

METHOD, DATA PROCESSING APPARATUS AND COMPUTER PROGRAM PRODUCT FOR GENERATING MAP DATA

TECHNICAL FIELD

[0001] Various embodiments relate to a method for generating map data, thus identifying uncharted roads of a road network in a map. Various embodiments relate to a data processing apparatus configured to carry out said method for generating map data. Various embodiments relate to a non-transitory computer-readable medium storing computer executable code including instructions for generating map data. Various embodiments relate to a computer executable code including instructions for generating map data.

BACKGROUND

[0002] The most important map feature for navigation and for estimated time of arrival (ETA) calculation is considered to be the road network. However, some missing roads from the road network could lead to wrong outcomes both on navigation and estimated time of arrival (ETA) calculation.

[0003] Global Positioning System (GPS) traces collected by mobile phones when they are used in motor vehicles can offer a good indication about whether or not there is a road in that region. Therefore, in theory, GPS traces collected by mobile phones can be used to infer the road network - based on the paths, where motor vehicles have been - and further detect the roads not present in the reference map. However, there are multiple sources of errors when computing the signals from GPS that prevents using GPS signals directly to infer road networks. These errors are due to atmospheric effects, e.g. ionospheric delay, humidity, atmospheric pressure, and multipath effects, e.g. radio signals reflect off surrounding terrain, and satellite's clock errors, space weather, and other artificial sources of interference, etc. All those sources of errors can lead to so called GPS jumps or GPS drifts.

[0004] Conventionally, for using GPS traces with the purpose to infer road networks, intersections are identified and the geometric representation of each road segment by track alignment is extract. Further conventionally, spatial-linear clustering is used to infer road segments from GPS traces. Further conventionally, intersection detection that uses a localized shape descriptor to represent the distribution of GPS traces around a point is used. However, in of these conventional techniques, above-mentioned GPS errors create false positive signals leading to

false positive roads detected. In particular, there is a high chance of encountering a Type I or Type II error when there are only a few GPS traces to support the detection of a road segment.

SUMMARY

[0005] In various aspects a method for generating map data is provided. The method may include generating of a plurality of cells of a real space area, each cell representing a two-dimensional sub-space of the real space area. Each of the cells include at least one global positioning system (GPS) point. The real space area may include a road network. The method may further include determining a number of GPS points for each cell of the plurality of cells. The method may further include determining a road network based on the numbers of GPS points using a Smoothed Particle Hydrodynamics (SPH) analysis. The cells of the plurality of cells may be particles of the SPH analysis and the number of GPS points per cell of the plurality of cells may be a scalar of each cell of the plurality of cells in the SPH analysis. The method may further include comparing the road network determined using the SPH analysis with the road network of the real space area, e.g. stored in map of the real space area.

[0006] Another aspect of the disclosure relates to various embodiments of a data processing apparatus configured to carry out the method of generating map data. As an example, the data processing apparatus may include at least one processor, and a memory having instruction stored therein that, when executed by the processor, cause the processor to: generate a plurality of cells of a real space area, each cell representing a two-dimensional sub-space of the real space area, wherein each of the cells comprise at least one global positioning system (GPS) point, and wherein the real space area comprises a road network; determine a number of GPS points for each cell of the plurality of cells; determine a road network based on the numbers of GPS points using a Smoothed Particle Hydrodynamics (SPH) analysis, wherein the cells of the plurality of cells are particles of the SPH analysis and the number of GPS points per cell of the plurality of cells are a scalar of each cell of the plurality of cells in the SPH analysis; and compare the road network determined using the SPH analysis with the road network of the real space area.

[0007] Another aspect of the disclosure relates to various embodiments of a non-transitory computer-readable medium storing computer executable code including instructions for the method of generating map data.

[0008] Another aspect of the disclosure relates to various embodiments of a computer executable code including instructions carrying out the method of generating map data.

[0009] Illustratively, roads are considered as continuous surfaces, and when roads are mapped, roads become continuous two-dimensional (2D) surfaces. Recorded GPS points can be considered as a cloud of discrete particles. In various aspects, a method to map the particle's cloud to the road continuous surface is provided. As an example, the method may be based on Smoothed Particle Hydrodynamics (SPH) applied to generate a road network.

[0010] The method may include splitting 2D spaces of a map into regular grids. Each grid cell, where at least one GPS point is available, may become a particle. The scalar property of a particle may be the number of GPS points recorded.

