Indian Monument 3D Reconstruction from images

Anshul Mittal
2019ANZ8717
IIT-Delhi
mittal.anshul@cse.iitd.ac.in

Aman Bhardwaj
2019SIY7580
IIT-Delhi
aman.bhardwaj@cse.iitd.ac.in

Abstract

Using multiple views of a camera can enable the perception of depth. This allows us to estimate the camera position as well as the 3D location of points on a certain structure. In this project, we used public sources of pictures (flickr, google images, bing images, etc.) of an Indian monument and reconstructed it in three dimensions by estimating camera poses and stitching together views from different cameras to infer the depth of corresponding points.

1. Introduction

Motivation: Heritage sites are historical key points which stores precious information about the past. From its inscriptions as well as its architecture we can learn so much about the culture, scientific developments of that time and socio-economic structures of the ancient society. Unfortunately with time or some other external factors these monuments often loses their structural integrity as a result they collapse. This is a huge loss to Mankind. To preserve the information either of two things can be done: 1) Government should do regular maintenance and keep detailed records and/or 2) Keep the structure intact digitally. Former costs a lot of time as well as effort and it can be inefficient. On the other hand making and keeping the digital copy of the monuments and heritage site not only allows us to keep it intact but also facilitates us to study it anywhere and anytime. Therefore in this work we used OpenSfM which is an open source tool to 3D reconstruct and preserve Indian Heritage site **Humayun's Tomb**.

OpenSfM [1, 2] is a Structure from Motion library written in Python. The library serves as a processing pipeline for reconstructing camera poses and 3D scenes from multiple images. It consists of basic modules for Structure from Motion (feature detection/matching, minimal solvers) with a focus on building a robust and scalable reconstruction pipeline. It also integrates external sensor (e.g. GPS, accelerometer) measurements for geographical alignment and robustness. A JavaScript viewer is provided to preview the point cloud and debug the pipeline. To view Dense Cloud .ply we have used MeshLab. [4] Our work can be divided into **five** major parts: *1) Data Collection, 2) Data Pre-processing 3) Model/Algorithm Setup 4) Experimentation and 5) Qualitative Analysis of results*. The same has been discussed in further sections in this paper.

2. Methodology

Figure 1 describes the general pipeline for 3D-reconstruction of the Indian Monument. Our contribution to this pipeline are as follows: a) Data collection of **high resolution** for Humayun's Tomb b) Setting up OpenSfM pipeline c) Image adjustments and feature matching experiments for better quality 3D reconstruction d) Qualitative analysis of reconstruction quality with different number of images and e) Model parameter adjustments to improve the 3D reconstruction.

3. Data Collection

Data has been collected from 4 sources: a) flickr.com b) google.com c) unsplash.com d) youtube.com. To download *HD* Images from the above sources following methods have been used.

Source	Method Used
flickr.com	*Bulkr* (Desktop) Trial version
google.com	*DownloadThemAll* Firefox Plugin
unsplash.com	*DownloadThemAll* Firefox Plugin
youtube.com	*XDM* Desktop Download video and Python3.7 to export HD Frames from Monument video.

Table 1. Methods used to download HD Images from sources.

Non-conventional Technique: we followed to create dataset was to download HD video of the monument from youtube vlog in **1080p** resolution, then export non-cluttered frames of the video as source images. This allowed us to get **1920*1080** resolution images for our dataset which is hard to find from open internet sources.

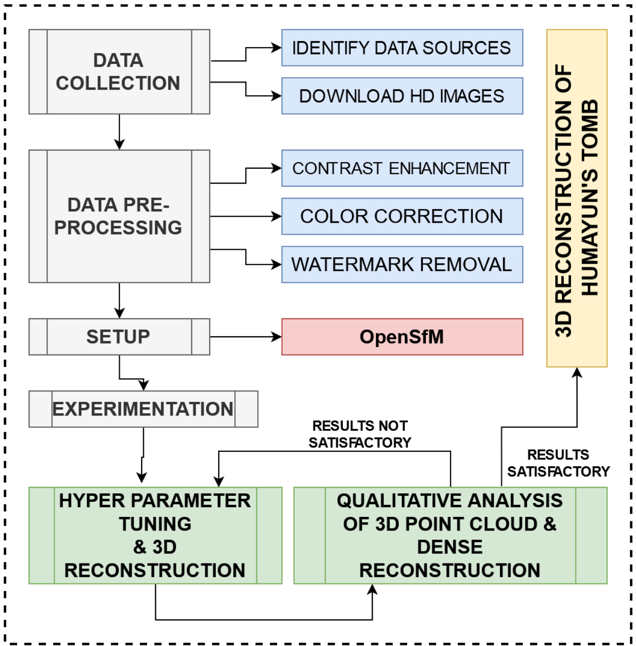

Figure 1. Methodology used for 3D Reconstruction of Indian Monument.

