Integrating Videos with LIDAR Scans for Virtual Reality

Giang Bui*, Brittany Morago*, Truc Le, Kevin Karsch^, Zheyu Lu, Ye Duan

University of Missouri-Columbia

Abstract

LIDAR range scans can be used to quickly create accurate 3D models for virtual reality and as a basis to visualize sets of photographs, videos, and virtual objects in a cohesive environment. The number of existing virtual reality programs that use LIDAR data as input has motivated our group to develop methods for fusing images with 3D scans and for augmenting the scans with both dynamic objects present in videos and virtual models. Bringing together as many data sources as possible increases users' abilities to present related information in one, intuitive venue. We demonstrate how to register a variety of 2D imagery with a range scan to construct photo-realistic models and to extract walking people captured in videos and model them in a 3D space. We also present a method for determining the sun position from a set of stitched photographs in order to apply correct lighting to virtual objects placed amongst real world data. Naturally lit objects can be inserted into original photographs using our 2D-3D registration information. These methods are all combined to display and study photos, videos, and virtual objects in a complete 3D environment.

1 Introduction

A central component of setting up many virtual reality systems is acquiring models and data to display and visualize. Using LIDAR (Light Detection and Ranging) range scans as a basis for generating 3D models has a great number of benefits, including quickly creating realistic representations of scanned objects and easily presenting 2D and 3D information from a variety of perspectives [1-4]. We present methods for registering 2D imagery with range scans in order to produce photo-realistic renderings of buildings and foliage and creating a virtual reality system. This system includes walking people captured in videos and static and dynamic virtual objects. The data collection process is quite simple. The user only needs to scan a scene once with a range scanner and take photographs and videos of a scene. With this information we can build a full 3D virtual representation displaying all the photographs and videos at once.

Using LIDAR range scans for virtual reality can provide a venue for studying many types of information in one cohesive environment. Images and videos obtained at different locations and times can be registered with a single scan and viewed from a variety of perspectives, including ones not originally photographed. Virtual objects can be added into this information-rich environment to help visualize how new objects would fit in a real setting. For instance, new foliage and buildings can be displayed with a range scan for urban planning and architectural design. 3D representations of people from videos registered with a scan can also be displayed in a virtual plan. Studies have shown that models of moving humans help users easily ascertain the correct scale of a scene which can be very important for displaying architectural models. The ability to register different data forms and augment them with new objects can also be extremely useful for archaeological study. A site can be scanned and scientist's photos and videos can be registered with the scan to create a 3D photorealistic model. Virtual representations of objects discovered at the site can then be added into the 3D scene at the relative location where they were actually found.

* Joint first authors

^ Currently at Lumenous Co©

Email: duanye@missouri.edu

2 Methodology

We use a Leica C10 HDS LIDAR scanner which provides a high resolution point cloud of a scene and 2D images of the scanned subject using a built-in camera. The output data from the scanner also consists of files containing the internal and external camera parameters for each image. The photographs and videos were taken with a variety of cameras including a Nikon D80 and several smartphones. In order to register an image with the LIDAR range scan, we must calculate the camera pose for the photograph in relation to the 3D point cloud. This entails matching an image of interest to other images whose 3D correspondences are known, such as our LIDAR camera images, and solving for the camera's projection matrix. Our 2D-3D registration method can be divided into two main phases. First, we perform 2D image matching between LIDAR photographs and new photographs to be registered and try to obtain as many keypoints matches as possible. We then move onto our 2D-3D phase in which we solve each camera's projection matrix to align it with the range scan and use depth information to refine our 3D pose estimate.

3 Application

3.1 Displaying Videos in 3D Environments

Video frames can be registered with a range scan using the approach described above. However, when an object that was not scanned is present in a video, such as a person walking around, it will be projected onto an incorrect location in the 3D space because there is no structure that corresponds to it. Though the visual result of an image's registration may look fine when the scene is viewed from the camera location, these errors are very apparent when the user starts changing perspectives as is demonstrated in Figure 1 Bottom Left. In order to handle many of these cases, we propose segmenting the motion in videos and adding 3D planes to the virtual environment to "catch" the projection of these new entities. We present a simple but effective approach here for modeling moving objects that are touching the ground and recorded by even a single, stationary camera.

In order to identify moving objects in the video stream, we use the Mixture of Gaussians (MOG) algorithm. This yields a binary image with the motion segmented from the background as shown in Figure 1 Top Right. The connected components algorithm is applied to the MOG image to create cohesive segments. We scan this image starting from the bottom row of pixels to find the lowest points in each moving segment. We then identify the 3D points in the range scan that match to these low points when the video frame is registered with the range scan. Assuming that the moving object is touching the ground, these 3D points are the correct locations for the bottom of the segmented objects. New 3D points with the same depth as the bottom points and varying heights are created and projected onto the MOG image. If they fall within the segmented portion of the image, they correspond to a moving object that was not scanned and are added into the 3D space with the corresponding color information from the original video frame. The result of performing these steps can be viewed in Figure 1 Bottom Right.

3.2 3D Video Authoring

We can also insert dynamic virtual objects into the 3D scene that either follow the trajectory of an object/person captured on video or a user-defined path. For the former case, the object is placed using the same bottom point found for the motion segment in each video frame. If the user instead wants to specify a path, he or she can define control points on the 3D point cloud outlining the shape of the route. A B-spline is fit to the final set of points and the virtual object is displayed moving along this curve.