[0011] The method may further include determining of a density field underlying the scalar property. The density field may be traced by particles using a SPH simulation. The method may further include providing a visualization image for the density field obtained after simulation.

[0012] The method may further include a classifying of each pixel from the resulting image, e.g. using a deep neural network, as a part of a road or not.

[0013] The method may further include to generate a network topology with geo-coordinates attributes from resulted binary mask classifying of each pixel. The method may further include a comparison of the resulting network with a reference road network. The method may further include identifying missing roads in the reference map.

[0014] The method may be applied when there is not a huge volume of GPS traces from the same road segment available. Illustratively, when there is a huge volume of GPS traces from the same road segment available, the average of the signals can give a good indication about the presence and the geometry of the road segment, due to the law of large numbers.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

- FIG. 1 to FIG. 3 show representations of a method of generating map data; and

- FIG. 4 exemplary shows a flow diagram of a method for generating map data according to various aspects.

DETAILED DESCRIPTION

[0016] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details and embodiments in which the disclosure may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the disclosure. Other embodiments may be utilized and structural, and logical changes may be made without departing from the scope of the disclosure. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

[0017] Embodiments described in the context of one of the enclosure assemblies, vehicles, or methods are analogously valid for the other enclosure assemblies, vehicles, or methods. Similarly, embodiments described in the context of an enclosure assembly are analogously valid for a vehicle or a method, and vice-versa.

[0018] Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments. Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments, even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0019] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0020] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0021] As used herein, the term “new road segments”, “missing road segments”, or “uncharted road segments” refer to real space road segments that are not considered in a road network of a map of the real space area including the real space road segments.

[0022] As used herein, the term smoothed-particle hydrodynamics (SPH) analysis refers to a computational method used for simulating the mechanics of continuum media and is a meshfree Lagrangian method, and the resolution of the method may

easily be adjusted with respect to variables such as the size of the cells or the number of GPS points per cell.

[0023] As used herein and in accordance with various embodiments, the expression “method for” may also include the meaning of “method of”, and vice versa.

[0024] As used herein and in accordance with various embodiments, the term “cell” may include geographical data, such as geospatial coordinate and may further include time, for example, as provided by the global positioning system GPS. The GPS coordinates may according to the World Geodetic System, WGS 84, for example, version G1674.

[0025] For the identification of uncharted road segments, the location data is real world data, for example real world GPS data. Correspondently, the geographical perimeter represents a perimeter on earth’s surface.

[0026] As used herein and in accordance with various embodiments, the terms ‘geographical’ and ‘geospatial’ may be used interchangeably.

[0027] The service of ride-hailing providers relies significantly on the quality of the digital map. The missing of a road segment from the map can lead to suboptimal routing decision or inaccurate estimation of a driver’s arrival time. Though OpenStreetMap (OSM) provides the community user-generated maps of the world, its data completeness and accuracy vary significantly in different cities. Considering the high cost of manual corrections of map data, using data automation may provide advantages for providing accurate maps. For example, data-driven approaches can learn from cities with relatively higher quality map data (e.g., Singapore) to automatically discover missing roads and intersections in cities with relatively poorer map data (e.g., Jakarta). To this end, various embodiments of the present disclosure use a multi-scale graph convolutional network that can effectively detect uncharted road segments from large-scale GPS trajectories.

[0028] While the disclosure has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. The scope of the invention is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

[0029] The method may include that a 2D space is split in cells, e.g. geohashes, e.g. with a length of 9, 4.77m×4.77m. Grid cells having no GPS points may have no

particle assigned. Grid cells having at least one GPS point may have a particle assigned. The number of GPS points, e.g. as they were collected by mobile phones, may be the scalar property of a particle. Alternatively, a density may be selected from one timeframe. In various aspects, all GPS points may have the same weight in the method regardless of their source or age. The method may further include, e.g. after this step, a scalar for each geohash representing its GPS points density may be provided. SPH may be configured to reconstruct a continuous field from a cloud of discrete particles which may be the property carriers.

[0030] SPH may be configured such that the volume of each particle is set to a constant, e.g. one. SPH may include a kernel that may be a Dome-shaped quadratic smoothing function.

[0031] In SPH, particles may interact through a kernel function (W) with characteristic radius, also denoted as "smoothing length". The smoothing length may be represented in equations by h . As an example, $W(R, h) = \alpha_d(1-R^2)$ with $\alpha_d=15/8(\pi h^3)$. Here, " h " is a smoothing length, " R " is the relative distance between two points: $R = r / h$, and " r " is the distance between the two points: x_i and x_{i+1} . The smoothing length may be variable. The value for each particle may be determined as the distance to its 16th closer neighbors. Here, the position of a particle would be the center of a geohash. The resulting field may be interpreted according to various aspects as a drivable area field. A high value of the area field in a certain point may be interpreted as a high probability of having a road there and vice-versa, as an example.

