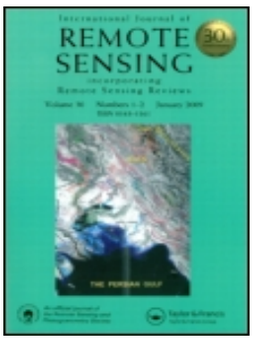

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

Robust approach for suburban road segmentation in high-resolution aerial images

D. Guo , A. Weeks & H. Klee

To cite this article: D. Guo , A. Weeks & H. Klee (2007) Robust approach for suburban road segmentation in high#resolution aerial images, International Journal of Remote Sensing, 28:2, 307-318, DOI: 10.1080/01431160600721822

To link to this article: http://dx.doi.org/10.1080/01431160600721822

International Journal of Remote Sensing
Vol. 28, No. 2, 20 January 2007, 307–318

Taylor & Francis
Taylor & Francis Group

Robust approach for suburban road segmentation in high-resolution aerial images

D. GUO*†, A. WEEKS‡ and H. KLEE‡

†Computer Science Program, Florida Gulf Coast University, Fort Myers, FL 33965, USA

‡Departments of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816, USA

(*Received 10 August 2005; in final form 28 March 2006*)

The goal of this research is to develop an algorithm that accurately segments high-resolution images, where linear features, such as roads, are corrupted by noise. In high-resolution images, there are two types of noises that are obstacles to road segmentation. Noise could be within road areas (such as cars, water, differences in composite road surface mix) or unwanted contents outside road areas, like buildings and trees. To remove unwanted contents, the geographical information from the United States Geographical Survey (USGS) is used. The USGS provides a collection of road centre line information that has been collected for many years and can be used to limit the area for road segmentations close to roads. In this paper, a standard process was developed to align the USGS geographical information with the high-resolution images. USGS geographical data is used to eliminate background clutter that is disjunct from roads. The road segmentation process is then reduced to dealing with automobile traffic, shadows and pavement colour discontinuity within road areas. In order to achieve reliable road segmentation in the presence of these objects, the mean-shift clustering approach is used within the hue-saturation-intensity (HSI) space. Conditional morphological image processing techniques are also used to significantly improve the segmentation results. The proposed method results in the average accuracy of road segmentation above 85%.

Keywords: road segmentation; USGS; High-resolution aerial photos

1. Introduction

Many methods have been designed for the automated extraction of roads from aerial photos, varying from 3 to 4 m per pixels (Lipari *et al.* 1989) to 0.25 m per pixel (Steger, 1998). The aerial images in this research have a resolution of about 0.25 m per pixel, giving a very fine resolution.

Some early research in road segmentation reported has focused on how to detect roads in images and retrieve the road centre line information (Vosselman and de Gunst 1997, Gruen and Li 1997, Kararzis *et al.* 2001, Lipari *et al.* 1989, Tupin *et al.* 1998, 2002, Wang *et al* 1988). This information is very important for segmentation in road areas, because it can limit the region for road segmentation to areas close to the centre line. The USGS, however, has been maintaining road centre line

*Corresponding author. Email: dguo@fgcu.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160600721822

information for many years (USGS-DLG, http://edc.usgs.gov/products/map/dlg.html#description, 2004, accessed 15 November 2004), as digital line graphs (DLGs). This type of data includes the centre line information of almost all roads within the continental USA. Consequently, instead of searching roads in an entire aerial image space using image processing techniques, the USGS DLG data can be used to limit road segmentation to regions close to the road areas. This idea is similar to that reported by Zhang *et al.* 2001), who used other geographical information to guide road segmentation, such as the digital surface model (DSM) and the digital terrain model (DTM). The search space for roads can be limited by the ground areas, defined by DSM/STM (Zhang *et al.* 2001). In this research, we also used some existing geographical information, which was directly road centre lines. Prior to using DLG data for road segmentation in an image, any distortion between DLG and the source image had to be identified and corrected.

Regarding road segmentation, some methods, such as those reported by (Laptev *et al.* 2000), are based on snakes (Kass *et al.*, 1988) or zip-lock snakes (Neuenschwander 1997). Both snakes and zip-lock snakes are curves that are moved within images to find object boundaries by calculating curve points iteratively based on image data. These methods are affected by lateral pavement colour discontinuities which make the road cross-sectional profile change at different points along the centre line. Lateral pavement colour discontinuities often occur with medians, especially short ones. Other pavement colour discontinuities may also be caused by wet areas or road surface composite materials. These elements increase the complexity of a road area.

