2000 so the rain-fed rice should start in stage after harvesting and the irrigated rice should start in maturing stage from the observed. When comparing the characteristics of the cosine, the initial phase φ_1 parameter is assumed to be 120 degree for rain-fed rice and 0 degree for irrigated rice.

Brute-force algorithm was selected to find best case of noise in mean, amplitude and phase from 0.01 to 5 and observation noise from 0.1 to 50. Here, we fix the noise of mean, amplitude, phase and observed be equal to 0.1, 0.05, 1 and 20, respectively in rain-fed rice area and 0.1, 0.05, 100 and 40, respectively in irrigated rice area.

The NDVI time-series data are submitted to the EKF model and the resulting estimated NDVI data are plotted together with the raw NDVI data in Figure 3.4 for rain-fed rice, and in

Figure 3.5 for irrigated rice. The raw NDVI data appears to be very noisy due to cloud and aerosol in the atmosphere during the image capturing. The predicted NDVI data from the EKF appears to be less noisy, and smooth. The EKF-derived mean is an average of NDVI time-series. It represents the averaged vegetation coverage of the pixel of interest. Hence, the mean parameter cannot be used to identify the cultivation of rice. The other two remaining parameters, the amplitudes and phases are generated to the seasonal variation that related to the cultivations of rice. Here, we observe that the amplitude parameter fluctuates up and down corresponding to the cultivation and un-cultivation period of the rice. In this work, we choose the threshold τ to be -0.02 and -0.07 for rain-fed and irrigated rice areas, respectively. Figures 6 and 7 display the seasonal terms and the corresponding cultivation dates for the rain-fed and irrigated rice, respectively.

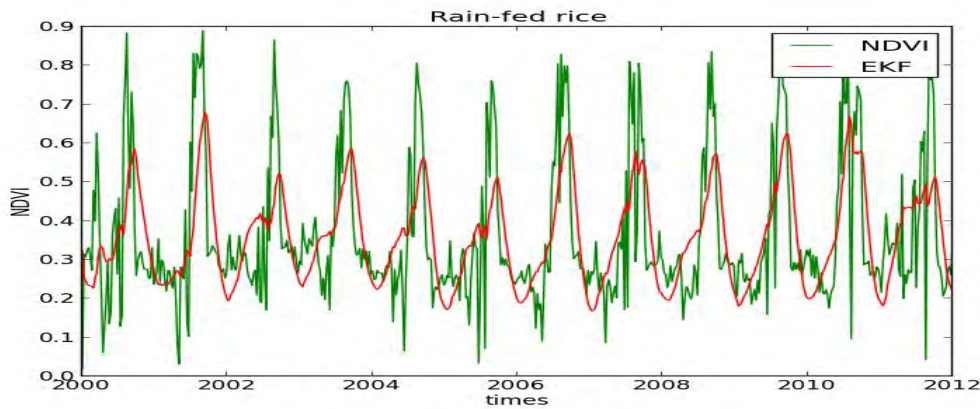


Figure 3.4: Comparing between the raw NDVI time-series data and EKF measurement data in rain-fed rice from 2000 to 2012.

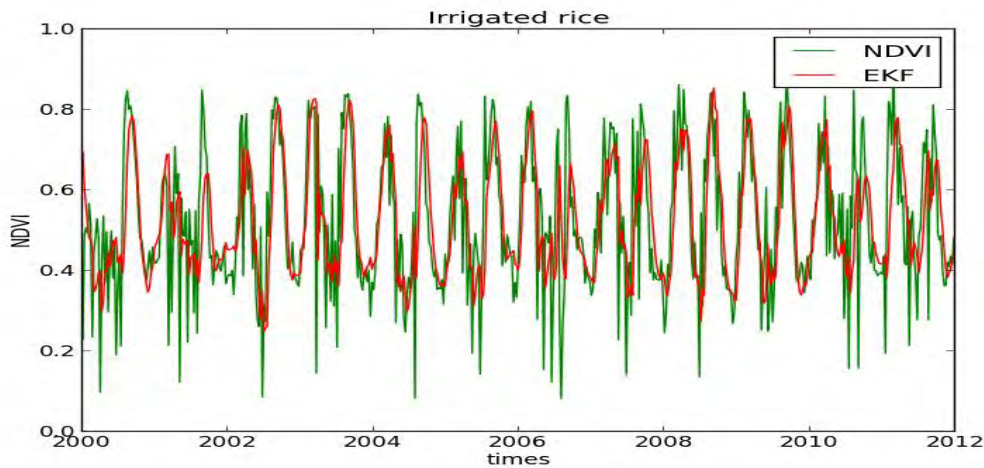


Figure 3.5: Comparing between the raw NDVI time-series data and EKF measurement data in irrigated rice from 2000 to 2012

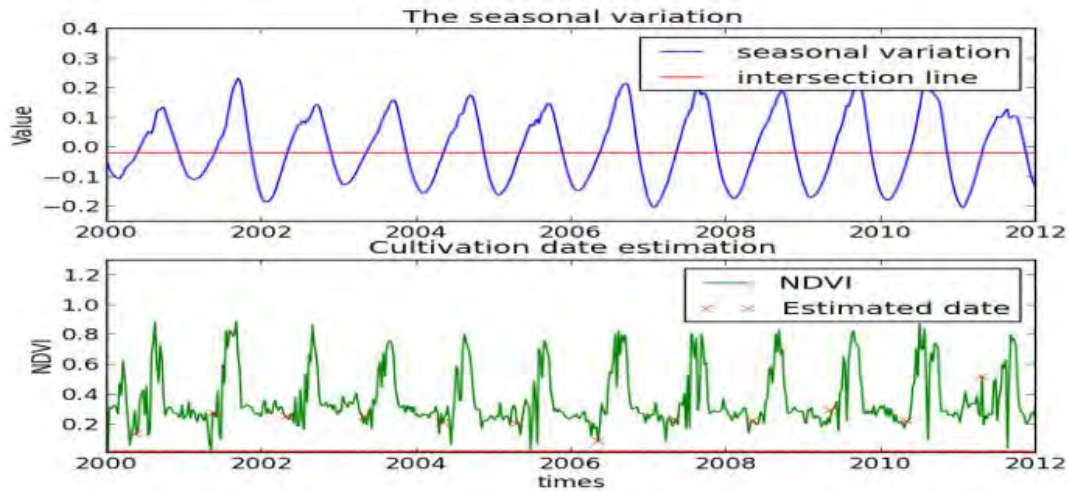


Figure 3.6: (top) Intersection of two lines between the seasonal variation and crossing line and (bottom) Rain-fed NDVI time-series with estimated cultivation dates

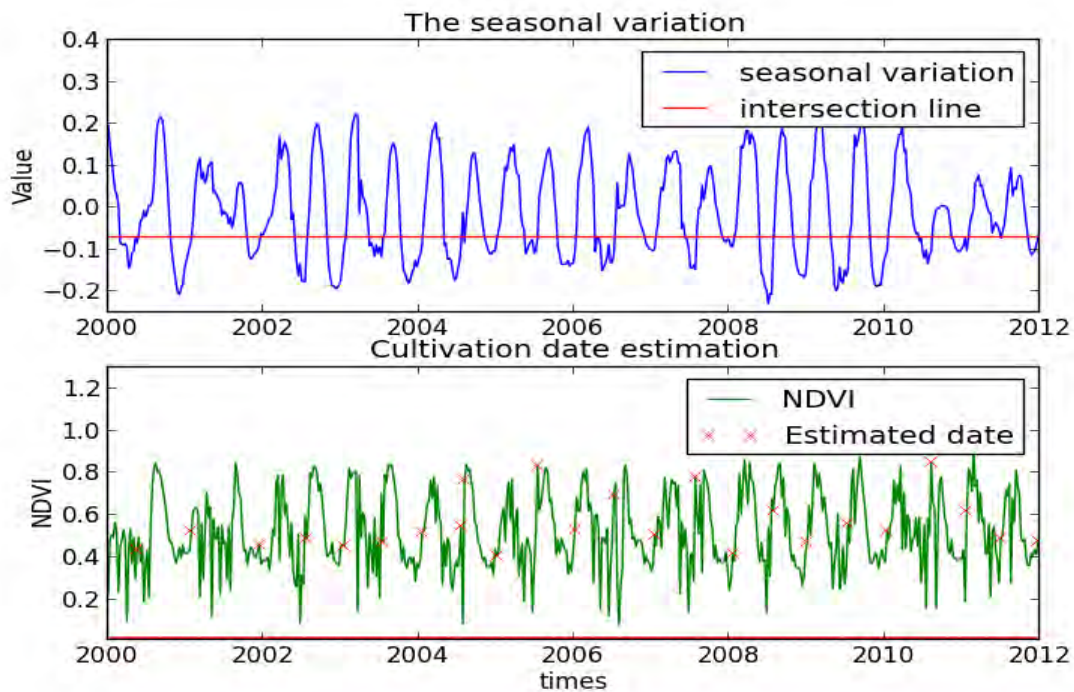


Figure 3.7: (top) Intersection of two lines between the seasonal variation and crossing line and (bottom) Irrigated NDVI time-series with estimated cultivation dates.

The results shown Table 3.1 is the estimation date error in terms of rice crop cycle. In single rice crop, the accuracy when errors are less than ± 8 and ± 16 equal 53.33 and 75.56 percent, respectively. The accuracy in double rice crop equal 54.17 and 83.33 percent. However, for area where rice is cultivated three times in a year, the estimation accuracy is low

accuracy because of NDVI time-series data had been rapidly changes and 8-day composite is too coarse to detect this rapid changes. .

Table 3.1: Error comparison between the ground truth and estimated rice cultivation date.

Rice crop cycle	Error ratio between ground-truth and estimation date		
	$\leq \pm 8days$	$\leq \pm 16days$	<i>Overall data</i>
Single rice crop	24 (53.33%)	34 (75.56%)	45 (100%)
Double rice crop	13 (54.17%)	20 (83.33%)	24 (100%)
Triple rice crop	1 (6.67%)	4 (26.67%)	15 (100%)

3.5 Discussion

The estimated rice cultivation date using the EKF based on a triply modulated cosine function can estimate the cultivation date in rain-fed rice and irrigated rice (two crops cycle) with reasonable accuracies. However, for irrigated rice which is cultivated three times per years, it has low estimation accuracies due to high cloud cover in the rainy season since clouds block the reflectance from the ground and result in incorrect NDVI data. Since the rice is cultivated three times per year, it has a high chance that the cultivation time is coincided with high cloud cover period. Moreover, since our data is eight-day composites, it cannot cope with the rapidly changes in the three crop-per-years rice.

In Figure 3.8, we display the cultivated dates for the rain-fed rice in June 2004 in Khonkaen provinces. We observe that the rain-fed rice in the areas near water sources was grown faster than other area. We also observe that the cultivation dates of neighboring proximity are likely to the same since all rain-fed rice will start cultivation after the first rain of the rainy season and the neighboring area are likely to have similar date for the first rain.

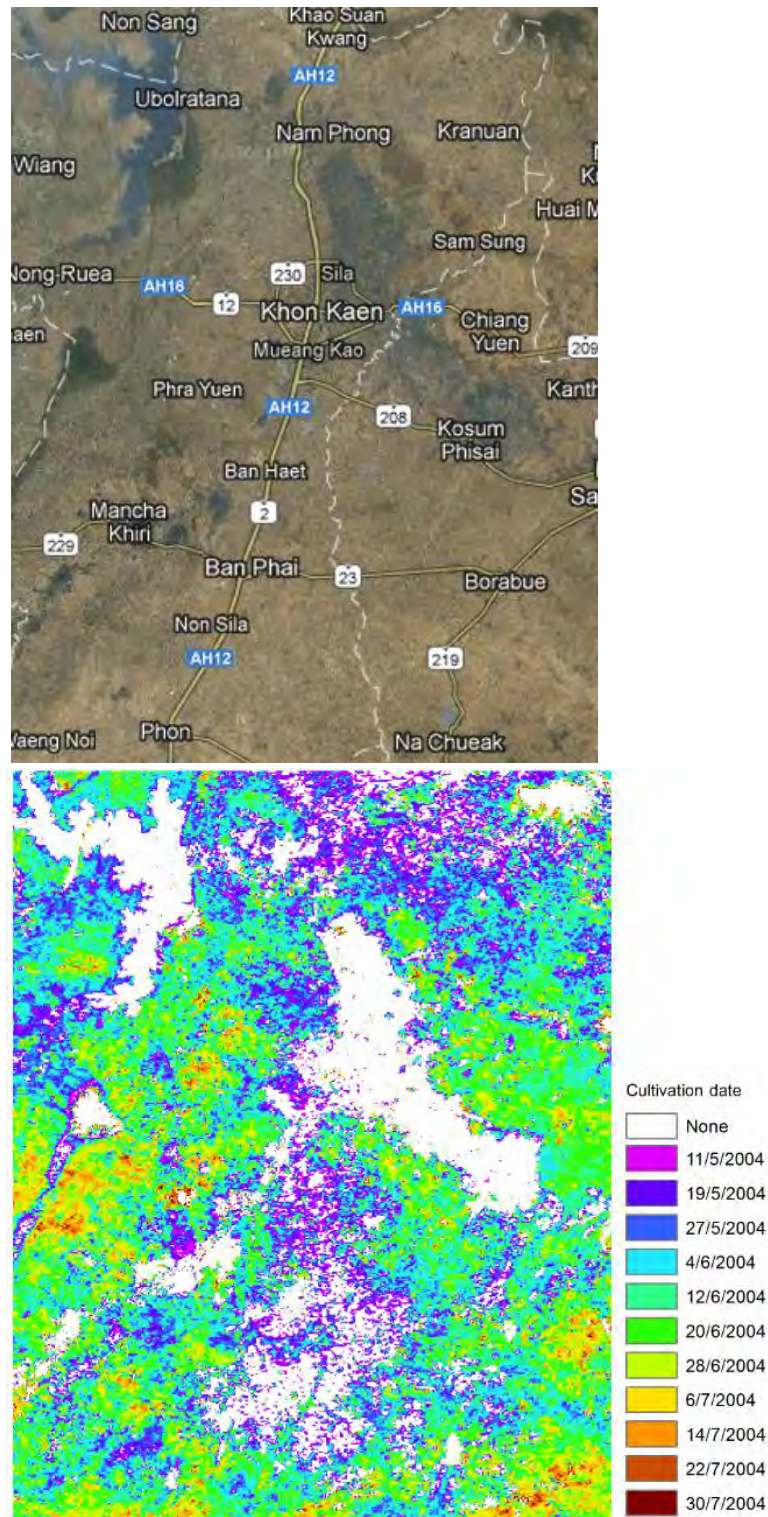


Figure 3.8: (top) Location of rain-fed rice in Khonkaen; and (bottom) Distribution of rain-fed rice cultivation date in Khonkaen and near province in 2004

4. References

- [1] D. Clausi and H. Deng, "Feature fusion for image texture segmentation," in *IEEE Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, vol. 1, pp. 580- 58, 2004.
- [2] X. Shi and R. Manduchi, "A study on Bayes feature fusion for image classification," in *Proceedings of the IEEE Workshop on Statistical Algorithms for Computer Vision*, Madison, Wisconsin., 2003.
- [3] W. J. Carper, T. M. Lillesand and R. W. Kiefer, "The use of intensity–hue–saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 5, p. 459–467, 1990.
- [4] P. S. Chavez and A. Y. Kwarteng, "Extracting spectral contrast in Landsat thematic mapper image data using selective component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, p. 339–348, 1989.
- [5] D. L. Civco, Y. Wang and J. A. Silander, "Characterizing forest ecosystems in Connecticut by integrating Landsat TM and SPOT panchromatic data," *Proc. ACSM/ASPRS Annu. Conv.*, vol. 2, p. 216–224, 1995.
- [6] R. Gomez, A. Jazaeri and M. Kafatos, "Wavelet-based hyperspectral and multi-spectral image fusion," *Proc. of SPIE*, vol. 4383, pp. 36-42, 2001.
- [7] M. Gonzalez-Audicana, J. L. Saleta, R. G. Catalan and R. Garcia, "Fusion of multi-spectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," *EEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 6, pp. 1291-1299, 2004.
- [8] K. Amolins, Y. Zhang and P. Dare, "Applications of wavelet transforms in image fusion," *Urban Remote Sensing Joint Event*, pp. 1-7, 2007.
- [9] Y. Zhang and M. He, "Multi-spectral and hyperspectral image fusion using 3-D wavelet transform," *Journal of Electronics*, vol. 24, no. 2, pp. 218-224, 2007.
- [10] Z. Liu, K. Tsukada, K. Hanasaki, Y. K. Ho and Y. P. Dai, "Image fusion by using steerable pyramid," *Pattern Recognition Letters*, vol. 22, no. 9, pp. 929-939, 2001.
- [11] V. K. Shettigare, "A generalized component substitution technique for spatial

- enhancement of multispectral images using a higher resolution data set," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 5, p. 561–567, 1992.
- [12] V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Processing*, vol. 13, no. 2, pp. 228-237, 2004.
- [13] Y. Xia and M. S. Kamel, "Novel cooperative neural fusion algorithms for image restoration and image fusion," *IEEE Trans. Image Processing*, vol. 16, no. 2, pp. 367-381, 2007.
- [14] R. C. Hardie, M. T. Eismann and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. on Image Processing*, vol. 13, no. 9, pp. 1174-1184, 2004.
- [15] Z. Li and H. Leung, "Fusion of multispectral and panchromatic images using a restoration-based method," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1482-1491, 2009.
- [16] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977-1000, 2003.
- [17] K. S. Arun, T. S. Huang and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *Trans. Pattern Anal. Machine Intell.*, vol. 95, pp. 698-700, Sep. 1987.
- [18] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 4, pp. 376-380, April 1991.
- [19] B. S. Reddy and B. N. Chatterji, "An FFT-Based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Processing*, vol. 5, no. 8, pp. 1266-1271, 1996.
- [20] H. Stone, B. Tao and M. McGuire, "Analysis of image registration noise due to rotationally dependent aliasing," *Journal of Visual Communication and Image Representation*, vol. 14, pp. 114-135, 2003.
- [21] P. A. Van Den Elsen, E. D. Pol, T. S. Sumanaweera, P. F. Her, S. Napel and J. R. Adler, "Grey value correlation techniques used for automatic matching of CT and MR brain and spine images. Visualization in Biomedical Computing," *Proc. SPIE*, vol. 235, 1994.

- [22] P. Thevenaz, U. E. Ruttimann and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans. Image Processing*, vol. 7, no. 1, pp. 27-41, Jan. 1998.
- [23] W. L. S. Costa, D. R. Haynor, R. M. Haralick, T. K. Lwellen and M. M. Graham, "A maximum-likelihood approach to PET emission attenuation image registration," *IEEE Nuclear Science Symposium and Med. Imag. Conference.*, pp. 1139-1143, Oct-Nov 1993.
- [24] S. Zheng, W. Z. Shi, J. Liu, G. X. Zhu and J. W. Tian, "Multisource image fusion method using support value transform," *EEE Trans. Image Processing*, vol. 16, no. 7, pp. 1831-1839, July 2007.
- [25] S. Chen, Q. Guo, L. H. and B. E., "A Maximum Likelihood Approach to Joint Image Registration and Fusion," *IEEE Transactions on Image Processing.*, vol. 20, no. 5, pp. 363-1372., 2011.
- [26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewoods Cliffs, NJ: Prentice-Hall, 1993.
- [27] P. Diaconis and L. Salo-Coste, "What do we know about the Metropolis Algorithm?," *J. Comp. and Syst. Science*, vol. 57, pp. 20-36, 1998.
- [28] D. S. Y. Mahapatra, "Integrating Segmentation Information for Improved MRF-Based Elastic Image Registration," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 170-183, Jan. 2012.
- [29] G. Winkle, *Image Analysis Random Fields and Dynamic Monte Carlo Methods*, New York: Springer-Verlag, 1995.
- [30] P. Bremaud, *Markov Chains Gibbs Field, Monte Carlo Simulation, and Queues*, New York: Springer-Verlag, 1999.
- [31] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.
- [32] A. Solberg, T. T. and A. Jain, "A Markov Random Field Model for Classification of Multisource Satellite Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, pp. 100-113, 1996.

- [33] T. Kasetkasem and P. Varshney, "An Image Change Detection Algorithm Based on Markov Random Field Models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 1815-1823, 2002.
- [34] L. Bruzzone and D. Prieto, "Automatic Analysis of the Difference Image for Unsupervised Change Detection," *IEEE Trans. Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1171-82, 2002.
- [35] H. Xie, L. Pierce and F. Ulaby, "SAR Speckle Reduction Using Wavelet Denoising and Markov Random Field Modeling," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, pp. 2196-2212, 2002.
- [36] B. Tso and P. Mather, "Classification of Multisource Remote Sensing Imagery Using a Genetic Algorithm and Markov Random Fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, pp. 1255-1260, 1999.
- [37] R. Nishii, "A Markov Random Field-Based Approach to Decision-Level Fusion for Remote Sensing Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, pp. 2316-2319, 2003.
- [38] A. Sarkar, M. Biswas, B. Kartikeyan, K. L. VMajumder and D. Pal, "A MRF Model-Based Segmentation Approach to Classification for Multispectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 1102-1113, 2002.
- [39] r. A. Dempste and N. R. D. Laird, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Soc. Stat.. ser. B.*, pp. 1-38, 1997.
- [40] T. Tanaka, "Mean Field Theory of Boltzmann Machine Learning," *Phys. Rev. E*, vol. 58, p. 2302-2310, 1998.
- [41] T. Tanaka, "A Theory of Mean Field Approximation," in *12th Annual Conference on Neural Information Processing Systems (NIPS'98)*, Denver, Dec. 1998.
- [42] J. Zhang, "The Mean Field Theory in EM Procedures for Markov Random Fields," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2570-83, 1992.
- [43] I. P. Bicke and K. Doksum, *Mathematical Statistics*, New Jersey: Prentice Hall, 1977.
- [44] H. Van Trees, *Detection, Estimation, and Modulation Theory*, New York: Wiley, 1968.
- [45] P. Varshney, *Distributed Detection and Data Fusion*, New York: Springer, 1997.

- [46] J. E. R. Kennedy, "Particle Swarm Optimization," *Proceedings of IEEE International Conference on Neural Networks IV*, pp. 1942-1948, 1995.
- [47] I. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Image Processing Letters*, vol. 85, pp. 317-325, 2003.
- [48] L. Tingting and L. Chuang, "Study on extraction of crop information using time-series MODIS data in the Chao Phraya Basin of Thailand," *Advances in Space Research*, vol. 45, no. 6, p. 2010, 775-784.
- [49] H. Zhao, Z. Yang, L. Li and H. Zhu, "Crop phenology date estimation based on NDVI derived from the reconstructed MODIS daily surface reflectance data," in *17th Intl. Conf. on Geoinformatics (Geoinformatics 2009)*, Fairfax, USA, August 12-14, 2009.
- [50] E. d. Silveira, L. d. Carvalho, F. Acerbi and J. d. Mello, "The assessment of vegetation seasonal dynamics using multitemporal NDVI and EVI images derived from MODIS," *Analysis of Multi-temporal Remote Sensing Images*, vol. 20, pp. 1-5, 2007.
- [51] F. Deng, G. Su and C. Liu, "Seasonal variation of MODIS vegetation indexes and their statistical relationship with climate over the subtropic evergreen forest in Zhejiang, China," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, p. April 2007, 236-240.
- [52] T. Sakamoto, N. V. Nguyen, H. Ohno, N. Ishitsuka and M. Yokozawa, "Spatio-temporal distribution of rice phenology and cropping systems in the Mekong Delta with special reference to the seasonal water flow of the Mekong and Bassac rivers," *Remote Sensing of Environment*, vol. 100, pp. 1-6, 2006.
- [53] T. Sakamoto, M. Yokozawa, H. Toritani, M. shibayama, N. Ishitsuka and H. Ohno, "A crop phenology detection method using time-series MODIS data," *Remote Sensing of Environment*, vol. 96, pp. 366-374, 2005.
- [54] C. F. Chen, N. T. Son and L. Y. Change, "Monitoring of rice cropping intensity in the upper Mekong Delta, Vietnam using time-series MODIS data," *Advances in Space Research*, vol. 49, no. 2, p. 292-301, Jan. 2012.
- [55] W. Kleynhans, J. C. Olivier, K. J. Wessels, F. v. d. Bergh, B. P. Salmon and K. C. steenkamp, "Improving land cover class separation using an extended Kalman filter on MODIS NDVI time-series data," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 381-

385, Apr. 2010.

- [56] W. Kleynhans, J. C. Olivier, K. J. Wessels, F. v. d. Bergh, B. P. Salmon and K. C. steenkamp, "Detecting land cover change using an extended Kalman filter on MODIS NDVI time-series data," *IEEE Geosci. Remote Sens. Lett.*, vol. 83, pp. 507-511, May. 2011.
- [57] J. W. Davis and M. A. Keck, "Modeling Pesestrain Behavior and Detecting Event anomalies using a Seasonal kalman Filter," *Computer Science and Engineering Technical Report*, pp. 1-8, 2005.

Outputs

1. T. Sritarapipat, **T. Kasetkasem**, and P. Rakwatin, “Fusion and Registration of THEOS Multispectral and Panchromatic Images,” under review at the *International Journal of Remote Sensing*
2. **T. Kasetkasem**, P. Rakwating, R. Sirirsommai and A. Eiumnoh, “a Joing Land Cover Mapping and Image Registration Algorithm Based on a Markov Random Field Model,” under review at *Remote Sensing*
3. **T. Kasetkasem**, P. Rakwatin, R. Sirisommai, A. Euimnoh, and T. Isshiki, “A MRF-Based Approach for a Multisensor Land Cover Mapping of Mis-Registered Images” *The 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS2012)*, July 22-27, Munich, Germany

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

T. Kasetkasem¹, P. Rakwatin², R. Sirisommai³, A. Eiumnoh⁴, and T. Isshiki⁵

¹Faculty of Engineering, Kasetsart University, Bangkok Thailand 10900, fengtsk@ku.ac.th

²GISTDA, Bangkok Thailand 10210, preesan@gistda.or.th

³ICTES Program, Kasetsart University, Bangkok Thailand 10900, ratchawits@gmail.com

⁴BIOTECH, NTSDA, Pathumthani 12120 THAILAND, apisit.eiu@biotec.or.th

⁵Dept. Communications and Integrated Systems, Tokyo Institute of Technology
Tokyo 152-8550 Japan, isschiki@vlsi.ss.titech.ac.jp

ABSTRACT

The traditional land cover mapping (LCM) algorithms assume that images are perfectly registered. In practice, this assumption may not always be valid since these images may be acquired from different sensor platforms, or at different time which may suffer small variations in platform flight paths. As a result, it is imperative to incorporate the registration error into the land cover mapping algorithm. In this paper, we propose a joint LCM and image registration algorithm under the Markov random field model. Here, the expectation-maximization algorithm is employed to search for the optimum LCM as well as the map parameters. Our result shows that the proposed MRF-Based approach can increase the accuracies of the classification maps as well as the map parameter estimation.

