Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Extracting Mobile Machine Routes from GPS Traces

Mikko Laurikkala * **Matti Vilkko** **

* *Department of Automation Science and Engineering, Tampere University of Technology, Finland. (email: mikko.laurikkala@tut.fi)*
** *Department of Automation Science and Engineering, Tampere University of Technology, Finland. (email: matti.vilkko@tut.fi)*

Abstract:
GPS data are available from a large amount of sources. Individuals and vehicles carry their receivers and are often willing to share their locations. These GPS traces are inexpensive compared to dedicated collection techniques. Therefore, using GPS traces in map extraction has attracted a number of researchers in the past decade. In the case of mobile machines, dedicated collection of GPS data is usually even impossible, so one has to confine to the data collected by the machine itself. This paper introduces an algorithm for extracting a map-like graph from mobile machine GPS traces with large uncertainties. The result is a topological map with intersections and route segments. It can be used for operator support instead of displaying the raw GPS traces, or information interchange between operators. In the future, even automatic route optimization is possible with the help of a graph like this.

Keywords: Mobile machines, Human-machine interface, Clustering, Data reduction, Forest harvesting, Topological map

1. INTRODUCTION

Mobile machines become more and more common in many applications throughout the Earth - and even in space. Examples of widely used mobile machines include harbour cranes, submarines, mine vehicles, and also ordinary cars. Somewhat more exotic applications involve autonomous aerial vehicles and Mars rovers. Some of these machines are unmanned, but a human operator is still needed in a variety of machines. In some applications, such as remotely controlled robots, the machine handles the low-level movement autonomously while a human operator performs localization and cognition activities (Siegwart et al., 2011).

GPS receivers are inexpensive and are thus used almost everywhere. Individuals carry GPS devices in their mobile phones or as separate devices and are often willing to share their location for free. Few vehicles lack GPS any more. Hence, a flood of geographical information is available from people and machines. This information can be used for creating maps. Collecting mapping data with dedicated personnel and vehicles is expensive – yet accurate – but using freely available GPS traces provides another effective way to refine maps.

In the past decade, a number of studies have been published that aim at automatically generating maps. Mobile robot research has yielded approaches using mainly other than GPS data, see (Bonin-Font et al., 2008) for a survey. GPS data then dominate in road traffic and other environments with good satellite coverage. The problem can be viewed as image processing, where the data are first converted to a bitmap (Shi et al., 2009; Fathi and Krumm, 2010). Another approach is to model the traces geometrically (Castro et al., 2006). Many papers, including this one, use some form of clustering to reduce the amount of data (Kasemsuppakorn and Karimi, 2013; Worrall and Nebot, 2007; Schroedl et al., 2004). Usually the mapping procedure results in a topological map with road crossings and sections between them.

In this paper, the mobile machine is a forest machine. In the cut-to-length method prevailing in European forest harvesting, two types of machines are used. First, a harvester fells trees, cuts them to desired length, and leaves the logs in small piles along its path. Then, a forwarder collects the logs and transfers them to a loading site along a road. The harvester usually creates its own paths, and the forwarder can only use paths created by the harvester. The operation of both machines is an extremely complicated optimization problem where several targets and constraints should be met. Automation helps in some low-level tasks but in general the productivity of harvesting depends highly on the operator skills. Different wood species and timber products, varying and mostly unknown terrain, and mechanical characteristics of the machine are some of the factors. The landowner interest and nature preservation aspects require that no excessive paths be created in the forest, which is also important from the mapping point of view. (Palmroth et al., 2009; Tervo et al., 2010)

Both harvester and forwarder have GPS receivers and store their locations at regular intervals. The accuracy of the data depends on trees, weather, and terrain and is often rather weak. The data collection is somewhat similar to collection from roads and streets by individuals: the data can be considered free, since it comes as a byproduct

Copyright © 2014 IFAC

Fig. 1. A forwarder. (John Deere)

of other activities. Yet in the forest, existing maps are seldom useful because they do not contain the paths created by the harvester. Also, using a dedicated vehicle for mapping a forest is not possible.

This paper presents an algorithm to extract a map from the unclear and uncertain GPS traces of forest machines. The result is a graph-like structure that can be used for operator support and information interchange between operators.

Section 2 of this paper introduces the GPS data, which makes it easier to understand the rest of the study. Section 3 gives a detailed description of the algorithm itself. Section 4 visualizes results and discusses the meaning of the only parameter of the algorithm. Finally, Section 5 concludes the paper and takes a look at the possibilities to use the algorithm.

