
# #100 EFFICIENT 3D ACCIDENT SCENE RECONSTRUCTION

Final Research Report

Luis E. Navarro-Serment, Ph.D.

The Robotics Institute

Carnegie Mellon University

## Disclaimer

## #100 Efficient 3D Accident Scene Reconstruction

# Problem

In this project, we propose to leverage recent developments in the field of computer vision to support the deployment of a system capable of reconstructing scenes in 3D using commodity cameras, such as the ones embedded in cell phones and tablets. With this system, a sequence of pictures is transformed into a three-dimensional representation, which can be used to visualize, understand, analyze and measure the geometry of the scene. This concept is illustrated in Fig. 1, where a 3D model is obtained from several pictures of the scene taken from different locations.

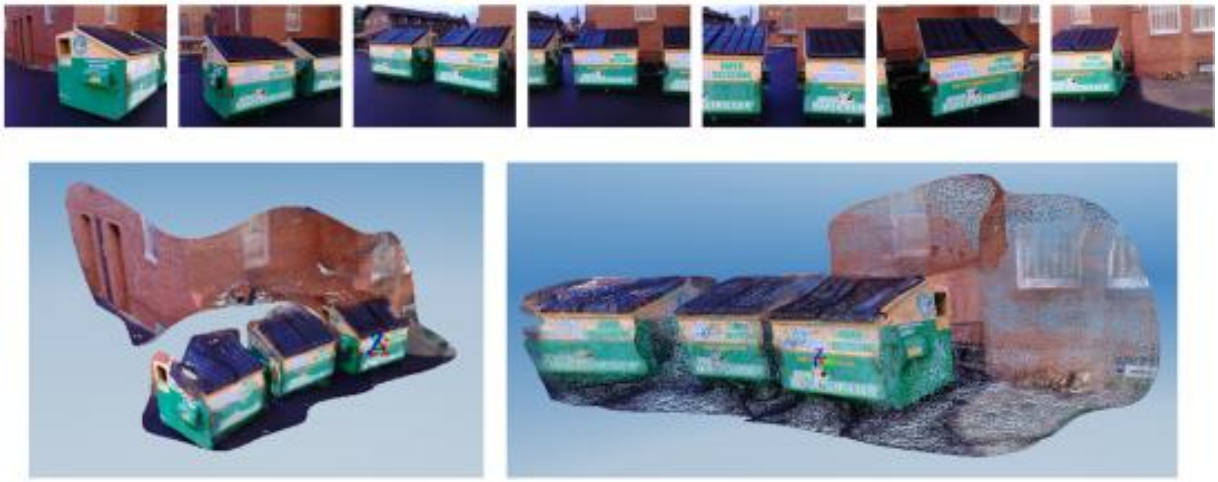

**Fig. 1: A sequence of pictures (top row) is transformed into a 3D model, which can be manipulated to see the scene from different viewpoints (bottom row). For this figure the 3D model is represented as a mesh. The mesh is clearly visible in the bottom-right figure.**

The process of estimating 3D geometry from multiple images is commonly known as the *Structure from Motion* (SFM) problem. Scientists have produced multiple algorithmic solutions to this problem in recent years. Typically these approaches start by locating 2D features in the scenes depicted in the images, such as corners of objects, contours, or other salient points. Then, a matching process attempts to associate these features to their corresponding occurrences in other images. By analyzing these correspondences, it is possible to recover measurements of the relative rigid poses between different cameras. Finally, a process known as *bundle adjustment* calculates a maximum likelihood estimate of the camera poses and feature point locations, after performing an initialization using a subset of the paired measurements.

The main requirement for a successful reconstruction is the quality of the image. For instance, incorrectly exposed photos typically contain regions that are either too dark or too bright, from where it is difficult to identify feature points. Likewise, finding good features is harder when images are out of focus. Additionally, images must be captured in a way that each picture contributes to the reconstruction by providing information about the scene through the projections of 3D points into the image planes of cameras placed at different locations. This concept is similar to the underlying principle of stereo vision, where depth is inferred by analyzing the disparities between images captured by the

stereo pair. Similarly, if the sequence of input images is not chosen carefully, the SFM algorithm may produce multiple models, which do not share enough images between each other. This is caused by the lack of feature matches between very different viewpoints. Fortunately, this situation can be averted by capturing enough images to cover the intervals between these viewpoints. This facilitates the detection of common feature points from images that are close to one another, so they can all be stitched together in a smooth and continuous manner.

