Augmented Reality Visualization of Building Information Model

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Yuchen Lai

Graduate Program in Civil Engineering

The Ohio State University

2017

Dissertation Committee:

Alper Yilmaz, Advisor

Lei Wang

Rongjun Qin
Abstract

Building Information Modeling (BIM) is an effective tool which widely used in construction industries. As its result, building information models (BIMs) serve an important role through project design, delivery, build and management stages, bringing many benefits. There are many reliable commercial BIM software on market using computers as their main platform. But the way they display the BIM and interact with the BIM are also limited by computers. On the other hand, Augmented Reality (AR), as a latest popular technique, shows a great potential of changing the way of people observing and interacting with the world. It provides a seamless way of combing virtual digital contents with the real world.

In this paper, we will discuss about the development of BIM and AR technique, and the possible benefits of combing them. In the last chapter we present an experimental system that is able to visualize BIM in AR. The results are demonstrated and the whole idea of our system can be served as a general framework of a wider range of AR-BIM system development.
Acknowledgments

First and most important, I would like to thank my advisor, Professor Alper Yilmaz for his valuable continuous help through my whole graduate study and the research of this thesis. The discussion with him greatly expand my horizon in my area and give me a right direction to my research.

I truly appreciate my committee members, Professor Lei Wang and Professor Rongjun Qin, for their precious time and suggestions on my thesis.

I’m very grateful to my lab mates Peng Liu, Changlin Xiao and Nima Ajam Gard for their help on my model and Hololens development. I also want to thanks my friend Jianming Xie for his suggestion of the Unity development.

Lastly, I would like to express my very profound gratitude to my parents and to my girlfriend for providing me with great support and encouragement throughout my years of graduate study in OSU.
Vita

2011.........................................................Shenzhen Senior High, China

2015..........................................................B.E. Remote Sensing Science and

Technology, Wuhan University, Wuhan,

China

2015 to present ........................................Graduate student, Department of Civil,

Environmental and Geodetic Engineering,

The Ohio State University

Fields of Study

Major Field: Civil Engineering
Table of Contents

Abstract ........................................................................................................................................... ii

Acknowledgments ......................................................................................................................... iii

Vita .................................................................................................................................................. iv

List of Figures ............................................................................................................................... viii

List of Table .................................................................................................................................. xi

Chapter 1: Introduction ................................................................................................................ 1

Motivations ..................................................................................................................................... 1

Building Information Modeling and its application ................................................................. 1

Advantage of Augmented Reality ............................................................................................... 3

Related Works .............................................................................................................................. 3

BIM System .................................................................................................................................. 3

Spatial Mapping ............................................................................................................................ 4

Augmented Reality Technologies ............................................................................................... 5

Objectives ...................................................................................................................................... 6

Chapter 2: 3D Modeling of Real World ......................................................................................... 7
Representation of 3D Model ......................................................................................... 7

Polygonal Modeling ........................................................................................................... 9

Basic Geometry .................................................................................................................. 9

Tessellation .......................................................................................................................... 10

3D Scanning Techniques .................................................................................................. 12

LiDAR Technique .............................................................................................................. 12

Structured Light Technique ............................................................................................. 15

Time of Flight Camera technique ................................................................................. 17

From Point Cloud to Meshes ........................................................................................... 18

Image-based Modeling ..................................................................................................... 20

Photometric Stereo .......................................................................................................... 20

Multi-view Stereo and Structure from Motion............................................................... 23

Chapter 3: Augmented Reality System ........................................................................ 26

System Components ........................................................................................................ 26

Display ................................................................................................................................. 26

Input Device ....................................................................................................................... 27

Tracking Device ................................................................................................................ 28

Image Registration Algorithm ....................................................................................... 28

Marker-based Method ....................................................................................................... 28
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Marker-less Method</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Pose Estimation</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Comparing with Virtual Reality System</td>
<td>33</td>
</tr>
<tr>
<td>Chapter 4: Augmented Reality Visualization of BIM</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic Principles of Hololens</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Main Components</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Spatial Mapping and Understanding</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Building Information Model Generation</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Manage Building Information Model in Unity</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Model Registration and Visualization</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Main Sources of Errors</td>
<td>47</td>
</tr>
<tr>
<td>Chapter 5: Conclusion and Future Work</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Future Work</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>54</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. Rule-based reasoning for assessing the impact of rule violation (Solibri 2010). 2
Figure 2. CSG and boundary representation (from paper [11]).............................................. 8
Figure 3. Polygon modeling of a human face (3D Modeling, in Wikipedia)......................... 10
Figure 4. Example of vertex-vertex meshes (by David Dorfman).................................... 10
Figure 5. A simple tessellation pipeline (by Romainbehar, in Wikipedia).......................... 11
Figure 6. Laser Scanner made by Topcon ................................................................. 13
Figure 7. Indoor model using LiDAR (Video and Image Processing Lab, UC Berkeley) 15
Figure 8. A simple structured light system (from slides of Guido Gerig, Univ. of Utah) 16
Figure 9. Hand model reconstruct using color-striped illumination (from paper [23]).... 16
Figure 10. Range Image of a human face captured by a ToF camera (Time-of-flight camera, in Wikipedia)........................................................................................................ 18
Figure 11. Delaunay triangulation ..................................................................................... 19
Figure 12. Comparison between classic photometric stereo and colored-light method. (a-c) are three gray-scale images used for classic photometric stereo reconstruction (d). Image (e) is the same object illuminated by three colored-lights and the only image used for reconstruction result (f). (from paper [31])................................................................. 22
Figure 13. (a) For calibrated camera, any construction must respect the angle between rays measured in the image. (b) For uncalibrated camera, reconstruction only respects the image points, which may not change even the object goes through a perspective transformation. (R.Hartley, A.Zisserman, Multiple view geometry) ........................................ 24


Figure 15. Head-mounted display in Microsoft Hololens (using optical see-through) .... 27

Figure 16. ArUco markers used in OpenCV toolbox ................................................................. 30

Figure 17. Feature matching result (Baggio, D. L. (2012). Mastering OpenCV with practical computer vision projects) ................................................................. 32

