CLAIMS

What is claimed is:

1. A method for object re-identification, the method comprising:

obtaining a first set of images from a first camera, and a second set of images from at least one second camera, wherein the at least one second camera is different from the first camera;

determining a first set of features based on the first set of images, the first set of features lying in a first feature space;

determining a second set of features based on the second set of images, the second set of features lying in a second feature space;

determining a first feature projection matrix and a second feature projection matrix that respectively map the first set of features and the second set of features to a shared feature space; and

determining a common dictionary based on the shared feature space, wherein the first set of features mapped to the shared feature space and the second set of features mapped to the shared feature space are represented by entries of the common dictionary,

wherein determining the first set of features based on the first set of images comprises:

generating first-level representations for the first set of images, wherein the first-level representations for the first set of images comprise image-level representations for the first set of images;

generating second-level representations for the first set of images, wherein the second-level representations for the first set of images comprise at least one of part-level representations for the first set of images or patch-level representations for the first set of images;

determining first-level features for the first set of images based on the first-level representations, wherein the first-level features for the first set of images comprise image-level features for the first set of images; and

determining second-level features for the first set of images based on the second-level representations, wherein the second-level features for the first set of images comprise at least one of part-level features for the first set of images or patch-

level features for the first set of images, the first set of features comprising the first-level features for the first set of images and the second-level features for the first set of images, and

wherein determining the second set of features based on the second set of images comprises:

generating first-level representations for the second set of images, wherein the first-level representations for the second set of images comprise image-level representations for the second set of images;

generating second-level representations for the second set of images, wherein the second-level representations for the second set of images comprise at least one of part-level representations for the second set of images or patch-level representations for the second set of images;

determining first-level features for the second set of images based on the first-level representations, wherein the first-level features for the second set of images comprise image-level features for the second set of images; and

determining second-level features for the second set of images based on the second-level representations, wherein the second-level features for the second set of images comprise at least one of part-level features for the second set of images or patch-level features for the second set of images, the second set of features comprising the first-level features for the second set of images and the second-level features for the second set of images.

2. The method of claim 1, wherein the second feature space is different from the first feature space.

3. The method of claim 1, wherein the first set of features mapped to the shared feature space and the second set of features mapped to the shared feature space are represented by respective linear combinations of the entries of the common dictionary.

4. The method of claim 1, wherein determining the first set of features based on the first set of images and determining the second set of features based on the second set of images comprises determining at least one of color histograms, dense scale-invariant

feature transform (SIFT) features, attributes, salient color features, mid-level feature maps, or polynomial kernel feature maps.

5. The method of any one of claims 1 to 4, further comprising:

obtaining a probe image from the first camera;

obtaining a plurality of gallery images from the at least one second camera;

determining a first sparse code corresponding to the probe image based on the common dictionary and the first feature projection matrix, the first sparse code representing features of the probe image in the shared feature space;

determining a plurality of second sparse codes corresponding to the plurality of gallery images based on the common dictionary and the second feature projection matrix, a respective second sparse code representing features of a respective gallery image in the shared feature space; and

determining a similarity between the probe image and each of the plurality of gallery images based on a comparison between the first sparse code and each of the plurality of second sparse codes.

6. The method of claim 5, further comprising generating a ranked list based on the comparison between the first sparse code and each of the plurality of second sparse codes, the ranked list comprising an indication of a level of similarity between the probe image and each of the plurality of gallery images.

7. The method of claim 5, wherein the comparison between the first sparse code and each of the plurality of second sparse codes is based on a Cosine similarity function.

8. The method of any one of claims 1 to 4, wherein a number of entries of the common dictionary is greater than or equal to 70.

9. A system for object re-identification, the system comprising:

a memory; and

at least one processor communicatively coupled to the memory and configured to perform operations comprising:

obtaining a first set of images from a first camera, and a second set of images from at least one second camera, wherein the at least one second camera is different from the first camera;

determining a first set of features based on the first set of images, the first set of features lying in a first feature space;

determining a second set of features based on the second set of images, the second set of features lying in a second feature space;

determining a first feature projection matrix and a second feature projection matrix that respectively map the first set of features and the second set of features to a shared feature space; and

determining a common dictionary based on the shared feature space, wherein the first set of features mapped to the shared feature space and the second set of features mapped to the shared feature space are represented by entries of the common dictionary,

wherein determining the first set of features based on the first set of images comprises:

generating first-level representations for the first set of images, wherein the first-level representations for the first set of images comprise image-level representations for the first set of images;

generating second-level representations for the first set of images, wherein the second-level representations for the first set of images comprise at least one of part-level representations for the first set of images or patch-level representations for the first set of images;

determining first-level features for the first set of images based on the first-level representations, wherein the first-level features for the first set of images comprise image-level features for the first set of images; and

determining second-level features for the first set of images based on the second-level representations, wherein the second-level features for the first set of images comprise at least one of part-level features for the first set of images or patch-level features for the first set of images, the first set of features comprising the first-level features for the first set of images and the second-level features for the first set of images, and

wherein determining the second set of features based on the second set of images comprises:

generating first-level representations for the second set of images, wherein the first-level representations for the second set of images comprise image-level representations for the second set of images;

generating second-level representations for the second set of images, wherein the second-level representations for the second set of images comprise at least one of part-level representations for the second set of images or patch-level representations for the second set of images;

determining first-level features for the second set of images based on the first level-representations, wherein the first-level features for the second set of images comprise image-level features for the second set of images; and

determining second-level features for the second set of images based on the second-level representations, wherein the second-level features for the second set of images comprise at least one of part-level features for the second set of images or patch-level features for the second set of images, the second set of features comprising the first-level features for the second set of images and the second-level features for the second set of images.