Other studies (Trinder *et al.* 1997, Vosselman and de Gunst 1997, Baumgartner *et al.* 1997, Wang *et al.* 1988) have constructed road models or rule bases with some parameters derived from training sets. The test images in these papers have relatively uniform areas, with same cross sectional profiles and similar colours. These methods need to be customized by adding more parameters to handle the high complexity of the road areas in this research. In this research, however, a method requiring as few parameters as possible without the need to build a training set is desirable.

The method described here is based on the mean-shift clustering approach (Comaniciu and Meer 1997) to segment the road area. The advantage of this method is that it globally finds the most significant feature in an image and does not rely on any training or query images. It uses a clustering method to group pixels with similar colours, regardless of their spatial relationships. When clustering, this method will first group pixels with the main colours of an image. These pixels account for most of the image. In a road image, they can belong to pavement and medians. The remaining pixels are probably due to small objects, such as traffic, shadows and different colours of road surface. The way to handle these small objects will be described later. In this study, several new rules are defined and combined with conditional morphological operations in order to improve the segmentation results and remove unwanted objects, such as shadows and occlusions.

2. Matching aerial photos with USGS DLG

Since the aerial images (source data) in this research and the digital road information from DLGs are from different sources, these images need to be georeferenced accurately before any recognition process is used. In this research,

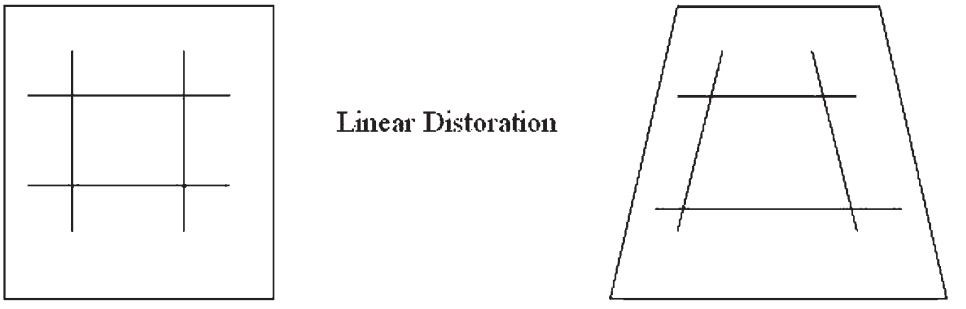

Figure 1. Linear distortion.

only the linear distortion was considered. An example of linear distortion is shown in figure 1.

Linear distortion is easily identified provided accurate correspondences in both types of data exist. It can be regarded as a planar projective transformation, which can be represented by a non-singular matrix (Hartley and Zisserman 2000) as shown below:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{1}$$

where (x, y, z) are coordinates in the USGS data and (x', y', z') are coordinates in the corresponding image. Since the source image and the DLGs data are planar, z and z' can be assigned a value of 1. Additionally, owing to the up-to-scale phenomenon (Hartley and Zisserman 2000), one of the nine elements in the non-singular matrix can be assigned a value of 1, e.g.

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix}$$

Therefore equation (1) becomes

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2}$$

The inhomogeneous form of equation (2) yields the corrected coordinates of (Hartley and Zisserman 2000):

$$\begin{aligned} x' &= (h_{11}x + h_{12}y + h_{13}/(h_{31}x + h_{32}y + 1)) \\ y' &= (h_{21}x + h_{22}y + h_{23}/(h_{31}x + h_{32}y + 1)) \end{aligned} \tag{3}$$

Given four pairs of measured data points, the transformation that maps the DLGs onto aerial images is:

$$\begin{pmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \\ x'_4 \\ y'_4 \end{pmatrix} = \begin{pmatrix} x_1, y_1, 1, 0, 0, 0, -x_1x'_1, -y_1x'_1 \\ 0, 0, 0, x_1, y_1, 1, -x_1y'_1, -y_1y'_1 \\ x_2, y_2, 1, 0, 0, 0, -x_2x'_2, -y_2x'_2 \\ 0, 0, 0, x_2, y_2, 1, -x_2y'_2, -y_2y'_2 \\ x_3, y_3, 1, 0, 0, 0, -x_3x'_3, -y_3x'_3 \\ 0, 0, 0, x_3, y_3, 1, -x_3y'_3, -y_3y'_3 \\ x_4, y_4, 1, 0, 0, 0, -x_4x'_4, -y_4x'_4 \\ 0, 0, 0, x_4, y_4, 1, -x_4y'_4, -y_4y'_4 \end{pmatrix} * \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{pmatrix} \quad (4)$$

