Road Network Extraction from High Resolution Airborne Digital Camera Data

Peng Gong* and Jinfei Wang†

*Center for Assessment and Monitoring of Forest and Environmental Resources
Department of Environmental Science, Policy and Management,
University of California, Berkeley, CA 94720-3310, USA
†Department of Geography
University of Western Ontario, London, Ontario, Canada, N6A 5B7

Abstract

Most road network extraction algorithms developed are based on linear analysis methods. These methods involve search of edges through edge filtering, morphological filtering, or gradient modelling. As image resolution increases from 10-30 m to 0.5-2 m, road networks will appear to be narrow areas rather than thin lines. This becomes a challenge for traditional linear analysis methods based on mask operations but creates an opportunity for classification based methods. We experimented with an advanced linear analysis, gradient direction profile analysis, and a few classification algorithms including a maximum likelihood classification, clustering and a contextual classifier for road network extraction using airborne digital camera data acquired over Livermore, California with approximately 1.6 m spatial resolution.

Results indicate that both the linear extraction and image clustering algorithms worked reasonably well. The linear extraction method requires some preprocessing such as filtering of the original image. Best road network results have been obtained by applying the linear extraction algorithm to a morphologically filtered image that was generated by combining the near infrared (NIR) and red (R) image bands through NIR/R+NIR. With this method, the correctly extracted road pixels account for 78.7% of the total road pixels obtained from image interpretation with field verification. The image clustering method resulted in 74.5% correctly extracted road pixels. The contextual classification resulted in relatively noise-free road networks in new residential areas but omitted some roads at older residential areas. When experimenting with the images resampled at approximately 3 m and 5 m resolution, the best overall accuracies for road extraction decreased to 74.6% and 61.6%, respectively.

摘 要

大部分道路提取的算法基于线性特征提取。这类方法大致由边界搜索、形态滤波或梯度模拟。当图象空间分辨率从10-30米提高到0.5-2米时，道路则由线状特征变成较狭窄的面。这既对传统线性提取算法带来了挑战，又给分类算法制造了机会。本文使用在加州Livermore 市获取的大约1.5米分辨率的多光谱航空数字摄影图象，试验了一种先进的线性特征提取方法－梯度方向剖面分析法，并与一些分类算法进行比较，其中包括最大似然法、簇分析法和一种上下文分类的方法。

试验结果表明线性提取和簇分析法对道路提取均有成效。使用线性提取法之前需对原始图象进行一些预处理。最佳结果是先对原始图象的近红外(NIR)和红波段(R)进行NIR/R+NIR 运算。然后使用亮度扩展的数学形态学方法对上述运算所得图象进行处理。再从该处理图象上运用梯度方向分析法进行道路提取，精确度可达78.7%。簇分析法所得道路网的精确度可达74.5%。当图象空间分辨率降低至大约3米或5米时，最佳精确度分别降低74.6%和61.6%。

I. INTRODUCTION

Road network changes constantly at many rural-urban fringe areas due to urban expansion. Urban planners and decision makers on land use development often have obsolete land use information because operational mapping methods based on manual interpretation of aerial photographs usually take a year or two to complete from the time of aerial photography. Research efforts have been made to develop computer analysis algorithms for road network extraction (e.g., Wang et al., 1992; Gruen and Li, 1995; Wang et al., 1996) and land-use mapping (e.g., Gong and Howarth, 1992) from satellite images. On 10-30 m spatial resolution satellite images, roads are linear features

1082-4006/97/0301~2-51$3.00
©1997 The Association of Chinese Professionals in Geographic Information Systems (Abroad)

represented by valleys or ridges of brightness. Wang and Liu (1994) grouped 4 types of line extraction methods that could be applied to road network extraction. They are (1) gradient operator and mask convolution method, (2) gradient direction profile analysis (GDPA) method, (3) mathematical morphology analysis method, and (4) knowledge-based method. Because the contrast between a road and the image background varies both spatially and spectrally, the use of multispectral data helps reduce road ambiguity (Wang and Liu, 1994).