[0032] In various aspects, a GPS density of any particle of the particles may be determined by summing the GPS density of all the particles that lie within the range of the kernel, the kernel being used as a weighting function W .

[0033] Hence, $A(r)$ field may be written as a convolution with W :

$$[0034] \quad A(r) = \int A(r')W(r, r', h)dV(r').$$

[0035] Here, $A(r)$ may be the integral interpolant of one or more reconstructed GPS traces density. $A(r')$ may be the local value for GPS traces density where the integration is over the entire circle having radius smoothing length.

[0036] For numerical work the integral interpolant may be approximated by a summation interpolant:

$$[0037] \quad A(r) = \sum_j V_j A_j W(r, r', h).$$

[0038] Here, the summation index " j " may denote a particle label, and the summation may be over all the particles localized in the circle of radius "smoothing length". V_j may be the volume of particle j . The volume V_j may be set to a constant

value of 1. A_j may represent the value of a GPS density, e.g. the scalar, in the point j .

[0039] The accuracy of the field reconstruction may be determined by a first factor and a second factor. A first factor may be how accurate particles are distributed over the 2D space. A second factor may be how good the estimated density of particles is. The first factor and second factor may be derived directly from measurements.

[0040] The properties of a kernel function used to interpolate the discrete particles may include a Dome-shaped quadratic smoothing function. A variable "smoothing length" may be used instead of a constant in case of a huge variability in GPS points density in similar roads.

[0041] In various aspects, the smoothing length of each particle of the particles may be determined, e.g. computed, as the distance to its "nb" closer neighbor. Here, nb may be a constant in a range of 2 to 20, e.g. having value of 16. The result of this procedure may be a field of GPS point density. The GPS point density may be interpreted as road existence probability density function, e.g. a probability of being a road in any point in 2D space. Illustratively, the input to the SPH may represent an unprocessed heatmap built from the GPS points, and the output may be a normalized heatmap. In the heatmap, the roads may be clearly visible.

[0042] In various aspects, a surface density of the distribution of particles may be plotted to generate visualizing images for the density field obtained after simulation, e.g. using a Py-SPHViewer library. A resulting image may be depicted in FIG. 1. The generated image 100 may act as an input into a semantic segmentation model, e.g. a deep learning machine learning. The generated image 100 may include first areas 120 that are associated with roads (in FIG.1 illustrated as bright areas), and may include second areas 110 that are associated with non-road areas 110, e.g. a building, a park, a pedestrian zone, etc. (in FIG.1 illustrated as dark area).

[0043] The purpose of the semantic segmentation model may be to learn the transformation of roads from feature space of road confidence that was generated before to the feature space of road-not road classification at pixel level.

[0044] In various aspects, the semantic segmentation model may include a dataset used for train, test and evaluation. The dataset may be generated by labeling all roads where received GPS points exist. Labels may be obtained by querying the OSM, and by retrieving all existing geometries.

[0045] The geometries in an image may then be converted by drawing the lines on an empty canvas, and, thus, a binary image may be obtained. In the binary image,

each pixel may have a first label and a second label. For example, the first label may be “road” and the second label may be “not road”. OSM may be queried on areas where OSM data has a very good quality.

[0046] The input images for a dataset may be obtained on selected OSM areas on top of already collected GPS points by the process steps described before. Thus, images may be generated as described before.

[0047] In various aspects, a Convolutional Network for Semantic Segmentation, e.g. a U-Net architecture-model may be trained. In the U-Net architecture, a resnet18 backbone may be used. The resnet18 backbone may provide an optimal accuracy-speed tradeoff.

[0048] An evaluation metric may be used in pixel-wise accuracy. The result may be a model able to predict at pixel level whether a road exists on that pixel or not.

[0049] When running an inference, each pixel may be classified as a part of a road. Alternatively, a binary mask having the same shape may result as an input image. An output from a neural network may be used, in which for each pixel there may be a probability of being a road, and the output may be binarized by applying a confidence threshold. The confidence threshold may be equal to 0.2. An example of an output is illustrated in FIG. 2. The generated image 200 may include first areas 220 that are associated with roads, and may include second areas 110 that are associated with non-road areas 110, e.g. a building, a park, a pedestrian zone, etc.

