Sparse 3D Model Reconstruction from photographs

Kanchalai Suveepattananont
ksuvee@seas.upenn.edu
University of Pennsylvania

Professor Norm Badler
badler@seas.upenn.edu
University of Pennsylvania

Advisor Aline Normoyle
alinen@seas.upenn.edu
University of Pennsylvania

Abstract

Nowaday, millions photos and videos are being shared and displayed on the internet everyday. Large photograph collections of buildings, statues and various tourist- attraction sites can be found easily on the internet. Ones no longer need to leave computers to "visit" the Notre Dame Cathedral in Paris or the Great Wall in China and many other sites in the world. Simply searching via large collections of photo on Google or Flickr, one can see all these places in large-scale and in details. However, images of these large constructions still lack of visualization technology to give users immerse experiences in browsing through them. Therefore to effectively and efficiently utilize this visual information, more and more efforts are put into developing visualization and user-interfaces tools to help users browsing through it.

In this project, I propose a system for interactively and immersively browsing and exploring large collection of unstructured photographs of sites using 3D-interface. The system allow users to see minute details of sites via photographs captured by other visitors of the sites. It will also enable users to visualize sites through sparse, 3D reconstructed scene which will give more immersive experiences, allow them to "walk" around the places as if they are there physically. The system will provide interactive users-interfaces which seamlessly transform between these 2D detail photograph and 3D scene. The system will use data-bases of photograph collected from the internet to create sparse 3D scene, employ computer-vision techniques such Structure from Motion to build the sparse 3D model in form of point cloud, and use non-photorealistic rendering in style of watercolor and combination of point and lines.

Project Blog: cis-497-3dreconstruction.blogspot.com

---

1 Introduction

Large collections of photograph of various tourists sites are available on the internet, giving users' experience to "visit" these sites. However, users still browse these large collection in traditional methods using grid-based layout, making the experiences less immersive. Ones can see details of many sites through these 2D collection of photographs; One can also easily find images of places in large scale and see overall pictures of them. However, there is a large discontinuity between these two categories of photographs, making it impossible for users to visualize where all details from these photography located on sites and how sites interact with their intricate details. This, then, makes user's experiences less immersive and effective.

I propose a system to enable users to experience sites in more immersive manners through interactive user-interfaces and sparse 3D scene visualization. The system will allow users to explore sets of photographs to visualize details collected by others people while at the same time able to visualize sites in large-scale and see how these intricate details fit with in overall scale of these sites. This system will provide transformation between 3D navigation and photographs exploration.

The project attempts to make following contributions:

- I will develop a system that construct sparse 3D scene composing of points clouds. The models will be constructed from photographs collections of sites taken from different perspectives and points of views.
- I will develope a non-photorealistic rendering of sparse 3D scene using points and lines and visualization of photographs and its relationship with sparse 3D model.
- I will build interactive and intuitive user-interfaces that seamlessly transform between photograph and sparse 3D models, providing immersive experiences.
- I will also learn various computer vision techniques particularly in the field of object recognitions, 3D reconstructions from images, image processing. In addition, I will be expose to non-photorealistic rendering.
- If time permit, I will also gather images from Facebook instead of simply query images from Google and Flickr and apply auto-annotation to organize images by dates, relationship and other social parameters. This will allow more social and group.

1.1 Design Goals

Main audience of the project is everyday computer users who would like to have immersive experience in browsing photographs of various places and being able to feel as if they visit these sites physically. Users will be able to navigation and transition between sparse 3D model and 2D photographs.

Goals of the project are to create visualization model that provide interactive and immersive experience in browsing and exploring collections of photographs, to provide architectural, historical, scenic details of various sites, and to allow me to be exposed to topics and application in both computer graphic and computer vision.

1.2 Project Proposed Features and Functionality

What features and functionality will you implement for your design project?

- Reconstruct sparse 3D model from collections of unstructured photographs obtained from the internet queries of various places and sites.
- Render 3D model in form of points clouds and line. If enough information is provided and time permits, non-photorealistic techniques in style of watercolor will be applied to provide better users experiences.
- Provide approximate geo-locations indicating where images are taken in relation to real geographic locations of the sites.
- Provide intuitive and immersive browsing and exploring user-interfaces which allow seamlessly transform between 3D navigation and 2d ones.
- If time permitting, instead of pulling photograph from query search engine, the system will use photographs from users social groups and organize images which reflect social relationship of owners integrate into already existed 3D visualization.

2 Related Work

Works in the area relied upon performing image processing, image analysis, rendering and visualization. They also relied upon advnacement in technology to identify location, date, and many metadata information of photographs. All of these are to be used to answers questions such as where images are taken (in the sense of GPS location and points of views), what cameras look at, what relationship between each image etc.

There are various work to organize large collection of photographs. For example, [Aris et al. 2004], which exploits location and time information of photographs to organize and create storytelling. Another similar works by [Naaman et al. 2004] also propose an algorithm to automatically organize digital photograph based on geographic coordinate and time when photographs are taken to provide an intuitive way which mimics how people perceive their photo collections. Earlier in 2003 [Naaman et al. 2003] also propose sharing geographic metadata information of digital photographs to facilitate such organization.