2. GPS DATA

The GPS traces were obtained from John Deere forwarders (Figure 1) as part of a large amount of other measurement data. Figure 3 shows a typical trace: the forwarder has moved along some roughly parallel paths along which the harvester has left the logs.

A location point with latitude, longitude, and time stamp was stored whenever the forwarder moved 10 metres or stood still for 30 seconds. As the GPS location suffers from rather large uncertainties, particularly the standing or slowly moving machine makes the trace rather messy (see Figure 3a).

All figures of this paper are plotted in metres instead of coordinates. This is done both to enhance readability and to maintain confidentiality. Each figure covers one working day. Splitting the data daily is not a necessary choice; it would also be possible to analyze traces of a shorter or longer period at once.

3. METHODS

The route extraction algorithm consists of clustering and connecting the clusters. In this section, the algorithm is described in detail. Figure 2 illustrates some of the phases.

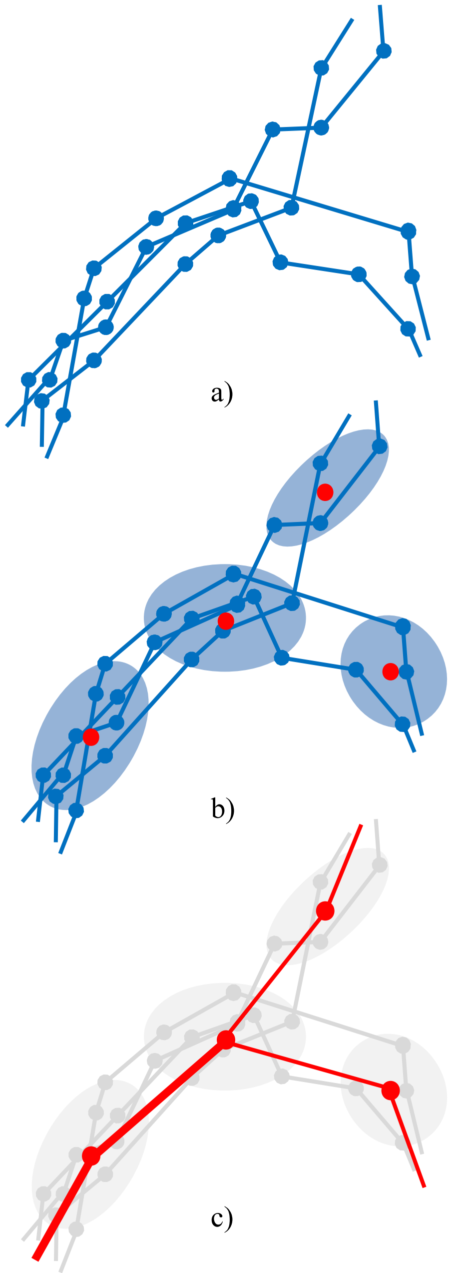

Fig. 2. Illustration of the mobile machine route extraction algorithm. a) Original GPS traces. b) Clustered GPS points and cluster centers. c) The final graph of extracted routes with line widths proportional to path utilization rate.

As a data-driven counterpart, also Figure 3 is referred to in the description.

The data come in as latitudes and longitudes, thus the first step is to convert coordinates to metres. As the felling area is only a few square kilometres at most, the question of using great circles instead of Euclidean distances is mainly of academic interest.

The clustering algorithm used in the next step needs initial clusters. The initialization is done with a simple sequential clustering algorithm (Theodoridis and Koutroumbas, 2008):