One additional step is required in placing the virtual object. We want the object to be oriented such that it is facing the direction in which it is moving. To accomplish this, we first assign an orientation for each frame by fitting a line through the object positions of the neighboring 50 frames (about 2 seconds worth of video). The orientation is set to be the angle of the line along the x-y plane. We include an averaging step to ensure that the object's orientation does not change drastically between frames. Figure 2 displays a virtual car in the 3D environment following a user-defined path. More examples of both of these types of video authoring are included in our supplementary material.

3.3 Realistically-Lit Object Insertion

The photo-LIDAR registration can also be used determine how to apply realistic lighting to a virtual 3D model and where to place it in photograph. We first estimate the sun location during the time of the scan and photo acquisition so that inserted objects can be lit realistically. We also must determine rough scene geometry to realistically insert new objects. Using the LIDAR point cloud and our segmentation information, the orientation of the ground plane and the camera location/height are known allowing us to cast shadows on the ground. Using the estimated lighting and geometry, we can insert synthetic 3D models into the photographs of the scanned environment after that have been registered with the point cloud. This allows us to transform our lighting/geometry into the image's coordinate system. Figure 3 shows stationary objects inserted into photographs using this approach.

4 Conclusion

LIDAR range scans can be used for displaying and navigating many types of imagery in virtual reality environments. Sets of images and videos taken in the same vicinity can be registered with a 3D scan and viewed together. Doing so allows the user to view 2D data from locations differing from where they were acquired. Visualizing images and videos in this manner also conveys the spatial relationships between camera locations in an intuitive manner since the user can see various images at once in a full 3D model of a scene. We presented a method for solving a camera's projection matrix and obtaining its relative pose to a 3D scan in order to accomplish this. By registering videos with range scans, we are able to create an augmented reality setup in which motion extracted from single videos can actually be displayed in its correct 3D location. We have also presented a method for extracting the sun position from a set of stitched images that have been registered with the LIDAR scan and inserting both static and dynamic virtual objects into the scene with correct lighting information. The combination of these methods for 2D-3D registration and augmenting range scans can be extremely useful for creating versatile virtual reality systems that use LIDAR data as a basis for 3D modeling.

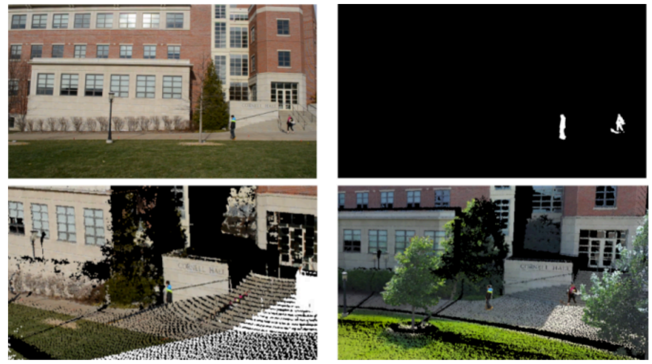

Figure 1: Displaying videos in 3D environments. Top Left: Original video frame. Top Right: Mask showing dynamic foreground (people) captured in video. Bottom Left: Video frame projected onto range scan without using our method for modeling moving objects. Right: 3D planes constructed for moving objects identified in video using our modeling method.

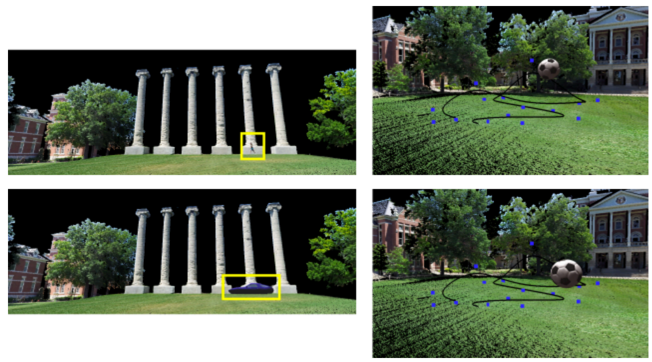

Figure 2: Dynamic virtual objects for video authoring. Left: Walking person captured in video shown in the 3D environment (highlighted in yellow) is replaced by virtual object. Right: Virtual object moving along user-defined path.

Figure 3: Virtual objects inserted into photographs with lighting from sun applied.

5 Reference

1. Bruder, Gerd, Frank Steinicke, and Andreas Nuchter. "Poster: Immersive point cloud virtual environments." 3D User Interfaces (3DUI), 2014 IEEE Symposium on. IEEE, 2014.
2. Nebiker, Stephan, Susanne Bleisch, and Martin Christen. "Rich point clouds in virtual globes–A new paradigm in city modeling?." Computers, Environment and Urban Systems 34.6 (2010): 508-517.
3. Zhao, Wenyi, David Nister, and Steve Hsu. "Alignment of continuous video onto 3D point clouds." Pattern Analysis and Machine Intelligence, IEEE Transactions on 27.8 (2005): 1305-1318.
4. Neumann, Ulrich, and Suya You. "Modeling and video projection for augmented virtual environments." U.S. Patent No. 7,583,275. 1 Sep. 2009.