Among various satellite and airborne sensing technologies, low-cost high spatial resolution digital CCD (Charge-Coupled Device) cameras are developing rapidly. The interest in multispectral imaging with digital CCD cameras has been increasing (King, 1995). High precision georeferencing techniques are available through integrating GPS (global positioning systems) and INS (inertial navigation systems) with digital photography (e.g., Schwarz et al., 1993). It is now possible to have high geometric and radiometric quality digital camera images on airborne platforms with spatial resolutions at the sub-meter level. In addition, 1-4 m resolution satellite imagery will soon become available (Fritz, 1996) and high spatial resolution digital image can be obtained by scanning aerial photographs. In a study of road networks extraction from scanned color-infrared films from aerial photography, Benjamin and Gaydos (1990) claim that 3 m spatial resolution is most suitable for road network extraction in Cupertino, California. They applied clustering and editing instead of the more sophisticated line extraction algorithms to scanned data resampled to different spatial resolution (1-5 m). Since most roads in urban areas are wider than 5 m, road networks become narrow areas rather than brightness valleys or ridges on images with pixel sizes smaller than 5 m by 5 m. At a spatial resolution better than 5 m, it is possible to extract road network with classification methods.

Our questions are:

- how well can traditional linear extraction algorithms perform when applied to those high spatial resolution images?
- to what extent, can classification methods be used for road network extraction purposes?
- how can methods in the two different paradigms be used to complement each other for improved road network extraction and land use classification?

This paper presents some of our efforts toward answering the first two questions. The objective was to compare classification methods with line extraction algorithms for road network extraction from digital camera images resampled at different spatial resolutions. A supervised per-pixel classification, a clustering, and a cover-frequency based contextual classification method were applied to classify an airborne digital camera image. A few road cover types were included in each classification. A gradient direction profile analysis algorithm was also applied to the same image. Road extraction results are presented and discussed.

II. STUDY SITE AND DATA

The study site is located on the east border (121°43' W, 37°41' E) of City of Livermore, California (Figure 1). The 1:24,000 USGS topographic map was last updated in 1981. Since then, new residential areas and road networks have been built. On June 30, 1995, an imaging system consisting of 4 Kodak DCS-200 cameras was used on board of an aircraft to acquire multispectral images over the study area. The four cameras simultaneously acquired images at 450 nm, 550 nm, 650 nm, and 850 nm, respectively, with a band width of 80 nm. The single-band images from individual cameras were then geometrically corrected and resampled to form a multispectral data set. Part of the red band image is shown in Figure 2. The spatial resolution is approximately 1.6 m.

Roads in the study area are mostly asphalt paved. Due to the fact that pavement ages are different, the asphalt road surface looks different in brightness. Older pavements appear brighter than the background among the visible bands but darker in the near-infrared band. Sidewalks are the brightest in the green and red bands of the image. There are also unpaved roads in the non-irrigated grass land in the study area. Grass lands in the build-up areas are usually irrigated.

III. ROAD EXTRACTION METHODS AND ACCURACY ASSESSMENT

Road network extraction methods generally involve five steps: image preprocessing, obtaining initial road network, noise removal, thinning, and pruning (Figure 3). The purpose of image preprocessing is to enhance road network features for subsequent analysis. This includes spatial filtering such as the

