UNIVERSITÉ DE STRASBOURG

THÈSE présentée par

Tran-Thanh NGO

soutenue le : **22 septembre 2015**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/Spécialité : Traitement du signal et des images

Shadow/Vegetation and Building Detection from single optical remote sensing image

THÈSE dirigée par :

    M. COLLET Christophe   Professeur des Universités, ICube, Université de Strasbourg

Rapporteurs :

    Mme TUPIN Florence   Professeur, Télécom ParisTech, Institut Mines-Télécom
    M. CHANUSSOT Jocelyn   Professeur des Universités, GISPA Lab, Grenoble INP

Autres membres du jury :

    M. MAZET Vincent (encadrant)   Maître de conférences, ICube, Université de Strasbourg

Remerciements

Tout d'abord, je voudrais dire un énorme merci à Christophe et à Vincent, pour votre encadrement, pour votre disponibilité, pour votre expertise scientifique et votre sympathie. Vous m'avez fait découvrir le monde de la recherche que je ne connaissais pas du tout. Particulièrement, vous étiez toujours disponible pour soutenir, conseiller, faire avancer dans mes travaux et vous avez toujours mis en valeur mon travail.

Je voudrais remercier les rapporteurs, Madame Florence Tupin, Professeur à Télécom Paristech (Laboratoire Traitement et Communication de l'Information), et Monsieur Jocelyn Chanussot, Professeur à Grenoble INP - Ense3 (laboratoire Gipsa-Lab) pour leur lecture de ce manuscrit ainsi que pour leurs points de vue et leurs remarques pertinentes sur cette thèse. Je voudrais remercier Monsieur Bernard Allenbach, ingénieur au Sertit, Université de Strasbourg, d'avoir accepté de participer à mon jury.

Ces trois années de thèse n'auraient pas été particulièrement réussies sans l'ensemble des permanents, des doctorants et des stagiaires que j'ai eu l'occasion de rencontrer à l'équipe MIV et qui ont contribué à créer une ambiance de travail agréable. Je tiens également à remercier les membre de Sertit et plus particulièrement Myldred, Arnaud et Bernard. Merci pour vos conseils et vos amitiés qui ont fait ces trois années des moments inoubliables.

Une énorme merci va à mes amis qui m'ont soutenu.

Je tiens également à remercier ma famille pour le soutien inconditionnel qu'ils m'ont toujours apporté et qui m'a permis de réaliser mes études et cette thèse dans les meilleurs conditions possibles.

Enfin, je remercie ma femme Minh-Ngoc, et ma fille Bao-Chau, pour leur confiance, leur amour, leur soutien et pour tous ces événements riches en émotions que nous avons connus pendant ces trois ans.

Résumé long en Français

Introduction

Ces travaux de thèse ont été réalisés au sein de l'équipe MIV, laboratoire ICube et sont financés par l'institut Carnot et la Région Alsace. Ils s'inscrivent dans le projet Kal-Haïti qui vise à construire une base de données d'images de télédétection et de données complémentaires relatives au séisme de janvier 2010 en Haïti. L'un des objectifs du projet Kal-Haïti est de créer et mettre à jour la base de données concernant l'état des bâtiments. La mise à jour est très coûteuse car elle s'effectue manuellement et commence par l'investigation des images. En conséquence, il y a un besoin réel pour développer une méthode automatique de mise à jour et de détection des changements dans la base de données des bâtiments.

Le processus de mise à jour est décrit ci-après. À partir d'une image prise au temps n, les bâtiments sont détectés puis classifiés parmi les trois états suivants :

- bâtiment construit, qui correspond à un bâtiment apparemment achevé et en bon état;

- bâtiment en construction, qui indique un bâtiment inachevé au temps n;

- bâtiment en ruine, qui fait référence à un bâtiment manifestement inhabitable et/ou gravement endommagé.

En utilisant une nouvelle image au temps $n + 1$, les emprises des bâtiments et leur état sont mis à jour. Pour maintenir la cohérence dans la base de données, un état "inexistant" est ajouté, car si un nouveau bâtiment apparaît au temps $n + 1$, alors il est identifié comme inexistant aux temps précédents. De même, il sera identifié comme inexistant aux temps suivants après une disparition complète (déblaiement). Ainsi, à chaque nouvelle image, les bâtiments sont détectés et leur état est déterminé. Enfin, en comparant l'état des bâtiments aux temps n et $n + 1$, le changement de la base de données est classé comme: nouvelle construction, rasé, pas d'évolution, en reconstruction, détruit, construction achevée, construction abandonnée, etc.
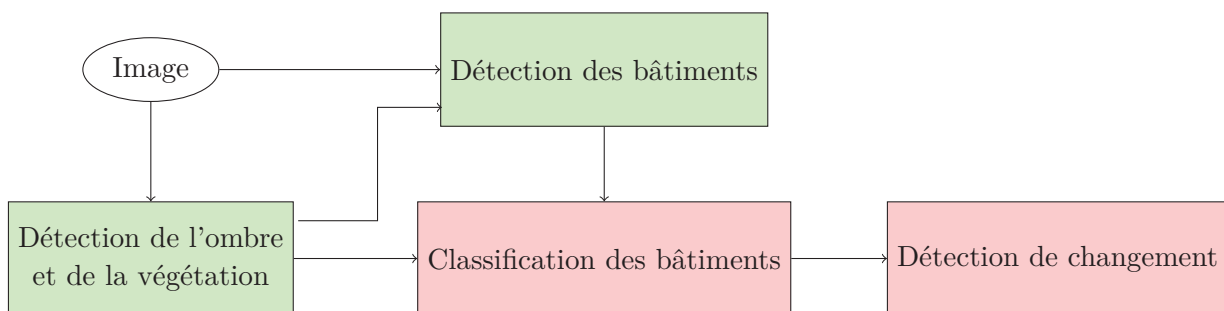

Figure 1: Chaîne de traitement: mise à jour de la base des données des bâtiments. Les blocs en vert sont étudiés dans le cadre de cette thèse.

La chaîne de traitement proposée est représentée figure 1.3. L'ombre portée est l'indice le plus important pour détecter les bâtiments. En outre, l'ombre et la végétation permettent d'obtenir des

indices géométriques et sémantiques sur l'état des bâtiments. C'est pourquoi une première étape de la chaîne de traitement consiste à détecter l'ombre et la végétation. Plusieurs indices de détection de l'ombre et de la végétation proposés dans la littérature ont été comparés, puis ces informations sont fusionnées grâce à la théorie de Dempster-Shafer (DS) [Dempster 1968, Shafer 1976] pour prendre en compte les incertitudes intrinsèques sur les données. Par ailleurs, une modélisation par champ de Markov permet de régulariser la solution. La deuxième étape de la chaîne de traitement consiste en la détection des bâtiments. L'image est tout d'abord segmentée en super-pixels homogènes ce qui redéfinit la grille d'échantillonnage d'origine sur un nouveau graphe. Les bâtiments détectés sont les super-pixels vérifiant certaines règles, comme par exemple la présence d'une ombre portée (détectée dans la première partie de la chaîne de traitement), la rectangularité, etc.

Détection de l'ombre et la végétation

Il existe une littérature abondante concernant la détection de l'ombre et de la végétation (par exemple [Adeline 2013, Tian 2012, Shorter 2009]). Cependant, la détection séparée de l'ombre et de la végétation ne nous semble toutefois pas optimale. En effet, un pixel de végétation ombragée sera classé comme étant de la végétation par l'algorithme de détection de végétation et comme une ombre par l'algorithme de détection de l'ombre. Dès lors, l'utilisation des approches classiques de la littérature n'est pas conseillée pour la détection simultanée de l'ombre et la végétation. Une inspection visuelle aboutira au même problème car l'information du pixel est imprécise et incertaine. Nous proposons une méthode originale, appelé SSVD ("simultaneous shadow vegetation detection"), pour détecter simultanément les zones d'ombres et de végétation. En d'autres termes, nous proposons une méthode de segmentation des images en trois classes : "ombre", "végétation" et "autre". La méthode proposée est résumée dans la figure 4.1.

En comparant les différents indices d'ombre de la littérature (notamment [Adeline 2013]), nous obtenons les meilleurs résultats avec l'indice c_3 de l'espace de couleur $c_1c_2c_3$ [Salvador 2004] :

$$c_3 = \arctan\left(\frac{B}{\max(R, V)}\right) \quad (1)$$

où R, V et B représentent respectivement les composantes rouge, verte et bleue de l'image. De même, il existe plusieurs indices pour détecter la végétation [Shorter 2009, Woebbecke 1995a]. Nous avons constaté que l'indice ExG est le plus performant [Woebbecke 1995a] :

$$ExG = \frac{2 \times V - R - B}{R + V + B}. \quad (2)$$

Par ailleurs, il est apparu que le résultat de la détection était amélioré en prenant en compte la luminance L :

$$L = \frac{R + V + B}{3} \quad (3)$$

Chacun des trois indices ci-dessous permet de construire l'une des images caractéristiques $Y^{(1)}$, $Y^{(2)}$, $Y^{(3)}$. La suite de la méthode consiste à segmenter l'observation en trois classes ω_1, ω_2 et ω_3 (correspondant à l'ombre, la végétation et le reste) à partir des images caractéristiques. Tout d'abord, un seuillage automatique d'Otsu [Otsu 1975] est appliqué aux images $Y^{(1)}$, $Y^{(2)}$ et $Y^{(3)}$ afin d'obtenir une segmentation initiale des zones d'ombre, des zones de végétation et des
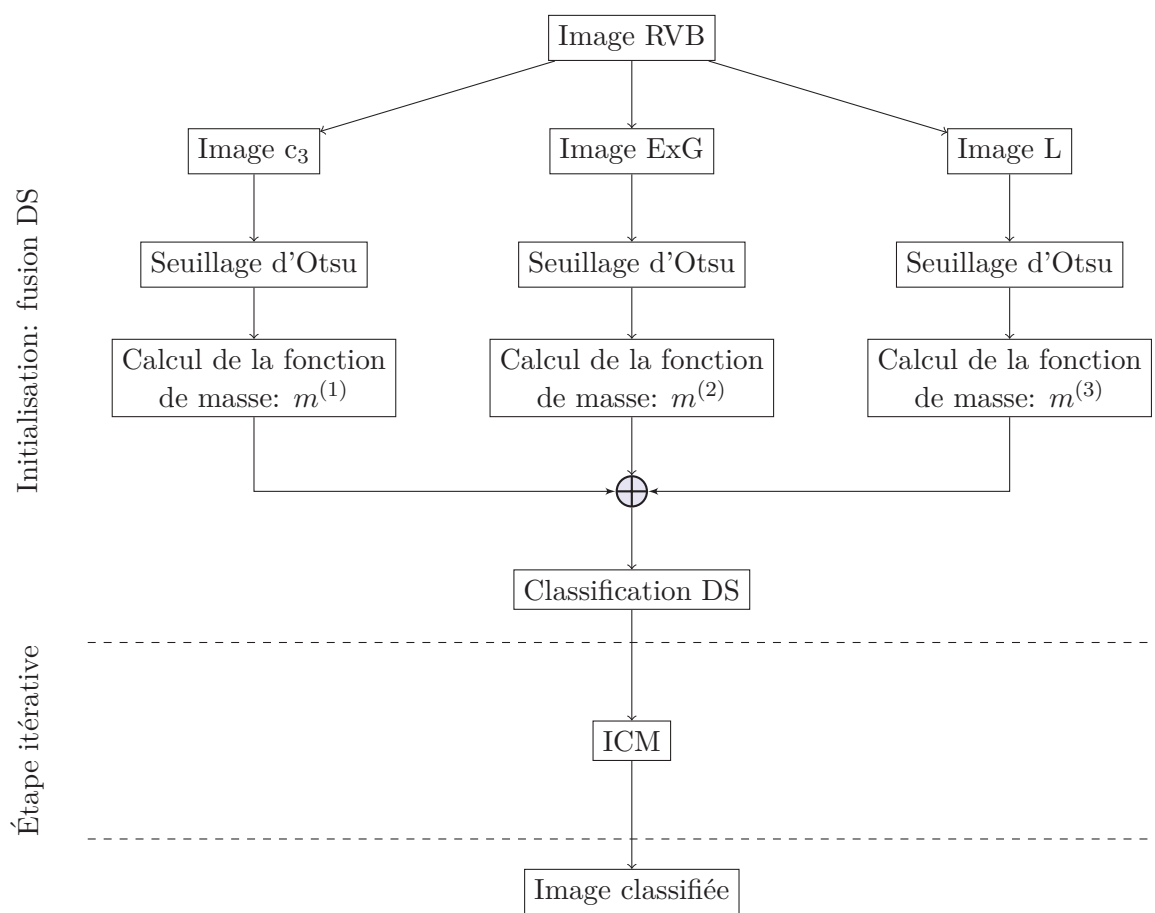
Image RVB

Image c$_3$ | Image ExG | Image L

Seuillage d'Otsu | Seuillage d'Otsu | Seuillage d'Otsu

Calcul de la fonction de masse: $m^{(1)}$ | Calcul de la fonction de masse: $m^{(2)}$ | Calcul de la fonction de masse: $m^{(3)}$

$\oplus$

Classification DS

Initialisation: fusion DS

Étape itérative

ICM

Image classifiée

Figure 2: Schéma pour détecter simultanément les zones d'ombre et de végétation.

zones sombres (qui inclut les zones d'ombre et de végétation). La méthode d'Otsu est d'ailleurs très utilisée pour détecter l'ombre [Tsai 2006] pour la qualité de ses résultats. Ainsi, l'image $Y^{(1)}$ est segmentée en deux classes ω_1 et $\{\omega_2, \omega_3\}$. De même, $Y^{(2)}$ est segmentée en les classes ω_2 et $\{\omega_1, \omega_3\}$ et $Y^{(3)}$ en les classes ω_3 et $\{\omega_1, \omega_2\}$. La théorie DS appliquée à la segmentation d'image permet alors de fusionner un par un les pixels provenant des trois segmentations et d'inférer sur les hypothèses (simples) H_i représentant les groupes simples : $H_i = \{\omega_i\}$. Le cadre de discernement Θ regroupe l'ensemble des classes : $\Theta = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$.

Ensuite, les fonctions de masse $m_s^{(1)}, m_s^{(2)}, m_s^{(3)}$ de la théorie DS sont estimées en faisant l'hypothèse de distributions gaussiennes. Leur combinaison est réalisée en utilisant la règle conjonctive prudente (notée $\varoslash$) de Denœux [Denœux 2006, Denœux 2008], qui permet de combiner des données dépendantes. La fusion des données est illustrée figure 3.

$$m_s^{(1)}(\{\omega_1\}), m_s^{(1)}(\{\{\omega_2\}, \{\omega_3\}\}), m_s^{(1)}(\Theta)$$

$$m_s^{(2)}(\{\omega_2\}), m_s^{(2)}(\{\{\omega_1\}, \{\omega_3\}\}), m_s^{(2)}(\Theta) \oplus \to \begin{array}{l} m_s(A) = m_s^{(1)} \varoslash m_s^{(2)} \varoslash m_s^{(3)} \\ s \in \mathcal{S}, A \subset \Theta \end{array}$$

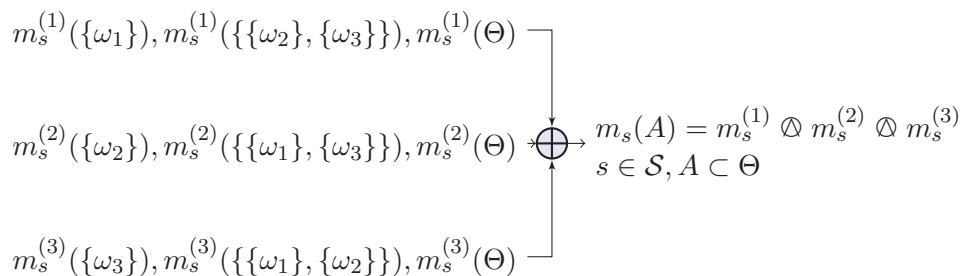$$m_s^{(3)}(\{\omega_3\}), m_s^{(3)}(\{\{\omega_1\}, \{\omega_2\}\}), m_s^{(3)}(\Theta)$$

Figure 3: Diagramme de fusion ($\mathcal{S}$ représente l'ensemble des pixels de l'image, A représente une hypothèse simple ou une union d'hypothèses simples).

Une fois que toutes les fonctions de masse des hypothèses simples et composées d'un pixel s sont déterminées, nous avons besoin d'un critère de décision pour décider quelle hypothèse est la plus réaliste. Pour la règle de combinaison prudente de Denoeux, la décision s'effectue au niveau dit pignistique qui impose une transformation des fonctions de masses en distributions de probabilité:

$$Betp(H_i) = \sum_{\{A : H_i \in A\}} \frac{m(A)}{|A|} \tag{4}$$

où $|A|$ est le cardinal de l'ensemble des éléments de A. La décision consiste à choisir l'hypothèse simple de plus grande probabilité pignistique. Un exemple de la classification est représenté figure 4.5.e.

Or, la fusion des informations se place au niveau du pixel : cela suppose implicitement que chaque pixel ne dépend pas de ses voisins, ce qui implique que la procédure est très sensible au bruit. Pour cette raison, une régularisation par champ de Markov [Geman 1984] de la classification est introduite dans la fusion de DS pour gérer au mieux la forte corrélation spatiale des pixels. L'utilisation conjointe de ces deux techniques a fait l'objet de plusieurs travaux en classification à partir d'observations multisource [Bentabet 2008]. Alors que le cadre probabiliste attribue une vraisemblance aux classes ω_i, la théorie DS remplace cette vraisemblance par une fonction de masse sur l'hypothèse simple $H_{x_s} = \{\omega_i\}$. Le terme de régularisation du champ de Markov peut alors être généralisé pour traiter l'hypothèse composée. Considérant la vraisemblance $m_s(A_p)$ et

la probabilité a priori $m_s(A_q|\hat{x}_{\mathcal{V}_s})$ comme deux sources d'information évidentielles, la probabilité a posteriori est remplacée par la somme orthogonale de DS de ces deux sources:

$$M_s(A|\hat{\mathbf{x}}_{\mathcal{V}_s}) = \frac{1}{1-\mathcal{K}} \sum_{A_p \cap A_q = A} m_s(A_p)m_s(A_q|\hat{x}_{\mathcal{V}_s}) \tag{5}$$

où $\mathcal{K} = \sum_{A_p \cap A_q = \emptyset} m_s(A_p)m_s(A_q|\hat{x}_{\mathcal{V}_s})$.

Une fois les fonctions de masse des hypothèses simples et composées déterminées, la classe de chaque pixel est estimée en maximisant la fonction de plausibilité sur l'information jointe:

$$\hat{x}_s = \arg\max_{x_s} \text{Pls}(H_{x_s}) = \arg\max_{x_s} \left(\sum_{A \cap H_{x_s} \neq \emptyset} M_s(A|\hat{\mathbf{x}}_{\mathcal{V}_s}) \right) \tag{6}$$

Une illustration de ces étapes est donnée figure 4.5. Par simple comparaison visuelle, on constate que les zones d'ombre et les zones de végétation sont obtenues avec une grande précision grâce à la régularisation spatiale : les petites zones isolées d'ombre et de végétation sont supprimées et la forme de l'ombre est préservée. De plus, nous évaluons la méthode proposée à la fois qualitativement mais aussi quantitativement. Pour l'évaluation quantitative, nous adoptons les mêmes métrique et table de précision que ceux proposés dans [Prati 2003] (table 3.4), en calculant le taux de vrais positifs (VP), faux négatifs (FN), faux positifs (FP) et vrais négatifs (VN). La somme VP + FN + FP + VN correspond donc au nombre total de pixels dans l'image. La précision du photointerpréteur indique comment les pixels des catégories connues sont correctement classés. La précision de l'utilisateur indique la probabilité de pixels correctement classés par rapport à la référence qui est ici le photointerpréteur. En combinant ces indices de précision, on obtient la précision générale τ qui permet d'évaluer la performance de l'algorithme. Des valeurs de paramètres élevées correspondent à de bons résultats. De plus, comme le fait [Rufenacht 2014], on calcule le coefficient de corrélation de Matthews (MCC) [Matthews 1975], qui est une mesure plus équilibrée si les deux classes ont des tailles différentes. La valeur de la MCC est entre $[-1, 1]$, où des valeurs élevées indiquent une meilleure prédiction.

Précision du photointerpréteur		Précision de l'utilisateur	
Ombre (rappel)	Non-ombre	Ombre (précision)	Non-ombre
$P_s = \dfrac{\text{VP}}{\text{VP} + \text{FN}}$	$P_n = \dfrac{\text{VN}}{\text{VN} + \text{FP}}$	$U_s = \dfrac{\text{VP}}{\text{VP} + \text{FP}}$	$U_n = \dfrac{\text{VN}}{\text{VN} + \text{FN}}$
Précision totale		Coefficient de corrélation de Matthews	
$\tau = \dfrac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}}$		$\text{MCC} = \dfrac{\text{VP} \times \text{VN} - \text{FP} \times \text{FN}}{\sqrt{(\text{VP} + \text{FP})(\text{VP} + \text{FN})(\text{VN} + \text{FP})(\text{VN} + \text{FN})}}$	

Table 1: Métrique et table de précision.

Les résultats donnés dans le table 2 montrent clairement que notre approche possède de très bonnes performances globales tant pour la détection d'ombre que pour la détection de végétation. Les précisions générales τ sont élevées (95.66% pour la détection de l'ombre et 98.38% pour la détection de végétation). De plus, une comparaison quantitative entre notre méthode et les méthodes de [Tian 2012] (pour la détection de l'ombre) et [Shorter 2009] (pour la détection de
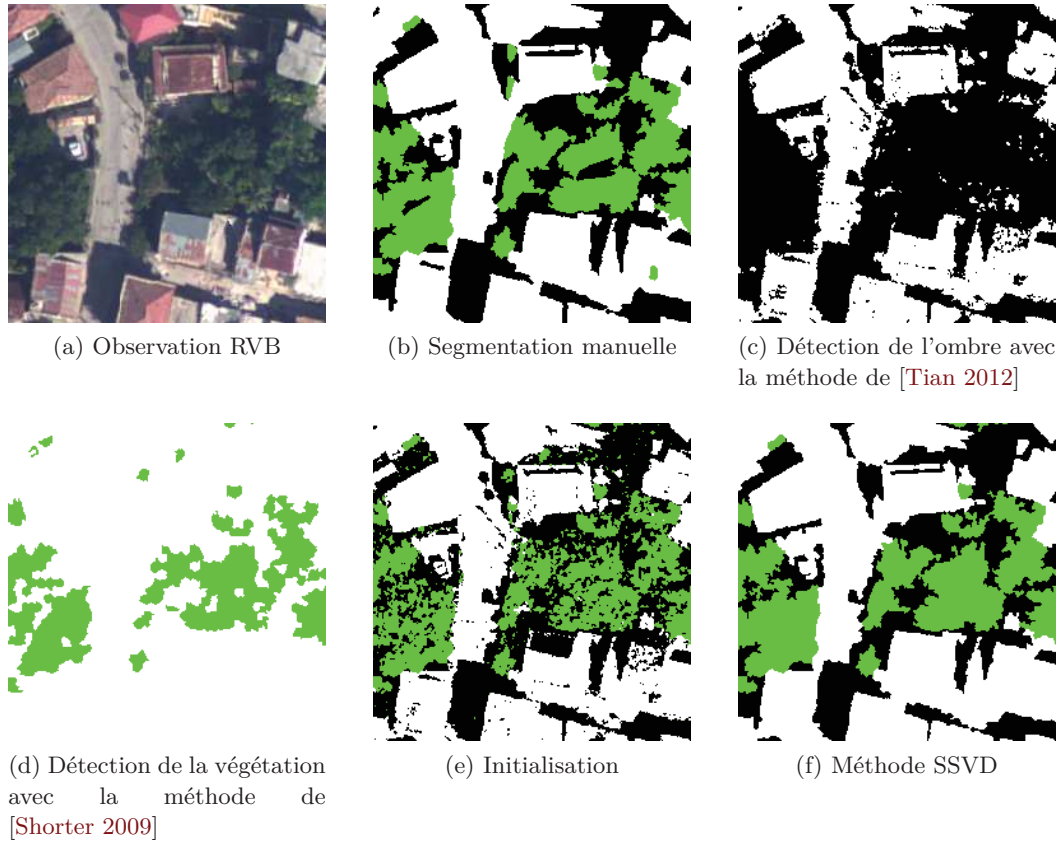
(a) Observation RVB

(b) Segmentation manuelle

(c) Détection de l'ombre avec la méthode de [Tian 2012]

(d) Détection de la végétation avec la méthode de [Shorter 2009]

(e) Initialisation

(f) Méthode SSVD

Figure 4: Segmentation des ombres (en noir), de la végétation (en vert) et du reste (en blanc)

Méthode	Détection des ombres					
	Précision du photointerpréeur		Précision de l'utilisateur		Précision totale	Coefficient de corrélation de Matthews
	Ombre	Non-ombre	Ombre	Non-ombre	τ	MCC
Méthode de [Tian 2012]	95.70	49.99	49.32	95.80	65.40	0.52
Initialisation	88.44	88.32	77.85	94.27	88.36	0.74
Méthode SSVD	89.13	98.98	97.81	94.71	**95.66**	**0.90**
Méthode	Détection de la végétation					
	Précision du photointerpréeur		Précision de l'utilisateur		Précision totale	Coefficient de corrélation de Matthews
	Végétation	Non-végétation	Végétation	Non-végétation	τ	MCC
Méthode de [Shorter 2009]	85.03	80.77	36.28	97.66	81.25	0.78
Initialisation	81.20	97.26	89.41	94.79	93.71	0.81
Méthode SSVD	93.25	99.04	92.60	99.13	**98.38**	**0.93**

Table 2: Qualité de la segmentation pour l'image représentée dans la figure 4.5.

végétation) montre que la méthode SSVD a la meilleure précision globale.

En conclusion, notre travail apporte trois contributions. Premièrement, il introduit un nouveau schéma pour détecter simultanément les zones d'ombre et de végétation. Deuxièmement, l'utilisation de la théorie DS permet de combiner différents indices d'ombre et de végétation afin d'obtenir une segmentation fiable et plus précise. Troisièmement, la modélisation par champ de Markov permet d'exploiter les informations de voisinage géométrique de l'image ; cette idée est largement utilisée en segmentation d'image, mais rarement dans le contexte de la détection de l'ombre et de la végétation.

Détection des bâtiments

Beaucoup de méthodes ont été proposées pour détecter les bâtiments à partir d'une image de télédétection [Ozgun Ok 2013, Femiani 2014, Li 2015]. Dans notre cas, la détection doit être utilisable sur des zones de densités très hétérogènes (de l'urbain au rural). Cela impose les hypothèses suivantes sur l'apparence des bâtiments :

1. les bâtiments ont une couleur homogène (en pratique, c'est cette hypothèse qui permet de séparer les bâtiments dans les zones très denses) ;

2. les bâtiments créent une ombre portée (d'où l'importance de l'heure de prise de vue et de la qualité de la détection de l'ombre) ;

3. les bâtiments ont une géométrie simple à angles droits (en forme de rectangle, de L, de U ou de I).

Ces hypothèses ont conduit à une méthode originale, appelé SBBD ("shadow-based building detection"), décrite ci-après. Le schéma de la méthode est représenté figure 5.3.

Tout d'abord, la carte des zones d'ombres obtenue dans la première partie de la thèse est nettoyée afin de supprimer les ombres générées par autre chose que les bâtiments : les ombres dues à la végétation sont supprimées à l'aide de la carte des zones de végétation, et les ombres très petites (dues à des objets de petite taille telle que les véhicules ou rochers) sont également éliminées. Ensuite, l'image est sur-segmentée en régions homogènes à l'aide de la méthode SLIC [Achanta 2012] permettant de remplacer avantageusement la structure rigide de la grille de pixels par un graphe de régions. Puis, une méthode de classification–fusion itérative est appliquée sur l'ensemble des régions. L'étape de classification affecte une classe à chaque région et les regroupe ensuite en groupes. Selon la position des ombres, une fusion permet de regrouper certaines régions d'un même groupe et de produire des régions de forme à peu près rectangulaire. Un exemple de la première itération de la segmentation par croissance de région est représenté figure 6.

Ces deux étapes de classification et de fusion sont répétées jusqu'à ce qu'aucune fusion ne soit possible. Enfin, les bâtiments sont déduits en approximant les régions fusionnées par un rectangle (un exemple est représenté figure 5.20).

La performance finale de la méthode SBBD est évaluée en comparant les résultats de la méthode avec la vérité terrain. Nous effectuons une évaluation quantitative au niveau pixellique
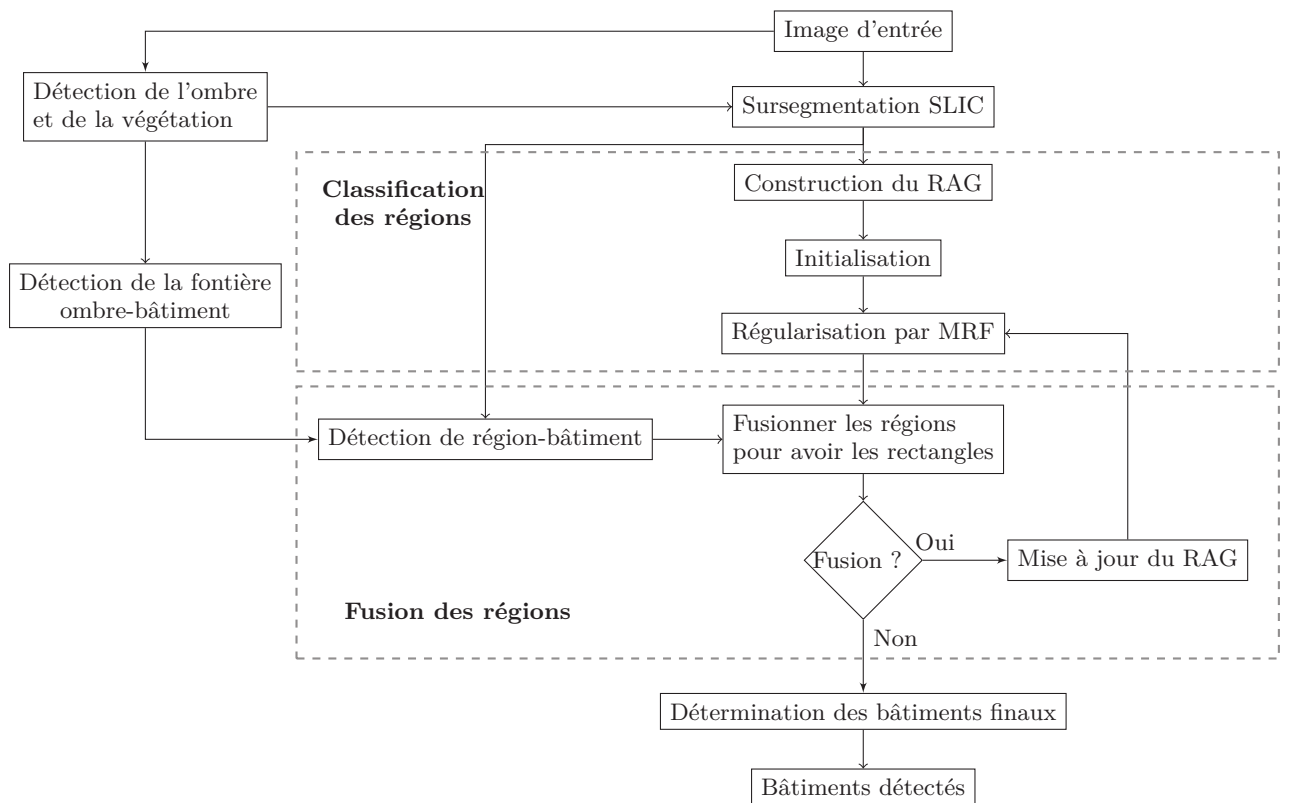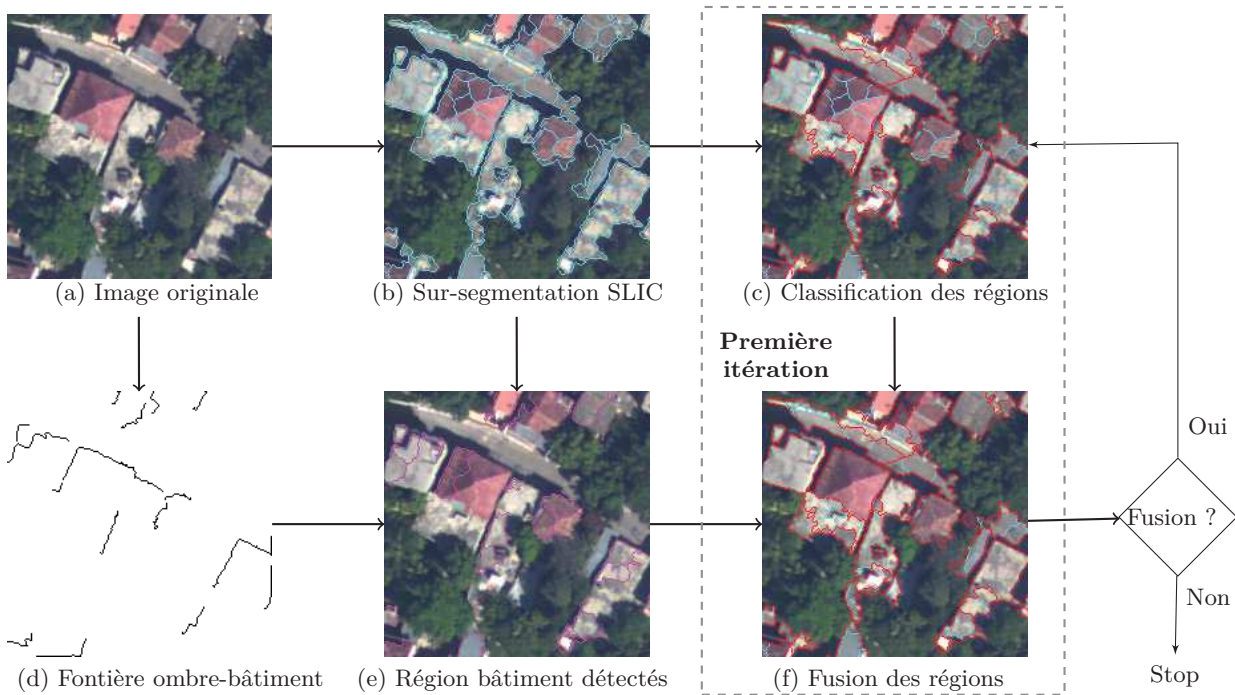
Image d'entrée

Détection de l'ombre et de la végétation → Sursegmentation SLIC

Classification des régions

Construction du RAG

Initialisation

Détection de la fontière ombre-bâtiment

Régularisation par MRF

Détection de région-bâtiment → Fusionner les régions pour avoir les rectangles

Fusion ? —Oui→ Mise à jour du RAG

Fusion des régions

Non

Détermination des bâtiments finaux

Bâtiments détectés

Figure 5: Schéma de la méthode SBBD.

(a) Image originale   (b) Sur-segmentation SLIC   (c) Classification des régions

Première itération

(d) Fontière ombre-bâtiment   (e) Région bâtiment détectés   (f) Fusion des régions

Oui

Fusion ?

Non

Stop

Figure 6: Un exemple de la première itération de la segmentation par croissance de région: (a) image originale, (b) sur-segmentation SLIC (les régions sont séparées par les lignes cyan), (c) classification de régions (les groupes sont séparés par les lignes rouges), (d) Frontière ombre-bâtiment, (e) région bâtiment détectés (les frontières sont délimitées par des lignes de violettes) , (f) Fusion de régions (quelques lignes cyan ont disparu).

(a) Observation RVB   (b) Segmentation d'image   (c) Bâtiments détectés
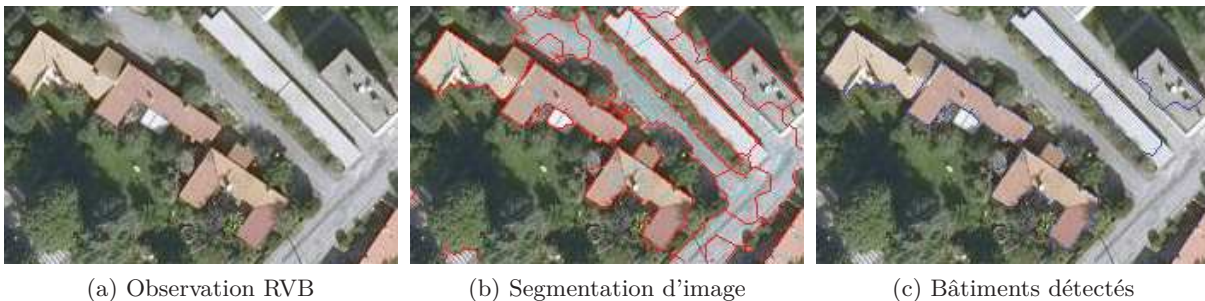
Figure 7: Détection de bâtiment à partir d'une image RVB.

et au niveau d'objet. Nous adoptons les mêmes métrique et table de précision que ceux proposés dans [Ozgun Ok 2013]. La précision P, le rappel R et la F-mesure F_1 sont utilisés pour la précision au niveau de pixel :

$$P = \frac{VP}{VP + FP}$$
$$R = \frac{VP}{VP + FN} \tag{7}$$
$$F_1 = \frac{2 \times P \times R}{P + R}$$

VP est le nombre de pixels de bâtiments correctement identifiés, FN est le nombre de pixels de bâtiments considérés comme n'étant pas des bâtiments, FP est le nombre de pixels hors des bâtiments considérés comme étant des bâtiments et VN est le nombre de pixels hors des bâtiments correctement identifiés. D'autre part, la performance peut être mesurée au niveau des objets eux-mêmes (et non des pixels), en utilisant également les mesures définies dans l'équation 6.1 : un bâtiment est un VP s'il a au minimum 60 % des pixels en commun avec le bâtiment de la vérité terrain ; un bâtiment est un FP s'il ne coïncide pas avec l'un des bâtiments de la vérité terrain ; un bâtiment est FN s'il correspond au bâtiment avec une quantité limitée de recouvrement (< 60%).

(a) Observation RVB   (b) Méthode de [Femiani 2014]   (c) Méthode SBBD
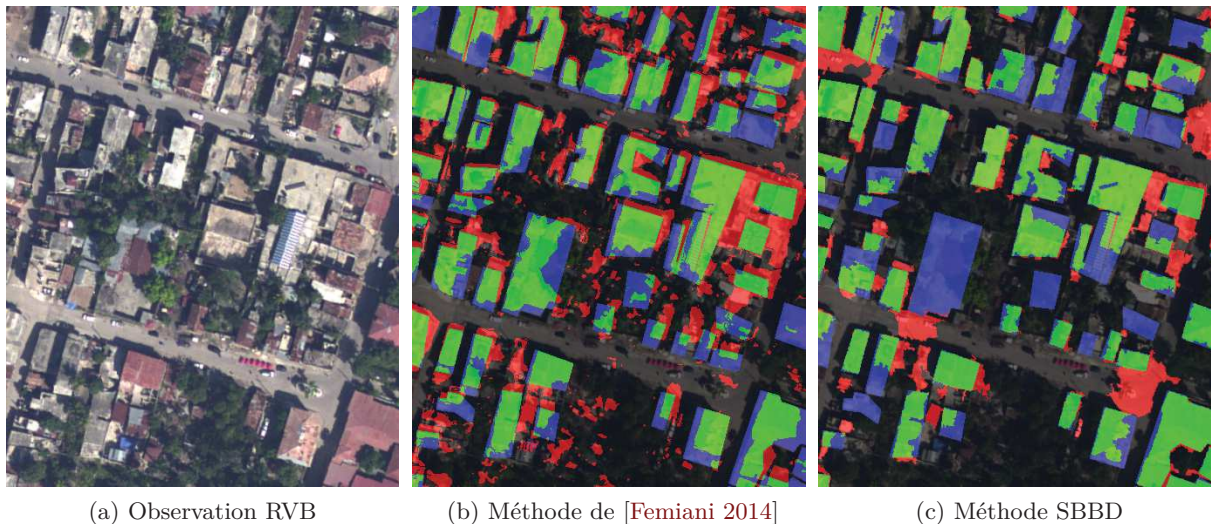
Figure 8 : Résultat de détection des bâtiments de la méthode de [Femiani 2014] et de la méthode SBBD dans la zone péri-urbaine (les VP, FP et FN apparaissent respectivement en vert, rouge et bleu).

Les résultats présentés tableaux 3, 4, 5 montrent clairement que la méthode SBBD a de très bonnes performances globales pour la détection de bâtiments, tant au niveau de pixel qu'au niveau d'objet. De plus, une comparaison quantitative avec trois méthodes de l'état de l'art ([Ozgun Ok 2013, Femiani 2014, Li 2015], qui utilisent également l'ombre), indique que la méthode SBBD donne de meilleurs résultats. De plus, la méthode SBBD est capable de distinguer les bâtiments même s'ils sont côte-à-côte, comme en milieu urbain où la densité des

bâtiments est importante.

Performance au niveau pixel(%)					
Méthode de [Femiani 2014]			Méthode SBBD		
P	R	F_1	P	R	F_1
60.7	62.6	61.6	74.0	61.9	67.4
Performance au niveau objet (%)					
Méthode de [Femiani 2014]			Méthode SBBD		
6.2	12.8	8.3	63.8	56.6	60.0

Table 3: Qualité de la détection pour l'image de la figure 8.

Performance au niveau pixel(%)					
Méthode de [Ozgun Ok 2013]			Méthode SBBD		
P	R	F_1	P	R	F_1
67.3	80.0	73.1	76.5	62.5	68.8
Performance au niveau objet (%)					
Méthode de [Ozgun Ok 2013]			Méthode SBBD		
78.4	69.2	73.5	93.5	76.3	84.1

Table 4: Qualité de la détection pour l'image de la figure 6.6 de la thèse.

Performance au niveau pixel(%)					
Méthode de [Li 2015]			Méthode SBBD		
P	R	F_1	P	R	F_1
81.6	66.4	73.2	78.7	69.9	74.0
Performance au niveau objet (%)					
Méthode de [Li 2015]			Méthode SBBD		
89.9	79.7	84.5	97.5	86.81	91.9

Table 5: Qualité de la détection pour l'image de la figure 6.9 de la thèse.

La méthode SBBD est paramétrée par plusieurs variables. Il n'existe pas de valeurs absolues de ces variables car celles-ci dépendent de la résolution de l'image, de la taille des bâtiments, de la densité d'ombre, etc. Ainsi, les paramètres ayant une signification physique peuvent être facilement fixés, alors que les autres paramètres doivent être déterminés par essai–erreur. Ces expérimentations sont détaillées dans le manuscrit.

En conclusion, notre travail apporte trois contributions. Premièrement, la méthode proposée a besoin d'une unique image (en haute résolution soit environ 50 cm par pixel). Les informations comme le proche-infrarouge, le lidar ou les données d'altitude ne sont pas nécessaires. Deuxièmement, la méthode est non supervisée, c'est-à-dire qu'elle ne nécessite pas de données pré-classifiées. Troisièmement, la méthode de segmentation de croissance de région prend en compte l'information radiométrique et géométrique du bâtiment.

Conclusions

Cette thèse est dédiée à la détection de l'ombre, de la végétation et des bâtiments à partir d'une unique image optique très haute résolution. La première partie présente la méthode SSVD pour détecter simultanément les ombres et la végétation : plusieurs indices d'ombre et de végétation sont comparés puis fusionnés grâce à la théorie de l'évidence de Dempster-Shafer afin d'obtenir une segmentation en trois classes : "ombre", "végétation" et "autre". Comme la fusion est une méthode pixellique, une modélisation par champ de Markov est utilisée pour mieux gérer la forte corrélation spatiale générique dans les images. Dans la deuxième partie, la méthode SBBD permet de détecter des bâtiments à partir d'une image optique monoculaire. L'ombre portée est un indice important pour détecter le bâtiment. Aussi, à partir de la carte de ombre/végétation obtenue dans la partie précédente, la frontière entre l'ombre et le bâtiment correspondant est détectée. Une nouvelle technique de segmentation d'images par croissance de région est ensuite proposée. Les bâtiments sont déterminés à partir du résultat de segmentation en utilisant une approximation par le rectangle à limite minimum. Les comparaisons quantitatives entre ces deux méthodes et les méthodes de l'état de l'art sur différents images avec différent caractéristiques (bande spectrale, résolution) montrent clairement que ces méthodes sont très robustes et précises.

