CIT. Journal of Computing and Information Technology, Vol. 31, No. 2, June 2023, 107–122
doi: 10.20532/cit.2023.1005752

RANSAC Algorithm and Distributed Framework for Point Cloud Processing of Ancient Buildings

Zhu Shen, Ni Luo, Wei Wang and Bo Yang

College of Engineering and Design, Hunan Normal University, Changsha, China

This paper introduces a comprehensive framework for the point cloud processing of traditional buildings. The framework includes segmentation using the RANSAC algorithm and distributed storage based on a fuzzy weighting approach. The methodology employs a height threshold parameter for segmentation to extract building structural elements effectively. Furthermore, an interactive access control model distributes tasks across nodes to achieve load balancing during point cloud matching and analysis. Experimental results demonstrate segmentation accuracy exceeding 99% and alignment time reduction to 967 ms for point cloud models. The distributed computation efficiency reaches 0.8, outperforming conventional methods. The proposed techniques enable accurate dimensional capture, efficient data storage, and information extraction from traditional buildings to support digital preservation.

ACM CCS (2012) Classification: Artificial intelligence → Computer Vision → Computer vision problems
Computer systems organization → Architectures → Distributed architectures

Keywords: RANSAC algorithm, Point cloud segmentation, Gaussian filtering, 3D NDT, Ancient architecture

1. Introduction

Chinese traditional architecture serves as a testament to the nation's historical evolution, encapsulating its economic, political, and cultural growth. Beyond their architectural and artistic significance, these structures hold profound historical, scientific, and societal value [1]. However, the vulnerability of these ancient buildings to environmental elements like water, fire, ice, and natural decay, coupled with the limitations of traditional conservation methods, has jeopardized their integrity and security.

Efforts to preserve and restore historically significant structures have been ongoing due to their importance and fragility. Despite accumulating extensive experience in this field, the predominant use of paper-based records and books for documentation poses significant challenges. These mediums are susceptible to damage, occupy considerable space, and demand substantial resources for maintenance, thereby impeding efficiency in storage and access to critical information [2–3].

Moreover, the rapid pace of China's economic and social development, especially the surge in urbanization, new rural development, and tourism, has heightened the threat to these ancient edifices. The prevailing reliance on traditional measurement techniques and manual documentation, largely through text and hand-drawn forms, not only proves inefficient and time-consuming but also fails to ensure the accuracy and preservation of architectural archives.

The advent of digital technology presents a promising solution to these challenges. Embracing digital technology offers multifaceted advantages. It enables comprehensive data collection and excavation, furnishing a robust foundation for informed decision-making in safeguarding ancient structures. Additionally, digital tools facilitate swift and precise information retrieval, significantly enhancing the efficiency of architectural data management.

Digital technology encompasses a spectrum of electronic computing facets, including big data, cloud computing, the Internet of Things, blockchain, and artificial intelligence. Each facet plays a distinctive role in preserving ancient buildings. Traditionally, measurements were conducted manually, often resulting in limited and cumbersome data management. Recent technological advancements offer alternatives that surpass single-point measurements. Methods such as total stations and leveling instruments transcend the limitations of traditional approaches. Nonetheless, these methods still pose challenges in terms of data accuracy and ease of management [4–5].

As the digital archiving of ancient buildings progresses, extracting implicit information from these structures has become a challenging aspect of their preservation. Extracting deformation data from point cloud information offers an efficient method to acquire accurate and reliable support data for conserving and restoring historical buildings. This approach is particularly valuable for assessing load-bearing structures like bases and columns [6].

Scholars researching digital archiving and conservation of ancient buildings, multi-source 3D model data fusion methods, cultural information extraction methods, and salvage information extraction methods, have achieved notable results. These fruitful research outcomes not only guide the conservation and restoration efforts for ancient buildings but also contribute valuable experience for future investigations across various topics [7–8]. However, current research still faces obstacles such as incomplete access to comprehensive information regarding ancient buildings, the need for comprehensive results in digital archiving, and insufficient exploration of the fusion of 3D data from multiple sources, cultural information extraction, and the analysis of salvage information detection pertaining to ancient buildings.

