

STUDY ON AUTOMATIC REGISTRATION METHOD OF SOURCE DATA FOR CIM BUILDING MODEL CONSTRUCTION

Shuqing Ran, Guoyin Cai,

School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture,
2108570021087@stu.bucea.edu.cn

KEY WORDS: CIM, building model, hierarchical and household, spatial registration, automation.

ABSTRACT:

The City Information Model (CIM) is a digital model designed to describe, analyze, and manage urban spatial information. It integrates various urban elements, providing decision support for urban planning, design, and management. This study focuses on the construction of building models in CIM, particularly the issue of automatic registration of source data. We propose an automatic registration method based on DWG-format architectural design drawings and SHAPE-format remote sensing building outline information. This method extracts the axis line information of the building from the plan design drawings, indirectly obtains feature points, and selects registration base points, thereby achieving the registration of real estate subdivision data. Meanwhile, by automatically fitting the outer bounding rectangle of the building and dynamically adjusting the angle, this study realizes the automatic registration of the aligned multilayer real estate subdivision data with the building outline data obtained from remote sensing images. This research provides an effective method for the automatic and batch accurate alignment of data for the construction of CIM building models, thereby enhancing the precision and reliability of the modeling.

1. INTRODUCTION

With the continuous increase in China's urbanization rate, rapid urban development and expansion have brought about a series of issues such as environmental pollution, resource shortage, and traffic congestion. The fine, scientific, and efficient governance of cities has become a hot topic in recent years. In the face of the continuous development of high-tech technologies such as the internet, big data, and artificial intelligence, people are beginning to explore the use of intelligent means to carry out urban planning, construction, and management, aiming to achieve a rational allocation of massive urban information resources, comprehensive control of the city's operating status, and integrated, refined management of the city. Against this background, a digital description and expression technology theory for cities, the City Information Modeling (CIM), came into being. There are various understandings of the definition and connotation of the CIM basic platform in the industry, but considering the consensus on the CIM basic platform, this paper believes that the CIM basic platform is an organic integration of three-dimensional urban spatial models and dynamic city information. It unifies the building information model (BIM) at the micro scale, geo-spatial data at the macro scale, and internet of things (IoT) data, forming a basic operation platform that supports urban planning, construction, management, and operation (Han, 2022).

In the construction of CIM, it involves the construction of a wide range of urban architectural models, and the construction of building models is one of its core components. Therefore, how to automatically and batch extract a wide range of building information from different sources and perform spatial alignment has become an important research topic.

In the process of urban three-dimensional modeling, there are currently four main methods for source data registration. The first is registration based on control points, which is

characterized by manually selecting common points with significant features as control points. This method is the mainstream registration method, but it is too dependent on labor and the efficiency is low. The second is feature-based registration, which uses feature extraction algorithms to automatically detect feature points, and then uses feature matching algorithms to match these feature points (Wang, 2021). This method is highly automated and mainly applicable to cases where the features between images are relatively significant. Since the feature points are extracted based on the algorithm, the accuracy of each feature point cannot be guaranteed for batch extraction. The third is model-based registration, which usually requires high-quality model data as support (Tao et al., 2022). This method has high accuracy, but the data volume is extremely large, inconvenient to operate, and particularly unsuitable for batch construction of architectural models. The fourth is registration based on deep learning. By training a deep neural network, the spatial relationship and feature mapping between the source data are automatically learned for registration (Li et al., 2023). Although this method is highly automated, it has problems such as large physical input and long development cycle.

In accordance with the industry requirements of CIM3 and above, the data used for building model construction should contain spatial information and attributes of layered division. The source data used in this study are DWG format architectural design drawings and SHAPE format building outline information automatically identified by remote sensing images. The former mainly provides spatial and attribute information of building division, while the latter is used to ensure that the building model is in the correct geographical location. The subsequent problem is that since the DWG format architectural design drawings do not have coordinate attributes, it is impossible to ensure the correct spatial relationship between layers. In addition, due to the large feature differences between real estate subdivision data and SHAPE format building outline data, the feature points at the corners are

unreliable, and the registration effect of the two is poor. Therefore, the goal of this study is to achieve feasible registration between the above two types of data. In response to the above two issues, this study adopts a new automatic feature point extraction method for multi-layer real estate subdivision data registration, and achieves rough registration between data from the same building but different sources. This method indirectly obtains feature points by extracting the axis line of the building from the plan design drawing and selects the registration base point, thereby achieving the registration of real estate subdivision data. And by automatically generating a bounding rectangle, after dynamic angle correction and matching, the registration of the corrected multi-layer real estate subdivision data with the building outline data obtained from remote sensing images of the same building is realized.

