Doctor of Philosophy

Automated BIM Data Generation Using Drawing Recognition for Construction Projects

August 2020

Department of Architecture & Architectural Engineering
The Graduate School of
Seoul National University

Juhee Rho

Abstract

Automated BIM Data Generation Using Drawing Recognition for Construction Projects

Juhee Rho
Department of Architecture & Architectural Engineering
The Graduate School
Seoul National University

The number of building information modeling (BIM)-applied construction projects has been increased with an expectation that it will shift labor-intensive industry into knowledge-based one. And as of 2018, many countries including Korea and UK made it mandatory to apply BIM data in the construction projects. Despite its advantages of applications that facilitates the continuous utilization of project information in terms of 3D shape and attribute values in one project's life cycle, a lot or workers and managers at the design and construction stage remain to work based on 2D drawings. Duplicated project information creating task, results from this gap

between the practical and the systemic situation, affects a negative impact on construction projects, that have limited cost and time.

Therefore, this research develops a framework for transforming project information of 2D drawings into an object model including drawing recognition and text-line extraction. In order to use the numerical values of quantity from the created model at the construction stage, the modeling specification that includes relationship between building components and priority between object types, which are based upon estimation standards, are established. To confirm the performance of the framework, text classification based on the Bayesian filter and object model generation are experimented. In addition, the applicability to the construction project is evaluated with the case experiment conducted with the generated BIM model targeting an office building and further quantity calculation with three different methods. The analysis results show the high accuracy of object model generation as well as more accurate computation to the current work approach. It is expected to the results of this research with the future technical development will improve the work process of data creation and further improve the productivity of construction project.

Keywords: Construction Management, Building Information Modeling (BIM), Text Classification, Automated BIM Data Generation, Quantity Take-off (QTO)

Student Number: 2017-37092

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Research Background and Motivation

Building information modeling (BIM), which first appeared in the 1990s as a form of three-dimensional computer-aided design (3D CAD), has been applied to construction projects since the late 2000s with improvement in computer performance (Hardin and McCool, 2015). Since then, the application of BIM has expanded, and, as of 2018, some countries such as Korea, Singapore, Hong Kong, and Australia have made it mandatory in construction projects (Mao, 2017). Being semantic-based and object-oriented, the application of BIM allows users to gradually accumulate project information, and use it continuously in management of construction projects for various purposes (Jung and Kim, 2016; Succar, 2009; Eastman et al., 2008). Furthermore, its implementation in construction projects influence more on the way to generate project information as well as to improve business process than adopting new technology (Howard et al., 2008). Accurate creation and utilization of data through BIM improves a project's productivity by enhancing design quality and further contributes to reducing wastage of material, resources and costs (Holzer, 2013; Azhar, 2011; Eadie etal., 2013).

The degree of BIM application is defined by the purposes for which it is used in the construction project and the detail level of the model created. Based on the purposes, BIM was initially used for visualization such as 3D design review and coordination, and its expanded use is data transformation with values of time, cost, and environmental review (Lee et al., 2017). In domestic(Korean) and overseas construction projects, BIM has been adopted for different purposes: in domestic projects, BIM is adopted mainly for transitional design (2D drawings to 3D BIM model); whereas, in overseas projects, the focus is more on improving design quality and constructibility (Koh, 2011).

Researchers have found that these differences are due to the efforts made by the academia as well as the outcome of the government's support programs in the early stage of BIM introduction (Park et al., 2009; Kang et al., 2007). For instance, in the United States where more than 90% architectural firms utilize BIM in their projects(as per 2009), the initial concept of a building description system and the specific descriptions were formed by researchers in 1970 and 1986 (Eastman et al., 1974; Aish, R. 1986). Since then, the technical development and the government's support was complemented by 50 pilot projects. This resulted in further development of BIM guidelines and also encouraged its application in construction projects (Kang et al., 2007).

Unlike in the case of foreign countries, the application of BIM domestically began with major construction firms, who encountered it while servicing the requirements of foreign motivation and lack of any systematic standards, application of BIM in domestic projects did not produce effective outcomes. Furthermore, due to the nature of the design phase in domestic construction projects that depend on subcontractors and do not encourage builders to participate in many cases, the gradual preparation of data BIM is difficult (Eadie et al., 2013; Choi, 2013).

Consequently, contrary to the expectations of the construction industry, data management of construction projects remains either in its infancy stage of visual application focused on design transformation or in a former phase of 2D drawing-based management. This has resulted in a conflict with the system, which has mandated the creation and use of BIM data. It has further led to problems such as BIM data creation only for project delivery submission, and the duplication of same information in 2D drawings and 3D models.

1.2 Problem Statement

Problems in Practice

According to the annual BIM utilization survey, 69% construction companies in United Kingdom are applying BIM to projects (NBS, 2019) and 59% of all domestic architectural firms, construction companies, and clients employ BIM to building projects (Lee and Lee, 2019). Among them, 49% of the respondents answered their adopting BIM to the extended application of quantity take-off (QTO) and cost estimation. As for the difficulty of BIM application and the limited specific areas, respondents indicated two fundamental causes: the first is an experts' inclination to work in conventional 2D drawing-based work methods (data creation and quantity calculation) and the second is data inconsistency in terms of modeling criteria and level of detail (LOD; incompatible data to QTO standards).

(1) Duplicated Task for Project Data Generation

To meet the national construction standards, design information recorded in the form of 2D drawings must be converted into 3D BIM data through manual modeling based on the drawings. According to experts, the time required to create a BIM model that has an LOD sufficient for 4D simulation (Boton et al., 2015) is about 5% of the total construction period.

As projects in the construction industry repeatedly suffer from delays and cost overruns (Enshassi et al., 2015), this time-consuming and repetitive data-writing task may further contribute to the problem.

(2) Non-utilization of Numeric Values from BIM Data

The advantages associated with the use of BIM include the ability to utilize data with expanded are of time and cost by automatically extracting and transforming data created with visual and geometric information. However, the data is not utilized in construction projects (Lee and Lee, 2019) because of the differences between the modeling criteria and QTO standards, which causes further time-consuming issues in the construction phase.

Limitations of Previous Efforts

The aforementioned problems can be addressed by developing a method for automatic generation of BIM data using 2D drawings. Some researchers suggested the creation of 3D shape data such as wall and room space using drawing and pattern recognition (Kim et al., 2009; Lu et al., 2005), while others focused on acquisition of attributes of quantity information, including area, through an automated simple model generation (Lawrence et al., 2014).

However, these research studies mostly focused on only one aspect of BIM—form creation or property acquisition—and, therefore, the created data

has limited use in actual construction projects. Even with recent research studies that facilitates object modeling through drawing recognition, the data generated can only by used for specific work trades or components (Gimenez et al., 2016; Janssen et al., 2016); therefore, automated BIM modeling that works regardless of the work trade remains unresolved.

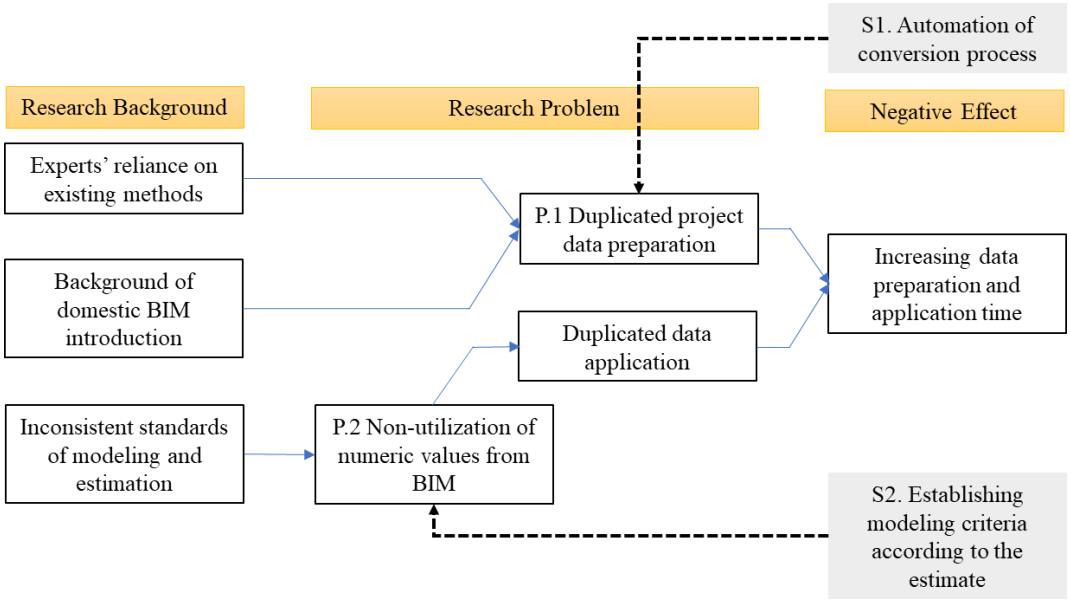


Figure 1-1. Description of research problem

1.3 Research Objective and Scope

This dissertation aims to address the aforementioned problems by developing an automatic BIM generation framework based on drawing recognition for generating data that can be utilized in the construction stage. To be specific, the framework includes the overall process, together with the theoretical description and technical methods, for converting the information in 2D drawing into 3D object model. In addition, a modeling criteria based on the relationships (priorities and deduction rules) between objects that accord with the estimation standards is defined so that the automatically calculated quantity from the project data completed with the set of generated objects can be utilized in the construction phase.

To achieve this goal and to develop this framework, the specific objectives are as follows:

(1) To generate an object model, including shape and properties, through recognition and conversion of line and text information in drawings that works regardless of the work trades.

(2) Establishing modeling criteria suitable for quantity calculation standards that can be used in the construction phase.

(3) Confirming feasibility and project applicability of these based on

experiments both at the object and project level.

By identifying the combinations of line and text in design drawings and extracting them into separate layers, an object model is created. To be specific, the line information is used as the planar shape of the object and the text information is employed for attributes such as its building components. The framework describes the overall process to create an object model through drawing recognition and the technical definitions in each stage. This study also identifies techniques applicable to the framework and highlights their strengths as well as potential synergies. With an aim to meet the QTO standards, this dissertation established the modeling criteria, which includes the reference level, level of modeling detail, and the priorities between objects. The proposed specification is verified by comparing the results with the drawing-based quantity calculation method.

The main scope of this research is automated BIM data generation for construction projects with 2D drawings, specifically regarding data conversion from 2D drawings to 3D object model as well as establishing relationships between object models to complete the project model. From the perspectives of design development, the 'detailed design' or 'construction document' phase during a design stage is the detailed stage of the project information to be targeted. Through the process of drawing recognition, line-

text extraction, text classification, and finally pattern recognition, a project data created in 2D drawings can be converted into an object model with 3D shape and attributes. Structural drawings prepared for construction projects are applied as case studies to confirm the proposed method. Finally, through implementation of a case project application, we explain the feasibility as well as the applicability of the proposed framework and the significance of this work.

The approach of creating 3D objects based on shape features and acquiring attribute information enables BIM data creation regardless of the construction work type of the status of design development. Compared to manual model creation, it can help improve the accuracy of project data by eliminating information loss and modeling errors. Further, it can enhance the efficiency and effectiveness during construction management through utilization of quantities that are automatically extracted from the project BIM data.

1.4 Dissertation Outline

This dissertation is structured in seven chapters (as described in Figure 1-2), and Figure 1-1 describes the process of employed to achieve the research objectives introduced above.

To begin with, as research background, the importance and necessity of a new project data delivery method using BIM is described, and then its difficulties in practical uses are indicated in this chapter. After explaining the challenge involved in application of BIM to construction projects, the ultimate goal and ways to address them are illustrated as following:

Chapter 2, Preliminary Research, examines the use of BIM in construction phase in the present scenario. It also examines material quantity calculation based on whether BIM data exists or not. An in-depth review of current research on automated 3D model generation is also conducted, specifically from a perspective of data extraction, data-utilized modeling, and the integrated process of these two. Finally, research hypotheses are established to address the knowledge gap that is unsolved problem by the existing body of knowledge.

Chapter 3, Research Methodologies, introduces the research framework and methods adopted to develop the research model that is required to explain the hypotheses introduced in the previous chapter. The research methods include information acquisition and utilization approaches for creating object model as well as the experimental models and result verification methods applied in object- and project-level experiments.

Chapter 4, BIM Data Generation Framework, describes the overall process of data transformation and the required technical methods for drawing-based BIM data generation. The framework is explained separately at the object level (which automatically converts drawing information created at the design stage into information of the object model) and at the project level (which describes the relationship between the types of object models that meet the quantity calculation criteria).

Chapter 5, Object-level Experiment, applies the proposed automated object modeling framework to a structural model and a duct model. Two separate experiments have been described to evaluate the proposed framework: for a text classifier using text resources from sample drawings and for object modeling using line from sample drawings.

Chapter 6, Project-level Experiment, explains the applicability of the

proposed framework to the case project and the significance of the work. Six complex building projects are employed for further application with modeling and further quantity calculation. Both conventional work method and the proposed method are implemented for case projects and the results are compared.

Chapter 7, Conclusions, summarizes the overall findings and discusses the implications, contributions, and limitations of this work. It also presents the possible directions for future works.

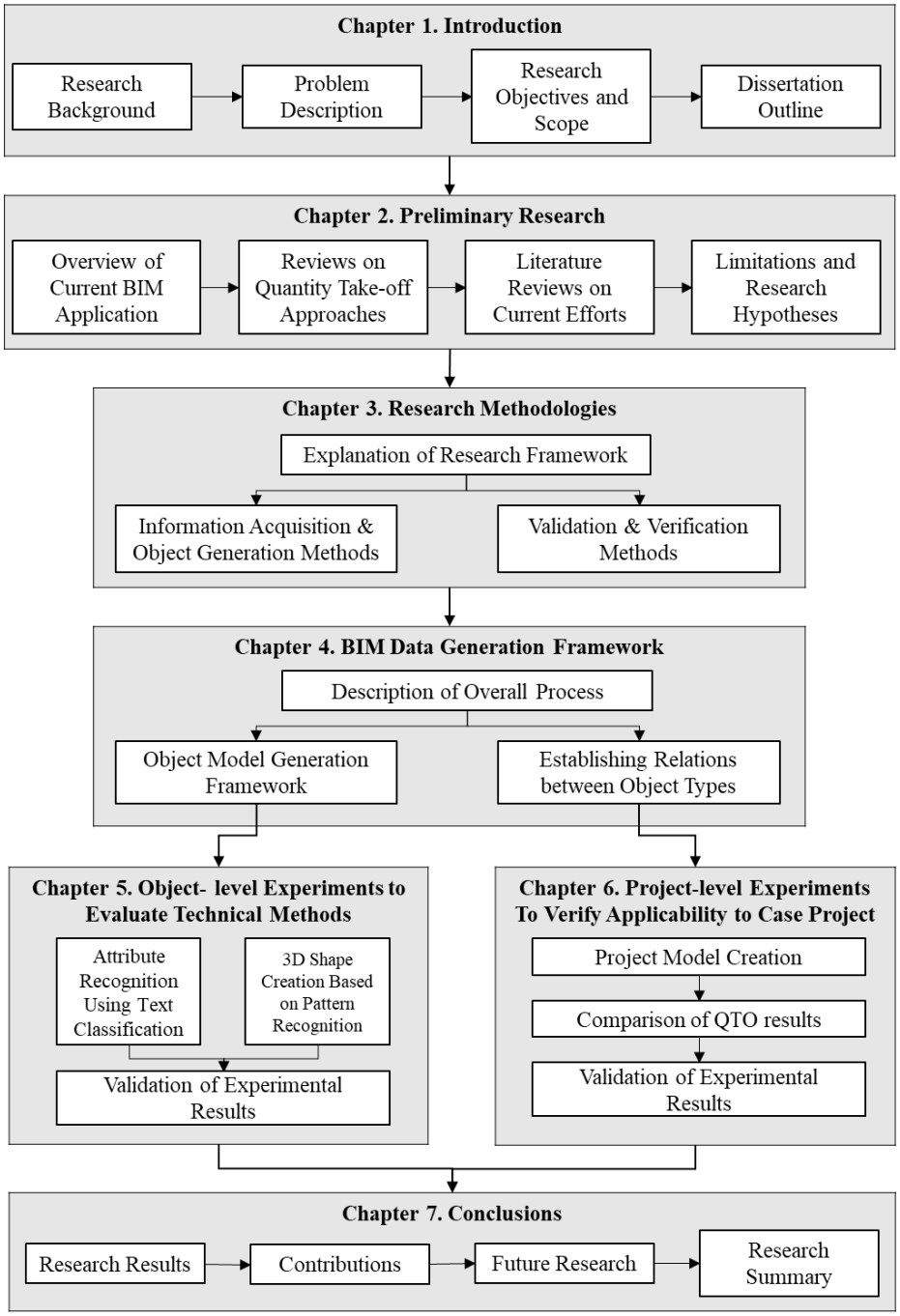


Figure 1-2. Dissertation outline

Chapter 2. Preliminary Research

The advantages relate to use of 3D data in construction projects have been identified and discussed in many previous research studies. For instance, Eadie (2013) quantitatively explained that the collaborative approach used as a result of BIM introduction improves project performance in the design and construction processes (Eadie, 2013). Due to the reluctance of key personnel involved in the project data generation and its application, duplication of tasks, wherein 3D models are prepared from 2D drawings sets has continued. In the construction stage, in particular, quantity calculation and cost estimation using BIM data is hardly deployed.

This chapter describes the difficulties in employment of this BIM-based project data and the current efforts to create and use it in an automated manner. To enable the extracted quantities from project BIM data in construction phase, the traditional project calculation methods are investigated; we then establish the 3D model preparation standards that can calculate the project material quantities automatically. After introducing the considerations involved in model creation, previous research studies that have focused on automated model generation are described.

The relevant methodologies—from traditional research works to the

recent advanced ones—are investigated. Their findings can be divided into two types: information extraction-based approaches and automatic modeling approaches. After an analysis of the technical merits and limitations using existing related research reviews, the research hypotheses are established. These assumptions include a framework with the overall process and techniques for object model generation based on drawing recognition, and relationships (priorities and deduction rules) between the objects to be considered for quantity calculation.

2.1 BIM in Construction

BIM represents the development and use of computer-generated 3D models in a building project's lifecycle. The result is a data-rich, object-oriented, and intelligent model created in a parametric manner that meets the needs and aims of various users (Azhar et al., 2008). To be specific, a BIM model is composed of the shapes and dimensions of the project; the space made from a relationship between objects; and non-geometric attributes such as material, zoning, and construction order (Gimenez et al., 2006).

In spite of the advantages of BIM, a preference for traditional methodology continues to lead to creation and utilization of 2D-based data (Lee et al., 2014). Hence, the mandatory creation and use of BIM data in construction projects has resulted in duplicated data creation; one 2D and the other 3D. The process of construction management utilizing BIM is shown in Figure 2-1. Figure 2-1 explains the application methods for BIM data according to the process of construction projects. As mentioned previously, 2D drawing based project data generation at the design stage requires BIM data generation task to be at the beginning of the construction stage. This leads to a duplicated data generation process, thereby leading to more time and efforts (expressed in the figure as gray solid fille).

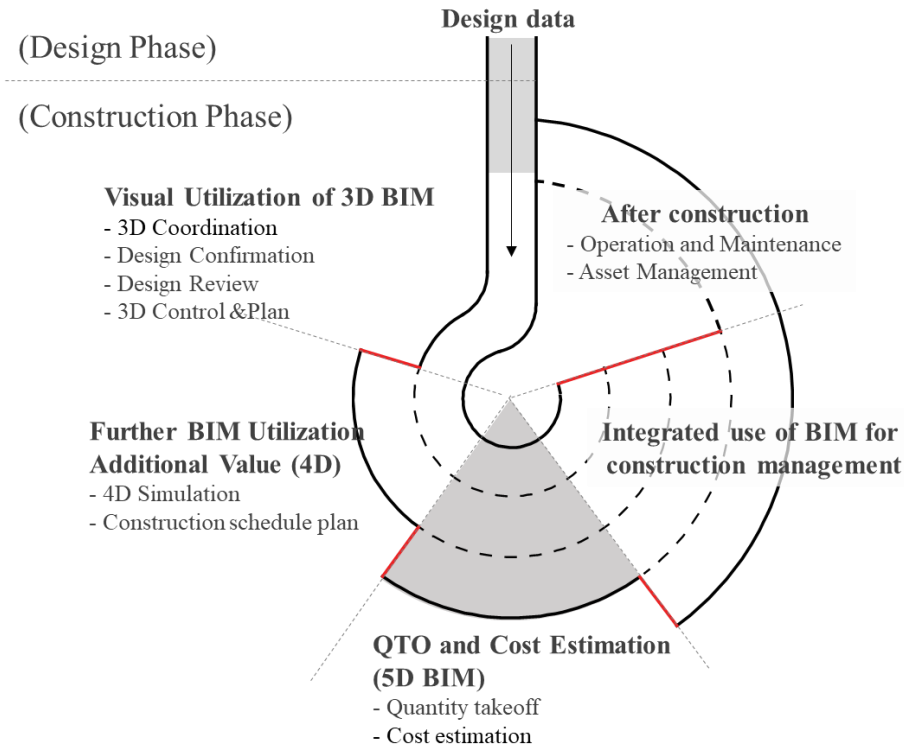


Figure 2-1. BIM application in construction management

Table 2-1 shows the modified results of a domestic survey conducted in 2019 (Lee and Lee, 2019). According to these results, visual utilization of design review, 3D coordination, 3D control and plan are the most actively applied application of BIM data during the construction phase. They are followed by quantity calculation, cost estimation, as-built model creation, code verification, and asset management. This is because of the technical difficulties as well as the professional knowledge required in data

modification (for its further utilization), which leads to concerns regarding return on investment and, subsequently reluctance to employ them.

However, when BIM data preparation is done correctly, the quantities can be estimated accurately as exact as the planned project without any separate calculating process (Lee et al., 2011). As there are no potential errors/omissions resulting from a lack of drawings or the estimator's assumptions as in drawing-based calculation process, it is convenient and the estimated results are accurate (Kim et al., 2009). In the following section, we consider modeled objects that are automatically generated using drawing recognition so that the quantities for the construction phase can be calculated accurately.

Table 2-1. The priority of current BIM application in construction project

BIM Applications	Priority Ranks (Rate of response, %)	Perceived Level of Difficulty (Rate of response, %)
3D Coordination	1 (87%)	15 (8%)
Design Authoring	2 (86%)	16 (4%)
Design Reviews	3 (62%)	14 (9%)
3D Controls and Planning	4 (56%)	12 (13%)
Cost Estimation	5 (49%)	7 (27%)
Site Utilization Planning	6 (33%)	12 (13%)
As-built Modeling	6 (33%)	11 (16%)
Digital Fabrication	8 (22%)	9 (24%)
Engineering Analysis (structure, energy)	9 (20%)	7 (27%)
Exsiting Condition Modeling (Laser scanning)	10 (18%)	10 (20%)
Building System Analyis	10 (18%)	4 (33%)
Building Operation and Maintenance	12 (13%)	2 (38%)
Code Validation	13 (11%)	3 (34%)
Sustainability Analysis	14 (10%)	4 (33%)
Disaster/Refuge Planning	15 (8%)	4 (33%)
Asset Management	16 (4%)	1 (42%)

2.2 QTO Methods and Requirements

2.2.1 Types of QTO Processes

2D Drawing-based Estimation Process

The traditional method of estimation in the domestic construction industry follows the method of classifying the numerical results obtained from 2D drawings by cost item for further cost estimation (Joo and Jun, 2009). The QTO and subsequent cost estimation processes based on BIM data and 2D drawings are shown in Figure 2-2. The distinguishing feature of the 2D drawing-based process is the role of estimator in terms of his/her involvement and his/her subjective judgement. In particular, drawing quality check and accumulation of construction items need to be performed before calculations can be made by estimator. The accumulated construction items also need to be re-arranged as cost items for cost estimation. In case there is insufficient project data from the drawing set, the estimator assumes the final set of items to arrive at an estimate.