In addition to organize sets of images based on geolocation and time, works have also been done in visualizing the collections in order to mimics real physical and geographical locations where images are taken. For instance, in 1995, [McMillan and Bishop 1995] proposes image-based rendering by sampling, reconstructing the plenoptic function to generate surface based on cylindrical projection. There are also various late works which similar but focus more on 3D applications. [Kadobayashi and Tanaka 2005], develops a method using 3D viewpoints to search and organize sets of photographs, as well as browse groups of images from similar points of views. Another one is [McCurdy and Griswold 2005] proposes a system to reconstruct architecture from video footages.

In 2006, [Snavely et al. 2006] propose another visualization and organization of photograph by creating sparse 3D scene constructed from collection of photograph from different points of view of the sites. As the project is mainly based on NOAH, some related technical details mentioned in the papers are included.

2.1 Keypoints feature and Image Matching

As NOAH proposes methods to reconstructions of sparse 3D scene, the first step in organizing and visualizing these images collections are to find feature points in each image. LOWE proposed an approach to perform this task of keypoint detector and localization, Scale Invariant Feature Transform (SIFT). However, to find similarity between images, [Snavely et al. 2006], proposes by [Arya et al. 1998], [Arya et al. 1998] develops algorithm to robustly approximate nearest neighbors. This general algorithm can then be applied in more specific context of the problem.

2.2 Bundle Adjustment

In order to reconstruction 3D model, [Snavely et al. 2006] also suggested using a software package developed [Lourakis and Argyros], which based on using levenberg-marquardt algorithm [Nocedal and Wright 1999], to perform bundle adjustment which is common process in 3D reconstruction to refine optical characteristic of camera and 3D coordinates of points.

3 Project Proposal

The objective of the project is to provide a new method to visualize unstructured collections of photographs and provide users an immersive experience in exploring and browsing through the collections. I propose a system which reconstructs sparse 3D models from unstructured photographs taken, uses non-photorealistic rendering compositing of points and lines as well as watercolor style when appropriated, and organize sets of images via their geographic location in which images are taken. Moreover, the system will provide an immersive and seamless user-interface which transition between 3D and 2D navigation, giving users an experience of physically visiting the sites.

3.1 Anticipated Approach

3.1.1 Keypoints Selection and Feature Matching

The first step to utilize sets of images is to detect key features in each image. The process of doing so is well studied as mentioned in 2.1. In order to perform robust keypoint detection which is invariant to image transformations, I will utilize existed SIFT package which will provides both locators and descriptors of keypoints. Once keypoints are selected, the next step is to find similarity between keypoint descriptors using approximate nearest neighbors as mentioned in 2.1. Similarly, such task can also be accomplished by utilizing existed software package.

3.1.2 Structure of Motions and Bundle Adjustment

In addition to information gained from previous part, I will have to parse more metadata stored in EXIF (Exchangeable Image File Format) of JPEG such as focal length, geographic locations, angles etc. The task can be leveraged using jHead library. With all these data, I am able to utilize them to reconstruct sparse 3D model, Structure of Motions (SfM). The process is an iterative process of continuously adding camera information to perform optimization. The core part of the SfM, performing Bundle Adjustment, will be done using the Sparse Bundle Adjustment Library as mention in **??**.

3.1.3 Rendering and Visualization

Main rendering technique of the project is to visualize the sparse 3D model as points clouds and lines. If applicable and time permitting, I will develope non-photorealistic rendering style of watercolor. In addition to visualize the model,I will also organize collections of photographs using geographic locations and estimated, relative camera positions resulted from SfM.

3.1.4 User-Interfaces

I will provide both 3D navigation of the model which is a standard motion controls and seamless transformation between 3D control and viewing of photographs. I will attempt to provide intuitive, effective, and user-friendly ways to visualize photographs and their relationship with 3D representation of the sites. If time permitting, it will also be in my interest to employ similar navigation techniques used by [Snavely et al. 2006] in moving between two photographs and their related views.

3.2 Target Platforms

I will mainly develop the project in C++ and use various libraries to perform certain tasks: ANN library for Approximate Neares Neighbor Searching, SIFT++ for Scale-Invariant Feature Transform detector and descriptor, jHead for reading EXIF of JPEG, Sparse Bundle Adjustment library for bundle adjustment, as well as refering to SfM libraries and frameworks. I also will use Eigen library for matrix computation if needed.

3.3 Evaluation Criteria

To evaluate the project, I will compare the sparse 3D model in relative to its original sites and whether it successfully provide accurate visualization (i.e users can identify the sites). I will compare to existed works of [Snavely et al. 2006] and see how close and accurate my application. Another criteria is to use database of photographs and compare them to similar areas of the model and observe how closely they link together.

4 Research Timeline

Below are my project timeline. I also have included a Gantt chart in the final page with more detail.