2. RESEARCH METHOD

2.1. Technical Roadmap

The source data used in this study are DWG-format architectural design drawings and SHAPE-format building outline information obtained through automated identification of remote sensing images. The DWG-format architectural design drawings primarily provide the spatial and attribute information of the building divisions on various floors, offering advantages such as low acquisition difficulty and suitability for large-scale modeling. The SHAPE-format building outline information, obtained through automated identification of remote sensing images, is used to ensure that the building model is located at the correct geographical position to satisfy the requirements of CIM applications. The technical route adopted in this study is as follows (Figure 1).

The objective of the source data preprocessing step is to extract the necessary real estate subdivision map from the architectural design drawings (DWG format files) using AutoCAD software for manual processing. After processing, the required building centerline data and axis line data for the study are extracted (Figure 2).

The attribute data batch processing step can convert the line data of DWG into spatial polygonal data and attach household numbers and room names for subsequent matching. Since the layered subdivision DWG file cannot determine the spatial correspondence of each layer, it is necessary to extract building line data and its annotation information from the plan design drawings and use standard lines to strictly align the spaces of each layer, which is the spatial data batch processing step.

The spatial and attribute information matching step involves matching the spatial data of the DWG files obtained in the previous steps with the attribute data of the DWG files. The spatial data of the real estate subdivisions, which have architectural line data and axis line data from the plan design drawings, are matched with the spatial polygon data of the real estate subdivisions with attribute information. This results in the acquisition of real estate subdivision data with correct room numbers and names, and coordinates of the alignment points and verification points.

The spatial data batch alignment processing step refers to the registration and alignment of spatial data from different floors of the same building, requiring their correct spatial relationships. The method used in this step contains one of the key technologies of this study, i.e., the automated alignment method for DWG spatial data based on the axis line.

The final step is the automatic matching of spatial data and vector information, which involves matching the aligned SHAPE format real estate subdivision data with the SHAPE format building outline data. This enables the transformation of the independent coordinates of the DWG real estate subdivision data to the CGCS 2000 coordinate system. The method used in this step contains another key technology of this study, i.e., the real estate data matching method based on dynamic angle correction.

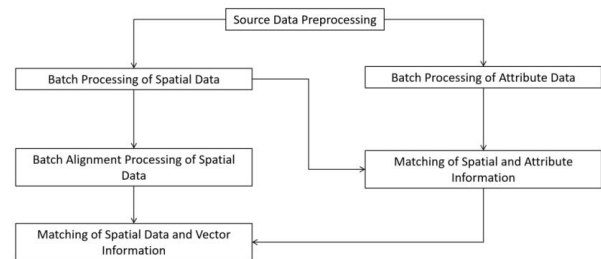


Figure 1. Overall technical route.

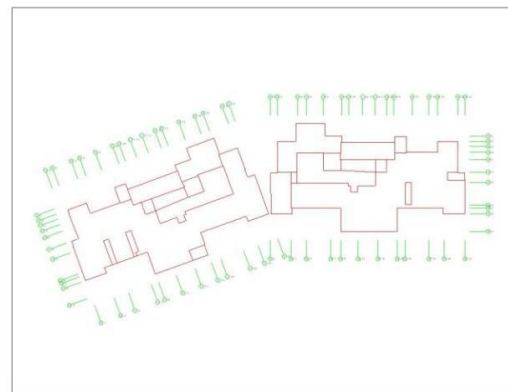


Figure 2. Building centerline data and axis line data.