If equation (4) is denoted by $X' = Mh$, both X' and h are 8×1 vectors, and M is an 8×8 matrix. Therefore, $h = M^{-1}X'$. A minimum of four points are chosen from the DLGs along with four points from aerial image. This gives the transformation between the USGS road data and the aerial image. Linear distortion between the two sets of photos for this project is represented by the non-singular matrix which is

$$H = \begin{pmatrix} 0.2504 & -0.0011 & -110.2432 \\ 0.0010 & 0.2491 & -117.4968 \\ 0 & 0 & 1 \end{pmatrix}$$

The most significant source of distortion is two-dimensional translation, because the first and second elements in the third row of H are zeros. Figure 2 shows the result of matching the USGS data to the aerial photography (the USGS roads are shown in red).

3. Segmentation

3.1 *Introduction to the mean-shift clustering method*

The mean shift algorithm for image segmentation was introduced in Comaniciu and Meer (1997). It is a non-parametric method for searching the real estimation of the

Figure 2. Matching of the USGS DLG and UCF aerial photos.

local maximum. The value of radius r decides the resolution of segmentation. Larger values of r result in lower segmentation resolution. In this paper, r is chosen to be the square root of the trace of the global covariance matrix in the HSI fields. The aforementioned trace is the summation of the variances of the three colour channels in an image. The square root of it approximately reflects how much the colour information of an image varies. The segmentation algorithm is defined as follows:

(1) Given an RGB aerial colour image, it is converted to an HSI colour image. The reason for this operation is that the correlation of the HSI three components (hue, saturation and intensity) is less than that of RGB (red, green, and blue). The converting process follows the one in Weeks (1996) as follows:

$$\begin{pmatrix} Y \\ R_y \\ B_y \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.701 & -0.587 & -0.114 \\ -0.299 & -0.587 & 0.886 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{5}$$

$$\begin{aligned} H &= \tan^{-1}\left(R_y / B_y\right) \\ S &= \sqrt{R_y^2 + B_y^2} \\ I &= Y \end{aligned} \tag{6}$$

(2) Next, a three-dimensional histogram of the HSI image is obtained with each dimension having 60 bins, which is an empirical choice.
(3) m points in the HSI image are chosen randomly.
(4) The point which corresponds to the largest value in the HSI histogram is identified as $\bar{x}=\{h, s, i\}$.
(5) Starting from this point, $\bar{x}=\{h, s, i\}$ in the HSI histogram, the shift vector $\Delta\bar{x}=(r^2\nabla p(\bar{x}))/(5p(\bar{x}))$ is found, where $p(\bar{x})$ is the value in the HSI histogram at $\bar{x}$.
(6) Next, $\bar{x}=\bar{x}+\Delta\bar{x}$ is computed. If $\bar{x}$ converges, go to step 7, otherwise go to step 5.
(7) Using $\bar{x}$ as the centre location, all points within the sphere of radius r in the HSI space are treated as one class and removed from the image.
(8) Steps 3–6 are repeated until all the points in the HSI image have been classified.

In step 5, a three-element shift vector is calculated. This method iteratively calculates the shift vector to find the local maximum. In this application, it is to find the cell in the three-dimensional histogram with the highest value.

After step 8, a number of classes are formed, each of which has an estimation vector, $\bar{x}$, of hue, saturation and intensity. Each class also has a number of pixels, which are classified belonging to it and have similar colour information. This algorithm can generate many classes. Some of them can have a relatively small number of pixels. They can be caused by noises, such as pavement markings and different pavement colours. Before these classes are further classified into pavement and non-pavement, as described in the next subsection, small classes should be merged to large classes. Therefore, the final step is to drop the classes whose number of pixels is less than 5% of the total number of pixels and to reclassify these pixels into one of the remaining classes which has the shortest Euclidean distance to these pixels. The threshold of 5% of the total number of pixels is used to determine whether a class is too small.