Furthermore, the process of acquiring point cloud data for traditional buildings presents challenges as the data volume and computational requirements dramatically increase. It becomes increasingly complex to achieve a simultaneous linear expansion of calculation and storage capacity. Currently, the solution often involves purchasing additional storage devices [9]. However, solely upgrading storage presents difficulties in unified device management. Multiple individuals accessing the data can give rise to security risks and significant cost escalation.

To address these challenges, a distributed intelligent control system can distribute the collected point cloud data to individual intelligent nodes for processing [10–12]. This system dynamically assigns tasks to different nodes based on factors such as computing resources, storage capacity, and processing capability, ensuring load balancing and efficient data processing. The exchange and collaboration of different data facilitate the fusion of point cloud data [13–14]. Distributed intelligent control systems facilitate operations such as data transfer, sharing of computation results, and collaborative processing among nodes through reliable communication mechanisms and protocols.

Traditional distributed storage methods for large-scale unstructured data primarily include NoSQL-based methods [15], Spark-based methods [16], and methods based on information dispersion algorithms [17]. These approaches are designed to achieve distributed storage of large-scale unstructured data by extracting feature quantities, performing clustering analysis through correlation information, and employing compression-aware techniques. However, these traditional methods often lack self-adaptability in the distributed storage of unstructured data and suffer from low data compression accuracy, high storage overhead, and suboptimal data storage performance, particularly over extended storage durations.

This paper aims to enhance the distributed storage of extensive unstructured data by developing methods that exhibit self-adaptability, improve data compression accuracy, minimize storage overhead, and optimize performance, especially for prolonged storage periods. The paper makes significant contributions in the following areas:

1. Operation system design and scheduling algorithm based on distributed heterogeneous processes. This innovative approach enhances operational efficiency and reduces costs by considering multiple perspectives. The design of the system incorporates improvements in operation modes, efficiency, and cost reduction.

2. Introduction of a RANSAC algorithm with a designated threshold for height parameter segmentation. This algorithm analyzes and tests the collected point cloud data, resulting in an enhanced segmentation model. The inclusion of the specified threshold improves both the completeness of the segmentation and the running time of the algorithm.

2. Related Works

The 3D laser scanning technology enables accurate acquisition of indoor spatial information [18], and the point cloud data obtained from it has recently been explored for its applicability in generating architectural floor plans. Oesau [19] utilized a point cloud density histogram to segment the point cloud of walls, employing the Hough transform to extract straight lines from the binary representation of the wall lines. Subsequently, the obtained wall lines were superimposed onto the original map, resulting in the final depiction of building walls.

In a similar vein, Yang [20] employed robust Principal Component Analysis (PCA) to compute the normal vectors of the point cloud, thereby extracting the boundaries of the walls to construct the interior floor plan. The study mainly concentrated on reducing noise and extracting wall line data from indoor point cloud information. However, they overlooked integrating location data associated with door and window openings. To address this, Zhou *et al.* [21] captured 2D images from the 3D point cloud model through photography, obtained an approximate range of doors and windows, and subsequently mapped this information back to the 3D point cloud [22]. By extracting contour lines from the localized point cloud data, the location information of door and window feature corner points was determined. Verbree [23] explored the integration of point cloud and SLAM-based laser scanner trajectories, identifying doors based on local minimum vertical distances. This feature was then applied to partition the subspace according to the location of the doors.

While these methods effectively detect door and window elements in 3D point cloud models, their complexity hinders seamless integration with wall extraction processes. As a result, ongoing research converting 3D point clouds into floor plans has primarily focused on generating wall lines, overlooking the critical task of pinpointing the exact locations of openings like doors and windows. Furthermore, there exists a dearth of efficient algorithms that unify the extraction of walls, doors, and windows, thereby streamlining the overall process.

