Urban road extraction from VHR images using a multiscale approach and a phase field model of network geometry

Ting Peng[1,2], Ian H. Jermyn[1], Véronique Prinet[2], Josiane Zerubia[1], BaoGang Hu[2]
[1]Ariana (joint research group INRIA/I3S), INRIA, B.P. 93, 06902 Sophia Antipolis Cedex, France
email: firstname.lastname@sophia.inria.fr
[2]LIAMA, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China
email: prinet@nlpr.ia.ac.cn

***Abstract*—This paper addresses the problem of automatically extracting the main road network in a dense urban area from a very high resolution optical satellite image using a variational approach. The model energy has two parts: a phase field higher-order active contour energy that describes our prior knowledge of road network geometry, *i.e.* that it is composed of elongated structures with roughly parallel borders that meet at junctions; and a multi-scale statistical image model describing the image we expect to see given a road network. By minimizing the model energy, an estimate of the road network is obtained. Promising results on 0.6m QuickBird Panchromatic images are presented, and future improvements to the models are outlined.**

I. Introduction

The commercial availability of sub-metric resolution optical satellite images (QuickBird, Ikonos, and Pléiades in the near future) provides new opportunities for the extraction of information from remote sensing data: qualitatively new categories of information are available, and the accuracy of previously extracted categories of information can be quantitatively improved. For example, road networks can be extracted as two-dimensional regions rather than as one-dimensional structures, and the geometric accuracy of the extracted road network can be greatly improved. Higher resolution brings with it new challenges however. The appearance of details invisible in lower resolution images, *e.g.* cars, road markings, shadows, and other linear but non-road features, can easily disrupt the recognition process, and demands more sophisticated modelling, both of the image and of the road network. For the former, the existence of phenomena at multiple scales suggests a multi-scale approach, while for the latter, the incorporation into the models of our prior knowledge of the geometry of the road network becomes critical.

There is a large body of work on extracting road networks [11] from images with resolutions significantly lower than 1m/pixel, particularly from rural or peri-urban sites where the network is readily visible, with less shade and occlusion artefacts than in inner cities. There is less work on VHR images of urban environments, where the problem is made more difficult by the complexity of the images. In [12], roads and junctions are extracted in two steps using two different types of active contours, given a topologically correct graph of the network. A multi-scale image model is used. In [15], a conventional multi-spectral classification of a Pan-Sharpened QuickBird image is combined with edge information from the Panchromatic image to eliminate misclassified objects. In [5], LIDAR data is used in conjunction with VHR imagery. Detection of road primitives is followed by an iterative Hough transform to select candidate road strips. A validation procedure and topological analysis are finally used to form the road network. In [7], good results are obtained, but using a complicated series of processing steps. [1] works on sub-metric SAR images. The algorithm is based on a Hough transform that localizes straight parts of roads in the scene, and a tracking algorithm that extracts the full road network. The result is improved by using knowledge of the road network context to complete the extraction in disturbed areas.

In the present work, we aim to extract the region in the image domain containing the main road network in dense urban areas from a single VHR QuickBird Panchromatic image (0.61m/pixel). We want to do so automatically (except for parameter estimation, which is a topic for future research). We work with a unified, analysable model of the road network and the image, rather than a series of processing steps. We combine a 'phase field higher-order active contour' model of the geometry of the road network, with a region-based image model. We describe this model in section II. We then introduce a multi-scale image model in section III. The optimization framework is discussed in section IV. In section V, we present results obtained using both the single-scale model at various resolutions, and the multi-scale model, and make a comparison between them showing the benefits of the multi-scale analysis. We conclude in section VI.

II. The model: prior and data energies

We model the region containing the road network R using the phase field formulation of HOACs introduced by [14]. Let ϕ be a phase field level set function in the image domain Ω.