3.2 *Saturation difference between pavement areas and non-pavement areas*

Based on our observation of the aerial photos in this research, pavements usually have lower saturation, compared with other areas. This phenomenon can be observed by a sample image shown in the top of figure 6 (the original image is shown at the top of figure 5). Hence, it is possible to use saturation values to segment aerial images. The difficulty is that a saturation image can be very noisy. Sometimes a saturation image is so noisy that the road boundary is not recognizable. In this case, the classification results by the mean-shift algorithm can drastically improve the road segmentation process. The mean-shift algorithm classifies an image into several classes. The average saturation values of the classes separate them into the pavement classes and non-pavement classes.

For the road segmentation process, the mean-shift algorithm is:

(1) Sort the classes based on the average saturation in a descending order.
(2) For class i, where i is from one to n, where n is the number of the classes, classify it to be non-pavement if $i \leqslant t$, otherwise classify it to be pavement. Index t must satisfy $(\overline{s_t} - \overline{s_{t+1}})/\overline{s_t} > 0.4$, where s stands for the average saturation within a class, where the threshold value of 0.4 is used to detect the first significant drop in the average saturations.

3.3 *Conditional morphological operations*

Morphological operations are used to improve the segmentation results to remove unwanted pixels in pavement areas and to join disconnected regions in non-pavement areas. The operations of closing and opening were used (Weeks 1996). The opening operation removes small connected areas while preserving the shape and size of large objects in the image and the closing operation fills small gaps between disconnected regions. It is important to correctly judge where to use the opening operation and where to use the closing operation.

Even though the HSI segmentation in section 3.2 can result in poor binary segmentation images (pavement and non-pavement), the pavement area and the non-pavement area can still be easily recognized. It is at this point in the segmentation algorithm that the USGS road information is used again. Along the USGS road centre line, given a valid lateral offset, a path parallel to the road centre line is specified. Some pixels along each path are classified as non-pavement and some are classified as pavement. This process is illustrated in figure 3, where the orange line is the centre line and the two blue lines are two paths. If the pavement pixels are numbered 0 and non-pavement pixels are numbered 1, the summation of all the pixels along a path reflects the likelihood of whether this path is in a pavement or non-pavement area. As shown in figure 3, the most pixels on the path

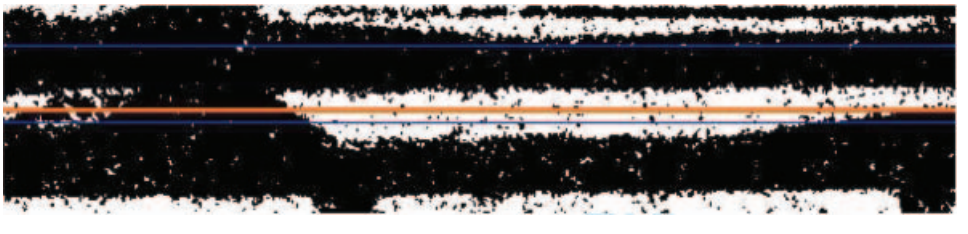

Figure 3. Illustration of the road centre line and paths.

above the centre line are classified as non-pavement and the aforementioned summation along this path should be relatively small. For the path below the centre line, most of its pixels are classified as pavement; therefore its summation should be considerably larger. Figure 4 shows a set of graphs, showing this value along the road lateral offset for the segmentation result on a road image.

The way to decide where the opening and closing morphological operations are to be used is to find two consecutive peaks and valleys, as shown in figure 4, where the crosses indicate these peaks and valleys. The middle points of these peaks and valleys give an approximation to the boundaries of the road (non-pavement, pavement, non-pavement, pavement, and non-pavement). In figure 4, these approximate boundaries are drawn as red lines. Closing operations should be performed in non-pavement areas and opening operations should be performed in pavement areas. The conditional morphological operations are used to remove the

