Automatic Extraction of Main Road Centerlines from High Resolution Satellite Imagery Using Hierarchical Grouping

Xiangyun Hu and Vincent Tao

Abstract
Automatic road centerline extraction from high-resolution satellite imagery has gained considerable interest recently due to the increasing availability of commercial high-resolution satellite images. In this paper, a hierarchical grouping strategy is proposed to automatically extract main road centerlines from high-resolution satellite imagery. Here hierarchical grouping means that, instead of grouping all segments at once, the selective segments are grouped gradually, and multiple clues are closely integrated into the procedure. By this means, the computational cost can be reduced significantly. Through the stepwise grouping, the detected fragmented line segments eventually form the long main road lines. The proposed method has been tested and validated using several Ikonos and QuickBird images both in open areas and build-up urban environments. The results demonstrate its robustness and viability on extracting salient main road centerlines.

Introduction
The demand for road database generation and updating from high-resolution satellite imagery is increasing dramatically due to its high spatial resolution (1 to 4 m), fast orbit repeatability, rich multi-spectrum information and stable, affordable acquisition cost. Such a promising end-user market drives the need for automating road information extraction from high-resolution satellite imagery.

Various linear feature extraction methods have been developed by the photogrammetry, remote sensing, and computer vision communities in the past decades. Most of them were developed mainly for general linear feature extraction, while some are particularly designed for road extraction. In general, these methods can be classified into four categories:

1. Linear feature detection methods: Roads appear as ridges or valleys of grey value function of an image. Road finding can be considered as a process consisting of ridge finding, ridge points linking, and road segments forming. Image filtering or image convolution using edge or ridge templates is a classic approach (Nevatia and Babu, 1980). Based on the analysis of the scale-space behavior of a line profile, Steger (1998) and Mayer and Steger (1998) presented an unbiased linear feature detector to detect road lines and their width. The road lines have different lateral contrast. Vosselman and de Knecht (1995) used a profile matching method derived from the road radiometric profile for road extraction.

2. Rule or knowledge-based methods: This is so called 'high-level' processing compared to pixel based low-level and intermediate level processing. In order to handle the issues of linear feature alignment and fragmentation, heuristic or rule-based methods were developed by Trinder and Wang (1998). Tönjes and Grove (1998) employed a semantic-net technique to combine information from multiple sensors for road extraction.

3. Optimization-based methods: In this category, road extraction is treated as an optimization problem. Dynamic programming is a popular method. It is applied to formulate the linkage of candidate points (Gruen and Li, 1995). Snakes (active contour) and least square template matching have also been applied to automatic (Laptev et al., 2000) or semiautomatic road extraction from mono and stereo images (Gruen and Li 1997; Hu et al., 2004), and image sequences (Tao et al., 1998).

4. Content-supported and map guided methods: Using image features alone, a road network can hardly be extracted completely and correctly. Contextual information, such as buildings, trees, rivers, is obviously valuable for road identification and road segment linking. Baumgartner et al. (1999) developed a method for road extraction from multi-scale images and discussed the role of grouping using contextual information. Map guided feature extraction is also considered in this category. Existing maps can provide important clues for road finding. Some algorithms have been developed to perform image-map matching for road identification. Fiset et al. (1998) used a multi-layer perceptron to extract road segments from swir-sav panchromatic images. Stilla (1995) described a method to find new roads based on the assumption that new roads are always connected to the old ones. Agouris et al. (2001) proposed a method that detects the changed road segments using differential snakes for image-based us updating.

It has been realized that robust feature extraction is largely dependent on the strategy developed and the constraints employed (Tao et al., 2001). Many factors such as the objective, image scale or resolution, image quality, and image complexity need to be taken into account. We have noticed that more and more methodologies are based on
hybrid strategies (Heipke, 1996; Mayer and Steger, 1998; Bazohar and Cooper, 1996; Laptev et al., 2000; Coulouigner and Ranchin, 2000; Hinz and Baumgartner, 2003; Zhang, 2004). Some recent reviews and progress on automatic road extraction can be found in Quackenbush (2004); Mena (2003), and a special issue of Photogrammetric Engineering & Remote Sensing, Volume 70, Number 12.