BIM-based Estimation Process

However, the BIM-based quantity calculation process excludes the steps involved between modeling and QTO. Specifically, the quantities are automatically calculated using the object model and therefore, the accuracy

of the model and the standards of creation are important to meet QTO criteria.

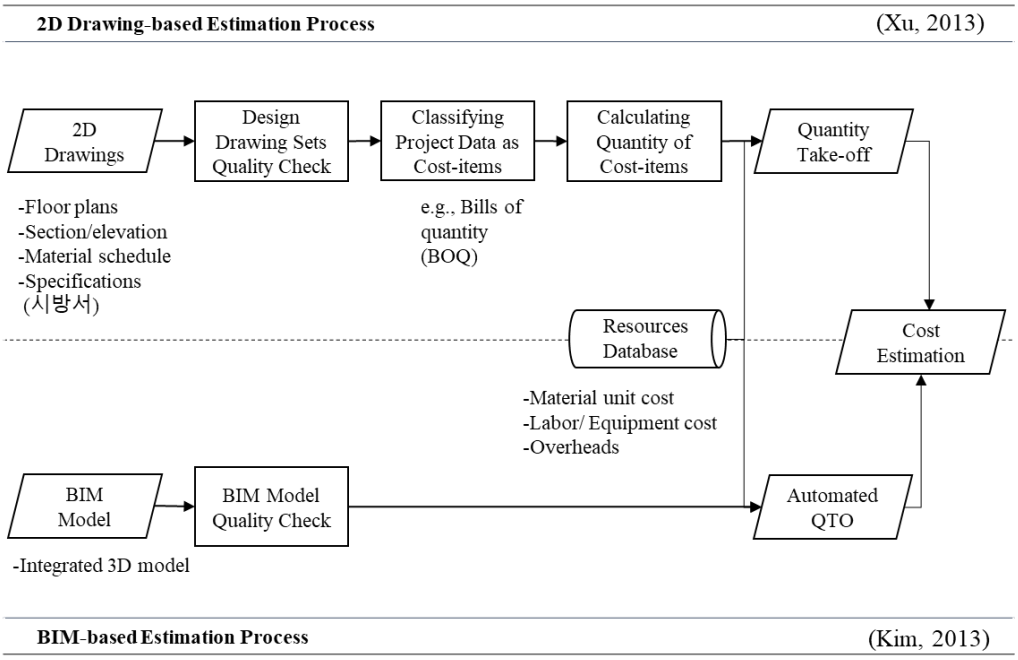


Figure 2-2. QTO and cost estimation process

2.2.2 Types of BIM-based QTO

To relieve the inconvenience of the traditional calculation process and to utilize information from BIM data, a BIM-based estimation method was proposed. There are three types of estimation methods depending on the calculation purpose and objects: finishing material table-based estimation, material take off, and schedule QTO (Joo and Jun, 2009).

The first method, using interior finishing material table, calculates only the interior materials. There are two approaches towards this method: one is to create all finishing material (single or complex) as object type and then complete the 3D model; and the other is to utilize room information of the completed BIM data by entering material information as room attributes.

The material take-off method is applied to building materials that require total volume, such as concrete or steel. By defining the modeling criteria for each object types and material, different results can be derived. Lastly, the schedule QTO is adopted when the number of objects, such as openings or mechanical electrical and plumbing (MEP) elements, is needed. Here, total number or length of each type can be calculated.

2.2.3 Required LOD for QTO

The abbreviation LOD was first adopted by the buildingSMART group and it describes the required accuracy of the BIM model at each stage of a construction project, with the level of detailed shape information and the level of detailed attribute information included (Adami et al., 2017). Table 2-2 compares the level of information required from LOD, building information level (BIL) that is suggested by Korean Public Procurement Service, and Singapore LOD promoted by Building and Construction

Authority (BCA) of Singapore. The level starts from 100, the conceptual design phase, and ends at 500, the as-built model.

In this dissertation, the steps from 200 to 400—from schematic design stages to construction drawings—are employed. Originally, LOD 200 defined the model detail at the schematic design, 300 was for developed design, and 400 was for construction document level. However, to the need for an additional level between 300 and 400, LOD 350 has been recently added to explain the detailed design stage. These same phases are indicated as 20 to 50 in BIL.

Although the level for construction documents is defined as LOD 400, from the specific contents of model requirements, it is assumed that LOD 350 or BIL 40 has sufficient information for QTO.

Table 2-2. Comparison of LOD, BIL, and Singaporean LOD

	LOD		BIL		BCA's LOD	
	Model / Estimation					
Schematic Design	200	- Graphic model with a generic system - Approximate quantity information	20	- Graphic model with main structure components without material info. - Approximate data for project review and conceptual estimate	2	- Graphic model with general building components - Approximate size and quantity
Design Development	300	- Graphic model as a specific system - Exact quantity as designed (size, shape)	30	- Graphic model with entire structural components - Quantity of elements, size, location and orientation for LCC analysis and decision making	3	
Detailed Design	350	- Graphic model with a specific system with assembles - Quantity data for objects as well as support or connection after coordination	40	- Graphic model with the entire structural and architectural elements - QTO	4	- Graphic model with specific elements - Quantity data of size, shape, location
Construction Document	400	- Graphic model with sufficient data and accuracy for fabrication	50	- Graphic model with specific data for construction - Quantity information for cost estimation and procurement management	5	- Graphic model with specific data for construction

2.3 Research Studies on Automated 3D Modeling Using Drawing Recognition

Automated model generation is based on the prerequisite that project-related information already exists, either drawings or existing structure of buildings, and that these can be obtained through data conversion. Therefore, the methods of drawing recognition to extract information or knowledge from existing drawings and the methods of transforming the extracted information are important. Many research studies and reviews have covered the topic of automated model generation, and they can be classified into two types based on the basic information used: as-planned model and as-built model creation.

As-built model creation has attracted considerable attention of researchers with the development of related technologies such as unmanned aerial vehicles (UAVs) (GIM international, 2018; Kwak et al., 2014), laser scanners, and point cloud data (Bosche and Haas, 2008; Brilakis et al, 2011; Aryaici and Gamito, 2006). Whereas due to expected decline of project information in the form of 2D CAD drawings in the future, as-planned model creation has been relatively ignored (Barki et al., 2015) and studies on it have focused on either specific technologies or on specific work trades like reinforced concrete or interior finishing work (Kim et al., 2009; Changsoft

I&I, 2019).

However, as mentioned previously, the use of 2D drawings remains common in the current scenario and, therefore, as-built models must be prioritized. This section reviews the previous research studies conducted on automatic as-planned model creation and the existing approaches on drawing recognition and information extraction (Figure 2-3, Table 2-3).

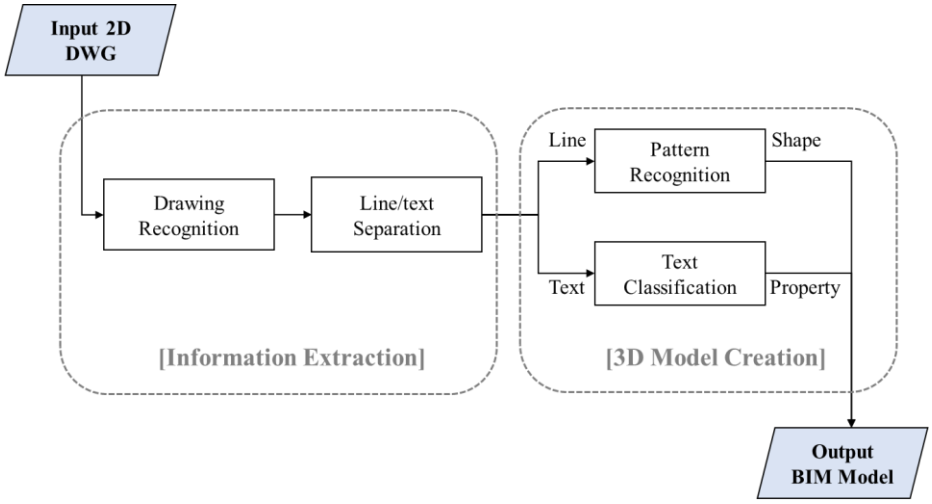


Figure 2-3. Process of converting 2D drawings into BIM model

Table 2-3. Existing approaches to information extraction and automated as-planned model creation based on 2D drawings

Researchers	Objective	Information Extraction and Separation	Model Creation	Process of Conversion from Drawings into 3D Model
		Method	**Method**	**Range of Application**
Fletcher and Kasturi (1998)	Automated text/graphic separation from documents	Hough transform	X	X
Tombre et al. (2002)	Segmenting the text and graphic layer from images	Heuristics rules with text bounding box	X	X
Zhang and Lu (2004)	Identifying image techniques for image retrieval	Pattern recognition/ Generic Fourier Descriptor (GFD)	X	X
Hoang and Tabbone (2010)	Text extraction from graphical document	Morphological component analysis (MCA) and post process	X	X
Pan et al. (2007)	Text extraction from complex background	Sparse Signal Representation and MCA	X	X
Lewis and Sequin (1985)	Generation of 3D model based on 2D plans	X	Contour extraction by following the borders	X

Researchers	Objective	Information Extraction and Separation	Model Creation	Process of Conversion from Drawings into 3D Model
		Method	Method	Range of Application
Kim et al. (2009)	Automated interior modeling for and quantity take-off	X	Determining the shape of rooms and model interior as composited material	X
Bredif et al. (2013)	Automatic solid model generation in city-level	X	Digital surface model generation by extracting contour lines and extrusion	X
Ahmed et al. (2011)	Automated floor plan analysis by finding outer walls	Text/graphics separation and optical character recognition (OCR) for space analysis	Contour extraction by, and polygonal approximation to get outlines	X
Barki et al. (2015)	Automatic conversion of 2D CAD into BIM models	Acquisition of priori information from CAD drawing	Simple 3D modeling using existing software	X

Researchers	Objective	Information Extraction and Separation	Model Creation	Process of Conversion from Drawings into 3D Model
		Method	**Method**	**Range of Application**
Changsoft I&I (2019)	Automatic BIM modeling for reinforced concrete	Drawing recognition and acquisition of location information	Modeling using rules of placement	Rebar modeling by recognizing rebar placing drawing
Gimenez et al. (2016)	3D model generation from 2D plans	Feature extraction for separation and OCR for text analysis	Pattern recognition (Hough method) and segment analysis	Wall and enclosed space modeling using scanned drawing
Proposed	Automated BIM model generation from 2D drawings	Text-line extraction for separation and the Bayesian filter for text analysis	Pattern recognition and linking acquired knowledge (textual) to individual object	Automated BIM object modeling based on information from digitalized drawing

2.3.1 Drawing Recognition and Information Extraction

Prior to the modeling process, project-related information (such as line and text data) from digitalized 2D drawings such as DWG or PDF needs to be extracted. Drawing recognition is one kind of pattern recognition or graphic recognition technical method that recognizes feature from a document in the computer engineering field (Kaneko, 1992; Chhabra, 2000). When the target subject is limited to engineering drawings, it is defined as drawing recognition, and it is the technique that classifies shape and text based on the relationship of vectorized pixel from the documents containing graphics. Geometric patterns from lines are acquired for understanding the shape of objects, and the attributes from texts are necessary to identify the material of the object, element type, and location information (Figure 2-3).

Because both data are required to complete the 3D shape of an object with properties, the process of distinguishing them as separated layers takes priority and a modeling process is employed consequently (Lu et al., 2005; Lu et al., 2008). Fletcher and Katsuri (1988) are well known for a research study that suggested text and line extraction using heuristic rules to separate data in an image (Hoang and Tabbone, 2008). This method is widely used due to its simplicity and scalability. Nevertheless, it is difficult to apply when

the line and text overlap or touch each other.

In order to complement this method and employ it for a graphic-rich images such as engineering drawings, Tombre (2002) proposed an extraction method that finds a text string's direction as well as its boundary box simultaneously. On the other hand, a morphological component analysis, suggested by Hoang (2010) and Pan (2007), uses a distinguishing feature between the text and line to separate them. Further, Zhang (2004) suggested the shape representation and description techniques using generic Fourier descriptor, and it became the most effective solution among general image analysis methods. After information extraction from the 2D drawing set, a BIM model can be created by linking the obtained information to a basic object model.

2.3.2 As-planned 3D Modeling Based on 2D Drawings

There have been many research studies on automatic 3D model generation using existing 2D graphic information. 2D drawings are composed of graphic elements of vertices and edges, and these elements can be utilized to generate 3D model accurately in terms of numerical value and shape.

However, in engineering drawings lines that overlapped or are very close to each are common; therefore, a technique that arrange them and leaves only necessary contour lines is important for further extrusion of the 3D model. In this direction, Lewis (1998) proposed building model generation (BMG), which creates the entire building plan in 3D by identifying lines in floor plans and leaving only closed curves to extrude. The output model can be employed for the design review stage with accurately created inner and external spaces and for the building operation plan, especially for building evacuation plan with walk-in function. Brendif (2013) suggested the digital surface model (DSM), which generates a city-level 3D model by extracting non-overlapped square-based outlines from 2D graphic images and constructing continuous boundaries with the array of line segments. It extracts outlines not only from CAD drawings but also from aerial

photographs as well, thereby demonstrating the use of multiple sources for 3D model generation. Kim (2009) proposed the use of pattern recognition for structural wall modeling. He suggested creating a surface based on line data first and then creating a complete 3D model by surrounding all sides with the surfaces. From the resulting model, exact values for length, area, and volume of a modeled object can be calculated (Kim et al., 2009). However, this surface model neither contains attributes nor has object-oriented features, and therefore, it is limited in practical use.

These previous studies have completed a shape-oriented surface model for design review of simple quantity calculation. Despite their lack of information in terms of properties of object-oriented features, the boundary construction and 3D extrusion skills defined may be applied in part or as is in this research this research study at the object shape creation stage.

2.3.3 Research Studies Using an Integrated Approach

The aforementioned research studies focused on two different themes: one distinguishes line and text information when they are mixed in a single image and the other creates 3D models automatically based on extracted line data from drawings. However, to utilize these methods in practical construction projects, these technologies must be integrated under one framework to ensure that work conducted in a continuous workflow.

There is a commercial software that supports automatic object modeling using drawing recognition (Changsoft I&I, 2019). Using this technology, project location information is defined and saved, and the reinforcement model is automatically generated after being recognized from the rebar placement drawings. Although this solution implements drawing recognition and subsequent automatic modeling, it focuses on a specific type of work trade.

Ahmed (2011) and Gimenez (2015) proposed a method that separates geometric elements and textual information through text-graphic segmentation and generates 3D wall model as well as enclosed space model surrounded by them (Ahmed et al., 2011; Gimenez et al., 2015). However,

they focused on wall modeling that constitutes space, and they did not consider generating the attributes of the objects. Further, a scanned engineering drawing was used as the source for modeling and, thus, cross-sectional relationships such as ceiling height or level were not considered.

2.3.4 Limitations of Previous Research Studies

Table 2-3 depicts the methods and findings of 12 research studies, organized in chronological order and classified under various subjects: information extraction (5 studies), 3D shape creation (3 studies), and the integrated process of drawing recognition and shape creation (4 studies).

Preliminary research including literature reviews and research on related technologies have been conducted to analyze to existing efforts with theoretical and technical similarities applicable to this study. Research studies that focused on model creation using drawing recognition showed a similarity with the purpose of this research study in terms of employing existing drawings as a basis for automatic model generation. However, most of their research results are suitable for a specific type of construction such as reinforcement (Changsoft I&I, 2019), or for a building component like walls and enclosed inner space (Ahmed, 2011; Gimenez, 2016). In particular,

Gimenez (2016) conducted his research along with software development and performed his experiments under research hypotheses of information separation and data transformation process. Although this is a similar to the research objectives of this study, their studies utilize a different approach based on the assumption that every structural objects are started from the inner space such as a roon space and enclosing walls, and then openings exist inside the wall.

2.4 Research Hypotheses

In order to generate an object model using drawing recognition regardless of the shape and work trades, this research study considered an approach that separates individually transforms line and text then integrates them later. The project data can be created by defining the relationships between independently created objects as priority and deduction rules comply with quantity calculation standards. This research model is defined by following research hypotheses (Fig. 2-4).

Hypothesis 1: Separate conversion process of shape and property creates object model regardless of shape or work trade.

The shape and property of the object model is a distinguishing feature of BIM data from the 3D model. This research study defines the framework that includes the overall process and related technologies of extracting line and text information separately, transforming data, and then integrating them. Based on experimental applications, it is confirmed that this framework enables to generate object model regardless of the object type and shape. Hypothesis 1 focuses on the individual object, whereas Hypothesis 2 considers relationships between created objects.

Hypothesis 2: Modeling criteria that complies with estimation standards facilitate quantity extraction during the construction phase.

Despite the availability of quantity information that is automatically extracted from BIM data, BIM-based quantity calculation is hardly utilized due to the inconsistent modeling standards with the traditional estimation rule. To confirm Hypothesis 2, the relationships between objects—priority and deduction when overlapped—are defined, and then the applicability of the resulted numerical information from BIM data is evaluated in accordance with the estimation criteria in the construction phase. To this end, the modeling criteria are established first, and then experiments are conducted.

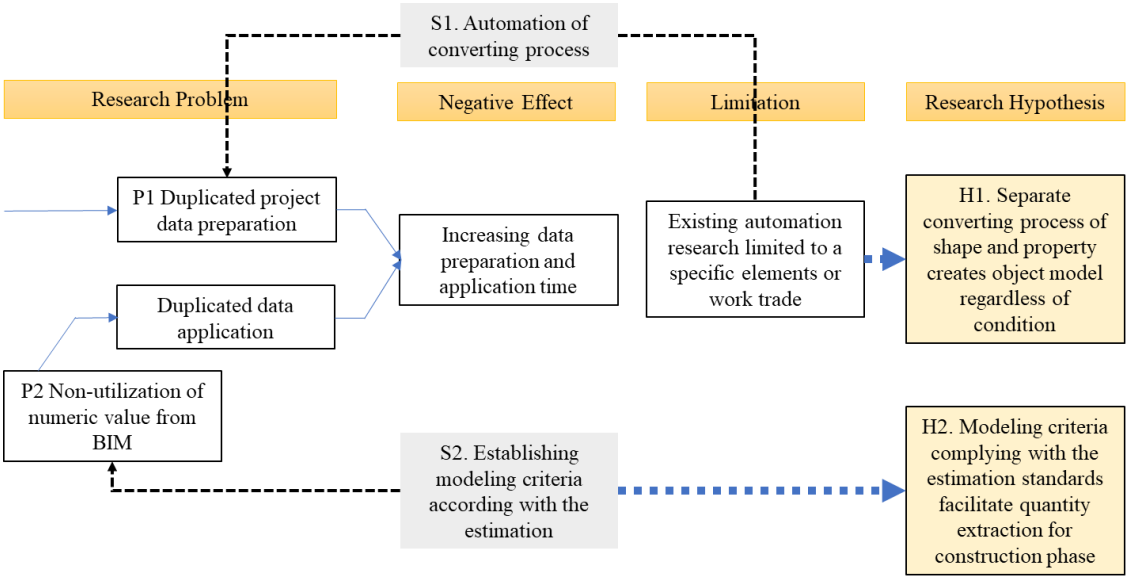


Figure 2-4. Description of research hypotheses

2.5 Summary

In this chapter, the advantages of BIM-based project data and the difficulties in applying it in the construction industry were addressed. Two fundamental causes as well as previous efforts to relieve them were presented. However, an integrated process of these two approaches is needed for practical use; in other words, an automated 3D object model generation that can be employed for QTO as well as cost estimation. In this context, this study compared 2D drawing–based QTO process with the BIM data–based one to identify points to be considered in model creation. To take advantage of the ability to automatically calculate the quantities from BIM data without the estimator's intervention, the required accuracy and LOD standards that are needed for model creation were confirmed.

Also, this chapter presented a comprehensive review of previous efforts to address the problem of repetitive data-writing works. Although some implications can be drawn by their outcomes, it is still difficult to employ them in practical applications. Specifically, these previous studies focused either on the process of drawing-based automatic model creation or on the extraction of drawing information. Some research studies also proposed an integrated process; however, they focused on a specific work trade, making

them difficult to employ in a construction project that includes complicated work trades and construction types. In order to overcome the limitations of these existing research studies on automated BIM data generation and to utilize the created data in the construction phase, two research hypotheses were established: the first hypothesized that a separate conversion process of shape and property creates object models regardless of shape or work trade; the other hypothesized that modeling criteria that comply with estimation standards facilitate quantity extraction during the construction phase.

Thus, this research study suggested the overall process and step-by-step methods including pattern recognition and text-line separation, proposed in the previous studies. To be specific, the automated object model creation was based on the modeling specification that meets the QTO standards, which is defined in consideration of the numerical data utilization—quantity and estimated cost—on construction project.

Chapter 3. Research Methodology

This chapter introduces the research framework and methods adopted to develop the research model required to explain the proposed hypotheses. The developed research framework, based upon specific research objectives, explains acquisition and utilization of information at the object level and establishes relationships between objects at the project level. For verification of the research framework and methodologies, further experiments are conducted in the next chapters. The research methods include information acquisition and utilization approaches for creating object model as well as the experimental models and result verification methods applied in object- and project-level experiments.

3.1 Research Framework

The research framework, including research methods, used to validate and verify the automated BIM generation framework is described in Figure 3-1. First, the automated object model generation framework based on drawing recognition, and then the priority and deduction rules between objects for quantity calculation were derived from related literature and technical reviews.

The experiments to confirm the research hypotheses were conducted separately at the object level and the project level. The former was performed with an aim to confirm the feasibility of the individual object generation framework. The latter was conducted to verify applicability of the completed project data, for example, it is confirmation of the quantity output result that has changed due to established relationships between object models. In the object-level experiment, a text classifier for collecting the properties of an object model and a line-based 3D shape creation environment were developed. Its effectiveness is evaluated using the results of the created 3D shape and the recognized object property as well as self-performance evaluation. In addition, F1-score and G-mean were calculated for cross validation.

In the project-level experiment, performance evaluation was conducted through quantitative comparison. To be specific, the quantity of the target case was extracted using three different methods: drawing-based calculation, BIM model-based calculation, and the BIM model considering priorities between objects-. Two relative errors in comparison with the drawing-based quantities were calculated, and the two results were comparatively evaluated to explain their applicability to construction projects.

