Bilkent University

Department of Computer Engineering

Senior Design Project

MoveIt, Indoor Manipulation : 3D Semantic Reconstruction, Display and Manipulation System

Project Specifications

Barış Can
Faruk Oruç
Mert Soydinç
Pınar Ayaz
Ünsal Öztürk

Supervisor: Associate Professor Selim Aksoy, Ph.D.

Jury Members: Assistant Professor Shervin Rahimzadeh Arashloo, Ph.D.

Assistant Professor Hamdi Dibeklioğlu, Ph.D.

Project Specifications Report

October 14, 2019
This report is submitted to the Department of Computer Engineering of Bilkent University in partial fulfillment of the requirements of the Senior Design Project course CS491/2.

Contents

Project Specifications

MoveIt, Indoor Manipulation : 3D Semantic Reconstruction, Display and Manipulation System

1. Introduction

The popularity of smartphones rises tremendously with each passing day. While they have many uses, it can be argued that taking and sharing photos is the most widespread use of today's incredibly powerful smartphones. One thing that is common to all photographs whether they depict people or animals in them, whether they capture a moment in nature or an indoor scene is that they all have information of objects contained within them. This amount of information increases as one adds additional angles and positions for the camera that is capturing the original image. This way, great amounts of information about the 3D scene can be interpreted from just a small set of images.

In our Senior Design Project, MoveIt: Indoor Manipulation, we aim to bring our homes into the virtual reality by recreating 3D indoor scenes using the camera of the smartphone. With MoveIt, a quick scan of the room will bring all the household appliances into the VR where they can be freely moved and interacted with. MoveIt aims to be a helpful tool that enables the user's aesthetic ability by allowing them to redecorate their room with ease while also providing functional help with its realistic object boundary detection. This 3D recreation of the room will be fully visualized in a VR environment that enables the user to freely move in the room using a VR Headset in their Personal Computer.

1.1. Description

MoveIt: Indoor Manipulation aims to develop an application which will semantically reconstruct the geometry of indoor areas, such as a living room or a kitchen, in 3D by using videos, live scans/feed obtained from the camera of a mobile device. The 3D reconstruction process will produce a 3D scene and will label the reconstructed objects and geometry semantically by attaching a semantic label to each object. It will also allow the users of the application to manipulate the scene by allowing the users to move, rotate, scale, and deform the geometry of the objects. Texture obtained from the live feed will also be mapped to the corresponding meshes. The application will store the 3D reconstruction information along with the texture, label, and mesh information of a given object on a database for later data analytics and machine learning purposes, and will be open to the public as a labelled set of 3D objects, which may be suitable for supervised learning purposes.

The main purpose of the application is to allow the users to manipulate an indoor scene without actually having to add, move, or alter the objects in the scene. For instance, if the user wishes to furnish a room, the user is able to use this application to reconstruct the geometry of the room and the current objects within the room as a 3D scene in VR. The application will allow the necessary facilities to manipulate the positions, the rotations, the scales, and the geometry of the scene, while also allowing the user to place other objects in this scene. These new objects can be fetched from the database that the application keeps or may be supplied by the user in the form of a 3D mesh and texture. Objects can be queried and retrieved from the database using semantic labels. As an example, the user will be able to request a 3D mesh representation of a chair, and if available, the application will present a set of chairs that the user can add to the scene present in the database. A user will not be able to upload objects directly to the database, and user-provided objects will only be stored locally. Another relevant feature is to save a scene with a given set of objects, geometry, and the texture locally on a device for later access and modification.

Another purpose, though not the main purpose, of this application is to create a publicly available dataset of labelled 3D objects and the textures corresponding to these objects. This dataset might be used to study various aspects of 3D reconstruction, texture analysis, and the relationship between these and the semantic labels provided alongside the objects. The dataset may also be used for supervised learning purposes, and the more the app is used, the larger the dataset will be.

As for the platforms and the hardware, on which the application will be deployed, MoveIt: Indoor Manipulation will rely on many other frameworks and paradigms. Retrieval of live scans/feed will depend on Android infrastructure; the texture analysis, texture to geometry mapping, 3D reconstruction, semantic segmentation of the live scan and semantic labelling of meshes will be carried out on a remote server with sufficient hardware. The models used for the semantic segmentation of images will be based on pre-trained convolutional neural networks, and for the 3D reconstruction and texture mapping step, popular frameworks such as OpenCV will be used. Once a model of the scene is created, the user may choose to have this scene rendered on a smartphone, VR headset, or on a personal computer. The user will be able to control and manipulate the scene using the touch screen of a smartphone, VR controllers, or keyboard and mouse.