As introduced, due to the complexity of the problem, it is extremely hard to develop a general system that applies to automatic road extraction from various images. On the other hand, demand for road database generation and updating from high-resolution satellite imagery is increasing dramatically due to the advantages. Our research objective is to automatically extract main road centerlines from high-resolution satellite imagery (i.e., 1 to 4 m ground resolution), in which the extraction result is expected to be comparable to that collected by a human operator. A main road is a major road for any form of transport and is a very important component of infrastructure information in GIS. This paper proposes a hierarchical perceptual organization strategy for grouping detected road primitives. In the next section we will give an overview of our approach, followed by the descriptions of the approach. Experimental results are demonstrated. Conclusions are then drawn and possible directions for future research are also discussed.

A Perceptual Grouping Strategy for Main Road Centerlines Extraction

Figure 1 shows the workflow of our method. Our strategy for main road centerlines extraction is based upon a fundamental question in computer vision: “How do we (i.e., the human visual system) perceive objects (i.e., roads) in an image?” When we look at (not see) an image, we can perceive most of main roads immediately. However, building a computational model to simulate the above process realized by our human visual system is very difficult; otherwise, feature extraction would have become straightforward in computer vision. In our common experience, we always perceive the salient linear features globally with little reasoning or high-level knowledge involved. Just looking at the drawing of the primitives in the middle of the figure, in a global and instant way, we can easily and reliably perceive/find the salient long collinear features: they mostly consist of the main road centerlines. Based on this observation, we developed a hierarchical approach for grouping or linking the fragmented lines into long collinear features globally and efficiently. As shown in Figure 1, the core step for road extraction is to automatically find the main road lines from the fragmented road segments that are detected from the image.

Main Road Detection

Main roads in high-resolution satellite images often appear as salient ribbon-like features. Taking into consideration geometric, radiometric, and topologic properties of features, a main road model is developed. We call this kind of road a ribbon road. A fast and reliable ribbon road detector (Hu and Tao, 2005) was employed to detect the main roads and form the candidate road segments.

A two-level image pyramid is generated according to the main road width so that, at the top level of the pyramid, the main roads are three to five pixels wide. Mean-value filtering is used for pyramid generation. For example, if the predefined main road width is about 20 m (20 pixels when the resolution is 1 m), we can use a 4 by 4 pixels mask to perform a moving average for generating the top level of the pyramid. The road detection and perceptual grouping are carried out at the top level with the lower resolution.

Multi-scale techniques are frequently used for image analysis. Using a “coarse-to-fine strategy” substructures and noise can be removed at the top level, while details at the higher resolution level are preserved to allow verification and precise delineation. However, one weak point is that when errors occur in higher levels, the errors will propagate to the higher resolution levels. To deal with wrong results in the top level, we remove isolated and short lines, because, according to our model, they are not likely to be a part of the main road network. By this relatively simple assumption, many incorrectly extracted lines can be eliminated. But there still might be incorrect elimination of lines in complicated scenes. Another drawback is missing main roads from the top level.

To handle it, one method is to develop a more detailed main road model which will guide the process of verifying the results, where more contextual information (i.e., lane markings, DEM/DSM, existing maps) could help find the missing roads. Due to the emphasis placed on grouping salient lines in this paper, we consider these issues as future work.

Perceptual Organization for Main Road Extraction

Perceptual organization is defined as an ability to explore a structuralized feature organization from sensory data (Sarkar and Boyer, 1993; Boyer and Sarkar, 1999). The Gestalt psychologists have found a set of important properties in perceptual organization, namely, proximity, continuity, similarity, closure, and symmetry (Rock and Palmer, 1990). Proximity, continuity, and similarity, can be used as primary constraints for linking fragmented road segments:

1. Continuity: Main roads are continuous and gaps between segments are likely to be bridged;
2. Proximity: Shorter gap length (closer segments) indicates higher probability for linking;
3. Similarity: Similar direction, image grey value, and other collinear characteristics make linking more likely. These constraints compromise both geometric and radiometric similarities.