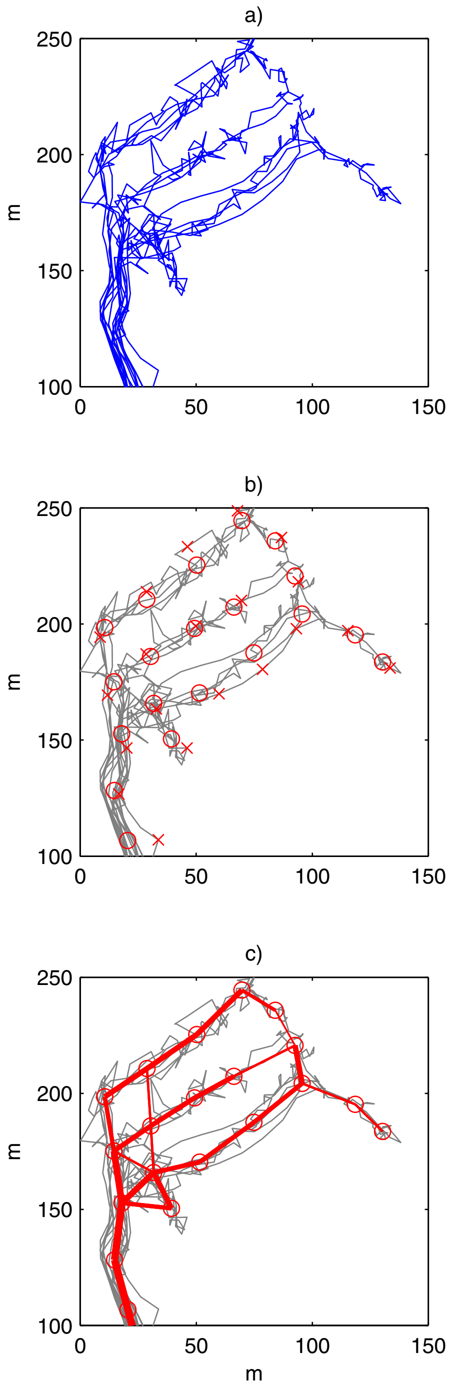

Fig. 3. Data-driven illustration of the mobile machine route extraction algorithm, cf. Figure 2. a) Original GPS traces. b) Original clusters obtained from sequential clustering (red crosses) and K-means tuned clusters (red circles). c) The final graph of extracted routes (red lines) with line widths proportional to path utilization rate.

(1) Follow the trace and add each point to the nearest cluster.
(2) If the distance to the nearest cluster is more than c, designate this point as a new cluster.

The initialization presented above results in a set of cluster centers on the trace. The clusters can and must be tuned with another clustering algorithm. K-means is adopted here mainly because of its simplicity. K-means has also the feature that each data point belongs to exactly one cluster, which is necessary for this application. The interested reader can find details of K-means clustering from e.g. (Xu and Wunsch II, 2009).

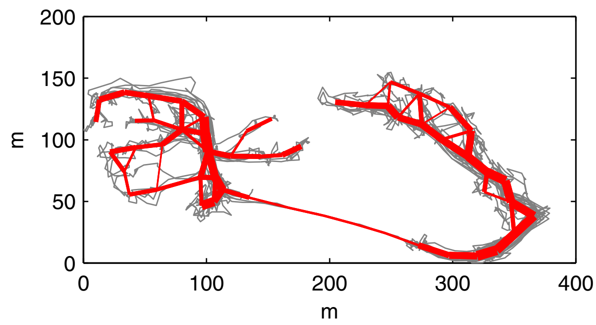

Fig. 4. Route extraction results from a forwarder.

Figure 3b shows an example of some initial clusters obtained from sequential clustering as well as clusters tuned with K-means. Only cluster centers are shown, though each and every data point belongs to the closest cluster and to exactly one cluster. The initial cluster centers lie on the trace that was created first among the nearby traces, while the tuned clusters represent better the average of nearby traces and are not located exactly on any trace.

Now each GPS point belongs to a cluster, and cluster centers represent the clusters. Next, the clusters should be connected by lines so that the result be a graph of cluster centers and lines connecting them. A line is drawn from cluster i to cluster j if there exists a trace from any point of cluster i to any point of cluster j. If more than one connection exist between clusters i and j, the line can be drawn thicker (see Figures 2c and 3c).

There is only one parameter in the algorithm: The distance c decides whether a data point should form a cluster of its own in the initialization phase. The parameter ensures that after the initialization no two clusters exist closer than c units apart. In the tuning phase, the clusters move slightly but still c can be considered a rough estimate of minimum cluster distance.

4. RESULTS

Some successful results of the extraction algorithm were already presented in Figure 3c. The red routes form a network that is much more readable than the original traces. In the following, some more results from other data files are presented.

Some typical results are presented in Figure 4. The machine has moved widely in two fairly separate areas. The algorithm successfully describes the main paths of both areas with thick lines. Between the areas, there is a transfer path that the machine has used only twice, probably once in each direction. The algorithm lumps these two traces together and shows one thin line instead.

Next, let us illustrate how deviations in GPS data may lead to faulty connections between clusters, see Figure 5.

