CLAIMS

1. A method comprising:

obtaining access to a three-dimensional point cloud representation of an object including poses of a scanning digital camera and corresponding video frames;

down-sampling said three-dimensional point cloud representation to obtain a set of region-of-interest candidates;

filtering said region-of-interest candidates to select those of said region-of-interest candidates having appearance changes, which distinguish different visual states, as selected regions of interest, based at least in part on said poses of said camera;

generating region of interest images for said selected regions of interest from corresponding ones of said video frames; and

training a deep learning recognition model based on said region of interest images.

2. The method of Claim 1, further comprising determining a visual state of said object with said digital camera coupled to a processor, using said trained deep learning recognition model.

3. The method of Claim 2, further comprising providing augmented reality instructions, related to said object, to a user, based on said determined state.

4. The method of Claim 3, further comprising re-localizing said digital camera with respect to said object prior to determining said state of said object.

5. The method of Claim 4, wherein said down-sampling of said three-dimensional point cloud representation to obtain said set of region-of-interest candidates includes removing outliers.

6. The method of Claim 5, wherein said filtering of said region-of-interest candidates to select those of said region-of-interest candidates having said appearance changes comprises image cropping and zooming.

7. The method of ~~any of claims~~claim 4 ~~to 6~~, wherein said determining of said state of said object comprises:

capturing images for those of said regions of interest having occupied voxels of interest;

providing a corresponding zoomed local image and viewing pose for each of said regions of interest having said occupied voxels to said recognition model.

8. The method of Claim 7, wherein:

said model comprises a multi-stream convolutional neural network long short-term memory model; and

said determining of said state of said object comprises:

separating said regions of interest into a plurality of branches;

extracting features from said corresponding zoomed local images using a convolutional neural network of said multi-stream convolutional neural network long short-term memory model;

aggregating spatial information by concatenating said extracted features;

aggregating temporal information by pooling said spatial information for a plurality of time steps and providing said pooled information to a long short-term memory layer to obtain a fully connected layer for classification to determine said state of said object.

9. The method of ~~either of claims~~claim 7 ~~or 8~~, further comprising refraining from capturing images for those of said regions of interest not having occupied voxels of interest, to reduce sensitivity to at least one of occlusions and camera instability.

10. The method of ~~any of claims~~claim 3 ~~to 9~~, wherein said object comprises an information technology asset to be repaired, said digital camera and said processor are part of a mobile device, and said providing of said augmented reality instructions comprises illustrating a repair procedure for said information technology asset.

~~11.~~ 11. A non-transitory computer readable medium comprising computer executable instructions which when executed by a computer cause the computer to perform a method of:

obtaining~~An apparatus comprising:~~

~~a memory; and~~

~~at least one processor, coupled to said memory, and operative to:~~

~~obtain~~ access to a three-dimensional point cloud representation of an object including poses of a scanning digital camera and corresponding video frames;

down-sampling said three-dimensional point cloud representation to obtain a set of region-of-interest candidates;

filtering said region-of-interest candidates to select those of said region-of-interest candidates having appearance changes, which distinguish different visual states, as selected regions of interest, based at least in part on said poses of said camera;

generating region of interest images for said selected regions of interest from corresponding ones of said video frames; and

training a deep learning recognition model based on said region of interest images.

12. The non-transitory computer readable medium of claim 11, wherein the method further comprises determining a visual state of said object with said digital camera coupled to a processor, using said trained deep learning recognition model.

13. The non-transitory computer readable medium of claim 12, wherein the method further comprises providing augmented reality instructions, related to said object, to a user, based on said determined state.

14. The non-transitory computer readable medium of claim 13, wherein the method further comprises re-localizing said digital camera with respect to said object prior to determining said state of said object.

15. The non-transitory computer readable medium of claim 14, wherein:

said down-sampling of said three-dimensional point cloud representation to obtain said set of region-of-interest candidates includes removing outliers; and

said filtering of said region-of-interest candidates to select those of said region-of-interest candidates having said appearance changes comprises image cropping and zooming.