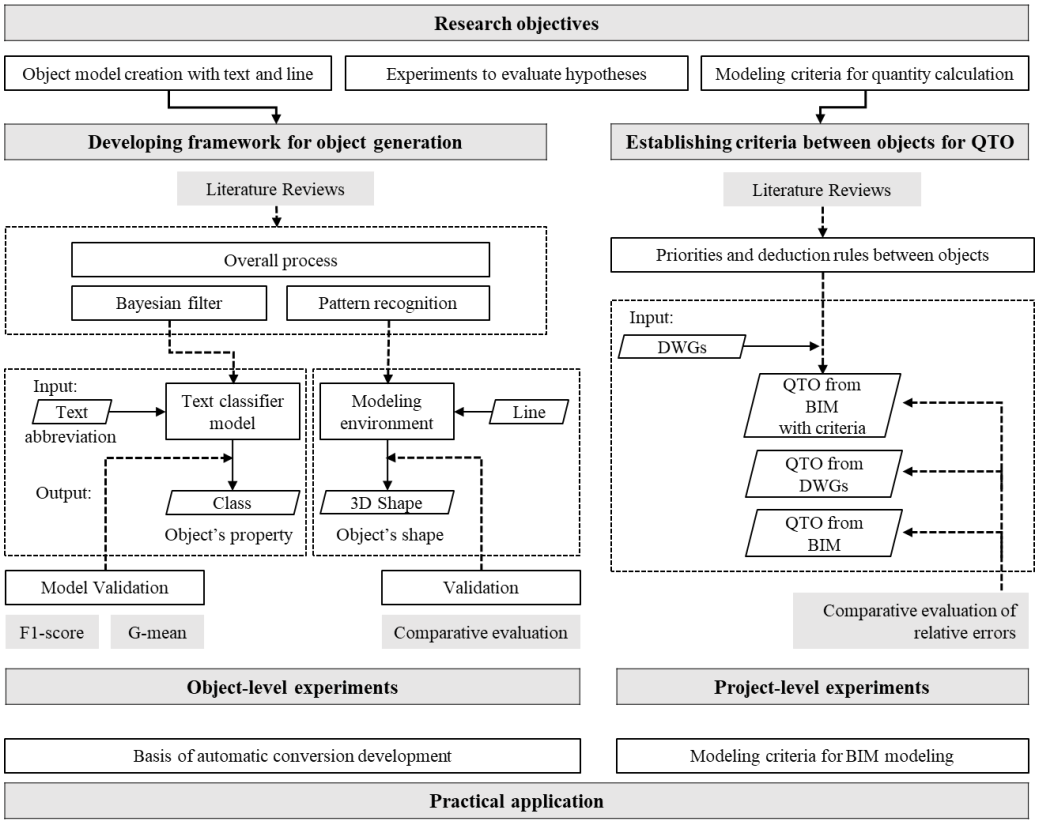


Figure 3-1. Research framework

3.2 Information Acquisition and Object Generation Methods

3.2.1 Line and Text Separation Method

The line and text separation algorithm is the basic method employed for information acquisition and object model creation using drawing recognition. The separation algorithm is based on the nature of map or drawings, wherein groups of graphic components (e.g., lines and circles) are connected continuously with large portions and clusters of text (e.g., annotations, room names) are arranged in rows with relatively small components (Tombre et al., 2002). The extracting procedure is conducted in the order of generated connected components: first creating area and ratio filter and then extracting text strings. Subsequent to creating an imaginary box surrounding the connected components, the layer is distinguished based on the composition of the connected objects in this box, that is, histogram and direction.

Distinguishing Criteria of Density and Direction

Text and line layers mixed in the document are classified based on the density and direction of components in the virtual box created in the previous step. Specifically, a set in case of having a direction with a high density is classified as the text layer, and a histogram with a large number of individual

components with low density is classified as the graphic layer. The distinguishing threshold is explained as following equation:

$$T1=n*max(Amp, Aavg) \quad \textit{(Eq. 3-1)}$$

In the above equation, A*mp* represents the components in the bounding box where the histogram is densest and A*avg* represents the number of components in the box where the histogram is of average density. It is classified as the text layer when it occupies an area smaller than T1 and the graphic layer in the opposite case.

Hough Transform and Domain for Word Segmentation

In addition, the Hough transform (HT) and the Hough domain (HD) are applied to explain the alignment of text groups and make semantic segmentation. HT is a line-to-point transformation that describes a group of components lying along a straight line with the centroid and directionality of the connected components (Flecther and Katsuri, 1988). In HD, a line of $x\cos\theta + y\sin\theta = \varrho$ in the Cartesian space (x,y) is translated as a point (ϱ, θ). To classify a word based upon meaning, it calculates the average height H*avg* of the bounding box and then finds the alignment with horizontal, vertical, or diagonal direction in HD. In terms of division, the average height *h* is figured and separated as words when it is smaller than $\mu*h$. (μ is a variable and is

considered as stable when it is near to 2.5) (Fletcher and Katsuri, 1988). Although this separation algorithm does not quite support the division of single text or dotted line, it works for abbreviations that specify the object's classification and for polylines that configurate the object's shape.

3.2.2 Text Classification Method Using Bayesian Theorem

Text components from the separated layer passes through the text classifier to determine the building element or object's category of the BIM model to which the object model to be created belongs. The simple Bayesian filter or simple Bayesian classifier (SBC) is known to be the most effective text classifier due to its capability that the words to be classified are suitable for supervised machine learning with objects labeled as a specific class (Diab and Hindi, 2017; Rennie et al., 2003). Specifically, the classification of the supervised learning model solves the problem deciding to which class a new input word belongs to, based upon trained model with existing labeled data.

Bayes' Theorem

Based on conditional probability, the Bayes' theorem explains the probability P(A|B) that A will occur after B occurs when P(A) and P(B) are defined as a probability of occurrences for A and B as follows:

$$P(A/B)= \frac{\{P(B|A)* \ P(A)\}}{P(B)} \qquad (Eq.\ 3\text{-}2)$$

By substituting both sides of the above equation, it can be transposed as follows:

$$P(A/B)*P(B)=P(B/A)*P(A) \quad (Eq.\ 3\text{-}3)$$

In case the simple variable A is replaced as a probability of vector, A*i* describes the *i*th class of the event B, and P(B|A*i*) can be decomposed as P(*vi*|A*i*)…P(*vj*|A*j*), where *va* is the value of the *a*th attribute in the example B (Domings and Pazzani, 1996).

Bayesian Text Classifier

When the probabilities are independent from the given class B, Domingos and Pazzani explained the classes that maximized the following equation should be predicted, and they called this procedure as the naïve Bayesian classifier:

$$P(Ai/B) = \{P(Ai)/P(B)\}\prod_{a=1}^{j}(P(va|Ai)) \qquad (Eq.\ 3\text{-}4)$$

In order to classify a word into a category, the classifier needs to confirm which class among the plurality of categories is more likely to be included. Using Eq. 3-2, P(A) presents the probability of the entire input, which are

replaced by 1, and then the word B is in the category is transformed as per Eq. 3-5:

$$P(B|A) = P(B)P(A|B) \quad \textit{(Eq. 3-5)}$$

The Bayesian classifier based on the above equations Eq. 3-3, Eq. 3-4, and Eq. 3-5 hase been developed in two different ways: the first is the multi-variate Bernoulli model to explain the occurrence or absence of a specific word, and the other is the multinominal model to calculate the number of times a specific word occurs (Ying, 2007).

Bernoulli Classification Model

In the Bernoulli model, the presence or absence of the word is explained as a vector of binary attributes (presence: q, absence: 0), and the Laplace's law that predicts the possibility of continuous occurrence is as follows:

$$P(wt|c = k) = 1 + nk\ (wt)\ Nk + \ |v| \quad \textit{(Eq. 3-6)}$$

In the above equation, *nk (wt)* is the number of documents of class *k* in which *wt* is available and *Nk* is the total number of documents of that class (Diab and Hindi, 2017).

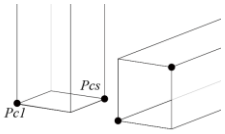
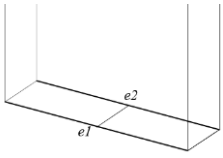
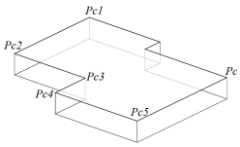
Multinominal Classification Model

The multinominal model, on the other hand, calculates the total occurrence or frequency of occurrences. The existing Bayesian classification assumes that the occurrence of each text data is independent, which is called as a naïve Bayesian assumption, and based on this assumption the classification model is trained and tested. In recent research studies, however, it was confirmed that the performance of the classifier is still effective when there is a dependency between words (Domingos and Pazzani, 1996; Diab and Hindi, 2017). Also a method of optimizing the decision boundary, by modifying the weight or using a complement class when the data among classes is imbalanced, was suggested (Rennie, 2013).

3.2.3 Pattern Recognition for Object Modeling

Pattern recognition is mainly employed in the field of image retrieval and it includes shape representation and shape description techniques (Zhang, 2004). Two approaches are applied in this recognition method: a contour-based approach of comprehending the target's outline to find a shape and a region-based approach of extracting shape from the entire area of the image composed of surfaces. For engineering drawing, where basic graphic shapes are drawn with lines, a contour-based approach is applicable. To create an object model, it is necessary to discover the shape/pattern of a plane or a cross section that is the basis of a 3D object shape; this can be defined using three different approaches (Table 3-1).

Table 3-1. Pattern recognition to configurate the basis plane/section shape

Recognition Subject	(1) Two Diagonal Points (Zhang, 2002)	(2) Two Parallel Lines at Intervals (Gimenez, 2016)	(3) Multiple Vertices
Applicable object type			
	Beam, column, girder	Wall	Slab, roof

First, in case of column and beam, planar or sectional rectangular shape can be completed by finding two diagonally separated points. In this case, the outline B*c* of the basic object's shape is expressed using the two points of *Pcl* (lower left corner point) and *Pcs* (upper right corner point) as follows:

$$Bc=\{Pcl(x, y), Pcs(x, y)\} \quad \textit{(Eq. 3-7)}$$

Secondly, there is an object type shaped with two parallel lines having a consistent gap with a longitudinal direction such as wall component. The outline B*w* is defined as follows (Gimenez, 2016):

$$Bw= \{e1, e2$$

$$e1shape = e2shape=segment$$

$$e1//e2$$

$$Smin < d(e1,e2)<Smax\} \quad \textit{(Eq, 3-8)}$$

In the above equation, *e* represents the geometry primitives of the longitudinal line of the bottom wall shape and *d* is the length of the wall created by parallels *e*1 and *e*2.

Lastly, the plane required for object model shape recognition is defined by multiple vertices:

$$Bs = \{\{Psi(x,y)\}\ i=1 \ldots n\} \quad \textit{(Eq. 3-9)}$$

The applicable object type has a 2D planar shape with multiple corner points like slabs and roof.

3.3 Validation and Verification Methods

In order to confirm the performance of the research model and to evaluate the experimental results, cross-validation methods and a comparative evaluation method were employed. For cross-validation, we employed confusion matrix, which describes the performance of the classification model, and further F1-score and G-mean, which calculate the effectiveness as well as the sensitivity to evaluate research model. In addition, we compared the experimental results with another method as well.

3.3.1 Confusion Matrix, F1-Score, and G-Mean

Confusion Matrix

The traditional metrics to evaluate a text classifier's effectiveness are accuracy and error rate (He and Garcia, 2009). The confusion matrix for a K-class classifier is presented by a k X k table, representing the frequency of observations between the actual class Ci and the expected class Cj. Although binary confusion matrix, which describes the positive and negative frequency for a class, is not often employed in practice, it was adopted in this study to calculate the performance index with the confusion matrix (Ruuska et al., 2018). Considering a binary classification question of "yes" or "no," the

model can predict “y” for a true “yes” class; in this case, it is called as TP. However, when the correct class is “no,” it will be called FP. Similarly, for the opposite cases, we have TN and FN.

This representation of the true class and the classifier’s hypothesis is described in Figure 5-7 as a confusion matrix.

Hypothesis output	True class: P	True class: n
Y	TP (True positive)	FN (False negative)
N	FP (False positive)	TN (True negative)
Column counts:	Pc	Nc

Figure 3-2. Confusion matrix for performance evaluation

From this definition, the accuracy and error rate are defined as

$$Accuracy = (TP+TN)/(TP+FN+TN+FP);$$

$$Error\ rate = 1-accuracy \qquad (Eq.\ 3\text{-}10)$$

This traditional model evaluation method is mainly engaged when the balance of the quantity and distribution of resource data employed for training and testing in the model development process is guaranteed. Compared to the amount of data applied in existing research studies that deal

with text classification problems where 45 to 3000 points are used, the data in this study was relatively small in number. Therefore, additional evaluation methods are required to verify the performance of the model developed in this study.

F1-Score and G-mean

When the amount of data is small or imbalanced, the F1-score and G-mean, which are comprehensive evaluation methods, are adopted for model measurement (He and Garcia, 2009; Tang, 2016)

$$Precision = TP/(TP + FP) \qquad \textit{(Eq. 3-11)}$$

$$Recall = TP/(TP + FN) \qquad \textit{(Eq. 3-12)}$$

$$F1\text{-}score = \frac{2 * Recall * Precision}{(Recall + Precision)} \qquad \textit{(Eq.3-13)}$$

$$G\text{-}mean = \sqrt{\frac{TP}{TP+FN} * \frac{TN}{TN+FP}} \qquad \textit{(Eq. 3-14)}$$

The F-score metric combines precision and recall to estimate the classifier's effectiveness. In particular, when precision and recall have equal weights, it is called as F1-score (He and Garcia, 2009). In addition to this, the G-mean method explains the sensitivity of the classifier with a positive and negative accuracy.

3.3.2 Comparative Evaluation of Relative Errors

Comparative Evaluation

The comparative evaluation method is an analysis method that is employed when evaluating criteria for direct comparison are clear because the analysis units between the research subjects are similar (Vartiainen, 2002). It is effective when trying to compare and evaluate the shape of an object or numeric values with a clear meaning because the results of comparison between objects are specific and micro-analysis (Etzioni-Halevy, 1990). It is considered a more efficacious way to make a rational decision between alternatives by identifying differences rather than confirming similarities between options (Scriven, 1967).

Relative Errors

An error represents a deviation from the actual value of the measuring variable (Helfrick and Cooper, 2005). Similarly, a relative error is used for making exact numerical comparisons and calculating the error rate between the target and the baseline. It is also called as relative uncertainty or approximate error.

In this study, the relative error rate was employed to confirm the degree

of deviation compared to the baseline for each object type. The following formula was applied.

Relative Error Rate = (Quantity based on 2D drawings - Calculated quantity) / (Quantity based on 2D drawings) X100 (%)

3.4 Summary

In this chapter, the research framework and methods to develop and validate the research models were introduced. Specifically, research framework and methods describing the information acquisition and the object model generation at the object-level, and the establishment of relationships between object types at the project level were explained. In addition, methods to validate the experimental results and to verify the performance of the research model were described.

First, to confirm the first research hypothesis, text-line separation, text classification, and patter recognition-based 3D shape creation methods were applied. Layer extraction based on the density and direction of individual components in drawing, text classifier based on the Bayesian filter, and shape creation with pattern recognition method were introduced. Second, methods to evaluate the research model performance and results of research experiments were described. These includes the confusion matrix, the ctoss-validation method, comparative evaluation as well as relative error rates

Chapter 4. BIM Data Generation Framework

This chapter describes the framework that includes the data transformation procedure and the required technical methods for drawing-based BIM data generation. This framework is explained separately at the object level (which automatically converts drawing information created at the design stage into information of the object model) and at the project level (which describes the relationship between object models that meet the quantity calculation criteria). The framework in this chapter is developed to explain the research hypotheses proposed in Chapter 2, and it will be validated through the research experiments described in Chapters 5 and 6.

4.1 Overall Process of Project BIM Data Generation

The overall process for the automated BIM data generation framework based on drawing recognition is described in Figure 4-1. To generate BIM data that includes numerical information that can be utilized during the construction phase, establishment of relationships between object types that follow the quantity calculation and an automated object model generation framework are necessary.

At the object model level, the design data from a design drawing is extracted using text/graphic separation techniques. Lines and texts in a drawing provide basic knowledge about a building project in terms of geometric and semantic information. Through project information recognition, the basic information in terms of planar center lines and the level information, which are location information of the project, is acquired. The text data, which is adjacent to a polygon, or object's shape, is passed through a Bayesian filter to identify what is the building component of the object to be modeled. Then, by mapping the geometric features of the planar size and height to a basic object for modeling defined from the previous step, the shape of the object model is completed. Finally, the properties of the created object, such as its location, size, and type, decided from the modeling process,

are recorded in the form of a lookup table for later use in construction management

To utilize the object model created in the previous step at the construction stage, the modeling criteria (that is, object model combination rules at the project level) are established based on the estimation method. Specifically, we searched the object types adopted in practice and the points to be considered for project data preparation. Based on these preliminary research findings, the modeling specification was defined, which is integrated with the priority between the objects and the deduction rule from calculation formula when they are overlapped.

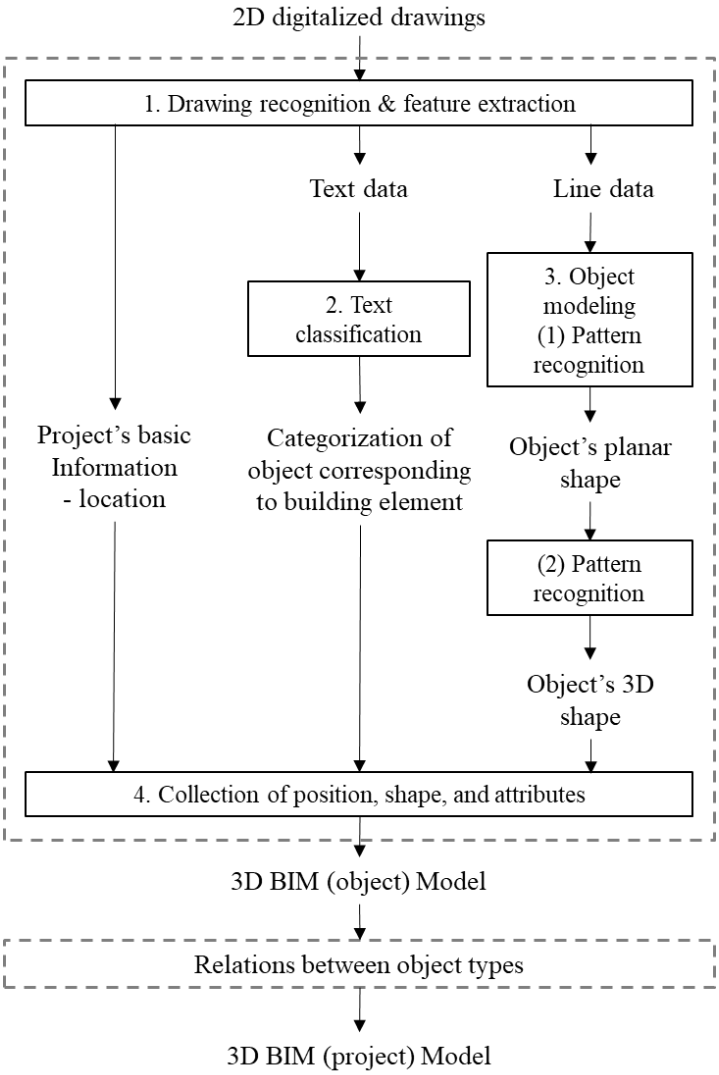


Figure 4-1. Overall process of conversion from 2D drawings to 3D BIM model

4.2 Object Model Generation Framework

This section describes the procedures and techniques for recognizing drawings, extracting requisite information to create an object model, and finally creating a model. To be specific, this includes the overall process and three information transforming processes: (a) separating information from digital format drawings completed in the previous step into text and line; (b) confirming the classification of each object to be built by identifying an abbreviation referring to the object; and (c) distinguishing line as a shape through pattern recognition and converting it into a 3D form or an object.

4.2.1 Line-text Extraction from 2D Drawings

First, plan and section drawings are imported to model the exact position (Fig. 4-2). In particular, the precise location of individual elements to be modeled is confirmed by comparing them with the center lines of two dimensions and the sectional level lines of three dimensions, which are the basic location information of a project. As geometric and semantic information of each building object used in the design process is described as not only a form of line but also as text, the recognition of the structural table should be accompanied.

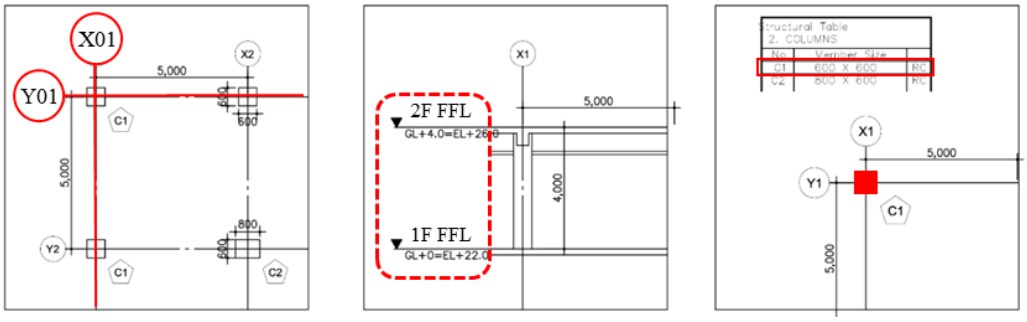


Figure 4-2. Confirming object's location with centerlines and level information from the structural table

Next, a process to separate the line and text information from the recognized drawings is employed. As engineering drawings display the project design information using lines and text, both data must be recognized and understood. The separation process of these two data was originally proposed and developed by Fletcher and Katsuri (1988) as text/graphic separation. However, with the recent developments in CAD software, these data can now be easily separated and stored.

The third step of this phase involves understanding the meaning of text located in proximity to the shaped object on the drawing. After storing the line and text in separated layers, it is imperative to identify an indication of that letter among the building components, such as column, wall, and slab, which also are elementary components of the BIM modeler. In other words, the modeling basic object, which is a building element, and the recognized

shape are connected to create an object with meaning and shape. To this end, we developed a text classifier using Bayesian filter based on conditional probabilities.

4.2.2 Classifying Texts of Abbreviation from 2D Drawings

The resources used for the training process included the extracted words from the structure list and the actual names of the structural elements. The reason that the classification training is performed with the character classifier is because the "object," the fundamental modeling element of BIM, is based on building components. That is the identified abbreviation refers to the basic model object type and the building element at the same time. For instance, the abbreviation "col" refers to a column that is a basic object type of column models, and a column that is a component of building structural system.

Different object types of classifications are applied depending on how the building components are classified or which object model system or platform is used by software developers. The classification system applied in this research study followed the industrial foundation classes (IFC), which is an interoperable data format developed by the buildingSMART organization

(shown in Figure 4-3). From the classification system, it is based on a "project" or a "building." The sub-category includes specific entities such as structural system, architectural system, and opening system, and in the case of the structural system, there are beams, columns, slabs, and walls.

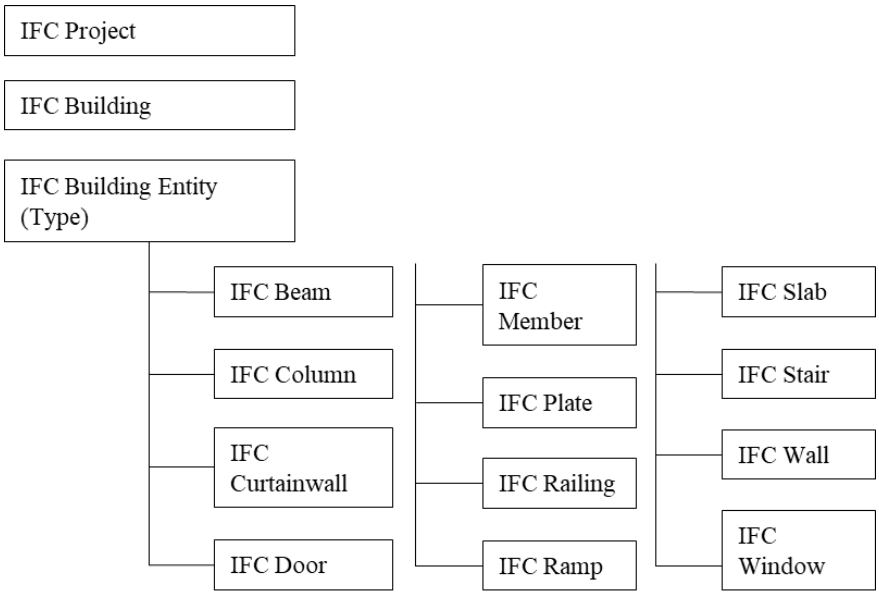


Figure 4-3. Building classification used in IFC

4.2.3 Object Modeling Using Pattern Recognition

(1) Acquisition of Project Location Information

To place objects accurately in the modeling environment based on the project's origin point, center lines and level reference lines must be recognized prior to the modeling step.