1.2. Constraints

In the next section, economical, ethical, implementation, feasibility and timeline constraints will be discussed.

Implementation Constraints

- Smartphones are not powerful enough to do the processing of the scan by themselves so a cloud computing service will be used.
- Semantic segmentation will be done on Python and the visual components of the application such as the movement of the objects will be done on Unity with C#.
- The mobile application will be developed on Android Studio with Java.
- The programming languages must be able to interact with each other.
- Github will be used for version control of the application during the development stage.
- We will use powerful pre-trained neural networks, such as GoogLeNet, ResNet or YOLO, for the semantic segmentation of the objects.

Feasibility Constraints

- The scanning devices will be Android smartphones which have at least one camera and preferably a Gyroscope with runs on at least Android 7.0. Otherwise, the scanning result might not be accurate enough to process [5].
- In order to implement a VR component to the project, we require a VR controller/renderer library.
- The depth and the object boundaries must be detected accurately enough not to cause any problems during the object manipulation stage.

Economical Constraints

- Microsoft Azure, a cloud computing service, provides 5000 free operations per month, however, if we need more operations we would need to pay around 5,6 TL per operation. Currently, we expect 5000 free operations to be sufficient [1].
- We are planning to use several libraries in Python such as Pytorch and OpenCV, which are open source.
- We do not own a VR headset that is necessary for the development thus, we plan to purchase HTC Vive VR headset which costs around 6800 TL [2].
- Visual Studio 2019 Community Edition, Spyder, Android Studio and Unity are freeware but we would need to purchase several assets from Unity Asset Store OpenCV for Unity ($95) and HTC Vive VR Controller Pack ($20) [3].
- The website is hosted on Github, which is free to use.
- The object meshes and their labels will be stored on Microsoft Azure data storage which costs 0.006TL per GB in a month. We currently cannot estimate the total cost, but we expect it to be 6TL per month at maximum.

Ethical Constraints

- Objects scanned by users will be stored in the database with their labels for further use by other users. However, these objects are to be stored anonymously without any user information.
- The application will follow the Code of Ethics [4].
- The users won’t have an account in the application so there won’t be any stored user data except the reconstructed objects if the user allows to do so.

Timeline Constraints

- The remaining reports which consist of Analysis, High Level Design, Low Level Design and the final report will be completed at least 1 week before the deadline.
- The 3D reconstruction part of the project which will only run on a PC will be done until February 2020.
- The VR implementation will be done after stabilizing the computer and the mobile phone parts which will be at the end of the April 2020.
- The final product will be ready in May 2020.

Health and Safety Constraints

- Since the user won’t be able to see around while wearing a VR headset, it should not make the user move around excessively to prevent injuries.
- The VR headsets cause nausea if the implementation is poor, therefore the implementation should ensure that it does not lead to any health issues.

Sustainability Constraints

- MoveIt: Indoor Manipulation will collect object meshes from willing users which means that the amount of data will increase as time passes. This would allow us to give better service with the provided data.
- Possible deals with furniture companies such as Ikea or Tepe, might provide us with their furnitures’ meshes or models. This would allow us to use them on our database, which in turn allow users to use them freely.

1.3. Professional and Ethical Issues

Since MoveIt requires scanned furniture of users' room that they want to change the appearance, we are planning to store every furniture scanned by every user in the database with their labels. However, these objects will be stored anonymously without any of their user information. This way, we will allow the use of non-owned furniture in the application while securing the users' privacy. The models and labels will not be shared by other users or third parties without getting permission in the case that companies want to use them for ads, and will just be kept in the database.

Required permissions such as accessing phone's storage and camera will be taken from users before using the application for the first time and the models will not be saved, or the camera will not open if they do not give the required permissions. The addition of the scanned models to the dataset will be asked after each scan. This would allow the user not to upload their specified models if they wish to do so. However, they will not be able to use the application if they do not give some permissions such as accessing the camera.

While constructing the datasets for meshes, copyright and privacy issues will be considered and for all purposes, the National Society of Professional Engineers' Code of Ethics will be followed.

2. Requirements

In this section, functional and non-functional requirements will be discussed.