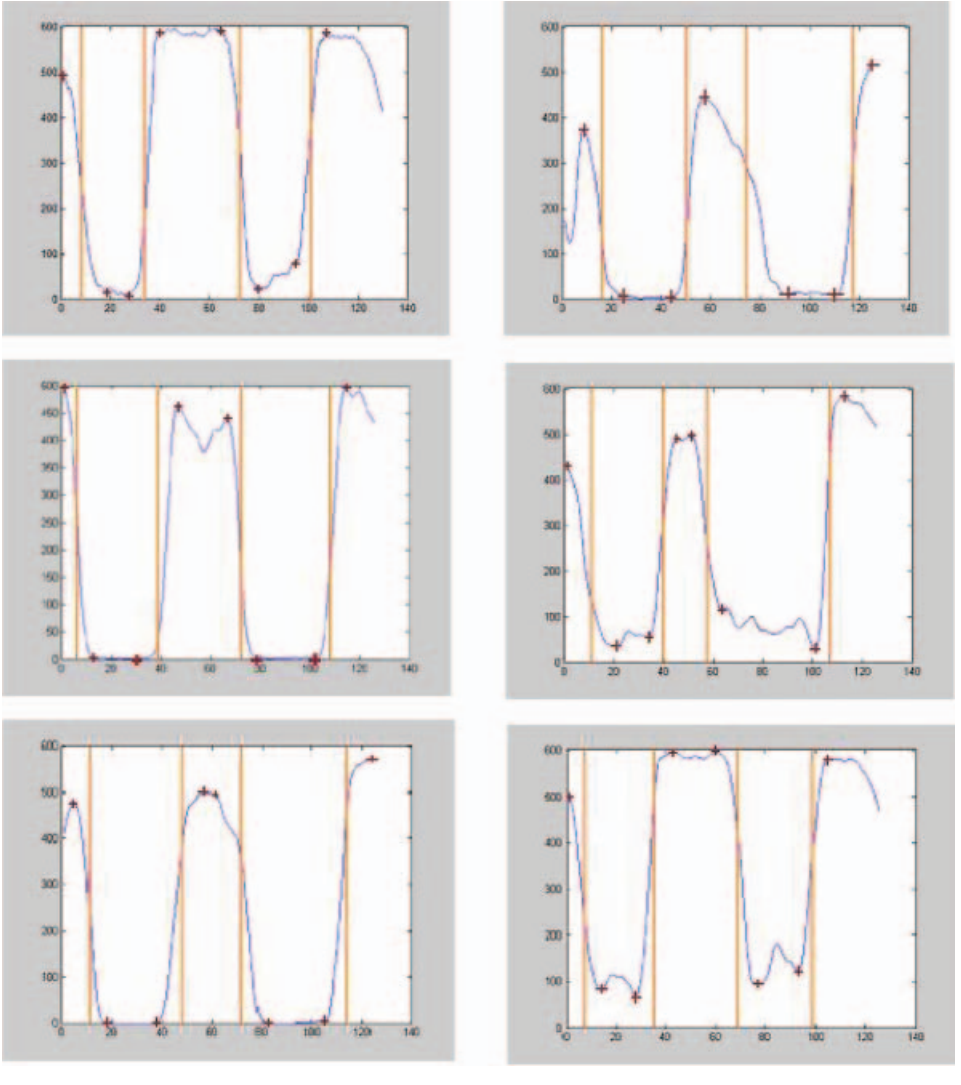

Figure 4. Numbers of non-pavement pixels versus lateral offsets.

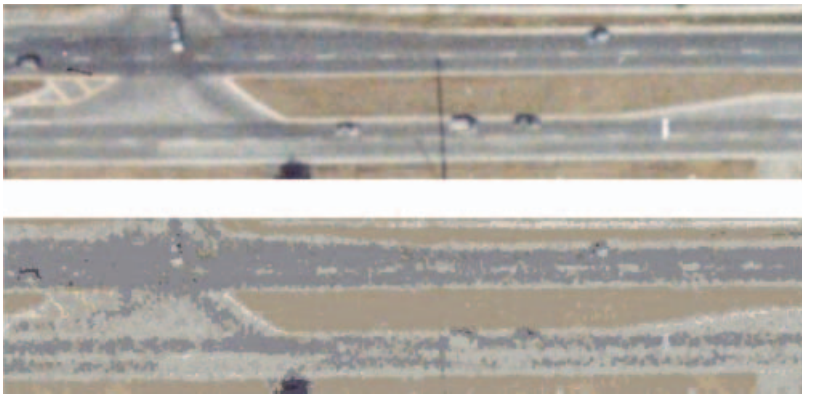

Figure 5. Road image and classification result by the mean-shift algorithm.

noise caused by overlapping colour information between pavement and non-pavement regions.

