INDEXING SATELLITE IMAGES WITH FEATURES COMPUTED FROM MAN-MADE STRUCTURES ON THE EARTH'S SURFACE

A. Bhattacharya[1,2*], *M. Roux*[1], *H. Maitre*[1]

[1]GET, Telecom Paris
46, rue Barrault
75013 Paris, France

I. H. Jermyn[2], *X. Descombes*[2], *J. Zerubia*[2]

[2]Ariana (joint research group INRIA/I3S)
INRIA, BP 93, 06902
Sophia Antipolis, Cedex France

ABSTRACT

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 this paper, we define several such environments, and classify them with the aid of geometrical and topological features computed from the road networks occurring in them. The relative failure of network extraction methods in certain types of urban area obliges us to segment such areas and to add a second set of geometrical and topological features computed from the segmentations. To validate the approach, feature selection and SVM linear kernel classification are performed on the feature set arising from a diverse image database.

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 [3]. 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. Image-based query characterizations are 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. 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 [9, 7], 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 [1] looked at the classification of a small image database into two classes, 'Urban' and 'Rural', using a small set of topological and geometrical road network features and kernel k-means. The study indicated that the idea had potential. The purpose of this paper is to describe further studies examining the classification of a much larger database into five classes ('Urban USA', 'Urban Europe', 'Mountains', 'Villages', 'Fields', shown in figure 1). Initially, the feature set was simply a larger set of road network features, but 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 obtain useful information from these parts of the images and improve the classification, a new set of features based on segmented urban areas was therefore introduced, and combined with the existing road network features. In order to reduce the dimensionality of the feature space, a suitable feature selection scheme was also tested, which, in combination with SVM linear kernel classification [2] on the combined set of features from road networks and urban areas, gave promising results for the classification of the different geographical

*This work was partially funded by the French Space Agency (CNES), ACI QuerySat, the STIC INRIA-Tunisia programme, and EU NoE Muscle (FP6-507752). The data was kindly provided by CNES and by Sup'com, Tunis. The work of the first author is supported by an INRIA PhD fellowship.

environments contained in a diverse image database.

In section 2 we describe the two network extraction methods considered in this study. We also describe the road network representation into which we convert the outputs of these methods before computing the features. In section 2.2 we describe the set of road network features introduced to classify the database into its classes. In section 3 we describe the features computed from segmented urban areas. In section 4 we describe the results of a number of classification experiments using our augmented feature set and an SVM linear classifier. In section 5 we conclude.

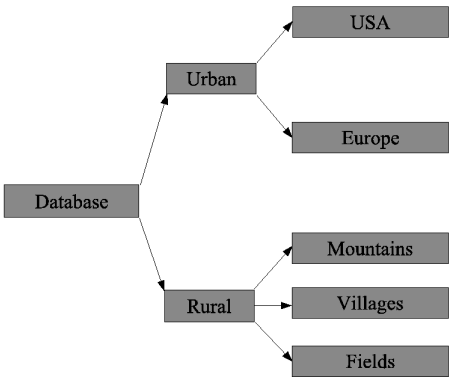


Fig. 1. Images categorized into five different classes.

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.

2.1. Extraction methods

In the present work, we consider two network extraction methods reported in [6, 4]. The output of the method described in [6] is a binary image, which after a distance function computation can serve as an input to our problem. Figure 3(b) shows an example of the extracted network.

The output of the method described in [4] is a list of multiply aligned segments. In order to have a suitable input for our problem, we convert the output of this method into a binary image and use some image processing techniques to obtain single connected segments. We then compute a distance function. Figure 3(d) shows an example of the extracted network.