Figure 18. Feature matching result after RANSAC (Baggio, D. L. (2012). Mastering OpenCV with practical computer vision projects) ................................................................. 32

Figure 19. System pipeline ........................................................................................................ 35

Figure 20. Optical sensors in Hololens ..................................................................................... 36

Figure 21. Fundamental model (view a) ................................................................................... 38

Figure 22. Fundamental model (view b) ................................................................................... 38

Figure 23. Detailed Model (view a) ......................................................................................... 39

Figure 24. Detailed Model (view b) ......................................................................................... 39

Figure 25. Detailed Model (view c) ......................................................................................... 40

Figure 26. Detailed model (with HVAC) .................................................................................. 41

Figure 27. Hierarchical structure in Unity ............................................................................... 42

Figure 28. Rendering material in Unity ................................................................................... 43
Figure 29. Room Mesh generated by Hololens ................................................................. 44
Figure 30. Final result (a), three electricity switch. .......................................................... 45
Figure 31. Final result (b), electricity socket and Ethernet socket................................. 45
Figure 32. Final result (c), HVAC duct in the ceiling ...................................................... 46
Figure 33. Final result (d), HVAC duct in the ceiling ...................................................... 46
Figure 34. Final result (e), electricity line and Ethernet. ............................................... 47
Figure 35. Spatial mapping error caused by the floor's reflection and color (holes in black region) ............................................................................................................. 48
Figure 36. Wall mesh ........................................................................................................ 49
Figure 37. Distance distribution histogram (0-0.005m: 61.26%, 0.005m-0.01m: 26.49%, >0.01m: 12.25%) .................................................................................................................. 50
Figure 38. Registration error caused by navigation error. (a) is the result right after registration and visualization. (b) is after user 2 minutes’ walk in the scene. ............... 51
List of Table

Table 1. Mesh point to estimated plane .............................................................................. 49
Building Information Modeling (BIM) is a digital representation of physical and functional characteristics of a facility [1]. As its result, building information models can be used through a building’s whole life-cycle, with a great number of benefits. In order to have better understanding and management of these models, the visualization of building information models becomes a very important part of using building information models. Currently, BIM software on market mainly use computer-based visualizations. However, there is still a large human perception gap between the graphic model on computer and real world. The emergence of Virtual Reality and Augmented Reality techniques in recent years bring a new way for people to perceive virtual world, which means a possibility of having a better understanding of the digital contents and a more immersive experience. The objective of this research is to find a way to bridge the gap between building information model and real world.

Motivations

*Building Information Modeling and its application*

Traditional architectural design mainly uses paper as its supporting platform. Resulting in the one-dimensional loss when passing the massage. With the development of computer
techniques, 3D modeling technique like BIM extends the designing process into a three-dimensional process, helping people demonstrate and analyze the structure of a building in a much easier way.

Building life cycle refers to the design, construction, operation, demolition and waste treatment stage of a building [2], where BIM can play an important role. For example, at the beginning of construction stage, BIM can be used to carry out the model checking process, analyzing the relationship between different components. Automatic clash detection is a common operation in this process. BIM can also be used to estimate the project costs and classify structural discrepancy severity levels during the construction stage (see Figure 1). More importantly, BIM is able to reduce the information loss between different stages of the project, allowing different teams to add and share their information in a more convenient way.

![Figure 1. Rule-based reasoning for assessing the impact of rule violation (Solibri 2010)](image)

Another important benefit of BIM is its advantage in education field. For structural engineering students and architecture students, understanding the structure of an
architecture is a vital part in their learning process. Because the structure of an architecture contributes to many properties of the architecture, like safety, energy saving performance, etc. Comparing with 2D technical drawings, the result of BIM is able to display the structure of a building in a much more direct way, helping students understand the structure in an easier way. Besides, many BIM software are capable of doing different structural analysis for architecture models. This can be used to demonstrate the principles of architectural design and improve students’ analysis skills.

**Advantage of Augmented Reality**

Augmented reality (AR) is a live direct or indirect view of a physical, real-world environment whose elements are augmented (or supplemented) by computer-generated sensory input such as sound, video, graphics or GPS data [3]. Comparing to traditional display, AR provides an enhanced perception for users. It superimposes digital contents on the real world perceived by user, creating a more unified experience. AR also provides a more natural way for people to interact with virtual objects, which is just like people interacting with their environment in daily lives. Since AR techniques provide a new way to visualize virtual objects and interact with them, these techniques can be ideal for visualizing 3D models and help people study them in a more natural, immersive way.

**Related Works**

**BIM System**

The earliest concept of BIM system can be traced back to 1960s, when Douglas C. Englebart suggested a system with object design, parametric manipulation and a relational database [4]. However, this was just a conceptual framework since there was
no applicable system with graphical user interface at that time, which was essential to visualize and interact with a building model. During the late 1970s and 1980s, solid modeling programs began to appear, using constructive solid geometry and boundary representation as chief shape displaying methods. In 1974, Chuck Eastman successfully designed a project named Building Description System (BDS) that was able to retrieve and add information to a model [5]. It has a graphical user interface and is capable of storing geometric data and attributes. The system can also provide perspective and orthographic projections.

After then, several BIM systems were built and successfully applied to the field, such as GDS, RUCAPS, Reflex, etc. In 1987, a software named ArchiCAD became the first BIM software that was able to run on a personal computer. Since then, many BIM software are developed on personal computer. Some of them uses architectural drafting tools like AutoCAD to cooperate in the whole BIM framework. Some BIM software like Revit allow user to add a fourth dimension, time into their component. This time attribute helps users create a simulation of construction process.

Spatial Mapping

In order to build a model, we need the measurement of our object. Normally we can start building modeling by using architectural files like layouts and plans. Another option is to ‘measure’ the building by using different spatial mapping (or depth mapping) techniques. Especially for those old architectures which lack related files. Nowadays, spatial mapping methods like light detection and ranging (LiDAR, time of flight (ToF) and structural light are already capable of generating depth map in real time.
LiDAR and ToF methods solve distance between sensor and object based on the speed of light, recording the time of light travels or measuring the phase of light wave. Structural light methods project coded-light patterns onto 3D surfaces. By using the distortion of light patterns caused by the shape of 3D surfaces, the system is able to reconstruct the shape of 3D objects.