3.4 *Removing shadows and occlusion*

Some of the pavement area is covered by the shadow caused by trees or occluded by trees completely. The colours of trees and shadows are typically dark compared with other objects in an aerial photo. An example is shown in the first image in figure 7, where shadows and occlusions are surrounded by green contours. The method for removing shadows/occlusions relies on the clustering result in section 3.2 and the intermediate result in section 3.3. In section 3.2, the mean-shift algorithm classifies all the pixels into several classes. Since shadow and occlusion areas will have lower intensity than other objects in the image, they will be separated into a class with the

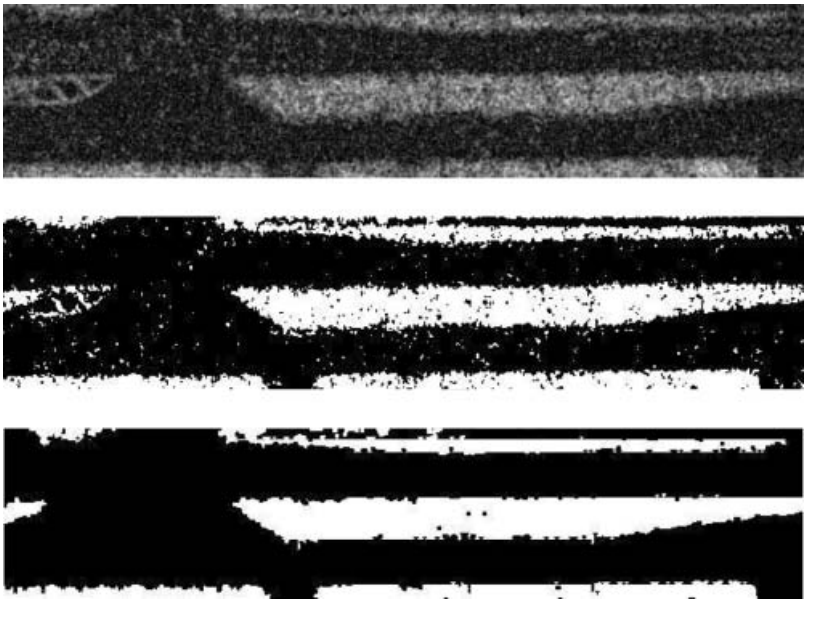

Figure 6. Saturation and segmentation results of the original image in figure 4.

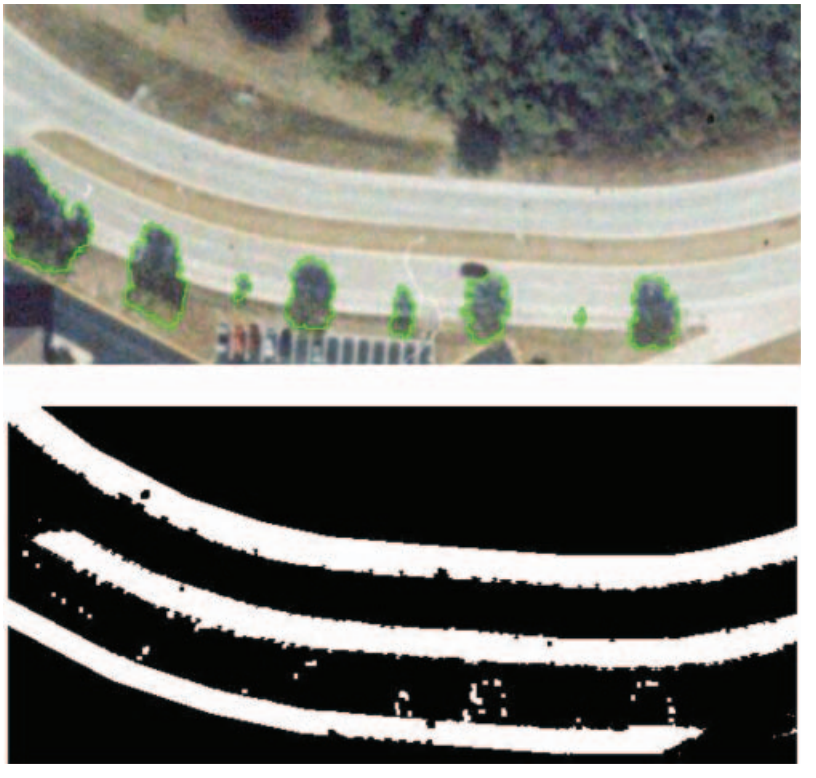

Figure 7. Road image with shadows and occlusions and its segmentation result.

