



**HAL**  
open science

## Indexing of satellite images using structural information

Avik Bhattacharya

► **To cite this version:**

Avik Bhattacharya. Indexing of satellite images using structural information. Signal and Image Processing. Télécom ParisTech, 2007. English. NNT: . pastel-00006275

**HAL Id: pastel-00006275**

**<https://pastel.hal.science/pastel-00006275>**

Submitted on 19 Aug 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Thèse

Présentée pour obtenir le grade de docteur  
de l'École Nationale Supérieure des Télécommunications

Spécialité : **Signal et Images**

**Avik BHATTACHARYA**

INDEXATION DES IMAGES SATELLITAIRES  
EN UTILISANT DES INFORMATIONS STRUCTURELLES

INDEXING OF SATELLITE IMAGES  
USING STRUCTURAL INFORMATION

Soutenue le 14 Décembre 2007 devant le jury composé de

M. Georges STAMON	Président
Mme. Josiane ZERUBIA	Examinatrice
M. Henri MAITRE	Examineur
M. Jordi INGLADA	Examineur
M. Matthieu CORD	Rapporteur
M. Nicolas PAPARODITIS	Rapporteur
M. Michel ROUX	Directeur de thèse
M. Ian JERMYN	Co-Directeur de thèse



*“You are never given a wish without also being  
given the power to make it come true.  
You may have to work for it, however.”*

*-Richard Bach*

---





## Acknowledgements

I would like to take this opportunity to thank the Department of Signal and Image processing, Telecom Paris and ARIANA Project group at INRIA, Sophia Antipolis for providing me the necessary facilities during my stay and work in France. I would like to thank the French Space Agency (CNES), ACI QuerySat, the STIC INRIA-Tunisia programme, and EU NoE Muscle (FP6-507752) for partially funding my Ph.D. work. I sincerely thank INRIA for the PhD fellowship.

I express my sincere gratitude to Professor Henri Maitre and Dr. Josiane Zerubia who provided me this great opportunity to work towards my Ph.D. degree. I appreciate very much their constant support and attention during my stay at INRIA, Sophia Antipolis and Telecom Paris. Their help and encouragement were crucial for my thesis.

I would like to express sincere thanks to my thesis director Dr. Michel Roux, my thesis co-director Dr. Ian Jermyn and my advisor Dr. Xavier Descombes for all their guidance and practical advices that helped me to complete my work.

I would like to thank the members of the jury of my thesis. I thank Professor Matthieu Cord and Dr. Nicolas Paparoditis for their reports and suggestions of my thesis. I thank Professor Georges Stamon and Dr. Jordi Inglada for accepting to participate in the jury of my thesis. I am thankful to CNES for providing me with the images.

I would like to thank Professor Mihai Datcu, Mr. Alain Giros, Dr. Marine Campedel and all other members of the CNES/DLR/ENST Competence Centre on Information Extraction and Image Understanding for Earth Observation for all their help during my work at Telecom Paris.

I would like to convey my thanks to all the members of ARIANA and TSI group for their constant support and help during my stay in France. It was a pleasure to work and communicate with them. I would like to specially thank Mr. Peter Horvath, my office mate at ARIANA for all his help during my thesis. I sincerely express my thanks to Mr. David Lesage, Mr. Nicolas Widynski, Mr. Julien Rabin and Ms. Marie Lienou for helping me write the abstract of the thesis in French.

I am very grateful to my family and to Smriti for their constant support. I thank them for all their care, support and emotional encouragement during my Ph.D. work.

*I dedicate this work to my grandparents.*

---



## Abstract

From the advent of human civilization on our planet to modern urbanization, road networks have not only provided a means for transportation of logistics but have also helped us to cross cultural boundaries. The properties of road networks vary considerably from one geographical environment to another. The networks pertaining in a satellite image can therefore be used to classify and retrieve such environments. In this work, we have defined several such environments, and classified them using geometrical and topological features computed from the road networks occurring in them. Due to certain limitations of these extraction methods there was a relative failure of network extraction in some urban regions containing narrow and dense road structures. This loss of information was circumvented by segmenting the urban regions and computing a second set of geometrical and topological features from them.

The small images forming our database were extracted from images acquired from the SPOT5 satellite with 5m resolution (each image of size 512x512 pixels). The set of geometrical and topological features computed from the road networks and urban regions are used to classify the pre-defined geographical classes. In order to avoid the burden of feature dimensionality and reduce the classification performance, feature selection was performed using Fisher Linear Discriminant (FLD) analysis and an one-vs-rest linear Support Vector Machine (SVM) classification was performed on the selected feature set. The impact of spatial resolution and size of images on the feature set have been explored in this work. We took a closer look at the effect of spatial resolution and size of images on the discriminative power of the feature set to classify the images belonging to the pre-defined geographical classes. Tests were performed with feature selection by FLD and one-vs-rest linear SVM classification on a database with images of 10m resolution. Another test was performed with feature selection by FLD and one-vs-rest linear SVM classification on a database with 5m resolution images each of size 256x256 pixels.

With the above mentioned approaches, we developed a novel method to classify large satellite images with patches of images each of size 512x512 extracted from them acquired by SPOT5 satellite of 5m resolution. There has been a large amount of work dedicated to the classification of large satellite images at pixel level rather than considering image patches of different sizes. Classification of image patches of different sizes from a large satellite image is a novel idea in the sense that the patches considered contain significant coverage of a particular type of geographical environment.

Road networks and urban region features were computed from these image patches extracted from the large image. A one-vs-rest Gaussian kernel SVM classification method was used to classify this large image. The classification labels the image patches with the one having the maximum geographical coverage of the area associated in the large image. The large image was mapped into a "region matrix", where each element of the matrix corresponds to a geographical class. In certain cases, this

---

produces anomalies, as a single patch may contain two or more different geographical coverages. In order to have an estimate of these partial coverages, the output of the SVM was mapped into probabilities. These probability measures were then studied to have a closer look at the classification accuracies. The results confirm that our method is able to classify a large image into various geographical classes with a mean error of less than 10%.

**Keywords :** Satellite images, road networks, urban regions, classification, indexing

---

## Résumé

### Introduction

Au cours des dernières décennies, la croissance de larges bases de données d'images, en conjonction avec les progrès des techniques d'acquisition, a attiré des chercheurs de domaines variés pour travailler sur les systèmes d'extraction d'information à partir d'images. Ces images, en provenance de sources diverses, doivent être systématiquement analysées pour mettre en avant des informations importantes, souvent négligées par la perception humaine. L'avancée technologique des capteurs d'imagerie satellitaire et les nouveaux systèmes de stockage ont rendu les images trop grandes et complexes. L'annotation manuelle, pour la description complète d'images d'une telle complexité, n'est pas envisageable en pratique. L'indexation et la recherche dans les bases de données de télédétection reposent sur l'extraction préalable d'informations pertinentes à l'objet d'intérêt [Dasc 05]. L'indexation d'images satellitaires [Mait 07] dépend du choix des caractéristiques extraites qui elles-mêmes dépendent du type et de la résolution des capteurs. Par exemple, les descripteurs SIFT sont largement utilisés dans le domaine multimédia [Lowe 04]. Les indices complexes, comme l'indice de végétation (Vegetation index, NDVI), l'indice de brillance (Brightness index, BI) ou l'indice urbain (Urban index, ISU) sont utilisés pour les images multi-spectrales ou hyper-spectrales. Les caractéristiques de texture sont connues pour être très discriminatives pour les images panchromatiques moyenne résolution [Schr 98]. Les caractéristiques structurelles décrivant des objets d'origine humaine dans les images de haute résolution sont efficaces pour décrire le contenu des images [Bhat 07]. Le réseau routier contenu dans une image en est un exemple. Les propriétés des réseaux routiers varient considérablement d'un environnement géographique à un autre. Les caractéristiques structurelles extraites peuvent donc être utilisées pour classifier de tels environnements [Bhat 07]. Afin de calculer les caractéristiques structurelles du réseau routier, on doit d'abord extraire le réseau et le convertir dans une représentation appropriée. Cette représentation doit être absolument indépendante de la méthode d'extraction. Les méthodes d'extraction de routes sont en général dépendantes de la résolution, et un algorithme précis et générique est très difficile à mettre en place. Les méthodes utilisées dans notre étude sont robustes pour de nombreuses caractéristiques de routes, mais échouent souvent face aux réseaux fins et complexes quasiment cachés dans les petites zones urbaines. Les difficultés d'extraction du réseau entraînent une mauvaise fiabilité des caractéristiques associées et donc une mauvaise classification des images contenant de telles zones. Afin d'obtenir des informations pertinentes sur de telles régions, une segmentation de ces régions est nécessaire. De nouvelles caractéristiques structurelles, calculées sur ces régions urbaines segmentées, ont été combinées avec les caractéristiques déjà disponibles sur le réseau routier. Cette combinaison a amené une amélioration de la classification des environnement géographiques.

Dans les images, les pixels fournissent l'information de plus bas niveau. Les valeurs

---

des pixels sont les mesures des capteurs satellites, correspondant à une région de la surface terrestre. L'information pixellique est à un niveau bien plus bas que la sémantique de l'objet ou la région d'intérêt. La classification d'images à partir des valeurs des pixels est difficile et coûteuse et n'est donc pas une stratégie efficace. Nous présentons une nouvelle méthodologie de classification d'images satellitaires à partir de patches extraits. L'originalité de la méthode réside dans le fait que les patches considérés sont significativement représentatifs d'une certaine classe géographique. La classification est réalisée par une approche de type "un contre tous" à l'aide d'une machine à vecteurs supports probabiliste à noyaux Gaussiens. Dans ce travail, nous définissons 7 classes: 2 classes urbaines (ville Américaine "Urban USA" et ville Européenne "Urban Europe"); 3 classes rurales (villages "Villages", montagnes "Mountains" et champs "Fields"); une classe pour les aéroports "Airports"; une classe commune "Common", qui peut être considérée comme une classe de rejet indiquant par exemple les images maritimes.

## Attributs structurels

Dans cette section, nous présentons quatre méthodes d'extraction de réseaux routiers, dont deux sont utilisées dans cette étude. Nous proposons une méthode pour représenter les réseaux extraits sous forme de graphes. Une méthode morphologique est proposée pour segmenter les zones urbaines. Enfin, nous décrivons les caractéristiques structurelles du réseau routier et des régions urbaines qui seront utilisées pour la classification et l'indexation des images satellitaires.

### Extraction du réseau routier et des régions urbaines

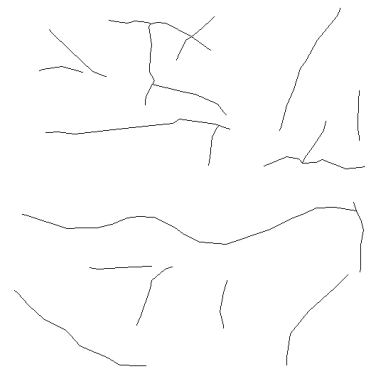
Afin de calculer les caractéristiques structurelles du réseau routier, nous devons d'abord extraire le réseau de l'image, et le convertir en une représentation appropriée. Cette représentation doit être indépendante du type de sortie de l'algorithme d'extraction, puisque nous ne voulons pas nous restreindre à une méthode unique. Dans les études préliminaires rapportées dans [Bhat 06], nous nous intéressons à deux méthodes d'extractions [Roch 03, Laco 05]. La méthode de [Roch 03] est basée sur des contours actifs d'ordre supérieur (Higher-Order Active Contours, HOACs) qui sont une généralisation des contours actifs classiques. La méthode de [Laco 05] modélise le réseau routier comme un processus objet, où les objets interagissant sont des segments linéiques. La sortie de l'algorithme est un ensemble de segments de droites de longueurs, orientations et positions variées. Cette sortie est transformée, via une dilatation, vers une représentation similaire à celle de [Roch 03], c'est-à-dire une image binaire. Malgré de bons résultats, ces méthodes n'ont pas été utilisées dans cette étude car elles ne sont pas adaptées aux grandes bases de données images. Une intervention manuelle est nécessaire pour régler les paramètres en fonction de la complexité de chaque image.

---

Dans cette étude, nous considérons les méthodes de [Fisc 81, Deso 00]. Ces méthodes sont plus aisées d'utilisation et furent facilement adaptées aux bases de données images. Les paramètres furent fixés avec succès suivant la résolution des images considérées. La sortie de la méthode décrite dans [Fisc 81] est une image binaire, qui, après une transformée en distance, peut être fournie directement comme entrée pour notre méthode. La Figure 1(b) donne des exemples de réseaux extraits. La sortie de la méthode de [Deso 00] est une liste de segments. Afin d'obtenir une entrée appropriée pour notre méthode, nous convertissons cette liste en une image binaire, puis traitons cette image pour obtenir des segments simplement connectés. La Figure 1(d) donne des exemples de réseaux extraits par cette méthode. Enfin, nous calculons une transformée en distance, utilisée comme entrée de notre méthode.



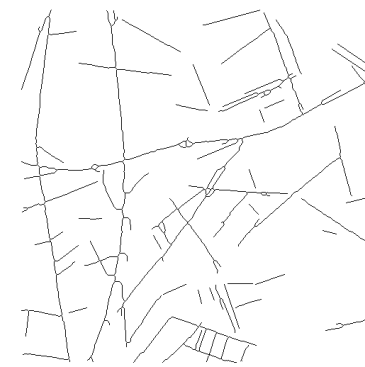
(a) Image originale ©CNES



(b) Résultats de l'extraction



(c) Image originale ©CNES



(d) Résultats de l'extraction

Figure 1: Résultats d'extraction obtenus avec les 2 méthodes considérées. L'exemple (b) est obtenu avec la méthode de [Fisc 81] et l'exemple (d) avec la méthode de [Deso 00]

Pour les deux méthodes, la fonction de distance obtenue est transformée en un graphe du réseau routier, pour le calcul des caractéristiques. Le graphe lui-même donne la topologie du réseau tandis que la géométrie est encodée via des attributs sur les noeuds et les arcs. Le graphe est produit par le calcul des points de choc de la fonction de distance, suivant la méthode de [Dimi 00, Sidd 02], étendue avec l'algorithme de recherche en profondeur (depth-first search, DFS [Corm 01]) pour gérer les com-



posantes connexes. La méthode extrait les points de choc grâce au caractère limitant du flux sortant moyen de la fonction de distance, lorsque les régions contenant les points de choc rétrécissent vers zéro. Un seuillage approprié de ce flux donne une approximation de la localisation des points de choc. Le graphe est construit en prenant les points triples (exceptionnellement, de plus haut degré) et les points terminaux comme noeuds. Ces points correspondent aux jonctions et terminaisons de routes. Les autres points forment les arcs, correspondant aux segments de routes. La Figure 2 donne un exemple de cette représentation. Le réseau routier de la Figure 2(b) est d'abord extrait de l'image de la Figure 2(a). Les méthodes citées précédemment sont alors utilisées pour générer les points de choc (Figure 2(c)), convertis en graphe (Figure 2(d)). Les noeuds et arcs sont renseignés avec des informations géométriques calculées à partir des points de choc. Les caractéristiques structurelles sont alors calculées grâce à ce graphe et ses attributs.

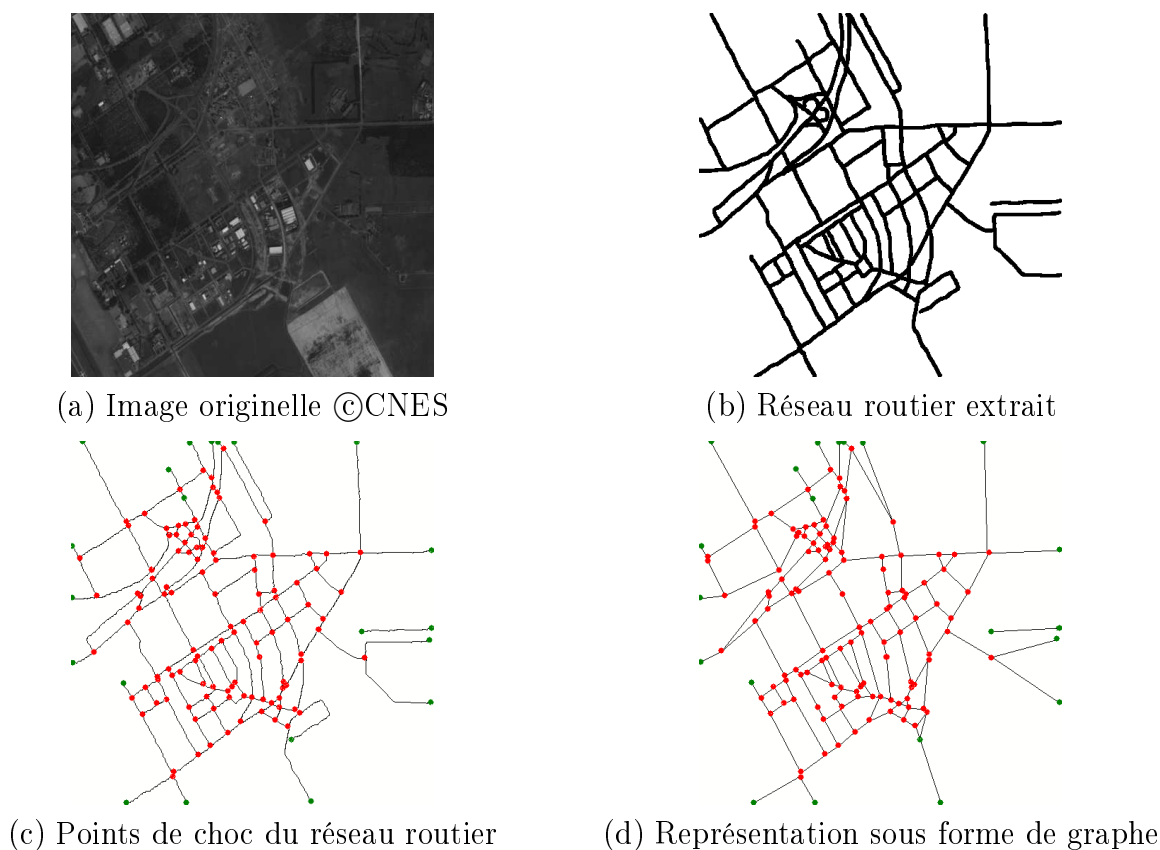


Figure 2: Exemple de représentation sous forme de graphe.

Les méthodes utilisées dans notre étude sont robustes pour de nombreuses caractéristiques de routes, mais échouent souvent face aux réseaux fins et complexes quasiment cachés dans les petites zones urbaines. Les difficultés d'extraction du réseau entraînent une mauvaise fiabilité des caractéristiques associées et donc une mauvaise classification des images contenant de telles zones. Afin d'obtenir des informations pertinentes sur de telles régions, une segmentation de ces régions est nécessaire.

L'hétérogénéité et la complexité géométrique des structures urbaines dans les images basse résolution (2 ou 5m) entraînent des effets de texture pour les objets de quelques pixels de large. Dans notre étude, nous utilisons le travail de [Roux 92] développé pour les régions urbaines dans les images SPOT. Dans les images SPOT, les zones urbaines apparaissent fortement texturées et l'extraction de ces régions est essentiellement un problème de différenciation de textures. La méthode utilisée ici est inspirée des travaux de [Sere 89] et [Khat 89]. L'idée principale est d'extraire les zones de fortes densités de pics brillants et sombres. Les techniques utilisées sont des opérations d'ouvertures et fermetures morphologiques. Les régions urbaines compactes segmentées sont illustrées par la Figure 3(b) et la Figure 3(d).

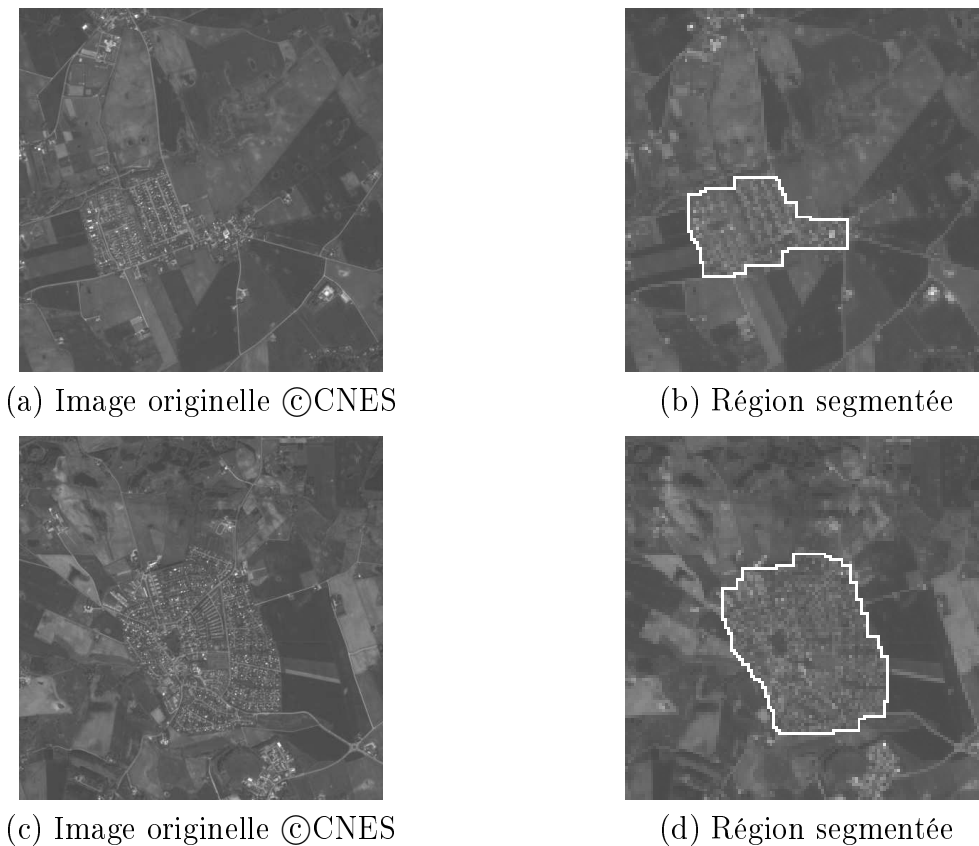


Figure 3: Images contenant des petites zones urbaines, et leurs segmentations.

## Caractéristiques des réseaux routiers

Nous étudierons dans cette section 16 caractéristiques résumées dans le tableau 1 (Les caractéristiques utilisées lors de la classification sont notées en gras). Ces caractéristiques peuvent être séparées en six groupes : six mesures de densité, quatre de courbures, deux d'homogénéité, une de longueur, deux de distributions et une d'entropie. Nous allons maintenant définir les caractéristiques des réseaux routiers.

Soient  $v$  un sommet et  $e$  une arête. Soit  $l_e$  la longueur du segment de route correspondant à  $e$ , et soit  $d_e$  la longueur de  $e$ , qui est la distance euclidienne entre deux sommets. Soit  $m_v = \sum_{e:v \in e} 1$  le nombre d'arêtes connectés à un sommet. Alors  $N_J = \sum_{v:m_v > 2} 1$  est le nombre total de jonctions de sommets et  $E_J = \sum_{m_v > 2} m_v$  est le nombre de jonctions d'arêtes. Soit  $\Omega$  l'aire de l'image en pixels. Nous définissons la densité de jonction de sommets par  $\tilde{N}_J = \Omega^{-1} N_J$  et la densité de jonction d'arêtes par  $\tilde{E}_J = \Omega^{-1} E_J$ . Ceux-ci forment une mesure utile pour séparer les zones urbaines des zones rurales : les zones urbaines ont vraisemblablement des valeurs  $\tilde{N}_J$  et  $\tilde{E}_J$  plus élevées que les zones rurales. De la même manière, nous définissons la longueur du réseau par  $L = \sum_e l_e$  et la densité de longueur par  $\tilde{L} = \Omega^{-1} \sum_e l_e$ . Encore une fois, nous estimons que la valeur  $\tilde{L}$  est plus élevée dans les zones urbaines. Notons qu'il peut arriver que la valeur  $\tilde{L}$  soit élevée et que  $\tilde{N}_J$  possède une faible valeur, ceci si les jonctions sont complexes and les segments de route sont "space-filling". Nous calculons également l'aire du réseau  $\Omega_L$  comme le nombre de pixels correspondant au réseau à partir des images binaires extraites and nous définissons la densité d'aire telle que  $\tilde{A} = \Omega^{-1} \Omega_L$ . Comme nous pouvons le constater sur la Figure 2, beaucoup de points de jonction sont regroupés dans une petite aire du réseau. Afin d'obtenir un caractéristique locale de la densité de jonction, nous définissons une mesure de densité de jonction dans une région circulaire telle que  $\tilde{N}_{r,j} = \Omega_{j,r}^{-1} \sum_{v \in \Omega_{j,r}: m_v > 2} 1$ . Ceci correspond à la densité de jonction des points appartenant à une région circulaire de rayon  $r$  centrée sur un point de jonction  $j$ . Nous calculons alors la moyenne et la variance de ces densités de jonction sur tous les points de jonction,  $\text{moyenne}(\tilde{N}_{r,j})$  et  $\text{var}(\tilde{N}_{r,j})$ . Une valeur élevée de la  $\text{var}(\tilde{N}_{r,j})$  signifie que les jonctions de route ont une structure étendue. Les structures de réseaux ruraux correspondent à une telle caractéristique. Une valeur faible indique que les points de jonction sont proches les uns de autres, ce qui est une mesure préminente des structures de réseaux urbains. La  $\text{moyenne}(\tilde{N}_{r,j})$  est également utilisée comme une mesure de densité.

Soient  $p_e = l_e/d_e$  et  $k_e = l_e^{-1} \int_e ds |k_e(s)|$ , c'est à dire, la valeur absolue de la courbure par unité de longueur du segment de route correspondant à  $e$ . Bien qu'il semble naturel de caractériser le réseau en utilisant la moyenne des valeurs par arêtes de ces quantités, il s'est révélé en pratique que les variances de ces quantités sont tout aussi utiles. Nous définissons ainsi le ratio de la variance des longueurs et le ratio de la moyenne des longueurs comme la variance et la moyenne de  $p_e$  sur les arêtes,  $\text{var}(p)$  et  $\text{moyenne}(p)$ , et la valeur moyenne de la variance de courbure et la valeur moyenne de la moyenne de courbure comme la variance et la moyenne de  $k_e$  sur les arêtes,  $\text{var}(k)$  et  $\text{moyenne}(k)$ . Notons qu'il est possible d'avoir une valeur  $p_e$  élevée pour une arête et en même temps une valeur  $k_e$  faible si le segment de route est composé de longs segments droits, et vice-versa, si une route 'serpente' rapidement autour de la ligne droite liant les deux sommets dans l'arête. Nous nous attendons à obtenir des valeurs fortes pour une de ces deux quantités dans le cas des zones rurales, alors que les zones urbaines auront probablement de faible valeurs, bien que ceci soit moins évident que dans les cas des mesures de densité.

Pour la mesure d'homogénéité de réseau, nous divisons chaque image en quatre

quadrants, étiquetés  $a$ . L'indice  $a$  indique les quantités évaluées pour le quadrant  $a$  plutôt que l'image entière. Soit  $M_{J,a} = \sum_{v \in a: m_v > 2} m_v$  le nombre d'arêtes émanant des jonctions dans le quadrant  $a$ . C'est très proche du double du nombre d'arêtes dans  $a$ , mais il est préférable de nous restreindre aux jonctions pour éviter les fausses terminaisons aux frontières de l'image. Soit  $\tilde{M}_{J,a} = \Omega_a^{-1} M_{J,a}$  la densité d'une telle arête dans le quadrant  $a$ . Nous définissons alors le réseau d'inhomogénéité comme la variance de  $\tilde{M}_{J,a}$  sur les quadrants,  $\text{var}(\tilde{M}_J)$ . Nous incluons également la caractéristique moyenne( $\tilde{M}_J$ ).

Dans le but de distinguer les deux classes urbaines (Etats-Unis et Europe), l'entropie de l'histogramme des angles des jonctions,  $H_\beta$ , ou  $\beta_j$  est le vecteur des angles entre les segments des routes au jonction  $j$ , est une bonne mesure. D'après les caractéristiques physiques des structures de réseaux routiers, les routes aux Etats-Unis tendent à être parallèles et à se croiser orthogonalement formants des jonctions en T ou des croisements de routes, alors que les routes européennes serpentent et se croisent entre elles dans les environs. Ainsi il semble naturel que  $H_\beta \leq 2$  bits soient nécessaires pour encoder l'information des segments de routes aux jonctions pour les réseaux routiers aux Etats-Unis, alors que dans le cas de l'Europe,  $H_\beta \geq 2$  bits sont nécessaires. La même mesure peut également être utilisée pour distinguer les montagnes des champs, lorsque la densité des caractéristiques distingue les réseaux ruraux des réseaux urbains.

Une mesure de distribution des arêtes à un sommet nous informe de la manière dont les arêtes à un sommet sont distribuées dans le réseau. Soit  $E_{D,i}$  la proportion des points de jonction avec  $i$  arêtes. Nous employons comme caractéristiques moyenne( $E_{D,i}$ ) et  $\text{var}(E_{D,i})$ . La variance de la distribution d'arêtes est plus faible dans le cas des réseaux urbains aux Etats-Unis par rapport à ceux existant en Europe.

## Caractéristiques des régions urbaines

Nous nous focalisons sur les quatre dernières caractéristiques du tableau 2 (Les caractéristiques utilisées lors de la classification sont notées en gras). Ces caractéristiques nous permettent de distinguer les classes rurales (villages et champs) des classes urbaines (Europe), qui autrement ne seraient pas bien classifiés par manque d'information du réseau extrait dans les petites régions urbaines denses des images, ce qui est visible dans les Figures 3(a) et 3(c). Soient  $\Omega$  l'aire de l'image et  $\Omega_R$  l'aire des régions extraites, soient  $L_\Psi$  la longueur du réseau dans  $\Psi = \Omega - \Omega_R$  et  $\Gamma_R$  le périmètre de la région extraite. Nous définissons deux descripteurs,  $\tilde{R}_A = \Omega^{-1} \Omega_R$ , la densité de région extraite et  $Cf_A = \Omega_R^{-1} \Gamma_R^2$ , le facteur de compacité de la région extraite. Ces deux caractéristiques nous aident à distinguer les classes de villages des autres classes: par exemple,  $\tilde{R}_A \simeq 1$  pour les classes urbaines et  $\tilde{R}_A \simeq 0$  pour les classes de montagnes et de champs.

La caractéristique  $\tilde{R}_v$ , le nombre de régions urbaine dans une image est utilisée pour distinguer les classes urbaines, villages, champs et montagnes. Une classe urbaine

Notation	Description
$m$	Nombre d'arêtes dans le graphe
$n$	Nombre de sommets dans le graphe
$\Omega$	Aire de l'image
$\Omega_L$	Aire du réseau
$a$	Label du quadrant
$l_e$	Longueur d'un segment de route correspondant à l'arête $e$
$m_v$	Nombre d'arêtes à un sommet $\sum_{e:v \in e} 1$
$N_J$	Nombre de jonctions de sommets $\sum_{v:m_v > 2} 1$
$\tilde{N}_J$	<b>Densité de jonction</b> $\Omega^{-1} N_J$
$L$	<b>Longueur du réseau</b> $\sum_e l_e$
$\tilde{L}$	<b>Densité de longueur</b> $\Omega^{-1} \sum_e l_e$
$\tilde{A}$	<b>Densité d'aire du réseau</b> $\Omega^{-1} \Omega_L$
$d_e$	Distance euclidienne entre les sommets dans une arête
$p_e$	Ratio des longueurs $l_e/d_e$
$\text{var}(p)$	<b>Ratio de la variance des longueurs</b> $m^{-1} \sum_e p_e^2 - (m^{-1} \sum_e p_e)^2$
$\text{moyenne}(p)$	<b>Ratio de la moyenne des longueurs</b> $m^{-1} \sum_e p_e$
$k_e$	Valeur moyenne de courbure $l_e^{-1} \int_e ds  k_e(s) $
$\text{var}(k)$	<b>Valeur moyenne de la variance de courbure</b> $m^{-1} \sum_e k_e^2 - (m^{-1} \sum_e k_e)^2$
$\text{moyenne}(k)$	<b>Valeur moyenne de la moyenne de courbure</b> $m^{-1} \sum_e k_e$
$E_{D,i}$	Nombre de jonctions avec $m_v = i$
$\text{var}(E_{D,i})$	<b>Variance de la distribution d'arêtes</b> $(1/\max(m_v)) \sum_i E_{D,i}^2 - ((1/\max(m_v)) \sum_i E_{D,i})^2$
$\text{moyenne}(E_{D,i})$	<b>Moyenne de la distribution d'arêtes</b> $(1/\max(m_v)) \sum_i E_{D,i}$
$E_J$	Nombre de jonctions d'arêtes $\sum_{m_v > 2} m_v$
$M_{J,a}$	Nombre de jonctions d'arêtes par quadrant $\sum_{v \in a: m_v > 2} m_v$
$\tilde{E}_J$	<b>Densité des arêtes de jonction</b> $\Omega^{-1} E_J$
$\tilde{M}_{J,a}$	Densité des arêtes de jonction par quadrant $\Omega_a^{-1} M_{J,a}$
$\text{var}(\tilde{M}_J)$	<b>Variance de la densité des arêtes de jonction</b> $(1/4) \sum_a \tilde{M}_{J,a}^2 - ((1/4) \sum_a \tilde{M}_{J,a})^2$
$\text{moyenne}(\tilde{M}_J)$	<b>Moyenne de la densité des arêtes de jonction</b> $(1/4) \sum_a \tilde{M}_{J,a}$
$\Omega_r$	Aire d'une région circulaire de $r$
$\tilde{N}_{r,j}$	Densité de jonction dans une région circulaire $\Omega_{j,r}^{-1} \sum_{v \in \Omega_{j,r}: m_v > 2} 1$
$\text{var}(\tilde{N}_{r,j})$	<b>Variance des densités de jonction sur toutes les régions circulaires</b> $(1/N_J) \sum_j \tilde{N}_{r,j}^2 - ((1/N_J) \sum_j \tilde{N}_{r,j})^2$
$\text{moyenne}(\tilde{N}_{r,j})$	<b>Moyenne des densités de jonction sur toutes les régions circulaires</b> $(1/N_J) \sum_j \tilde{N}_{r,j}$
$\beta_j$	Vecteur d'angles entre les segments à la jonction $j$
$H_\beta$	<b>Entropie de l'histogramme des angles de segments de route avec une taille de pas de discrétisation 30°</b>

Table 1: Une sommaire des caractéristiques calculées dans les réseaux routiers.

complète aura  $\tilde{R}_\nu = 1$ , alors qu'un village aura  $\tilde{R}_\nu > 1$  et les champs et montagnes auront  $\tilde{R}_\nu = 0$ . Une autre caractéristique,  $\Delta_\Omega$ , l'inverse de la densité de longueur, est également calculée pour séparer la classe village des classes urbaines, montagnes et champs. La longueur totale du réseau dans  $\Psi = \Omega - \Omega_R$  est dénoté par  $L_\Psi$ . Pour les classes urbaines complètes (Etats-Unis et Europe),  $L_\Psi = 0$  et pour les classes montagnes et champs  $\Psi = \Omega$ . Donc pour les classes montagnes et champs, l'inverse fractionnaire de la densité de longueur,  $\Delta_\Omega = 0$  alors que pour les classes urbaines complètes,  $\Delta_\Omega = \infty$ , et pour les classes villages  $0 < \Delta_\Omega < \infty$ . Ces caractéristiques de région urbaine viennent compléter les caractéristiques calculées à partir du graphe de représentation du réseau routier, comme décrit dans la section précédente, dans le but d'améliorer la classification des environnements géographiques qui, autrement, ne seraient pas bien classifiés à cause du manque d'information dans les petites régions urbaines denses.

Notation	Description
$\Omega$	Aire de l'image
$\Omega_R$	Aire des régions extraites
$L_\Psi$	Longueur du réseau dans $\Psi = \Omega - \Omega_R$
$\Gamma_R$	Périmètre des régions extraites
$\tilde{R}_A$	<b>Densité d'aire d'une région</b> $\Omega^{-1}\Omega_R$
$Cf_A$	<b>Facteur d'agglomération d'une région</b> $\Omega_R^{-1}\Gamma_R^2$
$\tilde{R}_\nu$	<b>Etiquette d'une région</b> $\#R$
$\Delta_\Omega$	<b>Inverse fractionnaire de la densité de longueur</b> $\frac{\Omega_R}{L_\Psi}$

Table 2: Sommaire des caractéristiques calculées pour les régions urbaines.

## Classification

Les 32 primitives (16 primitives pour chacune des deux méthodes d'extraction du réseaux routiers) décrites auparavant ont été calculées à partir d'une base de 497 images SPOT5, à une résolution de 5 mètres. Afin d'obtenir la vérité terrain, ces images ont été classifiées manuellement suivant 7 classes, représentant différents types de zones urbaines et rurales. La classification automatique a été réalisée avec cinq boucles de validation croisée sur les données, dont 80% sont utilisées pour l'apprentissage et les 20% restant pour les tests. Après avoir sélectionné les primitives via une analyse par FLD (Fisher Linear Discriminant) [Duda 00], la classification est effectuée par SVM à noyau linéaire sur l'ensemble des primitives sélectionnées. Le tableau 3 présente les résultats de cette classification. La classification SVM à noyau linéaire sur l'espace à 30 dimensions des primitives sélectionnées par FLD donne une erreur moyenne de 24.5% avec un écart type de 2.92%.

Comme nous pouvons clairement le constater grâce à la matrice de confusion du

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.55</b>	0.09	0.22	0.05	0.13	0.00	0.00
Mountains	0.10	<b>0.81</b>	0.00	0.00	0.05	0.00	0.02
Fields	0.19	0.05	<b>0.64</b>	0.05	0.18	0.00	0.00
USA	0.06	0.00	0.04	<b>0.82</b>	0.05	0.00	0.02
Europe	0.09	0.05	0.11	0.07	<b>0.60</b>	0.03	0.05
Airports	0.00	0.00	0.00	0.01	0.00	<b>0.97</b>	0.00
Common	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.91</b>

Table 3: Matrice de confusion de la classification SVM à noyau linéaire de 497 images avec 7 classes, 30 primitives parmi 32 étant sélectionnées par FLD.

tableau 3, la classe "Villages" est confondue avec la classe "Fields", et il y a également une légère confusion entre les classes urbaines "USA" et "Europe". Ces confusions apparaissent, comme nous l'avons évoqué précédemment, en raison de l'échec des méthodes d'extraction de réseaux routiers dans la détection des routes fines et densément structurées présentes dans certaines images. Le tableau 4 montre les résultats de la classification des mêmes images avec 20 primitives sélectionnées parmi 36 (32 primitives de réseaux routiers combinées avec 4 primitives calculées sur les régions urbaines segmentées). Il y a une nette amélioration de la matrice de confusion, notamment entre les classes Villages et Fields qui sont moins confondues qu'auparavant. L'erreur de classification SVM à noyau linéaire est très fortement réduite, passant de 24.5% avec les seules primitives de réseaux routiers à 12.9% avec les primitives combinées, l'écart-type étant de 3.29%. Ceci est dû au fait que l'information sur les zones urbaines (perdue avec les primitives de réseaux) est bien modélisée par les primitives structurelles calculées sur ces zones.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.83</b>	0.00	0.15	0.00	0.05	0.02	0.03
Mountains	0.04	<b>0.83</b>	0.01	0.00	0.00	0.00	0.00
Fields	0.04	0.08	<b>0.82</b>	0.01	0.00	0.00	0.01
USA	0.01	0.00	0.00	<b>0.92</b>	0.12	0.02	0.01
Europe	0.08	0.04	0.02	0.07	<b>0.84</b>	0.02	0.02
Airports	0.00	0.05	0.00	0.00	0.00	<b>0.96</b>	0.00
Common	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.93</b>

Table 4: Matrice de confusion de la classification SVM à noyau linéaire de 497 images avec 7 classes, 20 primitives parmi 36 étant sélectionnées par FLD.

## Dépendance des primitives élémentaires en fonction de la résolution et de la taille des images

Depuis l'avènement des capteurs modernes, l'acquisition des images satellitaires en termes de résolution et de taille est devenu un challenge pour l'extraction de l'information et l'interprétation des images à haute résolution. Les satellites à haute résolution tels que Quickbird2 et Ikonos avec une résolution panchromatique sub-métrique, ou les futurs satellites Pléiades avec une résolution spatiale proche du mètre, fournissent les informations adéquates sur les attributs des structures au sol des régions urbaines. Afin d'extraire ces informations sur une base d'images multi-résolution, une seule méthode ne sera pas suffisante pour cette tâche.

Plusieurs études ont été menées pour évaluer la résolution optimale pour la classification d'images satellitaires. Certains objets sont mieux classés avec une résolution plus fine, alors que d'autres nécessitent une résolution plus grossière. En particulier, une étude [Marc 94] explique les effets de la résolution sur la classification. Des travaux ont également porté sur la classification multi-résolution. Des primitives calculées à différentes résolutions spatiales ont été utilisées pour améliorer la classification de la couverture/utilisation des terres [Chen 02]. Dans notre approche, la méthode d'extraction des réseaux est optimisée pour une résolution à 5m, et va par conséquent extraire des structures redondantes dans la même scène à une résolution de 1m, et ne pas en extraire suffisamment dans la même scène à 10m de résolution. Les paramètres d'extraction sont en effet fixés pour une certaine résolution, et ne peuvent être dynamiquement adaptés pour des résolutions variables.

### **Influence de la résolution et de la taille des images sur l'extraction de réseaux routiers et de régions urbaines**

Dans toutes nos expérimentations, nous avons utilisé une base d'images SPOT5 fournie par le CNES à une résolution de 5m. La classification d'images aux différentes résolutions mentionnées n'est pas l'objet de cette étude, et de plus, le nombre d'images dont nous disposons pour ces résolutions n'est pas suffisant pour obtenir une base acceptable pour réaliser une classification. Afin de comprendre le rôle de la résolution et de la taille sur les primitives lors de la classification d'images appartenant aux classes mentionnées précédemment, nous construisons deux bases de données à partir des images SPOT5 dont nous disposons.

Les deux nouvelles bases sont obtenues à partir de 512x512 pixels à 5m de résolution. Les images sont sous-échantillonnées à une résolution de 10m pour la première base, et les images sont découpées en 4 sous-images de 256x256 pixels pour la seconde. Cette dernière base est donc quatre fois plus grande que l'originale, chacune des sous-images étant ré-attribuée manuellement aux classes définies pour la classification.

---



## Classification

Les primitives du graphe représentant le réseau routier et la zone urbaine sont calculées en suivant la méthode détaillée aux précédemment. Deux tests sont réalisés sur les bases respectives de 497 images SPOT5 256x256 pixels à 10m de résolution, et de 1988 images SPOT5 256x256 pixels à 5m de résolution. La classification automatique est réalisée avec cinq boucles de validation croisées sur les données, dont 80% sont utilisées pour l'apprentissage et les 20% restant pour les tests.

Les résultats de la classification SVM à noyau linéaire, utilisant une approche un contre tous, des 1988 images à 5m sur 7 classes, avec 22 primitives sélectionnées parmi 36 par FLD, sont donnés dans le tableau 5. L'erreur moyenne de classification est de 17.4% avec un écart-type de 2.19%. Les résultats de la classification SVM à noyau linéaire des 497 images à 10m sur 7 classes, avec 22 primitives sélectionnées parmi 36 par FLD, sont donnés dans le tableau 6. L'erreur moyenne de classification est de 25.4% avec un écart-type de 2.02%. Le tableau 8 résume les résultats de classification pour les 3 bases d'images.

Comme nous pouvions nous y attendre, les performances varient fortement avec le changement de taille et de résolution des images. Les quelques anomalies visibles dans le tableau 5 peuvent être expliquées ainsi : d'une part, le fait d'avoir diminué la taille des images de 512x512 à 256x256 pixels limite l'information extraite sur les réseaux routiers et les primitives associées sont donc moins discriminantes pour les classes Villages, Mountains, Fields, USA et Europe. D'autre part, la classe Airports est beaucoup mieux attribuée en raison des pistes d'envol qui sont plus facilement détectées que dans les images 512x512 pixels, où d'autres structures sont détectées dans leur voisinage et limite le pouvoir discriminant des primitives extraites.

Quelques anomalies sont également constatées lors de la classification d'images à 10m. Les pertes de performance globales constatées sont dues à l'extraction des primitives. En effet, la méthode d'extraction de réseaux [Deso 00] ne parvient pas à fournir suffisamment d'information fiable sur la structure du réseau routier à cette résolution. De même, la méthode d'extraction de zones urbaine de [Roux 92] peut faillir à cette résolution. Seule la classe Common (contenant des images de mer) voit son taux d'erreur diminuer. Ceci résulte du fait qu'à cette résolution, les structures linéaires formées par les vagues ne sont plus extraites avec la méthode de [Deso 00], ce qui n'était pas le cas avec les images à 5m.

## Indexation de grandes images SPOT5

Une image est indexée par un ensemble de mots-clés représentant son contenu. Ces mots-clés sont dépendants des scénarios d'application et leur nombre est en général limité. La classification est souvent utilisée comme une étape de pré-traitement pour

---

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.66</b>	0.06	0.19	0.02	0.15	0.00	0.00
Mountains	0.02	<b>0.82</b>	0.02	0.00	0.00	0.00	0.01
Fields	0.02	0.05	<b>0.74</b>	0.00	0.00	0.00	0.00
USA	0.17	0.00	0.00	<b>0.79</b>	0.13	0.00	0.00
Europe	0.12	0.06	0.04	0.19	<b>0.71</b>	0.00	0.00
Airports	0.02	0.00	0.00	0.00	0.00	<b>1.00</b>	0.00
Common	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.99</b>

Table 5: Matrice de confusion de la classification SVM à noyau linéaire de 1988 images de taille 256x256 pixels avec 7 classes, 22 primitives parmi 36 étant sélectionnées par FLD.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.53</b>	0.12	0.32	0.03	0.03	0.01	0.00
Mountains	0.08	<b>0.65</b>	0.03	0.02	0.01	0.04	0.00
Fields	0.19	0.08	<b>0.59</b>	0.00	0.00	0.00	0.00
USA	0.16	0.00	0.02	<b>0.82</b>	0.16	0.03	0.00
Europe	0.03	0.10	0.03	0.14	<b>0.79</b>	0.01	0.00
Airports	0.01	0.05	0.02	0.00	0.00	<b>0.90</b>	0.00
Common	0.00	0.00	0.00	0.00	0.00	0.00	<b>1.00</b>

Table 6: Matrice de confusion de la classification SVM à noyau linéaire de 497 images de résolution 10m avec 7 classes, 22 primitives parmi 36 étant sélectionnées par FLD.

	497/5m/512x512	1988/5m/256x256	497/10m/256x256
Villages	<b>0.83</b>	<b>0.66</b>	<b>0.53</b>
Mountains	<b>0.83</b>	<b>0.82</b>	<b>0.65</b>
Fields	<b>0.82</b>	<b>0.74</b>	<b>0.59</b>
USA	<b>0.92</b>	<b>0.79</b>	<b>0.82</b>
Europe	<b>0.84</b>	<b>0.71</b>	<b>0.79</b>
Airports	<b>0.95</b>	<b>1.00</b>	<b>0.90</b>
Common	<b>0.93</b>	<b>0.99</b>	<b>1.00</b>

Table 7: Tableau de comparaison des classes correctement identifiées.

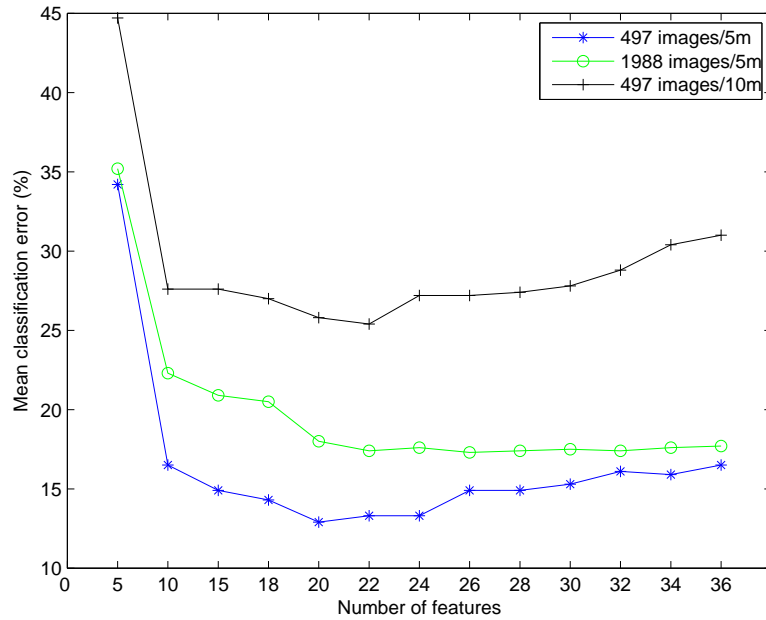


Figure 4: Taux d'erreur de classification en fonction du nombre de primitives.

