A Parameterization Model for Transportation Feature Extraction

Michael E. Hodgson, Xingong Li, and Yang Cheng

Abstract
This article presents one solution to the parameterization problem for automated road feature extraction models. Using a line-length buffer approach, a measure of agreement between extracted road features and reference road features is defined. An automated approach for deriving this agreement index is presented and implemented as a parameterization model. Coupling the implemented parameterization model with an existing transportation feature extraction model is demonstrated. Although the solution was designed for a road feature extraction model (FEM), the conceptual design could be applied to other linear FEMs, such as models for streams, fault lines, and isolines.

Introduction
Calibration of the parameters in a road feature extraction model is typically performed through an iterative staged approach using visual comparisons. The stages in this visual approach are (1) setting of the model parameters, (2) extraction of features from the model run, (3) visual comparison of results to a reference set, and (4) a repeat of the first three stages numerous times. The “best” parameterized model is concluded when the agreement (or accuracy) between model results and reference data is maximized. Visible comparisons are favored because of the endemic mis-registration problems and line-to-line correspondence problems plaguing an automated approach. A sparse sampling of the parameter space has been favored because of the combinatorial problems of multiple parameters and the limitation of human efficiency for evaluating model results by visual analysis.

Although automated methods for parameterizing road feature extraction models have been argued (Wang, et al., 1992), little research has focused on this issue. Conceptual and implementation impediments have prevented automated approaches for comparing the model results and reference linear features. An automated solution to the parameterization of road feature extraction models would enable a broad sensitivity analyses to be conducted using multiple parameters and range in values. For instance, how well the feature extraction model performs in other geographic areas, image types, and spatial resolution is of interest. As most road feature extraction models are evolutionary (i.e., under continual refinement), such an automated solution would also enable efficient assessments of each model refinement.

Automated approaches for parameterizing such models or conducting sensitivity analyses would obviously require an automated method for evaluating model performance (i.e., an accuracy assessment). This article presents one solution to the problem of an automated parameterization of road feature extraction models. This research builds on work in assessing spatial accuracy of linear features by Goodchild and Hunter (1997) using a line-length buffered approach. However, in the current research the assumptions and unknowns of Goodchild and Hunters’ buffered line method for accuracy assessment are inverted.

In this article, the conceptual nature of the accuracy assessment problem is first presented. Next, the proposed line-length buffered approach is presented along with the measure of agreement. The parameterization model was implemented in an embedded GIS approach and demonstrated using the road feature extraction model (FEM) from the Oak Ridge National Laboratory (ORNL). The algorithmic logic behind the ORNL FEM is briefly presented.

Parameterization and Accuracy Assessment Problem
Approaches for extracting transportation features from remotely sensed imagery may be categorized into autonomous methods and interactive (i.e., semi-autonomous) methods. Interactive methods, typically more successful, are guided by human operators that “train” the algorithms spectrally, spatially, or topologically. In some implementations, the interactive methods require frequent nudges and guidance for many features in an image. The parameters in interactive methods may be continually changed. Autonomous methods require no user input once they are parameterized. Autonomous methods, however, require careful prior parameterization for the geographic area and specific imagery.

Parameterizing a model is performed by conducting accuracy assessments on the output of different parameter combinations. The conceptual goal in the accuracy assessment is to compare one extracted feature to one reference feature (or set of extracted features to a set of reference features) and assess the level of “agreement” between features. If the reference feature is considered “truth” the analysis is an accuracy assessment, otherwise it is considered a measure of agreement. This distinction is important for those studies focused on change-detection or a better representation of the transportation features, as in versioning. This study assumes the transportation reference features are “truth.”