The primary challenge addressed by SFM algorithm is the automatic estimation of camera pose from a group of photographs collected from different locations. However, a successful automatic recovery of a 3D model is not always possible, and human intervention becomes necessary. Various software applications that implement algorithms for SFM are available in different modalities: from commercial products, to open source projects. These programs are capable of processing a group of pictures to produce a combined 3D representation. However, they frequently require human guidance to complete the reconstruction task; for this reason they usually include functionalities for the manual selection of feature points, which the software uses to reconstruct the 3D geometry of the scene. Despite the availability of such programs, there is still a clear need for a system capable of producing scene reconstructions consistently. However, our work in algorithms and techniques for scene reconstruction and understanding has provided insight and identified an element capable of satisfying this need: an assisted image capturing system, which guides users as they collect set of the images which will be merged into a 3D model. This element is the focus of the effort reported here.

## Our approach

### *Smartphone App*

As mentioned before, the input images must satisfy a set of requirements for a scene to be reconstructed properly. With this goal in mind, we developed a proof-of-concept application to help with the task of collecting the images that are used to model the scene. We have leveraged work done recently by colleagues in our lab to assist blind users in taking pictures during photographic documentation of transit problems and accessibility barriers. Their smart phone app used directional sounds and markers in the image to guide the photographer. Their user study demonstrated that this approach was helpful for sighted, visually impaired, and blind users. In particular, their approach for sighted users was the most relevant to our work.

Based on a similar concept, we developed a methodology for assisted image capturing. Its purpose is to help a non-expert to photograph the area to be documented in a way that a successful reconstruction of a 3D model can be accomplished later. We implemented a proof-of-concept application (App) on a smart phone using the free Android Software Development Kit[1]. We have made the source code for the App freely available from GitHub[2].

[1] https://developer.android.com/sdk/index.html
[2] https://github.com/cmu-navlab/scene-recon-android

This App, through a simple user interface, provides the photographer with a visual indication of where and how to aim the smart phone's camera, and where to move next to cover the entire scene. A screen capture of the App running is shown in Fig. 2.