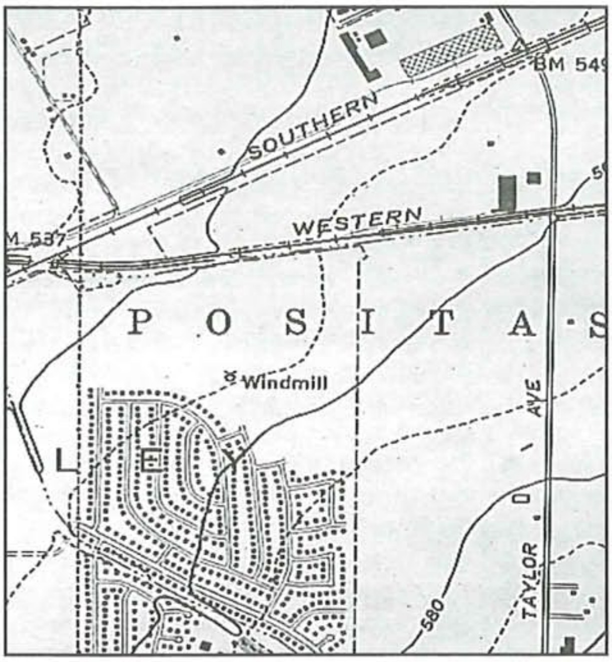

Figure 1. 1981 USGS 1:24,000 topographic map of the study area (note: presented map sacle may not be 1:24,000)

use of a median filter to preserve edges and remove speckle noise, morphological filtering, band ratioing and linear transformation such as the use of principal component analysis to enhance linear features in the original image (Jensen, 1996). Morphological filtering is usually conducted on binary images (Wang et al., 1996; Destival, 1986; Lee et al., 1987). We undertook grey-level dilation filtering to the original image. It is essentially an operation in search of maximum from a local neighborhood defined by the structuring element (Pratt, 1991; Sternberg, 1986). Moving a 3 X 3 kernel over a grey-level image, we assign the maximum grey-level value to the pixel at the kernel center. This is equivalent to a special version of the dilation filtering using a 3 X 3 structuring element with all its kernel values being 1. Noise removal is to remove from the initial road networks relatively small patches of pixels that have been identified as road segments. Pixel patches smaller than a certain size are removed from the initial road network image. Thinning reduces the detected road network from a few pixels wide to one pixel in width. Pruning removes short-branches of dead-end roads according to their lengths.

Gradient Direction Profile Analysis

The GDPA algorithm used in Wang et al. (1992) was

(a) (b)

Figure 2. Digital camera image acquired over Livermore, California on June 30, 1995. (a) Red band, and (b) Near infrared band.

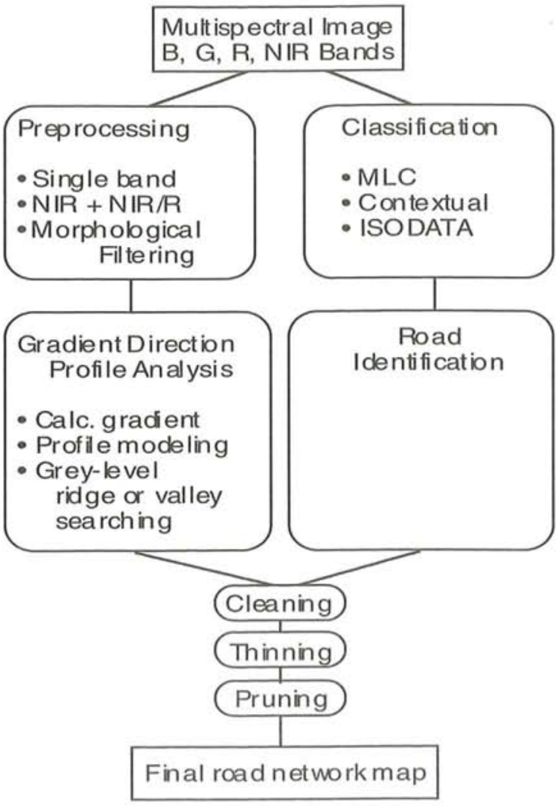

Figure 3. Procedures used in this study.

selected for use in this study. This algorithm first finds the greatest gradient direction for each pixel using the brightness values in a 3 by 3 neighborhood. A pixel is considered as a candidate road pixel if its greatest gradient exceeds a limit, T_d, specified by the analyst. The algorithm then searches among the candidate road pixels for pixels at ridge tops or valley bottoms of grey levels by modeling grey level profiles along the greatest gradient direction centered at each candidate pixel. The modeling is achieved through polynomial curve fitting along the gradient direction profile. The grey-level ridge top or valley-bottom positions are found through profile derivative and curvature analysis. The length of a profile, L, is specified by the analyst and the curvature of the polynomial function must be greater than a specified value, T_k. An initial road network is extracted from an image by adjusting the three parameters, T_d, L and T_k.