In the context of transportation feature extraction models, the three logical choices for conducting an accuracy assessment would be a: (1) visual comparison, (2) automated

Michael E. Hodgson is with the Department of Geography, University of South Carolina, Columbia, SC 29208 (hodgsomm@sc.edu).
Xingong Li is with the Department of Geography, University of Kansas. Lawrence, KS 66045.
Yang Cheng is with the Machine Vision Group, Jet Propulsion Laboratory, Pasadena, CA 91109.
comparison (in a raster model), or (3) automated comparison (in a vector model).

Visual accuracy assessments suffer from the same problem of quantitatively measuring success as do digital approaches; for example, determining which of the following characteristics define correctness:

- geometry (e.g., length of feature, number of nodes in feature curvature)
- number of distinct segments (e.g., road composed of n-segments with intersections)
- continuity (i.e., no gaps)
- spatial accuracy (i.e., deviation of extracted feature from reference feature)
- commission errors (i.e., incorrectly identified road segments)

The following example questions further elaborate on these issues. Is the correct identification of a 100 meter road segment considered equally important as the correct identification of a 1600 meter road segment? If the reference road is represented by two extracted segments with a small gap, is this considered a success, partial failure, or failure? If the reference feature is defined from a different scale (i.e., spatial resolution) of imagery than the imagery used for automated feature extraction, how should commensurate spatial differences in geometry and spatial accuracy be considered? Are these geometric or spatial differences caused by scale-considered failures? If more than one of the measures above are used are the measurements weighted equally? There is no agreement on above measures of success, particularly, the last question on weighting each measure.

A visual comparison between the model results and the reference data is the common approach used in parameterizing any extraction model. First, parameter settings are set, and the model is run. The output road features are visually compared to the reference source. Based on the results, a new set of parameter values is set, and the process is repeated. Such an approach is commonly used in selecting training sites for supervised classification procedures. The visual comparison is appropriate as a guide to parameterization if: (1) the number of parameters is relatively small, and (2) the “accuracy” of the extracted features can be adequately assessed. The sheer number of iterations involved with a large set of parameters, many of which are co-dependent, would prohibit the visual approach. Yet, the visual approach clearly dominates research in transportation feature extraction as alternatives are less desirable. Thus, the evaluation of model sensitivity to imagery and geography or the evaluation of a large range/combination in parameter values is compromised.

A digital comparison of the transportation pixels identified in the image and the reference transportation pixels would solve the computational problem associated with the number of parameters. This approach may be implemented as a map overlay in a raster data model. As in the visual approach, the iterative method of setting parameter values, comparison of extracted and reference features, and resetting parameters for a new run, is still followed. However, as with all pixel-to-pixel comparisons, it assumes the two maps are in perfect, spatial registration. In most real-world applications this spatial agreement is seldom true. Assessing accuracy of linear features using a raster approach is thus, seldom performed.

A digital comparison of extracted roads to reference roads in vector form is generally the ideal, as extracted roads typically populate a database in a vector model form. However, because the precision in representing the spatial location of critical points is so great, a comparison of vector themes using the critical points in corresponding lines will suggest nearly total disagreement. Coincident-paired points in the reference and FEM extracted line will rarely occur. One coverage may be more generalized than another, resulting in a different number of points for the reference line than the extracted line.

Some related work in feature extraction using “snakes” (Agouris, et al., 2001) is based on matching extracted features to pre-existing reference features using energy measures: spacing of nodes, curvature, and local gradient. It is possible that this approach could be used as the assessment approach within a parameterization model. The implicit assumption and resulting problem for applying these snake measures to a holistic accuracy assessment of multiple features is that each pair of reference road and extracted road must be compared in isolation from all other reference roads and extracted objects. In other words, the specific road portion to “match” to must be known. Problems may occur if the FEM road segment is incomplete. Issues of scale differences may also complicate an autonomous methodology for assessments.

