CLAIMS

1. A method for estimating a size of a target object in an image, the method comprising

extracting at least one reference object and the target object from the image;

generating a depth map of the image;

estimating a distance of the target object from an image-capturing device used to capture the image based on the depth map;

calculating the pixel size of each of the at least one reference object and the target object based on the extracted at least one reference object and the target object;

estimating a focal length of the image-capturing device based on the calculated size of each of the at least one reference object; and

estimating the size of the target object based on the focal length, the distance, and the pixel size of each of the at least one reference object and the target object.

2. The method according to claim 1, wherein generating the depth map of the image comprises using at least one deep learning model.

3. The method according to claim 2, wherein the at least one deep learning model comprises a monocular depth estimation model.

4. The method according to claim 3, wherein the monocular depth estimation model comprises at least one of Monodepth2, LeReS, and/or a high-resolution depth estimation model.

5. The method according to any one of claims 1 to 4, wherein extracting the at least one reference object and the target object includes using an object detection and image segmentation model.

6. The method according to claim 5, wherein the object detection and image segmentation model includes a mask-RCNN model.

7. The method according to claim 6, wherein the mask-RCNN model is configured to generate an output, the output comprises a first bounding box parameterizing each of the at least one reference object and a second bounding box parameterizing each of the target object.

8. The method according to claim 7, wherein the distance of the target object from an image-capturing device is estimated based on obtaining a center of the bounding box, or a median value of the depth information inside the bounding box.

9. The method according to any one of the preceding claims, wherein estimating the focal length of the image-capturing device includes assigning a known height to each of the at least one reference object.

10. The method according to claim 2 or 3, wherein the deep learning models are trained using at least one of a supervised learning model, an unsupervised learning model, and/or a semi-supervised learning model.

11. The method according to any one of the preceding claims, wherein estimating the size of the target object includes using a right triangle congruence theorem to determine the focal length, the distance, and the pixel size of each of the at least one reference object and the target object.

12. A system for estimating a size of a target object in an image, the system comprising at least one processor, the at least one processor configured to:

extract at least one reference object and the target object from the image;

generate a depth map of the image;

estimate a distance of the target object from an image-capturing device used to capture the image based on the depth map;

calculate the pixel size of each of the at least one reference object and the target object based on the extracted at least one reference object and the target object;

estimate a focal length of the image-capturing device based on the calculated size of each of the at least one reference object; and

estimate the size of the target object based on the focal length, the distance, and the pixel size of each of the at least one reference object and the target object.

13. The system according to claim 12, wherein the depth map is generated using at least one deep learning model.

14. The system according to claim 13, wherein the at least one deep learning model comprises a monocular depth estimation model.

15. The system according to claim 14, wherein the monocular depth estimation model comprises at least one of Monodepth2, LeReS, and/or a high-resolution depth estimation model.

16. The system according to any one of claims 12 to 15, wherein the extraction of the at least one reference object and the target object includes using an object detection and image segmentation model.

17. The system according to claim 16, wherein the object detection and image segmentation model includes a mask-RCNN model.

18. The system according to claim 17, wherein the mask-RCNN model is configured to generate an output, the output comprises a first bounding box parameterizing each of the at least one reference object and a second bounding box parameterizing each of the target object.

19. The system according to any one of claims 12 to 18, wherein the at least one processor is configured to assign a known height to each of the at least one reference object in the estimation of the focal length of the image-capturing device.

20. An event detection system comprising

an image-capturing device, the image-capturing device configured to capture an image of the event;

at least one processor configured to:

extract at least one reference object and a target object from the at least one image;

generate a depth map of the image;

estimate a distance of the target object from an image-capturing device used to capture the image based on the depth map;

calculate the pixel size of each of the at least one reference object and the target object based on the extracted at least one reference object and the target object;

estimate a focal length of the image-capturing device based on the calculated size of each of the at least one reference object; and

estimate the size of the target object based on the focal length, the distance, and the pixel size of each of the at least one reference object and the target object.

21. The event detection system of claim 20, wherein the image-capturing device comprises a surveillance camera.

22. The event detection system of claim 20 or 21, the event detection system further comprises a notification module configured to receive the estimated size of the target object and notify one or more first responders to the event.

23. A computer program, the computer program comprising instructions to execute the method according to any one of claims 1 to 11.