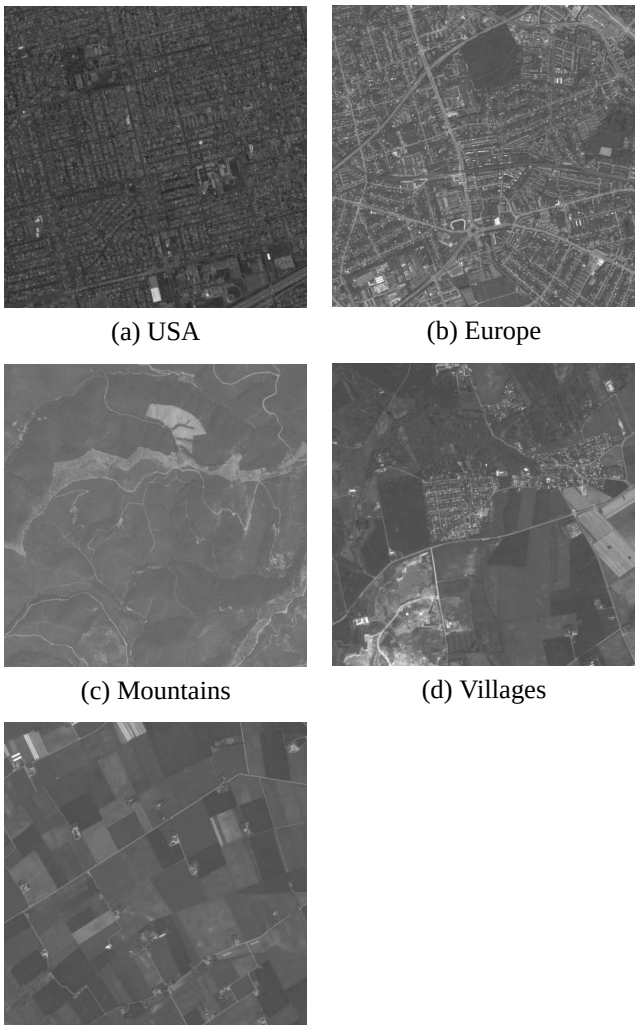
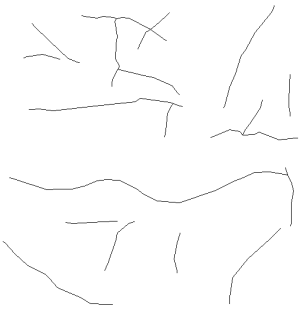


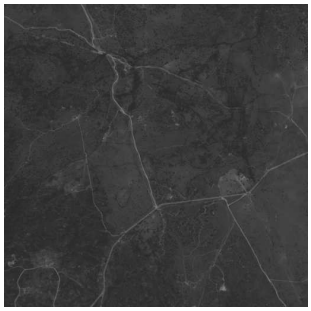
Fig. 2. An example of 2 urban and 3 rural classes ©CNES.

The distance function resulting from these methods is converted to a graph representation of the road network for feature computation purposes. The graph itself captures the network topology, while the network geometry is encoded by decorating the vertices and edges with geometrical information. The conversion is performed by computing the shock locus of the distance function using the method of [5, 8], extended to deal with multiple, multiply-connected components. The method identifies the shock points by finding out the limiting behaviour 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

(a) Original image ©CNES



(b) Extracted network.



(c) Original image ©CNES

(d) Extracted network.

Fig. 3. Example results from the two extraction methods: (b) method of [6], and (d) method of [4].

are composed of all other points, and correspond to road segments between junctions and terminals. Figure 4 shows an example of the representation graph. The road network (top right) is first extracted from the input image (top left). The methods cited above is then used to generate the shock locus (bottom left), which is then converted to the graph representation (bottom right). The vertices and edges are decorated with geometrical quantities computed from the shock locus. The features are then computed from the graph and its decorations. The features computed from the graph representation are described in section 2.2.

