

CLAIMS:

1. A method for image-based localization, comprising:
at a camera device, capturing a plurality of images of a real-world environment;
detecting a first set of image features in a first image of the plurality of images;
before additional sets of image features are detected in additional images of the plurality of images, transmitting the first set of image features to a remote device configured to estimate a pose of the camera device;
receiving, from the remote device, a first estimated pose of the camera device, the first estimated pose having been estimated based on the first set of image features detected in the first image of the plurality of images;
as the additional sets of image features are detected in the additional images of the plurality of images, transmitting the additional sets of image features to the remote device; and
receiving, from the remote device, a second estimated pose of the camera device, the second estimated pose having been estimated based on multiple sets of image features detected in two or more images of the plurality of images.
2. The method of claim 1, where the plurality of images of the real-world environment are captured substantially at once.
3. The method of claim 2, where each image of the plurality of images is captured by a different respective camera of the camera device, and the method further comprises transmitting, to the remote device, information indicating a spatial relationship of each of the different respective cameras of the camera device relative to one another.
4. The method of claim 1, where the plurality of images of the real-world environment are captured sequentially.
5. The method of claim 4, further comprising transmitting, to the remote device, information regarding movements of the camera device between capture of sequential images of the plurality of images.
6. The method of claim 1, where the first estimated pose of the camera device is estimated based on identified correspondences between image features transmitted to the remote device and map features in a digital environment map corresponding to the real-world environment.
7. The method of claim 6, where each detected image feature includes an indication of a two-dimensional (2D) pixel position at which the image feature was identified in a corresponding image of the real-world environment, and where the identified correspondences are between 2D pixel positions of image features transmitted to the remote device and three-dimensional (3D) map positions of map features in the digital environment map.

8. The method of claim 7, where the identified correspondences between the 2D pixel positions and 3D map positions are input into a random sample consensus (RANSAC) solver that outputs the estimated pose of the camera device.

9. The method of claim 1, where the first estimated pose of the camera device is a preliminary estimated pose and is received before the additional sets of image features are detected for every image of the plurality of images, and the preliminary estimated pose has an associated confidence value proportional to a quantity of image features transmitted to the remote device before the preliminary estimated pose was estimated.

10. The method of claim 9, where the second estimated pose is estimated based on more detected image features than the preliminary estimated pose and has a correspondingly higher confidence value.

11. The method of claim 10, further comprising receiving a final estimated pose having a corresponding confidence value that exceeds a final confidence threshold, and after receiving the final estimated pose, discontinuing transmitting additional sets of image features to the remote device.

12. A camera device, comprising:
one or more cameras; and
a logic machine, configured to:

- capture a plurality of images of a real-world environment via the one or more cameras;

- detect a first set of image features in a first image of the plurality of images;

- before additional sets of image features are detected in additional images of the plurality of images, transmit the first set of image features to a remote device configured to estimate a pose of the camera device;

- receive, from the remote device, a preliminary estimated pose of the camera device, the preliminary estimated pose having been estimated based on the first set of image features detected in the first image of the plurality of images, the preliminary estimated pose received before the additional sets of image features are detected for every image of the plurality of images, and the preliminary estimated pose having an associated confidence value that is proportional to a quantity of image features transmitted to the remote device before the preliminary estimated pose was estimated;

- as the additional sets of image features are detected in the additional images of the plurality of images, transmit the additional sets of image features to the remote device; and

- receive, from the remote device, a second estimated pose of the camera device, the second estimated pose having been estimated based on more detected image features than the preliminary estimated pose, and the second estimated pose having a correspondingly higher confidence value.

13. The camera device of claim 12, where the plurality of images of the real-world environment are captured substantially at once by different respective cameras of the camera device, and the logic machine is further configured to transmit, to the remote device, information indicating a spatial relationship of each of the different respective cameras of the camera device relative to one another.
14. The camera device of claim 12, where the plurality of images of the real-world environment are captured sequentially, and the logic machine is further configured to transmit, to the remote device, information regarding movements of the camera device between capture of sequential images of the plurality of images.
15. The camera device of claim 12, where the first estimated pose of the camera device is estimated based on identified correspondences between image features transmitted to the remote device and map features in a digital environment map corresponding to the real-world environment.
16. The camera device of claim 15, where each detected image feature includes an indication of a two-dimensional (2D) pixel position at which the image feature was identified in a corresponding image of the real-world environment, and where the identified correspondences are between 2D pixel positions of image features transmitted to the remote device and three-dimensional (3D) map positions of map features in the digital environment map.
17. The camera device of claim 12, where the logic machine is further configured to receive a final estimated pose having a corresponding confidence value that exceeds a final confidence threshold, and after receiving the final estimated pose, discontinue transmitting additional sets of image features to the remote device.
18. A computing device, comprising:
a logic machine configured to:
 receive, from a camera device, a plurality of images of a real-world environment of the camera device captured by one or more cameras of the camera device;
 as the plurality of images of the real-world environment are received, detect sets of image features in each image of the plurality of images;
 as the sets of image features are detected, identify correspondences between detected image features and map features of a digital environment map corresponding to the real-world environment;
 as the correspondences are identified, progressively output a plurality of estimated poses of the camera device based on the identified correspondences, where each estimated pose has a corresponding confidence value proportional to a

quantity of the correspondences identified when each respective estimated pose was output;

upon estimating a preliminary pose of the camera device having a corresponding confidence value that exceeds a preliminary confidence threshold, and before detecting the sets of image features for every image of the plurality of images, report the first pose to the camera device; and

upon estimating a second pose of the camera device having a corresponding confidence value that exceeds a final confidence threshold higher than the preliminary confidence threshold, report the second pose to the camera device.