

Roads Data Development in East Africa: Final Report

Report prepared by Alex de Sherbinin, Principal Investigator, Center for International Earth Science Information Network (CIESIN), the Earth Institute, Columbia University

Note: This work was sponsored under the NASA-funded Expansion of Regional Visualization & monitoring System (SERVIR) to other Geographic Areas including the Dominican Republic, Gulf of Mexico, Africa, Asia, and South America project. CIESIN gratefully acknowledges NASA's support of this project as well as the pro bono software development of the Center for Spatial Information Science of the University of Tokyo and the provision of ASTER imagery by the Japanese National Institute of Advanced Industrial Science and Technology (AIST). Without the vital contributions of our Japanese colleagues this project would not have been possible.

Introduction

This report documents the methods, results, and lessons learned from this pilot project to develop roads data from satellite imagery using software developed by the Center for Spatial Information Science (CSIS) of the University of Tokyo: the Global Road Mapping Tool (GRMT). The GRMT is a stand-alone software package that permits semi-automated digitization of ASTER imagery using a "snake" algorithm to connect seeded points along a road.

Project Background and Scope

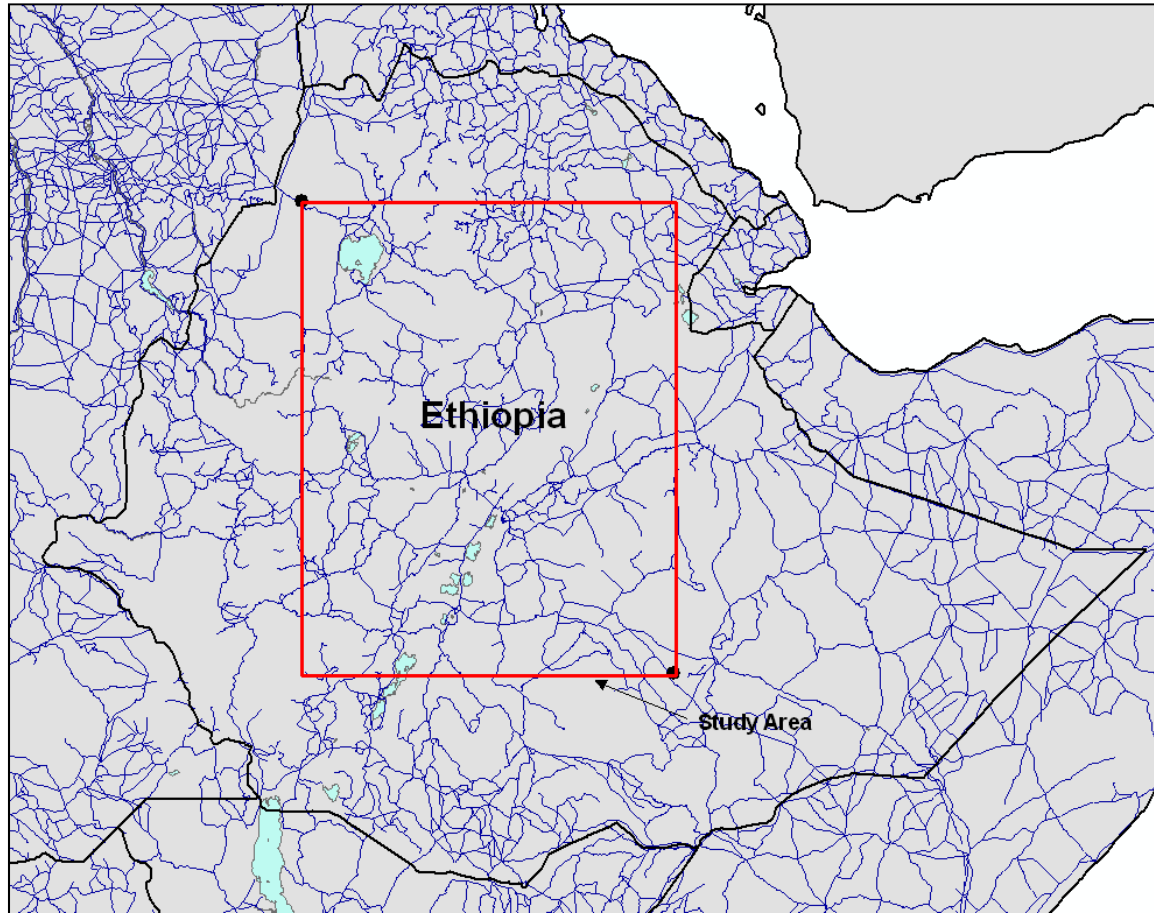
In January 2008, the International Council on Science (ICSU) Committee on Data for Science and Technology (CODATA) approved a new working group that will oversee the development of an improved global roads data set (CODATA 2008). The working group is composed of 11 roads data experts from four continents (two from Africa), and its secretariat is at CIESIN. Before its approval by CODATA, working group representatives and participants at a Global Roads Data Workshop (1-3 October 2007, Lamont Campus, Columbia University, Palisades, NY, USA) developed a strategy for developing an improved global roads data product (CIESIN et al. 2007). This project contributed to the strategy in two ways: First by testing and refining methods for roads data development, and secondly by beginning the process of compiling improved global roads data according to the specifications laid out in the strategy document.