**Parameterization Model**

The proposed solution to the spatial registration problem and point-matching problem is to use a line length comparison inside a buffered reference theme. This approach addresses many of the characteristics of linear accuracy listed earlier. Although not a perfect solution for a final accuracy assessment, it is a tractable solution to the problem of model parameterization. In the line-length buffer approach, a buffered distance commensurate with the spatial registration accuracy is specified (Figure 1). The total length of the extracted features located inside the buffered reference theme is compared to the total length of the reference lines inside the same buffer. A buffer is specified (Figure 1). The total length of the extracted features located inside the buffered reference theme is compared to the total length of the reference lines inside the same buffer. One index of accuracy is the traditional “overall accuracy” value (also referred to as producer’s accuracy) computed as:

$$\text{Overall Accuracy (\%)} = \frac{\text{Roads}_{\text{FEM}}}{\text{Roads}_{\text{Ref}}} \times 100$$

where,

- $\text{Roads}_{\text{FEM}}$ = Total road length found by Feature Extraction Model (FEM),
- $\text{Roads}_{\text{Ref}}$ = Total road length in Reference Data Set.

With line length as the measure of correctness, however, it is possible for the overall accuracy to be greater than 100 percent. Even if the extracted feature set of roads are indeed essentially the same as the reference features, the extracted features may

![Figure 1. The buffer-based calibration method for estimating the percentage of agreement between a reference linear transportation feature and candidate road segments extracted from an automated feature extraction program.](image-url)
have a length slightly greater or shorter than the length in the reference set because of scale differences. More importantly, the percentage may be over 100 percent if the extracted features include commission error.

The overall accuracy in Equation 1 is a measure of only omission error. A comparison of extracted features and reference features should also consider the commission error. In this study a measure of agreement is used that considers omission and commission errors. The formula used for considering both types of error simultaneously is defined as:

\[
\text{Road Feature Agreement} (%) = \frac{\text{Roads}_\text{FEM} \cap \text{Roads}_\text{Ref}}{\text{Roads}_\text{FEM} + \text{Roads}_\text{Ref} - \left(\text{Roads}_\text{FEM} \cap \text{Roads}_\text{Ref}\right)} \times 100
\]  

(2)

where,

- \(\text{Roads}_\text{FEM}\) = Total road length found by Feature Extraction Model (FEM)
- \(\text{Roads}_\text{Ref}\) = Total road length in Reference Data Set
- \(\text{Roads}_\text{FEM} \cap \text{Roads}_\text{Ref}\) = Total FEM road length in Buffered Area of Reference Data Set

This index for feature agreement is bounded between 0.0 and 1.0. Interpretation is as follows:

- A value of 0.0 means no features in the reference data set are represented in the set extracted by the FEM.
- A value of 1.0 means the FEM set is identical to the population in the reference data set and neither set contains additional transportation features not in the other set.

The index may also be interpreted as the percentage of the total features in both sets that is expressed in the common feature set.

The method of using a buffered reference line as the "true line location" was proposed by Goodchild and Hunter (1997) for evaluating the spatial accuracy of a linear feature. Their approach was to determine the spatial accuracy measured in circular statistics (e.g., circular error probable) for the linear features in a theme. The buffer distance in their method was varied in an iterative fashion to converge on the probability threshold of interest (e.g., 68 percent, 90 percent) for the agreement in line length. Our use of the buffered theme is an inversion of the assumptions and known variables of the features in the Goodchild and Hunter study. In Goodchild and Hunter’s work, the pairs of features (one in the extracted and one in the reference set) are known. The spatial accuracy is unknown. Our use of the buffered concept assumes the spatial accuracy is known while the pairs of features is unknown, i.e., an inversion of their assumptions. In the proposed accuracy assessment solution, the buffered approach is applied to the entire set of reference features simultaneously rather than individual features.

As noted earlier, determining the specific paired FEM road segment and reference road segment is problematic. Also, the application of the buffered approach to individual roads would result in double counting of extracted road lengths in the overlap buffer around road segment endpoints.

