<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *19th IEEE International Conference on Intelligent Transportation Systems, ITSC 2016; Windsor Oceanico Hotel Rio de Janeiro; Brazil; 1 November 2016 - 4 November 2016*.

Citation for the original published paper:

Fu, J., Jenelius, E., Haris, K. (2016)

Driving time and path generation for heavy construction sites from GPS traces.

In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* (pp. 1141-1146).
IEEE

<https://doi.org/10.1109/ITSC.2016.7795700>

N.B. When citing this work, cite the original published paper.

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-200584>

Driving Time and Path Generation for Heavy Construction Sites from GPS Traces

Jiali Fu, Erik Jenelius, Haris N. Koutsopoulos

Abstract— The paper presents a methodology for using GPS probe data to automatically extract the driving time between workstations and build a detailed representation of the paths between workstations in a construction environment. The inferred driving time distribution is aimed as input to construction simulation models to assess fleet performance, while the path information can be utilized to examine the performance of individual vehicles. A case study, using GPS data collected from a construction site, is used to demonstrate the capability of the proposed approach. The GPS data are processed without any prior knowledge about the underlying work environment. The results show that the proposed approach is capable of accurately inferring the driving time distribution and the paths between workstations.

I. INTRODUCTION

Heavy construction operations refer to the construction of highways, roads, dams, airports, industrial plants and related activities such as mining and quarries. Construction operations are in general complex systems and require the coordination of a wide variety of equipment. Heavy construction environments, also referred to as construction sites, are normally large and unstructured. It is crucial for the project management to plan and monitor the operations as accurately as possible. Nevertheless, the dynamic nature of operations and the frequently re-configured environment make the planning and monitoring process challenging. In practice, many heavy construction companies employ airplanes or helicopters equipped with laser scanning technology to generate maps of the construction site at regular intervals. However, this method is expensive and cannot capture the changes in the operating environment in between.

Accurate representation of vehicle travel times is essential for evaluating the productivity of heavy construction operations. Both internal and external factors influence the driving time of vehicles. Internal factors include the condition of vehicles while the external factors relate to the characteristics of the driving path. Manufacturers' performance charts are commonly employed to estimate the vehicles' travel times [1], [2]. This method is accurate but complex, and requires knowledge of the underlying environment that is not always available, and difficult to update. Further, the method does not take into account how

drivers interact with the topological environment and surrounding traffic. Researchers have developed various analytical methods for calculating travel time and duration of various construction activities [3]-[6]. These methods are simpler compared to using the performance charts, but still do not have the flexibility of adapting to geographical changes and taking drivers' performance into account.

As a result of rapid technological developments, GPS sensors have become common in commercial vehicles in recent years. A potential use of the vehicle trace data is to extract the travel times of vehicles traveling between tasks in the work area. Moreover, GPS data can be used to infer and update the layout and geometry of the underlying environment. A variety of map inference methods have been developed; a comprehensive overview and evaluation is given in [17]. In general, clustering methods group large datasets according to their similarity and are typically appropriate when vehicle traces are frequently sampled [7]-[12]. Incremental tracking insertion methods and intersection linking algorithms offer an advantage when no existing map information is available, which is often the case for heavy construction operations, as well as when the underlying environment has a frequently re-configured nature [13]-[16]. However, map inference for heavy construction operations is only applied in [9] from a safety perspective, with the aim of predicting and preventing collisions between vehicles.

Thus, this paper proposes a methodology for inferring the driving time distributions and generating detailed driving paths between workstations at heavy construction sites based on GPS data from hauling trucks. Compared to the performance charts and analytical methods, the inferred driving time is more realistic since it reflects the geographical condition of the construction site and also how operators actually drive. It is possible to manually collect the driving time distribution data at the construction site, but it is time-consuming, and not always feasible due to safety. Thus, the proposed method represents the actual conditions in a realistic manner, and at the same time is easier to apply on a continuous basis.