Index Terms— Remote sensing, joint land cover mapping and registration, Markov random fields, EM algorithm

1. INTRODUCTION

A Land cover map is an important application of remote sensing data. A number of image classification algorithms [1] have been developed to extract this information from a variety of remote sensing data. The performance of these classification algorithms depends on the quality of the input data, i.e., the classification accuracy is improved if features from different classes become more separable. One approach to increase the separation of classes is to use multi-temporal or multi-sensor data. Reports in [2, 3, 4, 5, 6] has shown that the accuracy of the resulting land cover map can be greatly increased when multi-sensor data are combined. The common assumption in these papers [2, 3, 4, 5, 6] is that images are perfectly registered, i.e., the corresponding pixels from different images occupy the identical area in the actual scene. In practice, this assumption may not always be valid since these images may be acquired from different sensor platforms. Even with images from the same platform, this assumption of the perfect registration can still be violated if images are acquired at different times due to the small variations of the platform flight path. As a result, it is imperative to incorporate the registration error into the land cover map algorithm.

As a result, in this paper, we consider the problem of land cover mapping of mis-registered images. Here, we model that the observed multispectral images are a remapped version of the registered images where the remap parameters are partially known, i.e., there are a

small errors in the remap parameters. Similar idea can be found in the work by Chen *et al.* [7] where the joint image fusion and registration was considered. In their paper, observed images are the remapped version of the original image with possibilities of polarity reverse and/or DC offset. However, their goal is different from us since their goal is to estimate the original image whereas the problem considered in this paper is to estimate the LCM.

Since the LCM is more likely to occur into patches of land cover classes, the Markov random field (MRF) is employed as the model of the LCM where the optimum LCM is chosen based on the maximum *a posteriori* (MAP) criteria. Under the perfect registration scenario, the resulting LCM has the minimum probability of error among all other mapping algorithms. However, if the images are not perfectly aligned the resulting probability of error is not optimum. To improve the mapping accuracy, the expectation-maximization (EM) algorithm is used as the parameter estimator. For every iteration of the EM algorithm, a new set of the map parameters are computed and the *a posteriori* probability is approximated. Our result shows that the accuracy in terms of both the percentage of correctly classified pixels and mis-registration errors can be significantly improved.

2. PROBLEM STATEMENT

Let S be a set of sites (pixels) and $\Lambda \in \{0, 1, \dots, L-1\}$ be the phase space (class labels). Furthermore, let $X \in \Lambda^S$ be the LCM of the size $M \times N$ pixels. Note that L is the number of land cover classes of interest. The LCM is assumed to satisfy the MRF properties with a Gibbs potential $V_C(X)$ [8]. Hence, we can write the marginal probability density function (PDF) of the LCM as

$$\Pr(X) = \frac{1}{Z_X} \exp \left(- \sum_{C \in S} V_C(X) \right) \quad (1)$$

where Z_X is a normalizing constant. Note that X is a realization of a LCM, C is a clique, and $\sum_{C \in S} V_C(X)$ is called the Gibbs energy function. A clique is a singleton or any subsets of whose two distinct elements are mutual neighbors. The popular model of the Gibbs energy function is the Ising model [8], given by,

$$V_{\{s,r\}}(X) = \begin{cases} -\beta; & \text{if } x_s = x_r \text{ and } r \in N_s \\ \beta; & \text{if } x_s \neq x_r \text{ and } r \in N_s \\ 0; & r \notin N_s \end{cases} \quad (2)$$

where N_s is a set of a neighboring site of s and β is a non-negative constant. In this paper, this model is also applied to describe the LCM because this model promotes the LCM with connected land cover classes. Let $Y_1 \in \mathcal{R}^{S \times B_1}$ and $Y_2 \in \mathcal{R}^{S \times B_2}$ be the observed registered images of the scene S capturing by two different sensors or the same sensor at two different time where B_1 and B_2 are the number of spectral bands of Y_1 and Y_2 , respectively. Here, we assume that observations from Y_1 and Y_2 are statistically independent given the LCM, i.e.,

$$p(Y_1, Y_2 | X) = p(Y_1 | X) p(Y_2 | X) \quad (3)$$

where p denotes the PDF. Furthermore, for a given LCM, the observations of Y_1 and Y_2 from different site (pixels) are assumed to be statistically independent, i.e.,

$$p(Y_a | X) = \prod_{s \in S} p(\mathbf{y}_{a,s} | x_s) \quad (4)$$

where $a \in \{1, 2\}$ is an index indicating the image of interest, and $\mathbf{y}_{a,s} \in \mathcal{R}^{B_a}$ and x_s are the observed vector of an image Y_a and the class label at a site s , respectively.

Due to imperfections in the image acquisition and capturing process, both images are not perfectly aligned to the LCM. Here, we denote θ_a as the unknown map parameters for Y_a over X . In other word, if θ_a is known, the observed image can therefore be registered with the LCM. Hence, the conditional PDF of the mis-registered Y_a given X and the mapping parameter is given by

$$p(Y_a | X, \theta_a) = \prod_{s \in S} p(\mathbf{y}_{a,s}(\theta_a) | x_s). \quad (5)$$

3. OPTIMUM SOLUTION AND PROPOSED ALGORITHM

For given map parameters, the LCM can be obtained under the maximum *a posteriori* (MAP) criteria as

$$X^{\text{opt}} = \arg \left\{ \max_X [\Pr(X | Y_1, Y_2, \theta_1, \theta_2)] \right\}. \quad (6)$$

where $\Pr(X | Y_1, Y_2, \theta_1, \theta_2)$ is the posterior probability of the LCM given observed images and map parameters. The resulting probability of error is minimum among all other classifiers [9]. By using the definition of conditional probability, eq. (6) can be rewritten as

$$X^{\text{opt}} = \arg \left\{ \max_X \left[\frac{p(Y_1, Y_2 | X, \theta_1, \theta_2) \Pr(X)}{p(Y_1, Y_2 | \theta_1, \theta_2)} \right] \right\}. \quad (7)$$

Since $p(Y_1, Y_2 | \theta_1, \theta_2)$ is independent of the choice of X , eq. (7) reduces to

$$X^{\text{opt}} = \arg \left\{ \max_X [p(Y_1, Y_2 | X, \theta_1, \theta_2) \Pr(X)] \right\} \quad (8)$$

Substituting eq. (1) and eq. (5) into eq.(8) the optimization problem becomes

$$X^{\text{opt}} = \arg \left\{ \min_X [E(\theta_1, \theta_2)] \right\}. \quad (9)$$

where

$$E(\theta_1, \theta_2) = E_1(\theta_1) + E_2(\theta_2) + \sum_{C \in S} V_C(X) \quad (10)$$

is called the energy function of a LCM given observations and map parameters, and $E_a(\theta_a) = \sum_{s \in S} \ln [p(\mathbf{y}_{a,s}(\theta_a) | x_s)]$ for $a \in \{1, 2\}$.

In general, $E(\theta_1, \theta_2)$ is a non-convex function and, therefore, conventional optimization algorithms cannot be applied to solve eq. (9). Furthermore, the number of possible LCMs is extremely large. Therefore, we approximate eq. (10) by using the mean field theory [10]. Basically, for a pixel s and a clique of type $C = \{s, r\}$, the potential function $V_{\{s,r\}}(x_s, x_r)$ is replaced by its expected value $\langle V_{\{s,r\}}(x_s, x_r) \rangle_{x_r | Y_1, Y_2, \theta_1, \theta_2}$ where $\langle F(x) \rangle_x$ denotes the expected value F of over a random variable x . From the above equation, the energy function defined in eq. (10) can be approximated as

$$E(\theta_1, \theta_2) \approx \sum_{s \in S} w_s(\theta_1, \theta_2) \quad (11)$$

where

$$w_s(\theta_1, \theta_2) = w_s^1(\theta_1) + w_s^2(\theta_2) + \sum_{r \in N_s} \langle V_{\{s,r\}}(x_s, x_r) \rangle_{x(r)},$$

and

$$w_s^a(\theta_a) = \ln [p(\mathbf{y}_{a,s}(\theta_a) | x_s)] \text{ for } a \in \{1, 2\}$$

Since $w_s(\theta_1, \theta_2)$ depends only on x_s , the optimization problem of X reduces to search for the individual x_s that minimizes $w_s(\theta_1, \theta_2)$. Hence, the optimum solution can be easily obtained. Furthermore, we observe that the right hand side of eq. (11) can be written as the addition of the configurations of LCM from an individual pixel. As a result, the posterior probability of a configuration at a pixel s , given the mis-registered observations is approximated as

$$\Pr(x_s | Y_1, Y_2, \theta_1, \theta_2) \approx B e^{-w_s(\theta_1, \theta_2)} \quad (12)$$

where B is a normalizing constant.

For the parameter estimation, the EM algorithm [11] is employed in this paper. The EM algorithm seeks the solution of the MLE by iteratively performs two steps. First, the EM algorithm find the expected value of the log-likelihood function of the observed images and the LCM given the map parameters over the conditional PDF of a LCM given observed images and the current estimated parameters, i.e., the EM algorithm computes the following equation,

$$Q(\theta_1, \theta_2 | \theta_1^t, \theta_2^t) = \langle \ln(p(Y_1, Y_2, X | \theta_1, \theta_2)) \rangle_{X | Y_1, Y_2, \theta_1^t, \theta_2^t} \quad (13)$$

This step is called the *E-step*. We observe that the computation of the expected values in the E-step requires the posterior probability of X given the observations and the estimated map parameters. By using the result, in eq. (4), eq. (5) and eq. (12), the expected value can be approximated as

$$Q(\theta_1, \theta_2 | \theta_1^t, \theta_2^t) \approx D \sum_{s \in S} q_s(\theta_1, \theta_2) e^{w_s(\theta_1^t, \theta_2^t)} \quad (14)$$

where $D = \langle \Pr(X) \rangle_{X | Y_1, Y_2, \theta_1^t, \theta_2^t}$ does not depend on θ_1 and θ_2 , and $q_s(\theta_1, \theta_2) = \ln [p(\mathbf{y}_{1,s}(\theta_1), \mathbf{y}_{2,s}(\theta_2) | x_s)]$.

After the E-step, the EM algorithm estimates a new set of the map parameters by maximizing eq. (13), i.e.,

$$(\theta_1^{t+1}, \theta_2^{t+1}) = \max_{(\theta_1, \theta_2)} [Q(\theta_1, \theta_2 | \theta_1^t, \theta_2^t)]. \quad (15)$$

The second step, here, is called the *M-step*. By using the approximation in eq. (14), eq. (15) becomes

$$\begin{aligned} & (\theta_1^{t+1}, \theta_2^{t+1}) \\ &= \max_{(\theta_1, \theta_2)} \left[\sum_{s \in S} \sum_{l=0}^{L-1} \ln [p(\mathbf{y}_{1,s}(\theta_1), \mathbf{y}_{2,s}(\theta_2) | x_s = l)] \right] \times e^{-w_s(\theta_1^t, \theta_2^t)} \end{aligned} \quad (16)$$

By combining the result for eq. (9) and eq. (16), our algorithm can be summarized as follows.

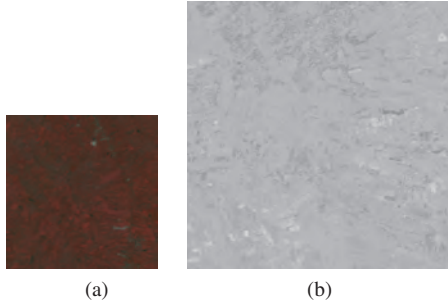


Fig. 1: Observed images; a) SMMS image and b) PALSAR Image

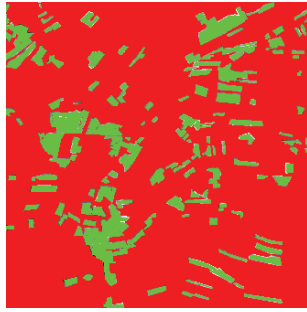


Fig. 2: Ground Data (green: sugarcane and red: others)

1. Set $\Pr(x_s = l | Y_1, Y_2, \theta_1, \theta_2) = \frac{1}{L}$, set θ_1^0 and θ_2^0 to some initial values, and let $t = 0$.
2. Compute the expected value of potential functions $V_{\{s,r\}}(x_s, x_r)$ or $\langle V_{\{s,r\}}(x_s, x_r) \rangle_{x_r | Y_1, Y_2, \theta_1, \theta_2}$.
3. Approximate a posterior probability using eq. (12) and choose the value of B such that the total probability is one.
4. Estimate a new set of mis-registration parameters using eq. (16).
5. Set $t = t + 1$. If t is greater than some predefined values, terminate the algorithm and compute the LCM using eq. (9) with the energy function defined in eq. (11). Otherwise go to Step 2.

4. EXPERIMENTAL RESULT

In this example, we use one optical image of size 335×332 pixels (Fig. (1a)) acquired by CCD sensor on board the small multi-mission satellite (SMMS) and one SAR image of size 670×665 pixels (Fig. (1b)) acquired by PALSAR. Both images covered the part of Sakaew province in Thailand. The SMMS image was taken on Nov 27th, 2010 while PALSAR image was taken on Aug 28th, 2010. In this example, only two classes are considered, namely, sugarcane and others, and the ground truth image is shown in (2) where the green and red colors are associated with sugarcane and others, respectively.

To test the performance of our algorithm, we assume that the PALSAR image is perfectly registered to the LCM. Only the map parameter of the SMMS image is unknown. From our dataset, we manually register the SMMS image with the PALSAR image to ob-

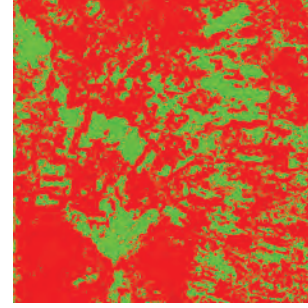


Fig. 3: The LCM without registration error correction

tain the map parameter which can be written in a matrix form as

$$\theta_2 = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix}$$

where r_{11} and r_{22} is a scale factors, r_{21} and r_{12} are the skewness, r_{13} and r_{23} are displacement, in column and row directions, respectively. The correct values are $r_{11} = r_{22} = 0.5$, and $r_{12} = r_{21} = r_{13} = r_{23} = 0$. Next, we modify the map parameters to new values of $r_{11} = 0.51$, $r_{22} = 0.49$, $r_{12} = r_{21} = 0$, $r_{13} = -2$, and $r_{23} = 2$ which results in the mis-registration of 6.7 pixels on average. Hence, the incorrect map parameter is given by

$$M = \begin{bmatrix} 0.51 & 0 & -2.0 \\ 0 & 0.49 & 2.0 \end{bmatrix}$$

Then, the LCM are obtained directly from the mis-registered image pair (Fig. 3), and we found that the initial LCM has many isolated pixels and the percentage of the correct classification is equal to 73.28%.

Next, the proposed algorithm is applied to the above dataset. The whole process is implemented using CUDA on NVIDIA GeForce GTS 450 with 1 GB memory. Each iteration takes between 3 and 4.5 seconds to perform. Here, the value of β is set to be between 0 and 1 (see eq. (2)), and the algorithm terminates when the number of iteration, t , is greater than 500 and the resulting LCMs are given in Fig. 4 for $\beta = 0, 0.3, 0.6$ and 1.0 , respectively. From eq. (2), the value of β controls the probability that the configurations of neighboring pixels are the same. In other words, the configurations of neighboring pixels are more likely to the same as the value of β increases. By comparing the resulting LCM for $\beta = 0$ and $\beta = 1.0$, we observe that the LCM for $\beta = 1.0$ is more connected whereas, for $\beta = 0$, there are a large number of isolated pixels in the LCM. Note here that $\beta = 0$ implies there is no interaction among neighboring pixels. The percentages of correct classification are equal to 74.09% (+0.81%), 77.49% (+4.21%), 78.17% (+4.89%) and 78.47% (+5.19%) for $\beta = 0, 0.3, 0.6$ and 1.0 , respectively. We note here that the number inside parenthesis is the improvement in percentage from the initial LCM. Again, these results further emphasize that the MRF model increases the performance of the classification process.

Another key performance of our algorithm is the accuracy of the estimated parameters. Here, if the algorithm performs perfectly, the original scale factor and the displacement should be obtained without additional modifications in skewness. Table 1 displays the estimated parameters, and the corresponding registration errors are 1.566, 1.013, 1.377, and 2.1667 pixels for $\beta = 0, 0.3, 0.6$ and 1.0 , respectively. We observe that there exists an optimum value of β that results in the minimum registration error. Furthermore, we also plot

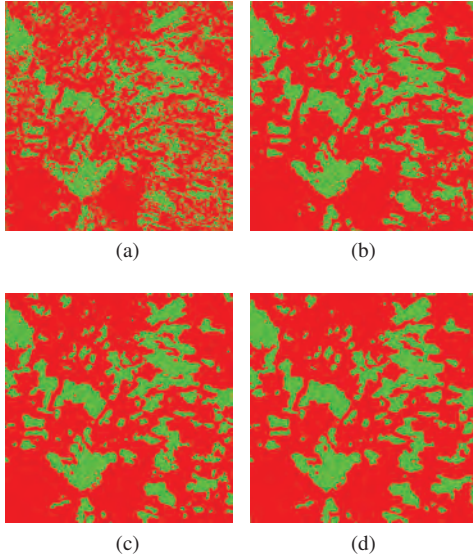


Fig. 4: The resulting LCM for the value of β equal to a) 0.0; b) 0.3; c) 0.6; and d) 1.0

β	0.0	0.3	0.6	1.0	Actual Value
r_{11}	0.4984	0.500	0.500	0.501	0.5
r_{12}	0.001	0.000	0.000	0.001	0.0
r_{21}	0.001	0.001	0.001	0.001	0.0
r_{22}	0.498	0.498	0.497	0.496	0.5
r_{13}	0.241	-0.301	-0.354	0.989	0.0
r_{23}	0.308	0.288	0.466	0.775	0.0
Error (pixels)	1.566	1.013	1.370	2.166	-

Table 1: The resulting estimated map parameters and mis-registration errors

the mis-registration error as the function of iterations number for all cases in Fig.5. We observed that higher value of β results in a slower convergence rate of the estimation of the map parameters.

5. SUMMARY

In this paper, a joint image classification and registration based on the MRF model is proposed. It is assumed that the observed images are remapped version of the registered image to the LCM. The results show that our proposed algorithm can successfully register images as well as produce more accurate LCMs.

6. ACKNOWLEDGMENT

This work is supported in part by the Thailand Research Fund under grant number RSA5480031, Thailand Advanced Institute of Science and Technology (TAIST), National Science and Technology Development Agency (NSTDA), Tokyo Institute of Technology, and Kasetsart University.

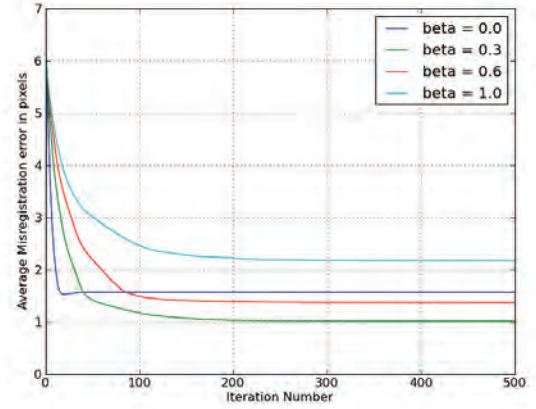


Fig. 5: The mis-registration errors as a function of the number of iterations for $\beta = 0.0, 0.3, 0.6, 1.0$.

7. REFERENCES

- [1] P. Mather, *Computer processing of remotely-sensed images: an introduction*. Wiley, 1999.
- [2] A. Bogdanov, S. Sandven, O. Johannessen, V. Alexandrov, , and L. Bobylev, "Multisensor approach to automated classification of sea ice image data," *IEEE Trans. on Geosci. and Remote Sensing*, vol. 43, pp. 1648–1664, 2005.
- [3] N. Ghoggali and F. Melgani, "Genetic svm approach to semisupervised multitemporal classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, pp. 212–216, 2008.
- [4] J. Munoz-Marf, L. Bruzzone, and G. Camps-Vails, "A support vector domain description approach to supervised classification of remote sensing images," *IEEE Trans. on Geosci. and Remote Sensing*, vol. 45, pp. 2683 – 2692, 2007.
- [5] A. Solberg, T. Taxt, and A. Jain, "A markov random field model for classification of multisource satellite imagery," *IEEE Trans. on Geosci. and Remote Sensing*, vol. 34, pp. 100–113, 1996.
- [6] P. Mathera and B. Tso, "Classification of multisource remote sensing imagery using a genetic algorithm and markov random fields," *IEEE Trans. on Geosci. and Remote Sensing*, vol. 37, no. 3, pp. 1255–1260, 1999.
- [7] S. Chen, Q. Guo, H. Leung, and E. Bosse, "A maximum likelihood approach to joint image registration and fusion," *IEEE Transactions on Image Processing*, vol. 20, pp. 1363–1372, May 2011.
- [8] G. Winkler, *Image Analysis Random Fields and Dynamic Monte Carlo Methods*.
- [9] P. Varshney, *Distributed Detection and Data Fusion*. Springer, 1997.
- [10] F. Forbes and N. Peyrard, "Hidden markov random field model selection criteria based on mean field-like approximations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 19, pp. 1089 – 1101, 2003.
- [11] J. Zhang, "The mean field theory in em procedures for blind markov random field image restoration," *IEEE Transactions on Image Processing*, vol. 2, no. 1, pp. 27–40, 1993.

Article

A Joint Land Cover Mapping and Image Registration Algorithm Based on a Markov Random Field Model

Teerasit Kasetkasem^{*1}, Preesan Rakwatin², Ratchawit Sirisommai³, and Apisit Eiumnoh⁴¹Faculty of Engineering, Kasetsart University, Bangkok, Thailand 10900, fengtsk@ku.ac.th²GISTDA, Bangkok Thailand 10210, preesan@gistda.or.th³ICTES Program, Kasetsart University, Bangkok Thailand 10900, ratchawits@gmail.com⁴BIOTECH, NTSDA, Pathumthani 12120 THAILAND, apisit.eiu@biotec.or.th* corresponding author; E-Mail: fengtsk@ku.ac.th

Tel.: +66-81-945-8899; Fax: +66-2-579-4490.

Received: / Accepted: / Published:

Abstract: Traditional multi-modal and multi-temporal land cover mapping algorithms assume that all images are perfectly aligned. However, since multi-modal and multi-temporal images are likely to be obtained from different satellite platforms and/or acquired at different times, perfect alignment is very difficult to achieve. As a result, a proper land cover mapping algorithm must be able to correct registration errors as well as perform an accurate classification. In this paper, we proposed a joint classification and registration technique based on a Markov random field (MRF) model to simultaneously align two or more images and obtain a land cover map (LCM) of the scene. The expectation maximization (EM) algorithm is employed to solve the joint image classification and registration problem by iteratively estimating the map parameters and approximate posterior probabilities. Then, the maximum *a posteriori* (MAP) criterion is used to produce an optimum land cover map. We conducted experiments on a set of four simulated images and one pair of remotely sensed images to investigate the effectiveness and robustness of the proposed algorithm. Our results show that, with proper selection of a critical MRF parameter, the resulting LCMs derived from an unregistered image pair can achieve an accuracy as high as when images are perfectly aligned. Furthermore, the registration error can be greatly reduced.

Keywords: Joint land cover mapping and registration; Markov random field; optimum classifier; mean field theory; EM algorithm

1. Introduction

Remotely sensed images captured from satellites have been widely used for land cover mapping applications because of their capability to allow 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. Some applications, especially agricultural field mapping, particularly benefit from using multi-temporal sequences of satellite images because vegetation appearance often changes according to the season. Moreover, multiple input images from different satellites can be used to further improve classification performance by providing better spectral separation characteristics that a single sensor alone cannot provide. A practical application is reported in [1] where multi-temporal sequences of synthetic aperture radar (SAR) images and a single optical image were used. The results from this study showed that overall discrimination performance was increased, consistent with other similar research where multi-sensor data have been combined. Skriver, *et al* [2] emphasized the benefits of using multi-temporal SAR images in short succession (weekly to monthly acquisitions) for crop classification. These authors reported improved classification accuracy by using multi-temporal information. The authors in [3¹] exploited the crop phenology information to determine the growth stages by using multi-temporal TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS. They reported a significant correlation between backscattering coefficient and the normalized vegetation index obtained from SPOT4-5 images.

For multi-sensor image classification, the registration process is a crucial initial step. Registration aligns multiple satellite images into a common coordinate system. Only when all of the input images are perfectly registered can a classification algorithm be applied. Otherwise mis-registration will produce classification errors. In practice, perfect registration may not always be achievable since there are some unknown variations on satellite platforms and flight paths when capturing images. As a result, the overall classification accuracy is likely to suffer from mis-registration effects.

