

A Study on Information Extraction Method of Engineering Drawing Tables

Riza Sulaiman
Faculty of Information Science
and Technology,
Universiti Kebangsaan
Malaysia, Selangor, Malaysia

Mohd Fahmi Mohamad
Amran
Faculty of Information Science
and Technology,
Universiti Kebangsaan
Malaysia, Selangor, Malaysia

Nur Amly Abd Majid
Faculty of Computer Science
and Information Technology,
Universiti Selangor, Malaysia

ABSTRACT

In getting an existing engineering drawing to be reuse, the first step to be done is to search for the appropriate file. To get the desired drawing, an engineer will have to access every file. This procedure consumed much time and energy as there might be lots of files and folders related to the drawings. To ease engineers search for old engineering drawings, the extraction of the table structure has to be done. This paper describes review on recent research in extraction of engineering drawing tables. Along the way of this reviewing process, the weakness and downside of previous researches has been identified.

General Terms

Information Extraction

Keywords

Computer Aided Design, Engineering Drawing Tables, Reuse Design

1. INTRODUCTION

When a designer designs a new engineering drawing, they will refer to the previous valid design documents that are the same or nearly match with their new design specification. This is due to the fact that the previous designs have been reviewed, analyzed and proven to be successful. By referring to the previous designs, it likely saves significant resources and manpower [1]. In getting an existing engineering drawing to be reuse, the first step to be done is to search for the appropriate file. To get the desired drawing, an engineer will have to access every file. This procedure consumed much time and energy as there might be lots of files and folders related to the drawings. The best way to search for an old drawing is by applying an electronic-based search in the operation system compared to the manual search which required an engineer to read the content of the drawings. The limitation in the search process based on the operation system, the information states only the names of the files. In engineering practices, the name of the files does not necessary match the drawings. It might not state information on the drawings such as the name of the drawings, the designer, the date, the version number and other information. By reading only the files name, it could not visualize the real picture to engineers about the content regarding the engineering drawings. It becomes more complicated if the files name is written by using the project code and engineers do not aware of the certain project code representing to related project. Engineers also face difficulties to obtain the latest version of the drawing especially which are related and dependent to

other several drawings [2]. Therefore, to gain information in the table, information extraction need to be done to access to the required information. Besides, by extracting information consists in the table itself, the information gained is accurate and free from problems.

To ease engineers search for old engineering drawings, the extraction of the table structure has to be done first, for text is not only written in the tables, but also in the whole of engineering drawings including the text which reflected the dimension, text notation and other component. On top of that, extraction also has to be done in differentiating the tables. The search information process is facing with difficulty if there are mistakes in the process of the extraction tables. In engineering drawings, there is a possibility of the existence of more than one table. Besides the title block, the table in engineering drawings is a table representing information such as the Bill of Materials (BOM), revision block, product structure and so on. Every schedule prepared different information and convey different meanings. For example, certain table explains about data regarding the engineering object whilst the other table explains data regarding the evaluation of the object [3].

The purpose of this paper is to review recent research in extraction of engineering drawing tables. This paper has been divided into two parts. The first part deals with review on related works in extraction methods of engineering drawings tables. For the second part, the weakness and the downside of the extraction methods will be analyzed and discussed.

2. OVERVIEW OF INFORMATION EXTRACTION IN ENGINEERING DRAWING

According to [4] the process of extraction is the process of extracting of CAD data in order to obtain details of products and processes that contribute to the production of the product. Basically, CAD data contains data on geometry and non geometry. Geometry data is a collection of objects described by CAD drawing of lines, arcs and circles. While non-geometry data are related to object attributes such as length, thickness, radius, diameter and label [5]. In this research, tables that will be studied is consider as non-geometry information. Geometry and non geometry data is closely related to information products [5]. [6] classified the automatic recognition of engineering drawings problems into three levels at which the first is the identification of low-level graphics primitives, the second stage of identification of high-level graphical objects, and the third level of high-level intelligence interpretation and analysis. There are several

types of tables in engineering drawing such as title block, bill of quantity/materials, revision number and so on.