Contents

Glossary

Part 1

Abbreviation	Concept
NOAA	National Oceanic and Atmospheric Administration
MRF	Markov Random Field
ICM	Iterated Conditional Mode
DS	Dempster-Shafer
SSVD	simultaneous shadow vegetation detection
TP, TN	True Positive, True Negative
FP, FN	False Positive, False Negative
Acc	Overall accuracy
MCC	Matthews Correlation Coefficient
SIs	Shadow indices
VIs	Vegetation indices

Notation	Signification
$\mathcal{S}$	Set of pixels on the lattice or set of nodes on the graph
s	Considered pixel (or node) on $\mathcal{S}$ ($s \in \mathcal{S}$)
Ω	Set of class labels
ω_i	Class label
$\mathbf{X} \ldots (\mathbf{x})$	Label field (realization of label field)
$X_s \ldots (x_s)$	Label of $\mathbf{X}$ at pixel s (realization of X_s)
$\mathbf{Y} \ldots (\mathbf{y})$	Observable field (realization of observable field)
$\mathbf{Y}_s \ldots (\mathbf{y}_s)$	Value of $\mathbf{Y}$ at pixel s (realization of $\mathbf{Y}_s$)
$Y^{(j)}$	The jth component of $\mathbf{Y}$
$Y_s^{(j)} \ldots (y_s^{(j)})$	Value of field $Y^{(j)}$ at pixel s (realization of $Y_s^{(j)}$)
$P(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y})$	Probability that variable field $\mathbf{X}$ will be $\mathbf{x}$ given the realization $\mathbf{y}$ of observable field $\mathbf{Y}$
$\hat{x}_{\mathcal{S}-\{s\}}$	Provisional estimate of the random field everywhere except at pixel s
$\mathcal{V}_s$	neighborhood of pixel s in S.
$\hat{x}_{\mathcal{V}_s}$	Provisional estimate of the neighborhood of pixel s
$P(X_s = x_s|\hat{x}_{X_{\mathcal{V}_s} = \mathcal{V}_s})$	Probability that X_s will be x_s given the estimate of the neighborhood of pixel s
$P(X_s = x_s|X_{\mathcal{S}-\{s\}} = \hat{x}_{\mathcal{S}-\{s\}})$	Probability that X_s will be x_s given the estimate of every pixels except pixel s
$P(\mathbf{Y}_s = \mathbf{y}_s|X_s = x_s)$	Probability that $\mathbf{Y}_s$ will be $\mathbf{y}_s$ given the estimate x_s of pixel s
$\hat{x}_s$	Provisional estimate of pixel s
β	Parameter of regularization term
δ	Kronecker's delta function
Θ	Frame of discernment (the set of all simple hypothesis)

Part 2

Abbreviation	Concept
NOAA	National Oceanic and Atmospheric Administration
VHR	Very High Resolution
RAG	Region Adjacency Graph
MBR	Minimum Bounding Rectangle
RMBR	Recursive Minimum Bounding Rectangle
SAR	Synthetic Aperture Radar
DSM	Digital Surface Models
DTM	Digital Terrain Models
RoI	Region of Interest
LiDAR	Light Detection And Ranging
CRF	Conditional Random Field
SBBD	shadow-based building detection

Notation	Signification
$\mathcal{S}$	Set of pixels on the lattice or set of nodes on the graph
M_S	Shadow mask
M_{SB}	Shadow mask after eliminating shadows cast by non-building objects

Introduction

Contents

1.1 Context

On January 12, 2010, a devastating earthquake with a magnitude of 7.3 struck Haiti. This massive earthquake, the biggest the region had seen in 200 years, lead more than 200 000 lives lost, tens of thousands injured, leaving close to 2 million people homeless and the whole country to be rebuilt. The response of the international community has been unprecedented. Hundreds of data were acquired during the emergency phase: optical and radar satellite images covering various spatial resolutions, aerial photography and in situ measurements. After photo-interpretation, these images were used by the civil protection agencies and rescue teams on the scene. The French Agence Nationale de la Recherche (ANR) has funded a project named KAL-Haiti which aims at gathering remote sensing imagery completed as possible with in-situ measurements and exogenous data into a knowledge base. This geo-referenced database, seen as a shareable resource, can serve as a basis for helping the reconstruction of the country, but also as a reference for scientific studies devoted to all phases of risk management.

Before the earthquake, the Sertit (SErvice Régional de Traitement d'Image et de Télédétection) was aware of the links between the urban community of Strasbourg (its home town), and Jacmel in Haiti, with which it was twinned. The historic city of Jacmel was largely destroyed by the earthquake. Construction and reconstruction of buildings after the disaster is one of the major challenges facing the Haitian nation and the international community. A lot of initiatives have been carried out: destroyed buildings have been removed or reconstructed, new buildings have been erected, etc. The town council of Jacmel thus used its connections with the University of Strasbourg to monitor this rapidly evolving situation, which would enable them to establish the criteria for taxation, the economy planning, health care, etc. However, updating the building database is very time-consuming and expensive since it is generally carried out manually, by visual inspection of images. This work was estimated to require up to 40% of the costs necessary to generate it from scratch [Champion 2010]. As a consequence, there is a growing need to automate this process, that is to propose an automatic method for detecting changes in a 2D

building database. As a partner of the KAL-Haiti project, Sertit wants to develop an automatic system to detect changes in building database. This PhD project, conducted at the MIV team of the ICube laboratory, is a part of the effort devoted to the study of 2D building database updating.
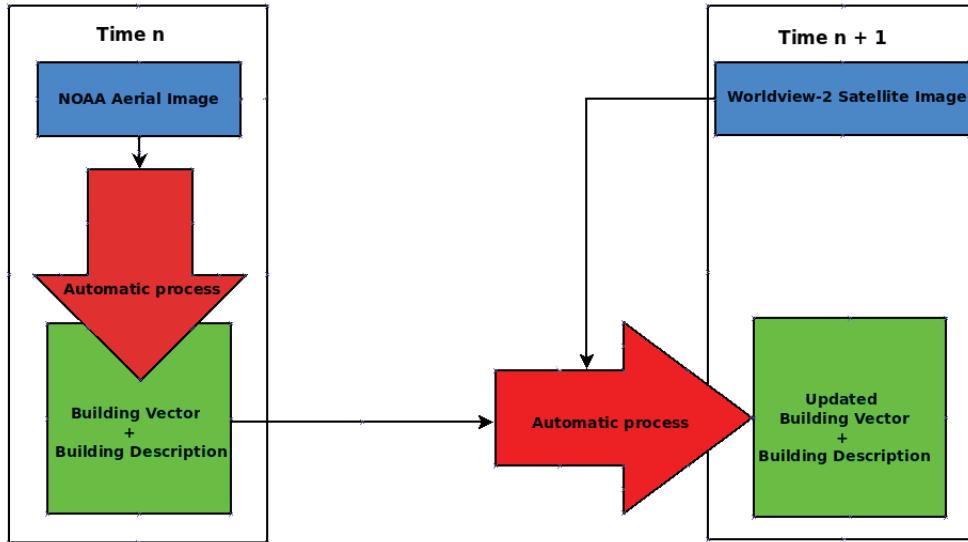

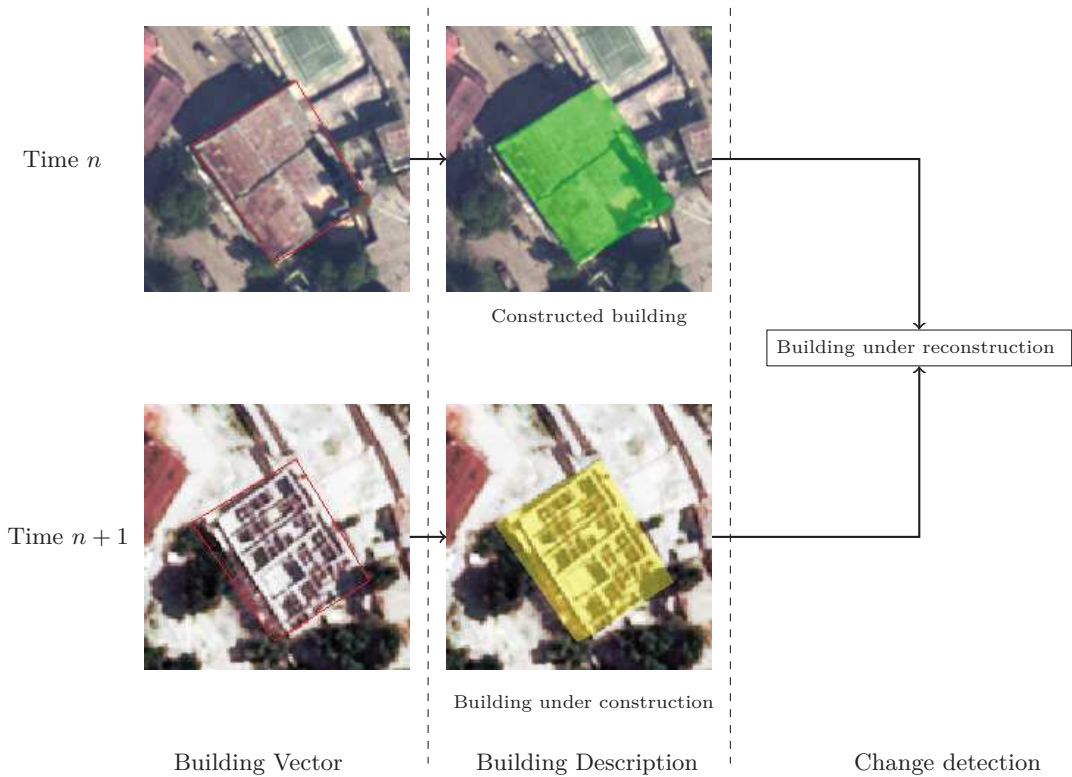Figure 1.1: Updating the building database of time n using new images of time $n+1$

Figure 1.2: An example of building database update (image credit: Sertit).

.

The updating process is described in Fig. 1.1. Buildings are first detected from images of time n and then, classified into three states:

- constructed building, which is apparently completed and in good condition;

- building under construction, which is not yet completed;

- building in ruin, which is clearly inhabitable and/or invaded by vegetation.

At time $n+1$, using new images, the new building footprints are estimated and the building state is updated. Moreover, new buildings are detected. Thus, to keep the consistency in the database, a 4th state corresponds to "non-existent" building is added: if at time $n+1$, a new building is detected, it is identified as "non-existent" building of time n. In the same way, the building will be identified as "non-existent" building of time $n+1$ after a completed disappearance.

Finally, by comparing the state of buildings at time n and time $n+1$, the changes on building database would be classified into different categories: new construction, building under construction, no evolution, completed construction, abandoned construction, building under reconstruction, destroyed building, damaged building. An example of building database update is shown in Fig. 1.2.

1.2 Problem Statement

This work takes place in the framework of high resolution remote sensing image analysis. The goal is to build a generic processing chain to create or update a cartographic building databases. Our work differs from the existing work in some respects as follows. First, it is designed to deal with optical images without any a priori knowledge such as DSM, DTM [Champion 2011] or SAR sensors [Poulain 2010]. Second, since a building database is not available at the input of the chain, the goal of this work is precisely to create one building database (building vector). Third, the state of buildings must be classified into different states (building description). The changes between two images of different time are not only the building footprints, but also their corresponding states. Thus, we propose a processing chain as shown in Fig. 1.3.
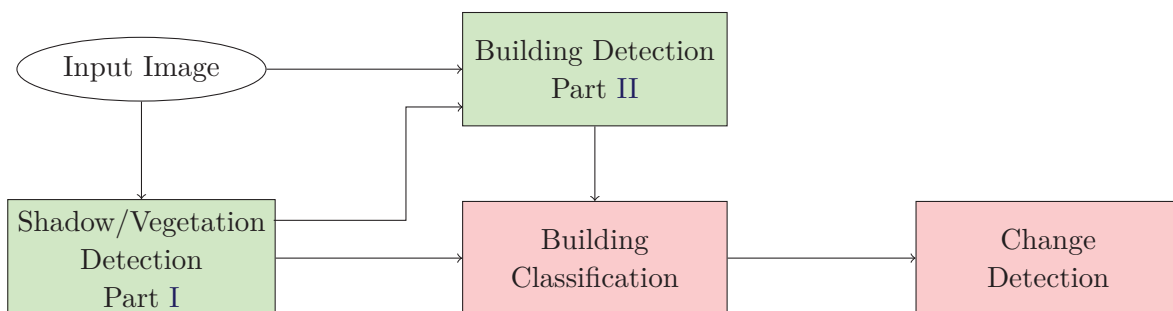

Figure 1.3: The workflow of building database update. The blocks in green are studied in the limit of this thesis.

To create the building database, an automatic algorithm for detecting buildings from single optical images is required. Moreover, shadows are valuable sources since a cast shadow is notably

strong evidence of an existence of a building structure. Besides, according to cartographic experts, shadows and vegetation can provide additional geometric and semantic clue about the state of buildings (constructed building, building under construction, or building in ruin). That is why we found it necessary to detect shadows and vegetation. These two issues are studied and presented in this thesis.

1.3   Thesis Contributions

Current shadow/vegetation detection methods in the literature detect separately shadow regions and vegetation regions [Shorter 2009], [Ozgun Ok 2013]. The first contribution of the thesis introduces a new algorithm that detects simultaneously shadows and vegetation from an high resolution optical image. In this thesis, several shadow and vegetation indices were investigated and merged using the Dempster-Shafer (DS) evidence theory so as to increase the reliability and accuracy of the segmentation. The principal advantage of this theory is its ability to take into account ignorance of the information by affecting a degree of confidence which is called a mass function to all simple and compound hypotheses. However, since the DS fusion processes at a pixel-level, the performance of the DS fusion is sensitive to noise. The Markov random field (MRF) is thus employed to model spatial information within the image. In short, the main contribution of this part is twofold.

- First, to the best of our knowledge, this is the first work on detecting simultaneously shadow regions and vegetation regions from optical (RGB aerial or satellite) image, by the use of Dempster-Shafer evidence theory.

- Second, Dempster-Shafer fusion is incorporated with a Markov random field, so that the image geometric information (spatial correlation) is taken into account.

Building detection from single image is, in general, a very difficult task due to the large scene complexity and irregular nature of the scenes [Izadi 2010], [Karadag 2015]. The second contribution of the thesis introduces a novel approach for the automated detection of buildings. Shadow analysis has been considered to be one of the most important clues of building objects in monocular image processing. In this thesis, from the shadow/vegetation mask detected in the previous part, boundaries between shadows and their corresponding buildings are detected by eliminating shadow regions generated by vegetation objects and other non-building objects. A novel region growing segmentation technique is then proposed. Image is first over-segmented into smaller homogeneous regions (superpixels) which can be used to replace the rigid structure of the pixel grid. An iterative region classification-merging is then applied over this set of regions. At each iteration, regions are first classified using a region-level MRF-based image segmentation. According to the position of shadows, a merging process is then performed over regions having the same class to produce regions whose shapes are appropriate to rectangles. In the third stage, from the results of region growing image segmentation, the final building regions are determined. In short, the main contribution of this part is threefold.

- The method requires only reasonably high resolution aerial image ($\simeq 50$ cm per pixel). Not required are multiple views, additional information such as near-infrared (NIR), lidar or any elevation data.

- The method does not require (a large amount of) prelabeled data as would be needed by a supervised learning.

- A novel Markov random field region growing technique for building detection is proposed, in which the radiometric and geometric information of building are combined.

These two algorithms have been tested on a variety of image datasets and demonstrates their efficiency.

1.4   Thesis Structure

The chapters of this thesis reflect the contributions stated in Section 1.3. The thesis is structured as shown below.

Chapter 2 outlines the Jacmel remote sensing dataset and three additional datasets used to evaluate the performance of the proposed algorithms in the thesis. The discussion concentrates on the image specification and ground truth generations of the datasets.

In Chapter 3, we review shortly the existing methods on shadow/vegetation detection. Then, among different shadow indices and vegetation indices proposed in the literature, we carry out the test to determine what indices are the most relevant regarding shadow detection task and vegetation detection task. Moreover, we explain what motivate us to propose a novel method detecting simultaneously shadows/vegetation from a high resolution optical image, instead of detecting sequentially shadows, then vegetation like other existing methods. Based on the indices chosen in Chapter 3, Chapter 4 develops a novel simultaneous shadow/vegetation detection method. The basic concepts of DS evidence theory and its applications for combining different shadow indices and vegetation indices are described. The DS fusion is carried out pixel by pixel and is incorporated in the Markovian context while obtaining the optimal segmentation with the energy minimization scheme associated with the MRF. The experimental results and discussion are also reported in the end of this chapter.

Coming now to the second part of the thesis, in Chapter 5, a novel approach for the automated detection of buildings from monocular VHR optical images is introduced. From the shadow/vegetation mask detected in the previous part, boundaries between shadows and their corresponding buildings are detected by eliminating shadow regions generated by vegetation objects and other non-building objects. In order to effectively extract building objects from image, an novel MRF region growing image segmentation is proposed. From the segmentation results, an algorithm for determining the final building objects is proposed. In Chapter 6, we first analyze the sensitivity of important parameters used in our algorithm. Then, we validate our method on multiple datasets described in Chapter 2 and compare to other existing methods in the literature. The advantages and the limitations of the proposed method are also discussed in the end of this chapter.

Finally, Chapter 7 gives concluding remarks on the proposed techniques, with some open issues and future research works.

1.5 List of Publications

The author's publications:

1. T.-T. Ngo, Ch. Collet, V. Mazet, Détection simultanée de l'ombre et la végétation sur des images aériennes couleur en haute résolution, *Traitement du signal, special issue, pp. 311–333, Vol. 32, Num. 2-3, 2015.*

2. T.-T. Ngo, Ch. Collet, V. Mazet, "Automatic rectangular building detection from VHR aerial imagery using shadow and image segmentation", *IEEE International Conference on Image Processing ICIP'15*, Québec City, Canada, september 2015.

3. T.-T. Ngo, Ch. Collet, V. Mazet, "MRF and Dempster-Shafer theory for simultaneous shadow/vegetation detection on high resolution aerial color images", *IEEE International Conference on Image Processing ICIP'14*, Paris, France, october 2014.

4. T.-T. Ngo, Ch. Collet, V. Mazet, "Détection simultanée de l'ombre et la végétation sur des images aériennes couleur en haute résolution", *RFIA'14 Reconnaissance de Formes et l'Intelligence Artificielle* , Rouen, juin 2014.

Experimental Data

Contents

2.1 Introduction

This chapter introduces data used in this thesis, including the Jacmel remote sensing dataset and three additional evaluation datasets. We will discuss about the image specification (spectral band, spatial resolution) and the ground truth generations of each dataset. These datasets with different image specifications and environmental settings are used to demonstrate the flexibility of the proposed algorithms.

2.2 Jacmel Remote Sensing Dataset

Project Motivation Taking into account the strong demand for helping reconstruction following the 12th January 2010 earthquake in Haiti, the principal aim of KAL-Haiti project is to produce and promote the use of a database of Earth observation data, in support of research and development activities for global risk management and sustainable reconstruction in Haiti. During the immediate response to the disaster, tens of satellite images, both optical and radar, covering various spatial resolutions were acquired by satellite operators whether they are national space agencies or private companies [Giros 2012]. Two images over the Jacmel area are studied in this thesis:

- a NOAA aerial image, acquired on 24th January 2010 (12 days after the earthquake);

- a Worldview-2 satellite image, acquired on 17th July 2011 (18 months after the earthquake).

These images represent respectively the image of state n and of state $n + 1$ in Diagram 1.1 of Chapter 1.

Figure 2.1: Survey area: the city of Jacmel

Ground Truth   Concerning the reconstruction monitoring in Jacmel area (as shown in Fig. 2.1), the KAL-Haiti project has produced a set of vector layers built from satellite and aerial imagery, and showing the buildings footprints at different periods. Three layers showing the footprints and states of 22 257 buildings over the Jacmel area (50 km^2) have been produced by photo-interpretation [Sertit 2013]. The set of building footprints produced by photo-interpretation is used as the ground truth to evaluate the performance of proposed building detection method. For shadow/vegetation detection method, the segmentation map with three classes (shadow, vegetation, other) is manually produced with the validation of the cartographic expert (Sertit).

2.2.1   NOAA Aerial Image

NOAA is an abbreviation of National Oceanic and Atmospheric Administration, scientific agency within the United States Department of Commerce focused on the conditions of the oceans and the atmosphere. For the Haiti earthquake disaster, the NOAA aerial images are acquired by the Cessna Citation II, a versatile twin-engine jet aircraft which is equipped with high-resolution digital cameras and other sensors that support a wide variety of remote sensing configurations, including large-format aerial photography as well as data collection for digital cameras, hyperspectral, multispectral and LIDAR systems.

Image Specification   The NOAA aerial image provided for this thesis conveys 3 spectral bands (red, green, blue) and is characterized by a spatial resolution of 24 cm. Sample location is shown in Fig. 2.2.

Figure 2.2: A sub-image (of size 600×720 pixels) of NOAA aerial image that covers a part of Jacmel

2.2.2 WorldView-2 Satellite Image

WorldView-2 is a commercial Earth observation satellite owned by DigitalGlobe, an American commercial vendor of space imagery and geospatial content, and operator of civilian remote sensing spacecraft. For the Haiti earthquake disaster, DigitalGlobe provided a free-of-charge high resolution images to aid the extensive relief and recovery efforts.

Image Specification The WorldView-2 satellite image provided for this thesis conveys 4 spectral bands (red, green, blue, near-infrared) and is characterized by a spatial resolution of 50 cm. Sample location is shown in Fig. 2.3.

2.3 Additional Dataset

These dataset were provided in the end of the PhD project and used to demonstrate the flexibility and the efficiency of proposed methods. The reference data (both for shadow/vegetation detection and building detection algorithm) are manually produced with the validation of the cartographic expert (Sertit).

(a) Visible RGB image   (b) NIR image

Figure 2.3: A sub-image (of size 420 × 504 pixels) of Worldview-2 satellite image that covers a part of Jacmel

2.3.1   BD ORTHO® Dataset

The Departmental Orthophotography of IGN (BD ORTHO®) is a database of the Digital aerial georeferenced photography, produced by the French national mapping agency (*Institut national de l'information géographique et forestière*, IGN France). The BD ORTHO® covers part of the French territory with Orthophotographs (aerial photographs in which all distortions are corrected). It is generated at a ground sample distance of 50 cm. The coverage by "département" (French administrative unit, mean area of around 6500 km^2) is renewed every 3 to 4 years.

Image Specification   The BD ORTHO® image provided for this thesis represents a part of Strasbourg city and is shown in Fig. 2.4. It conveys 3 spectral bands (red, green, blue) and is characterized by a spatial resolution of 50 cm.

2.3.2   Pléiades-HR Satellite Image

Pléiades is the CNES (French national space agency) program designed as the follow-on to its highly successful SPOT series of low Earth orbit multi-mission observation satellites, which has operated an uninterrupted service since 1986. The Pléiades constellation is composed of two very-high-resolution optical Earth-imaging satellites. Pléiades 1A and Pléiades 1B provide the coverage of Earth's surface with a repeat cycle of 26 days. Designed as a dual civil/military system, Pléiades will meet the space imagery requirements of European defence as well as civil and commercial needs.

Figure 2.4: BD ORTHO® image of size 2563 × 2563 pixels that covers the part of Strasbourg

Image Specification   The Pléiades-HR Satellite image image provided for this thesis represents a part of Strasbourg city and is shown in Fig. 2.5. It conveys 4 spectral bands (red, green, blue, NIR) and is characterized by a spatial resolution of 50 cm.

2.3.3   SZTAKI-INRIA Building Detection Benchmark

The SZTAKI-INRIA Benchmark [Benedek 2012] is an excellent resource for benchmarking building extraction algorithms. It contains the rectangular footprints of 665 buildings in 9 aerial or satellite images taken from Budapest and Szada (both in Hungary), Manchester (U.K.), Bodensee (Germany), Normandy and Côte d'Azur (both in France), as well as manually annotated ground truth data. Two datasets (Budapest and Szada) are aerial images, and the remaining four datasets are satellite images acquired from Google Earth. This benchmark is free for scientific use.

Image Specification   All of the images in the SZTAKI-INRIA Building Detection Benchmark used in this thesis, with authors' acceptance, convey 3 spectral bands (red, green, blue). Sample locations are shown in Fig. 2.6.

(a) Visible RGB image

(b) NIR image
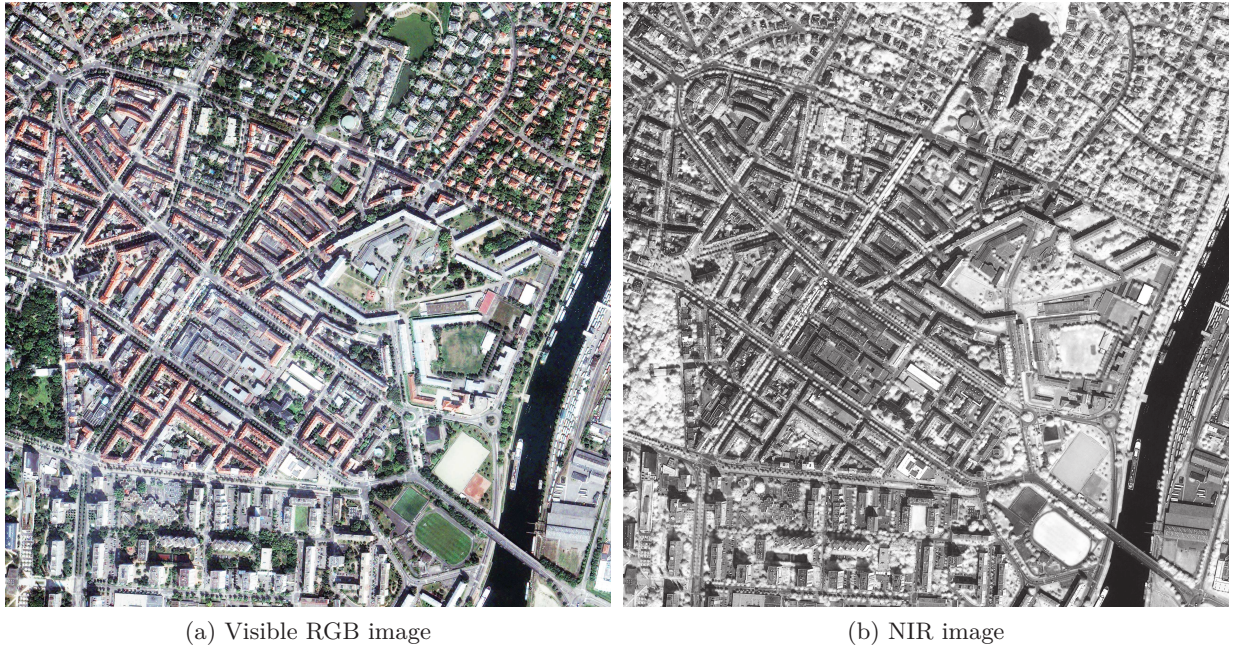
Figure 2.5: Pléiades-HR Satellite Image of size 2563 × 2563 pixels that covers the part of Strasbourg

(a) Manchester area

(b) Normandy area

Figure 2.6: Images from the SZTAKI-INRIA Building Detection Benchmark

2.4 Conclusions

In this work, four datasets will be used. They can be categorized into two groups in function of their spectral specification as follows:

- RGB image: NOAA image, BD ORTHO® image, SZTAKI-INRIA image;

- RGB-NIR image: Worldview-2 image, Pléiades-HR image.

These image datasets that cover different scenes and have different environmental settings, different specifications, different resolutions allow to demonstrate the reliability and flexibility of the proposed algorithms.

Part I

Shadow/Vegetation Detection

Related Research

Contents

3.1 Introduction

Shadows and vegetation in remote sensing images often result in problems for many applications, such as land-cover classification, change detection and damage detection in disasters. In recent years, research on these two issues has become very popular. In this chapter, we first review shortly the existing methods on shadow/vegetation detection. Among different shadow indices and vegetation indices proposed in the literature, we then run the test to determine what indices are the most relevant regarding shadow detection task and vegetation detection task. Moreover, we explain our motivation for a novel simultaneous shadow/vegetation detection method.

3.2 Shadow Detection in Remote Sensing Image

3.2.1 Related Works

Automatic shadow detection is a very important preprocessing step for many remote sensing applications, particularly for images acquired with high spatial resolution. Shadows can provide geometric and semantic information contained in images, including cues about the height, the shape [Kim 2007] and/or the position of buildings [Ozgun Ok 2013]. However, the existence of shadows also cause some undesirable problems. For example, shadows may cause objects to merge or shapes to distort, thus resulting in information loss or distortion of objects [Chung 2009]. Hence, detecting shadows is still a current topic and discussed widely in remote sensing literature.

Normally, shadows appear when objects occlude the direct light from the illumination source, usually the sun. But shadows are not all the same, they can be divided in two different classes:
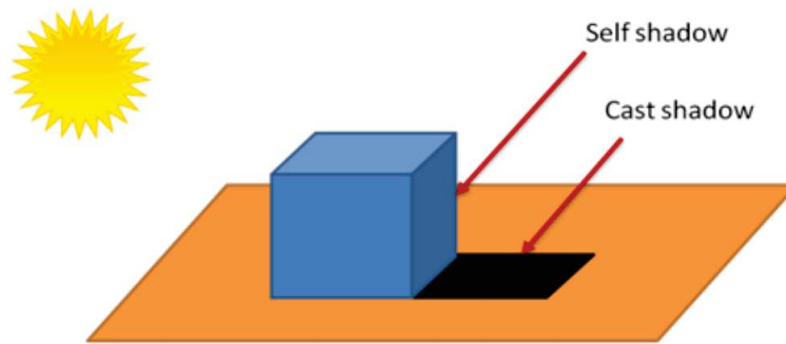
Figure 3.1: Illustration of cast and self shadows [Luca 2012].

cast and self shadows (as illustrated in Fig. 3.1). The former is caused by the projection of the light source in the direction of the object. The latter is still a shadow but represents the part of the object that is not illuminated directly by the light source. Consequently, in this thesis, we only focus on the detection of cast shadow, and we do not consider self shadow as shadow for detection. For the sake of simplicity, in the rest of the thesis, we use the word shadow to mean the cast shadow unless otherwise noted.

In the literature, different approaches have been proposed regarding shadow detection. They can loosely be classified into two categories, namely model-based [Tolt 2011, Nakajima 2002] and properties-based approaches [Arévalo 2008, Chen 2010, Cucchiara 2003, Tsai 2006, Salvador 2004, Sun 2010]. The former has limited applicability because they rely on some prior information like illumination conditions and 3D scene geometry. 3D models can be acquired using aerial or satellite photogrammetry (multi-view aerial images, stereoscopic pair of satellite images, etc.), airborne or terrestrial laser scanning, interferometric SAR (Synthetic Aperture Radar) or even from topographic data [Adeline 2013]. However, since usually such knowledge is not available, most of the detection algorithms are based on shadow properties, such as the fact that shadow areas have lower brightness, higher saturation and greater hue values. These shadow properties are exploited by different shadow indices (SIs) proposed in the literature.

For instance, in [Salvador 2004], a hypothesis is applied on the fact that cast shadows darken the surface which they are cast upon, and the validity of detected regions as shadows is further verified using the color feature c_3 of the $c_1c_2c_3$ color space (proposed by [Gevers 1999]), with geometric properties of shadows. Considering the atmospheric Rayleigh scattering effect, [M.Polidorio 2003] found that shadow areas in images have higher saturation and lower luminance, thus, they propose thresholding the difference image of the saturation and the intensity components for each pixel in a normalized HSI color space. Based on Phong illumination model [Phong 1975], [Huang 2004] presented a new imaging model of shadows, and employed thresholding on the hue, blue, and green-blue difference components sequentially to detect shadowed areas, which were then compensated by applying the Retinex technique to both shadowed and non-shadowed area separately. [Tsai 2006] uses different color spaces (HSV, HSI, HCV, YIQ, YCbCr) for shadow detection from aerial images where different ratio maps for each of these color spaces are calculated and then, Otsu method is used to segment shadow. But this method tends to misclassify the dark blue and dark green color surfaces as shadow regions.

Inspired by this method, a better approach was developed, which is based on a novel successive thresholding scheme [Chung 2009]. A recent paper from [Tian 2012] proposes an alternative method for shadow detection using a tri-color attenuation model (TAM). TAM describes the attenuation relationship between shadows and their non-shadow backgrounds in the three color channels. Finally, physical properties (e.g., temperature) of a blackbody radiator have been exploited in a recent method to detect shadows [Makarau 2011]. But since this method needs the prior measurement of the color temperature or the center wavelength for the sensors of each channel, it is not automatic enough for general data.

Our shadow detection method is devoted to the high resolution optical image without any prior information. Inspired of the existing methods, our approach exploits shadow properties through use of SIs. The question is now what indices are the most effective for shadow detection task. In the following, we investigate some SIs proposed in the literature over our image database.

3.2.2 Choice of Shadow Index

The main issue of this section is to give an experimental comparison of different indices regarding shadow detection on high resolution optical image. These SIs are produced from different color space models: RGB-NIR, $c_1c_2c_3$, HSI, HSV, HCV, YIQ, YC_bC_r, and presented in Table 3.1. The symbols stand for: R = Red, G = Green, B = Blue, NIR = Near Infrared, H = Hue, C = Chroma, S = Saturation, V = Value, I = Intensity; c_1, c_2, c_3 represent first, second and third chrominance respectively; in YIQ, Y stands for luminance, I for in-phase and Q for quadrature; in YC_bC_r, Y represents luminance and C_b, C_r represent chrominance. An overview of these color space models are presented in Appendix B.

Index	Formula	Color model	Author
c_3	c_3	$c_1c_2c_3$	[Salvador 2004]
NIR	NIR	RGB-NIR	[Nagao 1979]
SI_1	I - S	HSI	[M.Polidorio 2003]
SI_2	$\dfrac{H+1}{I+1}$	HSI	[Tsai 2006]
SI_3	$\dfrac{Q+1}{Y+1}$	YIQ	[Tsai 2006]
SI_4	$\dfrac{C_r+1}{Y+1}$	YC_bC_r	[Tsai 2006]

Table 3.1: Different shadow indices that are tested in this section. Among six color spaces used in [Tsai 2006], we test only three color spaces HSI, YIQ, YC_bC_r that are demonstrated in the author's work as the best for shadow detection task.

We have tested these SIs over all image datasets presented in Chapter 2. An example of three test images and their manually shadow masks is shown in Fig. 3.2. The efficiency of each shadow index can be observed in Table 3.2, where we plot the one dimensional marginal histograms of the shadow index values for manually marked shadowed and non-shadowed points. We observe that the c_3 and SI_1 are the most appropriate for a shadow detection method based on histogram thresholding, since the overlaps between the shadow and non-shadow histogram are less significant than others.
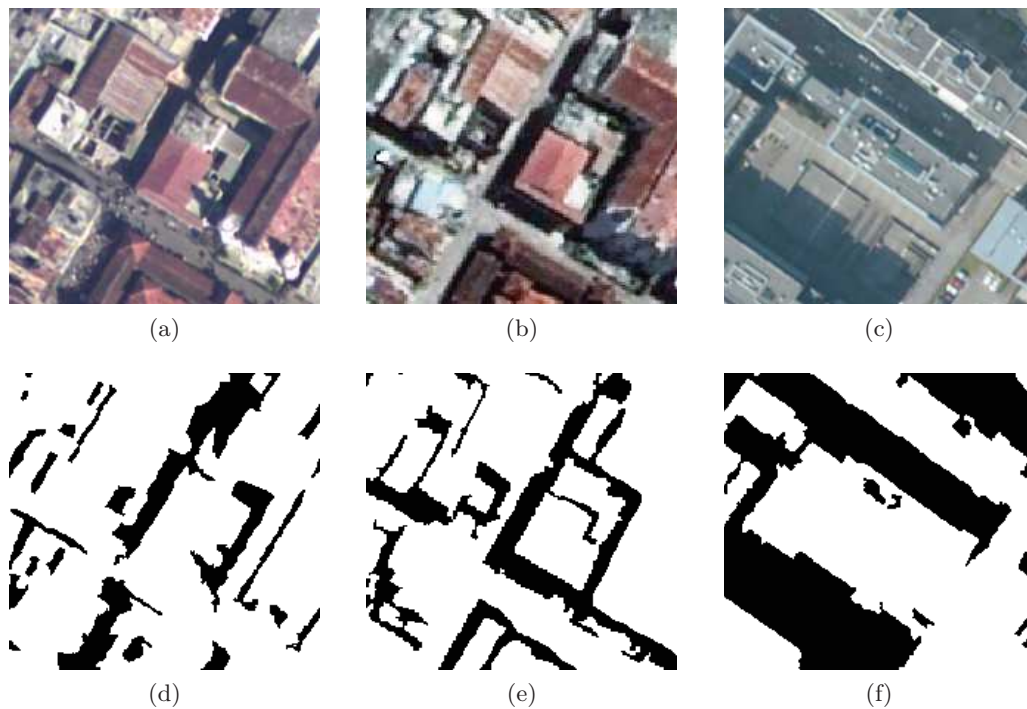
Figure 3.2: First line: RGB composition images, (a) NOAA aerial image, (b) Satellite Worldview-2 image, (c) BD ORTHO® aerial image. Second line: manually interpreted shadow masks (black area) as ground truth of shadow regions. These images have the size of 200 x 210 pixels.

In properties-based shadow detection method, shadow regions can be extracted by simply choosing an appropriate threshold. The search of that threshold in the histogram is usually based on some assumptions on the histogram: bimodal, Gaussian mixture model, the number of peaks and valleys, etc [Adeline 2013]. In this section, we apply Otsu's method [Otsu 1975] for automatically determining the threshold. The details of Otsu's method can be found in Appendix A. Resulting shadow masks are shown in Table 3.3. By comparing with the reference data (manually shadow masks) in Fig. 3.2, we visually find that c_3 is the best for shadow detection task.

Besides, the performance of each shadow index is also assessed by comparing the resulting shadow masks with the manually shadow masks. The evaluation metrics are defined based on true positive (TP), false negative (FN), false positive (FP), true negative (TN) (as shown in Table 3.4). Concerning shadow detection, TP is the number of shadow pixels correctly identified, FN is the number of shadow pixels identified as non-shadow, FP is the number of non-shadow pixels identified as shadows and TN is the number of non-shadow pixels correctly identified. Among these metrics, the producer's accuracies (also known as recall) indicate how well pixels of known categories are correctly classified. The user's accuracies (also known as precision) indicate the probabilities of pixels been correctly classified into actual categories on the ground. Combining the accuracies of the user and the producer, the overall accuracy (Acc) can be used to evaluate the correctness of the algorithm. These measures are given in percentage. For a good algorithm, values of these metrics should be high. Beside the overall accuracy, similar to [Rufenacht 2014], we compute Matthews Correlation Coefficient (MCC) [Matthews 1975], which is a more balanced

SI	Histogram of radiance for shadow index for image in Fig. 3.2a	Histogram of radiance for shadow index for image in Fig. 3.2b	Histogram of radiance for shadow index for image in Fig. 3.2c
c_3			
NIR	Unavailable		Unavailable
SI_1			
SI_2			
SI_3			
SI_4			

Table 3.2: Different shadow indices (first column) and histogram of radiance for shadow indices for images in Fig. 3.2. The black curve corresponds to shadow, the red curve corresponds to non-shadow. The overlaps between the shadow and non-shadow histogram of c_3 and SI_1 are less significant than others.
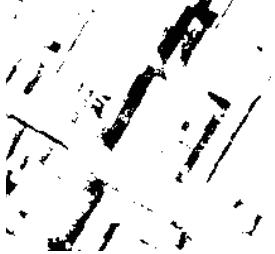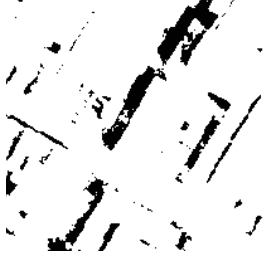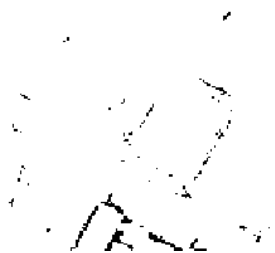
SI	Detected shadows (Fig. 3.2a)	Detected shadows (Fig. 3.2b)	Detected shadows (Fig. 3.2c)
Ground truth			
c_3			
NIR	Not available		Not available
SI_1			
SI_2			
SI_3			
SI_4			

Table 3.3:   Resulting shadow masks by Otsu thresholding method from different SIs for images in Fig. 3.2

measure if the two classes have different sizes. The value of the MCC is between $[-1, 1]$, where larger values indicate better prediction.

Producer's accuracies		User's accuracies	
Shadow (recall rate) $P_s = \dfrac{TP}{TP + FN}$	Non-shadow $P_n = \dfrac{TN}{TN + FP}$	Shadow (precision rate) $U_s = \dfrac{TP}{TP + FP}$	Non-shadow $U_n = \dfrac{TN}{TN + FN}$
Overall accuracy $Acc = \dfrac{TP + TN}{TP + TN + FP + FN}$		Matthews correlation coefficient $MCC = \dfrac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	

Table 3.4: Accuracy table and metrics for experimental evaluation.

	Shadow detection for image in Fig. 3.2a					
SI	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
c_3	93.33	91.69	72.70	98.30	**92.00**	**0.77**
NIR	-	-	-	-	-	-
SI_1	97.98	64.69	39.69	99.26	71.07	0.49
SI_2	56.35	99.24	94.64	90.55	91.02	0.68
SI_3	49.41	99.32	94.56	89.22	89.75	0.63
SI_4	47.49	99.33	94.40	88.86	89.39	0.62
	Shadow detection for image in Fig. 3.2b					
Method	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
c_3	98.09	76.40	58.21	99.17	**81.84**	**0.67**
NIR	99.08	65.39	46.71	99.57	73.29	0.54
SI_1	99.06	54.30	39.89	99.47	64.79	0.45
SI_2	34.81	60.75	21.35	75.27	54.67	-0.03
SI_3	30.04	74.42	26.45	77.65	64.02	0.04
SI_4	6.94	99.68	87.01	77.76	77.94	0.20
	Shadow detection for image in Fig. 3.2c					
Method	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
c_3	93.80	95.48	93.81	95.47	**94.77**	**0.89**
NIR	-	-	-	-	-	-
SI_1	88.69	94.38	92.02	91.96	91.98	0.83
SI_2	77.61	98.66	97.69	85.80	89.78	0.79
SI_3	81.95	98.28	97.21	88.18	91.39	0.82
SI_4	77.31	97.78	96.21	85.52	89.14	0.78

Table 3.5: Shadow detection accuracy measurements for images in Fig. 3.2

Table 3.5 shows the shadow detection accuracy measurements for images in Fig. 3.2 when

we apply the Otsu thresholding method on different shadow indices. It is shown that the best result is obtained with shadow index c_3 (overall accuracy are 92%, 81% and 94%). Three indices proposed by [Tsai 2006] (denoted as SI_2, SI_3, SI_4) perform well for the first and third image but they perform badly for the second image. In [Adeline 2013], the authors concluded that since sensor radiance received from shadowed regions decreases from short to long wavelengths due to scattering, so it is easier to distinguish shadows from non-shadows with NIR band rather than visible bands. But our test demonstrates that NIR band (overall accuracy 73.29%) outperforms other shadow indices but it is not better than shadow indice c_3. Besides, NIR band is not available for aerial images. These shadow indices are also tested for other image datasets and similar results are obtained. As a result, in the following, we will use index c_3 for shadow detection task.