Mahapatra and Sun [4] proposed an idea to incorporate the reduction of image registration error into an image classification tool. They attempted to integrate the segmentation information into an elastic image registration by using a Markov random field model. In their work, the configuration of a pixel contains both displacement of a pixel and a segmentation label. The multi-resolution graph-cut approach was employed to achieve sub-pixel registration accuracy. Although their results produced remarkable performance for non-rigid body image registration, this algorithm cannot be applied for rigid body image registration problems such as in the remote sensing since their model does not allow one set of the registration parameters to govern the remapping process of entire image. Furthermore, since they only consider the segmentation problem, their algorithm does not cover the multi-class scenarios that are often considered in the land cover mapping of remotely sensed images.

Another work by Chen *et al.* in [5] investigated the problem of joint image fusion and registration. In their paper, the observed images were remapped versions of the original images with possibilities of polarity reverse and/or DC offset. Chen *et al.* used an expectation maximization algorithm to solve the estimation problems of registration parameters and the true scene simultaneously. Different pairs of multi-sensor images were tested against the proposed joint process. Under the assumption that registration performance affects the quality of fusion result, the authors reported that better fusion performance can be achieved due to reduced registration errors. However, their work did not cover the problem of image classification in the presence of image registration errors.

¹ 6:37 PM

In this paper, we employ an approach similar to [5] to incorporate correction of mis-registration effects into the land cover mapping process. To do this, we assume that remotely sensed images are derived from a common unobservable land cover map (LCM), and then distorted, with unknown remapping parameters, into the observed remote sensing images. (Note that if these map parameters are known, the observed remote sensing images can be directly aligned with the land cover map.) Next, we assume that a land cover class of interest is more likely to occupy several connected patches than a number of isolated pixels. As a result, the Markov random field (MRF) is employed as the model of the LCM. MRF models have been used in various fields ranging from statistical physics [6] and [7] to remote sensing. The original work by Geman and Geman [8] on MRF-based statistical methodology in 1984 has inspired a continuous stream of remote sensing researchers to employ the MRF model for a variety of image analysis tasks (e.g., [9] – [15]). Solberg *et al.* [9] developed MRF-based algorithms for image classification and change detection using multi-source data. A significant increase in classification and change detection accuracy was obtained using an MRF based classification algorithm compared to other approaches. Kasetkasem and Varshney [9] and Bruzzone and Prieto [11] also applied MRF models for an image change detection problem. Similarly, Xie *et al.* [12] applied the MRF model to the recurring problem of speckle reduction in synthetic aperture radar (SAR) images. These promising results on image analysis problems have encouraged us to employ the concept of MRF models to the problem of generating an LCM.

Based on our image model, the registration and classification process can be performed in the following fashion. First, we estimate the unknown map transformation parameters based on the maximum likelihood (ML) criteria, and, then use these parameters to compute posterior probability for different arrangements of the land cover maps, where the MAP classifier selects the most likely LCM. However, in order to find the map parameters, the conditional probability of observed images given the map parameters is needed. This conditional probability can only be obtained by summing the joint probabilities of observed images and LCM associated with the map parameters, over all possible LCMs. This is impossible to obtain in most practical scenarios. As a result, the expectation-maximization (EM) algorithm [16] is also employed here. The EM algorithm iteratively searches for the most likely map parameters. The resulting parameters converge to one of the local optimum points of the likelihood function.

For a given iteration of the EM algorithm, our method computes the expected value of the logarithm of the probability of the observed images and land cover map given the map parameters, based on the *a posterior* probability of the LCM given observed remote sensing images and the current estimated map parameters. Then, new map parameters are obtained by maximizing the expected values. It has been shown in literature [16] that the new map parameters always correspond to a higher value of the likelihood function. Since each iteration of the EM algorithm calculates *a posterior* probability given the current estimated map parameter, an optimum LCM under MAP criteria can be easily obtained by choosing the LCM that maximizes *a posterior* probability. In other words, an optimum LCM for the most recent estimate of the map parameters under the maximum *a posteriori* (MAP) criterion is obtained on every iteration of the EM algorithm.

The remainder of this paper is organized as follows. The next section will define the problem and our model. In Section 3, we will derive the optimum land cover mapping and image registration process based on the model presented in Section 2. The optimization problem and its corresponding solution are presented in Section 4. Our experiments to evaluate our proposed approach are described in Section 5. Finally, Section 6 offers concluding remarks.

2. Problem Statement

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 $\Lambda \in \{0, 1, \dots, L-1\}$ be the class labels. Therefore, we can express the LCM as $X(\mathcal{S}) \in \Lambda^{\mathcal{S}}$. The label of LCM at pixel s is denoted by x_s which can also be called the configuration of $X(\mathcal{S})$ at the site s . Since land cover classes are more likely to occur in connected patches in the LCM than isolated pixels, the LCM is assumed to satisfy the MRF properties with Gibbs potential $V_C(X)$. Hence, the marginal probability density function (PDF) of a LCM can be written as

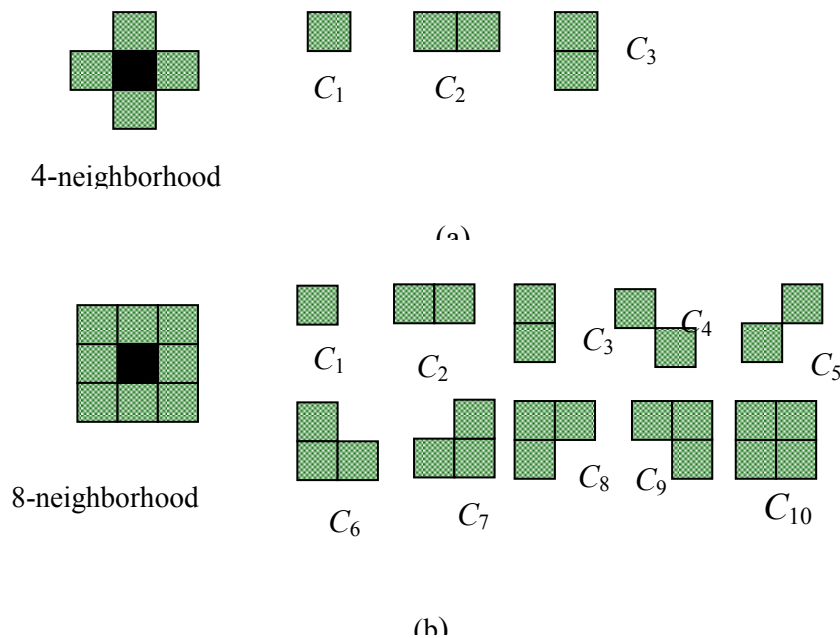
$$\Pr(X) = \frac{1}{Z_X} \exp \left(- \sum_{C \subset \mathcal{S}} V_C(X) \right) \quad (1)$$

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 [8]. Cliques are singleton or groups of pixels such that any two pixels are mutually neighbors. Figure 1 shows all possible clique types for 4- and 8-neighborhood systems. The value of the Gibbs potential function depends on the configurations of the entire LCM and the clique. Usually, low values of the potential function correspond to similar configurations whereas high values correspond to dissimilar configurations of a clique. For instance, the Ising model [5 and 7], given by,

$$V_{\{s,r\}}(X) = \begin{cases} -\beta; & \text{if } x_s = x_r \text{ and } r \in NG_s \\ \beta; & \text{if } x_s \neq x_r \text{ and } r \in NG_s \\ 0; & \text{if } r \notin NG_s \end{cases} \quad (2)$$

for any two sites r and s , has been used extensively by statistical physicists to explain why neighboring particles are more likely to rotate in the same direction (i.e., either clockwise or counterclockwise). Here the notation NG_s is a set of neighboring pixels of s . We can extend the above model to our problem by letting x_s and x_r be the class labels of pixels s and r in $\mathcal{S}$, respectively. With this modification, the Ising model can be applied to describe the LCM because land cover class distributions are similar to the phenomenon described above (i.e., classes occupying neighboring pixels are likely to be the same).

Figure 1: clique types for (a) 4-neighborhood; (b) 8-neighborhood



Furthermore, we assume that there are N remotely sensed images of the same scene acquired from different sensors and/or at different times. Here, $Y_n(\mathcal{T}_n) \in \mathfrak{R}^{\mathcal{T}_n \times B_n}$; $n = 1, 2, \dots, N$ denotes the n -th remotely sensed image where B_n denotes the number of spectral bands, and $\mathcal{T}_n$ is a map coordinate system to which the n -th remote sensing image is registered. Since all remotely sensed images and the LCM are from the same scene, the relationship between $\mathcal{S}$ and $\mathcal{T}_n$ can be determined. Let us denote a coordinate of a pixel s in the LCM as (x, y) where x and y are the column and row of x_s . Similarly, we can write $t_n = (u_n, v_n) \in \mathcal{T}_n$ where u_n and v_n are the column and row of the pixel t_n in Y_n . If the affine transformation is used, the relationship between s and t_n can be written as

$$\begin{bmatrix} u_n \\ v_n \end{bmatrix} = \begin{bmatrix} m_{1,n} & m_{2,n} \\ m_{3,n} & m_{4,n} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_{5,n} \\ m_{6,n} \end{bmatrix} \quad (3)$$

where $m_{1,n}$ and $m_{4,n}$ are scale parameters, $m_{2,n}$ and $m_{3,n}$ are skew parameters, and $m_{5,n}$ and $m_{6,n}$ are displacement parameters in column and row-direction, respectively. We refer to $M_n = [m_{1,n}, m_{2,n}, m_{3,n}, m_{4,n}, m_{5,n}, m_{6,n}]$ as the map parameter vector between coordinate systems $\mathcal{S}$ and $\mathcal{T}_n$.

When all the map parameter vectors are given, one can remap all remotely sensed images to perfectly align with the LCM. Let us denote

$$Z_n(\mathcal{S}) = f(Y_n(\mathcal{T}_n), M_n) \quad (4)$$

as the remapped and resampled version of the n -th remote sensing image. Here, we assume further that the remapped and resampled images are statistically independent for a given LCM, i.e.,

$$\Pr(\mathbf{Y}|\mathbf{X}(\mathcal{S}), \mathbf{M}) = \Pr(Z_1(\mathcal{S}), \dots, Z_N(\mathcal{S})|\mathbf{X}(\mathcal{S})) = \prod_{n=1}^N \Pr(Z_n(\mathcal{S})|\mathbf{X}(\mathcal{S})) \quad (5)$$

where $\mathbf{M} = \{M_1, \dots, M_N\}$ and $\mathbf{Y} = \{Y_1(\mathcal{T}_1), \dots, Y_N(\mathcal{T}_N)\}$ be collections of the map parameters and the observed multispectral images. Moreover, the intensity vectors from different pixels in $Z_n(\mathcal{S})$ are also assumed to be statistically independent when the LCM is given. Hence, the joint conditional PDF can be written as

$$\Pr(\mathbf{Y}|\mathbf{X}(\mathcal{S}), \mathbf{M}) = \prod_{n=1}^N \prod_{s \in \mathcal{S}} \Pr(\mathbf{z}_{n,s} | x_s) \quad (6)$$

where $\mathbf{z}_{n,s} \in \mathbb{R}^{B_n}$ denotes the intensity vector of the remapped image $Z_n(\mathcal{S})$ at a pixel s . We acknowledge that the assumption given in Eq.(6) may not always be true for all cases since some land cover classes have textural structure. One can incorporate texture information into our image model appropriately, which may further result in an increase in accuracy. This will, however, result in very complex problems which may not be desirable in practice.

If we assume further that the intensity vector at a pixel s of the remapped image $Z_n(\mathcal{S})$ given the class label x_s is a multivariate normal random vector with mean vector $\boldsymbol{\mu}_{x_s,n}$ and covariance matrix $\Sigma_{x_s,n}$, Eq. (6) can be rewritten as

$$\Pr(\mathbf{Y}|\mathbf{X}(\mathcal{S}), \mathbf{M}) = \prod_{n=1}^N \frac{1}{(2\pi)^{\frac{B_n}{2}} |\Sigma_{x_s,n}|^{1/2}} \exp \left[- \sum_{s \in \mathcal{S}} \frac{1}{2} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \Sigma_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) \right] \quad (7)$$

where T denotes the matrix transpose operation.

By using chain rule, the posterior probability of the LCM given the observed multispectral images and the map parameters can be written as

$$\Pr(\mathbf{X}|\mathbf{Y}, \mathbf{M}) = \frac{\Pr(\mathbf{Y}|\mathbf{X}, \mathbf{M}) \Pr(\mathbf{X})}{\Pr(\mathbf{Y}, \mathbf{M})} \quad (8)$$

Since $\Pr(\mathbf{Y}, \mathbf{M})$ is independent of the choice of X , it can be treated as a constant. Hence, we have

$$\Pr(X|\mathbf{Y}, \mathbf{M}) = C \Pr(\mathbf{Y}|X, \mathbf{M}) \Pr(X). \quad (9)$$

By substituting Eq. (1), and Eq. (7) into Eq. (9), we obtain

$$\Pr(X|\mathbf{Y}, \mathbf{M}) = \frac{1}{Z'} e^{-E(X|\mathbf{Y}, \mathbf{M})}. \quad (10)$$

where $Z' = \sum_{X \in \Lambda^S} e^{-E(X|\mathbf{Y}, \mathbf{M})}$ is a normalizing constant and independent of the choice of X , and

$$E(X|\mathbf{Y}, \mathbf{M}) = \frac{1}{2} \sum_{n=1}^N \left(\sum_{s \in S} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \boldsymbol{\Sigma}_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\boldsymbol{\Sigma}_{x_s,n}| \right) + \sum_{C \subset S} V_C(X) \quad (11)$$

is called a conditional Gibbs energy function. Since, in this paper, we consider cliques comprised of pixel pairs only, hence, the conditional Gibbs energy function can be written as

$$E(X|\mathbf{Y}, \mathbf{M}) = \frac{1}{2} \sum_{n=1}^N \left(\sum_{s \in S} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \boldsymbol{\Sigma}_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\boldsymbol{\Sigma}_{x_s,n}| \right) + \frac{1}{2} \sum_{s \in S} \sum_{r \in NG_s} V_{\{s,r\}}(x_s, x_r) \quad (12)$$

where NG_s denotes the set of neighboring pixels of s . The normalizing constant Z' cannot be computed in most practical scenarios due to the large number possible configurations (e.g., there are more than 2^{4096} possible configurations for binary LCM of size 64×64 .) As a result, we propose the use the mean field theorem [17]-[18] to remove the interaction between neighboring pixels defined in $V_C(X)$. The mean field theorem approximates the conditional Gibbs energy function as

$$E(X|\mathbf{Y}, \mathbf{M}) \approx \frac{1}{2} \sum_{s \in S} h_s(x_s|\mathbf{Y}, \mathbf{M}) \quad (13)$$

where

$$h_s(x_s|\mathbf{Y}, \mathbf{M}) = \sum_{n=1}^N \left\{ (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \boldsymbol{\Sigma}_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\boldsymbol{\Sigma}_{x_s,n}| \right\} + \sum_{r \in NG_s} E_{x_r}[V_{\{s,r\}}(x_s, x_r)]. \quad (14)$$

Here, $E_{x_r}[V_{\{s,r\}}(X)]$ is the expected value of the potential function with respect to the configuration of x_r . The expected value $E_{x_r}[V_{\{s,r\}}(x_s, x_r)]$ does not depends on x_r , and is equal to

$$E_{x_r}[V_{\{s,r\}}(X)] = \sum_{x_r \in \Lambda} V_{\{s,r\}}(x_s, x_r) p^{MF}(x_r|\mathbf{Y}, \mathbf{M}) \quad (15)$$

where $p^{MF}(x_r|\mathbf{Y}, \mathbf{M}) = \frac{1}{Z_r} e^{-\frac{1}{2} h_r(x_r|\mathbf{Y}, \mathbf{M})}$. Note here that $Z_r = \sum_{x_r \in \Lambda} e^{-\frac{1}{2} h_r(x_r|\mathbf{Y}, \mathbf{M})}$ is the normalizing constant for a pixel r . By using the approximation given in Eq. (13), the posterior probability can be written as

$$\Pr(X|\mathbf{Y}, \mathbf{M}) \approx \Pr^{MF}(X|\mathbf{Y}, \mathbf{M}) = \prod_{s \in S} p^{MF}(x_s|\mathbf{Y}, \mathbf{M}) \quad (16)$$

The approximation in Eq. (16) is closest to $\Pr(X|\mathbf{Y}, \mathbf{M})$ from all approximations of the form $\prod_{s \in S} p(x_s|\mathbf{Y}, \mathbf{M})$ when the Kullback-Leibler (KL) divergence [18]-[19] is used as a distance measure.

3. Optimum Image Registration and Land Cover Mapping Criteria

The standard approaches to multi-temporal and/or multi-modal image classification involve two steps. First images from different sources and/or times are registered to produce a set of images in a common coordinate system. Then, a land cover map is derived from this set of registered images. In this work, even though we propose an algorithm to simultaneously register and classify images, we still treat image registration and classification as two separate problems to follow standard approaches. As a result, we propose different optimization criteria for image registration and land cover mapping. However, we will show in Section IV that both image registration and land cover mapping can be combined into one algorithm so that the registration and land cover mapping can be performed simultaneously.

3.1 Optimum Image Registration

The maximum likelihood estimate (MLE) can be employed as the optimum map parameter estimator since the MLE is known to a consistent estimator [20]. The goal of the MLE is to determine the map parameters that maximize the joint probability density function (PDF) of all the observed images given the map parameters, i.e.,

$$(M_1, \dots, M_N)^{opt} = \arg \max_{M_1, \dots, M_N} \Pr(Y_1, \dots, Y_N | M_1, \dots, M_N) \quad (17)$$

In order to solve Eq. (17), the conditional PDF $\Pr(Y_1, \dots, Y_N | M_1, \dots, M_N)$ must be calculated and it is equal to

$$\begin{aligned} \Pr(Y_1, \dots, Y_N | M_1, \dots, M_N) &= \sum_{X \in \Lambda^S} \Pr(Y_1(\mathcal{T}_1), \dots, Y_N(\mathcal{T}_N), X(\mathcal{S}) | M_1, \dots, M_N) \\ &= \prod_{n=1}^N \sum_{X \in \Lambda^S} \Pr(Z_n | X(\mathcal{S})) \Pr(X(\mathcal{S})) \end{aligned} \quad (18)$$

Note here again that Z_n is the remapped and resampled version of Y_n . Since Eq. (18) is written as a multiplication of $\sum_{X \in \Lambda^S} \Pr(Z_n | X(\mathcal{S})) \Pr(X(\mathcal{S}))$, the solution of Eq. (17) can be individually obtain, i.e.,

$$M_n^{opt} = \arg \max_{M_n} \sum_{X \in \Lambda^S} \Pr(Z_n | X(\mathcal{S})) \Pr(X(\mathcal{S})), \quad (19)$$

for $n = 1, \dots, N$. Since $\mathcal{S}$ is also unknown, there are many possible sets of M_n that maximize Eq. (19). For instance, if $M_n = [1, 0, 0, 1, 0, 0]$ is the solution of Eq. (19) for $\mathcal{S} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, we have that $M'_1 = [1, 0, 0, 1, 1, 0]$ is also the solution of Eq. (19) for $\mathcal{S}' = \{(0, -1), (0, 0), (1, -1), (1, 0)\}$. As a result, it is imperative to limit the search space and number of possible solutions. Furthermore, in most practical situations, we may wish to produce the LCM registered to one of the input remote sensing images. Without loss of generality, we assume that the LCM is registered to Y_1 , i.e., we have $M_1 = M_1^* = [1, 0, 0, 1, 0, 0]$.

Next, let us consider a small LCM of size 100×100 pixels. In this case, there are $2^{10,000} \approx 2 \times 10^{3,010}$ possible binary LCMs. Therefore, the direct calculation of Eq. (19) is an impossible task, and hence, the solution of the MLE cannot be obtained in reasonable time. As a result, the expectation-maximization (EM) algorithm [16] is employed instead. The EM algorithm is an iterative parameter estimator which produces a new estimate for every iteration. It has been shown in [16] that this new

estimate always results in higher or at least the same value of the likelihood function. In other words, if we let $\mathbf{M}^t = \{M_1^t, M_2^t, \dots, M_N^t\}$ be the collection of all estimated parameters at the t -th iteration from the EM algorithm, we will have $\Pr(Y_1, \dots, Y_N | \mathbf{M}^t) \geq \Pr(Y_1, \dots, Y_N | \mathbf{M}^{t-1})$ where $\mathbf{M}^{t-1}$ is the collection of estimated parameters at $(t - 1)$ -th iteration. Here, and throughout the rest of the paper, we omit $\mathcal{S}$ and $\mathcal{T}_n$ for the sake of abbreviation. In Section 4, we will discuss the details of the EM algorithm employed in this work and how it can be combined with the land cover mapping process. However, before going into the detail of the proposed algorithm, let us state the optimization criterion for the land cover mapping considered in this paper.

3.2 Optimum Land Cover Map

The classifier based on the maximum *a posteriori* (MAP) criteria selects the most likely LCM given the observed data and the map parameters since the resulting probability of error is minimum among all other classifiers [21]–[22]. The optimum solution under the MAP criterion is expressed as

$$X^{opt} = \arg \max_X [\Pr(X | \mathbf{Y}, \mathbf{M})]. \quad (20)$$

In general, $\Pr(X | \mathbf{Y}, \mathbf{M})$ is non-concave function and, therefore, conventional gradient-based optimization algorithms are not applicable for the solution of Eq. (20). Furthermore, the number of possible solutions is also very large. A direct search for the solution of Eq. (20) is too expensive to be practically implemented. As a result, we propose the use the mean field theorem [17]–[18] to remove the interaction between neighboring pixels defined in $V_C(X)$. Hence, by substituting Eq. (16) into Eq. (20), the optimization problem becomes

$$X^{opt} = \arg \max_X \left[\prod_{s \in \mathcal{S}} p^{MF}(x_s | \mathbf{Y}, \mathbf{M}) \right]. \quad (21)$$

Since the optimizing function in Eq. (21) is written in the form of the multiplication of functions of an individual pixels, and $p^{MF}(x_s | \mathbf{Y}, \mathbf{M})$ is a non-negative function, the optimum solution can be solved from an individual function, i.e., for $s \in \mathcal{S}$,

$$x_s^{opt} = \arg \max_X [p^{MF}(x_s | \mathbf{Y}, \mathbf{M})]. \quad (22)$$

which is equivalent to

$$x_s^{opt} = \arg \min_X [h_s(x_s | \mathbf{Y}, \mathbf{M})]. \quad (23)$$

4. Joint Image Registration and Land Cover Mapping Algorithm

Since the EM algorithm is employed in this literature as the parameter estimator, we begin our discussion with the details of the EM algorithm. The EM algorithm [16] consists of two steps, namely the *expectation* (or *E*) and *maximization* (or *M*) steps. In the E-step, the EM algorithm finds the lower bound of the likelihood function given in the right hand side of Eq. (20) by calculating the expected value of the joint log-likelihood function of the observed images and the LCM. Here, the expected value is computed over the LCMs given the most recent estimate of the map parameter vectors and observed data, i.e.,

$$\begin{aligned} Q(\mathbf{M} | \mathbf{M}^{t-1}) &= E[\log \Pr(\mathbf{Y}, X | \mathbf{M}) | \mathbf{Y}, \mathbf{M}^{t-1}] \\ &= E[\log \Pr(\mathbf{Y} | X, \mathbf{M}) + \log \Pr(X) | \mathbf{Y}, \mathbf{M}^{t-1}] \end{aligned} \quad (24)$$

where $\mathbf{Y} = \{Y_1, \dots, Y_N\}$ is the set of all observed remotely sensed images, $\mathbf{M} = \{M_1, \dots, M_N\}$ is the set of all unknown map parameters, and $\mathbf{M}^t = \{M_1^t, \dots, M_N^t\}$ is the set of all estimated parameters from the t -

th iteration of the EM algorithm. Note here that $M_1^t = M_1^*$. By substituting Eq. (1) and Eq. (7) into Eq. (24), the expected value becomes

$$Q(\mathbf{M}||\mathbf{M}^{t-1}) = E \left[-\frac{1}{2} \sum_{n=1}^N \left(\sum_{s \in S} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \Sigma_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\Sigma_{x_s,n}| - \log(2\pi)^{B_n} \right) - \sum_{C \in S} V_C(X) - Z_X \middle| \mathbf{Y}, \mathbf{M}^{t-1} \right] \quad (25)$$

In the M-step, the expected value given in Eq. (25) is maximized and a new set of map parameter vectors is obtained, i.e.,

$$\mathbf{M}^t = \arg \max_{\mathbf{M}, M_1 = M_1^*} Q(\mathbf{M}||\mathbf{M}^{t-1}) \quad (26)$$

Clearly, the terms $\log |\Sigma_{x_s,n}|$, $\log(2\pi)^{B_n}$, $\sum_{C \in S} V_C(X)$, and Z_X in Eq. (25) do not depend on $\mathbf{M}$. Hence, Eq. (25) can be modified to

$$\mathbf{M}^t = \arg \max_{\mathbf{M}, M_1 = M_1^*} Q'(\mathbf{M}||\mathbf{M}^{t-1}) \quad (27)$$

where

$$Q'(\mathbf{M}||\mathbf{M}^{t-1}) = \arg \max_{\mathbf{M}, M_1 = M_1^*} E \left[-\frac{1}{2} \sum_{n=1}^N \sum_{s \in S} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \Sigma_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) \middle| \mathbf{Y}, \mathbf{M}^{t-1} \right]. \quad (28)$$

To find the solution of Eq. (28), the *a posteriori* probability of the LCM given the observed images and the map parameters from the $(t-1)$ -th iteration must be calculated in order to find the expected value. From the same reason as discussed in Section 2, the posterior probability cannot be practically calculated due to the huge number of possible LCMs. As a result, we employ the approximation given in Eq. (16), and hence, we have

$$\Pr(X|\mathbf{Y}, \mathbf{M}^{t-1}) \approx \prod_{s \in S} p_s^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1}) = \prod_{s \in S} \frac{1}{Z_s} e^{-\frac{1}{2} h_s(x_s|\mathbf{Y}, \mathbf{M}^{t-1})}. \quad (29)$$