The inferred driving time distribution between workstations may serve as input to a Fleet Performance Simulation (FPS) model, proposed in a previous study [19], to examine the overall performance of a given fleet of construction vehicles. The FPS model uses as inputs the operating logic at the construction site and the vehicles' driving time between workstations, and computes the performance of the operation in terms of productivity, costs, resource utilization, *etc.* As such, FPS provides a better understanding of system behavior, impact of various factors and constraints and operational bottlenecks. This knowledge

J. Fu is with the Department of Transport Science, KTH Royal Institute of Technology, Teknikringen 10, 10044 Stockholm, Sweden (phone: +46-8-7908007; e-mail: jiali@kth.se).

E. Jenelius is with the Department of Transport Science, KTH Royal Institute of Technology, Teknikringen 10, 10044 Stockholm, Sweden (e-mail: erik.jenelius@abe.kth.se).

H. N. Koutsopoulos is with the Department of Civil and Environmental Engineering, Northeastern University, 400 Snell Engineering, Boston, MA 02115 USA (e-mail: h.koutsopoulos@northeastern.edu).

can lead to more efficient allocation of resources and ultimately optimization of system performance [20].

The connected paths between pairs of workstations generated in the proposed methodology are composed of segments with detailed information such as driving distance, road grade, curve radius, and maximum driving speed. The road grade has a large influence on the travel time and fuel consumption for heavy vehicles due to their large mass and heavy loads. Both theoretical and experimental studies have shown promising results in reducing fuel consumption and travel time simultaneously for heavy-duty vehicles through careful operation planning if the grade of the paths used is known in advance [22]. Thus, the path generation process is designed to provide inputs to vehicle dynamic simulation software that equipment manufacturers use to compute fuel consumption and other performance metrics at a detailed level. The dynamic models are employed to study the design of various components on the overall performance of vehicles. Until now, there is no efficient means to acquire the road topological information except through manual measurements at the construction site, which is both expensive and time demanding. The proposed method is able to produce detailed road segment information in a process that is less time-consuming and expensive to construct and update.

The route identification framework is demonstrated using a case study from a construction site located in Stockholm, Sweden. The GPS data are collected from heavy construction vehicles (excavators, wheel loaders, trucks etc.) during normal operations, and processed without assuming any prior knowledge about the underlying construction environment. The case study shows that the proposed method is capable of extracting driving time distribution between workstations and producing a directed road map of the unknown terrain at a detailed level.

II. METHODOLOGY

In the route identification framework, the known locations of workstations and GPS traces from haulage trucks are used to first identify vehicle trajectories between pairs of workstations, and to extract the driving time distribution using these trajectories. Subsequently, the algorithm uses the trajectories to derive a representative path with detailed segment information between each workstation pair. The framework is illustrated in Figure 1. The two elements of the route identification methodology are described below.

A. Driving Time Distribution

The location inference method proposed in [18] computes the probability distribution of the locations of the workstations, and provides information about the workstation types as well as the center points of each workstation. The results are utilized here to extract the driving time distribution between workstations. The procedure of the driving time distribution module is separated into the following two steps.

1. Extract vehicle trajectories between each workstation pair.
2. Identify and filter trajectory outliers, and compute driving time distributions between workstations from the remaining trajectories.

Figure 1 The framework of route identification algorithm

In the first step, the algorithm goes through the GPS trace points of each haulage truck and detects when the truck leaves one workstation and arrives at another. The workstation pair is ordered as an origin/destination (OD) pair, for instance from a loading station (LS) to a dumping station (DS) and the reversed order (DS-LS). For each OD pair, a set of trajectories are thus identified, where each trajectory is a sequence of GPS trace points from a haulage truck traveling between the OD pair, and each trace point represents a position in a 3-dimensional space at a certain instant in time.

Haulage trucks may perform other tasks or take a detour in the transport process, and these “anomalous” trajectories should be excluded from the estimation of the driving time distribution. It is assumed that the anomalous trajectories are outliers in terms of travel times, and are removed using standard outlier detection methods (*e.g.* based on the percentiles of the full distribution). Finally the driving time distribution between each workstation OD pair is computed from the filtered trajectory database.

