CLAIMS

1. A method for processing images associated with a location of interest, the method comprising the steps of:

receiving a plurality of image data files associated with a location of interest comprising at least one road segment, each image data file associated with at least a part of the location of interest;

for each image

determining a position and orientation of an image capturing device relative to the location of interest and generating device pose information;

generating a corresponding depth map; and

generating a corresponding semantic segmentation;

grouping the image, the device pose information, the depth map and the semantic segmentation based on coordinates of the location of interest to form an image group; and

generating a voxel grid for each image group,

wherein the grouping the image comprises utilising a map-matching service to match coordinates of the image to the at least one road segment, so that two or more related images are grouped according to a matched segment.

2. The method of claim 1, further comprising steps of:

generating a voxel grid state array comprising one or more states of each voxel in the voxel grid; and

updating the voxel grid state array;

wherein the step of updating comprises obtaining a two-dimensional pixel and determining if the two-dimensional pixel is out of an image border specified by an image resolution.

3. The method of claim 2, further comprising a step of determining whether each voxel in the voxel grid is in a dynamically occluded state, wherein the step of determining whether each voxel in the voxel grid is in a dynamically occluded state includes selecting a state from a set of states comprising the following states: unseen, dynamically occluded, void, and occupied.

4. The method of claim 2 or 3, wherein the voxel grid state array is a one-dimensional array.

5. The method of claim 3 or 4, further comprising a step of initializing the voxel grid state array.

6. The method of claim 5, wherein the step of initializing the voxel grid state array comprises setting the state of every voxel to the unseen state.

7. The method of claim 6, further comprising a step of reprojecting the voxel onto a two-dimensional image plane based on the device pose information, and obtaining an associated two-dimensional pixel.

8. The method of claim 3, further comprising a step of assigning the dynamically occluded state to the voxel if the associated pixel is determined to be within the image border.

9. The method of claim 8, further comprising a step of comparing a first parameter d_v representing a depth of the voxel point with respect to the image capturing device, with a second parameter d_p representing the depth of the reprojected pixel, wherein if d_v is less than or equal to d_p, the voxel is assigned the void state.

10. The method of claim 9, further comprising a step of checking if the segmentation label of the reprojected pixel is a dynamic object, and if not, the voxel will be assigned the occupied state.

11. The method of any one of the preceding claims, wherein the step of grouping comprises matching the location of interest with at least one feature on a reference map.

12. The method of claim 11, wherein the step of generating a voxel grid comprises determining a length, a width and a height of the voxel grid based on the at least one feature on the reference map.

13. The method of any one of the preceding claims, wherein the step of generating the corresponding depth map of the image comprises using a trained deep learning model or a structure-from-motion (SfM) algorithm to estimate the depth map using the image as the only input.

14. The method of any one of the preceding claims, wherein the step of generating a corresponding semantic segmentation of the image comprises using a trained convolutional neural network model to generate semantic labels associated with one or more features on the image.

15. A device for processing images associated with a location of interest, the device comprising

an input module configured to receive a plurality of image data files associated with a location of interest comprising at least one road segment;

a device pose module configured to determine a position and orientation of an image capturing device relative to the location of interest and generating device pose information;

a depth map generation module configured to generate a corresponding depth map;

a segmentation module configured to generate a corresponding semantic segmentation;

an image aggregator module configured to group the image, the device pose information, the depth map and the semantic segmentation based on coordinates of the location of interest to form an image group; and

a voxel grid state estimator configured to generate a voxel grid for each image group,

wherein a map-matching service is utilised to match coordinates of the image to the at least one road segment so that two or more related images are grouped according to a matched segment.

16. A system for updating a voxel grid state array comprising:

a device for processing images associated with a location of interest, the device comprising:

an input module configured to receive a plurality of image data files associated with a location of interest comprising at least one road segment;

a device pose module configured to determine a position and orientation of an image capturing device relative to the location of interest and generating device pose information;

a depth map generation module configured to generate a corresponding depth map;

a segmentation module configured to generate a corresponding semantic segmentation;

an image aggregator module configured to group the image, the device pose information, the depth map and the semantic segmentation based on coordinates of the location of interest to form an image group; and

a voxel grid state estimator configured to generate a voxel grid for each image group, and generate a voxel grid state array comprising one or more states of each voxel in the voxel grid,

wherein a map-matching service is utilised to match coordinates of the image to the at least one road segment so that two or more related images are grouped according to a matched segment, and

wherein the system further comprises an updater to check if the voxel is previously detected to be in a dynamic occluded state and subsequently in a void state or occupied state.