There are potential problems with using this buffered reference theme approach for an accuracy assessment. As the buffer distance becomes large, the omission errors may be biased by commission errors. Incorrectly identified transportation features will occur inside large buffers, and their length will counteract the omitted transportation features; possibly resulting in an inflated accuracy. By using small buffer distances, this problem is minimized, although not eliminated.

**ORNL Feature Extraction Model**

As a demonstration, the proposed parameterization model using the line-length buffered theme approach was coupled with the ORNL road feature extraction model (FEM). The ORNL FEM is not the focus of this work as any road feature extraction model could be used. A brief presentation of the logic and parameters of the ORNL FEM follow.

The ORNL FEM for extracting transportation features from panchromatic imagery is not a supervised approach; it is, in fact, an autonomous approach in the purest sense (Cheng, et al., 1998). It was developed to support work by the U.S. Air Force in a collaborative effort with ORNL and Northrop Grumman Corporation. Imagery may be input to the system without any collateral information, such as spatial resolution, and without training sites. The model is designed to receive image input and extract transportation features.

The ORNL FEM identifies road segments in several stages (Figure 2). It contains several novel algorithms: ridge detection, ridge thinning, and road linking and cleaning. Instead of following the conventional road detection approach, which uses edge detection and parallel edge finding (edges of road), a ridge (or valley) detector is used to extract roads. Roads typically have a distinct appearance: a brighter (ridge) or darker (valley) strip with relative smooth and parallel borders. There are several advantages of detecting ridges over edges. First, a long, smooth strip in an image is very likely to be a manmade structure: a piece of road segment, canal, or runway. Therefore, the road extraction logic used after the ridge detection

**Figure 2.** Conceptual organization of processing steps in the transportation feature extraction model developed by the Oak Ridge National Laboratory.
will be simpler. Second, roads generally have a relatively consistent appearance regardless of the lighting and viewing directions. Therefore, the algorithm will be more stable. Twelve 5 × 5 templates representing 12 road directions were used to detect ridges and their associated orientations. The ridge likelihood index at a pixel is determined when the convolution of the twelve templates reaches a maximum. Two matrices, ridge likelihood and ridge direction, are obtained in this process. In order to detect road with different widths, a pyramid scheme is used (Couloigner and Ranchin, 2000; Baumgartner, et. al., 1999). A low pass 3 × 3 mean filter is used to create coarser level images. At each level, all single width, linear features are detected and then transferred to the finer level.

The ridge segments obtained after applying ridge detection are often represented by more than one pixel in width. The thinning (skeletonization) process obtains a simplified representation of the segments, probably including the positions of the features they represent in the original image. Unlike the conventional image thinning operation performed on ridge likelihood images, a new iterative thinning algorithm works on the ridge direction image. In order to do so, a ridge connectivity index, which represents how well a pixel is connected with its neighboring pixel along the ridge direction, is determined at each pixel. The connectivity index at a pixel is a function of its orientation and its neighbors’ orientations. If a pixel’s orientation agrees with its neighboring pixels, it will have a higher connectivity value. On a ridge cross profile, the pixel with the highest connectivity value will be selected as a skeleton point. The advantages of this image thinning algorithm are:

- It can better preserve a smooth image skeleton with fewer gaps.
- It does not require a threshold to cut off the weak and false pixels, and
- It is faster than most image thinning algorithms, which are iterative algorithms.

The ridge image is traced and linked by chain codes. The extracted ridges are tested and cleaned by three criteria: curvature, isolation, and length. Chains in a raster data model are then converted to road segments in a vector data model. The final road features are created by applying another tracing and linking algorithm to the vector segments. The tracing and linking process is applied in a conservative way based on spatial proximity, orientation, and local geometry of the segments. For a given segment, all the segments encountered in the scanning area, which is a buffer along the segment, are collected as potential candidates. Similarity in image intensity value, similarity in ridge index value, similarity in orientations at the extremities of the gap, and the length of the gap are used to evaluate the candidacy of the segments. Short and isolated segments are removed because they are very unlikely a part of a road network.