Some grouping methods use hierarchical, voting, and probabilistic methods (Sarkar and Boyer, 1994; Crevier, 1999). Concerning the hierarchical approach, it is assumed that coherent global structures gradually emerge from local features. Bold et al. (1989) apply hierarchical grouping for straight-line extraction. At each step, the grouped tokens are replaced by new tokens having additional emergent characteristics. A similar method is used by Dolan and Weiss (1988), where proximity and continuation are employed to
identify co-curving and collinear structures. Crevier (1999) used a probabilistic model to perform perceptual grouping of collinear chains of straight segments. This method enables combination of incommensurable information sources (e.g., geometric and radiometric). Jacot-Descombes and Pun (1997) presented an asynchronous grouping approach that processes a data flow of ordered contour primitives, with the rank in the flow determined according to a saliency measure. As a hybrid strategy from the above methods, we have developed a hierarchical grouping approach illustrated in Figure 2.

The grouping consists of two steps made up of three iterative procedures: (a) grouping straight collinear segments into longer straight segments and then into polyline segments; and (b) grouping additional curved segments that are close to the grouped polyline segments. The proposed strategy has clear advantages in terms of reliability and computational cost: (a) Major (or global) and more structuralized features are perceived prior to smaller and disorder ones. (This has been proven by psychological experiments (Chen, 1982 and 1989); we, first group the more structuralized tokens, i.e., straight segments, into longer salient structures, and then group the additional curves), and (b) At each grouping step, possible collinear chains are extracted and evaluated. New segments are formed by bridging the gaps, and a new link matrix can be created until there are no new links found. With each iteration, the segments get longer and longer, thus the perceptual range becomes farther and farther, the number of grouped tokens (segments) becomes fewer and fewer, and the global structure of the main road centerlines gradually emerges. The key component of this method is the computation of the link probability \( p_{ij} \). Based on Gestalt laws, perceptual properties (proximity \( pro \), continuity \( con \), and similarity \( sim \)) serve as the primary linkage clues \( f_{coll} \). In our method, the linking clues \( f_{coll} \) found in gaps between two segments are also considered. So, we compute the link probability by

\[
p_{ij} = f_{coll}(pro, con, sim) + f_{clu}.
\]

Figure 4 shows the geometric relationship between the two segments Seg and Segj that are used to compute the collinear characteristic \( f_{coll} \). A similar method was used by Crevier (1999) for grouping straight-line segments. The meanings of the symbols (i.e., \( \theta_1, \theta_2, g_1 + g_2, \ldots \)) can be directly referred from the figure. Proximity \( f_{prox} \) are described by the transverse gap \( g_t = g_1 + g_2 \), and the longitudinal gap \( g_l = g_1 + g_2 \). Mainly, continuity \( f_{cont} \) is described by the angle \( \theta \) between the two segments and the collinear measurement \( \theta = \theta_1 + \theta_2 \). Similarity \( f_{sim} \) also refers to the difference in segment characteristic \( \Delta p = |p_i - p_j| \). The \( p_i \) and \( p_j \) indicating the characteristic of Seg and Segj, respectively. The \( p_i \) and \( p_j \) are

**Implementation of Hierarchical Grouping**

Before the hierarchical grouping, the extracted segments are classified to two categories based on the geometric characteristic \( E_c \). \( E_c \) is a ratio of the distance from the start point to the end point of the segment and the actual length of the segment. If \( E_c > 0.95 \), hence the segment is a straight line; otherwise it is treated as a curve segment. The grouping consists of three procedures: grouping straight collinear segments, grouping polylines, and grouping additional curves.

---

**Figure 2. Hierarchical grouping for main road centerline extraction.**

**Figure 3. Collinear grouping workflow.**

**Figure 4. Geometric relationships used to compute collinear attributes (Crevier, 1999).**
evaluated by the model-based verification method (Hu and Tao, 2005) using their geometric and radiometric properties. Equation 2 computes the collinear characteristic \( f_{\text{coll}} \):

\[
 f_{\text{coll}} = c_0 \cdot f_{\text{pow}}(g_t, g_l, l_1 + l_2) + c_1 \cdot f_{\text{col}}(\alpha, |\theta_1 - \theta_2|)
 + c_2 \cdot f_{\text{sim}}(\Delta p).
\]  