In consultation with members of the CODATA working group and partners at the East African Regional Center for Mapping of Resources for Development (RCMRD), it was determined that the initial scope of this activity would be limited to development of roads data for the central "plateau" portion of Ethiopia (see Figure 1). Ethiopia was identified as one of the countries in East Africa with the least up-to-date and complete roads data. This rectangular area stretches North-South along the main plateau where road data are crucial for relief and development and available data are of poor quality. The region was also considered to be sufficiently small that the remote sensing processing could be achieved within the time constraints of this project, and it represented a useful test case because the country includes a range of

land cover types and a variegated topography that would test the automated detection algorithms utilized in the GRMT.

Figure 1. Map showing the study area in red

Bounding Box: upper left 12o44'11"N and 36o31'00"E, and lower right 6o19'21"N and 41o34'00"E.



Methodology

Our methodology was as follows. Using the GRMT, a prototype software developed by Dr. Shi Yun under the supervision of Dr. Ryosuke Shibasaki at the University of Tokyo, CIESIN staff extracted road vectors from orthorectified ASTER imagery contributed by AIST. The GRMT uses a “snake” algorithm to connect points seeded along a line (see Annex 1: GRMT Manual). In theory, given sufficient spectral contrast between the road and the surrounding land covers the software would be able to connect a sparse array of points by following the road edges. Initial road “seeds” are generated based on shape and structural analysis through segmentation, grouping, and differential geometry, and these are then integrated with the initial seeds using graph theory (Shi and Shibasaki, 2008). This algorithm was successfully tested in Japan, but until this project, it had not been tested in a developing region like Africa.

Over a period of three months (May-July 2009), CIESIN staff digitized roads for 27 ASTER scenes. The resulting output files (.ROD format) were converted to ESRI Shape files using a python script developed by CIESIN. Once the roads were digitized, CIESIN staff exported the Shape file to KML in order to visually inspect the roads in Google Earth and to code the road type (or functional class) according to Table 1.

Table 1. Functional Class (FCLASS) used for this project

1=highway	Highway (paved)
2=primary	Major road (in Africa, mostly unpaved) that connect major cities. In many cases, these can take the place of highways in rural Africa where no highways exist.
3=secondary	Regional roads that connect highways or major roads with smaller towns
4=tertiary	Minor roads that connect smaller towns and farms (away from highways and primary roads) with secondary roads
5=local/urban	Roads within cities or towns
6=trail	Trails
7=private	[not used]
0=unspecified	

Once the roads were assigned a functional class and a substantial number of ASTER scenes were digitized, the shape files for each scene were combined into one large File Geodatabase. These were then merged into a larger data set based on the UNSDI-Transport data model which combined recently collected data from a Gates Foundation-funded AGCommons project (in which field teams collected data using GPS-enabled personal digital assistants (PDAs) loaded with a custom software) and data from a World Food Programme base map for Ethiopia. Only those road segments from the ASTER imagery that represented new contributions to the overall Ethiopia road map were retained; this amounted to 111 out of a total of 5,148 unique segments (Figure 2).

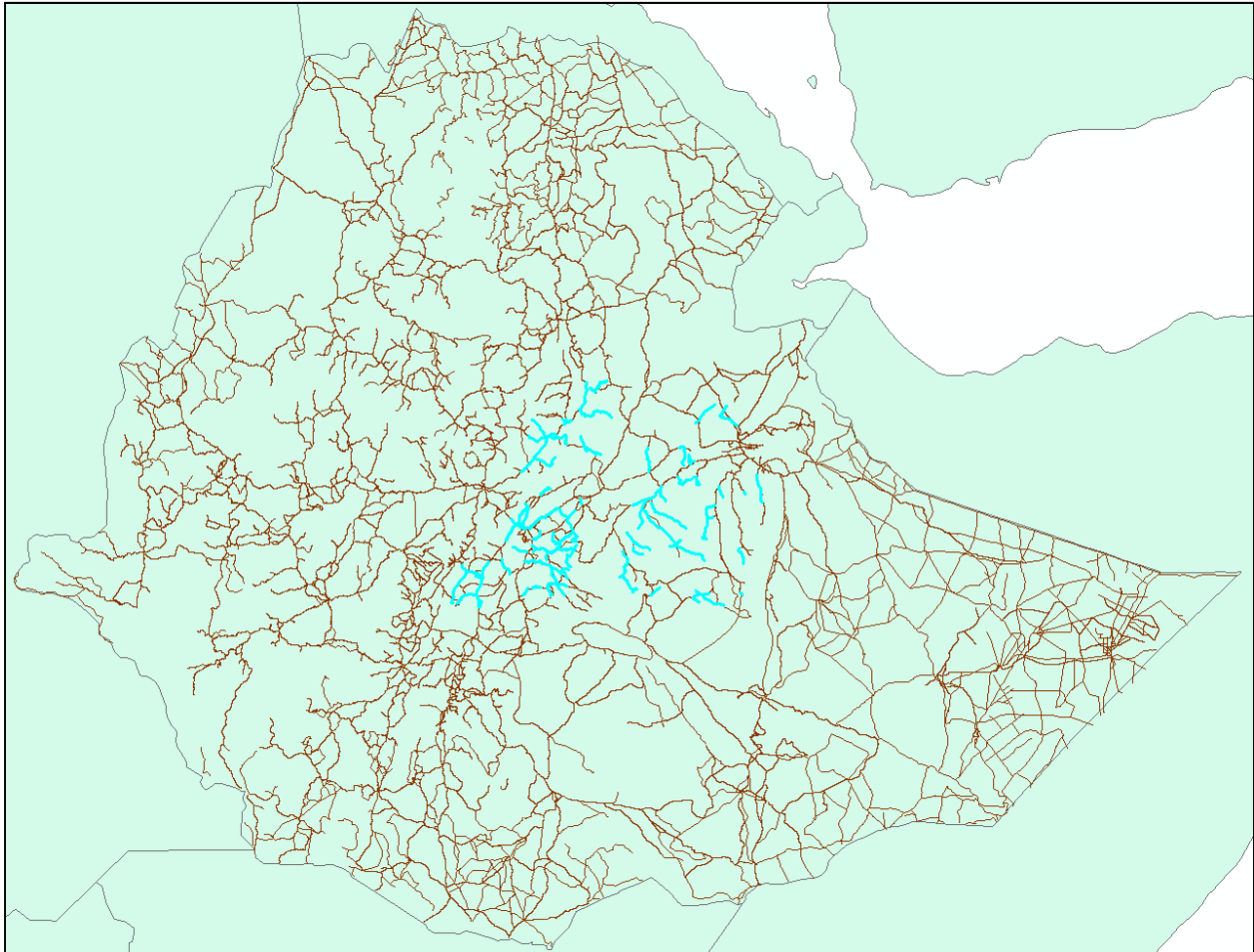
The merged data set along with ISO 19115 metadata are available for download from the Global Roads Open Access Data Set (gROADS) web site at <http://www.groads.org>. The data are being distributed free of copyright restrictions under Creative Commons 0 (CC0), v.1. The provisions of CC0 are available on the Web at <http://creativecommons.org/publicdomain/zero/1.0/>.