Details on GDPA is found in Wang et al. (1992). When applied to Landsat Thematic Mapper (TM) and SPOT multispectral and panchromatic data, the GDPA algorithm produced best road network results using visible spectral bands such as the green or red bands (Wang et al., 1992; Wang and Liu, 1994). This algorithm has also been successfully applied to extract various linear features from scanned topographic maps (Wang and Zhang, 1996).

Image Classification and Clustering

To classify roads from the digital camera image, we employed a clustering algorithm, a supervised maximum likelihood classifier and a cover-frequency based contextual classifier. The clustering algorithm is ISODATA (iterative self organizing data analysis technique) (Duda and Hart, 1974). The cover-frequency based contextual classifier is found in Gong and Howarth (1992). It first converts the original image to a land-cover map using a regular per-pixel maximum likelihood classification (MLC) algorithm, or a grey-level vector reduced image with a grey-level vector-reduction algorithm, or a cluster map through clustering. The algorithm then extracts frequencies of land cover, cluster, or grey-level vector from a neighborhood of a pixel and uses the frequencies in classification of land uses or discrimination of road types for that pixel. The size of the pixel neighborhood is specified by the analyst. The contextual classification requires supervised training to determine the frequencies for each class. The same training set was used for both the maximum likelihood classification and the contextual classification.

We used 8 classes in the supervised classification. They include four types of road covers according to different road surface colors and materials, new asphalt, older asphalt, concrete and railroad. The other 4 cover types are residential, industrial, well irrigated grass land and dry grass land. We selected training areas for the 8 land-use classes and used the maximum likelihood classifier and the contextual classifier to classify the imaged area.

Image Resampling

All image resampling methods can be achieved through image convolution. To resample the 1.6 m resolution data to coarser resolutions, one may use image averaging, nearest neighbor, bilinear or cubic convolution resampling methods (Shlien, 1979). These different methods can be realized through different design of weighting factors. For example, with image averaging a convolution kernel with equal weights is usually employed (Benjamin and Gaydos, 1990; Pratt, 1991). With nearest neighbor a unit weight is applied to the pixel closest to the newly transformed pixel location while a triangular function is used to determine the weights for the two pixels closest to the newly transformed pixel location along the row and column direction respectively.

In this study, we compared the four different image resampling methods. As expected image averaging had the greatest effects in blurring the edges or linear features as image resolution degrades. Edges and linear features were better preserved in the resampled images generated by the remaining three resampling methods and those resampled images looked very similar. Therefore, we tested the linear extraction algorithms using the images resampled to 3 m and 5 m resolutions with the nearest neighbor method. We did not apply classification methods to the resampled images because as discussed in the introduction images with coarser spatial resolution would be less useful for road network classification.

Accuracy Assessment

In order to compare the performances among the various road network extraction algorithms, we conducted field study and manually traced the road networks in the image. The manually traced road networks were used as the truth image (Figure 3). Three measures can be calculated for the quantitative comparison (Wang and Liu, 1994).

Let N_{ce} be the number of correctly extracted road pixels; N_{tr} be the number of true road pixels; and N_{te} be the total number of extracted road pixels:

overall accuracy = N_{ce} / N_{tr}

commission error = $(N_{te} - N_{ce}) / N_{tr}$

omission error = $1 - N_{ce} / N_{tr}$

The overall accuracy is the fraction of pixels correctly extracted as roads. The commission error is the number of pixels incorrectly extracted as road pixels divided by the number of true road pixels. The omission error is one minus the overall accuracy. Since it is not independent, we will only report the overall accuracy and commission error. Obviously a smaller number is better with the commission error and a larger number is better with the overall accuracy.