3.3   Vegetation Detection in Remoted Sensing Image

3.3.1   Related Works

Vegetation is of particular interest as it presents a versatile resource for effectively managing and moderating a variety of problems associated with urbanization, such as urban planning, disaster management or telecommunication planning. In the context of this thesis, the spatial distribution of vegetation can also provide the semantic information about the state of buildings. However, the detection of vegetation is a considerable challenge due to their complex nature and their interaction with other objects, such as buildings or shadows.

Previous studies document a variety of approaches for measuring the vegetation cover by image analysis techniques. Color/tone was intensively used to distinguish between vegetation and non vegetation. Since objects, including vegetation have their unique spectral signature, they can be identified according to their spectral characteristics, mostly by vegetation indices (VIs). VIs are spectral transformations of two or more bands designed to enhance the contribution of vegetation properties. As a simple transformation of spectral bands, they are computed directly without any bias or assumptions regarding land cover class, soil type, or climatic conditions [Huete 2002]. According to [Jackson 1983], the ideal vegetation index should be particularly sensitive to vegetative covers, insensitive to soil color and to soil brightness, little affected by atmospheric effects, environmental effects, solar illumination geometry and sensor viewing conditions. Numerous VIs have been proposed in the literature [Bannari 1997]: vegetation index number (VIN), normalized different vegetation index (NDVI), transformed vegetation index (TVI), to name a few. However, most of these VIs require the near-infrared band. Since our developed algorithm must be applicable in RGB color images, we focus on the VIs which utilized only the red, green and blue spectral bands. Among them, color index of vegetation extraction (CIVE) [Kataoka 2003], excess green index (ExG) [Woebbecke 1995b], redness index (RI) [Huete 1994] and excess green minus excess red (ExGR) [Neto 2006], are used. The advantages of using these indices is that they accentuate a particular color such as plant greenness, which should be intuitive for human comparison [Meyer 2008]. Moreover, they do not require the near-infrared band, which is not available for RGB color images.

Besides VIs, other classification algorithms have been applied on spectral features to classify

vegetation. They can be divided into supervised and unsupervised classification approaches. The most widely used supervised classification algorithm for satellite imagery is the Maximum Likelihood Classifier (MLC) [Xu 2005]. Methods for unsupervised classification most frequently used are the K-means and the ISODATA (Iterative Self-Organizing Data Analysis Technique) clustering algorithms [Xie 2008]. In the context of land-cover classification with multispectral satellite data, [Duda 2002] investigated several unsupervised classification (clustering) algorithms and compared with regard to their ability to reproduce ground data in a complex landscape. The clustering algorithms examined are K-means, extended K-means, agglomerative hierarchical, fuzzy K-means and fuzzy maximum likelihood and fuzzy clustering is found to perform best on satellite image. In more recent years, hyperspectral imagery is increasingly studied to extract vegetation cover as its wider spectral bands allow a better differentiation of individual species than multispectral data [Xie 2008, Wania 2007]. Note that these methods are mainly designed for land-cover classification (vegetation, sol, street, water, etc). To detect only vegetation, most used method is to exploit the NDVI index and employ a clustering algorithm such as K-means, Otsu's thresholding method, like [Ok 2013].

Our aim is to design a new unsupervised algorithm to detect vegetation, from optical high resolution image. Inspired of the existing methods, our approach exploits the spectral characteristics of vegetation through use of VIs. In the following, we investigate different VIs proposed in the literature over our image database to determine the most effective VI regarding vegetation detection.

3.3.2 Choice of Vegetation Index

Index	Abbreviation	Formula	Author
Excess Green	ExG	$\dfrac{2G - R - B}{R + G + B}$	[Woebbecke 1995b]
Excess Green minus Excess Red	ExGR	$\dfrac{3G - 2.4R - B}{R + G + B}$	[Neto 2006]
Color Index of Vegetation Extraction	CIVE	$\dfrac{0.411R - 0.811G + 0.385B}{R + G + B} + 18.787$	[Kataoka 2003]
Redness Index	RI	$\dfrac{R - G}{R + G}$	[Huete 1994]
Normalized Difference Vegetation Index	NDVI	$\dfrac{NIR - R}{NIR + R}$	[Rouse, JW 1974]

Table 3.6: Different vegetation indices that are tested in this section.

The main issue of this section is to give an experimental comparison of different indices regarding vegetation detection on high resolution optical image. Table 3.6 gives the most used VIs that are considered in this section. The first four indices employ only three bands R, G, B. The fifth index is the most frequently used vegetation index for vegetation detection: the Normalized Difference Vegetation Index (NDVI). But it requires the NIR band, which is not available in NOAA aerial image.
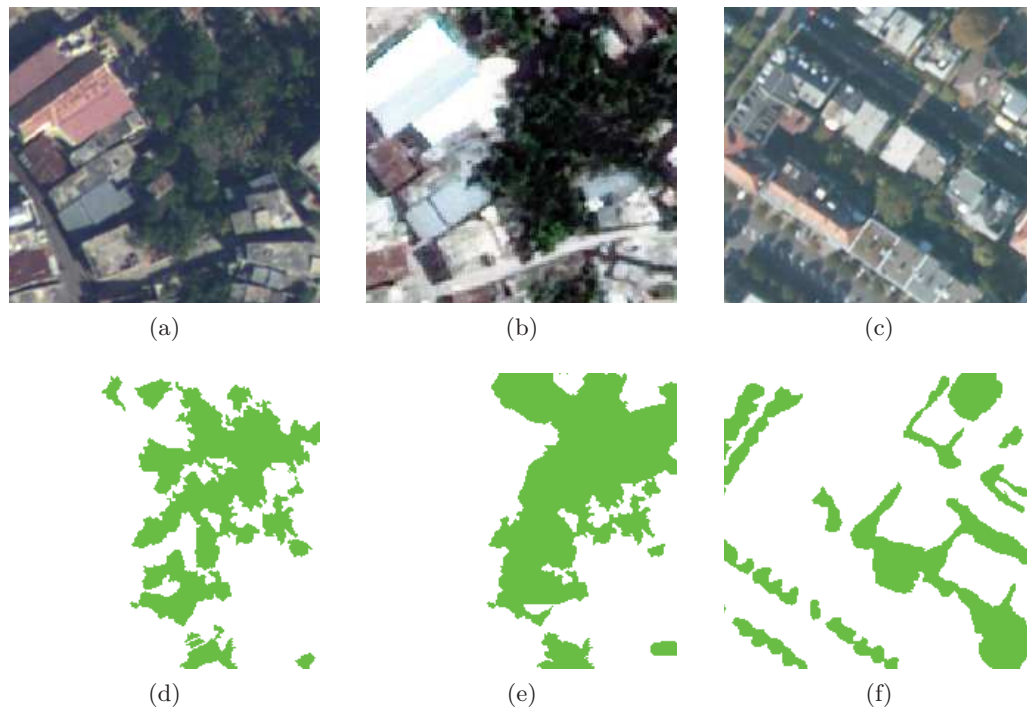
Figure 3.3: First line: RGB composition images, (a) NOAA aerial image, (b) Worldview-2 satellite image, (c) BD ORTHO® aerial image. Second line: corresponding manually interpreted vegetation masks (green area) as ground truth. These images have the size of 200 x 210 pixels.

We have tested these VIs over all image datasets presented in Chapter 2. An example of three test images and their manually vegetation masks is shown in Fig. 3.3. The efficiency of each vegetation index can be observed in Table 3.7, where we plot the one dimensional marginal histograms of the vegetation index values for manually marked vegetative and non-vegetative points. We observe that vegetation index ExG and NDVI (in case the NIR band is available) are the most appropriate for a vegetation detection method based on histogram thresholding, since their overlaps between the vegetation and non-vegetation histogram are less significant than others.

Similar to Section 3.2.2 in selecting the best shadow index, we apply the Otsu's thresholding method over the feature images computed from these VIs. Resulting vegetation masks are shown in Table 3.8. By comparing the resulting vegetation masks with the manually vegetation masks, we find visually that index ExG and NDVI (in case the NIR band is available) are the best. This observation can be confirmed by the quantitative evaluation. We evaluate quantitatively each detected vegetation mask by applying the same evaluation metric as described in Section 3.2.2. As shown in Table 3.9, for index ExG, the Acc (and MCC) is 92.88 (and 0.84) for Jacmel aerial image and 69.31 (and 0.34) for BD ORTHO® image. In case the NIR band is available, the NDVI is the most appropriate index (the Acc and MCC are 89.35 and 0.75 respectively). These vegetation indices are also tested for other image datasets and similar results are obtained. As a result, in the following, we use index ExG and index NDVI for vegetation detection task.
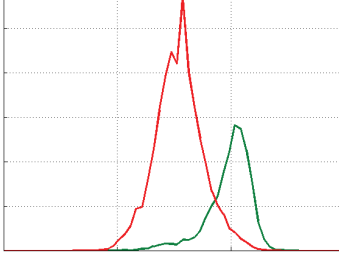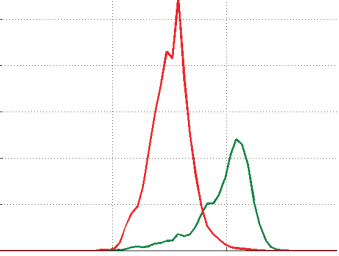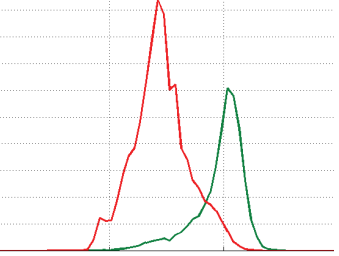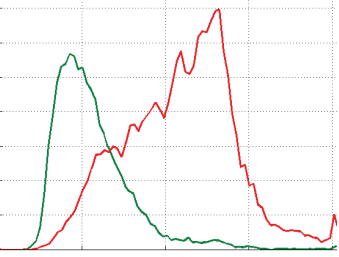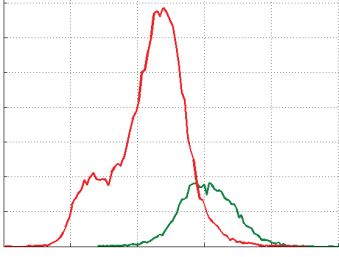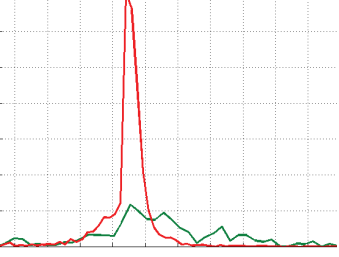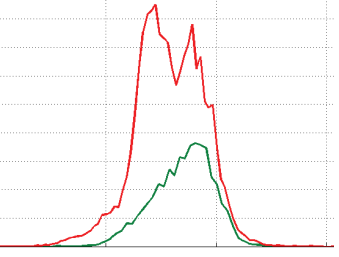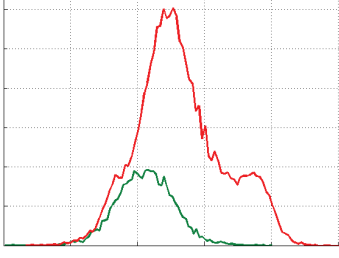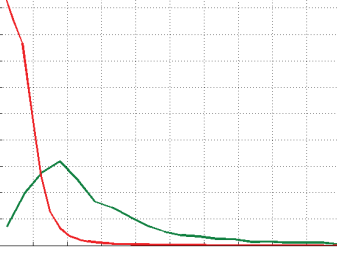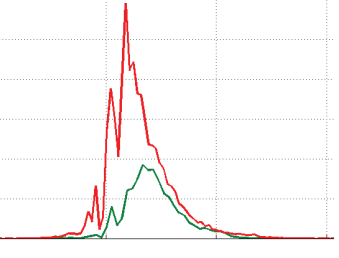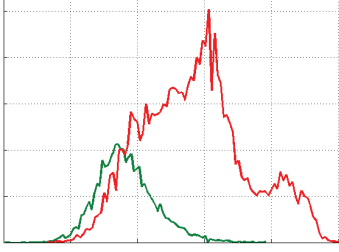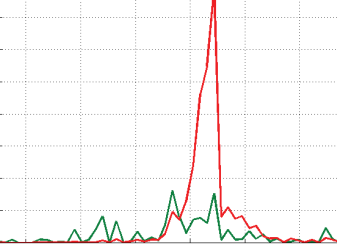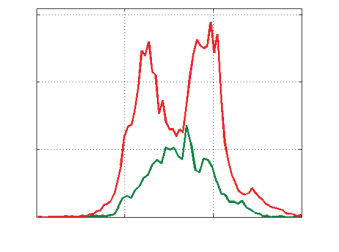
VI	Histogram of radiance for vegetation index for image in Fig. 3.3a	Histogram of radiance for vegetation index for image in Fig. 3.3b	Histogram of radiance for vegetation index for image in Fig. 3.3c
ExG			
NDVI	Not available		Not available
ExGR			
CIVE			
RI			

Table 3.7: Different VIs (first column) and histogram of radiance for vegetation indices for image in Fig. 3.3. The green curve corresponds to vegetation, the red curve corresponds to non-vegetation. The overlaps between the vegetation and non-vegetation histogram of ExG and NDVI are less significant than others.
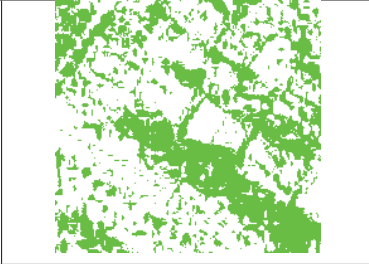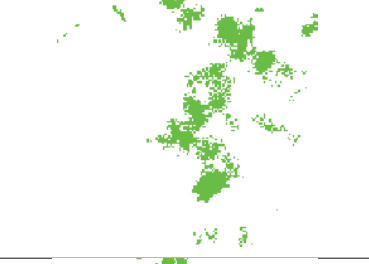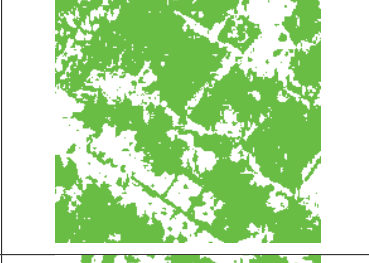
VI	Detected vegetation (Fig. 3.3a)	Detected vegetation (Fig. 3.3b)	Detected vegetation (Fig. 3.3c)
Ground truth			
ExG			
NDVI	Not available		Not available
ExGR			
CIVE			
RI			

Table 3.8: Resulting vegetation masks by Otsu's method from different VIs for image in Fig. 3.3

	Vegetation detection for image in Fig. 3.3a					
VI	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
ExG	99	91.27	77.82	99.32	**92.88**	**0.84**
NDVI	-	-	-	-	-	-
ExGR	88.48	57.01	38.65	94.17	64.38	0.38
CIVE	88.48	57.01	38.65	94.17	64.38	0.38
RI	96.14	57.30	40.80	97.98	66.40	0.45
	Vegetation detection for image in Fig. 3.3b					
VI	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
ExG	83.64	91.02	80.14	92.78	88.79	0.73
NDVI	84.29	91.54	81.18	93.08	**89.35**	**0.75**
ExGR	28.71	98.60	89.93	76.15	77.48	0.42
CIVE	36.19	98.00	88.73	78.00	79.33	0.47
RI	15.83	87.66	35.74	70.63	65.96	0.04
	Vegetation detection for image in Fig. 3.3c					
VI	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
ExG	72.12	68.49	40.02	89.39	**69.31**	**0.34**
NDVI	-	-	-	-	-	-
ExGR	67.37	52.37	29.19	84.63	55.76	0.16
CIVE	51.94	27.78	17.33	66.48	33.23	-0.18
RI	51.60	46.18	21.84	76.60	47.40	-0.01

Table 3.9: Vegetation detection accuracy measurements for images in Fig. 3.3

3.4   Motivation for simultaneous Shadow/Vegetation Detection

Current shadow/vegetation detection methods in the literature detect separately shadow regions and vegetation regions [Shorter 2009], [Ozgun Ok 2013]. In this section, we first demonstrate that sometimes these methods might not provide a sufficiently good segmentation map (shadow, vegetation, other). Then, we propose to use also the luminance L for shadow/vegetation detection task.

(a) Input image

(c) Shadow detection by [Tian 2012]'s method

(e) Shadow detection using c_3 and Otsu's method

(b) Manual segmentation

(d) Vegetation detection by [Shorter 2009]'s method

(f) Vegetation detection using ExG and Otsu's method
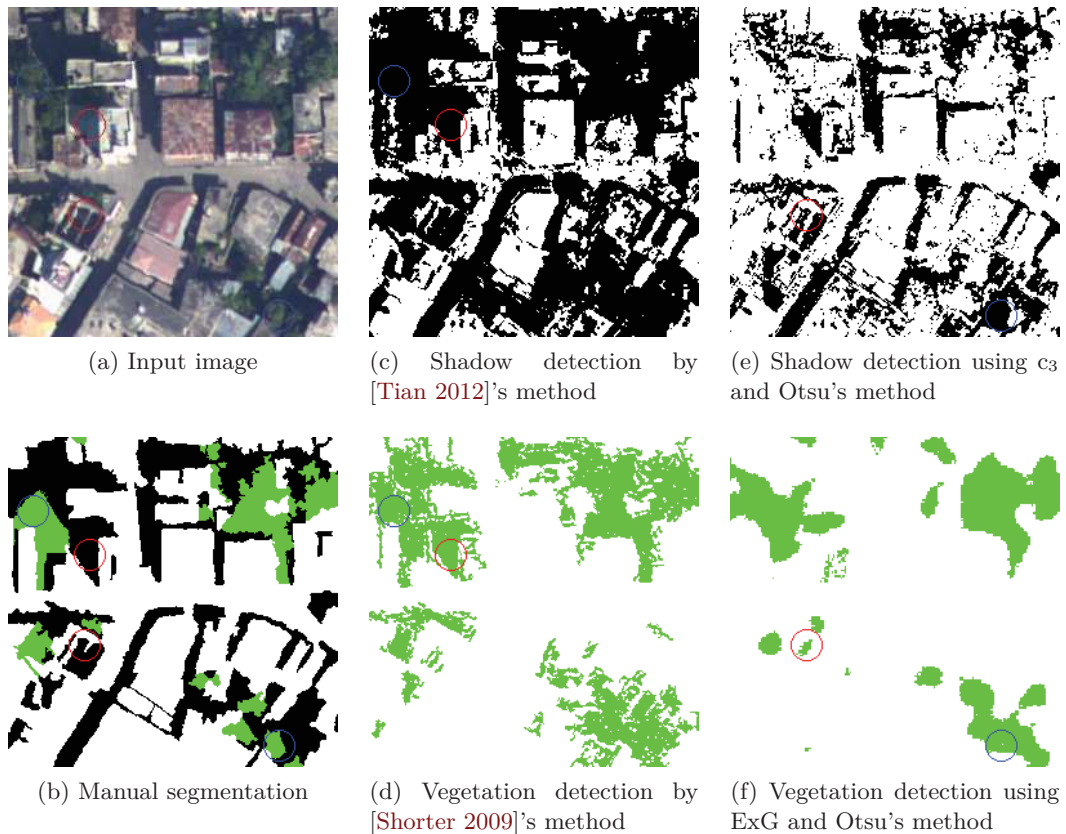
Figure 3.4: Sequential shadow/vegetation detection from a Jacmel aerial image

As shown in Figs. 3.4 and 3.5, a shadow detection algorithm (method of [Tian 2012]) and a vegetation detection algorithm (method of [Shorter 2009]) are applied on a test image. In these figures, the black areas correspond to shadows, the green areas correspond to vegetation and the white areas correspond to other. The methods of [Tian 2012] and [Shorter 2009] are detailed in Chapter 4 (Section 4.6). We observe that the method of [Tian 2012] detect quite well shadow regions (Figs. 3.4c and 3.5c). But it classifies some vegetation regions as shadows, for example, the region marked by blue circle. This region is correctly classified as vegetation region by the method of [Shorter 2009] (Figs. 3.4d and 3.5d). Again, the region marked by red circle is wrongly classified as vegetation region by [Shorter 2009], but [Tian 2012] detects correctly this region as shadow region. This observation is also obtained when we use index c_3 combined with Otsu's thresholding method for shadow detection (Figs. 3.4e and 3.5e) and index ExG combined with Otsu's thresholding method for vegetation detection (Figs. 3.4f and 3.5f).

(a) Input image

(c) Shadow detection by [Tian 2012]'s method

(e) Shadow detection using c_3 and Otsu's method

(b) Manual segmentation

(d) Vegetation detection by [Shorter 2009]'s method

(f) Vegetation detection using ExG and Otsu's method
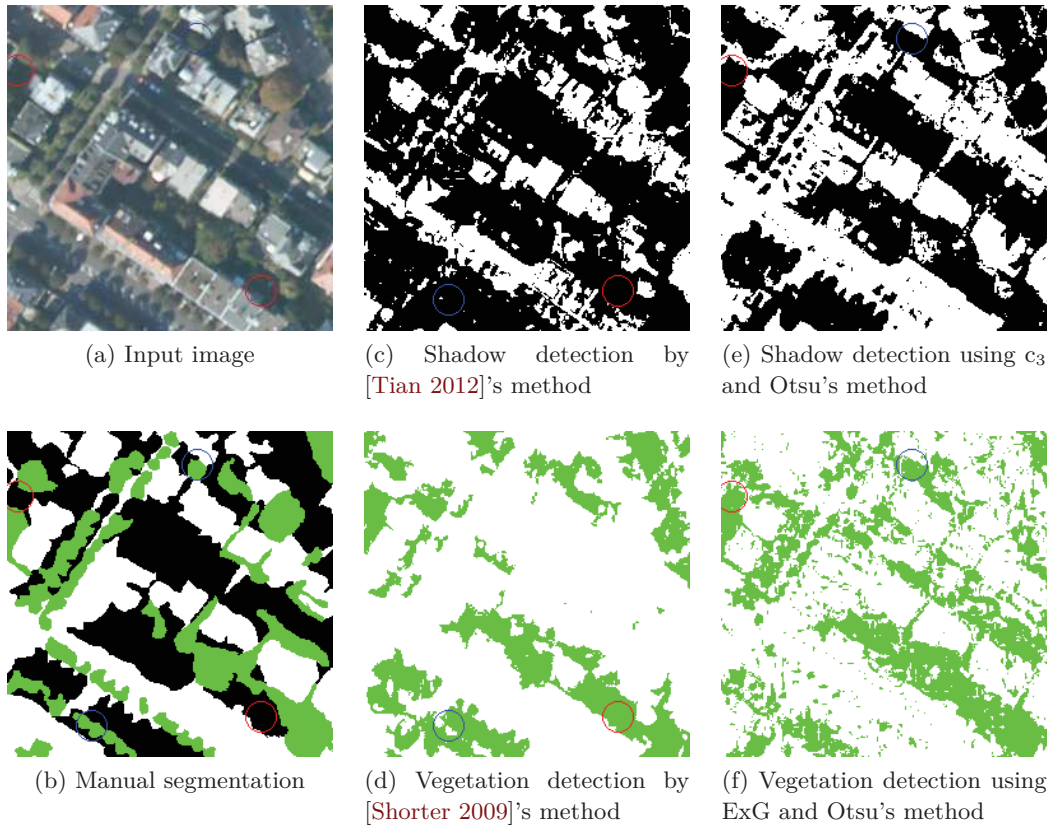
Figure 3.5: Sequential shadow/vegetation detection from a BD ORTHO$^®$ image.

In fact, the drawback of the sequential shadow/vegetation detection methods is that, for example, a vegetated pixel covered by shadow can be classified as vegetation (by a vegetation detection algorithm), and at the same time as shadow (by a shadow detection algorithm). So, depending on *how dark* these regions are, they will be classified as shadow or vegetation. In fact, visual inspection also has a similar problem. Because shadow and vegetation have the equally important effect on building classification task (see Chapter 1), we are motivated in developing a simultaneous shadow/vegetation detection algorithm, that allows to obtain a sufficiently good segmentation map with three classes: "shadow", "vegetation" and "other".

In the previous sections, we investigate different shadow indices (and vegetation indices) in order to choose the best index for shadow detection task (and vegetation detection task). Eventually, shadow index c_3 and vegetation index ExG (or NDVI in case the NIR band is available) are chosen for their best performance (visual and quantitative evaluation) comparing to other indices. However, one of the problems when using index c_3 is its instability for certain color values that lead to the misclassification of non-shadow pixels as shadow (false positives). As reported in [Gevers 1999, Salvador 2004], this occurs for pixels with extreme intensity values (as shown in Fig. 3.6, red circle). On the other hand, using only vegetation index ExG leads to classify all regions that are green and bright as vegetation (as shown in Fig. 3.7, blue circle). To solve these problems, we employ the luminance L, from the color space HSL, as proposed in [Yao 2006]. Shadow and vegetation regions are considered as having lower luminance than other regions [M.Polidorio 2003].

(a) RGB original image   (b) Manual segmentation   (c) c_3 and Otsu's method   (d) L and Otsu's method
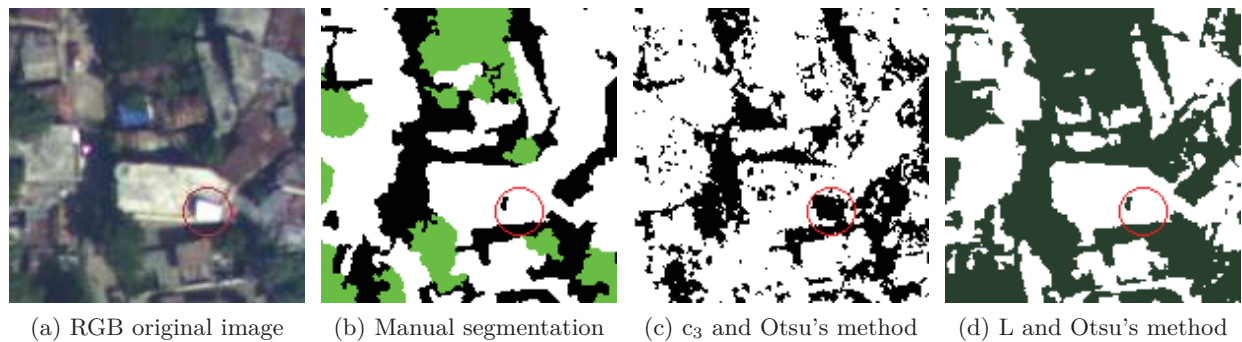
Figure 3.6: Otsu's thresholding method applied on image c_3 and luminance L: (c) using only index c_3 leads to classify a very bright object as shadow; (d) shadow regions and vegetation regions (dark green area) are well detected.

(a) RGB original image   (b) Manual segmentation   (c) ExG and Otsu's method   (d) L and Otsu's method
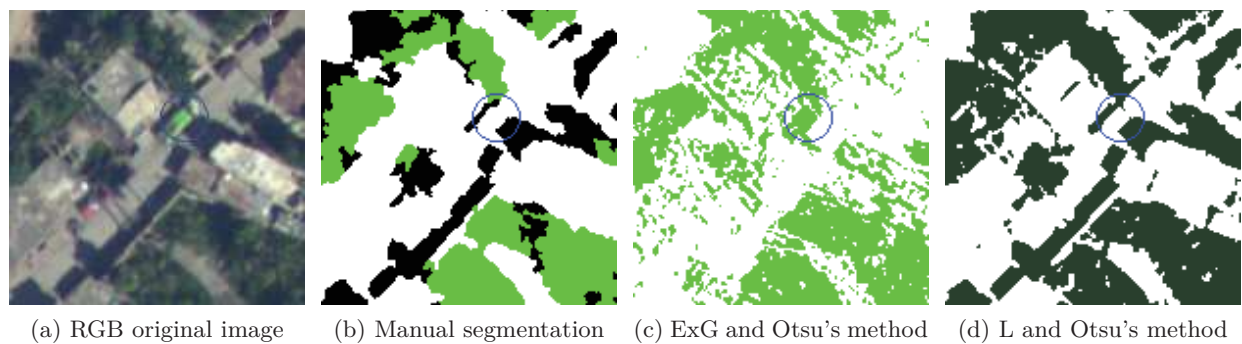
Figure 3.7: Otsu's thresholding method applied on image ExG and luminance L: (c) using only index ExG leads to classify a green and bright object as vegetation; (d) shadow regions and vegetation regions (dark green area) are well detected.

(a) a NOAA aerial image   (b) Shadow and vegetation   (c) a Worldview-2 image   (d) Shadow and vegetation
                            (in dark green color)   (in dark green color)
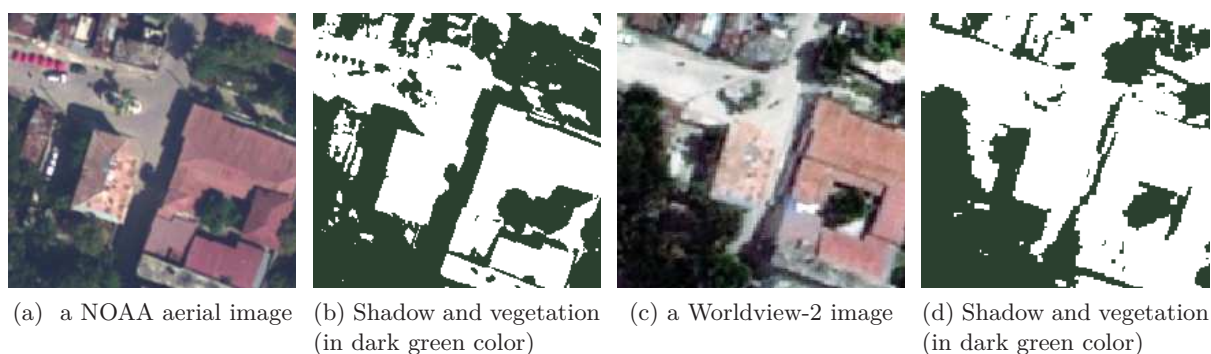
Figure 3.8: Illustration of shadow/vegetation detection using the luminance with two images from Jacmel dataset

Similar to the previous section, we test this index by first computing the luminance image and then applying the Otsu method to determine the appropriate threshold. As shown in Figs. 3.6, 3.7 and 3.8, index L allows to detect well shadow regions and vegetation regions (dark green area).

3.5 Conclusions

In this chapter, after reviewing the existing methods in shadow detection and vegetation detection from the remote sensing images in the literature, we test some most used shadow indices and vegetation indices on our image database. Index c_3 (resp. ExG (or NDVI)) is demonstrated as the most appropriate for shadow (resp. vegetation) detection task. Moreover, the false alarm due to the instability of these indices can be avoided by using the luminance L. Besides, we demonstrated that sometimes detecting separately shadows and vegetation might not provide a sufficient good result. Starting from these observation, a novel simultaneous shadow/vegetation detection is presented and detailed in the next chapter.

MRF and Dempster-Shafer Theory for simultaneous Shadow/Vegetation Detection

Contents

4.1 Introduction

In this chapter, we present our method for *simultaneously* detecting shadows and vegetation in remote sensing images. The main framework of the proposed method is summarized in Fig. 4.1. This method is based on Otsu's thresholding method and Dempster-Shafer (DS) fusion which aims at combining different indices: c_3, ExG, and luminance L; which are investigated in the previous chapter. Note that ExG is replaced with NDVI when the NIR band is available. However, for clarity of the proposed approach, we only consider the ExG in the following. The DS fusion processes at a pixel-level, assumes that each pixel in the image is independent of its neighbors and does not take into account spatial dependencies. Thus, the performance of the DS fusion is highly sensitive to noise. To overcome this shortcoming, a Markov Random Field (MRF) model [Geman 1984] may be employed. This chapter will firstly presents the application of DS fusion in shadow/vegetation detection, and then the link between MRF and DS fusion. The comparison to some existing shadow detection methods and vegetation detection methods is also reported at the end of the chapter.
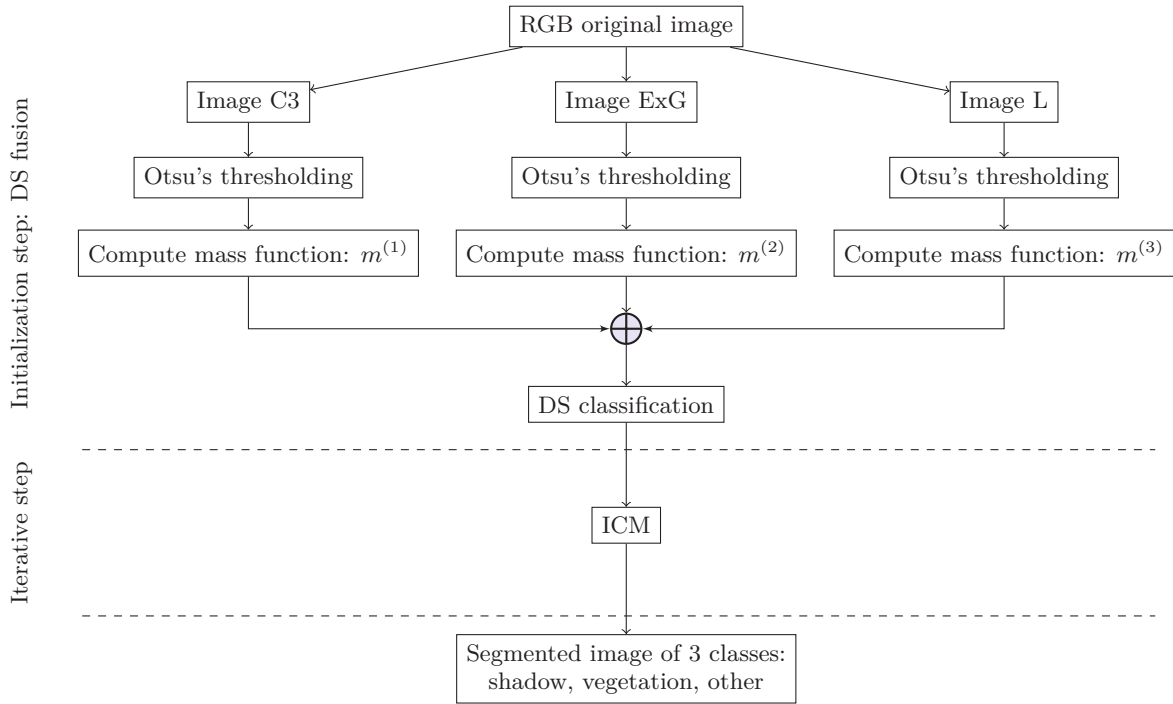
Figure 4.1: Global data fusion algorithm

4.2 Basic Concepts of Dempster-Shafer Evidence Theory

Dempster-Shafer (DS) evidence theory [Dempster 1968, Shafer 1976] makes inferences from incomplete and uncertain knowledge, provided by different independent knowledge sources. The theory allows strengthening or erosion of beliefs by combining additional sources of confidence, even in the presence of partly contradictory sensors. In DS evidence theory, there is a fixed set of N mutually exclusive and exhaustive elements, called the frame of discernment, which is symbolised by $\Theta = \{H_1, H_2, \ldots, H_N\} = \{H_i\}, 1 \leq i \leq N$.

The frame of discernment Θ defines the working space for the desired application since it consists of all propositions for which the information sources can provide confidence. Information sources can distribute mass values on subsets of the frame of discernment, $A \subset \Theta$. An information source assigns mass values only to those hypotheses, for which it has direct confidence. That is, if an information source can not distinguish between two propositions A and B, it assigns a mass value to the set including both propositions $(A \cup B)$. The mass distribution for all the hypotheses has to fulfill the following conditions:

$$
\begin{aligned}
&\forall A \subset \Theta, \quad 0 \leq m(A) \leq 1 \\
&\sum_{A \subset \Theta} m(A) = 1 \\
&m(\emptyset) = 0
\end{aligned}
\tag{4.1}
$$

If $m(A) > 0$, A is called a focal element.

The DS evidence theory provides a representation of both imprecision and uncertainty through

the definition of two functions: plausibility (Pls) and belief (Bel). They are defined as follows:

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B) \tag{4.2}$$

$$\text{Pls}(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{4.3}$$

In the Bayesian theory, uncertainty about an event is measured by a single value (probability) and imprecision about uncertainty measurement is assumed to be null. But in the DS theory, the belief value of hypothesis may be interpreted as the minimum uncertainty value about A, and its plausibility may be interpreted as the maximum uncertainty value of A. In other words, uncertainty about A is represented by the values of the interval $[\text{Bel}(A), \text{Pls}(A)]$, so-called "belief interval". The length of this interval measures the imprecision about the uncertainty value.

Another great advantage of DS theory is its robustness of combining information coming from various sources with the DS orthogonal sum. For $m^{(j)}$ being the mass function associated with source j, $(j = 1, 2, \ldots, n)$, this rule is written, for all non-empty subset A of Θ:

$$(m^{(1)} \oplus m^{(2)} \oplus \ldots \oplus m^{(n)})(A) = \frac{1}{1 - \mathcal{K}} \sum_{B_1 \cap \ldots \cap B_n = A} m^{(1)}(B_1) m^{(2)}(B_2) \ldots m^{(n)}(B_n)$$
$$(m^{(1)} \oplus m^{(2)} \oplus \ldots \oplus m^{(n)})(\emptyset) = 0 \tag{4.4}$$

where

$$0 \leq \mathcal{K} = \sum_{B_1 \cap \ldots \cap B_n = \emptyset} m^{(1)}(B_1) m^{(2)}(B_2) \ldots m^{(n)}(B_n) < 1 \tag{4.5}$$

To some extent, $\mathcal{K}$ can be interpreted as a measure of conflict between the sources and is directly taken into account in the combination as a normalization factor. It represents the mass which would be assigned to the empty set if masses were not normalized. The larger $\mathcal{K}$ is, the more the sources are conflicting and the less sense has their combination. Finally, the orthogonal sum does not exist when $\mathcal{K}$ is equal to 1. In this case, the sources are said to be totally or flatly contradictory, and it is no longer possible to combine them.

4.3 Use of DS Evidence Theory for Shadow/Vegetation Detection

4.3.1 DS Fusion Framework

Given the (R,G,B) color representation of a pixel, we use c$_3$ to build $Y^{(1)}$[1] , ExG to build $Y^{(2)}$, L to build $Y^{(3)}$. Let us denote ω_1, ω_2, ω_3, three clusters representing respectively "shadow", "vegetation" and "other". Our algorithm takes three feature images $Y^{(1)}, Y^{(2)}, Y^{(3)}$ as the input, and the output is a segmented image $\mathbf{x}$. The value of $\mathbf{x}$ at pixel s, denoted as x_s, is in $\Omega = \{\omega_1, \omega_2, \omega_3\}$.

---

[1]We use the uppercase letters (j) for all notations related to image $Y^{(j)}$, $j \in \{1, 2, 3\}$

Each feature image $Y^{(j)}$, $j \in \{1, 2, 3\}$ can be considered as an information source. In this context, source $Y^{(1)}$ is able to distinguish ω_1 from the two other classes but not ω_2 from ω_3. Source $Y^{(2)}$ is able to distinguish ω_2 from the two other classes but not ω_1 from ω_3. Similarly, source $Y^{(3)}$ is able to distinguish ω_3 from the two other classes but not ω_1 from ω_2. As illustrated in Fig. 4.2, we apply the Otsu's thresholding method over the images $Y^{(1)}, Y^{(2)}, Y^{(3)}$ to extract respectively shadow regions, vegetation regions and dark regions. After thresholding, image $Y^{(1)}$ (resp. $Y^{(2)}$; $Y^{(3)}$) is segmented into two classes ω_1 and $\{\omega_2, \omega_3\}$ (resp. ω_2 and $\{\omega_1, \omega_3\}$; $\{\omega_1, \omega_2\}$ and ω_3). Only the fusion of these three sources makes it possible to distinguish these three classes.

	Image $Y^{(1)}$: c$_3$	$\omega_1, \{\omega_2, \omega_3\}$
	Image $Y^{(2)}$: ExG	$\omega_2, \{\omega_1, \omega_3\}$
	Image $Y^{(3)}$: L	$\{\omega_1, \omega_2\}, \omega_3$
RGB image	Feature image	2-class segmentation map
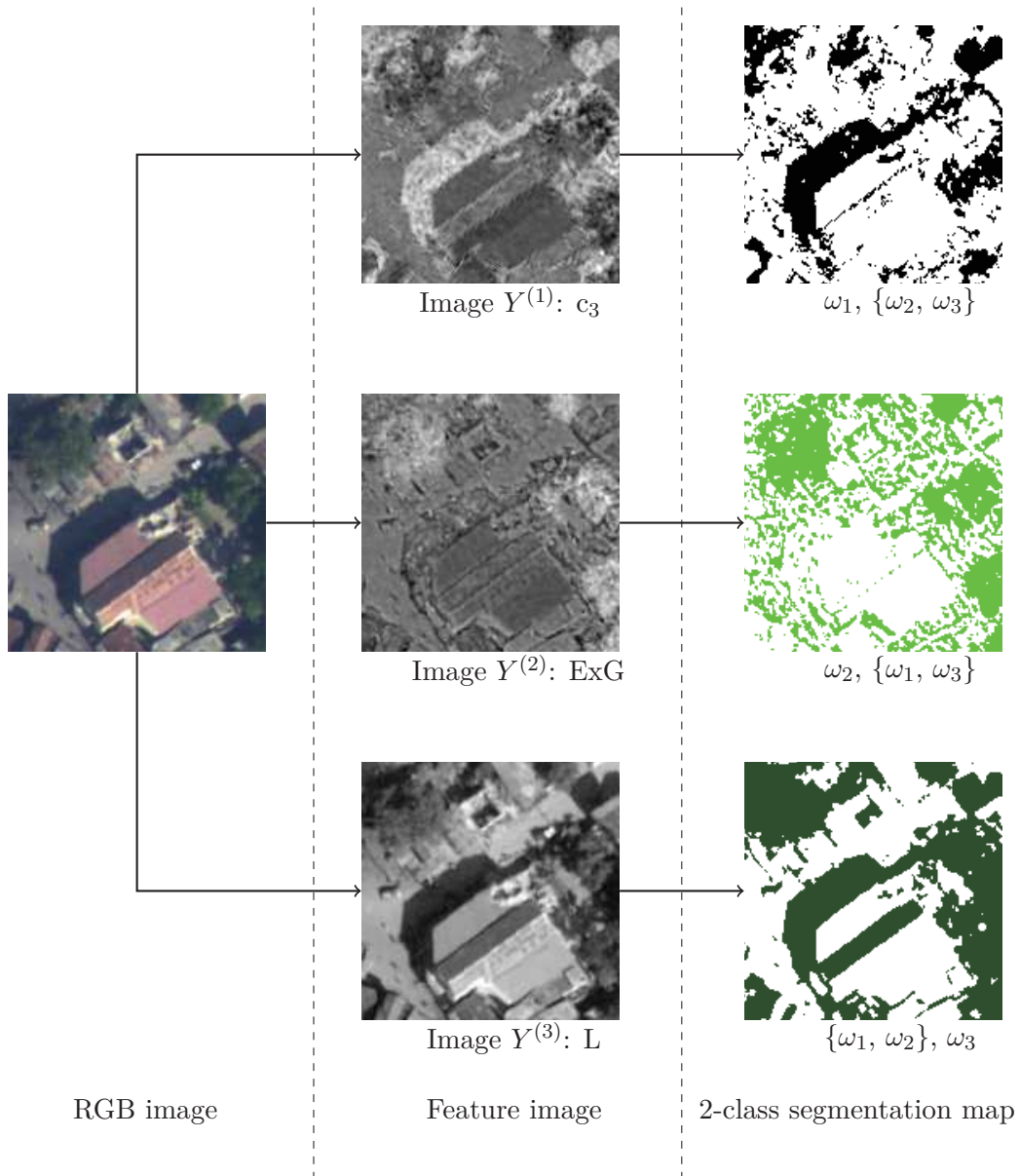
Figure 4.2: 2-class segmentation of each feature using Otsu thresholding method [Otsu 1975].

.