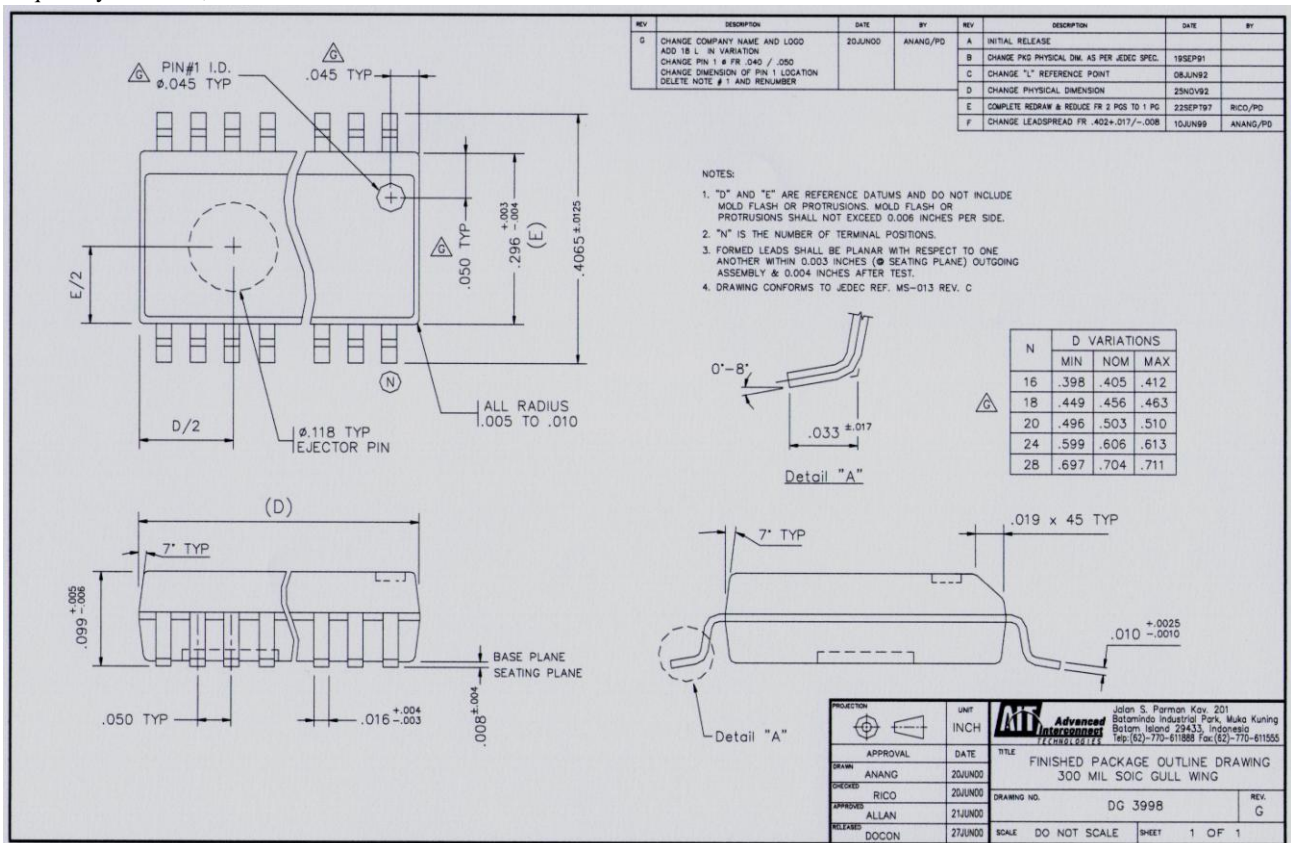


Fig 1: Example of engineering drawing

3. CHALLENGES ON TABLE EXTRACTION METHOD

As mentioned in [3] and [7], new methods are needed because of the great differences between engineering tables in engineering drawings and other documents. The first challenge to be resolved is to extract the information needed in engineering drawings that contain geometry and non-geometry information. This introduces the challenge of information extraction. These challenges are due to the various information such as the text, graphical primitive symbols and dimension located within tables in engineering drawings. Secondly, the problem is the recognition physical structure of tables. There are more than one table exist in engineering drawings, hence it creates the need to distinguish between a tables with another tables. For example, one engineering table describes section data of some engineering objects, while another gives their evaluation data [3]. Another challenge is the structure of the table. Many research focusing on detecting the intersections with regards of searching a set of perpendicular and parallel lines [7] and find all the enclosed rectangular regions [8], [9]. As one of the information type exist in tables is text, searching by text itself can be done to find the related information in engineering drawing tables. However, by searching information available in the tables using text based, it expose to time consuming because text field on a drawing composed of an abundance of line and text object [10]. Figure 1 shows example of engineering drawing.