Kuang *et al.* [24] introduced a novel approach for extracting image information, which automatically retrieves the necessary image details from point cloud data. The feasibility of this method was verified across various point cloud data objects, leading to the development of an automated extraction toolkit. On a similar note, Griffiths [25] proposed a processing method to classify and filter point cloud data, segregating sound and vegetation data based on their dimensions and rupture areas.

Currently, most scholars primarily concentrate on classifying or extracting 3D dimensional information such as building outlines, facades, and windows and doors, with limited research on categorizing and extracting cultural information from ancient structures. Through the research and practice of this system, some researchers have proposed that it can be transformed into relevant design application research in other art design fields, which has important practical significance for the future expansion of virtual reality space [26].

One of the most important issues in ancient architecture is the management of ancient buildings in order to monitor their conservation status or plan their maintenance actions. Thanks to the Scan-to-BIM process, this issue can now be resolved. This approach allows the creation of parametric models in a BIM environment, enabling enhanced geometric representation of build elements and integration of different types of data [27].

To facilitate a more convenient and expeditious acquisition of building features, most building extraction methods based on target features employ point cloud rasterization [28]. The point cloud is transformed into a Digital Surface Mode with a 2m interval through interpolation, as demonstrated by Guo *et al.* [29]. Subsequently, the ground model is derived from the DSM model, and both models are analyzed

using Laplace's algorithm and the pixel neighborhood-based boundary tracking algorithm to extract building information.

Similarly, Rückert [30] and Zhang [31] employ rasterization operations to convert point cloud data into regular grid images of varying resolutions. Bao *et al.* [32] explore building distribution through Gaussian kernel functions for building detection. Furthermore, Paiva *et al.* [33] employ the Watershed algorithm in digital image processing on regular grid images. This involves combining features such as higher elevation and distinct spectral characteristics from vegetation to facilitate the extraction of building point cloud data. A point cloud refers to a collection of points that depict the spatial distribution of a specific target within the same spatial reference system.

Various techniques, such as contact scanners, LIDAR, structured light, and triangulation, are employed to acquire point clouds. Consequently, it becomes crucial to employ suitable algorithms for background removal from the acquired 3D point cloud data. Yuan *et al.* [34] proposed a method that combines the region growth algorithm with the RANSAC algorithm, aiming to enhance the accuracy of segmentation in complex scenes. In a similar vein, Li *et al.* [35] initiated their approach based on the point cloud's normal vector. They employed the point normal equation to estimate the local plane by iteratively refining the plane segmentation through the angle of the normal vector as the weight. Zhao *et al.* [36], on the other hand, utilized local radial basis functions to calculate the curvature of the point cloud using RANSAC. They selected the point with the minimum curvature as the initial seed point, thereby improving the efficiency of point cloud segmentation.

However, their approach does not effectively segment the detailed location of the target. Rasterization for grid images, while aiding visualization and processing, risks information loss due to discretization, potentially compromising the accuracy of the original point cloud representation. Methods like Gaussian kernel functions heavily rely on parameter selection and may struggle with irregular building shapes, affecting the precision of detection. Algorithms such as the Watershed method could suffer from reduced performance when dealing with noise or ambiguous gradient information, impacting segmentation quality.

Approaches integrating RANSAC or normal vector-based segmentation might face challenges in scenes with high density or noisy data, potentially reducing accuracy. Similarly, employing radial basis functions for curvature calculation and seed point selection might encounter limitations in complex surface structures, affecting the segmentation efficiency. While many point cloud segmentation algorithms have reached a certain level of maturity, there is still room for improvement in their effectiveness across diverse scenarios. Most of these improvements rely on information such as the normal vector and curvature of the point cloud, which can potentially result in under-segmentation or over-segmentation.