Figure 2. Image samples from the dataset.

By using the above bulk data collection we were able to collect **1600+** images of Humayun's Tomb from various sources. Since we did bulk download, the images had to be analyzed for the usability. We filtered out the images on the basis of following criteria *a) min-resolution should be at least 600*600p b) minimum human distractions in the image around the monument c) less/no watermarks overlapped with the monument d) dataset must cover 360 degree view of the monument*. Post moderation we were left with **500+** high resolution, non cluttered images of Humayun's tomb which could be utilized for the 3D reconstruction. Please refer to figure 2 for few examples. Figure 3 summarizes our approach to collect and clean the data.

Figure 3. Images collection and moderation.

4. Experiments

In this section we discuss the techniques and tools we explored to enhance the 3D reconstruction of the monument.

4.1. Water Marks removal

Problem: Most of the public images consists of watermark of the owner. While it is important to protect the copyright for 3D reconstruction it can lead to false feature matching. If a watermark appears in too many images then it can appear in the reconstruction, therefore false positive features appear in the reconstruction. The same can be seen in fig 4. This figure shows the watermark from one viewing angle of reconstruction.

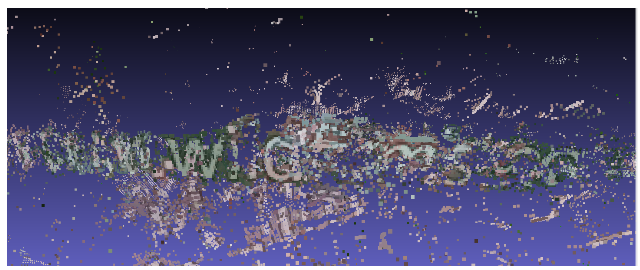

Figure 4. Adverse effect of watermark on Reconstruction.

Solution: We performed image inpainting [6] to get rid of the watermarks from the image, so that they don't contribute to false features for the monument and we get a clean images as source. (refer fig 5)

Figure 5. [*Left*] Image with watermark. [*Right*] Inpainted image, watermark removed.

4.2. Image correction

Problem: Since images where taken from different sources all over the internet each image had different lightning conditions. Please refer to figure 6(*left*) for example. Due to different lightning conditions features despite having obvious overlap where not able to match and those which did lead to the visible seam making 3D reconstruction look inferior.

Solution: To overcome the problem stated above, brightness enhancement and CLAHE [7] contrast stretching has been applied to images. These processed images are then used for further so as to get better matches and better end results. Please refer to figure 6(*right*) for an example.

Figure 6. [*left*] Original dull image. [*right*] CLAHE improved image.

4.3. Hyper-parameter Tuning

Problem: Running OpenSfM using default parameters leads to sub-optimal results. OpenSfM uses HaHoG as default feature matching algorithm and uses all images in the dataset folder.

Solution: First we optimized hyper-parameters by trying various feature matching algorithms (AKAZE, SURF, SIFT, HAHOG, ORB) available in the library. We found out that SURF [3] gave the best results. Then we varied the hyper-parameters `surf_hessian_threshold` & `surf_n_octavesrelated` related to SURF to get the optimum results for feature matching (refer figure 7). The front view results of SIFT and HaHoG were not clear as a lot of noise points can be observed in the figure. However SURF gave cleanest view of the monument as compared to the other two feature descriptors. The top view all the three were similar for all the three (refer figure 8).

4.4. Input Dataset Variation

Problem: Blindly using all images lead to excessive noise points around the monument cloud where no structure was there in the original images.

Solution: We gradually reduced number of images from 550+ to 55. We observed that randomly sub-sampling fewer images gave better results as compared to using large number of images. In reference to fig 9 the result *a)* is of 550 image set. The monument is cluttered with a lot of noise points. Improvement in the results can be see in *b)* 150 images and further in 85 images in *c)*. The best results obtained in Figure 10 from 55 images set. Minimal set of images was chosen such that they cover the monument from all the angles. This produced clean and non-cluttered 3D reconstruction (refer fig 10). This is analogous of choosing Minimum Description Length Principle [5].