4.1 Project Milestone Report (Alpha Version)

- Complete all background reading
- Collect all necessary images
- Developed software framework is functioning with simple base case such taking data set of images and performing keypoint detection
- Built keypoint detection and matching

4.2 Project Milestone Report (Beta Version)

- Built basic but functional SfM procedure although more tweeking may still require
- Devloped rendering system which will aid reconstruction process
- Built standard 3D navigation tools

4.3 Project Final Deliverables

- Complete sparse 3D recontruction from collections of photographs
- Video to demonstrate users interact with the application
- Live demo of the application
- Complete rendering of sparse 3D model in point clouds
- Compete final report

4.4 Project Future Tasks

If I have extra 6 months to work on the project, I would like to add following features

- 2D Navigation of views related to each others
- Non-photorealistic rendering in style of watercolor
- Providing options for quering database of image from social media
- Develop alternative organization models of images collection such as using time
- Provide functionality which users can manually select features on their images and see how such features interact with reconstruted model

---

You will fill in the following sections as you make progress on your project, particularly for the alpha review and the final deliverable. In these sections, list pseudo-code, charts, images, examples, etc. to show what youve done over the course of the semester.

5 Method

6 Results

7 Conclusions

8 Future Work

References

ARIS, A., GEMMELL, J., AND LUEDER, R. 2004. Exploiting location and time for photo search and storytelling. Tech. Rep. MSR-TR-2004-102, Microsoft Research, One Microsoft Way, Redmond, WA, 98052.

ARYA, S., MOUNT, D. M., NETANYAHU, N. S., SILVERMAN, R., AND WU, A. Y. 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. ACM 45*, 6 (Nov.), 891–923.

KADOBAYASHI, R., AND TANAKA, K. 2005. 3d viewpoint-based photo search and information browsing. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, New York, NY, USA, 621–622.

LOURAKIS, M., AND ARGYROS, A. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Institution of Computer Science-FORTH. Available from www.ics.forth.gr/ lourakis/sba.

MCCURDY, N. J., AND GRISWOLD, W. G. 2005. A systems architecture for ubiquitous video. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, ACM, New York, NY, USA, MobiSys '05, 1–14.

MCMILLAN, L., AND BISHOP, G. 1995. Plenoptic modeling: an image-based rendering system. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH '95, 39–46.

NAAMAN, M., PAEPCKE, A., AND GARCIA-MOLING, H. 2003. From wher to what: Metadata sharing for digital photogrphs with geographic coordinates. In *Proceedings of the Int. Conference on Cooperative Information Systems*, 196–217.

NAAMAN, M., SONG, Y. J., PAEPCKE, A., AND GARCIA-MOLINA, H. 2004. Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, ACM, New York, NY, USA, JCDL '04, 53–62.

NOCEDAL, J., AND WRIGHT, S. J. 1999. Numerical optimization. *Springer Series in Operations Research*.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: exploring photo collections in 3d. *ACM Trans. Graph. 25*, 3 (July), 835–846.

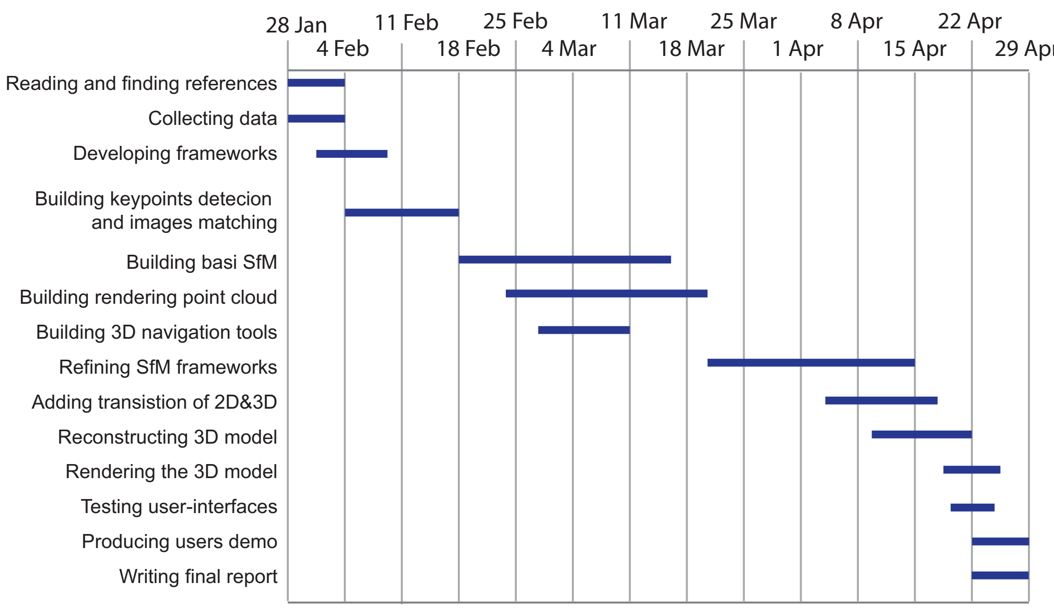

Figure 1: *Sample illustration.*