At each grouping procedure, to simplify the computation, obvious non-collinear links can be found by the collinear measures \( g_t \), \( g_l \), \( l_1 \), and \( l_2 \). For example, when grouping straight segments, a linkage pair is rejected (directly set \( p_{ij} \) to 0) if \( \alpha > \pi/12 \) or \( \theta > \pi/6 \). The weighted sum function Equation 2 is composed of three items in which the weight coefficients \( c_0 \), \( c_1 \), and \( c_2 \) are defined by:

\[
 c_0 = 0.8 - 0.5 \cdot \alpha/(\pi/12.0)
\]

\[
 c_1 = 0.5 \cdot \alpha/(\pi/12.0)
\]

\[
 c_2 = 0.2.
\]

Equations 3 and 4 indicate that when the angle similarity is high, the proximity is dominant because in this case only the transverse gap \( g_t \) and the longitudinal gap \( g_l \) is used for estimating the collinear characteristic (e.g., if \( \alpha = 0 \), but \( g_t \) or \( g_l \) is large, then the two segments are not likely to be collinear). Short segments removed before the grouping can serve as important linkage clues, particularly when the gap is long (see Figure 5). If the geometric relationship indicates it is possible to link the pair (if \( f_{\text{coll}} > 0.6 \), it works well in our experiments), Hough transform is applied to detect the short segments and their length. The linkage clue is evaluated by the length \( l \) of the detected short line. The link probability is defined as:

\[
 p_{ij} = 0.5 \cdot f_{\text{col}} + 0.5 \cdot f_{\text{clu}}(l), \quad \text{when } f_{\text{coll}} > 0.6
 p_{ij} = f_{\text{coll}} \quad \text{when } f_{\text{coll}} \leq 0.6.
\]

All items in Equations 2 and 6 are normalized to derive the linkage probability \( p_{ij} \). At each grouping step, we link the two segments with a high \( p_{ij} (p_{ij} > t, t = 0.65) \). We found that a large threshold \( t \) leads to less wrong linkages (higher correctness), but can lead to miss-linking. On the contrary, a small \( t \) results in more linkages (higher completeness), but they are probably wrong. A proper \( t \) is found empirically by seeking a balance of correctness and completeness. For simplicity, we only extract collinear chains along the route made up of the local maxima \( p_{ij} \). That means that when \( \text{Seg}_i \) has multiple links to other segments, only the linkage with maximal \( p_{ij} \) is chosen. Figure 6 shows the detected line segments and the corresponding grouping result at the top level. The black lines in Figure 6b are the extracted centerlines. It shows that the grouping successfully perceives the main road centerlines from a number of curves and straight segments (Figure 6a).

After the hierarchical grouping has been performed, topology analysis is used to remove short and isolated segments that are not likely to be a part of the main roads according to our model. The intersections among long lines can be detected by intersecting the extracted lines. The short segments (the length is less than a limit) not connected with long lines are directly removed. The image of the original resolution is not used, but might provide detailed information for the evaluation of the result and for a precise delineation.

**Experiments Using Ikonos and QuickBird Imagery**

Figure 7 displays a comparison between non-hierarchical and hierarchical grouping for main road extraction. Non-hierarchical grouping groups all tokens (straight and curve segments) in a single iterative procedure. An Ikonos panchromatic image of size 7,600 by 6,000 pixels has been used. The pixel resolution is 1 m. As shown in Figure 7a, the number of grouped tokens changes with iteration times.
Figure 7. A comparison between non-hierarchical and hierarchical grouping.
The computation time when using non-hierarchical and hierarchical grouping is 78 seconds and 18 seconds, respectively (the computer used is a PC with 1.9 GHZ CPU and 512 MB RAM, time for generating the image pyramid is not included). The final number of tokens is the result of the topology analysis by which the isolated short segments are removed. Compared to non-hierarchical grouping, hierarchical grouping reduces the computational cost in terms of space and time significantly. Figure 7b displays the comparison between results using the two different grouping strategies. It indicates that non-hierarchical grouping likely leads to more incorrect linkages, therefore finally misses some centerlines. Hierarchical grouping extracts most of the long and straight roads, as they are the ones with high scores. The scores are computed by the model-based method for road verification (Hu and Tao, 2005). This is illustrated by the numbers adhered to the long lines the image in Figure 7b. Based on visual inspection, the result is similar to that perceived by the human visual system from the extracted fragmental segments. However, as marked by an arrow in Figure 8, occlusion by clouds or other noise may cause some gaps.