The three main numerical approaches to data fusion are the probabilistic method, fuzzy set

theory, and DS evidence theory [Le Hegarat-Mascle 1997]. The fuzzy set theory provides a lot of combination operators, which allows the user to adapt the fusion scheme to the specificity of the data at hand but these operators are always selected in a supervised way. For Bayesian inference, the main limitation is that it fails to model imprecision about uncertainty measurement. An event is said to be uncertain if its probability is not equal to one or zero. However, there may be an imprecision on probability measurement. DS evidence theory allows to deal with ignorance and missing information by affecting a degree of confidence which is called a mass function to all simple and compound hypotheses. Another major advantage of DS evidence theory is that it can deal with any union of classes, in our context, $\{\omega_2, \omega_3\}$ (or $\{\omega_1, \omega_3\}$ or $\{\omega_1, \omega_2\}$). Because of these advantages, we choose the DS evidence theory for the data fusion task.

The use of DS evidence theory for shadow/vegetation detection is detailed in the following. In our application context, the frame of discernment Θ is the set of hypotheses about pixel class: $\Theta = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$. In DS theory, not only single classes (also called singletons), but also any union of classes can be represented. In the following, hypotheses about singletons and hypotheses about union of classes are respectively called simple hypothesis ($H_i = \{\omega_i\}, i \in \{1, 2, 3\}$) and compound hypotheses.

4.3.2 Mass Function Determination

The derivation of the mass function is the most crucial step since it represents the knowledge about the actual application as well as the uncertainty incorporated in the selected information source. However, there are no generic methods to determine the mass functions. Some existing methods can be cited here: method derived from probabilities [Le Hegarat-Mascle 1997, Vannoorenberghe 1999, Ben Chaabane 2008], fuzzy logic [Zhu 2002, Chaabane 2009], neural network classifier [Denoeux 2000], ...

In [Vannoorenberghe 1999], authors have shown through empirical studies that a good model of the mass functions is based on the assumption of Gaussian distributions and histogram thresholding. This method is used in our model to determine the mass function of each feature image in Diagram 4.2. The estimation of mass functions is detailed as follows. As shown in Fig. 4.2, the Otsu's thresholding method first partitions the feature image into two regions, denoted as A_1 and A_2. For image $Y^{(1)}$, $A_1 = \{\omega_1\}, A_2 = \{\{\omega_2\}, \{\omega_3\}\}$. The mass function $m_s^{(1)}$ for each pixel $s \in \mathcal{S}$ of pixels, defining on $\{\emptyset, A_1, A_2, \Theta\}$, is then estimated as follows:

$$m_s^{(1)}(\emptyset) = 0$$

$$m_s^{(1)}(A_i) = \frac{1}{\sigma_i^{(1)}\sqrt{2\pi}} \exp\left(-\frac{(y_s^{(1)} - \mu_i^{(1)})^2}{2\sigma_i^{(1)^2}}\right), \quad i \in \{1, 2\}$$

$$m_s^{(1)}(\Theta) = \frac{1}{\sigma_\Theta^{(1)}\sqrt{2\pi}} \exp\left(-\frac{(y_s^{(1)} - \mu_\Theta^{(1)})^2}{2\sigma_\Theta^{(1)^2}}\right)$$

$$(4.6)$$

These mass functions are then normalized so that their sum is equal to 1. Here, $y_s^{(1)}$ is the value of the considered pixel s (of image $Y^{(1)}$). $\mu_i^{(1)}$ (resp. $\sigma_i^{(1)^2}$) represents the mean (resp. the variance)

of pixels with hypothesis A_i present in $Y^{(1)}$. They are respectively estimated as follows:

$$\mu_i^{(1)} = \frac{1}{|A_i|} \sum_{s \in A_i} y_s^{(1)}, \quad i \in \{1, 2\}$$

$$\sigma_i^{(1)} = \sqrt{\frac{1}{|A_i| - 1} \sum_{s \in A_i} (y_s^{(1)} - \mu_i^{(1)})^2}, \quad i \in \{1, 2\}$$

$$\mu_\Theta^{(1)} = \frac{\mu_1^{(1)} + \mu_2^{(1)}}{2}$$

$$\sigma_\Theta^{(1)} = \max(\sigma_1^{(1)}, \sigma_2^{(1)})$$

(4.7)

where $|A_i|$ denotes the number of pixels verifying A_i.

The mass functions $m_s^{(2)}$ for each pixel s of image $Y^{(2)}$, defining on $\{\emptyset, A_1, A_2, \Theta\}$ where $A_1 = \{\omega_2\}$ and $A_2 = \{\{\omega_1\}, \{\omega_3\}\}$; and the mass functions $m_s^{(3)}$ for each pixel s of image $Y^{(3)}$, defining on $\{\emptyset, A_1, A_2, \Theta\}$ where $A_1 = \{\{\omega_1\}, \{\omega_2\}\}$ and $A_2 = \{\omega_3\}$; are estimated in the same way. Once the mass function of the three images are estimated, their combination is performed using the DS orthogonal sum. However, the DS combination assumes the independence of combined sources. In our context, the feature images are computed from the R, G ,B bands. So, they are dependent sources. In [Denœux 2006, Denœux 2008], Denoeux introduced the cautious rule of combination (denoted by Ⓐ) to combine dependent data. This rule has the advantage of avoiding double-counting of common evidence when combining non distinct mass functions. Diagram of data fusion for shadow/vegetation detection is illustrated in Fig. 4.3.

$$m_s^{(1)}(\{\omega_1\}), m_s^{(1)}(\{\{\omega_2\}, \{\omega_3\}\}), m_s^{(1)}(\Theta) \;—\;$$
$$m_s^{(2)}(\{\omega_2\}), m_s^{(2)}(\{\{\omega_1\}, \{\omega_3\}\}), m_s^{(2)}(\Theta) \;—\!\oplus\!\longrightarrow\; m_s(A) = m_s^{(1)} \otimes m_s^{(2)} \otimes m_s^{(3)}$$
$$s \in \mathcal{S}, A \subset \Theta$$
$$m_s^{(3)}(\{\omega_3\}), m_s^{(3)}(\{\{\omega_1\}, \{\omega_2\}\}), m_s^{(3)}(\Theta) \;—\;$$
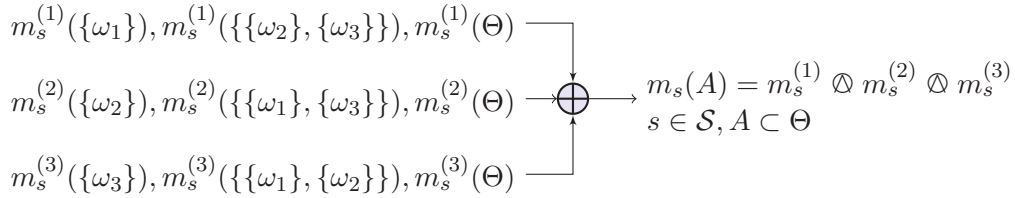
Figure 4.3: Data fusion diagram.

From the mass functions m_s, for all $A \subset \Theta$, two functions can be derived as follows [Dempster 1968]:

$$\text{Bel}_s(A) = \sum_{B \subseteq A} m_s(B)$$

(4.8)

$$\text{Pls}_s(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

(4.9)

(4.10)

4.3.3 Decision Making

Having computed the mass, plausibility, belief values for each simple and compound hypothesis, we need a criterion, which is called "decision rule", to decide which hypothesis is the more "realistic". Nowadays, the choice of this criterion remains application dependent. Several decision rules have

been proposed [Le Hegarat-Mascle 1997, Denoeux 1997]. For the cautious rule of combination, the decision making is carried out at the pignistic level, that maps the mass functions to probability measures:

$$Betp(H_i) = \sum_{\{A:H_i \in A\}} \frac{m(A)}{|A|} \tag{4.11}$$

where $|A|$ denotes the number of elements in A. And the decisional procedure is to choose the simple hypothesis that maximizes the pignistic probability. An example is shown in Fig. 4.4.

(a)   (b)   (c)
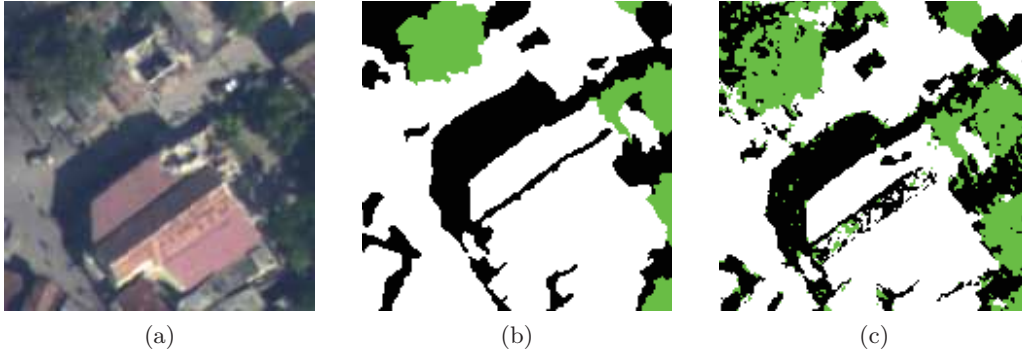
Figure 4.4: Decision making: (a) input image, (b) ground truth, (c) detected shadow/vegetation by maximizing the pignistic probability function.

4.4 DS Theory in Markovian Context

4.4.1 Markov Random Field Modeling

The DS fusion processes at a pixel-level, assumes that each pixel in the image is independent of its neighbors and does not take into account spatial dependencies. To overcome this shortcoming, a Markov Random Field (MRF) model [Geman 1984] may be employed, to consider not only the measurements at the pixel location, but also the class values among its closest neighbors. Let $\mathbf{X}$ be a random field over the set $\mathcal{S}$ of pixels, taking its values from a finite set of classes $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $\mathbf{x}$ a realization of $\mathbf{X}$. x_s denotes the value of $\mathbf{x}$ at the pixel $s \in \mathcal{S}$. x_s is a realization of the random variable X_s. $\mathbf{x}$ is estimated using the observed images $\mathbf{y}$. y_s denotes the value of $\mathbf{y}$ at the pixel s. $\mathbf{y}$ (resp. $\mathbf{y}_s$) is a realization of the observation field $\mathbf{Y}$ (resp. $\mathbf{Y}_s$). In our shadow/vegetation framework, $\mathbf{y}_s$ is a 3-dimensional feature vector: $\mathbf{y}_s = \{y_s^{(1)}, y_s^{(2)}, y_s^{(3)}\}$, where $y_s^{(j)}$ ($j \in \{1, 2, 3\}$) is the pixel value of the feature image $Y^{(j)}$.

In MRF model, the segmented image $\mathbf{x}$ can be obtained using the MAP estimate of $\mathbf{X}$. Thus,

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\max_{\mathbf{x}} P(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) \tag{4.12}$$

This problem can be solved by the iterated conditional modes (ICM) algorithm [Besag 1986]. In this method, a raster scan is used to iteratively visit all the pixels in field $\mathbf{X}$. We denote $\hat{\mathbf{x}}_{\mathcal{S}-\{s\}}$ as a provisional estimate of the segmentation field everywhere except at pixel s. Given the data $\mathbf{y}$ and the current segmentation $\hat{\mathbf{x}}_{\mathcal{S}-\{s\}}$, the algorithm updates the current segmentation $\hat{x}_s$ at pixel s, and replaces it with the new value $\hat{x}_s$ which maximizes $P(X_s = x_s | \mathbf{Y}_s = \mathbf{y}_s, X_{\mathcal{S}-\{s\}} = \hat{x}_{\mathcal{S}-\{s\}})$

with respect to x_s.

Since we consider $\mathbf{X}$ as an MRF, the latter follows from Bayes' theorem that

$$P(X_s = x_s | \mathbf{Y}_s = \mathbf{y}_s, X_{\mathcal{S}-\{s\}} = \hat{x}_{\mathcal{S}-\{s\}}) \propto P(\mathbf{Y}_s = \mathbf{y}_s | X_s = x_s) P(X_s = x_s | X_{\mathcal{S}-\{s\}} = \hat{x}_{\mathcal{S}-\{s\}}) \tag{4.13}$$

We adopt a second-order (eight neighbors) neighborhood system. Let $X_{\mathcal{V}_s}$ represent the set of labels assigned to the neighbors of s and $x_{\mathcal{V}_s}$ its realization. $\hat{x}_{\mathcal{V}_s}$ is the proximal estimate of $x_{\mathcal{V}_s}$. The prior probability term in Eq. 4.13 can be written:

$$P(X_s = x_s | X_{\mathcal{S}-\{s\}} = \hat{x}_{\mathcal{S}-\{s\}}) = P(X_s = x_s | X_{\mathcal{V}_s} = \hat{x}_{\mathcal{V}_s}) = \frac{\exp\left(-\beta \sum_{l\in\mathcal{V}_s} \{1 - \delta(x_s, \hat{x}_l)\}\right)}{\sum_{\omega_k\in\Omega} \exp\left(-\beta \sum_{l\in\mathcal{V}_s} \{1 - \delta(\omega_k, \hat{x}_l)\}\right)} \tag{4.14}$$

where $\delta(\cdot)$ stands for the Kronecker's delta function:

$$\delta(x_s, x_l) = \begin{cases} 1 & \text{si } x_s = x_l \\ 0 & \text{otherwise.} \end{cases} \tag{4.15}$$

and β is a regularized parameter, which represents the trade-off between fidelity to the observed image and the smoothness of the segmented image.

Estimation of β: The MRF parameter estimation method described in this section has been proposed by [Derin 1987]. But, unlike the MRF model in the paper of [Derin 1987], we restrict our attention to MRF's whose clique potentials involve pairs of neighboring nodes and we do not distinguish the different types of cliques (horizontal, vertical, diagonal), thus simplify the estimator. Referring to the prior term in Eq. 4.14, we can write:

$$P(X_s = x_s | X_{\mathcal{V}_s} = x_{\mathcal{V}_s}) \propto \exp\left(-\beta \sum_{l\in\mathcal{V}_s} \{1 - \delta(x_s, x_l)\}\right) \tag{4.16}$$

For any two distinct values of x_s, e.g., $x_s = \omega_1$ and $x_s = \omega_2$, with the same configuration of neighbors $x_{\mathcal{V}_s}$, we have:

$$\exp\left[-\beta\left(\sum_{l\in\mathcal{V}_s} \{1 - \delta(\omega_1, x_l)\} - \sum_{l\in\mathcal{V}_s} \{1 - \delta(\omega_2, x_l)\}\right)\right] = \frac{P(X_s = \omega_1 | X_{\mathcal{V}_s} = x_{\mathcal{V}_s})}{P(X_s = \omega_2 | X_{\mathcal{V}_s} = x_{\mathcal{V}_s})} \tag{4.17}$$

Taking the natural logarithm of Eq. 4.17, we obtain:

$$\left(\sum_{l\in\mathcal{V}_s} \{1 - \delta(\omega_2, x_l)\} - \sum_{l\in\mathcal{V}_s} \{1 - \delta(\omega_1, x_l)\}\right)\beta = \log\left(\frac{P(X_s = \omega_1 | X_{\mathcal{V}_s} = x_{\mathcal{V}_s})}{P(X_s = \omega_2 | X_{\mathcal{V}_s} = x_{\mathcal{V}_s})}\right) \tag{4.18}$$

Assuming the right-hand side of Eq. 4.18 is estimated, for each distinct pair (ω_1, ω_2) and all possible values of $x_{\mathcal{V}_s}$, Eq. 4.18 simply reduces to a set of linear equations. For all pixel s,

$P(X_s = \omega_i | X_{\mathcal{V}_s} = x_{\mathcal{V}_s})$ is estimated by use of histogram technique. Let us denote $N(x_s, x_{\mathcal{V}_s})$ is the number of appearances of a particular 3×3 configuration $(x_s, x_{\mathcal{V}_s})$. We have :

$$\frac{P(X_s = \omega_1 | X_{\mathcal{V}_s} = x_{\mathcal{V}_s})}{P(X_s = \omega_2 | X_{\mathcal{V}_s} = x_{\mathcal{V}_s})} = \frac{N(\omega_1, x_{\mathcal{V}_s})}{N(\omega_2, x_{\mathcal{V}_s})} \tag{4.19}$$

For a system of 8-nearest neighbors, there are 6561 ($= 3^8$) different configurations of $x_{\mathcal{V}_s}$, so we obtain 6561 equations. Applying 4.18 for all of the pixels in image, a linear equation can be obtained as:

$$\mathbf{A}\beta = b \tag{4.20}$$

The least square method allows to find β so that $\|\mathbf{A}\beta - b\|$ is minimum. And the result is:

$$\beta = (\mathbf{A}^\mathrm{T}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}b \tag{4.21}$$

where $\mathbf{A}^\mathrm{T}$ is the transposed matrix of $\mathbf{A}$.

4.4.2 Incorporation of MRF and DS evidence theory

In a Bayesian context, the interest of MRF which allows one to take contextual information into account, has been well known for more than 30 years. In recent years, both the MRF and DS evidence theory have been incorporated into more general methods for scene classification, often involving multisource analysis [Bendjebbour 2001, Foucher 2002, Bentabet 2008]. In fact, a major difference between the probabilistic framework and the DS evidence theory is the ability of the latter to quantify the imprecision through the assignment of confidence levels (by mass function value) to the union of hypotheses. The probabilistic framework assigns likelihood values $p(\mathbf{y}_s | x_s)$ only for each class $\omega_i \in \Omega$. From the view of evidence theory, the likelihood can be assimilated to a mass function which charges the singleton $H_{x_s} = \{\omega_i\}$. In [Bentabet 2008], authors propose to extend the likelihood term $p(\mathbf{y}_s | x_s)$ by defining mass values $m_s(A)$ for every $A \subset \Theta$. So, the new likelihood term is:

$$m_s(A) = m_s^{(1)} \quad \oslash \quad m_s^{(2)} \quad \oslash \quad m_s^{(3)} \tag{4.22}$$

for all $A \subset \Theta$. These mass functions are computed in the previous section (diagram 4.3).

Besides, the prior probability term $P(X_s = x_s | X_{\mathcal{V}_s} = \hat{x}_{\mathcal{V}_s})$, defined in Eq. 4.14, which assigns values only to single hypothesis, can be generalized to deal with the compound hypotheses:

$$m_s(A|\hat{x}_{\mathcal{V}_s}) = \frac{\exp\left(-\sum_{\{\omega_k\} \cap A \neq \emptyset} \beta \sum_{l \in \mathcal{V}_s} \{1 - \delta(\omega_k, \hat{x}_l)\}\right)}{\sum_{B \subset \Theta} \exp\left(-\sum_{\{\omega_k\} \cap B \neq \emptyset} \beta \sum_{l \in \mathcal{V}_s} \{1 - \delta(\omega_k, \hat{x}_l)\}\right)} \tag{4.23}$$

for all $A \subset \Theta$. Finally, considering the likelihood and the conditional membership components in Eqs. 4.22 and 4.23 as two "evidential" sources of information, the DS orthogonal sum is used to combine these two sources of information. The posterior probability in Eq. 4.13 can be replaced by the orthogonal sum $M_s(.)$, which carries the joint information:

$$M_s(A) = \frac{1}{1 - \mathcal{K}} \sum_{A_p \cap A_q = A} m_s(A_p) m_s(A_q | \hat{x}_{\mathcal{V}_s}) \tag{4.24}$$

where $\mathcal{K} = \sum_{A_p \cap A_q = \emptyset} m_s(A_p) m_s(A_q | \hat{x}_{\mathcal{V}_s})$.

Once all of the mass function of the single and compound hypotheses related to pixel s are determined, the estimation of $\hat{x}_s$ is obtained by maximizing the plausibility function, derived from the mass function M_s:

$$\hat{x}_s = \arg\max_{x_s} \mathrm{Pls}_s(H_{x_s}) = \arg\max_{x_s} \left(\sum_{A \cap H_{x_s} \neq \emptyset} M_s(A) \right)$$

	Probabilistic MRF, $x_s \in \Omega$	Evidential MRF, $A \subset \Theta$
Likelihood term	$p(\mathbf{y}_s \| x_s)$	$m_s(A)$
Prior term	$p(x_s \| \hat{x}_{\mathcal{V}_s})$	$m_s(A \| \hat{x}_{\mathcal{V}_s})$
Posterior term	$P(x_s \| \mathbf{y}, \hat{x}_{\mathcal{S}-\{s\}})$	$M_s(A) = m_s(A) \oplus m_s(A \| \hat{x}_{\mathcal{V}_s})$
Decision	$\hat{x}_s = \arg\max_{x_s} P(x_s \| \mathbf{y}, \hat{x}_{\mathcal{S}-\{s\}})$	$\hat{x}_s = \arg\max_{x_s} \left(\sum_{A \cap H_{x_s} \neq \emptyset} M_s(A) \right)$

Table 4.1: Extension of the classical ICM algorithm

Table 4.1 summarizes the extension of the classical ICM method to deal with the evidential case for every pixel $s \in \mathcal{S}$.

4.5   Overall Algorithm

The proposed iterative data fusion method is denoted as SSVD (simultaneous shadow/vegetation detection) method in the next. SSVD method takes as input the RGB image, and the output is a segmented image $\mathbf{x}$ with labels taking the value in $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The global scheme of SSVD method for shadow/vegetation detection is presented in Fig. 4.1 and detailed in Algorithm 1. The ICM labeling process is alternatively operated until the percentage of replacement between two consecutive iterations or the number of iterations reach preset values (ε and τ_{max} respectively). For all images in our database, the preset value is chosen as 0.2 % for the percentage of replacement and 100 for the maximum number of iterations.

An example of shadow/vegetation detection by DS fusion incorporated with MRF regularization is shown in Fig. 4.5. The quantitative results shown in Table 4.2 demonstrate the efficacy of MRF regularization. A more detailed evaluation of the proposed method is presented in the following section.
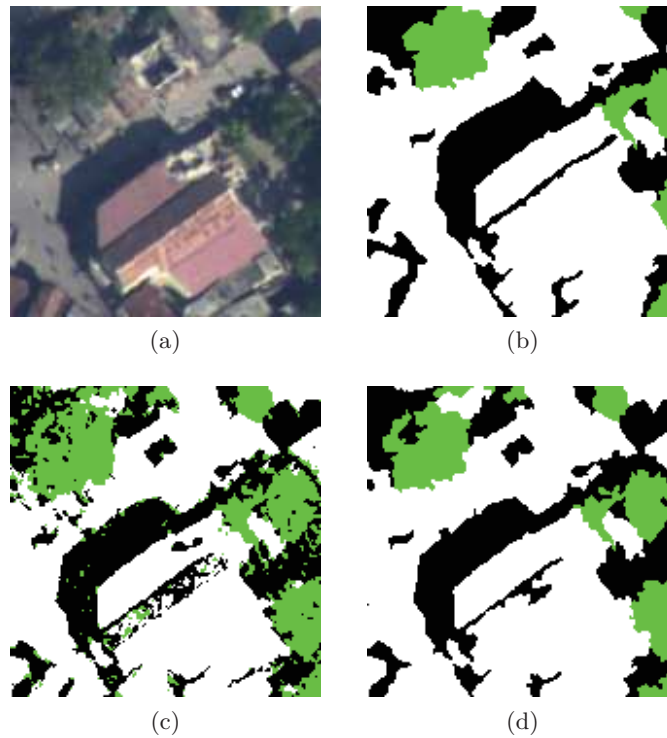
Figure 4.5: DS segmentation result with MRF regularization: (a) input image, (b) ground truth, (c) initialization (DS fusion), (d) after MRF regularization. We observe that the proposed method solves the problem of false dismissals, and improves the accuracy of segmentation. The MRF regularization removes the small objects in segmentation and preserves the shape of shadow region.

Method	Shadow detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
Without MRF	45.55	77.07	42.75	79.01	68.46	0.2219
With MRF	48.48	78.75	46.18	80.26	**70.48**	**0.2684**
Method	Vegetation detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
Without MRF	62.51	85.55	34.38	94.96	83.05	0.3755
With MRF	54.30	90.57	41.10	94.24	**86.65**	**0.3983**

Table 4.2: Shadow/vegetation detection accuracy measurements for images in Fig. 4.5. With MRF regularization, the performance of the method is improved (both overall accuracy and MCC are higher than the method without MRF).

Algorithm 1: SSVD method for simultaneous shadow/vegetation detection

---

/* ***Input*** */

- Input image.

- Maximum number of iterations τ_{max}.

- Residual error ε.

/* ***Initalization*** */

- Compute c_3, ExG, L which respectively correspond to $Y^{(1)}, Y^{(2)}, Y^{(3)}$.
- Apply Otsu thresholding method [Otsu 1975] for $Y^{(1)}, Y^{(2)}, Y^{(3)}$.
- For each $Y^{(j)}, j \in \{1, 2, 3\}$, compute $\mu_i^{(j)}, \sigma_i^{(j)}, i \in \{1, 2\}$ using Eq. 4.7.

for *each pixel $s \in \mathcal{S}$* **do**

> - For $j \in \{1, 2, 3\}$, and $i \in \{1, 2\}$, compute $m_s^{(j)}(A_i), m_s^{(j)}(\Theta)$ using Eq. 4.6.
> - Compute $m_s(A), \forall A \subset \Theta$, as shown in Fig. 4.3.
> - Decision making: $\hat{x}_s = \arg\max_{x_s} \left(\sum_{A \cap H_{x_s} \neq \emptyset} m_s(A) \right)$
>
> $\mathbf{x}^{[0]} \longleftarrow \hat{\mathbf{x}}$: initial configuration

/* ***ICM labeling process*** */

$k = 0$;   /* *number of iterations* */
repeat
> 1. Parameter estimation:
>
>   • Update $\mu_i^{(j)}, \sigma_i^{(j)}, \mu_\Theta^{(j)}, \sigma_\Theta^{(j)}$ for $j \in \{1, 2, 3\}, i \in \{1, 2\}$ from $Y^{(1)}, Y^{(2)}, Y^{(3)}, \mathbf{x}^{[k]}$ using Eq. 4.7.
>
>   • Update β using Eq. 4.21.
>
> 2. Compute $\mathbf{x}^{[k+1]}$ from $\mathbf{x}^{[k]}$:
> **for** *each pixel $s \in \mathcal{S}$* **do**
> > /* *We scan all sites s (by a zig-zag parcours)* */
> >
> > • For $j \in \{1, 2, 3\}$, and $i \in \{1, 2\}$, compute $m_s^{(j)}(A_i), m_s^{(j)}(\Theta)$ using Eq. 4.6.
> >
> > • Compute $m_s(A); \quad \forall A \subset \Theta$ (see Fig. 4.3).
> >
> > • Compute $m_s(A|\hat{x}_{\mathcal{V}_s}); \quad \forall A \subset \Theta$ using Eq. 4.23.
> >
> > • Compute $M_s(A); \quad \forall A \subset \Theta$ using Eq. 4.24.
> >
> > • Estimate: $\hat{x}_s = \arg\max_{x_s} \left(\sum_{A \cap H_{x_s} \neq \emptyset} M_s(A) \right)$
> >
> $\mathbf{x}^{[k+1]} \longleftarrow \hat{\mathbf{x}}$
> $k \longleftarrow k + 1$

until $(|\boldsymbol{x}^{[k]} - \boldsymbol{x}^{[k-1]}| \leq \varepsilon) \vee (k \geq \tau_{max})$;

/* ***Output*** */

- segmented image $\hat{\mathbf{x}} \longleftarrow \mathbf{x}^{[k]}$

4.6 Experiments

Experiments are performed using multiple datasets described in Chapter 2. The wide range of test images takes into account different factors including lighting conditions, concentration of vegetation, size of vegetation, shadow orientation and shape of the shadow projection. Different comparative experiment results for shadow/vegetation detection are shown, in which black pixels correspond to what the algorithms classify as shadow, the green pixels as vegetation, the white pixels as others. A quantitative evaluation is also presented. We categorize this section in function of spectral characteristics of test images: RGB image (without NIR band) and RGB-NIR image (with NIR band). The strategy for accuracy assessment is presented in Chapter 3.

4.6.1 State-of-the-art Methods for Further Comparison

4.6.1.1 Shadow Detection

As described in the review paper [Adeline 2013], the shadow chromaticity can be exploited by making use of RGB combinations, like our method with the use of c_3 feature, or working with shadow invariant images, like the tri-color attenuation (TAM) [Tian 2009]. TAM describes the attenuation relationship between shadows and their non-shadow backgrounds in the three color channels (R, G, B), and combines with intensity image (I from HSI color space) to detect shadows. In [Tian 2009], the authors first oversegment the image, and then apply the TAM model to decide whether a segment is in the shadow or not. The use of several thresholds, however, makes this approach fairly unstable. In [Tian 2012], they improve the TAM model, no longer use segmentation and only compute a single threshold, which yields much improved results.

In case the NIR band is available, we compare our SSVD method with the method of [Rufenacht 2014]. Starting from the observation that shadows are generally darker than their surroundings, in both the visible and the NIR, an the dark objects, which confound many shadow detection algorithms, often have much higher reflectance in the NIR, authors build an shadow candidate map based on image pixels that are dark both in the visible and NIR representations. The shadow map is further refined by incorporating ratios of the visible to the NIR image, based on the observation that commonly encountered light sources have very distinct spectra in the NIR band. It should be noted that both methods ([Tian 2012] and [Rufenacht 2014]) are designed to deal with shadow detection in different scenes (indoor, outdoor photographs, remote sensing), whereas SSVD method focus only on detecting shadows from remote sensing images.

Index c_3 is demonstrated among other shadow indices as the most appropriate for shadow detection task (see Chapter 3) and employed in our method together with index ExG and luminance L. It is successfully used by [Salvador 2001, Arévalo 2008, Yao 2006] to extract shadows from single RGB image. In this section, we use index c_3 to detect only shadow regions. For a fair comparison, we take into account spatial information by employing MRF model. The Otsu thresholding method is used to classify image into two classes: "shadow" and "non-shadow". The resulting classification map is then regularized by a MRF model, in which the likelihood term is modelled by a Gaussian distribution. For the prior term, we use the second-order neighborhood system. The Gaussian distribution is appropriate to model the likelihood since as observed in Fig. 3.2 (Chapter 3, histogram of radiance for different shadow indices), the realization of the observed field can be approximated by the mixture of two Gaussian distributions. The ICM

algorithm is used to optimize the MRF energy. The parameters are estimated at each iteration of ICM (μ and σ are easily estimated from the classification map and β is estimated using the method of [Derin 1987]). We denote this method as c_3-ICM in the rest of the section.

4.6.1.2   Vegetation Detection

The chosen vegetation detection method for comparison is the method of [Shorter 2009]. In their work, image is first segmented using a color quantization technique. Vegetation candidate pixels are identified by Otsu thresholding and a color invariant model. Then, if 60% of a given region, identified via the aforementioned color segmentation algorithm, contains vegetation candidate pixels, then that entire region is labeled as a vegetation region.

In case the NIR band is available, we can compute the NDVI, the most-used vegetation index in the literature. NDVI is used with the Otsu thresholding method for unsupervised vegetation detection [Ozgun Ok 2013]. In this section, similar to what is done with index c_3, we use the Otsu thresholding method over the NDVI feature image to classify image into two classes : "vegetation" and "non-vegetation". The resulting classification map is then regularized by MRF model (with the same assumption for likelihood and the prior probability). This method of vegetation detection is denoted as NDVI-ICM method. The NDVI index is replaced by the ExG index in case the NIR band is not available. This method is denoted as ExG-ICM for further comparison.

4.6.1.3   Shadow/vegetation detection

To the best of our knowledge, there exists some methods that focus on classifying the remote sensing image into different classes: shadow, vegetation, soil, water, street, etc [Xu 2005], but there is no other method that detect simultaneously shadows/vegetation. Through the histogram of radiance for index c_3 and ExG (chapter 3), we observed that if these indices are linearly stretched out into the interval $[0, 1]$, shadowed pixels have high c_3 value (near to 1) and low ExG value; vegetative pixels have high ExG value (near to 1) and low c_3 value. Therefore, a new index, called as SVI (shadow-vegetation index) is proposed: $SVI = c_3 - ExG$. Pixels having high SVI values (near to 1) correspond to shadow regions and pixels having low SVI values (near to -1) correspond to vegetation regions. The K-means clustering method is used to classify into three classes: "shadow", "vegetation" and "other". The resulting classification map is then regularized by a MRF model, in which the likelihood term is modelled by a Gaussian distribution and for the prior term, we use the second-order neighborhood system. The MRF model is used, in which the number of classes K is equal to 3. For the image in which the NIR band is available, EXG is replaced by NDVI. This new method of shadow/vegetation detection is denoted as SVI-ICM method for further comparison.

4.6.2   Experimental Results

4.6.2.1   Without NIR

Three examples of shadow/vegetation detection from RGB images are shown in Figs. 4.6, 4.7, 4.8. These images of size 256 x 256 are selected from three different image datasets: Jacmel, BD ORTHO® and SZTAKI-INRIA benchmark, that covers different scenes: Jacmel city, Strasbourg and Normandy. Each example is accompanied by a table of accuracy measurement (table 4.3,

(a) Input image   (b) Ground truth   (c) [Tian 2012]'s method   (d) [Shorter 2009]'s method

(e) c$_3$-ICM method   (f) ExG-ICM method   (g) SVI-ICM method   (h) SSVD method
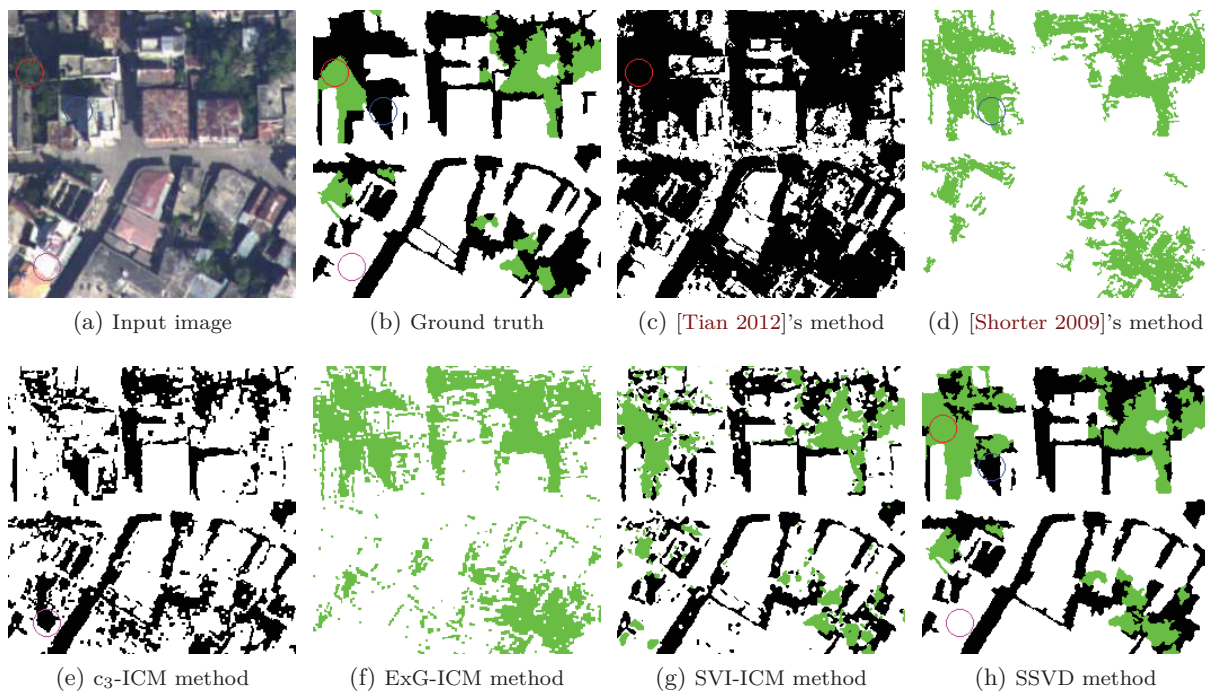
Figure 4.6: Shadow/vegetation detection from NOAA aerial image (Jacmel dataset)

Method	Shadow detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
[Tian 2012]	99.14	50.47	47.76	99.23	65.73	0.48
c$_3$-ICM	76.44	89.69	77.22	89.29	85.54	0.66
SVI-ICM	66.29	97.16	91.44	86.32	87.48	0.70
SSVD method	80.61	99.89	99.70	91.85	**93.84**	**0.85**
Method	Vegetation detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
[Shorter 2009]	86.71	79.17	31.66	98.16	79.93	0.44
ExG-ICM	82.35	74.49	26.43	97.43	75.27	0.36
SVI-ICM	78.39	95.18	64.45	97.53	93.50	0.67
SSVD method	99.25	93.72	63.94	99.94	**94.35**	**0.77**

Table 4.3: Shadow/vegetation detection accuracy measurements of Fig. 4.6.

(a) Input image   (b) Ground truth   (c) [Tian 2012]'s method   (d) [Shorter 2009]'s method

(e) c₃-ICM method   (f) ExG-ICM method   (g) SVI-ICM method   (h) SSVD method
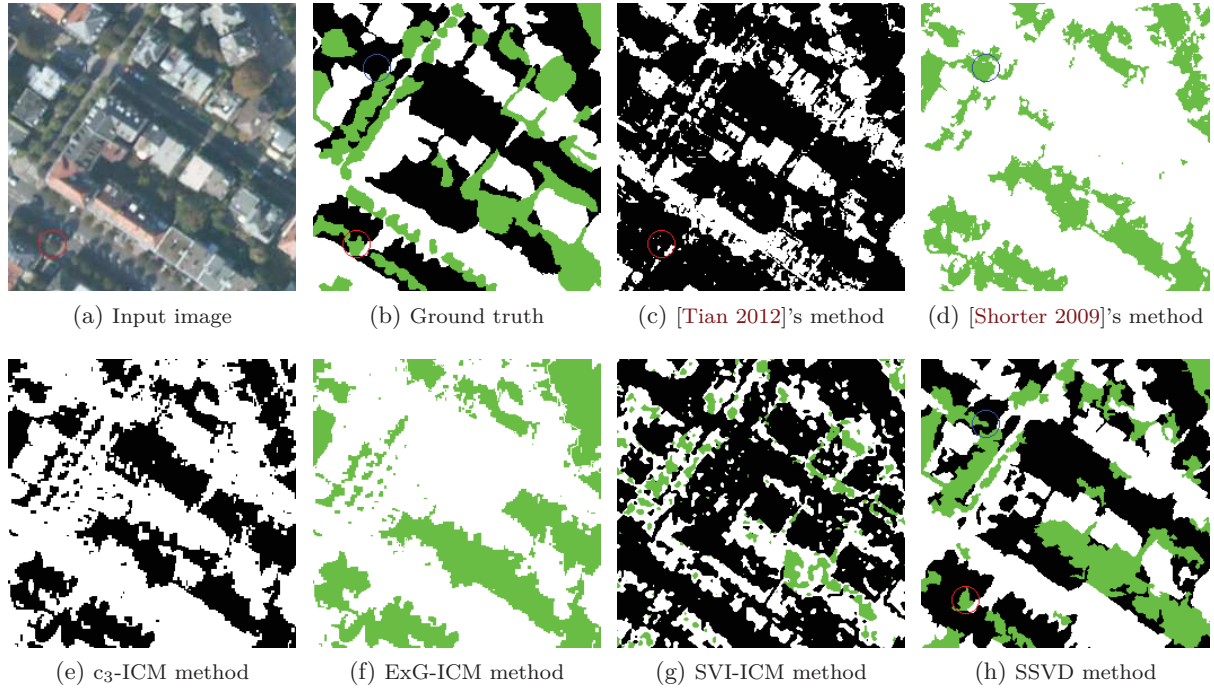
Figure 4.7: Shadow/vegetation detection from a RGB image (BD ORTHO® dataset) that covers a part of Strasbourg

Method	**Shadow detection**					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
[Tian 2012]	96.43	46.19	52.98	95.37	65.58	0.45
c₃-ICM	71.05	81.29	70.48	81.71	77.34	0.52
SVI-ICM	73.81	46.63	46.51	73.91	57.12	0.20
SSVD method	77.96	85.50	77.17	86.05	**82.59**	**0.63**
Method	**Vegetation detection**					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
[Shorter 2009]	25.86	73.67	22.55	77.02	62.74	0
ExG-ICM	41.89	68.76	28.44	79.97	62.62	0.09
SVI-ICM	27.26	98.44	83.89	82.03	82.17	0.41
SSVD method	61.25	91.52	68.17	88.85	**84.60**	**0.54**

Table 4.4: Shadow/vegetation detection accuracy measurements of Fig. 4.7.

(a) Input image　　　(b) Ground truth　　　(c) [Tian 2012]'s method　　　(d) [Shorter 2009]'s method

(e) c₃-ICM method　　　(f) ExG-ICM method　　　(g) SVI-ICM method　　　(h) SSVD method
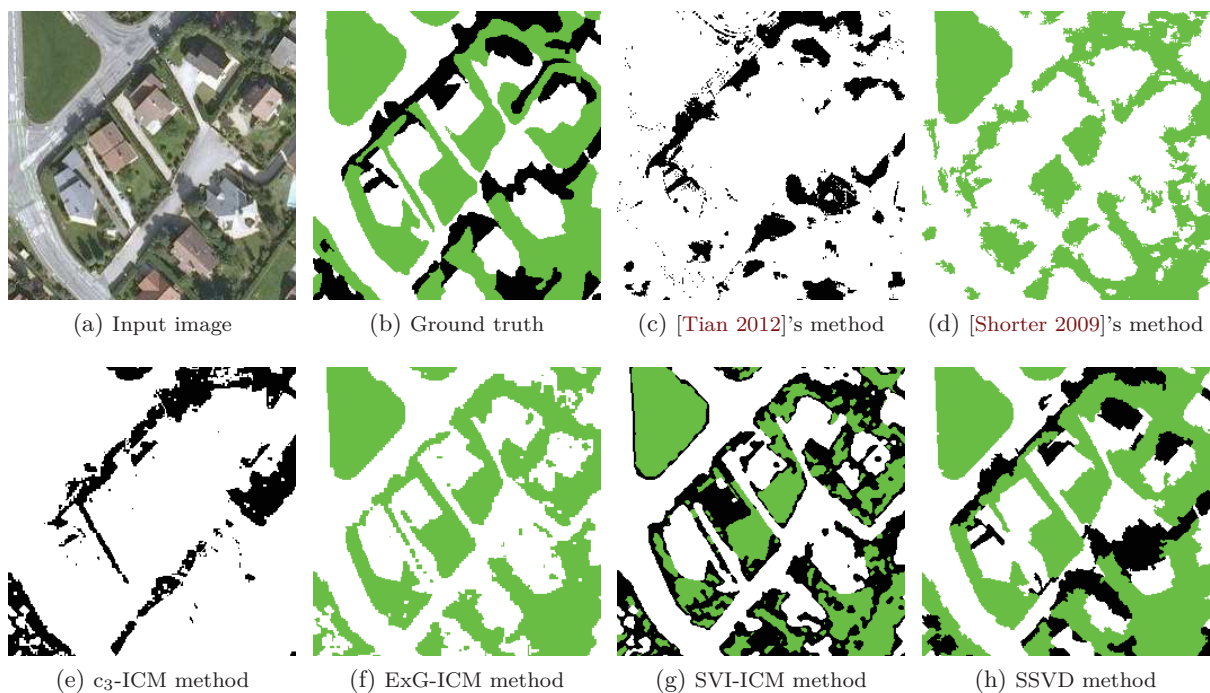
Figure 4.8: Shadow/vegetation detection from a RGB image (SZTAKI-INRIA benchmark) that covers a part of Normandy

Method	Shadow detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
[Tian 2012]	56.70	93.86	66.58	90.95	87.26	0.53
c₃-ICM	34.19	88.47	39.01	86.17	78.84	0.24
SVI-ICM	33.55	73.31	21.33	83.64	66.26	0.05
SSVD method	59.17	93.92	67.75	91.42	**87.75**	**0.56**
Method	Vegetation detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
[Shorter 2009]	69.84	83.14	74.67	79.49	77.61	0.53
ExG-ICM	90.38	86.34	82.48	92.66	88.02	0.76
SVI-ICM	55.93	94.82	88.49	75.15	78.66	0.56
SSVD method	91.17	85.88	82.12	93.18	**88.08**	**0.76**

Table 4.5: Shadow/vegetation detection accuracy measurements of the image in Fig. 4.8.