[0050] The binary mask may be used to generate a network topology. Each edge may represent a road using a skeletonization algorithm implemented in an image processing library. The result may be a road network having a graph representation where each node may have a correspondent geo-position, and edges may represent road segments.

[0051] The resulting network illustrated in a first map may be compared with a reference road network derived from a reference map in a second map, the second map may be the map of the real space area.

[0052] All segments presented in the first map and missing in the second map may be considered as being roads missing from the reference map/second map.

[0053] The comparison of the first map with the second map may be performed by taking each predicted road segment and checking whether there are existing roads in its proximity, for example.

[0054] In various aspects, a buffer around all the roads which already exist in the map may be used. Then, an intersection between a new missing roads segment and a road segment in the buffer may be computed. As an example, a confidence

value may be assigned to the intersection depending on the intersection percentage between the missing road and the road segment in the buffer. The confidence value may represent a probability of the road to be missing or not. As an example, Confidence 1 may have an intersection percentage < 0.25 and missing roads with high confidence, Confidence 2 may have an intersection percentage between 0.25 and 0.5 and missing roads with medium confidence, Confidence 3 may have an intersection percentage between 0.5 and 0.75 and missing roads with low confidence, and Confidence 4 may have an intersection percentage > 0.75 of existing roads. In an output 300, illustrated in FIG. 3, each segment may be color-coded by its degree of confidence, e.g. Confidence 1 330, Confidence 2 340, Confidence 3 350, and Confidence 4 320. Non-road areas 310 may have no color code.

[0055] In the illustrative Example depicted in FIG.3, a map of the metropolitan area of Bandung city from Indonesia is shown on the area defined by the following geo-coordinates: Top Left: -6.8553718, 107.5340093, Bottom Right: -6.9916266, 107.6685388, Geohashes: qqu2z qqu8b qqu8c qqu2x qqu88 qqu89 qqu2r qqu82 qqu83. FIG.3 illustrates a detected total a road network of 4069 km out of which: missing roads with high confidence 330: 495 km (Confidence 1), missing roads with medium confidence 340:196 km (Confidence 2), missing roads with low confidence 350: 160 km (Confidence 3), and existing roads 320: 3219 km (Confidence 4).

[0056] Illustratively, FIG. 4 depicts a flow diagram of a method 400 for generating map data according to various aspects. The method 400 may include generating 402 of a plurality of cells of a real space area, each cell representing a two-dimensional sub-space of the real space area. Each of the cells include at least one global positioning system (GPS) point. The real space area may include a road network. The method may further include determining 404 a number of GPS points for each cell of the plurality of cells. The method may further include determining 406 a road network based on the numbers of GPS points using a Smoothed Particle Hydrodynamics (SPH) analysis. The cells of the plurality of cells may be particles of the SPH analysis and the number of GPS points per cell of the plurality of cells may be a scalar of each cell of the plurality of cells in the SPH analysis. The method may further include comparing 408 the road network determined using the SPH analysis with the road network of the real space area.

[0057] In various aspects, the method may further include determining a first plurality of cells. The first plurality of cells may include the plurality of cells and at least one further cell. The at least one further cell may include no GPS point.

[0058] The GPS points may be provided from a plurality of mobile phone users in the real space area.

[0059] The cells of the plurality of cells may include that the cells may be spaces of a grid. Each cell may include the same size. Each cell may be a Geohash. Each of the GPS point of a cell may have the same weight as the other GPS points of the same cell.

[0060] Determining the road network may include a reconstructing of a continuous field from the plurality of cells. The continuous field may be interpreted as a drivable area field. A high value of the continuous field in a certain point of the real space area may be interpreted as a high probability of having a road in the point. A number of GPS points per cell may be interpreted as road existence probability density function. The input to the SPH analysis may represent an unprocessed heatmap built from the GPS points of each cell and the output may be a normalized heatmap. At least one visualization image for the distribution of number of GPS points per cell may be obtained after simulation. Each pixel from the resulting visualization image may be classified to determine whether the pixel may be a part of a road or not. The classifying may be using a deep neural network model with the visualization image as an input into a semantic segmentation. A dataset may be provided by labelling all roads where GPS points exist. Labels may be obtained by querying an open street map (OSM) and retrieving all existing geometries. A binary image may be generated in which every pixel may have a first label associated to a road or a second label associated to a non-road area by converting the geometries in the image by drawing a plurality of lines on an empty canvas. A convolutional network for semantic segmentation may be trained. The method may include the semantic segmentation. An inference may be run where each pixel may be classified having the first label or the second label resulting in a binary mask. The resulting binary mask may have the same shape as the input image. A network topology may be generated from the resulting binary mask. The network topology may include geo-coordinates attributes. The road network may include a graph representation with each node having a correspondent geo-position and edges represent road segments. The road segments presented in the road network determined by the SPH analysis may be considered as roads missing in the map of the real space area. It may be determined whether there may be existing roads in a proximity of new road segments determined by SPH analysis. A buffer around road segments known from the map of the real space area may also be determined. An intersection between uncharted road segments of the real space area road network and the buffer may be