4. RELATED WORKS

There are few researches particularly focusing on the extraction of information from engineering drawing tables. For example, [9] suggested an application in indexing engineering drawings whereby the method needs the design of the model. This research is using the rectangular detection algorithm which detects the existence of any rectangular which contains several dots. However, the algorithm used has several weaknesses, it only focuses on the extraction of the title block and not on other tables existed in the drawings. It recognizes the title block based on the starting registration point, which referring to the bottom-right corner of the drawing frame. Based on this registration-point, the algorithm then detect the rest set of rectangular.

A research by [11] presented an application that helps an extraction of information from drawings obtained from a telecommunication company. The application is focusing on the location of the title block. It is used by handler to help in correcting the title of a drawing. The first step taken by this application is by extracting the straight lines of the petite regions in the drawings by using the FAST method. The box with the biggest region that fulfilled certain criteria of the width-to-height ratio will be chosen as the title field. Since this research suggested a specific conclusion regarding problems they had to overcome such as telephones, the approaches could not be implemented to other problems.

Najman et al (2001) [12] suggested a method to identify automatically a location of a title block in a scanned technical drawing. It is based on a signal measurement which is being

calculated on the title block. There is a case rejected, especially because of the difficulty with the frames and the region of the title block, which part of it is being erased or perpendicular.

A research done [7] suggested an automatic system analysis in extracting structured knowledge from an engineering drawing table. The technique could analyse the layout of the drawings and extract the physical and logical structure. Based on the automatic analysis, this research suggested normalization method in the integration between the varieties of engineering tables compared to the engineering drawings in extracting the knowledge domain within. Besides that, this research successfully did search on straight lines from raster images with the use of Global Line Vectorization which suggested by previous researchers. With the application of this algorithm, the text could be recognized after identify the straight lines and symbols omitted from the drawing images. Based on the mean distance extracted from the straight lines nearby and parallel, the size could be determined in searching of the connected character box. The extraction of the straight lines nearby could be the guidance in search processes because of most of the text in an engineering drawing blocks are serial or parallel with the table lines. The weakness in this research is that it could not detect parallel and perpendicular lines which might contain errors.

Y. Cao et al. (2005) [8] had produced a methodology to control a version automatically which concentrate on the automatic version identification. The methodology applied is by analyzing the layout of a square region from a drawing frame and the information related to the version then is extract with the help of keywords defined in the database. Three major steps in analyzing a layout of a rectangular are by, firstly is to identify the rectangular positioned at the below left or right of the drawings. The rectangular is made from a group of rectangular and followed according to the keywords defined in the database found at the square. However, the weakness of the methodology is that it will analyze one-by-one, the region of the rectangular of the whole drawing and make sure of the keywords had been defined or not. This

method is time consuming as it has to scan the whole drawing where there are a lot of rectangular needs to be analyzed. The proposed technique in this research extracts version information in a drawing frame using a model-based method. When a new drawing is introduced to the system, it selects a predefined pattern most similar to the drawing frame used in the new drawing and extracts the information based on the selected pattern. However, the technique is not suitable for engineering drawings with different layout plan because predefined patterns should be prepared and implanted into the system before version extraction. It takes a lengthy time for the system to find the appropriate pattern. Moreover, if the frame pattern of new drawing does not exist in the predefined pattern located in the database, the system fail to do the extraction process which does not suit the variability of title block structures.

Meanwhile Mahmood (1999) [13] addressed the problem of extracting indexing keywords from engineering drawing images by localizing image structures called title block. In order to locate title block, they used technique of 2D pattern localization in unsegmented images called location hashing. The weakness of this technique is it unable to locate title block which has different layout plan as it requires a learning step to establish a model for each structure to locate.