**Augmented Reality Technologies**

In 1968, Ivan Sutherland invents the first augmented reality system, which uses an optical see-through head-mounted display [6] [7]. The system uses two 6DOF trackers to track user’s head movement. One is a mechanical tracker and the other one is an ultrasonic tracker. In 1992, Tom Caudell and David Mizell mentions the word “augmented reality (AR)” in their paper [8]. They think the primary difference between augmented reality and virtual reality (VR) is that AR does not need to render every object that user sees, which means AR system can be deployed to inexpensive microprocessors with less cost of computational power. They also point out that AR systems require more accurate registration with real world. Ronald Azuma defines AR as systems that should follow three characteristics: combines real and virtual, interactive in real time, registered in 3D [9].

An AR system usually contains four parts: displays, input devices, tracking sensors and processors. With the fast development of smartphones, assisting MEMS sensors like GPS, accelerometers, gyroscopes and magnetometers on the phone bring the possibility of developing AR applications on it. Besides the smartphone based AR systems, many head-mounted AR systems are developed and become popular on the market in recent
years, including Sony’s Glasstron, Google Glass and Microsoft Hololens. Comparing with AR systems based on smartphones, head-mounted AR systems are more likely to provide an immersive experience, which help people perceive and interact with virtual objects produced by the AR system.

Objectives

Since building information modeling has so many significant benefits, it is important to find a proper method to visualize the models generated from BIM in a more user-friendly way. Modern architectural software like Revit, Sketch Up, AutoCAD already have some built-in visualization tools or renderers that are able to provide some simple visualization result. With the help of extra visualization software like Lumion and 3ds Max, user can generate a more realistic result. However, these methods are also limited by way PC displays the content and the way it interacts with users.

The objective of this research is to develop a system which uses augmented reality techniques to provide a new way to visualize building information models, breaking the gap between BIM context and real world. By taking the advantages of AR and BIM, we hope to demonstrate the system’s potential in various applications, such education, facility operation and maintenance, fast decision making, etc.
Chapter 2: 3D Modeling of Real World

With the rapid development of computer graphic technologies, 3D modeling becomes a quite common terminology when we mention about computer-aided design or manufacturing. 3D modeling refers to the process of developing a mathematical representation of any three-dimensional surface of an object via specialized software [10]. The product of this process is called a 3D model.

Representation of 3D Model

3D models are usually represented in two different ways: Boundary (Shell) Model and Solid Model (see Figure 2). Boundary model represents the object by representing the surfaces. It is a more complex model comparing with wireframe model, which represents an object by specifying each edge of the physical object. Comparing with wire frame models, boundary models are ideal for rendering and mass property calculation. It also can avoid the ambiguity caused by wire frame models. Boundary models are widely used in architecture, game and film industries, because of its conveniences and fully capability of demonstrating the external aspect of a designed object, and only a small amount of geometric data is enough to satisfy the needs.
Solid model represents the object by defining the volume of the object. It focuses on ‘informationally complete’ representations of solid [11] and physical fidelity [12], allowing its geometrical property can be calculated automatically. Solid models are usually built with constructive solid geometry (CSG). CSG uses Boolean operators to combine simpler objects, such as intersection, union and difference [13]. The simplest objects used in this process are called primitives. Cuboids, cones, cylinders, spheres, prisms, pyramids can be served as primitives, which can be combined with certain operations and generate an object with more complexity. Another advantage of CSG is it can provide a result of good accuracy.

Due to the unique property of solid models, they are widely used in engineering and medical tests and applications, because they can lead to a better demonstration of how the product will perform under real situation. Besides, since solid models are parametric, users can easily find out what changes they have made and go back to an earlier state.
Polygonal Modeling

Polygonal modeling is a very simple way to model objects by using polygons to represent the object’s surfaces. Thus it is ideal for creating building or room models which do not need a high level of complexity.

Basic Geometry

The basic elements in polygonal structure are vertices, edges and faces [14]. Edges connect vertices and form faces (polygons). Since any three non-collinear points can determine a plane, triangle is the simplest polygon that can represent a single plane. Although polygons are planar, we can use multiple polygons to approximate curved surfaces, like Figure 3 below. These small polygons which share edges are called polygon meshes. Polygon meshes can be represented in many ways, like face-vertex meshes, vertex-vertex meshes (Figure 4), winged-edge meshes, etc.

Another important attribute of polygons is the normal, a unit vector that defines which way a face or vertex is pointing. When visualizing polygon models, the normal of a polygon can tell whether the polygon should be seen by users. Normals are also very important elements in rendering systems. Most of them use surface normals, while some rendering systems support using vertex normals to create a better result.
Tessellation

The process of using multiple polygons to represent curved surfaces, is called tessellation, which generates polygon meshes. The most commonly used polygons in polygon meshes are triangles and quadrangles [15]. Comparing with quadrangles, triangles are easier to rasterize because their projection is always convex [16]. In most
situation, it is not practical to build meshes manually, many tools are developed for the purpose of mesh building. Box modeling is a very widely used tool which uses primitive shapes to generate a basic shape, which is used later to sculpt out the final model\[17\]. Box modeling has two main tools. One is sub-divide tool which divides faces and edges into smaller pieces by adding new vertices. The other one is extrude which creates a new face that connected to each edges of an existing face \[18\].

The common pipeline of tessellation are showed in Figuer 5.

Tessellation has several benefits. Firstly it can convert a complicated 3D model to a simpler polygonal model, which is easier to analyze. Secondly, it has a much lower computer memory and bandwidth cost, which allows the computer to render a detailed object in a very fast speed. It also supports multi-level rendering by changing the number of polygons in the mesh on the fly.

Figure 5. A simple tessellation pipeline (by Romainbehar, in Wikipedia)
3D Scanning Techniques

Constructing an object model manually requires priori knowledge about the object. Most of the time these priori knowledges come from technical drawings, or traditional measurement, for those objects that do not have technical drawings. 3D scanning techniques can be used to construct an object model without requiring any prior knowledge about object’s profile.