4.4, 4.5). We observe that most of shadow regions are detected by the method of [Tian 2012] (the producer's accuracies are 99.14, 96.43). However, this method classifies vegetation regions as shadows (e.g. the region marked by red circle, Fig. 4.6.c and Fig. 4.7.c). The method using index c_3 and MRF regularization (c_3-ICM) detect well shadow regions. But it classifies some non-shadowed regions (which have the similar c_3 response as shadows) as shadows (e.g. the region marked by violet circle, Fig. 4.6.d). The method of [Shorter 2009] detects well vegetation regions (especially for image in Fig. 4.8). But it classifies some shadow regions as vegetation (e.g. the region marked by blue circle, Fig. 4.6.d and Fig. 4.7.d). The SVI-ICM method detect well shadows and vegetation for image in Fig. 4.6, but it performs badly for image in Fig. 4.7 and Fig. 4.8. SSVD method outperforms all other methods, with high overall accuracy and have the best Matthews correlation coeeficient.

4.6.2.2 With NIR

Two examples of shadow/vegetation from RGB-NIR images are shown in Figs 4.9, 4.10. These two images of size 256 x 256 are selected from two datasets: Worldview-2 and Pleiades, that covers respectively a part of Jacmel city and Strasbourg. Figs. 4.9, 4.10 show that the c_3 index and the NDVI index are very appropriate for shadow detection and vegetation detection. The overall accuracy of the combination of these indices with MRF regularization are high (above 80 %). We observe that the method of [Rufenacht 2014] performs well in two examples (Fig. 4.9.d and 4.10.d). Especially, for image in fig. 4.10, this method outperforms our method (Acc/MCC: 83.49%/0.55 versus 82.45%/0.46). That demonstrate the efficiency of using NIR band for shadow detection task. However, SSVD method performs well for these two examples, with the overall accuracies are above 80%.

(a) Visible image (b) NIR image (c) Ground truth (d)[Rufenacht 2014]'s method

(e) c$_3$-ICM method (f) NDVI-ICM method (g) SVI-ICM method (h) SSVD method
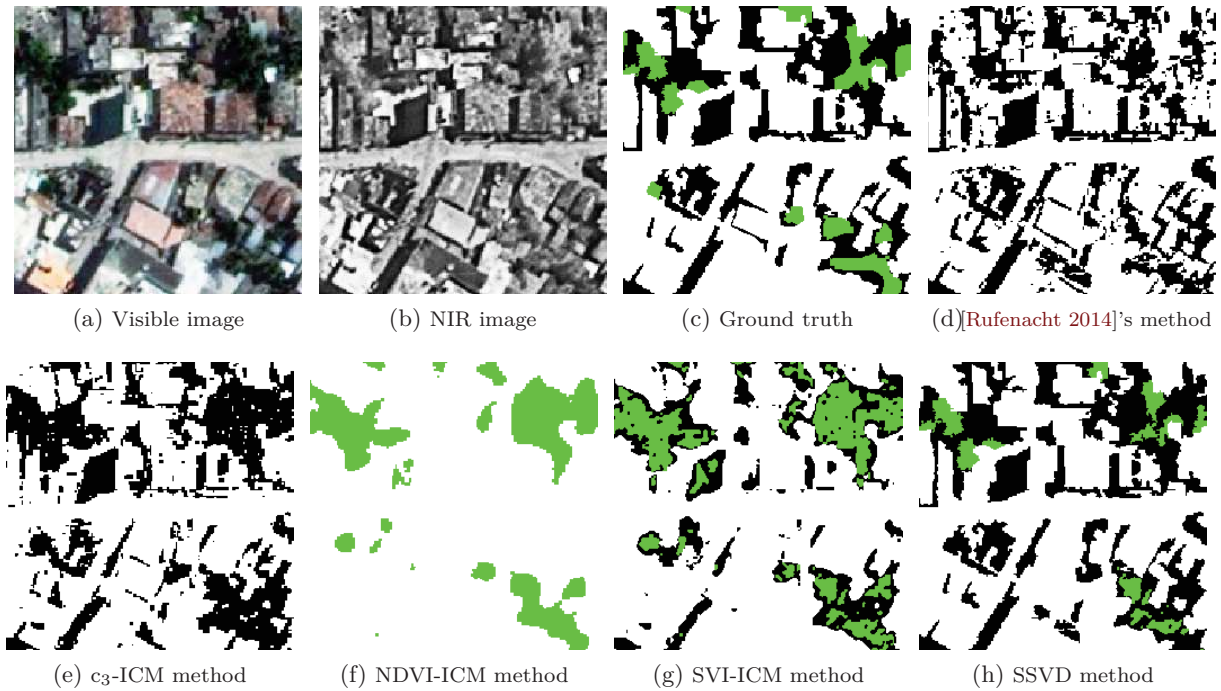
Figure 4.9: Shadow/vegetation detection from worldview-2 satellite image (Jacmel dataset)

Method	Shadow detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
[Rufenacht 2014]	86.36	86.92	72.74	94.03	86.76	0.70
c$_3$-ICM	79.16	82.95	65.24	90.78	81.86	0.59
SVI-ICM	46.05	93.34	73.66	81.06	79.73	0.46
SSVD method	96.34	94.38	87.39	98.46	**94.94**	**0.88**
Method	Vegetation detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
NDVI-ICM	94.91	89.56	42.44	99.54	89.96	0.60
SVI-ICM	73.28	92.79	45.18	97.71	91.32	0.53
SSVD method	57.88	99.48	90.03	96.68	**96.36**	**0.70**

Table 4.6: Shadow/vegetation detection accuracy measurements of Fig. 4.9.

(a) Visible image   (b) NIR image   (c) Ground truth   (d)[Rufenacht 2014]'s method

(e) c₃-ICM method   (f) NDVI-ICM method   (g) SVI-ICM method   (h) SSVD method
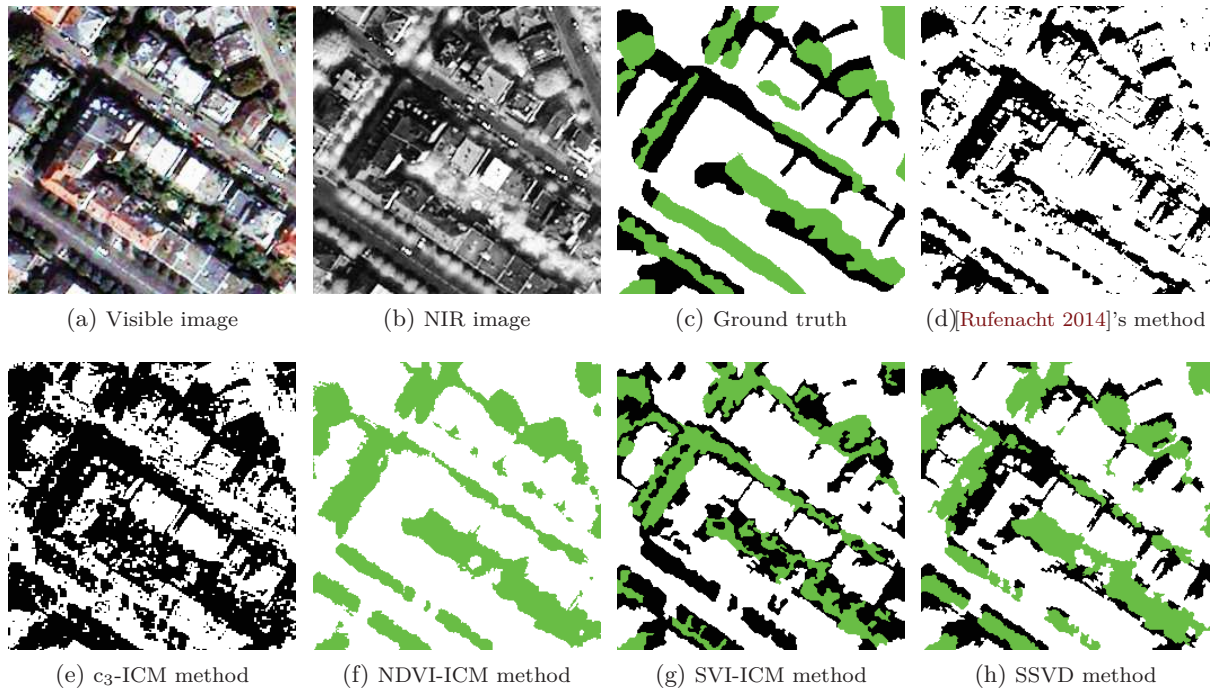
Figure 4.10: Shadow/vegetation detection from a pleiades satellite image that covers a part of Strasbourg

Method	Shadow detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Shadow	Non-shadow	Shadow	Non-shadow	Acc	MCC
[Rufenacht 2014]	70.76	87.12	61.10	91.25	**83.49**	**0.55**
c₃-ICM	92.80	58.58	39.03	96.60	66.19	0.42
SVI-ICM	32.06	75.24	27.01	79.49	65.64	0.07
SSVD method	51.15	91.40	62.95	86.75	82.45	0.46

Method	Vegetation detection					
	Producer's accuracies		User's accuracies		Overall accuracy	Matthews correlation coefficient
	Vegetation	Non-vegetation	Vegetation	Non-vegetation	Acc	MCC
NDVI-ICM	77.03	80.76	52.38	92.75	79.96	**0.51**
SVI-ICM	25.38	83.69	29.94	80.32	71.12	0.09
SSVD method	64.39	87.59	58.78	89.95	**82.59**	0.50

Table 4.7: Shadow/vegetation detection accuracy measurements of Fig. 4.10.

4.7 Conclusions

This chapter presents a novel shadow/vegetation detection strategy based on DS fusion procedure and MRF regularization. The fusion procedure makes it possible to combine different shadow index c_3, vegetation index ExG (or NDVI when the NIR band is available) and the luminance L. The spatial correlation between neighboring pixels is also taken into account using the MRF modelling in order to finally obtain a more reliable and efficient segmentation map with good accuracy, which is demonstrated in the experimental results.

This is the first contribution of this PhD and has been published in [Ngo 2014a] and [Ngo 2014b]. In the next part, we use the shadow/vegetation detection results for the automatic building detection from single optical image.

Part II

Building Detection

Building Detection Using Shadows and Image Segmentation

Contents

5.1 Introduction

Automatic detection of buildings in very high spatial resolution (VHR) remotely sensed imagery is of great practical interest for a number of applications; including urban monitoring, change detection, estimation of human population, among others. Researches in this domain started in the late 1980s, but due to the complexity and irregular nature of the scenes, building detection remains a difficult task. Since the initial research, building detection has been propelled by the development of new data sources and methods in the fields of photogrammetry and computer vision.

In this chapter, a novel approach for the automated detection of buildings from monocular VHR optical images is introduced. In Section 5.2, we review some related works in building detection. The workflow of the proposed method is presented in Section 5.3. The remaining sections detail our approach. The experimental results and discussion are presented in Chapter 6.

5.2   Related Works

There have been a significant amount of work on building detection from remote sensing images in the literature. The existing studies can be classified in different ways, with respect to:

- Degree of automation: automatic [Ok 2013], semi-automatic [Henricsson 1997], interactive [Gulch 1999].

- Input data: stereoscopic synthetic aperture radar (SAR) images [Tupin 2005], Li-DAR images [Sahar 2010], multi-view image (from the same geographical area [Fradkin 2001, Benedek 2012]), Digital Surface Models (DSM) [Lafarge 2010], grayscale image [Müller 2005], optical satellite and aerial images [Akcay 2010, Ok 2013, Katartzis 2008], hyperspectral image [McKeown Jr 1999].

- A priori knowledge: none, cadastral plan [Haala 2010], Laser Scanner Data [Rottensteiner 2007], topographic map [Hofmann 2002].

- Image features: points, edges, lines, regions [Huertas 1988, Lin 1998].

On the other hand, extensive reviews on building detection can be found in [Mayer 1999], [Baltsavias 2004], [Brenner 2005]. [Mayer 1999] discussed previous work published until the mid-1990s in an exceptional review in which the models and strategies of the developed approaches were deeply investigated and summarized. [Baltsavias 2004] presented results on buildings within the broader topic of object extraction using image analysis. More recently, [Ozgun Ok 2013] considered the previous studies aimed at automatically detecting buildings from monocular optical VHR image datasets.

Our aim is to develop an automatic building detection from single optical image without any a priori knowledge, using shadows and image segmentation. In this section, we limit the literature survey and discuss only the previous studies that involved in the proposed building detection framework, since it is impossible to mention all because of the variety of the developed methods.

5.2.1   Building Detection using Shadows

Several authors use shadows for rooftop segmentation from optical images. Shadows can be used in two ways. In the first hand, shadows are used after an initial building detection step, for building hypothesis verification and height estimation [Huertas 1988, Lin 1998, Sirmacek 2008]. For example, [Huertas 1988] utilized cast shadows to interpret the sides and corners of buildings. [Lin 1998] detected building roofs from oblique aerial images by assuming that the building shapes were rectilinear, and hypothesized rectangular buildings were verified both with shadow and wall evidences. In different study, [Sirmacek 2008] employed color invariant features and shadow information in a feature- and area-based approach. The shadows are detected by using the blue channel. Red roofs are identified by using the red channel. They first identify building candidate regions with searching shadow regions. If shadow regions are found, the regions, which are in opposite side of the shadowed regions with the illumination angle, are selected as candidate regions. Finally, a rectangle fitting method is used to align a rectangle with the Canny edges of the image.

In the other hand, if shadow regions can be accurately extracted from the images, they can support directly the detection steps. [Akcay 2010] detected candidate building regions using both shadow information and directional spatial constraints. The directional spatial relationship between buildings and their shadows are modeled by a probabilistic landscape approach. The final buildings were determined after clustering the candidate regions using minimum spanning trees. Recently, [Ozgun Ok 2013] used grabcut and shadows to segment rooftops from high-resolution imagery. Shadows were first detected and foreground (rooftop) pixels were labeled adjacent to them based on light direction. This was followed by iterative graph cuts (a modified version of the grabcut algorithm) on a region of interest (ROI) for each shadow. They employed an ROI determined by dilation of the shadow component with flat kernels opposite to the direction of light. Foreground pixels are determined by double thresholding a fuzzy region extended from shadows opposite to light direction. Then, they run grabcut in each ROI to label pixels inside it as rooftops or non-rooftops. Unlike the approach we propose, they required an additional near-infrared (NIR) band of data in order to locate shadows in the imagery.

More recently, inspired from the work of [Ozgun Ok 2013], the grabcut segmentation is also used in the work of [Femiani 2014]. But they run grabcut on the entire image (or overlapping tiles from the original image if the image is big) and not on an ROI for each shadow. Besides, they implement a self-correcting scheme that identifies falsely labeled pixels by analyzing the contours of buildings identified by the first pass of grabcut. The image is resegmented until a set of rooftops are extracted that are consistent with visible shadows. Their method requires only three-band RGB input data. In another study of [Li 2015], the input image is first segmented into perceptually homogeneous regions using the Gaussian mixture model (GMM) clustering method. Shadows and vegetation are then extracted from GMM labels. Remaining unlabeled regions are classified into probable rooftops and probable nonrooftops depending on shape, size, compactness and shadows. At last, a higher order multilabel conditional random field (CRF) segmentation is performed to get final results. These three methods ([Ozgun Ok 2013, Femiani 2014, Li 2015]) will be evaluated to compare with the proposed method.

5.2.2 Building Detection using Image Segmentation

In order to effectively retrieve objects from images, many methodologies partition image into smaller, usually homogeneous regions to enable region-based rather than global extraction. Image segmentation become a subsequent step in many object extraction algorithm, especially in building detection [Müller 2005, Liu 2005, Izadi 2010, Ozgun Ok 2013, Femiani 2014, Li 2015]. For example, [Liu 2005] used a probability function for detection of building regions. They first perform image segmentation by a multi-windows thresholding method. Afterwards, they extract features, e.g. shadow ratios, region entropy, contour edges and shape features in each region. Then, a confidence value is calculated by the probability function, which uses eight different features including distance to the straight lines, contour region entropy, contour including edges, grey level average value and standard deviation, shape, shadow ratio to define the region if it corresponds to a building. Some of the parameters are identified by manual interaction. The problems in the results, according to the authors, are mainly caused by shadows.

[Müller 2005]'s method is based on implementation of an area-based and feature-based algorithm for building detection from aerial imagery. They first convert the original image to a

grayscale image. Then, a region-growing process is applied for segmentation. A linear regression classifier detects the building regions using extracted features. However, their method works well only with buildings which have red roof.

In different study of [Izadi 2010], the input images are segmented at several levels. The line segments are detected in original images, and they are combined to generate roof hypotheses. The evidence of correct building is computed with a probability score for each hypothesis. They report that the proposed method is capable of detecting small gabled, residential roofs.

[Karadag 2015] proposed a domain-specific segmentation method for building detection, which integrates information related to the building detection problem into the detection system during the segmentation step. Buildings in a remotely sensed image are distinguished from the highly cluttered background, mostly, by their rectangular shapes, roofing material and associated shadows. The proposed method fuses the information extracted from a set of unsupervised segmentation outputs together with this a priori information about the building object, called domain-specific information (DSI), during the segmentation process. Finally, the segmentation output is provided to a two-layer decision fusion algorithm for building detection.

5.2.3   Building Detection using Shape Descriptors

Because of different materials of building rooftops, it is not easy to distinguish buildings from backgrounds using only spectral-based classification. In this case, shape analysis can be explored for extracting building rooftops. It is observed that the most common shape is naturally a rectangle or a combination of rectangles which are comprised of right-angled corners. Indeed, this building structure is often used because it needs less engineering effort in design, material and construction to achieve an equivalent or even an acceptable level of seismic performance [Lopez 1999]. [Cui 2011] used the Hough transform to extract the structure of buildings and then constructed a graph from those region information. A cycle detection on the graph is utilized finally to extract the boundary of buildings. [Benedek 2012] constructed a hierarchical framework to create various building appearance models from different elementary feature-based modules. The interaction between object extraction and local textural image-similarity information in their framework was exploited in a unified probabilistic model.

Another approach to deal with regular buildings is to use rectangular boundary fitting. Several rectangularity measures [Rosin 1999, Rosin 2003] are designed to evaluate, on their specific way, how much the object considered differs from a perfect rectangle. The standard method for calculating rectangularity is using the Minimum Bounding Rectangle (MBR) of the object [Toussaint 1983, Chaudhuri 2007, Chaudhuri 2012]. To the best of our knowledge, most of these approaches have been proposed for LiDAR images [Sahar 2010, Kwak 2014, Seo 2014], only two studies [Korting 2008, Korting 2011] have exploited the rectangularity measures to detect rectangular buildings in optical images.

5.3 Workflow of the Proposed Method

In this chapter, we present an original method for the automated detection of buildings from a single VHR optical image. We are motivated to develop a method that is applicable in various areas as shown in Fig. 5.1.

(a) Dense urban area   (b) Suburban area   (c) Rural area

Figure 5.1: Different study areas: high dense urban area with very high population density and attached buildings, suburban area with residential trip and detached buildings, rural area with very low population density, detached buildings and high vegetation density

(a) Rectangular roof   (b) L-shaped roof   (c) U-shaped roof   (d) I-shaped roof
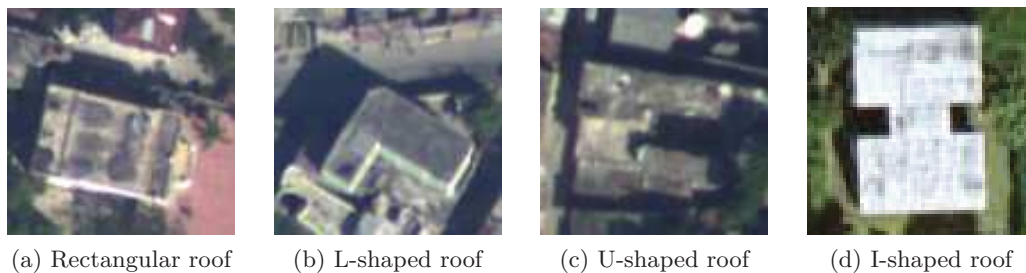
Figure 5.2: Examples of roof shapes that the proposed method can handle

This obliges us to follow certain assumptions about the appearance of the 3-D building objects:

1. We consider that buildings have a homogeneous color. Roof homogeneity have been exploited for building region detection in the literature [Müller 2005, Benedek 2012]. In practice, in high dense area, we separate the attached buildings based on their spectral features.

2. Our method is based on the fact that a 3D building structure should cast a shadow under suitable imaging conditions. Therefore, shadows generated from buildings must be detected. The images must be acquired in nadir views, and the time of acquisition favors the formation of shadows. Besides, correctness and precision of the shadow detection are strongly required.

3. Buildings may have similar spectral responses with their neighboring objects (e.g. street). Shape analysis is therefore used to distinguish buildings from backgrounds. In this study, we focus on the buildings with right-angled corners, such as: rectangular, L-shaped, U-shaped, I-shaped buildings (as shown in Fig. 5.2).

Starting from these assumptions, an original method for building detection task is designed and can be summarized as follows. The methodology begins with the detection and post-processing of shadow areas. For shadow detection, we employ our shadow/vegetation detection method presented in Chapter 4, that allows to divide the image into three distinct classes: *shadow, vegetation*, and *others* with good precision. The boundaries between shadows and their corresponding buildings are detected by eliminating shadow regions generated by vegetation objects and other non-building objects. In the second stage, a novel MRF region growing segmentation technique is proposed. Image is first over-segmented into smaller homogeneous regions (superpixels) which can be used to replace the rigid structure of the pixel grid. An iterative classification-merging is then applied over this set of regions. In region classification step, regions are classified into different classes and grouped into clusters. According to the position of shadows, a merging process is then performed to merge building segments with their neighboring regions which obtain the same class label in the classification stage to produce regions whose shapes are appropriate to rectangles. In the third stage, from the results of region growing image segmentation, the final building regions are determined using a recursive Minimum Bounding Rectangle (RMBR), proposed by [Kwak 2014]. The whole method is described in the flowchart in Fig. 5.3 . Details of the algorithm are presented in the following sections.
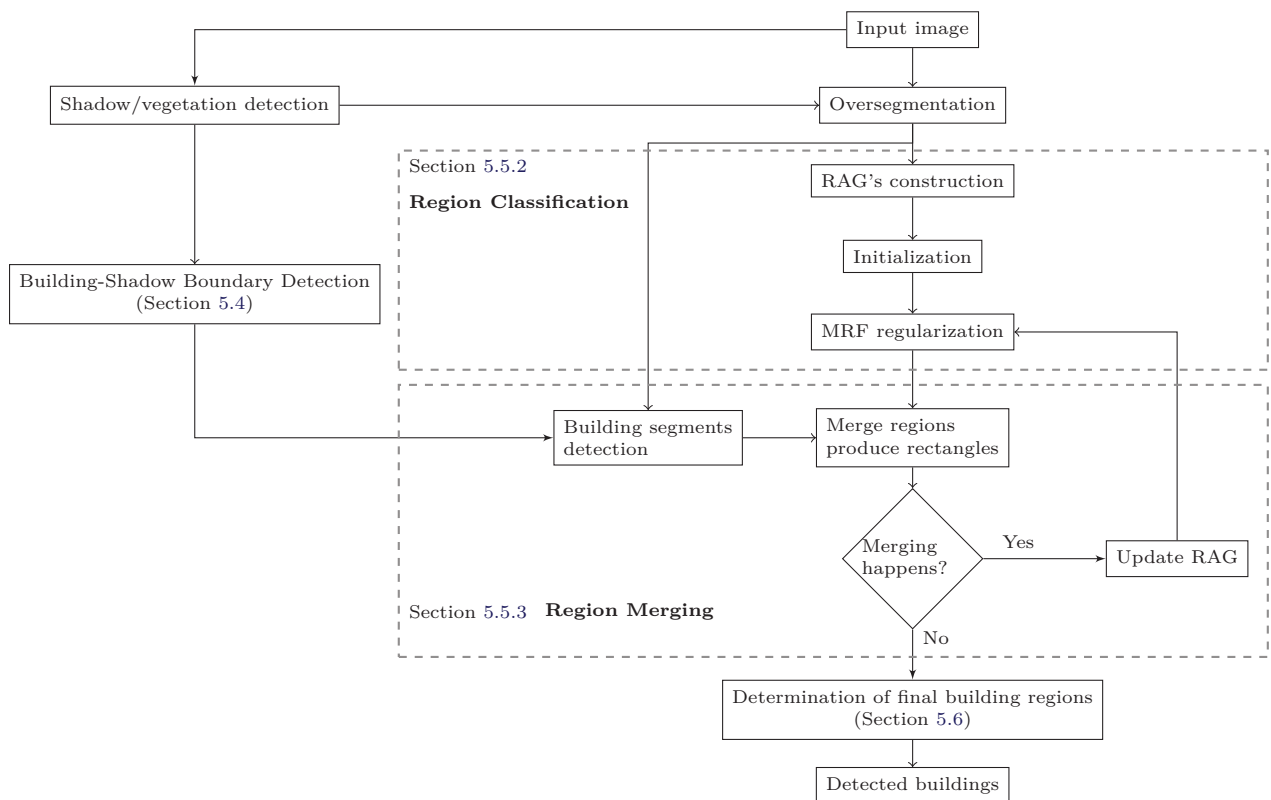
Figure 5.3: Flowchart of the proposed method and related sections

5.4 Building-Shadow Boundary Detection

A common feature that buildings possess is that they cast shadows on the ground. If shadows are present in the image, they can be used to identify the existence of buildings and to distinguish between buildings from flat, ground surface features. That is why our methodology begins with the detection of shadows and vegetation (Part I of the thesis). The shadow/vegetation detection step yields a vegetation mask M_V and a shadow mask M_S. Therefore, since building sides are usually surrounded by a variety of objects such as rock, vehicle, vegetation, the shadow may be cast by the building itself or by nearby objects (as shown in Fig. 5.4, red and blue circles). That is why it is essential for a building detection task to eliminate the shadow regions that occur due to non-building objects.

| (a) | (b) | (c) |
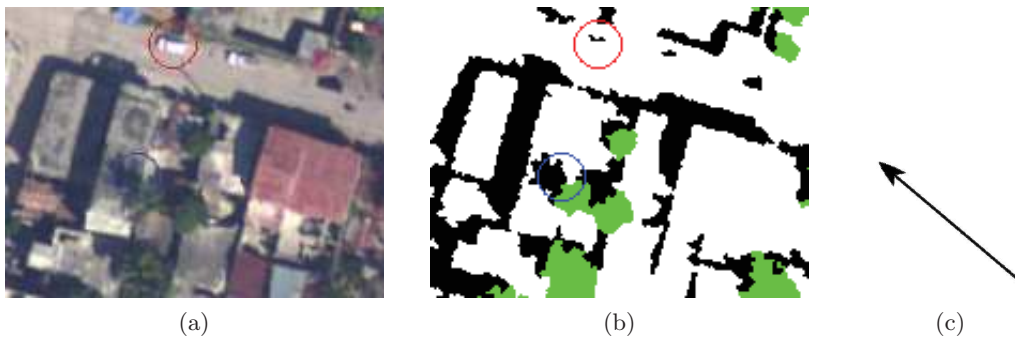
Figure 5.4: Shadow/vegetation detection: (a) input image, (b) detected shadow mask (black) and vegetation mask (green), (c) manually estimated illumination direction. Shadow is occluded by vehicles (red circle) or by vegetation (blue circle)

In reality, the detection of shadows cast by buildings is rather impossible since the only priori knowledge is the illumination angle (without any knowledge about the height of buildings or the zenith angle). However, as will be discussed in Subsection 5.5.3.1, what interests us is the boundary between buildings and their corresponding shadows. Therefore, the postprocessing step of shadow mask can be carried out as follows.

5.4.1 Elimination of Shadows cast by Vegetation

We suppose that the illumination angle θ is supplied. Otherwise, it can be empirically estimated from an image by counting the number of pixels horizontally and vertically from one corner of a rooftop (or from a corner of the footprint if lean is an issue), to the corresponding corner of a building shadow. To select shadows generated by distinct vegetation objects, we investigate, for each vegetation object of the vegetation mask M_V, the shadow evidence within the close neighborhoods of the vegetation object. To do that, binary morphological dilation is used, which allows expanding the shape of a vegetation object (as illustrated in Fig. 5.5b). The direction of the structuring element is determined by the illumination angle θ and its length l_{se} is empirically chosen. Once this expansion region is defined, we check for shadow evidence within the defined region with the help of the pre-computed binary shadow mask M_S. If there is more than one shadow region occurring in this expansion region, we select the shadow region that have a border with vegetation object (as shown in Fig. 5.5c).

(a)   (b)   (c)
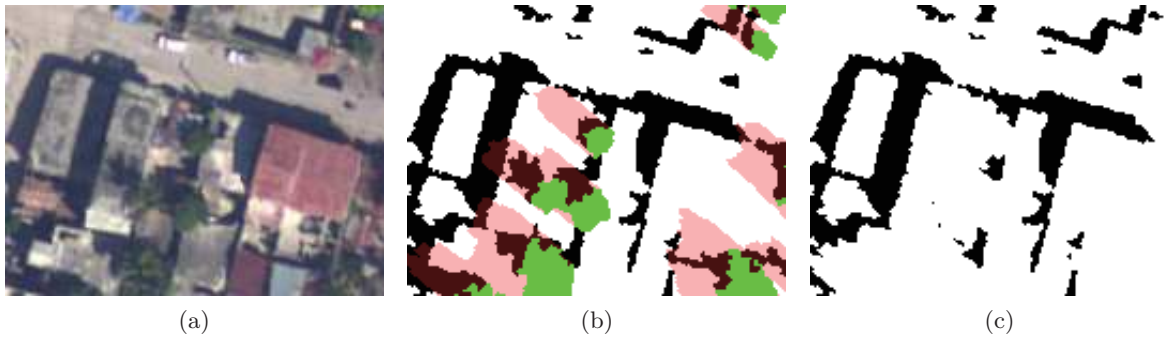
Figure 5.5: Removing shadows due to vegetation. (a) RGB aerial image, (b) The expansion regions (pink color) generated after the dilation of the vegetation objects overlaid with the shadow/vegetation map, (c) the shadow mask after eliminating shadows generated by vegetation

5.4.2   Elimination of Shadows cast by other non-Building Objects

Furthermore, shadows can be cast by other non-building objects as vehicles, rocks, . . . that are generally short objects. Thus, to to eliminate these shadow regions, we found it necessary to compute the length of boundary between each shadow object and their corresponding building and then filter out those building-shadow boundaries whose length is below the predefined threshold d_{sh}.

(a)   (b)   (c)
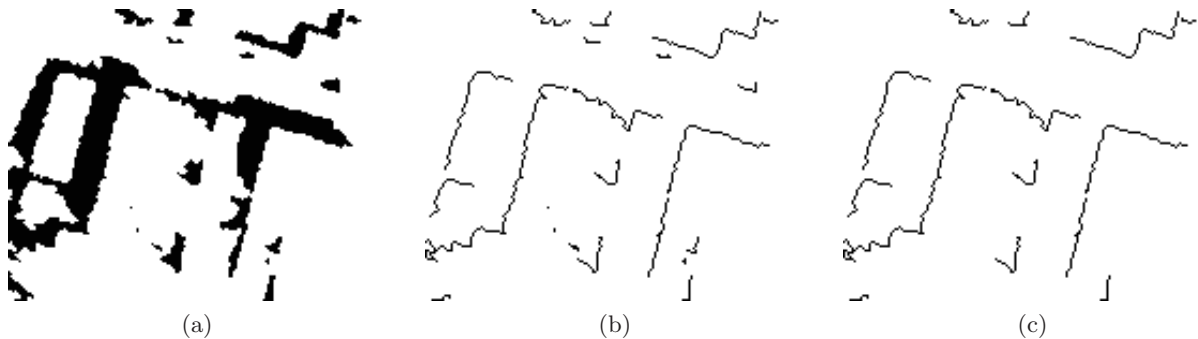
Figure 5.6: Removing shadows due to other non-building objects. (a) the shadow mask after eliminating shadows generated by vegetation, (b) potential building-shadow boundaries of (a), (c) final building-shadow boundary mask M_{SB} (black)

Potential boundaries between each shadow object and their corresponding building are detected as follows. The opposite direction of the illumination angle is quantized into one of eight directions (north (N), northeast (NE), east (E), southeast (SE), south (S), southwest (SW), west (W), and northwest) because we are working on a digital image grid. For example, for the illumination angle shown in Fig. 5.4c, the direction is southeast (SE). The contour of shadow objects are first detected. For each pixel on the contour, if it's pixel on the southeast is shadow, it will be removed from the contour. The remaining contour is considered as the potential building-shadow boundary (as shown in Fig. 5.6.b). Potential building-shadow boundaries whose length is below the predefined threshold d_{sh} are eliminated. The remaining shadow-building

boundary is shown in Fig. 5.6c and denoted as M_{SB}. The result achieved indicates that after the post-processing, all cast shadows except those that correspond to the main body of the building are successfully eliminated.

The objective now is that for each building-shadow boundary object, we identify the casting building region. In addition, when one shadow object is generated from a group of attached buildings, the developed algorithm must be capable to separate these attached buildings. To do that, we propose a novel region growing image segmentation, whose details are presented in the next section.

5.5 Region Growing Image Segmentation

Segmentation of remote sensing images is a difficult problem due to mixed pixels, spectral similarity, and the textured appearance of many land-cover types. A variety of segmentation techniques have been applied to remote sensing imagery with varying degrees of success [Dey 2010]. Among them, the region growing method is widely used for analyzing remote sensing imagery [Yu 2007, Qin 2010, Huang 2014]. In region growing technique, the input image is first tessellated into a set of homogeneous primitive regions (or superpixels), and an image segmentation is performed by iteratively merging the similar neighboring regions such that a certain homogeneity criterion is satisfied. In previous works, there are region-merging algorithms based on statistical properties [Nock 2004, Calderero 2010], graph properties [Chen 2013, Haris 1998], and spatio-temporal similarity [Moscheni 1998].

Our proposed region growing image segmentation method consists of two main stages: region classification and region merging, as shown in Fig. 5.3. An example of the first iteration of this iterative process is shown in Fig. 5.7 and can be summarized as follows. Image is first decomposed into different homogeneous regions by the oversegmentation algorithm SLIC [Achanta 2012] which allows to generate regular-sized regions with good boundary adherence. This algorithm is carried in the image in which shadow regions M_S and vegetation regions M_V are masked out. An iterative classification-merging is then applied over this set of regions. Regions are grouped into different clusters by the MRF-based region classification algorithm (Fig. 5.7.c). Building-Shadow boundary M_{SB} (Fig. 5.7.d) is used to determine the building segments (Fig. 5.7.e, the regions that are supposed to be a part of building) and a merging process is then performed to merge building segments with their neighboring regions which obtain the same class label in the classification stage to produce regions whose shapes are appropriate to rectangles (Fig. 5.7.f). This iterative classification-merging stops when no merging happens. The proposed algorithm is detailed as follows.

5.5.1 Oversegmentation SLIC

Oversegmentation occurs when image is segmented into smaller regions, each referred to as a "superpixel". A superpixel is a spatially-coherent, homogeneous, structure which preserves information over scales or sampling resolution [Moore 2009]. Superpixel algorithm allows us to capture image redundancy, provide a convenient primitive from which to compute image features, and greatly reduce the complexity of subsequent image processing tasks. Thus, they have become a common preprocessing step for many computer vision algorithms, for example, in

image segmentation [Yang 2008, Wang 2008, Ding 2010].

Different methods can be used to oversegment an image, including mean shift segmentation [Comaniciu 2002], watershed segmentation [Vincent 1991], turbopixel method [Levinshtein 2009], constrained Delaunay triangulation (CDT) method [Ren 2005], graph-based methods such as normalized cuts [Shi 2000] and efficient graph-based image segmentation [Felzenszwalb 2004]. In our proposed method, the oversegmentation algorithm SLIC [Achanta 2012] is employed. It is a simple and efficient method to decompose an image in visually homogeneous regions, based on the color similarity and proximity of pixels in the image plane. The SLIC is summarized in Algorithm 2 and is presented in the following.

Distance measure   The SLIC superpixel segmentation algorithm is a K-means-based local clustering of pixels in the 5-D $[Labxy]$ space defined by the L, a, b values of the CIELAB color space and the x, y pixel coordinates. CIELAB color space is chosen because it is perceptually uniform for small color distance. The SLIC algorithm takes as input a desired number of approximately equally-sized superpixels K, then for a image with N pixels, the approximate size of each su-

(a) Input Image   (b) Oversegmentation   (c) Region classification

First iteration

Yes

Merging happens?

No

(d) Building-Shadow Boundary M_{SB} (e) Building Segments   (f) Region merging   Stop

Figure 5.7: An example of the first iteration of the region growing image segmentation method: (a) input image, (b) SLIC oversegmentation (regions are separated by cyan lines) of the image in which shadow regions M_S and vegetation regions M_v are masked out, (c) MRF-based classification (clusters are separated by red lines), (d) Building-Shadow boundary mask M_{SB} (see Section 5.4), (e) detected building segments (boundaries are delineated by violet lines), (f) after merging (some cyan lines are disappeared).

.

perpixel is N/K. To produce roughly equally sized superpixels, the grid interval is S = $\sqrt{N/K}$. Let $[l_i, a_i, b_i, x_i, y_i]$ be the 5-D point of pixel i, cluster center C_k is represented by the same form $[l_k, a_k, b_k, x_k, y_k]$. Instead of directly using the Euclidean distance in the 5-D $[Labxy]$ space, SLIC introduce a new distance measure that considers superpixel size. For example, the distance D_k between pixel i and cluster center C_k is defined as follows:

$$
\begin{aligned}
d_{lab} &= \sqrt{(l_i - l_k)^2 + (a_i - a_k)^2 + (b_i - b_k)^2} \\
d_{xy} &= \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2} \\
D_k &= \sqrt{d_{lab}^2 + \left(\frac{m}{N_s}\right)^2 d_{xy}^2}
\end{aligned}
\tag{5.1}
$$

where D_k is the sum of the lab distance and the xy plane distance normalized by the grid interval S. The parameter N_s is the maximum spatial distance expected within a given cluster, $N_S = S = \sqrt{N/K}$. The parameter m is introduced to control the compactness of superpixels. When m is large, spatial proximity is more important and the resulting superpixels are more compact (i.e., they have a lower area to perimeter ratio). When m is small, the resulting superpixels adhere more tighly to image boundaries, but have less regular size and shape.

Algorithm 2: SLIC superpixel segmentation [Achanta 2012]

/ Initalization */*
Initialize cluster centers $C_k = [l_k, a_k, b_k, x_k, y_k]$ by sampling pixels at regular grid steps S.
Move cluster centers to the lowest gradient position in a 3×3 neighborhood (Eq. 5.2).
Set label $l(i) = -1$ for each pixel i.
Set distance $d(i) = \infty$ for each pixel i.
repeat
    / Assignment */*
    for *each cluster center C_k* **do**
        for *each pixel i in a 2S $\times$ 2S region around C_k* **do**
            Compute the distance D_k between C_k and i.
            if $D_k < d(i)$ **then**
                set $d(i) = D_k$
                set $l(i) = k$
    / Update */*
    Compute new cluster centers.
    Compute residual error E.
until $E \leq threshold$;

Algorithm: The SLIC superpixel algorithm is summarized in Algorithm 2. It begins by sampling K regularly spaced cluster centers and moving them to seed locations corresponding to the lowest gradient position in a 3×3 neighborhood. This is done to avoid placing them at an edge and to reduce the chances of choosing a noisy pixel. Image gradients are computed as:

$$
G(x, y) = \|\mathbf{I}(x+1, y) - \mathbf{I}(x-1, y)\|^2 + \|\mathbf{I}(x, y+1) - \mathbf{I}(x, y-1)\|^2
\tag{5.2}
$$

where $\mathbf{I}(x, y)$ is the *lab* vector corresponding to the pixel at position (x, y), and $\|.\|$ is the L_2 norm. Next, in the assignment step, each pixel i is associated with the nearest cluster center whose search region overlaps this pixel. Since the expected spatial extent of a superpixel is a region of approximate size S × S, the search for similar pixels is done in a region 2S × 2S around the superpixel center. Once each pixel has been associated to the nearest cluster center, the cluster centers are updated as the $[l, a, b, x, y]^T$ vector of all the pixels belonging to the cluster. The assignment and update steps are repeated iteratively until the stop criteria is fulfilled: a residual error E (L_2 norm) between the cluster center locations and previous cluster center locations is inferior to a predefined threshold. In the implementation, the processing stops when a predefined number of iteration is achieved. According to [Achanta 2012], 10 iterations are sufficient for most images.

Since the SLIC method does not enforce connectivity between pixels, there can be small and isolated segments resulted at the end of the clustering procedure. To obtain relatively uniform superpixels for further applications, the SLIC method adopts a post-processing which merges the isolated clusters with its largest neighbor. The stray segments whose size is smaller than a threshold (50 pixels) are eliminated (relabelled as the label of its nearest cluster).

In our implementation, the SLIC algorithm is applied on the image in which shadow regions M_S and vegetation regions M_V are masked out. The two parameters in SLIC are set as follows. The weighting factor m between colour and spatial differences is set to 10, as recommended by the authors, which can sufficiently preserve the boundaries of objects and the number of superpixels is set so as to have the superpixel size η_{sup}. Fig. 6.1 shows the examples of oversegmentation with different values of η_{sup}. We observe that oversegmentation generates regular-sized regions with high boundary adherence.

(a) $\eta_{sup} = 100$　　　　(b) $\eta_{sup} = 200$　　　　(c) $\eta_{sup} = 300$

Figure 5.8: An example of oversegmentation SLIC with different values of η_{sup} and the weighting factor m is chosen as 10.

5.5.2 Region Classification

MRF model [Besag 1986] provides a basis for modeling contextual constraints in visual processing and interpretation. Although MRF is mostly used on the pixel graph [Geman 1984], it is also proved to be a powerful model for feature-based graph (such as region adjacency graph (RAG) [Tupin 2005, Qin 2010] or line segment graph [Krishnamachari 1995]). In the proposed approach, a MRF model defined over a RAG is proposed as follows.