2.2. Key Techniques

2.2.1. Axis-based method for automatic alignment of spatial data

The previous section mentioned four feasible data registration methods in the process of urban three-dimensional modeling, of which the two methods mainly suitable for CIM batch and automatic application scenarios are feature-based registration and deep learning-based registration. Although deep learning-based registration has a high degree of automation, it also involves significant resource investment and a long development cycle. Facing the demand for large-scale model construction in CIM, this method is too costly and has therefore been discarded.

Feature-based registration methods are mainly suitable for cases where features between images are more pronounced. Since feature points are extracted based on algorithms, the accuracy of each feature point extraction cannot be guaranteed if batch extraction is to be carried out for precision. The DWG format architectural design drawings do not have coordinate attributes, so alignment points cannot be directly selected from the architectural design drawings. In response to this problem,

this study proposes a registration method based on architectural axis line features, i.e., using the intersection points generated by extending the axis line (Figure 3) as feature points for the real estate subdivision map. Since the building line data and the axis line data are highly correlated, and the accuracy of the axis line itself is extremely high, the intersection points extracted by this method as feature points have similarly high precision in feature point representation. Intersection points are automatically selected, and based on coordinate extremes, four intersection points are chosen as alignment points and verification points. After converting the independent coordinates of the data of the building's bottom layer to the CGCS 2000 coordinate system, the coordinates of other building layer data are corrected for coordinate translation based on the coordinate difference of the alignment points, unifying all real estate subdivision data to the CGCS 2000 coordinate system, and realizing strict alignment of data on all layers.

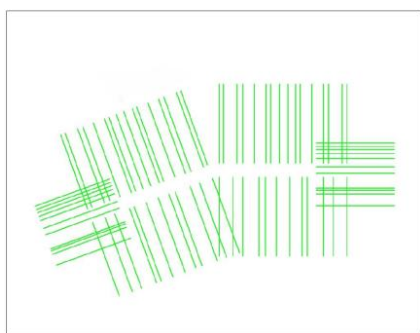


Figure 3.Extend the axis line and generate intersection points.

As illustrated in the schematic representation of the automatic alignment process for real estate parcel data (Figure 4), the data requiring alignment consists of three distinct types: alignment points, verification points with coordinate data, coordinate data of centerlines with associated coordinates, and multi-layered real estate parcel data.

The coordinates for alignment points and verification points are derived from the intersections generated by extending the centerlines. Alignment points are selected based on the criterion of having the minimum value of $x+y$, while verification points are chosen using criteria involving the maximum values of $x+y$, minimum values of $y-x$, and maximum values of $y-x$.

The coordinate data for centerlines with associated coordinates is directly extracted from architectural floor plans. This dataset is subjected to simplification, with centerlines featuring structural layouts being retained.

The multi-layered real estate parcel data originates from real estate parcel maps. It encompasses parcel attributes and geometric information, which must be matched with the centerline data featuring coordinates to simplify and retain centerline data with structural layouts. The associated attribute information is concurrently preserved within these centerline datasets.

The term "surface fusion" within the workflow refers to the comprehensive integration of single-layer real estate parcel data's geometric information. This step facilitates subsequent coordinate offsetting.

The calculation of offset values is contingent upon the automatic determination of the required offsets based on the alignment points of each layer. These offsets are then applied horizontally to achieve precise alignment between the base layer and other layers.