determined depending on the intersection percentage between the missing road (also denoted as uncharted road) and the one stored in the buffer (e.g. the road network stored in the buffer). A confidence value may be determined for the intersection percentage. A color-coding output representing missing roads may be provided with different labels, e.g. high confidence, medium confidence or low confidence.

[0061] As an example, the method may include a pre-processing of GPS data for the purpose of missing road detection, a splitting of 2D spaces into regular grids (also denoted as plurality of grid cells or grid cells), e.g. splitting in geohashes. The method may further include, for each grid cell, if there exists at least one GPS point, that the cell becomes a particle and that the grid cells having no GPS points will have no particle assigned. The method may further include a recording of the scalar property of particle being the number of GPS points such that all GPS points have the same weight or that, for each geohash, a scalar representing its GPS points density. The method may further include an estimating of the underlying density field traced by particles using Smoothed Particle Hydrodynamics (SPH) simulation by reconstructing a continuous field from a cloud of discrete particles. The resulting field may be interpreted as a drivable area field with the high value of the field in a certain point interpreted as a high probability of having a road there and vice-versa. The result of this step may be a field of GPS points density which can be interpreted as road existence probability density function (probability of being a road in any point in 2D space). The input to the SPH may represent an unprocessed heatmap built from the GPS points and the output may be a normalized heatmap in which the roads are much more clearly visible. The method may further include a generating of visualization images for the density field obtained after simulation by plotting surface density of particles distribution. The method may further include a classifying of each pixel from the resulting visualization image to determine whether the pixel is a part of a road or not using a deep neural network model with the visualization image as an input into a semantic segmentation. The method may further include a building of a dataset by labeling all roads where GPS points exist where the labels are obtained by querying the open street map (OSM) and retrieving all existing geometries, converting the geometries in the image by drawing the lines on an empty canvas, thus obtaining a binary image in which every pixel has two labels: road/not road, training a convolutional network for (7c1) semantic segmentation, and running inference where each pixel is classified as a part of a road or not resulting in a binary mask having the same shape as input image. The

method may further include, from a resulted binary mask, generating a network topology with geo-coordinates attributes where the road network having a graph representation with each node having a correspondent geo-position and edges represent road segments. The method may further include comparing the resulting network with a reference road network and identifying missing roads in the reference map where the road segments presented in the resulting network and missing in the reference network are considered as roads missing from the reference map, taking each predicted road segment and checking whether there are existing roads in its proximity, computing a buffer around all the roads which already exist in the map and then computing the intersection between a new missing roads segment and that buffer, and depending on the intersection percentage between the missing road and the buffer, assigning a confidence value to it, which represents the probability of the road to be missing or not. The method may further include a color-coding output representing missing roads with high/medium/low confidence.

[0062] The present disclosure also relates to a computer program product configured to carry out the method for generating map data according to various embodiments.

[0063] The present disclosure also relates to a data processing apparatus configured to carry out the method for generating map data according to various embodiments.

[0064] The data processing apparatus may include a processor, e.g. a microprocessor, a memory, communication devices, e.g. to enable access to a network.

[0065] The present disclosure also relates to a non-transitory computer-readable medium storing computer executable code including instructions for identifying road intersections according to various embodiments. Examples of non-transitory computer-readable medium are a flash memory, a hard disk drive, a CD-ROM, a DVD.

[0066] The present disclosure also relates to a non-transitory computer-readable medium storing computer executable code including instructions for training the classifier.

[0067] Examples of non-transitory computer-readable medium are a flash memory, a hard disk drive, a CD-ROM, a DVD.