lowest intensity. However, shadow areas or occlusion areas could cover both pavement areas and non-pavement areas, as shown in the first image of figure 7. In order to find out which part of a shadow/occlusion area is pavement, the intermediate result in section 3.3 is used. Based on the graphs in figure 4, the road cross sectional profile can be obtained. With this information, the shadow and occlusion areas can be classified. This process is to see whether a shadow/occlusion pixel falls in the pavement area or non-pavement area. The process is described as follows:

(1) The mean-shift algorithm in section 3.2 segments an aerial image into several classes. The class with the lowest average intensity is the shadow/occlusion class.
(2) Based on figure 4, a road profile is derived into non-pavement and pavement.
(3) For each pixel of the shadow/occlusion class, if its lateral position is within the lateral range of a pavement area, it is classified as a pavement pixel; otherwise, it is classified as a non-pavement pixel.

4. Results

4.1 *Segmentation results by the mean-shift algorithm*

The mean-shift algorithm classifies an image into several classes. The pixels in each class have the similar colour. In figure 5, two images are shown. The top image is the original image and the bottom image is the classification result, where each class is

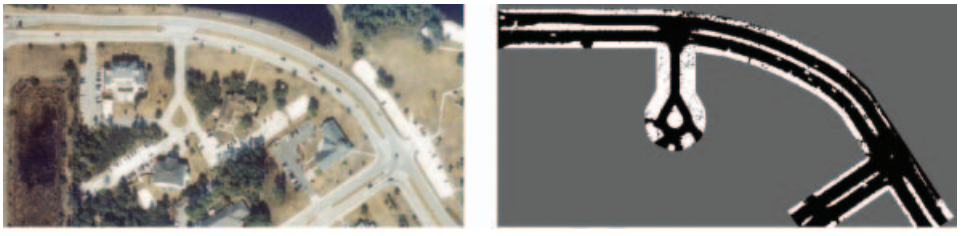

Figure 8. Overall view of road segmentation I.

demonstrated using the average RGB. To be specific, the mean-shift algorithm segments the image to several classes, each of which has a number of pixels. Each pixel in the second image of figure 5 is drawn using the average RGB of its class.

The saturation field is the next point of interest. The top of figure 6 shows the saturation of the original image in figure 5. As shown in this figure, the pavement area has a lower saturation than the non-pavement regions. Following the proposed segmentation algorithm given in section 3.2, the middle image of figure 6 is obtained. Finally, the morphological opening and closing operations are applied to the saturation segmented image as shown in the bottom of figure 6, resulting in a reduction in unwanted noise.

4.2 *Results of removing shadows and occlusions*

In the top of figure 7, there are several trees on the side of the road. Some of the pavement area is occluded by these trees or the shadow caused by them. The areas covered by shadows and occlusions are classified using the results of section 3.2 and the road cross sectional profile derived in section 3.3. The result of the proposed shadow and occlusion algorithm is shown in the bottom of figure 7.

4.3 *Results in the larger images*

The proposed algorithm was also tested in several larger images. This was to determine whether the algorithm can handle the variety of road appearances. The overall views of the road segmentation are shown in figures 8 and 9, where most of pavement and non-pavement are distinguished and all the medians and background are classified as non-pavement.

Figure 9. Overall view of road segmentation II.

Table 1. The accuracies of the segmentation results.

Image size	Accuracy
600×190	86.82%
600×190	86.87%
600×190	93.65%
600×190	93.78%
600×190	87.15%
600×190	88.05%
600×190	95.75%
2542×1475	91.78%
1652×806	90.56%

4.4 *Accuracy*

In order to measure the accuracies of the road segmentation results, they were compared with the results obtained manually using Acrobat PhotoShop. The results were used as the correct results. The accuracies are listed in table 1.

In table 1, all the accuracies are higher than 85%. The first seven images have relatively small sizes compared with the last two images, which are of a larger size. This comparison shows that image size does not affect the overall segmentation accuracies. When the size of an aerial photo increases, it is more likely that the image will have roads with different appearances, such as different colours, and more noise: traffic, trees and road colour discontinuities.