**LiDAR Technique**

Light detection and ranging (LiDAR) is a surveying method that measures the distance between the target and the system, by illuminating the target with a laser light. It is a fast 3D laser scanning technique with very high precision and resolution. Firstly, a pulse of light is emitted and the precise time is recorded. Then the reflection of that pulse is received and another precise time is recorded. The distance is obtained by using the equation (2.1) below

$$R = c \frac{t}{2}$$

where \(c\) is the speed of light (~299,792,458 meters/second), \(t\) is time interval between sending/receiving pulse. It is able to target a large variety of materials, from solid material like metal, rocks to non-solid material like aerosols.

LiDAR system usually has four main components: laser, scanner and optics, photodetector and receiver, positioning and navigation system [19]. Different wavelengths of lasers are used in different LiDAR systems according to the purpose of the system. For example, LiDAR systems used for bathymetry usually adopt 500-600 nm
lasers, since lasers at this range of wavelength penetrate water with less attenuation [20].

Near-infrared lasers around 1040-1060 nm, on the other hand, are mainly used in terrestrial mapping. Usually, lasers with shorter pulses can lead to a better scanning resolution. In addition, the maximum power of laser in LiDAR system is limited to satisfy the need of eye-safety. Figure 6 is a LiDAR product made by Topcon which has a rotated scanner.

Scanner controls the speed at which the resulting data can be generated. Different scanning methods are used for different purposes, like azimuth and elevation, including dual oscillating plane mirrors, dual axis scanner and polygonal mirrors [21]. For the simplest case, a rotating mirror is used to steer the laser beam thus it can scan around the scene. Angular resolution and measurement range of the system is mainly affected by the choice of optics. Position and navigation systems are usually included in LiDAR systems mounted on mobile platforms, like airplanes and cars. Global navigation satellite system (GNSS) will provide the location of LiDAR system, while Inertial Measurement Unit (IMU) provides the pose (orientation) of the sensor.

Figure 6. Laser Scanner made by Topcon
Advantages of LiDAR include a very high level of accuracy, capability of covering large areas in a relatively short time, highly automatic and low cost of turnaround. Also, by combining LiDAR with imaging system, target’s texture can be captured and mapped on corresponding 3D points, which can be used for a more realistic modeling. But for those materials that are too dense for lasers to penetrate, they may cause errors in the final model. Since the raw result produced by LiDAR is usually very large, it makes the result difficult to process and interpret.

The applications of LiDAR can be divided into two categories: aerial-based and land-based. Aerial-based systems are often used for producing digital terrain models (DTM), defoliation investigation, building segmentation, etc. Land-based LiDAR systems can be both static or mobile. Static platform is commonly used in surveying projects, with high accuracy but suffers from limited viewpoints. Mobile platform is when LiDAR system is mounted on a moving vehicle, coupling with GNSS and IMUs. It can capture 3D data from larger street sections with a fast speed.

Figure 7 shows the point cloud data of an indoor model generated by LiDAR.
Structured Light Technique

3D reconstruction using structured light is a process of projecting a known pattern onto a target object. A camera is set next to the projector to capture the pattern deformation caused by the shape of the target. Then a triangulation-based method is applied to get distance from the camera to the target, contributing to the surface information of the object (Figure 8). The patterns used in structured light can be binary or colorful, vary from simple stripes, fringes, speckles and grids [22].

Stripe pattern is the most commonly used pattern in structured light scanners. By projecting multiple stripes at a time, the total number of images required is reduced. In order to identify light stripes in captured image, coding techniques are added to the stripes [22] [23] (Figure 9). To increase capacity of codes, a sequence of patterns can be
used at the same static object successively, resulting in more detected 3D points and better result.

There are two major methods of generating stripe patterns: laser interference and projection [24]. Interference method uses two planar laser beam fronts, resulting in regular, equidistant line patterns. This method can produce very fine patterns but it is difficult to implement due to the nature of interference. The projection method uses a projector with a light modulator. Besides, digital projectors are usually adopted in many projects since they are more convenient and cost less.

Figure 8. A simple structured light system (from slides of Guido Gerig, Univ. of Utah)

Figure 9. Hand model reconstruct using color-striped illumination (from paper [23])
One advantage of structured light scanner is it can cover the entire scanning region at once. This can help reduce the errors caused by object’s motion. Another advantage of structured light scanner is that it is inexpensive comparing with some other laser scanners like LiDAR scanners. J. Bouguet and P. Perona propose a ‘weak structured light’ method which only uses a stick and a lamp to cast a moving shadow on object [25(Bouguet & Perona, 1998)]. Then the 3D shape of object can be reconstructed with an error less than 1%. However, structured light as an optical method, may confront difficulties with reflective and transparent surfaces.

*Time of Flight Camera technique*

Time of Flight (ToF) camera system can be considered as a scannerless LiDAR system. ToF calculates the distance by measuring the runtime of a traveled light pulse, or the phase difference between a reference signal and the reflected signal [26]. It captures the entire scene in a single shot, instead of using point-by-point scanning like LiDAR systems. A typical ToF camera system usually contains a light source like laser or LED, image sensor, optics focusing light onto the image sensor [27]. The operation rate of a ToF camera can reaches 120 images/second, which makes this technique ideal for capturing moving targets. It is widely used in many human-computer interfaces and gaming products (Figure 10). Since many ToF camera sensors use visible of near-infrared light, sunlight can be an error source to the system [27]. Even in an indoor scene where sunlight is largely suppressed, the pixel colors in the scene should still provide a relatively high dynamic range. Another disadvantage of ToF camera is that
since the system illuminates the whole scene at one shot, the light may travel along different paths, which cause the ‘multi-path effect’, creating an error in the calculated distance. This effect can be large when there are specular surfaces in the scene.