RAG's Construction

It is assumed that the image is initially over-segmented into a set $\mathcal{R}$ of disjoint regions. A RAG $G = (S, \mathcal{E})$ is defined over the set $\mathcal{R}$ of regions obtained from the oversegmentation step. Each node $s \in S$ represents each region $R_s \in \mathcal{R}$. The relationship between two regions is given by their adjacency, defining a set $\mathcal{E}$ of edges, as illustrated in Fig. 5.9. Suppose image is to be segmented into K classes. Let $\Omega = \{\omega_1, \ldots, \omega_K\}$ denote the set of class labels. x_s is a realization of the label X_s of region R_s. Also, let $\mathbf{X} = (X_s)_{s \in S}$ denote the joint random variable and the realization (configuration) $\mathbf{x} = (x_s)_{s \in S}$ of $\mathbf{X}$. $\mathbf{x}$ is estimated using $\mathbf{y} = (\mathbf{y}_s)_{s \in S}$ where $\mathbf{y}_s$ is the observation of all pixels in region R_s, and therefore $\mathbf{y}_s = \{\mathbf{y}_s(i), i \in R_s\}$. For input images, $\mathbf{y}_i(s)$ is a 3-dimensional feature vector (red, green, blue representation). $\mathbf{y}$ (resp. $\mathbf{y}_s$) is a realization of the observation field $\mathbf{Y}$ (resp. $\mathbf{Y}_s$).
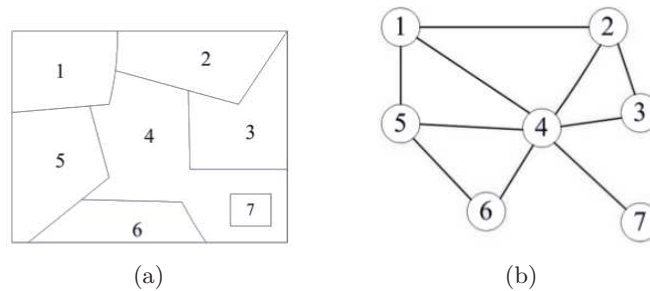
Figure 5.9: Example of RAG: (a) segmented image and (b) the corresponding RAG representation

MRF's Framework

We want to find an assignment $\hat{x}_s$ of all sites (nodes) $s \in S$ to Ω. To introduce contextual knowledge, $\mathbf{X}$ is supposed to be a MRF for the neighborhood defined by RAG. In MRF model, the optimal configuration $\hat{\mathbf{x}} = \{\hat{x}_s, s \in S\}$ which we are interested is the one that will be a maximum a *posteriori* probability (MAP) under the observation $\mathbf{Y}$. That is:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} P(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) \tag{5.3}$$

From Bayes'rule, we have:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \frac{P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{P(\mathbf{Y} = \mathbf{y})} \tag{5.4}$$

When the image is designed, $P(\mathbf{Y} = \mathbf{y})$ is constant. To maximize the *a posteriori* probability leads to minimize the posterior energy function:

$$
\begin{aligned}
U(\mathbf{x}, \mathbf{y}) &= -\log\left(P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})\right) \\
&= -\log\left(P(\mathbf{Y} = \mathbf{y}|\mathbf{X} = \mathbf{x})P(\mathbf{X} = \mathbf{x})\right) \\
&= -\log\left(P(\mathbf{Y} = \mathbf{y}|\mathbf{X} = \mathbf{x})\right) - \log\left(P(\mathbf{X} = \mathbf{x})\right) \\
&= U(\mathbf{y}|\mathbf{x}) + U(\mathbf{x})
\end{aligned}
\tag{5.5}
$$

The first term $U(\mathbf{y}|\mathbf{x})$ in Eq. (5.5) is the *likelihood* term, describing the probability of region R_s with its observation $\mathbf{Y}_s$ at the given region label x_s. Due to the independence assumption of the regions, the likelihood term can be written: $U(\mathbf{y}|\mathbf{x}) = \sum_{s\in S} U_s(\mathbf{y}_s|x_s)$. As Gaussian distribution is a usual and effective distribution for color images, this distribution is adopted to describe the image model. So, in cases where x_s takes the class label ω_k:

$$
U_s(\mathbf{y}_s|x_s) = \sum_{i\in R_s} \frac{1}{2} \times \left(\log(|\boldsymbol{\Sigma}_k|) + [\mathbf{y}_i(s) - \boldsymbol{\mu}_k]^T \boldsymbol{\Sigma}_k^{-1} [\mathbf{y}_i(s) - \boldsymbol{\mu}_k]\right)
$$

where $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$ are mean and covariance matrix of class ω_k.

The second term $U(\mathbf{x})$ in Eq. (5.5) is the *prior* term, describing what the likely labelings $\mathbf{x}$ should be like. Inside a building, the different parts should have a rather homogeneous color. This knowledge is introduced in the definition of the clique potential of the RAG. In order to reduce the computational complexity, we will restrict our attention to MRF's whose clique potentials involve pairs of neighboring nodes ($<s,t> \in \mathcal{E}$). The prior term is defined as follows:

$$
U(\mathbf{x}) = \sum_{s\in S} U_s(x_s) = \sum_{s\in S}\sum_{t\in \mathcal{V}_s} |R_s| \times \frac{b_{st}}{b_s} \times \frac{\beta}{|\bar{\mathbf{y}}_s - \bar{\mathbf{y}}_t|} \times (1 - \delta(x_s - x_t))
\tag{5.6}
$$

where:

- $\delta(\cdot)$: the Kronecker's delta function,

- $\mathcal{V}_s \subset S$: the neighbors of the node s,

- $|R_s|$: the number of pixels in region R_s,

- b_s : the length of boundary of region R_s,

- b_{st} : the length of common boundary of region R_i and region R_j,

- $\bar{\mathbf{y}}_s$: the mean intensity (taking values from 0 to 255) of region R_s.

Two constraints are introduced in the prior term: the normalized edge weight b_{st}/b_s and the inversed difference $|\bar{\mathbf{y}}_s - \bar{\mathbf{y}}_t|^{-1}$. They mean that if two regions share a long boundary and have similar mean intensity, they have high probability to obtain the same class label. β represents the tradeoff between fidelity to the observed image and the smoothness of the segmented image. In practice, the parameter β is empirically chosen.

Energy minimization: Finding a solution for for Eq. (5.5) represents a combinatorial optimization problem. Various combinatorial optimization techniques are known, including iterated conditional mode (ICM) [Besag 1986], simulated annealing (SA) [Geman 1984], graph cuts

[Kolmogorov 2004], loopy belief propagation [Yedidia 2005], and tree-reweighted message passing [Kolmogorov 2006]. The ICM [Besag 1986], the simplest among all of these, will be employed here. Each node is updated by choosing the class label x_s which minimize the local energy:

$$U_s(\mathbf{y}_s, x_s) = U_s(\mathbf{y}_s | x_s) + U_s(x_s) \tag{5.7}$$

This algorithm does not guarantee the convergence toward the optimal solution (only a local optimum depending on the chosen initialization is reached).

Parameter estimation: The parameters of MRF model are estimated at each iteration of ICM algorithm as follows:

$$\boldsymbol{\mu}_k = \frac{\sum\limits_{s \in S_k} \sum\limits_{i \in R_s} \mathbf{y}_s(i)}{\sum\limits_{s \in S_k} |R_s|} \tag{5.8}$$

$$\boldsymbol{\Sigma}_k = \frac{\sum\limits_{s \in S_k} \sum\limits_{i \in R_s} (\mathbf{y}_s(i) - \boldsymbol{\mu}_k)(\mathbf{y}_s(i) - \boldsymbol{\mu}_k)^T}{\sum\limits_{s \in S_k} |R_s|} \tag{5.9}$$

where S_k denotes the set of nodes whose class label is ω_k.

(a) Input image   (b) Oversegmentation   (c) Initialization   (d) MRF regularization
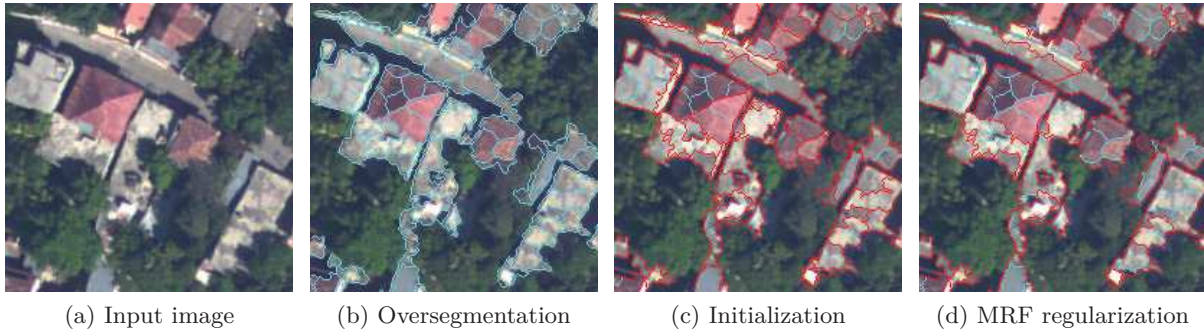
Figure 5.10: Experimental results of MRF-based region classification. Regions are separated by cyan lines and clusters are separated by red lines.

Initialization: Similar to the RKM method of [Qin 2010], our method seeks a set of optimal class mean vector $\boldsymbol{\mu}_k$, $k = 1, \ldots, K$ to minimize the following energy function:

$$E_{\text{init}} = \sum_{k=1}^{K} \sum_{s \in S_k} \sum_{i \in R_s} (\mathbf{y}_s(i) - \boldsymbol{\mu}_k)(\mathbf{y}_s(i) - \boldsymbol{\mu}_k)^T \tag{5.10}$$

The first classification $S_k, k = 1, \ldots, K$ is obtained by applying the K-means clustering on the mean values of regions: $\bar{\mathbf{y}}_s = \dfrac{\sum_{i \in R_s} \mathbf{y}_s(i)}{|R_s|}$. The class mean vectors $\boldsymbol{\mu}_k$, $k = 1, \ldots, K$ are computed by Eq. 5.8. An iterative process is carried out. At each iteration, the label field $\mathbf{x}$ is updated using the current class mean vectors according to the nearest center rule:

$$x_s = \arg\min_{\omega_k \in \Omega} \sum_{i \in R_s} (\mathbf{y}_s(i) - \boldsymbol{\mu}_k)(\mathbf{y}_s(i) - \boldsymbol{\mu}_k) \tag{5.11}$$

$\boldsymbol{\mu}_k$, $k = 1, \ldots, K$ is then recomputed using by Eq. 5.8. The energy E_{init} is updated using 5.10. The iterative procedure is terminated once the ratio of the absolute energy difference between consecutive iterations over the energy at the former iteration is below 10^{-6} or the number of iterations exceeds 100, to ensure that we obtain the best initialization as possible.

An example of region classification is shown in Fig. 5.10, in which regions are separated by cyan lines. After classification, connected regions that have similar spectral characteristics are grouped into clusters. In Fig. 5.10, clusters are separated by red lines.

5.5.3 Region Merging

The region merging procedure is designed based on three assumptions about the targeted buildings. First, buildings cast shadows on the ground, so regions to be merged have at least one region that is next to shadows in the opposite direction of the illumination angle. Second, we focus only on detecting the buildings with right-angled corners, a merging procedure is designed to produce new regions whose shapes are appropriate to rectangles. Third, we assume that building has a homogeneous color, so the regions to be merged need to have similar spectral characteristics. Hence, merging is only done between regions with the same class label. Starting from these three constraints, different steps of merging are designed and detailed in the following.

(a) Input image   (b) Oversegmentation   (c) Building Boundary M_{SB} (d) Building segments
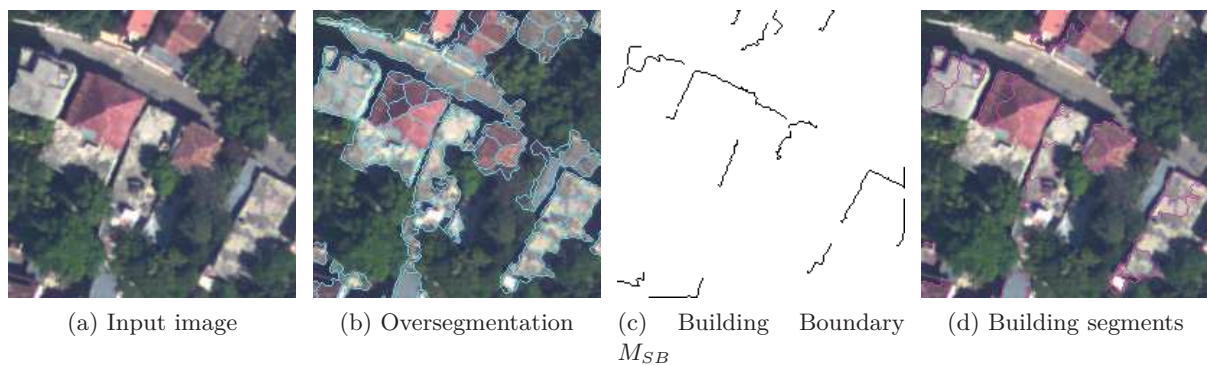
Figure 5.11: An example of building segments detection. We observe that each targeted building has one or several building segments.

5.5.3.1 Determination of Building Segments

The term "building segment" refers to regions that border with shadows cast by building, which are detected in Section 5.4, in the opposite direction of the illumination angle (as shown in Fig. 5.11, building segments are delineated by violet lines). To detect building segments, we measure the boundary between regions and the building-shadow boundary M_{SB}. Since region that shares a larger border with shadows is more likely to be a building segment, only regions whose boundary with M_{SB} is larger than a predefined threshold T_S is flagged as a building segment.

5.5.3.2 Measure the rectangularity

Minimum Bounding Rectangle (MBR) Several rectangularity measures already exist in literature. Basically, all of them are designed to evaluate, on their specific way, how much the shape considered differs from a perfect rectangle. The standard method for calculating rectangularity is using the MBR (Minimum Bounding Rectangle) of the segment. A well-known algorithm for finding the MBR is presented in [Toussaint 1983]. For an object, this technique first computes its convex hull, which can be done in $\mathcal{O}(n \log n)$ time, and then find the minimal rectangle in $\mathcal{O}(n)$ time (n is the number of edges of the convex hull). This depends on the observation that one side of the minimal rectangle must coincide with one edge of the convex polygon it must contain [Toussaint 1983]. Thus, one only has to consider the orientations given by edges of the convex polygon. For each orientation, we determine the minimum rectangle containing the corresponding edge of the convex polygon. We repeat this task for all orientations and the rectangle with the minimum area is chosen as the Minimum Bounding Rectangle. The illustration of the method is shown in Fig. (5.12).

<center>(a)   (b)   (c)   (d)</center>
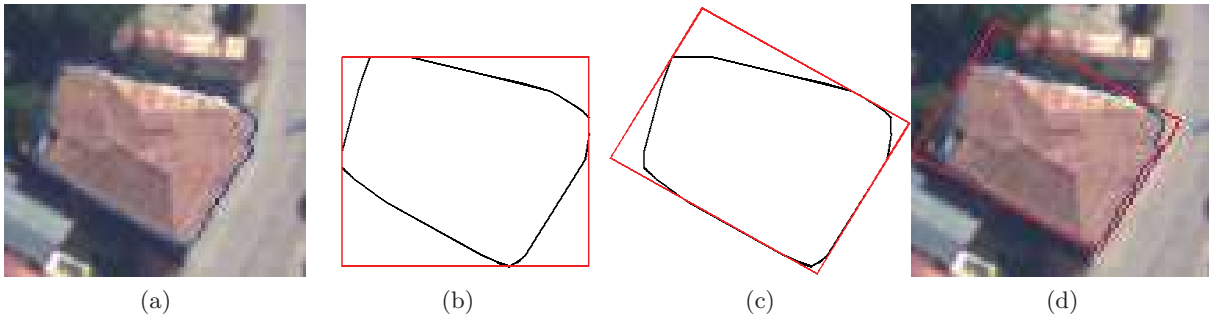
Figure 5.12: Illustration of the MBR estimation [Toussaint 1983]: (a) Object and its boundary (in blue), (b) convex hull and one bounding rectangle test (in red), (c) bounding rectangle with minimum area (in red), (d) object and its MBR

The rectangularity score of each segment is then computed as follows:

$$R_B = \frac{A_0}{A_{\text{MBR}}} \tag{5.12}$$

where A_0 is the region's area and A_{MBR} is the area of the MBR.

Rectangular Discrepancy A weakness of using the MBR is that it is very sensitive to protrusions from the region (as shown in Fig. 5.13). A narrow spike out of the region can vastly inflate the area of the MBR, and thereby produce very poor rectangularity estimates. Three new methods for measuring the rectangularity of regions are developed by Rosin [Rosin 1999]. They are tested together with the standard minimum bounding rectangle method on synthetic and real data. It is concluded that, while all the methods have their drawbacks, the best two are the bounding rectangle and discrepancy methods. In the discrepancy method, a rectangle is fitted to the region based on its moment. One way to measure the sides of rectangle is to find the best ellipse that corresponds to the region. This is an ellipse with the same first- and second-order

(a)   (b)

Figure 5.13: Building with protrusion artifact: (a) Building region (whose boundary is delineated by blue line) and its minimum bounding rectangle (red line). (b) Building region and the fitted rectangle estimated by discrepancy method.

moments as the region. From the semi-major and semi-minor axes α and β of the ellipse, the rectangle's measurement a and b are estimated as follows:

$$
\begin{aligned}
a = \sqrt{3}\alpha &= \sqrt{\frac{6\left[\mu_{02} + \mu_{20} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}\right]}{\mu_{00}}} \\
b = \sqrt{3}\beta &= \sqrt{\frac{6\left[\mu_{02} + \mu_{20} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}\right]}{\mu_{00}}}
\end{aligned}
\tag{5.13}
$$

The centroid $(\bar{x}, \bar{y})$ and orientation θ of the rectangle are estimated as follows:

$$
\bar{x} = \frac{m_{10}}{m_{00}}, \qquad \bar{y} = \frac{m_{01}}{m_{00}}, \qquad \theta = \frac{1}{2}\tan^{-1}\frac{2\mu_{11}}{2\mu_{20} - 2\mu_{02}}
\tag{5.14}
$$

where μ_{pq} and m_{pq} are respectively the central moments and the raw moment of the region [Prokop 1992].

Rectangularity is then measured as the normalised discrepancies between the areas of the rectangle and the region. More precisely, given the following area: A_1 the difference between the rectangle and the region, A_2 the difference between the region and the rectangle, and A_3 the rectangle's area. The rectangularity score is computed as follows:

$$
R_D = \frac{A_1 + A_2}{A_3}
\tag{5.15}
$$

We test these two rectangularity measures for a building with protrusion (as shown in Fig. 5.13). The obtained rectangular score is $R_B = 0.59$ for the bounding box method and $R_D = 0.67$ for the discrepancy method. The discrepancy measure method ranks this building more rectangular than the standard MBR method. As a result, it is less sensitive to protrusions from the region than the MBR method and will be used in the next to measure the rectangularity of a shape. For the building with instrusion (as shown in Fig. 5.14), we found that these two rectangularity measure are not sensitive to intrusion. The obtained rectangular score for both bounding box measure and discrepancy measure are respectively 0.83 and 0.85. In the next, we choose the discrepancy measure as the rectangularity measure to accomplish region merging.
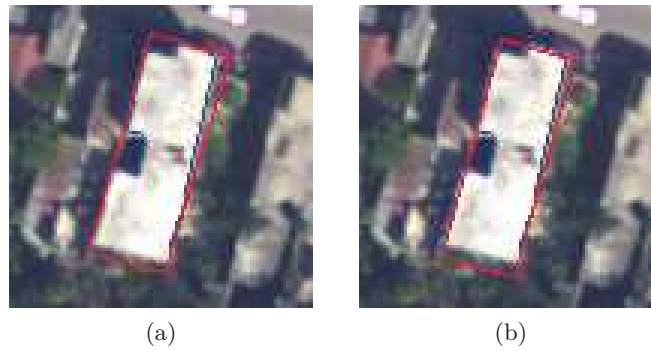
Figure 5.14: Building with intrusion artifact: (a) Building region (whose boundary is delineated by blue line) and its minimum bounding rectangle (red line). (b) Building region and the fitted rectangle estimated by discrepancy method.

5.5.3.3 Region merging procedure

Since merging is only done between regions with the same class label, we deal with each cluster (group of connected regions having the same class label) independently. Some notations are defined as follows. For each cluster, BdS denotes the list of building segments that are not "visited" (not merged with its neighbors) in cluster. We define a possible merging as a group of connected regions that includes at least one building segment. $\mathcal{L}p$ denotes the list of possible merging. T_R is the predefined minimum rectangularity degree. The merging procedure is described in Algorithm 3. The criteria of merging is to merge building segments with their neighboring regions while increasing the rectangularity degree of building segments. After merging, the new building segments are obtained and used in the next iteration. An example of the result of merging process is shown in Fig. 5.15.e, in which some cyan lines that separate the regions are disappeared.

Algorithm 3: Merging procedure for each cluster

Input:
- BdS: list of building segments that are not "visited" in cluster.
- $\mathcal{L}p$: list of possible merging in cluster.
- $\mathcal{L}$: list of regions to be merged. For initialization, $\mathcal{L} \leftarrow \emptyset$.

while BdS $\neq \emptyset$ **do**
  1. $i^* \leftarrow \arg\max_i R_D[\mathcal{L}p(i)]$.
  /* best of possible merging */
  if $R_D[\mathcal{L}p(i^*)] \geq T_R$ **then**
    1. Bd $= \mathcal{L}p(i^*) \cap$ BdS.
    /* building segments in $\mathcal{L}p(i^*)$ */
    2. **if** $R_D[\mathcal{L}p(i^*)] \geq \max(R_D[\text{Bd}])$ **then**
      $\lfloor$ Add $\mathcal{L}p(i^*)$ to $\mathcal{L}$.
    3. Update: BdS $=$ BdS $\setminus$ Bd.
  else
    $\lfloor$ break ;

Merge regions in $\mathcal{L}$.

Output: new building segments

| (a) Oversegmentation | (b) Building segments | (c) Region classification | (d) After merging |
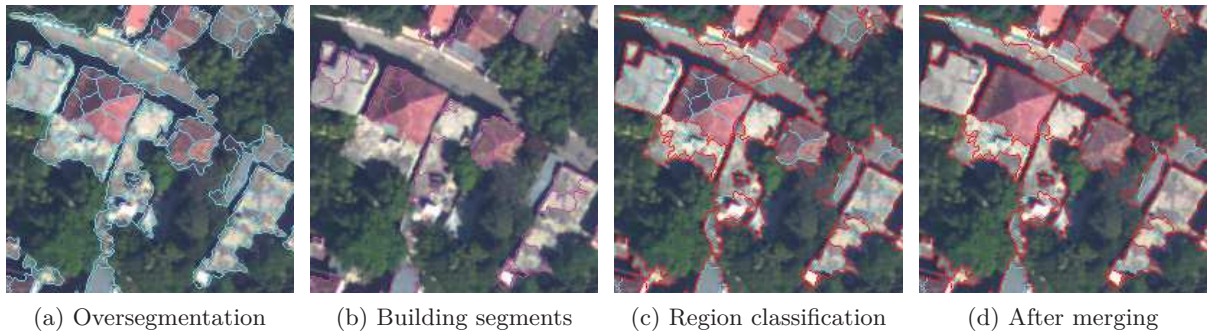
Figure 5.15: An example of region merging. Building segments are delineated by violet lines. Regions are separated by cyan lines and clusters are separated by red lines.

5.5.4   Iterative Classification and Merging

The main idea of our iterative classification-merging in respect to building detection is to increase the rectangularity degree R_D of building segments by merging them with their neighboring regions while the region classification helps to avoid merging between parts of different objects (e.g. two adjoining rectangle buildings). In the end of this iterative process, the building segments are expected to converge toward their full building object outlines.

| (a) Input image | (b) 1st classification | (c) 1st merging | (d) 2nd classification | (e) 2nd merging |
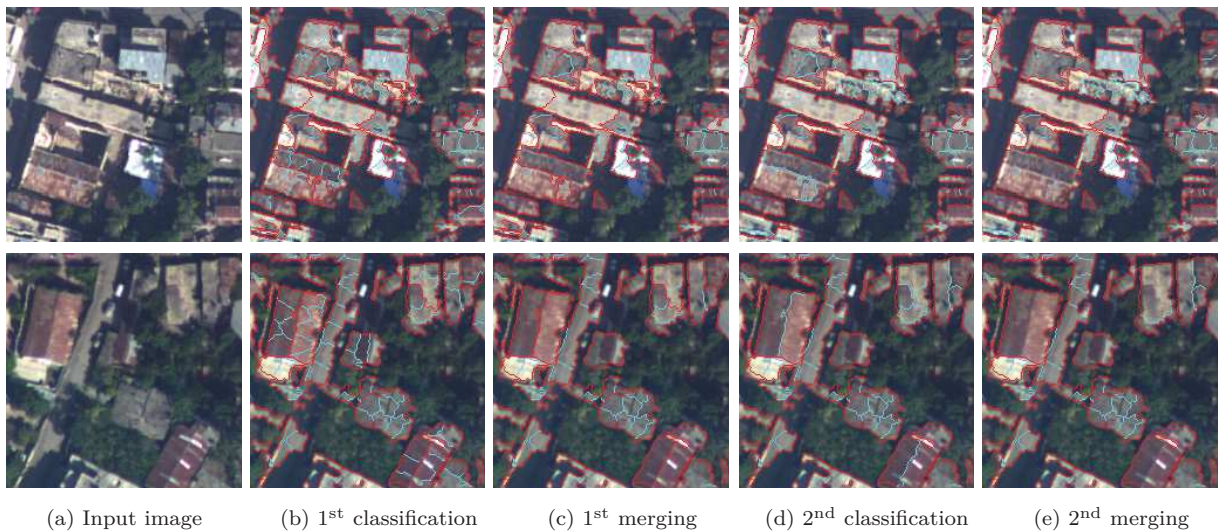
Figure 5.16: Examples of the first two iterations.

After merging, the RAG is updated. The graph node labels do not change since the previous merging procedure is only done between regions having the same class label. As a result, the feature model class statistics ($\boldsymbol{\mu}$ and Σ) do not change. The parameter β of prior term is kept constant. The ICM process continues with this new RAG topology to search for its new suboptimal solution of MRF energy minimization. The merging procedure starts when the

ICM stops, and so on. This iterative segmentation merging ends when there is no merging happens.

Fig. 5.16 illustrates the first two iterations of the iterative classification-merging process. SLIC over-segmentation decomposes image into different regions (boundaries between different regions are delineated using cyan lines) and the MRF-based region classification group these regions into different clusters (boundaries between clusters are delineated using red lines), as shown in Fig. 5.16b. Within each cluster, region merging procedure merge regions to generate rectangular shapes (Fig. 5.16c, some cyan lines are disappeared). The ICM continues with this new configuration. When the ICM stops (as illustrated in Fig. 5.16d, some red lines become cyan), the merging procedure merges region to increase the rectangularity degree of building segments (Fig. 5.16e).

5.6 Determination of Final Building Regions

The proposed iterative region growing results in a segmentation map, in which the building segments have the best possible rectangular score. But it is not evident that these building segments are the final buildings. First, because of the strict constraint of merging, a building segment whose form is very close to a perfect rectangle can not to be merged with its neighboring regions (the other parts of the same building), as shown in Fig. 5.17a. Second, we focus on the buildings with right-angled corners, which are characterized as a collection of rectangles (e.g: L-shape, U-shape, T-shape, gable roofs, and more complex building shapes which are combinations of the aforementioned shapes), not only the rectangular buildings. These types of buildings are partitioned into different part in the segmentation map, as shown in Fig. 5.17b-c. In this section, we describe how to determine the final building regions from the results of region growing image segmentation.

(a)   (b)   (c)
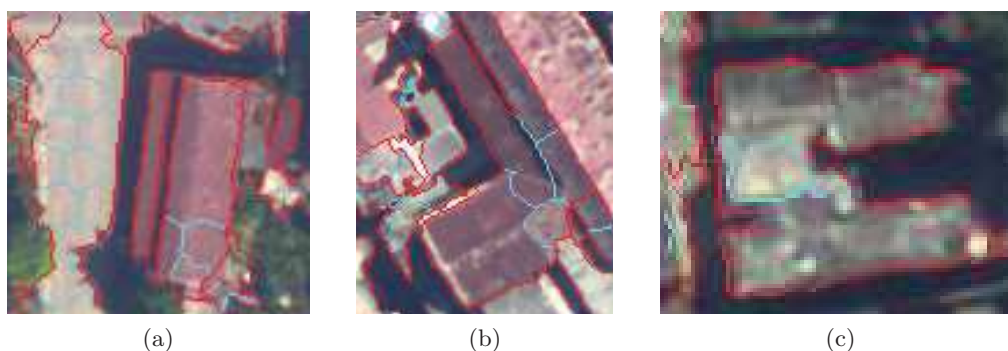
Figure 5.17: Results of region growing image segmentation: (a) a rectangular building can not achieve its final shape, (b) and (c) L-shaped and U-shaped buildings are partitioned into different regions.

5.6.1 Recursive MBR

In a recent paper [Kwak 2014], the authors introduce a new framework for automatically detecting and reconstructing accurate right-angle corner building models from LiDAR data. In their approach, building boundaries are first approximated by LiDAR data and then regularized using rectangular model through a Recursive Minimum Bounding Rectangle (RMBR) process. Given a boundary shape, we apply the RMBR algorithm in the way as follows:

1. Apply the MBR algorithm to the boundary shape and the generated MBR is denoted as $\text{MBR}_{(1)}$, as illustrated in Fig. 5.18b and 5.19b.

2. Track the non-overlapping boundary segments with $\text{MBR}_{(1)}$ (Fig. 5.18c and 5.19c).

3. If the length of the detected non-overlapping segments is larger than 20 pixels, apply MBR algorithm again on the non-overlapping segments and their projection onto the MBR sides (Fig. 5.18d and 5.19d) to derive $\text{MBR}_{(2)}$ (Fig. 5.18e and 5.19e).

4. Track the non-overlapping boundary segments with $\text{MBR}_{(2)}$.

5. If the length of the detected non-overlapping segments is larger than 20 pixels, apply MBR algorithm again on the non-overlapping segments and their projection onto the MBR sides to derive $\text{MBR}_{(3)}$.

In our method, we limit the level of recursive MBR to 3. The final shape, denoted as RMBR, is determined as follows:

$$\text{RMBR} = \sum_{k=1}^{3}(-1)^{k+1}\text{MBR}_{(k)} \tag{5.16}$$

Note that if the condition in step 3 (step 5) is not satisfied, we have $\text{MBR}_{(2)}$ ($\text{MBR}_{(3)}$) is an empty set. Based on this notation, Fig. 5.18f and 5.19f show a RMBR with 2 levels of details ($\text{RMBR} = \text{MBR}_{(1)} - \text{MBR}_{(2)}$).

5.6.2   Procedure of Determining Final Building Regions

For the sake of simplicity, we denote RMBR as an operator to measure how much the object considered differs from its RMBR approximation. This operator is defined as follows:

$$\text{RMBR} = \frac{\text{Area of object}}{\text{Area of RMBR}} \tag{5.17}$$

Similar to the merging algorithm (Algorithm 3), the determination of final buildings is done within the limit of each cluster, obtained from the results of iterative classification-merging process. The main idea is to check the possibility of having a building with right-angled corners in the cluster. For each cluster, we denote a candidate building as a group of connected regions that includes all building segments. Then, we verify if the candidate building is a building with right-angled corners by measuring its RMBR score. The procedure of determining final building regions is described in Algorithm 4. An example of determining final building regions from the image segmentation results is shown in Fig. 5.20.
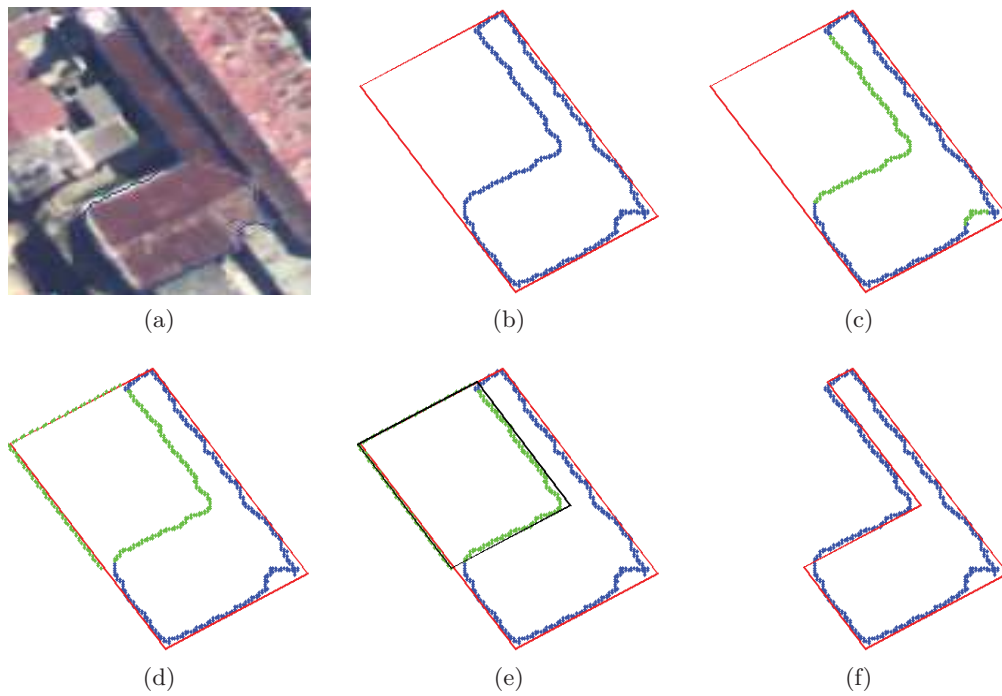
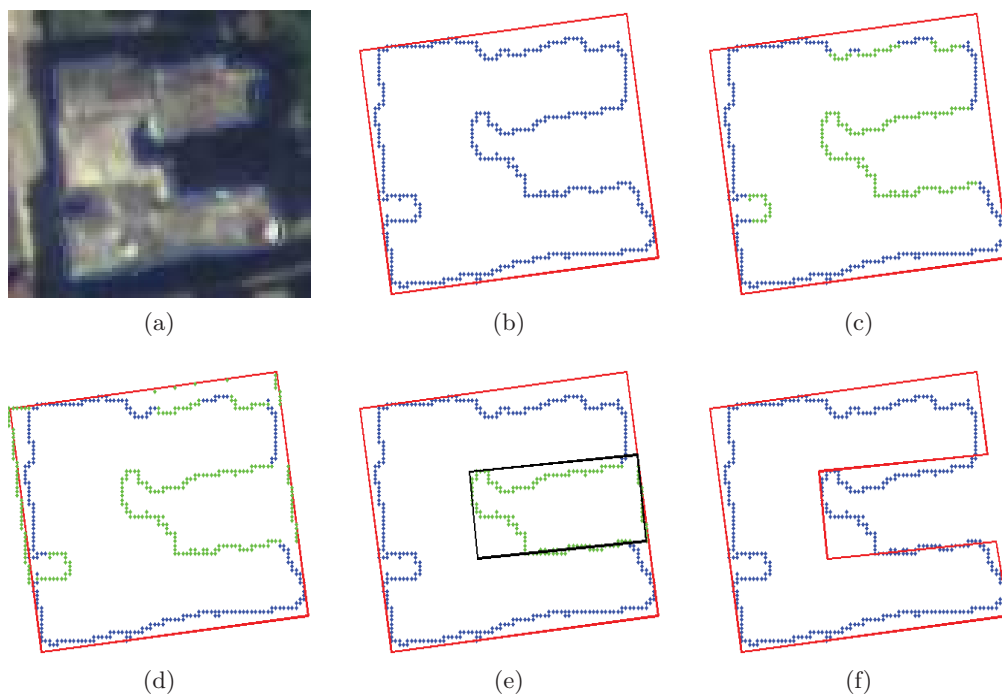Figure 5.18: The step-by-step illustration of L-shaped bounding box estimation

Figure 5.19: The step-by-step illustration of U-shaped bounding box estimation

---

Algorithm 4: Determining the final building regions in a cluster

Input:

- T_B: minimum RMBR-score.

- BdS: list of building segments in cluster.

- N: number of regions in cluster.

for $k \leftarrow N$ **to** $|\text{BdS}|$ **do**

    1. Determine $\mathcal{L}_k$: list of candidate buildings merged from k regions. $\mathcal{L}_k(j)$ is the j-th element in $\mathcal{L}_k$.

    2. $j^* \leftarrow \arg\max\limits_{j} \text{RMBR}[\mathcal{L}_k(j)]$.

    /* best of candidate building */

    if $\text{RMBR}[\mathcal{L}_k(j^*)] \geq T_B$ **then**

        1. Merge regions in $\mathcal{L}_k(j^*)$ and get the final building region of cluster.

        2. Break ;

Output: final building region

(a) Input image   (b) Segmentation result   (c) Final building
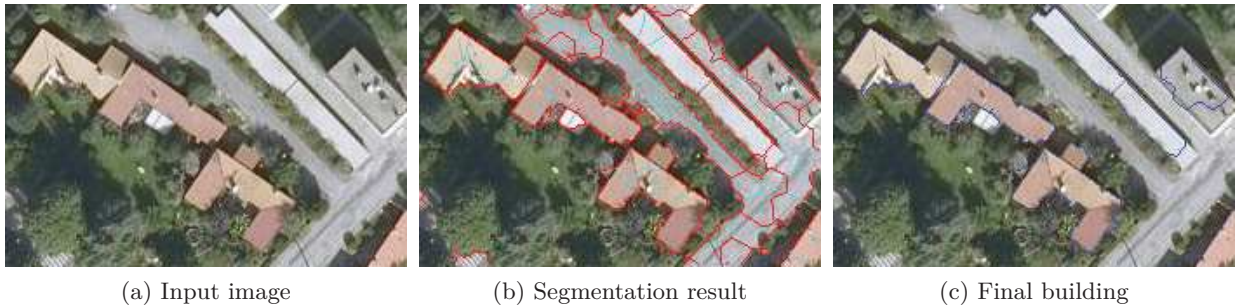
Figure 5.20: Determination of final buildings from the region growing image segmentation result: 5 buildings are detected, in which 3 buildings (left) are approximated by the recursive MBR of level 2 and other 2 buildings (right) are rectangular. These detected buildings have RMBR-scores of 0.87, 0.91, 0.84, 0.93, 0.81 respectively (from left to right). The value of T_B is chosen as 0.8.

5.7   Conclusions

In this chapter, we review some existing building detection methods that are related to the proposed method. The methodology begins with the detection and post-processing of shadow areas. For shadow detection, we employ our shadow/vegetation detection method presented in Chapter 4, that allows to divide the image into three distinct classes: *shadow*, *vegetation*, and *others* with good precision. The boundaries between shadows and their corresponding buildings are detected by eliminating shadow regions generated by vegetation objects and other non-building objects from shadow mask. A novel region-growing image segmentation algorithm is then proposed. This segmentation algorithm consists of two main stages: a MRF- based region classification, and a region merging procedure that exploits the position of shadows and the rectangularity measure. Final buildings are extracted from the obtained iterative classification-merging result. The evaluation of the proposed method will be presented in the next chapter.

Result Evaluation

Contents

6.1 Introduction

This chapter presents the experimental evaluation of the proposed building detection method. This method will be denoted as SBBD, that stands for shadow-based building detection. The strategy for accuracy assessment is first introduced. Then, we analyse the sensitivity of important parameters used in SBBD method, which will be tested on the Jacmel dataset and the SZTAKI-INRIA Building Detection Benchmark [Benedek 2012]. At last, we discuss the advantages and the limitations of SBBD method.

6.2 Strategy for Accuracy Assessment

The final performance of the proposed approach is assessed by comparing the results of the proposed approach with the ground truth. We perform quantitative evaluation both at object and pixel levels. The buildings that are partially visible at image boundaries are also included in the analysis. In this work, we use the same metrics and accuracy table from [Ozgun Ok 2013]. The

common measures of precision (i.e., P), recall (i.e., R), and the F_1-score (i.e., F_1) are used to measure pixel level accuracy, where:

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN} \qquad (6.1)$$
$$F_1 = \frac{2 \times P \times R}{P + R}$$

Here, TP stands for true positives and refers to the number of pixels assigned as rooftop in both ground truth and segmentation result. FP stands for false positives and refers to the number of pixels assigned as building in result but not in the ground truth. FN stands for false negatives and refers to the number of pixels assigned as building in ground truth but not in result. Among these metrics, the producer's accuracies (also known as recall) indicate how well pixels of known categories are correctly classified. The user's accuracies (also known as precision) indicate the probabilities of pixels been correctly classified into actual categories on the ground truth. The F-score (F_1) captures both precision and recall as single metric that gives each equal importance.

On the other hand, the object-based performance can also be evaluated with the measures given in Eq. 6.1, similar to [Ok 2013] . An output building object is labeled as TP if it has at least a 60% pixel overlap ratio with a building object in the ground truth. However, we label an output building object as FP if it does not coincide with any of the building objects in the ground truth, and we label an output building object as FN if it corresponds to a reference object with a limited amount of overlap ($< 60\%$). Thus, it is possible to compute Precision (P), Recall (R) and F_1-score (F_1) for object-based performance measurement.

6.3   Parameters

To initialise SBBD method, different parameters are required. However, there are no parameter setting that can be applicable for all images since these images have different building sizes, different shadow density, etc. Table 6.1 lists the default settings of the parameters used in SBBD method for NOAA aerial image of Jacmel dataset.

Some parameters have physical meaning and can be intuitively defined. Otherwise, other parameters need to be determined by test and trial. Fig. 6.2 illustrates the effects of choosing different values for detecting building rooftops. We performed a large number of tests on different parameters and investigated the effects of each parameter on the detection performance using the quality measures given in Eq. 6.1.

6.3.1   Detection of Shadows cast by Buildings

Two parameters (l_{se}, d_{sh}) are required for post-processing of the building-shadow boundary mask M_{SB} (c.f. Section 5.4). The first parameter l_{se} is the length of structuring element of the directional binary morphological dilation. This parameter is used to define the size of the

Step	Parameter	Value
Detection of Shadows Cast by Buildings	l_{se}	60
(cf. Section 5.4.2)	d_{sh}	5m
Oversegmentation SLIC [Achanta 2012]	η_{sup}	175
	m	10
	K	12
ICM optimization - Region Classification	β	150
	τ_{max}	500
	ε	0.01
Region Merging	T_S	15
	T_R	0.65
Determination of Final Building	T_B	0.8

Table 6.1: Parameter settings of the proposed approach for NOAA aerial image

expansion region, starting from each vegetation object. If there is more than one shadow regions occur in this expansion region, we select the shadow region that have a border with vegetation object. That is why parameter l_{se} is not important and just need to be set large enough. In our algorithm, we choose l_{se} as 60 pixels.

The second parameter is the minimum length of the boundary between building objects and their corresponding shadows d_{sh}. 10 different values of d_{sh} (2.5 meters - 7 meters) are tested. The effects of these values on the computed performances are shown in Fig. 6.2a. The results indicate improvements as the d_{sh} values increase for pixel-based precision ratios, whereas at the same time, performance decreases are observed for the recall ratios. When d_{sh} gets low values, shadows from non-building objects are also detected. The non-building objects having the rectangular form are detected and considered as buildings. As a result, the FP pixels are high and the precision ratio is therefore low. Conversely, when d_{sh} is high, we do not detect some building shadows and therefore miss some buildings, the FN pixels are high and the recall ratio is low. In view of the computed performance measures, we select the value of d_{sh} as 5 meters (that corresponds to 21 pixels for NOAA aerial image of 24cm resolution), which maximizes the F_1-score computed for the pixel-based case.

6.3.2   Oversegmentation SLIC

Since oversegmentation SLIC [Achanta 2012] is considered as a preprocessing step of the region classification task, it is important to reduce error propagation from this step to the final building validation. In this step, the parameter m controls the tradeoff between superpixel compactness and boundary adherence. The greater the value of m, the more spatial proximity is emphasized and the more compact the cluster. In our algorithm, m is set as 10, as recommended by the authors, that allows to regularize regular-sized regions with good boundary adherence. The sole parameter of SLIC is the number of desired superpixels. This parameter is set so that the initial superpixel size is η_{sup}. The effects of different values of η_{sup} are shown in Fig. 6.2b. Both precision and recall ratios do not significantly change for the low values of η_{sup}. However, when η_{sup} is high, the number of FP pixels and FN pixels increase since the oversegmentation does not adhere well to

building boundaries. An example of oversegmentation SLIC with different values of η_{sup} is shown in Fig. 6.1. If the initial superpixel size η_{sup} is low (Fig. 6.1a), the image is segmented into more regions and we can not benefit from the interest of oversegmentation. The number of superpixels increase and therefore the computation time increases. However, the image is under-segmented (Fig. 6.1c, red circle) when η_{sup} is high. Since the oversegmentation SLIC is fast, the good choice of this parameter can be obtained by varying its value on some test images and observing the oversegmentation results. In our algorithm, η_{sup} is set to 175, which maximizes the F_1-score as shown in Fig 6.2b and can sufficiently preserve the boundaries of objects and structures without under-segmentation errors in natural images.