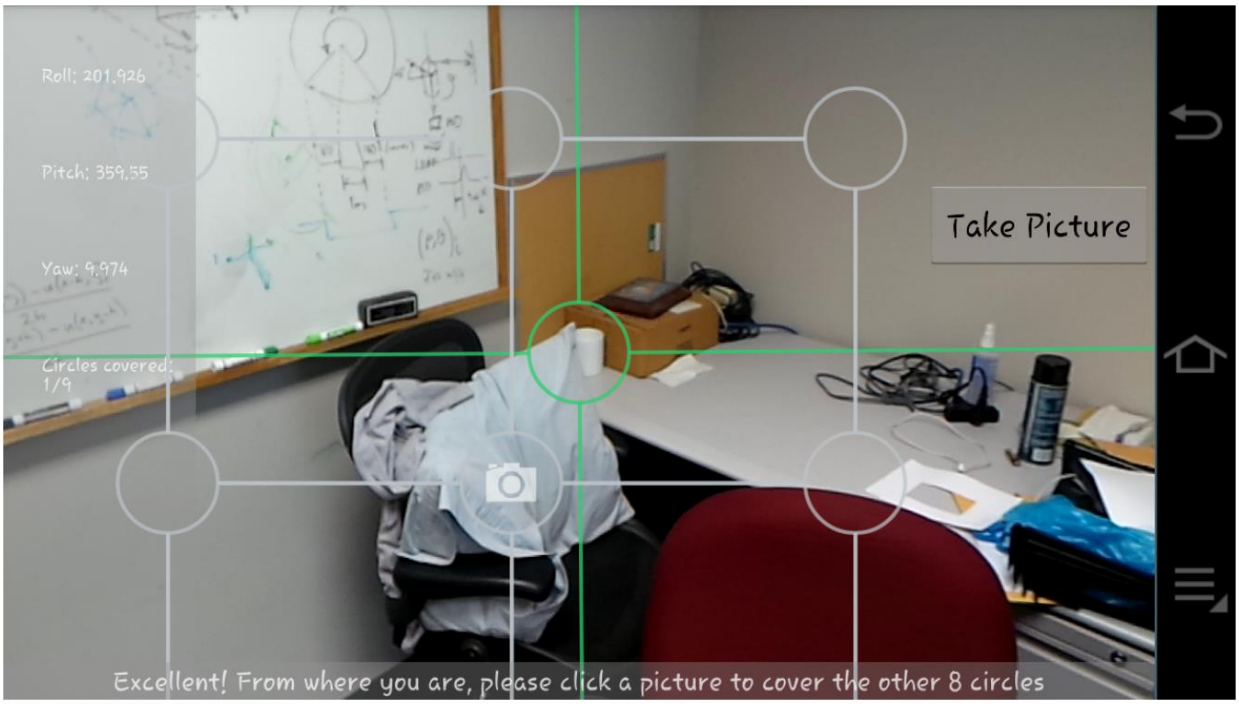


**Fig. 2: Screen capture of the App, showing its user interface. Markers are overlaid on top of a live camera preview.**

The application displays a live camera preview. Its user interface uses colors and numeric indica- tors overlaid on top of the image to alert the photographer of undesirable smartphone poses. The user interface provides visual indications in the form of colored cross-hairs to ease alignment, as seen in Fig. 3. When the smart phone is not horizontal, the "Take Picture" button is disabled. As soon as the device is horizontal, the button is enabled, and the user may take a picture. The camera is forced to operate continuously in auto-focus mode to reduce the possibility of blurred images. This mechanism effectively prevents users from collecting undesirable pictures.

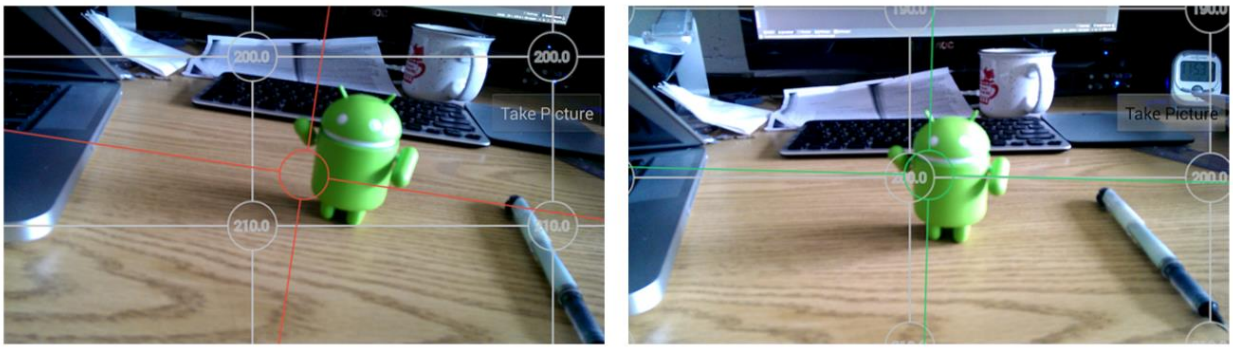


**Fig. 3: Visual indicators assist the photographer. The "Take Picture" button is disabled when the camera is tilted (left). It becomes enabled once that the camera is horizontal (right).**

An important element of the assisted image capturing App is a visual marker used to indicate whether a picture has been taken at a particular orientation (i.e. matching a certain alignment circle). If a picture has been taken, a camera icon–clearly visible in Fig. 2–will appear in the center of the alignment circle. Otherwise, the circle will be empty. The application uses the smart phone's internal sensors (magnetometer, accelerometers, and gyroscopes) to detect changes in the camera's pose. Typically, the pan and tilt values for each viewpoint—given the default focal length of the camera used—should not vary by more than 30°, and the distance to the object should change by no more than 30% from one step to the next. These settings can be adjusted for different focal lengths.

The user interface also provides instructions in the form of a text message at the bottom of the screen. The same instructions are also provided verbally using the smart phone's voice synthesizer.

### *Results*

We measured the consistency of the reconstructions obtained by several users documenting the same scene. In this particular study we were not as interested in the quality of the reconstructions, as much as we were in assessing the utility of the App to assist the photographers and its ability to produce consistent reconstruction results.

With the help of 5 persons who volunteered for this experiment, we collected five sets of images from a test scene staged in our lab. These participants had different levels of experience as photographers. We also collected LIDAR data from the same scene using a low-cost 3D scanner developed in our lab.