2.1. Functional Requirements

This system has 4 main components, these are Mobile phone components, Computer Vision components, Cloud computing and VR Headset together with a personal computer. For smartphone, we will employ the Android API set for gathering the required data and passing it along to a centralized server. This server will utilize the Microsoft Azure cloud computing service for semantically segmenting the image into known objects and storing database objects. Computer vision component will take care of the semantic segmentation by employing either GoogLenet, ResNet or YOLO libraries that are already trained for this purpose. At last, a VR headset will be responsible for the 3D presentation of the original set of images by using the Unity VR Game Engine.

Mobile Phone Component

For mobile phones, the system's required functionalities include:

- Scan the contents of the indoor area by the use of its camera.
- Pass this information along to the cloud server.

Cloud Computing

The cloud server will provide the following functionalities:

- Provide enough computational power to computer vision algorithms.
- Provide the necessary storage for the already scanned household items.

Computer Vision

The system will use pre-trained neural networks that are specialized on semantic segmentation for labeling the reconstructed 3D environment of the room, which will follow these steps:

- Detect surfaces and objects in the indoor setting.
- Create titles and bounding boxes for the objects.
- Clearly, separate the boundaries of the room from the boundaries of the interior objects using semantic labels.

- Use certain assumptions in order to reconstruct the obstructed angles and views of the room and the objects.
- Create a detailed 3D map of the indoor environment with distinct separate objects.

VR Headset with a personal computer

In the final step, this component will provide these functionalities:

- Fetch the required room data from the cloud server.
- Recreate the room in 3D Virtual Reality with fully interactable objects.
- Place the user in the created room.
- Allow the user to create additional copies of the existing furnitures or simply copy - paste some parts of them to elsewhere.

2.2. Non-Functional Requirements

In the following subsections, usability, reliability, security, smartphone friendliness and availability will be discussed.

2.2.1. Usability

- The MoveIt application should have a user-friendly interface that would enable different users from various age and knowledge groups to use the application with ease.
- The application should include a user manual or an instructions page that would explain the users some key concepts such as how to use the application in general, how to record a room, how to move objects around etc.
- Users should be able to move around objects with natural controls that are easy to perform and comes as natural.

2.2.2. Reliability

- 3D reconstruction of the objects in a room must be as detailed as possible and the application should be able to recreate obscure areas of the objects with high accuracy by making correct assumptions.
- Different components of the application should work together harmoniously. To elaborate, MoveIt will have a smartphone application as well as a background side where 3D reconstruction will take place. Since the user interacts with the smartphone application, the background computation times should be optimized as much as possible to guarantee responsiveness.
- When the reconstructed objects are manipulated by the user (e.g. moved around), the previous place of the object as well as the new one should be altered and displayed correctly by the application.

2.2.3. Security

- The application should be able to protect the data that the user uploads to the system such as recordings since these data can be classified as personal.

2.2.4. Smartphone Friendliness

- The users are expected to use smartphones as an interface to interact with the MoveIt application. Therefore, the application should not be very taxing on smartphones and the power usage as well as the mobile data usage should be optimized to ensure user satisfaction.

2.2.5. Availability

- The system should be available for the users and function at all times. But since server failures are inevitable, we are aiming for a SLA level of 99.9%. (Yearly 8h46m of downtime)

3. References

[1] “Fiyatlandırma - Görüntü İşleme API'si: Microsoft Azure,” *Fiyatlandırma - Görüntü İşleme API'si | Microsoft Azure*. [Online]. Available: https://azure.microsoft.com/tr-tr/pricing/details/cognitive-services/computer-vision/. [Accessed: 09-Oct-2019].

[2]“HTC Vive Sanal Gerçeklik Gözlüğü,” *Hepsiburada*. [Online]. Available: https://www.hepsiburada.com/htc-vive-sanal-gerceklik-gozlugu-pm-HB000003KBLK. [Accessed: 09-Oct-2019].

[3] “OpenCV for Unity,” *Asset Store*. [Online]. Available: https://assetstore.unity.com/packages/tools/integration/opencv-for-unity-21088. [Accessed: 09-Oct-2019].

[4] “Code of Ethics,” *Code of Ethics | National Society of Professional Engineers*. [Online]. Available: https://www.nspe.org/resources/ethics/code-ethics. [Accessed: 10-Oct-2019].

[5] “Nougat,” *Android*. [Online]. Available: https://www.android.com/intl/en-gb/versions/nougat-7-0/. [Accessed: 09-Oct-2019].

4. Website

https://barisc22.github.io/MoveIt/