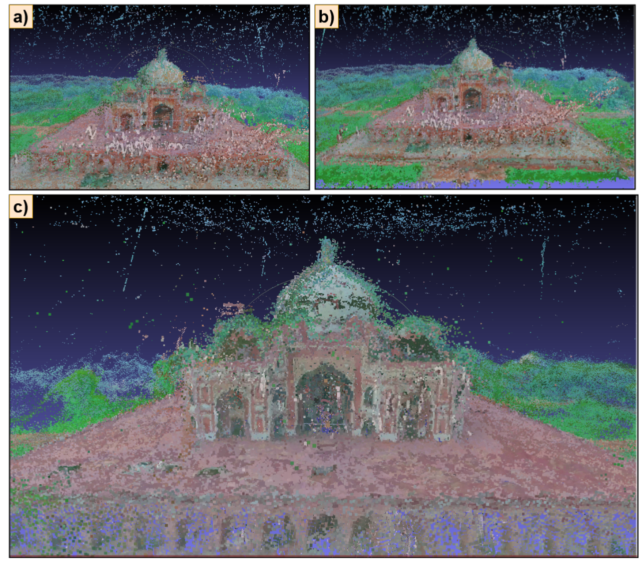

Figure 7. Front view results on default setting but different feature descriptor for *500+* images a) HaHoG b) SIFT c) SURF

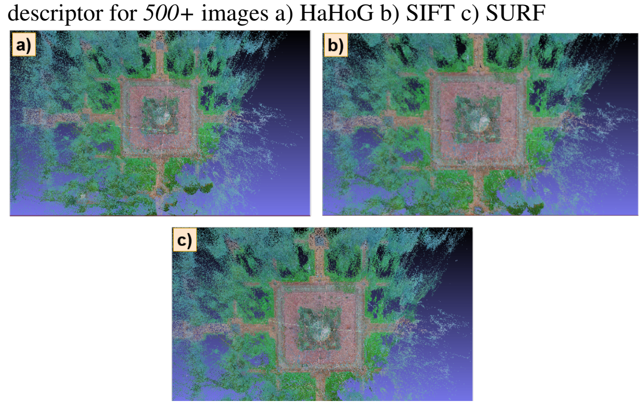

Figure 8. Top view results on default setting but different feature descriptor for *500+* images a) HaHoG b) SIFT c) SURF

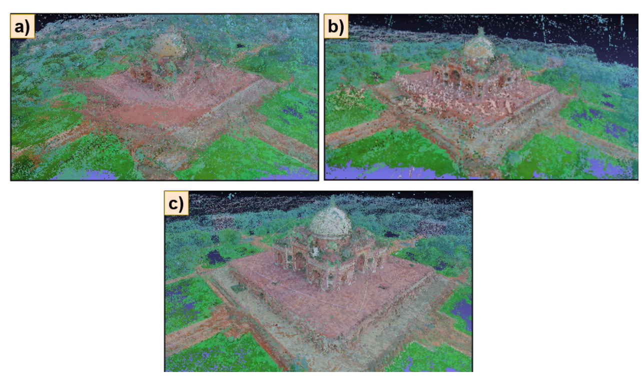

Figure 9. Improvement in results by varying number of images in the input dataset. Number of images in a, b and c are 550, 150, 85 respectively.

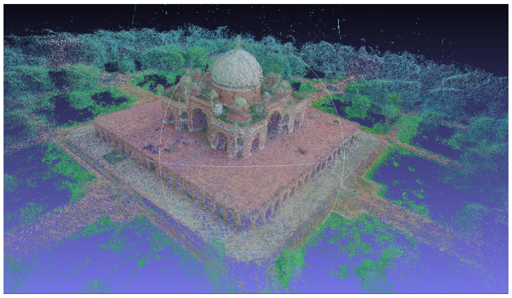

Figure 10. Results of final dataset in consensus to Minimum Description Length Principle. [5]

5. Final Results

For our final output we used 43 images randomly sampled from 550 images. All these images were color corrected and water mark was removed as described in the section 4

Highlights from our reconstruction.

- Complete 360 degree reconstruction of Humayun's Tomb.

- Monument features like pillars, dome, entrance gate *etc*. are clearly visible. refer fig. 11 and fig. 12

- No image from inside the monument were included in our dataset, yet the bottom view has constructed close enough architecture of the monument from inside. refer fig. 13

- Nearby pathways, fences, trees can also be seen. refer fig. 14

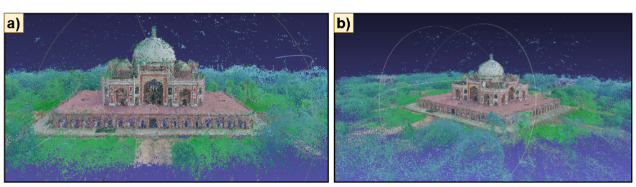

Figure 11. Wide angle view of Humayun's Tomb. a) front b) diagonal view.