In parallel to the Ethiopia roads data development, with supplementary funding from the NASA-funded Socioeconomic Data and Applications Center (SEDAC), CIESIN evaluated and integrated national-level roads data sets for inclusion in an East Africa regional integrated roads data set. This data set is built upon the best public-domain data and features an integrated attribute database (using the UNSDI-

12 March 2010

Transport data model) and vector matching at border crossings. We were able to combine data for Tanzania, Kenya, Uganda, and Somalia.

Figure 2. Combined Ethiopia roads data set with ASTER-derived data in turquoise



Lessons Learned

Several issues arose during the process of digitizing and were discussed at length during a project debriefing workshop from 22-23 June 2009 at CIESIN offices in Palisades, NY (see Annex 2 for agenda). The workshop included representation from CIESIN (Alex de Sherbinin, Steffen Foerster, Valeryia Fiodarava, Paola Kim Blanco, and Greg Yetman), CSIS at the University of Tokyo (Shi Yun and Hiroyuki Miyazaki), RCMRD (Erick Khamala), and the University of Georgia's Information Technology Outreach Services (Karen Payne). The issues that arose in digitization and suggested improvements to the GRMT software tool are described in greater detail in Annex 3. Although in theory digitization was supposed to be accomplished in the close-up window using the ASTER imagery, in practice much of the actual digitization occurred in the Google Map reference window because the road features were easier to identify in that window. Furthermore, it was found that the seed points needed to be very closely

12 March 2010

spaced in order for the algorithm to track the road – close enough, in fact, that the process more resembled “heads up” digitizing than it did semi-automated extraction. This may have been due to the lack of spectral contrast between the roads and the surrounding territory. Whatever the reason, the snake algorithm was rarely able to follow the road, and after some trial and error, parameters were set such that the lines generally represented straight segments between points (to avoid spikes and other anomalies).

In addition, for each ASTER scene representing 60 x 60 km it took approximately 8 hours to complete the digitization, even in relatively rural areas with sparse road networks. For a country the size of Ethiopia, it would probably be cost prohibitive to digitize the whole country using this method. Finally, the actual determination of functional class was often based on personal judgment, given that some road classes are not easy to distinguish from satellite imagery. The fact that further detail on road surface type, surface condition, etc., as specified under the UNSDI-T model, could not generally be obtained (these were generally left “unspecified”), means that the data may be of limited operational use for some purposes. On the other hand, surface type can often be surmised from functional class.

Additional suggestions were provided to CSIS for improvement of the software, including the potential to use other remote sensing imagery. These recommendations are currently being addressed by CSIS, and they are developing a web-enabled version.

Conclusions

The project was designed to be a pilot effort to test a novel road extraction algorithm. The literature in the area of road extraction from remote sensing imagery suggests that this is not an easy task, especially in developing countries where road surfaces often have similar spectral signatures to surrounding areas owing to wind-blown dust or the exposure of substrate (for dirt roads). CSIS conducted tests of a commercial software – Tracking Analyst by Overwatch Geospatial – and it did not perform any better in the Ethiopian context. Most such software packages are designed for use in developed country suburban contexts, where road surfaces, curbs, sidewalks, and lawns are all spectrally distinct. So, the fact that some difficulties were encountered is to be expected. Overall, the team worked successfully together, a significant number of roads were digitized, and important feedback was provided to CSIS for the refinement of the GRMT. This NASA-SERVIR project also complemented work being carried out under the AGCommons project and enabled us to fill important gaps in the overall Ethiopia road map, as well as to clean and compile some additional road maps for the region of interest to RCMRD.

