The Power of Indoor Crowd: Indoor 3D Maps from the Crowd

Si Chen, Muyuan Li, Kui Ren

Department of Computer Science and Engineering

The State University of New York at Buffalo

E-mail: {schen23, muyuanli, kuiren}@buffalo.edu.

Abstract—**Remarkable progress was made with smartphones in the last few years. Modern smartphones are now equipped with high-resolution cameras and various micro-electrical sensors that open up new mobile application possibilities. In this work, we address a critical task of reconstruct indoor large-scale 3D model from crowd-sourced images. We propose, design, and implement IndoorCrowd, a smartphone empowered crowdsourcing system for large-scale indoor 3D scene reconstruction. IndoorCrowd fills a gap in current cloud-based 3D reconstruction systems as it ensures at mobile side that the captured image set fulfills desired quality for indoor large-scene 3D reconstruction. At the cloud side, we deploy an automated image-based 3D reconstruction pipeline, which generates 3D models from images and sensor data. Moreover, we provide an intuitive online annotation tool that allows easy image labeling. We present that these labeling information combined with sensor data helps IndoorCrowd reduce the total processing time greatly.**

Keywords—Indoor 3D map, crowdsourcing, 3D reconstruction.

Fig. 1: Initial Data Acquisition using Smartphone App

I. INTRODUCTION

Recently, several online map services provide a 3D mode that enables you to fly around cities that are rendered with amazing detail. These 3D models are constructed by a combination of aerial photography and radar technology. However, most online maps focus only on outdoor environments rather than indoor, although most of our time, we spent indoors.

Unlike outdoor environment, reconstruct indoor 3D maps are much more difficult. This is due to the reason that indoor environment contains a lot of private areas that cannot be accessed by the public. Therefore, we cannot send vehicles or hiring volunteers to photograph the indoor scenes. What's more, no accurate GPS signal available inside the building, and thereby, it is difficult to acquire the geographic location for each image in indoor environment. Hence, existing outdoor 3D reconstruction techniques cannot be directly apply to indoor environment.

To conquer these difficulties, we present IndoorCrowd, a smartphone empowered crowdsourcing system for large-scale indoor 3D scene reconstruction. Our approach is based on crowdsourcing, smartphone users using our mobile application to capture their preference indoor environment and upload key frames to the cloud, along with real-time sensory data and labeling information. The cloud infers photographer's location and orientations from the uploaded images combined with sensory data and produces a sparse 3D geometric representation of the scene using the state-of-the-art 3D reconstruction algorithm. Once finishe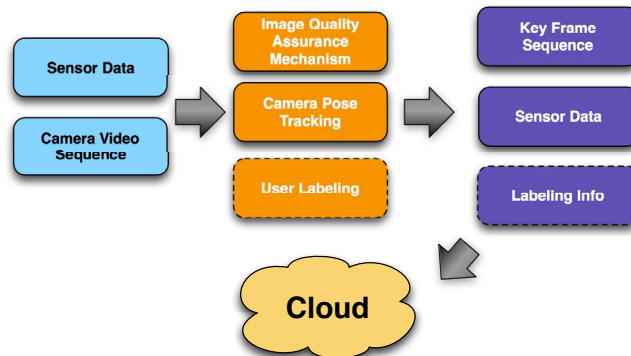d, these sparse 3D points will go through the remaining 3D reconstruction pipeline. The final output of our system is a dense 3D model which represents the indoor scene.

These dense 3D models will connect with each other based on its geographic location and form an indoor 3D maps. These crowdsourcing indoor 3D maps can be used for various applications, such as 3D visualization, indoor navigation or routing, public participation, and emergency response.

Compared with several related works [1] [2], IndoorCrowd is designed based on regular smartphones without supporting from any external infrastructures and additional devices and our cloud-based architecture is capable to reconstruct large-scale indoor environment. To ensure our system have the ability to handle large scale indoor environment in widely different light conditions, such as the complex reflectance, shading effects and ambient occlusions, our mobile application provides real-time feedback while user capturing the indoor environment.

II. SYSTEM ARCHITECTURE OVERVIEW

We adopt a three-phase solution to accomplish the whole system: initial data acquisition, 3D reconstruction pipeline, and user labeling phase.