1-4241-4244-0712-5/07/$20.00

We define our model energy in the following form:

$$E(\phi) = \theta E_P(\phi) + E_D(I, \phi) \ ,$$

where θ is a constant that balances the contribution of the prior energy $E_P(\phi)$ and the data energy $E_D(I, \phi)$.

A. Prior Energy

Conventional active contours [2]–[4], [6], [10] are defined by linear functionals, expressible as single integrals. In contrast, HOACs [13] are defined by polynomial functionals. Expressed as multiple integrals, they describe non-local interactions between points in the region boundary ∂R. Via these long-range interactions, HOACs allow the inclusion of complex prior geometrical constraints. For this reason, HOACs are more robust to noise than conventional active contours, and permit a generic initialization that renders them automatic.

Phase fields model the region R using a level set function ϕ defined over the entire image, rather than using its bounding contour. Phase fields differ from most level-set methods in that ϕ is not constrained to be a distance function. Phase fields have several advantages over parametric active contours or standard level sets: a linear representation space; ease of implementation; and a neutral initialization. In addition, they allow greater topological freedom, which is critical when the topology of the region is not known *a priori*. In the context of the present application, the objects to be segmented, *i.e.* road networks, may have several connected components and many loops. Dealing with this topological complexity is arguably one of the most difficult aspects of automatic road network extraction; phase fields handle it "naturally". *Phase field HOACs* are phase field models that also include the long-range interactions characteristic of HOACs.

The prior energy E_P is the same as the one used by [14]:

$$E_P(\phi) = \int_{x\in\Omega} \Big\{ \frac{D}{2} \nabla\phi(x) \cdot \nabla\phi(x) + V(\phi(x)) \Big\} - \frac{\beta}{2} \iint_{(x,x')\in\Omega^2} \nabla\phi(x) \cdot \nabla\phi(x') \, \Psi(|x - x'|) \ . \quad (1)$$

where x is a point in the image domain Ω and $D, \beta, \lambda, \alpha$ are constant parameters. The potential V is

$$V(z) = \lambda(\frac{1}{4}z^4 - \frac{1}{2}z^2) + \alpha(z - \frac{1}{3}z^3) \ ,$$

By definition, $R = \{x \in \Omega : \phi(x) > \alpha/\lambda\}$, but in addition the potential V effectively constrains $\phi(x) \simeq 1$ for $x \in R$ and $\phi(x) \simeq -1$ for $x \notin R$. The local derivative product $\nabla\phi(x) \cdot \nabla\phi(x)$ penalizes large gradients and ensures that ϕ makes a smooth transition from -1 to 1 across the boundary ∂R. The interaction function Ψ is a function of the distance between the interacting points.

It was demonstrated in [14] that the first line of equation 1 is equivalent to an active contour model with an energy $E = \lambda_C L(\partial R) + \alpha_C A(R)$, where L is the length of the contour ∂R and A its interior surface area. The second line introduces the long-range HOAC interactions that encode prior geometric knowledge. It causes pairs of points with antiparallel normal vectors to repel each other, and with parallel normal vectors to attract each other. This has two effects: it prevents pairs of points with anti-parallel normal vectors from coming too close, and it encourages the growth of armlike structures. Consequently, the effect is to assign low energy to (and hence favour) regions composed of long arms of a certain width and with roughly parallel sides that join together at junctions. In other words, it makes the generation of a network possible. E_P thus constitutes an appropriate prior term for our problem.

B. Data energy

In order to segment road networks in the image, the data term of the functional must be defined in terms of the phase field function ϕ. Since, as we mentioned earlier, the potential V constrains $\phi(x) \simeq 1$ for $x \in R$ and $\phi(x) \simeq -1$ for $x \notin R$, $\phi_\pm = (1 \pm \phi)/2$ are approximate characteristic functions for R and $\bar{R} = \Omega - R$ respectively.