IV. RESULTS

We applied the GDPA algorithm to each band of the original image. Among the visible bands, the road is relatively brighter than the background vegetation. We used a ridge-searching algorithm. However, the sidewalks are readily observable and are the brightest in the scene. Therefore, the road networks extracted from the visible bands are mostly sidewalks and are far from perfect. In the near infrared band, road networks and house roofs are the darkest. Since the roads are a few pixels wide, we applied grey-level dilation filtering to the original image using a 3 X 3 structuring element for 5 times (Pratt, 1991; Sternberg, 1986). This filtering expanded bright component in the image such as vegetation while shrank the dark components such as the road and the house roofs. We then applied the GDPA algorithm to the filtered image by searching for valleys (Td=3.0, L=3, Tk=5.0). The road network result has been considerably improved. However, a lot of house roofs have been extracted as roads and they are too big to be removed using the noise removal method. The final road network extraction results after thinning and pruning are shown in Figure 4. Similar results have been achieved by applying the GDPA algorithm to the second principal component image obtained from the original 4 bands.

The best result with the GDPA algorithm has been achieved using an image that combines the ratio between the near infrared and red band with the

Figure 4. Final road networks obtained using GDPA from the NIR band.

Figure 5. The image obtained through NIR R+NIR.

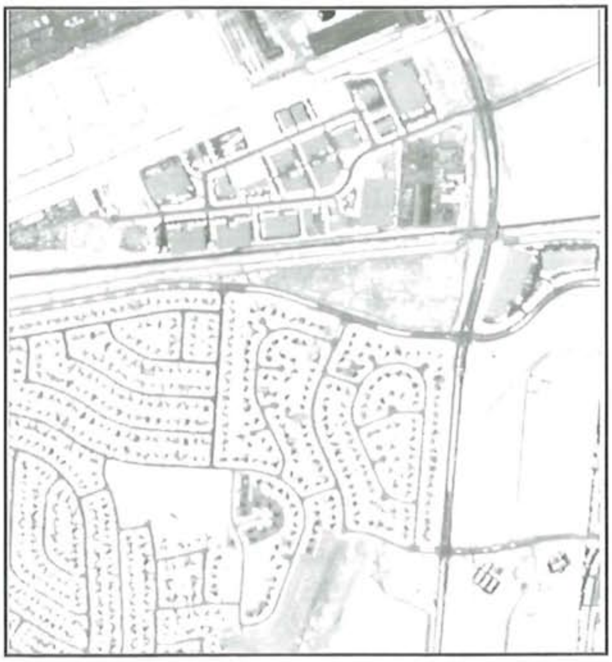

Figure 6. The image after 3 iterations of dilation filtering from the image shown in figure 5.

near infrared band, [band 4/band 3] + band 4 (Figure 5). The newly generated image suppressed the shade and shadow in the original image through image ratioing, while the brightness of the vegetation portion of the image was enhanced. We then applied dilation filtering to the band combination image for 3 iterations (Figure 6). The initial road network was extracted from the dilated image (Td=1.0, L=5, Tk=3.0). The final road extraction result is shown in Figure 7.

With the image classification approaches, no preprocessing was applied to the original images. The four bands of image were clustered using ISODATA resulting in 49 clusters. After a cluster by cluster examination, we selected those clusters corresponding to road networks. All road network clusters were merged to form an initial road network map, which was further processed by noise removal, thinning and pruning, and resulted in a final road map (Figure 8). Supervised MLC classification was also performed on the original four bands. Similarly, noise removal, thinning and pruning were applied to produce a final road map. The final road network results obtained from MLC and from the clustering algorithm are similar. The clustering algorithm picked up more road details, particularly the rail roads, than the MLC method. It also included more non-road artifacts in the final result.