Database and Resolution	Feature Dimension	Selection	Classification Error (%)
497/5m	36,7 classes	Fisher LD, 20	Linear SVM, $12.9 \pm 3.29$
1988/5m	36,7 classes	Fisher LD, 22	Linear SVM, $17.4 \pm 2.19$
497/10m	36,7 classes	Fisher LD, 22	Linear SVM, $25.4 \pm 2.02$

Table 8: Performance de classification des différentes bases.

l'indexation. Indexer soigneusement une base d'images aide à récupérer efficacement le contenu de l'image. Le schéma de notre méthode d'indexation comporte les trois étapes suivantes :

### **Etape 1 : la base de données**

La base de données images peut être vue comme deux ensembles disjoints : l'un contenant les images ou les images segmentées, et l'autre, les primitives extraites des images. Soient  $S_I$ , l'ensemble d'images, et  $S_F$ , l'ensemble des primitives; un pointeur est utilisé entre  $S_I$  et  $S_F$  pour adresser une image aux primitives qui lui sont associées. Les images utilisées appartiennent à une grande archive de 497 images, chacune de taille 512x512 pixels et catégorisée en 7 classes. Les informations extraites de ces images, en termes de caractéristiques structurales, sont conservées dans un fichier de données. Le processus off-line de création de ce fichier de données est effectué une seule fois, et en cas de nouvelle entrée, les informations extraites de la nouvelle image sont ajoutées dans le fichier de données existant. L'adresse de la nouvelle entrée est convenablement affectée au pointeur. Ceci sera utilisé plus tard comme ensemble d'"apprentissage" pour la tâche de classification.

### **Etape 2: le fichier de primitives**

Etant donnée une image de grande taille, le processus off-line pour l'utilisateur est le suivant : l'image de taille 5120x5120 pixels est automatiquement découpée en morceaux de taille 512x512 pixels, ceci sans recouvrement. Pendant ce processus, un pointeur d'association est extrait des petites images, définissant leur position spatiale dans la grande image. L'extraction du réseau routier, sa représentation par graphes et les méthodes de segmentation de zones urbaines sont appliquées parallèlement sur l'ensemble des sous-images (100 images). Les caractéristiques structurales issues de la représentation par graphes et des zones urbaines sont stockées dans un fichier. Les images sont étiquetées *a priori* de manière aléatoire, les classes allant de 1 à 7. Ceci sera utilisé plus tard comme ensemble de "tests" pour la tâche de classification.

### **Etape 3: la classification**

Dans plusieurs travaux de classification d'images satellitaires, les informations *a priori* sur la configuration des étiquettes des classes sont disponibles et il est essentiel et crucial de combiner ces informations dans le processus de classification pour obtenir une réponse fiable. La méthode SVM classique ne fournit aucune estimation de la confiance de classification et par conséquent, ne nous permet pas de comprendre les informations *a priori*. La méthode SVM probabiliste nous fournit une solution consistant à constru-

---

ire un classificateur pour produire une probabilité a posteriori  $P(class = c|input)$  nous permettant de prendre une décision quantitative à propos de la classification [Plat 99]. Dans ce travail, nous utilisons un classificateur SVM à noyau Gaussien avec  $\sigma = 10$  et une approche un contre tous. Le choix du noyau Gaussien  $\sigma$ , qui contrôle la largeur du noyau, est difficile à estimer dans des situations pratiques. Dans cette étude, nous avons considéré la valeur du noyau,  $\sigma$ , qui minimisait l'erreur d'apprentissage.

Les résultats obtenus en sortie du SVM probabiliste peuvent être interprétés de la manière suivante : la sortie du classificateur doit être une probabilité a posteriori calibrée. L'apprentissage SVM est d'abord effectué, puis les paramètres  $A$  et  $B$  d'une fonction sigmoïde (voir Equation 1) sont estimés à partir de l'ensemble d'apprentissage  $(f_i, y_i)$  pour transformer la sortie du SVM en probabilités. L'étiquette prédite d'une image est celle qui a la plus grande probabilité. La sortie de la classification est représentée par une matrice ("Matrice de régions"), dans laquelle chaque élément correspond à une classe d'images. La grande image SPOT5 de Los Angeles, à 5m de résolution Figure 5(a) est bien classifiée, le pourcentage de bonne classification étant d'environ 85%.

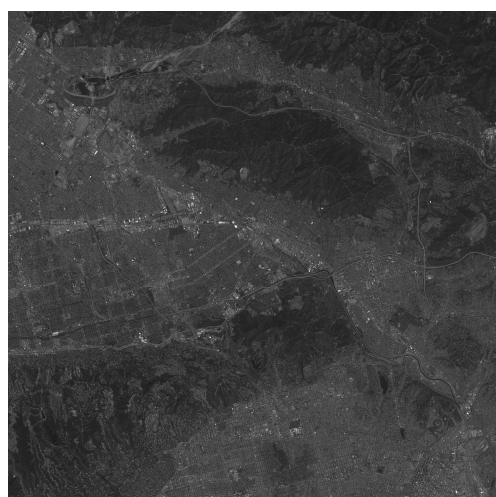
$$P(y = c|f) = \frac{1}{1 + \exp(Af + B)} \quad (1)$$

La sortie du SVM probabiliste dans la matrice de régions, Figure 5(b) montre que certaines zones sont classifiées comme des zones urbaines européennes. Ceci peut s'expliquer par le fait que soit les probabilités de classification sont faibles, soit les probabilités des régions voisines sont comparables. L'autre explication réside dans le fait que les structures du réseau dans ces zones sont similaires à celles qu'on trouve dans plusieurs régions urbaines d'Europe. La validation des régions classifiées avec les vérités de terrain de la Figure 5(d) est représentée par l'image de la Figure 5(c).

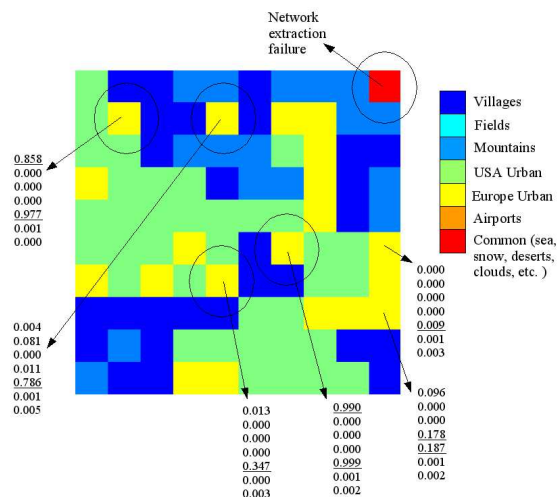
## Conclusion

La classification de grandes images satellitaires en découpant l'image en morceaux est une nouvelle idée dans le sens où les morceaux d'image considérés contiennent une couverture significative d'un type particulier d'environnement géographique. Le SVM probabiliste nous fournit une analyse quantitative de la classification. Cette méthode apporte une base pour des analyses plus complexes d'images satellitaires de grande taille. L'effet de recouvrement des morceaux d'images dans la classification n'est pas mentionné. Ceci pourrait être une étude intéressante dans la mesure où elle peut aider à mieux classifier les images. De plus, les morceaux d'images de tailles différentes peuvent aussi être utilisés pour améliorer les performances de la classification. Notre méthode d'indexation avec les perspectives mentionnées ci-dessus peut être adaptée à des systèmes actuels et futurs de fouille d'information dans les images pour les archives

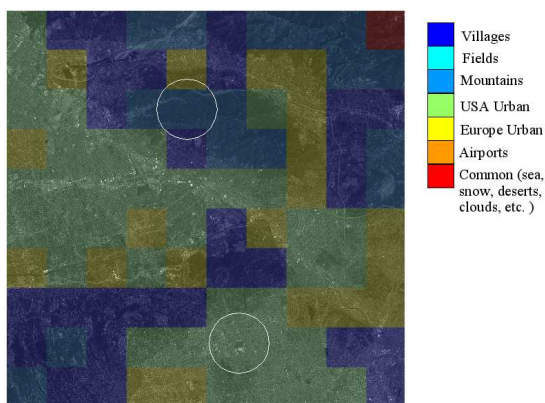
---



(a)



(b)



(c)



(d)

Figure 5: Classification d'une grande image de Los Angeles: (a) Image originale de Los Angeles ©CNES; (b) Résultats de classification pour Los Angeles; (c) Résultats de classification superposés à l'image originale de Los Angeles; (d) Vérité de terrain avec Google maps ©2007 Google-Imagery ©2007 NASA, TerraMetrics, Map data ©NAVTEQ<sup>TM</sup>, LeadDog Consulting

d'Observation de la Terre.

---

# Contents

<b>Contents</b>	<b>27</b>
<b>List of figures</b>	<b>31</b>
<b>List of tables</b>	<b>37</b>
<b>1 Introduction</b>	<b>41</b>
1.1 Image content retrieval and mining overview . . . . .	43
1.2 Which information ? . . . . .	45
1.3 The methodology . . . . .	46
1.4 Organization of the thesis . . . . .	46
<b>2 Image Information Mining Systems</b>	<b>49</b>
2.1 Systems . . . . .	49
2.2 The IKONA system . . . . .	50
2.3 The KIM system . . . . .	51
2.4 Image representation and similarity . . . . .	53
2.5 Image content . . . . .	59
2.5.1 Shape content of an image . . . . .	59
2.5.2 Texture content of an image . . . . .	60
2.6 Image classification . . . . .	61

---



---

2.6.1	Classification tools . . . . .	62
2.7	Discussion . . . . .	64
<b>3</b>	<b>Road Network Structural Information</b>	<b>65</b>
3.1	Introduction . . . . .	65
3.2	Network extraction and representation . . . . .	69
3.3	Preliminary study . . . . .	76
3.3.1	Feature selection and clustering . . . . .	78
3.4	Larger image database with refined classes and features . . . . .	81
3.4.1	New features from the graph . . . . .	83
3.4.2	Classification . . . . .	85
3.5	Discussion . . . . .	89
<b>4</b>	<b>Region Structural Information</b>	<b>91</b>
4.1	Urban region structures . . . . .	92
4.2	Segmentation of urban regions . . . . .	96
4.2.1	Features from the urban regions . . . . .	97
4.2.2	Classification . . . . .	100
4.2.3	2-Level classification . . . . .	103
4.2.4	Classification and retrieval with relevance feedback . . . . .	103
4.3	Discussion . . . . .	105
<b>5</b>	<b>Dependence of primitive features on image resolution and size</b>	<b>107</b>
5.1	Introduction . . . . .	107
5.2	Road network and urban region extraction for different resolutions . . .	108
5.3	Classification . . . . .	114
5.4	Discussion . . . . .	118

---

---

<b>6</b>	<b>Indexing of large satellite images</b>	<b>119</b>
6.1	Introduction . . . . .	120
6.2	Indexing . . . . .	121
6.2.1	Step 1: The database . . . . .	121
6.2.2	Step 2: The feature file . . . . .	121
6.2.3	Step 3: The classification . . . . .	122
6.2.4	Detailed analysis of 4 regions . . . . .	132
6.3	Indexing results on large satellite images . . . . .	134
6.3.1	Barcelona . . . . .	135
6.3.2	Detailed analysis of 2 regions . . . . .	139
6.3.3	Los Angeles . . . . .	141
6.3.4	Detailed analysis of 2 regions . . . . .	145
6.3.5	Madrid . . . . .	146
6.3.6	Detailed analysis of 2 regions . . . . .	150
6.3.7	Paris . . . . .	152
6.3.8	Detailed analysis of 2 regions . . . . .	156
6.4	Discussion . . . . .	157
<b>7</b>	<b>Conclusion</b>	<b>159</b>
<b>A</b>	<b>Personal Biography</b>	<b>163</b>
	<b>Bibliography</b>	<b>164</b>

---



---

## List of Figures

1	Résultats d'extraction obtenus avec les 2 méthodes considérées. L'exemple (b) est obtenu avec la méthode de [Fisc 81] et l'exemple (d) avec la méthode de [Deso 00] . . . . .	11
2	Exemple de représentation sous forme de graphe. . . . .	12
3	Images contenant des petites zones urbaines, et leurs segmentations. . .	13
4	Taux d'erreur de classification en fonction du nombre de primitives. . .	22
5	Classification d'une grande image de Los Angeles:(a) Image originale de Los Angeles ©CNES; (b) Résultats de classification pour Los Angeles; (c) Résultats de classification superposés à l'image originale de Los Angeles; (d) Vérité de terrain avec Google maps ©2007 Google-Imagery ©2007 NASA, TerraMetrics, Map data ©NAVTEQ <sup>TM</sup> , LeadDog Consulting . . . . .	25
1.1	Images from different satellites with different resolutions. . . . .	42
1.2	The schematic diagram of the large image indexing process. . . . .	47
2.1	Generic image database. Left: Selection of a query image(the red border indicates the selected image). Right: Query results for the selected image. Courtesy ( <a href="http://www-rocq.inria.fr/imedia/">http://www-rocq.inria.fr/imedia/</a> ) . . . . .	51
2.2	Specific Image database. Left: Selecting a query image with relevance feedback mechanism in IKONA. Right: Query results for the selected image. Courtesy ( <a href="http://www-rocq.inria.fr/imedia/">http://www-rocq.inria.fr/imedia/</a> ) . . . . .	52
2.3	The hierarchical image content description. Courtesy KIM project [Datc 99] . . . . .	53
2.4	Image selection. Courtesy (ESA, DLR, ETH Zurich and ACS) : Courtesy <a href="http://kes.esrin.esa.int/kes/">http://kes.esrin.esa.int/kes/</a> . . . . .	54

---

---

2.5	Results of query from the image. Courtesy (ESA, DLR, ETH Zurich and ACS) : Courtesy <a href="http://kes.esrin.esa.int/kes/">http://kes.esrin.esa.int/kes/</a> . . . . .	55
2.6	Interactive learning. Courtesy (ESA, DLR, ETH Zurich and ACS) : Courtesy <a href="http://kes.esrin.esa.int/kes/">http://kes.esrin.esa.int/kes/</a> . . . . .	56
2.7	Metadata based search applet. Courtesy (ESA, DLR, ETH Zurich and ACS) : Courtesy <a href="http://kes.esrin.esa.int/kes/">http://kes.esrin.esa.int/kes/</a> . . . . .	57
2.8	An example of road network extraction and graph representation . . . . .	58
3.1	Synthetic images and their corresponding graph representation . . . . .	67
3.2	Images categorized into different classes. . . . .	68
3.3	An example of an urban and a rural class ©CNES. . . . .	68
3.4	An example of 2 urban, 3 rural, airport and common (sea) classes ©CNES. . . . .	70
3.5	Examples from the two extraction methods: (b) method of [Roch 03]; (d) and (f) method of [Laco 05]. . . . .	72
3.6	Extraction results with the method of [Fisc 81]. . . . .	73
3.7	Extraction results with the method of [Deso 00]. . . . .	74
3.8	Graph representation . . . . .	75
3.9	An example of the graph representation. . . . .	76
3.10	Scatter plots of selected pairs of features from manually extracted networks. Red stars correspond to rural areas, blue circles to urban areas. From left to right, top to bottom: $\tilde{N}_J$ versus $var(k)$ ; $\tilde{L}$ versus $var(k)$ ; $\tilde{N}_J$ versus $var(p)$ ; $\tilde{L}$ versus $\tilde{N}_J$ ; $var(p)$ versus $var(k)$ ; $var(\tilde{M}_J)$ versus $\tilde{L}$ . . . . .	79
3.11	Classification error vs number of features from the two extraction methods of [Deso 00] and [Fisc 81]. . . . .	87
3.12	Classification error vs number of features combined from the two extraction methods of [Deso 00] and [Fisc 81]. . . . .	89
4.1	An example of network extraction failure in a small urban region. . . . .	92
4.2	Examples of small urban regions ©CNES. . . . .	93
4.3	Image texture analysis with window size variation. . . . .	95

---

---

4.4	Example of urban area extraction using the probabilistic method of [Desc 93] ©CNES/ENST. . . . .	96
4.5	Examples of opening and closing operators. . . . .	97
4.6	Examples of opening and closing operators ©CNES/ENST. . . . .	98
4.7	Images containing small urban areas and their segmentations. . . . .	99
4.8	Feature selected from the feature group. . . . .	101
4.9	Classification error vs number of features . . . . .	102
4.10	2-level hierarchical model . . . . .	103
4.11	Classification error vs number of features . . . . .	104
5.1	Images and network extracted from them at different resolutions. The method of [Deso 00] was used as an extraction method. . . . .	111
5.2	Images containing small urban areas and their segmentations at different resolutions. The method of [Roux 92] was used for the segmentation. . . . .	113
5.3	Feature selected from the feature group. Left: Feature selected from 36 features from 497 images with 5m resolution. Right: Feature selected from 36 features from 1988 images with 5m resolution. . . . .	115
5.4	Feature selected from the feature group. Left: Feature selected from 36 features from 497 images with 5m resolution. Right: Feature selected from 36 features from 497 images with 10m resolution. . . . .	115
5.5	Classification errors vs number of features . . . . .	118
6.1	Classification error vs number of features. . . . .	123
6.2	Image patches of size 512x512 pixels from a large SPOT5 image of Copenhagen with 5m resolution of size 5120x5120 pixels ©CNES. . . . .	125
6.3	Classification of image patches of size 512x512 pixels from a large image of Copenhagen of size 5120x5120 pixels ©CNES. Each colored pixel of the 10x10 matrix corresponds to a class to which an image patch of size 512x512 belongs. SVM with Gaussian kernel of $\sigma = 10$ . . . . .	126

---

---

6.4	Classification of image patches of size 512x512 pixels from a large image of Copenhagen of size 5120x5120 pixels ©CNES. Each colored pixel of the 10x10 matrix corresponds to a class to which an image patch of size 512x512 belongs. A probabilistic SVM with Gaussian kernel of $\sigma = 10$ . The values show the probabilities of each image patches belonging to a class. . . . .	127
6.5	Examples of ambiguous cases in classifications. . . . .	128
6.6	The algorithm for changing the labels. . . . .	128
6.7	Classification modified using neighborhood properties. Top: The modified labels, Bottom: The probabilistic SVM classification. . . . .	129
6.8	Classification results superimposed on the original image of Copenhagen.	130
6.9	Ground truth with Google maps ©2007 Google-Imagery ©2007 Terra-Metrics, Map data ©2007 Tele Atlas. . . . .	131
6.10	Probabilistic classification output with labels. . . . .	132
6.11	Probabilistic classification output with labels. . . . .	132
6.12	Probabilistic classification output with labels. . . . .	133
6.13	Probabilistic classification output with labels. . . . .	133
6.14	The experimental structure. . . . .	134
6.15	Original image of Barcelona ©CNES. . . . .	135
6.16	Classification results for Barcelona. . . . .	136
6.17	Classification results superimposed on the original image of Barcelona.	137
6.18	Ground truth with Google maps ©2007 Google-Imagery ©2007 Terra-Metrics, Map data ©2007 Tele Atlas. . . . .	138
6.19	Probabilistic classification output with labels. . . . .	139
6.20	Probabilistic classification output with labels. . . . .	139
6.21	Original image of Los Angeles ©CNES. . . . .	141
6.22	Classification results for Los Angeles. . . . .	142
6.23	Classification results superimposed on the original image of Los Angeles.	143

---

---

6.24	Ground truth with Google maps ©2007 Google-Imagery ©2007 NASA, TerraMetrics, Map data ©NAVTEQ <sup>TM</sup> , LeadDog Consulting. . . . .	144
6.25	Probabilistic classification output with labels. . . . .	145
6.26	Probabilistic classification output with labels. . . . .	145
6.27	Original image of Madrid ©CNES. . . . .	146
6.28	Classification results for Madrid. . . . .	147
6.29	Classification results superimposed on the original image of Madrid. . .	148
6.30	Ground truth with Google maps ©2007 Google-Imagery ©2007 TerraMetrics, Map data ©2007 Tele Atlas. . . . .	149
6.31	Probabilistic classification output with labels. . . . .	150
6.32	Probabilistic classification output with labels. . . . .	150
6.33	Original image of Paris ©CNES. . . . .	152
6.34	Classification results for Paris. . . . .	153
6.35	Classification results superimposed on the original image of Paris. . . .	154
6.36	Ground truth with Google maps ©2007 Google-Imagery ©2007 TerraMetrics, Map data ©2007 Tele Atlas. . . . .	155
6.37	Probabilistic classification output with labels. . . . .	156
6.38	Probabilistic classification output with labels. . . . .	156

---





---

## List of Tables

1	Une sommaire des caractéristiques calculées dans les réseaux routiers. . .	16
2	Sommaire des caractéristiques calculées pour les régions urbaines. . . .	17
3	Matrice de confusion de la classification SVM à noyau linéaire de 497 images avec 7 classes, 30 primitives parmi 32 étant sélectionnées par FLD.	18
4	Matrice de confusion de la classification SVM à noyau linéaire de 497 images avec 7 classes, 20 primitives parmi 36 étant sélectionnées par FLD.	18
5	Matrice de confusion de la classification SVM à noyau linéaire de 1988 images de taille 256x256 pixels avec 7 classes, 22 primitives parmi 36 étant sélectionnées par FLD. . . . .	21
6	Matrice de confusion de la classification SVM à noyau linéaire de 497 images de résolution 10m avec 7 classes, 22 primitives parmi 36 étant sélectionnées par FLD. . . . .	21
7	Tableau de comparaison des classes correctement identifiées. . . . .	21
8	Performance de classification des différentes bases. . . . .	22
3.1	Summary of features computed from road networks . . . . .	80
3.2	Kernel k-means ( $\sigma = 0.5$ ) clustering result with 5 features selected from the scatter plot analysis of pairs of feature vectors. . . . .	81
3.3	Kernel k-means clustering ( $\sigma = 0.5$ ) result with 6 out of 11 features selected by FLD analysis. A kNN classification gives 7.45% classification error. . . . .	81
3.4	Kernel k-means clustering ( $\sigma = 0.5$ ) result with 5 out of 11 features selected by the RelieF algorithm. A kNN classification gives 5.69% classification error. . . . .	81

---

---

3.5	The 5 features selected from the scatter plot analysis of pairs of features.	82
3.6	The 6 out of 11 features selected by FLD analysis. . . . .	82
3.7	The 5 out of 11 features selected by the ReliefF algorithm. . . . .	82
3.8	Summary of some more road network features . . . . .	84
3.9	Summarization of the feature groups. . . . .	85
3.10	Confusion matrix of an SVM linear kernel classification of 497 images with 7 classes with 15 out of 16 features from the graph representation of the road networks extracted with the method of [Deso 00]. . . . .	86
3.11	Confusion matrix of an SVM linear kernel classification of 497 images with 7 classes with 13 out of 16 features from the graph representation of the road networks extracted with the method of [Fisc 81]. . . . .	86
3.12	Confusion matrix of an SVM linear kernel classification of 497 images with 7 classes with 32 features from the graph. . . . .	88
3.13	Confusion matrix of an SVM linear kernel classification of 497 images with 7 classes with 30 out of 32 features selected by FLD. . . . .	88
3.14	Classification performance. . . . .	88
4.1	Summary of features computed for urban areas. . . . .	98
4.2	Confusion matrix of a SVM linear kernel classification on 497 images with 7 classes with 36 features. . . . .	101
4.3	Confusion matrix of a SVM linear kernel classification on 497 images with 7 classes with 20 out of 36 features selected by FLD. . . . .	102
4.4	Classification performance. . . . .	102
4.5	Confusion matrix of an SVM linear kernel classification on 497 images with 3 classes with 22 out of 36 features selected by FLD. . . . .	103
4.6	Classification with relevance feedback and memory. . . . .	104
5.1	Confusion matrix of a SVM linear kernel classification on 497 images with 7 classes with 20 out of 36 features selected by FLD. The original database, with images of resolution 5m. . . . .	116

---

---

5.2	Confusion matrix of a SVM linear kernel classification on 1988 images with 7 classes with 22 out of 36 features selected by FLD. The database of 1988 images of size 256x256 pixels. . . . .	116
5.3	Confusion matrix of a SVM linear kernel classification on 497 images with 7 classes with 22 out of 36 features selected by FLD. The images were downsampled to 10m resolution. . . . .	117
5.4	An overall comparison table of "good" classified classes. . . . .	117
5.5	Classification performance. . . . .	117
6.1	Confusion matrix of an Gaussian SVM classification of 497 images with 7 classes with 20 out of 36 features selected by FLD. The original database, with images of resolution 5m. . . . .	122

---



# Chapter 1

## Introduction

The recent advances in large data storage and new satellite technology acquiring high volumes of images have enabled the creation of large image archives. In order to deal with these high volumes of data, it is necessary to develop appropriate information systems to efficiently manage these collections. Image indexing and retrieval from these large image databases is one of the most important services that need to be supported by such systems. In remote sensing and astronomy, large amounts of image data are received daily by ground stations for processing, analyzing and archiving. The need of processing, analyzing and archiving of images has been identified in applications such as cartography and meteorology for instance. Users from various domains require information or services that are precise, low cost and timely and where the forms and formats of the data are compatible with user's need. Earth Observation (EO) archiving centres generally offer data, images and other low level products. User's needs are partially satisfied by a number of, usually small, companies applying time consuming processing techniques which are mostly manual. The end products rely largely on the knowledge of certain experts to extract information from these data or images. The manual processes will be even more difficult with the greater sophistication of user's needs, requiring, for example, the fusion of multi-sensors or EO and non-EO data, and the exponential increase in the volume and the complexity of data archives due to the rapid increase in number of sensors, kinds of sensed data, sensor resolution, number of spectral bands, number of data formats and type and size of data archives.

Today's Synthetic Aperture Radar (SAR) and optical sensors generate 10 to 100 gigabytes of data per day, so that in a multi-sensor spacecraft scenario the volume of data to be archived annually reaches 10 terabytes. The Shuttle Radar Topography Mission (SRTM) provides about 18 terabytes of SAR data in just 11 days and ESA's Envisat spacecraft collects about 80 terabytes of multi-sensor data per year. After 30 years of remote sensing, there are large volumes of valuable information for a single site on the Earth that have not been fully exploited because of the lack of automated tools. New technologies are required for the automatic extraction and classification of information from such EO data which could otherwise remain hidden forever or be

---

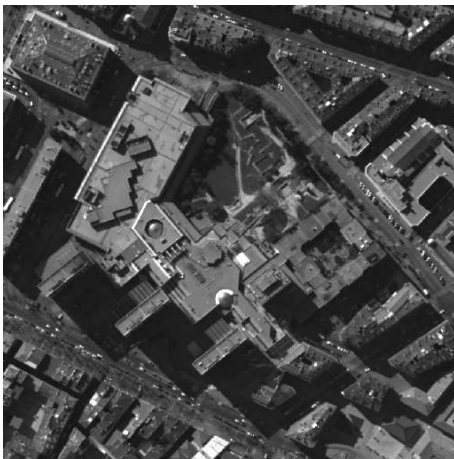
detected only by chance.



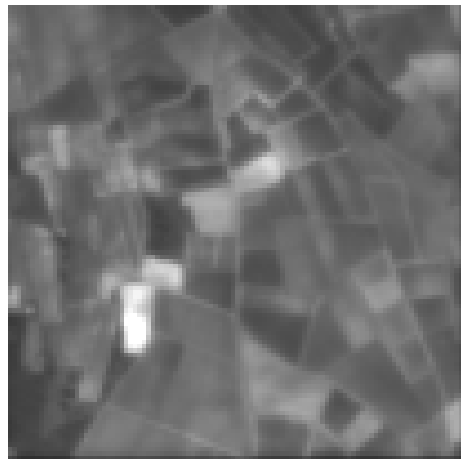
(a) QuickBird2, 61cm ©DigitalGlobe



(b) IKONOS, 1m ©Space Imaging



(c) Simulated from Aerial images, 0.5m ©CNES



(d) Simulated from Aerial images, 10m ©CNES

Figure 1.1: Images from different satellites with different resolutions.

In recent years, the capability to store large volumes of data has outperformed the capability to extract meaningful information from it. Remote sensing data access systems, particularly for images, allow queries by geographical location, time of acquisition or type of sensors. This information is often less relevant than the actual content of the scene, i.e. structures, objects or scattering properties of the sensors. Many applications of remote sensing require the knowledge of complicated spatial and structural information and the abstract relationships between objects within the image. This information is usually hidden in the image structure and must be mined to retrieve meaningful spectral information and higher level abstraction of objects in the image, such as cities, roads, rivers and forests, etc. Knowledge driven information mining from large EO archives needs the utilization of a family of methods for knowledge discovery, learning and information extraction from large volumes of data. This mining of information is performed with the identification of a specific feature and

---

certain applications in mind or it is performed to extract a key feature without any specific application requirement at the time of extraction. Companies and research institutes around the world are devoting large efforts to the design and production of Content Based Image Retrieval (CBIR) systems for EO archives. Several attempts have been made to have a CBIR approach to EO archives, but difficulties have been encountered in applications, as the basic concept of CBIR tools apply for searching the archives by global image similarity by scrutinizing the entire image archive to detect new features. This approach is cumbersome and time expensive for large volumes of images.

## 1.1 Image content retrieval and mining overview

CBIR technology has been used in several applications, for example fingerprint identification, biodiversity information systems, digital libraries, medicine, historical research, etc. The creation of CBIR systems involves research on databases and image processing, storage handling problems and user friendly interfaces. In these systems, image processing algorithms are used to extract features that represent the image properties such as color, texture and shape. The CBIR approach retrieves images similar to one chosen by the user (query-by-example) based on a similarity metric. Images are particularly complex to manage besides the volume they occupy: retrieval of images from databases is a context-dependent task. The first step towards retrieval requires a translation of high-level user perception into low-level image features. Moreover, image indexing is not just a string processing method: images or objects in images are represented as points in a high-dimensional space.

CBIR systems work have two main functionalities: data insertion and query processing. The data insertion system is responsible for extracting features from the images and storing them in a feature database. The query processing module is organized with an interface that allows the user to specify a query image and visualize the retrieved similar images. The query processing module extracts feature vectors from the query image and applies a metric (such as the  $L^2$  or the  $L^1$  norm) to evaluate the similarity between the query and the database images. The database images are ranked in decreasing order of similarity to the query image and the most similar images are shown in the output interface.

A CBIR system requires the construction of an image descriptor, which is characterized by the extraction algorithm to code the image features in a vector and a similarity measure to compare two images. A feature vector  $\vec{v}_I$  of an image  $I$  is defined as a point in  $\mathbb{R}^n$ . An example of possible feature vectors are the color histogram, the set of Fourier coefficients, geometrical patterns of shapes of objects, texture features, etc. An image content descriptor  $D_I$  is defined by a tuple  $(\xi_{D_I}, \delta_{D_I})$ ; where  $\xi_{D_I}$  is a function which maps an image  $I$  into its feature vector  $\vec{v}_I$  and  $\delta_{D_I}$  is a similarity function that computes the distance between the features from two images.

---



Several CBIR systems have been proposed in the recent past, some of which have become commercial products, while many others were proposed as research prototypes developed at universities and research laboratories. To name a few of the existing CBIR systems: QBIC (Query by Image Content), developed by IBM; Photobook, developed at the Massachusetts Institute of Technology (MIT); Chabot; Netra; and VisualSEEK. These systems retrieve image content based on color, shape, texture and the spatial location of segmented regions of the image. A few other systems, e.g. the PicHunter system, are based on a Bayesian framework that models user's needs during information query, while the Blobworld system is a region-based image retrieval system in which the pixels are clustered according to their color and texture properties. In this work, we will describe in detail the IKONA CBIR system developed at the IMEDIA research group at INRIA Rocquencourt, France.

An Image Information Mining (IIM) paradigm is dependent on the ability to evaluate retrieval systems and image understanding functions and methods. In order to evaluate such a system one should look at the technicalities of the system as well as user related concepts. These two approaches can be described as 'objective' and 'subjective' concepts respectively. The main problem arising in CBIR performance evaluation is the lack of a common image database and the validation of the retrieval process. In the literature, a considerable amount of work can be found on approaches based on query by image example. These approaches mainly validate system performance based on precision-recall graphs. On the other hand, in a knowledge driven image information mining system, the evaluation principle is divided into two main categories. The first one can be seen as objective and the process involves the technicalities of information content extraction, information quality and the complexity of the system. The second category is subjective and is based on the concept of relevance feedback and user satisfaction in the retrieval operation.

The growing complexity of image information content has made CBIR a weaker system for image retrieval. The CBIR system hardly allows for any adaptivity to user's needs as they are highly computer-centered approaches. A man-machine interaction system is the new state-of-the-art in image information mining systems. The next generation man-machine architecture of information mining systems is based on a stochastic link between the user defined semantic image content and the unsupervised content index. This stochastic link allows the users to query the image archive for relevant images and the system in turns returns a probabilistic classification of the entire image archive. The field of information mining opens a brand new perspective of information extraction from large volumes of heterogeneous data and the correlation of these data with various application scenarios. The Knowledge-Driven Image Information Mining (KIM) project is a next-generation architecture for man-machine interaction via the internet that adaptively incorporates application specific interests. The information representation of such a system is based on image feature extraction using a library of algorithms. These features are then clustered in an unsupervised manner followed by data reduction and supervised learning of users' semantics.

---

## 1.2 Which information ?

The rapid advances in image acquisition methods and the storage technology have led to the growth of large image databases. The image information content is implicitly embedded in the image and is usually hard to detect. These images, if analyzed would reveal useful knowledge to human users. The extraction of this implicit knowledge from the images has given rise to the field of image information mining. Image information mining systems are more than that of data mining systems in the image domain. A collaborative approach with an overall expertise in the fields of computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence is required for mining information from images.

Indexing and retrieval from remote sensing image databases relies on the extraction of appropriate information from the data about the entity of interest (e.g. land cover type) and on the robustness of this extraction to nuisance variables. Other entities in an image may be strongly correlated with the entity of interest and their properties can therefore be used to characterize this entity. The road network contained in an image is one example. The properties of road networks vary considerably from one geographical environment to another, and they can therefore be used to classify and retrieve such environments. In order to compute geometrical and topological features of the road network, we first need to extract the road network from the image, and then convert the output to an appropriate representation. This representation must be independent of any extraction methods. Features computed from the representation of the road network can then be used to classify the geographical environments.

The road extraction methods are in general resolution dependent. An extraction method used specifically for metre resolution images will extract many ‘road-like’ linear structures which in turn will provide us with redundant features for the proper classification of the geographical environment. The same logic holds otherwise, when an extraction method is meant for sub-metre resolution images. Figure 1.1 shows some examples of satellite and aerial images of different resolutions. An optimal road network extraction algorithm that accurately delineates road structures for practical use is very hard to design. The real world scenario is too complex to handle and many sophisticated algorithms for extracting pertinent road structure information rely on the specificity of the applications. In almost all cases, the road extraction algorithms are tuned to consider only a limited set of characteristics, and the algorithms fail to identify new characteristics when encountered in an application domain.

The methods used in our study are robust and can accurately extract the road networks in a SPOT5, 5m resolution image but they often failed to extract the narrow and finely structured road networks which are almost hidden in small dense urban areas. The lack of structural information available from these images containing such areas hinders the classification performance. In order to obtain useful structural information from these parts of the images and improve the classification, a new set of

---

features based on the segmented urban areas was introduced. These features were then combined with the existing road network features. Further study examined the classification performance with the structural features from the road networks augmented with the new urban regions structural features.

The question that still remains to be answered is this: what is the optimal size and resolution of an image in order to obtain the “best” classification of geographical regions on the Earth’s surface? The performance analysis of classification with features extracted from images of a certain size and resolution is an important issue concerning image database indexing. A study concerning this has been reported in this thesis.

### 1.3 The methodology

Satellite image classification has been a major research field for many years with its varied applications in the fields of Geography, Geology, Archaeology, Environmental sciences and for Military purposes. Many different techniques have been proposed, including stochastic methods, genetic algorithms, fuzzy theory and neural networks. The satellite classification works reported in the last twenty years have shown a significant increase in classification accuracy with the different approaches and new technology of image acquisition. There has been a large amount of work dedicated to the classification of large satellite images at pixel level rather than considering image patches of different sizes. Classification of image patches of different sizes from a large satellite image is a novel idea in the sense that the patches considered contain significant coverage of a particular type of geographical environment.

A large satellite image of size 5120x5120 pixels is cut into 100 non-overlapping image patches each of size 512x512. The road networks and the urban areas are extracted from these smaller images. Geometrical and topological features are then computed from the road network representation and the segmented urban areas. These images along with the features are then tested against the existing image database to classify each image belonging to a certain pre-defined geographical class. Probability measures are associated with the labelled classes to quantitatively infer about the confidence of classification. Few image patches are studied in details to analyze the classification results with ground truths from Google maps. The schematic diagram of the indexing processes is shown in Figure 1.2.

### 1.4 Organization of the thesis

In **Chapter 2** we provide an overview of image information mining systems in general. We study in detail a CBIR system and a Knowledge Discovery and Data Mining (KDD) system. We then talk about shape and texture as image content used in the scope of

---

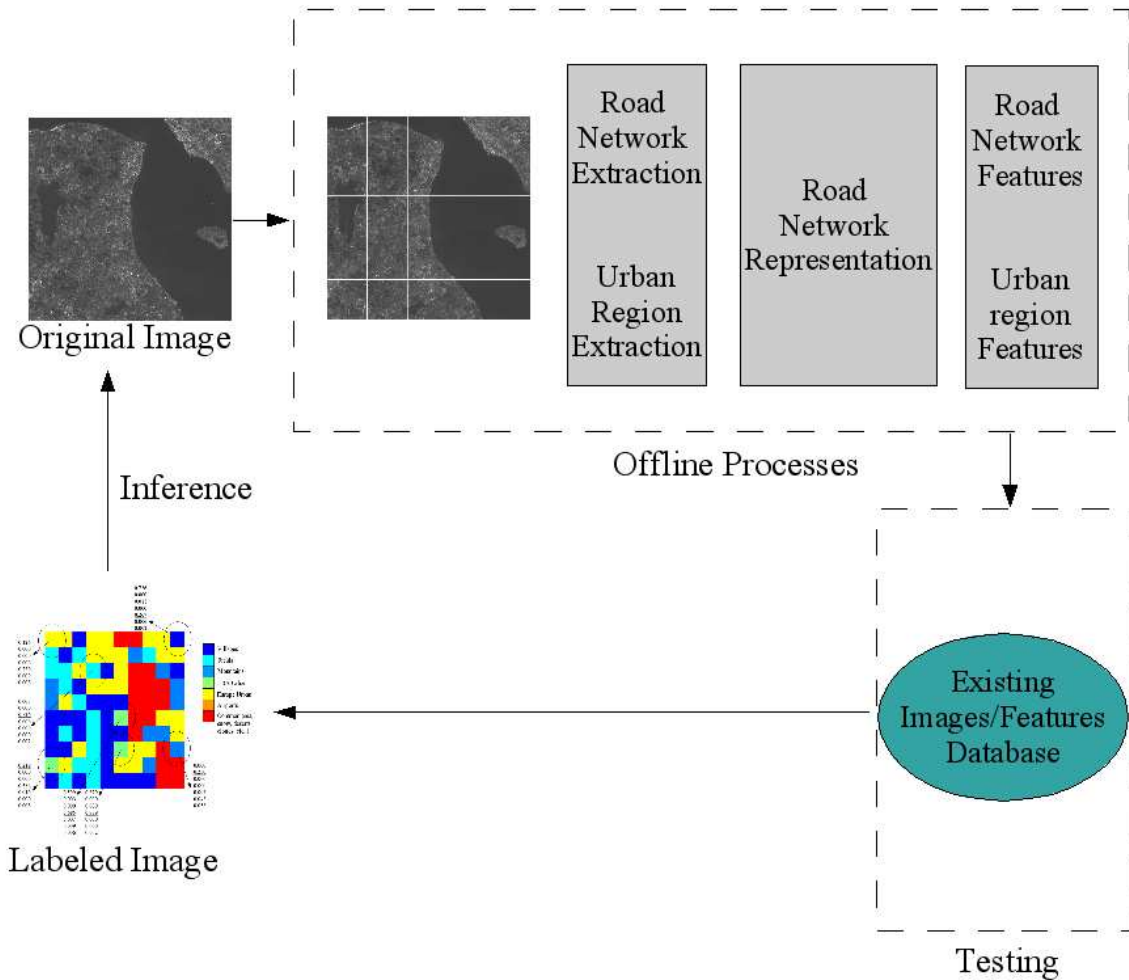


Figure 1.2: The schematic diagram of the large image indexing process.

this work. In order to convert features into classes, we review classification tools in general and Support Vector Machines (SVM) in particular.

In **Chapter 3**, we introduce the road network extraction methods used on SPOT5, 5m resolution images. A database of 497 images is created with seven geographical classes with each image of size 512x512 pixels. The output of the extraction is converted into a suitable representation. Geometrical and topological features are computed from the representation. These features are then used to classify the images with linear SVM method. Feature subset selection is done with Fisher Linear Discriminant analysis to reduce the redundancy of some features and hence lower the classification error.

Due to certain limitations of the road network extraction methods, small regions with dense road structures are not appropriately extracted. In order to circumvent this problem and to correctly classify these images, **Chapter 4** introduces a region extraction method. The geometrical and topological features computed from these

regions are then augmented with the road network features to better classify the geographical environments. Linear SVM and feature subset selection with Fisher Linear Discriminant analysis is done to further reduce the classification error.

**Chapter 5** shows the resolution and size dependency of images on the extraction outputs and hence the classification performance. A resolution and size dependency test on classification is done on two image databases and the results are compared with the previous classification results.

In **Chapter 6** we show the indexing results on large satellite images of size 5120x5120 pixels with image patches of size 512x512 pixels each extracted from it. Here we use probabilistic SVM in order to have an estimate of the classification results. The probabilities are then studied in details to have a closer look at the classification accuracies.

In **Chapter 7** we conclude by giving some perspectives for our state-of-the-art implementation in future image information mining systems for EO archives.

---

## Chapter 2

# Image Information Mining Systems

The rapid advances in image acquisition methods and storage technology have led to the growth of large image databases. The image information content is implicitly embedded in the image and is usually hard to detect. These images, if analyzed will reveal useful knowledge to human users. The extraction of this implicit knowledge from the images has given rise to the field of image information mining. Image information mining systems are more than data mining systems in the image domain; they require an interdisciplinary approach with collaborative expertise in the fields of computer vision, image processing, image retrieval, data mining, machine learning, database management, and artificial intelligence.

## 2.1 Systems

There is an explosion in the amount of data being produced by various means, and mining from this data can uncover important information useful for researchers across the world. Data mining [Fayy 96, Rao 02, Datc 03, Dasc 05] has made great strides in the past years in terms of the technologies and methodologies being developed and researched. The work in the field of data mining includes methods for analyzing complex forms of data, as well as specific techniques and methods. The categories of data mining include hypertext mining, multimedia, spatial, and time series data mining. In this work we are concerned with spatial data mining in terms of the implicit information and the spatial relationships of patterns observed in satellite images. The topological information contained in spatial and geographic databases differentiate them from other databases. The extraction of spatial relationships and patterns which are not explicitly stored in the spatial database is the basic concept behind spatial data mining [Este 95, Hans 00, Chaw 00].

Image mining denotes the union of data mining and image processing in order to

---

help in the understanding of information in the image domain. Image mining differs from low-level computer vision and image processing in the sense that image mining focuses on the extraction of patterns from large collections of images whereas computer vision and image processing techniques are used in the understanding and extraction of patterns from a single image. There seems to be some commonalities between image mining and content-based image retrieval as both deal with large image collections, but image mining goes far beyond the problem of retrieving relevant images. A good image mining system is one which can provide users' with effective access into the image database and also generate knowledge about the patterns hidden in the images. Several image mining systems have been developed for various applications ranging from art image databases [BAL] to remote sensing image databases [ESA]. In this chapter we will consider two such systems as motivation for our work which can be incorporated in these systems for various application scenarios.

## 2.2 The IKONA system

The IMEDIA research group at INRIA Rocquencourt [INRI] has developed a content-based indexing technique and interactive retrieval methods for browsing large multimedia databases by content. In order to design an effective image retrieval system, the database is divided into two categories. The first category is concerned with specific databases with known ground truth. During the process of indexing the user will consider these ground truths in order to tune the models or the parameters which in turn maximize the system efficiency. The group has developed specific signatures for face recognition and fingerprint identification. The second category is concerned with databases containing heterogeneous images without any ground truth. In this context, generic image features are computed which describe a general visual appearance such as color and texture [Bouj 01, Goue 01].

The IKONA system architecture is based on a client-server architecture where the server is written in C++ and includes image feature extraction algorithms, user interaction (retrieval with visual similarity, relevance feedback mode, partial query mode, points of interest mode, etc...) and a network module for communication with the client. The client needs to be portable and is written in Java. It runs on machines with the Java Runtime Environment (JRE). The user is presented with a Graphical User Interface (GUI), the query mode is set for the server and the relevant images are displayed as the search results. The system is flexible in the sense that functionalities can be easily added without disturbing the overall architecture. In order to deal with generic databases, IKONA includes a relevance feedback technique which allows the user to refine their query by specifying a set of relevant and non-relevant images. IKONA has a Relevance Feedback (RF) mode [Fere 05b, Fere 05a] for category search in image databases which assists the user in rapidly finding images in a large database. The IKONA system also has a region based query mode, in which the user can select a part of the image and the system will search images or part of images that are visually

---

similar to the selected part [Fauq 02, Fauq 04, Fere 04]. In this case the query is focussed and the system response is enhanced with regard to the user's objective since background image properties are not considered. Several segmentation methods and point of interest methods have been developed to achieve partial queries. The generic image database with query image selection and query results is shown in Figure 2.1.



Figure 2.1: Generic image database. Left: Selection of a query image (the red border indicates the selected image). Right: Query results for the selected image. Courtesy (<http://www-rocq.inria.fr/imedia/>)

The research interests of the IMEDIA team also include the combination of text and image features for indexing and retrieval. The image database is partially annotated with keywords and the IKONA system in turn can use these keywords for very fast retrieval. Research is being carried out on keyword propagation, semantic concept search and hybrid text-image retrieval mode. In the category of specific databases, IKONA also deals with databases containing faces in a complex background. Face detection is performed using a hierarchical algorithm based on coarse-to-fine support vector classifiers and the system assigns for each face a set of indexes with Dynamic Space Wrapping (DSW) [Sahb 02a, Sahb 02b]. The face image database with query image selection and query results is shown in Figure 2.2.

## 2.3 The KIM system

Information mining, knowledge discovery and data management are changing the paradigm of user/data interaction by providing simpler and wider access to Earth Observation (EO) data archives. The Knowledge-Driven Image Information Mining



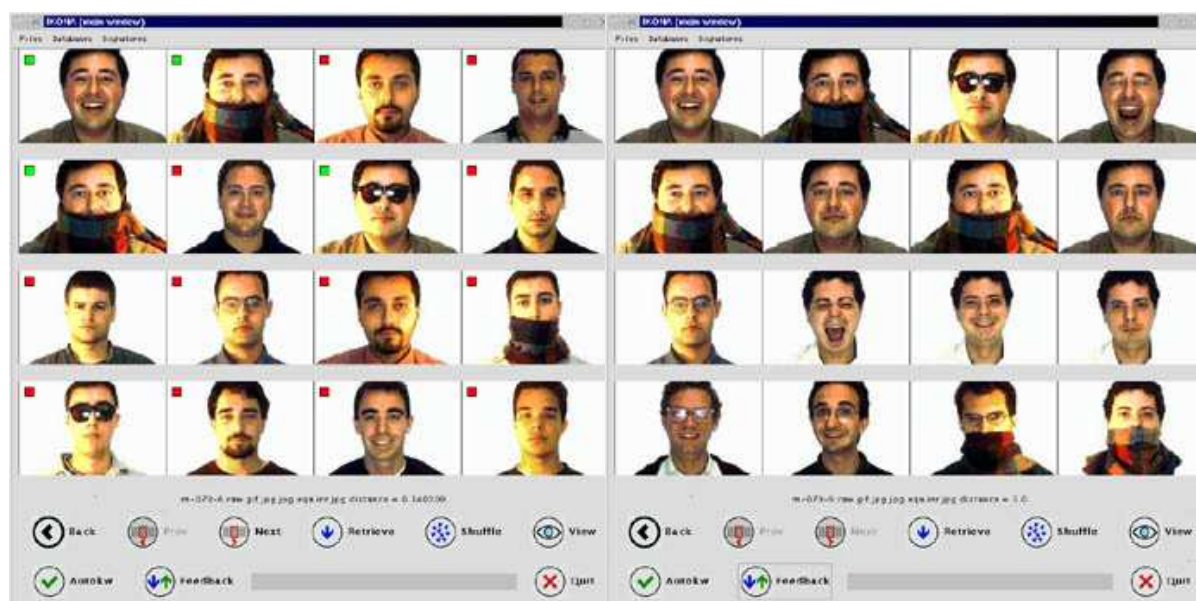


Figure 2.2: Specific Image database. Left: Selecting a query image with relevance feedback mechanism in IKONA. Right: Query results for the selected image. Courtesy (<http://www-rocq.inria.fr/imedia/>)

(KIM) project is the next-generation architecture for man-machine interaction via the internet and for adaptively incorporating application-specific interests [Datc 99, Datc 02]. The exponentially growing volumes of EO data archives needs to be fully exploited and the role of human-centered computing will play a major role in the design of EO data exploitation. The KIM system addresses these scenarios and is a solution that satisfies the requirements of the end users, for example, the scientific community, civil protection agencies and educational institutions.