Initial Data Acquisition The first step that has to be done is the acquisition of a valid data set. In particular, IndoorCrowd adopt to the crowdsourcing approach. Therefore, we proposed a mobile application on the iOS system to let the crowd contribute their data. Fig.1 illustrates the basic
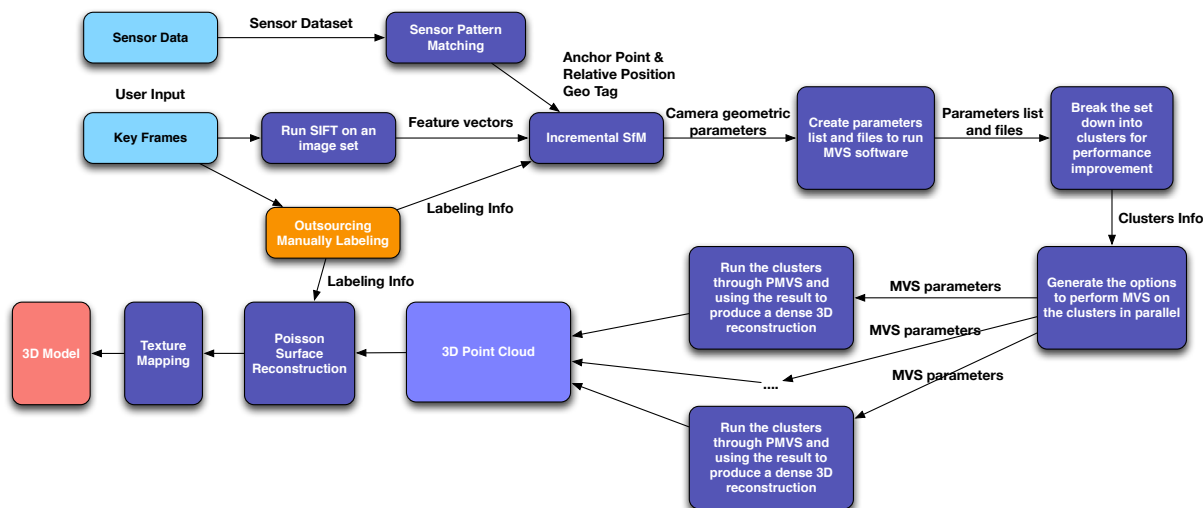
Fig. 2: 3D Reconstruction Pipeline on the cloud

working flow of the initial data acquisition phase, and the functionality of detailed components is as follows. User record video and upload key frames using our iOS app, which pairs the image with contextual data from sensors, including last available GPS, WiFi, accelerometer, and gyroscope. When user capturing the indoor environment, our iOS app also provides an useful real-time feedback based on sensor readings and image pre-processing result.

3D Reconstruction Pipeline Generation of 3D models for real-world environments from 2D images has been a long-term goal in computer vision community. Recently, large-scale 3D reconstruction from multiple images using the structure from motion (SfM) and multi-view stereo (MVS) pipeline has reached a certain level of maturity. There has been significant success with several algorithms [4]–[6]. By leveraging advanced techniques from computer vision, we choose the state-of-the-art 3D reconstruct pipeline for indoor 3D scene reconstruct. Fig. 2 illustrates the total 3D reconstruction pipeline.

At the beginning of the 3D reconstruction pipeline, incremental Structure from Motion (SfM) is used to reconstruct a 3D representation model of a physical space from 2D images. The model SfM pipeline generated consists of many points, a point cloud, in 3D Euclidean space. We choose OpenMVG [7], an open source software package for incremental SfM, both for the initial model construction and later image-to-model alignment.

While the SfM technique is complex, its usage is straightforward. Simply, the input is several photographs of a physical space (the minimum number is four). With these images as input, SfM operates in a pipeline: (1) extraction of the salient characteristics of a single image using SIFT [3] algorithm, (2) comparison of these characteristics across images to match shared points of reference, and (3) minimizing the projection error of the 3D key points into the cameras planes. Step (1) and (2) combined together also known as key-point matching process, step (3) refers to local bundle adjustment process.

The goal of multi-view stereo pipeline is to reconstruct a

complete 3D object model from a collection of images taken from known camera viewpoints, it performs an optimization on these reference points and the output is a reconstructed 3D point cloud. For processing a large number of input images, we integrates Clustering Views for Multi-view Stereo (CMVS) [6] algorithm to decomposed the input images and then use Patch-based Multi-view Stereo (PMVS) [5] algorithm to process each cluster independently and in parallel.

The poisson surface reconstruction algorithm convert the output 3D point cloud into a 3D surface triangle mesh. The last step is texture mapping, where we mapping the surface texture on to our 3D model.

User Labeling Phase Indoor environment contains a lot of smooth textureless objects such as walls, desktops and floors. Existing 3D reconstruction pipeline are vulnerable to these textureless objects and will produce disconnected components. To solve this problem, we provide an intuitive online annotation tool that allows easy image labeling. Then we use the object labels to fix errors in the 3D reconstruct models.

REFERENCES

[1] Tanskanen, Petri, et al. *Live Metric 3D Reconstruction on Mobile Phones.*, ICCV, December 2013

[2] Newcombe, Richard A., et al. *KinectFusion: Real-time dense surface mapping and tracking.* Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on. IEEE, 2011.

[3] Lowe, David G, *Distinctive Image Features from Scale-Invariant Keypoints*, Int. J. Comput. Vision, November 2004

[4] Snavely, Noah and Seitz, Steven M. and Szeliski, Richard, *Photo Tourism: Exploring Photo Collections in 3D*, ACM Trans. Graph., July 2006

[5] Yasutaka Furukawa and Jean Ponce, *Accurate, Dense, and Robust Multi-View Stereopsis*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010

[6] Yasutaka Furukawa and Brian Curless and Steven M. Seitz and Richard Szeliski, *Towards Internet-scale Multi-view Stereo*, CVPR, 2010

[7] OpenMVG, *http://imagine.enpc.fr/ moulonp/openMVG/index.html*