Ondrejcek et al. (2009) [14] had suggested an automatic found system from an unknown relationship between engineering drawing files and 3D CAD models. In this research, engineering drawings is in the form of images where the scanning process is done where the resolution applied is less than 300dpi. The title block is identified manually on the engineering drawings and contains a cropped title block. The cropped title block is identified according to the type and is compared to the existing template. However, the system has limitation due to the usage of the template in determining the type of a title block. If it is found out that the title block is not suitable to the template, the title block could not be identified and the information extracting process failed to execute. Table 1 compares the information extraction methods of engineering drawing tables from previous researchers.

ITEM NO.	PART NUMBER	DESCRIPTION	C0/QTY.	C1/QTY.	C2/QTY.
1	8112156	Handle	1	1	-
2	8112174	Handle-shoulder	1	1	-
3	8113199	End cap	1	1	-
4	9113155	Embedded bolt	1	1	1
5	8112992	Housing	1	1	1
6	8116170	Gearshaft screw	1	-	1
7	112-135	Crescent washer	1	1	1
8	112-139	Shaft screw	1	1	1
9	113-144	Shaft spring	1	1	1
10	8112188	Shaft	1	1	1

Fig 2: Example of bill of materials

REV	ZONE	REVISION	DESCRIPTION	DATE	APPR
A	ALL SHTS	INITIAL RELEASE		12/05/05	
B	PG2 C-2	ADDED ADDITIONAL MOUNTING POINTS		12/05/05	
C	PG2 A-1	ADDED ACCESS PANEL IN BULKHEAD		01/02/06	

Fig 3: Example of revision block

Table 1. Comparative information extraction methods from engineering drawing tables.

Author	Method/Approach	Problem(s)
L. Najman et al (2001)	Detects the existence of any rectangular which contains several dots. Recognize based on the starting registration point by referring to bottom-right corner point.	Focus on title block extraction only.
J.F. Arias et. al (1997)	Extracting the straight lines of the petite regions in the drawings by using the FAST method. The box with the biggest region that fulfilled certain criteria of the width-to-height ratio will be chosen as the title field.	Case-based problem. Not suitable to implement on other type of engineering drawing. Focus on title block extraction only.
L. Najman et al (2001)	Based on a signal measurement which is being calculated on the title block.	There is a case rejected, especially because of the difficulty with the frames and the region of the title block, which part of it is being erased or perpendicular.
T. Lu et. al (1997)	Extracting structured knowledge from an engineering drawing table by implement Global Line Vectorization algorithm.	Could not detect parallel and perpendicular lines which might contain errors.
Y. Cao et al. (2005)	Analyzing the layout of a square region from a drawing frame and the information related to the version then is extract with the help of keywords defined in the database.	Time consuming Need to scan the whole drawing where there are a lot of rectangular needs to be analyzed. Not suitable for engineering drawings with different layout plan because predefined patterns should be prepared and implanted into the system.
T. S. Mahmood (1999)	Implement technique of 2D pattern localization in unsegmented images called location hashing.	Unable to locate title block which has different layout plan as it requires a learning step to establish a model for each structure to locate.
M. Ondrejcek et al. (2009)	The title block is identified manually on the engineering drawings and contains a cropped title block. The cropped title block is identified according to the type and is compared to the existing template.	Using template in determining the type of a title block. If it is found out that the title block is not suitable to the template, the title block could not be identified and the information extracting process failed to execute. Focus on title block extraction only.

5. RESULT & DISCUSSION

As mentioned before, there is a possibility of the existence of more than one table in engineering drawing. A typical bill of material is shown in Figure 2 and revision block in Figure 3. Looking at the majority of the existing information extraction technique and algorithm for table in engineering drawing, we can observe three things. First, most published works reported for engineering drawings with different layout plan of drawing frames, predefined patterns should be prepared and implanted into the system before information extraction take place. Therefore, the method unable to extract information if the layout is not same with predefined layout in the database.