5. Conclusion

In this paper, a segmentation approach is introduced to segment roads from aerial images. The segmentation process is augmented using the USGS road centre line data projected onto the aerial images, allowing a considerable portion of unwanted image content ignored during the segmentation process; however, attention must be paid when applying the USGS data, because of linear distortions that might exist.

Even though the USGS data can help limit the area to be segmented, there are still many types of noise that can be present that reduce the accuracy of the final segmented image. The mean-shift algorithm is chosen to pre-classify all the pixels into several classes along with the saturation information to improve the overall segmentation process. Shadow areas are identified and unwanted noise is corrected using morphological operations.

Acknowledgement

This research is part of the ongoing research project of fast three-dimensional geo-specific road modeling for driving simulation at the University of Central Florida.

References

Baumgartner, A., Echstein, W., Mayer, H., Heipke, C. and Ebner, H., 1997, Context-supported road extraction. *International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Image*, 5–9 May 1997, Ascona, Switzerland.

Comaniciu, D. and Meer, P., 1997, Robust analysis of feature spaces: colour image segmentation. *International Conference on Computer Vision and Pattern Recognition*, San Juan, PR, pp. 750–755.

GRUEN, A. and LI, H., 1997, Linear feature extraction with 3-D LSB-snake. *International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Image*, 5–9 May 1997, Ascona, Switzerland.

HARTLEY, R.I. and ZISSERMAN, A., 2000, *Multiple View Geometry in Computer Vision*, pp. 69–74 (London: Cambridge University Press).

KARARZIS, H., SAHLI, H., PIZURICA, V. and CORNELIS, J., 2001, A model-based approach to the automatic extraction of linear features from airborne images. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(9), pp. 2073–2079.

KASS, M., WITKIN, A. and TERZOPOULOS, D., 1988, Snakes: active contour models. *International Journalof Computer Vision*, **1**(4), pp. 321–331.

LAPTEV, I., MAYER, H., LINDEBERG, T., ECKSTEIN, W., STEGER, C. and BAUMGARTNER, A., 2000, Automatic extraction of roads from aerial images based on scale-space and snakes. *Machine Vision and Applications*, **12**, pp. 23–31.

LIPAROL, C., TRIVEOI, M. and HARLOW, C., 1989, Geometric modelling and recognition of elongated regions in aerial images. *IEEE Transactions on System, Man and Cybernetics*, **19**(6).

NEUENSCHWANDER, W.M., 1997, Ziplock snakes. *International Journalof Computer Vision*, **25**(3), pp. 191–201.

STEGER, C., 1998, An unbiased detector of curvilinear structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, pp. 113–125.

TRINDER, J.C., WANG, Y., SOWMYA, A. and PALHANG, M., 1997, Artificial intelligence in 3-D feature extraction. *International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Image*, 5–9 May 1997, Ascona, Switzerland.

TUPIN, F., MAITRE, H., MANGLN, J.-F., NICOLAS, J.M. and PECHERSKY, E., 1998, Detection of linear features in SAR images: application to road network extraction. *IEEE Transactions on Geoscience and Remote Sensing*, **36**(2), pp. 434–453.

TUPIN, F., HOUSHMAND, B. and DATCU, M., 2002, Road detection in dense urban areas using SAR imagery and the usefulness of multiple views. *IEEE Transactions on Geoscience and Remote Sensing*, **40**(11), pp. 2405–2414.

VOSSELMAN, G. and DE GUNST, M., 1997, Updating road maps by contextual reasoning. In *International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Image*, 5–9 May 1997, Ascona, Switzerland.

WANG, F. and NEWKIRK, R., 1988, A knowledge-based system for highway network extraction. *IEEE Transactions on Geoscience and Remote Sensing*, **26**(5), pp. 525–531.

WEEKS, A.R., 1996, *Fundamentals of Electronic Image Processing*, Chap. 7.4, pp. 228–294, pp. 333–347 (Bellingham, WA: SPIE Optical Engineering Press/Piscataway, NJ: IEEE Press).

ZHANG, C., BALTSAVIAS, E.P. and GRUEN, A., 2001, Updating of cartographic road databases by image analysis. *International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Image*, 10–15 June 2001, Ascona, Switzerland.