Figure 10. Range Image of a human face captured by a ToF camera (Time-of-flight camera, in Wikipedia)

From Point Cloud to Meshes

All the 3D scanning methods that mentioned above generate 3D point cloud as their result. A point cloud is a set of data points in a coordinate system. Although point clouds can be rendered with corresponding texture mapping, directly using them in 3D applications is still not applicable. Thus point cloud data is usually converted to polygon mesh or NURBS surface models. Delaunay triangulation is a common technique used to construct triangle mesh from point cloud.
Triangulation of a set of points $P$ in $\mathbb{R}^d$ is a simplicial complex that covers the convex hull of $P$, and whose vertices belong to $P$ [28]. If $n$ is the number of points in $P$, $k$ is the number of points in $P$ on the convex hull of $P$, then the triangulation of $P$ will have $2n-2-k$ triangles and $3n-3-k$ edges. According to Boris Delaunay’s work in 1934, a Delaunay triangulation for a set of points $P$ in a plane is a triangulation $DT(P)$ such that no point in $P$ is inside the circumcircle of any triangle in $DT(P)$ [29]. The minimum angle of each triangle in $DT(P)$ is maximized during this process, which means Delaunay triangulation tends to avoid sliver triangles (see Figure 11).

![Delaunay triangulation](image)

Figure 11. Delaunay triangulation

Delaunay triangulation has several benefits: It can largely avoid the potential numerical precision problems caused by sliver triangles [30]. Since the triangulation process won’t change the location of input points, the final model is able to preserve the precision. Also, the result of triangulation is independent from the sequence that points are processed.
However, the Delaunay triangulation is a relatively complex algorithm, it may take a very long time to process large data. An appropriate data preprocess becomes very important in this case, like noise removal and data down-sampling.

**Image-based Modeling**

Image-based modeling methods, not like those active reconstruction methods mentioned above, are passive 3D reconstruction methods. This kind of method does not interfere with the target, only capturing the radiance reflected or emitted by the target’s surface. Digital camera with CMOS sensor is a common choice for passive 3D reconstructions. It is cost effective and has the potential to be used for crowd-source data. But these image-based methods usually come with large computational cost.

*Photometric Stereo*

Photometric stereo is a well-established computer vision technique used for 3D reconstruction. Multiple images of one static object are taken under different illumination conditions, then the surface normals are estimated using the intensity variation of each pixel [31]. The object’s surface geometry can be obtained by integrating all surface normals.

Based on Woodham’s paper [32], early photometric stereo papers use Lambertian model as the reflectance model, which means uniform albedo. Albedo is the ratio of radiation reflected from the surface to incident radiation. A perfectly black object has an albedo of 0, while a perfectly white object has an albedo of 1. Lambertian surface model means the brightness of an object is the same no matter where the observer is at. This kind of assumption simplifies the calculation of photometric stereo, but limits its application.
Assume there is no ambient illumination, the brightness of a point $x$ on the surface is

$$B(x) = \rho(x)N(x) \ast S_1$$  \hspace{1cm} (2.2)

where $\rho(x)$ is the albedo, $N$ is the surface normal, $S_1$ is the light source vector. So the intensity value of pixel at location $x$ is

$$I(x) = kB(x) = k\rho(x)N(x) \ast S_1 = g(x) \ast V_1$$ \hspace{1cm} (2.3)

where $g(x) = \rho(x)N(x)$ and $V_1 = kS_1$, $k$ is a constant between camera response and input radiance. The equation can be solved using least-square solution.

One of the key limitation of classic photometric stereo is the object needs to maintain static when taken multiple images under different illuminations. For those moving or deforming objects, this is not very practical.

In recent years, some state-of-art papers like [31] [33] use colored-light photometric stereo to alleviate such limitation (Figure 12). The basic principle behind is, when a Lambertian surface is illuminated by red, green and blue light from different directions, it will reflect each of these light simultaneously and independently. The quantity of each colored-light reflected is a linear function of surface normal [31].
Figure 12. Comparison between classic photometric stereo and colored-light method. (a-c) are three gray-scale images used for classic photometric stereo reconstruction (d). Image (e) is the same object illuminated by three colored-lights and the only image used for reconstruction result (f). (from paper [31])

Comparing with laser scanning methods, photometric methods are much less expensive and easier to operate. Laser scanning methods may face difficulties when applying to reflective objects like porcelain. Photometric methods are able to handle this problem by using specular reflection model or bidirectional reflectance distribution function model.

There are two problems of photometric stereo that have been widely discussed, one is the shadow problem, the other one is the single-view problem. The easiest solution to shadow problem is increasing different illumination directions thus we can use the residual of least square solution to judge whether shadow has occurred. As for the situation that no spare observation is available, solutions to the shadow problem usually make use of the smoothness of the recovered shape, and the integrability of gradient field.
As for the single-view problem, many researchers are trying to solve this problem by developing multi-view photometric stereo methods. This is a quite challenging topic especially for those objects with a uniform texture. We will not discuss about these methods for now.

*Multi-view Stereo and Structure from Motion*

Multi-view stereo (MVS) is a well-established topic in computer vision area and it has been studied for a very long time. It computes the 3D geometric representation of an object by using multiple images taken from different views. Images can be obtained from different cameras on different location, or from multiple frames of a video sequence.

A typical multi-view stereo reconstruction usually contains following steps: Image acquisition, camera calibration, feature extraction and matching, triangulation. The camera calibration step can be carried out before the image acquisition step. It calculates the camera intrinsic (focal length, principle point, distortion parameter) and extrinsic parameters (rotation, translation), which are necessary to avoid the reconstruction ambiguity (see Figure 13). After image acquisition, feature points are extracted from each image and corresponding feature points are matched. For dense multi-view stereo, matches are iteratively expanded to points around them, leading to a denser set of matched points. Some constraints like homography geometry are used here to check matched points and get rid of the false matches. After setting up the point correspondences, 3D points can be obtained via triangulation.
While MVS assume the camera intrinsic and extrinsic parameters are known, structure from motion (SFM) constructs the object’s 3D model with unknown camera calibration. As we mentioned above, there will be reconstruction ambiguity if the camera calibration is unknown. Thus additional constraints are needed to recover the appropriate 3D shape. Let’s take affine SFM as example. Affine projection is the combination of linear mapping and translation in inhomogeneous coordinates. First, an affine shape and corresponding projection matrices are constructed by using at least two views. Then, the property of image axis (perpendicular to each other and of unit length) is used as extra constraint to recover the rigid Euclidean shape. Thus the affine ambiguity is reduced to a similarity ambiguity.