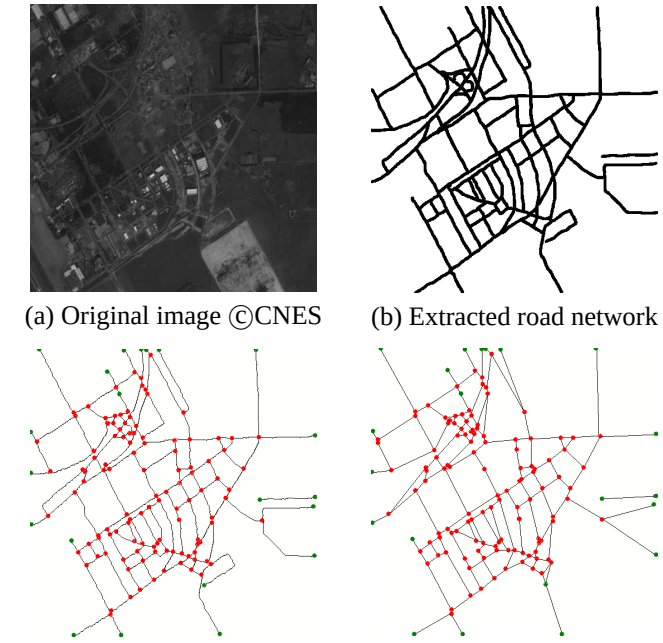
2.2. Features from the graph

In this section we focus on 15 features summarized in table 1. These features can be categorized into six groups: five 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 now define the road network features.

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 and $E_J = \sum_{m_v>2} m_v$ is the number of junction

Notation	Description
m	Number of edges in graph
n	Number of vertices in graph
Ω	Area of image
Ω_L	Network area
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 $\sum_{v:m_v>2} 1$
$\tilde{N}_J$	**Junction density** $\Omega^{-1} N_J$
L	**Network length** $\sum_e l_e$
$\tilde{L}$	**Length density** $\Omega^{-1} \sum_e l_e$
$\tilde{A}$	**Network area density** $\Omega^{-1}\Omega_L$
d_e	Euclidean distance between vertices in an edge
p_e	Ratio of lengths l_e/d_e
var(p)	**Ratio of lengths variance** $m^{-1}\sum_e p_e^2 - (m^{-1}\sum_e p_e)^2$
mean(p)	**Ratio of lengths mean** $m^{-1}\sum_e p_e$
k_e	Average curvature $l_e^{-1}\int_e ds\,\lvert k_e(s)\rvert$
var(k)	**Average curvature variance** $m^{-1}\sum_e k_e^2 - (m^{-1}\sum_e k_e)^2$
mean(k)	**Average curvature mean** $m^{-1}\sum_e k_e$
$E_{D,i}$	Proportion of junctions with $m_v = i$
var$(E_{D,i})$	**Variance of edge distribution** $(1/\max(m_v))\sum_i E_{D,i}^2 - ((1/\max(m_v))\sum_i E_{D,i})^2$
mean$(E_{D,i})$	**Mean of edge distribution** $(1/\max(m_v))\sum_i E_{D,i}$
E_J	Number of junction edges $\sum_{m_v>2} m_v$
$M_{J,a}$	Number of junction edges per quadrant $\sum_{v\in a:m_v>2} m_v$
$\tilde{E}_J$	**Density of junction edges** $\Omega^{-1}E_J$
$\tilde{M}_{J,a}$	Density of junction edges per quadrant $\Omega_a^{-1}M_{J,a}$
var$(\tilde{M}_J)$	**Variance of density of junction edges** $(1/4)\sum_a \tilde{M}_{J,a}^2 - ((1/4)\sum_a \tilde{M}_{J,a})^2$
mean$(\tilde{M}_J)$	**Mean of density of junction edges** $(1/4)\sum_a \tilde{M}_{J,a}$
Ω_r	Area of a circular region of r
$\tilde{N}_{j,r}$	Junction density in a circular region $\Omega_{j,r}^{-1}\sum_{v\in\Omega_{j,r}:m_v>2} 1$
$max_j\{\tilde{N}_{j,r}\}$	**Maximum of the junction densities** $max_j\{\tilde{N}_{j,r}\}$
β_j	Vector of angles between segments at junction j
H_β	**Entropy of histogram of road segment angles with bin size** 30°

Tab. 1. A summary of the features computed from road networks.



(a) Original image ©CNES (b) Extracted road network

(c) Shock locus of road network (d) Graph representation

Fig. 4. An example of the graph representation.