3. Model Design

3.1. Point Cloud Segmentation

Initially, the raw point cloud data is downsampled to reduce the data transfer rate or data size, and voxel filters are used to preserve the shape characteristics while reducing the overall point count. The process involves utilizing the RANSAC algorithm to fit a plane, computing distances between points and the plane, assessing the deviation between the model and additional data, and filtering out data that meets specific threshold criteria for parameter estimation. Points exceeding the defined threshold range are then eliminated from the point cloud. Finally, the Gaussian filter is used to further denoise the segmented point cloud, and the final result is obtained. The entire segmentation process is shown in Figure 1.

The RANSAC (Random Sample Consensus) algorithm, as an iterative method for estimating model parameters from data containing outliers, operates based on the following principles. Initially, a minimal number of elements required to form a mathematical model is randomly selected from the input point cloud, and these elements are utilized to calculate the corresponding model's parameters. In the context of fitting a plane to a spatial point cloud using

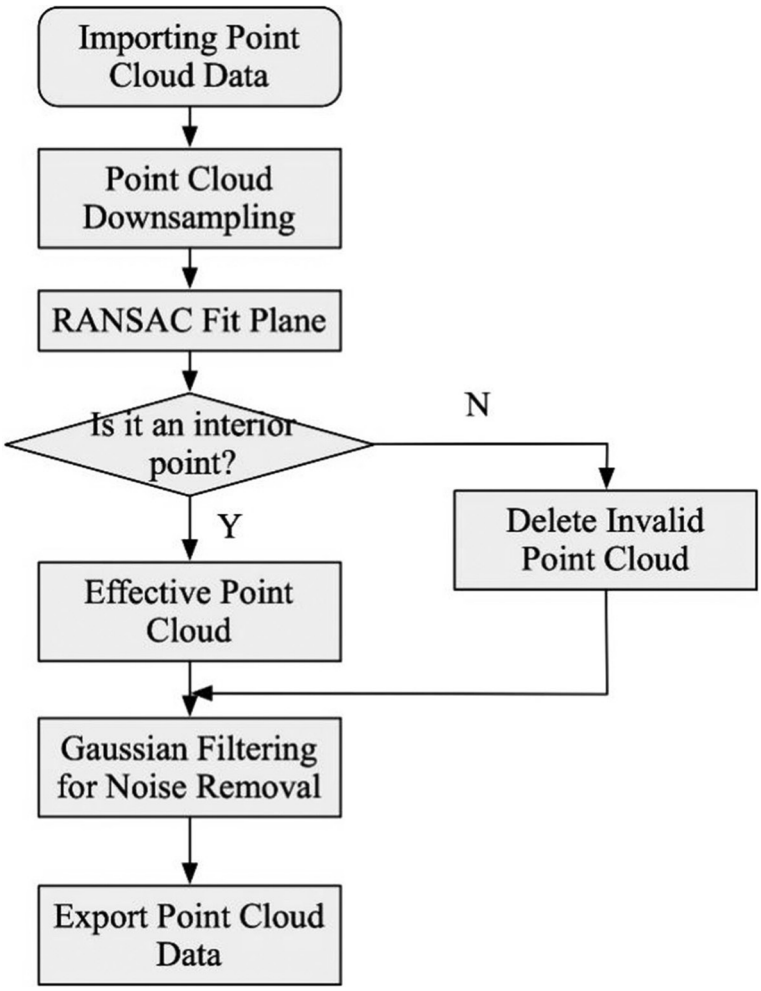

Figure 1. Point cloud segmentation process.

the RANSAC algorithm, three randomly chosen points from the cloud are used to define a plane. Subsequently, the distances between the remaining points and the plane are calculated, and a threshold is applied to determine whether the remaining points lie within the same plane. If a sufficient number of points are found to reside within the same plane, the set of points forming that plane is saved and marked as a match. The iteration continues, extracting the plane with the largest number of matching points until no further matches are found.