B. Road Segment Description

In this module, identified vehicle trajectories are used to derive a directional path between each OD pair, with detailed segment information. The procedure consists of three steps: cluster generation, road segment linkage, and road segment computation.

1. Generate clusters (representing node) along vehicle trajectories between each workstation OD pair.
2. Link clusters into directional edges for each OD pair.
3. Compute detailed information of path segments (segment length, gradient, curve radius, *etc.*).

1) Cluster generation along vehicle trajectories

The path generation methodology is illustrated in Figure 2. Vehicle trajectories are shown as grey lines in Figure 2(a) with GPS trace points depicted as squares. In the first step (Figure 2(b)), the algorithm examines the GPS points and places initial cluster centroids along a random vehicle trajectory (In Figure 2(b), the trajectory to the right with GPS points marked with black filled squares) at equal distances d_{init} . The initial cluster centroids are shown as a blue dot, a red square, and a purple cross in the figure.

The cluster generation procedure proceeds in a similar fashion as the k -means algorithm. In step 2, the algorithm calculates the distances between each GPS point and all cluster centroids and finds the candidate clusters within a certain distance range d_{max} . Each trace point is assigned to the closest candidate cluster such that the heading difference between the GPS point and the cluster is less than a threshold θ_{max} . In Figure 2(c), the GPS points assigned to each cluster are marked out in similar fashion as the initial cluster centroid. The heading difference of a GPS point x_t at timestamp t is defined as the change in moving direction between timestamps t and $t - 1$ [18]. The reason of keeping track of the heading information of the vehicles is that the clusters will be connected into road segments as directional links. If no existing cluster centroid satisfies both criteria a new cluster seed is created using this GPS point (step 2.1). As illustrated in Figure 2(c), a new cluster centroid (the dark green dot) is created and some of the GPS points (light green dots) are assigned to this cluster.

The choice of distance and heading difference thresholds, d_{max} and θ_{max} , depends mainly on the quality of the data, characteristics of the underlying application and use of the graph. In construction operations, the underlying road networks rarely have parallel lanes in the same direction due to space constraints, so the map does not have to include lane definition. However, if the resulting digital map is intended to be the input to dynamic simulation models that calculate driving time and fuel consumption, the distance threshold should be relatively small for accuracy. For the choice of heading difference threshold, roads in construction sites typically have large curvatures, and the threshold can be set to a relatively large value.

After initiating the cluster centroids (step 1) and assigning GPS points to clusters (step 2), the average position and heading of each cluster are recomputed in step 3 (new locations of cluster centroids with heading information in Figure 2(d)) as well as the distance between the old and new centroid position Δ_{dist} . Δ_{dist} reveals how “similar” the members are to each other. Due to the change in position and heading of the centroids, some of the members may no longer satisfy the position and heading criteria, thus are removed and assigned to another cluster. If Δ_{dist} is larger than the critical distance value ψ_{dist} , the algorithm re-assigns the GPS points to new clusters obtained from the last iteration (step 4.1), and compute new cluster centroids (step 4.2). The algorithm iterates until the average centroid distance moved is no longer above the critical value ψ_{dist} , and the clusters reach their “steady state” positions [21]. The full algorithm is summarized in Table 2.

Figure 2 GPS trajectories and clusters generation procedure

TABLE 1 CLUSTER GENERATION ALGORITHM

Step 1	Place initial cluster seeds at equal distance d_{init} along the vehicles' trajectories between each OD pair
Step 2	Assign the GPS points to the clusters using the position and heading criteria Step 2.1 If a GPS point cannot be assigned to the existing clusters, generate a new cluster using this GPS point
Step 3	Compute the centroids of clusters (location and heading) and calculate the average distance moved from the initial seeds, Δ_{dist}
Step 4	While $\Delta_{dist} > \psi_{dist}$, where ψ_{dist} is the given maximum distance threshold of cluster centroids, or the number of iterations is less than the maximum iteration threshold, execute Step 4.1 Assign GPS points to the clusters according to the position and heading criteria Step 4.2 Re-compute the cluster centroids (location and heading criteria) and calculate the average distance moved from last iteration, Δ_{dist}
Step 5	Stop