The technology from Changsoft I&I, which recognizes a specific drawing layer and loads it as separate information, enables a mechanism to import and utilize guide lines and level information as the project's basic information by importing the relative position from the origin of the drawing (Changsoft I&I, 2019). To be specific, we can gain the center lines from the planar drawing and the level information from the sectional drawing

(2) Planar-sectional Geometry Mapping

Subsequent to the acquisition of the positioning guides, the position, shape, and abbreviation indicating the object are simultaneously identified by means of dragging the area including the object to be recognized from the floor plan.

Pattern recognition is applied to understand an object's planar shape by grasping the positions of both diagonal end points that determine the size and shape of the rectangle. In the case of a perpendicular column C1, the object's lower center point (Pcl; where the structural center lines intersect the object's footprint) and a left-top point (Plt) and a right-below one (Prb; which are the two diagonally located end-points of the polyline at the planar section) need to be ascertained (Figure 4-4).

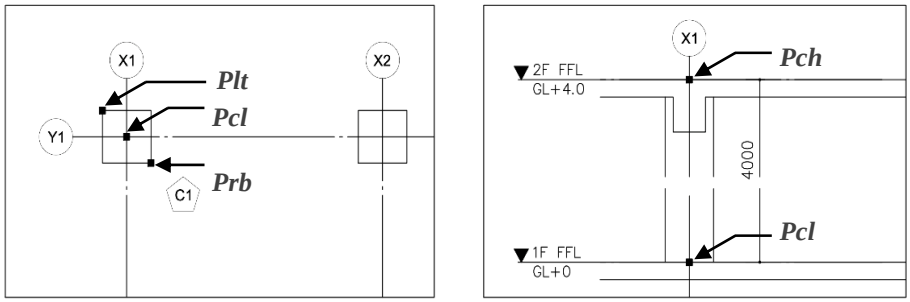


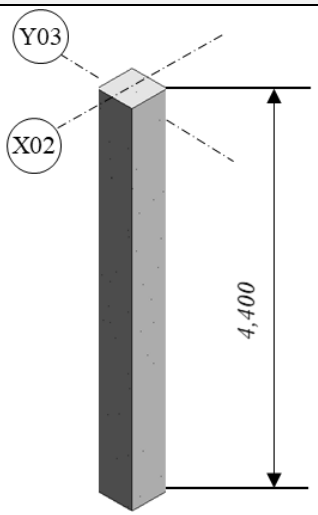
Figure 4-4. Example of extracted points from imported drawings and recognized column, C1

After the acquisition of these three points, the geometric information of position, shape, and size can be mapped to a basic column component. The sectional position is also identified from the level reference obtained in the previous section. In a similar way, a 3D shape creation is finished by identifying the planer center point (Pcl) and the other center point on an upper reference (Pch) of the object, which is the height value for an extraction.

(3) Property Assignment on Modeled Objects

Property information of the modeled object is saved as a look-up table and continuously used throughout the lifecycle of the project. This attribute is a distinguishing feature of the BIM model compared to other 3D models, and it is filled in two different ways. One is the given information that is obtained from the design phase, which comprises design information such as location, size, and material. The other is created after the model is completed, e.g., specific area and volume of a certain material. Further, manually recorded information such as construction sequence, zoning, and a person-in-charge are included. The example of these properties assignment with lookup table is explained as Table 4-1.

Table 4-1. Example of completed column object with lookup table

Object	Parameter	Variable (Given)
	Type	Column
	Type Name	C1
	Constraints	
	Base Level	B1F
	Top Level	1F
	Material	
	Structure Material	Concrete
	Parameter	**Variable (Calculated)**
	Dimension	
	B	400.0 mm
	H	400.0 mm
	Volume	0.640㎥

4.3 Establishing Relationships between Object Types

To complete project data facilitated in construction projects based on the object model prepared in the previous section, the criteria for object model combinations that meets the estimation standards should be established and the project BIM model created accordingly. This section describes the object model, which is the basis for BIM data preparation and examines its various types being adopted in practical applications. By comparing object types and building components, the modeling criteria for quantity calculation in the construction phases are explained. To this end, the calculation equation used for estimation is investigated and then the object modeling standards, which include priority and deduction when overlapping, are suggested. The modeling specification in the last section describes the modeling considerations or relationships between object types and is applicable for the experiments described in Chapter 6.

4.3.1 BIM Objects Definition

The classification system of BIM data is basically composed of architecture, structure, and MEP system, which is similar to the classification of building systems. An “object type” or “family” is a basic element for BIM modeling, and its classification system is comparable to categorization of building components. For instance, the most basic structural elements of building components are slabs, columns, and beams; at the same time, these are also object types in BIM models. An “object” is a basic element in modeling and it also indicates individual object indicate the exact building component in the BIM environment. According to the NBS National BIM Library, “an object is detailed information that defines the product and geometry that represents the product’s physical characteristics. The visualization data that gives the object a recognizable appearance and behavioral data, such as detection zones, enables the object to be positioned or to behave in exactly the same way as the product itself” (NBS, 2020).

An element’s name or the number of sub-classes are different depending on the BIM solution (Table 4-2). According to IFC, there are eighteen composing entities. Whereas the .pla format from Graphicsoft and .rvt format from Autodesk have 16 elements and 25 families, respectively. Although

different names and classification criteria are used, they are defined based on the building construction system.

Table 4-2. Designations and various number of BIM objects

Basic Objects		
IFC Format, 18 Entities	**Graphicsoft (.pla format), 16 Elements**	**Autodesk (.rvt format), 25 Families**
(S) Beam, (S) Column, (S) Wall, (S) Pile, Railing, Ramp, Stair, (S) Footing, Member, (S) Slab, Window, Door, (S) Reinforcement element, (S) Plate, Ramp flight, Stair flight, Building element Proxy, Covering	(S) Wall, (S) Slab, (S) Column, (S) Beam, Door, Window, Object, Stair, Roof, Shell, Skylight, Curtain wall, Morph tool, Zone, Mesh, Railing	(S) Beam, (S) Wall, (S) Column, (S) Floor/Slab, (S) Truss, (S)Brace, (S) Beam System, Connection, (S) Foundation, (S) Reinforcement Component, (S) Steel, Wall, Door, window, roof, ceiling, etc.

Despite the 3D model generation based on the definition of a system and the basic elements being similar to the building components, utilization of the BIM data for quantity calculation remains unexploited. To identify the problems and to address these limitations in applications, interviews with BIM experts and estimators from construction companies were conducted.

The experts pointed out that models created at the design or construction

stage were irrelevant for calculation criteria. In order to utilize the prepared BIM data in the quantity estimation stage, the following three conditions should be considered (Figure 4-5):

1. Different reference level location as per object type
2. Calculating priority when different object types overlapped
3. Duplicated calculation for overlapping surfaces when plural objects meet

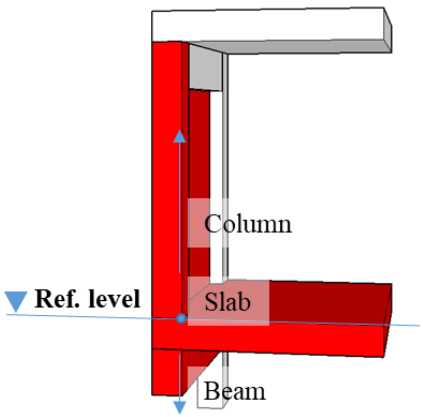
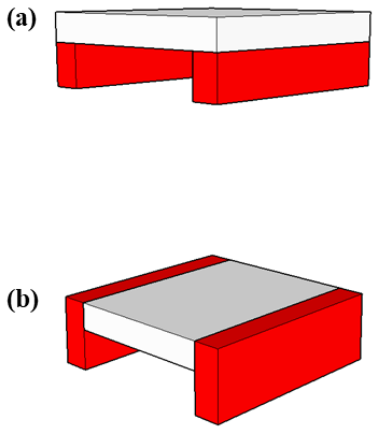
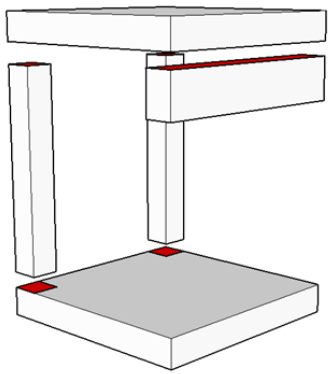
(1) Reference level location	(2) Priority between overlapped objects	(3) Duplicated calculation
		
- Deviated reference level location as per object type	- Calculating priority when different object types overlap: (a) Slabs over beams or (b) beams over slabs	- Duplicated calculation for overlapping surfaces when plural objects meet

Figure 4-5. List of consideration for object modeling

4.3.2 Current Quantity Estimation Approaches

Estimation and quantity calculation methods differ depending on the subject of the application but in Korea, the formula is defined in the architectural quantity survey (Jang, 1969). In a different way from object categorization, for example of structural system, classification of objects for QTO follows building components classification of slab, column, beam, foundation, and wall. The calculation methods defined in architectural quantity survey is based on the required dimensions between the objects depicted in Figure 4-6. As the total number of objects are counted regardless of the relationships among them, in this study we considered only the formula for area and volume calculation.

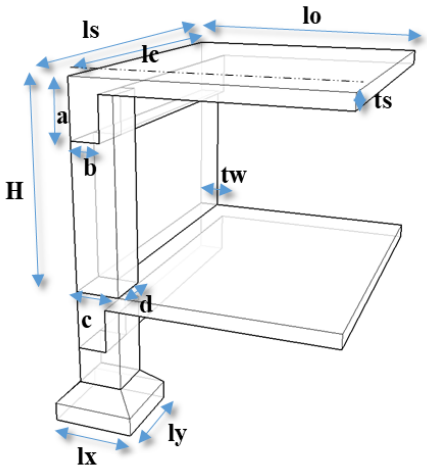


Figure 4-6. Dimensions of object model for quantity calculation

Volume Calculation for Slabs

The quantity of object is classified into object type with the same size, and then the total quantity is estimated by multiplying the calculated value of one object by the number. The object that obtains the highest priority among the structural system is the slab. Volume calculation for slab is done using the equation below:

$$Vs = ts*(ls * lo) \qquad \textit{(Eq. 4-1)}$$

In the above equation, *Vs* represents the volume of slab, *ts* represents the thickness of slab, *ls* represents its length and *lo* represents its width.

Volume Calculation for Columns

The second priority is the column, which is calculated:

$$Vc = (c*d)*(H-ts) \qquad \textit{(Eq. 4-2)}$$

In the above equation, *Vc* represents the volume of column, *c* represents the length of cross-section of column, *d* represents its width, and *H* is the floor height.

Volume Calculation for Beams

Then, the volume of beams is calculated:

$$Vb = (a-ts)*b*(ls-d) \quad \textit{(Eq. 4-3)}$$

In the above equation, *Vb* represents the volume of beam, *a* represents its height, *b* represents its width, and *d* is length.

Volume Calculation for Walls

For walls, Vw is obtained by:

$$Vw = (H-ts) *tw*(ls-d) \quad \textit{(Eq. 4-4)}$$

In the above equation, *Vw* represents the volume of wall, *H* represents the floor height, *tw* represents the wall thickness, and *ls* represents wall length.

Lastly, the relationships with other structures are not considered for foundation. Using these formulas, priority among structural elements is confirmed as: slab→column→beam→and wall.

4.3.3 Object Modeling Specification for QTO

An additional consideration in generating object models for quantity calculation is the difference in reference level by object type. The Korean architectural documents and information standards, published in 2006 by the Korean Institute of Architects, describes the standards for building drawings and is commonly employed (KIA, 2006).

To indicate a building's reference level, finishing level (FL), which is based on the finishing level of each floor, or engineering level (EL), which is based on the structural slab surface excluding finishing materials, are employed. The reference level for each object type is shown in Figure 4-5 (1). For slab, reference level is approximately on the top surface and beam follows its connected slabs' level. Wall and column identically refer the level based on their bottom surface. Table 4-3 shows the criteria for model preparation that reflects the accuracy, priority, and reference level of the model defined in the previous chapters.

Table 4-3. Object modeling specification table considering QTO

LOD	(1) Reference Level	(2) Priority					
			Foundation	Slab	Column	Beam	Wall
LOD350 /BIL4	OBB	Foundation		-	-	-	-
	OTB	Slab	-		-1	-1	-1
	OAB	Column	-	+1		-1	-1
	OBB	Beam	-	+1	+1		-1
	OAB	Wall	-	+1	+1	+1	
(3) Model for Formwork		Creating a separate surface-based model for calculating formwork quantity					

In describing the reference level, the relational position between the object and the reference level is considered. When an object faces the reference level, it is called as OTB (object touching baseline). When the object is positioned higher than the baseline, it is OAB (object is above baseline), and in case the reference level is higher than the object, it is OBB (object is below the baseline).

To deal with the intersection relationship, a structure–item–priority matrix (SIP matrix) was established by comparing the relative priorities between the five structural items. The relative priority between the structural items in the row and items in the column is expressed in values as follows:

(1) ‘1’ represents the item in the row has a higher priority than the item in the column

(2) ’0’ means both items in the row and column have an equal priority

(3) ‘-1’ represents a lower priority of the item in the row than the item in the column.

The accuracy level of the object model is defined in LOD and BIL to indicate the minimum detail required for quantity calculation. BuildingSMART’s LOD is 350, and BIL is 4. In a case of area calculation such as concrete formwork, only visible surface is calculated on the internal space.

4.4 Summary

This chapter described the proposed framework, which includes the overall process of data transformation and the required technical methods for drawing-based BIM data generation. Next, the following were explained in sequence: The purpose behind deriving the output of the object model by importing the completed drawings, the process of drawing recognition and following line/text separation, classifying object to be modeled using text classifier, and pattern recognition to complete the shape. The relationships among types of object model were also established according to the quantity calculation criteria so that the results can be employed during the construction phase.

It was also described how to the line-based graphic information constituting the engineering drawings and the text information describing the graphic were extracted and separated into different layers, using the line'text separation method. The abbreviation for the object was automatically recognized through the text classifier. This classifier could be developed based on Bayes' theorem, and the object, which is the basis element of BIM model, simultaneously follows the category of building components. This research utilized the IFC format of building component.

The line layer separated in the previous step was used to configure the object model. In order to discover the planar shape that is basis of the 3D extrusion, pattern recognition based finding two diagonal points, finding two parallel lines, and the discovery of a multi vertices methods were employed to determine its outline. For extrusion of the 2D sketch created to complete a 3D shape, the pattern recognition was re-applied to retrieve the height value of the cross-section and applied as the extrusion value.

In addition, priority between object types and deduction rule for overlapped objects were established to complete project BIM data based on the object model created. The completed project BIM model according to the established modeling criteria can help reducing extra quantity calculating task by automatically extracting numerical information along with the configuration information displayed.

Chapter 5. Object-level Experiments

In order to confirm the feasibility of the proposed framework for engineering drawings of real construction projects and to prove Hypothesis 1, this study conducts two separate experiments: (a) an experiment using a text classifier to automatically identify an abbreviation that refers to a building component so as to recognize an object's attribute with a form of text; (b) an experiment on configurating a 3D object model using graphic information recognized from drawings so as to confirm the 3D shape creation with a form of graphic/lines (Fig. 5-1). Through performance evaluation and quantitative analysis of these two experimental applications, the feasibility of the framework was confirmed and then the project-level experiment regarding relationships between objects was performed, as described in Chapter 6.

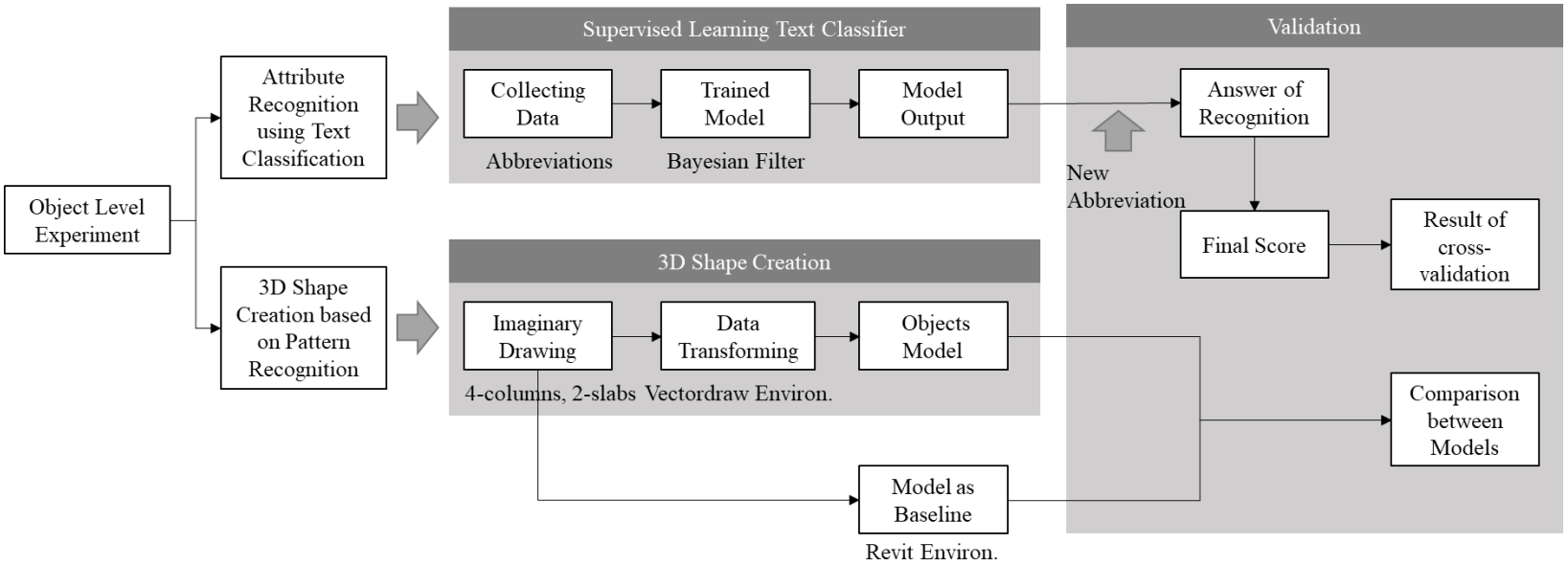


Figure 5-1. Design of experiment

5.1 Text Classification and Attribute Recognition Experiment

5.1.1 Design of Experiment

As described previously, this section describes a text classifier experiment for confirming the recognition of object attributes written in text form. To be specific, abbreviations that refer to building components used in engineering drawings were employed for the first test. For this, the abbreviation data for creating and training the classifying model was extracted from existing drawings. After creating the classifier of a supervised learning machine-based on the Bayesian filter, new words that follow the abbreviation generation method were input to the trained model. By confirming the recognition result and the final score of model performance, its performance as well as the feasibility of the proposed framework was validated.

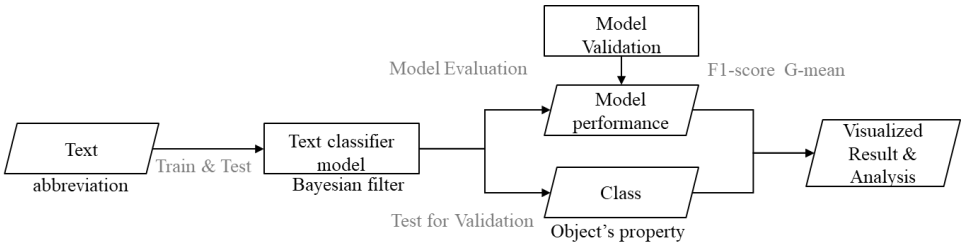


Figure 5-2. Model development algorithm

Furthermore, the results were cross validated re-evaluate model performance due to insufficient data amount (Figs. 5-2, 5-3).

The Bayesian filter was developed using Python 3.6 in Anaconda, ver. of 4.6. The source codes developed and applied in this experiment are as shown in Figure 5-4 and further supplementary details are included in Appendix 1.

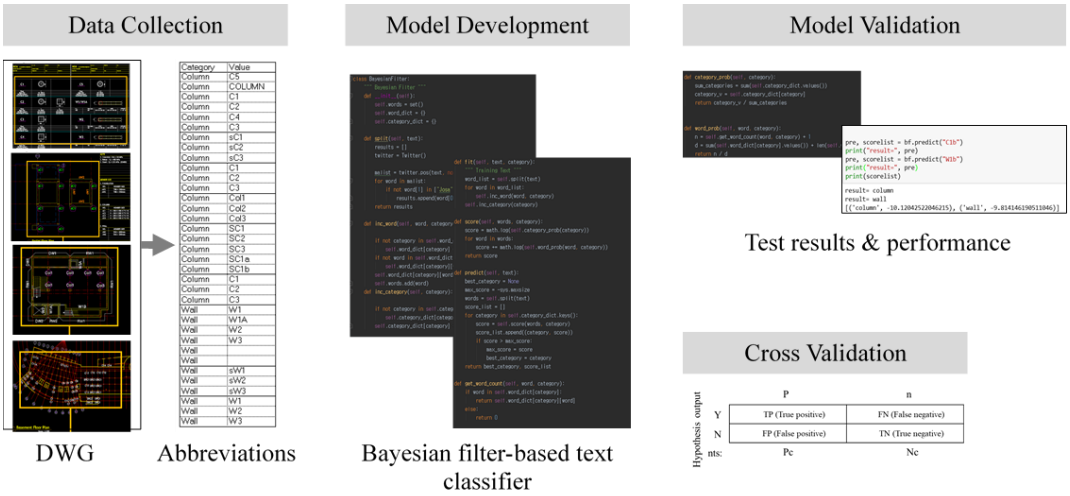
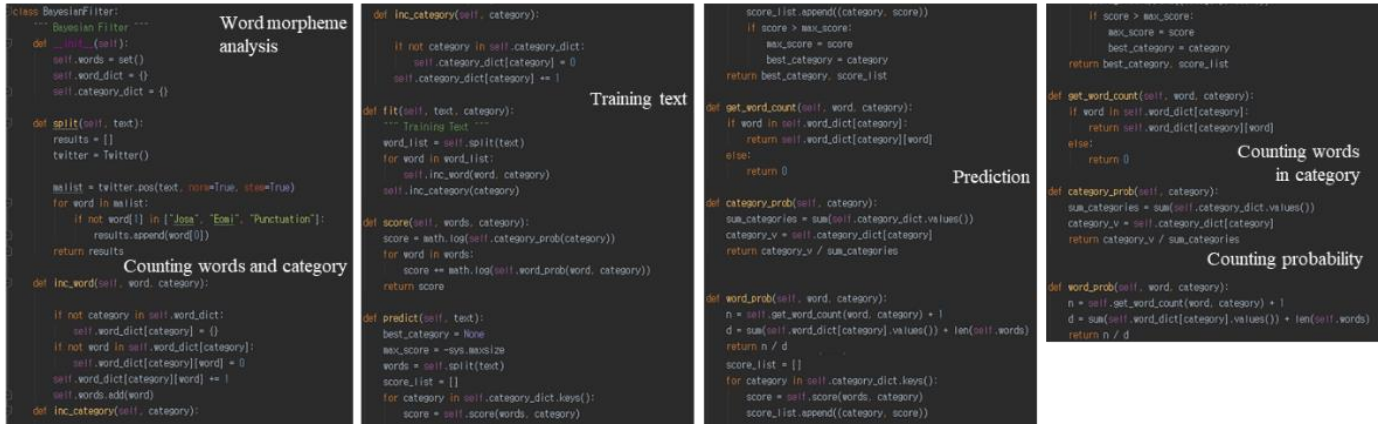


Figure 5-3. Data collection and model development



Algorithm 1 Bayesian Filter Algorithm - Training

Train Multi-class BF (C, D)
2: T ← Extracted Text (D)
3: N ← Count Text (D)
4: for each c ∈ C
5: do Nc ← Count Texts in Class (D, c)
6: prior[c] ← Nc/N
7: textc ← Concatenate Text of All Texts in Class (D, c)
8: for each t ∈ T
9: do Pct ← Count Tokens of Term (textc, t)
10: for each t ∈ T
11: do condprob[t][c] ← $\frac{\boldsymbol{Pct+1}}{\sum_{t'}(\boldsymbol{Pct'+1})}$
12: return T, prior, condprob

Algorithm 2 Bayesian Filter Algorithm - Testing

Applying Multi-classes BF (C, T, prior, condprob, d)
1: W ← Extracted Tokens from Text (T, d)
2: for each c ∈ C
3: do score[c] ← log prior[c]
4: for each t ∈ W
5: do score[c]+ = log condprob[t][c] t)
6: return argmax{c∈C}score[c]

Figure 5-4. Model environment and algorithms for the initial experiment of text classification based on the Bayesian filter

5.1.2. Data Collection and Model Development

Prior to text classification model development, the abbreviation data used in engineering drawings was collected. Specifically, by employing text-line extraction through AutoCAD, abbreviations labeled with the corresponding building elements (Fig. 5-5) were extracted. All extracted data were saved in extensible markup language (.xml) format for later text classification experiments (Table 5-1).