The ORNL FEM also included an autonomous image rectification step. The extracted road features are used to automatically rectify the image to reference roads. The only assumption is the original image is “in the general study area,” and some common road reference features exist. After the rectification process is performed, the rectified road features are then used to populate a transportation feature database.

Since an assumption in the parameterization model is that the spatial accuracy of the image is known a priori, the model can only be used if the imagery is rectified. In the case of the ORNL FEM, the model would first be parameterized, and then the parameterized model run on un-rectified imagery. In our use of the ORNL FEM, we omitted the autonomous rectification step; rectified imagery rather than un-rectified imagery was used as input.

**Implementation of Parameterization Model**

The parameterization model presented above was implemented using a combination of UNIX shell scripts, C programs, and Arc Macro Language® (AML) programs. The implementation is a tightly-coupled approach between a GIS and ORNL FEM. The structure of this approach is graphically displayed in Figure 3.

The model was implemented as a client/server system, where the parameterization model is a client and ARC/INFO® is a server. The communication between the model and ARC/INFO® server was based on a client/server C-library provided by ARC/INFO®. The library itself is based on Remote Procedure Call (RPC), which is one of many mechanisms of UNIX process communication. Generating parameter configurations, executing the ridge detection and road segment tracing programs, converting from image coordinates to map coordinates, and calculating accuracy indexes are the main tasks in the client part. The ARC/INFO® server is responsible for generating buffers around the reference roads and, initially, overlaying the extracted roads with the buffered reference roads to calculate the length inside and outside the buffer. All thinning algorithm used the overlay and length calculation were re-implemented as a C-program which is about 30 times faster than the ARC/INFO® server.

The results of the sensitivity model and all parameter values are stored in a separate ASCII file (Table 1). In the current implementation thirteen of the fields are the parameter values used in ridge detection and road segment tracing. The next three fields are total length of the reference road, the road length inside and length outside reference buffers. The last two fields are the overall accuracy and the feature agreement indexes. Each record represents a unique parameter configuration. The output file contains the results of every parameter combination. The “best” parameter value set is easily found through a table sort operation. Because the search for the best set is an exhaustive search, alternative near-best solutions are easily identified.

The parameterization model shown in Figure 3 creates an ASCII file containing the parameter configuration and accuracy results for each unique parameter set. On a full-scale test, this data set may contain over 12,000 records. After examining the model results, it is often desirable to then select a specific model configuration and view the results of one of the model runs (Figure 4). While the sensitivity model analyzes all parameter combinations, it does not save the extracted roads from each model run (the data storage requirements would be exceedingly large). An ArcView® project with accompanying Avenue® and C-code were implemented to allow the user to select a specific parameterization and automatically execute the model to produce the extracted roads from this parameterization.

**Example Calibration**

Image data and reference data of various types were assembled for performance testing of the ORNL FEM. Digital orthophotography (1 m × 1 m) of the Oak Ridge Reservation (ORR) was used as an example. Coarser resolutions (1 m to 10 m by 1 m steps) of the imagery were artificially created through spatial averaging. Class one through five transportation features were defined as in the USGS Digital Line Graph (DLG) classification scheme. Reference data were compiled through visual interpretation of the photography for the ORR. The buffer radius commensurate with the approximate spatial accuracy of the 3 m × 3 m imagery used as an example in this article was 1 m. Based on available information on the expected performance of the FEM, we conducted numerous calibration tests with selected values for the seven key parameters of the FEM. These seven parameters, their ranges, and examined values are

---

**References**


Baumgartner, et. al., 1999. A low pass 3 × 3 mean filter is used to create coarser level images.
shown in Table 1. The number of unique parameterized models for each image was 10,368.