Assuming that the percentage of interior points in the data is t, the model is calculated using n points, iterating k and sampling outliers each time. The probability of collecting the correct n points to compute the correct model is:

$$P = 1 - (1 - t^n)^k. \tag{1}$$

This formula essentially represents the likelihood of obtaining the necessary n points after k iterations, given the interior point percentage t and the total number of points n. It quantifies the success rate of the iterative process in collecting the correct data points.

Moreover, to determine the number of iterations required (represented by k), an inverse calculation can be employed using the formula

$$k = \frac{\log(1 - P)}{\log(1 - t^n)}. \tag{2}$$

To mitigate the impact of noise and outliers surrounding the sampled points on the segmentation outcomes, an enhanced version of the RANSAC algorithm incorporates a Gaussian function for filtering high-frequency noise points. This is achieved by utilizing Fourier

transform techniques to assign weights following a Gaussian distribution within a specified region.

Employing a statistical approach to noise removal, the point cloud density is determined through statistical analysis of the point cloud data and the spacing between neighboring points. The density of the point cloud in proximity to the sampled point is then utilized to assess the presence of noise. This algorithm is built upon the KD-Tree structure, which iteratively searches for k nearest neighbors for each point in the point cloud.

The average Euclidean distance from the sampled point to its k nearest neighbors is computed, providing a measure of the average Euclidean distance between the sampled points and their neighboring points. Given the set $P = \{p_i (x_i, y_i, z_i \mid i = 1, 2, ..., n)\}$ as the original point cloud data, after performing a KD-Tree search, the dataset obtained is $P_{ij} = \{p_i (x_{ij}, y_{ij}, z_{ij} \mid j = 1, 2, ..., k)\}$:

$$d_i = \sqrt{\frac{\left(x_{ij} - x_i\right)^2 + \left(y_{ij} - y_i\right)^2 + \left(z_{ij} - z_i\right)^2}{k}} \quad (3)$$

$$\sigma = \frac{\sum_{i=1}^{n}\left(d_i - \bar{d}_i\right)^2}{n} \quad (4)$$

where d_i is the average distance of a point p_i to its k neighborhood points, $\bar{d}_i$ is the mean value of d_i, σ is the standard deviation of d_i.

The distances of all points in the point cloud should form a Gaussian distribution, given the mean and variance, to achieve noise reduction.

3.2. Point Cloud Matching

The 3D Normal Distributions Transform (NDT) algorithm is employed to represent the observed point cloud as a Gaussian probability distribution and subsequently match it with another point cloud [37]. The algorithm follows these steps:

1. The target point cloud model space is subdivided into uniformly sized cubes.
2. For each cube that contains at least 6 points, the mean vector q and the covariance matrix C are calculated based on the points within that cube:

$$q = \frac{1}{n}\sum_{k=1}^{n} x_k \quad (5)$$

$$C = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - q\right)\left(x_k - q\right)^T \quad (6)$$

where x_k is the number of points contained in the cube, and n is the number of points contained in the cube.

3. For each point in the cube x_k, model the normal distribution $N(q, C)$. The probability density function of x_k can be calculated, as shown in Formula (7).

$$p(x) = \frac{1}{\sqrt{2\pi \mid C \mid}} \exp\left[-\frac{\left(x-q\right)^T C^{-1}\left(x-q\right)}{2}\right] \quad (7)$$

A set of normal distributions $\{N(\mu, \Sigma)\}$ is used to represent the point cloud space, forming a segmented smooth spatial representation.

4. Each point of the point cloud to be registered is transformed according to the transformation matrix T (transform matrix), and the probability distribution function of the response is calculated according to the probability density distribution function of the cube where the point $T(x_i)$ lies:

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad (8)$$

where on the top left is the rotation matrix R, the upper right is the translation matrix (vector) t, the lower left is the scaling vector, and the lower right is 1.