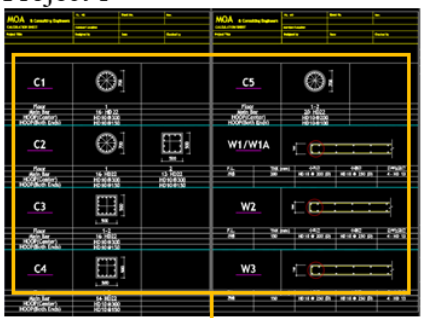
Table 5-1

Abbreviations for structural column from four sample engineering drawing sets

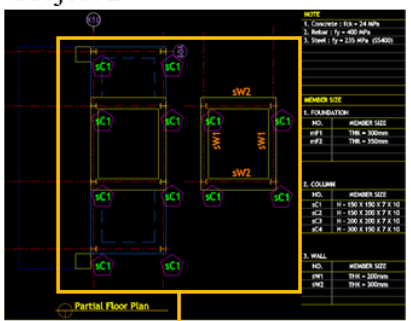
Project 1		Project 2		Project 3		Project 4	
Column	Wall	Column	Wall	Column	Wall	Column	Wall
C1	W1	sC1	sW1	Col2	RW1	SRC1	SW1
C2	W1A	sC2	sW2	Col3	RW2	SRC2	SW2
C3	W2	sC3		Col4	DW1	SRC3	SW3
						C1	SW1a
C4	W3	sC4			DW2	C2	SW1b

<DWG>

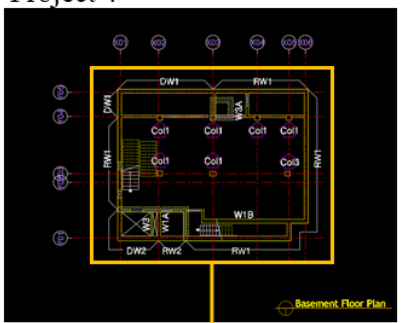
Project 1



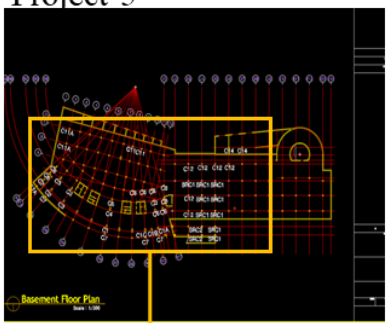
Project 2



Project 4



Project 5



<XML>

Project	Type	Category	Value	Category	Value
1. N_School	Text	Column	C5	Wall	W1
1. N_School	Text	Column	COLUMN	Wall	W1A
1. N_School	Text	Column	C1	Wall	W2
1. N_School	Text	Column	C2	Wall	W3
1. N_School	Text	Column	C4	Wall	
1. N_School	Text	Column	C3	Wall	
2. B_School	Text	Column	sC1	Wall	sW1
2. B_School	Text	Column	sC2	Wall	sW2
2. B_School	Text	Column	sC3	Wall	sW3
3. T_Hotel	Text	Column	C1	Wall	W1
3. T_Hotel	Text	Column	C2	Wall	W2
3. T_Hotel	Text	Column	C3	Wall	W3

3. T_Hotel	Text	Column	C2	Wall	W2
3. T_Hotel	Text	Column	C3	Wall	W3
4. B_APT	Text	Column	Col1	Wall	DW
4. B_APT	Text	Column	Col2	Wall	SW
4. B_APT	Text	Column	Col3	Wall	RW
5. T_APT	Text	Column	SC1	Wall	SW1
5. T_APT	Text	Column	SC2	Wall	SW2
5. T_APT	Text	Column	SC3	Wall	SW3
5. T_APT	Text	Column	SC1a	Wall	SW1a
5. T_APT	Text	Column	SC1b	Wall	SW2a
6. I_Tower	Text	Column	C1	Wall	W1
6. I_Tower	Text	Column	C2	Wall	W2
6. I_Tower	Text	Column	C3	Wall	W3

Figure 5-5. Process of text data extraction from structure elements table and conversion to extensible markup language file

5.1.3 Results and Validation

Test Using New Abbreviation Data

For this evaluation, we tested the developed and trained text classifier with a new abbreviation and then, monitored the resulting value to identify whether it indicates the building elements correctly. As shown in Figure 5-5, abbreviations used in construction drawings are generated using a combination of numbers and alphabets ‘a’ or ‘b’ together with part of the corresponding building component name spelling. Specifically, only consonants in the spelling are used or only the first letter or the front syllable unit with numbers 1, 2, and 3 with ‘a’, or ‘b’ (if necessary) are found. Therefore, hypothetical abbreviations for the test were created as C1b, W1b, and so on (first letter + number + ‘a’ or ‘b’).

This evaluation of building elements was more focused on the resulting output from the classification and following model creation at the building level. As the aim of our research is to provide an automated BIM modeling framework with its environment, this is more compliant with the objectives of a BIM model that has geometric and topological information with its attributes.

The text classification results using the Naïve Bayes filter showed a high level of correct answering rate (Figure 5-6). Regarding abbreviations C1b and W1b, the classifier found that there is 89.8% probability of "column" and 90.1% probability of "wall."

```
from bayes import BayesianFilter
bf = BayesianFilter()

bf.fit("c1", "column")
bf.fit("c2", "column")
bf.fit("c4", "column")
bf.fit("c5", "column")
bf.fit("SC1", "column")
bf.fit("sC1", "column")
bf.fit("sc3", "column")
bf.fit("C1a", "column")
bf.fit("w1", "wall")
bf.fit("w1a", "wall")
bf.fit("sw1", "wall")
bf.fit("sw2", "wall")
bf.fit("sw1a", "wall")
bf.fit("벽", "wall")
bf.fit("C9", "column")
bf.fit("wall", "wall")
bf.fit("기둥1", "column")
bf.fit("기둥2", "column")
bf.fit("co.", "column")
bf.fit("wa1.", "wall")
bf.fit("sW1.", "wall")


pre, scorelist = bf.predict("C1b")
print("result=", pre)
pre, scorelist = bf.predict("W1b")
print("result=", pre)
print(scorelist)
```

```
result= column
result= wall
[('column', -10.12042522046215), ('wall', -9.814146190511046)]
```

Figure 5-6. Text classification test results with a new abbreviation input

Cross-Validation

Considering the amount of data, we applied the macro-F1 score for cross-validation of our results. It gives the same weight to each class among the two types of F1-score calculation methods (Yang, 1999). The cross-validation results are as shown in Table 5-2.

Table 5-2. Cross-validation results of the text classifier

Precision	Recall	F1 Score	G-Mean
0.97 (+/-0.07)	0.98 (+/-0.09)	0.974	0.975

5.2 3D Shape Creation Experiment

5.2.1 Design of Experiment

Next, we conducted the experiment to confirm the feasibility of an object's geological and topological shape created using pattern recognition. Specifically, the creation of an object model's 3D shape by utilizing the graphic information of line obtained from the drawing at the correct location was confirmed. After creating an experimental environment where it was possible to create a model using pattern recognition, two sets of virtual drawing for test were drawn (Figure 5-7). For validation, we adopted a baseline in terms of the model by the existing object modeler and compare two results model in terms of created objects' shape, their location, and quantity (Fig. 5-8).

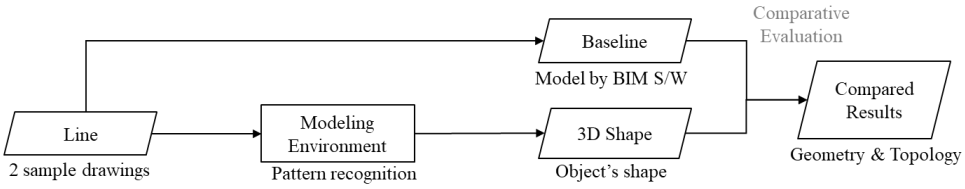
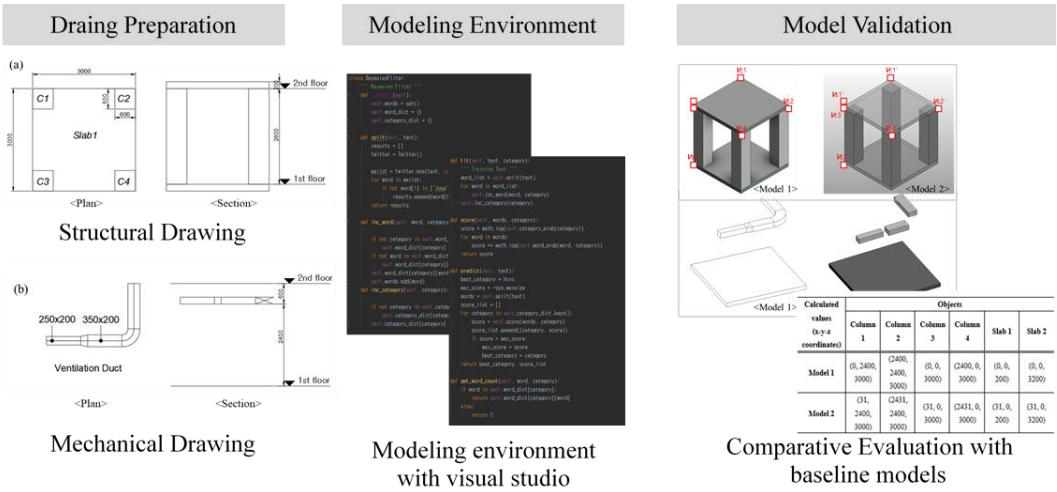


Figure 5-7. Model development algorithm



Calculated values (x-y-z coordinates)	Objects					
	Column 1	Column 2	Column 3	Column 4	Slab 1	Slab 2
Model 1	(0, 2400, 3000)	(2400, 2400, 3000)	(0, 0, 3000)	(2400, 0, 3000)	(0, 0, 200)	(0, 0, 3200)
Model 2	(31, 2400, 3000)	(2431, 2400, 3000)	(31, 0, 3000)	(2431, 0, 3000)	(31, 0, 200)	(31, 0, 3200)

Figure 5-8. Data preparation and validation with baseline

For acquisition of location information, a commercial modeling software from Changsoft (Changsoft, 2019) was used. The modeling environment for automatic object creation was built in Microsoft Visual Studio, as shown in Figure 5-9.

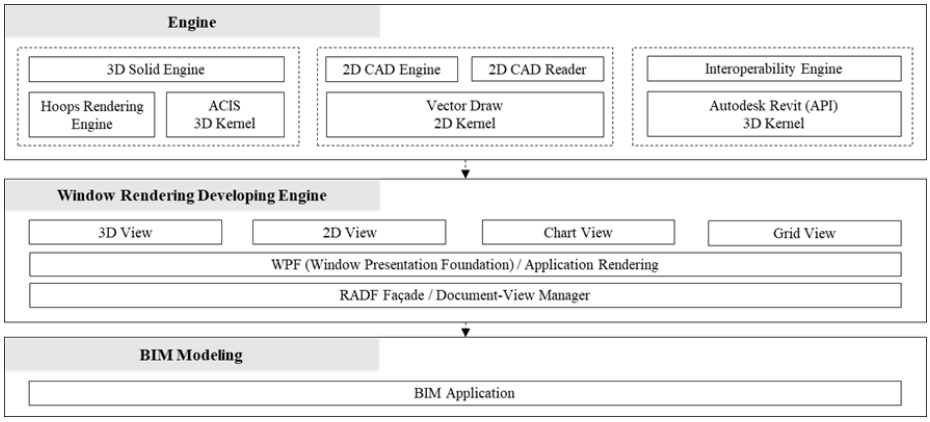


Figure 5-9. Building environment for the second experiment of automated object generation built in Microsoft Visual Studio

5.2.2. Test Drawings Preparation

Prior to the experiment, two sets of test drawings were prepared to test the 3D shape creation experiments. Considering that this was only a preliminary test, sets of plane and section drawings of a simple structure with two- slabs and four- columns that are capable of basic spatial composition was created. Similarly, a mechanical drawing with two types of ventilation ducts was created (Fig. 5-10).

Structure drawing (a) includes four concrete columns with a square planar shape and a size of 600 × 600 mm and height of 2,800 mm. The two slabs at top and bottom are identical, having a planar size of 3,000 × 3,000 mm with a thickness of 200 mm and a height difference of 3,000 mm. In order to test objects with different shapes, another mechanical drawing that included a duct located in a ceiling was used (b). The duct is located at a height of 2,400 mm from the floor. The sectional plane sizes are 250 × 200 mm and 350 × 200 mm, and the length is identically 1000 mm(excluding the fitting units).

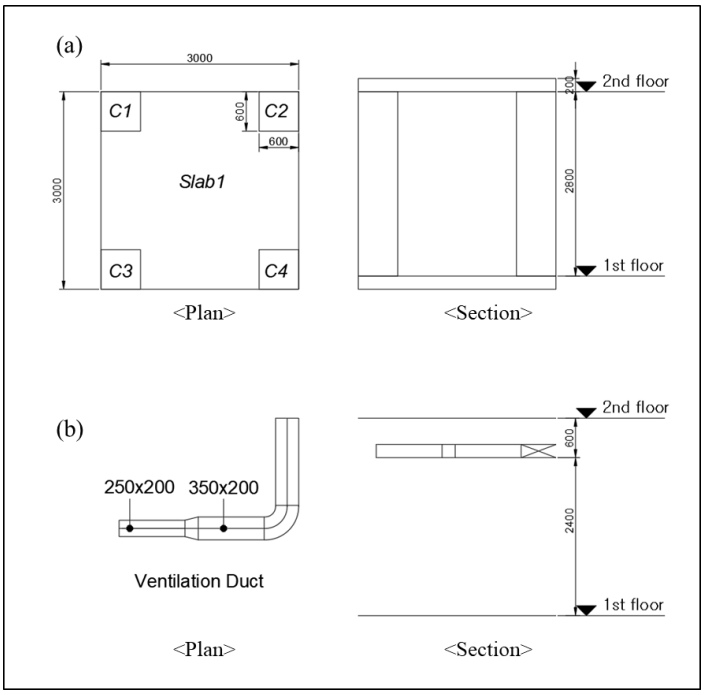


Figure 5-10. Test sample drawings for 3D creation experiment

5.2.3 Results and Validation

Baseline for Comparative Evaluation

To validate the shape created through pattern recognition, we compared the results with a baseline model using the previously discussed test drawings. The baseline model was created manually using a conventional object modeler, Revit ver. 2017 from Autodesk. The results are compared in Figure 5-11 and Table 5-3. In this study, the experiment was conducted under the assumption that the building element, which is the result of the text classifier, is brought into the basic object for modeling.

Results

This evaluation of building elements was more focused on the resultant output from the model creation at the building level. As the aim of our research is to provide an automated BIM modeling framework with its environment, this is more compliant with the objective of BIM model that has geometric and topological information. To be specific, to compare the results of these two models, the shape displayed on the screen and the origin of the generated object models were compared, and theit area and volume of were also validated.

Test Structure Drawing

The shapes of the two models for the test structural drawing are shown in Figure 5-12. The origin points of object model from the baseline are referred to as <vs1>, <vs2> for slabs, and as <vc1, vc2,..vc4> for columns. We added on apostrophe in front of object names to signify objects of the proposed model. Each origin position value converted to Cartesian coordinates is shown in Table 5-3. To check the numerical information associated with the create objects, the area and volume of the generated object in each model area was confirmed; the results are shown in Table 5-4.

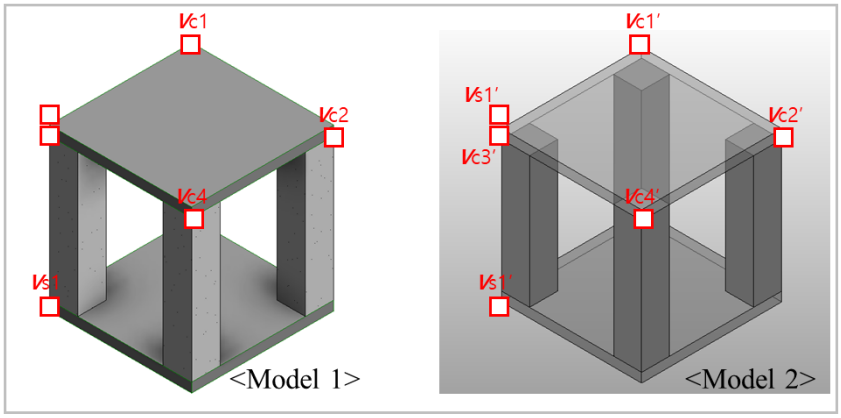


Figure 5-11. Comparison of created structure models: model 1 is the baseline and model 2 is the research model

Table 5-3. Converted coordinates of each object's origin points

Calculated values (x-y-z coordinates)	Objects					
	Column 1	**Column 2**	**Column 3**	**Column 4**	**Slab 1**	**Slab 2**
Model 1	(0, 2400, 3000)	(2400, 2400, 3000)	(0, 0, 3000)	(2400, 0, 3000)	(0, 0, 200)	(0, 0, 3200)
Model 2	(31, 2400, 3000)	(2431, 2400, 3000)	(31, 0, 3000)	(2431, 0, 3000)	(31, 0, 200)	(31, 0, 3200)

Table 5-4. Quantity of area and volume of generated object models

		Objects					
		Column 1	**Column 2**	**Column 3**	**Column 4**	**Slab 1**	**Slab 2**
Model 1	Area	7.44	7.44	7.44	7.44	20.4	20.4
	Volume	1.008	1.008	1.008	1.008	1.8	1.8
Model 2	Area	7.44	7.44	7.44	7.44	20.4	20.4
	Volume	1.008	1.008	1.008	1.008	1.8	1.8

Test MEP Drawing

Figure 5-10 shows the size and location for test mechanical duct drawing. Due to the feature of objects that are not on the reference level, the origin points of each created model is converted into x-y-z coordinates and the length of the three duct items are measured (Table 5-5). As the 3D shape creating environment for this experiment did not develop a duct fitting object type, the reducer and elbow connectors between duct objects were omitted as shown in model 2 in Figure 5-12.

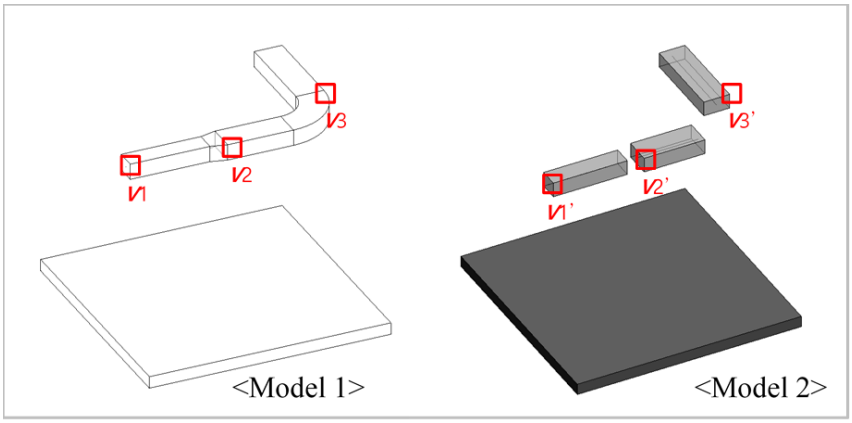


Figure 5-12. Comparison of created duct models: model 1 is the baseline and model 2 is the research model

Table 5-5. Converted coordinates of object's origin points and length

		Objects		
		Duct 1	**Duct 2**	**Duct 3**
Model 1	**x-y-z coordinates**	(275, -3000, 2100)	(1475, -3000, 2100)	(3000, -24000, 2100)
	Length	1000	1000	1000
Model 2	**x-y-z coordinates**	(275, -3000, 2100)	(1475, -3000, 2100)	(3000, -24000, 2100)
	Length	1000	1000	1000

5.3 Discussion

Two separate experiments were conducted to confirm Hypothesis 1, and the following results of experiment were derived.

Result Analysis

The first experiment's result showed data acquisition and storage without any loss of design data. Additionally, the cross validation results (Table 5-2) for the classifying model showed the high stability. The standard score that evaluates a classifier to have stable classification performance is 0.8 (Zhang, 2015); therefore, our text classifier with a result of 0.974 F1-measure and 0.975 G-mean can be determined to be stable. The result of the second experiment (Table 5-3, Figure 5-9) showed the location deviation compared to the baseline. The comparison between the created models using the basic location and line information of the project with the baseline indicated, six total mapped objects with the correct size. However, there was a difference between these two models in terms of their positioning. The relative position of each object was created correctly; however, the object from our developed environment had a position deviation of 31mm along the x-axis and it was identical for all objects, which is considered an error in origin point recognition. In addition, for the 3D shape creation test with the

duct sample drawings, even though the fitting items were omitted, the created duct items with the baseline were confirmed to be identical. Hence, based on these results the feasibility of the proposed framework was confirmed.

Research Hypothesis and Expected Practical Application

From the aforementioned experiments, it was confirmed that the generating object model is not limited to a specific shape or work type. These experiments were planned to test the first research hypothesis, and the results show that this approach will resolve the problems of construction projects that prepare data in duplicate and thereby suffer from time shortages and cost overruns. The developed framework, which includes the data separation-conversion-integration process with related technologies is expected to be used as the basis for automatic data conversion development from 2D drawing to BIM data in practical applications. However, this research conducted research hypotheses and experiments in separate procedures, and the automation of the entire process was not fulfilled. For technical supplementation to achieve theoretical research hypotheses and objectives, future research should carry out research proposal and following experiment on platform that can incorporate individual processes.

5.4 Summary

With the purpose to confirm the applicability of the proposed framework to engineering drawings of the real construction projects, this chapter conducted object identification and pattern recognition-based object generation experiments with an imaginary drawings. Environment for two experiments were prepared respectively: one is for text classification based on the Bayesian classifier with four sets of abbreviation applied in the construction drawings in the Python environment; and the other is to configurate object model with the graphic information from drawings in the visual studio environment.

The developed text classifier showed high performance for model evaluation and cross validation, and the accurate test result. The 3D shape creation experiments with two sample drawings also showed the results of geometric and topological shape acquisition for the main objects. The proposed framework and experiment process to test the first research hypotheses showed effective results, and the individual object model generation based on the drawing recognition was confirmed.