(a) $\eta_{sup} = 125$ (b) $\eta_{sup} = 175$ (c) $\eta_{sup} = 225$
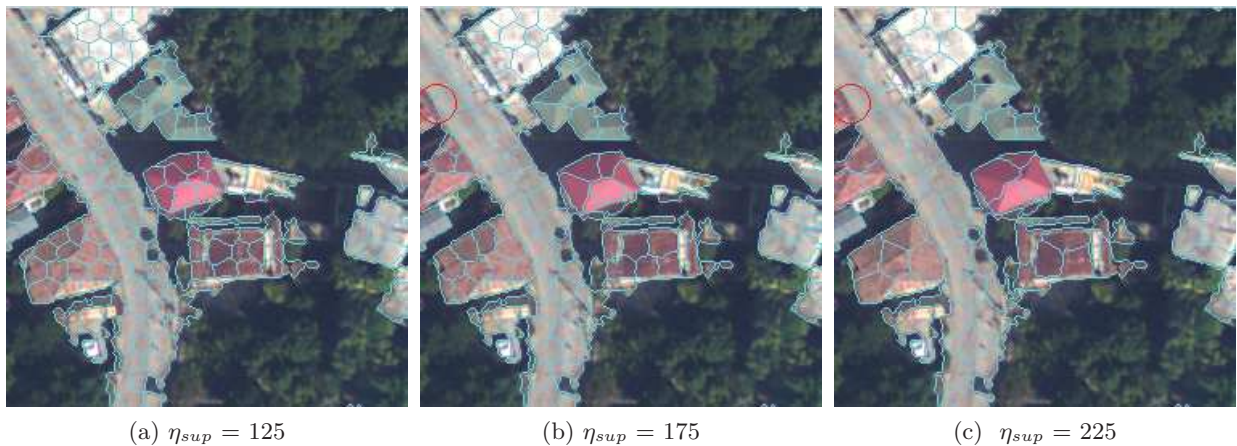
Figure 6.1: An example of oversegmentation SLIC with different values of η_{sup}. Low value of η_{sup} produces lots of regions and high value of η_{sup} makes the oversegmentation not adhere well to building boundaries.

6.3.3   Region Classification

In this study, one major task is to consider the parameters used for region classification. The two most important parameters are the number of classes K and the coefficient of regularization β. To select K, we first ignore the MRF regularization model. It means that the region growing image segmentation is reduced as the initialization of region classification followed by a merging process. For initialization, the iterative procedure is terminated once the ratio of the absolute energy difference between consecutive iterations over the energy at the former iteration is below 10^{-6} or the number of iterations exceeds 100 to ensure that we obtain the best initialization as possible. The effects of different values of K are shown in Fig. 6.2c. When K is small, we can not distinguish building regions and their neighboring soil. As a result, the number of FP pixels is high and the precision ratio is low. Conversely, when K is high, building is segmented into two or more regions. Only the region neighboring shadows is detected as building. As a result, the number of TP and FN pixels are both low. As shown in Fig. 6.2c, we choose K as 12 for the best performance.

The parameters of ICM algorithm are chosen as follows. Maximum number of iterations τ_{max} is set to 500, which is normally not exhausted since the early stopping criterion is met with the stopping criteria ε is set to 0.01. The effect of different values of the coefficient parameter β are shown in Fig. 6.2d. The general rule for selecting β is that it should be set to a large value for

Figure 6.2: Pixel-based performance in the case of different parameter settings. In each plot, the nonvarying coefficients are kept at their optimal settings (Table 6.1).
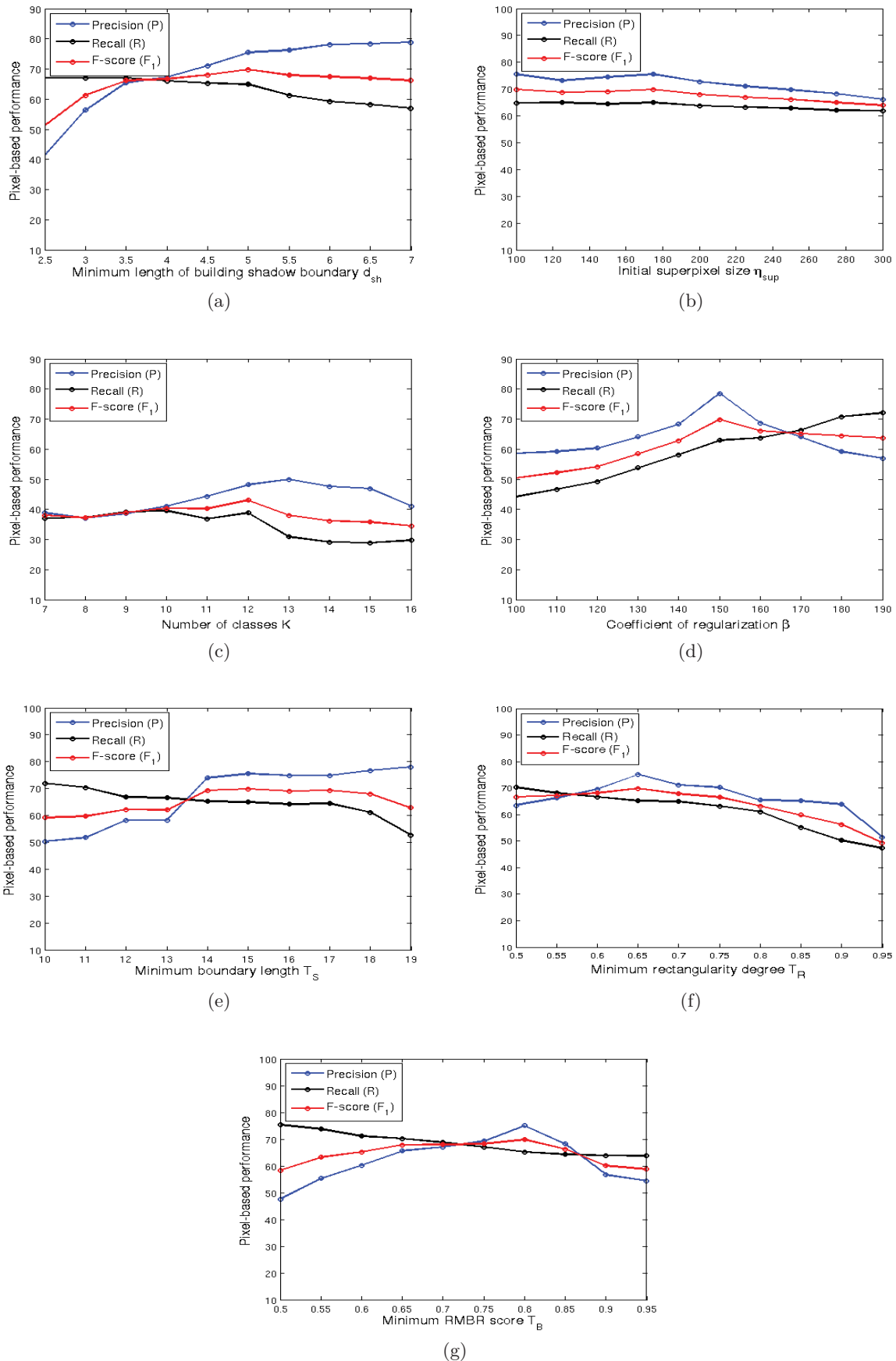
simple scenes and small for complex scenes. The results indicate improvements as β increases for both the precision and recall ratio. But when β is high, the neighboring regions of building are classified the same class label as building. This causes more FP pixels and the precision ratio decreases. β is therefore chosen as 150, which maximizes the F_1-score computed for the pixel-based case.

6.3.4   Region Merging

Two parameters are required for the region merging step of the iterative region growing image segmentation. The first parameter is the minimum length of boundary T_S between region (superpixel), produced by the oversegmentation, and the building-shadow boundaries M_{SB} to determine what region is building segment. The effects of different values of T_S on the computed performances are shown in Fig. 6.2e. The results indicate that low values of T_S significantly reduce the computed precision ratio since FP pixels are high. The false positive building segments cause the false positive buildings. The larger values of T_S do not significantly change the results since if we can detect one building segment, we can detect the completed building. We found that the optimal value of T_S is 15 pixels.

The second parameter is the minimum rectangularity degree T_R. The effects of different values of T_R are shown in Fig. 6.2f. The results indicate performance decrease are observed for both the precision and recall ratios as the T_R values increases. Indeed, large values of T_R prevent the merging and therefore reduce the number of TP pixels and increase the number of FN pixels. It should be noted that this parameter take effects for only the first iteration of the region growing algorithm. Then, the merging is decided in order to increase the rectangularity degree of building segments. In view of the computed performance measures, we select the value of T_R as 0.65, which maximizes the F_1-score computed for the pixel-based case.

6.3.5   Final Determination of Buildings

For the determination of final building regions from the segmentation results, the minimum RMBR-score T_B is required. The effects of different values of T_B are shown in Fig. 6.2g. For the low value of T_B, we detect the maximum number of TP pixels, minimum number of FN pixels, but get lots of FP pixels, therefore, the precision ratio is low and the recall ration achieves the maximum. Conversely, if T_B is high, some buildings will be missed. As a result, all three measures are low. Thus, we select the value of T_B as 0.8, which maximizes the F_1-score computed for the pixel-based case.

6.4   Experiments

6.4.1   NOAA Aerial Image

We first compare our SBBD method with the method proposed by [Femiani 2014]. The authors have processed their algorithm on our NOAA aerial images. An example is shown in Fig. 6.3, in which true positives (TP) are tinted green, false negatives (FN) are tinted red, and false positives (FP) are tinted blue. Table 6.2 shows precision, recall and F_1-score for three test images shown in Figs. 6.3, 6.4, 6.5 based on manually entered ground truth. For the image in Fig. 6.3, SBBD method gives a highly competitive result with precision 74.0%, recall 61.9% and F_1-score

67.4% compared with precision 60.7%, recall 62.6% and F_1-score 61.6% concerning pixel-level performance.

(a) Input image   (b) Method of [Femiani 2014]   (c) SBBD method

Figure 6.3: Comparisons with method in [Femiani 2014] on NOAA aerial images of Jacmel dataset.

Concerning object-level performance, the method of [Femiani 2014] produces more false positives (small red patches) than SBBD method and therefore have lower object-level performance. These false positives are due to the bad performance of their shadow detection algorithm. If we remove all false negative objects that are bigger than 80 pixels, the new precision, recall and F_1-score for object-level performance are respectively 17.0%, 11.3% and 13.6%. Regarding SBBD method, the false positives (tinted blue) are due to the missing shadows (blue circle). Besides, in red circle, a long wall generates a long shadow patch, which can not be eliminated in the detection of building shadow step, and cause a false positive patch.

Image	Pixel level performance						Object level performance					
	[Femiani 2014]			SBBD method			[Femiani 2014]			SBBD method		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Fig. 6.3	60.7	62.6	61.6	74.0	61.9	67.4	6.2	12.8	8.3	63.8	56.6	60.0
Fig. 6.4	38.7	66.3	48.8	61.7	58.0	59.8	2.5	10.9	4.1	67.0	75.3	70.9
Fig. 6.5	57.2	65.2	60.9	72.0	71.4	71.7	12.7	51.9	20.5	62.8	57.4	60.0

Table 6.2: Numerical pixel-level and object-level building detection evaluation of applying the method of [Femiani 2014] and our method on Jacmel aerial images, shown in Figs. 6.3, 6.4, and 6.5.

Fig. 6.4 shows the building detection results of [Femiani 2014]'s method and SBBD method on rural areas. As shown in Table 6.2, concerning pixel-level performance, SBBD method has better results with precision 61.7%, recall 58.0% and F_1-score 59.8% compared with precision 38.7%, recall 66.3% and F_1-score 48.8%. In this high vegetation density area, [Femiani 2014]'s method produces lots of false positives, that make their method very low in object-level performance.

(a) Input image
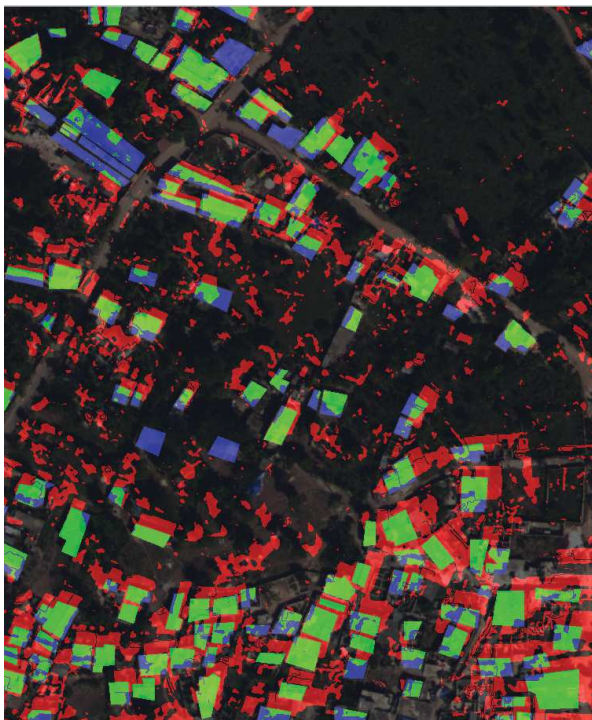
(b) Method of [Femiani 2014]

(c) SBBD method

Figure 6.4: Comparisons with method in [Femiani 2014] on NOAA aerial images of Jacmel dataset.

Regarding SBBD method, lots of buildings are dismissed because their shadows can not be detected. In particular, in the high dense building area (half below of the image), [Femiani 2014]'s method detects mostly all buildings, whereas SBBD method dismiss some buildings.

(a) Input image
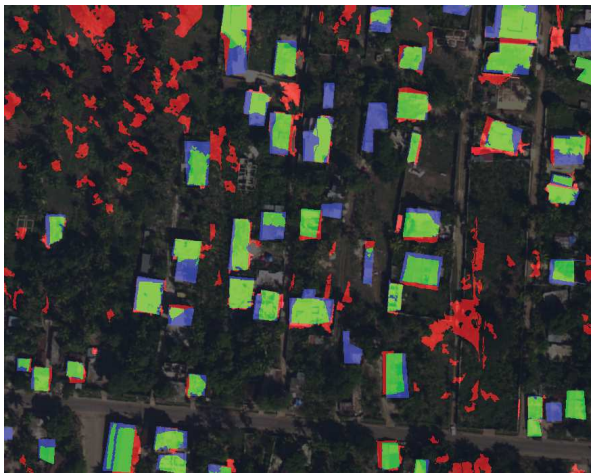
(b) Method of [Femiani 2014]
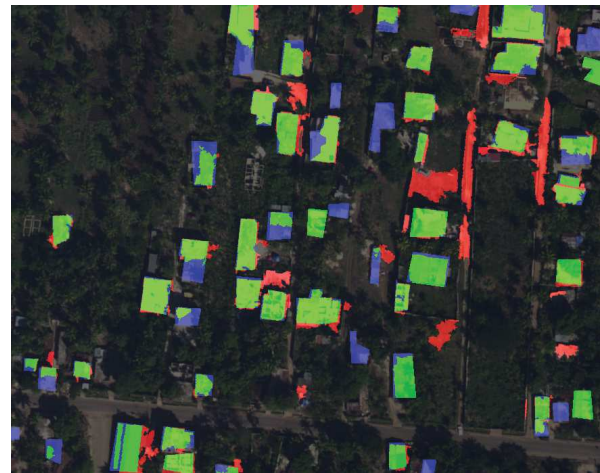
(c) SBBD method

Figure 6.5: Comparisons with method in [Femiani 2014] on NOAA aerial images of Jacmel dataset.

An another example is shown in Fig. 6.5. Similar conclusions can be drawn for the method of [Femiani 2014] (many false positive patch). SBBD method detects most buildings but some false positives are produced (red circle), because there are the soils that are higher than others and generate some shadow regions. These shadow regions make our method classify these soils as buildings.

6.4.2 Worldview-2 Satellite Image

We also compare SBBD method with the automated rooftop detection method proposed by [Ozgun Ok 2013] on Worldview-2 images of Jacmel dataset as shown in Figs. 6.6 and 6.7. The Worldview-2 satellite images have four bands: red, green, blue, and NIR, which allows Ok's

method to be compared.

(a) Input image   (b) Method of [Ozgun Ok 2013]   (c) SBBD method

Figure 6.6: Comparisons with method in [Ozgun Ok 2013] on Worldview-2 images of Jacmel dataset.

(a) Input image   (b) Method of [Ozgun Ok 2013]   (c) SBBD method
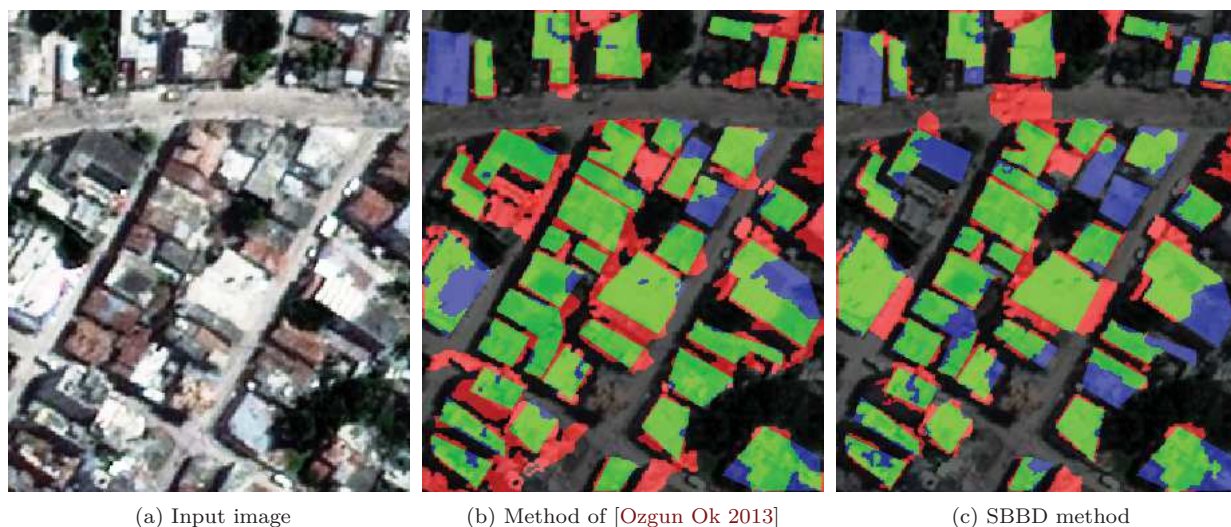
Figure 6.7: Comparisons with method in [Ozgun Ok 2013] on Worldview-2 images of Jacmel dataset.

For the image in Fig. 6.6, we visually observe that Ok's method produces more false positive pixels but less false negatives pixels than SBBD method. Concerning pixel-level performance, as shown in Table 6.3, Ok's method is slightly better than ours with precision 67.3%, recall 80.0% and F_1-score 73.1% compared with precision 76.5%, recall 62.5% and F_1-score 68.8%. However, concerning object-level performance, SBBD method is better than Ok's method (F_1-score 73.5% for Ok's method and F_1-score 83.8% for SBBD method). An another example is given in Fig. 6.7. As shown in Table 6.3, both methods have equally high F_1-scores for pixel- and object-level

performance. Visually, we observe that most of buildings are partially or fully detected.

Image	Pixel level performance						Object level performance					
	[Ozgun Ok 2013]			SBBD method			[Ozgun Ok 2013]			SBBD method		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Fig. 6.6	67.3	80.0	73.1	76.5	62.5	68.8	78.4	69.2	73.5	93.5	76.3	84.1
Fig. 6.7	63.0	81.8	71.2	71.3	71.6	71.4	78.1	73.2	75.5	86.3	82.6	84.4

Table 6.3: Numerical pixel-level and object-level building detection evaluation of applying the method of [Ozgun Ok 2013] and our method on Worldview-2 images of Jacmel dataset, shown in Figs. 6.6 and 6.7.

6.4.3 SZTAKI-INRIA Building Detection Benchmark

Besides, we compare SBBD method with the rooftop extraction method using higher order CRF proposed by [Li 2015]. The experiments are carried out on SZTAKI-INRIA Building Detection Benchmark. The results on Manchester and Normandy images are provided by the authors.

Image	Pixel level performance						Object level performance					
	[Li 2015]			SBBD method			[Li 2015]			SBBD method		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Normandy	77.9	66.3	71.7	74.2	68.4	71.2	89.3	78.3	83.4	95.1	68.3	79.5
Manchester	81.6	66.4	73.2	78.7	69.9	74.0	89.9	79.7	84.5	97.5	86.81	91.9

Table 6.4: Numerical pixel-level and object-level building detection evaluation of applying the method of [Li 2015] and our method on SZTAKI-INRIA Building Detection Benchmark, shown in Fig. 6.8 and Fig. 6.9 .

Table 6.4 shows the quantitative evaluation of both methods. On the Normandy site, the method of [Li 2015] gives a slightly better results with precision 77.9%, recall 66.3% and F_1-score 71.7% compared with precision 74.2%, recall 68.4% and F_1-score 71.2% concerning pixel-level performance. This method also detects more buildings than SBBD method. SBBD method dismiss a lot of dark building rooftops, which are classified as shows by shadow/vegetation detection method (an example is shown in Fig. 6.10).

On the Manchester site, SBBD method surpasses the method of [Li 2015] (precision 78.7%, recall 69.9% and F_1-score 74.0% compared with precision 81.6%, recall 66.4% and F_1-score 73.2%) in term of pixel-level evaluation. SBBD method detect most of buildings on the site (precision 97.5%, recall 86.81% and F_1-score 91.9% compared with precision 89.9%, recall 79.7% and F_1-score 84.5% in term of object-level evaluation) than the method of [Li 2015].
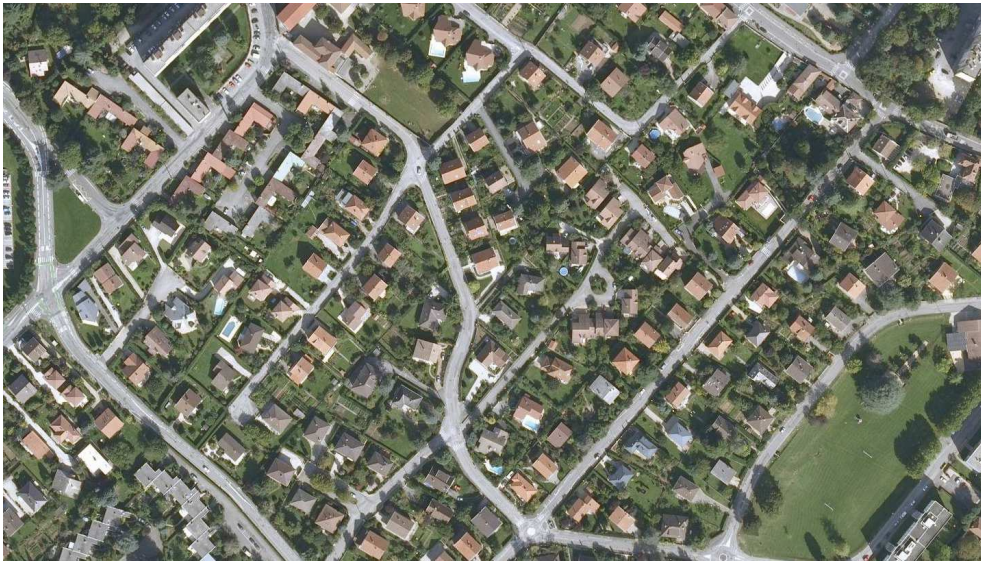
(a) Input image

(b) Method of [Li 2015]

(c) SBBD method

Figure 6.8: Comparisons with method in [Li 2015] on Normandy image. True positives are tinted green, false negatives are tinted red, and false positives are tinted blue.
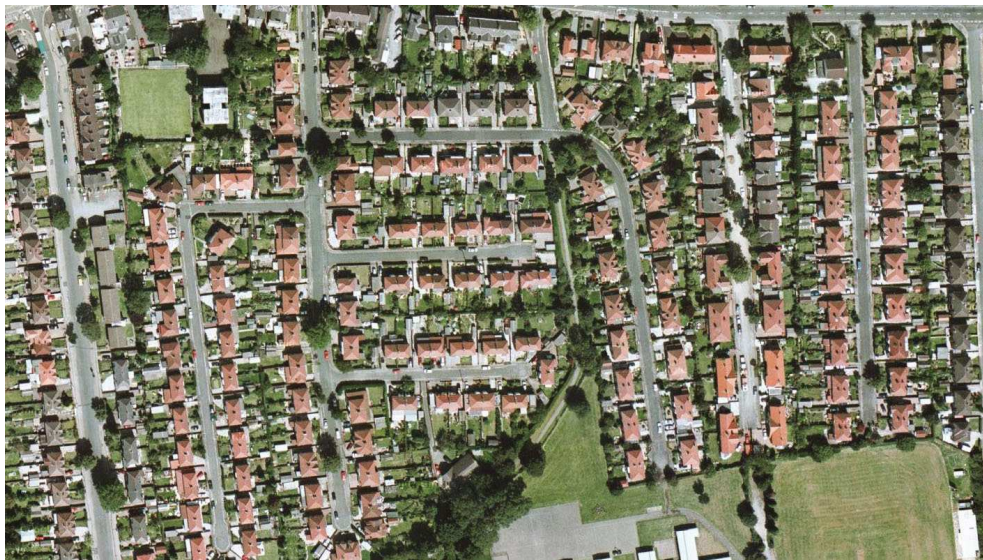
(a) Input image

(b) Method of [Li 2015]

(c) SBBD method

Figure 6.9: Comparisons with method in [Li 2015] on Manchester image. True positives are tinted green, false negatives are tinted red, and false positives are tinted blue.

(a) Image   (b) Final segmentation   (c) Detected buildings
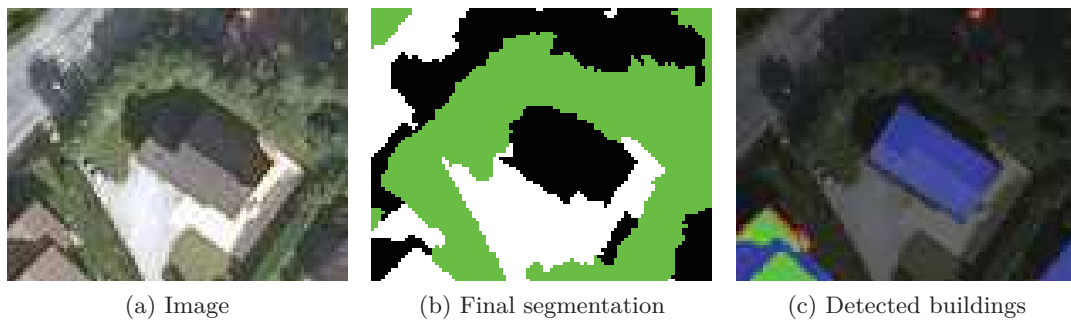
Figure 6.10: An entire face of the building rooftop is classified as shadows.

6.5   Discussion

6.5.1   Advantages

The advantages of the proposed building detection method are stated as follows:

- In an urban environment, buildings are generally dense and complex. A shadow region can be cast by a group of connected buildings. The results prove that the approach has the ability to separate these attached buildings since it takes into account the spectral features through the image segmentation stage. An example of detected building in a dense urban area is shown in Fig. 6.11.

(a) Image   (b) Final segmentation   (c) Detected buildings
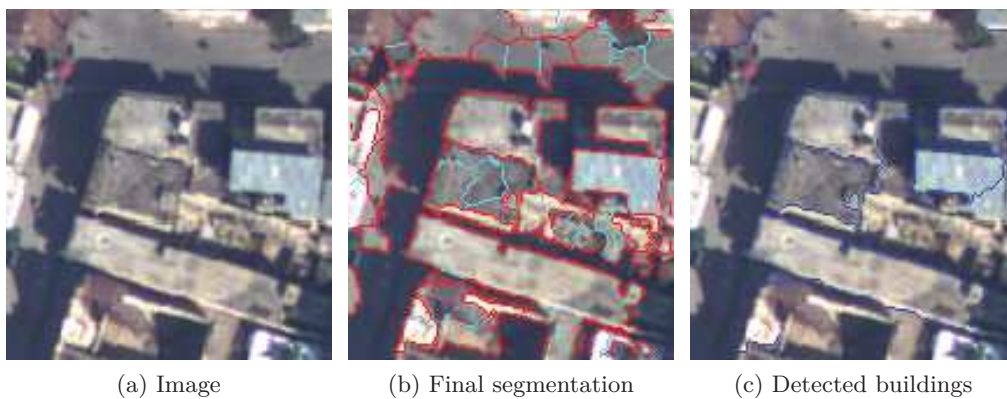
Figure 6.11: Detected building in a dense urban area. Attached buildings are separated by blue lines.

- The approach is not limited to urban areas but can also be used for building detection in rural environments (as shown in Fig. 6.5). Provided that the shadows cast by buildings are not completely occluded, the approach can recover building regions located in relatively dense vegetation canopies.

- The method requires only reasonably high resolution aerial image. Not required are multiple views, additional information such as near-infrared (NIR), lidar or any elevation data. Users must only estimate the illumination angle and supply the parameters. The determination

of these parameters is simple (as explained in Section 6.3). The method does not require a large amount of prelabeled data as would be needed by a supervised learning.

- The decision-making is robust and powerful. Since each cluster obtained from the region classification step is independently treated, large number of image data sets can be utilized and evaluated in the same manner. The parallel computing can be used. As a result, the computation time can be reduced.

6.5.2 Limitation and Failures Cases

In spite of those advantages, the proposed approach still contains some major limitations. Since shadow regions are used to support directly the building detection step, this approach can not detect buildings whose shadow is not visible or missing (as shown in Fig. 6.12), similar to the limitations exist in some state-of-the-art methods like [Femiani 2014, Li 2015, Ozgun Ok 2013]. Besides, under oblique lighting, a gabled rooftop may exhibit significantly different intensities on the sloped portions of the roof, so the proposed method may lose its efficiency as it detects only one side of the rooftop as is shown in Fig. 6.13. In an extreme case, an entire face of the rooftop may be in shadow.

(a) Image	(b) Shadow/vegetation detection	(c) Detected buildings
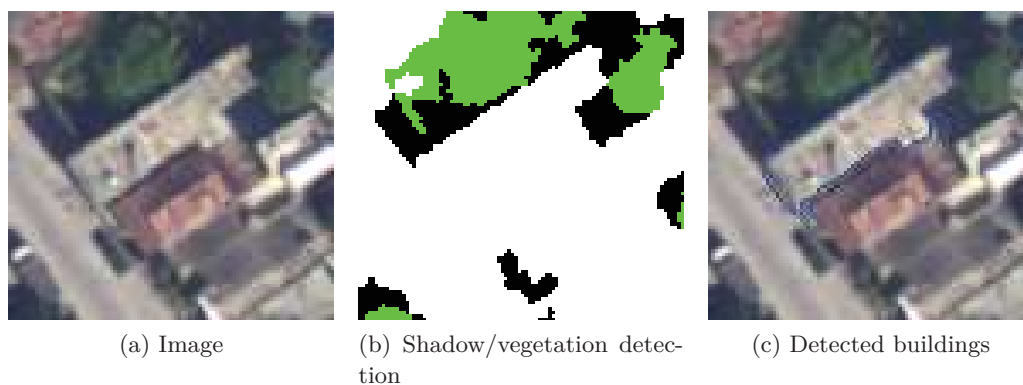
Figure 6.12: A building is not detected since its shadow can not be detected.

(a) Image	(b) Shadow/vegetation detection	(c) Detected building
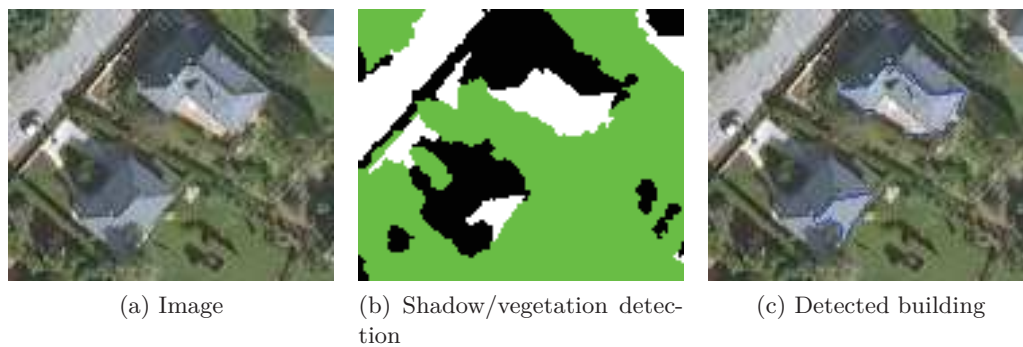
Figure 6.13: (Self-) occlusion occurs on the shadow areas.

| (a) Image | (b) Detected building | (c) Evaluation |
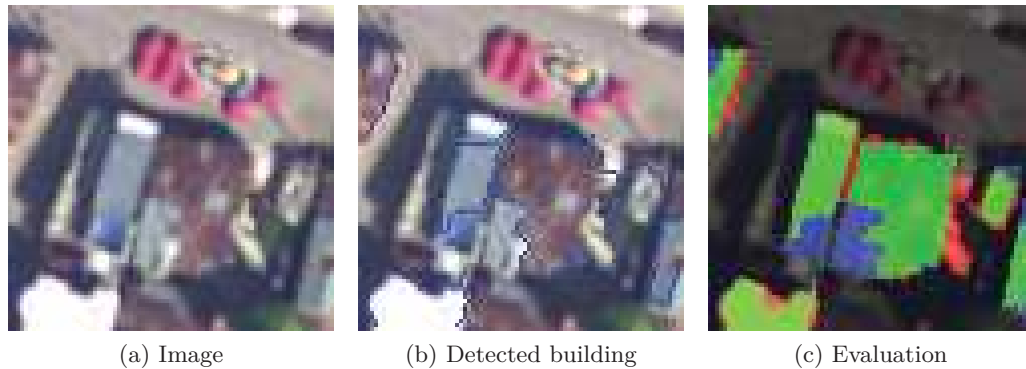
Figure 6.14: A part of building can not be recovered since building rooftop contains several components with different colors.

Our approach is designed based on the assumption that buildings have a homogeneous color. If the rooftop contains several components with different colors then our method will fail to obtain the entire rooftops. Only the homogeneous part bordering the corresponding shadow region is detected (as shown in Fig. 6.14).

6.6   Conclusions

This chapter presents the experimental evaluation of the proposed building detection method, in which both pixel-level and object-level performances are considered. The SBBD method requires some parameters to be specified by a user. In Section 6.3, the sensitivity of important parameters are analyzed. Performances performed on Jacmel dataset and the SZTAKI-INRIA Building Detection Benchmark demonstrate that SBBD method detects the building regions with a high success rate in comparison with some building detection method in the literature. Besides, the advantages and limitations of SBBD method are discussed. In general, the experiment shows that SBBD method is reliable and robust. This is the second contribution of this PhD and has been published in [Ngo 2015].

Concluding Remarks and Future Directions

Contents

    The aim of this thesis was to develop a system capable of detecting shadow/vegetation and building from single high resolution optical images. This chapter gives a review of what has been accomplished in this thesis (cf. Section 7.1). Besides, perspectives and future directions will be drawn in Section 7.2.

7.1   Conclusions

First, this thesis presents SSVD method, a novel shadow/vegetation detection method based on Dempster-Shafer fusion and Markov random fields modelling. Current shadow/vegetation detection methods in the literature detect separately shadow regions and vegetation regions. In some cases, these methods might not provide a sufficiently good segmentation map (shadow, vegetation, other). For example, a vegetated pixel covered by shadow can be classified as vegetation (by a vegetation detection algorithm), and at the same time as shadow (by a shadow detection algorithm). Therefore, we are motivated for a simultaneous shadow/vegetation detection. Different shadow indices and vegetation indices are investigated. Shadow index c_3, vegetation index ExG (or NDVI if the NIR is available) and the luminance L are chosen. Each feature image computed from these indices can be considered as an information source. In this context, DS fusion that aims at merging different data sources is employed. The principal advantage of this theory is its ability to take into account ignorance of the information by affecting a degree of confidence which is called a mass function to all simple and compound hypotheses. Moreover, the spatial correlation between neighboring pixels is also taken into account using the MRF modelling in order to finally obtain a more reliable and efficient segmentation map with good accuracy. SSVD method is tested on a variety of image datasets and demonstrates its efficiency [Ngo 2014a, Ngo 2014b].

    Second, a novel approach is developed for the detection of buildings from a single optical image. This SBBD method does not require any elevation or other additional data. In this method, we

focus on buildings with right-angled corners, that can be characterized as a collection of rectangles, with the assumption that shadows cast by building objects can be detected and buildings have a homogeneous color. From the shadow/vegetation mask detected in the previous part, boundaries between shadows and their corresponding buildings are detected. In order to effectively extract building objects from image, an novel MRF region growing image segmentation is proposed. The input image is first decomposed into a set of homogeneous primitive regions. Regions are then grouped into different clusters by a MRF-based region classification. According to the position of the shadows, a merging procedure is performed over regions that belongs to the same cluster to produce regions whose shapes are appropriate to rectangles. This iterative classification-merging stops when no merging happens. From the resulting segmentation, an algorithm for determining the final building objects is proposed. The experimental results prove that SBBD method is applicable in various areas (high dense urban, sub-urban, and rural) and is highly robust and reliable [Ngo 2015].

7.2   Perspectives

7.2.1   Shadow/Vegetation Detection

Conventional image classification techniques assume that all the pixels within the image are pure, that they represent an area of homogeneous cover of a single class. This assumption is often untenable with pixels of mixed composition abundant in an image. Future work on shadow/vegetation detection can consider not only pure class (shadow or vegetation) but also mixed class (shadow and vegetation). Moreover, since DS evidence theory can deal with any union of classes, it is particularly useful to represent the "mixed" pixels in classification problems.

$$m_s^{(1)}(\{\{\omega_1\}, \{\omega_4\}\}), m_s^{(1)}(\{\{\omega_2\}, \{\omega_3\}\}), m_s^{(1)}(\Theta)$$

$$m_s^{(2)}(\{\{\omega_2\}, \{\omega_4\}\}), m_s^{(2)}(\{\{\omega_1\}, \{\omega_3\}\}), m_s^{(2)}(\Theta) \quad \oplus \quad m_s(A) = m_s^{(1)} \oplus m_s^{(2)} \oplus m_s^{(3)}$$
$$s \in \mathcal{S}, A \subset \Theta$$

$$m_s^{(3)}(\{\omega_3\}), m_s^{(3)}(\{\{\omega_1\}, \{\omega_2\}, \{\omega_4\}\})), m_s^{(3)}(\Theta)$$
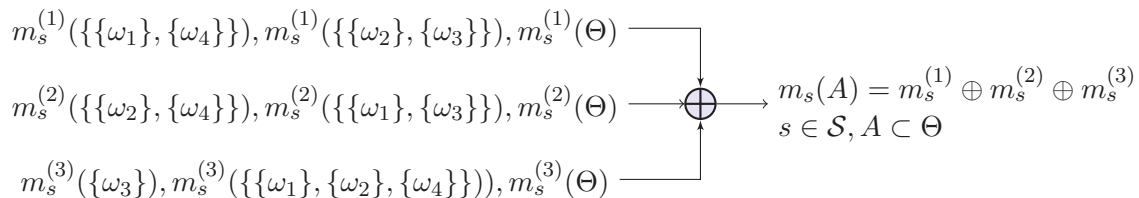
Figure 7.1: DS data fusion diagram for 4-class image segmentation. ω_4 denotes the new class ("shadow-vegetation")

We have run the test on this new model, in which the DS data fusion diagram (Fig. 4.3, Chapter 4) is modified as shown in Fig. 7.1 and other steps of the proposed method (Algorithm 1) are kept the same. Fig. 7.2 shows a preliminary result of 4-class segmentation. We observe that shadows and vegetation are very well detected. There are no false alarm on the mixed class (shadow-vegetation). However, some mixed class regions are not detected. Two possible avenues to improve the performance are replacing the Otsu method by other thresholding methods [Mehmet Sezgin 2004] or investigating other methods for determining the mass function, such as fuzzy logic [Zhu 2002, Chaabane 2009], neural network classifier [Denoeux 2000].

(a) RGB image (b) Ground truth (c) Our segmentation result

Figure 7.2: Results of segmentation into four classes: shadow (black), vegetation (green), shadow-vegetation (cyan) and other (white).

7.2.2 Building Detection

There are several directions in which one could take to further the research on building detection. For a prior MRF model, we intend to express the input data in different color spaces (HSV, LUV, YIQ, XYZ and LAB) and then co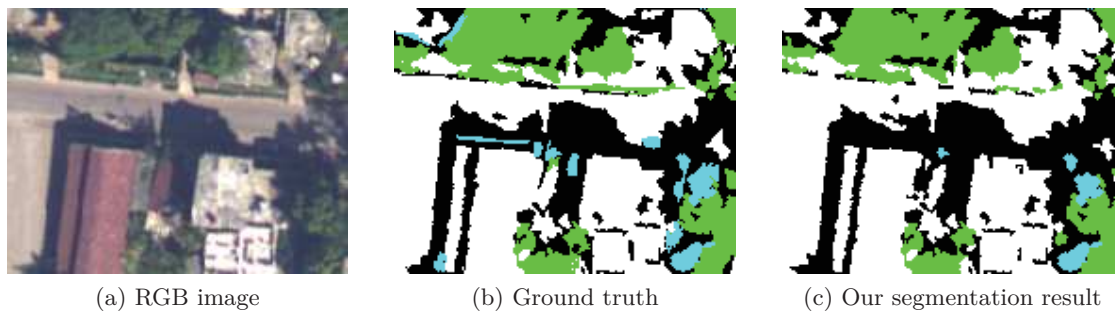mbine the set of color values to obtain the multidimensional feature descriptor vector. In fact, each color space has an interesting property, which can efficiently be taken into account in order to render image segmentation more reliable [Banks 1991].

Moreover, inspired from [Yu 2007], we are interested in defining a MRF energy function in which three terms: data term, regularization term and form are included. The neighboring regions are classified as the same class if they have similar spectral information and their fusion are appropriate to rectangles. As a result, the segmentation performance may be improved. The merging may be done between two regions having the same class label so as to increase the MRF energy. Instead of merging regions at the end of segmentation and repeating the region classification-merging like our method, the merging is run at each iteration of ICM. This may make the algorithm faster than our method.

7.2.3 Building Classification

Another stage of the processing chain for building database updating described in Chapter 1 should be studied: building classification. In order to do a good job of building classification, we need to select the appropriate set of features. An example of rules to distinguish between constructed building and building under construction is shown in Table 7.1.

	Constructed building	Building under construction
Proportion of shadow pixels inside the building	Few or none	Many
Shadows (of wall inside) are parallel with the building boundaries	No	Yes

Table 7.1: Distinguish between a building under construction and a constructed building

A preliminary result is shown in Fig. 7.3 in which we use only the proportion of shadow pixels. More rules to classify buildings need to be defined. For example, the high density of vegetation regions inside the building footprint is a good index to classify a building as the building in ruin.