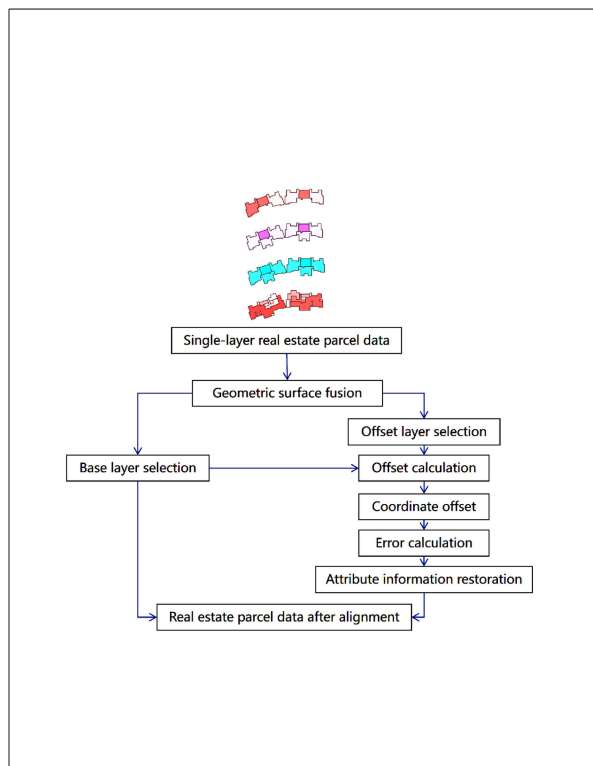


Figure 4.Automatic alignment process for real estate parcel data.

2.2.2. Dynamic angle correction method for real estate matching

Currently, the demand for the subdivision of building models is primarily focused in the field of real estate information management. Existing two-dimensional real estate property management data is simple, contains limited information, lacks correlation, and has limited spatial representation and analytical capabilities (Li, 2019). It cannot provide an intuitive understanding of the three-dimensional perception of a building and struggles to meet the demand for fine, accurate, and intuitive management of three-dimensional real estate information (Hou et al., 2017). This has prompted the transition of real estate registration from two dimensions to three. In 2022, Xiong and others proposed a real estate data coordinate correction method based on specific edge matching. This method uses large-scale DLG data extracted from the original data as the outline of the house and the real estate subdivision data. By selecting specific edges to correct the coordinates, and fitting the real estate subdivision data into the house outline with real coordinate values (Xiong et al., 2022).

CIM involves the construction of large-scale urban building models, and the source data must be plentiful and not too difficult to obtain overall. Therefore, the source data used in this study are SHAPE format building outline data extracted from remote sensing images. Since the shape representation accuracy of this source data is low, there is a significant shape difference when fitting with the real estate subdivision data,

and it is impossible to correct the coordinates of the two in correlation. Therefore, only a rough registration can be performed.

To address this issue, this study proposes a dynamic angle correction method. This method automatically fits the building outline data and real estate subdivision data to generate the external rectangle, calculates the angle difference between the two rectangles, and performs angle correction. The corrected external rectangle is then fitted, indirectly achieving the registration effect of the building outline data and real estate subdivision data. This method is different from traditional coordinate correction registration methods and is an indirect registration method. As the SHAPE format building outline information obtained by automated identification through remote sensing images is only used to ensure the correct geographical location of the building model, this method is mainly implemented for automation and batching.

3. RESEARCH FINDINGS

The main achievements of this study are two-fold. The first is the alignment results of multi-layer real estate subdivision data (Figure 5), the presentation method of which is visual interpretation. The accuracy validation of these results will be discussed later. The other achievement is the matching results of real estate data with building outline data. When this result has not undergone angle correction (Figure 6), it is clearly deviated compared to the results after angle correction (Figure 7).

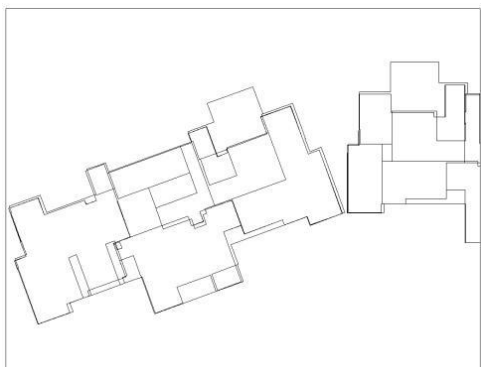


Figure 5. Results of the alignment of multilayer real estate subdivision data.

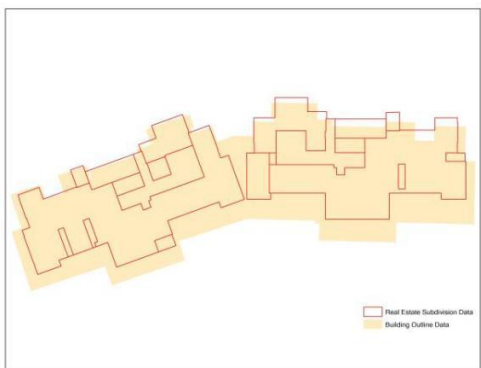


Figure 6. Results of the matching process without angle correction.