Second, most current information extraction methods assume that the table in engineering drawing is perfectly structured and the lines are well connected between them. Third, there is no available method that can solve problem if there have an error found in the table where the table lines are not interconnected with each other.

In determining the existence of tables, algorithms that detect rectangular assume that horizontal and vertical lines are perpendicular [9]. Perpendicular and parallel lines that make

intersections are connected and the degree is always 90 degrees. However, most of the commonly used algorithms in table recognition in engineering drawing are unable to make a correction if the intersections which supposedly connected are not connected. This raises issues and brings challenges to the information extraction of table in engineering drawing.

6. SUMMARY

This paper describes review on recent research in extraction of engineering drawing tables. Along the way of this reviewing process, the weakness and downside of previous researches has been identified. This paper began with brief scenario in information extraction on engineering drawing and ends with analysis and discussion.

7. ACKNOWLEDGMENTS

The authors honourably appreciate Universiti Kebangsaan Malaysia for the research university grant HEJIM-FTSM-FKAB-MTDC-101101005. Universiti Selangor (UNISEL) Scholarship Programme also acknowledged with gratitude.

8. REFERENCES

- [1] T. Lu, C.L. Tai, F. Su, S. J. Cai. A New Recognition Model for Electronic Architectural Drawings. *Computer-Aided Design* 37 (10), pp. 1053-1069, (2005).
- [2] Amran, Mohd. Fahmi Mohamad; Sulaiman, Riza; Kahar, Saliyah; Marjudi, Suziyanti; Abdullah, Khairul Annuar; Adnan, Zuraidy; , "A comparative study on extraction method of non-geometry information in engineering drawing," *Business, Engineering and Industrial Applications (ISBEIA)*, 2011 IEEE Symposium on , vol., no., pp.5-9, 25-28 Sept. 2011.
- [3] T. Lu, R.Yang, F. Su, S. J. Cai. Understanding of complex engineering tables for automatic 3D modeling. *International Workshop on Graphics Recognition (GREC'07)*, pp. 1-3. (2007).
- [4] Y. Cao, S. H. Wang, H. Li. Automatic Recognition of Tables in Construction Tender Documents. *Automation in Construction* 11, 573– 584 (2002).
- [5] M. F. A. Jabal et al. A Comparative Study on Extraction and Recognition Method of CAD Data from CAD Drawings. *International Conference on Information Management and Engineering*, pp. 709-913 (2009).
- [6] Z. Lu. Detection of Text Regions from Digital Engineering Drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 4, 431-439 (1998).
- [7] T. Lu et al. Knowledge Extraction from Structured Engineering Drawings. *Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 415-419 (2008).
- [8] Y. Cao et al. Using Engineering Drawing Interpretation for Automatic Detection of Version Information in CADD Engineering Drawing. *Automation in Construction* 14, pp. 361-367 (2005).
- [9] L. Najman, O. Gibot, S. Berche. Indexing Technical Drawings using Title Block Structure Recognition. In *Proceedings of 6th International Conference on Document Analysis and Recognition*, Seattle (USA), pp. 587–591 (2001).
- [10] N. Weber, A. Henrich. Retrieval of technical drawings in DXF format - concepts and problems. *LWA 2007*, pp. 213-220 (2007).
- [11] J.F. Arias, A.K. Chhabra, V. Misra, A Practical Application of Graphics Recognition: Helping with the Extraction of Information from Telephone Company Drawings. In *Proc. GREC*, pp.314-321 (1997).
- [12] L. Najman, O. Gibot, M. Barbey. Automatic Title Block Location in Technical Drawings. In *Proceedings of 4th IAPR International Workshop on Graphics Recognition*, Kingston, Ontario (Canada), pp. 19-26 (2001).
- [13] T. S. Mahmood. Extracting Indexing Keywords from Image Structures in Engineering Drawings. *Fifth International Conference on Document Analysis and Recognition (ICDAR'99)*, pp.471-474 (1999).
- [14] M. Ondrejcek et al. Information Extraction from Scanned Engineering Drawings. *Image Spatial Data Analysis Group, National Center for Supercomputing Applications*. Urbana, IL, Tech. Rep. NCSA-ISDA09-001 (2009).