Input image   Shadow/Vegetation detection   Detected building and its classification
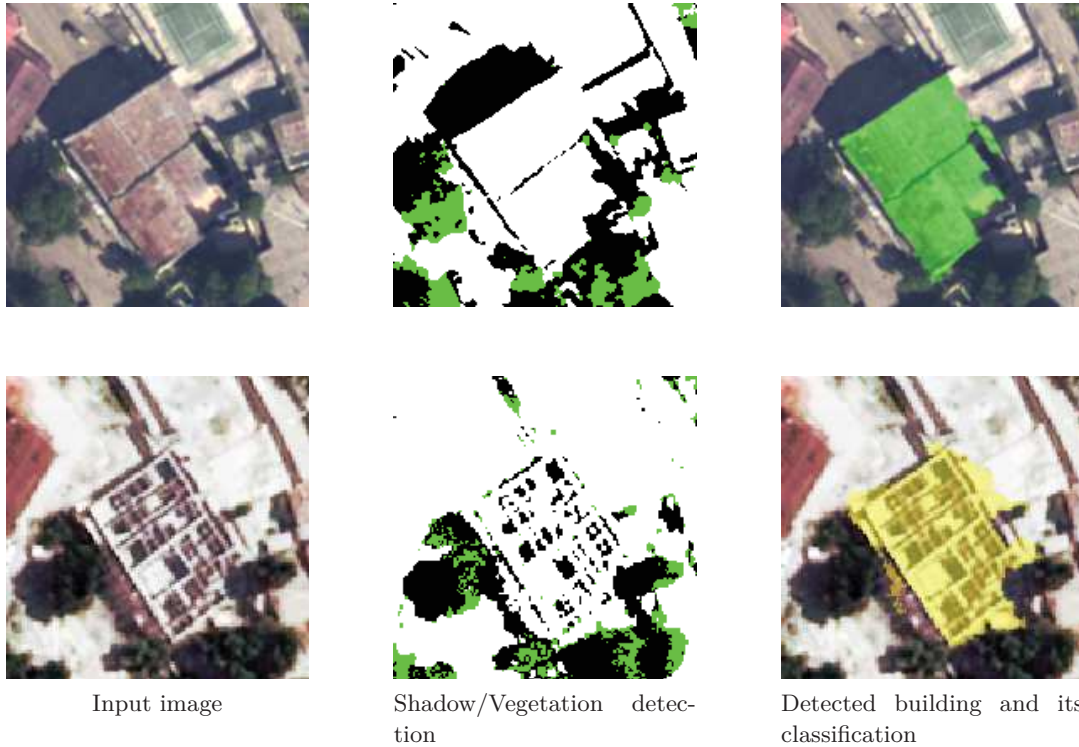
Figure 7.3: Change detection on building states: the first line represents a constructed building detected from NOAA aerial image (of time n); the second line represents the same building under construction detected from Worldview-2 satellite image (of time $n+1$) of Jacmel datasets. Hence, this building is classified as building under reconstruction between two different times.

Otsu Thresholding Method

Thresholding is an important technique in image segmentation applications. The basic idea of thresholding is to select an optimal gray-level threshold value for separating objects of interest in an image from the background based on their gray-level distribution. Thresholding creates binary images from gray-level ones by turning all pixels below some threshold to zero and all pixels about that threshold to one. If $g(x, y)$ is a thresholded version of $I(x, y)$ at some global threshold t:

$$g(x, y) = \begin{cases} 1, & \text{if } I(x, y) \geq t \\ 0, & \text{otherwise} \end{cases}$$

Otsu method [Otsu 1975] is a global thresholding selection method, proposed by Nobuyuki Otsu in 1975. This method is represented as follows. Assuming image $I(x, y)$ is represented in L gray levels $[0,1,\ldots,L-1]$. The number of pixels at level i is denoted by n_i, and the total number of pixels is denoted by $N = n_0 + n_1 + \ldots + n_{L-1}$. The probability of gray level i is denoted by:

$$p_i = \frac{n_i}{N}, \quad n_i \geq 0, \quad \sum_{i=0}^{L-1} p_i = 1 \tag{A.1}$$

In the bi-level thresholding method, suppose that the pixels of image are divided into two classes C_0 and C_1 by a threshold at level t; C_0 denotes pixels with levels $[0, 1, \ldots, t]$ and C_1 denotes pixels with levels $[t + 1, \ldots, L - 1]$ by the threshold t. The gray level probability distributions for the two classes are :

$$w_0 = \sum_{i=0}^{t} p_i$$
$$w_1 = \sum_{i=t+1}^{L-1} p_i \tag{A.2}$$

The means of class C_0 and C_1 are:

$$u_0 = \sum_{i=0}^{t} \frac{ip_i}{w_0}$$
$$u_1 = \sum_{i=t+1}^{L-1} \frac{ip_i}{w_1} \tag{A.3}$$

The total mean of gray levels is denoted by u_t

$$u_t = w_0 u_0 + w_1 u_1 \tag{A.4}$$

The class variances are:

$$\sigma_0{}^2 = \sum_{i=0}^{t} \frac{(i - u_0)^2 p_i}{w_0}$$
$$\sigma_1{}^2 = \sum_{i=t+1}^{L-1} \frac{(i - u_1)^2 p_i}{w_1}$$

$$(A.5)$$

The within-class variance is:

$$\sigma_W{}^2(t) = w_0 \sigma_0{}^2 + w_1 \sigma_1{}^2 \tag{A.6}$$

The between-class variance is:

$$\sigma_B{}^2(t) = w_0 (u_0 - u_t)^2 + w_1 (u_1 - u_t)^2 = w_0 w_1 (u_1 - u_0) \tag{A.7}$$

So, the algorithm is for each potential threshold t:

1. Separate the pixels into two clusters according to the threshold.

2. Find the mean of each cluster (u_0, u_1).

3. Square the difference between the means.

4. Multiply by the number of pixels in one cluster times the number in the other (eq. A.7).

This depends only on the difference between the means of the two clusters, thus avoiding having to calculate differences between individual intensities and the cluster means. The optimal threshold is the one that maximizes the between-class variance $\sigma_B{}^2(t)$ (or, conversely, minimizes the within-class variance $\sigma_W{}^2(t)$ because the total variance of gray levels $\sigma_t{}^2 = \sigma_W{}^2 + \sigma_B{}^2$ is constant for different partitions.)

$$T = \arg\max_{0 \leq t \leq L-1} \sigma_B{}^2(t) = \arg\min_{0 \leq t \leq L-1} \sigma_W{}^2(t) \tag{A.8}$$

Color Space Transformation

The use of color in image processing is motivated by two principal factors. First, color is a powerful descriptor that often simplifies object identification and extraction from a scene. Second, humans can discern thousands of color shades and intensities, compared to few shades of gray. All colors are seen as variable combination of the three primaries in the RGB color model, which is usually used in representing and displaying images. Besides, several color models that decouple luminance and chromaticity are briefly described in the following in terms of their relations with the RGB model.

RGB Model   In this model, each color appears in its primary spectral components of red (R), green (G) and blue (B). This model is based on Cartesian coordinate system.

HSI Model   When humans view a color object, the object is described by its hue, saturation and brightness. Hue (H) is a color attribute that describes a pure color (rainbow color). Saturation (S) gives a measure of the degree to which a pure color is diluted by white light. The intensity (gray level) (I) is a most useful descriptor of monochromatic images. The HSI model manipulates color images with the following transformation from the RGB model [Gonzalez 2004]:

$$
\begin{aligned}
\theta &= \cos^{-1}\left[\frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\left[(R-G)^2+(R-B)(G-B)\right]^{\frac{1}{2}}}\right] \\
H &= \begin{cases} \theta, & \text{if } B \leq G \\ 360-\theta, & \text{otherwise} \end{cases} \\
S &= 1 - \frac{3}{R+G+B}\left[\min(R,G,B)\right] \\
I &= \frac{R+G+B}{3}
\end{aligned}
\tag{B.1}
$$

HSV model   The HSV model describes colors in terms of hue (H), saturation (S), and value (V). Hue (H) represents the wavelength of a color and varies from 0 to 1 when color goes from red to green then to blue and back to red (H is defined modulo 1). Saturation (S) represents the amount of white color mixed with the monochromatic color. Value (V) represents the brightness.

[Smith 1978] describes a triangle-based HSV model in the following relations with the RGB model:

$$V = \max(R, G, B)$$
$$S = \frac{V - \min(R, G, B)}{V}$$
$$H = \begin{cases} \dfrac{G - B}{6S}, & \text{if } V = R \\ \dfrac{1}{3} + \dfrac{B - R}{6S}, & \text{if } V = G \\ \dfrac{2}{3} + \dfrac{R - G}{6S}, & \text{if } V = B \end{cases} \tag{B.2}$$

HCV model   The HCV model describes the dominant frequency, the amount of color, and luminance, respectively, in the following relations with the RGB model [Gonzalez 2004]:

$$V = \frac{R + G + B}{3}$$
$$H = \tan^{-1}\left[\frac{R - B}{\sqrt{3}(V - G)}\right]$$
$$C = \begin{cases} \dfrac{V - G}{\cos H}, & \|\cos H\| > 0.2 \\ \dfrac{R - B}{\sqrt{3}\sin H}, & \text{otherwise} \end{cases} \tag{B.3}$$

YIQ model   In this scheme, Y is proportional to the gamma-corrected luminance, which corresponds roughly with intensity, I and Q jointly describe the chroma, which corresponds with hue and saturation, of a color image in the following relations with the RGB model [Gonzalez 2004]:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{B.4}$$

YC_bC_r **model**   This color space is useful in compression applications [Kumar 2002]. It has the following relations with the RGB model:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{B.5}$$

$c_1c_2c_3$ **model**   The color space $c_1c_2c_3$ is defined by [Gevers 1999] as follows:

$$c_1 = \arctan\left(\frac{R}{\max(G, B)}\right)$$
$$c_2 = \arctan\left(\frac{G}{\max(R, B)}\right)$$
$$c_3 = \arctan\left(\frac{B}{\max(G, R)}\right) \tag{B.6}$$

Bibliography

[Achanta 2012] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua and Sabine Susstrunk. *SLIC superpixels compared to state-of-the-art superpixel methods.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pages 2274–2282, 2012. (Cited on pages ix, 71, 72, 73, 74 and 89.)

[Adeline 2013] K.R.M. Adeline, M. Chen, X. Briottet, S.K. Pang and N. Paparoditis. *Shadow detection in very high spatial resolution aerial images: A comparative study.* ISPRS Journal of Photogrammetry and Remote Sensing, vol. 80, pages 21–38, 2013. (Cited on pages iv, 22, 24, 28 and 51.)

[Akcay 2010] Huseyin Gokhan Akcay and Selim Aksoy. *Building detection using directional spatial constraints.* In IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pages 1932–1935, 2010. (Cited on pages 64 and 65.)

[Arévalo 2008] V Arévalo, J González and G Ambrosio. *Shadow detection in colour high resolution satellite images.* International Journal of Remote Sensing, vol. 29, no. 7, pages 1945–1963, 2008. (Cited on pages 22 and 51.)

[Baltsavias 2004] EP Baltsavias. *Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems.* ISPRS Journal of Photogrammetry and Remote Sensing, vol. 58, no. 3, pages 129–151, 2004. (Cited on page 64.)

[Banks 1991] Stephen P Banks. Signal processing, image processing and pattern recognition. Prentice Hall PTR, 1991. (Cited on page 105.)

[Bannari 1997] Abderrazak Bannari, Denis Morin and Dong-Chen He. *Caractérisation de l'environnement urbaina l'aide des indices de végétation dérivés des données de hautes résolutions spatiale et spectrale.* Télédétection des milieux urbains et périurbains, pages 47–64, 1997. (Cited on page 28.)

[Ben Chaabane 2008] Salim Ben Chaabane, M. Sayadi, F. Fnaiech and E. Brassart. *Color image segmentation using automatic thresholding and the fuzzy C-means techniques.* In IEEE Mediterranean Electrotechnical Conference, MELECON'2008, Ajaccio-France, pages 857–861, 2008. (Cited on page 43.)

[Bendjebbour 2001] Azzedine Bendjebbour, Yves Delignon, L. Fouque, V. Samson and W. Pieczynski. *Multisensor image segmentation using Dempster-Shafer fusion in Markov fields context.* IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 8, pages 1789–1798, 2001. (Cited on page 47.)

[Benedek 2012] Csaba Benedek, Xavier Descombes and Josiane Zerubia. *Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 1, pages 33–50, 2012. (Cited on pages 15, 64, 66, 67 and 87.)

[Bentabet 2008] Layachi Bentabet and Jiang Maodong. *A combined Markovian and Dirichlet submixture modeling for evidence assignment: Application to image fusion.* Pattern Recognition Letter, vol. 29, no. 13, pages 1775–1783, 2008. (Cited on pages vi and 47.)

[Besag 1986] Julian Besag. *On the Statistical Analysis of Dirty Picture.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 48, no. 3, pages 259–302, 1986. (Cited on pages 45, 75, 76 and 77.)

[Brenner 2005] Claus Brenner. *Building reconstruction from images and laser scanning.* International Journal of Applied Earth Observation and Geoinformation, vol. 6, no. 3, pages 187–198, 2005. (Cited on page 64.)

[Calderero 2010] Felipe Calderero and Ferran Marques. *Region merging techniques using information theory statistical measures.* IEEE Transactions on Image Processing, vol. 19, no. 6, pages 1567–1586, 2010. (Cited on page 71.)

[Chaabane 2009] S Ben Chaabane, Mounir Sayadi, Farhat Fnaiech and Eric Brassart. *Dempster-Shafer evidence theory for image segmentation: application in cells images.* International Journal of Signal Processing, vol. 5, no. 1, 2009. (Cited on pages 43 and 104.)

[Champion 2010] Nicolas Champion, Didier Boldo, Marc Pierrot-Deseilligny and Georges Stamon. *2D building change detection from high resolution satellite imagery: A two-step hierarchical method based on 3D invariant primitives.* Pattern Recognition Letter, vol. 31, no. 10, pages 1138–1147, July 2010. (Cited on page 5.)

[Champion 2011] Nicolas Champion. *Détection de changement 2D à partir d'imagerie satellitaire: Application à la mise à jour des bases de données géographiques.* PhD thesis, 2011. (Cited on page 7.)

[Chaudhuri 2007] D Chaudhuri and A Samal. *A simple method for fitting of bounding rectangle to closed regions.* Pattern recognition, vol. 40, no. 7, pages 1981–1989, 2007. (Cited on page 66.)

[Chaudhuri 2012] D Chaudhuri, NK Kushwaha, I Sharif and A Samal. *Finding best-fitted rectangle for regions using a bisection method.* Machine Vision and Applications, vol. 23, no. 6, pages 1263–1271, 2012. (Cited on page 66.)

[Chen 2010] Chun-Ting Chen, Chung-Yen Su and Wen-Chung Kao. *An enhanced segmentation on vision-based shadow removal for vehicle detection.* In International Conference on Green Circuits and Systems, pages 679–682, 2010. (Cited on page 22.)

[Chen 2013] Dongyue Chen, Ting Zhou and Zongwen Chen. *A Novel Image Segmentation Algorithm: Region Merging Using Superpixel-based Local CRF Model.* Journal of Information and Computational Science, pages 5145–5153, 2013. (Cited on page 71.)

[Chung 2009] Kuo-Liang Chung, Yi-Ru Lin and Yong-Huai Huang. *Efficient Shadow Detection of Color Aerial Images Based on Successive Thresholding Scheme.* IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 2, pages 671–682, February 2009. (Cited on pages 21 and 23.)

[Comaniciu 2002] Dorin Comaniciu and Peter Meer. *Mean shift: A robust approach toward feature space analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pages 603–619, 2002. (Cited on page 72.)

[Cucchiara 2003] Rita Cucchiara, Costantino Grana, Massimo Piccardi and Andrea Prati. *Detecting moving objects, ghosts, and shadows in video streams*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 10, pages 1337–1342, 2003. (Cited on page 22.)

[Cui 2011] Shiyong Cui, Peter Reinartz and Qin Yan. Graph search and its application in building extraction from high resolution remote sensing imagery. INTECH Open Access Publisher, 2011. (Cited on page 66.)

[Dempster 1968] Arthur P Dempster. *A generalization of Bayesian inference*. Journal of the Royal Statistical Society. Series B (Methodological), pages 205–247, 1968. (Cited on pages iv, 40 and 44.)

[Denoeux 1997] Thierry Denoeux. *Analysis of evidence-theoretic decision rules for pattern classification*. Pattern Recognition, vol. 30, no. 7, pages 1095–1107, 1997. (Cited on page 45.)

[Denoeux 2000] Thierry Denoeux. *A neural network classifier based on Dempster-Shafer theory*. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 30, no. 2, pages 131–150, 2000. (Cited on pages 43 and 104.)

[Denœux 2006] Thierry Denœux. *The cautious rule of combination for belief functions and some extensions*. In Information Fusion, 2006 9th International Conference on, pages 1–8. IEEE, 2006. (Cited on pages vi and 44.)

[Denœux 2008] Thierry Denœux. *Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence*. Artificial Intelligence, vol. 172, no. 2, pages 234–264, 2008. (Cited on pages vi and 44.)

[Derin 1987] Haluk Derin and Howard Elliott. *Modeling and segmentation of Noisy and Textured Images Using Gibbs Random Fields*. IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 1, pages 39–55, 1987. (Cited on pages 46 and 52.)

[Dey 2010] V Dey, Y Zhang and M Zhong. *A review on image segmentation techniques with remote sensing perspective*. In Proceedings of ISPRS Congress, volume 38, pages 5–7, 2010. (Cited on page 71.)

[Ding 2010] Lei Ding and Alper Yilmaz. *Interactive image segmentation using probabilistic hypergraphs*. Pattern Recognition, vol. 43, no. 5, pages 1863–1873, 2010. (Cited on page 72.)

[Duda 2002] Tanja Duda and Mort Canty. *Unsupervised classification of satellite imagery: choosing a good algorithm*. International Journal of Remote Sensing, vol. 23, no. 11, pages 2193–2212, 2002. (Cited on page 29.)

[Felzenszwalb 2004] Pedro F Felzenszwalb and Daniel P Huttenlocher. *Efficient Graph-Based Image Segmentation*. International Journal of Computer Vision, vol. 59, no. 2, pages 167–181, 2004. (Cited on page 72.)

[Femiani 2014] John Femiani, Er Li, Anshuman Razdan and Peter Wonka. *Shadow-Based Rooftop Segmentation in Visible Band Images*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014. (Cited on pages ix, xii, xiii, 65, 92, 93, 94, 95 and 101.)

[Foucher 2002] Samuel Foucher, Mickaël Germain and Jean-Marc Boucher. *Multisource Classification Using ICM and Dempster–Shafer Theory*. IEEE Transactions on Instrumentation and Measurement, vol. 51, no. 2, pages 277–281, 2002. (Cited on page 47.)

[Fradkin 2001] M Fradkin, H Maıtre and Michel Roux. *Building detection from multiple aerial images in dense urban areas*. Computer Vision and Image Understanding, vol. 82, no. 3, pages 181–207, 2001. (Cited on page 64.)

[Geman 1984] S Geman and D Geman. *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 6, pages 721–41, 1984. (Cited on pages vi, 39, 45, 75 and 76.)

[Gevers 1999] Theo Gevers and Arnold W M Smeulders. *Color-based object recognition*. Pattern Recognition, vol. 32, no. 3, pages 453–464, 1999. (Cited on pages 22, 35 and 110.)

[Giros 2012] A Giros, D Fontannaz, B Allenbach, D Treinsoutrot and M De Michele. *KAL-Haiti: A research database for risks management and sustainable reconstruction in Haiti*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pages 39–B8, 2012. (Cited on page 11.)

[Gonzalez 2004] Rafael C Gonzalez, Richard Eugene Woods and Steven L Eddins. Digital image processing using matlab. Pearson Education India, 2004. (Cited on pages 109 and 110.)

[Gulch 1999] E Gulch, H Muller and T Labe. *Integration of automatic processes into semi-automatic building extraction*. International Archives of Photogrammetry and Remote Sensing, vol. 32, no. 3; SECT 2W5, pages 177–186, 1999. (Cited on page 64.)

[Haala 2010] Norbert Haala and Martin Kada. *An update on automatic 3D building reconstruction*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 65, no. 6, pages 570–580, 2010. (Cited on page 64.)

[Haris 1998] Kostas Haris, Serafim N Efstratiadis, Nikolaos Maglaveras and Aggelos K Katsaggelos. *Hybrid image segmentation using watersheds and fast region merging*. IEEE Transactions on Image Processing, vol. 7, no. 12, pages 1684–1699, 1998. (Cited on page 71.)

[Henricsson 1997] Olof Henricsson and Emmanuel Baltsavias. 3-d building reconstruction with aruba: a qualitative and quantitative evaluation. Monte Verità. Springer, 1997. (Cited on page 64.)

[Hofmann 2002] Alexandra D Hofmann, Hans-Gerd Maas and André Streilein. *Knowledge-based building detection based on laser scanner data and topographic map information*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34, no. 3/A, pages 169–174, 2002. (Cited on page 64.)

[Huang 2004] Jianjun Huang, Weixin Xie and Liang Tang. *Detection of and compensation for shadows in colored urban aerial images.* Fifth World Congress on Intelligent Control and Automation, vol. 4, no. l, pages 3098–3100, 2004. (Cited on page 22.)

[Huang 2014] Zhijian Huang, Jinfang Zhang, Xiang Li and Hui Zhang. *Remote sensing image segmentation based on Dynamic Statistical Region Merging.* Optik-International Journal for Light and Electron Optics, vol. 125, no. 2, pages 870–875, 2014. (Cited on page 71.)

[Huertas 1988] Andres Huertas and Ramakant Nevatia. *Detecting buildings in aerial images.* Computer Vision, Graphics, and Image Processing, vol. 41, no. 2, pages 131–152, 1988. (Cited on page 64.)

[Huete 1994] AR Huete, H Liu, GR De Lira, K Batchily and R Escadafal. *A soil color index to adjust for soil and litter noise in vegetation index imagery of arid regions.* In Geoscience and Remote Sensing Symposium, 1994. IGARSS'94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation., International, volume 2, pages 1042–1043, 1994. (Cited on pages 28 and 29.)

[Huete 2002] A Huete, Kamel Didan, Tomoaki Miura, E Patricia Rodriguez, Xiang Gao and Laerte G Ferreira. *Overview of the radiometric and biophysical performance of the MODIS vegetation indices.* Remote sensing of environment, vol. 83, no. 1, pages 195–213, 2002. (Cited on page 28.)

[Izadi 2010] Mohammad Izadi and Parvaneh Saeedi. *Automatic building detection in aerial images using a hierarchical feature based image segmentation.* In 20th International Conference on Pattern Recognition (ICPR), pages 472–475, 2010. (Cited on pages 8, 65 and 66.)

[Jackson 1983] RD Jackson, PN Slater and PJ Pinter. *Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres.* Remote sensing of environment, vol. 13, no. 3, pages 187–208, 1983. (Cited on page 28.)

[Karadag 2015] Ozge Oztimur Karadag, Caglar Senaras and Fatos T Yarman Vural. *Segmentation Fusion for Building Detection Using Domain-Specific Information.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pages 1–11, 2015. (Cited on pages 8 and 66.)

[Kataoka 2003] T. Kataoka, T. Kaneko, H. Okamoto and S. Hata. *Crop growth estimation system using machine vision.* In Proceedings of the 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pages 1079–1083, 2003. (Cited on pages 28 and 29.)

[Katartzis 2008] Antonis Katartzis and Hichem Sahli. *A stochastic framework for the identification of building rooftops using a single remote sensing image.* IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 1, pages 259–271, 2008. (Cited on page 64.)

[Kim 2007] Taejung Kim, Ts Javzandulam and Tae-Yoon Lee. *Semiautomatic reconstruction of building height and footprints from single satellite images.* In Geoscience and Remote Sensing Symposium, IGARSS 2007, pages 4737–4740, 2007. (Cited on page 21.)

[Kolmogorov 2004] Vladimir Kolmogorov and Ramin Zabin. *What energy functions can be minimized via graph cuts?* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pages 147–159, 2004. (Cited on page 77.)

[Kolmogorov 2006] Vladimir Kolmogorov. *Convergent tree-reweighted message passing for energy minimization.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pages 1568–1583, 2006. (Cited on page 77.)

[Korting 2008] T.S. Korting, L.M.G. Fonseca, L.V. Dutra and F.C. Da Silva. *Image re-segmentation - A new approach applied to urban imagery.* In VISAPP 2008 - 3rd International Conference on Computer Vision Theory and Applications, Proceedings, volume 1, pages 467–472, 2008. (Cited on page 66.)

[Korting 2011] Thales Sehn Korting, Luciano Vieira Dutra and Leila Maria Garcia Fonseca. *A resegmentation approach for detecting rectangular objects in high-resolution imagery.* IEEE Geoscience and Remote Sensing Letter, vol. 8, no. 4, pages 621–625, 2011. (Cited on page 66.)

[Krishnamachari 1995] Santhana Krishnamachari and Rama Chellappa. *Delineating buildings by grouping lines with MRFs.* IEEE Transactions on Image Processing, vol. 5, no. 1, pages 164–168, 1995. (Cited on page 75.)

[Kumar 2002] Pranaw Kumar, Kuntal Sengupta and Albert Lee. *A comparative study of different color spaces for foreground and shadow detection for traffic monitoring system.* In The IEEE 5th International Conference on Intelligent Transportation Systems, pages 100–105. IEEE, 2002. (Cited on page 110.)

[Kwak 2014] Eunju Kwak and Ayman Habib. *Automatic representation and reconstruction of DBM from LiDAR data using Recursive Minimum Bounding Rectangle.* ISPRS Journal of Photogrammetry and Remote Sensing, vol. 93, pages 171–191, 2014. (Cited on pages 66, 68 and 83.)

[Lafarge 2010] Florent Lafarge, Xavier Descombes, Josiane Zerubia and Marc Pierrot-Deseilligny. *Structural approach for building reconstruction from a single DSM.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pages 135–147, 2010. (Cited on page 64.)

[Le Hegarat-Mascle 1997] S. Le Hegarat-Mascle, I. Bloch and D. Vidal-Madjar. *Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing.* IEEE Transactions on Geoscience and Remote Sensing, vol. 35, no. 4, pages 1018–1031, 1997. (Cited on pages 43 and 45.)

[Levinshtein 2009] Alex Levinshtein, Adrian Stere, Kiriakos N Kutulakos, David J Fleet, Sven J Dickinson and Kaleem Siddiqi. *Turbopixels: Fast superpixels using geometric flows.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 12, pages 2290–2297, 2009. (Cited on page 72.)

[Li 2015] Er Li, John Femiani, Shibiao Xu, Xiaopeng Zhang and Peter Wonka. *Robust Rooftop Extraction From Visible Band Images Using Higher Order CRF.* IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 8, pages 4483–4495, 2015. (Cited on pages ix, xii, xiii, 65, 97, 98, 99 and 101.)

[Lin 1998] Chungan Lin and Ramakant Nevatia. *Building detection and description from a single intensity image.* Computer vision and image understanding, vol. 72, no. 2, pages 101–121, 1998. (Cited on page 64.)

[Liu 2005] Wei Liu and Véronique Prinet. *Building detection from high-resolution satellite image using probability model*. In International Geoscience and Remote Sensing Symposium, volume 6, page 3888, 2005. (Cited on page 65.)

[Lopez 1999] Oscar A Lopez and Elizabeth Raven. *An overall evaluation of irregular-floor-plan-shaped buildings located in seismic areas*. Earthquake spectra, vol. 15, no. 1, pages 105–120, 1999. (Cited on page 66.)

[Luca 2012] Lorenzi Luca. *Innovative methods for the reconstruction of new generation satellite remote sensing images*. PhD thesis, University of Trento, 2012. (Cited on page 22.)

[Makarau 2011] Aliaksei Makarau, Rudolf Richter, Rupert Muller and Peter Reinartz. *Adaptive Shadow Detection Using a Blackbody Radiator Model*. IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 6, pages 2049–2059, June 2011. (Cited on page 23.)

[Matthews 1975] Brian W Matthews. *Comparison of the predicted and observed secondary structure of T4 phage lysozyme*. Biochimica et Biophysica Acta (BBA)-Protein Structure, vol. 405, no. 2, pages 442–451, 1975. (Cited on pages vii and 24.)

[Mayer 1999] Helmut Mayer. *Automatic object extraction from aerial imagery—a survey focusing on buildings*. Computer vision and image understanding, vol. 74, no. 2, pages 138–149, 1999. (Cited on page 64.)

[McKeown Jr 1999] David M McKeown Jr, Steven D Cochran, Stephen J Ford, J Chris McGlone, Jeffery A Shufelt and Daniel A Yocum. *Fusion of HYDICE hyperspectral data with panchromatic imagery for cartographic feature extraction*. IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 3, pages 1261–1277, 1999. (Cited on page 64.)

[Mehmet Sezgin 2004] Bulent Sankur Mehmet Sezgin. *Survey over image thresholding techniques and quantitative performance evaluation*. Journal of Electronic Imaging, vol. 13, no. 1, pages 220–220, January 2004. (Cited on page 104.)

[Meyer 2008] George E. Meyer and João Camargo Neto. *Verification of color vegetation indices for automated crop imaging applications*. Computers and Electronics in Agriculture, vol. 63, no. 2, pages 282–293, October 2008. (Cited on page 28.)

[Moore 2009] Alastair P Moore, Simon JD Prince, Jonathan Warrell, Umar Mohammed and Graham Jones. *Scene shape priors for superpixel segmentation*. In 2009 IEEE 12th International Conference on Computer Vision, pages 771–778. IEEE, 2009. (Cited on page 71.)

[Moscheni 1998] Fabrice Moscheni, Sushil Bhattacharjee and Murat Kunt. *Spatio-temporal segmentation based on region merging*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 9, pages 897–915, 1998. (Cited on page 71.)

[M.Polidorio 2003] Airton M.Polidorio. *Automatic Shadow Segmentation in Aerial Color Images*. Computer Graphics and Image Processing, 2003. SIBGRAPI 2003. XVI Brazilian Symposium, pages 270–277, 2003. (Cited on pages 22, 23 and 35.)

[Müller 2005] Sönke Müller and Daniel Wilhelm Zaum. *Robust building detection in aerial images*. International Archives of Photogrammetry and Remote Sensing, vol. 36, no. B2/W24, pages 143–148, 2005. (Cited on pages 64, 65 and 67.)

[Nagao 1979] Makoto Nagao, Takashi Matsuyama and Yoshio Ikeda. *Region Extraction and Shape Analysis in Aerial Photographs.* Computer Graphics and Image Processing, 1979. (Cited on page 23.)

[Nakajima 2002] Takashi Nakajima, Guo Tao and Yoshifumi Yasuoka. *Simulated recovery of information in shadow areas on IKONOS image by combing ALS data.* In Proc. Asian Conference on Remote Sensing, 2002. (Cited on page 22.)

[Neto 2006] J.C. Neto, G.E. Meyer and D.D. Jones. *Individual leaf extractions from young canopy images using Gustafson-Kessel clustering and a genetic algorithm.* Computers and Electronics in Agriculture, vol. 51, pages 66–85, 2006. (Cited on pages 28 and 29.)

[Ngo 2014a] Tran-Thanh Ngo, Christophe Collet and Vincent Mazet. *Détection simultanée de l'ombre et la végétation sur des images aériennes couleur en haute résolution.* In Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014, 2014. (Cited on pages 59 and 103.)

[Ngo 2014b] Tran-Thanh Ngo, Christophe Collet and Vincent Mazet. *MRF and Dempster-Shafer theory for simultaneous shadow/vegetation detection on high resolution aerial color images.* In IEEE International Conference on Image Processing (ICIP), 2014. (Cited on pages 59 and 103.)

[Ngo 2015] Tran-Thanh Ngo, Christophe Collet and Vincent Mazet. *Automatic rectangular building detection from VHR aerial imagery using shadow and image segmentation.* In IEEE International Conference on Image Processing (ICIP), 2015. (Cited on pages 102 and 104.)

[Nock 2004] Richard Nock and Frank Nielsen. *Statistical region merging.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, pages 1452–1458, 2004. (Cited on page 71.)

[Ok 2013] Ali Ozgun Ok, Caglar Senaras and Baris Yuksel. *Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery.* IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 3, pages 1701–1717, 2013. (Cited on pages 29, 64 and 88.)

[Otsu 1975] Nobuyuki Otsu. *A Threshold Selection Method from Gray-Level Histograms.* Automatica, vol. 11, no. 285-296, pages 23–27, 1975. (Cited on pages iv, 24, 42, 50 and 107.)

[Ozgun Ok 2013] Ali Ozgun Ok. *Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts.* ISPRS Journal of Photogrammetry and Remote Sensing, vol. 86, pages 21–40, 2013. (Cited on pages ix, xii, xiii, 8, 21, 34, 52, 64, 65, 87, 95, 96, 97 and 101.)

[Phong 1975] Bui Tuong Phong. *Illumination for computer generated pictures.* Communications of the ACM, vol. 18, no. 6, pages 311–317, 1975. (Cited on page 22.)

[Poulain 2010] Vincent Poulain. *Fusion d'images optique et radar à haute résolution pour la mise à jour de bases de données cartographiques.* PhD thesis, 2010. (Cited on page 7.)

[Prati 2003] a. Prati, I. Mikic, M.M. Trivedi and R. Cucchiara. *Detecting moving shadows: algorithms and evaluation.* IEEE Trans. on Pattern Anal. Mach. Intell, vol. 25, no. 7, pages 918–923, July 2003. (Cited on page vii.)

[Prokop 1992] Richard J Prokop and Anthony P Reeves. *A survey of moment-based techniques for unoccluded object representation and recognition.* CVGIP: Graphical Models and Image Processing, vol. 54, no. 5, pages 438–460, 1992. (Cited on page 80.)

[Qin 2010] AK Qin and David A Clausi. *Multivariate image segmentation using semantic region growing with adaptive edge penalty.* IEEE Transactions on Image Processing, vol. 19, no. 8, pages 2157–2170, 2010. (Cited on pages 71, 75 and 77.)

[Ren 2005] Xiaofeng Ren, Jitendra Malik and Charless C Fowlkes. *Cue integration for figure/ground labeling.* In Advances in neural information processing systems, pages 1121–1128, 2005. (Cited on page 72.)

[Rosin 1999] Paul L Rosin. *Measuring rectangularity.* Machine Vision and Applications, vol. 11, no. 4, pages 191–196, 1999. (Cited on pages 66 and 79.)

[Rosin 2003] Paul L Rosin. *Measuring shape: ellipticity, rectangularity, and triangularity.* Machine Vision and Applications, vol. 14, no. 3, pages 172–184, 2003. (Cited on page 66.)

[Rottensteiner 2007] Franz Rottensteiner, John Trinder, Simon Clode and Kurt Kubik. *Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis.* ISPRS Journal of Photogrammetry and Remote Sensing, vol. 62, no. 2, pages 135–149, 2007. (Cited on page 64.)

[Rouse, JW 1974] Rouse, JW, Haas, RH, Schell, JA, Deering, DW and Harlan, JC. Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. Texas A & M University, Remote Sensing Center, 1974. (Cited on page 29.)

[Rufenacht 2014] Dominic Rufenacht, Clément Fredembach and Sabine Susstrunk. *Automatic and accurate shadow detection using near-infrared information.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 8, pages 1672–1678, 2014. (Cited on pages vii, 24, 51, 56, 57 and 58.)

[Sahar 2010] Liora Sahar, Subrahmanyam Muthukumar and Steven P French. *Using aerial imagery and GIS in automated building footprint extraction and shape recognition for earthquake risk assessment of urban inventories.* IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 9, pages 3511–3520, 2010. (Cited on pages 64 and 66.)

[Salvador 2001] E. Salvador, A. Cavallaro and T. Ebrahimi. *Shadow identification and classification using invariant color models.* In IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 3, pages 1545–1548, 2001. (Cited on page 51.)

[Salvador 2004] Elena Salvador, Andrea Cavallaro and Touradj Ebrahimi. *Cast shadow segmentation using invariant color features.* Computer Vision and Image Understanding, vol. 95, no. 2, pages 238–259, 2004. (Cited on pages iv, 22, 23 and 35.)

[Seo 2014] Suyoung Seo, Jeongho Lee and Yongil Kim. *Extraction of Boundaries of Rooftop Fenced Buildings From Airborne Laser Scanning Data Using Rectangle Models*. IEEE Geoscience and Remote Sensing Letter, vol. 11, no. 2, 2014. (Cited on page 66.)

[Sertit 2013] Sertit. *Mise à Jour d'une base de données du bâti avec des images satellitaires*. Rapport technique, 2013. (Cited on page 12.)

[Shafer 1976] Glenn Shafer. A mathematical theory of evidence, volume 1. Princeton university press Princeton, 1976. (Cited on pages iv and 40.)

[Shi 2000] Jianbo Shi and Jitendra Malik. *Normalized cuts and image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pages 888–905, 2000. bibtex: shi2000normalized. (Cited on page 72.)

[Shorter 2009] Nicholas Shorter and Takis Kasparis. *Automatic Vegetation Identification and Building Detection from a Single Nadir Aerial Image*. Remote Sensing, vol. 1, no. 4, pages 731–757, 2009. (Cited on pages iv, vii, viii, 8, 34, 35, 52, 53, 54, 55 and 56.)

[Sirmacek 2008] B Sirmacek and Cem Unsalan. *Building detection from aerial images using invariant color features and shadow information*. In Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on, pages 1–5. IEEE, 2008. (Cited on page 64.)

[Smith 1978] Alvy Ray Smith. *Color gamut transform pairs*. In ACM Siggraph Computer Graphics, volume 12, pages 12–19. ACM, 1978. (Cited on page 110.)

[Sun 2010] Bangyu Sun and Shutao Li. *Moving cast shadow detection of vehicle using combined color models*. In Chinese Conference on Pattern Recognition, pages 1–5, 2010. (Cited on page 22.)

[Tian 2009] Jiandong Tian, Jing Sun and Yandong Tang. *Tricolor attenuation model for shadow detection*. IEEE Transactions on Image Processing, vol. 18, no. 10, pages 2355–63, October 2009. (Cited on page 51.)

[Tian 2012] Jiandong Tian, Linlin Zhu and Yandong Tang. *Outdoor shadow detection by combining tricolor attenuation and intensity*. EURASIP Journal on Advances in Signal Processing, pages 1–8, 2012. (Cited on pages iv, vii, viii, 23, 34, 35, 51, 53, 54, 55 and 56.)

[Tolt 2011] G. Tolt, M. Shimoni and J. Ahlberg. *A shadow detection method for remote sensing images using VHR hyperspectral and LIDAR data*. 2011 IEEE International Geoscience and Remote Sensing Symposium, pages 4423–4426, July 2011. (Cited on page 22.)

[Toussaint 1983] Godfried T Toussaint. *Solving geometric problems with the rotating calipers*. In Proc. IEEE Melecon, volume 83, page A10, 1983. (Cited on pages 66 and 79.)

[Tsai 2006] Victor JD Tsai. *A comparative study on shadow compensation of color aerial images in invariant color models*. IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 6, pages 1661–1671, June 2006. (Cited on pages vi, 22, 23 and 28.)

[Tupin 2005] Florence Tupin and Michel Roux. *Markov random field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications.* IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 8, pages 1920–1928, 2005. (Cited on pages 64 and 75.)

[Vannoorenberghe 1999] Patrick Vannoorenberghe, Olivier Colot and Denis de Brucq. *Color image segmentation using Dempster-Shafer's theory.* In International Conference on Image Processing, ICIP 1999, volume 4, pages 300–304, 1999. (Cited on page 43.)

[Vincent 1991] Luc Vincent and Pierre Soille. *Watersheds in digital spaces: an efficient algorithm based on immersion simulations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, pages 583–598, 1991. (Cited on page 72.)

[Wang 2008] Jingdong Wang, Yangqing Jia, Xian-Sheng Hua, Changshui Zhang and Long Quan. *Normalized tree partitioning for image segmentation.* In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1–8. IEEE, 2008. (Cited on page 72.)

[Wania 2007] Annett Wania. *Urban vegetation–detection and function evaluation for air quality assessment.* PhD thesis, Universität Mainz, 2007. (Cited on page 29.)

[Woebbecke 1995a] D. M. Woebbecke, G. E. Meyer, K. Von Bargen and D. A. Mortensen. *Color indices for weed identification under various soil, residue, and lighting conditions.* T. ASAE, vol. 38, pages 259–269, 1995. (Cited on page iv.)

[Woebbecke 1995b] D. M. Woebbecke, G. E. Meyer, K. Von Bargen and D. A. Mortensen. *Color indices for weed identification under various soil, residue, and lighting conditions.* Transactions of the ASAE, vol. 38, no. 1, pages 259–269, 1995. (Cited on pages 28 and 29.)

[Xie 2008] Yichun Xie, Zongyao Sha and Mei Yu. *Remote sensing imagery in vegetation mapping: a review.* Journal of plant ecology, vol. 1, no. 1, pages 9–23, 2008. (Cited on page 29.)

[Xu 2005] Min Xu, Pakorn Watanachaturaporn, Pramod K Varshney and Manoj K Arora. *Decision tree regression for soft classification of remote sensing data.* Remote Sensing of Environment, vol. 97, no. 3, pages 322–336, 2005. (Cited on pages 29 and 52.)

[Yang 2008] Allen Y Yang, John Wright, Yi Ma and S Shankar Sastry. *Unsupervised segmentation of natural images via lossy data compression.* Computer Vision and Image Understanding, vol. 110, no. 2, pages 212–225, 2008. (Cited on page 72.)

[Yao 2006] Jian Yao and Zhonefei (Mark) Zhang. *Hierarchical shadow detection for color aerial images.* Computer Vision and Image Understanding, vol. 102, no. 1, pages 60–69, 2006. (Cited on pages 35 and 51.)

[Yedidia 2005] Jonathan S Yedidia, William T Freeman and Yair Weiss. *Constructing free-energy approximations and generalized belief propagation algorithms.* IEEE Transactions on Information Theory, vol. 51, no. 7, pages 2282–2312, 2005. (Cited on page 77.)

[Yu 2007] Qiyao Yu, David A Clausi*et al. SAR Sea-Ice Image Analysis Based on Iterative Region Growing Using Semantics.* IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 12, page 3919, 2007. (Cited on pages 71 and 105.)

[Zhu 2002] Yue Min Zhu, Layachi Bentabet, Olivier Dupuis, Daniel Babot, Michele Rombaut and others. *Automatic determination of mass functions in Dempster-Shafer theory using fuzzy c-means and spatial neighborhood information for image segmentation.* Optical Engineering, vol. 41, no. 4, pages 760–770, 2002. (Cited on pages 43 and 104.)

Tran Thanh NGO

**Shadow/Vegetation and Building
detection from single optical
remote sensing image**

Abstract: This PhD thesis is devoted to the detection of shadows, vegetation and buildings from single high resolution optical remote sensing images. The first part introduces a new method for simultaneously detecting shadows and vegetation. Several shadow and vegetation indices were investigated and merged using the Dempster-Shafer evidence theory so as to obtain a segmentation map with three classes : "shadow", "vegetation" and "other". However, the performance of the fusion is sensitive to noise since it processes at a pixel-level. A Markov random field (MRF) is thus integrated to model spatial information within the image. In the second part, a novel region growing segmentation technique is proposed. The image is oversegmented into smaller homogeneous regions which replace the rigid structure of the pixel grid. An iterative region classification-merging is then applied over these regions. At each iteration, regions are classified using a MRF-based image segmentation, then, according to the position of shadows, regions having the same class are merged to produce shapes appropriate to rectangles. The final buildings are estimated using the recursive minimum bounding rectangle method from the final classification. These two algorithms have been validated on a variety of image datasets and demonstrate their efficiency.

Keywords: remote sensing, Dempster-Shafer evidence theory, Markov Random Fields, pattern recognition, image segmentation, rectangularity measure, building detection, region growing.

Résumé: Cette thèse est dédiée à la détection de l'ombre, de la végétation et des bâtiments à partir d'une unique image optique très haute résolution. La première partie présente une nouvelle méthode pour détecter simultanément les ombres et la végétation : plusieurs indices d'ombre et de végétation sont comparés puis fusionnés grâce à la théorie de l'évidence de Dempster-Shafer afin d'obtenir une segmentation en trois classes : "ombre", "végétation" et "autre". Comme la fusion est une méthode pixellique, elle est incorporée dans un contexte markovien pour régulariser la segmentation. Dans la deuxième partie, une nouvelle technique de segmentation d'images par croissance de region est proposée. L'image est tout d'abord sur-segmentée en régions homogènes afin de remplacer la structure rigide de la grille de pixels. Une classification-fusion itérative est ensuite appliquée sur ces régions. À chaque itération, les régions sont classées en utilisant une segmentation markovienne, puis regroupées entre elles en fonction de la position des ombres, de leur classe, et de la rectangularité de la forme fusionnée. Les bâtiments sont estimés à partir de la classification finale comme étant les rectangles d'emprise minimale. Ces deux algorithmes ont été validés sur plusieurs images de télédétection et ont permis de démontrer leur efficacité.

Mots-clefs: télédétection, théorie de Dempster-Shafer, champ de Markov, reconnaissance de forme, segmentation, mesure de rectangularité, détection de bâtiment, croissance de région.