Virtual View Networks for Object Reconstruction

João Carreira, Abhishek Kar, Shubham Tulsiani and Jitendra Malik
University of California, Berkeley

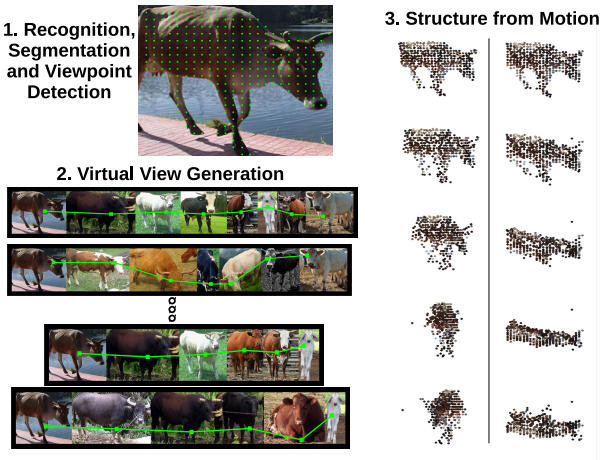

Figure 1: We reconstruct an object from a single image using factorization-based SfM on virtual views. These views are obtained by aligning a regular grid of points on the object (shown on top) with points on similar grids defined on objects in a reusable collection. Accurate alignment is achieved by computing geodesics on *virtual view networks* (VVN) which induce smooth rotations through the class object collection and simplify matching.

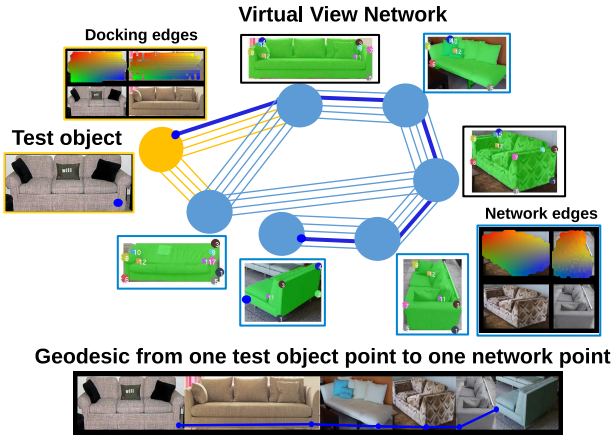

Figure 2: A test object is docked to the VVN by matching it to a few network objects with similar viewpoint and then it is aligned with all other objects based on geodesic distances in the network.

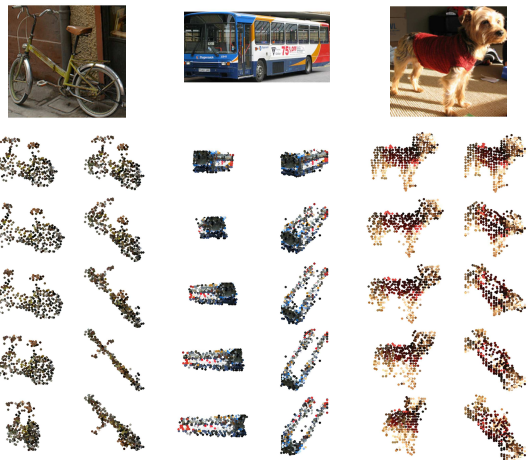

Figure 3: Example reconstructions produced by VVN on 3 PASCAL VOC categories from a single image. The first column below each image shows shapes from increasing camera azimuths, the second from different elevations as in fig. 1. All reconstructions can be better visualized online at `http://goo.gl/8Xzy3m`

Modern structure from motion (SfM) and multiview stereo approaches [1, 2] are widely used to recover viewpoint and shape information of objects and scenes in realistic settings due to their accuracy and flexibility, but require multiple images with overlapping fields of view. Here we aim to extend such techniques to cases where only a single image of the target object is available, or when multiple ones are available but from viewpoints far apart. We do this by incorporating recognition: once an object is recognized as a "car" or an "aeroplane", a collection of reusable images of similar objects from those classes can be used to provide additional *approximate* or *virtual* views.

Our idea, illustrated in fig. 1, is to *synthesize* such virtual "SfM" views of the target object by aligning it with images of different instances from the same class, then reconstructing its visible surfaces using robust rigid SfM techniques on both real and virtual views. The main technical challenge we face is the need to align the target object with different objects, which may be pictured with arbitrary viewpoint displacements. Instead of attempting to match to each collection object one by one, we predict the pose of the target object and identify a subset of objects from the collection with similar poses – the intuition is that these will be easier to align with. Following this, we propagate the correspondences to all other collection objects along geodesics on a *virtual view network* of the class, as shown in fig. 2.

An additional difficulty in our setting is that standard RANSAC-based SfM approaches do not apply due to tight rigidity assumptions made by epipolar geometry and fundamental matrix estimation – we triangulate points in objects that do not have exactly the same shape. We pursue instead scaled-orthographic factorization techniques [3] that make fewer assumptions and optimize fewer parameters. We also introduce methodology for a) increasing robustness to the multitude of noise sources by extrapolating synthetic inliers using domain knowledge and b) increasing the specificity of the resulting reconstructions, by emphasizing information in images that more likely related to the target object.

We tested the proposed method on PASCAL VOC and obtained convincing reconstructions on a variety of classes, including animals – a few samples are shown in fig. 3.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

[1] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8):1362–1376, 2010.

[2] Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. *ACM transactions on graphics (TOG)*, 25(3):835–846, 2006.

[3] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.