Claims

1. A computer assisted method comprising:

storing a training dataset including a plurality of geotagged candidate images and a plurality of query images, each query image having at least one corresponding candidate image having the same geolocation;

applying a quasi-random or random azimuth rotation to each of the plurality of query images, and storing the azimuth rotation for each of the plurality of rotated query images;

training a machine learning model, including:

extracting features from the plurality of rotated query images;

estimating the azimuth rotation of the rotated query image based on an inference of the extracted features of the rotated query image and extracted features from the candidate images, and

using an objective function including a first loss function based on a weighted soft-margin triplet loss, and a second loss function based on an error loss between the stored azimuth rotation and the estimated azimuth rotation for the stored dataset.

2. The method of claim 1 further comprises cropping each of the plurality of rotated query images to a restricted field of view.

3. The method of claim 1 or 2 wherein the training a machine learning model further comprising ranking the correlation of the extracted features of the plurality of candidate images to the extracted features of the rotated query image, selecting the highest ranked candidate image, and estimating the geolocation of the query image by the geolocation of the highest ranked candidate image.

4. The method of claim 3 wherein the training a machine learning model further comprising adjusting the orientation of the plurality of candidate images based on

the estimated azimuth rotation of the query image and/or cropping the field of view of the plurality of candidate images depending on the field of view of the query image.

5. The method of claim 3 or 4 further comprising storing an approximate geolocation for one or more of the plurality of query images, and selecting a subset of the candidate images based on proximity to the approximate geolocation of the query image to correlate against the query image.

6. The method of any preceding claim wherein the objective function is defined as $L = L_{\text{match}}(A, P, N) + \beta L_{\text{angle}}(A, P)$,

7. The method of claim 6 wherein the first loss function is defined as

$$L_{\text{match}} = \ln(1 + \exp(\alpha (D(F_A, F'_P) - D(F_A, F'_N))))$$

$$D(F_1, F_2) = 2(1 - \cos(F_1, F_2))$$

8. The method of claim 6 or 7 wherein the second loss function is defined as $L_{\text{angle}} = (0.5W - |w_{\text{gt}} - w_{\text{est}}| - 0.5W)/0.5W$

9. The method of claim 8 wherein the error loss is determined using an absolute angle error.

10. The method of claim 9 wherein the absolute angle error is calculated using $\theta_{\text{err}} = 180^{\text{O}} - ||\theta_{\text{gt}} - \theta_{\text{est}}| - 180^{\text{O}}|$.

11. The method of any of claims 6 to 10 wherein β is between 0.1 to 0.5 or is substantially similar to 0.3.

12. The method of any preceding claim further comprising applying one or more metrics to the machine learning model selected from the group consisting of a fine-

grained histogram, a mean angle error, an accuracy below specific threshold and any combination thereof.

13. The method of claim 12 wherein the fine-grained histogram is calculated using

$$\theta_{\text{err}} = 180^\circ - \left| |\theta_{\text{gt}} - \theta_{\text{est}}| - 180^\circ \right|,$$
$$H(\theta_i) = H(\theta_i) + 1, \quad \text{where } \theta_i - 1 \leq \theta_{\text{err}} < \theta_i.$$

14. The method of claim 12 wherein the accuracy below specific threshold is calculated using

$$\text{r@x}^\circ = \sum_{i=1}^{x} H(\theta_i) / \sum_{i=1}^{180} H(\theta_i).$$

15. The method of any preceding claim further comprising applying a polar transform to each of the plurality of candidate images.

16. The method of any preceding claim wherein the applying a random azimuth rotation comprises cropping a portion of one side of the image and appending it to the other side of the image.

17. The method of any preceding claim wherein the training dataset is based on a south aligned coordinate system, the plurality of query images corresponding to street-view images and the plurality of candidate images corresponding to aerial images.

18. The method of any preceding claim wherein the training a machine learning model further comprising interpolating the extracted features of the rotated query image and extracted features from the candidate images by a scaling factor, and correlating the interpolated extracted features of the rotated query image and the interpolated extracted features from the plurality of candidate images using the first loss function.

19. The method of any of claims 1 to 17 wherein the training a machine learning model further comprising correlating the interpolated extracted features of the rotated query image and the interpolated extracted features from the plurality of candidate images using the first loss function, and smoothing a curve associated with the correlation using a scaling factor.

20. The method of claim 19 wherein the smoothing the curve comprises:

Fast Fourier Transforming (FFT) the correlation curve to the frequency domain;

zero-padding of predetermined number of times to the middle of the transformed curve; and

Inverse Fast Fourier Transforming (FFT) the zero padded curve.

21. A method comprising using a trained machine learning model in an inference phase, wherein the machine learning model was trained using the method of any of claims 1 to 20.

22. A system comprising

a communication server;

at least one mobile communication device; and

communication network equipment configured to establish communication with the communications server, and the at least one mobile communication device;

wherein the mobile communication device comprises a first processor and a first memory, the mobile communications device being configured, under control of the first processor, to execute first instructions stored in the first memory to:

capture a query image;

transmit the query image to the communication server;

and wherein the communication server comprises a second processor and a second memory, the communication server being configured, under control of the second processor, to execute second instructions stored in the second memory to:

operate in an inference phase, using a machine learning model trained according to any of claims 1 to 20, to estimate the azimuth rotation and/or geolocation of the query image.

23. A mobile communication device according to the at least one mobile communication device in claim 22.