Chapter 6. Project-level Experiment

In order to confirm Hypothesis 2, we conducted this second experiment with the specifications as defined in Chapter 4. To apply the modeling criteria as the only independent variable under the controlled experimental environment, all project BIM data was built with commercial software. Three quantity results were calculated: one with a manual calculation based on 2D drawing; one with the existing modeling method and inherent standards; and the other was modeled with the modeling specification. Quantitative numerical comparison is performed by calculating the relative error and error rate between the other two quantities based on the results of manual calculation (Fig.6-1.). To improve the reliability of the impact of modeling criteria, expanded case application and their results are compared.

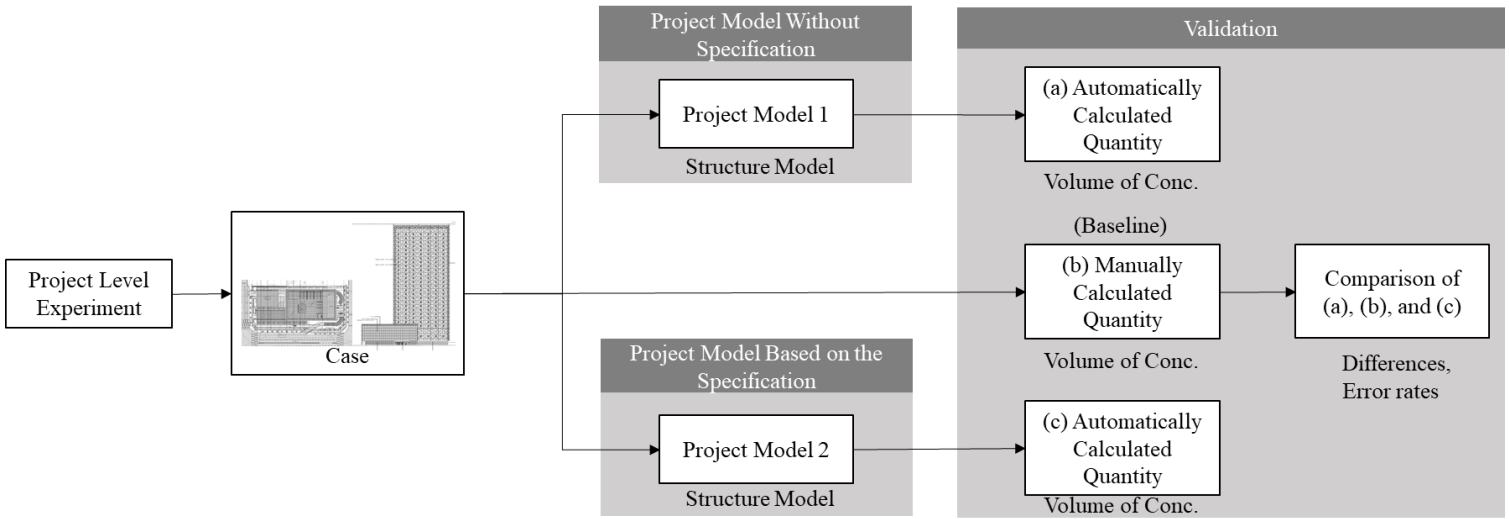


Figure 6-1. Design of experiment

6.1 Design of Experiment

We conducted the experiment based on the existing building to confirm the applicability of the framework and modeling criteria proposed in the previous chapter. Based on the results of the experiments in Chapter 5 on attribute recognition and shape creation, this experiment confirmed the feasibility of calculating quantities, which is a critical information for construction management tasks such as scheduling, planning, and commencement. Two comparison targets were adopted to confirm the adequacy of the numerical values derived from the project model generated based on the framework and criteria in this research study. To be specific, there were two results of estimated quantities: one was the automatically derived from the created BIM model that did not follow the modeling specification, while the other was manually derived using 2D drawings that is the currently employed method in the construction industry (Fig. 6-2).

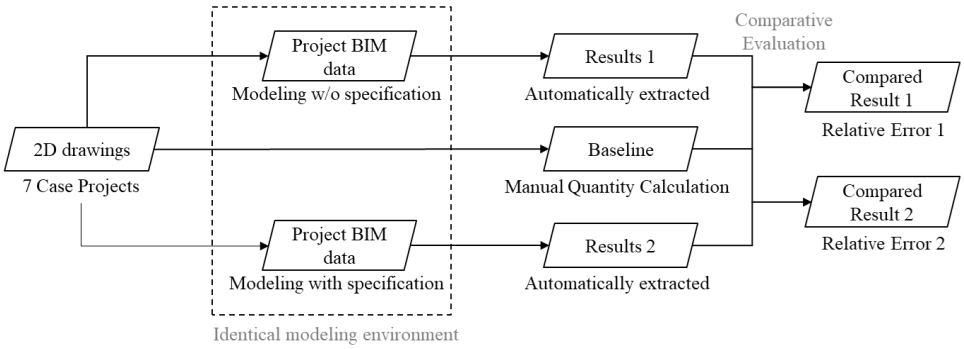


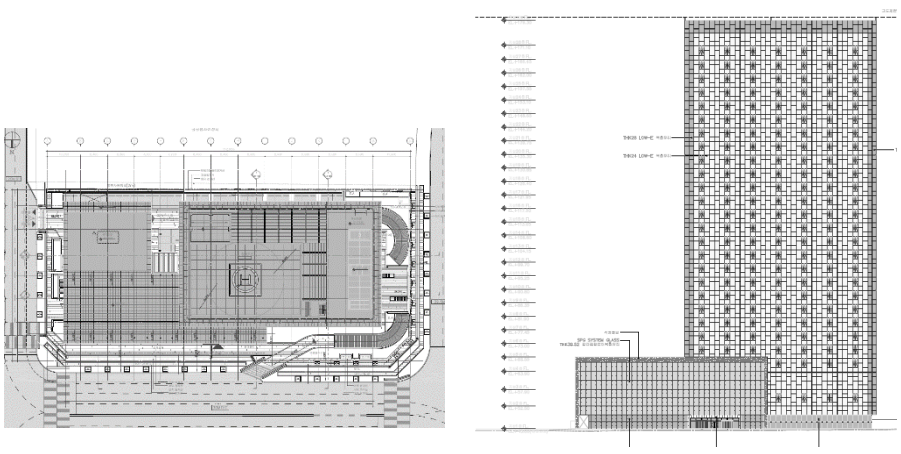
Figure 6-2. Model development algorithm

Considering the initial experiment circumstances, the target was limited to building elements with concrete as material among the structural systems, and it was assumed that there was no additional model modification process like steps on slab or staircases. After performing automatic QTO using the model created by the proposed method for the target project and the model created by the existing modeler, both were compared with the baseline (manually calculated result based on the same project drawings) to confirm their applicability.

The calculation of BIM-based quantities in this chapter was conducted through the material QTO and schedule QTO (among the methods described in Section 2.2). The calculated results were described as the total number and area or volume based on the object type and for each floor of the building.

6.1.1 Case Project Description

The target subject for the case experiment was a building with 2-underground and 28- overground stories completed in 2010 in Seongnam, Gyeonggi-do. It is a high-rise facility with a total floor area of 96,972 ㎡ and the maximum height of 135.3 m; the main structure is made of steel-and- reinforced concrete (Fig. 6-3). To measure precise numerical values, and further comparison, only concrete structure elements among the structural systems were tested.



(a) (b)

Figure 6-3 (a) Site plan drawing and (b) façade of target project

6.1.2 Project BIM Data Creation

Three concrete volume estimations were made to compare the results between the calculation methods and the modeling methods.

- Two different modeling methods: One used a commercial BIM modeling method and the other was based on the modeling specification proposed in this research study.
- Three calculation results: Two automatically calculated concrete volumes from the above-mentioned BIM models, and the third was manually estimated concrete volume using 2D drawings.

The three results were arranged in the order of object types and as floor levels, and then the total amounts were compared. The modeling scope was defined to be 30 stories and we excluded any modeling work that required additional modifications such as stairs and steps on the floor. The modeling was based on the detail level of construction drawings, and the total number of created object models (instance of model creation based on the specification from this research study) were 33 slabs, 805 columns, 779 beams, and 1736 walls (Table 6-1).

Due to the different deduction standards in case of overlap between objects, the two models have different display and number of objects as a result of deduction or Boolean split.

Table 6-1 Number of generated objects in different modeling standards

Building Elements	Traditional Modeling Method	Proposed Modeling Method
Slab	33	33
Column	805	805
Beam	799	779
Wall	1,729	1,736
Total Object Amount	3,366	3,353

6.2 Quantity Comparison by Calculation Methods

6.2.1 QTO with Traditional BIM Method

(1) Deduction of Overlap between Objects in Traditional Method

Commercial BIM solutions in the market have been developed to make it easy to configurate the model easily and quickly, and to create and modify it immediately for use in various stages of one project. Unless otherwise specified from the project team, overlaps between building elements created for immediate communication follow the basic overlapping reaction and further calculation deduction rules defined by each solution (Kim, 2011).

Tables 6-2 and 6-3 explain the priorities for calculating the quantities according to the reaction against the overlapped elements from two BIM solutions that are popular in Korea. Model (1), which is mainly employed for apartments and general buildings with regular-shaped buildings, confirmed that the priority among objects was in the order of Beam>Slab=Wall>Column (Kim, 2011). Modeler (2), which has been recently applied in the majority of cases has a different priority of deduction, and follows the order of Slab = Wall >Beam > Column.

(1) Modeler 1

Properties between object type: Beam>Slab=Wall>Column

Table 6-2 Priority and deduction basis for overlapped objects

	Slab	Column	Beam	Wall
Slab		Column<Slab	Beam>Slab	Wall=Slab
Column	Slab>Column		Beam>Column	Wall>Column
Beam	Slab<Beam	Column<Beam		Wall<Beam
Wall	Slab=Wall	Column<Wall	Beam>Wall	

(2) Modeler 2

Properties between object type: Slab=Wall>Beam>Column

Table 6-3 Priority and deduction basis for overlapped objects

	Slab	Column	Beam	Wall
Slab		Column<Slab	Beam<Slab	Wall=Slab
Column	Slab>Column		Beam>Column	Wall>Column
Beam	Slab>Beam	Column<Beam		Wall>Beam
Wall	Slab=Wall	Column>Wall	Beam<Wall	

(2) QTO of Models in Traditional Modeling Method

In this research study, we utilized Modeler (2), and the calculation results of the created model are shown in Table 6-4. The derived material volume of concrete for 3,366 object models resulted in 29,528.58 m³ for slab; 3,360.8 m³ for column; 471.22 m³ for beam or girder; 11,433.07 m³ for wall; and a total sum 44,792.6 m³.

Table 6-4 Quantities with the traditional modeling method

	Slab		Column		Beam		Wall	
Floor	Counts (ea)	Volume (m³)	Counts (ea)	Volume (m³)	Counts (ea)	Volume (m³)	Counts (ea)	Volume (m³)
B1	1	2219.65	65	347.37			85	1330.24
1	2	1319.39	34	247.07	29	15.9	90	760.18
2	1	1092.93	34	178.85	31	17.03	103	581.76
3	2	883.12	34	190.26	27	16.64	84	605.39
4	1	1355.66	32	124.99	25	15.86	72	394.27
5	1	1277.46	26	101.33	25	15.86	52	313.1
6	1	899.08	26	98.12	26	15.54	52	313.1
7	1	899.08	26	98.12	26	15.54	52	313.1
8	1	899.08	26	98.12	26	15.54	52	313.1
9	1	899.08	26	98.12	26	15.54	52	313.1
10	1	899.08	26	98.12	26	15.54	52	313.1
11	1	899.08	26	98.12	26	15.54	52	313.1
12	1	899.08	26	98.12	26	15.54	52	313.1
13	1	899.08	26	98.12	26	15.54	52	313.1

14	1	899.08	26	98.12	26	15.54	52	313.1
15	1	899.08	26	98.12	26	15.54	52	313.1
16	1	899.08	26	98.12	26	15.54	52	313.1
17	1	899.08	26	98.12	26	15.54	52	313.1
18	1	899.08	26	98.12	26	15.54	52	313.1
19	1	899.08	26	98.12	26	15.54	52	313.1
20	1	899.08	26	98.12	26	15.54	52	313.1
21	1	899.08	26	98.12	26	15.54	52	313.1
22	1	899.08	26	98.12	26	15.54	52	313.1
23	1	899.08	26	98.12	26	15.54	52	313.1
24	1	899.08	26	98.12	26	15.54	52	313.1
25	1	899.08	26	98.12	26	15.54	52	313.1
26	1	899.08	26	98.12	26	15.54	52	313.1
27	1	899.08	26	102.53	26	15.54	54	293.28
28	1	884.99	7	7.08	27	25.78	51	352.32
28.5	1	334.29	1	0.8	21	12.1	46	227.43
Roof	1	380.33	0	0	42	10.17	0	0
Total	33	29527.58	805	3360.8	799	471.22	1729	11433.07

6.2.2 2D Drawing-based Calculated QTO

(1) Manual Estimating Method Based on 2D Drawings

The method of calculating quantities using 2D drawings is described in Chapter 3, and its equation for each object type when the structure system is reinforced concrete is shown in Figure 6-4. This method is the basic formula used for quantity calculations currently in the construction industry. The concrete quantity of the case project is derived in the next section using the formula that manually checks and calculates the numerical dimensions identified in the drawings.

Calculating volume of RC elements

1. Slab
   Vs = ts*(ls * lo)
2. Column
   Vc = (c*d)*(H-ts)
3. Beam
   Vb = (a-ts)*b*(ls-d)
4. Wall
   Vw = (H-ts) *tw*(ls-d)

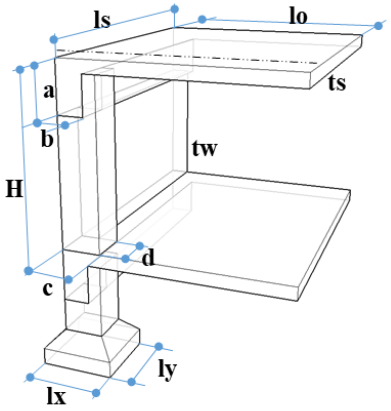


Figure 6-4 Reinforced concrete volume calculation equation

(2) QTO of Models in Traditional Modeling Method

The calculated results for the created model using the traditional modeling method is shown in Table 6-5. The derived volume for total 1,736 concrete material are 29,527.58 m³ Slab, 3,032.97 m³ column, 471.22 m³ beam or girder, 10,964.46 m³ wall, and a total sum 45,996.23 m³.

Table 6-5 Quantities with manual calculation

	Slab		Column		Beam		Wall	
Floor	Counts (ea)	Volume (m³)	Counts (ea)	Volume (m³)	Counts (ea)	Volume (m³)	Counts (ea)	Volume (m³)
B1	1	2219.65	65	329			85	1254.17
1	2	1319.39	34	234	29	15.9	93	673.37
2	1	1092.93	34	160.5	31	17.03	103	550.45
3	2	883.12	34	147	27	16.64	85	472.45
4	1	1355.66	32	115.6	25	15.86	72	385.24
5	1	1277.46	26	92.8	25	15.86	53	292.5
6	1	899.08	26	88.3	26	15.54	52	292.5
7	1	899.08	26	88.3	26	15.54	52	292.5
8	1	899.08	26	88.3	26	15.54	52	292.5
9	1	899.08	26	88.3	26	15.54	52	292.5
10	1	899.08	26	88.3	26	15.54	52	292.5
11	1	899.08	26	88.3	26	15.54	52	292.5
12	1	899.08	26	88.3	26	15.54	52	292.5
13	1	899.08	26	88.3	26	15.54	52	292.5
14	1	899.08	26	88.3	26	15.54	52	292.5
15	1	899.08	26	88.3	26	15.54	52	292.5
16	1	899.08	26	88.3	26	15.54	52	292.5

17	1	899.08	26	88.3	26	15.54	52	292.5
18	1	899.08	26	88.3	26	15.54	52	292.5
19	1	899.08	26	88.3	26	15.54	52	292.5
20	1	899.08	26	88.3	26	15.54	52	292.5
21	1	899.08	26	88.3	26	15.54	52	292.5
22	1	899.08	26	88.3	26	15.54	52	292.5
23	1	899.08	26	88.3	26	15.54	52	292.5
24	1	899.08	26	88.3	26	15.54	52	292.5
25	1	899.08	26	88.3	26	15.54	52	292.5
26	1	899.08	26	88.3	26	15.54	52	292.5
27	1	899.08	26	92.7	26	15.54	55	292.5
28	1	884.99	7	6.37	27	25.78	52	348.69
28.5	1	334.29	1	0.7	21	12.1	46	348.69
Roof	1	380.33	0	0	42	10.17	0	203.9
Total	33	29527.58	805	3032.97	799	471.22	1736	10964.46

6.2.3 QTO with the Proposed BIM Method

(1) Deduction of Overlap between Objects

The object modeling standards for quantity calculation are described in Chapter 3, and the priority of each object type is: Slab>Column>Beam>Wall.

(2) QTO of Models in Proposed Modeling Method

Object models of experiment target project were created excluding the overlapped area to be subtracted by this procedure, and the results of the quantity are shown in Table 6-6.

Table 6-6 Quantities from the model with the specification

	Slab		Column		Beam		Wall	
Floor	Counts (ea)	Volume (m^3)	Counts (ea)	Volume (m^3)	Counts (ea)	Volume (m^3)	Counts (ea)	Volume (m^3)
B1	1	2219.65	65	324.21			85	1330.24
1	2	1319.39	34	235.05	29	4.83	93	718.07
2	1	1092.93	34	168.75	31	18.21	103	555.63
3	2	883.12	34	175.03	27	17.07	85	579.17
4	1	1355.66	32	116.7	25	16.92	72	377.07
5	1	1277.46	26	92.85	25	17.56	53	282.66
6	1	899.08	26	88.39	26	17.83	52	282.66
7	1	899.08	26	88.39	26	17.82	52	282.66
8	1	899.08	26	88.39	26	17.82	52	282.66
9	1	899.08	26	88.39	26	17.82	52	282.66

10	1	899.08	26	88.39	26	17.82	52	282.66
11	1	899.08	26	88.39	26	17.82	52	282.66
12	1	899.08	26	88.39	26	17.82	52	282.66
13	1	899.08	26	88.39	26	17.82	52	282.66
14	1	899.08	26	88.39	26	17.82	52	282.66
15	1	899.08	26	88.39	26	17.82	52	282.66
16	1	899.08	26	88.39	26	17.82	52	282.66
17	1	899.08	26	88.39	26	17.82	52	282.66
18	1	899.08	26	88.39	26	17.82	52	282.66
19	1	899.08	26	88.39	26	17.82	52	282.66
20	1	899.08	26	88.39	26	17.82	52	282.66
21	1	899.08	26	88.39	26	17.82	52	282.66
22	1	899.08	26	88.39	26	17.82	52	282.66
23	1	899.08	26	88.39	26	17.82	52	282.66
24	1	899.08	26	88.39	26	17.82	52	282.66
25	1	899.08	26	88.39	26	17.82	52	282.66
26	1	899.08	26	88.39	26	17.82	52	282.66
27	1	899.08	26	93.71	26	17.82	55	287.65
28	1	884.99	7	6.61	27	26.52	52	346.38
28.5	1	334.29	1	0.7	22	11.89	46	197.7
Roof	1	380.33	0	0	21	11.89	0	0
Total	33	29527.58	805	3069.8	779	516.94	1736	10610.43

6.3 Results and Validation

Table 6-7 shows the results of calculating the concrete quantities of the project using the three different methods: (a) is the result of automatic calculation from the project model created by using the existing modeler, (b) is the drawing-based quantity calculation result, and (c) is the estimate based on the proposed modeling specification for quantity estimation. The deduced deviation between the results of the existing modeling method and the method proposed in this research study are +764 and -303, respectively, with the 2D drawing-based manual calculation as the baseline.

In addition, to confirm the degree of deviation compared to the baseline for each object type, the following relative error rate calculation formula was applied (Table 6-8).

Relative Error Rate = (Quantity based on 2D drawings - Calculated quantity) / (Quantity based on 2D drawings) X100 (%)

Table 6-7 Comparison of total concrete amount for each element

Concrete Volume for Each Element	(a) Traditional Modeling Method	(b) Manual Calculation Method	(c) Proposed Modeling Method
Slab	29,527.58	29,527.58	29,527.58
Column	3,360.8	3,032.97	3,069.8
Beam	471.22	503.37	516.94
Wall	11,433.07	10,964.46	10,610.43
Total	44,792.67 (+764)	44,028.38	43,724.75 (-303)

Table 6-8 Relative error rate of quantity comparison based on the manual approach

Concrete Volume for Each Element	Traditional Modeling Method (%)	Manual Calculation Method (Baseline, m^3)	Proposed Modeling Method (%)
Slab	0	29,527.58	0
Column	+10.8	3,032.97	+1.21
Beam	-6.38	503.37	+2.7
Wall	+4.27	10,964.46	+3.2
Total	+1.73	44,028.38	-0.63

6.4 Results of Extended Cases Experiment

In order to improve the reliability of the numerical results, the number and variety of case projects were increased and they were then processed the same way as described earlier to calculate the relative error and error rate for each calculated quantity. These case projects included: two residential buildings, an office building, a hotel, and a hospital projects (Figure 6-5). Tables 6-9 to 6-18 describe the total estimation per object types and comparative evaluation of relative error and error rates for each building.