We also tested our method using QuickBird images. Figure 8 presents extraction results from 2.44 m multispectral images. The image content is more complex as they are built-up areas. However, our approach extracts most of the main road centerlines. Figure 8a shows that the grouping method correctly produces a correct result even in a relatively complex environment. We also noticed that properly adjusting the processing parameters (i.e., the thresholds in the center point detection and linkage probability) leads to missing linkages or wrong linkages of relatively non-salient main road parts, yet the most salient centerlines were still extracted reliably.

To compare the results with those collected by a human operator, we evaluated completeness and correctness of the road extraction. The main road centerlines extracted manually serve as the reference data. The completeness and correctness have been used for the evaluation of road extraction (Fischler and Heller, 1998; Wiedemann 2002). The completeness measure \( m_{\text{completeness}} \) is the percentage of the reference roads extracted by our method:

\[
\text{m}_{\text{completeness}} = \frac{L_{\text{match}}}{L_{\text{ref}}},
\]

where \( L_{\text{match}} \) is the total length of matched (correctly extracted) roads, and \( L_{\text{ref}} \) is the total length of all reference roads. The correctness measure \( m_{\text{correctness}} \) is the percentage of correctly extracted roads with respect to all extracted ones:

\[
\text{m}_{\text{correctness}} = \frac{L_{\text{match}}}{L_{\text{ext}}},
\]

Figure 8. Main road extraction from QuickBird 2.44 m resolution imagery.
TABLE 1. QUALITY MEASURE ON MAIN ROAD EXTRACTION COMPARED WITH HUMAN COLLECTED DATA

<table>
<thead>
<tr>
<th>Image Figure</th>
<th>Image Figure 8a</th>
<th>Image Figure 8b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>Correctness</td>
<td>0.94</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Here, $L_{\text{ref}}$ is the total length of all automatically extracted roads. To get the matched roads, we set a tolerance distance (5 pixels) from an extracted centerline to a reference centerline, as there might be alignments for the extracted centerline. Table 1 shows the quality measures of the extraction from the three tested images in Figure 7, Figure 8a, and Figure 8b. Geometric accuracy has not been analyzed. From Table 1, we can see that concerning completeness and correctness, the automatically extracted results are close to the ones collected by a human operator. The comparable similarity to manual collection indicates the potential of the proposed method. Figure 8b shows the limits of the approach. There is a relative lower completeness, caused by the missing roads (pointed out by black arrows) and a lower correctness, caused by the incorrectly extracted roads (pointed out by white arrows). It results from undetected road segments due to the complexity of the image scene. On the other hand, segments close to the image borders or corners are more likely to be removed because they have relatively less linkages than those in the center of the image.

Conclusions and Future Work
In this paper, we proposed a hierarchical grouping strategy to extract main road centerlines from high-resolution satellite imagery. We attempt to reliably find the main roads by making use of perceptual organization. Hierarchical grouping plays a key role in extracting the collinear chains from the fragmented segments. Compared to non-hierarchical grouping, a selective and sequential grouping of tokens is more reliable, and it reduces the computational cost significantly. By the stepwise grouping of the fragmented primitives, the global structure of salient road centerlines eventually emerges. Multiple clues are also combined to evaluate the linkage probability. All the above algorithms are fully integrated so as to maximize the reliability of the extraction. We tested an Ikonos 1 m and a QuickBird 2.44 m image both in open areas and residential environments. Based on these tests, our approach is capable of automatically extracting main road centerlines. The results are comparable with the results collected by a human operator.

Regarding future work, we are particularly interested in:

- Developing a more detailed road model including contextual information will be helpful for improving the results. We used a relatively simple model and method for verifying the extraction results of the top level of the image pyramid, which leads to some incorrect results and missed roads. Integrating the detection results from different image levels should provide more evidences for forming a more complete road network. Information fusion is one of the key problems for future research;
- Making use of additional sources and integrating more clues for grouping and verification, such as existing map information, spectral information (both Ikonos and QuickBird come with spectral information), digital elevation/surface information (e.g., from airborne lidar), road texture, as well as contextual information (e.g., vehicles).

Acknowledgments
The authors thank the support of a GEOIDE project of Canada “Automating 3D Feature Extraction and Change Detection from High Resolution Satellite Imagery with Lidar and Maps” and a Nature Science Foundation of China project (No. 40337055) “Modeling and methodologies for updating GIS spatial database.”

References