The cluster map was also used as the basis for the contextual classification. Classified road results were used as the initial road networks for subsequent noise removal, thinning and pruning. The final road network results obtained from the contextual classification are shown in Figure 9. We used a 9 by 9 pixel neighborhood size to generate cover frequencies in the contextual classification. We did not attempt to find an optimal window size for the classification. Generally, a large window size would remove more road details particularly when the road is relatively narrow. While a small window size would results in more unwanted details. At the relatively new residential area, the contextual classifier resulted in the cleanest road network. However, like the MLC results, the algorithm did not pick up roads on the upper left corner of the image. The roads are within the oldest residential area at that particular section of the image where roads are narrower than the newer residential sections. Some trees cover part of the road network causing the difficulties for the contextual classification methods. The supervised MLC method and the contextual classification methods may also be sensitive to training sample selection.

Table 1 lists some of the accuracy assessment results. With the 1.6 m resolution image, the best overall accuracy, 78.7%, is obtained by the GDPA

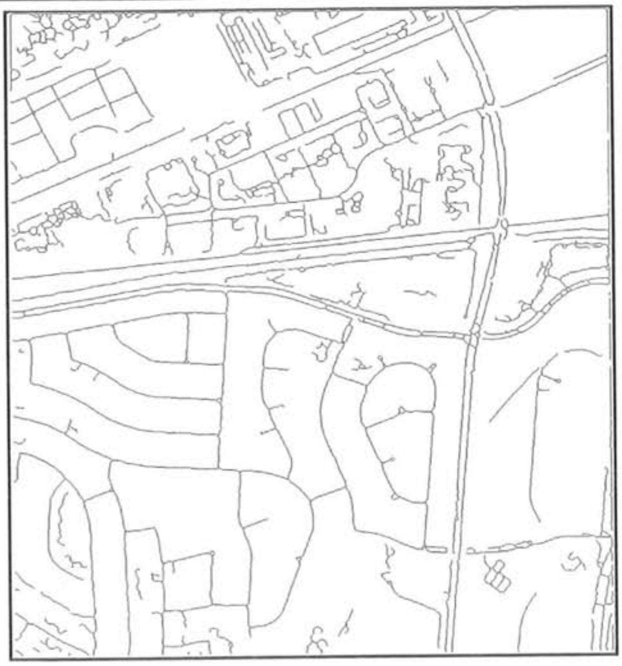

Figure 7. Final road networks obtained using GDPA from figure 5.

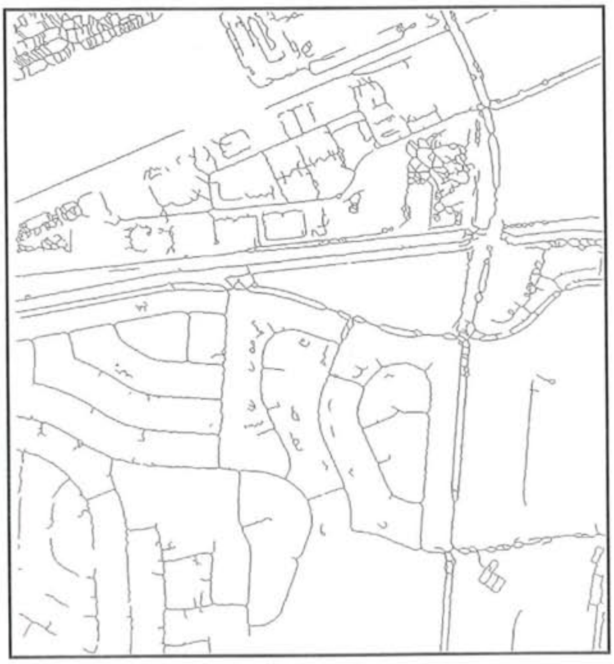

Figure 8. Road networks obtained from clustering-based method.

method from the image specially enhanced (b4 + b4/b3). The second best is achieved by the clustering method (74.5%). However, both methods have relatively high commission errors. The lowest overall accuracy is from the contextual classification results but the commission error is the lowest.