The advantage of SFM that it does not require preliminary calibration makes it capable of being used in a much larger scene. The scene model can be constructed using images.
taken by different devices (Figure 14). Usually a bundle adjustment method like [34] is used for refining structure and motion iteratively, at the goal of minimizing reprojection error.

Figure 14. Colosseum, reconstructed using crowd-source data. (Snavely, N., Seitz, S. M., & Szeliski, R. (2008). Modeling the world from internet photo collections)
Chapter 3: Augmented Reality System

System Components

Main components of an augmented reality system usually are displays, input devices, tracking devices, and processors.

Display

Usually three types of displays are used in AR systems: head mounted displays (HMD), handheld displays and spatial displays. HMD is a display device that worn on head or serves as part of helmet. According the number of displays in HMD, a HMD can be a monocular HMD or binocular HMD. Depending on the way how HMD “see” the world, HMD can be optical see-through or video see-through. Video see-through system first captures the real world using one or several cameras, then the video is transferred to the display inside HMD. This is useful when enhanced videos are used in the AR system, like thermal imaging and night vision imaging. Optical see-through system uses optical mixer which is made of partly silvered mirrors. Digital imagery can be projected onto the optical mixer, without blocking user’s view into the real scene (Figure 15).

Handheld displays are small displays combined with sensors and processors that can be hold in users’ hands. With the development of smart phone, it becomes a very popular platform for AR. However, the size of handheld displays are limited and users need to
hold the device in front of them all the time. Spatial displays usually use digital projectors to project digital imagery onto real objects. One characteristic of spatial display is that it is separated from users, making it capable of collaborating with different user groups easily.

Figure 15. Head-mounted display in Microsoft Hololens (using optical see-through)

**Input Device**

Input devices are devices that can help user input commands to the AR system, such as microphones, cameras, controllers. Microphones can be used for speech recognition and cameras can be used for hand gesture detection and recognition. Some AR systems use electronic gloves or wrist band as input controller. In addition, smart phones can also serve as input devices since they already equipped with different kinds of sensors.
Tracking Device

Tracking devices in AR system include digital cameras, global navigation satellite system (GNSS), inertial measurement unit (IMU), accelerometers, solid state compass, gyroscopes, wireless modules, etc. GNSS provides the geographical location of the system, while other wireless modules like Wi-Fi and Bluetooth can provide relative location in the indoor scene. Pose estimation is a crucial part in AR system, it provides the necessary geometry information for correct projection of virtual imagery. Errors in pose estimation may not only cause wrong projection, but may also cause user’s adverse reaction like dizziness for head-mounted AR system. Pose estimation and head movement tracking can be accomplished by using IMU and accelerometer.

Image Registration Algorithm

Since the main purpose for augmented reality is to enhance the reality with virtual content, it is important to make sure that virtual objects are correctly registered to the real scene. This can help users view the virtual content as part of the real world. Correct registration can be obtained by estimating the pose of the camera (for video see-through) or user’s view point (for optical see-through). The registration process usually consists of two parts. In the first part, fiducial markers or feature points are detected, using marker-based methods or marker-less methods. Then the second part estimates the pose and map 3D virtual objects through proper projective geometry.

Marker-based Method

Markers used in computer vision are patterns that can be detected with relatively simple algorithm in a very short time. They are widely used in applications that relative pose and
position between camera and object are required. Circular and square markers are the most common types in computer vision, because their geometric property makes them easy to detect and more robust than other shapes. Marker-based method generally can be divided into two steps, marker candidate detection and marker identification.

In marker Candidate detection step, algorithms search for the region that matches the perspective distortion of the marker geometry. For square markers, this means search for matching quadrilaterals, and for circles this means ellipses. Since most markers only contain white and black, images are converted into grayscale images. Then binarization and edge detection are usually used as preprocess to help extract candidate regions.

For the marker identification step, the perspective distortion of candidate regions are often removed by homography transform. Depending on the way that markers are encoded, there are several ways to identify markers. Correlation-based markers are correlated with reference grids. This kind of markers can be identified by calculating the correlation coefficient. For digital-code-based markers, marker ID can be directly read from the marker. Usually the marker ID is made of binary digits. There are different ways for encoding the marker, some divide the marker into square subregions (Figure 16), some use triangles pointing at several direction, and some use encoding techniques based on fourier transform like [35]. As for the topology-based markers, marker identification relies on topoligocal adjacency. Region adjacency graph is computed using the binary image. Topology-based markers are robust to geometry distortions, but the topology itself can be changed by occlusion.
Figure 16. ArUco markers used in OpenCV toolbox

Marker-less Method

Marker-less method is based on the feature extraction and recognition in the real world. It may be less accurate than marker-based method, but it does not need any marker to be placed ahead. It helps the AR system detects the scene in a more natural way.

In the feature extraction step, the goal is to find areas of interest in input image that can be served as unique and reliable markers. There are lots of feature detection and extraction algorithms based on different single or combination of features, such as edges, corners, blobs, ridges. For example, corner-based detectors focus on the rapid change of image gradients, using the first or second derivative of gradients along different directions.

After features are detected, a small image patch around the interest point (or area) will be extracted using specific feature extraction algorithm and feature descriptors will be generated. Feature descriptors are important for the following feature matching part.

Since relative position between camera and the object is always changing, it is necessary
to use descriptors that are invariant to scale and rotation, such as SIFT and SURF. In order to achieve rotation invariant, the direction of dominant gradient in the feature point region usually serves as the main direction. Then feature descriptors are computed in scale-space to achieve scale invariant.

Not like marker-based methods, in which markers have their own ID, marker-less methods find corresponding point by feature matching process. Firstly, a distance function that compares features is defined. Then, for a given feature in image A, use the distance function to test all the features in image B. The one with minimum distance can be considered as the corresponding feature point. There are many kinds of distance functions, like Sum of Squared Difference (SSD), Sum of Absolute Difference (SAD), etc. Besides the exhaustive search method that mentioned above, there are many other search methods using different searching strategy. For example, Flann-based matcher uses a fast approximate nearest neighbor search algorithm.

