CLAIMS

1. A method for detecting dynamic occlusion on one or more 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 the position and orientation of an image capturing device relative to the location of interest and generating device pose information;

generating a corresponding depth map;

generating a corresponding semantic segmentation;

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

generating a voxel grid associated with the image group; and

determining whether each voxel in the voxel grid is in a dynamically occluded state.

2. The method of claim 1, 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.

3. The method of claim 1 or 2, further comprising the step of generating a voxel grid state array comprising the states of the each of the voxel in the voxel grid.

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

5. The method of claim 3 or 4, wherein at an initialization of the voxel grid state array, the state of every voxel is set to an unseen state.

6. The method of claim 5, further comprising the step of reprojecting the voxel onto a two-dimensional image plane based on the ~~camera~~ device pose information, and obtaining an associated two-dimensional pixel.

7. The method of claim 6, further comprising a step of determining if the pixel is out of an image border specified by the image resolution, and assigning the dynamic occluded state to the voxel if the associated pixel is determined to be within the image border.

8. The method of claim 7, further comprises 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.

9. The method of claim 8, further comprises checking if the segmentation label of the reprojected pixel is a dynamic object, and if not, the voxel will be assigned the occupied state.

10. 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.

11. The method of claim 10, 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.

12. 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.

13. 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.

14. A device for detecting dynamic occlusion on one or more 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 the 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, ~~camera~~ device pose information, depth map and semantic segmentation based on coordinates of the location of interest to form an image group;

a voxel grid state estimator configured to generate a voxel grid associated with the image group; and determine whether each voxel in the voxel grid is in a dynamically occluded state.

15. The device of claim 14, wherein determination of 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: unseen state, dynamically occluded state, void state, and occupied state.

16. The device of claim 14 or 15, wherein the voxel grid state estimator is further configured to generate a voxel grid state array comprising the states of the each of the voxel in the voxel grid.

17. The device of claim 16, wherein the voxel grid state array is a one-dimensional array.

18. A system for updating a voxel grid state array comprising the device of claim 16 or 17, further comprising 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.

19. The system of claim 18, wherein if the voxel is detected to be in a void state or occupied state, the system is configured to update the voxel grid state array associated with the change of state(s).

20. A non-transitory computer-readable storage medium comprising instructions, which, when executed by one or more processors, cause the execution of the method according to any one of claims 1-13.