For the 3 m and 5 m resolution images, we conducted similar experiments as for the 1.6 m resolution data using the GDPA method. For each resolution, better accuracies have been achieved with the band 4 image and the specially enhanced image by taking the sum of band 4 and the ratio between band 4 and band 3. These were obtained from the preprocessed images with only 1 iteration of grey-level dilation. With the 3 m resolution, we obtained better overall accuracy and smaller commission error from the band 4 image in comparison with the 1.6 m resolution data. An overall accuracy of 74.6% has been achieved with the specially enhanced image. The commission error (0.832) from this image is smaller than that (0.984) obtained from the 1.6 m image with the same enhancement. When the resolution decreases to 5 m, the overall road network extraction accuracies are considerably lower.

V. DISCUSSION AND CONCLUSIONS

With the gradient direction profile analysis algorithm, previous studies using 10 - 30 m resolution Landsat TM or SPOT images indicated that visible spectral bands are more effective than the near infrared bands for the purpose of road network extraction . As the spatial resolution of digital imagery improves to 1 - 3 m, road networks are no longer lines of one pixel in width but narrow areas of several pixels wide. This makes it harder to directly apply line extraction algorithms to high spatial resolution images. In order to effectively use line extraction algorithms, it is necessary to preprocess the original image. Due to their spectral similarity with road networks, house roofs and drive ways are the primary disturbing factors to road network extraction. It would be desirable to completely remove the house roofs while keeping the road network at narrow widths. We found that a morphological filtering based on a dilation process applied to the near infrared band can accomplish this to a large extent. The visible spectral bands may not be the most effective bands for road network detection. The near infrared band exhibits the greatest contrast between road networks and their background. The best result using the GDPA method was achieved using an image produced by summing up the near infrared image and the ratio

Figure 9. Road networks obtained from the contextual classification method.

between the near infrared and the red band.

We had expected that better accuracies might be obtained by line extraction algorithms such as the GDPA method if the original high spatial resolution of the data was degraded. This expectation was not met in our experiments with 3 m and 5 m resolution images resampled from the 1.6 m resolution data. The best overall accuracy achieved with the 1.6 m data is only 4.1% higher than that from the 3 m data, while the commission error is also higher. We consider the best road network extraction results are comparable with the 1.6 m and the 3 m resolution data. However, the best overall accuracy obtained with the 5 m data drops considerably in comparison with those obtained from the 1.6 m and 3 m data. For similar types of landscapes, it seems to us that the procedure developed for the GDPA method would perform consistently well if image resolution is better than 1.6 m. If the resolution is finer than 1.6 m, all one needs to do is to apply the morphological dilation for a few more times. In conclusion, for similar types of landscapes the GDPA in combination with the preprocessing methods developed in this research works well for images with resolutions better than 3 m. For road network extraction 3 m resolution images may be more desirable from a computational point of view than finer resolutions because the image file is smaller.

Table 1. Performance evaluation of different road network extraction results

Images	Overall accuracy(%)	Commission error
band4	59.8	1.010
clustering	74.5	0.845
contextual	49.5	0.476
MLC	68.4	0.659
b4+b4/b3	78.7	0.984
band4 (3m)	66.0	0.951
b4+b4/b3 (3m)	74.6	0.832
band4 (5m)	53.5	0.882
b4+b4/b3 (5m)	61.6	0.693

With the classification based approaches, road networks can be obtained reasonably well, particularly with the clustering method. The variability among different road surfaces makes it impossible to treat road as a single class in image classification. Training must be carefully done if supervised classification is used. A large number of clusters is needed if clustering is used to extract road networks. With the cover-frequency based contextual classification algorithm, the most noise-free road networks are obtainable for relatively new residential areas. It works less effectively in older residential areas where trees block the visibility of road surfaces.