[0068] The data processing apparatus includes at least one processor, and a memory having instruction stored therein that, when executed by the processor, cause the processor to: generate a plurality of cells of a real space area, each cell

representing a two-dimensional sub-space of the real space area, wherein each of the cells comprise at least one global positioning system (GPS) point, and wherein the real space area comprises a road network; determine a number of GPS points for each cell of the plurality of cells; determine a road network based on the numbers of GPS points using a Smoothed Particle Hydrodynamics (SPH) analysis, wherein the cells of the plurality of cells are particles of the SPH analysis and the number of GPS points per cell of the plurality of cells are a scalar of each cell of the plurality of cells in the SPH analysis; and compare the road network determined using the SPH analysis with the road network of the real space area.

[0069] While several aspects and embodiments refer to road segments, the present aspects and embodiments may as well be applied to other kinds of way segments, for example, bikeways segments, pedestrian ways segments, shared use intersections.

[0070] Shared use may be, e.g., selected from at least two of: bikeways intersections, pedestrian ways intersections, motorized vehicles roads intersections. Thus, the term 'road' as used herein, and in accordance with various embodiments, may also mean a way or a path.

CLAIMS

1. A method (400) for generating map data, the method (400) comprising:
generating (402) of a plurality of cells of a real space area, each cell representing a two-dimensional sub-space of the real space area, wherein each of the cells comprise at least one global positioning system (GPS) point, and wherein the real space area comprises a road network;
determining (404) a number of GPS points for each cell of the plurality of cells;
determining (406) a road network based on the numbers of GPS points using a Smoothed Particle Hydrodynamics (SPH) analysis, wherein the cells of the plurality of cells are particles of the SPH analysis and the number of GPS points per cell of the plurality of cells are a scalar of each cell of the plurality of cells in the SPH analysis; and
comparing (408) the road network determined using the SPH analysis with the road network of the real space area.
2. The method (400) of claim 1, further comprising:
determining a first plurality of cells, wherein the first plurality of cells includes the plurality of cells and at least one further cell, wherein the at least one further cell comprises no GPS point.
3. The method (400) of claim 1 or 2, further comprising:
providing GPS points from a plurality of mobile phone users in the real space area.
4. The method (400) of any one claims 1 to 3,
wherein the cells of the plurality of cells comprise are spaces of a grid,
wherein each cell comprises the same size.
5. The method (400) of claim 4,
wherein each cell is a Geohash.
6. The method (400) of any one of claims 1 to 5,
wherein each of the GPS point of a cell has the same weight as the other GPS points of the same cell.

7. The method (400) of any one of claims 1 to 6,
wherein determining (406) the road network comprises a reconstructing of a continuous field from the plurality of cells, wherein the continuous field is interpreted as a drivable area field.
8. The method (400) of claim 7,
wherein a high value of the continuous field in a certain point of the real space area is interpreted as a high probability of having a road in the point.
9. The method (400) of any one of claims 1 to 8,
wherein determining (406) the road network comprises a number of GPS points per cell interpreted as road existence probability density function.
10. The method (400) of any one of claims 1 to 9,
wherein the input to the SPH analysis represents an unprocessed heatmap built from the GPS points of each cell and the output is a normalized heatmap.
11. The method (400) of any one of claims 1 to 10, further comprising:
providing at least one visualization image for the distribution of number of GPS points per cell obtained after simulation.
12. The method (400) of claim 11, further comprising:
classifying each pixel from the resulting visualization image to determine whether the pixel is a part of a road or not.
13. The method (400) of claim 12, further comprising:
wherein the classifying is using a deep neural network model with the visualization image as an input into a semantic segmentation.
14. The method (400) of claim 12 or 13, further comprising:
providing a dataset by labelling all roads where GPS points exist.
15. The method (400) of claim 14,
wherein labels are obtained by querying an open street map (OSM) and retrieving all existing geometries, and generating a binary image in which every pixel has a first label associated to a road or a second label associated to a non-road area

by converting the geometries in the image by drawing a plurality of lines on an empty canvas.

16. The method (400) of claim 1,
wherein the road segments presented in the road network determined by the SPH analysis are considered as roads missing in the map of the real space area.
17. The method (400) of any one of claims 1 to 16, further comprising:
determining whether there are existing roads in a proximity of new road segments determined by SPH analysis.
18. A data processing apparatus configured to carry out the method (400) of any of the claims 1 to 17.
19. A non-transitory computer-readable medium storing computer executable code comprising instructions for generating map data according to the method (400) of any one of claims 1 to 17.
20. A computer executable code comprising instructions for generating map data according to the method (400) of any one of claims 1 to 17.