The evaluation principles of the KIM system are divided into two categories: objective and subjective. The objective part involves the technicalities of the system to extract information content and information quality [Datc 98, Schr 98]. The subjective part includes relevance feedback, user satisfaction and semantic confusions. The KIM system is based on a hierarchical Bayesian representation of image content. The information representation on hierarchical levels of different semantic abstraction is based on a 5-level Bayesian learning model shown in Figure 2.3. The first level deals with the extraction of primitive features from the data  $D$  with various stochastic signal models  $M$ . The parametric data model  $p(D|\theta, M)$  assigns the probability of the data  $D$  for a certain value of the parameter vector  $\theta$  and a certain data model  $M$ . The information is extracted by maximum a-posteriori (MAP) estimation of  $\theta$ . The Bayesian concept is used to select the most evident model describing the data. This approach introduces the next level of features called metafeatures. With the existing features from the two levels a set of signal classes  $w_i$  are derived describing the clusters of points in the  $\theta$  space of different models  $M$ . The classes  $w_i$  are obtained by unsupervised clustering of the entire set of data points using a  $k$ -means algorithm with

predefined number of classes depending on the size of the image archive. At the last level, the users' interests  $L_\nu$ , i.e., the semantic interpretation of the image content, are linked to the signal classes  $w_i$  by probability  $p(w_i|L_\nu)$ . The screen shots of random image selection, Figure 2.4, the KIM system working interface, Figure 2.5, and the search applet, Figure 2.6 and Figure 2.7 are shown accordingly.

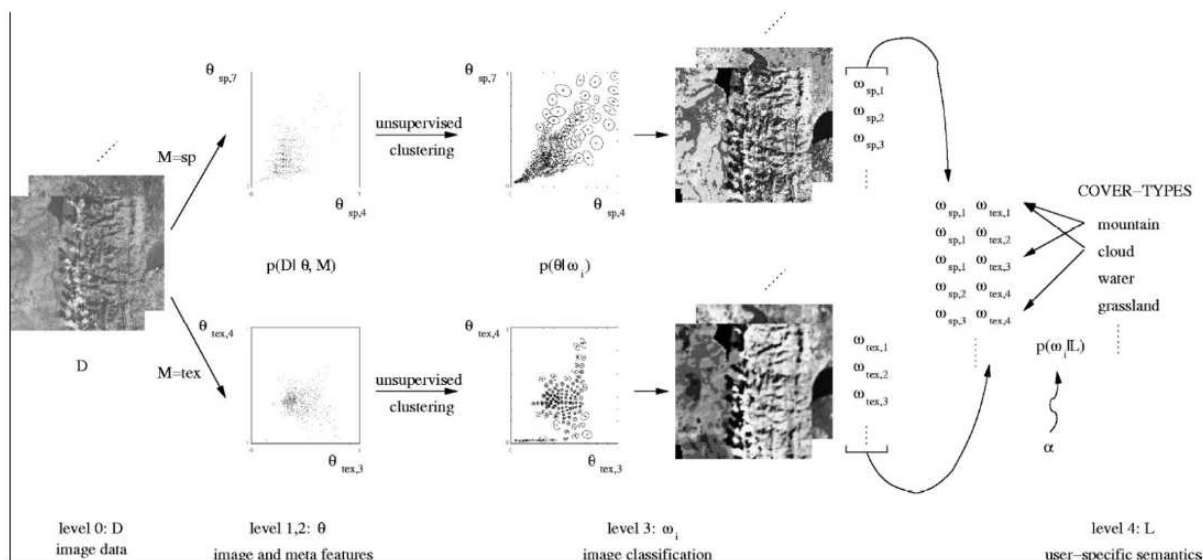


Figure 2.3: The hierarchical image content description. Courtesy KIM project [Datc 99]

The KIM system is based on a server-client architecture. The system is designed to operate over the internet. The User Management Interface (UMI) provides the facility to register new users and login existing users to access KIM services. The Interface Learning (IL) module is a Java applet. This applet is used to define or refine a label. Users can perform search with a pre-defined label or can set some parameters in the image database. The search results are presented as HTML pages.

Among others, the framework proposed by [Durb 04] is based on content and semantic based information retrieval from remote sensing image archives. A method for knowledge discovery and data mining based on hierarchical segmentation is proposed in [Tilt 00].

## 2.4 Image representation and similarity

There are various levels of image representation. Image data structures can be represented as matrices, chains, graphs, relational databases, complex hierarchical methods,

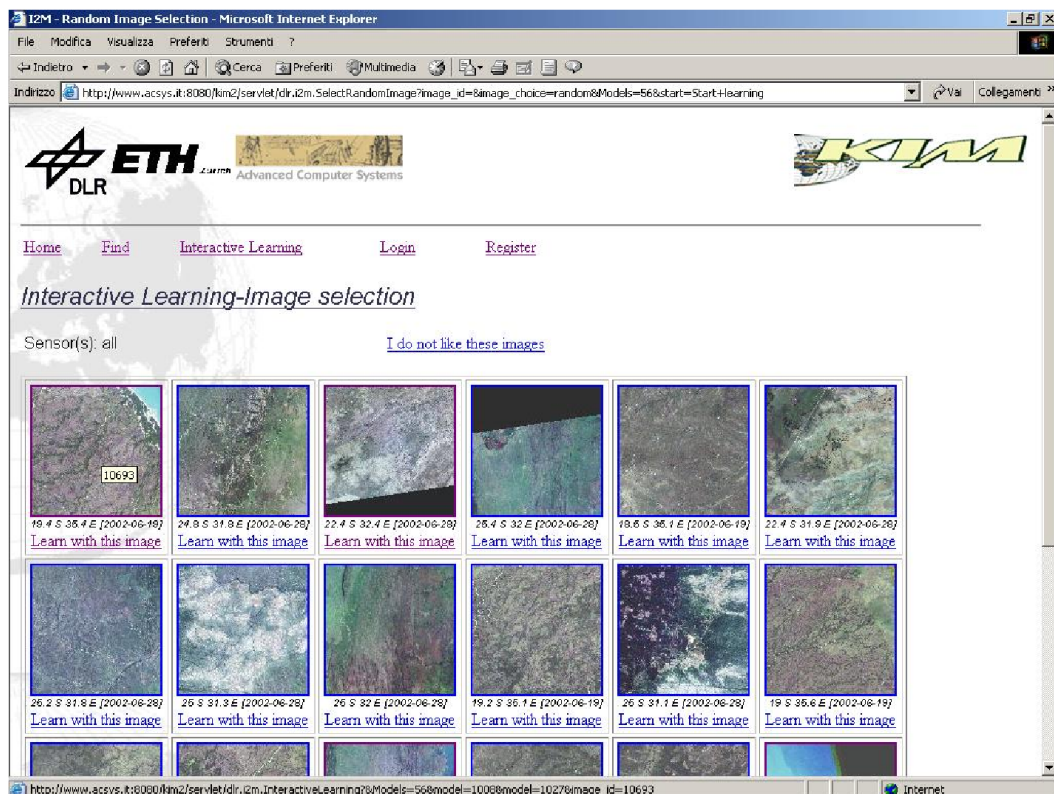


Figure 2.4: Image selection. Courtesy (ESA, DLR, ETH Zurich and ACS) : Courtesy <http://kes.esrin.esa.int/kes/>

etc. The most naive way of image representation is by matrix. Elements of the matrix are numbers defining the pixel intensities or their associations. Chain codes are used to describe properties of objects present in the image [Free 61, Kane 85]. The chain code is a symbolic representation which represents the neighborhood association of primitive objects in the image. This representation is best suited for extracting local information of an object. In the case of global information retrieval, this method poses certain problems. Topological data structures are used to represent elements of an image and their relations. Simple graphs or adjacency graphs are widely used in computer vision to represent images. Another very commonly used representational schemes in computer vision are the hierarchical data structures [Niem 80, Peur 01]. The two typical common structures used are the pyramids and quadtrees.

A general methodology for designing a system for image indexing and retrieval has not yet been developed. Particular application domains demand the appropriate techniques of image analysis, image content description, storage and retrieval of image content. Image descriptors are usually defined in terms of properties of objects or regions and the relationships among such objects and regions contained in the image. Image descriptors can be represented by relational structures such as graphs.

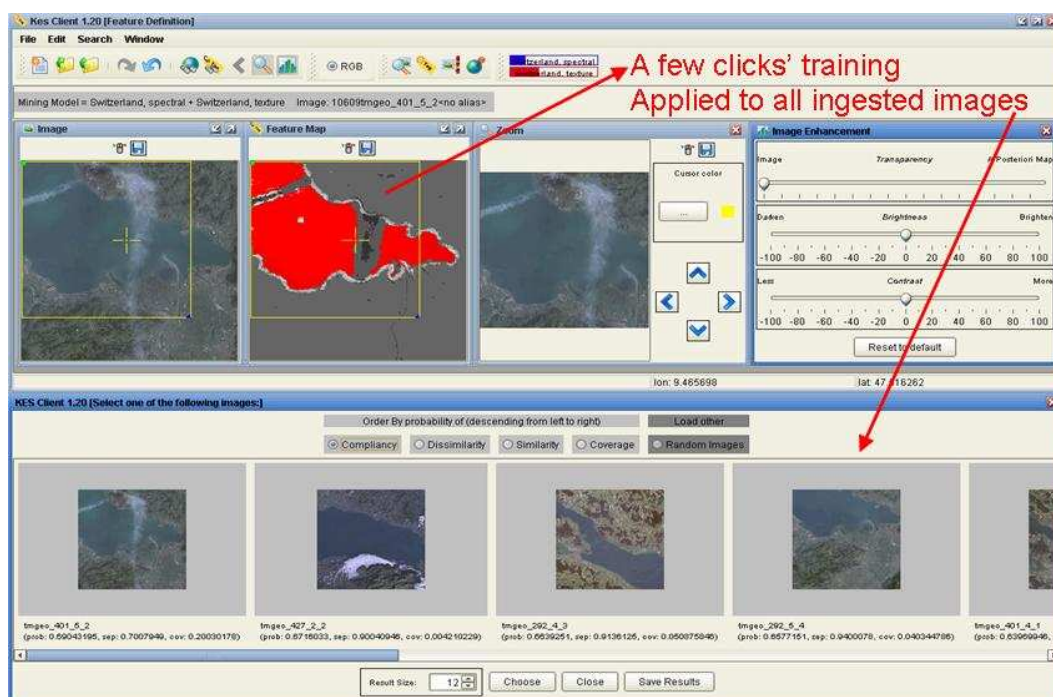


Figure 2.5: Results of query from the image. Courtesy (ESA, DLR, ETH Zurich and ACS) : Courtesy <http://kes.esrin.esa.int/kes/>

In general the descriptors that are used are derived from the raw image data and can be geometric, statistical, or textural. These kinds of descriptors are generally termed low-level descriptors. High-level descriptors find the correspondence between these low-level descriptors of a specific image and a representation of the image, e.g., a graph structure, to understand and interpret the image content. The characteristics of both low/high level descriptors is highly dependent on the segmentation of meaningful objects from images. Prior to storage, the derived image descriptors need to be appropriately represented. The particular choice of a representation can reduce the burden of storage and data processing. Representation of image semantic or structural content is of particular importance in image indexing and retrieval and depends on the specific application scenario. Graph based representation is one of the representation structures used in this work. Graphs are data structures with a great deal of expressive power which are widely used in the fields of pattern recognition and computer vision. The graph provides a simple and compact way to represent the spatial and structural content of the image. Image representation by strings can also be found in the literature [Euri 93]. The string representation is used in applications where images contain mutually disjoint objects and have rather simple shapes. A better form of the string representation for representing images consisting of complex and overlapping objects is reported in [Euri 93]. Fractal codes are also used as a representation method [Krup 95]: they are used to represent arbitrary shapes. Fractal codes are particularly popular in the geographic community, as natural structures are irregular

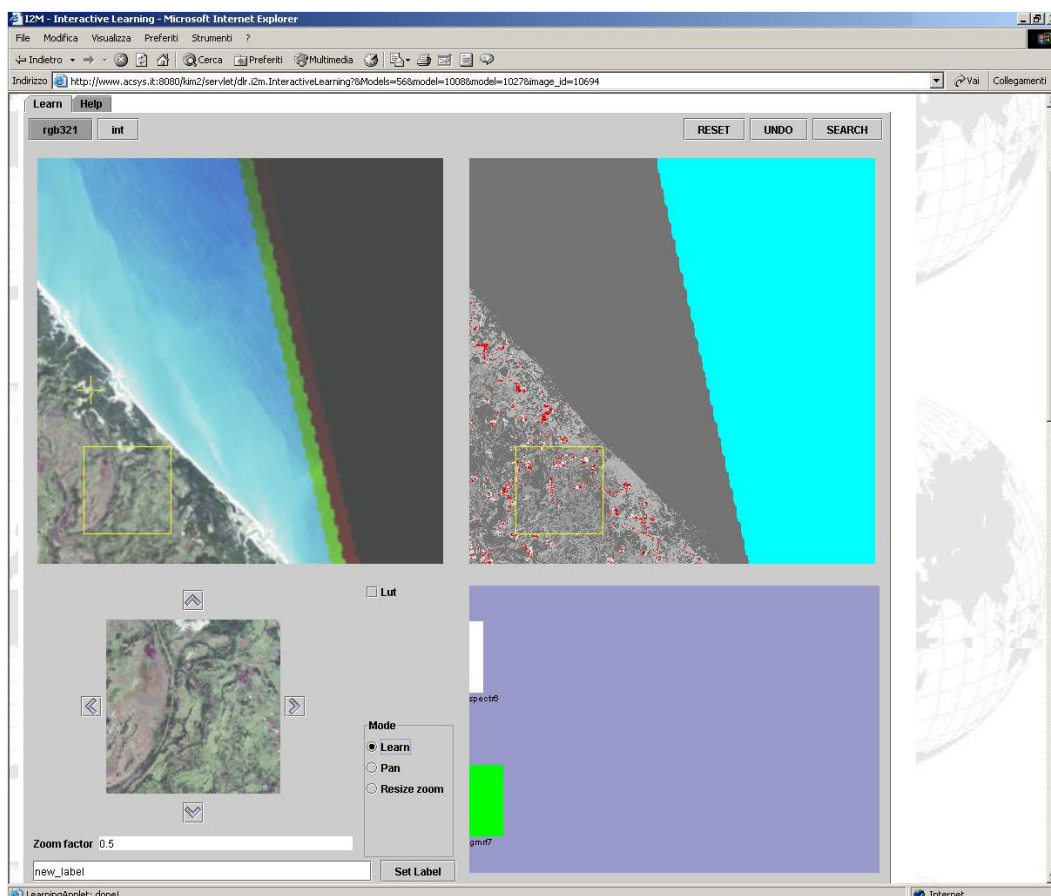


Figure 2.6: Interactive learning. Courtesy (ESA, DLR, ETH Zurich and ACS) : Courtesy <http://kes.esrin.esa.int/kes/>

and complex.

Prior to the representation of an image, the image must be segmented into regions or objects of interest. The segmentation of regions or objects or both is a complex task and varies according to application scenarios. Although accurate and robust segmentation techniques can be found in the literature, the accuracy of the segmentation method can be relaxed in case of image content indexing as opposed to the case of image understanding or information mining.

However, image representation must be unique and should be tolerant to small variations in the object properties. Apart from the uniqueness criterion, a representation must also be complete, i.e., a representation must include all the disjoint objects or regions in the image considered.

In the literature, image data is often considered as a part of multimedia document. A multimedia document [Jain 98, Doer 98, ODoc 91] is a structured collection of im-



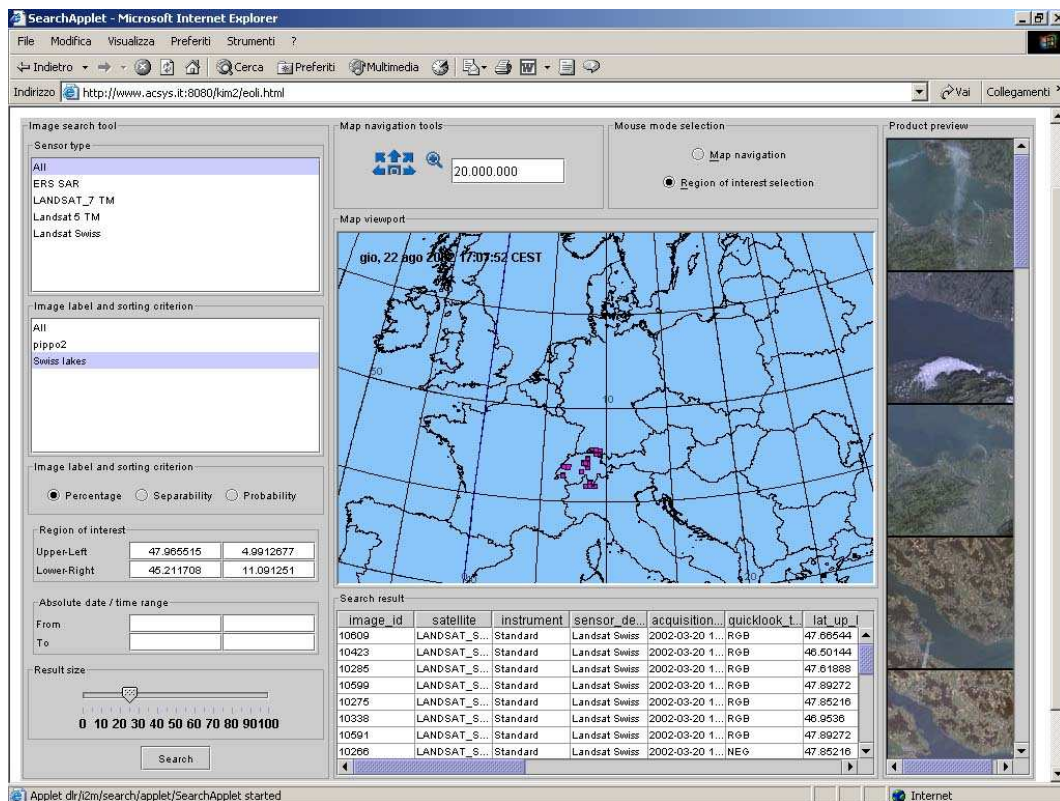


Figure 2.7: Metadata based search applet. Courtesy (ESA, DLR, ETH Zurich and ACS) : Courtesy <http://kes.esrin.esa.int/kes/>

age, sound, text or other data. In this work, we will only consider image data as an entity of the database. An image indexing and retrieval system also contains an image similarity measure for the performance of retrievals. Similarity criteria used in a content based search engine can be either global or local. The similarity measure can also be characterized as a distance measure, metric or non-metric, of the content of the two images in some representational forms. The notion of the distance measure varies with the representation schemes of the image content. In the case of a graph representation, a distance between the contents of two images can be defined as the minimum cost of transforming one graph into another. Other similarity metrics used in content based image retrieval are the  $L^1$  or the  $L^2$  norm [Agga 01]. A number of recent approaches in computer vision compare images using measures that are non-metric, [Howa 05] in that they do not obey the triangle inequality. The triangle inequality is rather difficult to implement in complex matching algorithms. The use of non-metric measures is also motivated by studies in psychology which show that human judgement of similarity is non-metric in nature. Work has also been done on normalization methods for feature sets for similarity measures, because the most popular measures like the  $L^2$  norm are highly susceptible to features with long ranges than those with small ranges. Probabilistic methods for image retrieval have also been widely studied. Probabilistic

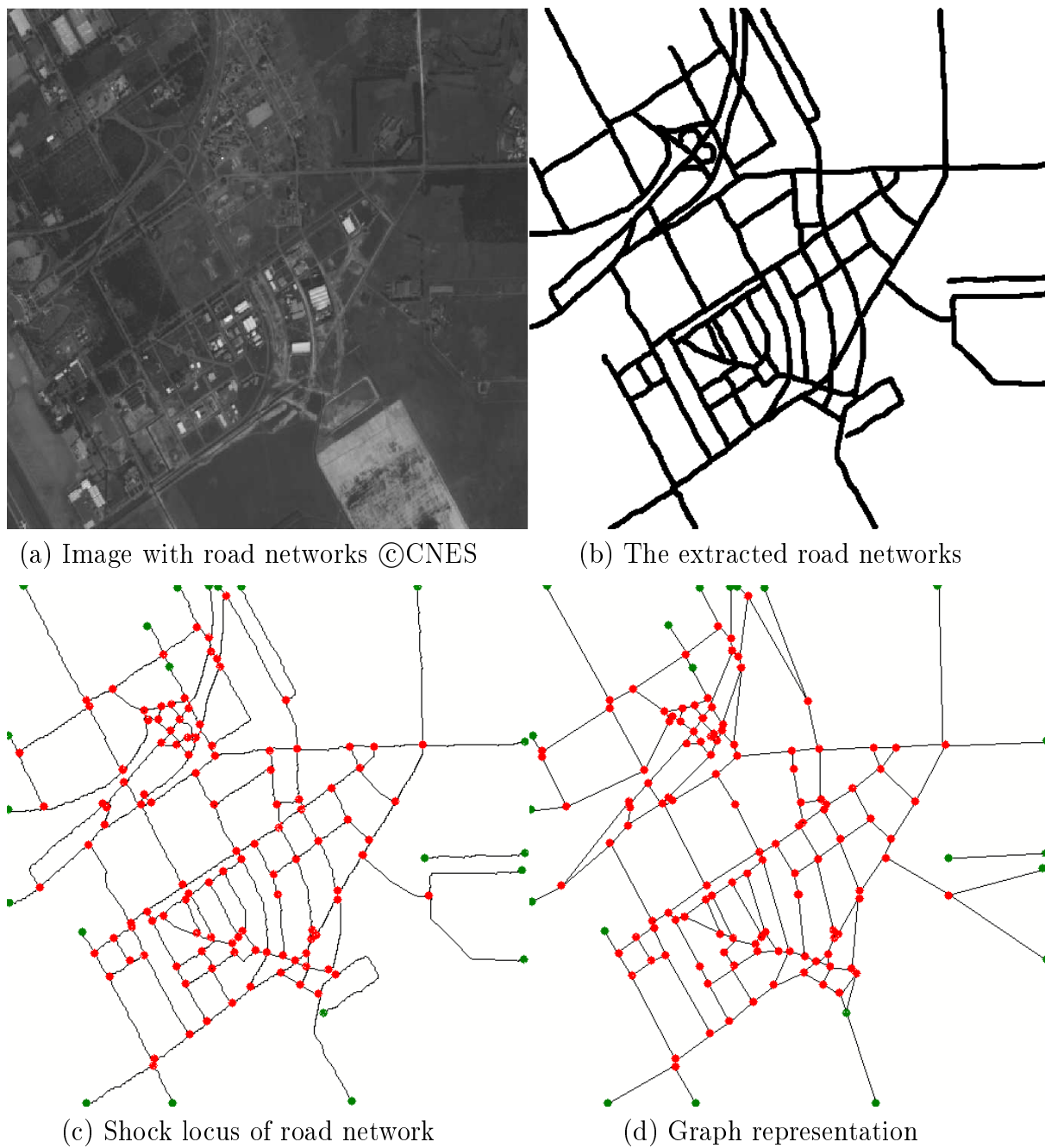


Figure 2.8: An example of road network extraction and graph representation

measures such as the Mahalanobis distance and Bhattacharyya distance are used for image retrieval [Rahm 05]. Likelihood-based similarity measures [Akso 00, Akso 01] are also used for image retrieval. The measure used likelihood ratios that were derived from a Bayesian classifier that measured the relevancy of a query image and a database image.

## 2.5 Image content

Image content is dependent on human perception. A huge amount of information in terms of color, texture and shape is hidden in images. A meaningful way to extract this information in order to better understand an image is a challenging task in the field of pattern recognition and computer vision. In the scope of this work, we will only focus on the shape and the texture properties of an image.

### 2.5.1 Shape content of an image

The objective of content-based image query is to efficiently find and retrieve images from the database that satisfy the criteria of similarity to the user's query image. A query image may in general specify objects or parts of objects and the retrieval must find images which contains similar objects or a query image may specify structural or semantic content of an image. The problem of image retrieval by content is well know in the literature of computer vision research and a large number of such techniques exist in the literature of object recognition and classification [Duda 00, Scha 92, Bish 95]. Generally there are two types of shape descriptors; contour based and region based shape descriptors. These two kinds of descriptors can still be sub-divided into global and structural descriptors. In the case of contour-based types, structural descriptors consists of chain codes, polygons and splines, and global descriptors include Fourier and wavelet descriptors and scale space. In the case of region based descriptors, global features include area, Euler number and moments, and structural features include the medial axis and convex hull. Many previous works have used chain codes for shape descriptors, but chain codes are not invariant to shape size and orientation. Fourier descriptors and moment invariants have also been used successfully for shape representation. Fourier descriptors use the Fourier transformed boundary as the shape feature [Pers 77, Chel 84]. Work has also been done on modified Fourier descriptors which are robust to noise and invariant to geometric transformation [Rui 96]. Moment invariants are used as shape features, [Hu 62] defining seven such moments using region-based moments which are invariant to transformation.

Some more recent work using the finite element method (FEM) [Pent 96], the turning function [Arki 91], and wavelets [Chua 96] have also been used in shape representation. The PhotoBook system, a set of application tools for image browsing and

---



retrieval uses FEM as the main framework for image representation and similarity. In the FEM method, the connectivity of each point on an object with other points is defined by a stiffness matrix and the eigenvectors of this matrix defines the feature space. The method is also used for matching objects. Similarity is computed based on the eigenvalues of this stiffness matrix. A turning function based method has also been developed for comparing both convex and concave shapes. The work of [Chua 96] uses a wavelet transform to describe object shapes. This is a stable representation and accommodates multiresolution and spatial localization properties. There are many review papers in shape representation [Li 95, Meht 97]. The article [Meht 97] on shape representation compares contour-based representations, region-based representations and combined representations. In [Li 95], it is shown that there is a simple linear relationship between the geometric moments method (region-based) and Fourier descriptors (contour-based). Mathematical morphology has also been widely used in shape representation. The work of [Pita 91] uses mathematical morphology in combination with simple object structures perceivable by people. A comparison study of two morphological shape representations is reported in [Rein 94].

The work of [Dese 98] introduces a new concept called "Veinerization" for describing shapes with skeletons. The work reported in [Dimi 00, Sidd 02] consider the computation of the subpixel skeleton of an object which preserves its topology. In our work we use this technique to represent road networks extracted from a satellite image as a skeletal graph shown in Figure 2.8. The graph is computed from the divergence of the gradient function along with a thinning process.

### 2.5.2 Texture content of an image

Texture is an implicit property of all surfaces, natural or man-made. For instance, clouds have a different textural property to roads. Texture also provides us with the structural arrangement of objects and their relationships with the surroundings. A vast amount of research has been done in this field over the past few decades. The co-occurrence matrix representation of texture [Hara 73] three decades ago opened a new direction in the field of pattern recognition. This approach considers texture as grey level spatial dependency. The matrix is constructed by taking the distance between image pixels. Various statistics are then computed from this matrix in order to represent texture properties. Some of these statistics lack the essence of human visual perception of texture. Motivated by this fact a new texture representation was developed by [Tamu 78]. These features are closer to human perception of texture.

Wavelets were introduced by many researchers [Kund 92, Lain 93, Chan 93, Smit 94] for texture representation. Later, wavelet were used in combination with Kohonen maps [Gros 94] and the co-occurrence matrix [Thya 94] to perform texture analysis. Other kinds of texture representation were compared with the co-occurrence matrix representation, including Markov Random Field representations [Cros 83], fractal

---

based representations [Pent 84] and multichannel filtering representations [Bovi 90].

An evaluation can be found in [Ma 95], where texture analysis with various wavelet transforms was performed. The evaluation study showed that the Gabor wavelet transform was one of the best texture representations among others.

The IBM's CBIR system, QBIC [Equi 94] and the University of California at Irvine's Database Research Group project MARS [Orte 97] use the Tamura texture features for image content representation and retrieval.

In satellite images, urban areas are strongly textured [Desc 93, Lore 00] and the problem of extraction of these regions is essentially a problem of differentiation of textures. In our study, we use the work of [Roux 92] developed for the extraction of urban regions, using the morphology operations of opening and closing. This difference gives prominence to textured regions such as urban areas in a satellite image.

## 2.6 Image classification

Classification is the most challenging task in the field of computer vision due to its application specific definition. In remote sensing, broadly speaking, image classification is a means to convert a finite set of features computed from the image into a finite set of classes that represent the surface types seen in the imagery. The classification task can be divided into two sets: supervised or unsupervised classification.

In supervised classification, the classes are defined *a priori*. The features from the image are computed and a classification model is chosen. This model is chosen according to the application type and the data set. In the case of supervised classification, the most common models are the Parallelepiped, Maximum Likelihood [Math 04] and Support Vector Machine (SVM) [Vapn 97]. The output of the classification can be further processed according to application needs. In unsupervised classification, there are no *a priori* class labels, the classification clusters data based on statistics only. The commonly used models for unsupervised classification are *k*-means [Kanu 00] and kernel *k*-means [Dhil 04]. In almost all cases of unsupervised classification, the output requires a great deal of post-processing in order to obtain meaningful results.

Classification of images has been a broadly studied research area for years. The semantic organization of an image database becomes meaningful with an appropriate classification mechanism. Many classification methods have been studied and successfully implemented within the premises of pattern recognition and content based image retrieval (e.g., *k*-nearest neighbor, Bayesian nets, maximum likelihood analysis, neural networks, support vector machines, linear discriminant analysis, etc.). Previous work in the field of image classification includes [Yu 95], using Hidden Markov Model (HMM) to classify scenes in natural images. The work of [Hara 73] uses tex-

---

ture features based on gray-level spatial dependences to classify aerial, panchromatic and multispectral images. A supervised classification model based on a variational approach was proposed by [Sams 01]. Neural networks to classify radar images is reported in [Hara 94]. Support vector machine classification work is reported in [Chap 99b, Vapn 97, Merc 03] and many others. In this work we will use the feature selection and SVM classification work reported in [Camp 05], with features computed from representations as described in Chapter 3 and Chapter 4.

### 2.6.1 Classification tools

Support Vector Machines (SVM) have been widely used in pattern recognition, and with its growing success, researchers from different fields are using it to analyze the data at their disposal. A special interest of SVM is in the field of CBIR [Chap 99a, Chen 01, Chan 03]. The goal of SVM classification is to find the best hyperplane separating the relevant and irrelevant data which maximizes the margin between the classes. In its simplest case, it assumes the relevant and irrelevant vectors to be linearly separable. In a real world scenario, data are almost always non linearly separable. To deal with such problems, kernel methods have been introduced to deal with the non-linearity of the data space. Kernel functions are used to map the data into a higher-dimensional space, where they can be separated by a hyperplane. It also incorporates a tolerance to noisy data by introducing a bound. The optimization problem can be expressed as follows:

$$\beta = \underset{\alpha}{\operatorname{argmax}} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (2.1)$$

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad (2.2)$$

$$\forall i \in [1, N] \quad 0 \leq \alpha_i \leq C \quad (2.3)$$

The training data is labeled as  $\{\mathbf{x}_i, y_i\}$ ,  $i = 1, \dots, N$ ,  $y_i \in \{-1, 1\}$ ,  $\mathbf{x}_i \in \mathbb{R}^d$ . Suppose that we have some hyperplane which separates the positive from the negative examples. The points  $\mathbf{x}$  which lies on the hyperplane satisfy  $\mathbf{w} \cdot \mathbf{x} + b = 0$ , where  $\mathbf{w}$  is the normal to the hyperplane,  $|b|/\|\mathbf{w}\|$  is the perpendicular distance from the hyperplane to the origin, and  $\|\mathbf{w}\|$  is the Euclidean norm of  $\mathbf{w}$ .  $\alpha_i$  are the positive Lagrange multipliers and the constant  $C$  is used to reduce the effect of outliers on classification.

In case of non separable data, we map the data  $\mathbf{x}$  into some infinite dimensional Euclidean space  $H$  using a mapping  $\Phi$ :

$$\Phi : \mathbb{R}^d \mapsto H \quad (2.4)$$

Then the training algorithm depends on the data through dot products in  $H$ , i.e., on the function of the form  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ . Now if there is a “kernel function”  $K$  such that  $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ . We then use this function  $K$  in the training algorithm without without explicitly knowing the function  $\Phi$ . Some common kernel functions are:

$$\begin{aligned} \text{Homogeneous polynomial kernel: } & K(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}')^d \\ \text{Radial basis function: } & K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \quad \text{for } \gamma > 0 \\ \text{Radial Gaussian Basis function: } & K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \end{aligned} \quad (2.5)$$

The optimal value  $\beta$ , which measures the distance of a vector  $\mathbf{x}$  and the separating hyperplane is used to compute the relevance of  $\mathbf{x}$  as:

$$f(\mathbf{x}) = \sum_{i=1}^N \beta_i y_i K(\mathbf{x}_i, \mathbf{x}') + b \quad (2.6)$$

The separating hyperplane  $f(x)$  is computed with the value of  $b$  from the Karush-Kuhn-Tucker (KKT) conditions [Camp 00].

The output of an SVM does not provide any estimation of the classification confidence and thus does not allow us to include any *a priori* information. Probabilistic SVM provides us with a way to construct a classifier producing a posterior probability  $P(\text{class} = c | \text{input})$ , which allows us to take a quantitative decision about the classification. In this work we use the method reported in [Plat 99], which uses a two parameter sigmoid function as a post-processing agent. The SVM along with the sigmoid function performs better than the standard SVM with kernels.

$$P(y = c | f) = \frac{1}{1 + \exp(Af + B)} \quad (2.7)$$

The parameters  $A$  and  $B$  are computed by maximum likelihood estimation from the training data set  $(f_i, y_i)$ .

---

## 2.7 Discussion

The growth of large image databases during the last few decades with the advancement in image acquisition technologies has attracted many researchers from different fields to work in the domain of image information mining systems. These images coming from various sources must be systematically analyzed to render important information which are often less relevant to human perception. A vast literature can be found in the fields of CBIR and KDD. The two systems which are at our disposal have motivated us to work in the field of indexing and retrieval of satellite images from large image repositories. In the future our work can be integrated with these two systems to provide some additional features along with the existing ones.

This chapter particularly looks at two different systems, one being a CBIR system and the other a KDD system. A brief literature overview on image content reveals that shape and texture have been widely used to capture information from satellite images. Classification with shape or texture features extracted from the image helps us to understand the spatial structure of the Earth's surface. The subsequent Chapters will systematically analyze and use such features to characterize some geographical environments as observed in mid-resolution satellite images.

---

## Chapter 3

# Road Network Structural Information

Indexing and retrieval from remote sensing image databases relies on the extraction of appropriate information from the data about the entity of interest (e.g. land cover type) and on the robustness of this extraction to nuisance variables. Other entities in an image may be strongly correlated with the entity of interest and their properties can therefore be used to characterize this entity. The road network contained in an image is one example. The properties of road networks vary considerably from one geographical environment to another, and they can therefore be used to classify and retrieve such environments. In our work, we define several such environments, and classify them with the aid of geometrical and topological features computed from the road networks occurring in them. In order to compute geometrical and topological features of the road network, we first need to extract the road network from the image, and then convert the output to an appropriate representation. This representation must be independent of any extraction method. Features computed from the representation of the road network are then used to classify the geographical environments. Fisher Linear Discriminant (FLD) analysis is used to select the best feature subset and a linear kernel support vector machine (SVM) classification is performed on the feature set arising from a diverse image database.

### 3.1 Introduction

The retrieval of images from large remote sensing image databases relies on the ability to extract appropriate information from the data, and on the robustness of this extraction [Dasc 05]. Most queries do not concern, for example, imaging modality, but rather information that is invariant to imaging modality, for instance the land cover type of a region. Illumination is another example of such a nuisance parameter. Despite all the work that has been done on classifying geographical environments using, for example, the texture properties of images of those environments, image-

---

based query characterizations are still far from invariant to changes in such nuisance parameters, and they thus fail to be robust when dealing with a large variety of images acquired under different conditions. Query characterizations based on semantic entities detected in the scene, however, are invariant to such nuisance parameters, and thus inferences based on such entities can be used to retrieve images in a robust way. Road networks extracted from an image provide one example: their topological and geometrical properties vary considerably from one geographical environment to another. For example, road networks arising from urban USA will be structurally different from that of urban Europe; similarly road network structures in fields will be different from those of mountains. A set of geometrical and topological features computed from an extracted road network can therefore in principle be used to characterize images or parts of images as belonging to different geographical environments. This differs from much previous work, for example [Wils 97, Luo 01], in that the aim is not to identify the same network in different images, or in a map and an image, and produce a detailed correspondence, but rather to use more general road network properties to characterize other properties of an image, in this case, its geographical environment.

A preliminary study described in [Bhat 06] looked at the classification of a small image database into two classes ‘Urban’ and ‘Rural’, shown in Figure 3.2(a) with image examples shown in Figure 3.3. A small set of topological and geometrical road network features were computed for this study. An unsupervised kernel  $k$ -means clustering was performed to validate the feature set coming from the images of the two classes. The study indicated that the idea had potential, and so a further study went on to examine the classification of a much larger database with seven classes: ‘Urban USA’, ‘Urban Europe’, ‘Mountains’, ‘Villages’, ‘Fields’, ‘Airports’, and ‘Common’ (sea, cloud, desert, snow, etc.) shown in Figure 3.2(b) with image examples shown in Figure 3.4. A larger set of road network features were computed for this study. In order to compute these features, we first need to separate out the road networks occurring in the image. In this study we considered two network extraction methods [Roch 03, Laco 05]. The method of [Roch 03] is based on Higher-Order Active Contours (HOACs). Higher-order active contours are a generalization of standard active contours that use long-range interactions between contour points to include non-trivial prior information about region shape. The method of [Stoi 00, Laco 05] models the line network as an object process, where the objects are interacting line segments. The results obtained from these methods are robust but are not optimal for large image databases as they require manual expertise to set the parameters in the algorithms.

In subsequent studies, we considered the two network extraction methods reported in [Fisc 81, Deso 00]. These methods were rather easier to handle and could easily be adapted to large image database. The method of [Fisc 81] extract roads and linear structures in aerial images using information from multiple image operators. The output of the method is a binary image, which after a distance function computation can serve directly as an input to our method. The method of [Deso 00] detect geometric structures in an image without any *a priori* information. The method focus on

---

extracting aligned structures in images. The output of the method is a list of multiply aligned segments. In order to have a suitable input for our method, we convert the output of this method into a binary image, and use some image processing techniques [Serr 82] to obtain single connected segments.

We then compute a distance function. The distance function resulting from these methods is converted to a graph representation of the road network for feature computation purposes. The conversion is performed by computing the shock locus of the distance function using the method of [Dimi 00, Sidd 02], extended to deal with multiple, multiply-connected components with the depth-first search (DFS) algorithm [Corm 01]. The method identifies the shock points by finding out the limiting behavior of the average outward flux of the distance function as the region enclosing the shock point shrinks to zero. A suitable thresholding on this flux yields an approximation to the shock locus. The graph is constructed by taking triple (or, exceptionally, higher degree) points and end points as vertices, corresponding to junctions and terminals, while the edges are composed of all other points, and correspond to road segments between junctions and terminals.

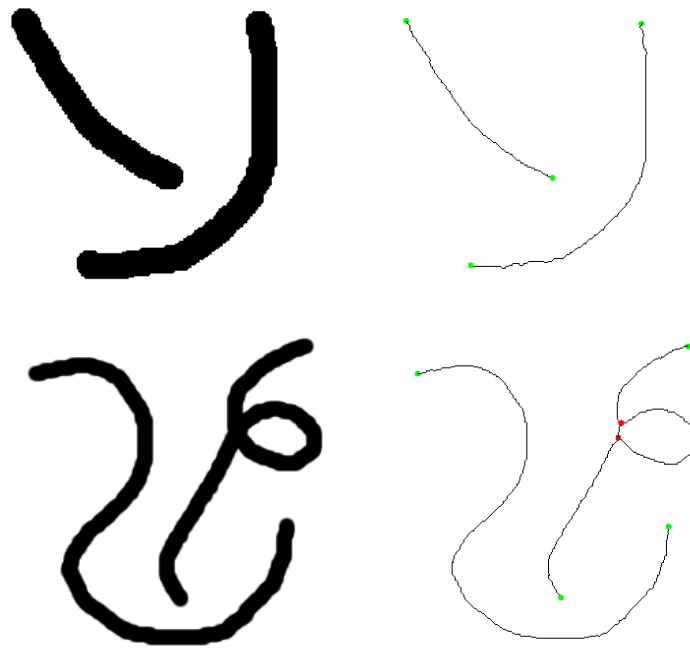


Figure 3.1: Synthetic images and their corresponding graph representation

The shock locus constitutes a graph, where the vertices and the edges contain certain geometrical and topological information. Some examples of graph representations of synthetic images are shown in Figure 3.1. In our study, we represent the road network extracted from an image as a graph. The graph itself captures the network topology, while the network geometry is encoded by decorating vertices and edges with

---



geometrical informations.

Geometrical and topological features were computed from the graph representation of the road network. In order to reduce the dimensionality of the feature space, a suitable feature selection scheme was used, which, in combination with SVM linear kernel classification [Camp 05] on the set of features from road networks, gave promising results for the classification of the different geographical environments contained in a diverse image database.

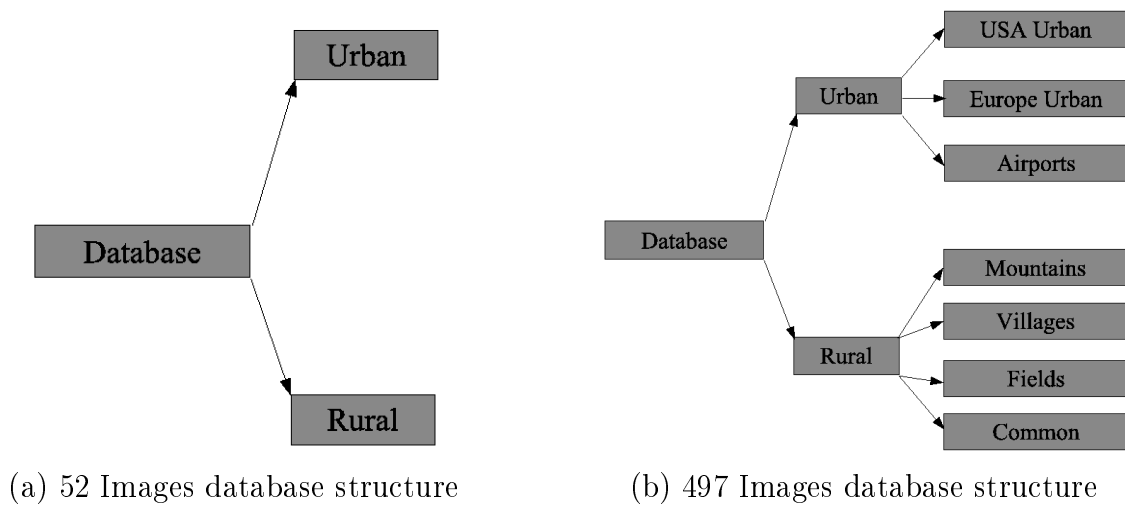


Figure 3.2: Images categorized into different classes.

This chapter will address the preliminary study with a small image database of two classes. The promising results of the unsupervised classification with few geometrical and topological features motivated us to consider a larger image database for further study with a more refined set of classes.



(a) Urban



(a) Rural

Figure 3.3: An example of an urban and a rural class ©CNES.

In Section 3.2, we present a small literature survey of linear structures and road network extraction methods and describe the four network extraction methods considered in our work. Two methods were used in the preliminary study, and two were used in the follow-up studies. We also describe the road network representation into which we convert the outputs of these methods before computing the features. In Section 3.3, we describe the road network features that were used in the preliminary study and later we go on to show the scatter plots of selected pairs of the features. In Section 3.3.1 we evaluate these features with selection and clustering methods. In Section 3.4, we describe the expanded set of road network features introduced to classify the larger database into a more refined set of classes. In Section 3.4.2, we describe the results of a number of classification experiments using feature selection and the SVM linear classifier that was used to perform classification. In Section 3.5, we discuss.

## 3.2 Network extraction and representation

In order to compute geometrical and topological features of the road network, we first need to extract the road network from the image, and then convert the output to an appropriate representation. This representation should be independent of the output of the extraction algorithm, since we do not want to be committed to any single such method. A vast literature is available on the extraction of linear structures in general and road networks in particular for various applications. Mathematical models with variational calculus, differential geometry, fuzzy theory, neural networks, stochastic geometry, morphology and probability theory have been used to formulate the road network extraction algorithms. Research in the field of linear structures or road segment extraction can be traced back more than twenty years or so, with the work of [Cann 86] and the method developed by [Marr 80]. These methods are used to filter out low frequency information in the image while preserving the high frequency information which corresponds to road networks. Other methods which were developed rely on image texture, shapes and geometry, local intensity and changes of patterns. Road information extraction from optical and Synthetic Aperture Radar (SAR) images has been an important research field. In this section we cite a few references from the vast list. There are some methods particularly developed for the extraction of road networks from aerial images [Neva 80, Groc 82, Lipa 88, McKe 88, Maye 98], satellite images [Bajc 76, Dest 86, Paru 91, Wang 92, Chau 96, Dial 01, Roch 03, Laco 05] and Synthetic Aperture Radar (SAR) images [Tupi 98, Jeon 00, Chen 06]. There has also been quite a lot of work in the field of road network extraction from satellite images with varying resolution [Baum 96, Baum 97]. Some of the more complicated algorithms takes into account the combination of various properties of the image under consideration. A very recent work on automatic road extraction from high resolution satellite images is reported in [Chri 07].

A very classical approach in the context of contour based methods is that of [Neva 80]. The work of [Wang 92] introduced a method called Gradient Direction

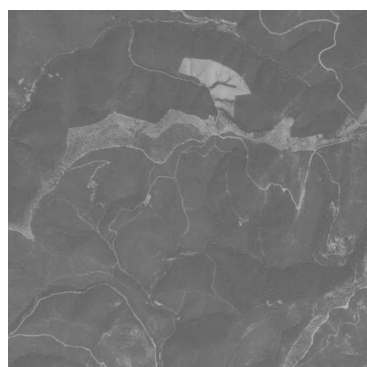
---



(a) USA



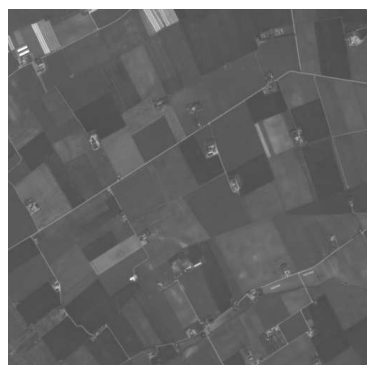
(b) Europe



(c) Mountains



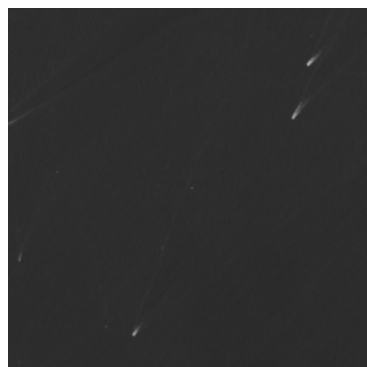
(d) Villages



(e) Fields



(e) Airports



(e) Common (sea)

Figure 3.4: An example of 2 urban, 3 rural, airport and common (sea) classes ©CNES.

Profile Analysis (GDPA) to represent contours. The novel work of [Roch 03] is based on Higher Order Active Contours (HOACs), which are a generalization of standard active contours. The work of [McKe 88] uses road texture and road edge properties independently to track roads in aerial images. There has also been quite a lot of work on road extraction with dynamic programming: a review is given in [Grue 95]. The work of [Fisc 81, Merl 96] and [Gema 96] also used the dynamic programming approach to delineate curves in noisy images. Statistical models to extract road structures also became very popular with the work of [Coop 79] using Markov processes. Further development in this field came with the work of [Barz 96]. Recent work in this field includes the study of [Yuil 00] with improvement of [Barz 96] method. The work of [Laco 05] models line network as an object process.

In the preliminary studies reported in [Bhat 06] we considered two network extraction methods [Roch 03, Laco 05]. The method of [Roch 03] is based on Higher-Order Active Contours (HOACs). HOACs are a generalization of standard active contours that use long-range interactions between contour points to include non-trivial prior information about region shape, in this case the region should be network-like, that is composed of arms with roughly parallel sides meeting at junctions. The output of this method is a distance function defining the region corresponding to the road network. Figure 3.5(b) shows an example of the extracted network. The method of [Laco 05] models the line network as an object process, where the objects are interacting line segments. The output is a set of line segments of varying lengths, orientations, and positions. Figure 3.5(d) and (f) shows examples of the extracted network. This output is converted to the output of [Roch 03] by performing a dilation and then a distance function computation on the resulting binary image. As can be seen in the figures, the methods produces good extraction results, but inspite of this, these methods were not used in the subsequent studies due to the fact that they are not optimal for use with large image databases: manual expertise is needed to set the parameters in the algorithms according to image complexities.