Road networks possess the following properties:

- they are mainly built from the same materials (concrete, asphalt) and thus tend to have somewhat homogeneous spectral properties. In contrast, the background (*i.e.* non-road regions) has no single photometric characteristic.
- The surfaces of main roads as seen in the image are not entirely uniform due to the presence of deterministic noise, such as zebra crossings, over-bridges, vehicles, shadows, road signs, etc. Nevertheless, they still show much less variability than the background.

These properties discriminate roads from the background, and are taken into account in the data term of our model, $E_D(I, \phi)$:

$$E_D(I, \phi) = - \int_{x\in\Omega} \Big\{ \big[\ln P_{\mathrm{i}+}(I(x)) + \theta_{\mathrm{v}} \ln P_{\mathrm{v}+}(V(x))\big] \phi_+(x) + \big[\ln P_{\mathrm{i}-}(I(x)) + \theta_{\mathrm{v}} \ln P_{\mathrm{v}-}(V(x))\big] \phi_-(x) \Big\} \ . \quad (2)$$

The $P_{\mathrm{i}\pm}$ and $P_{\mathrm{v}\pm}$ terms are the density functions of the grey level I (one-point statistics model) and of the variances V computed in a window centred at point x (two-point statistics model) respectively, estimated in the main roads $(+)$ and in the background $(-)$. θ_{v} is the weight of the variance.

To learn the parameters for this model, Geographic Information System (GIS) road layer data was used as a mask covering the main roads (see figure 1). For each class (road and background), we then computed the histograms of the pixel intensities I and the variances V, from which we estimated the parameters of Gaussian mixture distributions $(a_\pm; \mu_{1\pm}, \sigma^2_{1\pm}; \mu_{2\pm}, \sigma^2_{2\pm})$ and of Gamma distributions $(b_\pm, c_\pm, d_\pm)$ respectively:

$$P_{\mathrm{i}\pm}(I) = a_\pm N(I; \mu_{1\pm}, \sigma^2_{1\pm}) + (1 - a_\pm) N(I; \mu_{2\pm}, \sigma^2_{2\pm})$$
$$P_{\mathrm{v}\pm}(V) = \frac{V^{b_\pm}}{d_\pm} e^{-\frac{V}{c_\pm}} \ , \quad (3)$$

where $a_\pm \in [0, 1]$ and N is the normal distribution. Examples of histograms and fitted models are shown in figure 2.

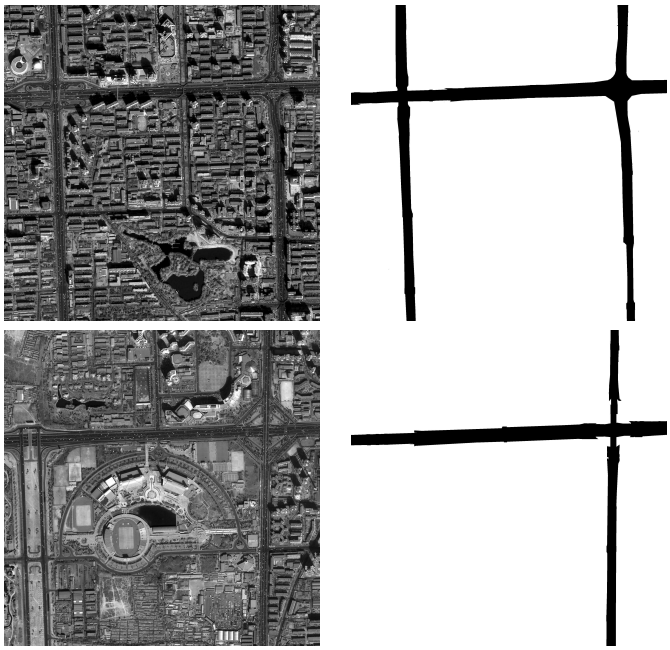


Fig. 1: Two QuickBird©images (size: 2560×2560) and the corresponding GIS masks of the main roads.