References

CIESIN, CGIAR-CSI, ICSU-CODATA, and WRI (Center for International Earth Science Information Network, Columbia University; Consultative Group for International Agricultural Research-Consortium for Spatial Information; International Council for Science-CODATA; and World Resources Institute). (2007). A

12 March 2010

Strategy for Developing an Improved Global Roads Data Set. Palisades, NY: CIESIN. Available from <http://www.ciesin.columbia.edu/confluence/display/roads/>

CODATA. 2008. Working Group: Global Roads Data Development. Available from <http://www.codata.org/taskgroups/WGglobalroads/>

Nelson, A., A. de Sherbinin, and F. Pozzi. (2006). Towards Development of a High Quality Public commons Global Roads Database. *Data Science Journal*, Vol. 5, pp. 223-265. Available from http://www.jstage.jst.go.jp/article/dsj/5/0/223/_pdf.

Shi, Y., and R. Shibasaki. (2008). "Extraction of Roads Data Using Remotely Sensed Imagery." Presentation made at the 25th Asia-Pacific Advanced Network (APAN) Forum, 24 January 2008, Honolulu, Hawaii. Available from <http://www.ciesin.columbia.edu/confluence/pages/viewpageattachments.action?pageId=19726351>

12 March 2010

Annex 1. Global Road Mapping Tool (GRMT) Manual

Produced by Dr. Shi Yun, CSIS, University of Tokyo

Edited by Steffen Foerster, CIESIN, Columbia University

System requirements

OS: Windows XP, Vista

Internet Explorer 6, 7 installed








Installation and startup

Extract GRM.zip to a folder of your choice. Execute the RoadExtractor.exe to start the program.

Interface



Toolbar

-  Open ASTER/VNIR image file.
-  Zoom in (close-up window)
-  Zoom out (close-up window)
-  INPUT mode: Input initial points for road extraction algorithm
-  EXTRACT mode: Apply road extraction algorithm
-  SELECT: Extracted roads can be selected by left-click and deleted with [Delete] key
-  PARAMETERS: Configure parameters of the road extraction algorithm that affect the form and location of the line created by the SNAKE algorithm. Alpha and gamma are calculated based on the digital numbers of the reflectance data and beta is based on the geometry of the line.

Description of parameters

- Alpha: greater values push the line towards the “lowest point” or the area with the lowest digital number, which is seen as the darkest spot on the image.
- Beta: greater values will create a smoother line, while lower values will keep the line closer to the vertices determined based on the alpha and gamma settings.
- Gamma: greater values push the line towards the area of the image with the greatest contrast (edge detection). This is found by calculating the difference between the digital numbers of two pixels: the area with the largest difference is the area of the greatest contrast.

Overview window

The window displays an overview of the opened ASTER/VNIR image. A red rectangle represents the extent currently visible in the CLOSEUP window. You can move the region shown in the CLOSEUP window by left-clicking anywhere in the OVERVIEW window.

Close-up window

This window displays a close-up of the region represented by the red rectangle in the OVERVIEW window and serves as the main input window. You can move the view by left-click-and-dragging, and synchronize it with the Google Map window by right-clicking.