16. The non-transitory computer readable medium of claim 14, wherein:

said determining of said state of said object comprises:

capturing images for those of said regions of interest having occupied voxels of interest;

providing a corresponding zoomed local image and viewing pose for each of said regions of interest having said occupied voxels to said recognition model;

said model comprises a multi-stream convolutional neural network long short-term memory model; and

said determining of said state of said object further comprises:

separating said regions of interest into a plurality of branches;

extracting features from said corresponding zoomed local images using a convolutional neural network of said multi-stream convolutional neural network long short-term memory model;

aggregating spatial information by concatenating said extracted features; and

aggregating temporal information by pooling said spatial information for a plurality of time steps and providing said pooled information to a long short-term memory layer to obtain a fully connected layer for classification to determine said state of said object.

17. The non-transitory computer readable medium of claim 16, wherein the method further comprises refraining from capturing images for those of said regions of interest not having occupied voxels of interest, to reduce sensitivity to at least one of occlusions and camera instability.

18. The non-transitory computer readable medium of claim 13, wherein said object comprises an information technology asset to be repaired, and said providing of said augmented reality instructions comprises illustrating a repair procedure for said information technology asset.

19. An apparatus comprising:

a memory; and

at least one processor, coupled to said memory, and operative to:

obtain access to a three-dimensional point cloud representation of an object including poses of a scanning digital camera and corresponding video frames;

down-sample said three-dimensional point cloud representation to obtain a set of region-of-interest candidates;

~~——~~filter said region-of-interest candidates to select those of said region-of-interest candidates having appearance changes, which distinguish different visual states, as selected regions of interest, based at least in part on said poses of said camera;

~~——~~generate region of interest images for said selected regions of interest from corresponding ones of said video frames; and

~~——~~train a deep learning recognition model based on said region of interest images.

~~12.~~ 20. The apparatus of ~~Claim 11~~claim 19, wherein said at least one processor is further operative to determine a visual state of said object with said digital camera, said digital camera being coupled to said processor, using said trained deep learning recognition model.

~~13.~~ 21. The apparatus of ~~Claim 12~~claim 20, wherein said at least one processor is further operative to provide augmented reality instructions, related to said object, to a user, based on said determined state.

~~14.~~ 22. The apparatus of ~~Claim 13~~claim 21, wherein said at least one processor is further operative to re-localize said digital camera with respect to said object prior to determining said state of said object.

~~15.~~ 23. The apparatus of ~~Claim 14~~claim 22, wherein:

~~-~~

said down-sampling of said three-dimensional point cloud representation to obtain said set of region-of-interest candidates includes removing outliers;

~~-~~said filtering of said region-of-interest candidates to select those of said region-of-interest candidates having said appearance changes comprises image cropping and zooming; and

~~-~~said determining of said state of said object comprises:

~~capturin images for those of said regions of interest having occupied voxels of interest;~~

~~providing a corresponding zoomed local image and viewing pose for each of said regions of interest having said occupied voxels to said recognition model.~~

~~16. The apparatus of any of claims 14 to 15, wherein said determining of said state of said object comprises:~~

~~-~~capturing images for those of said regions of interest having occupied voxels of interest;

~~-~~providing a corresponding zoomed local image and viewing pose for each of said regions of interest having said occupied voxels to said recognition model.

~~17.~~ 24. The apparatus of ~~Claim 16~~claim 23, wherein:

~~-~~said model comprises a multi-stream convolutional neural network long short-term memory model; and

~~-~~said determining of said state of said object comprises:

~~-~~separating said regions of interest into a plurality of branches;

~~-~~extracting features from said corresponding zoomed local images using a convolutional neural network of said multi-stream convolutional neural network long short-term memory model;

extracting features from said corresponding zoomed local images using a convolutional neural network of said multi-stream convolutional neural network long short-term memory model;

aggregating spatial information by concatenating said extracted features; and

aggregating temporal information by pooling said spatial information for a plurality of time steps and providing said pooled information to a long short-term memory layer to obtain a fully connected layer for classification to determine said state of said object; and

said at least one processor is further operative to refrain from capturing images for those of said regions of interest not having occupied voxels of interest, to reduce sensitivity to at least one of occlusions and camera instability.

25. The apparatus of claim 21, wherein said object comprises an information technology asset to be repaired, said digital camera and said processor are part of a mobile device, and said providing of said augmented reality instructions comprises illustrating a repair procedure for said information technology asset.