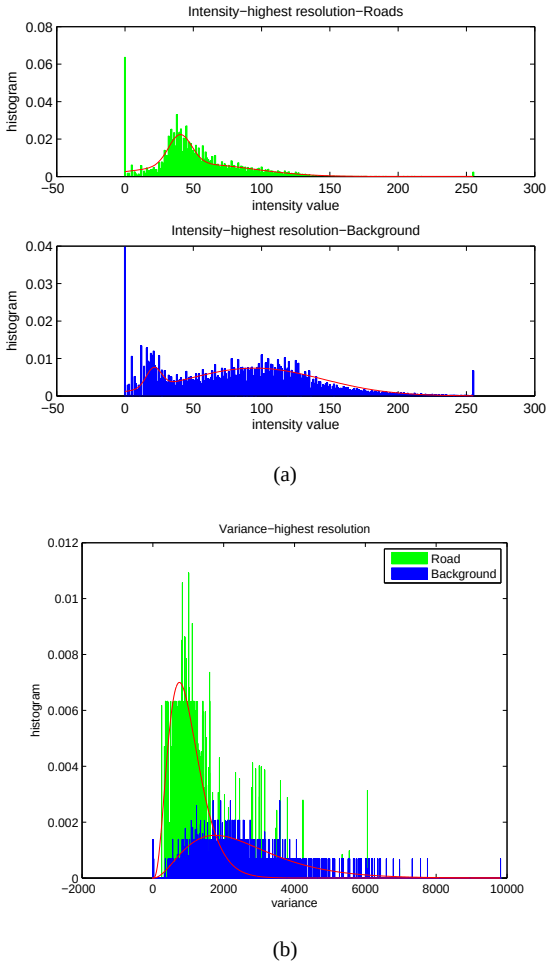
III. Multi-resolution data energy

A multi-resolution analysis is motivated by the following three factors. First, as noted in section I, VHR images contain objects, *e.g.* roads, buildings, at many different scales. In order to capture this complicated behaviour, it makes sense to analyse an image at several resolutions simultaneously. Second, the large size of VHR images compels us to "optimize" the computation speed. Multi-resolution algorithms can be significantly faster than single resolution algorithms. Last, the same object observed at high or low resolutions presents different characteristics. In particular, at low resolutions, the background can be viewed as noise, while the larger roads are still clearly distinguished as homogeneous areas. Road segmentation is thereby facilitated, but is also less precise. In contrast, higher resolutions can provide a more precise location and width for the road. The use of several resolutions thus allows the combination of coarse data, in which details in the image that can disrupt the recognition process have been eliminated, with fine data to increase precision.

To generate a multi-scale version of the image, a wavelet transform [9] is used to generate a series of scaling coefficients at different scales (levels). Since the model energy with a single level works well in level 3 (after three levels of wavelet transform), as we will show in section V, we start with information from this level. Our data energy is a sum of energies computed at four different levels:

$$E_{D\text{MUL}}(I,\phi) = \sum_s E_{D,s}(I_s,\phi) \ ,$$

where I_s, $s \in \{0,1,2,3\}$, are the scaling coefficients at level s of a wavelet transform and $E_{D,s}$ is the data energy at a single



(a)

(b)

Fig. 2: 2(a): histograms of the pixel intensity I in the main road (top) and the background (bottom); 2(b): histograms of the pixel intensity variances V in the main road (green/light grey) and the background (blue/dark grey). Fitted models, equations 3, are shown as solid lines.