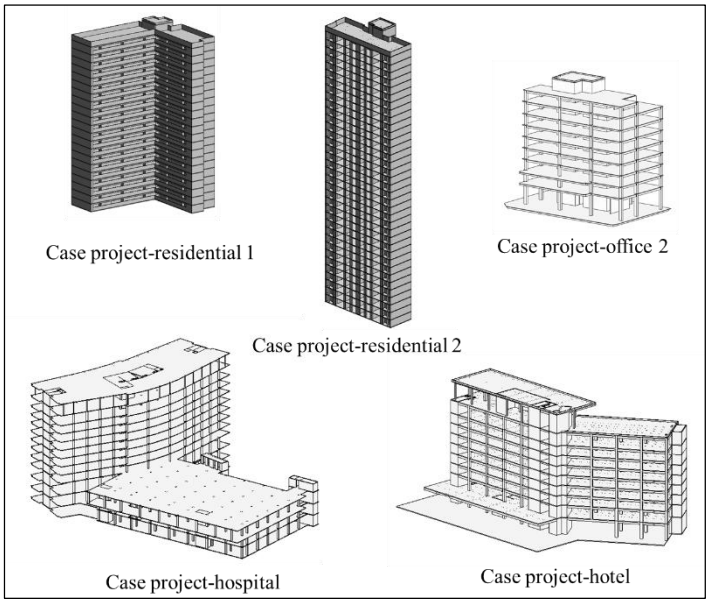


Figure 6-5. 3D model for case project experiments

Case Project 2. Residential Building 1

Table 6-9 Comparison of total concrete amount for each element

Concrete Volume for Each Element	(a) Traditional Modeling Method	(b) Manual Calculation Method	(c) Proposed Modeling Method
Slab	2,721.66	2,721.66	2,721.66
Wall	3,399.51	3,218.06	3,153.9
Total	6,121.17 (+181.45)	5,939.72	5,875.56 (-64.16)

Table 6-10 Relative error rate of quantity comparison based on the manual approach

Concrete Volume for Each Element	Traditional Modeling Method (%)	Manual Calculation Method (Baseline, m³)	Proposed Modeling Method (%)
Slab	0	2,721.66	0
Wall	+5.64	3,218.06	-1.99
Total	+5.64	5,939.72	-1.99

Case Project 3. Residential Building 2

Table 6-11 Comparison of total concrete amount for each element

Concrete Volume for Each Element	(a) Traditional Modeling Method	(b) Manual Calculation Method	(c) Proposed Modeling Method
Slab	2,582.43	2,582.43	2,582.43
Wall	4,529.58	4,238.18	4,214.36
Total	7,112.01 (+291.3)	6,820.71	6,796.79 (-23.92)

Table 6-12 Error rate of quantity comparison based on the manual approach

Concrete Volume for Each Element	Traditional Modeling Method (%)	Manual Calculation Method (Baseline, m³)	Proposed Modeling Method (%)
Slab	0	2,582.43	0
Wall	+6.87	4,238.18	-0.56
Total	+6.87	6,820.71	-0.56

Case Project 4. Office Building 2

Table 6-13 Comparison of total concrete amount for each element

Concrete Volume for Each Element	(a) Traditional Modeling Method	(b) Manual Calculation Method	(c) Proposed Modeling Method
Slab	1,297.52	1.297.52	1.297.52
Column	181.48	181.48	181.47
Beam	970.39	958.47	956.56
Wall	1,036.53	1,008.11	950.65
Total	3,485.92 (+37.34)	3,445.58	3,386.2 (-59.35)

Table 6-14 Relative error rate of quantity comparison based on the manual approach

Concrete Volume for Each Element	Traditional Modeling Method (%)	Manual Calculation Method (Baseline, m³)	Proposed Modeling Method (%)
Slab	0	1.297.52	0
Column	0	181.48	0
Beam	+1.24	958.47	-0.2
Wall	+2.82	1,008.11	-5.6
Total	4.06	3,445.58	-5.8

Case Project 5. Hospital

Table 6-15 Comparison of total concrete amount for each element

Concrete Volume for Each Element	(a) Traditional Modeling Method	(b) Manual Calculation Method	(c) Proposed Modeling Method
Slab	10,060.42	10,060.42	10,060.42
Column	1,543.21	1542.9	1542.71
Wall	2,875.18	2,774.9	2,753.04
Total	14,478.81 (+100.59)	14,378.22	14,356.17 (-22.05)

Table 6-16 Relative error rate of quantity comparison based on the manual approach

Concrete Volume for Each Element	Traditional Modeling Method (%)	Manual Calculation Method (Baseline, m³)	Proposed Modeling Method (%)
Slab	0	10,060.42	0
Column	+0.2	1542.9	-0.01
Wall	+3.6	2,774.9	-0.78
Total	+3.8	14,378.22	-0.79

Case Project 6. Hotel

Table 6-17 Comparison of total concrete amount for each element

Concrete Volume for Each Element	(a) Traditional Modeling Method	(b) Manual Calculation Method	(c) Proposed Modeling Method
Slab	3,336.15	3,336.15	3,336.15
Column	441.42	440.6	440.43
Wall	1,134.35	1,087.22	1,076.69
Total	4,911.92 (+74.95)	4,836.97	4,853.27 (-10.7)

Table 6-18 Relative error rate of quantity comparison based on the manual approach

Concrete Volume for Each Element	Traditional Modeling Method (%)	Manual Calculation Method (Baseline, m³)	Proposed Modeling Method (%)
Slab	0	3,336.15	0
Column	+0.19	440.6	-0.03
Wall	+4.33	1,087.22	-0.97
Total	+4.52	4,836.97	-0.1

6.5 Discussion

Result Analysis

From the results of the first case experiment, it was confirmed that the relative error rate were +1.73% and -0.63% using the traditional modeling method and the proposed modeling method, respectively, compared to the calculation results, and that the model for the purpose of the calculation based on the estimation standards also showed a difference from the 2D drawing-based manual method. Despite the more accurate quantity value with the proposed method, some numerical deviation remains. These differences are considered to be a consequence of the difficulty in understanding of 3D information with the manual calculation process that is based on 2D graphic of planar and/or cross sectional engineering drawings.

As the research limited its target of experiment to be construction structure system of concrete material, the error that would occur when the scope and target are extended to the entire project is expected to be greater. In addition, considering that the numerical standard employed for procurement and implementation of the current construction project is based on the result of manual work, the error that is expected to occur continuously should be deemed as an additional considerations so that the construction manager can understand the deviation and prepare for an unexpected

situation resulting from numerical differences.

Table 6-19 Comprehensive relative error rate of quantity comparison for case projects

Case Projects	Traditional Modeling Method (%)	Proposed Modeling Method (%)
Case 1	+1.73	-0.63
Case 2	+5.64	-1.99
Case 3	+6.87	-0.56
Case 4	+4.06	-5.8
Case 5	+3.8	-0.79
Case 6	+4.52	-0.1

From the results of additionally performed case experiments, it was confirmed that the calculated quantities with the modeling criteria are close to manual calculation, except in one case project. From Table 6-19, with a minimum deviation of 0.1% and a maximum deviation of 1.99%, the reliability of calculated quantities from BIM data at construction projects was confirmed and consequently, the second research hypothesis was confirmed through these experimental results. The derived error rates should be considered as a relative extra error rate. The automatically derived quantities based on the model created in consideration of the specification will help construction managers working with traditional calculation methods.

6.6 Summary

This chapter described the experiments conducted to test Hypothesis 2 and to confirm the applicability of the modeling specification defined in Chapter 3. The quantities were calculated in three different ways to confirm the applicability of the modeling criteria in this research study for QTO-based modeling. For the 30-story office building constructed in 2010, the results of automatic volume calculation of the model according to the existing object creation method as well as the proposed method were compared with the drawing-based manual calculations.

The comparison of three results can be seen in Table 6-19. As shown in the table, the modeling method without QTO consideration showed an error of +1.7~6.87% compared to the current calculation method, but the quantity based on the proposed modeling method showed -0.6~1.99%, which is a much smaller error rate range. It can be seen that the difference in numerical values still exists because drawing-based calculation of quantities cannot take into account the 3D relationship between objects. However, because the details used as the basis for procurement and implementation used at construction projects are estimates based on drawing-oriented methods, the error rates should be understood and considered accordingly.

Chapter 7. Conclusions

The objective of this dissertation was to develop a framework that includes a detailed process with applicable techniques for the automated generation of a 3D object model for construction projects based on drawing recognition and text-line extraction. This would have helped address the following problems: (a) Used of traditional 2D-based project information creation and utilization; (b) duplicity in project data creation task due to the mandatary use of BIM in construction projects; and (c) consequent delays in construction commencement.

In this chapter, we summarize the findings derived from the results of this dissertation. Then the expected contributions, both in academic and in practice, are discussed. Finally, the last section concludes with the limitations that will be handled with further studies.

7.1 Research Results

The question addressed in this research study was how to shorten and further automate the data transformation process from 2D drawing to 3D object models with properties that can be adopted in practice. Furthermore, the aimed was to have a discriminative ability to extract the numerical values required in a construction project regardless of the shape or type of the target object. Unlike the previous researches with similar research objectives, the research studies scope was not limited to a specific work trade or building elements. This was achieved by using an independent conversion processes for text and lines. This dissertation was based on two research hypotheses: 1) ***that a separate conversion process of shape and property creates object model regardless of shape or work trade***; and 2) ***that modeling criteria with estimation standards facilitate quantity extraction during the construction phase***.

To achieve this, current efforts in terms of extracting information from text or graphic–oriented documents as well as how to create a 3D model from 2D shape through a pattern recognition process were contemplated. Then, it was confirmed that the existing drawing-based automated 3D model creation and its applications are limited in term of applicability. Therefore, this study developed a framework for creating objects based on drawing recognition,

considering the used of precise numerical values in the construction stage. This included: – (a) the overall process of automated modeling based on drawing recognition, (b) drawing recognition and line-text extraction, (c) text classification to obtain object's properties, (d) pattern recognition to configurate object shape, and (e) object modeling specification for quantity calculation.

In addition, the performance of the model and the applicability of the framework were verified by conducting separate tests. The first tested extraction of abbreviations from engineering drawings through the text classifier. The second tested 3D shape generation using pattern recognition. From the text classifier experiment, it was confirmed that the building components were distinguished with a high probability for a new input term generated by similar abbreviation creating rule as a result of trained and tested object-referenced abbreviations applied in construction drawings. In the second experiment, 3D shape generation based on the pattern recognition was performed with two sets of test drawings and the feasibility of the proposed framework and technology was confirmed.

Based on this discussion, it was confirmed that (a) the calculated numerical values required in the construction phase were derived. To be specific, case experiments were conducted to confirm the project

applicability in the construction stage with the numerical values of data generated based on the model preparation criteria. The calculated quantities (b) through automatic overlap deduction from the existing modeling process and (c) based on the two-dimensional drawing-were compared with the deductive value from the model based upon the modeling specification.

While the quantities derived from (a) and (b) showed numerical differences, the deviation was smaller than the deviation between (b) and (c) and it was in the acceptable range. This difference between (a) and (b) was analyzed as a non-considered priority relationship among multiple objects occurring in three-dimensions. However, the current construction planning and management is based on (b), so it can be effective to consider the estimated error rate with (a).

7.2 Research Contribution

This study was conducted with the purpose of solving the problem of duplicated tasks carried out due to the difference between the government-mandated system and the workers' preference. It is hoped that through this work, it is possible to reduce the workload of the projects and further contribute to resolving the chronic problems of time shortcoming and cost overruns that are common in the construction industries.

The developed framework that generates shape and identifies attribute information and then aggregates these as an integrated object information, is a distinctive feature of our study. From a construction management perspective, this research study develops a modeling specification capable of an accurate quantity calculation. The output quantity values provided by the proposed methodology were found to be similar to the drawing-based quantity estimation, thereby providing practically applicable numerical information. In addition, it is possible to skip the separated drawing-based calculation process by providing an error rate to be considered for numerical deviation due to the lack of precision in the drawing-based quantity process.

As an academic contribution, this research study has developed the framework by integrating knowledge and techniques from various fields.

Specifically, machine learning technology was used for text classification, and pattern recognition was used for the 3D object creation process. In addition, knowledge of architectural arithmetic was applied to the modeling specification, in the form of relationships between objects, as the criteria for preparing the project BIM data. The study also contributed by enlarging the body of knowledge in the field of construction BIM research.

By reducing the time and efforts required to convert drawing data into BIM data, it is expected that the study will help improve the efficiency and effectiveness of construction management. Specifically, it will improve productivity in the early construction stage or at the construction documents preparation stage. Project data created in 2D drawings and 3D BIM data will ensure smoother communication among various participants in construction projects. For example, by automatically creating identical project data, the construction manager, who may bes familiar with the BIM environment will be able to understand the project information without communication problem with the designers or sub-consultants who may be accustomed to a 2D-base work environment.

7.3 Future Research

It is acknowledge that the proposed framework needs to be improved in order to make it applicable various circumstances. For instance, in buildings that include cross-sectionally irregular building components, these object models do not correspond to the pattern recognition to configurate, so it is difficult to create their shape using an automated object model framework. As another example, model creation and quantity calculation cannot be made in the case of finishing materials in which construction and modeling bases are composite materials rather than individual objects.

To address these limitations, the creation of an object shape considering the 3D configuration of an object that is not in a faceted shape and further calculating for these atypical shaped objects needs to be considered. For example, a supplementary equation of iterative integral can be adapted for area and volume calculation methods or for shape creation. Otherwise in case of composite materials, a rule for object creation based on the finishing material specification should be considered and then be integrated to the current framework.

In order to take an advantage of existing approaches, future research

studies can complementary integrate the framework and modeling specification for expanding the applicability of this research. Future research would expand the framework and specification for further applicability for automatic BIM generation, and would support technical platform that integrates conceptual and technical development including specifications that can contribute both academically and practically. Considering the scalability to specific information required at the field level in addition to the basic project data in this research study, future research should include more detailed numerical information combined with construction schedule, daily progress management, cost information, and role of sub-consultants.

It can also be said that this research has followed a passive approach in terms of solving the research problem at the construction stage. Thus, to enable a positive feedback effect that can contribute to improve productivity in the design phase, further research should lead to an active conversion between drawings and BIM data from the early design stage. This implies that the results of this research can be continuously applied in an alternative project delivery method environment, such as of design build (DB) system where the design and the construction services are contracted simultaneously by one entity.

Bibliography

Ahmend, S., Liwick, M., Weber, M., Dengel, A. (2011), “Improved automatic analysis of architectural floor plans.” International Conference on Document Analysis and Recognition 2011, 864-869

Ahn, S., Lee, H., Park, M., and Kim, W. (2009) “Object-oriented Modeling of Construction Operations for Schedule-Cost Integration, Based on BIM.” Journal of the Architectural Institute of Korea Structure &Construction, Vol.25(12), 165-174

Aish, R. (1986), “Building modeling the key to integrated construction CAD.” THE FIFTH INTERNATIONAL SYMPOSIUM ON THE USE OF COMPUTERS FOR ENVIRONMENTAL ENGINEERING RELATED TO BUILDINGS. 55-67

Anul, E., Tang, P., Akinci, B., and Huber, D. (2013), “Deviation analysis method for the assessment of the quality of the as-is Building Information Models generated from point cloud data.” Automation in Construction, Vol.35, 507-516

American Institute of Architects. (2013), “AIA BIM Protocol”

Azhar, S., Hein, M., Sketo, B. (2008), “Building Information Modeling (BIM): Benefits, Risks, and challenges.” Pro., 44th Associated Schools of construction National Conference, Auburn, AL.

Azhar, S. (2011), “Building Information Modeling (BIM): trends, benefits, risks and challenges for the AEC industry.” Leadership and Management in Engineering, Vol.11(3), 241-252

Barki, H., Fadli, F., Shaat, A., and Boguslawski, P. (2015), “BIM models generation from 2D CAD drawings and 3D scans: an analysis of challenges and opportunities for AEC practitioners.” Building Information Modeling(BIM) in Design, Construction and Operations, Vol.149, 369-380

Bazjanac, V. (2005). “Model based cost and energy performance estimation during schematic design.” Lawrence Berkely National Laboratory, Berkeley

Be Purple et al. (2013), “Research on BIM Development Strategy for Daelim Industrial.” 8-28

Bosche, F., Haas, C. (2008). “Automated retrieval of 3D CAD model objects in construction range images.” Automation in Construction 17. 499-512

Bosche, F., Ahmed, M., Turkan, Y., Haas, C., and Haas, R. (2015) “The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components.” Journal of Automation in Construction, Vol.24, 201-213

Brilzkis, I., Park, M., Jog, G. (2011), “Automated vision tracking of project related entities.” Advanced Engineering Informatics 25 (2011) 713-724

BuildingSMART, Industry Foundation Classes (IFC). Accessed: May 13, 2020. [online]. Available: https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/

Bulgakov, A., Engenov, A., and Weller, C. (2015), “Automation of 3D Building Model Generation Using Quadrotor.” Journal of Creative Construction Conference (CCC), 101-109

Chhabra, A., and Phillips, I. (2000), “Performance evaluation of line drawing recognition systems.” Proceedings of International conference of Pattern Recognition, Vol. 15(4), 864-867

Changsoft I&I. C. Kim. BuilderHub User's Guide Book v1.0. (2015). Accessed: Apr. 16, 2019. [online]. Available: http://chang-soft.com/wp-content/uploads/2015/11/BuilderHub_UsersGuideBook_20151014.pdf

Choi, C. (2011), "What is the problem in introducing BIM technology?" KIMIN Magazine, Vol.1(0), 16-19

Cho, Y. BIM Use Analysis for Establishing BIM Strategic Plan of Domestic Construction company. M.S. Thesis, Hanyang University, 2017.

Choi, J., Shin, J., Kim, M., and Kim, I. (2016), "Development of openBIM-based Energy Analysis Sofrware to Improve the Interoperability of Energy Performance Assessment." Journal of Automation in Construction, Vol. 72, 52-64

Chun, s., and Kim, H. (2012), "A Study on the 3D-based System of Automatic Quantity-calculation and Its Applications." Journal of KIBIM, 124-126

Construction & Economy Research Institute of Korea. (2018), "A Study on Performance Evaluation of Turnkey Ordering Method."

Construction Association of Korea. (2017), "Construction Order Analysis

Czmoch, I., Pekala, A. (2014), “Traditional Design versus BIM Based Design.” (2014). Journal of Procedia Engineering Vol.91, 210-215

Dick, A., Torr, P., and Cipolla, R. (2004), “Modeling and Interpretation of Architecture from Several Images.” International Journal of Computer Vision 60(2), 111-134

Eadie, R, Browne, M., Odeyinka, H., McKeown, C., and McNiff, S. (2013), “BIM implementation throughout the UK construction project lifecycle: An analysis.” Automation in Construction, Vol.36, 145-151

Eastman, C. Fisher, D., Lafue, G., Lividini, J., Stoker, D., and Yessios C. (1974), “An Outline of the Building Description System.” Carnegie-Mellon Univ., Pittsburgh.

Eastman, C., Teicholz, P., Sacks, and R., Liston, K. (2008), BIM Handbook: A guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors. Wiley.

Etzionin-Halevy, E. (1990), “Comparing semi-corruption among parliamentarians in Britain and in Australia.” Comparative Methodology: Theory and Practive I International Social Research, 113-133

Fletcher, L., Katsuri, R. (1988), “A robust algorithm for text string separation from mixed text/graphics images.” IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Vol.10 (6), 910-918

GIM International. (2014). https://www.gim-international.com/content/news/point-clouds-photogrammetry-or-lidar. Accessed May 1, 2019

Gimenez, L., Hippolyte, J., Robert, S., Suard, F. (2015), “Review: reconstruction of 3D building information models from 2D scanned plans.” Journal of Building Engineering 2 (2015), 24-35

Gimenez, L., Robert, S., Suard, F., and Zreik, K. (2016), “Automatic reconstruction of 3D building models from scanned 2D floor plans.” Automation in Construction 63 (2016), 48-56

Grabowski, R. (2010), “CAD & BIM: Is there a Free Pass?.” A resport on Grapjosoft

Hardin, B., and McCool D. (2015). BIM and Construction management, 2nd ed. Indiana, United States: SYBEX

He, H. and Zhang, M. (2016), “Research on Intelligent Construction and Integration Management of a Large-Scale Construction Enterprise Based on BIM Technology.” ICCREM, 89-96

Helfrick, A., and Cooper, W. (2005), “Modern Electronic Instrumentation and Measurement Techniques.” Englewood Cliffs, N.J.: Present Hall

Hoang, T., Tabbone, S. (2008), “Text extraction from graphical document images using sparse representation.” International Workshop on Document Analysis Systems, 143-150

Holzer, D., Are you talking to me? Why BIM alone is not the answer, available online at https://epress.lib.uts.edu.au/conferences/index.php/AASA/2007/paper/viewFile/48/52 2007 (accessed June 2020).

Howard, R., and Bjork.B. (2008). “Building information modeling-Experts’ views on standardization and industry deployment.” Advanced Engineering Informatics, Vol.22(2), 271-280

Huang, J., Kim, H., and Lee, J. (2016), “Prototyping a BIM-enabled Design Tool for the Auto-arrangement of Interior Design Panels.” Society of Design Convergence, Vol.15(5), 71-83

Hungu, C. (2010), Utilization of BIM from Early Design Stage to Facilitate Efficient FM Operations. M.S. Thesis, Chalmers University of Technology.

Hwang, Y. (2004). “Automatic Quantity Takeoff from Drawing Through IFC Model.” Journal of the Architectural Institute of Korea Planning & Design 20(12), 89-97

Jang, (1969). Architectural quantity surveying, Seoul, Korea. Wiley.

jbknowledge, Annual Construction Technology Report, https://jbknowledge.com/2018-construction-technoloty-report-survey. Accessed September 27, 2019

Jeong, W., and Kim, K. (2016), “A Performance Evaluation of the BIM-Based Object-Oriented Physical Modeling Technique for Building Thermal Simulations: A Comparative Case Study.” Journal of Sustainability, Vol.8(648), 1-27

Joo, S., and Jun, H. (2009), “A Study on the BIM-based Material Take Off That is Usable as 4D Simulation.” Conference Paper of Architectural Institute of Korea, Vol.29(1), 335-338

Jung, W., and Kim, K. (2016), “Object-oriented physical modeling technique for building thermal simulations: a comparative case study.” Sustainability 2016, Vol.8, 648, 1-27

Kaneko, T. (1992), “Line structure extraction form line-drawing images.” Journal of Pattern Recognition Vol.25(9), 963-973

Kang, H., and Lee, M. (2007), “A Study on the Development of Construction Business Integrated Model through case studies of BIM.” Conference Paper of Architectural Institute of Korea, Vol.27(1), 61-64

Khaddaj, M., and Srour, I. (2016), “Using BIM to Retrofit Existing Buildings.” International conference on sustainable Design, Engineering and Construction, Vol.145, 1526-1533

Kim, C. “A Study on the Moedling Method for Material Taks-off BIM Based Framework.” Ph.D. Dissertation, The University of Gyeongsang National University, 2011.

Kim, C., Lee, J., Lee, J., and Lee, S. (2010), “Analysis of the Work Breakdown Structure and Cost-Rate for Small Interior Works in Apartment.” Journal of KIBIM, Vol.10(1), 213-219

Kim, H. Development of Performance Assessment Model and System for BIM by Application Phases Using Data Mining Method. M.S. Thesis, Gyeongsang National University, 2016.

Kim, I., Bae, J., and Choi, J. (2017), “Development of Property Mapping Contents between Open BIM Data and Automated Code Checking System.” Korean Journal of Computational Design and Engineering, 118-128

Kim, J., Clayton, M., and Yan, W. (2013), “Parameterize Unban Design Codes with BIM and Object-Oriented Programming.” International Conference on Computer-Aided Architectural Design Research in Asis (CAADRIA), 33-42

Kim, K., and Park, S. (2013), “Sustainable Construction through Integration of IPD and BIM.” Project Management Review, Vol.3(1), 17-34

Kim, M. (2013), “Research on the Collective Space Where the Contemporary Publicness is Presented.” Conference Paper of Architectural Institute of Korea, Vol.29(1), 479-482

Kim, S., Chin, S., Yoon, S., Shin, T., Kim, Y., and Choi, C. (2009), “Automated building information modeling system for building interior to improve

productivity of BIM-based quantity take-off." Information and Computational Technology 2009, 492-496

Kim, U. Deriving functional requirements for BIM-based design e-collaboration. M.S. Thesis, Seoul National University, 2015.
Korean Ministry of Construction and Transportation".

Klein, L., Li, N., and Becerik-Gerber, B. (2012), "Image-based Verification of As-built Documentation of Operational Buildings." Journal of Automation in Construction, Vol.21, 161-171

Koskela, L., Howell, G., Ballard, G., and Tommelein, I. (2002), "The Foundation of Lean Construcion." Design and Construction: Building in Value, 211-226

Lawrence, M., Pottinger, R., Staub-French, S., and Nepal, M. (2014), "Creating flexible mappings between Building Information Models and cost information." Automation in Construction 45 (2014), 107-118

Lee, G., and Lee, K. (2019). "2018 Survey on the Status of BIM Adoption." The BIM 2019, 20, 8-11

Lee, H, Kim, J., Choi, C., and Song, S. (2017). “BIM-based cost estimation by integration with BIM model data and cost Information-Case study on economy evaluation of apartment in Sejong special self-governing city-.” Korean BIM Journal (2017), Vol. 7 (3).