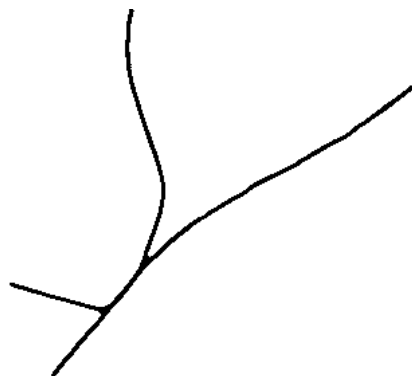
In subsequent studies, we considered the two network extraction methods reported in [Fisc 81, Deso 00]. These methods are rather easier to handle and are well adapted to large image databases. The parameters, once set in the algorithms, work well with images of a certain resolution. The method of [Fisc 81] is a line detector which is based on the homogeneity of the road and the contrast with the neighbourhood. The output of the method is a binary image, which after a distance function computation can serve directly as an input to our method. Figure 3.6 shows examples of the extracted network. The method of [Deso 00] is based on a *contrario* method for the detection of alignments in images. The output of the method is a list of multiply aligned segments. In order to have a suitable input for our method, we convert the output of this method into a binary image, and use morphological closing to obtain single connected segments. Figure 3.7 shows examples of the extracted network. We then compute a distance function which acts as an input to our method.

We use the algorithm of [Dimi 00, Sidd 02] to compute a subpixel skeleton from

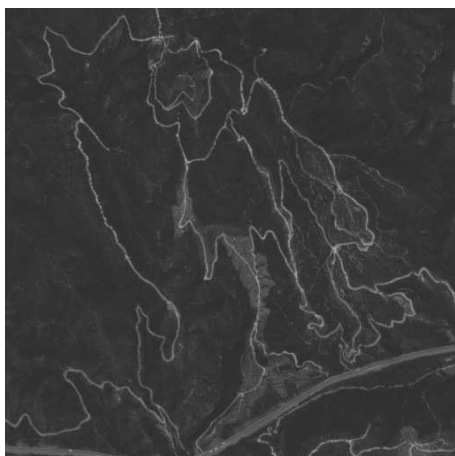
---



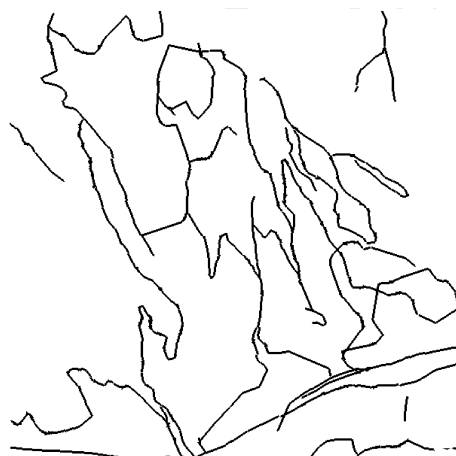
(a) Original image ©IGN



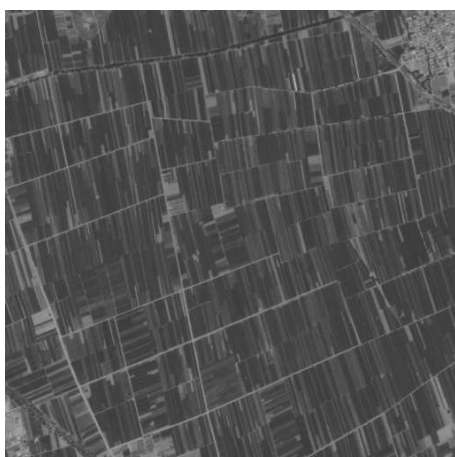
(b) Extracted network



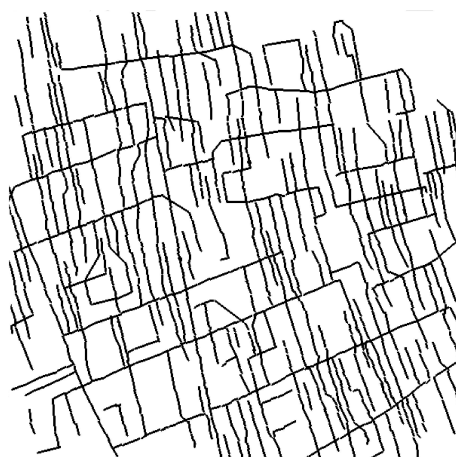
(c) Original image ©CNES



(d) Extracted network



(e) Original image ©CNES

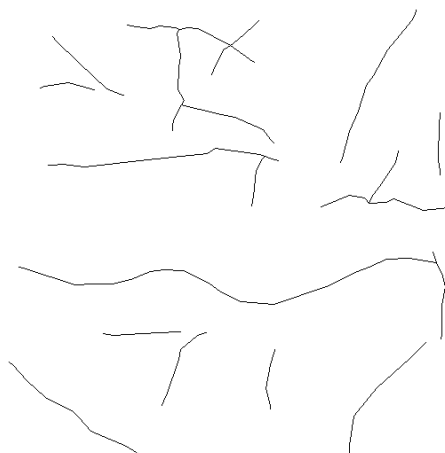


(f) Extracted network

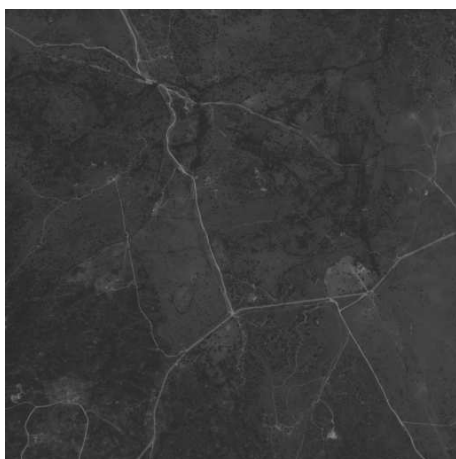
Figure 3.5: Examples from the two extraction methods: (b) method of [Roch 03]; (d) and (f) method of [Laco 05].



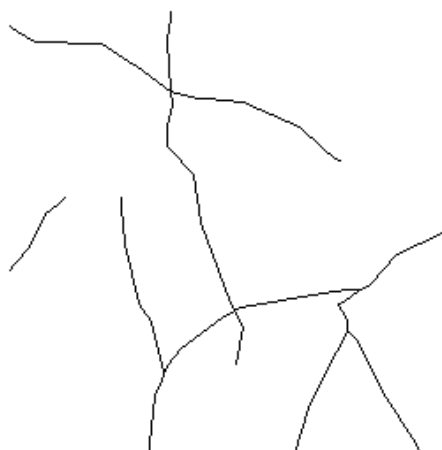
Original image ©CNES



Extracted network



Original image ©CNES



Extracted network.



Original image ©CNES



Extracted network.

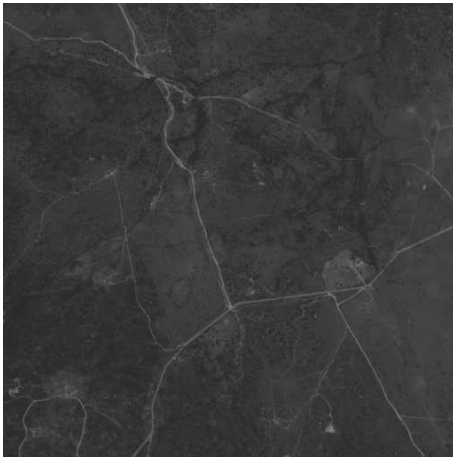
Figure 3.6: Extraction results with the method of [Fisc 81].



Original image ©CNES



Extracted network



Original image ©CNES



Extracted network.



Original image ©CNES



Extracted network.

Figure 3.7: Extraction results with the method of [Deso 00].

the distance function defining the road network region. This algorithm is used because it accurately preserves the region topology. The original version of the algorithm in [Dimi 00, Sidd 02] does not consider multiply connected components. We extend it to deal with multiple, multiply-connected components using a depth-first search (DFS) algorithm [Corm 01]. The main idea of the algorithm is to distinguish the skeletal points of an object from the non-skeletal ones. The approach is to consider the limiting behavior of the average outward flux of the gradient of the distance function,  $\nabla D$ , as the region enclosing the skeleton shrinks to zero. The measure of the average outward flux of the gradient of the distance function is a hyperbolic partial differential equation. A hyperbolic system can be considered to be a conservation law. In the case of hyperbolic systems, there is a notion of weak solution. For instance, a smooth initial curve may develop singularities or shocks as it progresses. These are exactly the points where classical derivatives do not exist.

Considering the limiting values of the outward flux and the average outward flux of the vector field  $\nabla D$  through a convex curve bounding a region  $\Omega$  reveals the property, that whereas the limiting value of the outward flux is zero for both skeletal and non-skeletal points, the average outward flux has different limiting behaviors at the skeletal points than at the non-skeletal ones. The method identifies the shock points by finding the limiting behavior of the average outward flux of the distance function as the region enclosing the shock point shrinks to zero. A suitable thresholding on this flux yields an approximation to the shock locus. The graph is constructed by taking triple (or, exceptionally, higher degree) points and end points as vertices, corresponding to junctions and terminals, while the edges are composed of all other points, and correspond to road segments between junctions and terminals. Figure 3.8(A) shows a shock locus and Figure 3.8(B) shows an Euclidean graph with each edges,  $e$  and vertex  $v$  containing information:  $v = \{\text{position, terminal or junction point, number of edges at junctions, etc.}\}$  and  $e = \{\text{length of original road segment, average curvature of road segment, etc.}\}$

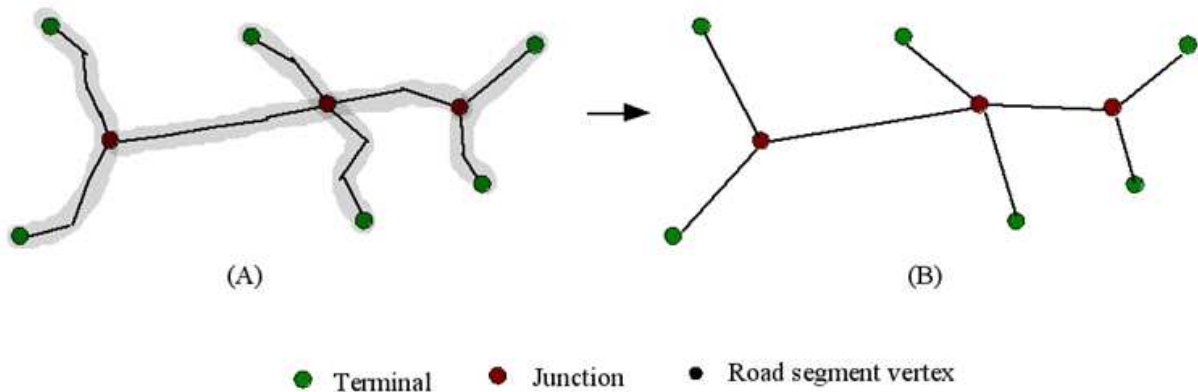


Figure 3.8: Graph representation

Figure 3.9 shows an example of the graph representation of a road network. The road network Figure 3.9(b) is first extracted from the input image Figure 3.9(a). The



methods cited above are then used to generate the shock locus Figure 3.9(c), which is then converted to an Euclidean graph representation Figure 3.9(d). The vertices and edges are decorated with geometrical quantities computed from the shock locus. The features are then computed from the Euclidean graph and its decorations. The features computed from the graph representation are described in section 3.3 and section 3.4.1.

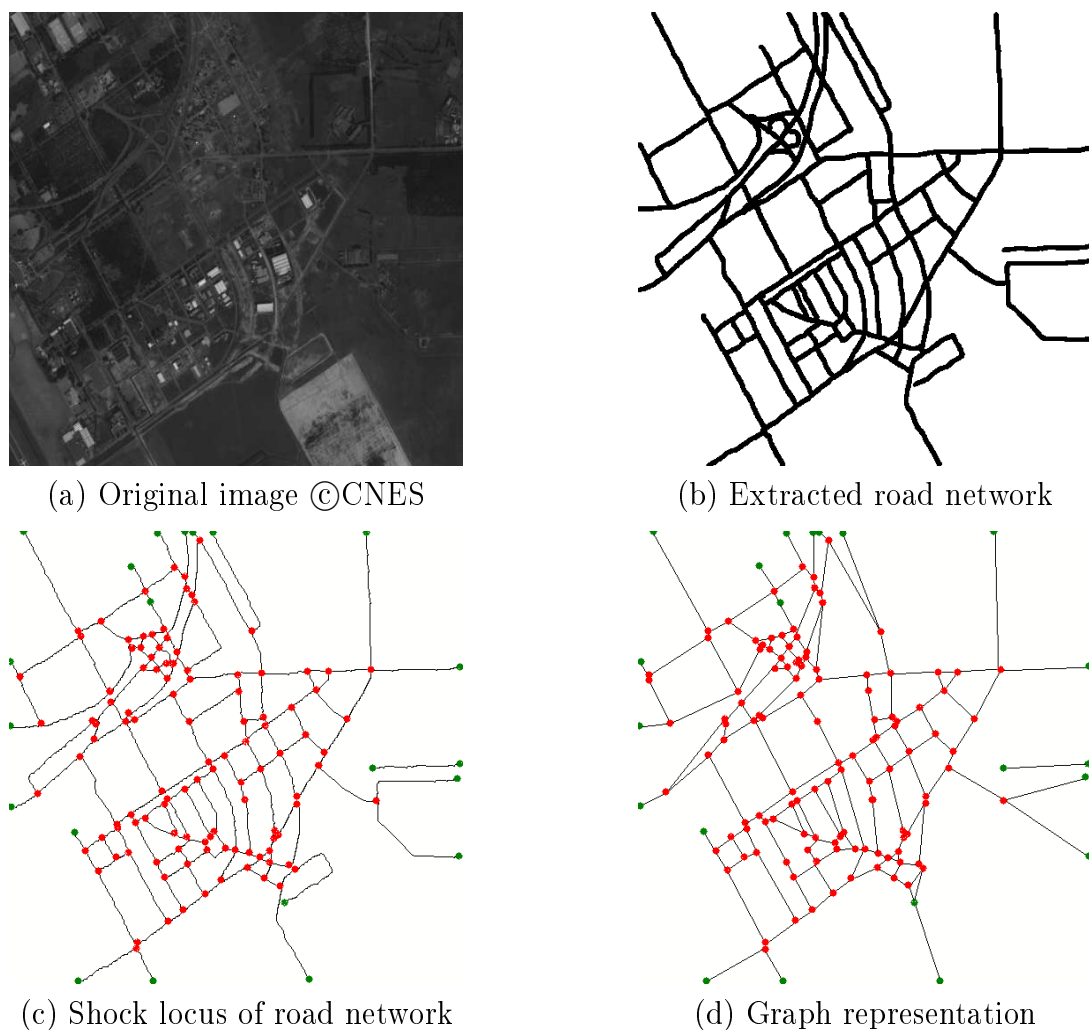


Figure 3.9: An example of the graph representation.

### 3.3 Preliminary study

The preliminary study uses a database of 52 images, and attempts to classify them into two classes, ‘Urban’ and ‘Rural’, using five topological and geometrical road network features. These features, along with the definitions and notations used for quantities involved in their definition, are described in Table 3.1 (the features are marked in bold). They fall into three groups: two measures of ‘density’, two measures of ‘curviness’,

and one measure of ‘homogeneity’. Let  $v$  be a vertex, and  $e$  be an edge. Let  $l_e$  be the length of the road segment corresponding to  $e$ , and let  $d_e$  be the length of  $e$ , that is the Euclidean distance between its two vertices. Let  $m_v = \sum_{e: v \in e} 1$  be the number of edges at a vertex. Then  $N_J = \sum_{v: m_v > 2} 1$  is the number of junction vertices. Let  $\Omega$  be the area of the image in pixels. We define the ‘junction density’ to be  $\tilde{N}_J = \Omega^{-1} N_J$ . This is intuitively a useful measure to separate urban and rural areas: we expect urban areas to have a higher value of  $\tilde{N}_J$  than rural areas. Similarly, we define the ‘length density’ to be  $\tilde{L} = \Omega^{-1} \sum_e l_e$ . Again, we expect urban areas to have a higher value of  $\tilde{L}$  than rural areas. Note that one can have a high value of  $\tilde{L}$  and a low value of  $\tilde{N}_J$  if junctions are complex and the road segments are ‘space-filling’.

Let  $p_e = l_e/d_e$ , and  $\bar{k}_e = l_e^{-1} \int_e ds |k_e(s)|$ , i.e., the absolute curvature per unit length of the road segment corresponding to  $e$ . Although it may seem natural to characterize the network using the average values per edge of these quantities, in practice we have found that more useful features are obtained by using their variances. We thus define the ‘ratio of lengths variance’ to be the variance of  $p_e$  over edges,  $\text{var}(p)$ , and the ‘variance of average curvature’ to be the variance of  $\bar{k}_e$  over edges,  $\text{var}(k)$ . Note that it is quite possible to have a large value of  $p_e$  for an edge while having a small value of  $\bar{k}_e$  if the road segment is composed of long straight segments, and vice-versa, if the road ‘wiggles’ rapidly around the straight line joining the two vertices in the edge. We expect rural areas to have high values of one of these two quantities, while urban areas will probably have low values, although this is less obvious than for the density measures.

To measure network homogeneity, we divide each image into four quadrants, labelled  $a$ . Subscript  $a$  indicates quantities evaluated for quadrant  $a$  rather than the whole image. Let  $M_{J,a} = \sum_{v \in a: m_v > 2} m_v$  be the number of edges emanating from junctions in quadrant  $a$ . This is very nearly twice the number of edges in  $a$ , but it is convenient to restrict ourselves to junctions to avoid spurious termini at the boundary of the image. Let  $\tilde{M}_{J,a} = \Omega_a^{-1} M_{J,a}$  be the density of such edges in quadrant  $a$ . Then we define the ‘network inhomogeneity’ to be the variance of  $\tilde{M}_{J,a}$  over quadrants,  $\text{var}(\tilde{M}_J)$ .

In the experiments reported in the next section, all the images have the same resolution. However, more generally we need to consider the scaling of the above quantities with image resolution. We assume that changing the resolution of the image does not change the extracted road network. This can happen, for example, if the network extracted from a lower resolution image lacks certain roads contained in the network extracted from a higher resolution image because they are less than one pixel wide. This effectively limits the range of the resolutions that we can consider simultaneously. Having assumed this, invariance to image resolution is easily accomplished by converting quantities in pixel units to physical units using the image resolution.

Figure 3.10 shows scatter plots of selected pairs of the features described above as computed from a database of 52 SPOT5, 5m resolution images of size 512x512 pixels

with 26 images of each class, representing various types of urban and rural landscapes. The plots show, from left to right, top to bottom:  $\tilde{N}_J$  versus  $var(k)$ ;  $\tilde{L}$  versus  $var(k)$ ;  $\tilde{N}_J$  versus  $var(p)$ ;  $\tilde{L}$  versus  $\tilde{N}_J$ ;  $var(p)$  versus  $var(k)$ ;  $var(\tilde{M}_J)$  versus  $\tilde{L}$ . Blue circles correspond to images of urban areas, red stars from images of rural areas.

As can be seen from the plots, the junction densities,  $\tilde{N}_J$ , for urban areas are for the most part higher and more varied than those for rural areas, where the values are small. The network length density,  $\tilde{L}$ , behaves similarly. The behavior of the average curvature variance,  $var(k)$ , is perhaps less expected. Urban areas show generally higher values of this feature, and there is also a wide spread of values, while rural areas demonstrate, with a few exceptions, very little curvature variance. The ratio of lengths variance,  $var(p)$  is also interesting. Both classes cluster around low values, again with a few exceptions in the case of rural areas. The average curvature variance varies widely w.r.t ratio of length variance for urban areas, whereas the ratio of length variance varies widely w.r.t average curvature variance for rural areas. The network inhomogeneity  $var(\tilde{M}_J)$  for rural areas is low and does not vary with the network length density, whereas for urban areas the network inhomogeneity is low but varies widely with the network length density. Perhaps the most intriguing plot is length density against junction density, in which both rural and urban data points follow a well defined curve, well approximated by  $\tilde{L} = 1.4\tilde{N}_J^{1/2}$ . Naively, if there is on average one junction for every  $a^2$  pixels, the junctions will be separated by a distance  $O(a)$ . If each junction has  $r$  edges, there will be on average  $r/2$  segments of length  $O(a)$  for every  $a^2$  pixels, and thus  $\tilde{L} \simeq (r/2)\tilde{N}_J^{1/2}$ . For a square lattice  $\tilde{L} = 2\tilde{N}_J^{1/2}$ , so that in some sense road networks are ‘less connected’ than a square lattice. However, this analysis effectively assumes a uniform distribution of junctions and no symmetry-breaking ‘clustering’ effects due to dependencies between different junction positions. In general, there seems to be no reason *a priori* why even the exponent  $1/2$  should be consistent across images and classes, let alone the pre-factor, but this is true for images of a certain resolution even for a well-detected network. It remains to be seen whether this relation is preserved in a larger data set. Finally, and most importantly, note that the points from the two classes are quite well separated in many of the plots, making it reasonable to use these features for classification.

### 3.3.1 Feature selection and clustering

As a first step towards classification, the features are selected from the scatter plot analysis of pairs of features. The scatter plot results indicate that the selected features represent suitable choices for classification based on road network properties. Apart from the features mentioned in Table 3.1, we introduce 6 other features: the network length,  $L$ , average curvature mean,  $mean(k)$ , network area density,  $\tilde{A}$ , ratio of length mean,  $mean(p)$ , density of junction edges,  $\tilde{E}_J$  and mean of density of junction edges,  $mean(\tilde{M}_J)$ . A detailed description and analysis of these features are reported in Section 3.4.1. These 6 features along with the 5 features defined in Table 3.1 are then

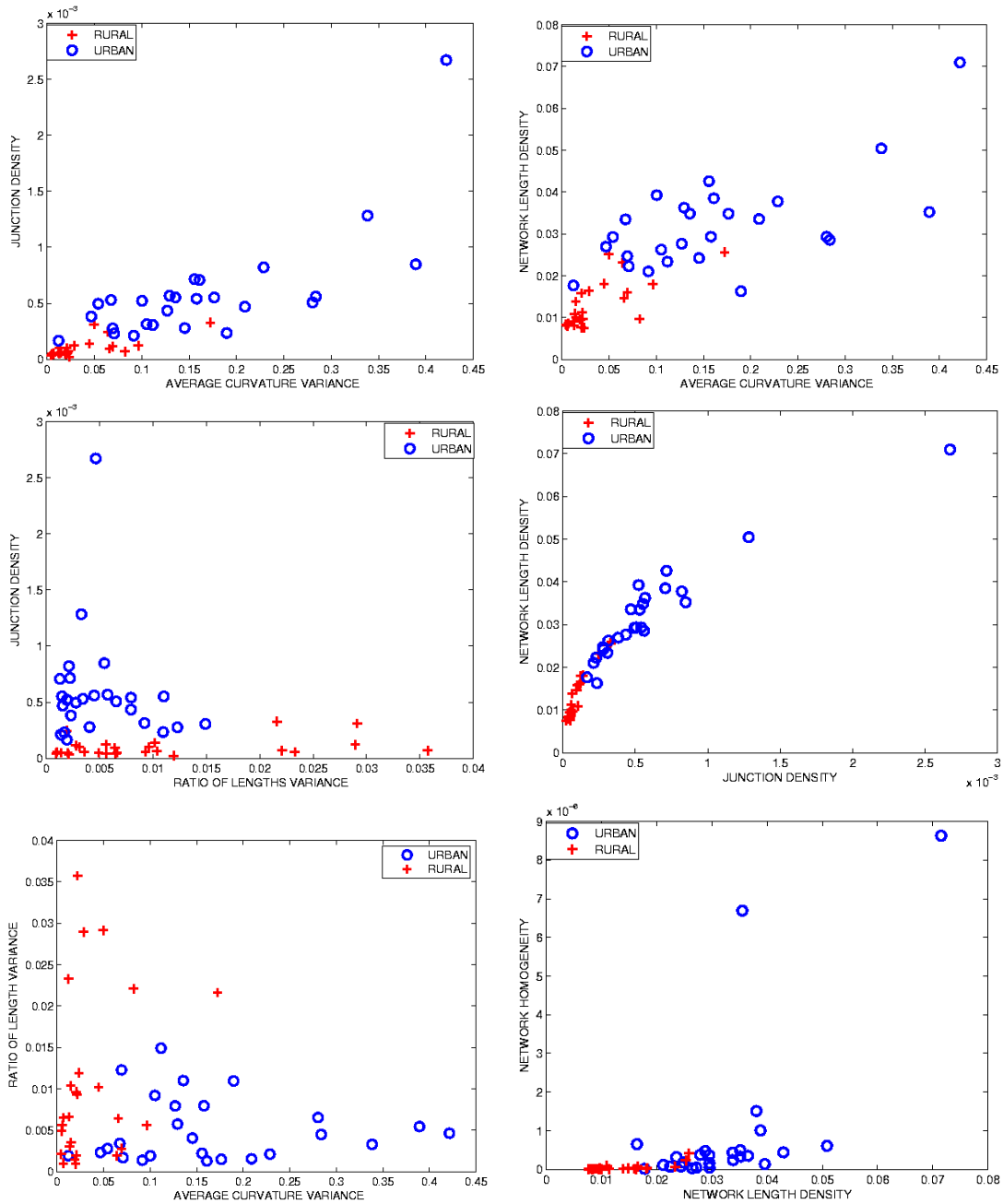


Figure 3.10: Scatter plots of selected pairs of features from manually extracted networks. Red stars correspond to rural areas, blue circles to urban areas. From left to right, top to bottom:  $\tilde{N}_J$  versus  $\text{var}(k)$ ;  $\tilde{L}$  versus  $\text{var}(k)$ ;  $\tilde{N}_J$  versus  $\text{var}(p)$ ;  $\tilde{L}$  versus  $\tilde{N}_J$ ;  $\text{var}(p)$  versus  $\text{var}(k)$ ;  $\text{var}(\tilde{M}_J)$  versus  $\tilde{L}$ .

Notation	Description
$m$	Number of edges in graph
$\Omega$	Area of image
$a$	Quadrant label
$l_e$	Length of road segment corresponding to edge $e$
$m_v$	Number of edges at a vertex $\sum_{e: v \in e} 1$
$N_J$	Number of junction vertices $\sum_{v: m_v > 2} 1$
$\tilde{N}_J$	<b>Junction density</b> $\Omega^{-1} N_J$
$\tilde{L}$	<b>Length density</b> $\Omega^{-1} \sum_e l_e$
$d_e$	Euclidean distance between vertices in an edge
$p_e$	Ratio of lengths $l_e/d_e$
$var(p)$	<b>Ratio of lengths variance</b> $m^{-1} \sum_e p_e^2 - (m^{-1} \sum_e p_e)^2$
$\bar{k}_e$	Average curvature $l_e^{-1} \int_e ds  k_e(s) $
$var(k)$	<b>Average curvature variance</b> $m^{-1} \sum_e \bar{k}_e^2 - (m^{-1} \sum_e \bar{k}_e)^2$
$M_{J,a}$	Number of junction edges per quadrant $\sum_{v \in a: m_v > 2} m_v$
$\tilde{M}_{J,a}$	Density of junction edges per quadrant $\Omega_a^{-1} M_{J,a}$
$var(M_J)$	<b>Variance of density of junction edges</b> $(1/4) \sum_a \tilde{M}_{J,a}^2 - ((1/4) \sum_a \tilde{M}_{J,a})^2$

Table 3.1: Summary of features computed from road networks

used for classification.

Apart from selecting features from the scatter plot, we move on to select features in a more robust way. Fisher discriminant analysis and the Relief feature selection [Kira 92] algorithm were used to select the most pertinent features and a kNN classification scheme was used on these features for classification. A  $d$ -dimensional feature vector was selected for which the classification error is minimum. We then move on to classify images from the two classes with the selected  $d$ -dimensional feature vector using the kernel  $k$ -means algorithm since the feature data are not linearly separable. We use a Gaussian kernel,

$$\psi(X_1, X_2) = e^{-\frac{\|X_1 - X_2\|^2}{2\sigma^2}} \quad (3.1)$$

where  $X_1$  and  $X_2$  are two feature vectors. The clustering result, displayed in Table 3.2, shows that the two classes can be well partitioned using the 5 features (Table 3.5) selected from the scatter plots. 19 and 25 images from ‘rural’ and ‘urban’ classes respectively were correctly classified, while 1 and 7 images from ‘urban’ and ‘rural’ classes respectively were incorrectly classified. Table 3.3, shows the class partition using 6 features selected (Table 3.6) by FLD analysis. 21 and 25 images from ‘rural’ and ‘urban’ classes respectively were correctly classified, while 1 and 5 images from ‘urban’ and ‘rural’ classes respectively were incorrectly classified. Table 3.4, shows the class partition using 5 features (Table 3.7) selected by the Relief algorithm. 22

and 25 images from ‘rural’ and ‘urban’ classes respectively were correctly classified, while 1 and 4 images from ‘urban’ and ‘rural’ classes respectively were incorrectly classified. The feature selection with different methods in Table 3.5, Table 3.6 and Table 3.7 shows that, the features like, the network length densities,  $\tilde{L}$  and average curvature variance,  $var(k)$  have the most discriminative capacity to classify the two classes followed by features like, the variance of density of junction edges,  $var(\tilde{M}_J)$ , the network length,  $L$  and the network area density,  $\tilde{A}$ .

	Urban	Rural
Class 1	1	19
Class 2	25	7

Table 3.2: Kernel k-means ( $\sigma = 0.5$ ) clustering result with 5 features selected from the scatter plot analysis of pairs of feature vectors.

	Urban	Rural
Class 1	1	21
Class 2	25	5

Table 3.3: Kernel k-means clustering ( $\sigma = 0.5$ ) result with 6 out of 11 features selected by FLD analysis. A kNN classification gives 7.45% classification error.

	Urban	Rural
Class 1	1	22
Class 2	25	4

Table 3.4: Kernel k-means clustering ( $\sigma = 0.5$ ) result with 5 out of 11 features selected by the Relief algorithm. A kNN classification gives 5.69% classification error.

### 3.4 Larger image database with refined classes and features

Motivated by the results obtained in the preliminary study on a small database, we move beyond this simple example and construct a larger image database with a more refined set of classes. The new database has 497 SPOT5, 5m resolution images of size 512x512 pixels. Some examples are shown in Figure 3.4. The aim is to classify them into 7 classes: three urban classes, ‘Urban USA’, ‘Urban Europe’ and ‘Airports’ and four rural classes, ‘Mountains’, ‘Villages’, ‘Fields’ and a ‘Common’ class (more appropriately could be identified as a rejection class with images consisting of sea, clouds, snow, desert, etc.).

Notation	Description
$\tilde{N}_J$	Junction density: $\Omega^{-1}N_J$
$\tilde{L}$	Length density: $\Omega^{-1}\sum_e l_e$
$var(p)$	Ratio of lengths variance: $m^{-1}\sum_e p_e^2 - (m^{-1}\sum_e p_e)^2$
$var(k)$	Average curvature variance: $m^{-1}\sum_e \bar{k}_e^2 - (m^{-1}\sum_e \bar{k}_e)^2$
$var(\tilde{M}_J)$	Variance of density of junction edges: $(1/4)\sum_a \tilde{M}_{J,a}^2 - ((1/4)\sum_a \tilde{M}_{J,a})^2$

Table 3.5: The 5 features selected from the scatter plot analysis of pairs of features.

Notation	Description
$L$	Network length $\sum_e l_e$
$\tilde{L}$	Length density: $\Omega^{-1}\sum_e l_e$
$\tilde{A}$	Network area density: $\Omega^{-1}\Omega_L$
$var(k)$	Average curvature variance: $m^{-1}\sum_e \bar{k}_e^2 - (m^{-1}\sum_e \bar{k}_e)^2$
$mean(k)$	Average curvature mean: $m^{-1}\sum_e \bar{k}_e$
$\tilde{E}_J$	Density of junction edges: $\Omega^{-1}E_J$

Table 3.6: The 6 out of 11 features selected by FLD analysis.

Notation	Description
$L$	Network length $\sum_e l_e$
$\tilde{L}$	Length density: $\Omega^{-1}\sum_e l_e$
$\tilde{A}$	Network area density: $\Omega^{-1}\Omega_L$
$var(k)$	Average curvature variance: $m^{-1}\sum_e \bar{k}_e^2 - (m^{-1}\sum_e \bar{k}_e)^2$
$var(\tilde{M}_J)$	Variance of density of junction edges: $(1/4)\sum_a \tilde{M}_{J,a}^2 - ((1/4)\sum_a \tilde{M}_{J,a})^2$

Table 3.7: The 5 out of 11 features selected by the Relief algorithm.

To help in the more complex classification problem, and also to provide input for the feature selection procedure to be used later, we define some extra road network extraction features, described in Table 3.8. The addition of these extra features brings the total number of road network features to 16. These features can be categorized into six groups: six measures of ‘density’, four measures of ‘curviness’, two measures of ‘homogeneity’, one measure of ‘length’, two measures of ‘distribution’ and one measure of ‘entropy’. We will define the new features in the following section. The total set of 16 features are summarized in 6 groups as shown in Table 3.9.

### 3.4.1 New features from the graph

In this section we will describe the new features which will be used in the following section for the classification of the new database. Let  $m_v = \sum_{e:v \in e} 1$  be the number of edges at a vertex. Then  $E_J = \sum_{m_v > 2} m_v$  is the number of junction edges. Let  $\Omega$  be the area of the image in pixels. Similarly to the definition of ‘junction density’, we define the ‘density of junction edges’ as  $\tilde{E}_J = \Omega^{-1} E_J$ . This is intuitively a useful measure to separate urban and rural areas: we expect urban areas to have a higher value of  $\tilde{E}_J$  than rural areas.

The network area density,  $\tilde{A}$ , is computed from the binary image as the number of pixels corresponding to the road network from the extracted binary image divided by the image area,  $\Omega$ . This measure is useful in classification as its value is high for urban networks as compared to rural networks. The network length,  $L$ , is computed from the graph as the total length of road segments,  $\sum_e l_e$ . This feature, like the network area density feature, is useful in characterizing urban and rural network structure.

As can be seen in Figure 3.9(d), many junction points may be clustered around a small area in the network. To obtain a local characterization of the junction density, we define a measure called ‘junction density in a circular region’,  $\tilde{N}_{r,j} = \Omega_{j,r}^{-1} \sum_{v \in \Omega_{j,r}: m_v > 2} 1$ . This is the density of junction points falling in a circular region of radius  $r$  centered at junction point  $j$ . We then compute the mean and the variance of these junction densities over all junction points,  $mean(\tilde{N}_{r,j})$  and  $var(\tilde{N}_{r,j})$ . Rural network structures will show a low  $var(\tilde{N}_{r,j})$ , indicating a very sparse network. A high  $var(\tilde{N}_{r,j})$ , indicates a range of densities in the same image. The  $mean(\tilde{N}_{r,j})$  is also used as a measure of sparsity of network structures.

In order to distinguish between the two urban classes (USA and Europe), the entropy of the histogram of angles at junctions,  $H_\beta$ , where  $\beta_j$  is the vector of angles between road segments at junction  $j$ , is a good measure. As is evident from the physical characteristics of these road network structures, roads in the USA tend to be parallel and cross each other orthogonally forming T-junctions or crossroads, whereas European roads have broader angle distribution at junctions. Thus it seems natural that  $H_\beta \leq 2$  bits are necessary to encode information about road segments at junctions for road networks in the USA, whereas for road networks in Europe,  $H_\beta \geq 2$  bits are



Notation	Description
$L$	<b>Network length</b> $\sum_e l_e$
$\tilde{A}$	<b>Network area density</b> $\Omega^{-1}\Omega_L$
$mean(p)$	<b>Ratio of lengths mean</b> $m^{-1}\sum_e p_e$
$mean(k)$	<b>Average curvature mean</b> $m^{-1}\sum_e k_e$
$E_{D,i}$	Number of junctions with $m_v = i$
$var(E_{D,i})$	<b>Variance of edge distribution</b> $(1/\max(m_v))\sum_i E_{D,i}^2 - ((1/\max(m_v))\sum_i E_{D,i})^2$
$mean(E_{D,i})$	<b>Mean of edge distribution</b> $(1/\max(m_v))\sum_i E_{D,i}$
$E_J$	Number of junction edges $\sum_{m_v>2} m_v$
$M_{J,a}$	Number of junction edges per quadrant $\sum_{v\in a:m_v>2} m_v$
$\tilde{E}_J$	<b>Density of junction edges</b> $\Omega^{-1}E_J$
$mean(\tilde{M}_J)$	<b>Mean of density of junction edges</b> $(1/4)\sum_a \tilde{M}_{J,a}$
$\Omega_r$	Area of a circular region of $r$
$\tilde{N}_{r,j}$	Junction density in a circular region $\Omega_{j,r}^{-1}\sum_{v\in\Omega_{j,r}:m_v>2} 1$
$var(\tilde{N}_{r,j})$	<b>Variance of the junction densities over all circular region</b> $(1/N_J)\sum_j \tilde{N}_{r,j}^2 - ((1/N_J)\sum_j \tilde{N}_{r,j})^2$
$mean(\tilde{N}_{r,j})$	<b>Mean of the junction densities over all circular region</b> $(1/N_J)\sum_j \tilde{N}_{r,j}$
$\beta_j$	Vector of angles between segments at junction $j$
$H_\beta$	<b>Entropy of histogram of road segment angles with bin size <math>30^\circ</math></b>

Table 3.8: Summary of some more road network features

necessary. The same measure can also be used to distinguish between Mountains and Fields, while the ‘density’ features distinguish rural networks from urban networks.

A ‘distribution’ measure of edges at a vertex provides us with information as to how the edges at a vertex are distributed in the network. Let  $E_{D,i}$  be the proportion of junction points with  $i$  edges at them. We use  $mean(E_{D,i})$  and  $var(E_{D,i})$  as features. The variance of the edge distribution is lower in the case of networks in urban areas as opposed to rural, and it is lower also in the case of urban networks in the USA as opposed to in Europe.

Feature groups	Features
6 Density measures	Length, Junction, Area, Junction in circular region (mean, variance), Junction edges
4 Curviness measures	Ration of lengths (mean, variance), Average curvature (mean, variance)
2 Homogeneity measures	Density of edges in quadrants (mean, variance)
1 Length measure	Network length
2 Distribution measures	Edges at vertex (mean, variance)
1 Entropy measures	Histogram of angles at junctions

Table 3.9: Summarization of the feature groups.

### 3.4.2 Classification

As before we need to consider the scaling of the above quantities with image resolution. A detailed view of the resolution dependency of the extraction methods and the feature set will be discussed in Chapter 5. For the moment we assume that the feature set is invariant to image resolution.

The features described in the above sections were computed for a database of 497 SPOT5, 5m resolution images. To provide ground truth, these images were manually classified into the seven classes described above representing various kinds of urban and rural environments. Machine classification was done with a five-fold cross validation on the data set, with 80% of data for training and the remaining 20% for testing in each fold. First we show the results in Table 3.10 of the one-vs-rest SVM linear kernel classification of 497 images into 7 classes, using 15 features selected by FLD [Duda 00] from 16 features with the extraction method of [Deso 00] with mean error of 30.2% and a standard deviation of 3.0%. In this result we see that the Villages, Fields and Europe Urban classes are highly confused with each other. Then we show the results

in Table 3.11 on the same set of images using 13 features selected by FLD from 16 features with the extraction method of [Fisc 81] with mean error of 39.2% with a standard deviation of 1.81%. In this result we see that the apart from Villages, Fields, Europe Urban and USA Urban classes being highly confused with each other, the Mountains class is also confused with Fields, Europe Urban and USA Urban classes. The classification error dependency on features from the two extraction methods is shown in Figure 3.11.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.154</b>	0.090	0.351	0.097	0.198	0.000	0.000
Mountains	0.073	<b>0.812</b>	0.007	0.000	0.000	0.000	0.000
Fields	0.093	0.053	<b>0.488</b>	0.077	0.205	0.000	0.000
USA	0.133	0.012	0.014	<b>0.720</b>	0.087	0.000	0.000
Europe	0.547	0.033	0.140	0.097	<b>0.510</b>	0.014	0.000
Airports	0.000	0.000	0.000	0.009	0.000	<b>0.986</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>1.000</b>

Table 3.10: Confusion matrix of an SVM linear kernel classification of 497 images with 7 classes with 15 out of 16 features from the graph representation of the road networks extracted with the method of [Deso 00].

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.483</b>	0.015	0.146	0.091	0.186	0.000	0.000
Mountains	0.084	<b>0.354</b>	0.153	0.109	0.129	0.000	0.023
Fields	0.282	0.268	<b>0.470</b>	0.122	0.187	0.017	0.000
USA	0.009	0.132	0.098	<b>0.556</b>	0.067	0.000	0.022
Europe	0.142	0.231	0.133	0.113	<b>0.431</b>	0.000	0.000
Airports	0.000	0.000	0.000	0.009	0.000	<b>0.983</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.955</b>

Table 3.11: Confusion matrix of an SVM linear kernel classification of 497 images with 7 classes with 13 out of 16 features from the graph representation of the road networks extracted with the method of [Fisc 81].

Finally we show the results of the one-vs-rest SVM linear kernel classification of 497 images into 7 classes, using 32 features, in Table 3.12 (16 features each from the graph for 2 network extraction methods). There is a mean error of 24.7% with standard deviation of 2.93%.

With such a large number of features, and with some similarity between different features, it seems likely that there is some redundancy in the feature space. This redundancy can be reduced by feature selection. In the final classification experiment,

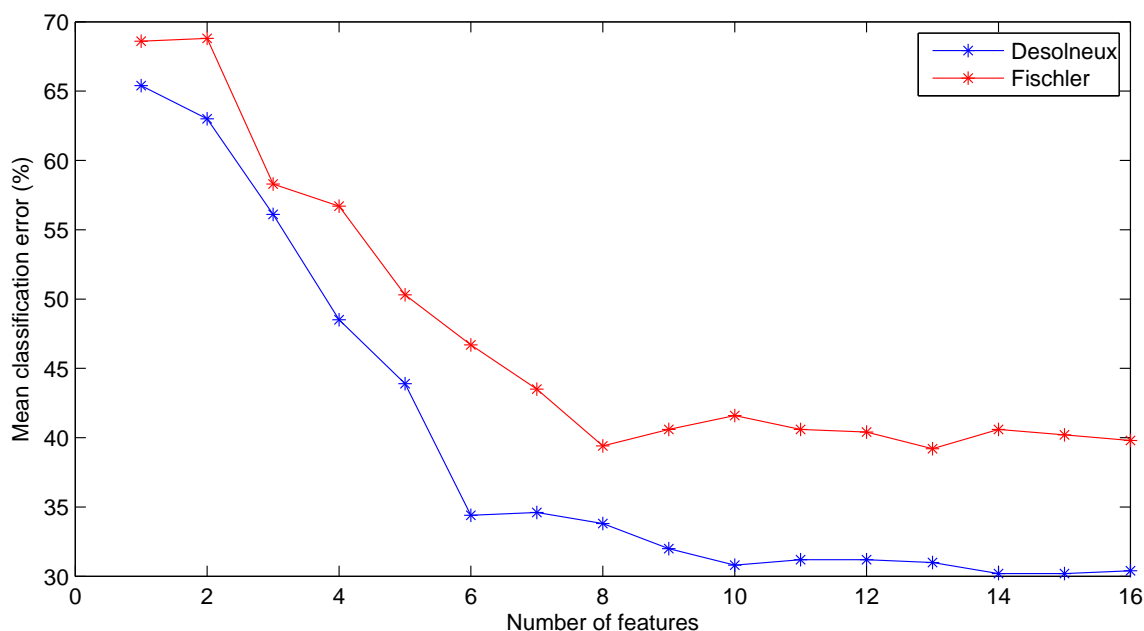


Figure 3.11: Classification error vs number of features from the two extraction methods of [Deso 00] and [Fisc 81].

we performed feature selection using a FLD analysis, followed by an SVM linear kernel classification on the selected feature set. The result of the classification is shown in Table 3.13. The SVM linear kernel classification on the 30-dimensional feature space selected by the FLD shows a mean error of 24.5% with a standard deviation of 2.92%. The classification error dependency on features is shown in Figure 3.12. An overall classification performance summary is depicted in Table 3.14, where classification error in % is given as “mean  $\pm$  standard deviation” error. Not much of an improvement can be seen in the confusion matrix except for the USA urban and Europe urban classes. The selection of features does not decrease the classification error as there is no redundancy in the feature set. In an ideal case the Common class (images of sea body) should show a 100% classification as the images in this class contain no structural information. The ambiguity of other classes being confused with this class (shown in Figure 3.4) is due to the fact that at 5m resolution, SPOT5 images shows contours of waves in seas. The network extraction methods thus extract these wave contours as linear structures. The classification results in Chapter 5 clarify this anomaly.

As can be clearly seen in the confusion matrix, the Villages class is confused with the Fields class and also there is a slight confusion between the Fields and Urban Europe classes. These confusions arise because the road extraction methods fail to detect the fine and densely structured roads present in some regions. The loss of information from these regions reduces the discriminative power of the features and hence results in poor classification. In the next chapter we will discuss a method to circumvent this problem and will show another set of classification results.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.544</b>	0.100	0.187	0.047	0.142	0.000	0.000
Mountains	0.056	<b>0.803</b>	0.000	0.021	0.053	0.000	0.020
Fields	0.242	0.050	<b>0.668</b>	0.025	0.150	0.000	0.000
USA	0.061	0.000	0.026	<b>0.801</b>	0.076	0.000	0.019
Europe	0.085	0.047	0.119	0.092	<b>0.579</b>	0.032	0.053
Airports	0.012	0.000	0.000	0.014	0.000	<b>0.968</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.908</b>

Table 3.12: Confusion matrix of an SVM linear kernel classification of 497 images with 7 classes with 32 features from the graph.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.553</b>	0.097	0.215	0.047	0.128	0.000	0.000
Mountains	0.101	<b>0.808</b>	0.000	0.000	0.047	0.000	0.019
Fields	0.189	0.047	<b>0.635</b>	0.047	0.175	0.000	0.000
USA	0.059	0.000	0.038	<b>0.819</b>	0.047	0.000	0.019
Europe	0.097	0.047	0.113	0.072	<b>0.603</b>	0.032	0.052
Airports	0.000	0.000	0.000	0.014	0.000	<b>0.968</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.909</b>

Table 3.13: Confusion matrix of an SVM linear kernel classification of 497 images with 7 classes with 30 out of 32 features selected by FLD.

Feature Dimension	Selection	Classification Error (%)
16 (Desolneux), 7 classes	Fisher LD, 15	Linear SVM, 30.2±3.0
16 (Fischler), 7 classes	Fisher LD, 13	Linear SVM, 39.2±1.81
32, 7 classes	No selection	Linear SVM, 24.7±2.93
32, 7 classes	Fisher LD, 30	Linear SVM, 24.5±2.92

Table 3.14: Classification performance.

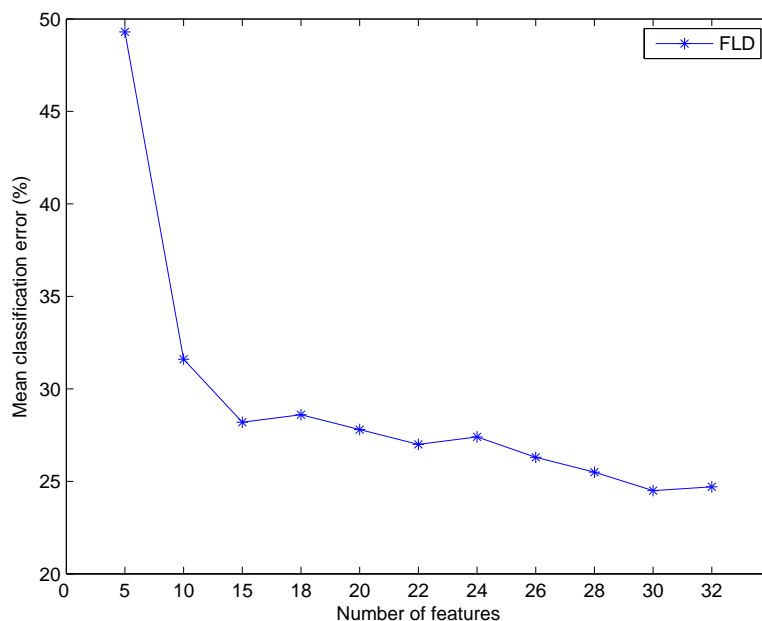


Figure 3.12: Classification error vs number of features combined from the two extraction methods of [Deso 00] and [Fisc 81].

## 3.5 Discussion

The properties of road networks vary considerably from one geographical environment to another, and they can therefore be used to classify and retrieve different geographical environments. We have defined several such environments, and classified them with the aid of geometrical and topological features computed from the road networks occurring in them. In order to compute these features, we have used a graph representation of the road networks. This representation is robust and captures well the network topology.