edges. Let Ω be the area of the image in pixels. We define the 'junction density' to be $\tilde{N}_J = \Omega^{-1} N_J$ and 'density of junction edges' to be $\tilde{E}_J = \Omega^{-1} E_J$. These are intuitively a useful measure to separate urban and rural areas: we expect urban areas to have a higher value of $\tilde{N}_J$ and $\tilde{E}_J$ than rural areas. Similarly, we define the 'network length' $L = \sum_e l_e$ and 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 than 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'. We also compute the network area Ω_L as the number of pixels corresponding to the network from the extracted binary image and define the 'area density' as $\tilde{A} = \Omega^{-1}\Omega_L$. As can be seen in figure 4, many junction points are clustered around a small area in the network. To obtain a local characteristic of the junction density, we define a measure called 'junction density in a circular region' $\tilde{N}_{j,r} = \Omega_{j,r}^{-1} \sum_{v \in \Omega_{j,r} : m_v > 2} 1$. This is the density of junction points falling in a circular region $\Omega_{j,r}$ of radius r centred at a junction point. We then compute the maximum of these junction densities over all junction points, $max_j\{\tilde{N}_{j,r}\}$. A high value indicates that junction points are clustered close to many other junction points, which is a prominent measure of urban network structure. Rural areas will have a lower value of this feature, indicating the sparse structure of road junctions.

Let $p_e = l_e/d_e$, and $k_e = l_e^{-1} \int_e ds\, |k_e(s)|$, *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 the variances of these quantities are equally useful. We thus define the 'ratio of lengths variance' and the 'ratio of lengths mean' to be the variance and mean of p_e over edges, var(p) and mean(p), and the 'average curvature variance' and 'average curvature mean' to be the variance and mean of k_e over edges, var(k) and mean(k). Note that it is quite possible to have a large value of p_e for an edge while having a small value of 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, var$(\tilde{M}_J)$. We also include mean$(\tilde{M}_J)$ as a feature.

In order to distinguish between the two urban classes (USA and Europe), the entropy of the histogram of angles at junctions, H_β, where β_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 USA tend to be parallel and cross each other orthogonally forming T-junctions or crossroads, whereas European roads tend to wiggle and meet or cross each other at roundabouts. 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 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.

3. EXTRACTION AND CHARACTERIZATION OF URBAN REGION

Classification experiments show that the above features are not sufficient for images that contain a significant proportion of small urban areas. This is because the extraction methods frequently fail to extract the dense road network structures in these areas. Some example images are shown in figure 5(a) and figure 5(c). In order to circumvent this problem and to extract useful information from these parts of the image, we instead segment the urban area itself, and then compute some geometrical features of the resulting region. These features will be combined with the road network features described above for classification purposes.

3.1. Extraction of urban region

We use a sequence of morphological operators to extract the region of interest from the image. A difference is computed between a morphologically closed and opened image. This difference gives prominence to textured regions, like urban areas. Then an alternated sequential filter aggregates neighbouring components and eliminates small isolated components. We compute two geometrical features from these regions as shown in table 2.

(a) Original image ©CNES (b) Segmented region

(c) Original image ©CNES (d) Segmented region

Fig. 5. Images containing small urban areas and their segmentations.

Notation	Description
Ω	Area of image
Ω_R	Area of extracted regions
Γ_R	Perimeter of extracted regions
$\tilde{R}_A$	**Region area density** $\Omega^{-1}\Omega_R$
Cf_A	**Region compactness factor** $\Omega_R^{-1}\Gamma_R^2$

Tab. 2. Summary of features computed for urban areas.

3.2. Features from the region

We focus on the last two features in table 2. These two features enable us to distinguish between Villages and Fields classes, which otherwise were misclassified due to the lack of extracted network information from the small compact urban region in the images, shown in figure 5(a) and figure 5(c). Let Ω and Ω_R be the area of the image and the area of the extracted regions respectively and Γ_R be the perimeter of the extracted regions. 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.

4. FEATURE SELECTION AND CLASSIFICATION

All the images in our database 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.

The features described in the above sections were computed for a database of 355 SPOT5, 5m resolution images. To provide ground truth, these images were hand classified into the five 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.