2) Linking clusters

Once the clusters are placed along the driving path, the next step is to connect the centroids into a directed graph

using the sequence of the GPS trajectories extracted between workstation OD pairs. However, many links are superfluous and are removed. A transition count matrix and transition probability matrix as in [9] are computed to reduce the complexity of the graph. A transition count matrix of size $N \times N$ (N is the total number of clusters) records the number of vehicle trajectories going from cluster i to cluster j . The diagonal elements in the matrix are zero. The transition count matrix is further row-normalized to form the transition probability matrix, A . The probability matrix provides a measure of importance of each edge. Thus, the less frequent used paths are disconnected from the original graph in order to reduce the complexity of the graph.

As illustrated in Figure 3, the transition probability from cluster 2 to cluster 4, A_{24} , has a lower value than the likelihood A_{23} , and it is possible to travel from cluster 2 to cluster 4 via cluster 3. Thus, the link that connects cluster 2 and 4 is regarded as redundant and removed from the graph.

Figure 3 Linking cluster centroids

3) Extracting road segment information

After constructing the driving path between each workstation pair, the length of each segment in the (x,y) -space is calculated as the Euclidean distance between the two connected cluster centroids, Δd . Since the segments are relatively short by design, it is assumed that the gradient and curvature of each segment are constant.

The GPS dataset also includes altitude information. The altitude of each cluster centroid along the driving path is first computed by taking the average altitude value of all trace points assigned to the cluster, and the altitude difference, Δh , between two connected clusters is calculated. The grade of a road segment is calculated as $\alpha = \tan^{-1}(\Delta h / \Delta d)$. Moreover, taking altitude into account, the actual distance travelled is $\sqrt{\Delta d^2 + \Delta h^2}$.

Finally, the curvature κ of the road segment is estimated from the GPS coordinates in (x, y) -plane of several connected points using the method proposed in [23]. The assumption of plane curves on the ground plane is used, *i.e.* the vertical dimension of the road is ignored. The curvature of a road limits the speed of a vehicle and is an important input to any vehicle dynamic simulation.

III. CASE STUDY

The proposed route identification method is applied in a case study using data from a large quarry site located north of Stockholm, Sweden. The quarry site produces gravel,

aggregate, and sand. There are around 20 construction vehicles, including wheel loaders, excavators and dump trucks working at the site.

A. Experimental Design

Three GPS Garmin GPSmap 62 devices are used to collect GPS trace points from one wheel loader and two haulage trucks. The GPS receivers register timestamp, latitude, longitude and altitude at 1 Hz. Four hours of GPS data were collected, with over 14,000 probes for each vehicle.

Figure 4(a) shows the recorded GPS trace points from the three construction vehicles in different colours. The GPS coordinates are omitted for confidentiality reasons. It is observed that the wheel loader with traces marked in pink dots drives between two loading stations. The two dump trucks (blue for truck 1 and green for truck 2) travel mostly between the two loading stations and the dumping station, and occasionally to other locations. The locations of loading and dumping stations were estimated from the GPS dataset using the method of [18] and are depicted in purple in Figure 4(b), 4(c). Two loading stations and one dumping station were detected.

Figure 4 GPS traces and the probabilities of workstations

B. Results

1) Driving time extraction

Given the locations of workstations, the vehicle trajectories between each workstation OD pair are extracted from the haulage trucks' GPS trace points, and the driving times of all trajectories are computed. From these trajectories, it is clear that the trucks occasionally travel to other places at the construction site to assist with other work tasks, and the driving times of these unusual activities are much lower or higher than the regular trips. A simple detection method using Tukey's test [24] is used to filter outliers, and the remaining trajectories are used to compute the driving time distribution. The final trajectories from the two identified loading stations to the dumping station are shown in dark green in Figure 5. In total, there are 20 trajectories from

loading station 1 (LS1) to the dumping station (DS), and 21 trajectories from loading station 2 (LS2) to DS.