To help understand the performance of different road network extraction methods, the study area was purposely selected not to be as complicated as many old built-up areas could be. Old areas with high density of tree coverage present a problem to all road network extraction approaches. As an important component on large scale urban land use maps, road networks also have the potential in improving classification accuracies. It is worthwhile to explore the complementary roles that classification and traditional line extraction methods could play for improving both the road network extraction and image classification results.

The procedure described here may not work well for non-residential areas as can be seen from the industrial area in the image. The preprocessing method works well only when there are vegetation surrounding houses in relatively newer suburban areas. The morphological dilation procedure will not work well in areas where trees covering road surfaces and leaving the road networks to be imaged in different widths. For high resolution image data, image preprocessing is necessary and different procedures are needed for different areas.

REFERENCES

[1] Benjamin, S., and L. Gaydos, 1990. Spatial resolution requirements for automated cartographic road extraction, *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 1: 93-100.

[2] Destival, I., 1986. Mathematical morphology applied to remote sensing, *Acta Astronautica*, Vol. 13, No. 6/7, 371-385.

[3] Duda, R.O. and P.E. Hart, 1973. *Pattern Classification and Scene Analysis*, Wiley-Interscience, New York.

[4] Fritz, L.W., 1996. The era of commercial earth observation satellites, *Photogrammetric Engineering and Remote Sensing*, Vol. 62, No. 1: 39-45.

[5] Gong, P. and P. J. Howarth, 1992. Frequency-based contextual classification and grey-level vector reduction for land-use identification. *Photogrammetric Engineering and Remote Sensing*, 58(4):423-437.

[6] Gruen, A. and H.H. Li, 1995. Road extraction from aerial and satellite images by dynamic programming, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 50, No. 4:11-20.

[7] Jensen, J.R., 1996. *Introductory Digital Image Processing, a Remote Sensing Perspective*, 2nd Ed., Prentice Hall, Upper Saddle River, New Jersey.

[8] King, D.J., 1995. Airborne multispectral digital camera and video sensors: a critical review of system designs and applications, *Canadian Journal of Remote Sensing*, Vol. 21, No. 3, 245-273.

[9] Lee, J.S., R.M. Haralick, and L.D. Shapiro, 1987. Morphologic edge detection, *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 2:142-156.

[10] Pratt, W., 1994, *Digital Image Processing*, John Wiley and Sons: Toronto.

[11] Schwarz, K-P., M.A. Chapman, M.E. Cannon, and P. Gong, 1993. An integrated INS/GPS approach to the georeferencing of remotely sensed data, *Photogrammetric Engineering and Remote Sensing*, Vol. 59, No. 11, 1667-1674.

[12] Shlien, S., 1979. Geometric correction, registration, and resampling of Landsat Imagery, *Canadian Journal of Remote Sensing*, 5(1):74-87.

[13] Sternberg, S.R., 1986. Grayscale morphology, *Computer Vision, Graphics, and Image Processing*, 35:333-355.

[14] Wang, D, D.C. He, L. Wang, and D. Morin, 1996. Extraction of urban road networks from SPOT HRV images. *International Journal of Remote Sensing*, 17(4):827-833.

[15] Wang, J.F., P. M. Treitz and P.J. Howarth, 1992. Road Network detection from SPOT imagery for updating geographical information systems in the rural- urban fringe, *Int. J. Geographical Information Systems*, Vol. 6, No. 2:141-157.

[16] Wang, J.F. and W. Liu, 1994. Road detection from multispectral satellite imagery, *Canadian Journal of Remote Sensing*, Vol. 20, No. 2:180-189.

[17] Wang J.F. and K. Zhang, 1996. Linear cartographic data acquisition from scanned topographic maps. *Geographic Information Sciences*, Vol. 2(1):12-23.