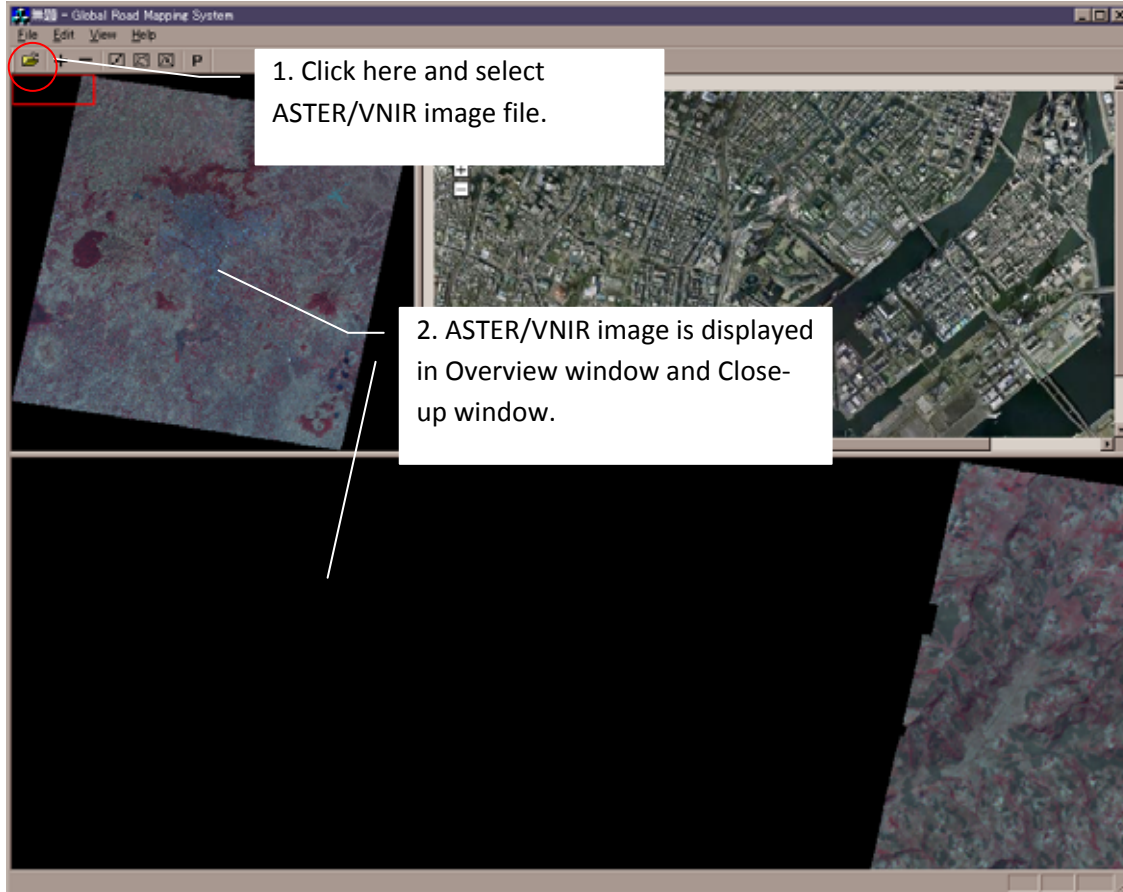
Google Map window

This window displays a portion of the satellite image in Google Map, centered at the location right-clicked in the CLOSEUP window.

Steps for extracting road features

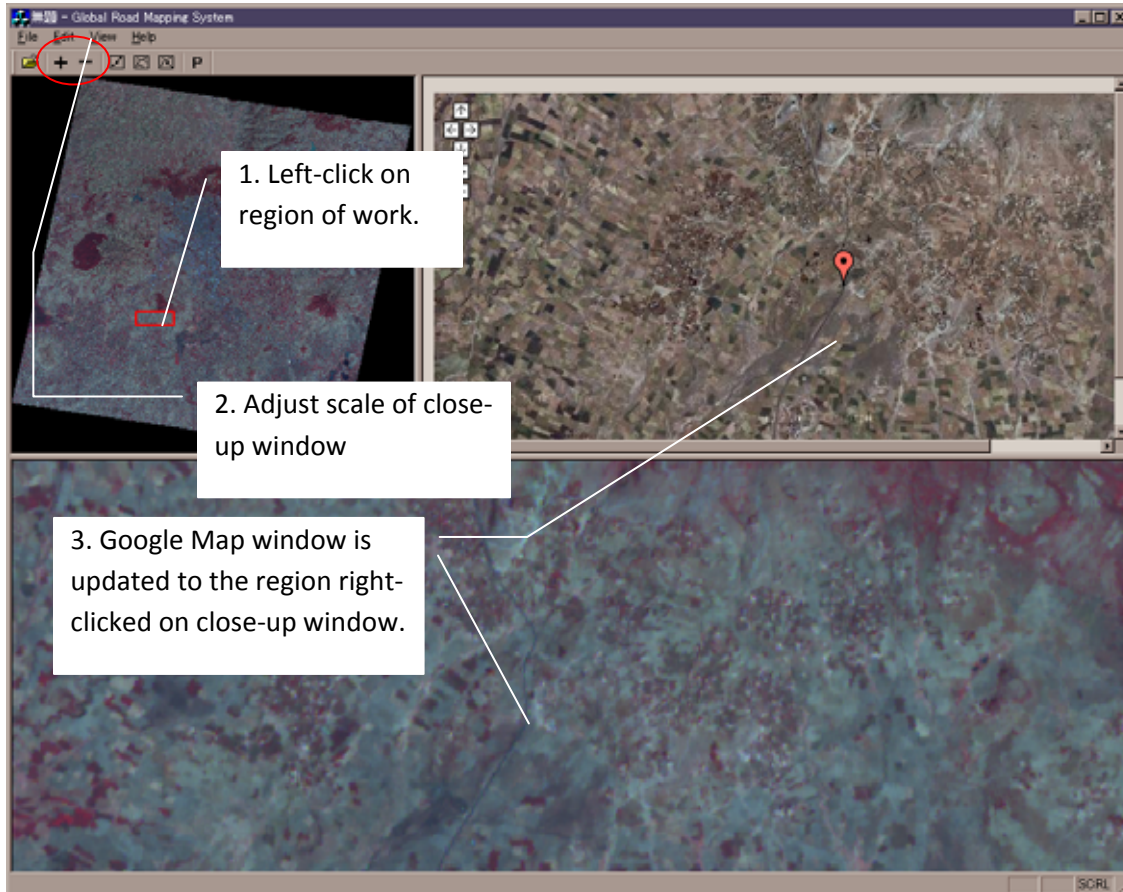
Open ASTER/VNIR image (TIFF format)

Note: Be sure that a GH file of the same name as the image file is located in the folder containing the ASTER/VNIR image, because it contains the geographic coordinates needed to synchronize the image with Google Map.