The classification results reported in this chapter indicate that geometrical and topological features computed from road networks can serve as robust characterizations of a number of geographical environments found in remote sensing images. The road network extraction methods often failed to extract the finely structured road networks in small urban areas, with the consequence that the features computed from road networks poorly classify images containing such areas. In order to circumvent the loss of information from these areas a method must be adopted to incorporate the properties of these areas in our study. The next Chapter outlines a method to characterize geometrical and topological features of urban regions.

A comparison study of the features from the graph representation of the network extracted by different methods particularly designed for mid-resolution satellite images could be interesting. These methods can then be tested and efficiently adapted to deal

with large satellite image databases, where time of execution is an important factor.

---

## Chapter 4

# Region Structural Information

The road extraction methods are in general resolution dependent. A road extraction method originally developed for metre resolution images will extract many 'road-like' linear structures in a sub-metre resolution image. These structures in turn will provide us with redundant information for the proper classification of geographical environments. An optimal road network extraction algorithm that accurately delineates road structures for all practical purposes is very hard to achieve. The real world scenario is too complex to handle and many sophisticated algorithms for the extraction of pertinent road structure information rely on the specificity of the applications. In many application scenarios, road extraction algorithms can be tuned to consider only a limited set of characteristics, and the algorithms fail to identify new characteristics whenever encountered in an application domain. There has been a great deal of study of the extraction of structures of urban regions from high resolution images. New sensor technology has provided us with high resolution images of the Earth's surface which in turn has induced many researchers to work in the field of road and man-made structure extraction from high-resolution satellite and Synthetic Aperture Radar (SAR) data. The methods used in our study are robust to many such road characteristics but often fail to extract the narrow and finely structured road networks which are almost hidden in small urban areas. An example of this is shown in Figure 4.1.

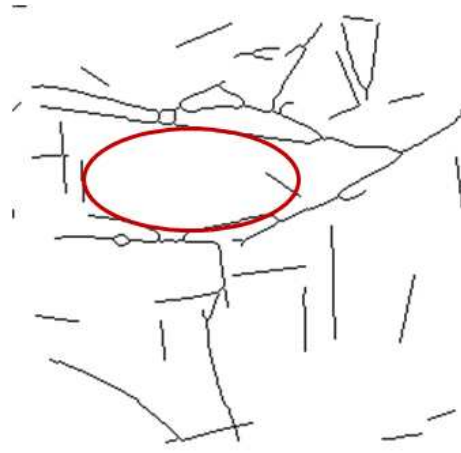
This failure of the extraction methods and hence the features computed from road networks results in a poor classification of images containing such areas. In order to obtain some meaningful information from these regions, we need to segment such areas occurring in the images. Geometrical information can be extracted from these parts of the images to improve the classification of the geographical environments. A new set of features were thus computed on segmented urban areas. This set of features was then combined with the existing road network features. To validate the feature set, Fisher Linear Discriminant (FLD) analysis was used to select the best features from the combined set and linear kernel support vector machine (SVM) classifications were performed on the feature set arising from a diverse image database.

---





Original Image ©CNES



Extracted network

Figure 4.1: An example of network extraction failure in a small urban region.

## 4.1 Urban region structures

Recent advances in sensor technology has provided us with images of sub-metre resolution where man-made objects such as roads are clearly visible, and as such could be detected reliably with numerous methods available in the literature. The study presented here uses metre resolution (5m) images, where the road networks in a dense urban structure are latent and hence remain undetected. This limitation of the road network extraction methods used in our study prevents us from extracting the potential information in these urban regions.

Studies of urban region structures have been carried out by researchers from different domains for the past few decades. These studies involve urban settlement management, disaster management and the detection of changes in land cover and land use. The main problem arising in these studies is the proper definition of urban areas. The definition varies from country to country. Urban areas are in general defined by their boundaries and population density. Satellite images are used to define urban areas in a consistent way.

There is a vast literature on the extraction of urban regions from images coming from different sources, for example Landsat TM images, Multispectral SPOT images and Polarimetric Synthetic Aperture Radar. Urban region classification work can be dated back to the study of texture for classifying regions by [Wesz 76] and the region extraction methods on aerial images introduced by [Naga 79]. Since then, people have used different mathematical models to extract urban regions from images. In general, texture analysis is used to characterize urban morphology with the spatial distribution of ground spectral and radiometric variations. The most common texture analysis methods use fixed size windows to compute the textural properties within this window and then move the window over the entire image. These methods are



Figure 4.2: Examples of small urban regions ©CNES.

strongly dependent on the choice of the window size and shape. Figure 4.3 shows an example of the window size dependency factor. The work of [Lore 99] uses texture with probabilistic and fuzzy classification to detect urban region in satellite images. The work of [Rell 02] also uses classical texture analysis to characterize urban areas. Texture analysis with Gaussian Markov models in order to extract urban areas from SPOT and ERS images is reported in [Vive 03]. Another texture-based method to analyze urban structures in satellite and SAR images can be found in [Dell 03]. A study on urban area analysis with a combination of different operators is reported in [Bess 96]. A method is proposed in [Yang 06] to extract urban regions from SAR images with Gaussian Markov Random Field (GMRF) models.

Geographers often describe an urban spatial structure with mathematical models taken from fractal theory. Mathematically, a fractal is the limit set of a repeatedly replicating geometrical object. Natural objects have complex shapes and hence classical geometry often fails to correctly describe their forms. Geographers often use fractals to measure the urban morphology [Tann 05].

Wavelet transforms have been used in the context of urban region extraction from panchromatic, multispectral and SAR images. The works of [Garz 02, Yu 99] uses wavelet transforms to detect and extract urban environments from images. Grey scale mathematical morphology formulated by [Serr 82] and [Math 75] has also been widely used to characterize and extract the topological aspects of urban areas in images. The work of [Merg 85] uses mathematical morphology to analyze structural forms in satellite images. [Maup 97] uses mathematical morphology in combination with fuzzy logic to capture the spatial changes in urban settlements. The method presented a new approach to spatial change detection. Watershed segmentation was introduced in image analysis by [Buec 79], and was later used for finding the structures in an image, based on the detection of edges. A watershed segmentation is the thinning of a gradient image, resulting in the formation of regions and region contours. The standard watershed method applied to images causes over-segmentation. This is mainly due to the presence of the textural content in the image caused by varying object size and also due to the sensor resolution.

The work by [Welc 82] highlights the dependence on spatial resolution of the characterization of urban areas. Compact urban morphology and narrow and dense road network patterns will require higher spatial resolution to explore the necessary information, as these are less prominent in lower resolution images. This fact will be addressed in Chapter 5, where we will show the resolution dependence of the features used for classifying several geographical environments.

---

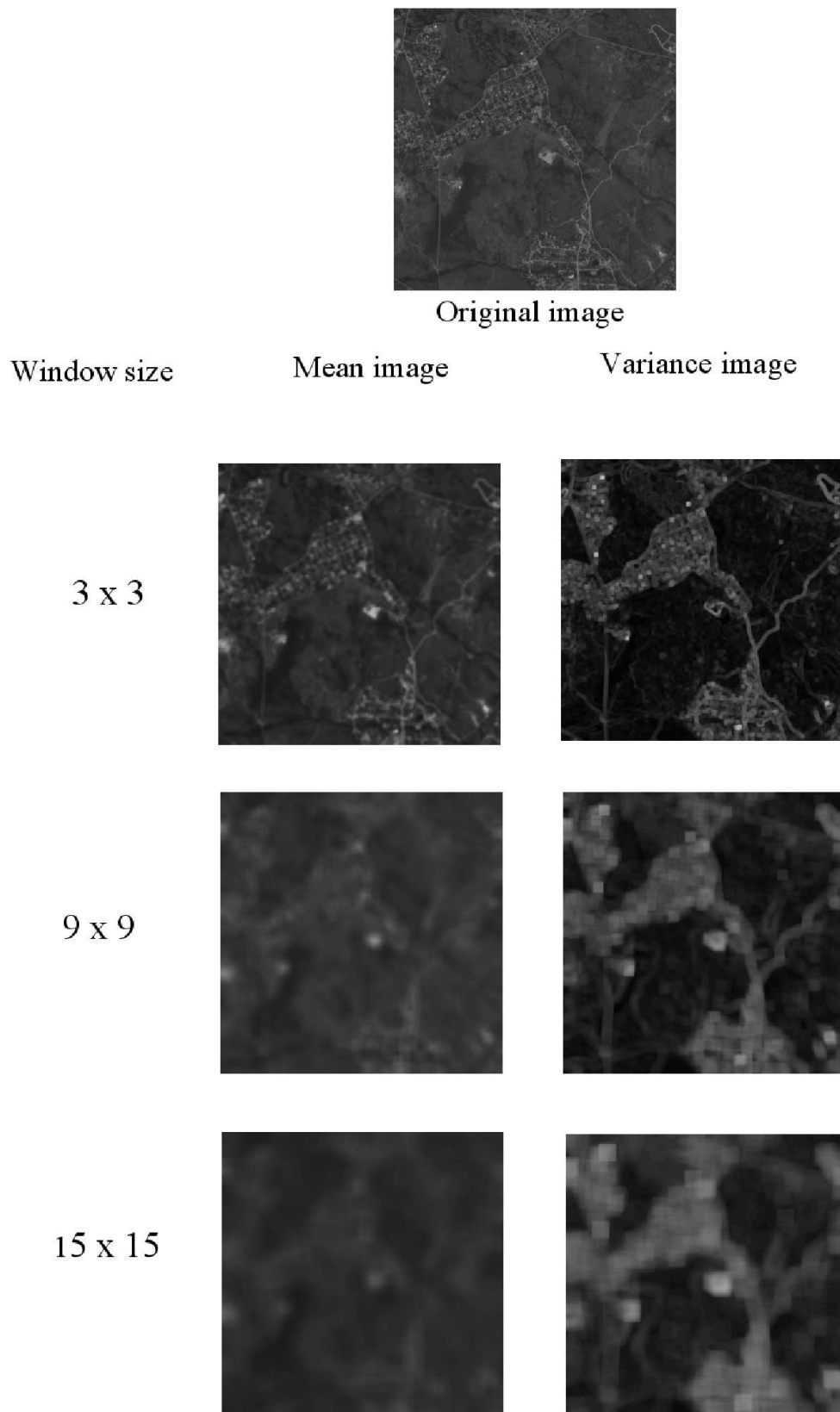


Figure 4.3: Image texture analysis with window size variation.

## 4.2 Segmentation of urban regions

The heterogeneity and the geometrical complexity of urban structures in low radiometric and metre resolution (2m or 5m) images show textural effects for objects with a width of a few pixels. Texture information has been widely used and is a very popular method to infer useful information about an urban area. The texture statistics developed by [Hara 73] has been widely used for image classification. According to Haralick, texture is 'the spatial distribution of tonal variation'. Studies by [Gong 92] showed the usefulness of textural statistics for distinguishing between land cover classes. The work of [Zhan 03] evaluates the performance of different texture features and their combination for mapping urban structures using SPOT images. Work has been done on the extraction of urban regions from SPOT5 images using texture analysis with probabilistic approaches [Desc 93]. In this work, the urban area extraction process is divided into two steps. The first step consist of extracting some discriminant texture parameters from the image with a 2D Markov Random Field (MRF) model. The original image is divided into simple parameter images. In the second step, the parameter images are clustered and false alarms are removed to obtain the segmented urban areas. Figure 4.4 shows an example from this method.

A work very close to our method using mathematical morphology has been reported in [Pesa 01], where urban region detection is based on the residual of the grey value function and a composition of morphological opening and closing transforms using the geodesic metric.

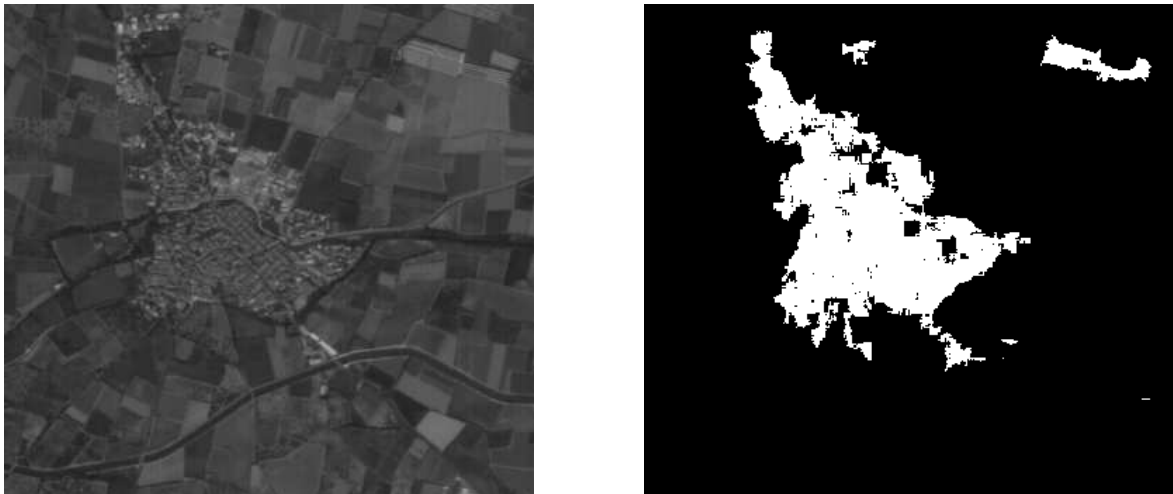


Figure 4.4: Example of urban area extraction using the probabilistic method of [Desc 93] ©CNES/ENST.

In our study we use the work of [Roux 92], developed to extract urban regions from SPOT images. In SPOT images, urban zones appear to be strongly textured and the problem of extraction of the regions is essentially a problem of differentiation of textures. The method used here is inspired by the work of [Sere 89] and [Khat 89].

---

The principle idea is to extract the zone of high density of light and dark peaks. The techniques used are the mathematical morphology operations of opening and closing.

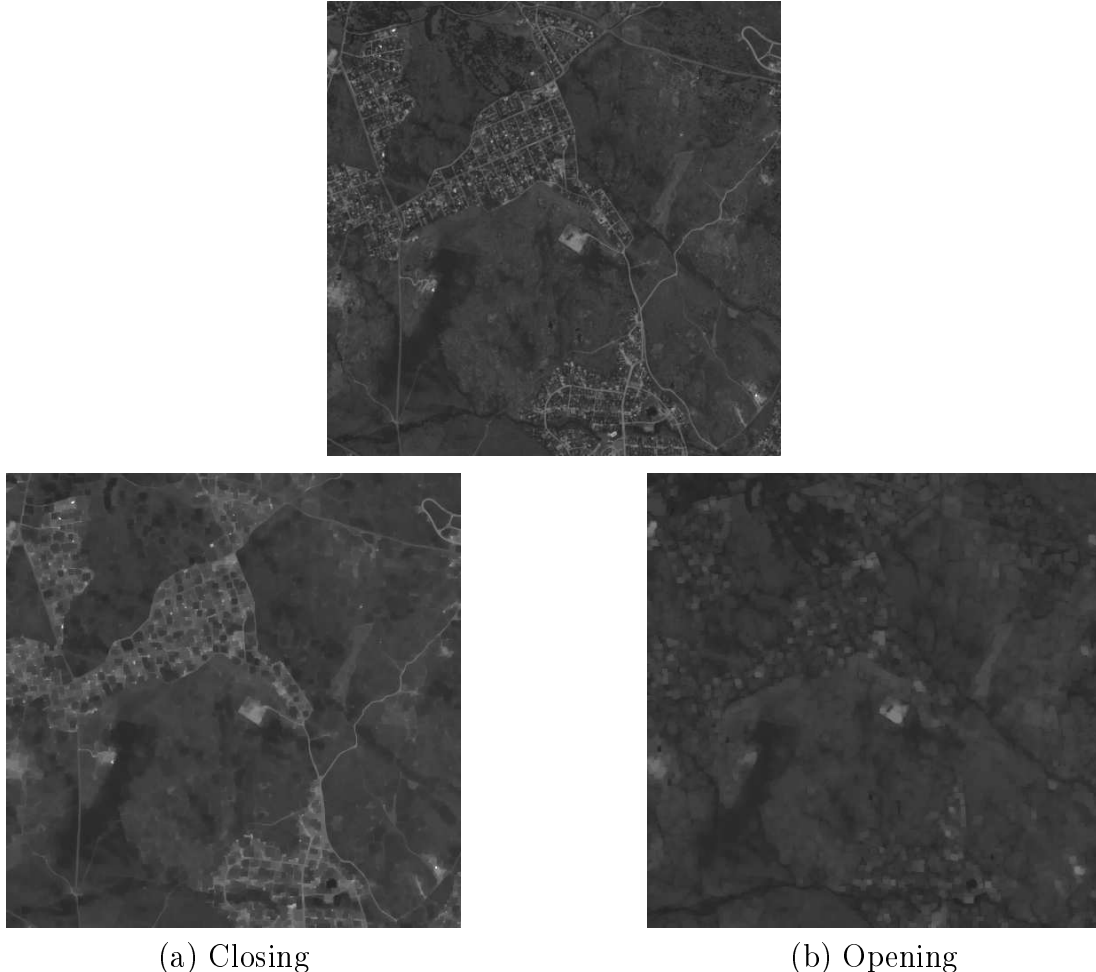


Figure 4.5: Examples of opening and closing operators.

A difference is computed between morphologically closed (Figure 4.5(a)) and opened images (Figure 4.5(b)). This difference gives prominence to textured regions, like urban areas. An example of an opening and a closing profile is shown in Figure 4.6. The difference image is then reduced to a lower resolution with pixel intensity averaged over a window of 4x4 pixels. Then an alternated sequential filter aggregates neighboring components and eliminates small isolated components. We compute four geometrical features from these regions as shown in Table 4.1.

### 4.2.1 Features from the urban regions

We focus on the last four features in Table 4.1. These features enable us to distinguish between rural classes (Villages and Fields) and the urban Europe class, which other-

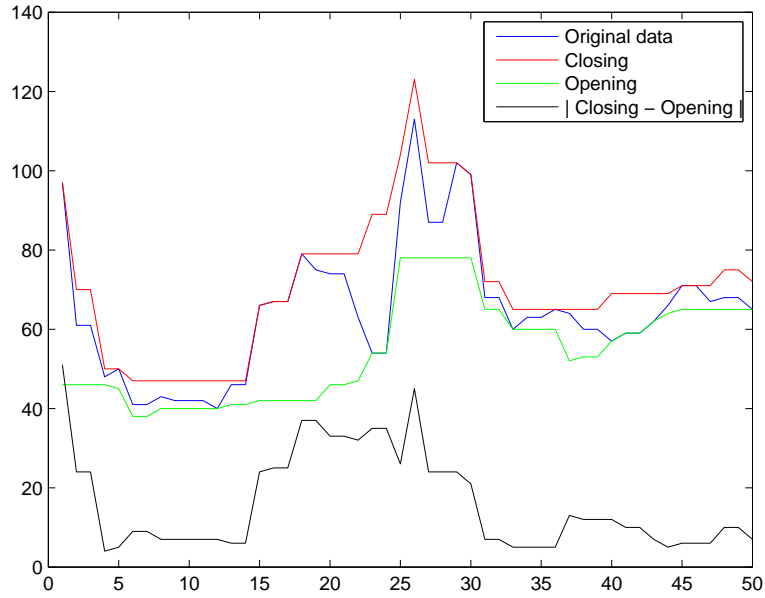


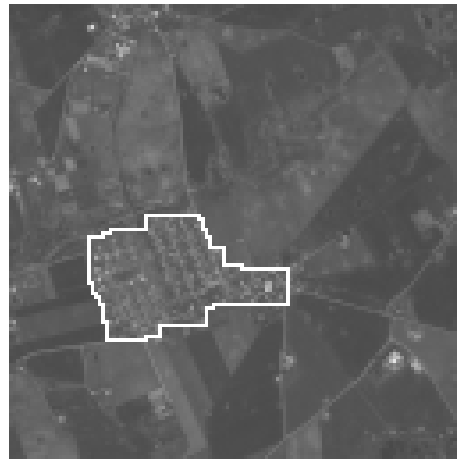
Figure 4.6: Examples of opening and closing operators ©CNES/ENST.

Notation	Description
$\Omega$	Area of image
$\Omega_R$	Area of extracted regions
$L_\Psi$	Network length in $\Psi = \Omega - \Omega_R$
$\Gamma_R$	Perimeter of extracted regions
$\tilde{R}_A$	<b>Region area density</b> $\Omega^{-1}\Omega_R$
$Cf_A$	<b>Region compactness factor</b> $\Omega_R^{-1}\Gamma_R^2$
$\tilde{R}_\nu$	<b>Region labels</b> $\#R$
$\Delta_\Omega$	<b>Inverse fractional length density</b> $\frac{\Omega_R}{L_\Psi}$

Table 4.1: Summary of features computed for urban areas.



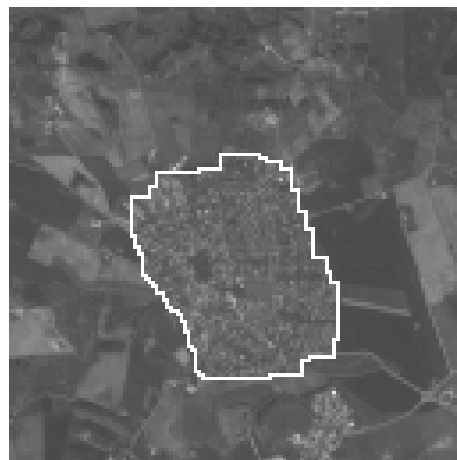
(a) Original image ©CNES



(b) Segmented region



(c) Original image ©CNES



(d) Segmented region

Figure 4.7: Images containing small urban areas and their segmentations.



wise were misclassified due to the lack of extracted network information from the small compact urban regions in the images, shown in Figure 4.7(a) and Figure 4.7(c). Let  $\Omega$  and  $\Omega_R$  be the area of the image and the area of the extracted regions respectively and  $L_\Psi$  and  $\Gamma_R$  be the network length in  $\Psi = \Omega - \Omega_R$  and the perimeter of the extracted regions respectively. We define two descriptors,  $\tilde{R}_A = \Omega^{-1}\Omega_R$ , the extracted region density and  $Cf_A = \Omega_R^{-1}\Gamma_R^2$ , the extracted region compactness factor. These two features help us to distinguish the Villages class from the rest of the classes: for example,  $\tilde{R}_A \simeq 1$  for Urban classes and  $\tilde{R}_A \simeq 0$  for Mountains and Fields classes. The feature  $\tilde{R}_\nu$ , the number of urban regions in an image is used to distinguish between complete Urban, Villages, Fields and Mountains. A complete Urban (USA and Europe) will have  $\tilde{R}_\nu = 1$ , whereas, a Villages will have  $\tilde{R}_\nu > 1$  and Fields and Mountains will have  $\tilde{R}_\nu = 0$ . Another feature  $\Delta_\Omega$ , the inverse fractional length density, is also computed to separate the Village class from Urban and Mountains and Fields. The total network length in  $\Psi = \Omega - \Omega_R$  is denoted by  $L_\Psi$ . For complete Urban classes (USA and Europe),  $L_\Psi = 0$  and for Mountains and Field classes  $\Psi = \Omega$ . Hence for Mountains and Fields classes, the inverse fractional length density,  $\Delta_\Omega = 0$ , while for complete Urban classes,  $\Delta_\Omega = \infty$ , and for the Village class  $0 < \Delta_\Omega < \infty$ . We augment these urban region features with the features computed from the graph representation of the road network as described in Chapter 3 to improve the classification of the geographical environments which otherwise were misclassified due to the loss of information from small dense urban regions.

## 4.2.2 Classification

All the images in our database have the same resolution. The features described in the above sections were computed for a database of 497 SPOT5, 5m resolution images. Machine classification was done with a five-fold cross validation on the data set, with 80% of data for training and the remaining 20% for testing in each fold. The features computed from the urban areas were augmented with the features computed from the road networks, as described in Chapter 3. In this study, 36 features (16 features each from the graph for 2 network extraction methods and 4 features from the urban region) are used for the classification of the same 7 classes mentioned in Chapter 3.

The results of the one-vs-rest SVM linear kernel classification of 497 images into 7 classes, using 36 features, is shown in Table 4.2. As can be seen, the mean classification error is drastically reduced from 24.7%, with only road network features to 16.5%, with the combined feature set. This is due to the fact that the of information from the urban areas is well captured with the geometrical and topological features described in Section 4.2.1.

With such a large number of features, and with some similarity between different features, it seems likely that there is some redundancy in the feature space. This redundancy can be reduced by feature selection. In the next classification experiment,

we performed feature selection using a FLD analysis, followed by SVM linear kernel classification on the selected feature set. Figure 4.8 shows the number of features selected from each feature group. The results of the classification are shown in Table 4.3. The SVM linear kernel classification on the 20-dimensional feature space selected by the FLD shows a mean error of 12.9% with a standard deviation of 3.29%. The dependence of classification error on the number of features is shown in Figure 4.9. Overall classification results with all features and selection is shown in Table 4.4.

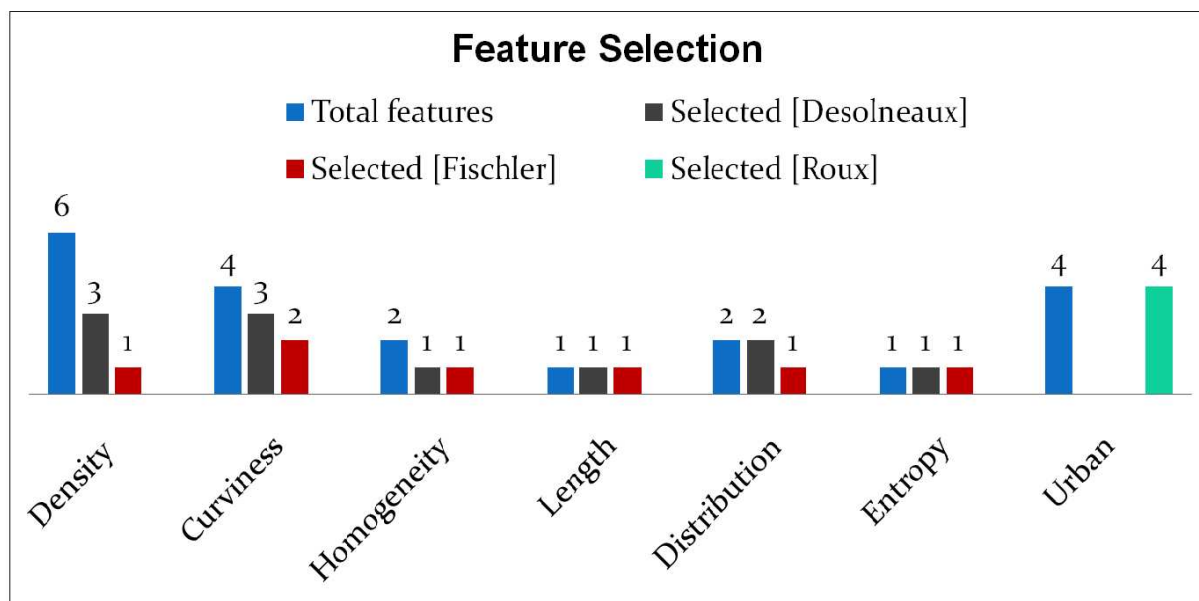


Figure 4.8: Feature selected from the feature group.

As can be clearly seen in Table 4.3, there is an improvement in the confusion matrix. The Villages class is less confused with the Fields and the Europe class than before. The new features augmented with the road network features provide good discriminative power to classify images which otherwise were misclassified.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.677</b>	0.022	0.119	0.015	0.051	0.015	0.044
Mountains	0.038	<b>0.888</b>	0.034	0.012	0.022	0.000	0.000
Fields	0.145	0.022	<b>0.817</b>	0.000	0.000	0.000	0.015
USA	0.032	0.000	0.000	<b>0.902</b>	0.193	0.000	0.010
Europe	0.085	0.067	0.029	0.054	<b>0.734</b>	0.030	0.029
Airports	0.022	0.000	0.000	0.015	0.000	<b>0.953</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.902</b>

Table 4.2: Confusion matrix of a SVM linear kernel classification on 497 images with 7 classes with 36 features.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.832</b>	0.000	0.148	0.000	0.049	0.017	0.033
Mountains	0.038	<b>0.832</b>	0.008	0.000	0.000	0.000	0.000
Fields	0.038	0.077	<b>0.823</b>	0.013	0.000	0.000	0.010
USA	0.010	0.000	0.000	<b>0.920</b>	0.116	0.016	0.010
Europe	0.083	0.041	0.021	0.067	<b>0.835</b>	0.015	0.020
Airports	0.000	0.049	0.000	0.000	0.000	<b>0.951</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.928</b>

Table 4.3: Confusion matrix of a SVM linear kernel classification on 497 images with 7 classes with 20 out of 36 features selected by FLD.

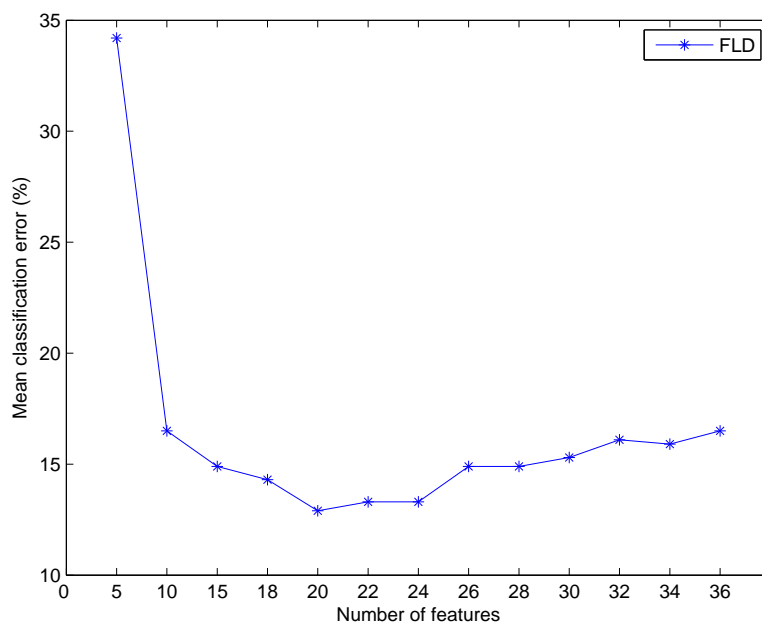


Figure 4.9: Classification error vs number of features

Feature Dimension	Selection	Classification Error (%)
36, 7 classes	No selection	Linear SVM, $16.5 \pm 2.64$
36, 7 classes	Fisher LD, 20	Linear SVM, $12.9 \pm 3.29$

Table 4.4: Classification performance.

### 4.2.3 2-Level classification

In order to have a better understanding of the classes, we divided our class representation into a 2-level hierarchical model as shown in Figure 4.10. The 1-level consists of three classes: "Urban", "Semi-Urban, and "Non-Urban". Classification with the same set of features was done on the 1-level. Depending upon this classification, we wanted to move to the 2-level for a finer classification. The results of the 1-level classification are shown in Table 4.5.

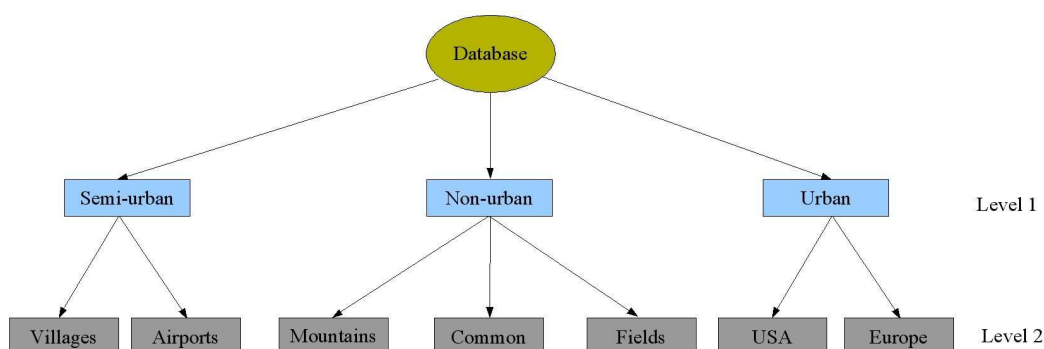


Figure 4.10: 2-level hierarchical model

	Class 1	Class 2	Class 3
Semi-urban	<b>0.879</b>	0.063	0.049
Non-urban	0.069	<b>0.913</b>	0.000
Urban	0.051	0.024	<b>0.951</b>

Table 4.5: Confusion matrix of an SVM linear kernel classification on 497 images with 3 classes with 22 out of 36 features selected by FLD.

As can be seen from the confusion matrix, the "Semi-urban" and "Non-urban" classes are classified with 87.9% and 91.3% respectively with 22 features selected with FLD, Figure 4.11. With this classification rate it is rather difficult to move on to the next level, as the error will grow multiplicatively in the subsequent process. This hinders us from proceeding any further with the 2-level classification. We therefore restrict ourselves to a 1-level classification with 7 classes.

### 4.2.4 Classification and retrieval with relevance feedback

Relevance feedback is used in the retrieval process by a user annotating images as relevant or irrelevant based on his/her judgement. This process can be seen as the training stage of a supervised classification. In many cases the importance of an image

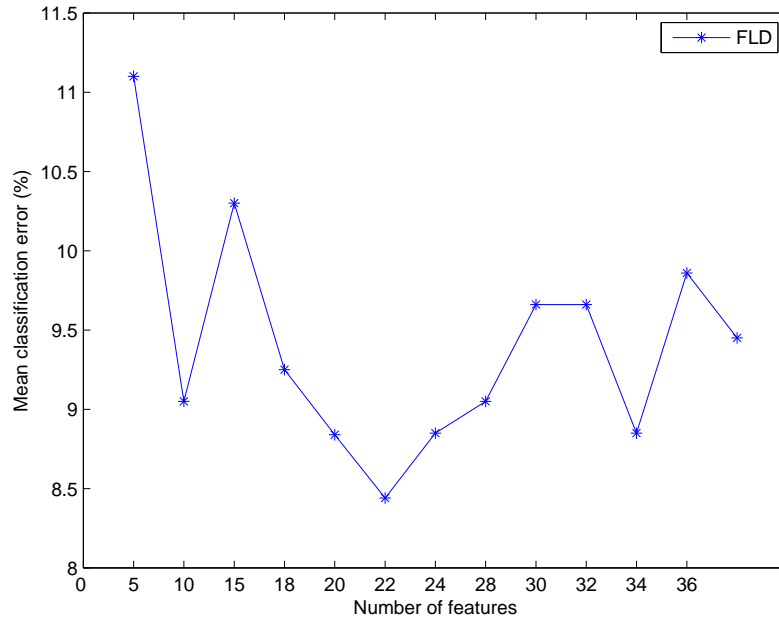


Figure 4.11: Classification error vs number of features

for the interest of a user is not taken into consideration, hence affecting the retrieval process. Several methods have been proposed to indicate the importance of an image in the RF protocol. A important drawback of most RF systems is the loss of memory of the annotated examples at the end of a search session. This is costly in terms of time when dealing with a large database. Many different methods have been adopted to create RF systems with memory.

In this work, we will test our features with the existing system of [Cost 06], which uses an SVM classification process to compute the separating surface between the relevant and irrelevant examples. In practice, a number of RF steps are required to attain a good level of retrieval. This number is significantly reduced by adding a memory which takes into account the results of past queries. Table 4.6 shows the results with 8 RF steps without memory and 1 RF step with memory. As can be seen, the results are comparable, with a slight change in the case with memory. This can be explained by the fact that the SV set chosen during classification is not perfect.

Feature set	Classification method	RF steps	Memory	Retrieval (%)
36 features/7 classes	SVM	8	No	93.1
36 features/7 classes	SVM	1	Yes	92.3

Table 4.6: Classification with relevance feedback and memory.

SVM has already been proved to be an efficient classification tool in CBIR systems, but this is not enough, the features must also have good discriminative power for

efficient classification. This test establishes the fact that the feature set used in this study has good discriminative power to classify different geographical environments with only few RF steps.

### 4.3 Discussion

The road extraction methods used in our study are robust, but they often fail to extract the narrow and finely structured road networks which are almost hidden in small urban areas. This failure of the extraction methods and hence the features computed from road networks results in poor classification of images containing such areas. In order to obtain some meaningful information from these regions, we need to segment such areas when they occur in images. Urban structures in low radiometric and metre resolution images show textural effects for objects size of few pixels. Morphological methods used to extract urban regions and the features computed on them provide useful information for characterizing geographical environments. The urban region statistics augmented with the road network statistics provide a meaningful classification. A retrieval test using the feature set confirms its discriminative power for classifying different geographical environments. Based on this, more statistics could be computed from the shape analysis of urban areas, e.g., moments, Fourier analysis of the urban contours, etc. to incorporate more information in the classification process.

---



## Chapter 5

# Dependence of primitive features on image resolution and size

The classification studies reported in Chapter 3 show some confusion between classes with the primitive features extracted from the graph representation of road networks. These confusions arise because the road extraction methods fail to detect the fine and densely structured roads present in some regions. The loss of information from these regions reduces the discriminative power of the features and hence results in poor classification. In the subsequent chapter, these regions were detected in the images. A new set of geometrical and topological features were then computed from these areas. A new classification test was done with the road network features augmented with the new features computed from these areas with the same set of classes. The result shows improvement in distinguishing the classes.

So far, we have not explored the impact of the spatial resolution and size of images on the feature set. The question which still remains to be answered is, what is the optimal resolution and size of an image for the classification of geographical regions on the Earth's surface? In this Chapter, we take a closer look at the effect of the spatial resolution and size of images on the discriminative power of the feature set. We perform a one-vs-rest linear SVM classification on two image databases with the same set of features used in Chapter 3 and Chapter 4.

### 5.1 Introduction

With the advent of modern sensor technology, the increase in the resolution and size of the images acquired from satellites has exposed a major challenge in information extraction and understanding. High resolution satellites, namely QuickBird2 with sub-metre panchromatic images and the future Pleiades and Ikonos with a spatial resolution

---



of 1m provide us with images that contain adequate information on the attributes of the ground structure of an urban area. In order to extract this information from a range of multiresolution images, however one single method will not suffice.

Several studies have been done to assess the optimal spatial resolution for classification. Some objects are better classified at finer resolution while others require coarser resolutions. A study reported in [Marc 94] explains the effect of resolution on classification. Work has also been done on multiresolution classification. Features computed at different spatial resolutions are integrated to improve land use/cover classification [Chen 02].

In our work, the network extraction method suitably tuned for a 5m resolution image will extract redundant structures in an image of the same ground area with 1m resolution, whereas it will extract insufficient structures in an image with 10m resolution. The extraction parameters are fixed for a certain resolution and cannot be dynamically adapted for varying resolutions.

Image size also plays an important role in classification. For instance, a small image (e.g., 64x64 pixels) will be devoid of the completeness of a road network structure (e.g., highways and national routes) whereas a large image (e.g., 1024x1024) may contain road networks from a region containing urban structures, mountains and fields. In these two cases it is always difficult to assign a specific class to an image. The power of the feature set to classify geographical environments is also dependent on an 'optimal' image size.

In this chapter, we will study the effect of the size and the resolution of images on the features extracted by a single method, originally developed for a spatial resolution of 5m, SPOT5 images. The studies reported in Chapter 3 and Chapter 4 use a database of these images of size 512x512 pixels.

## 5.2 Road network and urban region extraction for different resolutions

In this section we show the network extracted from the complete urban structure of Toulouse, France ©CNES, with the method of [Deso 00] at different resolutions. Figure 5.1 shows the extraction results. The method extracts almost all linear structures (e.g., Building edges) from a 1m resolution image, while at the other extreme, it merely detects any linear structure from an 8m resolution image. The images with resolutions 3.175m, 4m and 5m show an optimal number of correctly extracted road network segments. The urban area extraction method [Roux 92] discussed in the previous chapter also extracts redundant areas in the 1m and 2m resolution images as the morphological operations fail to discriminate the texture information of the 'true' urban areas from the surroundings. Figure 5.2 shows the urban area extraction results.

---

A good extraction of urban regions can be seen in the images with 3.175m, 4m and 5m resolutions.

In our previous studies we have particularly used a database of SPOT5 images with 5m resolution provided by CNES. The classification of images with all the mentioned resolutions is out of scope of this work. The number of images at our disposal with all the mentioned resolutions is not adequate to form an acceptable database for classification. In order to have a glimpse at the size and resolution dependence of the features for classification of images belonging to the classes mentioned in Chapter 3, we construct two databases.

One database corresponds to the same set of images studied in Chapter 3, but with images downsampled to 10m resolution. The other database is constructed from the original database studied in Chapter 3, by cutting each image of size 512x512 pixels into four equal parts, each of size 256x256 pixels. This database is now four times larger than the original database. The images were then rearranged according to the pre-defined classes for further study.

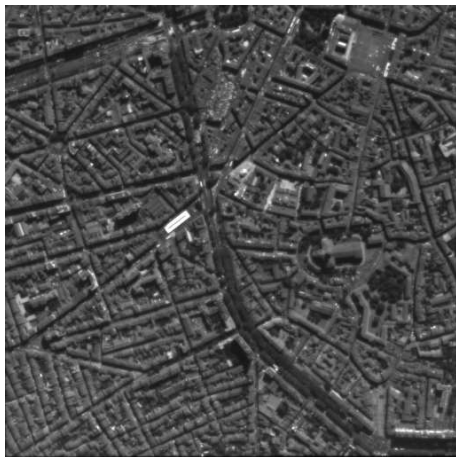
---



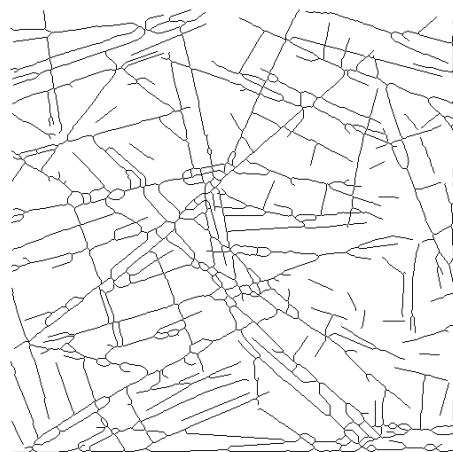
(a) Original image, 1m ©CNES



(b) Extracted network



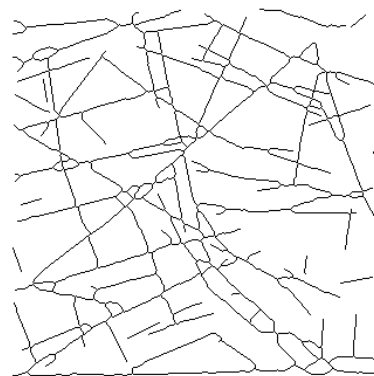
(c) Original image, 2m ©CNES



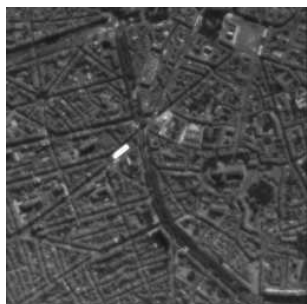
(d) Extracted network



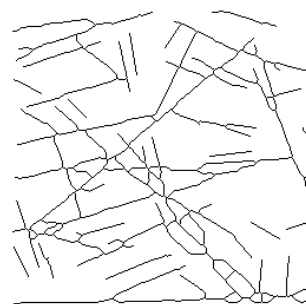
(e) Original image, 3.175 ©CNES



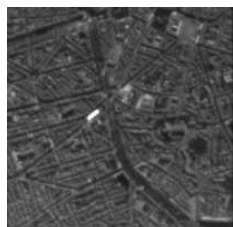
(f) Extracted network



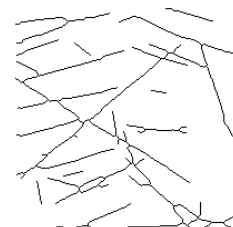
(g) Original image, 4m ©CNES



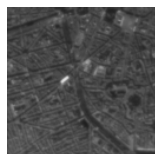
(h) Extracted network



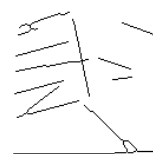
(i) Original image, 5m ©CNES



(j) Extracted network



(k) Original image, 8m ©CNES



(l) Extracted network

Figure 5.1: Images and network extracted from them at different resolutions. The method of [Deso 00] was used as an extraction method.



(a) Original image, 1m ©CNES



(b) Extracted region



(c) Original image, 2m ©CNES



(d) Extracted region

---



(e) Original image, 3.175 ©CNES



(f) Extracted region



(g) Original image, 4m ©CNES



(h) Extracted region



(i) Original image, 5m ©CNES



(j) Extracted region



(k) Original image, 8m ©CNES



(l) Extracted region

Figure 5.2: Images containing small urban areas and their segmentations at different resolutions. The method of [Roux 92] was used for the segmentation.

### 5.3 Classification

The features from the graph representation of the network and the urban area were computed as mentioned in Chapter 3 and Chapter 4. Three tests were done on databases as follows:

- 497 SPOT5 images of 5m resolution with each image of size 512x512 pixels.
- 497 SPOT5 images of 10m resolution with each image of size 256x256 pixels. This image database were constructed by downsampling the images from the above mentioned images.
- 1988 SPOT5 images of 5m resolution with each image of size 256x256 pixels. This image database were constructed by relabelling the 4 images obtained by cutting each image of size 512x512 pixels.

Machine classification was done with a five-fold cross validation on the data set, with 80% of data for training and the remaining 20% for testing in each fold.

The results of one-vs-rest linear SVM classification of 497/5m images into 7 classes with 20 features selected from 36 features by FLD is shown in Table 5.1. The mean classification error is 12.9% with a standard deviation of 3.29%.

The results of one-vs-rest linear SVM classification of 1988/5m images into 7 classes with 22 features selected from 36 features by FLD is shown in Table 5.2. The mean classification error is 17.4% with a standard deviation of 2.19%. Figure 5.3 shows the number of features selected from each feature group.

The results of one-vs-rest linear SVM classification of 497/10m images into 7 classes with 22 features selected from 36 features by FLD is shown in Table 5.3. The mean classification error is 25.4% with a standard deviation of 2.02%. Figure 5.4 shows the number of features selected from each feature group.

Figure 5.5 and Table 5.5 show an overall classification result with all the features and the selected set of features respectively.

As expected, the classifications varies dramatically with changes in image size and resolution. The few anomalies that can be seen in Table 5.2 could be explained as follows: changing the image size from 512x512 pixels to 256x256 pixels, causes a poorer classification of Villages, Mountains, Fields, USA and Europe than in Table 5.1, as too little information is available from the road network features to discriminate the classes. On the other hand, the Airports class is well classified because a small area captures the airport runways better than in the images with 512x512 pixels, where non-runway areas from the surroundings lower the discriminative capacity of the features.

---

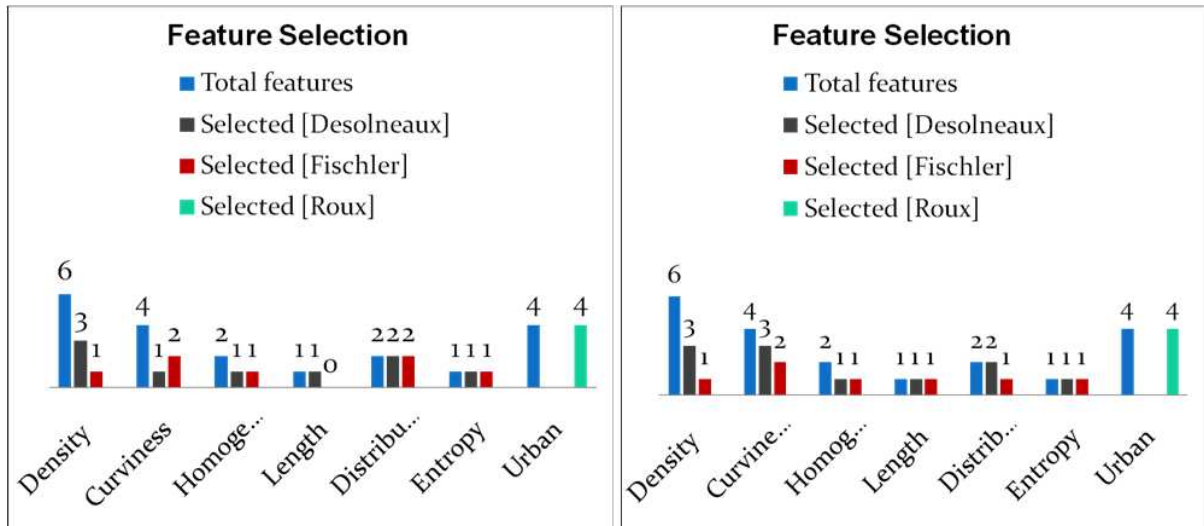


Figure 5.3: Feature selected from the feature group. Left: Feature selected from 36 features from 497 images with 5m resolution. Right: Feature selected from 36 features from 1988 images with 5m resolution.