The results of SVM linear kernel classification of 355 images into 5 classes, using 30 features, is shown in table 3 (15 features each from the graph for 2 network extraction methods). There is a mean error of 36.1% with standard de-

	Class 1	Class 2	Class 3	Class 4	Class 5
Villages	**0.527**	0.094	0.245	0.055	0.131
Mountains	0.048	**0.805**	0.000	0.015	0.059
Fields	0.218	0.000	**0.593**	0.063	0.129
USA	0.065	0.020	0.046	**0.771**	0.144
Europe	0.140	0.086	0.117	0.102	**0.536**

Tab. 3. Confusion matrix of an SVM linear kernel classification of 355 images into 5 classes with 30 features

	Class 1	Class 2	Class 3	Class 4	Class 5
Villages	**0.726**	0.047	0.151	0.030	0.055
Mountains	0.035	**0.876**	0.027	0.000	0.000
Fields	0.142	0.018	**0.822**	0.018	0.025
USA	0.035	0.000	0.000	**0.818**	0.137
Europe	0.069	0.058	0.000	0.135	**0.783**

Tab. 4. Confusion matrix of an SVM linear kernel classification of 355 images into 5 classes with 32 features.

viation of 8.49%. 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 Urban USA and Urban Europe classes. These confusions arise because, as stated above, the road extraction methods fail to detect the fine and densely structured roads present in some images. Table 4 shows the results of classification of the same set of images, this time with 32 features: 30 road network features plus the two features computed from the segmented urban areas. As can be seen, there is an improvement in the confusion matrix. The Villages class is less confused with the Fields class than before. The SVM linear kernel classification in this case gives us a mean error of 20.3% with a standard deviation of 7.75%.

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, we performed feature selection using a Fisher linear discriminant (FLD) analysis, followed by SVM linear kernel classification on the selected feature set. The results of classification are shown in table 5. The SVM linear kernel classification on the 15-dimensional feature space selected by the FLD shows a mean error of 17.5% with a standard deviation of 3.81%. An overall classification performance summary is depicted in table 6, where classification error in % is given as "mean $\pm$ standard deviation" error.

	Class 1	Class 2	Class 3	Class 4	Class 5
Villages	**0.751**	0.051	0.139	0.011	0.059
Mountains	0.034	**0.896**	0.014	0.000	0.000
Fields	0.074	0.015	**0.826**	0.012	0.000
USA	0.028	0.000	0.000	**0.897**	0.189
Europe	0.112	0.037	0.022	0.080	**0.752**

Tab. 5. Confusion matrix of an SVM linear kernel classification of 355 images into 5 classes with 15 features selected by FLD.

Feature Dimension	Selection	Classification Error (%)
30	No	36.1$\pm$8.49
32	No	20.3$\pm$7.75
32	Fisher	17.5$\pm$3.81

Tab. 6. Classification performance.

5. CONCLUSION

The classification results reported above indicate that geometrical and topological features computed from road networks and urban areas can serve as robust characterizations of a number of geographical environments found in remote sensing images. Future work will involve a two-level hierarchical model with three top-level categories ('Urban', 'Semi-Urban' and 'Non-Urban'), with several subcategories within each main category ('Urban/USA', 'Urban/Europe', 'Urban/Asia', 'Semi-Urban/Sparse Regions', 'Semi-Urban/-Non-Sparse Regions', 'Non-Urban/Mountains', 'Non-Urban-/Fields'); computing feature statistics; and experimenting with different classifiers to improve the classification and hence retrieval results. Future work will also involve a construction of a larger database with few additional classes.

6. REFERENCES

[1] 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.

[2] M. Campedel and E. Moulines. Classification et sélection automatique de caractéristiques de textures.

Revue des Nouvelles Technologies de l'Information, pages 25–37, 2005.

[3] H. Daschiel and M. Datcu. Image information mining system evaluation using information-theoretic measures. *EURASIP Journal on Applied Signal Processing*, 14:2153–2163, 2005.

[4] A. Desolneux, L. Maison, and J-M. Morel. Meaningful alignments. *International Journal of Computer Vision*, 40(1):7–23, 2000.

[5] P. Dimitrov, C. Phillips, and K. Siddiqi. Robust and efficient skeletal graphs. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1417–1423, Hilton Head Island, USA, 2000.

[6] 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*, 15(3):201–223, 1981.

[7] B. Luo and E. R. Hancock. Structural graph matching using the EM algorithm and singular value decomposition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(10):1120–1136, October 2001.

[8] K. Siddiqi, S. Bouix, A. Tannenbaum, and S. W. Zucker. Hamilton-jacobi skeleton. *International Journal of Computer Vision*, 48(3):215–231, 2002.

[9] R. C. Wilson and E. R. Hancock. Structural matching by discrete relaxation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(6):634–648, 1997.