We present a scalable and incremental approach for creating interactive image-based walkabouts for visualizing dynamic photo collections of landmarks such as the ones available through photo-sharing sites like Flickr. Prior approaches, such as [3], perform a global scene reconstruction as they require the knowledge of all the camera poses. These are recovered via batch processing involving pairwise image matching and the subsequent incremental 3D reconstruction stage where multiple rounds of global non-linear optimization referred to as bundle adjustment are performed. Both steps can become computational bottlenecks for large image collections. Instead of computing a global reconstruction and all the camera poses, our system utilizes several partial reconstructions, each of which is computed from only a small subset of overlapping images. This makes our approach immune to difficulties faced while using an incremental SfM reconstruction, viz. the sensitivity of the choice of the initial pair, and compounding of catastrophic errors in computing camera pose as more images with overlapping views get added to the sequence. It is possible for our system to initialize with a few of images and incrementally add more images as they become available, whereas [2, 3] require all the photographs to be present during processing.

Our system takes a set, \( S \), of uncalibrated images as input. First, we try to determine for every image \( i \in S \), a set of neighbouring images, \( N_i \), as the ones having similarity in appearance with \( i \). For this we extract SIFT features from all the images and use Bag of words based image matching technique using a visual word vocabulary. A Vocabulary tree based image search used in [1] can also be used to avoid the overhead of creating a vocabulary. The matches obtained above can be false so we use a geometry based test to refine the matches. For this we estimate the pairwise epipolar geometry of every image in \( N_i \) by estimating a Fundamental matrix using RANSAC which best explains the feature matches between the images. If the number of inliers to this fundamental matrix is good then the match is accepted into the set \( V_i \), the set of verified images. Next for every image \( i \in S \), we send the images \( i \cup V_i \), through a Sfm pipeline to calibrate \( i \) with respect to all images in \( V_i \). This gives a partial reconstruction, \( P_i \), which comprises of reconstructed cameras corresponding to the images in the set \( i \cup V_i \), and a sparse set of 3D points visible in these images, in a common coordinate frame. Some of the images may fail to get registered by the Sfm pipeline. We call the set of registered images as \( R_i \).

Figure 1: [left] An input image collection. [middle] Our interactive image navigation interface. [right] One of the multiple partial reconstructions of the scene, computed from the images shown in [middle].

Our Image-based rendering framework makes use of these partial reconstructions to provide an image-based walkthrough in a virtual setting. We start with one of the images, \( i \), and display its corresponding partial reconstruction as in Figure 1[right]. The virtual camera is placed congruent to the camera parameters recovered for image \( i \) and image \( i \) is projected on it proxy surface computed from the points visible to \( i \). Images in \( R_i \) are shown as wireframes of their corresponding projections on their proxy planes(computed from the points visible to the images) as in Figure 1[middle]. The user can point and click at any of the wireframes to move to the partial reconstruction corresponding the new image, \( j \). This is done by showing a transition within \( P_i \) during which the virtual camera moves from the camera parameters corresponding to image \( i \) to the camera parameters corresponding to image \( j \) accompanied by the respective fading in and fading out of the destination and source images. Note that for every pair of images \( i \) and \( j \), if \( j \) is a registered neighbour of \( i \), then \( i \) can be made a registered neighbour of \( j \) by obtaining the required parameters from \( P_i \).

As shown in Figure 2, our experiments with datasets of different sizes show a speed up as compared to [2, 3] which involves a global reconstruction. The speedup increases with the size of the dataset as the time complexity of our approach is approximately linear in the number of images making our approach very suitable for incremental addition of images as they become available.

Figure 2: A comparison of running time between our system and Bundler [2] for datasets of various sizes.

Figure 3: Distribution of vertex degrees in the final connectivity graph on which browsing is performed after running our system on the FORT dataset with nearly 6k images.

Our experiments with the FORT dataset with nearly 6k images resulted in a final connectivity graph having the largest connected components of sizes 4249 and 453 images. The average vertex degree for the same graph was 7.1 (Figure 3). In another experiment with a 687 image dataset, we initialized the system with 200 images and incrementally inserted the rest 487 images to obtain a final connectivity graph having a connected component of 674 images. We conclude that a framework based on partial reconstructions can effectively be used to navigate through large collections of images and also incorporate more images as they become available.