By substituting d Eq. (26) into Eq. (29), we have

$$\begin{aligned} Q'(\mathbf{M}||\mathbf{M}^{t-1}) &\approx Q^{MF}(\mathbf{M}||\mathbf{M}^{t-1}) \\ &= -\frac{1}{2} \sum_{n=1}^N \sum_{s \in S} \sum_{x_s \in \Lambda} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \Sigma_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) p_s^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1}) \end{aligned} \quad (30)$$

Hence, in the M-step, the new map parameters can be obtained by maximizing the approximation given Eq. (30), i.e.,

$$\mathbf{M}^t = \arg \max_{\mathbf{M}, M_1 = M_1^*} Q^{MF}(\mathbf{M}||\mathbf{M}^{t-1}) \quad (31)$$

Since $\mathbf{z}_{n,s}$ depends only M_n and the right hand side of Eq. (30) is written as the summation of $\mathbf{z}_{n,s}$ from different images, the above optimization problem can be rearranged into the optimization of each individual mapping parameters, i.e.,

$$M_n^t = \arg \max_M Q_n^{MF}(M_n||\mathbf{M}^{t-1}); n = 2, \dots, N \quad (32)$$

where

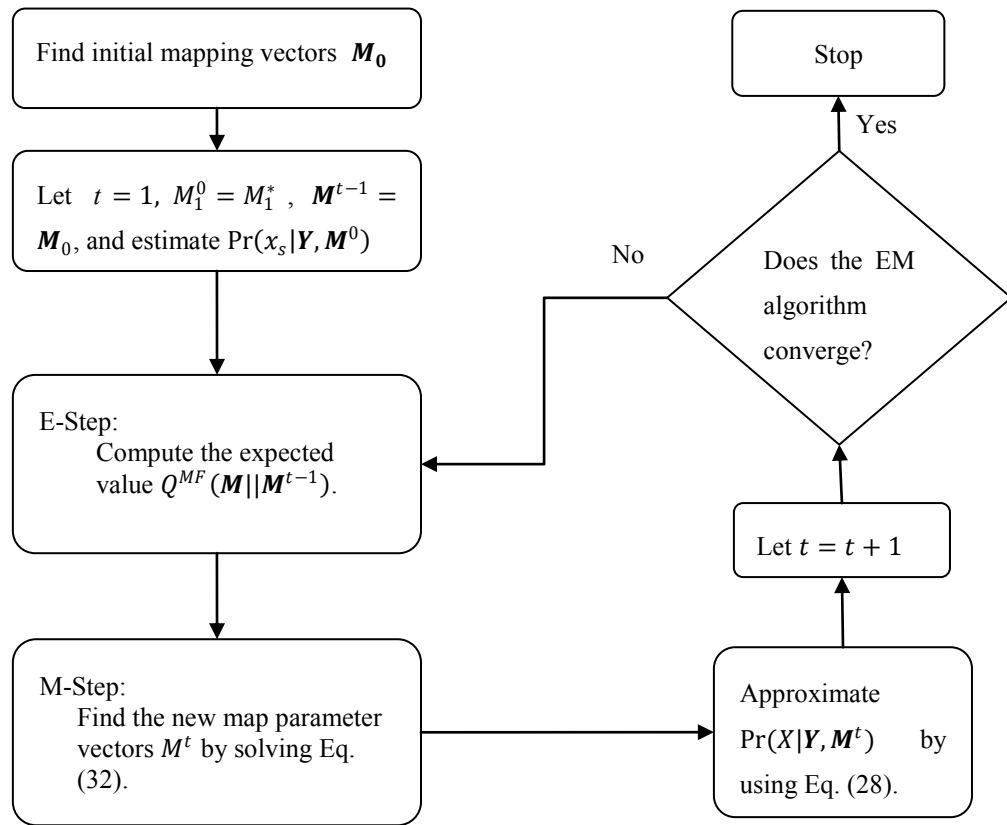
$$Q_n^{MF}(M_n||\mathbf{M}^{t-1}) = -\frac{1}{2} \sum_{s \in S} \sum_{x_s \in \Lambda} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \Sigma_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) p_s^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1}). \quad (33)$$

Using the approximations given above, the modified EM algorithm is displayed in Figure 2. For each iteration, the posterior probability $\Pr(X|\mathbf{Y}, \mathbf{M}^t)$ is approximated by recalculating $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$. We follow the work by Zhang [19] which suggested that $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$ can be obtained from

$$\begin{aligned}
h_s(x_s|\mathbf{Y}, \mathbf{M}^t) &= h_{obv}(x_s|\mathbf{Z}) + h_{ng}(x_s|X_{NG}) \\
&= \sum_{n=1}^N \left\{ (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n})^T \Sigma_{x_s,n}^{-1} (\mathbf{z}_{n,s} - \boldsymbol{\mu}_{x_s,n}) + \log |\Sigma_{x_s,n}| \right\} \\
&\quad + \sum_{x_r \in \Lambda} V_{\{s,r\}}(x_s, x_r) p^{MF}(x_s|\mathbf{Y}, \mathbf{M}^{t-1}).
\end{aligned} \tag{34}$$

where $h_{obv}(x_s|\mathbf{Z})$ and $h_{ng}(x_s|X_{NG})$ are the potential functions depending upon the observation and neighboring pixels, respectively.

Figure 2: Block diagram of the modified EM algorithm



Since $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$ is recalculated for every iteration of the EM algorithm, we can choose a land cover class that minimizes $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$, and obtain the optimum LCM based on criterion given in Eq. (23). By combining the EM algorithm given in Figure 2 and the land cover mapping process by minimizing Eq. (23), the joint image registration and land cover mapping algorithm is given as

1. Initialize map parameters, i.e., $M_1^0 = M_1^*$ and $\mathbf{M}^0 = \{M_1^0, \dots, M_n^0\}$, let $t = 1$, and assign $p^{MF}(x_s|\mathbf{Y}, \mathbf{M}^0)$ based on some prior knowledge.
2. Compute $Q_n^{MF}(M_n||\mathbf{M}^{t-1})$ for $n = 2, \dots, N$.
3. Obtain M_n^t by solving Eq. 32 for $n = 2, \dots, N$, and assign $M_1^t = M_1^*$ and $\mathbf{M}^t = [M_1^t, \dots, M_N^t]$
4. Compute $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$ by using Eq. (34).
5. Find the new LCM that minimizes $h_s(x_s|\mathbf{Y}, \mathbf{M}^t)$ for all $s \in \mathcal{S}$.
6. Let $t = t + 1$, and go to Step 2 if a convergence criterion is not satisfied.

The critical step in the successful implementation of the joint image registration and land cover mapping algorithm proposed above is how to solve Eq. (32) efficiently. Here to find the maxima, we employ the particle swarm optimization (PSO) algorithm [23] since the traditional gradient search approaches are likely to fall into one of the local optimum points of $Q^{MF}(\mathbf{M}||\mathbf{M}^{t-1})$ due to its non-convexity. The PSO exploit the cooperative behavior for a group of animals such as birds and insects. In the PSO, an individual animal is called a particle and a group of animals is called swarm. These particles are initially distributed throughout the search space, and move around the search space. Based on some social and cooperative criteria, these particles will eventually cluster in the regions where the global optima can be found.

In our work, for a given image Y_n , each particle represents a mapping parameter and we denote the i -th particle as $M_{n,i}$. At each iteration, the i -th particle moves by a velocity V_i which is a function of the best-known positions (mapping parameter) discovered by the i -th particle (P_i) itself, and from all particles (G), i.e.,

$$V_i^r = \omega V_i^{r-1} + \varphi_1 u_1 (P_i - M_{n,i}^{r-1}) + \varphi_2 u_2 (G - M_{n,i}^{r-1}) \quad (35)$$

and

$$M_{n,i}^r = M_{n,i}^{r-1} + V_i^r \quad (36)$$

for $n = 2, \dots, N$. where ω is the inertial weight, φ_1 and φ_2 are acceleration constants, and u_1 and u_2 are uniform random numbers between zero and one. The velocity is usually kept in the range of $[V_{min}, V_{max}]$ to make sure that $M_{n,i}^r$ is in the valid regions. Note here that the performance of the PSO depends on the selection of, ω , φ_1 and φ_2 , and the number of iterations. In this paper, we set the number of particles to 80 and the maximum number of iterations to be 200 as a suitable setup for our experiment. We acknowledge that different setups of these parameters may result in different convergence rate. However, the investigation of the optimum parameter selection of the PSO in term of convergence rate is out of scope of this paper. We refer to the work by [24] for more details.

5. Experiments

In this section, we provide the results of two experiments based on the methodology derived in Section 4 to jointly register and classify a set of remotely sensed images. The first experiment is conducted over a simulated dataset in order for us to investigate many aspects of our proposed algorithm. Next, we will examine the performance of our algorithm in the actual remote sensing image. For both examples, the goal is to examine the performance of algorithm to different degrees of initial registration errors. If our algorithm performs perfectly, it should be able to align images together and produce a LCM from unregistered images as accurate as when images are registered.

Experiment 1:

In the experiment, we examine the performance of the proposed algorithm in term of classification performance and registration accuracy by attempting to produce a land cover map from a set of four simulated images. All the simulated images have an equal size of 512×512 pixels (Figure 3) and contain four land cover classes (Classes 1-4) with intensity values of zero, one, two and three for black, dark gray, light gray and white areas, respectively. Based on the noiseless image, the ground truth image in this example is given in Figure 4 where the blue, black, green and red colors correspond to Classes 1-4, respectively. Next, all of the input images are added with the independent and identical Gaussian noise with zero mean and standard deviation of $\sigma = 1$ to examine the performance of our proposed algorithm to the image noise. Figure 5 shows an example of the input image for $\sigma = 1$. We observe that the observed image appears to be very noisy.

Figure 3: Noiseless Simulated Image in Example 1



Figure 4: The ground data of Example 1

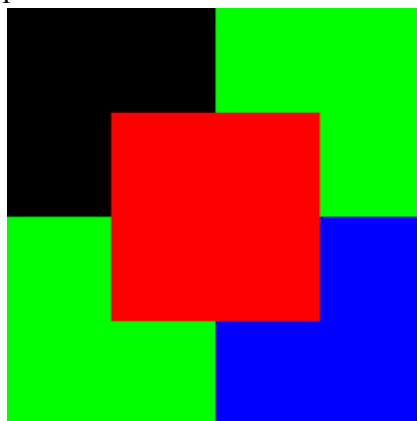
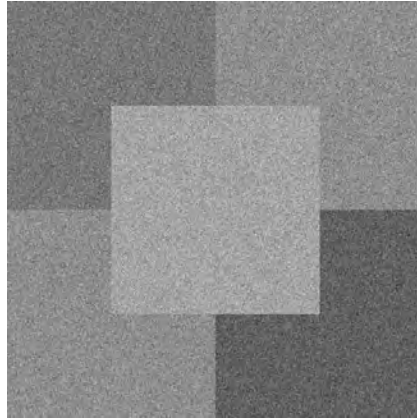


Figure 5: An example of the noisy input image at $\sigma=1$ in Example 1

Since our algorithm performs both image registration and land cover mapping at the same time, the performance of our algorithm can be evaluated in terms of how much the resulting LCM deviates from the reference LCM, and the estimation error between our calculated map parameters and the actual parameters that relate the LCM to the simulated images. If our algorithm performs perfect registration and land cover mapping, the resulting percentages of mis-classified pixels will be zero, and the registration error between images and LCM will be zero. In this example, the correct mapping parameters for all observed images are the same and equal to $M_{perfect} = [1,0,0,1,0,0]$ which correspond to unit scale, zero skew, and zero displacement. Next, since we want to examine the effect of the initial registration errors to the performance of our algorithm, we investigate different scenarios of initial registration errors by varying the initial mapping parameters between the observed images and LCM at different values of displacement, scale and skew parameters. In particular, we investigate three scenarios for the only displacement, only scale and only skew errors, respectively. Table 1 shows the initial mapping parameters for all three scenarios. Here, δ , ρ and η are the initial displacement, scale, and skew parameter errors. Note that the initial mapping parameter errors for Image 1 for all scenarios are zero since we assume that the first image is registered to the LCM as mentioned in Section 3.1.

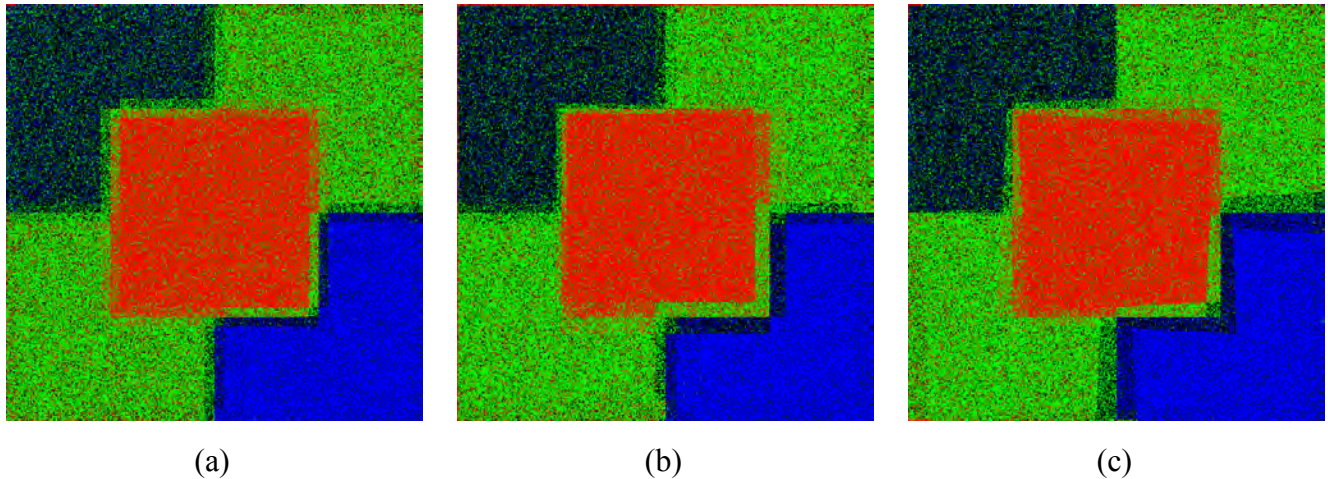
Before examining the performance of our proposed algorithm, we examine the effect of registration errors to the performance of image classification. This value can be viewed as the worst case scenario where the LCM is derived directly from the set of mis-registered images. Here, we employ the maximum likelihood classifier (MLC) [21] to the set of four remapped images, and the LCM is obtained from

$$x_s^{MLC} = \arg \min_x \left[\sum_{n=1}^4 \left\{ (z_{n,s} - \mu_{x_s,n})^T \Sigma_{x_s,n}^{-1} (z_{n,s} - \mu_{x_s,n}) + \log |\Sigma_{x_s,n}| \right\} \right] \quad (37)$$

where the subscript n denotes the n -th remapped image. We note here that Eq. (37) is the special of the optimum LCM obtained from Eq. (22) when $\beta = 0$. Figure 6 (a), (b) and (c) display the resulting LCM for $\delta = 12$ and $\sigma = 1$ for Scenario I, $\rho = 0.05$ and $\sigma = 1$ for Scenario II, and $\eta = 0.05$ and $\sigma = 1$ for Scenario III. The averaged percentages of misclassified pixels after a hundred independent runs are equal to 28.66%, 31.93 and 27.03, for Scenarios I, II and III given above, respectively.

Table 1: Mapping parameter errors are three scenarios in Example 1

	Image	Mapping parameters					
		m_1	m_2	m_3	m_4	m_5	m_6
Scenario I: Displacement error (δ)	1	1	0	0	1	0	0
	2	1	0	0	1	δ	0
	3	1	0	0	1	0	$-\delta$
	4	1	0	0	1	$-\delta$	δ
Scenario II: Scale error (ρ)	1	1	0	0	1	0	0
	2	$1 + \rho$	0	0	1	0	0
	3	1	0	0	$1 + \rho$	0	0
	4	$1 - \rho$	0	0	$1 - \rho$	0	0
Scenario III: Sheer error (η)	1	1	0	0	1	0	0
	2	1	η	0	1	0	0
	3	1	0	η	1	0	0
	4	1	$-\eta$	$-\eta$	1	0	0

Figure 6: Examples of the MLC-Based LCMs for (a) Scenario I with $\delta=12$ and $\sigma=1$; (b) Scenario II with $\rho=0.05$ and $\sigma=1$; and (c) Scenario III with $\eta=0.05$ and $\sigma=1$ 

Next, the proposed algorithm is applied to the above datasets. The whole process was implemented using CUDA on NVIDIA Tesla M2090 with 1 GB memory. Here, we assign $p^{MF}(x_s|Y, \mathbf{M}^0) = \frac{1}{4}$ as the most extreme case where no prior information is given. In different trials, the value of β is set to be 0.00, 0.25, 0.50, and 0.75 (see Eq. (2)). Since our algorithm performs both image classification and registration, the termination criteria must ensure the convergences in both the estimated posterior probability and mapping parameters. As a result, we define

$$p_{changes} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \sum_{x_s \in \Lambda} |p_s^{MF}(x_s | \mathbf{Y}, \mathbf{M}^t) - p_s^{MF}(x_s | \mathbf{Y}, \mathbf{M}^{t-1})|, \quad (38)$$

to measure changes in the posterior probabilities from two consecutive iterations. We also define

$$d_{movement, n} = \frac{1}{|\mathcal{T}_n|} \sum_{(u_n, v_n) \in \mathcal{T}_n} \sqrt{(x_n^t - x_n^{t-1})^2 + (y_n^t - y_n^{t-1})^2} \quad (39)$$

to characterize the movement of coordinates of the remapped image Z_n from two consecutive iterations where

$$\begin{bmatrix} x_n^t \\ y_n^t \end{bmatrix} = \begin{bmatrix} m_{1,n}^t & m_{2,n}^t \\ m_{3,n}^t & m_{4,n}^t \end{bmatrix}^{-1} \begin{bmatrix} u_n - m_{5,n}^t \\ v_n - m_{6,n}^t \end{bmatrix}. \quad (40)$$

Here, $m_{i,n}^t$ denotes the mapping parameter m_i from the n^{th} at the t^{th} iteration. In this example, the algorithm terminates when $p_{changes}$ is less than $p_{min} = 10^{-5}$, and $d_{movement, n}$ is less than 0.1 pixels for five consecutive iterations for $n = 2, 3, 4$. To create a benchmark for our proposed algorithm, we examined two extreme cases where LCMs are derived directly from the unregistered image pairs and from perfect registered image pair. The LCMs from these extreme cases are classified using our proposed algorithm by fixing $\mathbf{M}^t = \mathbf{M}^*$. For perfect registration, we have $\mathbf{M}^* = \mathbf{M}_{perfect}$ whereas, for unregistered image pairs, we set $\mathbf{M}^*$ equal the values given in Table 1 for the respective scenarios. The first extreme case can be considered as the lower limit on the classification accuracy if we perform the land cover mapping without alignment of images first. The second case is an upper bound on the classification accuracy when we produce a map from a registered image pair. By setting up our experiment in this fashion, we can investigate how much improvement our algorithm can gain by integrating the registration and classification together, and how far the performance of our algorithm is from the upper limit where all uncertainties in registration are removed. To ensure the statistical significance of our experiment, all experiments are repeated ten times.

Table 2 displays the averaged percentages of misclassified pixels (PMP) of the LCMs for different values of β and for Scenario I with $\delta = 12$, Scenario II with $\rho = 0.05$ and Scenario III with $\eta = 0.05$ when $\sigma = 1$. Note here that, in this example, we employ the percentages of mis-classified pixels as the performance metric to evaluate the classification performance rather than the overall accuracy to highlighted small differences in the classification performance between LCMs derived from image datasets without registration error and LCMs obtained from our proposed algorithm. From Table 2, it is clear that, from all scenarios, the PMPs derived from image datasets without registration errors corrections are always significance poorer than those derived from registered image datasets. These results support our claims that it is important to consider lack of alignments in performing image classification. We also observe that, for $\beta = 0.25, 0.5$ and 0.75 , our proposed algorithm produced the LCM with the accuracy similar to those obtained from image dataset without any registration error. These results imply that our proposed algorithm attain the upper-bound accuracy with proper selection of MRF parameter. To ensure the statistical significance, we compute the pairwise t -statistics for unequal variance populations [20] of the PMPs obtained from LCMs derived from the proposed algorithm for various initial registration errors against those obtained from image dataset with no registration error, and the resulting p -values [20] of the t -statistics are given in Table 3. The p -value

represents the probability that there is no difference in PMPs. Hence, a smaller p -value implies that PMPs from two experiments are different. We also compute the t -statistics comparing LCMs obtained from image dataset with and without registration errors. The resulting p -values of these t -statistics are also summarized in Table 3. It is clear from Table 3 that there is significant differences in term of PMPs from LCMs obtained from image dataset with and without registration errors. Furthermore, the p -values also support our claim that for $\beta = 0.25, 0.5$ and 0.75 , our proposed algorithm produced the LCM with the accuracy similar to those obtained from image dataset without any registration error. However at $\beta = 0$, our proposed algorithm perform significantly poorer than those of perfect registration. In fact, at $\beta = 0$, our proposed algorithm achieves roughly the same performance as in situation where there is no registration error correction since at $\beta = 0$, our proposed algorithm cannot correctly estimate the map vectors. Figure 7 shows examples of the resulting LCMs at $\beta = 0.75$ for all scenarios. We observe that all the LCMs appeared to be more connected than the MLC-based LCMs given in Figure 6: Examples of the MLC-Based LCMs for (a) Scenario I with $\delta=12$ and $\sigma=1$; (b) Scenario II with $\rho=0.05$ and $\sigma=1$; and (c) Scenario III with $\eta=0.05$ and $\sigma=1$

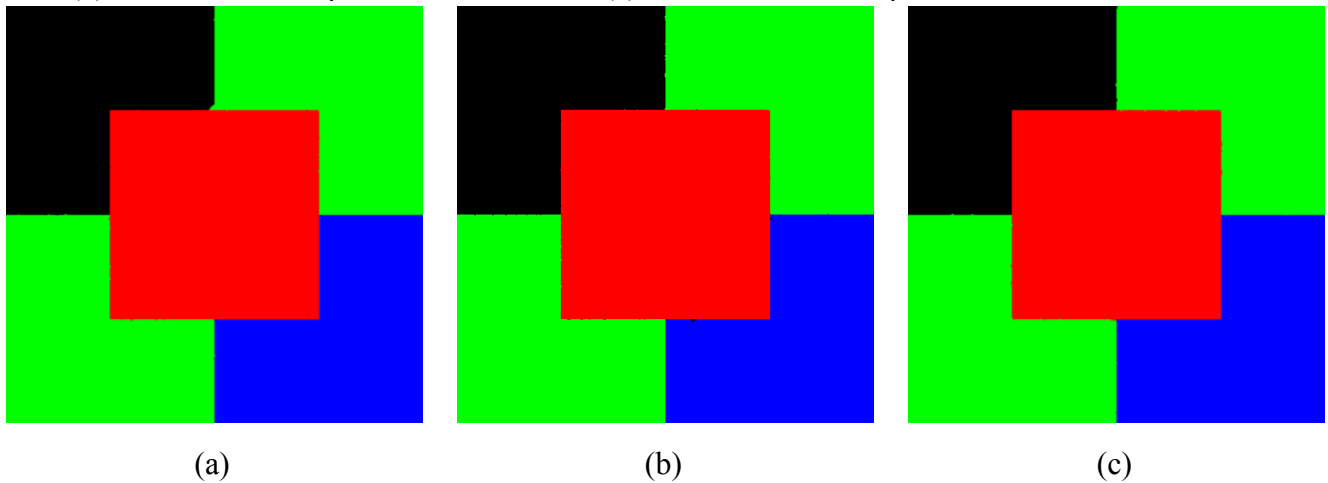
Table 2: Comparison of the averaged percentages of misclassified pixels (PMP) between two extreme cases and our proposed algorithm

β	No registration Error	No registration error correction			Proposed Algorithm with initial registration errors		
		Scenario I with $\delta = 12$	Scenario II with $\rho = 0.05$	Scenario III with $\eta = 0.05$	Scenario I with $\delta = 12$	Scenario II with $\rho = 0.05$	Scenario III with $\eta = 0.05$
0.0	25.65%	28.66%	26.87%	27.05%	28.65%	26.07%	27.12%
0.25	0.43%	4.81%	5.96%	6.45%	0.45%	0.43%	0.43%
0.5	0.039%	4.24%	5.65%	6.21%	0.039%	0.041%	0.043%
0.75	0.021%	4.19%	5.56%	6.13%	0.024%	0.032%	0.026%

Table 3: the p -values of the pairwise t -test with unequal variances of our proposed algorithm to the perfect registration cases, and no registration error correction to the perfect registration cases

β	No registration Error	No registration error correction			Proposed Algorithm with initial registration errors		
		Scenario I with $\delta = 12$	Scenario II with $\rho = 0.05$	Scenario III with $\eta = 0.05$	Scenario I with $\delta = 12$	Scenario II with $\rho = 0.05$	Scenario III with $\eta = 0.05$
0.0	1	1.5×10^{-22}	1.6×10^{-14}	4.0×10^{-18}	1.9×10^{-23}	4.0×10^{-15}	3.9×10^{-15}
0.25	1	2.0×10^{-17}	3.5×10^{-19}	3.6×10^{-18}	0.457	0.717	0.500
0.5	1	1.5×10^{-15}	2.8×10^{-17}	1.8×10^{-16}	0.712	0.167	0.401
0.75	1	1.5×10^{-14}	1.4×10^{-15}	6.2×10^{-17}	0.060	0.033	0.079

Figure 7: Examples of the resulting LCMs from our proposed algorithm (a) Scenario I with $\delta=12$ and $\sigma=1$; (b) Scenario II with $\rho=0.05$ and $\sigma=1$; and (c) Scenario III with $\eta=0.05$ and $\sigma=1$



Since at $\beta = 0.75$, our proposed algorithm achieves highest performance, we examine the effect of the initial registration errors to the performance of our algorithm by varying values of δ , ρ , and η for Scenarios I, II and III, respectively for $\beta = 0.75$. Again, ten independent runs are performed to ensure the statistical significance and the results are given in Table 4. We observe that, for all scenarios, the PMPs are roughly the same. In other words, the initial registration errors have little effect on the performance of our algorithm. These results imply the robustness of our proposed algorithm to the initial mis-registration errors if the proper value of β is chosen.