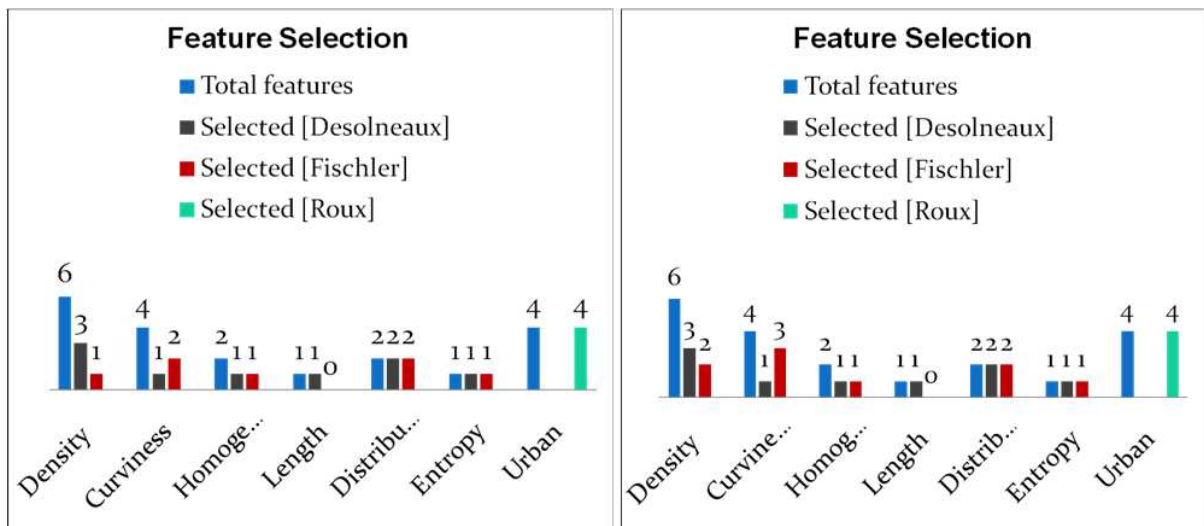


Figure 5.4: Feature selected from the feature group. Left: Feature selected from 36 features from 497 images with 5m resolution. Right: Feature selected from 36 features from 497 images with 10m resolution.



A few anomalies can also be seen in the classification of images with 10m resolution. The poorer classification results in Table 5.3 compared to Table 5.1 could be explained as follows: the network extraction methods fail to provide reliable information from the road network structures at this resolution with the method of [Deso 00] used in Chapter 3. The urban extraction method of [Roux 92] possibly fails to extract any urban area at this resolution. All the classes show poorer classification except for the Common (sea) class. This could be explained by the fact that, at this resolution, the linear structures of waves in the sea could not be extracted with the method of [Deso 00], as opposed to images with 5m resolution.

The overall classification performance results are shown in Table 5.5. As can be seen, the mean classification error is doubled for images of 10m resolution compared to images of 5m resolution. This fact cannot be generalized to all resolutions. The number of images at our disposal with different resolutions is not enough to form databases for a comparative analysis.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.832</b>	0.000	0.148	0.000	0.049	0.017	0.033
Mountains	0.038	<b>0.832</b>	0.008	0.000	0.000	0.000	0.000
Fields	0.038	0.077	<b>0.823</b>	0.013	0.000	0.000	0.010
USA	0.010	0.000	0.000	<b>0.920</b>	0.116	0.016	0.010
Europe	0.083	0.041	0.021	0.067	<b>0.835</b>	0.015	0.020
Airports	0.000	0.049	0.000	0.000	0.000	<b>0.951</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.928</b>

Table 5.1: Confusion matrix of a SVM linear kernel classification on 497 images with 7 classes with 20 out of 36 features selected by FLD. The original database, with images of resolution 5m.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.658</b>	0.059	0.195	0.019	0.153	0.000	0.000
Mountains	0.018	<b>0.823</b>	0.020	0.002	0.000	0.000	0.009
Fields	0.019	0.053	<b>0.740</b>	0.000	0.004	0.000	0.000
USA	0.167	0.000	0.002	<b>0.793</b>	0.129	0.000	0.000
Europe	0.116	0.060	0.043	0.185	<b>0.714</b>	0.000	0.000
Airports	0.021	0.004	0.000	0.000	0.000	<b>1.000</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.991</b>

Table 5.2: Confusion matrix of a SVM linear kernel classification on 1988 images with 7 classes with 22 out of 36 features selected by FLD. The database of 1988 images of size 256x256 pixels.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.530</b>	0.117	0.320	0.026	0.031	0.014	0.000
Mountains	0.082	<b>0.649</b>	0.032	0.019	0.014	0.040	0.000
Fields	0.187	0.081	<b>0.585</b>	0.000	0.000	0.000	0.000
USA	0.155	0.000	0.015	<b>0.820</b>	0.163	0.029	0.000
Europe	0.033	0.101	0.029	0.136	<b>0.792</b>	0.014	0.000
Airports	0.012	0.051	0.017	0.000	0.000	<b>0.902</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>1.000</b>

Table 5.3: Confusion matrix of a SVM linear kernel classification on 497 images with 7 classes with 22 out of 36 features selected by FLD. The images were downsampled to 10m resolution.

	497/5m/512x512	1988/5m/256x256	497/10m/256x256
Villages	<b>0.832</b>	<b>0.658</b>	<b>0.530</b>
Mountains	<b>0.832</b>	<b>0.823</b>	<b>0.649</b>
Fields	<b>0.823</b>	<b>0.740</b>	<b>0.585</b>
USA	<b>0.920</b>	<b>0.793</b>	<b>0.820</b>
Europe	<b>0.835</b>	<b>0.714</b>	<b>0.792</b>
Airports	<b>0.951</b>	<b>1.000</b>	<b>0.902</b>
Common	<b>0.928</b>	<b>0.991</b>	<b>1.000</b>

Table 5.4: An overall comparison table of "good" classified classes.

Database/Resolution	Feature Dimension	Selection	Classification Error (%)
497/5m	36,7 classes	Fisher LD, 20	Linear SVM, $12.9 \pm 3.29$
1988/5m	36,7 classes	Fisher LD, 22	Linear SVM, $17.4 \pm 2.19$
497/10m	36,7 classes	Fisher LD, 22	Linear SVM, $25.4 \pm 2.02$

Table 5.5: Classification performance.

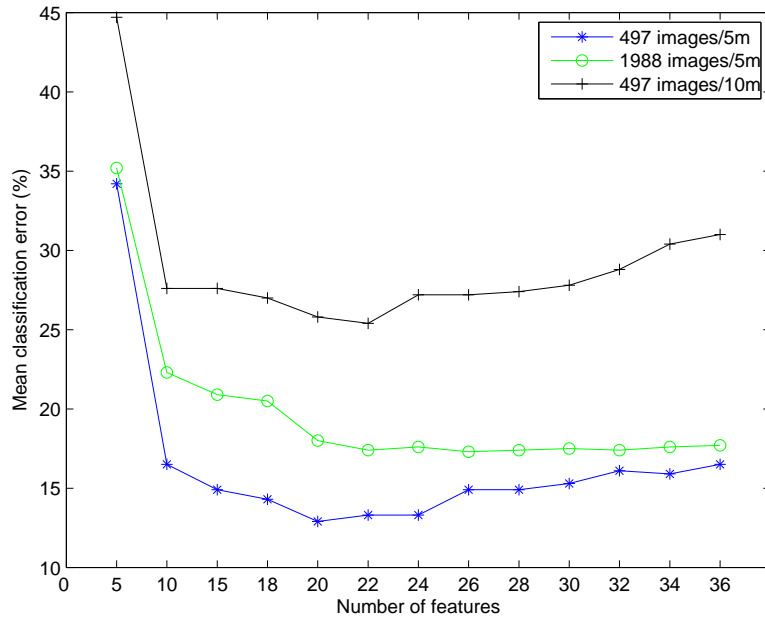


Figure 5.5: Classification errors vs number of features

## 5.4 Discussion

In this study we have shown the dependence on image size and resolution of the feature set computed from a single extraction method developed for a particular resolution. Features computed from a lower resolution show a greater amount of confusion between classes. This confusion is also visible with images of smaller sizes. However, a few anomalies can be explained on the basis of the road network and urban area extraction methods. Based on this study, more detailed analysis could be done on image databases of different resolutions. The search for an optimal resolution for the categorization of man-made structures on the Earth's surface could not be performed based on this study as the number images at our disposal with different resolutions is not enough to form databases for a comparative analysis. The image size factor, which is also crucial for delineating class boundaries, remains to be solved. This study provides a base for answering such queries in the future.

## Chapter 6

# Indexing of large satellite images

Satellite image classification has been a major research field for many years with its varied applications in the fields of geography, geology, archaeology, environmental sciences and for military purposes. Many different techniques have been proposed, including texture analysis, stochastic methods, genetic algorithms, fuzzy theory and neural networks. Work on satellite image classification as reported in the last twenty years has shown a significant increase in classification accuracy due to new approaches and new image acquisition technology. There has been a large amount of work dedicated to the classification of large satellite images at pixel level rather than by considering image patches of different sizes. Classification of image patches of different sizes from a large satellite image is a novel idea in the sense that the patches considered contain significant coverage of a particular type of geographical environment. In this Chapter, we will show how to classify large satellite images with geometrical and topological features computed on small image patches extracted from it.

A one-vs-rest Gaussian kernel SVM classification method is used to classify the large image shown in Figure 6.2. The classification labels the image patches with the label having the maximum geographical coverage of the associated area in the large image. The large image will be mapped into a "region matrix", where each element of the matrix corresponds to a geographical class. In certain cases, this provides some anomalies, as a single patch may contain two or more different geographical environments. In order to have an estimate of these partial coverages, the output of the SVM is mapped into probabilities. These probability measures are then studied to have a closer look at the classification accuracies.

---

## 6.1 Introduction

Technologically advanced satellite sensors and new storage systems have made image data vast and complex. Complete manual annotation of a complex image is not feasible. In images, pixels provide the most basic level of information. Pixel values are the measurements by the satellite sensor of a region on the Earth's surface. The information from these pixels is at a level far below the semantic meaning of the desired object or region. The classification of images based on the pixel values is thus tedious and expensive. Although the inter-class boundaries in an image are at the pixel level, the classification of images at the pixel level is not an efficient browsing strategy. Class anomalies can also be found in classification using pixel values. Image processing and computer vision techniques are used to convert the pixel values to a meaningful representation corresponding to human perception of the entity. This high-level information can then be used for image indexing and retrieval.

Satellite images are geographical data and each pixel of the image corresponds to some region on the Earth's surface. The classification of large satellite images with patches of small images segmented from it is a better and a computationally less expensive method than the classical pixel-by-pixel approach. Moreover, classification of patches of images instead of pixels helps us to incorporate more freedom in complex querying. The reason is that the image patches from a large satellite image contain significant coverage of a particular type of environment, such as urban regions or mountains.

In the first part of this chapter we perform a hard classification, where each image patch is given a single label. In the subsequent part, we map the output of this classification into probabilities. The probability measures help us to understand better the geographical content of the large image.

The large image  $I$  is divided into equal sized non-overlapping images  $p_i$ ,  $i = 1 \dots N$ , e.g., for a large image of size 5120x5120 pixels/5m resolution with each image patch of size 512x512 pixels/5m resolution, we have  $N = 100$  image patches. For each patch  $p_i$  we compute the features from the graph representation of the extracted road network and the urban regions in it. The segmentation of the image patches  $p_i$  from the large image  $I$  along with the extraction and computation of the road network and urban region features are done off-line. The features are kept in a file with each image randomly labelled from 1 to 7 classes. This file is then used as a "testing" set for the classification of each image patch.

The large image is indexed as a "Region matrix", where each element of the matrix corresponds to a class label. The one-vs-rest Gaussian SVM classifier gives us an overall error of 7%. The misclassification of certain image patches is due to the feature properties pertaining to it. For example, an image patch from the large satellite image of Copenhagen is classified as an USA urban area. This is due to the fact that the road network structures in this image patch are similar to those found in most Amer-

---

ican urban regions. The modern urban structures ("grid") of many European cities are similar to conventional American urban structures. In other cases of misclassification, the image patches failed to find similar image patterns in the "training" dataset. Anomalies can also be seen in some image patches consisting of two or more geographical environments. Our hard classification method labels the image patch with the label having the maximum geographical coverage. This does not provide any measure of confidence of the labelling process. No measurement is inferred about the 'distance' to the labelled class or the proximity of any other class. In order to circumvent these drawbacks and allow us to take a definitive decision on the classification results, we map the output of the SVM into probabilities [Plat 99]. These probabilities provide us with the classification confidence of the labelled class and the proximity of any other classes.

## 6.2 Indexing

An image is indexed by a set of descriptors representing the content of the image. These descriptors, often termed indices, are usually limited in number and are dependent on application scenarios. Classification is often used as a pre-processing step for indexing. A careful indexing of an image database assists efficient retrieval of image content. The workflow of our indexing method is divided into three steps as follows:

### 6.2.1 Step 1: The database

The image database can be viewed as two sets disjointly partitioned to contain images or segmented images in one set and features extracted from images in another set. We will indicate the image set as  $S_I$ , and the feature set as  $S_F$ . A pointer is used between  $S_I$  and  $S_F$  to attach an image to its associated feature set. The information extracted in terms of geometrical and topological features from the large image archive of 497 images, each of size 512x512 pixels, categorized into 7 classes are kept in a data file. The off-line process of data file creation is done only once, and in the case of a new entry, the information extracted from this image is added to the existing data file. The pointer is appropriately assigned the address of this new entry. This will be used as the "training" set later in the classification task.

### 6.2.2 Step 2: The feature file

The off-line process for the user given a large image is as follows: the large image of size 5120x5120 pixels is automatically divided into non-overlapping image patches, (see Figure 6.2) each of size 512x512 pixels. During this process a pointer is associated with

---

each image patch, defining its spatial position in the large image. The road network extraction, its graph representation and the urban area segmentation methods are applied in parallel on the image set (100 images). The geometrical and topological features from the graph representation and the urban areas are stored in a file. The images are *a priori* randomly labelled with classes from 1 to 7. This will later be used as a "testing" set against the above defined "training" set in the classification task.

### 6.2.3 Step 3: The classification

The on-line classification process is done with a one-vs-rest Gaussian SVM classifier with  $\sigma = 10$ . The best choice of the Gaussian kernel variance  $\sigma^2$ , which controls the width of the kernel, is hard to assess in practical situations. In this study we used the value of  $\sigma$  that gave us the least training error. A result is shown in Table 6.1 of the one-vs-rest Gaussian SVM classification of 497 images into 7 classes, using 20 features selected by FLD from the set of 36 features (see Figure 6.1). The mean classification error is 10.5% with a standard deviation of 2.5%.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Villages	<b>0.867</b>	0.013	0.121	0.000	0.022	0.000	0.000
Mountains	0.014	<b>0.831</b>	0.008	0.000	0.014	0.000	0.000
Fields	0.022	0.065	<b>0.863</b>	0.000	0.014	0.000	0.000
USA	0.009	0.000	0.000	<b>0.933</b>	0.114	0.000	0.011
Europe	0.087	0.047	0.008	0.067	<b>0.835</b>	0.014	0.000
Airports	0.000	0.043	0.000	0.000	0.000	<b>0.986</b>	0.000
Common	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.989</b>

Table 6.1: Confusion matrix of an Gaussian SVM classification of 497 images with 7 classes with 20 out of 36 features selected by FLD. The original database, with images of resolution 5m.

The output of the classification is represented as a matrix ("Region Matrix"), where each element of the matrix corresponds to an image class. The classified images were compared with the ground truth image. The errors in the classification are shown in Figure 6.3, where the circles mark the misclassified images.

In much satellite image classification work, prior information about the class label configuration is available, and it is essential to include this information in the classification process to obtain a reliable answer. Standard SVM does not provide any estimation of the classification confidence and thus does not allow us to include prior information. Probabilistic SVM provides us with a way to construct a classifier that produces a posterior probability  $P(class = c|input)$  which allows us to take a quantitative decision about the classification. The probabilistic SVM has the additional

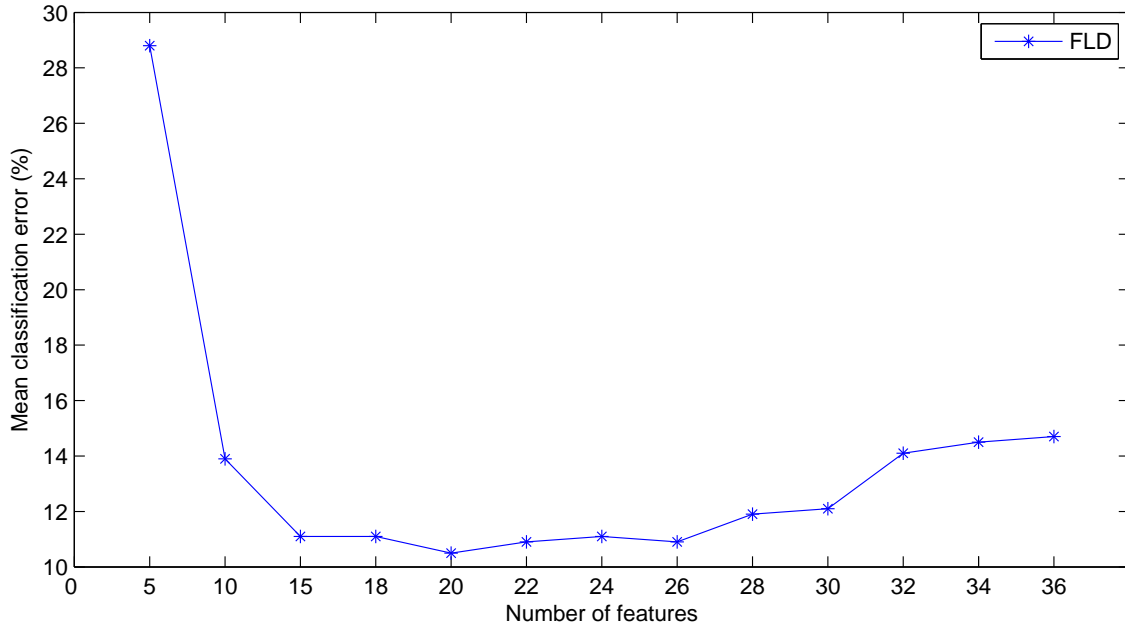


Figure 6.1: Classification error vs number of features.

advantage of being less dependent on the choice of the parameters.

The results of the probabilistic SVM output can be interpreted as follows: The classifier output should be a calibrated posterior probability. First the SVM is trained and then the parameters  $A$  and  $B$  of a sigmoid function (see Equation 6.1) are estimated from the training set  $(f_i, y_i)$  to map the output of the SVM into probabilities. The predicted label of an image is the one with the largest probability value. Figure 6.4 shows the results of an probabilistic SVM classification. In certain cases, there could be an ambiguity in the assignment of the class label, e.g., an image with probability 98% of being USA can be ambiguous, if the same image is classified with probability 88% to be Villages. In such cases it is difficult to assign a class to an image. A detailed analysis of four regions with classification probabilities is reported in subsection 6.2.4.

$$P(y = c|f) = \frac{1}{1 + \exp(Af + B)} \quad (6.1)$$

In order to avoid such uncertainties and correctly assign a label to an image, we consider neighborhood properties. For the above mentioned case, where the classification probabilities are comparable, we look at its 8-neighbors. We choose the label to assign to an image to be the one which appears most frequently amongst the neighbors. With the example shown in Figure 6.5, we can modify some of the misclassified labels from Figure 6.4, where the images were classified as USA urban due to the road network structures, to Figure 6.7 with the algorithm shown in Figure 6.6. The



changed labels are marked by circles. As can be clearly seen, there are a few more ambiguous labels which can be post-processed by the algorithm. Coastal regions are mostly misclassified due to the fact that the waves are extracted as linear structures thereby confusing them with road structure. These regions are then labelled as Fields or Mountains according to the structures extracted. The region matrix superimposed on the original image shown in Figure 6.8 provides us with a visual inspection of the overall classification results.

The method of changing the labels according to neighborhood properties is not implemented for all the other cases studied in this chapter.

---

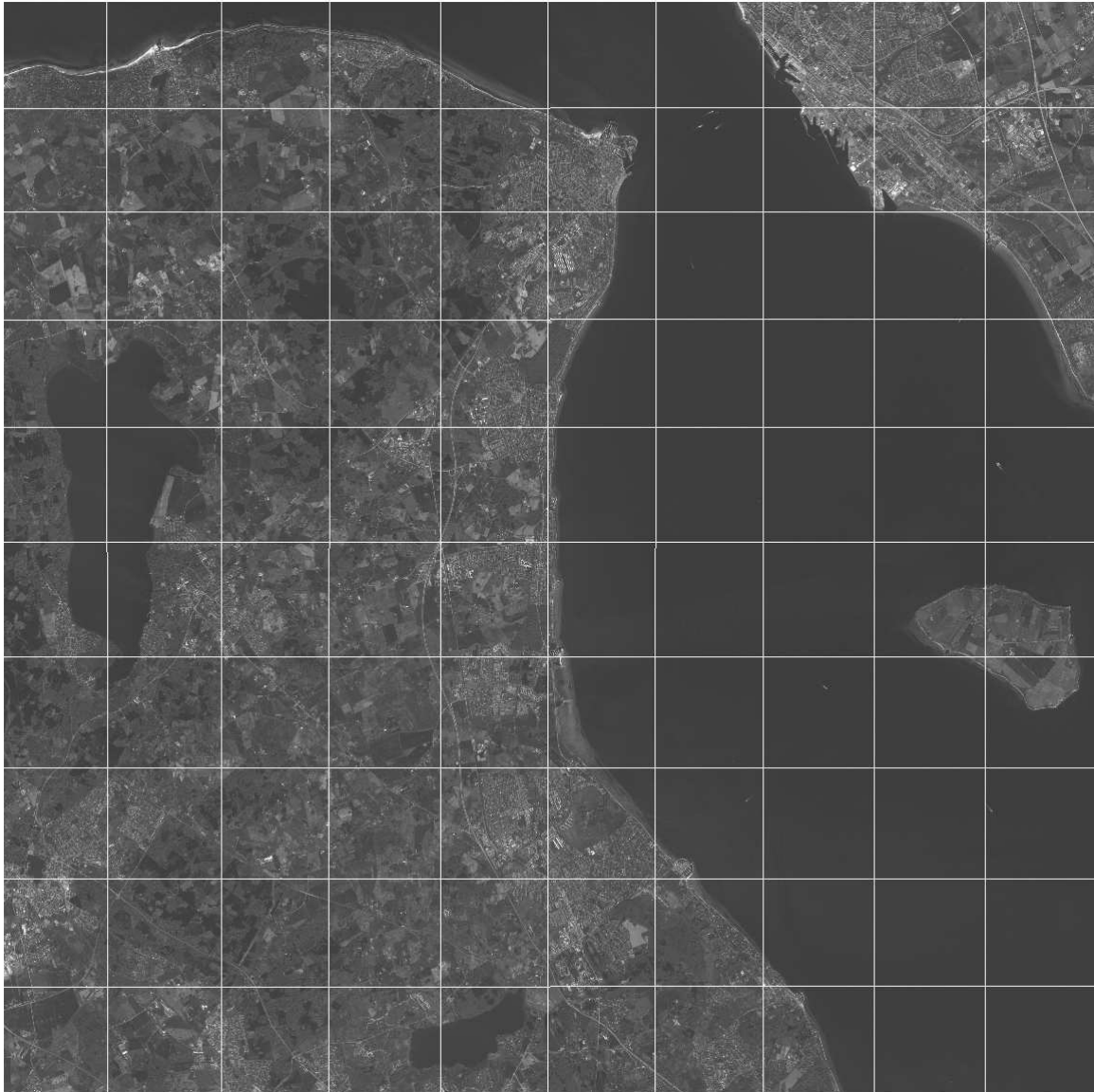


Figure 6.2: Image patches of size 512x512 pixels from a large SPOT5 image of Copenhagen with 5m resolution of size 5120x5120 pixels ©CNES.

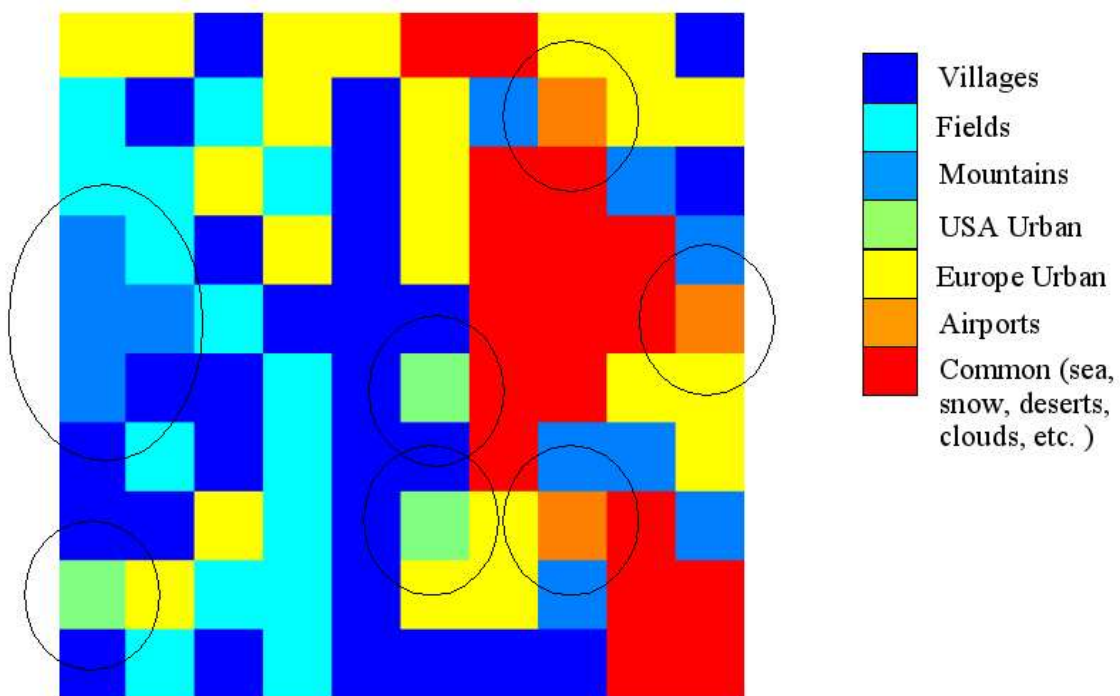


Figure 6.3: Classification of image patches of size 512x512 pixels from a large image of Copenhagen of size 5120x5120 pixels ©CNES. Each colored pixel of the 10x10 matrix corresponds to a class to which an image patch of size 512x512 belongs. SVM with Gaussian kernel of  $\sigma = 10$ .

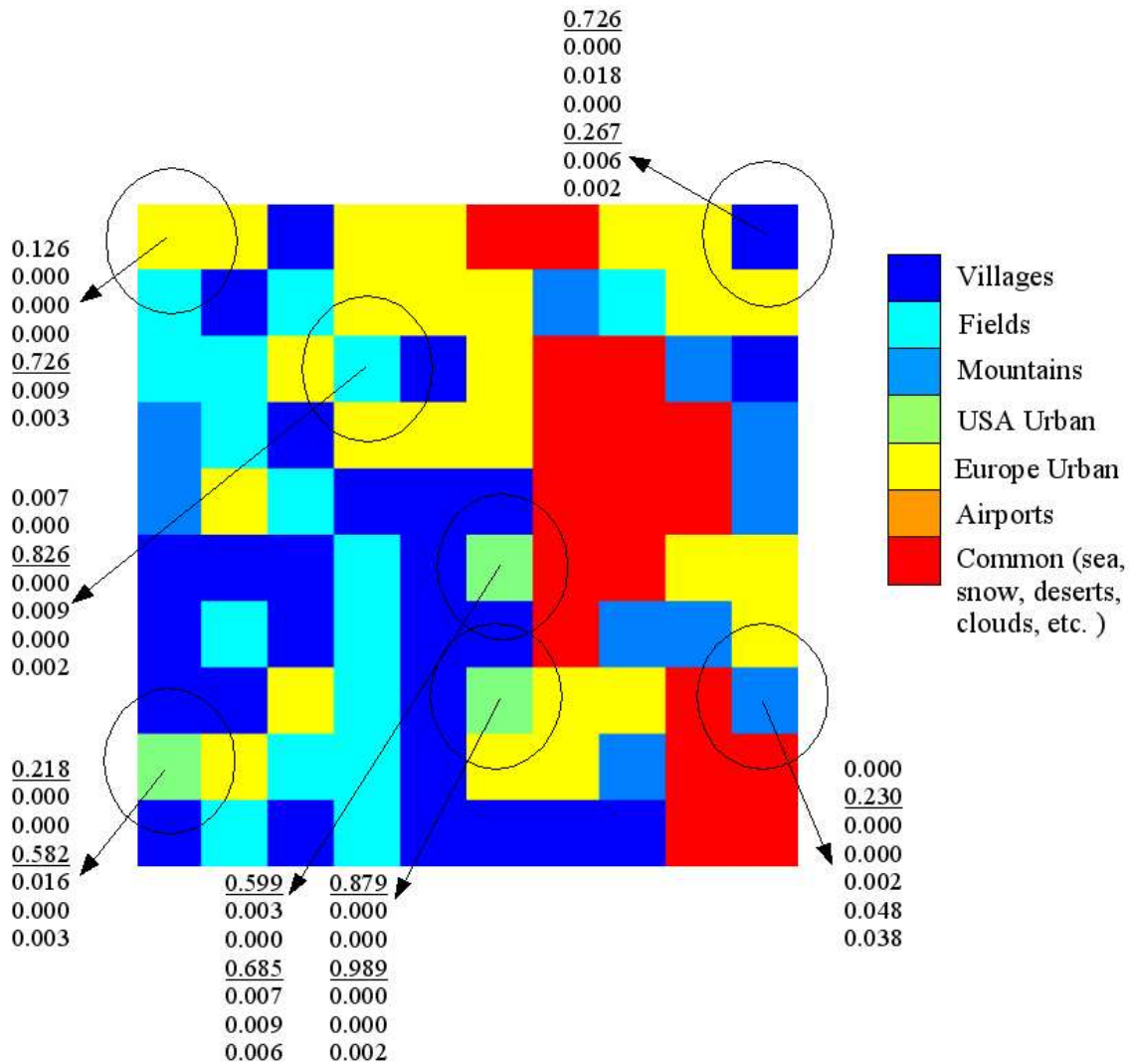


Figure 6.4: Classification of image patches of size 512x512 pixels from a large image of Copenhagen of size 5120x5120 pixels ©CNES. Each colored pixel of the 10x10 matrix corresponds to a class to which an image patch of size 512x512 belongs. A probabilistic SVM with Gaussian kernel of  $\sigma = 10$ . The values show the probabilities of each image patches belonging to a class.

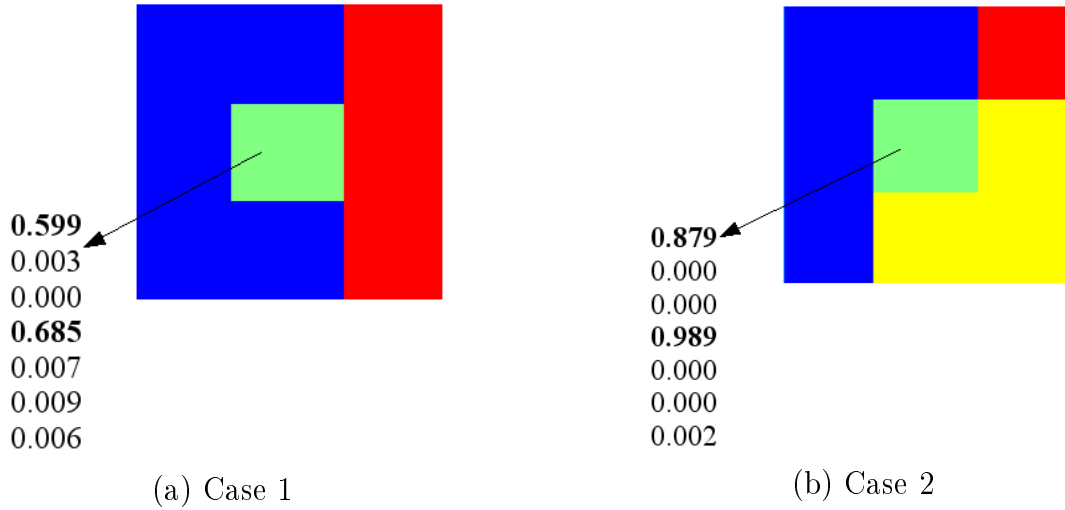


Figure 6.5: Examples of ambiguous cases in classifications.

**Input:** The probability vector  $[p]$  of an image

**Output:** The changed label of the image

$p_l = \max(p_i), i = 1 \dots 7$  ;

$L_0 = \text{Label}(p_l)$  ;

**foreach**  $i$  **do**

$z_i = \max(p_i) - p_i$ ;

**if**  $z_i \leq 0.5$  **then**

$L_i = \text{Label}(p_i)$ ;

$NL_0 = \text{Count 8-neighbors}(L_0)$ ;

$NL_i = \text{Count 8-neighbors}(L_i)$ ;

**if**  $NL_0 \geq NL_i$  **then**

| Put  $L_0$ ;

**else**

|  $L_0 = L_i$ ;

**end**

**end**

**end**

Figure 6.6: The algorithm for changing the labels.

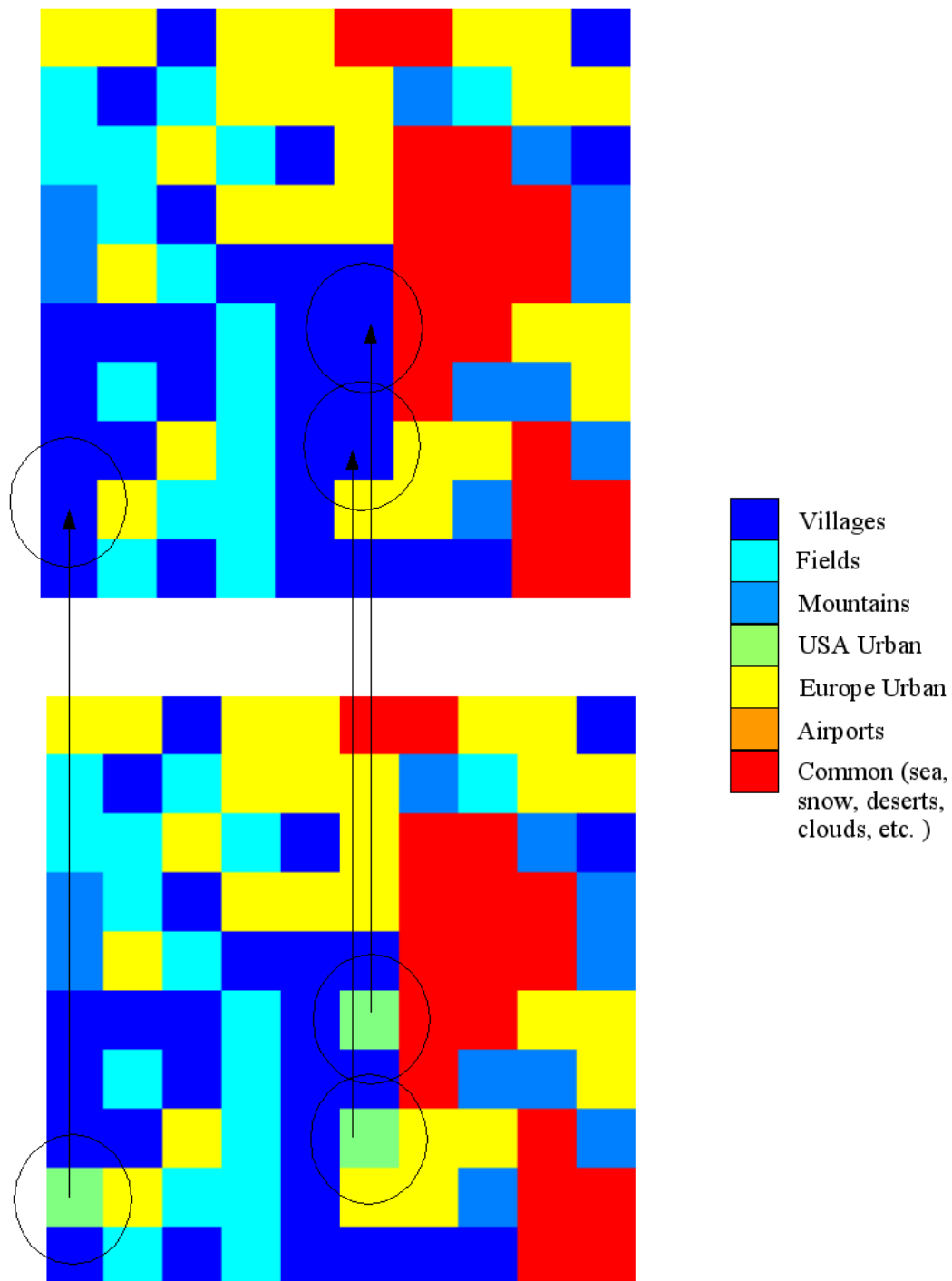


Figure 6.7: Classification modified using neighborhood properties. Top: The modified labels, Bottom: The probabilistic SVM classification.

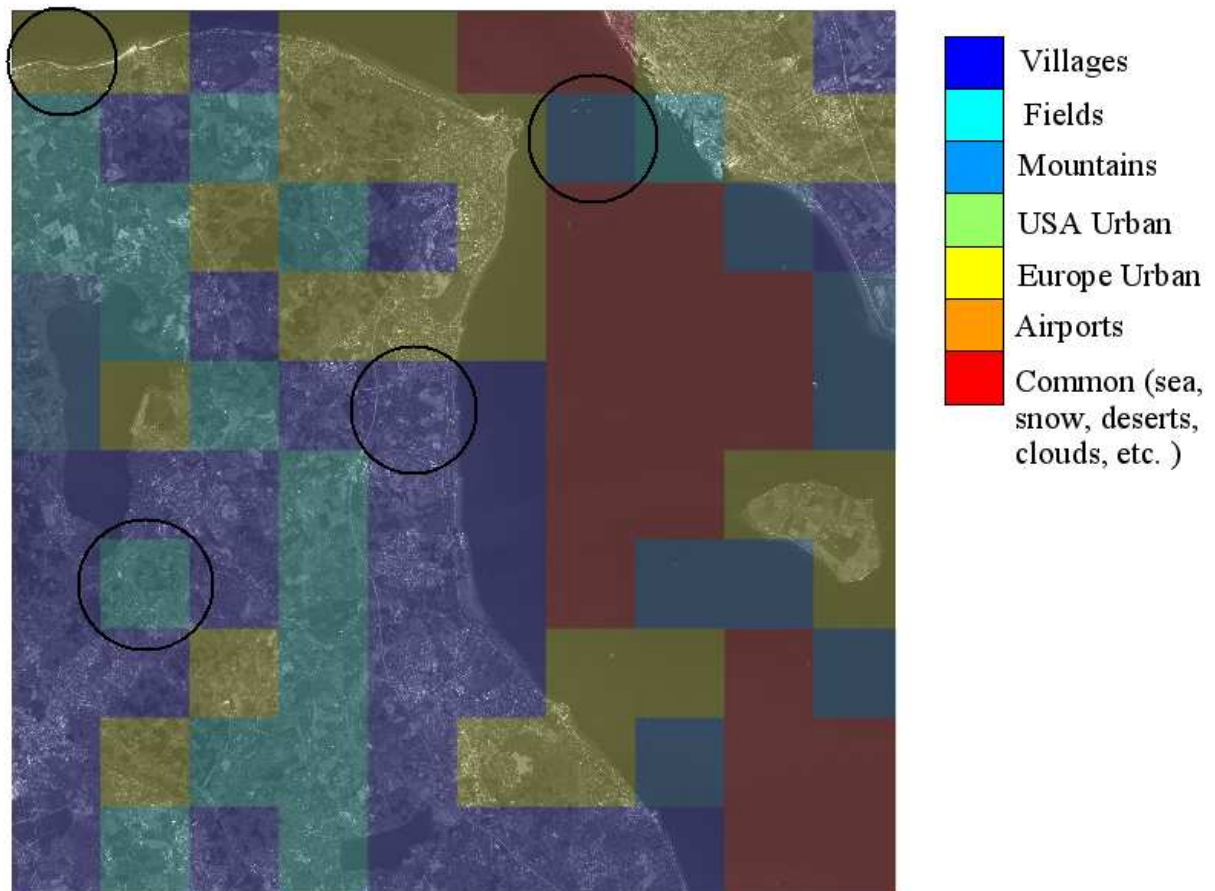


Figure 6.8: Classification results superimposed on the original image of Copenhagen.



Figure 6.9: Ground truth with Google maps ©2007 Google-Imagery ©2007 Terra-Metrics, Map data ©2007 Tele Atlas.



As can be seen in Figure 6.8, 80% of the regions are well classified, with a few regions having ambiguous label and a few regions misclassified due to insufficient or wrong information collected from the extracted networks. A detailed study of the four regions marked in Figure 6.8 is given in subsection 6.2.4.

#### 6.2.4 Detailed analysis of 4 regions

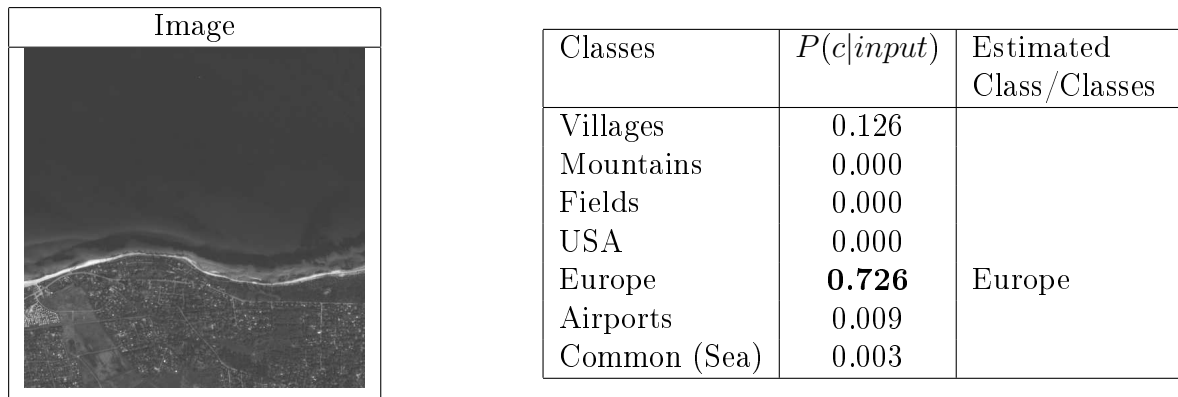


Figure 6.10: Probabilistic classification output with labels.

This is the region of Hornbaek, around 50km north of Copenhagen, Figure 6.9. This is a coastal town with geographic coordinates  $55^{\circ}5'N$ ,  $12^{\circ}27'E$ . The road structures pertaining in this region are well extracted and the features from them classify it correctly to be a Europe urban region.

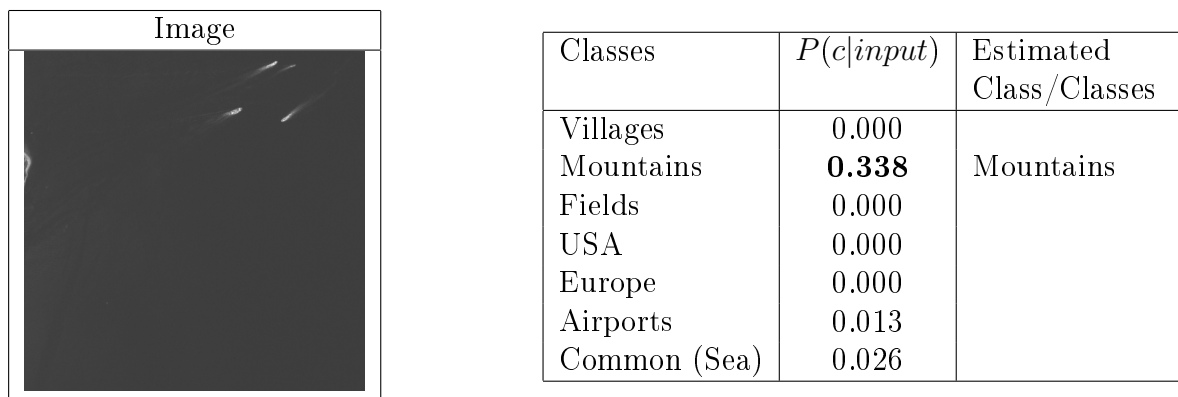


Figure 6.11: Probabilistic classification output with labels.

This is the strait of water that separates Denmark from Sweden, and that connects the North Sea with the Baltic Sea, Figure 6.9. This is a ferry route from Helsingor in Denmark to Helsingborg in Sweden. The constant movement of ferries in this region forms linear streaks on the water surface, which are extracted by the methods. Features from them then create confusion with road network structures. The region is hence classified as Mountains with a probability of 34%.

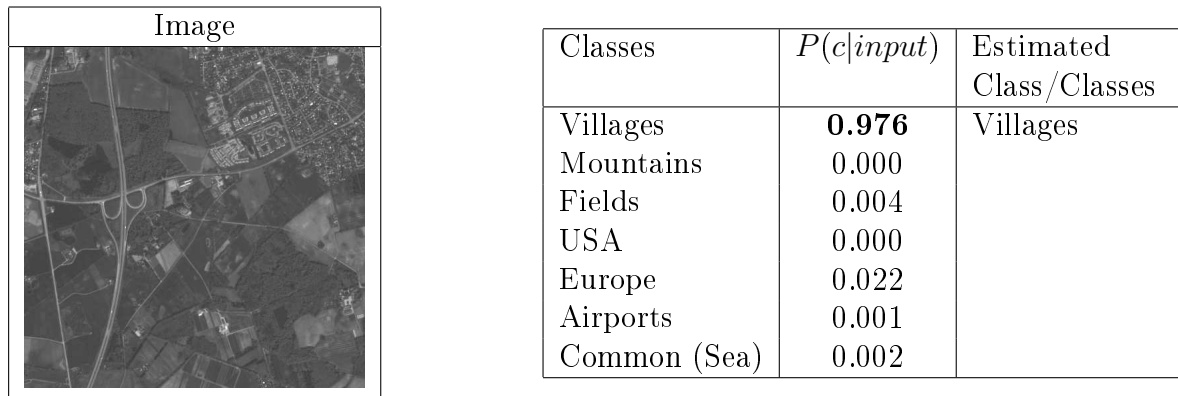


Figure 6.12: Probabilistic classification output with labels.

This is the small town of Espergaerde, located in the north of Copenhagen with geographic coordinates 56°0'N, 12°34'E, Figure 6.9. The urban region and the roads are well extracted to classify it as a Villages class.

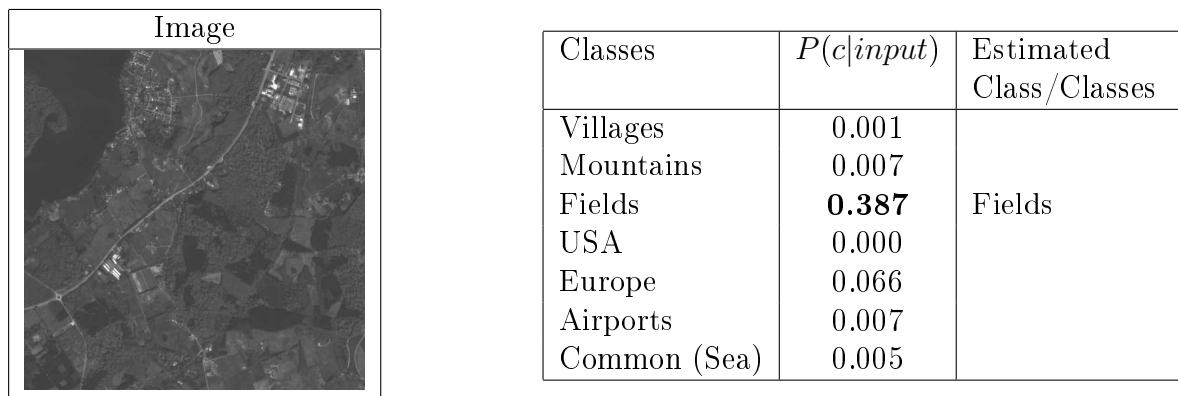


Figure 6.13: Probabilistic classification output with labels.

This region lies in the south of the Esrum See, Figure 6.9. This is an agricultural region with large stretches of fields. The straight and curved road structures from the fields are well extracted and the features computed from them classify this area correctly as Fields class.

### 6.3 Indexing results on large satellite images

In this section we present some more examples of indexing of large satellite images. The "Region Matrix" shows the confidence of the classification performance. The quantitative measures of these classifications with probabilities are then analyzed to infer the certainty of the class labels. The region matrix is then superimposed on the original image to give a visual indication of the overall indexing result. The indexing scheme is shown in Figure 6.14. Detailed analysis of some specific regions is performed using ground truth from satellite images together with maps available from Google maps.