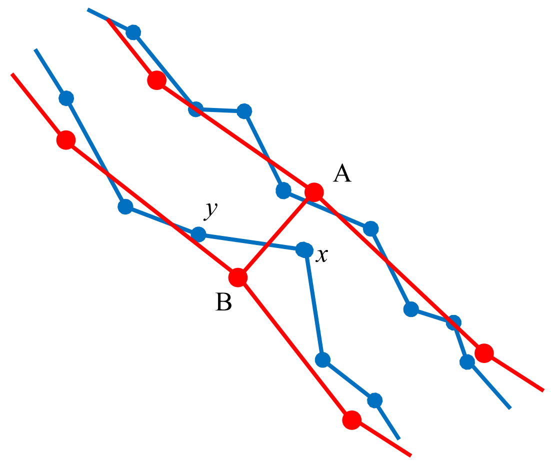

Fig. 5. How deviation in GPS data leads to a faulty connection. Original GPS data in blue, extracted routes in red. Note that the illustration is a simplified version and does not fully correspond to true data and clustering.

Because of GPS uncertainty, data point x is slightly closer to cluster A than cluster B. Therefore, it gets clustered to cluster A although it should belong to the route going through cluster B. Now there is a trace from point x in cluster A to point y in cluster B, and the algorithm draws an erroneous connection between A and B.

The results so far were generated with $c = 20$ (metres). Figure 6 explains the effect of c with three different values. The underlying traces in the figure form a back-and-forth pattern that the forwarder has used when collecting a given area. There are no connections between parallel paths, so the extraction algorithm should not add such connections. Yet, the apparently erroneous deviations in the original GPS traces makes it almost impossible not to make errors in the clustering.

In the top plot of Figure 6, $c = 15$. This means that cluster centers are relatively close to each other, the number of clusters is larger and the size of clusters smaller. Accordingly, the algorithm can follow smaller turns in the original traces. The smaller clusters result in more accurate connections, thus there are only a few faulty connections (one of them is marked in the figure with a black ellipse). On the other hand, attempting to create too fine-grained connections causes too many red lines where only one would suffice. The readability of the figure suffers from the small c value in the densest region of traces.

Choosing $c = 20$ causes more errors in connecting parallel paths but then again makes the graph simpler and more readable. The best choice of the parameter c might be somewhere between 15 and 20. Moreover, a less dense pattern of traces would most obviously produce different results.

The value $c = 30$ in the lowest plot of Figure 6 is more or less a disaster. The clusters are too sparse and connections between them do not follow the original traces. The northernmost trace does not get a red line at all because its points are less than 30 metres from the neighbouring trace.

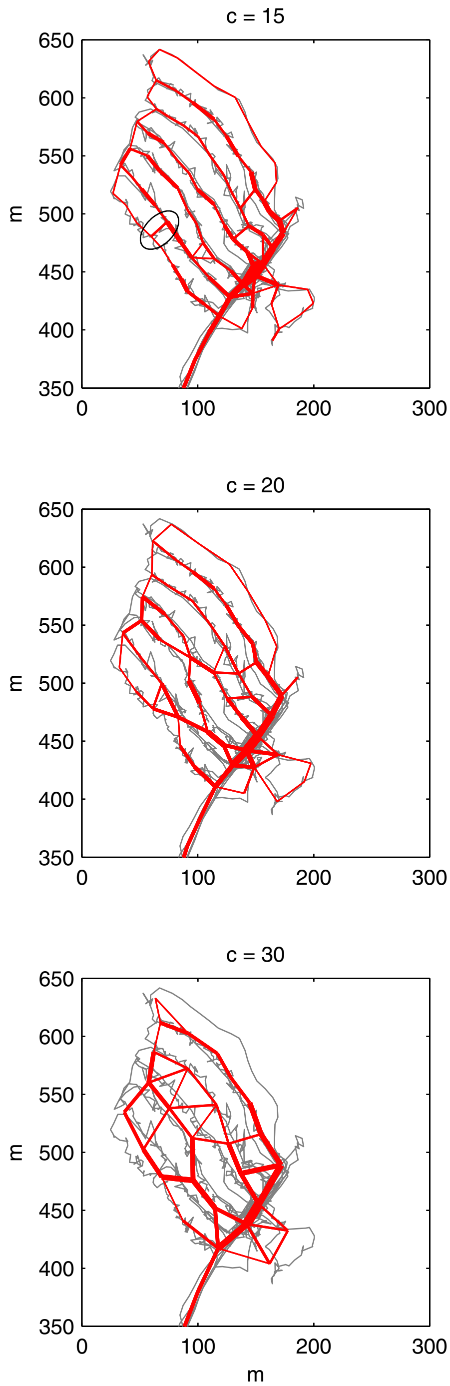

Fig. 6. Comparison of extraction results with different values of minimum cluster distance c.