Table 4: The averaged percentages of mis-classified pixels as the function of the initial registration error for all Scenarios

Scenario I		Scenario II		Scenario III	
δ	PMP	ρ	PMP	η	PMP
0	0.019%	-0.05	0.035%	-0.05	0.036%
4	0.032%	-0.03	0.035%	-0.03	0.029%
8	0.029%	-0.01	0.022%	-0.01	0.043%
12	0.026%	0.01	0.030%	0.01	0.040%
		0.03	0.024%	0.03	0.036%
		0.05	0.032%	0.05	0.026%

Another key performance metric in this example is the residual registration errors after processing. Table 5 displays the means and standard deviations of the root mean square errors (RMSEs) from ten independent runs between each simulated images and the reference LCM. The RMSE of the n -th image is computed from

$$RMSE_n = \frac{1}{|\mathcal{T}_n|} \sum_{(u_n, v_n) \in \mathcal{T}_n} \sqrt{(x_n^{gt} - x_n^{est})^2 + (y_n^{gt} - y_n^{est})^2} \quad (41)$$

where (x_n^{gt}, y_n^{gt}) and (x_n^{est}, y_n^{est}) are the ground truth and estimated coordinates. Here, the ground truth coordinates obtained by letting $M_n = M_{perfect}$. Clear, for $\beta = 0.25, 0.5$, and 0.75 , our algorithm can successfully registered all images with the LCMs. However, at $\beta = 0$ our algorithm cannot align these images with the LCM. The results in Table 5 emphasize the importance of parameter selection. Note here that the RMSE of Image 1 is not shown in the Table 5 since it is assumed to be perfectly aligned (registration error is zero.) with the LCM. Next, we examine the effect of image noise to the registration accuracy by varying the noise variance σ^2 from -30dB to 0 dB and the resulting averaged RMSEs for $\beta = 0.0$ and 0.75 are given in Table 6 and Table 7, respectively. We observe here that there are slight performance differences in term of the RMSEs for σ^2 of -30, -20 and -10 dB for both $\beta = 0.00$ and 0.75 . However, for the noise variance equal to 0 dB, our algorithm can only correctly aligned Images 2-4 to the LCM at $\beta = 0.00$. This result emphasizes the importance of a parameter selection to the convergence of our algorithm. For the performance comparison, we compare the registration accuracy of our proposed algorithm for various scenarios and $\beta = 0.75$ with a traditional image-to-image registration technique. Here we employ the mean square error criteria (MSEC) [25] since the MSEC is suitable for register images with the same modality and suffered from additive Gaussian noise. For the traditional image-to-image registration, we register Images 2-4 with Image 1 since Image 1 is assumed to be aligned with the LCM. The averaged RMSEs from ten independence runs for various noise variances are given in Table 8. Again the particle swarm optimization algorithm with eighty particles is employed to ensure global optimality. As expected, the registration accuracy decreases as the noise variance increase. By comparing Tables 6 and 8, the RMSEs from our proposed algorithm seem to be lower (better) than those obtained from the MSEC for noise variances equal to -20, -10 and 0 dBs. Next, we again perform the pairwise t -test to determine whether there are significant differences in RMSEs obtained from our proposed algorithm and the MSEC, and the resulting p -values [20] are shown in Table 9. From the p -values, we can conclude that our proposed algorithm achieves significantly better registration accuracies than those obtained from the MSEC for the noise variances of -20, -10 and 0 dBs. Note here that, for a noise variance equal to -30 dB, the registration errors from our proposed algorithm and the MSEC are roughly zero and, therefore there is no different in term of registration accuracy.

Table 5: The residue registration errors of our proposed algorithm for various scenarios and values of β .

Scenario			No registration Error Correction	$\beta = 0.0$	$\beta = 0.25$	$\beta = 0.50$	$\beta = 0.75$
I ($\delta = 12$)	Image 2	Mean	12	11.99	0.111	0.295	0.280
		STD	-	0.0015	0.259	0.139	0.100
	Image 3	Mean	12	11.99	0.031	0.192	0.312
		STD	-	0.0018	0.020	0.120	0.156
	Image 4	Mean	16.97	16.96	0.213	0.338	0.212
		STD	-	0.0017	0.566	0.088	0.136
II ($\rho = 0.05$)	Image 2	Mean	14.06	13.56	0.028	0.281	0.327
		STD	-	0.072	0.010	0.130	0.113
	Image 3	Mean	14.06	13.49	0.020	0.353	0.312
		STD	-	0.032	0.080	0.102	0.106
	Image 4	Mean	21.97	20.97	0.253	0.245	0.315
		STD	-	0.095	0.636	0.120	0.082
III ($\eta = 0.05$)	Image 2	Mean	14.76	14.71	0.025	0.295	0.296
		STD	-	0.204	0.020	0.149	0.098
	Image 3	Mean	14.76	14.73	0.017	0.415	0.350
		STD	-	0.182	0.006	0.090	0.136
	Image 4	Mean	21.72	22.04	0.350	0.312	0.371
		STD	-	0.0325	0.983	0.155	0.088

Table 6: The residue registration errors for various noise variances and $\beta = 0.75$.

Noise variance (dB)	Average root mean square errors								
	Scenario I, $\delta = 12$			Scenario II, $\rho = 0.05$			Scenario III, $\eta = 0.05$		
	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4
-30	0.007	0.011	0.009	0.006	0.010	0.019	0.012	0.019	0.013
-20	0.010	0.012	0.009	0.023	0.016	0.012	0.017	0.016	0.011
-10	0.036	0.035	0.037	0.028	0.018	0.029	0.028	0.030	0.022
0	0.244	0.280	0.185	0.119	0.138	0.071	0.078	0.053	0.200

Table 7: The residue registration errors for various noise variances and $\beta = 0$.

Noise variance (dB)	Average root mean square errors								
	Scenario I, $\delta = 12$			Scenario II, $\rho = 0.05$			Scenario III, $\eta = 0.05$		
	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4
-30	0.016	0.08	0.010	0.015	0.007	0.019	0.009	0.011	0.019
-20	0.017	0.012	0.014	0.015	0.018	0.015	0.010	0.015	0.017
-10	0.014	0.018	0.015	0.018	0.018	0.023	0.019	0.016	0.014
0	11.99	11.99	16.97	11.91	11.89	20.28	12.75	12.79	20.61

Table 8: The residue registration errors using the minimum mean square error criteria for various noise variances.

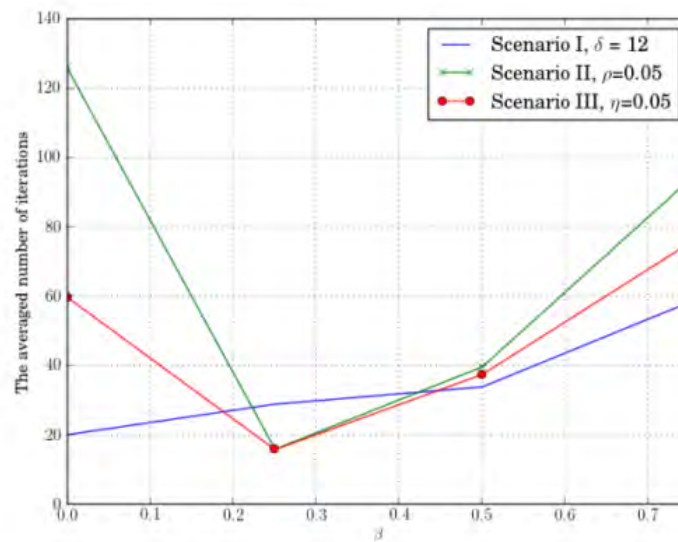
Noise variance (dB)	Image 2		Image 3		Image 4	
	Mean	STD	Mean	STD	Mean	STD
-30	0.008	0.0029	0.007	0.0041	0.010	0.0054
-20	0.422	0.0040	0.425	0.0033	0.423	0.0049
-10	0.663	0.0037	0.665	0.0014	0.664	0.0017
0	0.875	0.516	1.637	1.441	1.352	0.9744

Table 9: The p -value from the pairwise t-test between the traditional registration method and our proposed algorithm for various Scenario at $\beta = 0.75$.

Noise variance (dB)	Average root mean square errors								
	Scenario I, $\delta = 12$			Scenario II, $\rho = 0.05$			Scenario III, $\eta = 0.05$		
	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4	Image 2	Image 3	Image 4
-30	0.829	0.402	0.883	0.413	0.413	0.201	0.507	0.092	0.407
-20	1×10^{-18}	4×10^{-14}	2×10^{-21}	1×10^{-13}	1×10^{-13}	3×10^{-15}	2×10^{-13}	2×10^{-13}	5×10^{-17}
-10	3×10^{-14}	2×10^{-14}	3×10^{-14}	3×10^{-15}	3×10^{-15}	5×10^{-16}	2×10^{-23}	1×10^{-14}	7×10^{-17}
0	0.004	0.016	0.004	0.001	0.001	0.003	0.0010	0.007	0.004

Figure 8 shows the averaged numbers of iterations that the algorithm requires before the convergence criterion is satisfied for different scenarios and β . For $\beta = 0.25, 0.5$ and 0.75 , more iterations are needed as the value of β increases. However, at $\beta = 0$, our algorithm terminates at the higher numbers of iterations for Scenarios II and III. The main reason to the slow convergences is due to the small changes in the mapping parameters from one iteration to another and since $\beta = 0$, this small changes in the mapping parameters have significant influence on the posterior probability.

Figure 8: The averaged number of iterations required before the termination criteria are satisfied for different scenarios in Example 1



Experiment 2

A QuickBird dataset consisting of one multispectral image (MI) of size 150×300 pixels and one panchromatic image (PAN) of size $600 \times 1,200$ pixels was used in this experiment (Figure 9). The MI and PAN have resolutions of 2.4 and 0.6 meters, respectively. Both images captured a part of Kasetsart University in Bangkok, Thailand, covering around 0.2592 km^2 on July 10th, 2008. By visual interpretation, we classified the area into five classes, namely, water, shadows, vegetations and impervious type 1 and impervious type 2, and the ground truth image is shown in Figure 10 where blue, black, green, red and white colors correspond to water, shadow, vegetation, impervious type 1 and impervious type 2, respectively. Here, the impervious is divided into two types due to different roof and pavement colors in the scene. By using both PAN and MI images, we randomly select 1000 samples for each land cover classes.

Figure 9: QUICKBIRD dataset of a part of Kasetsart University (a) False color composite MI; and (b) PAN

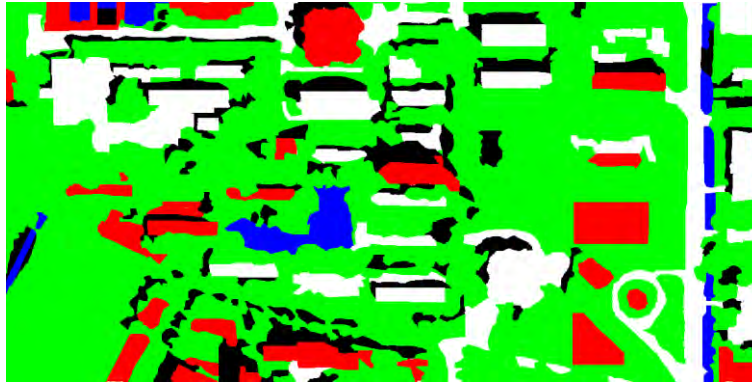


(a)



(b)

Figure 10: Ground truth image for Example 2 (green, blue, black, red and white colors for vegetation, water, shadow, impervious type 1 and impervious type 2, respectively)



In Experiment 2, we focused on the robustness of the proposed algorithm with different degrees of the initial displacement, scale and rotation errors. In fact, there are six displacement errors in x -direction and y -directions, four scale errors and six rotational errors used in this experiment. The termination criteria used in this example is similar to those in Example 1, i.e., our algorithm is terminated if $p_{changes}$ (see Eq. (38)) is less than 10^{-5} and $d_{movement,MI}$ (see Eq. (39)) is less than 0.1 pixels for five consecutive iterations. Before examining the robustness of our algorithm, we determined the benchmark performance of the MRF-based land cover mapping when MI and PAN are perfectly registered. The resulting LCMs are shown in Figure 11. Again, as we progress to greater values of β , more connected LCMs are obtained. The overall accuracy graph shown in Figure 12 agrees with the visual inspection that the classification performance increases as the values of β increases. In this example, we employ the overall accuracy rather than the percentages of misclassified pixels used in Example 1 since overall accuracy is more widely used performance metrics in remote sensing image classification.

Figure 11: LCMs for the perfect registration case for (a) $\beta = 0$; (b) $\beta = 0.25$; (c) $\beta = 0.50$; and (d) $\beta = 0.75$

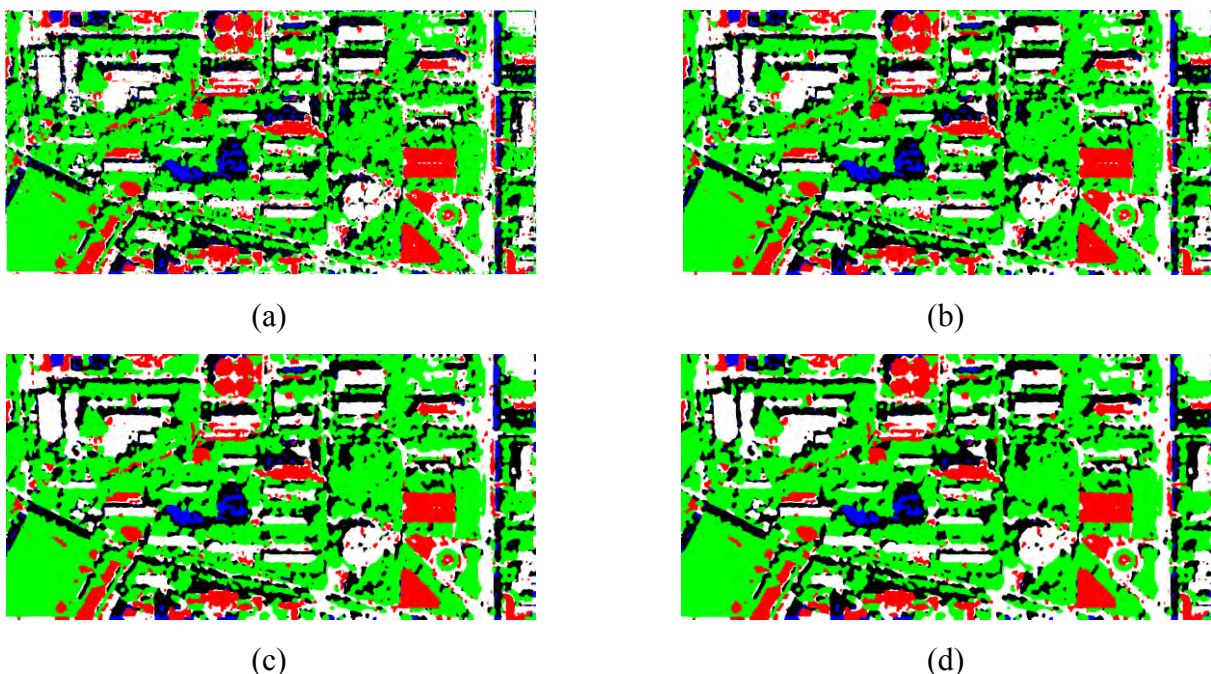
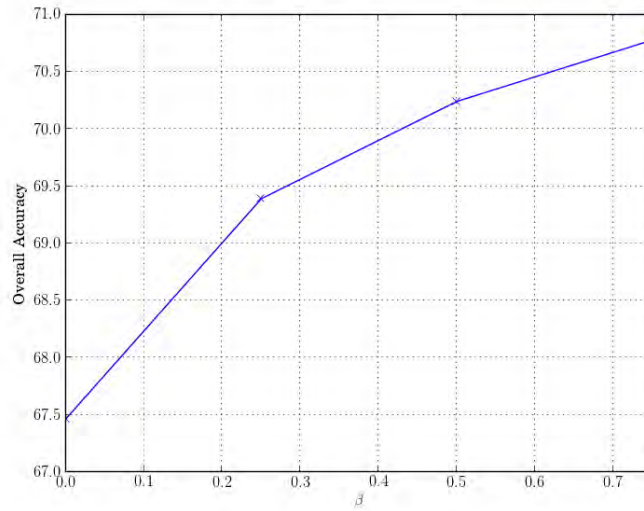


Figure 12: Overall accuracies for different values of β when MI and PAN are perfectly aligned

Since the PAN has a higher resolution, we assume that it is aligned with the LCM, and we only need to find map parameters of the MI. Here, the PAN has a higher resolution than the multispectral image by the factor of four, and both MI and PAN are obtained from the same satellite, the optimal map parameter vector relating the two images should be equal to

$$M^{opt} = [0.25, 0, 0, 0.25, 0, 0].$$

To investigate the robustness of our algorithm to displacement, scale and rotation errors, we introduce the registration errors in displacement in x -direction, displacement in y -direction, scale and rotation into the MI and PAN pair. If we introduce the displacements into the image pair, the initial map parameter is set to be equal to

$$M^0 = [0.25, 0, 0, 0.25, \Delta x, 0]$$

and

$$M^0 = [0.25, 0, 0, 0.25, 0, \Delta y],$$

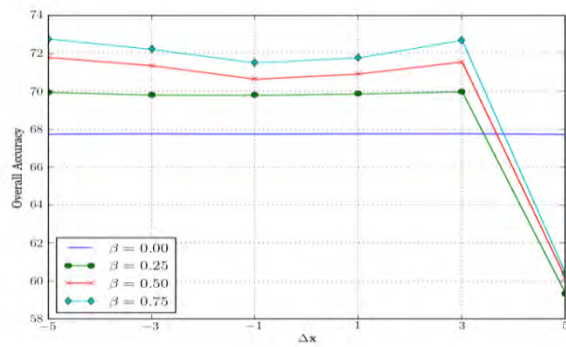
for displacement errors in x - and y -direction, respectively. Here, the values of Δx and Δy are set to be $-5, -3, -1, 1, 3$, and 5 . For the scale and rotation errors, we assume that the both images are aligned at mid points, i.e., the pixel (75,150) of MI is at a pixel (300,600) of PAN. Next, we apply the initial scale errors, Δs , of $-5\%, -2.5\%, 0\%, 2.5\%$ and 5% when comparing with the PAN image to the multispectral image. Here, the scale errors of $-5\%, -2.5\%, 0\%, 2.5\%$ and 5% correspond to the initial scales of PAN to MI of $3.8, 3.9, 4.0, 4.1$ and 4.2 , respectively. For rotation errors, we rotate the MI by $\Delta\theta$ degrees in the counter clockwise direction. Here, the initial rotation errors $\Delta\theta$ are set to be $-3^\circ, -2^\circ, -1^\circ, 1^\circ, 2^\circ$, and 3° . The initial $RMSE_{MI}$ (see Eq. (41)) for all cases are given in Table 10. Again, if our algorithm performs perfectly, the estimated map parameter will converge back to M^{opt} . In other words, we will eventually have $M^t = M^{opt}$. Once the correct map parameter vector is obtained, the classification accuracies of the LCMs should be equal to that in the perfect registration cases (Figure 11 (a)-(d)). In this example, we again assign $p^{MF}(x_s|Y, M^0) = \frac{1}{5}$, the most extreme case where no prior information is given.

Table 10: The initial $RMSE_{MI}$ in meters (pixels in LCM) for various cases in Example 2

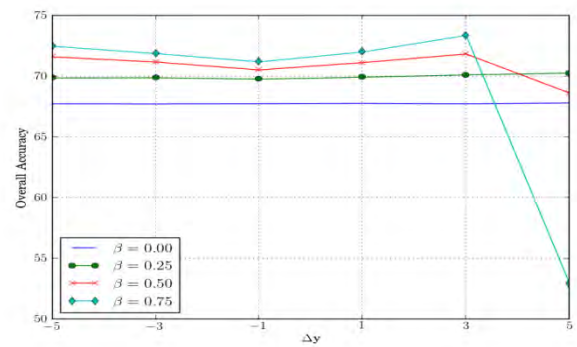
Error in x- direction		Error in y- direction		Error in scale		Error in Rotation	
Δx	$RMSE_M$	Δy	$RMSE_M$	Δs	$RMSE_M$	$\Delta \theta$	$RMSE_M$
-5	12 (20)	-5	12 (20)	-5%	21.3 (36)	-3	11.12 (19)
-3	7.2 (12)	-3	7.2 (12)	-2.5%	10.7 (18)	-2	7.45 (12)
-1	2.4 (4)	-1	2.4 (4)	0%	0.0 (0)	-1	3.72 (6.2)
1	2.4 (4)	1	2.4 (4)	2.5%	10.7 (18)	1	3.72 (6.2)
3	7.2 (12)	3	7.2 (12)	5%	21.3 (36)	2	7.45 (12)
5	12 (20)	5	12 (20)			3	11.12 (19)

The overall accuracies as the function of Δx , Δy , Δs , and $\Delta \theta$ are shown in Figure 13 (a)-(d), respectively. From all most all scenarios, the overall accuracies increase as the value of β increase since the MRF model promote more connected land cover maps, and, therefore, remove the isolated misclassified pixels. However, for $\Delta x = 5$ and $\Delta y = 5$, the overall accuracies of our algorithms decrease as β increases. The main reason to these performance degradations are due to the fact that our algorithm terminate on one of the local optima since the EM algorithm employed in our work cannot guarantee the global optimum solution. The evidence can be seen in Figures Figure 14 (a) and (b). In Figure 14 (a), we observe that, in almost all of the initial values of Δx , the number of iterations increases as the value β increases. However, for $\Delta x = 5.0$, our algorithm terminates at only 75, 109 and 129 iterations for $\beta = 0.25, 0.5$ and 0.75 whereas, for $\beta = 0.0$, our algorithm terminates after 180 iterations. Similarly, we observe the same phenomenal in Figure 14 (b) for $\Delta y = 5.0$ where the algorithm terminates at the lower number of iterations for $\beta = 0.75$ than $\beta = 0.0, 0.25$ and 0.50 . This result shows the effect of the initial registration errors to the convergence of our algorithm. In most practical situation, such an large initial registration errors is unlikely to occur since most remote sensing images are embedded with coordinate information from a producer.

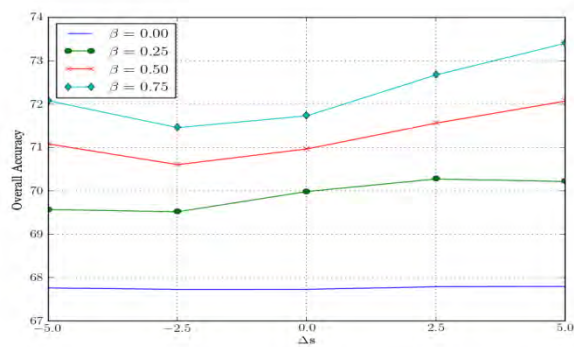
Figure 13: The effect of initial registration errors to the overall accuracies



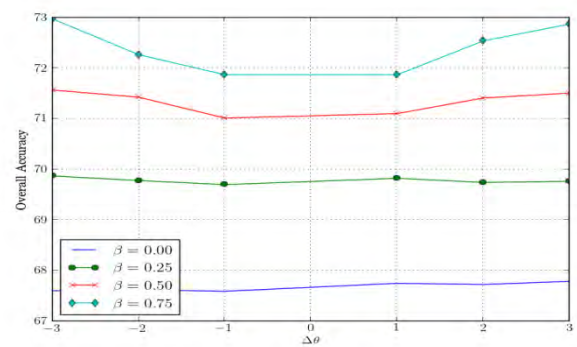
(a)



(b)

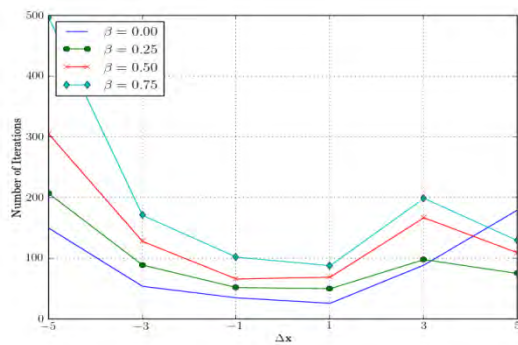


(c)

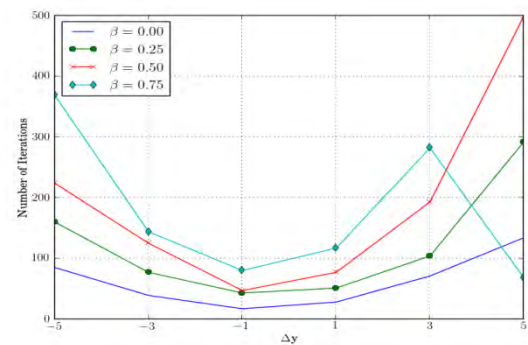


(d)

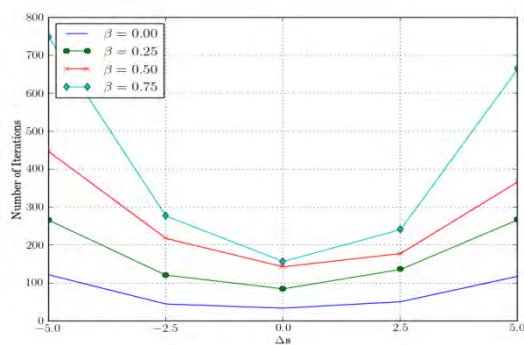
Figure 14: The effect of the initial registration errors to the number of iterations



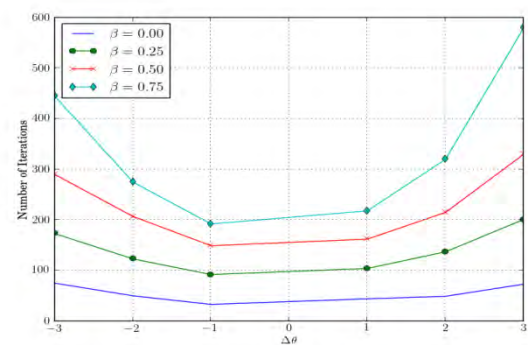
(a)



(b)



(c)



(d)

Similar to previous example, we also compare the performance of our proposed algorithm (PA) with two extreme cases where images are perfectly registered (PR) and there is no registration error correction (NC) and the results are given in Tables 11–14. From this comparison, we observe that if our algorithm converges to the global optimum solutions, the resulting overall accuracies from our proposed algorithm are similar to those of the perfect registration cases, and the significant improvements are obtained from the cases where there is no registration error correction. The maximum performance improvements from no registration error correction for each cases are 12.6% for $\Delta x = -5$ and $\beta = 0.75$, 12.4% for $\Delta y = -5$ and $\beta = 0.75$, 17.4% for $\Delta s = -5\%$ and $\beta = 0.75$, and 14.9% for $\Delta\theta = 3^\circ$ and $\beta = 0.75$. We observe that the maximum improvements are achieved at $\beta = 0.75$. This observation suggests that a higher performance gain can be obtained by increasing the value of β . Next, we also notice that our proposed algorithm can sometimes achieve even higher accuracies than those of the perfect registration cases. The reason is due to the fact that our algorithm requires more iterations than the scenarios where image pair is perfectly registered since our algorithm terminates if both the estimated map parameters and the resulting LCM converge whereas, in the perfect registration case, the process terminates if only the resulting LCM converges. Hence, our algorithm may terminate at lower percentages of changes in the LCM, and result in more accurate LCM which results in higher precision.