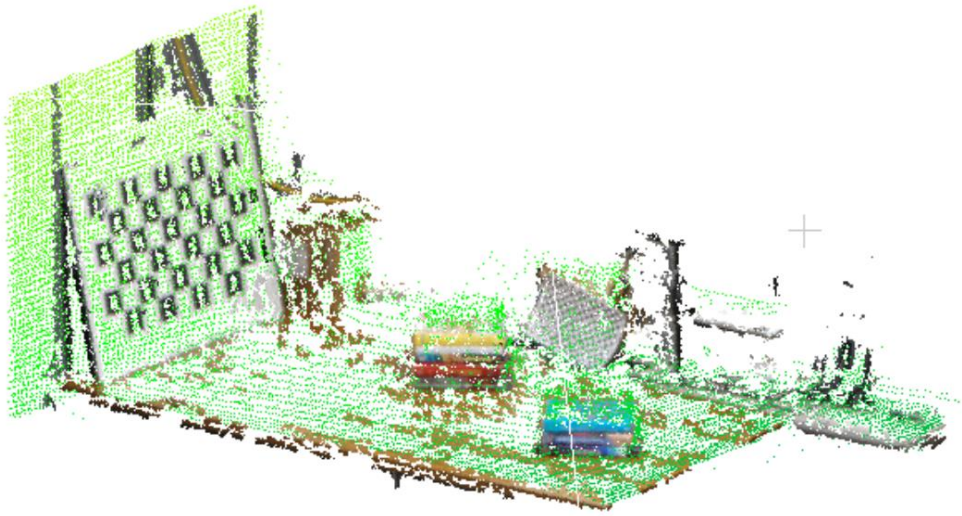

**Fig. 4: Sample data- the SFM reconstruction is compared against the point cloud ground truth. Green dots = ground truth, others = sparse 3D model from SFM.**

None of the test subjects had expertise in computer vision, SFM, or 3D modeling. The users' ages varied from 18 to 49 years old. All the participants used the App described before, after a brief demonstration of how to operate it. The participants instructed to start their collections at approximately the same location with respect to the scene. This was the only guidance provided by the researchers; the App

provided instructions about where to aim the camera, and also when to move after a certain number of pictures had been collected from a certain location.

The point cloud from the LIDAR was used as ground truth. We then obtained SFM reconstructions from each of the image sets to evaluate whether the 3D models derived from image sets captured by different people are of similar quality, using accuracy and completeness as performance metrics[3]. For each dataset from the individual subjects, we first carried out a dense reconstruction process using VisualSFM. The evaluation was done using CloudCompare. We started by aligning the reconstructed model to the scanned data. Next, we compared the reconstruction result to the ground truth for accuracy; finally, we compared against the ground truth to the reconstruction result for completeness. Fig. 4 illustrates this concept, where the ground truth is overlaid on top of the 3D model obtained from SFM.

The evaluation results are shown in Table 1. In terms of accuracy, the average mean distance d for the five individuals was 0.018 m. Similarly, the average completeness for five individuals was 43.80%. The results indicate that a consistent reconstruction of 3D models can be obtained through the use of the image capture assistance App.

**Table 1: Results from reconstruction experiment involving 5 test subjects.**

| | Subject 1 | Subject 2 | Subject 3 | Subject 4 | Subject 5 | All Images |
|---|---|---|---|---|---|---|
| Mean [m] | 0.023 | 0.02 | 0.018 | 0.014 | 0.016 | 0.016 |
| Standard deviation [m] | 0.02 | 0.025 | 0.013 | 0.01 | 0.012 | 0.012 |
| Completeness | 43.4% | 44.2% | 44.8% | 43.2% | 43.4% | 53.4% |

## Findings

Our study has investigated the potential of SFM techniques to be used and adopted as a regular tool for the documentation of crime and accident scenes. The main aspect to consider was the ability of these techniques to consistently produce acceptable reconstructions. Our results provide significant evidence that a consistent performance can be achieved through the use of a specialized App that guides the investigators during the image collection process.

This study has also shown that a successful 3D reconstruction using SFM still requires some ex-pertise to consolidate all the information into a single model which is both accurate and complete. However, our results also indicate that at least the data collection part can be ensured to be done properly. This is very important, since a crime or accident scene is only temporary, and it would be undesirable for investigators to have to return to it to acquire new data.

[3] Geometric accuracy measures how close the reconstructed model *R* is to the ground truth model *G*. Similarly, completeness measures how much of *G* is modeled by *R*.

## Participants

Luis E. Navarro-Serment, PI
Jinhang Wang, Visualization specialist

**Students:**
Venkatesh Manikavasegam
Andrew Orobator

## Deployment Partners/Participating Organizations

• Allegheny County Department of Emergency Services.
• Pennsylvania Turnpike Commission.
**Provide**: discussions with experts, recommendations on practical issues.

## Outcomes

We have developed a video processing pipeline to detect people from images, which is customized for operation with the type of cameras currently used to monitor vehicular traffic.
We have also developed an approach to calibrate traffic cameras on-site, which is inexpensive in terms of time and logistics; does not require expensive instruments or software packages; uses a low-cost custom-made laser scanner; and can be performed by personnel with minimal training.
Finally, building upon the calibration approach, we have developed a methodology to determine the location of a person in an image with respect to the geometry of the traffic intersection, and considering all the cameras covering the intersection.

## Other Products associated

The source code for the image capture assistant App is freely available from GitHub:

https://github.com/cmu-navlab/scene-recon-android