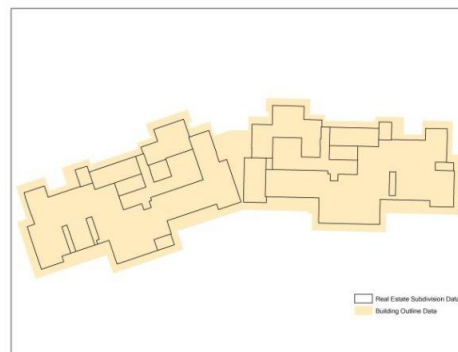


Figure 7. Results of the matching process using angle correction

4. DISCUSSION

To verify the precision of alignment for multi-layer real estate subdivision data, the method adopted was to substitute the corrected alignment points and three additional verification points respectively into the two-dimensional Euclidean distance formula (1), and take the average as the alignment error of the alignment point for the single-layer real estate subdivision data. In this way, the alignment errors of four alignment points can be obtained. The final alignment error for single-layer real estate subdivision data is calculated by taking the average of these four alignment point errors. According to experiments with multiple source data samples, the alignment method used in this study can keep the actual offset (Figure 8) between different layers of real estate subdivision data below ± 3 centimeters, fully suitable for bulk construction of CIM models. However, this study still has its shortcomings. The main defect is that because the automated alignment process is based on FME software development, the process cannot limit the x and y values at the same time when automatically selecting alignment and verification points, and needs to choose a priority for the x or y values. Therefore, there may be multiple extremal points when limiting a single coordinate value. The current method of adjustment in this study is to choose the point with the smallest sum of x and y as the alignment point, and the points with the largest sum of x and y, the smallest difference of y-x, and the largest difference of y-x as verification points (Figure 9). After adjusting the method, theoretically, the accuracy and uniqueness of the selected intersecting points of the axial line have improved, but more application scenarios are still needed for further verification in the future.

At present, the primary limitation of this study lies in the partial inconsistency of the axis lines among floor plans for the same building. This issue may potentially lead to discrepancies in the extracted alignment points and validation points for various floors. In subsequent research, we intend to address this by employing a method that selects identical axis lines across different layers to ensure uniformity in the axis lines of each layer.

To begin with, we will expose the annotation information for axis line data (Figure 10) within the attributes. By comparing the encoding of this annotation information, we can determine the maximum number of annotations shared across floor plans for a particular building. Subsequently, we will preserve the axis line data associated with these matching annotations. Since these axis line data are present in the floor plans of all layers,

this approach ensures consistency in the intersections of axis lines on each floor.

$$\text{Distance} = \pm \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

where X_n, Y_n = image coordinates

111202377-2	2.7268762813...
111202377-2	2.7268762813...
111202377-2	2.7268762813...
111202377-2	2.7268762813...
111202377-2	2.7268762813...
111202377-2	2.7268762813...
111202377-2	2.7268762813...
111202377-2	2.7268762813...
111202377-29	1.7472491608...
111202377-29	1.7472491608...
111202377-29	1.7472491608...
111202377-29	1.7472491608...
111202377-29	1.7472491608...
111202377-29	1.7472491608...
111202377-29	1.7472491608...
111202377-29	1.7472491608...
111202377-29	1.7472491608...
111202377-3	2.3969837434...
111202377-3	2.3969837434...
111202377-3	2.3969837434...
111202377-3	2.3969837434...
111202377-3	2.3969837434...
111202377-3	2.3969837434...
111202377-3	2.3969837434...
111202377-3	2.3969837434...
111202377-3	2.3969837434...
111202377-3	2.3969837434...

Figure 8. Actual offsets of real estate subdivision data on each floor.

Test Condition	
If	@Value(x+y) = @Value(...
Else If	@Value(x+y) = @Value(...
Else If	@Value(y-x) = @Value(y...)
Else If	@Value(y-x) = @Value(y...)