10. The system of claim 9, wherein the second feature space is different from the first feature space.

11. The system of claim 9, wherein the first set of features mapped to the shared feature space and the second set of features mapped to the shared feature space are represented by respective linear combinations of the entries of the common dictionary.

12. The system of claim 9, wherein determining the first set of features based on the first set of images and determining the second set of features based on the second set of images comprises determining at least one of color histograms, dense scale-invariant feature transform (SIFT) features, attributes, salient color features, mid-level feature maps, or polynomial kernel feature maps.

13. The system of any one of claims 9 to 12, the operations further comprising:

obtaining a probe image from the first camera;

obtaining a plurality of gallery images from the at least one second camera;

determining a first sparse code corresponding to the probe image based on the common dictionary and the first feature projection matrix, the first sparse code representing features of the probe image in the shared feature space;

determining a plurality of second sparse codes corresponding to the plurality of gallery images based on the common dictionary and the second feature projection matrix, a respective second sparse code representing features of a respective gallery image in the shared feature space; and

determining a similarity between the probe image and each of the plurality of gallery images based on a comparison between the first sparse code and each of the plurality of second sparse codes.

14. The system of claim 13, the operations further comprising generating a ranked list based on the comparison between the first sparse code and each of the plurality of second sparse codes, the ranked list comprising an indication of a level of similarity between the probe image and each of the plurality of gallery images.

15. The system of claim 13, wherein the comparison between the first sparse code and each of the plurality of second sparse codes is based on a Cosine similarity function.

16. The system of any one of claims 9 to 12, wherein a number of entries of the common dictionary is greater than or equal to 70.

17. A non-transitory computer-readable medium comprising instructions that are operable, when executed by a data processing apparatus, to perform operations comprising:

obtaining a first set of images from a first camera, and a second set of images from at least one second camera, wherein the at least one second camera is different from the first camera;

determining a first set of features based on the first set of images, the first set of features lying in a first feature space;

determining a second set of features based on the second set of images, the second set of features lying in a second feature space;

determining a first feature projection matrix and a second feature projection matrix that respectively map the first set of features and the second set of features to a shared feature space; and

determining a common dictionary based on the shared feature space, wherein the first set of features mapped to the shared feature space and the second set of features mapped to the shared feature space are represented by entries of the common dictionary,

wherein determining the first set of features based on the first set of images comprises:

generating first-level representations for the first set of images, wherein the first-level representations for the first set of images comprise image-level representations for the first set of images;

generating second-level representations for the first set of images, wherein the second-level representations for the first set of images comprise at least one of part-level representations for the first set of images or patch-level representations for the first set of images;

determining first-level features for the first set of images based on the first-level representations, wherein the first-level features for the first set of images comprise image-level features for the first set of images; and

determining second-level features for the first set of images based on the second-level representations, wherein the second-level features for the first set of images comprise at least one of part-level features for the first set of images or patch-level features for the first set of images, the first set of features comprising the first-level features for the first set of images and the second-level features for the first set of images, and

wherein determining the second set of features based on the second set of images comprises:

generating first-level representations for the second set of images, wherein the first-level representations for the second set of images comprise image-level representations for the second set of images;

generating second-level representations for the second set of images, wherein the second-level representations for the second set of images comprise at least one of part-level representations for the second set of images or patch-level representations for the second set of images;

determining first-level features for the second set of images based on the first-level representations, wherein the first-level features for the second set of images comprise image-level features for the second set of images; and

determining second-level features for the second set of images based on the second-level representations, wherein the second-level features for the second set of images comprise at least one of part-level features for the second set of images or patch-level features for the second set of images, the second set of features comprising the first-level features for the second set of images and the second-level features for the second set of images.

18. The non-transitory computer-readable medium of claim 17, the operations further comprising:

obtaining a probe image from the first camera;

obtaining a plurality of gallery images from the at least one second camera;

determining a first sparse code corresponding to the probe image based on the common dictionary and the first feature projection matrix, the first sparse code representing features of the probe image in the shared feature space;

determining a plurality of second sparse codes corresponding to the plurality of gallery images based on the common dictionary and the second feature projection matrix, a respective second sparse code representing features of a respective gallery image in the shared feature space; and

determining a similarity between the probe image and each of the plurality of gallery images based on a comparison between the first sparse code and each of the plurality of second sparse codes.

19. The non-transitory computer-readable medium of claim 18, the operations further comprising generating a ranked list based on the comparison between the first sparse code and each of the plurality of second sparse codes, the ranked list comprising

an indication of a level of similarity between the probe image and each of the plurality of gallery images.