Lee, J. and Um, S. (2013), “BIM Applicability Analysis of the PMIS Modules using Component Business Model.” Architectural Institute of Korea, Vol.29(10), 125-132

Lee, M., Kim, J, Kim, J., and Ham, N. (2010), “A Study on 3D Data Model Development by Normalizing and Method of its Effective Use: Focused on Building Interior Construction.” Journal of the Korean Digital Architecture and Interior Association, Vol.10(3), 11-14

Lee, S., Kim, K., and Yu, J. (2014). “BIM andontology-based approach for building cost estimation.” Automation in Construction 41 (2014), 96-105

Lewis, D., and Crotf, W. (1990), “Term Clustering of Syntactic Phrases.” Proceedings of the 13th annual international ACM SIGIR conference on Research and development in information retrieval. 385-404

Lu, Q., and Lee, S. (2017) “A Semi-automatic Approach to Detect Structural Components from CAD Drawings for constructing As-is BIM Objects.” Journal of Computing in Civil Engineering, 84-91

Lu, T., Tai, C., Bao, L., Su, F., and Cai, S. (2005), “3D Reconstruction of detailed buildings from architectural drawings.” Computer-Aided Design & Applications, Vol.2, Nos, 1- 4, 527-536

Lu, T., Yang, Y., Yang, R., and Cai, S. (2008), “Knowledge extraction from structured engineering drawings.” IEEE 2008 (184), 415-419

Mao, R. (2017) “BIM adoption and Trend in UK, China & HK.” build SMART Conference.

McCallum, A., and Nigam, K. (1988), “A Comparison of Event Models for Naïve Bayes Text Classification.” In Proc. Of the AAAI-98 Workshop on Learning for Text Categotization, 41-48

Min, Y., and Kim, K. (2010), “A Study on the good use of Construction Management Model Based on BIM.” Journal of the Korean Digital Architecture, Interior Associatoin Vol.10(3), 87-95

Oh, H. Kim, J., Kim, M., Lee, J., and Jeong, J. (2011), “The Study on the Structure Design Process and Modeling Guidelines Based on BIM.” Journal of the Architectural Institute of Korea Structure & Construction Vol.27(12), 107-114

Pan, W., Bui, T., Suen, C. (2007), “Text Segmentation from Complex Background Using Sparse Representations.” Ninth International conference on Document Analysis and Recognition (ICDAR 2007)

Pardis, P. (2017), “Future Smart Facilities: State-of-the-Art BIM-Enabled Facility Management.” Journal of Construction engineering and Management, 143(9)

Park, E., Lee, K., and Kim, J. (2000), “3D object-based modeling method for design information sharing and resuing.” Journal of the Architectural Institue of Korea, JAIK. Vol.16(12), 207-216

Park, J., Kim, S., Lee, S., and Song, H. (2009), “Suggesting Solutions when applying Builiding Information Modeling (BIM) to the Korean Construction Industry through Case Studies.” Journal of the Korean Institues of Building Construction, KIC. Vol.9(4), 93-102

Park, S., Kim, K., Kim, J., and Kim, J. (2012), “Customized interior design business process analysis with BIM.” Journal of KIBIM, Vol.2(1), 18-26

Public Procurement Service. (2015), “Basic BIM Guideline for Public Projects in Korea.” V.1(30) 5-96

Public Procurement Service. (2016), “Basic BIM Guideline for Public Projects in Korea.” V.1 (31)

Rahman, C., Sohel, F., Naushad, P., and Kamruzzaman, S. (2001), “Text Classification using the Concept of Association Rule of Data Mining.”

Royal Institute of British Architects. (2012), “BIM Overlay to the RIBA Outline Plan of Work.”

Ruuska, S., Hamalainen, W., Kajava, S., Maghal, M., Matilainen, P., and Mononen, J. (2018), “Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behavior of cattle.” Journal of Behavioural Processes Vol.148, 56-62

Scriven, M. (1967), “The methodology of evaluation.” Perspectives of curriculum evaluation, Vol. 1, 39-83. Chicago, IL: Rand McNally.

Shin, J., Hwang, J., Lee, S., and Lee, S. (2012), “BIM based Reliability Analysis and Automated Quantity Calculation.” Conference paper of Korean Society of Hazard Mitigation Vol.12(3), 46-55

Sohn, Y., Kim, U. (1994), “A Study on the Development of an Integrated Steel structure Modeling System based upon Object oriented Method.” Conference paper of Architectural Institute of Korea, Vol.14(2), 535-1069

Succar, B. (2009) “Building information modeling framework: a research and delivery foundation for industry stakeholders.” Automation in Construction 18 (2009), 357-375

Suzuki, S., Abe, K. (1985), “Topological Structural Analysis of Digitized Binary Images by Border Following.” Computer Vision, Graphics, and Image Processing, Vol.30, (1), 32-46

Takim, R., Harris, M., and Nawawi, A. (2013). “Building Information Modeling (BIM): A new paradigm for quality of life within Architectural, Engineering and Construction(AEC) industry.” Journal of Social and Behavioral Sciences 101, 23-32

Tang, P., Huber D., Akini, B, Lipman, R., and Lytle, A. (2010) “Automatic Reconstruction of As-built Building Information Models from Laser-scanned Point clouds: A Review of Related Techniques.” Journal of Automation in Construction, Vol.19, 829-843

Tombre, K., Tabbone, S., Pelissier, L., Lamiroy, B., and Dosch, P. (2002), “Text/graphics separation revisted.”

Troncoso-Pastoriza, F., Equia-Oller, P., Diaz-Redondo, D., and Granada-Alvarez, E. (2018), “Generation of BIM Data Based on the Automatic Detection, Identification and Localization of Lamps in Buildins.” Sustainale Cities and Society Vol.36, 59-70

Vartiainen, P. (2002), “On the principles of comparative evaluation." Sage Journals, Vol.8(3), 359-371

Xie, H., Tramel, J., and Shi, W. (2011), “Building Information Modeling and Simulation for the Mechanical, Electrical, and Plumbing Systems.” International Conference of Computer Science and Automation Engineering, Vol.3, 77-80

Xion, X., Adam, A., Akinci, B., and Huber, D. (2013), “Automatic Creation of Semantically Rich 3D Building Models from Laser Scanner Data.” Journal of Automation in Construction Vol.31, 325-337

Yin, X., Wonka, P., and Razdan, A. (2009), “Generating 3D Building Models from Architectural Drawings: A Suevery.” Journal of the Computer Society, 20-30

Yun, J. A Study on the Anticipated Changes of Architectural Design Process and Software under IPD (Integrated Project Delivery) Environment. M.S. Degree Thesis, 2017, Seoul National University

Zhang, D., Lu, G. (2004), “Review of shape representation and description techniques.” The journal of the pattern recognition society 37 (2004) 1-19

Appendices

Appendix A: Terminology

Term	Acronym	Explanation
As-built Model	-	As-built model refers to the creation of a digital 3D model based on the current state of construction or an existing structure. With a development of related technologies, photogrammetry or laser scanning is mainly adopted to generate as-built 3D models (Wang et al., 2015).
As-planned Model	-	It means the model created as planned in the construction project, and mainly refers to a 3D model created based on 2D drawings from the design stage. It is compared with as-built model to confirm the construction status comparing to the planning.
Bayesian Filter	-	It is the simplest and most widely used classification model based on the Baye's therorem (McCallum and Nigam, 1988). It is practically used with a feature that classifies text regardless of their dependency (Diab and Hindi, 2017; Rennie et al., 2003).
Baye's Theorem	-	It is a conditional probabilities that deals with decision-making problems mathematically under uncertainty. Unlike other probabilistic inferences that are based upon the deductive reasoning, this method is based on inductive and empirical reasoning.

Term	Acronym	Explanation
Building Components	-	Building components mean the basic elements of a building that includes building structure with foundation, floors, walls, beams, column, roof, stair with an aim to support and enclose it (Building Technology Guid, 2020).
Building Information Level	BIL	Definition of the detailed BIM level data published by the Public Procurement Service in Korea (Public Procurement Service, 2016). Different level of detail is defined according to the purpose of utilization, and it is divided into 10 units from BIL 10 at the planning level to BIL 60 at the facility management level.
Building Information Modeling	BIM	Building Information Modeling(BIM) represents the developed work process by adopting a computer generated model based upon the object in a building lifecycle (Azhar et al., 2008; Eastman et al., 2008). It also means Integration as well as utilization of information for building project.
Cost Estimation	-	Cost estimation is an approximation of project cost and accurate calculation helps avoid project overruns. Based on the quantity from the design phase, the total construction cost can be estimated considering direct cost of material, labor, equipment and indirect cost as well (Xu and Liu, 2013).

Term	Acronym	Explanation
Drawing Recognition	-	It refers to the process of recognizing line and text information in a 2D drawing to reconstruct 3D model. Although there is a difference in methods and procedures for classifying an recognizing information, the purpose to use line that includes geometric and topological information and to use text of annotation, dimension, and table (Lu et al., 2007; Tombre et al., 2002).
Hough Domain	HD	It means a virtual spatial area applied for the Hough Transform, and refers to a space that contains the converted point when transforming an expression describing a specific line into a single vertex.
Hough Transform	HT	Hough transform is a feature extraction technique used in image analysis, in specific, a line to point transformation in which, when applied to the centroids of connected components in an image, can be used to detect sets of connected components that lie along a given straight line (Fletcher and Katsuri, 1988).
Industry Foundation Classes	IFC	IFC is an open, international standard BIM data format developed by buildingSMART group. Interoperable feature of the data promotes vendor-neural and usable capabilities regardless of device, software platform, and interfaceses for manu different cases (buildingSMART, 2020).

Term	Acronym	Explanation
Level of Detail (Development)	LOD	It means the details of 3D model in computer graphic, and explains the complexity of model representation. In the case of BIM model defined by buildingSMART, the level 500 means the 3D model similar to the actual building and 100 is model in planned level.
Line/text Separation	-	It is originated on the graphic/text separation that is suggested and developed by Tombre (2013) and Fletcher (1998). It refers to a technique of classifying and storing two into separate layers based on the feature of line and text from one drawing.
Object Model	-	It is a basic element of BIM data generation, architectural elements are created with their shape and size attributes, with information and knowledge regarding the elements constituting the building are expressed as an object model (Park et al., 2000).
Pattern Recognition	-	It is a technique of shape representation and description generally employed in image search. In 2D drawings, region-based approach and contour-based approach can be applied according to the graphic's feature (Zhang, 2004).
Project BIM Data	-	It means data model including information related to one construction project. It is made of a set of object models, which are building components, and is utilized by purpose.

Term	Acronym	Explanation
Property (of object model)	-	Property of BIM is the most distinguished feature from the general 3D model. It comprehensively means information such as materials, architectural codes, zoning for construction, schedule, and structure required for a construction project, and any information can be added according to project characteristics (Kim et al., 2017).
Quantity Take Off	QTO	QTO is a detailed measurement of materials and manpower required to complete a construction project. It is conducted by the estimator at the pre-construction stage traditionally, and with the BIM data, it can be automatically calculated without a separate calculation process (Eastman et al.).
Relative Error	-	Relative error means the ratio of the error to the true value when an error occurs. It is used in absolute value to show the difference and expressed as a percentage(%). When multiple errors happen, it is used to explain accurate information of the approximation.
Shape (of object model)	-	Shape information is component of the object model with property information. It shows the model in three-dimensional environment with accurate shape, size, and material. In order to construct BIM project data, the method of joining between objects should be considered and displayed in the display environment.

Term	Acronym	Explanation
Text Classification	-	Text classification is a technique that automatically allocates natural texts according to the predefined categories. The basic method is to convert text to the method of answering to the user's problem and extracting data (Rahman et al., 2001). Algorithms for developing classification model varies by purpose, and the Naïve Bayes-based classifier is used often (Lewis et al., 1992).

Appendix B: Source of Code for Text Classifier

Python # for creating text classifier using Bayesian filter and training

```
import math, sys
from konlpy.tag import Twitter
class BayesianFilter:
    def __init__(self):
        self.words = set()
        self.word_dict = {}
        self.category_dict = {}
    def split(self, text):
        results = []
        twitter = Twitter()
        malist = twitter.pos(text, norm=True, stem=True)
        for word in malist:
            if not word[1] in ["Josa", "Eomi", "Punctuation"]:
                results.append(word[0])
        return results
    def inc_word(self, word, category):
        if not category in self.word_dict:
            self.word_dict[category] = {}
```

```
        if not word in self.word_dict[category]:
            self.word_dict[category][word] = 0
        self.word_dict[category][word] += 1
        self.words.add(word)
    def inc_category(self, category):
        if not category in self.category_dict:
            self.category_dict[category] = 0
        self.category_dict[category] += 1
    def fit(self, text, category):
        word_list = self.split(text)
        for word in word_list:
            self.inc_word(word, category)
        self.inc_category(category)
    def score(self, words, category):
        score = math.log(self.category_prob(category))
        for word in words:
            score += math.log(self.word_prob(word, category))
        return score
    def predict(self, text):
        best_category = None
        max_score = -sys.maxsize
        words = self.split(text)
```

```
        score_list = []
        for category in self.category_dict.keys():
            score = self.score(words, category)
            score_list.append((category, score))
            if score > max_score:
                max_score = score
                best_category = category
        return best_category, score_list
    def get_word_count(self, word, category):
        if word in self.word_dict[category]:
            return self.word_dict[category][word]
        else:
            return 0
    def category_prob(self, category):
        sum_categories = sum(self.category_dict.values())
        category_v = self.category_dict[category]
        return category_v / sum_categories
    def word_prob(self, word, category):
        n = self.get_word_count(word, category) + 1
        d = sum(self.word_dict[category].values()) + len(self.words)
        return n / d
```

Python # for calculating the precision rate of the created model

```
import numpy as np

from sklearn import cross_validation

from sklearn import datasets

from sklearn import svm

from sklearn.model_selection import StratifiedKFold

from sklearn.metrics import precision_score, recall_score


iris = datasets.load_iris()

skf = StratifiedKFold(n_splits=10)

clf = svm.SVC(kernel='linear', C=1)


X = iris.data

y = iris.target

precision_scores = []

recall_scores = []

for train_index, test_index in skf.split(X, y):

    X_train, X_test = X[train_index], X[test_index]

    y_train, y_test = y[train_index], y[test_index]


    y_pred = clf.fit(X_train, y_train).predict(X_test)

    precision_scores.append(precision_score(y_test,y_pred,average='macro'))
```

```
recall_scores.append(recall_score(y_test, y_pred, average='macro'))

print(precision_scores)
print("Precision: %0.2f (+/- %0.2f)" % (np.mean(precision_scores), np.std(precision_scores) * 2))
print(recall_scores)
print("Recall: %0.2f (+/- %0.2f)" % (np.mean(recall_scores), np.std(recall_scores) * 2))
```

Python # for test the created text classifier

```
from bayes import BayesianFilter
bf = BayesianFilter()

bf.fit("c1", "column")
bf.fit("c2", "column")
bf.fit("c4", "column")
bf.fit("c5", "column")
bf.fit("SC1", "column")
bf.fit("sC1", "column")
bf.fit("sc3", "column")
bf.fit("C1a", "column")
bf.fit("w1", "wall")
bf.fit("w1a", "wall")
bf.fit("sw1", "wall")
bf.fit("sw2", "wall")
bf.fit("sw1a", "wall")
```

```
bf.fit("벽", "wall")

bf.fit("C9", "column")
bf.fit("wall", "wall")

bf.fit("기둥 1", "column")


bf.fit("기둥 2", "column")

bf.fit("co.", "column")
bf.fit("wa1.", "wall")
bf.fit("sW1.", "wall")


pre, scorelist = bf.predict("C1b")
print("result=", pre)
pre, scorelist = bf.predict("W1b")
print("result=", pre)
print(scorelist)
```

Appendix C: Case Project Description

Overview of the project

■ 건축개요

구분	내용			
사 업 명	분당 NHN벤처타워 신축공사			
대지위치	경기도 성남시 분당구 정자동 178-1번지			
대지면적	6,600.00 m2 (1,996.50평)			
지역/지구	중심상업지역, 지구단위계획구역	도로현황	서측39.5m, 동측20m, 북측30m도로	
건축면적	3,955.32 m2 (1,191.84평)	건 폐 율	59.67%	법정:80%
지상연면적	61,094.28 m2 (18,481.02평)	용 적 률	925.67%	법정:1000%
지하연면적	37,667.76 m2 (11,394.50평)			
전체연면적	98,762.04 m2 (29,875.52평)			
공개공지	451.52 m2 (7%) > 법정:5% (330.00m2)			
조경면적	1,114.70 m2 (17%) > 법정:15%(990.00m2)			
층 수	지하6층, 지상28층			층고4.1m기준
주차대수	810 대 (지상69대, 장애인 16대 포함) > 법정:667 대			
최고높이	135.8m < 법정:135.6m (항공고도제한 150.3m, 대지기준해발고도44.7m기준)			
용 도	업무시설, 문화및집회시설, 근린생활시설			
구 조	철골철근콘크리트조			
외장재료	THK24 로이복층유리 + AL커튼월			
승 강 기	승용-12대(24인승), 비상용-2대(24인승)			
	셔틀용-2대(지하6층~지상2층)			
설 비	냉난방설비 -냉열원방식 : 빙축열 + 터보냉동기			
	-온열원방식 : 지역난방, 관류형보일러, VAV(각층 UNIT방식)			
	급수설비 -직수(비상시부스터펌프)+부스터급수펌프상향공급			

■ 주차대수산정

단위:m2 / 평

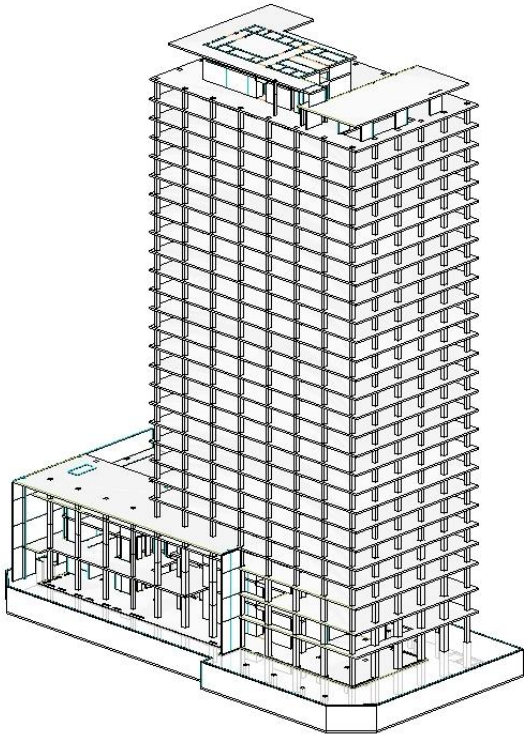
구 분	업무시설		문화및집회시설(공연장)		근린생활시설				총계	
면 적	62,650.90	18,951.90	1,867.06	564.79	2,957.91	894.77			67,475.87	20,411.46
산 출 근 거	1대/100㎡		1대/100㎡		1대/135㎡					
법정주차대수	627 대		19 대		22 대				667 대	
계획주차대수	755 대		25 대		30 대				810 대	

■ 용도별 면적개요

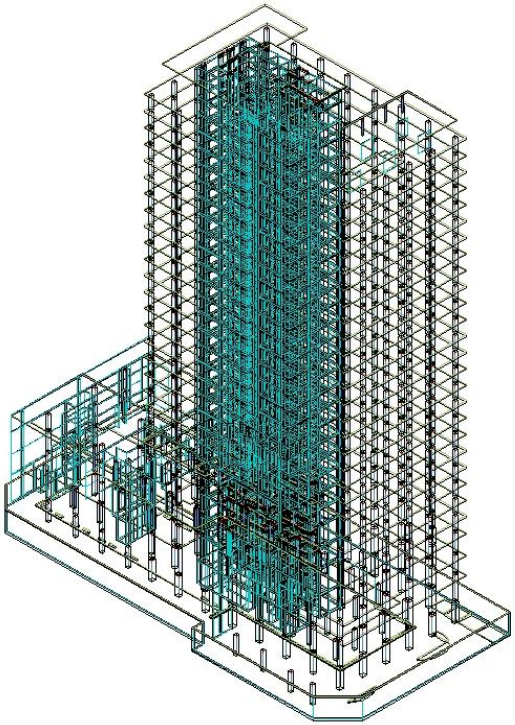
단위:m2 / 평

구분	업무시설		문화및집회시설		근린생활시설				총계	
시설면적	56,377.53	17,054.20	1,680.11	508.26	2,661.75	805.17			60,719.37	18,367.61
기계실면적	6,273.37	1,897.69	186.95	56.55	296.18	89.60			6,756.50	2,043.84
주차장면적	29,049.00	8,787.32	865.69	261.87	1,371.48	414.87			31,286.17	9,464.07
총계	91,699.90	27,739.22	2,732.75	826.66	4,329.39	1,309.64			98,762.04	29,875.52
비율	92.85%		2.77%		4.38%				100.00%	

Generated Project Models



(a) Project model with current modeling standards



(b) Project model with the proposed specification

國 文 抄 錄

도면 인식 방법을 활용한 건설 공사를 위한 BIM 정보 자동 작성 방안

정해진 비용으로 시공하는 건설 프로젝트는 공사비 예측과 조달, 실행을 위해 설계 정보를 바탕으로 하는 정확한 수량 산출과 비용 계산이 필요하다. 건설 프로젝트 정보를 삼차원 디지털 환경에서 작성하는 Building Information Modeling(이하 BIM)은 건물의 생애주기 전 단계에서 활용 가능한 프로젝트의 설계, 시공 관련 정보를 삼차원 형상과 속성 정보로 저장한다. 데이터 작성 이후 속성 정보를 활용한 수치적 계산이 가능하다는 점에서 건설관리에서의 BIM 적용은 많은 관심을 받아왔으며, 2018년 기준 한국, 영국, 싱가포르 등 여러 국가에서 그 적용을 의무화하였다.

그러나 데이터 작성과 활용의 장점에도 불구하고, 작업자들은 여전히 기존 작업 방식인 이차원 도면 작성과 이를 활용한 물량산출 등을 시공에 적용하고 있어 동일한 정보를 도면과 삼차원 모델로 중복 작성하는 이중 작업을 하고 있다. 공사 기간의 단축과 비용 절감이 매우 중요한 건설산업 분야에서 불필요한 작업을 최소화하기 위하여 도면을 기반으로하여 자동으로 작성되는 BIM 데이터가 요구된다.

본 연구에서는 그 방법으로 디지털화된 도면 정보를 객체 모델로 변형하는 프레임워크를 제안하였다. 그것은 도면에 포함된 선과 문자 정보를 객체 모델의 형상과 속성 정보로 자동 변환하기 위한 일

종의 패턴 인식 기술과 문자 분류 기술이다. 또한, 작성된 객체 모델의 물량 정보를 시공 단계에서 활용하기 위하여 견적계산의 기준이 되는 빌딩 구성 요소 간의 관계와 모델 작성 기준을 수립하였다.

연구에서 제안된 프레임워크의 성능-문자 분류기와 객체 모델 생성, 그리고 물량산출의 정확성-을 확인하기 위하여 베이지안 필터 기반의 지도기계학습 모델인 문자 분류기를 생성하였고, 형상 인식 기반 삼차원 모델 생성과정을 통하여 '도면 인식 방법'을 기반으로 한 객체 모델 작성 프레임워크의 적용 가능성을 확인하였다.

마지막으로, 제시된 물량 산출을 기준으로 실제 건설 프로젝트를 대상으로 작성된 모델에서 도출된 전체 수량과 견적 기반 수작업으로 계산된 결과물, 그리고 일반 BIM 모델도구에서 산출된 수치 정보를 비교하여 타당성을 검증하였다. 또한 결과 분석을 통한 수치적 오차를 바탕으로 건설 프로젝트에 활용하는 경우 고려하여야 하는 할증을 분석하였다.

향후 사용자 개입이 최소화되는 기술적 발전을 이루어 건설 프로젝트에서의 시간 절약과 비용의 절감, 나아가 생산성 향상을 이룰 수 있을 것으로 기대한다.

주요어: 건설관리; 빌딩정보모델링(빔); 선-문자 추출; 문자분류; 패턴인식; 물량산출

학번: 2017-37092 **이름:** 노 주 희