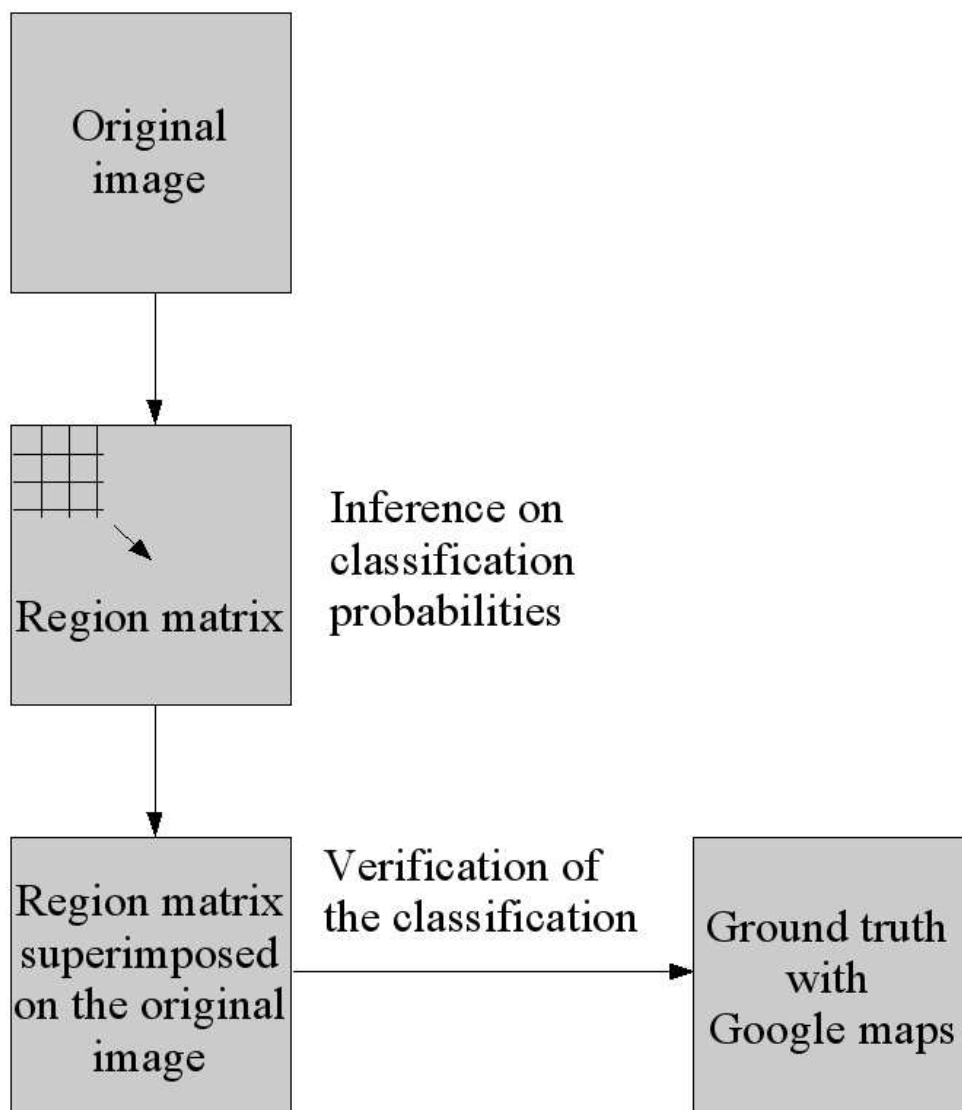


Figure 6.14: The experimental structure.

### 6.3.1 Barcelona

The region matrix in Figure 6.16 shows the classification probabilities of an area located in the north of Barcelona. As can be seen in the original image in Figure 6.15, a large part of it is a mountainous region, which is correctly labeled in the region matrix. The south-eastern portion of the image consists of big and small urban structures which are correctly classified as Europe urban and Villages classes. Two areas, shown with red labels, corresponding to a Common class are misclassified due to a network extraction failure in these regions.



Figure 6.15: Original image of Barcelona ©CNES.

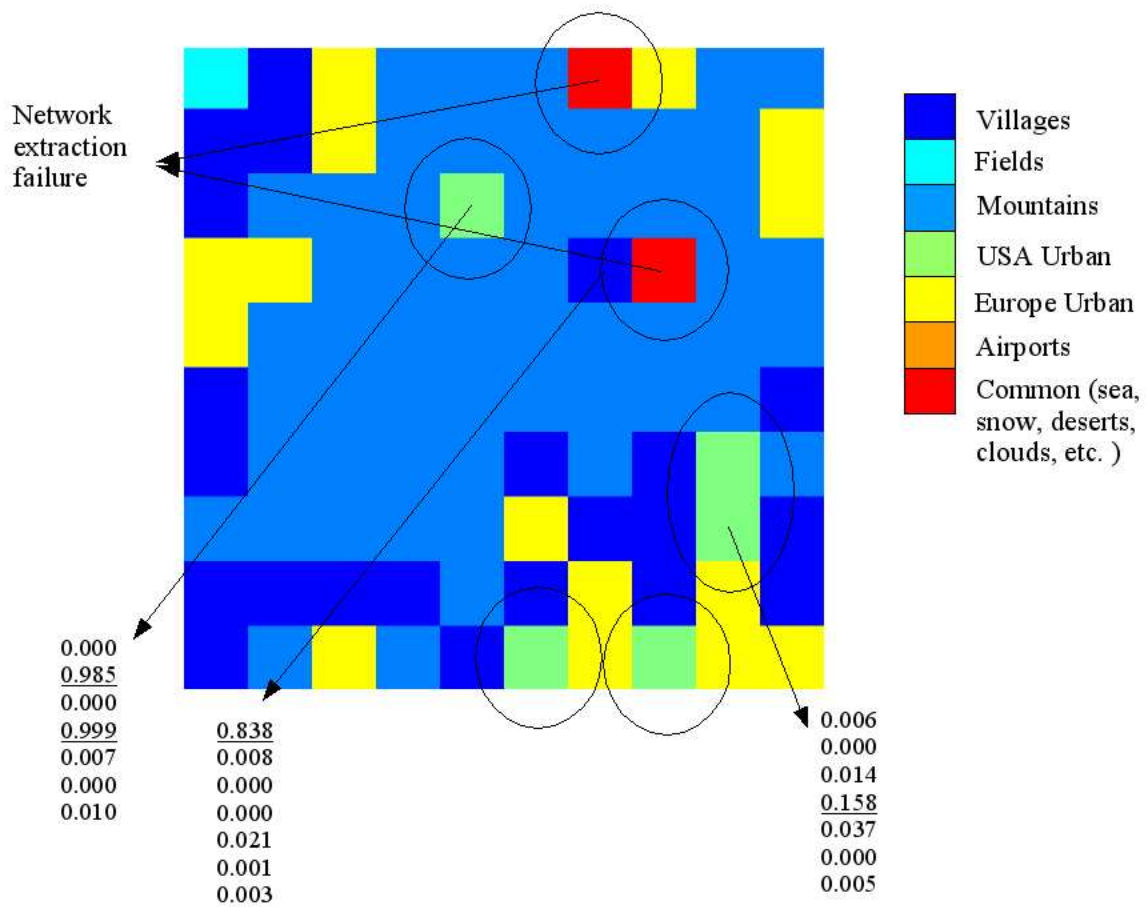


Figure 6.16: Classification results for Barcelona.

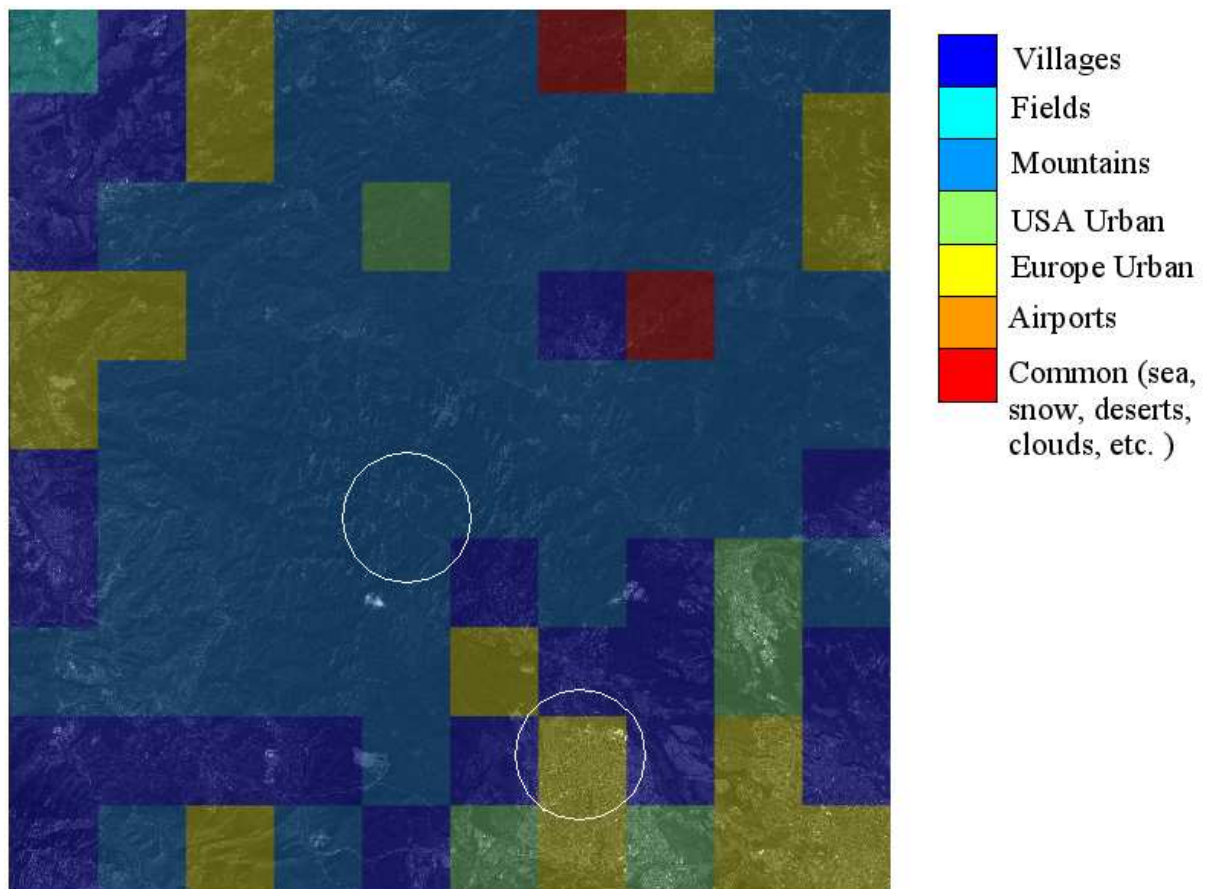


Figure 6.17: Classification results superimposed on the original image of Barcelona.



Figure 6.18: Ground truth with Google maps ©2007 Google-Imagery ©2007 Terra-Metrics, Map data ©2007 Tele Atlas.

It can be seen in the matrix that a few regions are labelled as USA urban. A closer look at the matrix reveals that either the classification probabilities of these regions are low or they are comparable to other neighboring classes. The superimposed image shown in Figure 6.17 gives us an overall qualitative measure of the classification. A detailed analysis of the two regions marked in Figure 6.17, with ground truth from Figure 6.18, is reported in subsection 6.3.2.

### 6.3.2 Detailed analysis of 2 regions

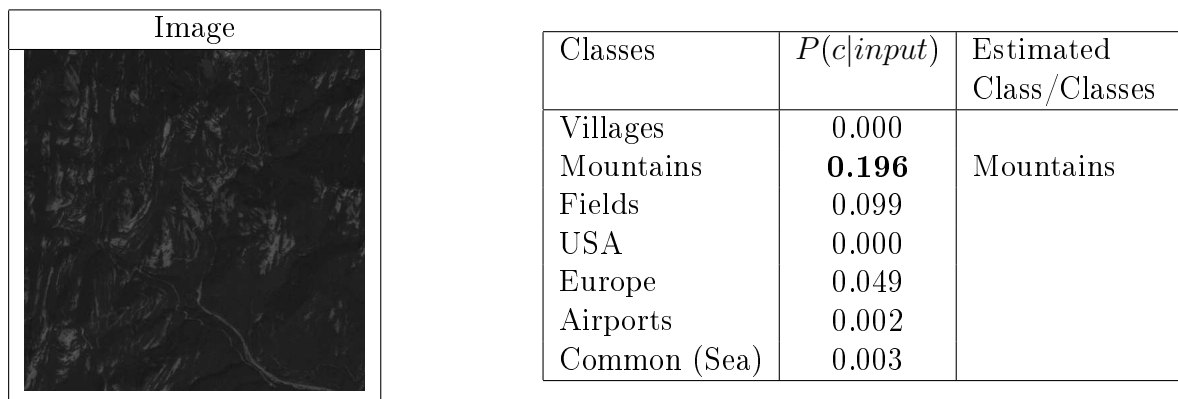


Figure 6.19: Probabilistic classification output with labels.

This region lies close to a small urban area la Barata, north of the city of Barcelona, Figure 6.18. This is a mountaneous region covered with dense forest. The curved and winding road network structure provides a good discriminative property for the classification of this region. The region is correctly classified as Mountains class as shown in the superimposed image, Figure 6.17.

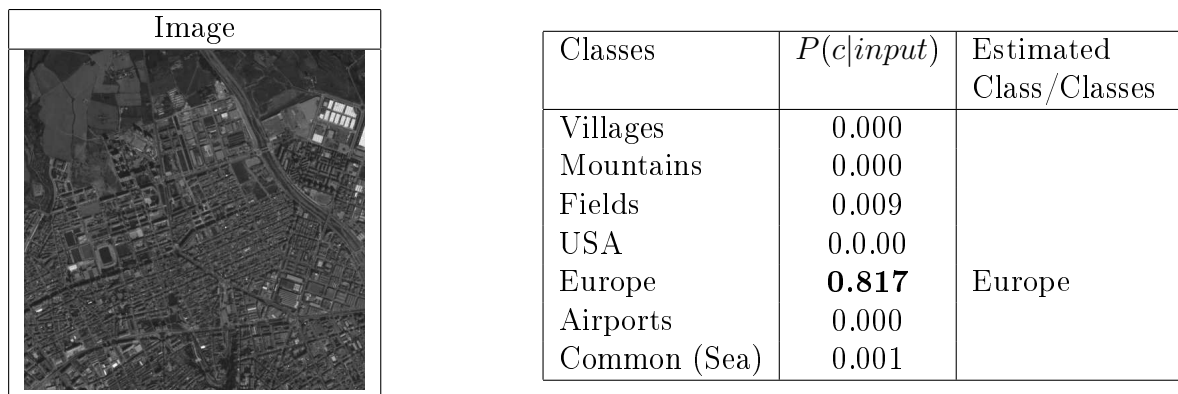


Figure 6.20: Probabilistic classification output with labels.

This is the industrial town of Tarrasa which lies to the north of the city of Barcelona, Figure 6.18. Its geographic coordinates are 41°34'N, 2°00'E. The features from the road



network structures in this region correctly classify it as Europe urban class, as shown in Figure 6.17.

---

### 6.3.3 Los Angeles

The original image of a part of Los Angeles city in Figure 6.21 shows an urban coverage on the western, central and south-eastern regions. The north-eastern and south-western regions are mountainous, and are correctly labelled as Mountains class.



Figure 6.21: Original image of Los Angeles ©CNES.

The region matrix in Figure 6.22 shows that certain areas are classified as Europe urban. This can be explained from the fact that either the classification probabilities are low or they are comparable with the neighboring classes. The other reason for this is that the network structures in these areas is similar to the one found in many

---

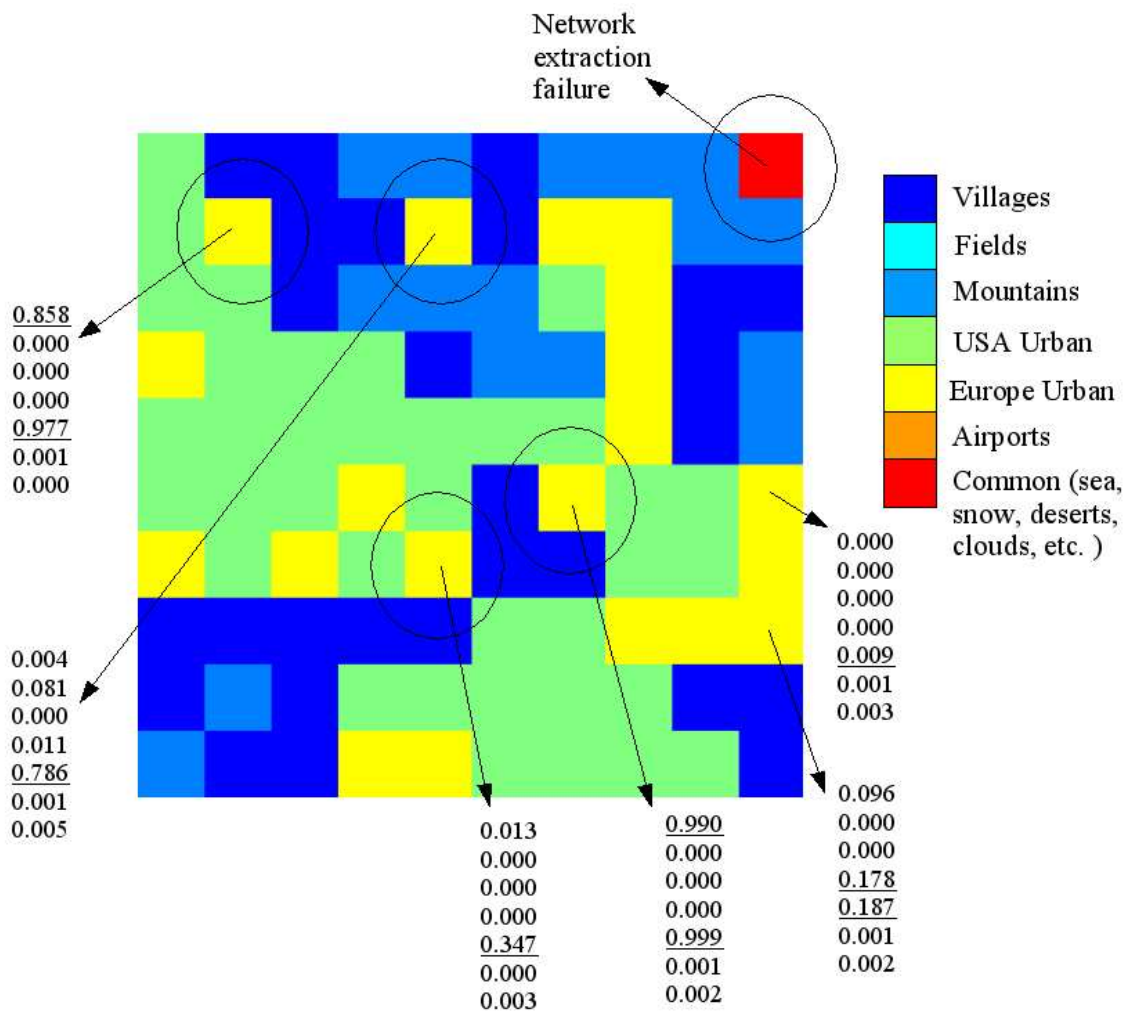


Figure 6.22: Classification results for Los Angeles.

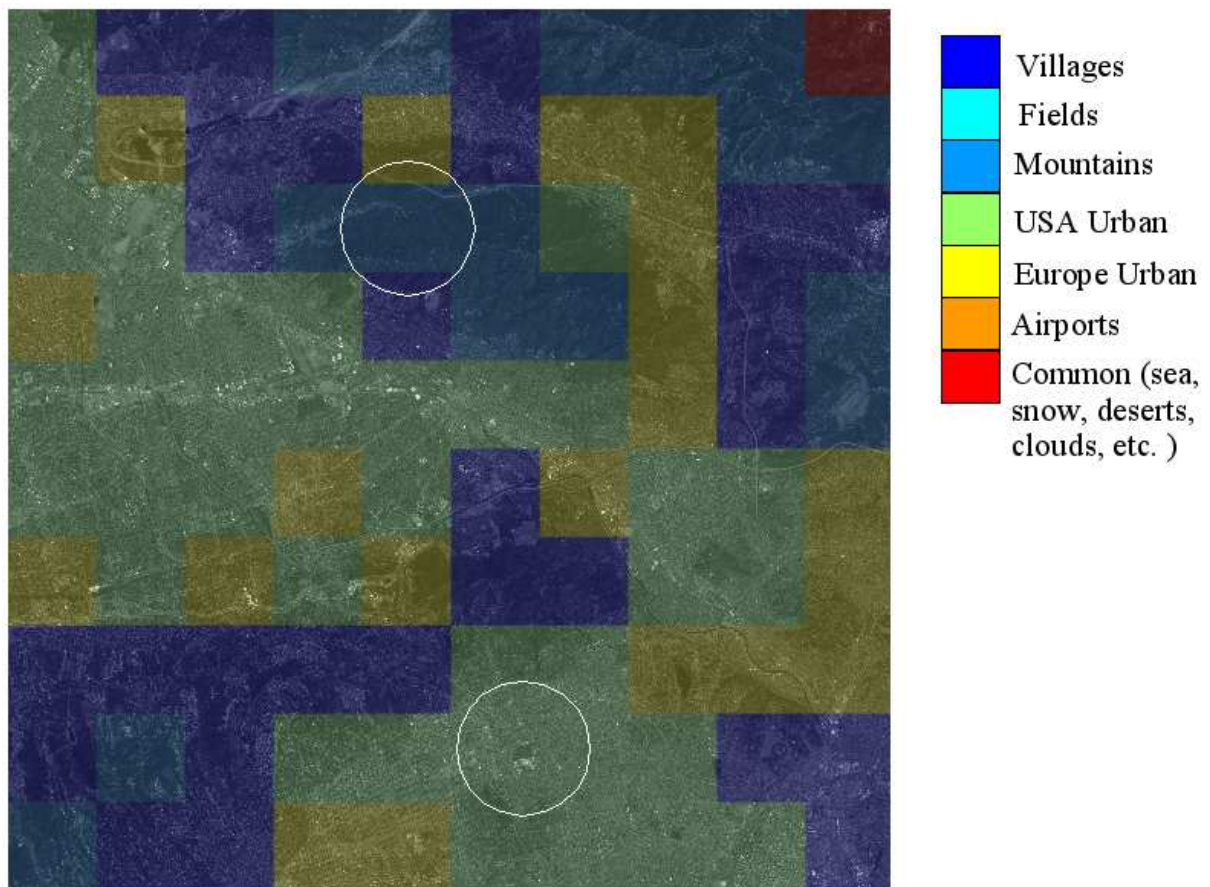


Figure 6.23: Classification results superimposed on the original image of Los Angeles.





Figure 6.24: Ground truth with Google maps ©2007 Google-Imagery ©2007 NASA, TerraMetrics, Map data ©NAVTEQ<sup>TM</sup>, LeadDog Consulting.

European urban structures. The superimposed image in Figure 6.23 validates the classified regions. A detailed analysis of the two regions marked in Figure 6.23, with ground truth from Figure 6.24, is reported in subsection 6.3.4.

### 6.3.4 Detailed analysis of 2 regions

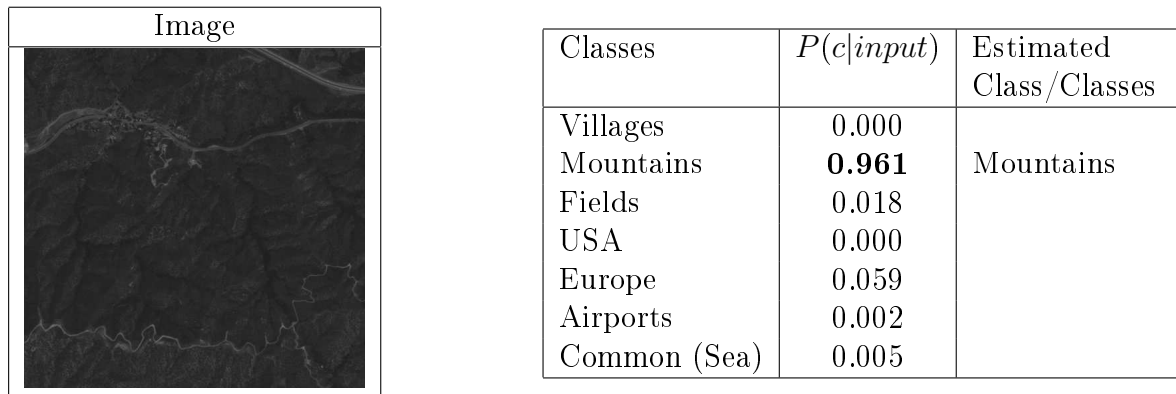


Figure 6.25: Probabilistic classification output with labels.

This region lies to the south of the two small urban areas of Sunland and Tujunga, Figure 6.24. This is a hilly region covered with green patches of grassland. There is a big highway across this region. There are many winding narrow roads in the hills. The features from these road network structures correctly label this region as Mountains class, as shown in the superposed image, Figure 6.23.

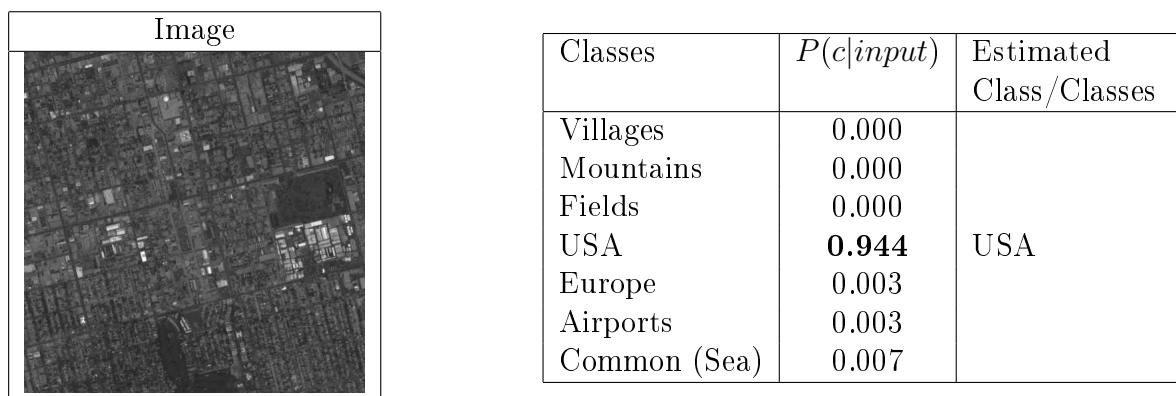


Figure 6.26: Probabilistic classification output with labels.

This is the central part of the city of Los Angeles, Figure 6.24. This is the largest city in the state of California. The dense grid-like medium and broad road structures found in this region can be used to identify the geographical location of such urban settlements in satellite images. The features extracted from the road structures pertaining in this region are used to correctly classify this region as a USA urban region. Figure 6.23 validates the classification results.

### 6.3.5 Madrid

The original image corresponds to a area located in the north of the city of Madrid. Figure 6.27 shows a large coverage of highlands with some big and small scattered urban areas. As can be seen in the region matrix in Figure 6.28, the central and the north-eastern parts are highland areas with some Villages and Fields structures in between. The eastern part is covered with small urban regions which are correctly labelled as Villages.

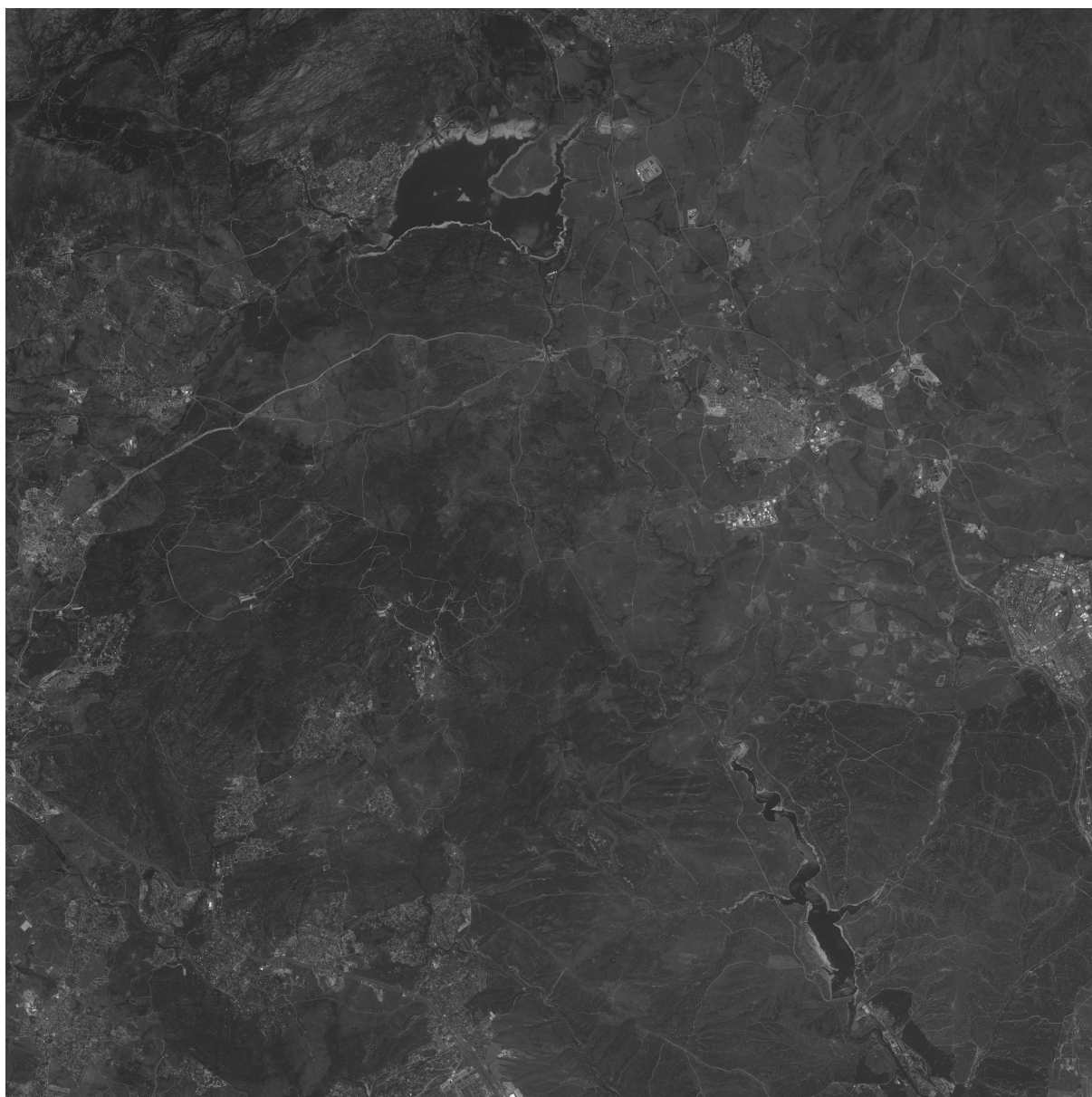


Figure 6.27: Original image of Madrid ©CNES.

The probabilities in the matrix show that the classification confidence in the marked

---



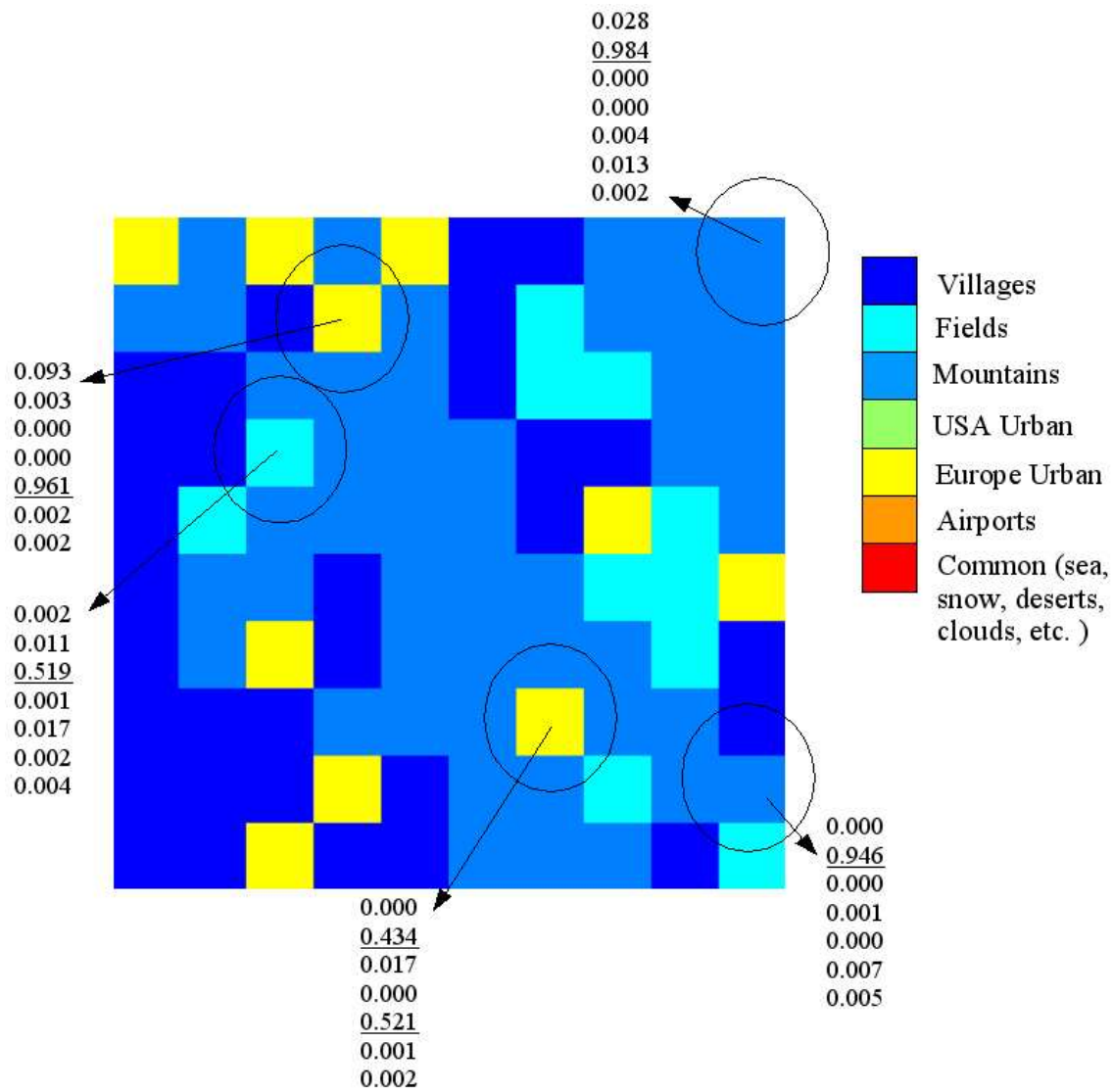


Figure 6.28: Classification results for Madrid.



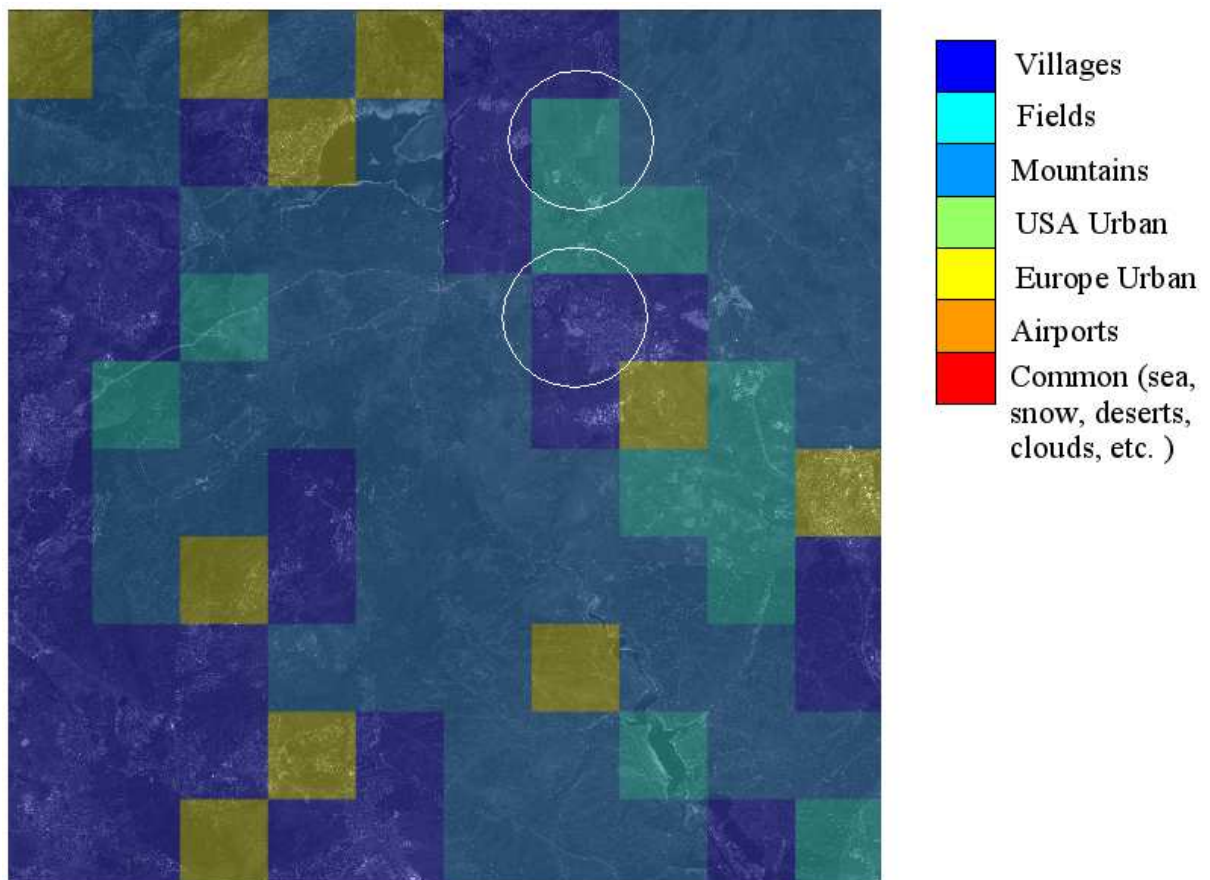


Figure 6.29: Classification results superimposed on the original image of Madrid.



Figure 6.30: Ground truth with Google maps ©2007 Google-Imagery ©2007 Terra-Metrics, Map data ©2007 Tele Atlas.

areas is high, i.e. there is a near-certainty in the class labelling. A European urban area marked in the region matrix has classification probability comparable to that of Mountains. A closer look at the neighbors of this region and the network structure pertaining to it affirm that this region is misclassified as a Europe urban region. A closer analysis with Google maps in Figure 6.30 of the two regions marked in Figure 6.29 is reported in subsection 6.3.6.

### 6.3.6 Detailed analysis of 2 regions


Image	Classes	$P(c input)$	Estimated Class/Classes
	Villages	0.003	Fields
	Mountains	0.085	
	Fields	<b>0.331</b>	
	USA	0.000	
	Europe	0.139	
	Airports	0.008	
	Common (Sea)	0.002	

Figure 6.31: Probabilistic classification output with labels.

This region lies to the north of the small urban settlement of San Pedro, Figure 6.30. This is an agricultural region with a large stretch of fields. The curved road structures are well extracted from the image and the features from them classify it correctly as Fields class.


Image	Classes	$P(c input)$	Estimated Class/Classes
	Villages	<b>0.987</b>	Villages
	Mountains	0.000	
	Fields	0.003	
	USA	0.000	
	Europe	0.756	
	Airports	0.001	
	Common (Sea)	0.002	

Figure 6.32: Probabilistic classification output with labels.

This is the small village of Colmenar Viejo, located to the north of Madrid with geographic coordinates of 40°39'N, 3°46'E, Figure 6.30. The road networks and the urban region features correctly classify it as Villages class. As can be seen in the

classification probabilities, the road features and the size of the urban structures also classify this region to be Europe urban with a probability of 76%.

---

### 6.3.7 Paris

The image in Figure 6.33 shows a part of a suburb of Paris with a major coverage of Fields areas. The south-eastern portion of the image is covered with big and small urban areas. The white regions are clouds which cover certain portions of the urban regions. The region matrix in Figure 6.34 depicts the good classification of the Fields class which comprises of around 70% of the image.

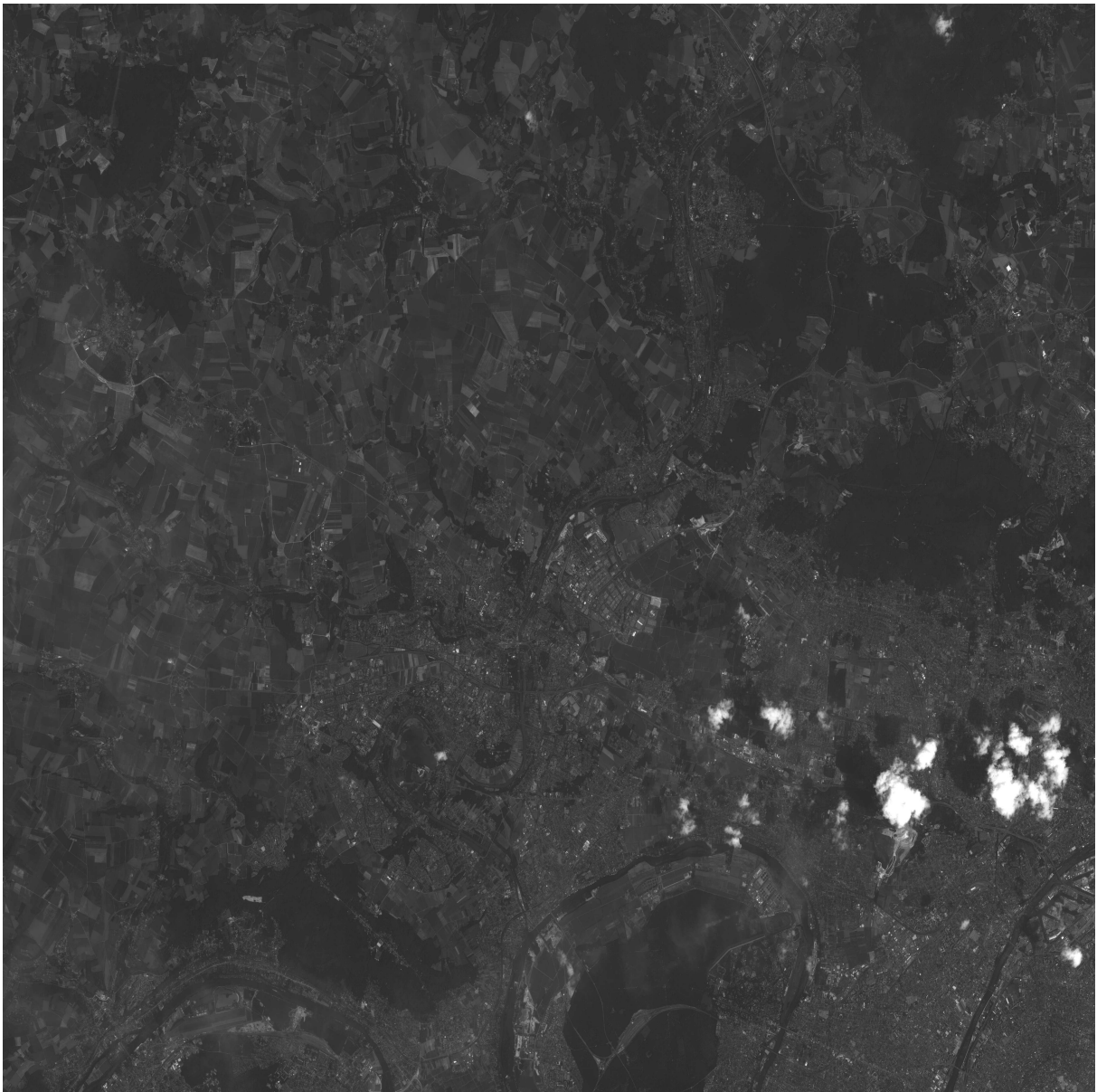


Figure 6.33: Original image of Paris ©CNES.

Some small urban areas are misclassified as Mountains due to poor extraction of the road network and urban regions pertaining to that image patch. The extraction

---

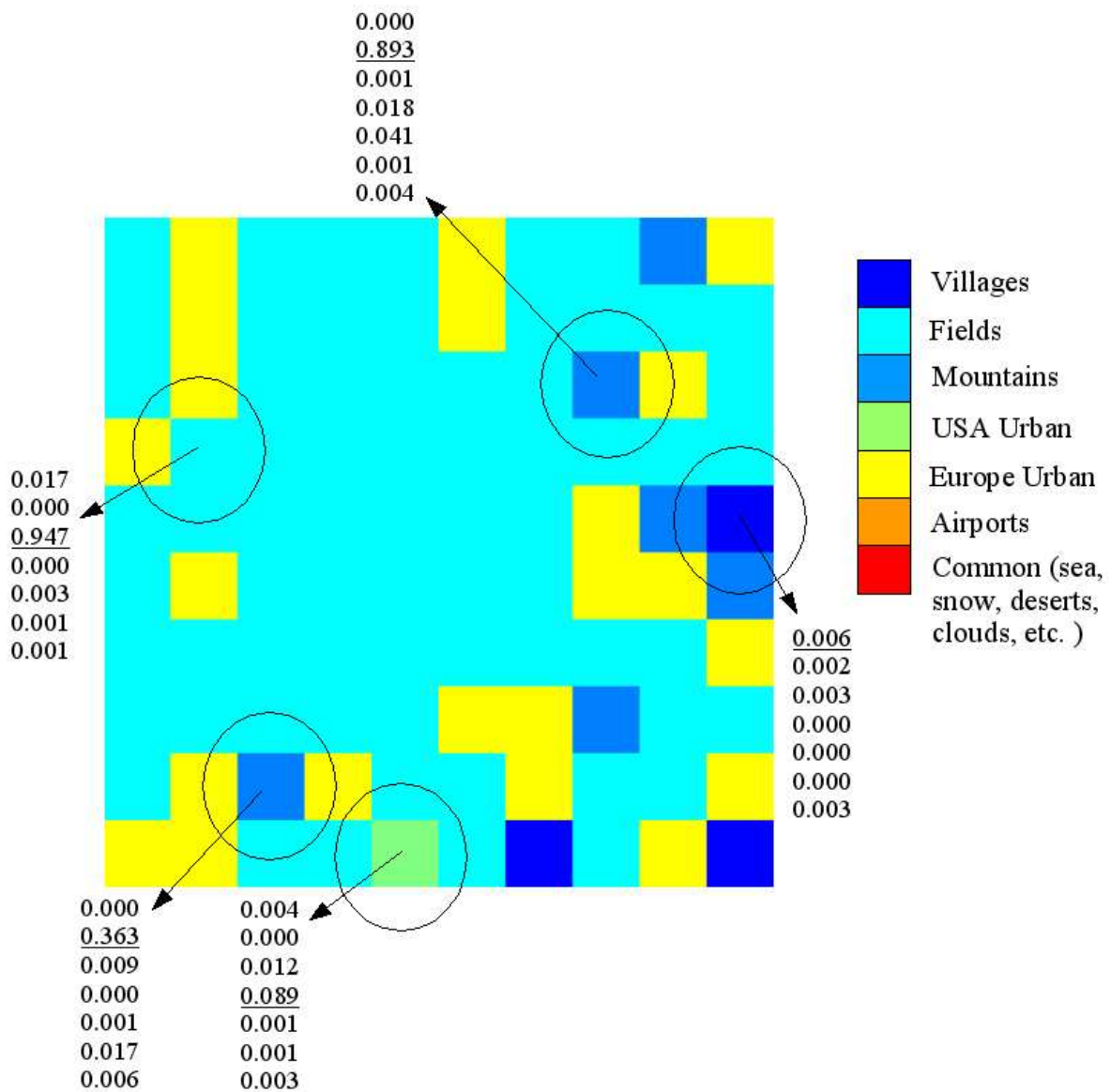


Figure 6.34: Classification results for Paris.