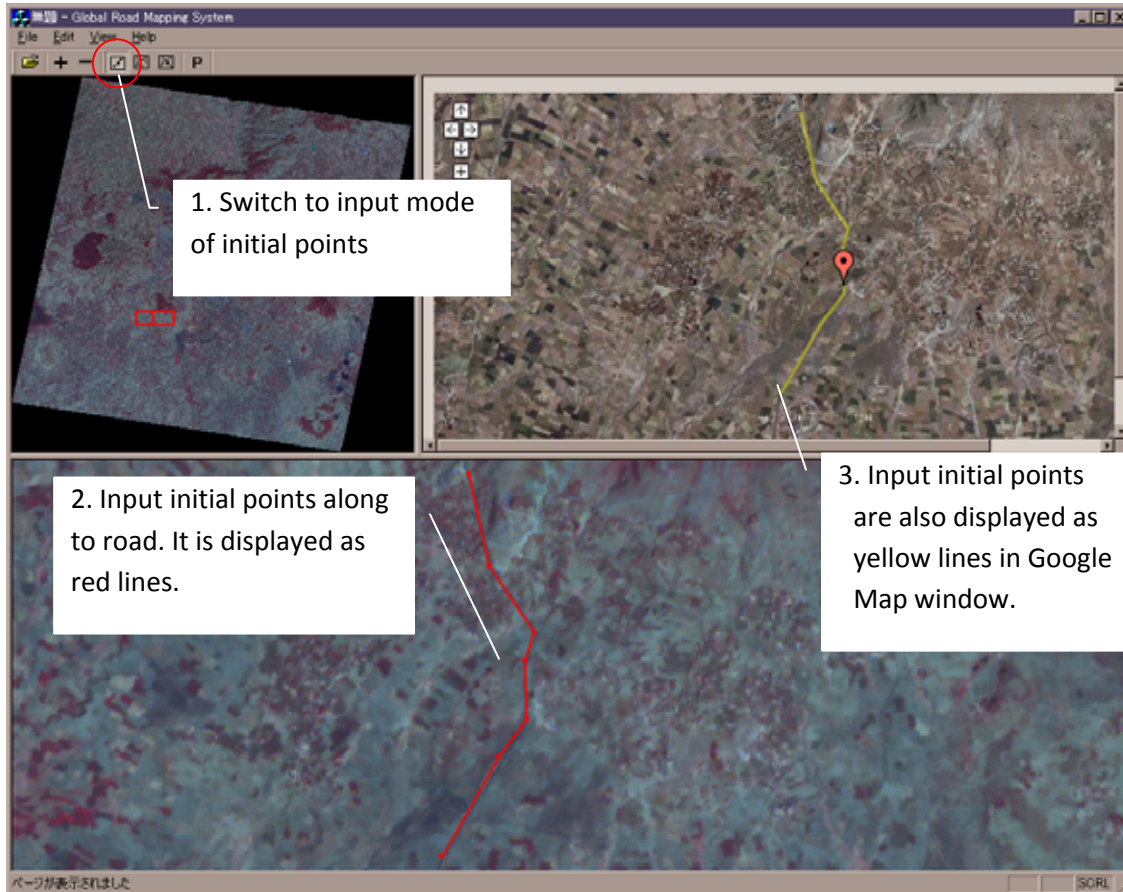
Display region of work

By right-clicking on the region of work in the overview window, the corresponding region is displayed at its optimal resolution in the close-up window. Further adjustments can be made with the ZOOM buttons. The Google Map window is updated by right-clicking in the close-up window at the desired location.