Table 11: Overall accuracies for different values of β on two extreme cases and our proposed algorithm for different initial displacement error in x-direction Δx where PA and NC denote the cases of the proposed algorithm and no registration error correction, respectively.

β	PR	$\Delta x = -5.0$		$\Delta x = -3.0$		$\Delta x = -1.0$		$\Delta x = 1.0$		$\Delta x = 3.0$		$\Delta x = 5.0$	
		PA	NC	PA	NC	PA	NC	PA	NC	PA	NC	PA	NC
0.0	67.5	67.7	57.6	67.8	62.2	67.7	66.9	67.8	66.7	67.7	61.8	67.8	57.0
0.25	69.4	70.0	58.8	69.8	63.7	69.8	68.6	69.9	68.3	70.0	63.4	59.3	58.4
0.5	70.3	71.8	59.7	71.4	64.6	70.6	69.6	70.9	69.2	71.5	64.4	60.2	59.2
0.75	71.1	72.8	60.2	72.2	65.2	71.5	70.3	71.8	70.0	72.7	65.0	60.4	59.9

Table 12: Overall accuracies for different values of β on two extreme cases and our proposed algorithm for different initial displacement error in y-direction Δy where PA and NC denote the cases of the proposed algorithm and no registration error correction, respectively.

β	PR	$\Delta y = -5.0$		$\Delta y = -3.0$		$\Delta y = -1.0$		$\Delta y = 1.0$		$\Delta y = 3.0$		$\Delta y = 5.0$	
		PA	NC	PA	NC	PA	NC	PA	NC	PA	NC	PA	NC
0.0	67.5	67.7	57.6	67.7	62.2	67.7	66.9	67.7	66.7	67.7	61.8	67.8	57.0
0.25	69.4	69.9	58.8	69.9	63.7	69.8	68.6	70.1	68.3	70.1	63.4	70.3	58.4
0.5	70.3	71.6	59.7	71.2	64.6	70.5	69.6	71.8	69.2	71.8	64.4	68.6	59.2
0.75	71.1	72.5	60.1	71.9	65.2	71.2	70.3	73.4	70.0	73.4	64.9	62.9	59.9

Table 13: Overall accuracies for different values of β on two extreme cases and our proposed algorithm for different initial scale error Δs where PA and NC denote the cases of the proposed algorithm and no registration error correction, respectively.

β	PR	$\Delta s = -5\%$		$\Delta s = -2.5\%$		$\Delta s = 0\%$		$\Delta s = 2.5\%$		$\Delta s = 5\%$	
		PA	NC	PA	NC	PA	NC	PA	NC	PA	NC
0.0	67.5	67.8	52.7	67.7	61.0	67.7	67.5	67.8	64.9	67.8	57.8
0.25	69.4	69.6	53.4	69.5	62.4	70.0	69.4	70.3	66.1	70.2	58.9
0.5	70.3	71.1	54.2	70.6	63.3	71.0	70.3	71.6	67.0	72.1	59.7
0.75	71.1	72.1	54.7	71.5	64.2	71.1	71.1	72.7	67.6	73.4	60.1

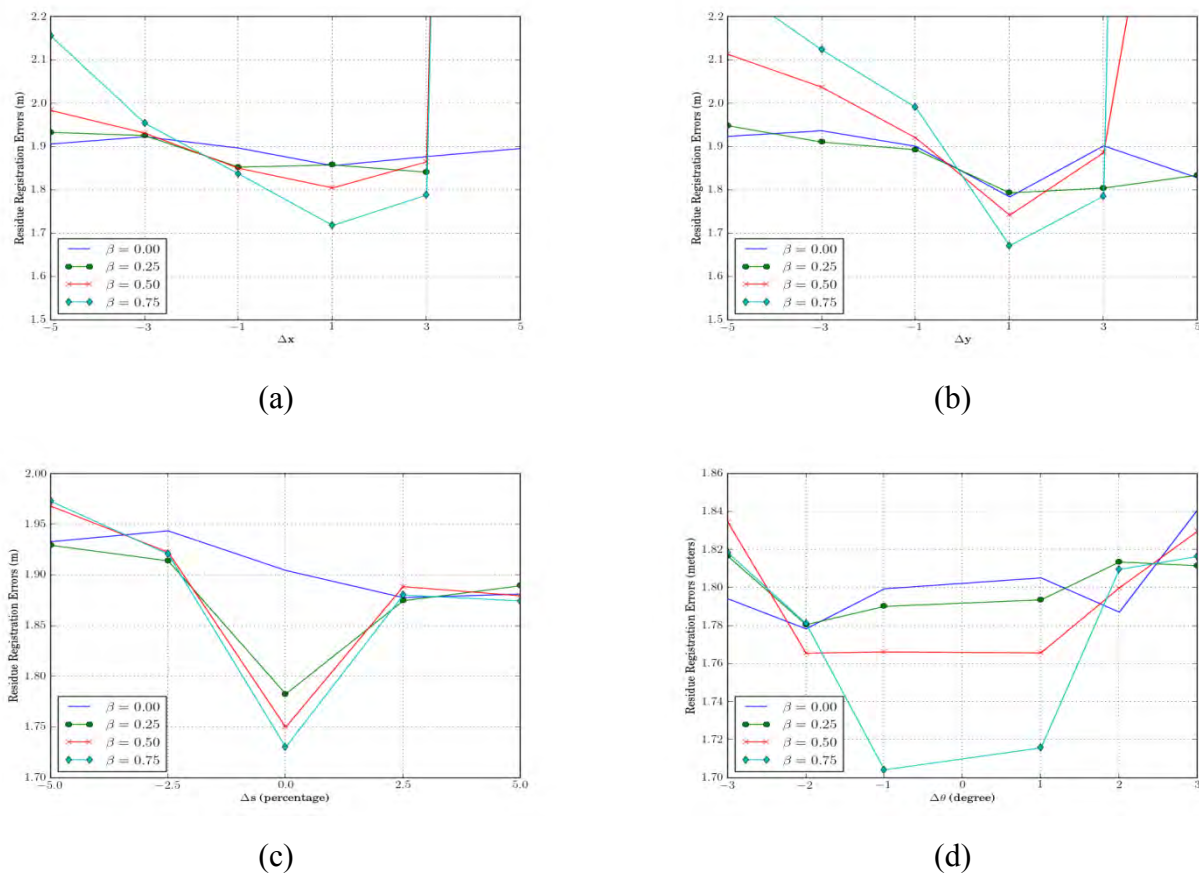
Table 14: Overall accuracies for different values of β on two extreme cases and our proposed algorithm for different rotation error $\Delta\theta$ where PA and NC denote the cases of the proposed algorithm and no registration error correction, respectively.

β	Perfect registration	$\Delta\theta = -3^\circ$		$\Delta\theta = -2^\circ$		$\Delta\theta = -1^\circ$		$\Delta\theta = 1^\circ$		$\Delta\theta = 2^\circ$		$\Delta\theta = 3^\circ$	
		PA	NC	PA	NC	PA	NC	PA	NC	PA	NC	PA	NC
0.0	67.5	67.6	57.3	67.6	60.8	67.6	65.3	67.7	64.8	67.7	59.8	67.8	55.5
0.25	69.4	69.9	58.5	69.8	62.2	69.7	66.9	69.9	66.5	69.7	61.1	69.8	56.6
0.5	70.3	71.6	59.3	71.4	63.0	71.0	67.8	71.1	67.4	71.4	62.0	71.5	57.4
0.75	71.1	73.0	59.7	72.3	63.6	71.9	68.4	71.9	68.1	72.5	62.6	72.9	58.0

Another key performance metric of our algorithm is the resulting registration errors. Figure 15(a)-(d) show the residue registration errors in term of RMSE (in meters) between the MS image and the LCM for different initial registration errors. We observe that, if our algorithm converges to the global optimum solutions, it can successfully reduce the registration error down to around 1.8 meters in the LCM (or equivalently 0.75 pixels on MS image and 3 pixels on PAN image and LCM). These results imply that our algorithm can align images together to the accuracy less than those of the lowest resolution (here is MS image). For each initial registration error cases, the minimum RMEs of 1.718 (2.86 pixels in the LCM) meters for $\Delta x = 1.0$, 1.672 (2.79 pixels in the LCM) meters for $\Delta y = 1.0$, 1.730 (2.88 pixels in the LCM) meters for $\Delta s = 0\%$ and 1.704 (2.84 pixels in the LCM) meters for $\Delta\theta = -1^\circ$ occurs at $\beta = 0.75$. These results suggest that, if our algorithm converges, the larger value of β increases the accuracy of registration as well as the classification. However, for the cases of $\Delta x = 5.0$ and $\Delta y = 5.0$, our algorithm cannot register the MS image to the LCM since our algorithm is stuck in one of the local optima. The residue registration errors for $\Delta x = 5.0$ are 1.896 (3.16), 10.96 (18.3), 11.14 (18.6), and 11.41 (19.0) meters (pixels in the LCM), and for $\Delta y = 5.0$ are 1.827 (3.05), 1.834 (3.06), 3.133 (5.22) and 11.57 (19.3) meters (pixels in the LCM) for $\beta = 0.0, 0.25, 0.50$, and 0.75 , respectively. Here, the initial displacement error corresponds to the RME of 20 pixels in the LCM. Such a large initial RME are only found when remote sensing images have significant different in spatial resolutions. LCMs derived from remote sensing image dataset with such a large scale difference are often unreliable and not often found in practice.

For performance comparison, we apply the normalized cross correlation method [25] to register PAN and MS images together and the resulting RME is equal to 1.836 meters or 3.06 pixels in the LCM. From Figures 15(a)–(d), we found that, with proper parameter selections and the initial registration errors, our proposed algorithm can achieved higher registration accuracy than those from the normalized cross correlation method. For example, our algorithm obtains the registration errors of 1.718 (2.86 pixels) meters for $\beta = 0.75$ and $\Delta x = 1$, 1.671 (2.79 pixels) meters for $\beta = 0.75$ and $\Delta y = 1$, or 1.702 (2.84 pixels) meters for $\beta = 0.75$ and $\Delta \theta = -1^\circ$.

Figure 15: The effect of the initial registration errors to the residue registration error of our proposed algorithm in Example2



6. Conclusion

In this paper, we propose a joint image registration and land cover mapping algorithm based on a Markov random field model. The algorithm assumes that observed remote sensing images are derived from a hidden LCM and captured with an unknown misalignment. Two adjacent pixels of the LCM are more likely to belong to the same land cover class than different classes. By integrating this fact into the model, a large number of misclassified pixels, which often appear as isolated pixels, are removed from the resulting LCM. Since the map parameter vector relating the different images is unknown, we employ the EM procedure to simultaneously estimate the map parameters and use mean field theory to approximate the posterior probability.

We performed an experimental study using one simulated dataset, and one real remote sensing data set of 2.4m QUICKBIRD multispectral and 0.6m QUICKBIRD panchromatic images. Our results show that, for the first data set, our algorithm can successfully classify image pairs and align them in different initial registration errors with proper selection of the MRF parameter. In fact, if the MRF parameter is chosen properly, our algorithm can classify mis-registered image pair with similar accuracy to the situation where images are perfectly aligned. For the real remote sensing dataset, we focused the investigation on the robustness of our algorithm to the initial alignment of image pair. The study showed that our algorithm is less sensitive to the initial alignment when value of the MRF parameter, β is small since the EM algorithm tends to converge faster. However, if the degree of misalignment is beyond a certain level, our algorithm cannot estimate the map parameter vector accurately since the EM algorithm employed here tends to become trapped in a local optimum. Hence, in the future we plan to investigate how to incorporate another variation of the EM algorithm that can escape from local optima in order to make our algorithm more robust.

Acknowledgments

This work is supported in part by the Thailand Research Fund under Grant RSA5480031.

Conflict of Interest

The authors declare no conflict of interest.

References and Notes

- 1 Lombardo P., Oliver C.J., Pellizzeri T.M., Meloni M., A New Maximum-Likelihood Joint Segmentation Technique For Multitemporal SAR and Multiband Optical Images, *IEEE Transactions on Geoscience and Remote Sensing*, **2003**, 41, 2500-2518.
- 2 Skriver H., Mattia F., Satalino D., Balenzano A., Pauwells V.R.N., Verhoest N.E.C., Davison M., Crop Classification Using Short-Revisit Multitemporal SAR Data, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **2011**, 4, 423-431.
- 3 Baghdadi N., Todoroff P., Hjj M. E., Begue A., Potential of SAR Sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for Monitoring Sugarcane Crops on Reunion Island, *Remote Sensing of Environment*, **2009**, 113, 1724-1738.
- 4 Mahapatra D., Sun Y., Integrating Segmentation Information for Improved MRF-Based Elastic Image Registration, *IEEE Transactions on Image Processing*, **2012**, 21, 170-183.
- 5 Chen S., Guo Q., Leung H., Bosse E., A Maximum Likelihood Approach to Joint Image Registration and Fusion, *IEEE Transactions on Image Processing*, 2011, 1363-1372.
- 6 Winkle G., *Image Analysis Random Fields and Dynamic Monte Carlo Methods*, Springer-Verlag, New York, 1995.
- 7 Bremaud P., *Markov Chains Gibbs Field, Monte Carlo Simulation, and Queues*, Springer-Verlag, New York, 1999.
- 8 Geman S., Geman, D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1984, 6, 721-741.

- 9 Solberg A.H.S., Taxt T., Jain, A.K., A Markov Random Field Model for Classification of Multisource Satellite Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, **1996**, 34, 100-113.
- 10 Kasetkasem T., Varshney P.K., An Image Change Detection Algorithm Based on Markov Random Field Models, *IEEE Transactions on Geoscience and Remote Sensing*, **2002**, 40, 1815-1823.
- 11 Bruzzone L., Prieto D.F., Automatic Analysis of the Difference Image for Unsupervised Change Detection, *IEEE Trans. Geoscience and Remote Sensing*, **2002**, 38, no. 3, 1171-82.
- 12 Xie H., Pierce L.E., Ulaby F.T., SAR Speckle Reduction Using Wavelet Denoising and Markov Random Field Modeling, *IEEE Trans. on Geoscience and Remote Sensing*, 2002, 40, 2196-2212.
- 13 Tso B.C.K., Mather P.M., Classification of Multisource Remote Sensing Imagery Using a Genetic Algorithm and Markov Random Fields, *IEEE Transactions on Geoscience and Remote Sensing*, **1999**, 37, 1255-1260.
- 14 Nishii R., A Markov Random Field-Based Approach to Decision-Level Fusion for Remote Sensing Image Classification, *IEEE Transactions on Geoscience and Remote Sensing*, **2003**, 41, 2316-2319..
- 15 Sarkar A., Biswas M.K., Kartikeyan B., Kumar VMajumder, K. L, Pal D.K., A MRF Model-Based Segmentation Approach to Classification for Multispectral Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, **2002**, 40, 1102-1113.
- 16 Dempster A.P., Laird N.M., Rubin D.B., Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. Roy. Soc. Stat.. ser. B.*, **1997**, 1-38.
- 17 Tanaka T., Mean Field Theory of Boltzmann Machine Learning, *Phys. Rev. E*, **1998**, 58, 2302–2310.
- 18 Tanaka T., A Theory of Mean Field Approximation, *12th Annual Conference on Neural Information Processing Systems (NIPS'98)*, Denver, Dec. **1998**.
- 19 Zhang J., The Mean Field Theory in EM Procedures for Markov Random Fields, *IEEE Transactions on Signal Processing*, **1992**, 40, 2570-83..
- 20 Bickel P.J., Doksum K.A., *Mathematical Statistics*, Prentice Hall, New Jersey, **1977**.
- 21 Van Trees H.L., *Detection, Estimation, and Modulation Theory*, Wiley, New York, **1968**.
- 22 Varshney P.K., *Distributed Detection and Data Fusion*, Springer, New York, 1997.
- 23 Kennedy J., Eberhart R., Particle Swarm Optimization, *Proceedings of IEEE International Conference on Neural Networks IV*, **1995**, 1942-1948.
- 24 Trelea I.C., The particle swarm optimization algorithm: convergence analysis and parameter selection, *Image Processing Letters*, **2003**, 85, 317-325.
- 25 Zitova B. and Flusser J., Image registration: A survey, *Image and computing*, **2003**, 21, 977-1000.

Research Article**Fusion and Registration of THEOS Multispectral and Panchromatic Images**

Tanakorn Sritarapipat [†], Teerasit Kasetkasem ^{* ‡}, Preesan Rakwatin [†]

[†]Geo-Informatics and Space Technology Development Agency (Public Organization), 120 The Government Complex, Building B, 7th Floor, Chaeng Wattana Road, Lak Si, Bangkok, 10210, Thailand

[‡]Faculty of Engineering, Kasetsart University, 50 Pahonyothin Road, Lat Yao, Chatuchak, Bangkok, 10900, Thailand

(Received Day Month Year; in final form Day Month Year)

This paper presents a new method for fusion and registration of THEOS (Thailand Earth Observation Satellite) multispectral and panchromatic images in a single step. In the usual procedure, fusion is an independent process separated from registration. However, both image registration and fusion can be formulated as estimation problems. Hence, the registration parameters can be automatically tuned so that both fusion and registration can be optimized simultaneously. Here, we concentrate on the relationship between low-resolution multispectral and high-resolution panchromatic imagery. The proposed technique is based on a statistical model. It employs the maximum a posteriori (MAP) estimator to solve the fusion problem, and applies the Metropolis algorithm to solve the joint optimization problem. A closed-form solution to find the fused multispectral image with correcting registration is also derived here. In our experiment, a THEOS multispectral image with high spectral resolution and a THEOS panchromatic image with high spatial resolution are combined to produce a multispectral image with high spectral and spatial resolution. The results of our experiment show that the proposed fusion and registration algorithm can produce high quality high-resolution multispectral images from low-resolution multispectral and high-resolution panchromatic images even when they are severely mis-registered.

1. Introduction

Many applications such as vegetation mapping, environmental monitoring, mineral mapping,

^{*} Corresponding author. Email: fengtsk@ku.ac.th

oil exploration, hydrology, disaster response and agricultural yield prediction cannot be successful without the information provided by remote sensing images. The use of remote sensing data is increasing due to the increasing availability of high quality images with a reasonable cost and the declining price of computational power.

Current applications continue to demand data with higher spatial and spectral precision to generate even more spatial detail in maps with many different land cover classes. These requirements can be fulfilled either by building new satellites with superior resolution or by utilizing image processing techniques. The main advantages of the second approach are its lower cost and greater timeliness. It is not necessary to wait for a new satellite to be designed, built and launched.

The fusion of multi-sensor image data is a widely used procedure for enhancing information extraction. Fusion uses already-existing data from different sources, acquired at different times, or captured by different sensors, to produce higher resolution data. When high spatial resolution data are merged with the high spectral resolution data, the fused image has both finer spatial and finer spectral resolution.

Image fusion algorithms can be divided into three categories, namely, feature-based, pixel-based and transform-based methods. The feature-based methods first segment input sources into homogeneous regions by using some segmentation techniques (Clausi and Deng, 2004; Shi and Manduchi, 2003) according to texture characteristics. Next, a data combination algorithm combines texture information from different images together to produce a texture-enhanced image. The transformation-based methods convert the input images into a common transformed domain, such as intensity-hue-saturation (IHS) (Carper, 1990), principle components (PCA) (Chavez and Kwarteng, 1989), Brovey (Civco et al., 1995), 2D wavelet (Gomez et al., 2001;

Gonzalez et al., 2004; Amolins et al., 2007), 3D wavelet (Zhang and He, 2007), pyramids (Liu et al., 2001) or high pass filtering (Shettigare, 1992). Image fusion is performed by combining the transform coefficients from all images. After that, the combined image is transformed by the inverse-transformation into the original image space. In the pixel-based methods (Petrovic and Xydeas, 2004; Xia and Kamel, 2007), a pixel in the fused image is determined from a set of pixels from the input sources. For example, Hardie (Hardie et al., 2004) and Li (Li and Leung, 2009) applied statistical models to a pixel-based approach where intensity values from corresponding pixels from fusing images statistically depend on the intensity value of a fused image. The approach detailed in this paper is also pixel-based and use a statistical model.

Image registration is the process of transforming different sets of data into a single coordinate system. Data may come from different sensors, from different times, or from different viewpoints. Current registration methods can be classified into three categories, feature-based, transform-based, and intensity-based methods (Zitova and Flusser, 2003). The feature-based approaches usually extract geometric features, also known as control points, such as intersections and landmarks, and use a least square criterion to estimate the registration parameters (Arun et al., 1987; Umeyama, 1991). The transform-based methods work with images in the frequency domain (Reddy and Chatterji, 1996; Stone et al., 2003) to compute the registration parameters by utilizing the properties of translation and rotation under the Fourier transform. In intensity-based methods, the registration parameters are estimated by maximizing some similarity measure between pixel values of the input images. Normalized cross-correlation (Van Den Elsen, 1994), LS (Thevenaz, 1998), and maximum likelihood (ML) (Costa, 1993) are some popular criteria used to measure similarity.

Although many algorithms have been proposed for image registration and image fusion, the

operations are traditionally viewed as two independent processes. More precisely, image registration is normally performed first, followed by image fusion, which usually assumes that perfect registration has been achieved (Petrovic and Xydeas, 2004; Xia and Kamel, 2007; Zhengnd et al., 2007). However, in practice, the registration process is not guaranteed to be error-free, and this registration error can negatively affect the fusion performance. The quality of fused images degrades significantly when the multispectral and panchromatic images are not perfectly registered.

The work by Siyue (Siyue et al., 2010) attempts to join the fusion and registration processes together by using the expectation maximization (EM) algorithm. However, they assume that the intensity value in each pixel of the high spatial resolution image can be selected from any one of the fused images. In the case of spatial enhancement, the intensity values of the fused image should follow the intensity values from the multispectral image only. As a result, the algorithm proposed in (Siyue et al., 2010) is not suitable for spatial enhancement by combining multispectral and panchromatic images (so called “pan-sharpening”). In pan-sharpening, the fused image must have the same color distribution as the multispectral image but contain the spatial detail from the higher resolution panchromatic image.

In this paper, we develop the maximum a posteriori (MAP) estimator (Kay, 1993) to simultaneously fuse and register multispectral and panchromatic images to produce a high-resolution multispectral image. Here, we focus on the use of high-resolution panchromatic data to enhance multispectral imagery. However, the estimation framework developed here can be extended for any number of spectral bands in the primary and auxiliary sensors. The proposed technique is suitable for applications where some correlation, either localized or global, exists between the auxiliary image and the image being enhanced. A spatially varying statistical model

is used to help exploit localized correlations between the primary and auxiliary image. Another important aspect of the proposed algorithm is that it uses an accurate observation model relating the “true” scene to the low-resolution observations. This means that a potentially wavelength-dependent spatially-varying system point spread function (PSF) can be incorporated into the estimator.

The paper is organized as follows. Section I describes the study area and dataset for our experiment. Section II gives the details of the observation model relating low-resolution multispectral and high-resolution panchromatic images. Section III introduces the maximum a posteriori (MAP) estimator for fusing images and the Metropolis algorithm for solving the joint optimization problem. In Section IV, experimental results are presented and discussed. Finally in Section V, we offer our conclusions.

2. Observation Model

Let $X(\mathcal{S}) \in \mathbb{R}^{M \times B}$ be a fine resolution multispectral image (FRMI) having M pixels and B bands where $\mathbb{R}$ is a real number. Here, $\mathcal{S} = \{s_1, \dots, s_M\}$ denotes a set of pixels registered to some real world scene. The fine resolution multispectral image is usually represented in a vector form so that, $\mathbf{x}(s_i) \in \mathbb{R}^B$ is a fine resolution multispectral vector (FRMV) containing intensity values of a pixel s_i in the FRMI from all spectral bands. Here, we assume that the FRMVs from different pixels are statistically independent, and have identical multivariate Gaussian distribution with a mean vector, $\boldsymbol{\mu}_X$, and a covariance matrix, C_X . The marginal probability density function (PDF) of the FRMI can therefore be written as

$$\Pr(X) = \prod_{s_i \in \mathcal{S}} \frac{\exp\left[-\frac{1}{2}(\mathbf{x}(s_i) - \boldsymbol{\mu}_X)^T \mathbf{C}_X^{-1} (\mathbf{x}(s_i) - \boldsymbol{\mu}_X)\right]}{(2\pi)^{\frac{B}{2}} \sqrt{|\mathbf{C}_X|}} \quad (1)$$

We omit $\mathcal{S}$ for the sake of abbreviation. In this paper, we assume that the direct observation of the FRMI is impossible, but we, instead, observe the coarse resolution multispectral image (CRMI), and the panchromatic image (PI) of the same scene.