To improve the matching result, false matches can be removed by certain filter algorithms. One popular method is using the random sample consensus (RANSAC) method. RANSAC estimates the homography transform matrix between matched feature points, using only part of the feature points. Then it calculates the reprojection error of feature points using current homography transform matrix and mark inliers and outliers. After several loops, the algorithm picks the transformation that has most number of inliers and remove the outliers (see Figure 17 and 18). The final homography matrix is generated by using least-square method with inliers.
Figure 17. Feature matching result (Baggio, D. L. (2012). Mastering OpenCV with practical computer vision projects)

Figure 18. Feature matching result after RANSAC (Baggio, D. L. (2012). Mastering OpenCV with practical computer vision projects)
Pose Estimation

In order to place a virtual object in 3D, we need to know the relative pose between the object and the camera (or user’s view point). Then we can calculate the object’s corresponding 2D projection from user’s view. The general equation of such transformation is:

\[ x = K[R|t]X \]

where \( X \) is a point in 3D space, \( x \) is the 2D projection of \( X \) on image plane, \( K \) is the camera intrinsic matrix, and \([R|t]\) is the extrinsic matrix (rotation and translation). Camera matrix \( K \) can be obtained by camera calibration, or we can use the structure from motion method mentioned above. Then, with the location of detected markers of feature points, pose \([R|t]\) can be computed using simple least square method. Once the pose is estimated, virtual objects can be easily projected to the 2D image plane.

Comparing with Virtual Reality System

Not like Augmented Reality, Virtual Reality (VR) puts user in a totally immersive virtual world with head-mounted device. Although using different techniques, both AR and VR systems aim at enriching people’s experience of perceiving the world. Many of these devices have been used in industries like entertainment, education, architectural design, etc. VR is ideal for the situation that the application needs a simulation of completely different environment, like pilots doing flight training on the ground. AR can be used for the situation that the real environment is a necessary part to the application, like those AR assisting systems used in surgery. In our case, AR is more ideal because only by
combining the real environment with BIM can we realize the potential benefits that mentioned above.

As for the experience of current products on market, VR devices are able to meet most customer’s expectation slight more, comparing with AR devices. Most VR devices on market can provide satisfying immersive experience for users. The common expectations of current VR products are higher resolution and eye tracking capability. On the other hand, due to the high computational cost of AR algorithms, such as depth mapping and feature tracking, current AR devices are still trying to find a balance between system performance and physical size. Taking Microsoft Hololens as example, as an AR devices, it shows a very impressive ability to map the environment and overlay virtual objects on the environment. But its field of view (FOV) is largely limited and the whole device is much larger than normal glasses.

The price of VR products on market ranges from $50 to $700, while AR products range from $1000 to $3000. Due to the high price of current AR products, their customers are usually labs and big companies, instead of normal people. Since VR creates a fully immersive environment for users, it is more likely to be mainly used in indoor scenes. In our application, it is important to enable user walking in a certain are with the system on, which is another reason why we choose AR. But the ultimate goal of AR products should be small enough that can be hidden in normal glasses or other daily wears, and can be used in public.
In this chapter, we design an experiment using Microsoft Hololens to implement the augmented reality visualization of BIM. The proposed idea of our experimental system is demonstrated in Figure 19.

Figure 19. System pipeline
Basic Principles of Hololens

*Main Components*

Hololens use an optical-see-through head-mounted display, with a resolution of 1268*720 per eye. Processing components contain a 32-bit CPU, a custom-built Microsoft Holographic Processing Unit (HPU) and 2GB RAM. As for tracking sensors, Hololens uses IMU as its main tracking sensor. The input devices of Hololens contain four microphones, one ambient light sensor, one two-million-pixel photo/video camera, one depth camera and four environmental understanding cameras (Figure 20). Using these input devices, Hololens supports many human understanding functions such as voice recognition, gaze tracking, gesture input, etc.

Figure 20. Optical sensors in Hololens
**Spatial Mapping and Understanding**

The most outstanding feature of Hololens is its capability of mapping the environment through its depth sensors. Firstly, a depth map is generated by its depth camera. Then the processors generate corresponding mesh grids according to the depth map. By using the spatial mapping result and images taken by environmental understanding cameras, Hololens can rebuild the entire environment and locating itself in this environment.

When users try to interact with the environment in Hololens, it is important to analyze the environment and understand it. The mesh grids generated by spatial mapping can be used to help Hololens understand the environment. According to the geometric properties of meshes, such as area and topology, object detection stage can be performed to find continuous planes, small objects, etc. By using surface normals from meshes, we can divide planes into floors, ceilings, walls.

**Building Information Model Generation**

In this experiment, we chose our lab as test site. Since currently we don’t have building information model for our lab, we choose to create a simple model of our lab by manually measurement, for the sake of demonstration of our method. We choose SketchUp as the model editing software because it is easy to learn and operate for new users.

In the first step, we create the fundamental model of the lab, including the floor, ceiling, walls and columns (see Figure 21 and Figure 22). Then specific components are added to the fundamental model to make a detailed model (see Figure 23-26). In this experiment, we choose electric lines, HVAC, Ethernet cables as detailed components for
demonstration. The ceiling of following models are hidden in figures for a better demonstration.

Figure 21. Fundamental model (view a)

Figure 22. Fundamental model (view b)
Figure 23. Detailed Model (view a)

Figure 24. Detailed Model (view b)
Figure 25. Detailed Model (view c)
Manage Building Information Model in Unity

For visualization purpose, it is necessary to have appropriate rendering settings for the room model. Rendering settings include many features like viewpoint, texture, lighting, shading, reflection, etc. Rendering software performs the calculation of rendering process and generate corresponding 2D scene according to the 3D model. Since Hololens
supports developing apps in Unity, we choose Unity as the platform to perform rendering process.