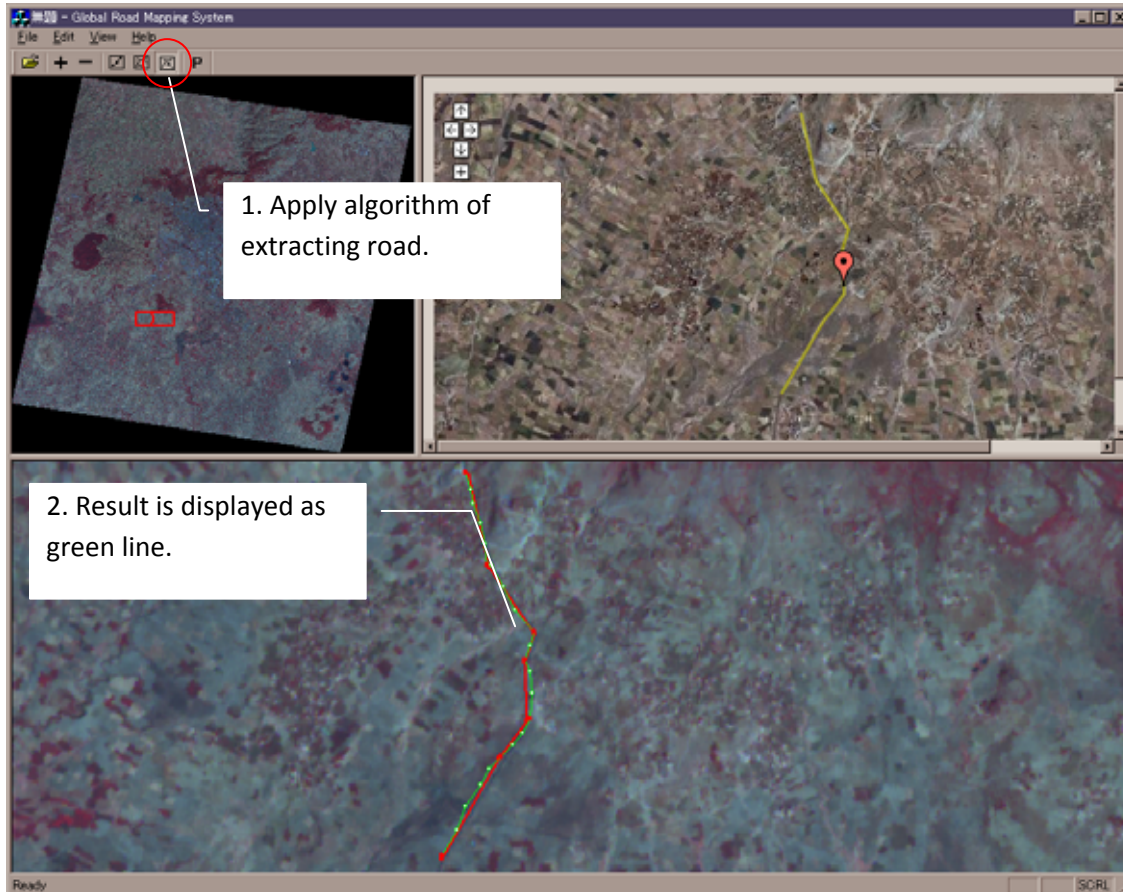
Input initial points

Switch to INPUT mode by clicking  in the toolbar. During this mode, initial point can be inputted by left-clicking in the close-up window. The traced segments are displayed as a red line in close-up window and appear in real time as a yellow line in the Google Map window. Wrong input points can be deleted with the [Backspace] key.



Apply algorithm of extracting road

After inputting the initial points for a road, apply the extraction algorithm by clicking the EXTRACT button. The result is displayed as a green line.



Modify extraction result and adjusting parameters

If the extraction result is not appropriate, you may switch to SELECT mode. Green lines can be selected by left-click and will turn blue. Selected lines can be deleted by hitting the [Delete] key. You can also apply the algorithm multiple times with modified extraction parameters, and compare the resulting extracts (multiple green lines) side by side before deleting all but the best extract.



12 March 2010

Save and restore as working file

You can save the results by clicking [File] -> [Save Result]. The working file is restored by clicking [File] -> [Open Result].



Format of output file

Total_Object 5	:	Number of line segments recorded in the file
000000 89	:	Number of vertices of the first ('000000' is dummy)
38.544970 8.875044 0.000000	:	Longitude & latitude of vertex ('000000' is dummy)
38.546472 8.875994 0.000000	:	Same as above
38.547974 8.876809 0.000000		
(Continued)		
38.644526 8.918481 0.000000		
38.645209 8.919159 0.000000		
38.646028 8.919838 0.000000	:	End of vertices of the first
000000 86	:	Number of vertices of the second ('000000' is dummy)
38.597821 8.904771 0.000000	:	Longitude & latitude of vertex ('000000' is dummy)
38.596455 8.905993 0.000000	:	Same as above
38.596182 8.906671 0.000000		
(Continued)		

Annex 2. Agenda of the CODATA Global Roads Data Development Working Group Workshop

22-23 June 2009, Lamont Campus, Columbia University, Palisades, NY

The goal of this workshop is multiple:

- To convene the CODATA-Roads Working Group (WG) in person and via teleconference on Tuesday, June 23, in order to:
 - Report on progress regarding data cleaning, data cataloging, and collaboration with partners such as GIST-ITOS, OSM, and Google Mapmaker
 - Refine the CODATA-Roads strategy and implementation plan in light of recent developments, assess priorities, and identify next steps
- To convene the partners in the NASA-SERVIR Roads Data Development in East Africa Project (CIESIN, University of Tokyo, and RCMRD) in order to:
 - Assess progress in compiling improved roads data for East Africa
 - Assess the Global Road Mapping Tool (GRMT) using ASTER imagery, including suggestions for improvements/refinements
 - Identify a strategy for promoting the use of the GRMT in the wider community
- To convene the partners in the AGCommons project (CIESIN, IMMAP, and RCMRD) in order to:
 - Report on the testing of the PDA tool for road mapping
 - Assess project implementation to date
 - Assess next steps and the way forward

Monday, 22 June

9:00-10:15	Welcome Introduction to CIESIN activities <ol style="list-style-type: none">1. Road data catalog2. Data evaluation (multiple collections)3. Data set cleaning and merging
10:15-10:30	Coffee Break
10:30-11:30	AGCommons Project – presentation of the PDA tool and discussion among partners
11:30-12:30	NASA-SERVIR project: <ol style="list-style-type: none">1. alpha testing the Global Road Mapping Tool (GRMT)<ol style="list-style-type: none">a. suggestions for improvements/enhancementsb. edits to documentation2. potential for using the tool for non-ASTER imagery
12:30-14:00	Lunch

12 March 2010

- 14:00-17:00 NASA-SERVIR project:
3. the script for converting to shape file (bring in Greg Yetman)
 4. work flow process for digitizing, adding attribute information, and editing topology
 5. steps to wider dissemination of the GRMT
 6. next steps in project implementation

Tuesday, 23 June

- 9:00-10:15 Meeting between Karen Payne and CIESIN staff on data cleaning and compilation (bring in Greg Yetman by phone/Skype)
- 10:15-10:30 Coffee break
- 10:30 -12:00 CODATA Roads WG telecon
- 12:00-14:00 Lunch
- 14:00-15:00 Bilateral meetings (Karen Payne and Alex de Sherbinin, Erick Khamala and IRI)

12 March 2010

Annex 3. Comments on the Global Road Mapping Tool

By Steffen Foerster, Valeryia Fiodarava, and Alex de Sherbinin (CIESIN, Columbia University)

September 2009

These comments were mostly compiled and circulated prior to the June 2009 workshop in Palisades NY, and **red type** indicates that a topic was brought up during the workshop and how it was resolved, if at all. Additional edits and amendments were made by Alex de Sherbinin in September 2009.