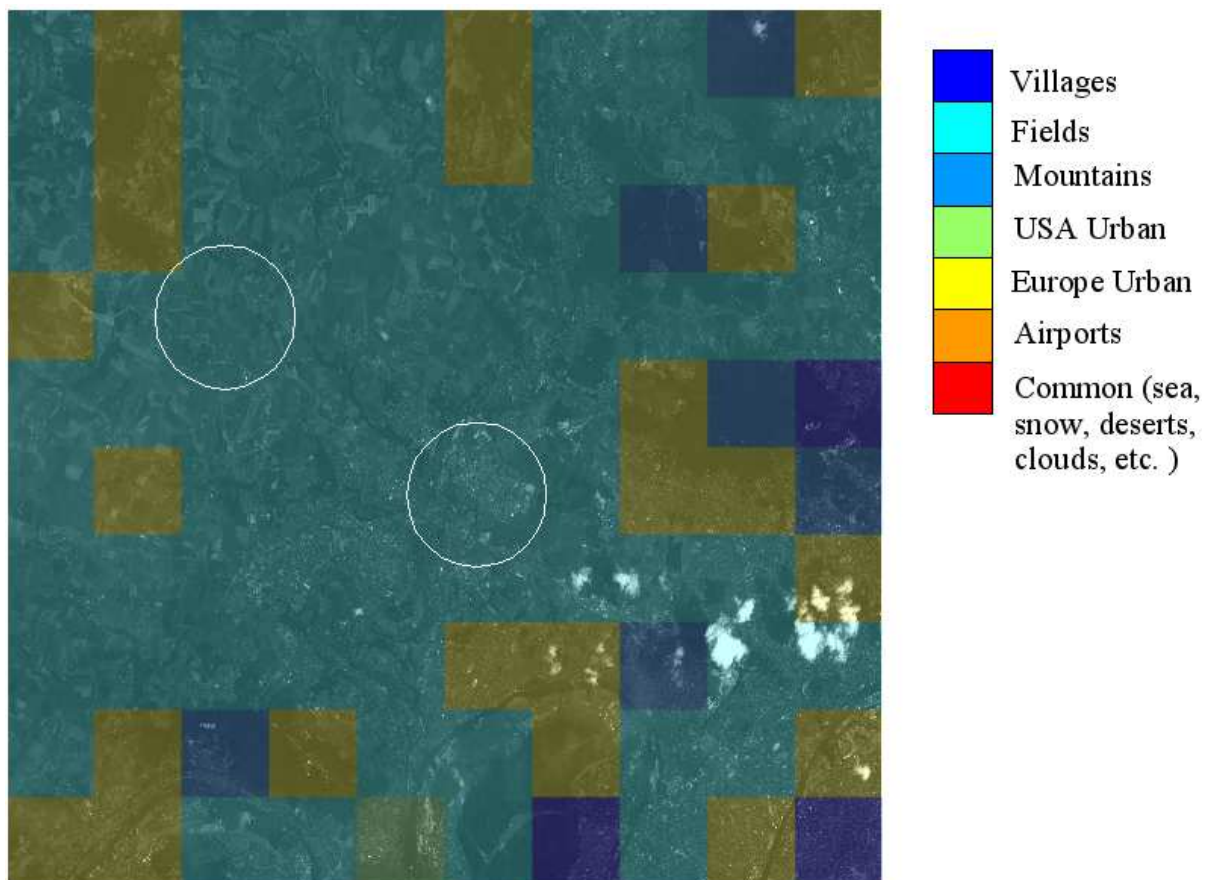


Figure 6.35: Classification results superimposed on the original image of Paris.



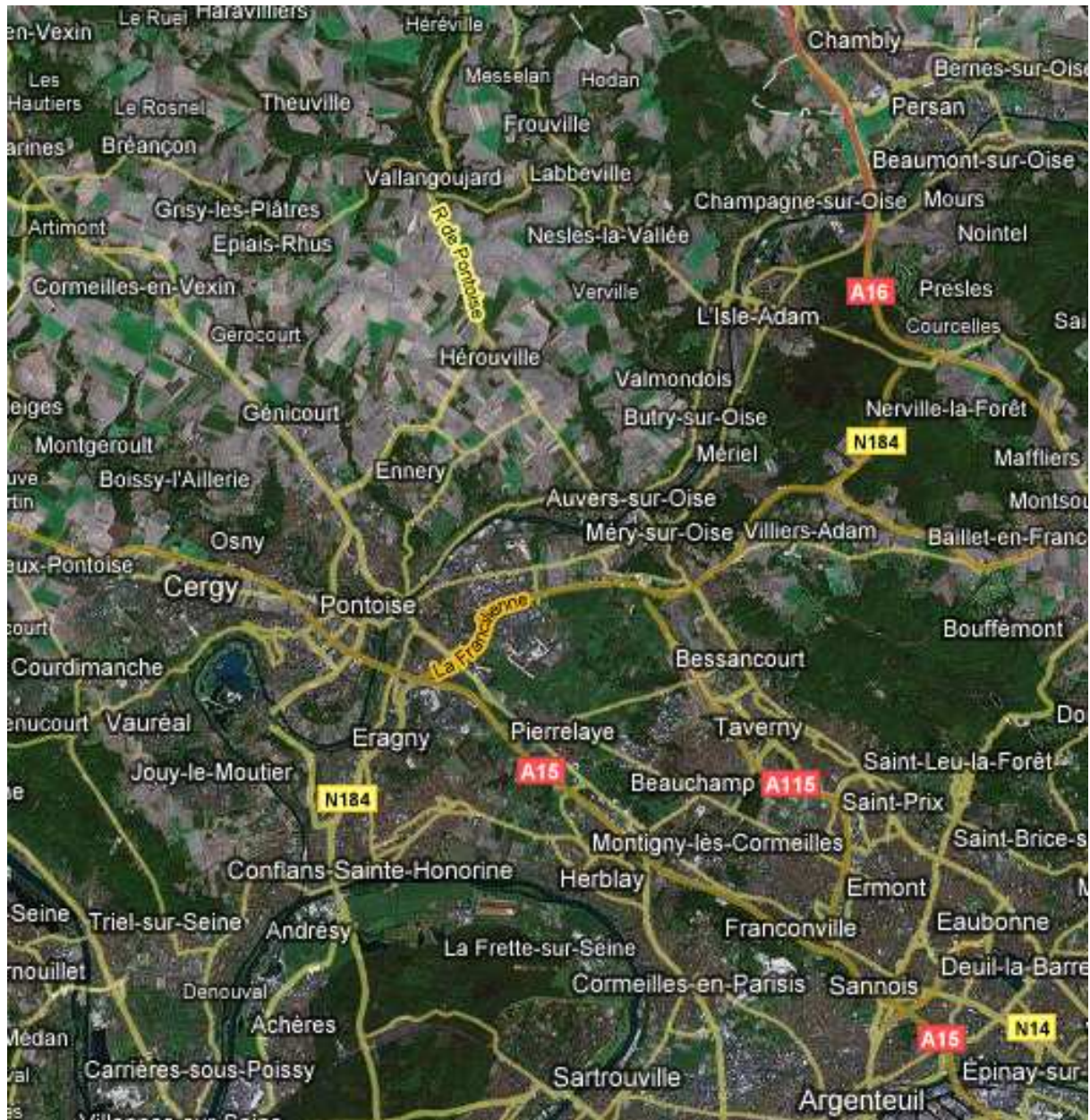


Figure 6.36: Ground truth with Google maps ©2007 Google-Imagery ©2007 Terra-Metrics, Map data ©2007 Tele Atlas.



of networks from certain urban regions is hindered by cloud coverage over that area, resulting in poor classification. A detailed analysis with Google maps in Figure 6.36 of the two regions marked in Figure 6.35 is reported in subsection 6.3.8.

### 6.3.8 Detailed analysis of 2 regions

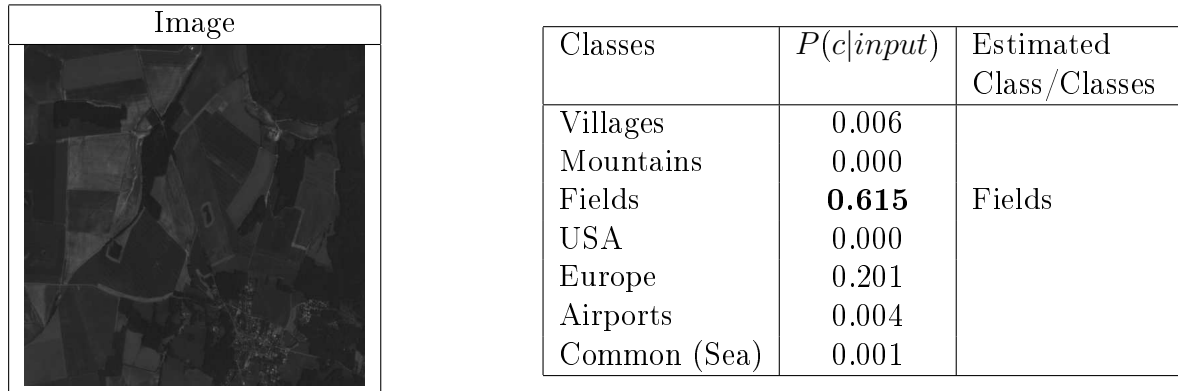


Figure 6.37: Probabilistic classification output with labels.

This area lies in the north-west of the region of Cergy as can be seen in Figure 6.36. This is an agricultural region with large fields. The road network structures from these field areas are well extracted, and the features from them correctly classify it as Fields class.

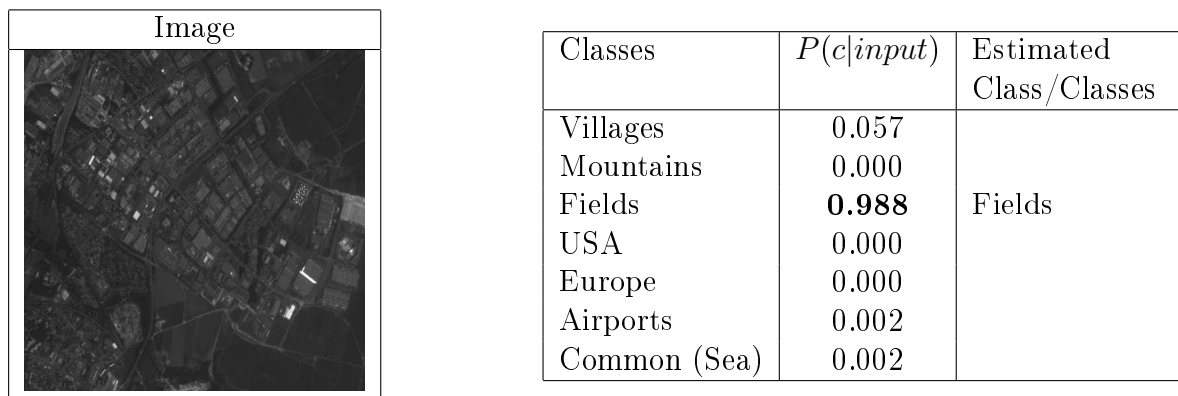


Figure 6.38: Probabilistic classification output with labels.

This is the region of Pontoise, a commune of the Val-d'Oise located in the north-western suburbs of Paris. Its geographic coordinates are 49°3'N, 2°6'E. The road network structures are not well extracted from the image and hence the features from them hinders correct classification of this area. A good extraction of the road network would have classified this area as Europe urban, but instead it is classified as Fields.

## 6.4 Discussion

Classification of large satellite images with patches of images extracted from them is a novel idea in the sense that the patches considered contain significant coverage of a particular type of geographical environment. Probabilistic SVM provides us with a quantitative analysis of the classification. The results show that the method is able to classify a large image into various geographical classes with a mean error less than 10%. The few classification errors that can be seen in the results can be rectified using the neighborhood information from the misclassified image. This method provides a basis for more complex analysis of large satellite images. There are some issues which have not been tackled in this work. The effect of overlapping patches on classification is not reported. This may be an interesting study, as it can help to better classify the images. Moreover, image patches of different sizes can also be used to improve the classification performance.

The detailed analysis of the regions marked in the superposed images confirms the classification results with ground truth from Google maps. A closer look at the urban structures and the road networks pertaining to these regions also confirm our classification results.

Further post processing of the region matrix with Stochastic models, e.g., Markov Random Fields (MRF), can also be used to regularize the region matrix for a finer classification.

---



## Chapter 7

### Conclusion

The growth of large image databases during the last few decades with the advancement in image acquisition technologies has attracted researchers from different fields to work in the domain of image information mining systems. The images coming from various sources must be systematically analyzed to render important information. A vast literature can be found in the field of CBIR and KDD. A brief overview of the literature on image content reveals that shape and texture have been widely used to capture information from satellite images. Classifications with shape or texture features extracted from the image helps us to understand the spatial structure of the Earth's surface. The two retrieval systems which were at our disposal have motivated us to work in the field of indexing and retrieval of satellite images from large images repositories. In the future, our work can be integrated with these two systems to render some additional features along with the existing ones.

Indexing and retrieval of images from remote sensing data archives relies on the extraction of appropriate information. The road networks contained in an image represent one type of information. The properties of road networks vary considerably from one geographical environment to another, and they can therefore be used to classify and retrieve different geographical environments. In our work, we have defined several such environments, and classified them with the aid of geometrical and topological features computed from the road networks occurring in them. In order to compute these features, we have used a graph representation of the road networks. This representation is robust and captures the network topology. There are certain limitations of the road network extraction methods. The road network extraction methods often fail to extract the finely structured road networks in small urban areas, with the consequence that the features computed from road networks poorly classify images containing such areas. In order to circumvent the loss of information from these areas, a method is adopted to incorporate the properties of these areas in our study. In SPOT images, the urban zones appear to be strongly textured and the problem of extracting the regions is essentially a problem of differentiating textures. The urban regions are segmented from the image and a new set of geometrical and topological features are computed

---

from these regions. These new features augmented with the road network features better classified the geographical environments.

A detailed study showed the dependence on image size and resolution of the features set. Features computed from a lower resolution show a greater amount of confusion between classes. This confusion is also visible with images of smaller sizes. However, a few anomalies can be explained on the basis of the road network and urban area extraction methods. Based on this study, more detailed analysis could be done on image databases of different resolutions. The optimal resolution for the categorization of man-made structures on the Earth's surface could be determined based on this study. The image size factor, which is also crucial for delineating class boundaries, remains to be solved. The study reported in this work provides a foundation for answering several such queries in the future.

Classification of large satellite images with patches of images extracted from them is a novel idea in the sense that the patches considered contain significant coverage of a particular type of geographical environment. The probabilistic SVM provides us with a quantitative analysis of the classification. The results confirm that the method is able to classify a large image into various geographical classes with a mean error of less than 10%. The few misclassifications that can be seen in the results can be rectified using neighborhood information from the classified image. The method provides the basis for a more complex analysis of large satellite images. There are some issues which have not been tackled in this work. The effect of overlapping patches on classification is not reported. This may be an interesting study, as it can help to better classify the images. Moreover, image patches of different sizes can also be used to improve the classification performance.

The detailed analysis of selected regions confirms the classification results using ground truth from Google maps. A closer look at the urban structures and the road networks pertaining to these regions confirm our classification results. Further post-processing of the region matrix with stochastic models, e.g., Markov Random Fields (MRF), can also be used to regularize the region matrix for a finer classification.

Future studies can use operators to detect not only man-made structures like roads and urban areas, but natural entities like rivers, forests, etc. The work of [Papa 98] could also be used to detect and reconstruct buildings from aerial imagery of different resolutions for the analysis of urban and suburban areas. In this work, we have restricted ourselves to a single resolution, but our methodology can be adapted to consider images of other resolutions from QuickBird or future Pleiade satellites. At a higher resolution, it may be possible to extract different structures such as buildings, gardens, cross-roads, etc. This in turn will allow us to incorporate more classes and appropriately classify any geographical environment. At an image resolution of 1m, we may imagine to have sub-classes of an existing class, for e.g., classes such as urban Europe and urban USA can be divided into downtown, residential and industrial classes.

---

At a lower resolution, there will be a profusion of features and the interaction between these features will be complex to handle. A method should be adapted to deal with such cases. In order to have a better understanding of images we should be able to incorporate texture features in our existing system. The results of this work, together with the above mentioned perspectives, can be adapted for use with existing and future image information mining systems for EO archives.

---



# Appendix A

## Personal Biography

### JOURNALS

**Computing Statistics from Man-Made Structures on the Earth's Surface for Indexing Satellite Images**, A. Bhattacharya and M. Roux and H. Maitre and I. Jermyn and X. Descombes and J. Zerubia. *International Journal of Simulation Modelling, Special issue on CompIMAGE - Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications*, Vol. 6, No. 2, pp. 73-83, 2007.

### CONFERENCES

**Indexing of Mid-Resolution Satellite Images with Structural Attributes**, A. Bhattacharya and M. Roux and H. Maitre and I. Jermyn and X. Descombes and J. Zerubia. Submitted to *XXI ISPRS Congress, 2008, Beijing, China*

**Indexing Satellite Images with Features Computed from Man-Made Structures on the Earth's Surface**, A. Bhattacharya and M. Roux and H. Maitre and I. Jermyn and X. Descombes and J. Zerubia. In Proc. *Fifth International Workshop on Content-Based Multimedia Indexing (CBMI 2007), Bordeaux, France*. pp: 244-250

**Computing Statistics from a Graph Representation of Road Networks in Satellite Images for Indexing and Retrieval**, A. Bhattacharya and I. H. Jermyn and X. Descombes and J. Zerubia. In Proc. *CompIMAGE - Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications, Coimbra, Portugal, October 2006*. pp: 97-100

**Shape Moments for Region-Based Active Contours**, P. Horvath and A. Bhattacharya and I. H. Jermyn and J. Zerubia and Z. Kato. In Proc. *Hungarian-Austrian Conference on Image Processing and Pattern Recognition, Szeged, Hungary, May 2005*. pp: 187-194

---



**POSTERS**

**Extraction of Structural Primitive Features from Satellite Images for Indexation and Retrieval**, A. Bhattacharya and M. Roux and H. Maitre and I. Jermyn and X. Descombes and J. Zerubia. *Poster at the 1st Symposium of the CNES/DLR/ENST Competence Centre on Information Extraction and Image Understanding for Earth Observation, Paris, France.*

**INVITED TALKS/SUMMER SCHOOL**

**Computing Statistics from Man-Made Structures on the Earth's Surface for Indexing Satellite Images**, At the *7th CNES/DLR Workshop on Information Extraction and Scene Understanding for Meter Resolution Images, DLR, Oberpfaffenhofen, Germany.*

**Features from Graph Representation of Road Networks in Satellite Images for Indexation**, Cross-Seminar *ARIANA/MAESTRO* research groups, *INRIA, Sophia Antipolis, France.*

**Features from Graph Representation of Road Networks in Satellite Images for Indexation**, At *IMEDIA* research group, *INRIA Rocquencourt, France.*

**Graph Representation of Road Networks in Satellite Images for Indexation**, At *École Supérieure des Communications de Tunis (Sup'Com), Tunisia, under Action Concertée Incitative Massées de Donnes (ACI) QuerySat Project.*

**Participant at the 3rd British Computer Society Summer School on Pattern Recognition**, *Portland Square, Plymouth, UK, July, 2005.*

---

## Bibliography

- [Agga 01] C. Aggarwal, A. Hinneburg, and D. Keim. “On the Surprising Behavior of Distance Metrics in High Dimensional Space”. *Lecture Notes in Computer Science*, Vol. 1973, pp. 420–434, 2001.
- [Akso 00] S. Aksoy and R. Haralick. “Probabilistic vs. Geometric Similarity Measures for Image Retrieval”. In: *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 2357–2362, 2000.
- [Akso 01] S. Aksoy and R. Haralick. “Feature normalization and likelihood-based similarity measures for image retrieval”. *Pattern Recognition Letters*, Vol. 22, No. 5, pp. 563–582, 2001.
- [Arki 91] E. Arkin, L. Chew, D. Huttenlocher, K. Kedem, and J. Mitchell. “An efficiently computable metric for comparing polygonal shapes”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 3, pp. 209–216, 1991.
- [Bajc 76] R. Bajcsy and M. Tavakoli. “Computer Recognition of Roads from Satellite Pictures”. *IEEE Trans. Syst., Man, and Cybern.*, Vol. 6, No. 9, pp. 623–637, Sep. 1976.
- [BAL] BAL. “Bridgeman Art Library”.
- [Barz 96] M. Barzohar and D. Cooper. “Automatic finding of main roads in aerial images by using geometric stochastic models and estimation”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 7, pp. 707–721, July 1996.
- [Baum 96] A. Baumgartner, C. Steger, C. Wiedemann, H. Mayer, W. Eckstein, and H. Ebner. “Update of Roads in GIS from Aerial Imagery: Verification and Multi-Resolution Extraction”. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXI, Part B3, pp. 53–58, 1996.
- [Baum 97] A. Baumgartner, C. Steger, H. Mayer, and W. Eckstein. “Multi-Resolution, Semantic Objects, and Context for Road Extraction”. In: *TopoInfo97*, pp. 140–156, 1997.
-

- 
- [Bess 96] V. Bessettes, J. Desachy, and E. Cubero-Castan. “Applying co-operative operators for urban-area detection using SPOT imagery”. In: *SPIE, Image and Signal Processing for Remote Sensing III, Jacky Desachy; Ed.*, pp. 106–117, 1996.
- [Bhat 06] A. Bhattacharya, I. H. Jermyn, X. Descombes, and J. Zerubia. “Computing statistics from a graph representation of road networks in satellite images for indexing and retrieval”. In: *Proc. CompIMAGE - Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications*, Coimbra, Portugal, 2006.
- [Bhat 07] A. Bhattacharya, M. Roux, H. Maitre, I. H. Jermyn, X. Descombes, and J. Zerubia. “Indexing Satellite Images with Features Computed from Man-Made Structures on the Earth’s Surface”. In: *Proc. International Workshop on Content-Based Multimedia Indexing*, pp. 244–250, 2007.
- [Bish 95] C. Bishop. *Neural networks for pattern recognition*. Oxford, Clarendon, 1995.
- [Bouj 01] N. Boujemaa, J. Fauqueur, M. Ferecatu, F. Fleuret, V. Gouet, B. Le Saux, and H. Sahbi. “IKONA: Interactive Generic and Specific Image Retrieval”. In: *International workshop on Multimedia Content-Based Indexing and Retrieval*, Rocquencourt, France, 2001.
- [Bovi 90] A. Bovic, M. Clark, and W. Geisler. “Multichannel texture analysis using localized spatial filters”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, pp. 55–73, 1990.
- [Buec 79] S. Buecher and C. Lantuejoul. “Use of watershed in contour detection”. In: *Proc. Int. Workshop Image Processing, Real-Time Edge and Motion Detection/Estimation*, pp. 17–21, Rennes, France, 1979.
- [Camp 00] C. Campbell. “Algorithmic approaches to training support vector machines: A survey”. 2000.
- [Camp 05] M. Campedel and E. Moulines. “Classification et sélection automatique de caractéristiques de textures”. *Revue des Nouvelles Technologies de l’Information*, Vol. RNTI-C-1, pp. 25–37, 2005.
- [Cann 86] J. Canny. “A computational approach to edge detection”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, pp. 679–698, 1986.
- [Chan 03] E. Chang, B. Li, G. Wu, and K. Goh. “Statistical learning for effective visual information retrieval”. In: *Proc. IEEE International Conference on Image Processing (ICIP)*, Barcelona, Spain, Sep. 2003.
-

- 
- [Chan 93] T. Chang and C.-C. Kuo. "Texture analysis and classification with tree-structured wavelet transform". *IEEE Trans. Image Processing*, Vol. 2, No. 4, pp. 429–441, 1993.
- [Chap 99a] O. Chapelle, P. Haffner, and V. Vapnik. "Support Vector Machines for Histogram-Based Image Classification". *IEEE Trans. Neural Networks*, Vol. 10, No. 5, pp. 1055–1064, 1999.
- [Chap 99b] O. Chapelle and V. Vapnik. "Model Selection for Support Vector Machines". In: *Neural Information Processing Systems*, pp. 230–236, 1999.
- [Chau 96] B. Chaudhuri, A. Mukherjee, S. Parui, B. Chaudhuri, and R. Krishnan. "An Efficient Algorithm for Detection of Road-Like Structures in Satellite Images". In: *Proc. International Conference on Pattern Recognition (ICPR)*, pp. 875–879, Washington, DC, USA, 1996.
- [Chaw 00] S. Chawla, S. Shekhar, W. Wu, and U. Ozesmi. "Extending Data Mining for Spatial Applications: A Case Study in Predicting Nest Locations". In: *Proc. Int. Conf. ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, Dallas, USA, May 2000.
- [Chel 84] R. Chellappa and R. Bagdazian. "Fourier coding of image boundaries". *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 6, No. 11, pp. 102–105, 1984.
- [Chen 01] Y. Chen, X. Zhou, and T. Huang. "One-class SVM for learning in image retrieval". In: *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 34–37, Thessaloniki, Greece, Oct. 2001.
- [Chen 02] D. Chen, D. Stow, and A. Getis. "Multi-resolution classification framework for improving land use/cover mapping". *Linking People, Place, and Policy: A GIScience Approach*, edited by Steve Walsh and Kelly Meyer, Kluwer Academic Publisher, pp. 235–262, 2002.
- [Chen 06] Y. Chen, Q. Yang, Y. Gu, and J. Yang. "Detection of Roads in SAR Images using Particle Filter". In: *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 2337–2340, Atlanta, GA, USA, 2006.
- [Chri 07] E. Christophe and J. Inglada. "Robust Road Extraction for High Resolution Satellite Images". In: *Proc. IEEE International Conference on Image Processing (ICIP)*, Sep. 2007.
- [Chua 96] G.-H. Chuang and C.-C. Kuo. "Wavelets descriptors of planar curves: Theory and applications". *IEEE Trans. Image Processing*, Vol. 5, No. 1, pp. 56–70, 1996.
- [Coop 79] D. Cooper. "Maximum Likelihood Estimation of Markov Process Blob Boundaries in Noisy Images". *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, pp. 372–384, 1979.
-

- 
- [Corm 01] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press and McGraw-Hill, 2001.
- [Cost 06] M. Costache, H. Maitre, and M. Datcu. “Categorization based relevance feedback search engine for Earth observation images repositories”. In: *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Denver, Colorado, 2006.
- [Cros 83] G. Cross and A. Jain. “Markov random field texture models”. *IEEE Trans. Pattern Recognition and Machine Intelligence*, Vol. 5, No. 1, pp. 25–39, 1983.
- [Dasc 05] H. Daschiel and M. Datcu. “Image Information Mining System Evaluation Using Information-Theoretic Measures”. *EURASIP Journal on Applied Signal Processing*, Vol. 2005, No. 14, pp. 2153–2163, 2005.
- [Datc 02] M. Datcu and K. Seidel. “An Innovative Concept for Image Information Mining”. In: *Proc. of the 3rd International Workshop on Multimedia Data Mining*, pp. 11–18, Edmonton, AL, Canada, 2002.
- [Datc 03] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. Marchetti, and S. D’Elia. “Information Mining in Remote Sensing Image Archives- Part A: System Concepts”. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 41, No. 12, pp. 2923–2936, 2003.
- [Datc 98] M. Datcu, K. Seidel, and M. Walessa. “Spatial Information Retrieval from Remote Sensing Images: Part I. Information theoretical perspective”. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 36, pp. 1431–1445, Sep. 1998.
- [Datc 99] M. Datcu, K. Seidel, and G. Schwarz. “Information Mining in Remote Sensing Image Archives”. In: I. Kanellopoulos, G. Wilkinson, and T. Moons, Eds., *Machine Vision and Advanced Image Processing in Remote Sensing*, pp. 199–212, Springer, 1999.
- [Dell 03] F. Dell’Acqua and P. Gamba. “Texture-based characterization of urban environments on satellite SAR images”. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 41, No. 1, pp. 153–159, 2003.
- [Desc 93] X. Descombes. *Champs Markoviens en analysis d’images*. PhD thesis, École National Supérieure des Télécommunications, Paris, France, Dec. 1993.
- [Dese 98] M. Deseilligny, G. Stamon, and C. Suen. “Veinerization: A New Shape Description for Flexible Skeletonization”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 20, No. 5, pp. 505–521, 1998.
-

- 
- [Deso 00] A. Desolneux, L. Maisson, and J.-M. Morel. “Meaningful Alignments”. *International Journal of Computer Vision*, Vol. 40, No. 1, pp. 7–23, 2000.
- [Dest 86] I. Destival and H. Le Men. “Detection of Linear Networks on Satellite Images”. In: *Proc. International Conference on Pattern Recognition (ICPR)*, pp. 856–858, Paris, France, 1986.
- [Dhil 04] I. Dhillon, Y. Guan, and B. Kulis. “Kernel k-means: spectral clustering and normalized cuts”. In: *The tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 551–556, New York, NY, USA, 2004.
- [Dial 01] G. Dial, L. Gibson, and R. Poulsen. “IKONOS Satellite Imagery and Its Use in Automated Road Extraction”. In: E. Baltsavias, A. Gruen, and L. Van Gool, Eds., *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*, pp. 357–367, Zurich, Switzerland, 2001.
- [Dimi 00] P. Dimitrov, C. Phillips, and K. Siddiqi. “Robust and efficient skeletal graphs”. In: *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 1417–1423, Hilton Head Island, USA, 2000.
- [Doer 98] D. Doermann. “The Indexing and Retrieval of Document Images: A Survey”. *Computer Vision and Image Understanding*, Vol. 70, No. 3, pp. 287–298, June 1998.
- [Duda 00] R. Duda, P. Hart, and D. Stork. *Pattern Classification, 2ed.* Wiley-Interscience, 2000.
- [Durb 04] S. Durbha and R. King. “Knowledge Mining in Earth Observation Data Archives: A Domain Ontology Perspective”. In: *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2004.
- [Equi 94] W. Equitz and W. Niblack. “Retrieving Images from a Database using Texture - Algorithms from the QBIC System”. Tech. Rep. RJ 9805, IBM Research, 1994.
- [ESA] ESA. “European Space Agency”.
- [Este 95] M. Ester, H. Kriegel, and X. Xu. “Knowledge Discovery in large Spatial Databases and Focussing Techniques for Efficient Class Identification”. *Lecture Notes in Computer Science*, Vol. 951, pp. 67–82, 1995.
- [Euri 93] G. M. P. Euripides and C. O. Stelios. “Methodology for the representation, indexing and retrieval of images by content.”. *Image Vision Comput.*, Vol. 11, No. 8, pp. 504–521, 1993.
- [Fauq 02] J. Fauqueur and N. Boujemaa. “Image Retrieval by Range Query Composition of Region Categories”. Research Report RR-4686, INRIA, Rocquencourt, 2002.
-

- 
- [Fauq 04] J. Fauqueur and N. Boujemaa. “Region-Based Image Retrieval: Fast Coarse Segmentation and Fine Color Description”. *Journal of Visual Languages and Computing (JVLC), special issue on Visual Information Systems*, Vol. 15, No. 1, pp. 69–95, 2004.
- [Fayy 96] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*. MIT Press, New York, NY, USA, 1996.
- [Fere 04] M. Ferecatu, M. Crucianu, and N. Boujemaa. “Sample selection strategies for relevance feedback in region-based image retrieval”. *Lecture Notes in Computer Science*, Vol. 3332, pp. 497–504, 2004.
- [Fere 05a] M. Ferecatu, N. Boujemaa, and M. Crucianu. “Active Relevance Feedback for Interactive Satellite Images retrieval”. In: *Image Information Mining - Theory and Application to Earth Observation*, ESRIN, Frascati, Italy, Oct. 2005.
- [Fere 05b] M. Ferecatu, N. Boujemaa, and M. Crucianu. “Hybrid visual and conceptual image representation within active relevance feedback context”. In: *ACM SIGMM International Workshop on Multimedia Information Retrieval*, Singapore, Nov. 2005.
- [Fisc 81] M. Fischler, J. Tenenbaum, and H. Wolf. “Detection of Roads and Linear Structures in Low-resolution Aerial Imagery Using a Multisource Knowledge Integration Technique”. *Computer Graphics and Image Processing*, Vol. 15, No. 3, pp. 201–223, 1981.
- [Free 61] H. Freeman. “On the Encoding of Arbitrary Geometric Configurations”. *IRE Trans. Electron. Comput.*, Vol. EC-10, pp. 260–268, June 1961.
- [Garz 02] A. Garzelli. “Wavelet-Based Fusion of Optical and SAR Image Data over Urban Area”. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, Vol. 34, No. 3/B, pp. 59–62, 2002.
- [Gema 96] D. Geman and B. Jedynek. “An active testing model for tracking roads in satellite images”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 1, pp. 1–14, 1996.
- [Gong 92] P. Gong, D. J. Marceau, and P. J. Howarth. “A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data”. *Remote Sensing of Environment*, Vol. 40, pp. 137–151, 1992.
- [Goue 01] V. Gouet and N. Boujemaa. “Object-based queries using color points of interest”. In: *IEEE Workshop on Content-Based Access of Image and Video Libraries*, Hawaii, USA, 2001.
-

- 
- [Groc 82] W. Groch. “Extraction of Line Shaped Objects from Aerial Images Using a Special Operator to Analyze the Profiles of Functions”. *Computer Graphics and Image Processing*, Vol. 18, No. 4, pp. 347–358, 1982.
- [Gros 94] M. Gross, L. Koch, R. Lippert, and A. Dreger. “Multiscale image texture analysis in wavelet spaces”. In: *Proc. IEEE International Conference on Image Processing (ICIP)*, Austin, Texas, USA, 1994.
- [Grue 95] A. Gruen and H. Li. “Road extraction from aerial and satellite images by dynamic programming”. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 50, No. 4, pp. 11–21, 1995.
- [Hans 00] U. Hans and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan-Kaufman, New York, NY, USA, 2000.
- [Hara 73] R. Haralick, K. Shanmugam, and I. Dinstein. “Texture features for image classification”. *IEEE Trans. Syst., Man, and Cybern.*, Vol. SMC-3, No. 6, 1973.
- [Hara 94] Y. Hara, R. Atkins, S. Yueh, R. Shin, and J. Kong. “Application of neural networks to radar image classification”. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 32, No. 1, pp. 100–109, 1994.
- [Howa 05] P. Howarth and S. Rüger. “Fractional Distance Measures for Content-Based Image Retrieval”. In: *27th European Conference on Information Retrieval*, pp. 447–456, Santiago de Compostela, Spain, 2005.
- [Hu 62] M. Hu. “Visual pattern recognition by moment invariants, computer methods in image analysis”. *IRE Transaction on Information Theory*, Vol. 8, 1962.
- [INRI] INRIA. “IMEDIA, INRIA Rocquencourt”.
- [Jain 98] R. Jain. “Multimedia Documents”. *IEEE MultiMedia*, Vol. 5, No. 3, p. 1, 1998.
- [Jeon 00] B. Jeon, J. Jang, and K. Hong. “Road Detection in SAR Images Using Genetic Algorithm with Region Growing Concept”. In: *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 688–691, 2000.
- [Kane 85] T. Kaneko and M. Okudaira. “Encoding of Arbitrary Curves Based on the Chain Code Representation”. *IEEE Trans. Communications*, Vol. 33, No. 7, pp. 697–707, July 1985.
- [Kanu 00] T. Kanungo, D. Mount, N. Netanyahu, P. Piatko, R. Silverman, and A. Wu. “The analysis of a simple k -means clustering algorithm”. In: *Symposium on Computational Geometry*, pp. 100–109, 2000.
-



- 
- [Khat 89] L. Khatir. *Recherche d'algorithmes de localisation de routes dans les images satellite à haute résolution. Application à l'imagerie SPOT*. PhD thesis, Université Paul Sabatier de Toulouse, France, 1989.
- [Kira 92] K. Kira and L. Rendell. "A practical approach to feature selection". In: *Proc. 9th Int. Work. on Machine Learning*, pp. 249–256, Aberdeen, Scotland, United Kingdom, 1992.
- [Krup 95] H. Krupnik, D. Malah, and E. Karnin. "Fractal Representation of Images via the Discrete Wavelet Transform". In: *IEEE 18th Conv. of EE in Israel, Tel-Aviv*, 1995.
- [Kund 92] A. Kundu and J.-L. Chen. "Texture classification using qmf bank-based subband decomposition". *CVGIP: Graphical Models and Image Processing*, Vol. 54, No. 5, pp. 369–384, 1992.
- [Laco 05] C. Lacoste, X. Descombes, and J. Zerubia. "Point Processes for Unsupervised Line Network Extraction in Remote Sensing". *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, pp. 1568–1579, 2005.
- [Lain 93] A. Laine and J. Fan. "Texture classification by wavelet packet signature". *IEEE Trans. Pattern Recognition and Machine Intelligence*, Vol. 15, No. 11, pp. 1186–1191, 1993.
- [Li 95] B. Li and S. Ma. "On the relation between region and contour representation". In: *Proc. IEEE International Conference on Image Processing (ICIP)*, Washington DC, USA, 1995.
- [Lipa 88] C. Lipari, M. Trivedi, and C. Harlow. "Representation and Recognition of Elongated Regions in Aerial Images". *SPIE*, Vol. 937, pp. 557–567, Apr. 1988.
- [Lore 00] A. Lorette. *Analyse de Texture par Méthodes Markoviennes et par Morphologie Mathématique : Application à l'Analyse des Zones Urbaines sur des Images Satellitales*. PhD thesis, Université de Nice Sophia Antipolis, Nice, Sep. 2000.
- [Lore 99] A. Lorette, X. Descombes, and J. Zerubia. "Texture Analysis Through Markov Random Fields: Urban Areas Extraction". In: *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 430–434, Kobe, Japan, 1999.
- [Lowe 04] D. Lowe. "Distinctive image features from scale-invariant keypoints". *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [Luo 01] B. Luo and E. R. Hancock. "Structural Graph Matching Using the EM Algorithm and Singular Value Decomposition". *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 10, pp. 1120–1136, Oct. 2001.
-

- 
- [Ma 95] W. Ma and B. Manjunath. “A comparison of wavelet transform features for texture image annotation”. In: *Proc. IEEE International Conference on Image Processing (ICIP)*, Washington DC, USA, 1995.
- [Mait 07] H. Maitre. “Indexing and retrieval in large satellite image databases”. In: *Proc. 5th Int. Symp. Multispectral Image Processing and Pattern Recognition (MIPPR)*, p. , 2007.
- [Marc 94] D. Marceau, P. Howarth, and D. Gratton. “Remote sensing and the measurement of geographical entities in a forest environment 2: The optimal spatial resolution”. *Remote Sensing of Environment*, Vol. 49, pp. 105–117, 1994.
- [Marr 80] D. Marr and E. Hildreth. “Theory of edge detection”. In: *Proc. of the Royal Society of London. Series B, Biological Sciences*, pp. 187–217, 1980.
- [Math 04] P. Mather. *Computer Processing of Remotely Sensed Images*. John Wiley and Sons Ltd., 2004.
- [Math 75] G. Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, New York, 1975.
- [Maup 97] P. Maupin, P. Le Quere, R. Desjardins, M.-C. Mouchot, B. St-Onge, and B. Solaiman. “Contribution of mathematical morphology and fuzzy logic to the detection of spatial changes in an urbanized area: towards a greater integration of image and geographical information systems”. In: *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 207–209, Singapore, 1997.
- [Maye 98] H. Mayer, A. Baumgartner, and C. Steger. “Road Extraction from Aerial Imagery”. In: *CV-Online: On-line Compendium of Computer Vision*, 1998.
- [McKe 88] D. McKeown and J. Denlinger. “Cooperative methods for road tracking in aerial imagery”. In: *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, Santa Barbara, CA, USA, 1988.
- [Meht 97] B. Mehtre, M. Kankanhalli, and W. Lee. “Shape measure for content based image retrieval: A comparison”. *Information Processing and Management*, Vol. 33, No. 3, 1997.
- [Merc 03] G. Mercier and M. Lennon. “Support Vector Machines for hyperspectral image classification with spectral-based kernels”. In: *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Toulouse, France, 2003.
- [Merg 85] Y. Merghoub. *Reconnaissance et analyse de formes sur des images de télédétection : Modélisation par la morphologie mathématique*. PhD thesis, Université Paul Sabatier de Toulouse, France, 1985.
-

- 
- [Merl 96] N. Merlet and J. Zerubia. “New Prospects in Line Detection by Dynamic Programming”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 4, pp. 426–431, 1996.
- [Naga 79] M. Nagao and T. Matsuyama. “Region extraction and shape analysis in aerial photographs”. *Comput. Graph. Image Process.*, Vol. 10, No. 3/B, pp. 195–223, 1979.
- [Neva 80] R. Nevatia and K. Babu. “Linear feature extraction and description”. *Computer Graphics and Image Processing*, Vol. 13, pp. 257–269, 1980.
- [Niem 80] H. Niemann. “Hierarchical Graphs in Pattern Analysis”. In: *Proc. International Conference on Pattern Recognition (ICPR)*, pp. 213–216, Miami, USA, 1980.
- [ODoc 91] M. O’Docherty and C. Daskalakis. “Multimedia Information Systems—The Management and Semantic Retrieval of all Electronic Data Types”. *The Computer Journal*, Vol. 34, No. 3, pp. 225–238, 1991.
- [Orte 97] M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, and T. Huang. “Supporting similarity queries in MARS”. In: *Proc. of the 5th ACM international conference on Multimedia*, pp. 403–413, New York, NY, USA, 1997.
- [Papa 98] N. Paparoditis, M. Cord, M. Jordan, and J. P. Cocquerez. “Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery”. *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 122–142, 1998.
- [Paru 91] S. Parui, B. Uma Shankar, A. Mukherjee, and D. Dutta Majumder. “A Parallel Algorithm for Detection of Linear Structures in Satellite Images”. *Physical Review Letters*, Vol. 12, pp. 765–770, 1991.
- [Pent 84] A. Pentland. “Fractal-based description of natural scenes”. *IEEE Trans. Pattern Recognition and Machine Intelligence*, Vol. 6, No. 6, pp. 661–674, 1984.
- [Pent 96] A. Pentland, R. Picard, and S. Sclaroff. “Photobook: Content-based Manipulation of Image Databases”. *International Journal of Computer Vision*, Vol. 18, No. 3, pp. 233–254, 1996.
- [Pers 77] E. Persoon and K. Fu. “Shape discrimination using Fourier descriptors”. *IEEE Trans. Syst., Man, and Cybern.*, Vol. 7, pp. 170–179, 1977.
- [Pesa 01] M. Pesaresi and J. Benediktsson. “A new approach for the morphological segmentation of high-resolution satellite imagery”. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 39, No. 2, pp. 309–320, 2001.
-

- 
- [Peur 01] M. Peura. “Flexible Heuristic Matching of Attribute Trees”. In: *Proc. of the 12th Scandinavian Conference on Image Analysis*, pp. 207–213, Bergen, Norway, 2001.
- [Pita 91] I. Pitas and A. Venetsanopoulos. “Morphological shape representation”. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2381–2384, Toronto, Ont., Canada, 1991.
- [Plat 99] J. Platt. “Probabilities for Support Vector Machines”. In: A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds., *Advances in Large Margin Classifiers*, pp. 61–74, MIT Press, 1999.
- [Rahm 05] M. M. Rahman, P. Bhattacharya, and B. Desai. “Similarity Searching in Image Retrieval with Statistical Distance Measures and Supervised Learning”. In: *International Conference on Advances in Pattern Recognition*, pp. 315–324, Bath, UK, 2005.
- [Rao 02] A. Rao, R. Srihari, L. Zhu, and A. Zhang. “A method for measuring the complexity of image databases”. *IEEE Trans. Multimedia*, Vol. 40, No. 2, pp. 160–173, 2002.
- [Rein 94] J. Reinhardt and W. Higgins. “Shape Representation: Comparison between the Morphological Skeleton and Morphological Shape Decomposition”. In: *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 91–95, Austin, Texas, USA, 1994.
- [Rell 02] G. Rellier, X. Descombes, J. Zerubia, and F. Falzon. “A Gauss-Markov Model for Hyperspectral Texture Analysis of Urban Areas”. *Proc. International Conference on Pattern Recognition (ICPR)*, Vol. 1, pp. 692–695, 2002.
- [Roch 03] M. Rochery, I. H. Jermyn, and J. Zerubia. “Higher Order Active Contours and their Application to the Detection of Line Networks in Satellite Imagery”. In: *Proc. IEEE Workshop Variational, Geometric and Level Set Methods in Computer Vision*, at ICCV, Nice, France, October 2003.
- [Roux 92] M. Roux. *Recalage d’images multi-sources. Application au recalage d’une image SPOT et d’une carte*. PhD thesis, École Nationale Supérieure des Télécommunications, Paris, France, Sep. 1992.
- [Rui 96] Y. Rui, A. She, and T. Huang. “Modified Fourier descriptors for shape representation – a practical approach”. In: *Proc of First International Workshop on Image Databases and Multi Media Search*, 1996.
- [Sahb 02a] H. Sahbi and N. Boujemaa. “Coarse to fine face detection based on skin color adaptation”. In: *Biometric Authentication, Proc. of International ECCV 2002 Workshop*, pp. 112–120, Copenhagen, Denmark, 2002.
-

- 
- [Sahb 02b] H. Sahbi and N. Boujemaa. “Robust Face Recognition Using Dynamic Space Warping”. In: *Biometric Authentication, Proc. of International ECCV 2002 Workshop*, pp. 121–132, Copenhagen, Denmark, 2002.
- [Sams 01] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. “Two Variational Models for Multispectral Image Classification.”. In: *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 344–358, 2001.
- [Scha 92] R. Schalkoff. *Pattern Recognition : Statistical, Structural, and Neural Approaches*. John Wiley and Sons Inc., 1992.
- [Schr 98] M. Schroeder, H. Rehrauer, K. Seidel, and M. Datcu. “Spatial Information Retrieval from Remote Sensing Images: Part II. Gibbs Markov Random Fields”. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 36, pp. 1446–1455, Sep. 1998.
- [Sere 89] M. Serendero. *Extraction d’informations symboliques en imagerie SPOT: réseaux de communication et agglomérations*. PhD thesis, Université de Nice, France, 1989.
- [Serr 82] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.
- [Sidd 02] K. Siddiqi, S. Bouix, A. Tannenbaum, and S. W. Zucker. “Hamilton-jacobi skeleton”. *International Journal of Computer Vision*, Vol. 48(3), pp. 215–231, 2002.
- [Smit 94] J. Smith and S. Chang. “Transform features for texture classification and discrimination in large image database”. In: *Proc. IEEE International Conference on Image Processing (ICIP)*, Austin, Texas, USA, 1994.
- [Stoi 00] R. Stoica, X. Descombes, and J. Zerubia. “Road extraction in remote sensed images using a stochastic geometry framework”. In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Gif sur Yvette, France, July 2000.
- [Tamu 78] H. Tamura, S. Mori, and T. Yamawaki. “Texture features corresponding to visual perception”. *IEEE Trans. Syst., Man, and Cybern.*, Vol. SMC-8, No. 6, 1978.
- [Tann 05] C. Tannier and D. Pumain. “Fractals in urban geography : a theoretical outline and an empirical example”. *Cybergeo, Systèmes, Modélisation, Géostatistiques*, No. 307, 2005.
- [Thya 94] K. Thyagarajan and T. Nguyen. “A maximum likelihood approach to texture classification using wavelet transform”. In: *Proc. IEEE International Conference on Image Processing (ICIP)*, Austin, Texas, USA, 1994.
-

- 
- [Tilt 00] J. Tilton, M. Giovanni, and M. Datcu. “Knowledge Discovery and Data Mining Based on Hierarchical Segmentation of Image Data”. 2000. A research proposal submitted in response to NRA2-37143 from NASA’s Information Systems Program.
- [Tupi 98] F. Tupin, H. Maitre, J. Mangin, J. Nicolas, and E. Pechersky. “Detection Of Linear Features In SAR Images: Application To Road Network Extraction”. *IEEE Trans. Geoscience and Remote Sensing*, Vol. 36, No. 2, pp. 434–453, March 1998.
- [Vapn 97] V. Vapnik. “The Support Vector Method”. In: *International Conference on Artificial Neural Networks*, pp. 263–271, Lausanne, Switzerland, 1997.
- [Vive 03] O. Viveros-Cancino. *Analyse Des Zones Urbaines Par Fusion de Données En Télédétection*. PhD thesis, University of Nice Sophia Antipolis, France, 2003.
- [Wang 92] J. Wang, P. Treitz, and P. Howarth. “Network detection from SPOT imagery for updating geographical information systems in the rural- urban fringe”. *Int. J. Geographical Information Systems*, Vol. 6, No. 2, pp. 141–157, 1992.
- [Welc 82] R. Welch. “Spatial resolution requirements for urban studies”. *International Journal of Remote Sensing*, Vol. 3, No. 2, pp. 139–146, 1982.
- [Wesz 76] J. Weszka, C. Dyer, and A. Rosenfeld. “A Comparative Study of Texture Measures for Terrain Classification”. *IEEE Trans. Syst., Man, and Cybern.*, Vol. 6, No. 4, pp. 269–286, 1976.
- [Wils 97] R. C. Wilson and E. R. Hancock. “Structural Matching by Discrete Relaxation”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 6, pp. 634–648, 1997.
- [Yang 06] Y. Yang, H. Sun, and Y. Cao. “Unsupervised urban area extraction from SAR imagery using GMRF”. *Journal of Pattern Recognition and Image Analysis*, Vol. 16, No. 1, pp. 116–119, 2006.
- [Yu 95] H. Yu and W. Wolf. “Scenic classification methods for image and video databases”. In: *Proc. SPIE Int’l Conf. on Digital Image Storage and archiving systems*, pp. 363–371, Philadelphia, PA, USA, 1995.
- [Yu 99] S. Yu. “Urban area detection in SPOT images using multi-scale technique and map knowledge”. In: *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 62–64, Hamburg, Germany, 1999.
- [Yuil 00] A. Yuille and J. Coughlan. “Fundamental limits of Bayesian inference: Order parameters and phase transitions for road tracking”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 2, pp. 160–173, 2000.
-

- [Zhan 03] Q. Zhang, J. Wang, P. Gong, and P. Shi. "Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis". *International Journal of Remote Sensing*, Vol. 24, No. 21, pp. 4137–160, 2003.
-