level as defined in equation 2. In practice, since the size of the image varies by a factor of 2 from level s to level $s+1$, we up-sample all I_s to the finest resolution. The total energy for the multi-scale model is then defined as

$$E = E_P + E_{DMUL} \ .$$

IV. Optimization framework

To minimize the energy E, we performed a gradient descent with the neutral initialization [14]: the initial value of ϕ is set to be a constant equal to α/λ everywhere in Ω, which corresponds to the local maximum of the potential V. During the algorithm, no re-initialization or *ad hoc* regularization is required. Parameter values and weights are for the moment set by hand, but they are constrained by stability conditions [14]. In the case of the single-scale model energy, the gradient

descent equation is

$$\frac{\partial \phi}{\partial t} = \theta\big[D\nabla^2\phi - \lambda(\phi^3 - \phi) - \alpha(1-\phi^2) - \beta\nabla^2\Psi * \phi\big] + \frac{1}{2}\big[\ln P_{\mathrm{i}+}(I(x)) + \theta_{\mathrm{v}} \ln P_{\mathrm{v}+}(V(x))\big] - \frac{1}{2}\big[\ln P_{\mathrm{i}-}(I(x)) + \theta_{\mathrm{v}} \ln P_{\mathrm{v}-}(V(x))\big] \ .$$

where $*$ indicates convolution. The equation for multiple scales involves adding a copy of the last two lines for each scale.

V. Experimental results

Two pieces of a QuickBird Panchromatic image are shown in figure 1. It is a typical dense urban scene in Beijing. In the same figure are shown the corresponding pieces of the associated GIS main road layer. The image dates from several years after the GIS data were acquired, however, so it is a measure of the robustness of the approach that the model parameters are learned from these outdated GIS masks.

A. Results using a single scale

In this section, we apply the single-scale model to the scaling coefficients of the images in figure 1 at different levels of the wavelet decomposition. Figure 3(a) shows one of these images at level 3 of the wavelet transform, while figure 3(b) shows a zoom on part of this image illustrating that even after three levels of smoothing and down-sampling, the data is still rather complex. The results of applying the single-scale model to the images at level 3 are shown in figure 3(c) and figure 3(d). The road networks in the GIS masks are nearly perfectly retrieved. Thanks to the incorporation of strong geometric prior knowledge, at the coarser resolution, the model is able to bridge gaps resulting from shadows cast on the road, remove errors along the boundary of the roads and ignore erroneously detected roads in the background. Moreover, figure 3(d) shows that the model also has the ability to find roads that do not appear in the GIS masks, *i.e.* that have been built more recently. This is an important issue in the application to map updating. On the other hand, if the radiometric properties of the new roads are quite different from those computed from the given GIS mask, they cannot always be found (see figure 3(c)).

The level 3 image is already quite complex, and we find that if we try to use the same model at finer resolutions, using the images at levels 2, 1, or 0, the details of the scene in the image make road extraction more difficult (see figure 4). The erroneous detections in the background result from regions of poor contrast between the main roads and the buildings or areas of vegetation, and also from the smaller roads, which have statistical properties similar to the main roads. The shadows of high buildings, cars, road markings and bridges lead to jagged borders or gaps along the roads. The former indicates a lack in the image model, while the latter seem more likely to be due to a weakness in the prior model, which therefore needs to be improved in order to enforce the road geometry more effectively.