An example of the extraction accuracy by variation in values of these seven parameters for a 3 m image within the ORR is illustrated in Figure 4. Maximum feature agreement for the 3 m imagery was 87 percent. When the goal is maximizing feature agreement (high accuracy with low commission errors), the ideal histogram power ranged from 0.5 to 1.0. The ideal sensitivity index ranged from 14 to 16 for this same image resolution. For this 3 m image, slight changes in the

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Range</th>
<th>Initial Default</th>
<th>Values Examined in Sensitivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image enhancement</td>
<td>hist_power</td>
<td>0.8–2.5</td>
<td>1.0</td>
<td>0.5, 1.0, 1.5</td>
</tr>
<tr>
<td></td>
<td>hist_tilesize</td>
<td>9–2048</td>
<td>2048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>hist_remap_size</td>
<td>1–2048</td>
<td>2048</td>
<td></td>
</tr>
<tr>
<td>Line filter</td>
<td>sensi_index</td>
<td>0–16</td>
<td>8</td>
<td>2, 4, 6, 8, 10, 12, 14, 16</td>
</tr>
<tr>
<td></td>
<td>max_levels</td>
<td>1–3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>min_bv</td>
<td>80–120</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>min_count</td>
<td>0–10</td>
<td>5</td>
<td>1, 3, 5</td>
</tr>
<tr>
<td></td>
<td>std_ratio</td>
<td>1–25</td>
<td>10</td>
<td>1, 5, 10, 15, 20, 25</td>
</tr>
<tr>
<td></td>
<td>max_link_index</td>
<td>0–12</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Chain code tracing</td>
<td>min_ridge_index</td>
<td>2–30</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>max_curv</td>
<td>2–3.0</td>
<td>2.5</td>
<td>2.5, 3.0, 3.5</td>
</tr>
<tr>
<td></td>
<td>min_curv</td>
<td>0–0.5</td>
<td>0.1</td>
<td>0, 0.02</td>
</tr>
<tr>
<td></td>
<td>min_length</td>
<td>20–40 (pixel)</td>
<td>40</td>
<td>5, 10, 20, 30</td>
</tr>
<tr>
<td>Vector gap filling</td>
<td>split_dist</td>
<td>2–10 (pixel)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>split_angle</td>
<td>150–180</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td></td>
<td>fan_angle</td>
<td>0–90</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>fan_radius</td>
<td>0–50 (pixel)</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>seg_angle</td>
<td>120–180</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean_diff</td>
<td>0–50</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
other parameter values have little effect on extraction agreement.

Artificial degradation of the ORR 3 m × 3 m image to coarser resolutions (5 m to 10 m) produced new findings. Small changes in the optimized parameter values for the 5 m imagery were found to result in large changes in extraction agreement with 10 m imagery. The maximum extraction agreement for the 5 m imagery would use a sensitivity index between 10 and 12. The best values for the histogram power vary dramatically with image resolution, from 0.8 to 2.3. For the 10 m imagery, the range in histogram power and sensitivity index values is much greater although the incremental values within this range result in dramatic differences in extraction agreement.

Summary
The proposed buffered reference theme approach solves several problems in model calibration, sensitivity analysis, and accuracy assessment phases. First, it allows for a digital extraction accuracy assessment without visual analysis. This digital accuracy assessment then provides for an autonomous accuracy assessment. Third, because the assessment is completely autonomous, this approach addresses the basic problem of a large number of parameter combinations. The approach uses a vector data model and incorporates the known spatial error in the reference data. Finally, once a collection of images and reference themes has been established, evolutionary changes in the feature extraction model may be easily evaluated.

The parameterization model does have limitations, however. Commission errors can occur within the buffered area. This may be observed when nearby non-transportation features, such as powerlines, are inadvertently identified as dirt roads. The model does not include other characteristics of agreement, such as segments or gaps in road segments. Future studies using imagery of a variety of geographic environments and imagery types could be performed to assess the stability of the buffered reference theme approach for model calibration.

Acknowledgments
We would like to express our appreciation for the assistance in this work from John Smyre and Richard Durfee of the Oak Ridge National Laboratory.

References