Let $Y(\mathcal{T}) \in \mathbb{R}^{N \times B}$ be the CRMI with $N < M$ pixels and B bands where $\mathcal{T} = \{t_1, \dots, t_N\}$ is a set of pixels belonging to the CRMI. Since both CRMI and FRMI are taken from the same scene, the relationship between $\mathcal{S}$ and $\mathcal{T}$ can be determined. Let us denote a coordinate of a pixel s in a FRMI as (x, y) where x and y are the row and column of $\mathbf{x}(s)$. Similarly, we can write $t = (u, v)$ where u and v are the row and column of the pixel t in the CRMI. Hence, the relationship between s and t can be written as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \eta \cos \theta & \eta \sin \theta \\ -\eta \sin \theta & \eta \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d_x \\ d_y \end{bmatrix} \quad (2)$$

where η , θ , d_x , and d_y are scale, rotation angle, translation in column direction and translation in row direction between a pixel coordinates η and t , respectively. We denote $\omega = [\eta \ \theta \ d_x \ d_y]^T$ as the parameter vector. In practice, since the panchromatic and multispectral images are captured from the same satellite, they are already somewhat pre-registered to each other within the predefined accuracy range. As a result, the parameter vector can be assumed to be uniformly distributed within the defined set, i.e.,

$$\Pr(\boldsymbol{\omega}) = \begin{cases} \frac{1}{|\Omega_{\omega}|}, & \omega \in \Omega_{\omega} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where Ω_{ω} is the set of all possible values in the parameter space. If the parameter vector is known, the CRMI can be remapped and resampled into the coordinate system $\mathcal{S}$. As a result, let us denote $Y_m(\mathcal{S})$ as the remapped version of CRMI into $\mathcal{S}$. Clearly both $X(\mathcal{S})$ and $Y_m(\mathcal{S})$ have equal numbers of pixels and spectral bands. However, $X(\mathcal{S})$ has finer and clear detail than $Y_m(\mathcal{S})$ since all the remapping algorithms use only information from the coarser resolution image. We model this loss of finer and clearly detail as additive noise, i.e.,

$$\mathbf{y}_m(s) = \mathbf{x}(s) + \mathbf{n}_c(s) \quad (4)$$

where $\mathbf{y}_m(s) \in \mathfrak{R}^B$ is the vector containing intensity values from all bands of the remapped CRMI, and $\mathbf{n}_c(s) \in \mathfrak{R}^B$ is the additive noise vector. We assume further that noise vectors from all pixels are independent and Gaussian distributed with a zero mean vectors and a covariance matrix, C_c . Hence, the conditional PDF of $Y(\mathcal{S})$ give $X(\mathcal{S})$ and the map parameter is given by

$$\begin{aligned} \Pr(Y|X, \boldsymbol{\omega}) &= \prod_{s \in \mathcal{S}} \Pr(\mathbf{y}_m(s) | \mathbf{x}(s)) \\ &= \prod_{s \in \mathcal{S}} \frac{\exp\left[-\frac{1}{2}(\mathbf{y}_m(s_i) - \mathbf{x}(s_i))^T C_c^{-1} (\mathbf{y}_m(s_i) - \mathbf{x}(s_i))\right]}{(2\pi)^{\frac{B}{2}} \sqrt{|C_c|}} \end{aligned} \quad (5)$$

Next, let $Z(\mathcal{S}) \in \mathfrak{R}^{M \times 1}$ be the observed panchromatic image (PI) having M pixels and only one band. Here, we assume that the PI has the same resolution as the FRMI and is perfectly

registered with the coordinate system of FRMI. We assume further that the PI is the weighted band summation of the FRMI plus an additive noise, i.e.,

$$z(s) = \mathbf{w}^T \mathbf{x}(s) + n_z(s) \quad (6)$$

where $z(s)$ is the observation of the PI at the pixel s , $\mathbf{w} = [w_1 \ \cdots \ w_B]^T$ is weight vectors, and $n_z(s)$ is the additive noise to the PI. Again, we also assume that noise values for all the pixels are independent and Gaussian distributed with a zero mean and a variance, σ_z^2 , and that they are also independent of $\mathbf{n}_c(s)$. Hence, the conditional PDF of $Z(S)$ give $X(S)$ is given by

$$\begin{aligned} \Pr(Z|X) &= \prod_{s \in S} \Pr(z(s)|\mathbf{x}(s)) \\ &= \prod_{s \in S} \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{1}{2\sigma_z^2}(z(s) - \mathbf{w}^T \mathbf{x}(s))^2} \end{aligned} \quad (7)$$

We formulate the image fusion problem as an M-ary hypothesis testing problem where each hypothesis corresponds to a different FRMI. Furthermore, since we formulate our problem as an M-ary hypothesis testing problem, techniques developed to solve signal detection problems can be employed. We provide our methodology in the next section.

3. Optimum Image Fusion

The maximum *a posteriori* (MAP) criterion (Kay, 1993) is used for solving the above problem in our work. This criterion is expressed as

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} [\Pr(X, \omega | Y, Z)] \right] \quad (8)$$

From Bayes' rule and assuming conditional independence of Y and Z given X and ω , Eq. (8) can be rewritten as

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[\frac{\Pr(Y|X, \omega) \Pr(Z|X) \Pr(X, \omega)}{\Pr(Y, Z)} \right] \right] \quad (9)$$

Since $\Pr(Y, Z)$ is independent of the choice of (X, ω) , it can be omitted and above equation reduces to

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} [\Pr(Y|X, \omega) \Pr(Z|X) \Pr(X, \omega)] \right] \quad (10)$$

By assuming further that X and ω are statistically independent, and substituting Eq. (1), Eq. (5), and Eq. (7) into Eq. (10), we have

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[\left(\prod_{s \in \mathcal{S}} \Pr(\mathbf{y}_m(s)|\mathbf{x}(s)) \Pr(z(s)|\mathbf{x}(s)) \Pr(x(s)) \right) \Pr(\omega) \right] \right] \quad (11)$$

Eq. (11) can be rewritten as

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[- \sum_s \{E_Y(\mathbf{y}_m, \mathbf{x}) + E_P(Z, \mathbf{x}) + E_X(\mathbf{x})\} - \ln |\Omega_\omega| \right] \right] \quad (12)$$

where

$$E_Y(\mathbf{y}_m, \mathbf{x}) = \frac{1}{2} (\mathbf{y}_m - \mathbf{x})^T C_C^{-1} (\mathbf{y}_m - \mathbf{x}) + \frac{B}{2} \ln 2\pi + \frac{1}{2} \ln |C_C| \quad (13)$$

$$E_z(z, \mathbf{x}) = \frac{1}{2\sigma_z^2} (z(s) - \mathbf{w}^T \mathbf{x}(s))^2 + \frac{1}{2} \ln 2\pi + \frac{1}{2} \ln \sigma_p^2 \quad (14)$$

and

$$E_X(x) = \frac{1}{2}(\mathbf{x}(s_i) - \boldsymbol{\mu}_X)^T C_X^{-1}(\mathbf{x}(s_i) - \boldsymbol{\mu}_X) + \frac{B}{2} \ln 2\pi + \frac{1}{2} \ln |C_X| \quad (15)$$

Again, we omit the term s for sake of abbreviation. Since the terms, $\frac{B}{2} \ln 2\pi$, $\frac{1}{2} \ln |C_C|$, $\frac{1}{2} \ln \sigma_P^2$,

$\frac{B}{2} \ln 2\pi$, $\frac{1}{2} \ln |C_X|$ are just constants, they can be ignored. Therefore, the optimum image fusion

becomes

$$\begin{aligned} (\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[-\frac{1}{2} \sum_s \left\{ (\mathbf{y}_m - \mathbf{x})^T C_c^{-1} (\mathbf{y}_m - \mathbf{x}) + \frac{(z - \mathbf{w}^T \mathbf{x})}{\sigma_z^2} \right. \right. \right. \\ \left. \left. \left. + (\mathbf{x} - \boldsymbol{\mu}_X)^T C_X^{-1} (\mathbf{x} - \boldsymbol{\mu}_X) \right\} - \ln |\Omega_\omega| \right] \right] \end{aligned} \quad (16)$$

The above equation can be modified to

$$(\hat{X}, \hat{\omega}) = \arg \left[\max_{(X, \omega)} \left[-\frac{1}{2} \sum_s (\mathbf{x} - \mu_s(\mathbf{y}_m, z, \boldsymbol{\mu}_X))^T C_s^{-1} (\mathbf{x} - \mu_s(\mathbf{y}_m, z, \boldsymbol{\mu}_X)) - \ln |\Omega_\omega| \right] \right] \quad (17)$$

where

$$C_s = \left[C_X^{-1} + C_C^{-1} + \frac{\mathbf{w}\mathbf{w}^T}{\sigma_z^2} \right]^{-1}, \quad (18)$$

and

$$\boldsymbol{\mu}_s(\mathbf{y}_m, z, \boldsymbol{\mu}_X) = C_s \left[C_X^{-1} \boldsymbol{\mu}_X + C_C^{-1} \mathbf{y}_m + \frac{\mathbf{w}}{\sigma_z^2} z \right] \quad (19)$$

Next, we define the energy function $E(X, \omega)$ as the negative of the argument inside Eq. (19),

i.e.,

$$E(X, \boldsymbol{\omega}) = \frac{1}{2} \sum_s (\mathbf{x} - \mu_s(\mathbf{y}_m, z, \boldsymbol{\mu}_X))^T C_s^{-1} (\mathbf{x} - \mu_s(\mathbf{y}_m, z, \boldsymbol{\mu}_X)) + \ln |\Omega_{\boldsymbol{\omega}}| \quad (20)$$

The optimum joint image fusion and registration problem becomes the energy minimization problem, i.e.,

$$(\hat{X}, \hat{\boldsymbol{\omega}}) = \arg \left[\min_{(X, \boldsymbol{\omega})} E(X, \boldsymbol{\omega}) \right] \quad (21)$$

For a given map parameter vector $\boldsymbol{\omega}$, the optimum FRMI can be obtained by taking derivative of Eq. (20) with respect to $\mathbf{x}$, and the optimum FRMV is given by

$$\mathbf{x} = \boldsymbol{\mu}_s(\mathbf{y}_m, z, \boldsymbol{\mu}_X) \quad (22)$$

We note here that our result in Eq. (22) is similar to the work by Hardie (Hardie, 2004). However, in their work, images to be fused are assumed to be perfectly aligned.

In order to obtain the map parameter, we need to find minimize Eq. (20) with respect to $\boldsymbol{\omega}$. However, the derivative of Eq. (21) with respect to the $\boldsymbol{\omega}$ map parameter vector is very difficult to obtain analytically. Furthermore, the energy function $E(X, \boldsymbol{\omega})$ is an extremely non-convex function. Hence, a gradient based approach cannot be used since the resulting map parameter will very likely become stuck in one of many local optima. Instead, we employ the Metropolis algorithm (Diaconis and Salo-Coste, 1998) for determining the optimum map parameter vector.

The Metropolis algorithm is a stochastic search method where, at each iteration, a new value of $\boldsymbol{\omega}_n$ is randomly proposed. If $\boldsymbol{\omega}_n$ results in a better fit (lower energy function), it is accepted, and the Metropolis algorithm sets $\boldsymbol{\omega} = \boldsymbol{\omega}_n$. However, even if $\boldsymbol{\omega}_n$ corresponds to a higher energy value, $\boldsymbol{\omega}_n$ is still accepted with some probability. These random moves allow the Metropolis

algorithm escape from local optima. By employing the Metropolis algorithm, the optimization algorithm can be written as

1. Set $(X, \boldsymbol{\omega}) = (X_0, \boldsymbol{\omega}_0)$, $h=1$, and $T = \frac{T_0}{\log(h+1)}$ where X_0 and $\boldsymbol{\omega}_0$ are the initial FRMI and map parameters, h is the iteration number, T_0 is the initial temperature.
2. Find a FRMI by using Eq. (23), and Let X be the resulting FRMI.
3. Use Eq. (21) to compute energy function and assign $E_{old} = E(X, \boldsymbol{\omega})$.
4. A new map parameter $\boldsymbol{\omega}_n$ is drawn from Eq. (3) and the corresponding energy function $E_n(X, \boldsymbol{\omega}_n)$ is computed using Eq. (22).
5. Assign $\boldsymbol{\omega} = \boldsymbol{\omega}_n$ with probability $P = \max\left(1, \exp\left[-\frac{E_n(X, \boldsymbol{\omega}_n) - E(X, \boldsymbol{\omega})}{T}\right]\right)$.
6. Set $h = h+1$ and $T = \frac{T_0}{\log(h+1)}$. Go to (2) if $h < h_{max}$.

We observe that as the number of iteration increases to infinity, the temperature T decreases to zero. This implies that, after a small number of iterations, the Metropolis algorithm is likely to accept almost any proposed map parameter. However, when the number of iterations becomes large, the Metropolis algorithm prefers to accept the map parameter that results in a better fit (lower energy function). The convergence of the Metropolis algorithm to the global optimum regarding the choice of the initial map parameter and FRMI is guaranteed if T_0 is sufficiently large. However, this value is often too large to be practically implemented. Nevertheless, in our problem, the initial map parameters and FRMI are not far from the global optimum since both

multispectral and panchromatic sensors are mounted on the same satellite. Figure 1 summarizes the proposed optimization algorithm in this paper.

[Figure 1 about here.]

4. Experimental Results

To examine the performance of the proposed algorithm, we used various image scenes from many parts of Thailand, namely, city (Figure 2), drought (Figure 3), agriculture (Figure 4), mountain (Figure 5) and seashore (Figure 6) areas. Table 1 summarizes the detailed information for all the scenes used in this experiment. Multispectral and panchromatic images were acquired by THEOS, an Earth observation mission of Thailand, developed at EADS Astrium SAS, Toulouse, France. THEOS was launched from Dombarovski, Russia at 06:37 GMT on 1 October 2008, by a Dnepr rocket.

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Table 1 about here.]

The THEOS imagery products include optical and near infrared spectroscopy as four

multispectral bands, plus panchromatic images. The THEOS satellite orbit parameters are summarized in Table 2. Panchromatic products provide 2 meter resolution (at nadir) and 8 bits information depth. The output scene is a square scene of 22 km. x 22 km, and covers wavelengths ranging from 450 to 900 nm as shown in Figure 7. The multispectral products provide 15 meter resolution (at nadir) and 8 bits information depth. All four bands are delivered as one file. The output scene is a square scene of 90 km. x 90 km. Wavelength ranges are 450-520 nm for the blue band, 530-600 nm for green, 620-690 nm for red, and 770-900 nm for near infrared as shown in Figure 7. Table 3 summaries the information on multispectral and panchromatic images provided by the THEOS satellite.

[Table 2 about here.]

[Figure 7 about here.]

[Table 3 about here.]

Since our algorithm performs both image fusion and registration at the same time, the performance of our algorithm can be evaluated in terms of how far off the fused image is from the actual FRMI, that is, the estimation error between the actual parameters that map between the panchromatic and multispectral image pair and the estimated ones. If our algorithm performs perfect registration and fusion, the fused image will be exactly equal to FRMI and the estimation error will be zero. In order to measure this key performance metric, we need to know both actual FRMI and the map parameters. However, since we use observed THEOS multispectral and panchromatic images in this experiment, the actual FRMI is unknown. As a result, we treat the observed multispectral image as the reference FRMI, and perform the proposed image fusion algorithm on resampled multispectral and panchromatic image pairs. In this experiment, we

resample multispectral and panchromatic images to 112.5 and 15 meters resolution, respectively. The resampled images have a resolution of 7.5 times lower than the original ones. After fusing a 112.5 meter multispectral image with a 15 meter panchromatic image, we obtain the fused multispectral image at 15 meter resolution which can be compared with the observed multispectral image.

Next, since our algorithm is designed to fuse unregistered image pairs, we would like to investigate the performance of our algorithm for different scenarios of initial registration errors. We would like to vary the relationship between the multispectral and panchromatic images in a pair by applying different values of displacements and rotations. However, since multispectral and panchromatic sensors are located at different locations on the THEOS satellite, their images are not perfectly aligned to start with. To be sure that we can precisely measure the amount of mis-registration we introduce, we manually register all multispectral and panchromatic image pairs first, using nine or more ground control points with the second-order polynomial transformation. The resulting root mean square errors were less than 0.0002 pixels for all images.

For the sake of brevity, we provide the details of our experiment only for the case of the city area image. The other cases followed the same steps. The multispectral (Figure 2) and the panchromatic (Figure 8) image pair were degraded by resampling to 112.5 (Figure 9) and 15 (Figure 10) meters resolution respectively. Then we fused the degraded images with our proposed algorithm to obtain a 15 meter-resolution multispectral image. We use the original 15-meter resolution multispectral image as the “ground truth” – that is, the (actually unobservable) FRMI. Hence, in our experiment, the resulting fused image is compared with the original multispectral image. If our algorithm performs perfect registration and fusion, the resulting mean

square error (MSE) between the resulting FRMI and the original multispectral image will be zero.

In all of our experiments, the original multispectral and panchromatic registered images are of size 750×750 and 5625×5625 , respectively. Since the resolution ratio of the multispectral to the panchromatic is $15/2$ or 7.5 , we need 7.5 times as many panchromatic pixels as multispectral pixels to cover the same spatial extent.

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

Mis-registration errors were deliberately introduced into the degraded image pairs. Here, the affine transformation was used and there were four parameters, namely scale factor η , rotation angle θ in degree, displacement in column direction d_x in pixel, and displacement in row direction d_y in pixel. Table 4 summarizes all mis-registration errors introduced in this experiment. Note here that mis-registration in the scale dimension was not investigated in this paper since images are taken from the same satellite at the same time. Figure 11 shows a panchromatic image distorted by 3 degrees of rotation, 3 pixels of displacement in the column and 3 pixels of displacement in the row.

[Table 4 about here.]

[Figure 11 about here.]

To find the appropriate benchmark for our algorithm, we first fuse the registered image pair by using Eq. (23) and compare it to the FRMI from manual registration as shown in Figure

12. We note here that our algorithm and the work by Hardie are equivalent if no registration error is considered. In other words, the benchmark scenario is when we apply Hardie algorithm directly to multispectral and panchromatic image pairs. The corresponding MSE compared to the original multispectral image was equal to 7.2465.

To generate the initial FRMI for our algorithm for the non-benchmark case, we fused a mis-registered image pair using Eq. (23) and used it as the initial FRMI. An example of the initial FRMI for the case of 3 degrees of rotation, 3 pixels of displacement in column and 3 pixels in row is illustrated in Figure 13. It is obvious that the initial FRMI appears to be blurred and does not contain any sharp edges. Next, we submit the initial FRMI to our algorithm and iteratively refine it by applying different sets of transformation parameters as selected by the algorithm. Here, we set $T_0 = 0.1$, $h_{max} = 500$ iterations. The covariance matrices of the noise introduced by resampling to create the CRMI from different spectral bands are assumed to be statistically independent with variance of 25, i.e., $C_C = 25I$ where I is the identity matrix. Figure 14 displays the resulting FRMI for the case of 3 degrees of rotation, 3 pixels of displacement in column and 3 pixels in row. By visual inspection, we can observe that the resulting FRMI looks very similar to FRMI when perfect registration is obtained.

The results summarized in Table 5 clearly show that our algorithm can move the map parameters closer to the correct values. We note here that, for all cases, our algorithm achieved lower RMSE values than those of the manual registration. These results imply that the manual registration that we performed may not have been perfect after all. Some tiny registration errors may still be present in the registered image pair.

[Figure 12 about here.]

[Figure 13 about here.]

[Figure 14 about here.]

[Table 5 about here.]

Before testing our proposed algorithm with other image scenes, we examined our hypothesis that the gradient search optimization approach cannot be applied to this problem by varying the displacement in column direction, d_x , and computing the energy function from Eq.21. The result is shown in Figure 15. We observed that there are many local optima as expected. Next, we plotted the energy function produced from our algorithm against the number of iterations as shown in Figure 16. It is obvious that our approach continuously achieved lower energy values as the number of iterations increased. This result illustrates the success of the Metropolis algorithm employed in this research.

[Figure 15 about here.]

[Figure 16 about here.]

Next, we investigated the other scenes. The resulting MSE and map parameters are summarized in Tables 6-9, for drought, agriculture, mountain, and seashore scenes. It is clear that the results from all cases are similar. Our algorithm can successfully move the map parameters closer to correct values regardless of the size of mis-registrations. Furthermore, in most cases, the resulting FRMIs from the proposed algorithm yield lower RMSE values compared to the perfect registration case. This suggests that our algorithm finds errors remaining after manual registration.

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

5. Conclusion

This paper has developed a method to simultaneously register and fuse THEOS multispectral and panchromatic images. The fusion and registration are executed in a single step.

Our algorithm is based on a statistical model. In the estimation strategy, a MAP criterion is employed for estimation from the observation model to produce the fused image. The Metropolis algorithm is used to solve the joint optimization problem. A closed-form solution of the fused images with registration is also derived in this paper.

In our experiment, we compared the performance of our proposed fusion and registration technique with a fusion algorithm without correcting registration. The experimental results confirm that our proposed fusion and registration technique provides good performance in both of spatial and spectral resolution. The root mean square errors between the simulated FRMI and the fused image are lower than for the Hardie algorithm for the case of perfect registration for all image scenes. We believe that the manual registrations of panchromatic and multispectral image pairs are not perfect and these registration errors affect the performance of the Hardie algorithm for all image scenes although our proposed algorithm and Hardie algorithm are similar for the case where no registration is performed.

Acknowledgements

This work is supported in part by the Thailand Research Fund under Grant RSA5480031.

References

- Amolins, K., Zhang, Y., Dare, P., 2007. Applications of wavelet transforms in image fusion. Urban Remote Sensing Joint Event., pp. 1-7, Apr. 2007.
- Arun, K. S., Huang, T. S., Blostein, S. D., 1987. Least-squares fitting of two 3-D point sets. Trans. Pattern Anal. Machine Intell., vol. 9, no. 5, pp. 698-700, Sept. 1987.
- Carper, W. J., Lillesand, T. M., Kiefer, R. W., 1990 The use of intensity–hue–saturation transformations for merging SPOT panchromatic and multispectral image data. Photogramm. Eng. Remote Sens., vol. 56, no. 4, pp. 459–467.
- Chavez, P. S., Kwarteng, A. Y., 1989. Extracting spectral contrast in Landsat thematic mapper image data using selective component analysis. Photogramm. Eng. Remote Sens., vol. 55, no. 3, pp. 339–348.
- Civco, D. L., Wang, Y., Silander, J. A., 1995. Characterizing forest ecosystems in Connecticut by integrating Landsat TM and SPOT panchromatic data. in Proc. ACSM/ASPRS Annu. Conv., Charlotte, NC, 1995, vol. 2, pp. 216–224.
- Clausi, D. A., Deng, H. W., 2004. Feature fusion for image texture segmentation. IEEE Proceedings of the 17th International Conference on Pattern Recognition (ICPR), vol. 1, pp. 580-583, Aug. 2004.
- Costa, W. L. S., Haynor, D. R., Haralick, R. M., Lwellen, T. K., Graham, M. M., 1993 A maximum-likelihood approach to PET emission attenuation image registration," IEEE Nuclear Science Symposium and Med. Imag. Conference., pp.1139-1143, Oct-Nov 1993.
- Diaconis, P., Salo-Coste, L., 1998. What do we know about the Metropolis Algorithm?, J.

Comp. and Syst. Science, vol.57, pp.20-36.

Gomez, R., Jazaeri, A., Kafatos, M., 2001. Wavelet-based hyperspectral and multi-spectral image fusion. Proc. of SPIE, vol.4383, pp.36-42.

Gonzalez-Audicana, M., Saleta, J. L., Catalan, R. G., Garcia, R., 2004. Fusion of multi-spectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," IEEE Trans. on Geoscience and Remote Sensing, vol. 42, no. 6, pp. 1291-1299.

Hardie, R. C., Eismann, M. T., Wilson, G. L., 2004. MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor. IEEE Trans. on Image Processing, vol.13, no.9, pp.1174-1184.

Kay, S. M., 1993. Fundamentals of Statistical Signal Processing: Estimation Theory. Englewoods Cliffs, NJ: Prentice-Hall.

Li, Z., Leung, H., 2009. Fusion of multispectral and panchromatic images using a restoration-based method. IEEE Trans. Geosci. Remote Sens., vol. 47, no. 5, may. 2009.

Liu, Z., Tsukada, K., Hanasaki, K., Ho, Y. K., Dai, Y. P., 2001. Image fusion by using steerable pyramid. Pattern Recognition Letters, Vol. 22, no. 9, pp. 929-939, July 2001.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines, J. Chem. Phys, 21, 1087-1092.

Petrovic, V. S., Xydeas, C. S., 2004. Gradient-based multiresolution image fusion. IEEE Trans. Image Processing, vol. 13, no. 2, pp. 228-237, Feb. 2004.

Reddy, B. S., Chatterji, B. N., 1996. An FFT-Based technique for translation, rotation, and scale-invariant image registration. IEEE Trans. Image Processing, vol. 5, no. 8, pp. 1266-1271.

Shettigare, V. K., 1992. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. Photogramm. Eng.

Remote Sens., vol. 58, no. 5, pp. 561–567.

Shi, X., Manduchi, R., 2003. A study on Bayes feature fusion for image classification. Proceedings of the IEEE Workshop on Statistical Algorithms for Computer Vision, Madison, Wisconsin.

Stone, H.S., Tao, B., McGuire, M., 2003. Analysis of image registration noise due to rotationally dependent aliasing," Journal of Visual Communication and Image Representation, vol. 14, pp. 114-135.

Thevenaz, P., Ruttimann, U. E., Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. IEEE Trans. Image Processing, vol. 7, no. 1, pp. 27-41, Jan. 1998.

Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Machine Intell., vol.13, no. 4, pp. 376-380, April 1991.

Van Den Elsen, P. A., Pol, E. D., Sumanaweera, T. S., Her, P. F., Napel, S., Adler, J. R., 1994. Grey value correlation techniques used for automatic matching of CT and MR brain and spine images. Visualization in Biomedical Computing, Proc. SPIE, vol. 2357, pp. 227-237.

