CLAIMS

[Claim 1] A method for point of interest (POI) detection clustering using image embeddings, the method comprising:

obtaining, using a neural network, the image embeddings from POI detections comprising portions of one or more images that are captured in an area of a map, in which at least one POI is detected;

calculating an embedding distance between the obtained image embeddings;

recognizing texts included in the POI detections;

calculating an optical character recognition (OCR) distance between the recognized texts;

calculating a final distance between the POI detections, by:

determining whether the calculated OCR distance is less than a first predetermined distance,

determining whether the calculated embedding distance is less than a second predetermined distance,

based on the calculated OCR distance being determined to be less than the first predetermined distance and the calculated embedding distance being determined to be less than the second predetermined distance, setting the final distance to be zero, and

based on the calculated OCR distance being determined to be greater than or equal to the first predetermined distance and/or the calculated embedding distance being determined to be greater than or equal to the second predetermined distance, setting the final distance to be the calculated embedding distance; and

clustering the POI detections, based on the calculated final distance.

[Claim 2] The method of claim 1, further comprising calculating a coordinate distance between geographic coordinate pairs of a respective pair of the POI detections.

[Claim 3] The method of claim 2, further comprising determining whether the calculated coordinate distance is greater than or equal to a predetermined threshold distance.

[Claim 4] The method of claim 3, further comprising, based on the coordinate distance being determined to be greater than or equal to the predetermined threshold distance, filtering the POI detections by removing the respective pair of the POI detections from the POI detections.

[Claim 5] The method of claim 4, wherein the obtaining the image embeddings comprises obtaining the image embeddings from the filtered POI detections.

[Claim 6] The method of any one of claims 4 and 5, wherein the recognizing the texts comprises recognizing the texts included in the filtered POI detections.

[Claim 7] The method of any one of claims 4 to 6, wherein the clustering the POI detections comprises clustering the filtered POI detections, based on the calculated final distance between the filtered POI detections.

[Claim 8] The method of any one of claims 1 to 7, wherein the embedding distance is a cosine distance between the obtained image embeddings, and is calculated based on a following equation:

$\text{similarity} = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$, and

A and **B** denote the obtained image embeddings.

[Claim 9] The method of any one of claims 1 to 8, wherein the calculating the OCR distance comprises removing common words from a pair of the recognized texts.

[Claim 10] The method of claim 9, wherein the calculating the OCR distance further comprises calculating the OCR distance between an ordering of words in both of the pair of the texts from which the common words are removed.

[Claim 11] The method of any one of claims 1 to 10, wherein the clustering the POI detections comprises determining whether the calculated final distance is less than a predefined threshold distance.

[Claim 12] The method of claim 11, wherein the clustering the POI detections further comprises, based on the calculated final distance being determined to be less than the predefined threshold distance, clustering the POI detections.

[Claim 13] The method of any one of claims 1 to 12, further comprising:

updating the map to comprise a POI corresponding to the clustered POI detections; and

providing the updated map to a user and/or a controller configured to generate control signals for a vehicle based on the updated map.

[Claim 14] A server configured to perform the method of any one of claims 1 to 13.

[Claim 15] A computer program element comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 13.

[Claim 16] A computer-readable medium comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 13.