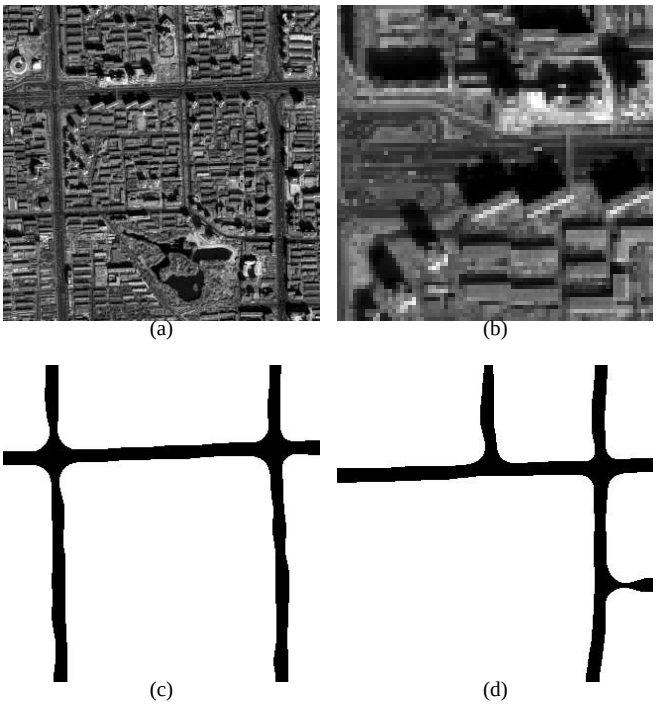


Fig. 3: Image data and extraction results at level 3 (size: 320×320) using the single-scale model energy. Road width $\simeq 12$ pixels. 3(a): image data at level 3 for one of the two test images. 3(b): zoom to illustrate the complexity remaining even at this level. 3(c) and 3(d): the results obtained using the single-scale model energy at level 3.

B. Results using multiple scales

In order to improve the segmentation result at finer resolutions, we applied the multi-scale model (see section III). A result is shown in figure 5. The result is not perfect, but is very promising considering the complexity of the image. The use of a multi-scale model improves the result obtained from the original image at a single level (see figure 4(c)). However, there are still some false detections in the background and the road borders are rather inaccurate due to geometric noise along the boundaries of the road. The result indicates that a simple sum of data energies at several different scales, while helpful, is not sufficient to solve the problem completely. It suggests that the model should include more complicated relationships among the scaling coefficients and maybe the wavelet coefficients, within and between scales.

VI. Conclusion

'Phase field HOACs' are powerful models for image region segmentation. They have many advantages with regard to conventional models: a linear representation space; ease of implementation; neutral initialization (initialization is chosen as a constant function); greater topological freedom; and the inclusion of sophisticated prior knowledge of region geometry. In this paper, we have proposed a new multi-resolution image

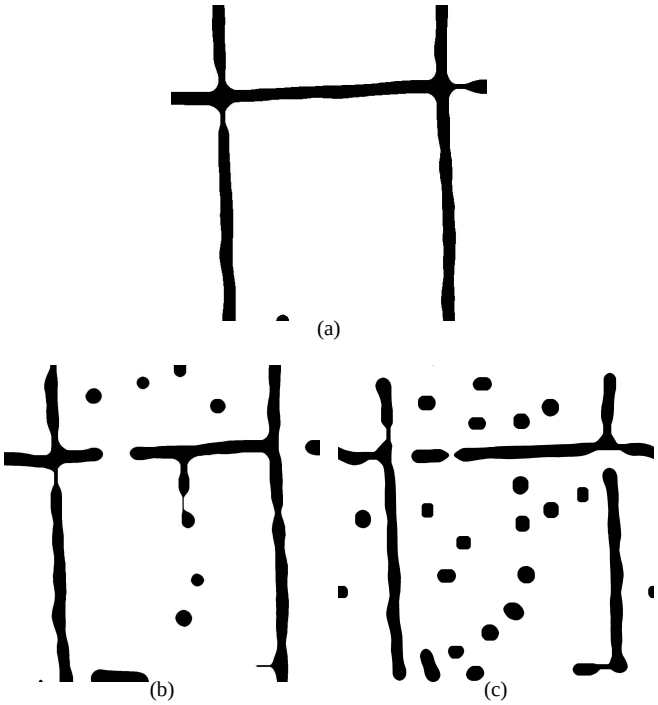


Fig. 4: Extraction results using the single-scale model energy at finer resolutions. 4(a): result at level 2 (size: 640×640), road width $\simeq 24$ pixels. 4(b): result at level 1 (size: 1280×1280), road width $\simeq 48$ pixels. 4(c): result at original resolution (size: 2560×2560), road width $\simeq 96$ pixels.