Figure 5 Trajectories from the loading stations to the dumping stations

The driving time distributions between the workstation OD pairs, LS1-DS, S2-DS, DS-LS1, DS-LS2, are shown in Figure 6 as boxplots with the top and bottom of each box at the 25th and 75th percentiles of the samples, respectively. The outliers are marked with a red “+” sign. It is observed that the average driving times between LS1 and DS are shorter than between LS2 and DS in both directions due to shorter travel distances. Further, the trips from the loading stations to the dumping station are generally longer than the return trips because of the carried load. The outliers show that there are more unusual activities in the return trips compared to the hauling trips. This is reasonable since it is more likely for a truck to carry out other tasks after emptying its load at the dumping station.

Figure 6 Driving time distribution between workstation OD pairs

2) Road segment description

In the road segment generation procedure, the computed characteristics of the road segments are designed as inputs to detailed vehicle dynamic simulations. The distance between initial cluster centroids, d_{init} , is set to 10 meters. The heading difference δ_θ is chosen as 90° and the maximum distance threshold ψ_{dist} for convergence is 0.5 meters.

The clustering generation procedure converges after the 4th iteration for both directions. The centroids show a clear pattern of two parallel lanes in opposite directions at the relative straight road section because the equipment operators drive on separate lanes. On the more curved road section, meanwhile, the cluster centroids lie almost along a single line. It is worth pointing out that the resulting driving paths

represent the average of all trajectories between each OD pair, but not necessarily the most frequently travelled paths.

Figure 7(a) shows the produced driving paths between LS1 and DS, with GPS points plotted in light blue and green in the background. The black line is the LS-DS path with 94 segments, and the red line is the return DS-LS path (81 segments). The obtained driving paths follow intuition. During hauling trips drivers stay on the left side of the opposite lane after the sharp curve, which is consistent with manual observation at the site. The length of this section is short, and the trucks turn north when entering the dumping station, with the drivers on the left side of the road after the sharp curve. Hence, it is concluded that the underlying topology between the workstations is correctly captured.

Figure 7 Inferred paths between loading station 1 and the dumping station

A zoom-in plot of the sharp curve is shown in Figure 7(b), with the travel direction of each road segment indicated by arrows. On the return trip, the trucks turn away from the main road at the curve, possibly yielding to oncoming traffic. Normally empty trucks give way to loaded trucks for safety reasons. It is observed that there are not many trace points around the generated path. This is because the generated path is the average of all trajectories, even though the trucks give way only on a few return trips.

The computed altitude profile of the driving path between OD pair LS1-DS is depicted in Figure 8. The driving path consists of frequent up- and downhill slopes that are common in construction environments. In particular, a steep uphill

slope approximately 100 meter after the loading station can be seen, which is consistent with observations at the site.

Figure 8 Altitude profile of the driving path between the OD pair LS1-DS

IV. CONCLUSIONS AND FUTURE WORK

The topology of a construction site changes over time. The paper proposes an approach to automatically extract driving times and detailed paths between workstations using GPS data from vehicles. The approach first generates clusters based on location and travel direction from the GPS probes, and then links the clusters into a directed graph. The algorithm was tested using GPS data from a quarry site in Sweden. The results suggest that the proposed algorithm is capable of extracting the driving time distribution between workstations, as well as accurately capturing the shape and topology of the construction site. However, evaluation using different amounts of GPS data and sample resolutions is still needed to evaluate the sensitivity of the proposed approach.

Combined with a method to extract the locations of various workstations in a construction environment, the map inference method represents an integrated framework that can automatically generate a complete and detailed map of the locations of various workstations and the underlying directed road network using only GPS measurements. The digitalized map may be continuously updated using GPS and other data sources collected from daily operations to monitor changes in the construction environment.

The proposed method may be extended to extract durations of other construction activities such as loading times, resource idle times, etc. In addition to GPS data, large amounts of data are increasingly becoming available from other sources such as the vehicles' CAN buses and other on-board sensors. Utilizing such information to extract more useful representations of the construction environment and to achieve more efficient heavy construction operations is an interesting direction for future research.

ACKNOWLEDGMENT

This project is financially supported by the Swedish Government Agency for Innovation Systems (Vinnova). The authors would like to express their gratitude to the industry collaborator, Volvo Construction Equipment, for their constant support and providing us with access to job sites

and project data. We would also like to thank the equipment operators at Skanska for their help with obtaining the data used in the paper.