Xia, Y., Kamel, M. S., 2007. Novel cooperative neural fusion algorithms for image restoration and image fusion", IEEE Trans. Image Processing, vol. 16, no. 2, pp. 367-381, Feb. 2007

Zheng, S., Shi, W. Z., Liu, J., Zhu, G. X., Tian, J. W., 2007. Multisource image fusion method using support value transform. IEEE Trans. Image Processing, vol. 16, no. 7, pp. 1831-1839, July 2007.

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. Image and Vision Computing, vol. 21, pp. 977-1000.

Zhou, J., Civco, D. L., Silander, J. A., 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. Int. J. Remote Sens., vol. 19, no. 4, pp. 743–757, Mar. 1998.

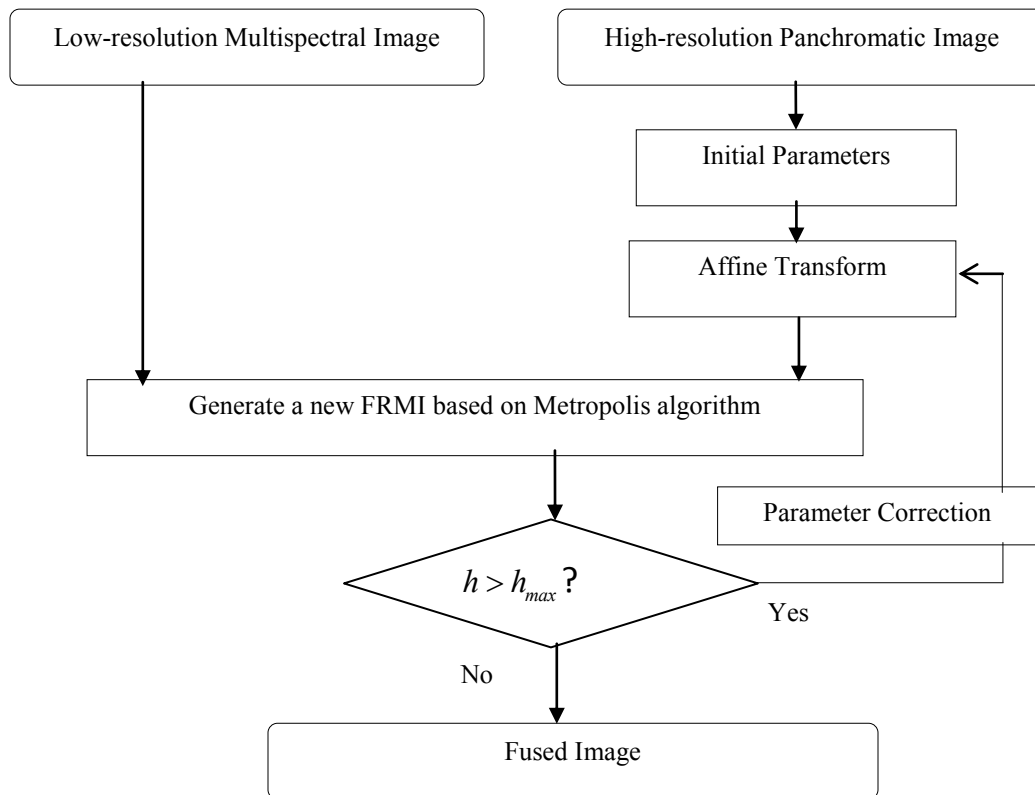


Figure 1. Flowchart of the proposed algorithm

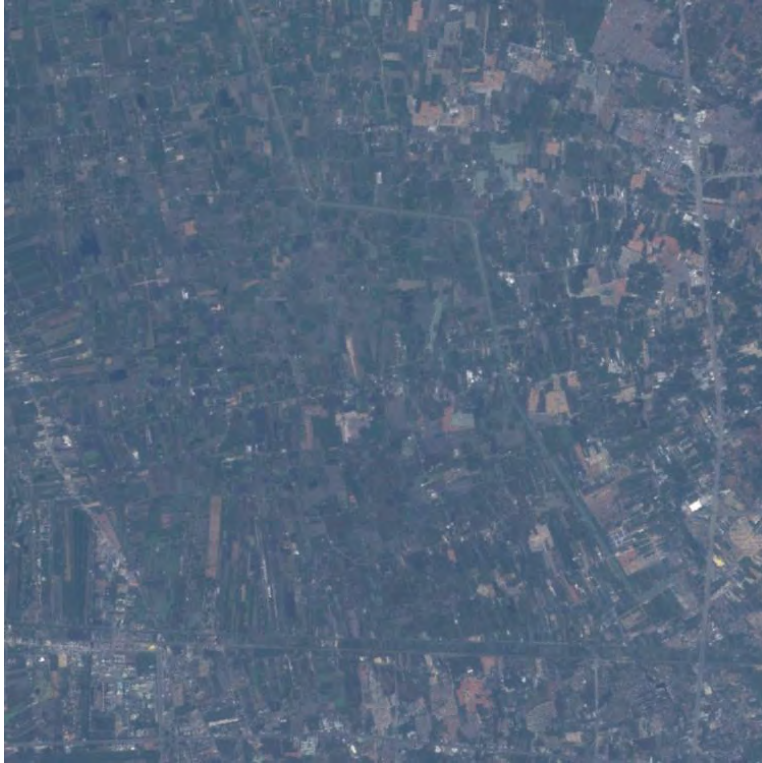


Figure 2. 750x750 THEOS multispectral image of city area

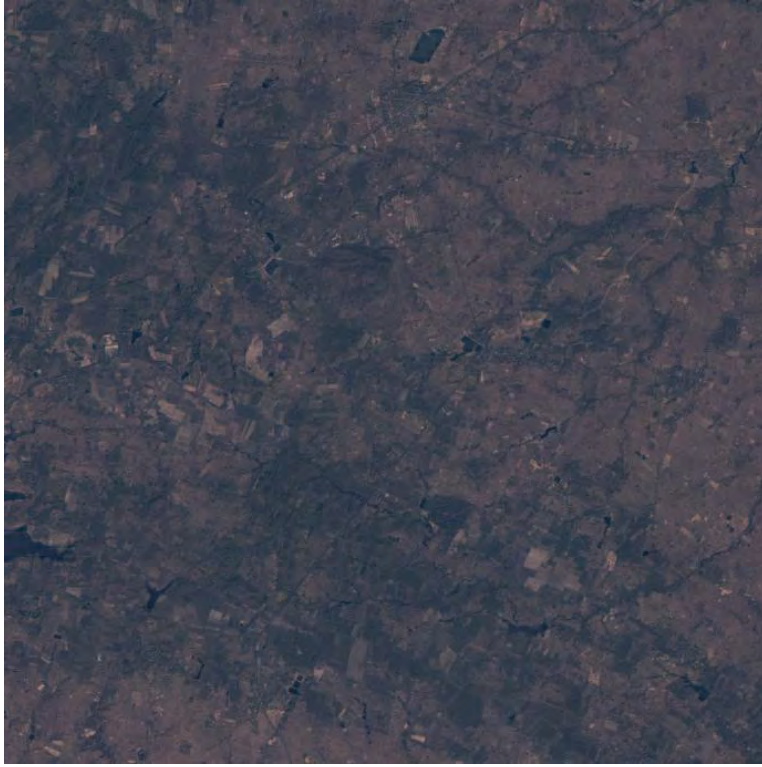


Figure 3. 750x750 THEOS multispectral image of drought area

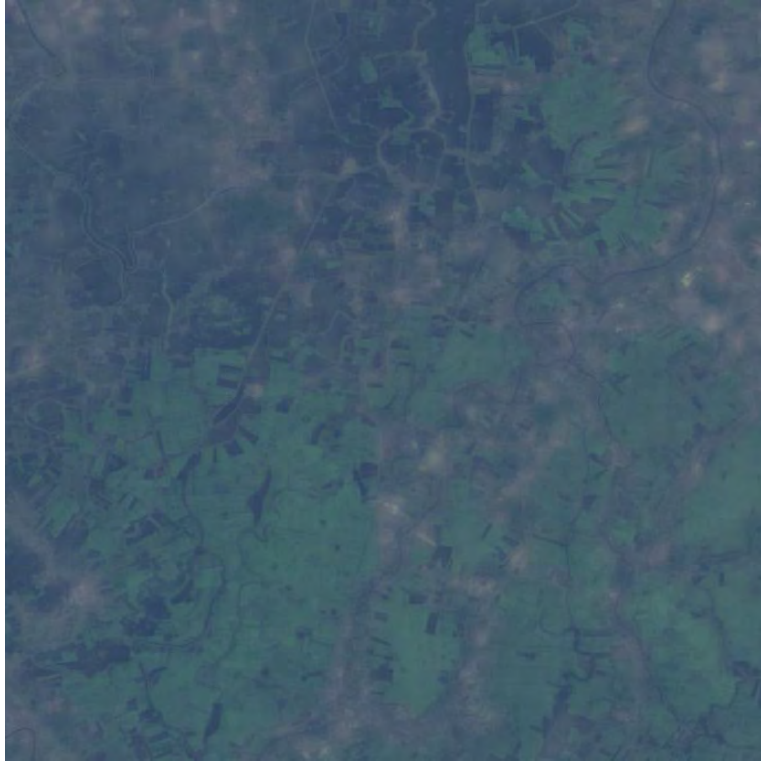


Figure 4. 750x750 THEOS multispectral image of agriculture area



Figure 5. 750x750 THEOS multispectral image of mountain area

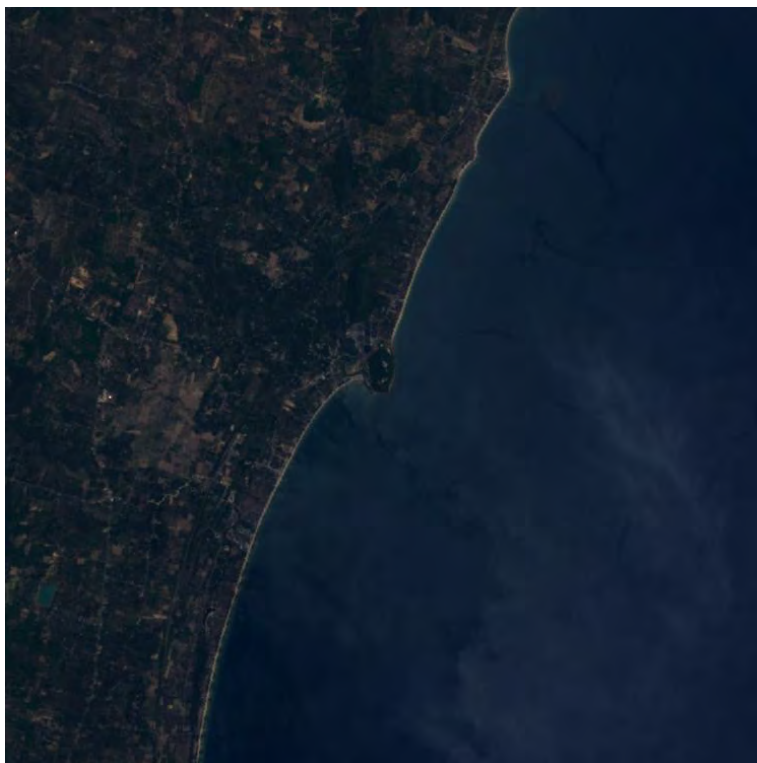


Figure 6. 750x750 THEOS reference multispectral image of seashore area

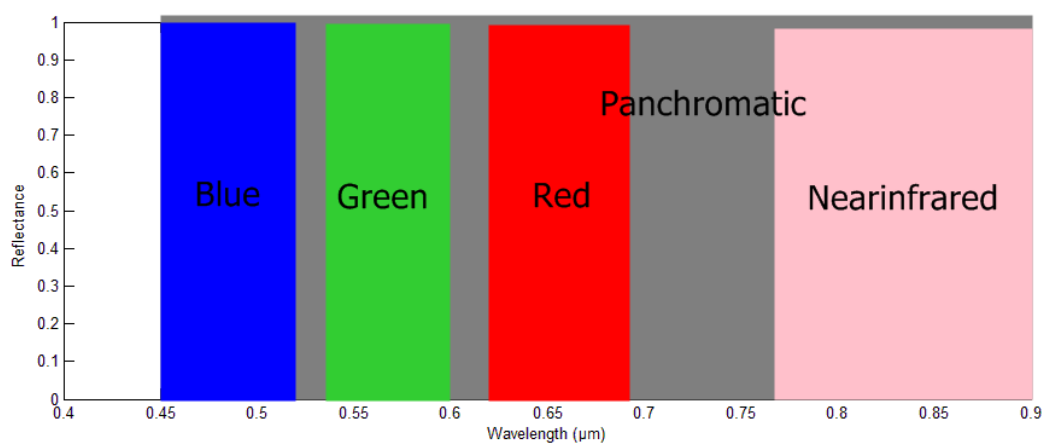


Figure 7. Ideal spectral response of THEOS

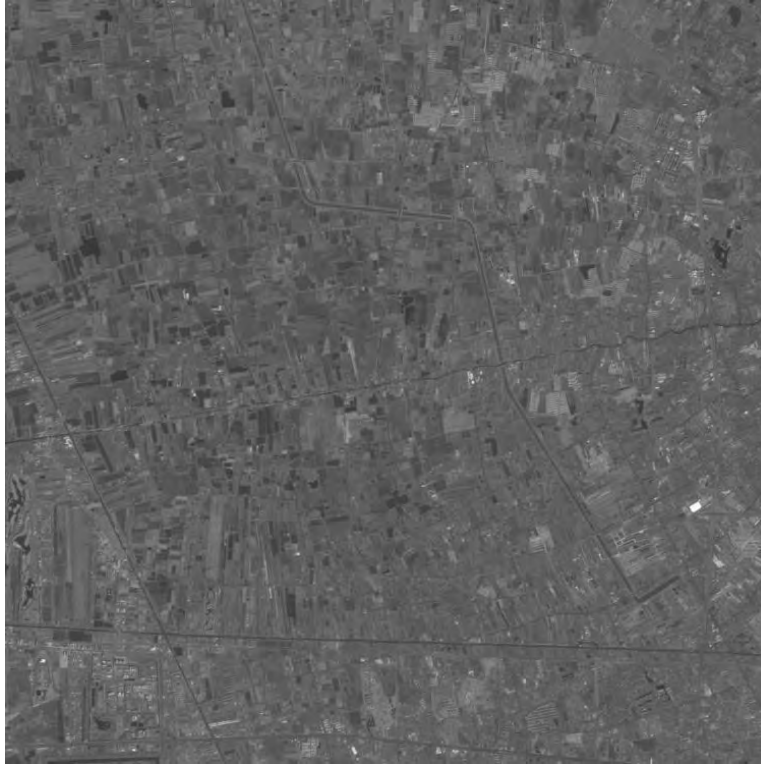


Figure 8: Panchromatic image of the city scene of size 5625×5625 pixels



Figure 9: Degraded multispectral image of city scene at 112.5×112.5 meters resolution of size 100×100 pixels

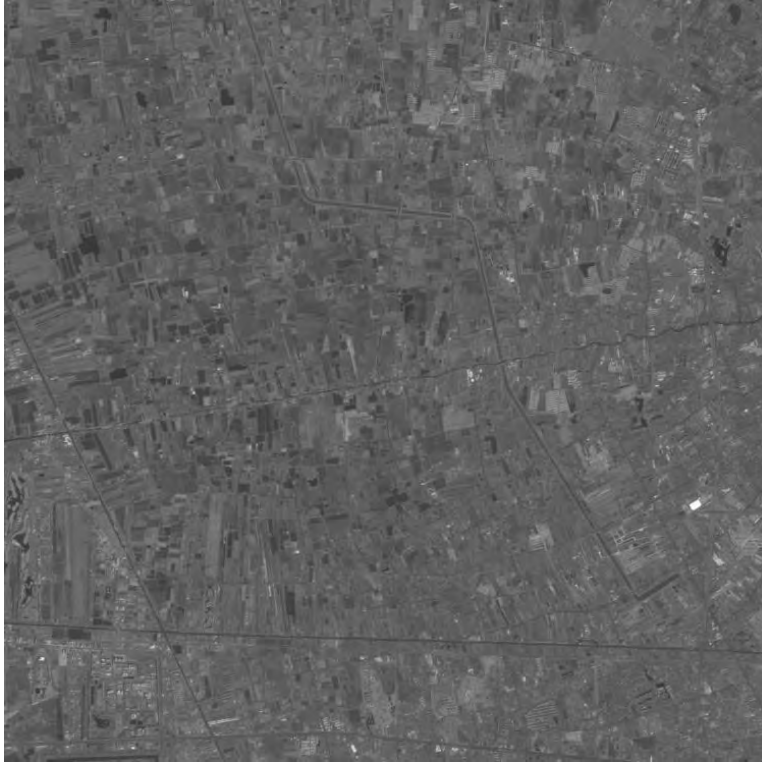


Figure 10: Degraded panchromatic image of the city scene at 15×15 meters resolution of size 750×750 pixels

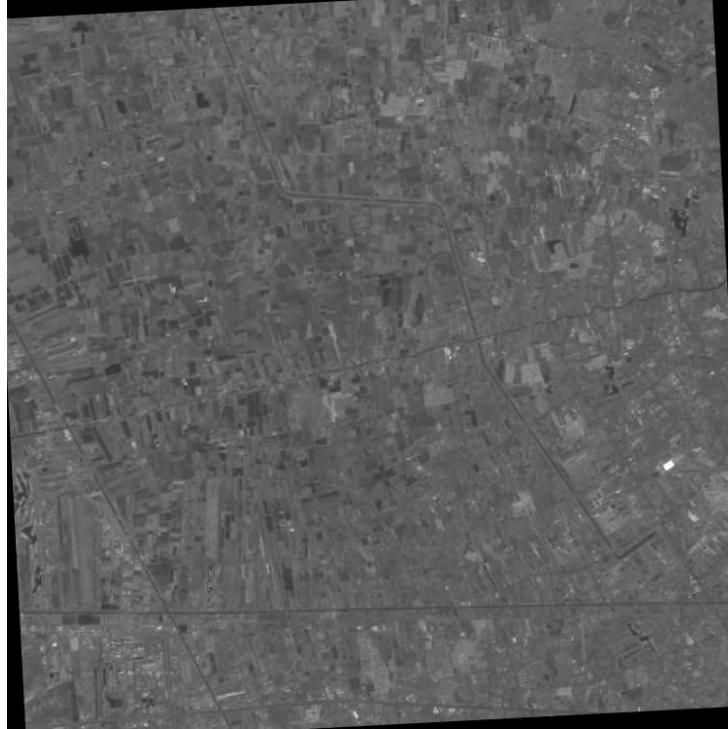


Figure 11: Degraded panchromatic image with mis-registration error of 3 degrees of rotation, 3 pixels of displacement in column and 3 pixels in row



Figure 12. THEOS fused multispectral image with city area and 750x750 pixels



Figure 13. THEOS initial fused multispectral image obtained by applying Hardie algorithm to the unregistered image pair with 3 pixels displacement and 3 degrees rotation with city scene



Figure 14. The resulting fused image for city area

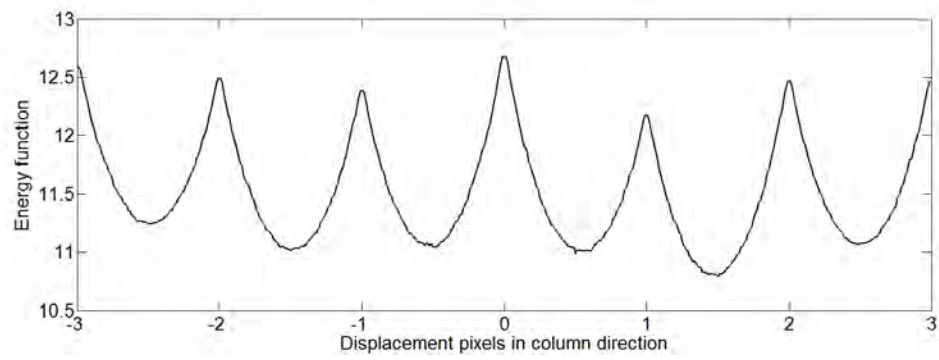


Figure 15. Energy function defined in Eq. 21 of the displacement pixels in row column direction.

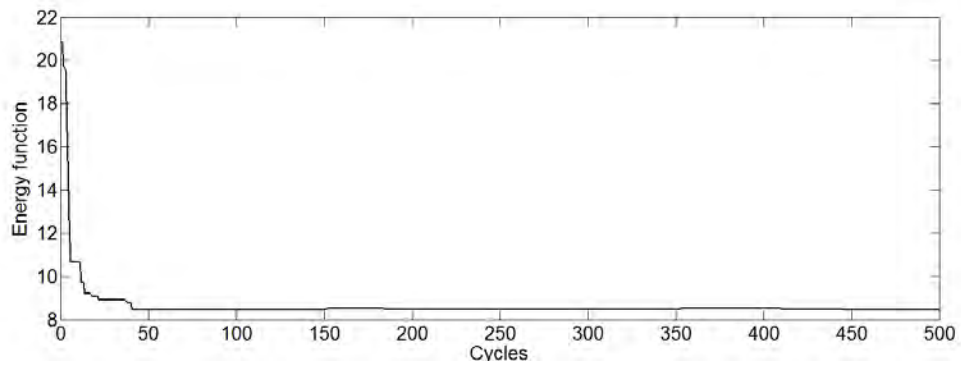


Figure 16. Energy function for each iteration of the Metropolis algorithm

Table 1: Details of THEOS images in our experiment

Area	Longitude	Latitude	Date	Time
City	100.3845251902455N	13.82391143621505E	2009/12/13	10.32 AM
Drought	105.058671289676N	15.92080873025695E	2010/01/04	10.31 AM
Agriculture	100.5728597866535N	14.6219909729372E	2009/12/13	10.32 AM
Mountain	100.561439575278N	19.69589619804945E	2010/02/13	10.31 AM
Seashore	99.59914701590165N	11.3390399265943E	2010/03/06	10.33 AM

Table 2: Orbital characteristics of THEOS

Satellite	THEOS
Orbit	Sun synchronous recurrent frozen orbit
Altitude	822 km
Inclination	97.95 degree
Repetition cycle	31 days
Descending node (local time)	10:30 AM
Orbital period	101.4 minutes
On-board capacity	16 Gbits

Table 3: Technical specifications of THEOS payload

	Panchromatic	Multispectral
Wavelength	450-900 nm	Blue: 450-520 nm
		Green: 530-600 nm
		Red: 620-690 nm
		Near Infrared: 770-900 nm
Resolution	2 m	15 m
Swath width	22 km	90 km
Pixels	12000	6000

Table 4: The mis-registration errors examined in this experiment

	Scale	Ratio angle	Displacement in column direction	Displacement in row direction
Case I	1	1	1	1
Case II	1	-2	-2	-2
Case III	1	3	3	3
Case IV	1	-4	-4	-4
Case V	1	5	5	5

Table 5: Comparison of fusion performance for images of city area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	7.2465	1	1	1	10.2043	0.4315	0.4195	0.1826	6.9368
0	0	0	7.2465	-2	-2	-2	9.7011	0.4426	0.4225	-0.0597	6.3584
0	0	0	7.2465	3	3	3	9.2295	1.5293	0.3358	-0.0037	6.9519
0	0	0	7.2465	-4	-4	-4	9.0814	1.2999	0.6848	-0.1954	7.4333
0	0	0	7.2465	5	5	5	8.8802	1.4716	0.6873	-0.0114	7.1084

Table 6: Comparison of fusion performance for images of drought area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	4.4493	1	1	1	6.7294	0.4178	1.6116	-0.0425	4.4184
0	0	0	4.4493	-2	-2	-2	6.3793	0.4561	0.5670	-0.0832	3.5014
0	0	0	4.4493	3	3	3	5.9539	1.4802	0.6065	0.0960	4.0669
0	0	0	4.4493	-4	-4	-4	5.7613	1.5929	0.4242	-0.1787	4.3226
0	0	0	4.4493	5	5	5	5.5656	0.6040	0.4540	0.0551	3.7374

Table 7: Comparison of fusion performance for images of agriculture area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	8.0216	1	1	1	11.3885	0.5491	1.3608	0.0223	5.6583
0	0	0	8.0216	-2	-2	-2	11.3888	0.4865	0.5077	0.0182	5.7027
0	0	0	8.0216	3	3	3	10.9708	0.3072	0.5173	0.0277	5.8182
0	0	0	8.0216	-4	-4	-4	10.6829	0.5111	0.7775	-0.0304	5.6377
0	0	0	8.0216	5	5	5	10.4727	0.4003	0.6520	0.0465	5.9665

Table 8: Comparison of fusion performance for images of mountain area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	3.6175	1	1	1	4.4446	0.3589	0.6566	-0.2068	3.2774
0	0	0	3.6175	-2	-2	-2	4.5898	0.4614	0.4915	-0.2390	3.3056
0	0	0	3.6175	3	3	3	4.6388	0.6505	-0.5470	-0.0401	3.2220
0	0	0	3.6175	-4	-4	-4	4.5726	0.8501	-0.4960	0.0323	3.1460
0	0	0	3.6175	5	5	5	4.6244	1.5957	0.4378	-0.0534	3.3172

Table 9: Comparison of fusion performance for images of sea shore area

Hardie algorithm with perfect registration				Fusion Result using Hardie algorithm without registration error correction				Proposed joint image fusion and registration algorithm			
d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE	d_x	d_y	θ	RMSE
0	0	0	5.1446	1	1	1	7.4105	0.6091	0.6279	0.0773	3.9202
0	0	0	5.1446	-2	-2	-2	7.2712	0.4977	0.5948	0.0856	3.9134
0	0	0	5.1446	3	3	3	7.0384	0.3827	1.3204	0.0796	4.1890
0	0	0	5.1446	-4	-4	-4	6.7605	0.6050	0.6162	0.0149	3.8124
0	0	0	5.1446	5	5	5	6.6767	0.5031	0.6723	-0.0195	4.0117