Figure 9. Constraints on the selection of verification points.

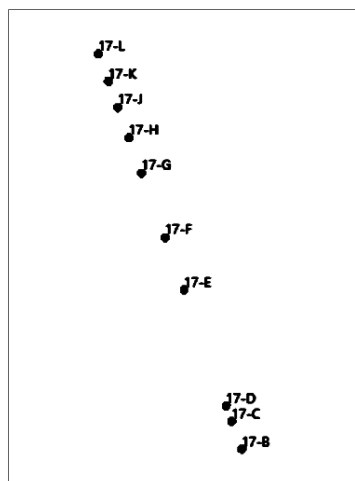


Figure 10. Annotation information for axis line data..

5. CONCLUSIONS

This study has explored methods for automatic data alignment suitable for the construction of CIM building models, conducting research in two aspects: the automatic alignment of real estate subdivision data, and the automatic matching of real estate subdivision data with building contour data. The research methods are based on the axial line space data automatic alignment method and the dynamic angle correction property matching method. The results indicate that the source data, namely the DWG-formatted building design diagrams and the SHAPE-formatted building contour information obtained through automated recognition of remote sensing images, can realize the automatic alignment of source data for building model construction in specific workflows when applying the above two methods. This research implies that due to the strong correlation between building line data and axial line data, indirect alignment of building line data is possible based solely on axial line data, with strict alignment results. In addition, when the precision of building contour data representation is not high enough, matching with real estate subdivision data can be realized through angle correction, not just coordinate correction.

There are some deficiencies in the research methods of this paper. The most significant one is that because the automation workflow is based on FME software development, a single top priority restriction may lead to the selection of multiple extremal points in the axial line intersection, which may raise doubts about the uniqueness of the alignment points. Although this paper has improved extraction restrictions, theoretically, the probability of multiple extremal points is negligible, but it still needs to be verified in large-scale model construction application scenarios for its accuracy.

At present, there are still some key issues related to the results of this study that need to be resolved, namely, the problem that the source data used in this study cannot be automatically extracted temporarily. Extracting the required real estate subdivision diagrams from the building design diagrams (DWG-formatted files) currently requires manual processing using AutoCAD software. There is no automatic extraction method yet, leading to a serious disconnect with the automatic alignment workflow of this research. This issue provides a preliminary direction for follow-up research in this study.

REFERENCES

- Li, X., Ai, W., Feng, R., et al. (2023). A review of deep learning registration methods for remote sensing images [J]. *Journal of Remote Sensing*, 27(02), 267-284.
- Tao, J., Chen, X., Yu, C., et al. (2022). Surface flatness inspection method of segmental beam based on spatial model registration [J]. *Building Structure*, 52(S1), 1962-1967. DOI: 10.19701/j.jzjg.22S1594.
- Wang, S. (2021). Research on model registration method based on 3D point cloud data features [D]. Beijing Institute of Printing. DOI: 10.26968/d.cnki.gbjyc.2021.000027.
- Xiong, C., Gong, X., Shen, Z., et al. (2022). Key technologies for three-dimensional building modeling of real estate registration data based on building stratification and household division [J]. *Bulletin of Surveying and Mapping*, No. 541(04), 149-153. DOI: 10.13474/j.cnki.11-2246.2022.0127.
- Hou, Y., Lu, X., Xiao, F., et al. (2017). Aggregation and segmentation method of property model considering immovable property rights. *Bulletin of Surveying and Mapping*, (488)11, 115-117. DOI: 10.13474/j.cnki.11-2246.2017.0359.
- Li, F. (2019). Key technology of 3D real estate cadastre surveying and mapping production. *Bulletin of Surveying and Mapping*, (510)09, 73-76+114. DOI: 10.13474/j.cnki.11-2246.2019.0288.
- Han, W., Xing, Y., Wang, F., et al. (2022). Research on CIM basic platform for data engineering system construction. *Intelligent Building and Smart City*, (305)04, 6-10. DOI: 10.13655/j.cnki.ibci.2022.04.001.