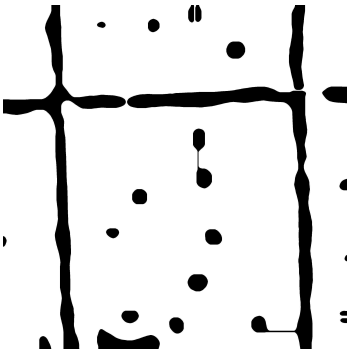


Fig. 5: Extraction result at the original resolution (size: 2560×2560) using the multi-scale model energy.

model to adapt 'phase field HOACs' to the problem of urban road extraction from VHR QuickBird Panchromatic images. We have shown that the model is robust: it is able to retrieve roads in any image samples in which the roads have similar statistical properties to those used for the parameter learning. Nevertheless, the multi-scale approach needs further improvements in order to eliminate false detections and improve the accuracy of road border delineation.

Our current work is focused on a further study of the two-point statistics of the wavelet and scaling coefficients to improve the data model; we are also working on a coherent multi-scale prior energy term (as opposed to data term) to reduce computational load. In addition, with the map updating application in mind, we plan to incorporate even more prior knowledge in the form of outdated GIS maps, which should help greatly to improve the results.

ACKNOWLEDGMENTS

The work of the first author is supported by an MAE/Alcatel/LIAMA grant. The authors would like to thank J. G. Planes and B. Serra (Alcatel Alenia Space) for interesting discussions. They also thank the Beijing Institute of Surveying and Mapping for providing the GIS data.

REFERENCES

[1] V. Amberg, M. Spigai, and Y. Le Roy, *A tracking algorithm for road extraction in dense urban areas: importance of contextual information*, EUSAR, Dresden, Germany, May 2006.

[2] T. Chan and L. A. Vese, *Active contours without edges*, IEEE Transactions on Image Processing, 10(2): 266–277, February 2001.

[3] Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. Gopinath, R. Briggs, and E. Geiser, *Using prior shapes in geometric active contours in a variational framework*, International Journal of Computer Vision, 50(3): 315–328, 2002.

[4] D. Cremers, F. Tischhauser, J. Weickert, and C. Schnorr, *Diffusion snakes: introducing statistical shape knowledge into the Mumford-Shah functional*, International Journal of Computer Vision, 50(3): 295–313, 2002.

[5] X. Hu, C. V. Tao, and Y. Hu, *Automatic road extraction from dense urban area by integrated processing of high resolution imagery and lidar data*, In Proc. International Society for Photogrammetry and Remote Sensing (ISPRS), Istanbul, Turkey, July 2004.

[6] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: active contour models*, International Journal of Computer Vision, 1(4), pp. 321–331, 1988.

[7] H. Long and Z. Zhao, *Urban road extraction from high resolution optical satellite images*, International Journal of Remote Sensing, 26(22): 4907–4921, November 2005.

[8] S. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, pp. 674–693, July 1989.

[9] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1997.

[10] H. Mayer, I. Laptev, and A. Baumgartner, *Multi-scale and snakes for automatic road extraction*, European Conference on Computer Vision (ECCV), Freiburg, Germany, June 1998.

[11] J. B.Mena, *State of the art on automatic road extraction for GIS update: A novel classification*, Pattern Recognition Letters, 24(16): 3037–3058, December 2003.

[12] R. Péteri, J. Celle, and T. Ranchin, *Detection and extraction of road networks from high resolution satellite images*, In Proc. IEEE International Conference on Image Processing (ICIP), Barcelona, Spain, September 2003.

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

[14] M. Rochery, I. H. Jermyn, and J. Zerubia, *Phase field models and higher-order active contours*, IEEE International Conference on Computer Vision (ICCV) 2005, Beijing, China, October 2005.

[15] R. Wanga and Y. Zhang, *Extraction of urban road network using QuickBird pan-sharpened multispectral and panchromatic imagery by performing edge-aided post-classification*, ISPRS Joint Workshop on 'Spatial, Temporal and Multi-Dimensional Data Modeling and Analysis', Quebec City, Canada, October 2003.