5. CONCLUSION

This paper presented a method of converting uncertain GPS data of a mobile machine into a map-like graph. The resulting graph is extremely useful for the human operator. When the graph is displayed on the screen of the machine, the operator can easily see the shape of the area and the location of the main paths. Furthermore, the graph can serve as information interchange between operators. Often the same machine has several operators, and with the help of the graph the evening shift operator gets an overview of what has been done in the morning shift. In the case of forest machines, the forwarder operator sees where the harvester has been moving.

One of the premises of the study was that no data other than GPS traces are available. The true route of the machine is not known and cannot thus be compared with the results. Yet using additional information would obviously improve the results and is a potential topic of further research. A useful map of the terrain is typically not available but, for example, sensors could measure slopes that hinder the moving machine. Information on GPS accuracy combined with the movement would incorporate a means of treating uncertainties of the map.

Furthermore, the algorithm could be used as a part of route optimization. Route optimization of a forest machine is an extremely complicated problem. The logs should be cut and delivered out of the forest with least possible costs. Costs are induced by fuel and time consumption. Moreover, logs should be collected in an appropriate order that takes into account the variety of timber products. Movement of the machine is affected by several features of the terrain, either known or unknown. The problem remains largely unsolved, that is why operator skills are a key resource in harvesting. The main function of data analysis is to provide the operator assistance and support in decision making.

The route extraction presented in this paper could be a part of route optimization of a forwarder. As the forwarder can only use tracks created by the harvester, the route map extracted from harvester GPS data would serve as a basis for optimization.

6. ACKNOWLEDGEMENTS

The research was funded by the Data to Intelligence programme of DIGILE (Finnish Strategic Centre for Science, Technology and Innovation in Digital Business), project No. 264428 of the Academy of Finland, and John Deere Forestry.

REFERENCES

Bonin-Font, F., Ortiz, A., and Oliver, G. (2008). Visual navigation for mobile robots: A survey. *Journal of Intelligent and Robotic Systems*, 53(3), 263–296.

Castro, M., Iglesias, L., Rodríguez-Solano, R., and Sánchez, J.A. (2006). Geometric modelling of highways using global positioning system (GPS) data and spline approximation. *Transportation Research Part C: Emerging Technologies*, 14(4), 233–243.

Fathi, A. and Krumm, J. (2010). Detecting road intersections from GPS traces. In S.I. Fabrikant, T. Reichenbacher, M. Kreveld, and C. Schlieder (eds.), *Geographic Information Science. 6th International Conference, GIScience 2010, Zurich, Switzerland. Proceedings*, Lecture Notes in Computer Science, 56–69. Springer-Verlag.

John Deere (2013). John Deere 1010E -kuormatraktori. `http://www.deere.fi/wps/dcom/fi_FI/products/equipment/forwarders/1010e/1010e.page`. Retrieved November 11, 2013.

Kasemsuppakorn, P. and Karimi, H.A. (2013). A pedestrian network construction algorithm based on multiple GPS traces. *Transportation Research Part C: Emerging Technologies*, 26, 285–300.

Palmroth, L., Tervo, K., and Putkonen, A. (2009). Intelligent coaching of mobile working machine operators. In *Intelligent Engineering Systems, 2009. INES 2009. International Conference on*, 149–154.

Schroedl, S., Wagstaff, K., Rogers, S., Langley, P., and Wilson, C. (2004). Mining GPS traces for map refinement. *Data Mining and Knowledge Discovery*, 9(1), 59–87.

Shi, W., Shen, S., and Liu, Y. (2009). Automatic generation of road network map from massive GPS vehicle trajectories. In *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*, 48–53.

Siegwart, R., Nourbakhsh, I.R., and Scaramuzza, D. (2011). *Autonomous Mobile Robots.* The MIT Press, 2nd edition.

Tervo, K., Palmroth, L., and Koivo, H. (2010). Skill evaluation of human operators in partly automated mobile working machines. *Automation Science and Engineering, IEEE Transactions on*, 7(1), 133–142.

Theodoridis, S. and Koutroumbas, K. (2008). *Pattern Recognition.* Academic Press, 4th edition.

Worrall, S. and Nebot, E. (2007). Automated process for generating digitised maps through GPS data compression. In *Australasian Conference on Robotics and Automation.*

Xu, R. and Wunsch II, D. (2009). *Clustering.* Wiley-IEEE Press.