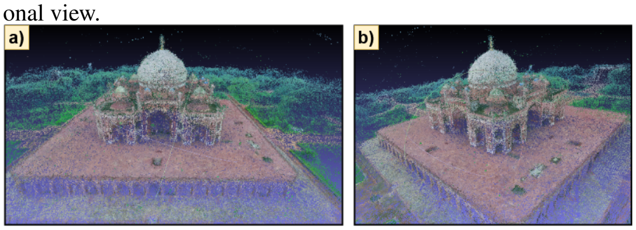

Figure 12. Close view of Humayun's Tomb. a) front view b) slightly elevated diagonal view.

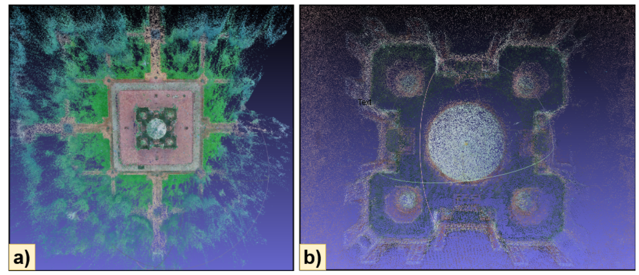

Figure 13. Top view and bottom view of the dense could showing the layout of the entire monument.

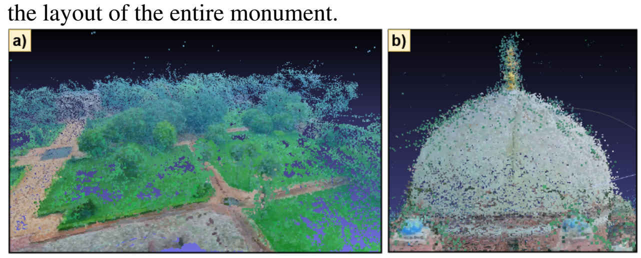

Figure 14. Utilities like trees and dome is clearly visible in the dense cloud.

6. (BONUS) Local Monument reconstruction using same pipeline

Pipeline Validation: To validate the robustness of the pipeline setup for the 3D reconstruction of Humayun's Tomb, we will have collected images of **Laxmi Niwas Palace, Bikaner, Rajasthan (*Hometown*)** and processed them using the same pipeline. A total of 14 images of the palace were taken such that we had a clear view of the palace. The results for the same are as follows. Refer figure 15 and figure 16

Figure 15. Front view of Palace.

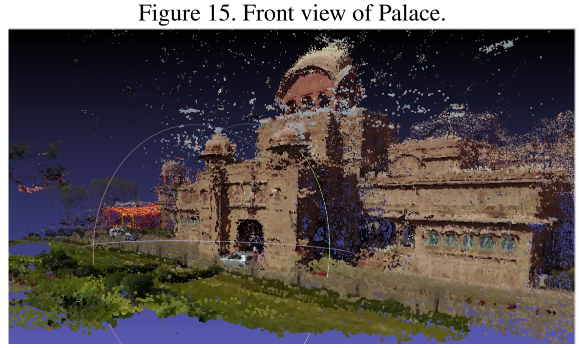

Figure 16. Diagonal View of Palace

7. Discussion

We presented 3D re-construction of an Indian heritage site (Humayun's Tomb) and proposed a simple and easy to use data collection pipeline. Additionally our proposed techniques further improved the reconstruction strategy. It should be noted that this reconstruction pipeline is not unique to Humayun's Tomb and can be used for any monument site. For future work, we will try feature matching using deep learning which could easily accommodate complex structures and novel patterns unlike HaHoG and SURF which focus only on corner or corner like objects.

References

[1] P. F Alcantarilla. Fast explicit diffusion for accelerated features in nonlinear scale spaces. 2011.

[2] P. F. Alcantarilla, A. Bartoli, and A. J Davison. Kaze features. In *European Conference on Computer Vision*. Springer, 2012.

[3] H. Bay, A. Ess, T. Tuytelaars, and Luc Van G. Speeded-up robust features (surf). *Computer vision and image understanding*.

[4] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. MeshLab: an Open-Source Mesh Processing Tool. In Vittorio Scarano, Rosario De Chiara, and Ugo Erra, editors, *Eurographics Italian Chapter Conference*. The Eurographics Association, 2008.

[5] Jorma Rissanen. *Minimum Description Length Principle*, pages 666–668. Springer US, Boston, MA, 2010.

[6] A. Telea. An image inpainting technique based on the fast marching method. *Journal of graphics tools*.

[7] G. Yadav, S. Maheshwari, and A. Agarwal. Contrast limited adaptive histogram equalization based enhancement for real time video system. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE.