Computing statistics from a graph representation of road networks in satellite images for indexing and retrieval*

Avik Bhattacharya & Ian H. Jermyn & Xavier Descombes & Josiane Zerubia
Ariana (joint research group INRIA/I3S), INRIA, BP 93, 06902 Sophia Antipolis, Cedex France
Email: FirstName.LastName@inria.fr

Retrieval from remote sensing image archives relies on the extraction of pertinent information from the data about the entity of interest (*e.g.* land cover type), and on the robustness of this extraction to nuisance variables (*e.g.* illumination). Most image-based characterizations are not invariant to such variables. However, other semantic entities in the image may be strongly correlated with the entity of interest and their properties can therefore be used to characterize this entity. Road networks are one example: their properties vary considerably, for example, from urban to rural areas. This paper takes the first steps towards classification (and hence retrieval) based on this idea. We study the dependence of a number of network features on the class of the image ('urban' or 'rural'). The chosen features include measures of the network density, connectedness, and 'curviness'. The feature distributions of the two classes are well separated in feature space, thus providing a basis for retrieval. Classification using kernel k-means confirms this conclusion.

1 INTRODUCTION

The retrieval of images from large remote sensing image archives relies on the ability to extract pertinent information from the data, and on the robustness of this extraction (Daschiel and Datcu 2005). In particular, most queries will not concern, for example, the atmospheric conditions or illumination present when the images were acquired, but instead information that is invariant to these quantities, for instance land cover type. Most image-based characterizations are, however, far from invariant to changes in such nuisance variables, and thus fail to be robust when dealing with a large variety of images acquired under different conditions. Characterizations based on semantic entities detected in the scene, in contrast, are invariant to such changes, and inferences based on such entities can thus be used to retrieve images in a robust way. For this to work, the properties of the entity in question must be strongly dependent on the query. Road networks provide one example of such an entity: their topological and geometrical properties vary considerably, for example, from urban to rural areas. Features computed from an extracted road network can therefore in principle be used to characterize images, or parts of images, as belonging to one of these classes. Our work differs from much previous work, for example (Wilson and Hancock 1997; Luo and Hancock 2001), in that we are not interested in identifying the same network in different images, or in a map and an image, and producing a detailed correspondence, but rather in using more general network properties to characterize other entities in an image. The work to be described in this paper takes the first steps towards classification (and hence retrieval) based on such inter-semantic dependencies. Specifically, we study the dependence of a number of road network features on the class of the image, which for the moment we restrict to be either 'urban' or 'rural'. The chosen features when taken in isolation measure the density of the network and its 'curviness' in various ways. Taken together they also measure its connectedness. We find that the feature distributions of the two classes are well separated in feature space, and thus provide a basis for retrieval given an appropriate feature metric. Classification based on kernel k-means confirms that this is the case. We also discover an unexpected relation between two of the features that is consistent across images and classes.

*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.

2 NETWORK EXTRACTION AND REPRESENTATION

In order to compute topological and geometrical features of the network, we first need to extract the road network from an 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. We consider two network extraction methods (Lacoste et al. 2005; Rochery et al. 2006). The method of (Rochery et al. 2006) is based on 'higher-order active contours'. 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, in this case that 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. The method of (Lacoste et al. 2005) models the line network as an object process, where the objects are interacting line segments. The output of (Lacoste et al. 2005) is a set of line segments of varying length, orientation, and position. This output is converted to the output of (Rochery et al. 2006) by performing a dilation and then a distance function computation on the resulting binary image.

The distance function resulting from these methods is then 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 vertices and edges with geometric information. The conversion is performed by computing the shock locus of the distance function using the method of (Dimitrov et al. 2000; Siddiqi et al. 2002), extended to deal with multiple, multiply-connected, components. The method identifies shock points by examining the limiting behaviour of the average outward flux of the distance function as the region enclosing the shock point shrinks to zero. A threshold on this flux yields an approximation to the shock locus. The graph is then constructed by taking triple points and end points as vertices, corresponding to junctions and termini, while the edges are composed of all other points, and correspond to road segments between junctions and termini. Figure 1 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 are 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 geometric quantities computed from the shock locus. The features are then computed from the graph and its dec-

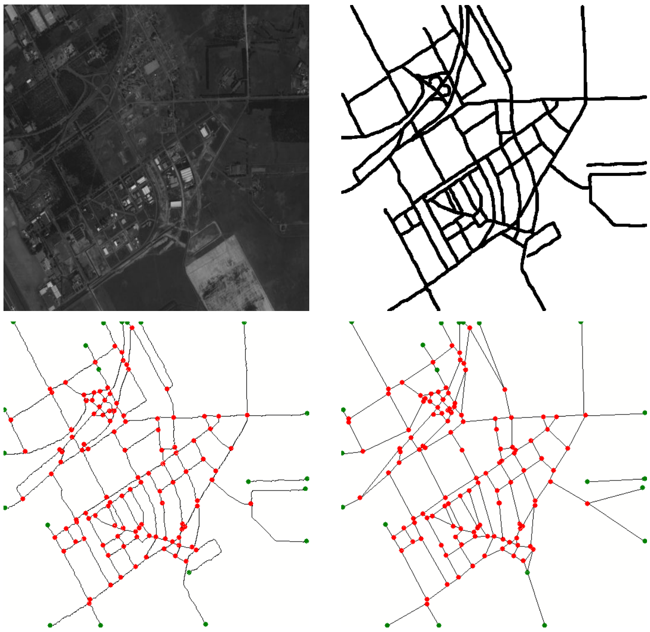

Figure 1: An example of a graph representation. Top left: original image ©CNES; top right: extracted road network; bottom left: shock locus of road network; graph representation.

Notation	Description
m	Number of edges in graph
Ω	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 $\sum_{v:\,m_v>2} 1$
$\tilde{N}_J$	Junction density $\Omega^{-1} N_J$
$\tilde{L}$	Length density $\Omega^{-1}\sum_e l_e$
d_e	Euclidean distance between vertices in an edge
p_e	Ratio of lengths l_e/d_e
var(p)	Ratio of lengths variance $m^{-1}\sum_e p_e^2 - (m^{-1}\sum_e p_e)^2$
k_e	Average curvature $l_e^{-1}\int_e ds\, \lvert k_e(s)\rvert$
var(k)	Average curvature variance $m^{-1}\sum_e k_e^2 - (m^{-1}\sum_e k_e)^2$
$M_{J,a}$	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($\tilde{M}_J$)	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 1: Summary of features

orations. They are described in the next section.

2.1 Feature vectors

We focus on five features, summarized in table 1. 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 Ω 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 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'.

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 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, var(p), and the 'average curvature variance' to be the variance of k_e over edges, 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 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$(\tilde{M}_J)$.

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 2 shows scatter plots of selected pairs of the features described above as computed from a database of 52 SPOT5, 5m resolution images, 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 behaviour 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 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

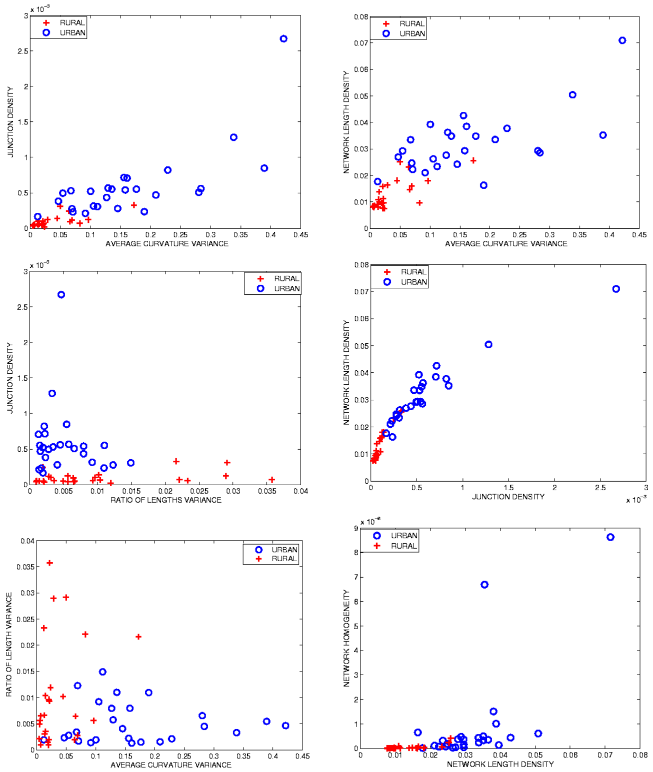

Figure 2: Scatter plots of selected pairs of features. Red stars correspond to rural areas, blue circles to urban areas. From left to right, top to bottom: $\tilde{N}_J$ versus var(k); $\tilde{L}$ versus var(k); $\tilde{N}_J$ versus var(p); $\tilde{L}$ versus $\tilde{N}_J$; var(p) versus var(k); var$(\tilde{M}_J)$ versus $\tilde{L}$.

no reason *a priori* why even the exponent $1/2$ should be consistent across images and classes, let alone the pre-factor. 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 CLUSTERING

The above results indicate that the selected features represent suitable choices for classification based on road network properties given an appropriate feature metric. In order to classify the images from the two classes, we use the kernel k-means algorithm since the feature data are not linearly separable. We use a Gaussian kernel,

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

where X_1 and X_2 are two feature vectors. The clustering result, displayed in table 2, shows that the two classes can be well partitioned using the above five features. 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.

	Urban	Rural
Class 1	1	19
Class 2	25	7

Table 2: Kernel k-means clustering result with $\sigma = 0.5$.

4 CONCLUSION

The preliminary studies reported above indicate that certain features of road networks can serve as characterizations for various image classes that are robust to nuisance parameters and in principle also to imaging modality. Future work will involve extracting road networks from many more images and studying the features described above and others (*e.g.* area density, number and angles of roads at junctions, network connectivity) on the resulting database of graphs. Probabilistic models of selected features will be developed, both for the above coarse classes, and refinements of them. These models will then enable classification and retrieval based on road networks.

REFERENCES

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

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

Lacoste, C., X. Descombes, and J. Zerubia (2005). Point processes for unsupervised line network extraction in remote sensing. *IEEE Trans. Pattern Analysis and Machine Intelligence 27*(10), 1568–1579.

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

Rochery, M., I. H. Jermyn, and J. Zerubia (2006). Higher-order active contours. *International Journal of Computer Vision 69*(1), 27–42.

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

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