Unity is a cross-platform game engine used to develop games for PC, mobile devices and websites. It uses C# as its main scripting language. The first step is model input. Unity accepts .fbx file and .obj file as 3D model input files. So we use Sketch Up to export our lab model to .fbx file. After the model is imported in Unity, group operations can be performed to keep different model objects in an appropriate hierarchical structure. This can be helpful for applying different rendering materials to different classes of objects in the following steps.

![Hierarchical structure in Unity](image)

**Figure 27. Hierarchical structure in Unity**

For demonstration, we set the walls to be half-transparent and use solid colors for other components: green for HVAC, red for electricity, blue for Ethernet. A typical setting panel of rendering material is showed in Figure 28. Each class of objects only needs one
rendering material. In this stage, none of the model is instantiated. It will be instantiated in the next model registration step.

Figure 28. Rendering material in Unity

Model Registration and Visualization

In the model registration step, the spatial mapping function of Hololens is used to retrieve the 3D information of our scene. In this experiment, only one wall in real scene is used for registration purpose. Firstly, the wall is scanned by Hololens. Then the depth map is converted to polygon meshes (Figure 29). Higher mesh density can lead to a better result,
but may increase scanning time. Every mesh vertex has corresponding 3D coordinates in the virtual world coordinate system in Unity.

Figure 29. Room Mesh generated by Hololens

After the polygon meshes are generated, we can obtain 3D information we need for registration. In this experiment, we adopt a plane fitting module provided by Hololens-Unity toolkit. The plane fitting module analyzes the existing meshes and fit plane game objects in meshes. Since the result of plane fitting is a game object, it is easy to obtain its 3D information by using Unity APIs. Once we obtain the location and rotation quaternion of the plane in the virtual world coordinate system, the room model can be instantiated and registered in the virtual world. Registered model will be automatically rendered
because rendering materials have already been set in Unity. In our experiment, we use red for electricity lines, blue for Ethernet and green for HVAC (Figure 30-34).

Figure 30. Final result (a), three electricity switch.

Figure 31. Final result (b), electricity socket and Ethernet socket.
Figure 32. Final result (c), HVAC duct in the ceiling.

Figure 33. Final result (d), HVAC duct in the ceiling
Main Sources of Errors

In this experiment, there are three main sources of errors. The first one is the model error. The room model is created from the data of manual tape measurement. This measurement is a low precession data source, comparing with the high precession building information model used in professional construction teams.

Another error source is the registration error, which is also the initiation error in the system. In this experiment, registration error can be caused by spatial mapping step and plane fitting step. Although Microsoft does not reveal the detail information of Hololens’ depth camera, it is widely believed that it is a time-of-flight (ToF) camera. The
disadvantage of ToF camera is that it can be influenced by the background light and multi-path reflection in the scene, causing errors in the depth map (Figure 35).

Figure 35. Spatial mapping error caused by the floor's reflection and color (holes in black region)

To analyze the system noise of ToF camera, we take the partial mesh result of a wall (Figure 36), and estimate corresponding plane equation using a RANSAC-based least square method. Then the distance from every mesh node to the estimated plane are calculated. The distance distribution is showed in Figure 37, mean distance and standard distribution are showed in Table 1. Since Hololens only provides mesh data to developers, instead of point cloud data from the depth map, the conversion from point cloud to polygon meshes can also introduce error in the final meshes. As for plane fitting
algorithm, it takes the meshes as input, thus these errors can be passed to the final result of plane fitting.

Figure 36. Wall mesh

<table>
<thead>
<tr>
<th>Number of points</th>
<th>Mean distance</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1363</td>
<td>0.0048 m</td>
<td>0.0047 m</td>
</tr>
</tbody>
</table>

Table 1. Mesh point to estimated plane
The navigation system in Hololens which use IMU and visual as navigation solution can lead to locating error. One main disadvantage of IMU is that it has an accumulated error along time. After the model registration and visualization step complete, if no extra algorithm is used, the IMU will be the only sensor that tracks user’s movement. When the accumulated error in IMU increases, it is very likely that user will find the model are not registered to the real world as before (see Figure 38).

Figure 37. Distance distribution histogram (0-0.005m: 61.26%, 0.005m-0.01m: 26.49%, >0.01m: 12.25%)
Figure 38. Registration error caused by navigation error. (a) is the result right after registration and visualization. (b) is after user 2 minutes’ walk in the scene.
Chapter 5: Conclusion and Future Work

Conclusion

In this paper, we analyze the advantage of using AR visualization on BIM and propose an implementation method by using Microsoft Hololens. Comparing with the traditional way of visualizing BIM on computer, our experimental result shows that BIM can be displayed in a more direct and understandable way by using our proposed method. It demonstrates a potential of improving the manipulation efficiency of BIM and expanding BIM’s user group, from current project managers to field workers. The idea of our method is also applicable to other AR systems under different circumstances. Since the AR products on market are still at the beginning stage, we highly believe that the incoming fast development of such products in the future will bring more possibility and application to our proposed framework.

Future Work

In this research we successfully developed an AR visualization system that can be used for building information models. However, there are still some issues need to be studied and some modification that can be accomplished in the future, to make the result better.

(1) Dynamic registration
To handle the IMU accumulated error and registration error in the system, developing a dynamic registration algorithm can be useful. Comparing with one-time registration, a dynamic registration can dynamically update the relative 3D correspondence between AR device and real scene during the scene roaming. One possible way of doing this is to implement a matching algorithm that can dynamically match the BIM to the room mesh that are currently in front of user. But matching two kinds of 3D data could be a challenging task.

(2) Model Improvement

To achieve a better result, a professional construction-level BIM should be used as source data. Its high precision can be used for registration accuracy evaluation. Besides, a professional BIM also comes with large amount of information that can be utilized in a later application stage.

(3) Interaction function

AR technique not only provides a new way of visualizing virtual contents, but also provides a new way of interacting with them. In our experiment, we only focused on the visualization part as demonstration. The next step should be allowing user to interact with BIM. The interaction can take multiple inputs, like users’ gaze, gestures and voice input. With the interaction with BIM, specific information in BIM can be displayed on request. Such as object detail, operation guidance, diagnosis information, etc. It can also allow user to edit the BIM by using interaction function.
References