REFERENCES

- [1] *Volvo Site Simulation Program Manual*. Volvo Construction Equipment, 2015.
- [2] *Caterpillar Performance Handbook*, Edition 45. Caterpillar Inc. 2015.
- [3] S. M. AbouRizk and D. W. Halpin, "Statistical properties of construction duration data," *Journal of Construction Engineering and Management*, vol. 118, no. 3, pp. 525-544, 1992.
- [4] J. C. Hicks, "Haul-unit performance," *Journal of construction engineering and management*, vol. 119, no. 3, pp. 646-653, 1993.
- [5] M. Marzouk, *Optimizing earthmoving operations using computer simulation*. PhD thesis. Concordia University, 2002.
- [6] C. Schexnayder, K. Knutson, and J. Fente, "Describing a beta probability distribution function for construction simulation," *Journal of construction engineering and management*, vol. 131, no. 2, pp. 221-229, 2005.
- [7] S. Edelkamp and S. Schrödl, "Route planning and map inference with global positioning traces," *Computer Science in Perspective*, Springer Berlin Heidelberg, pp. 128-151, 2003.
- [8] S. Schroedl, K. Wagstaff, S. Rogers, P. Langley, and C. Wilson, "Mining GPS traces for map refinement," *Data mining and knowledge Discovery*, vol. 9, no. 1, pp. 59-87, 2004.
- [9] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Robust inference of principal road paths for intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 298-308, 2011.
- [10] J. Biagioni, and J. Eriksson, "Map inference in the face of noise and disparity", In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, pp. 79-88, 2012.
- [11] J. J. Davies, A. R. Beresford and A. Hopper, "Scalable, distributed, real-time map generation", *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 47-54, 2006.
- [12] D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial-temporal data", *Data & Knowledge Engineering*, vol. 60, no. 1, pp. 208-221, 2007.
- [13] L. Cao, and J. Krumm, "From GPS traces to a routable road map", In *Proc. of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 3-12, 2009.
- [14] S. Rogers, P. Langley, and C. Wilson, "Mining GPS data to augment road models", In *Proc. of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 104-113, 1999.
- [15] A. Fathi, and J. Krumm, "Detecting road intersections from GPS traces", In *Geographic Information Science*, pp. 56-69, 2010.
- [16] S. Karagiorgou, and D. Pfoser, "On vehicle tracking data-based road network generation", In *Proc. of the 20th international conference on advances in geographic information systems*, pp. 89-98, 2012.
- [17] M. Ahmed, S. Karagiorgou, D. Pfoser, and C. Wenk, "A comparison and evaluation of map construction algorithms using vehicle tracking data", *GeoInformatica*, vol. 19, no. 3, pp. 601-632, 2015.
- [18] J. Fu, E. Jenelius and H. N. Koutsopoulos, "Identification of workstations in earthwork operations from vehicle GPS data", In *Proc. of the Transportation Research Board Annual Meeting*, 2016.
- [19] J. Fu, "A microscopic simulation model for earthmoving operation", In *Proc. of the International Conference on Sustainable Design and Construction*, pp. 218-223, 2012.
- [20] J. Fu, E. Jenelius, and H. N. Koutsopoulos, "Optimal fleet selection for earthmoving operations", In *Proc. of 7th International Structural Engineering and Construction Conference*, pp. 1261-1266, 2013.
- [21] A. K. Jain, "Data clustering: 50 years beyond K-means", *Pattern recognition letters*, vol. 31, no. 8, pp. 651-666, 2010.
- [22] J. Fu and G. Bortolin, "Gear shift optimization for off-road construction vehicles", *European Journal of Transport and Infrastructure Research*, vol. 14, no. 3, pp. 214-228, 2014.
- [23] *Tools and methods to measure and analyze data needed for construction productivity*, Volvo internal report, 2013.
- [24] J. W. Tukey, *Exploratory Data Analysis*. Addison-Wesley, 1977.