Overall the tool is useful, and even though the road extraction algorithm does not work 100% as expected, perhaps owing to the resolution of the ASTER imagery, or perhaps due to the lack of spectral contrast between roads and surrounding land cover in Ethiopia, it could still be a very useful tool for “heads up” digitizing of areas for which there are no data. The combination of the Google Map pane makes the tool even more powerful. The fact that there are no easy ways to extract data from Google Earth or Google Mapmaker means that this tool still could fill a significant niche.

Our understanding is that the tool could be used with other imagery provided the proper header file is created. Thus, one recommendation would be to enable this tool to easily function with other imagery such as Landsat, CBERS (which is free to anyone in Africa), or perhaps SPOT and IRS. This might mean creating an “import imagery” file menu, with multiple choices that would accommodate multiple satellite image types.

Below please find some more specific suggestions for improvements.

1. Main suggestions for Usability Improvements

Tracing

Once a road had been extracted, single input points cannot be deleted or modified. Being able to modify single input would be helpful to improve the extraction. Currently, the only alternative is to retrace the entire road.

Parameters

The description and function of the alpha and gamma parameters are not very clear. It is hard to figure out what these do differently. After playing with the settings for a while, we came to the conclusion that in order to avoid random spikes in the road (see Figure 1) and not have overly undulating extracts (zigzagging around the real road line) one has to set beta to high values for smoothness, and both alpha and gamma to low values. As a result, the algorithm only works well if roads are traced accurately by frequent input points placed relatively close together. This means that the tool is largely used for “heads up” digitizing rather than semi-automated road feature extraction.

Note: This was discussed during the June workshop and Dr. Shi presented some documentation on the alpha, beta, and gamma parameters (a powerpoint presentation “Snakes based road mapper” and a paper by Kass et al., 1987, “Snakes: Active Contour Models”, in *International Journal of Computer Vision*), but it was still difficult to understand and document how these work. Note that the edited version 2 manual includes an effort to describe the three parameters based on our discussions, but CIESIN is not sure that these are correct interpretations.



Figure 1: Random spikes produced in areas where roads are not easily detectable (much of East Africa) and when using default parameter settings

Google Map window

It would be nice if right-clicking in the Google Map window would take one to the appropriate location in the close-up window of the ASTER image. Sometimes, one ends up searching around in Google Map for roads, and then easily loses track of the location in the CLOSEUP window. It is very hard to then find the location on the ASTER image.

2. Minor issues

Toolbar: Instead of graphical icons, could these be changed to INPUT, EXTRACT, and SELECT buttons for a more intuitive use? It is sometimes possible to hit the wrong button and undo one’s work. Also, it would be good if the name of the ASTER image currently loaded could be displayed somewhere, for reference.

Google Map window: Since in practice, a lot of the digitizing occurs in the Google Map pane, instead of the hand, can there be a cross-hair or something that lends itself more easily to pinpointing an exact

location? Is it possible to maximize/change the size of the Google Map window? Much of the digitizing actually takes place in this window when using high resolution imagery, and working in a larger window, especially when on a big screen, would really be an advantage. Maybe there could even be a way to maximize either of the windows to full-screen view?

3. Bugs

Roads caching problem

All roads extracted in a session are cached, so even after new scenes are opened, extracted roads from previous scenes are loaded and saved with every result file. This leads to overlaps in each consecutive result file that gets created. When converted to shapefile format, the latest file will contain ALL roads from previous scenes edited during the same program session. This is not desirable because it leads to multiple overlaps in the shapefiles that need to be edited out later during topology cleaning.

Workaround: For now, we found a workaround to this problem by following this procedure:

- Open the road extractor tool
- Extract roads in the scene as usual
- Save the result file in a different location, i.e. in a folder reserved for all result files. Name each result file with the last 7 digits of the scene number.
- Close the program
- Reopen program, load a different scene. Repeat procedures. Close program after every scene is completed and saved, and then reopen with an empty cache.
- After each extracts from a scene are converted into shapefile format using the python script, delete or move the result files to a different folder, so that they do not get re-created the next time the script is run.

Unexpected crash

Sometimes program crashes after inputting points and accidentally pressing “road select” instead of “road extract”. Repeatedly pressing IGNORE on the error message that pops up usually prevents the program from shutting down, but it may remain unstable and refuse to extract afterwards. (Note this problem could be reduced if the icons were changed to plain text, per suggestion above.)

Inputting points do not go where intended

In some instances, input points are not placed exactly where one intends to, i.e. the point is shifted away from the position of the mouse pointer, and nothing can be done to put it where it is supposed to go. In most instances, however, the extraction algorithm would trace the road correctly, i.e. recognize the actual location of the road even if input points are slightly offset.

12 March 2010

Note: Dr. Shi indicated that if you zoom in beyond a certain level in Google maps that you cannot place a point just anywhere – Google maps forces the point just to the side.



Figure 2: Screenshot showing Google map window with pointer location where input point is supposed to be located (green arrow), and the location of the input placed by the extractor (red arrow).