5. Multiply the response probability distribution of each point to be aligned as the transfer matrix T as the fractional value of $s(p)$.

$$s(p) = \prod_{i=1}^{n} p\left[T\left(x_i\right)\right] \tag{9}$$

6. The Hessian matrix method is used to find the optimal value of $s(p)$.

Several techniques can be employed in the NDT algorithm to improve the efficiency of the matching process. First, the source point cloud is filtered within the NDT to reduce the data volume, thereby reducing the computational load. Second, if there is a significant disparity in orientation between the two point clouds, initializing an initial transformation matrix can enhance the alignment results. Setting a rotation matrix can significantly reduce the angular difference between the two point cloud models, in particular. Furthermore, during the alignment process, certain parameters related to measurement accuracy need to be adjusted to fit the dataset. The primary objective of the coarse alignment is to swiftly reduce the distance between the two point clouds, focusing on achieving a substantial reduction in distance and minimizing the angular disparity between the two models in the shortest possible time.

3.3. Data Fusion

Figure 2 illustrates the construction of an interactive access control model for unstructured data, utilizing a fuzzy central weighting approach, where the steps of the model are explained as follows.

Based on the aforementioned interactive access control model, the linear weighting control processing is conducted within unstructured data clusters. In this process, the calculation of weighting coefficients is essential.

The point cloud data was fused based on the feature extraction results to calculate the weighting coefficients. The feature space was subsequently divided, and the autocorrelation fusion clustering analysis method was employed to obtain the distribution of the sample space.

$$Q_n = \int_0^j Y_j(n)dn - \frac{1}{X(n)} \tag{10}$$

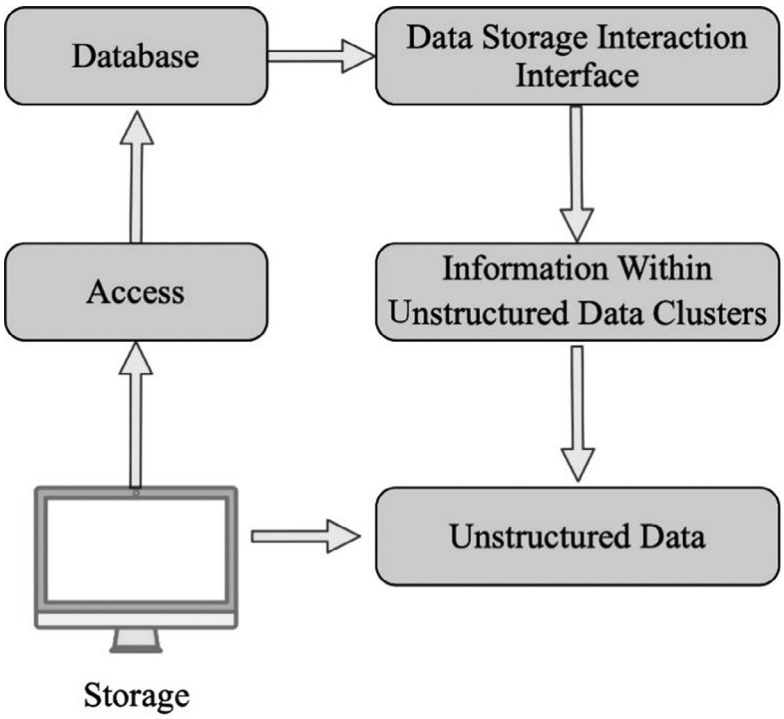

Figure 2. Point cloud data storage process.

The optimal storage of the data is achieved by reorganizing the individual structure information of the large-scale unstructured point cloud data in accordance with the sample spatial distribution, thereby obtaining a fuzzy spatial distribution.

$$w_j(n+1) = \sum_{n=1} Q_n + \left[Y_j(n) + Y_{j-1}(n) \right]^2 \tag{11}$$

Based on the data pre-processing results and the fuzzy spatial distribution, the expected value of the data storage space occupation is shown in Formula (12).

$$\eta = \int_0^j w_j(n+1)\, dn - \sqrt{Q_n} \tag{12}$$

Assuming $w_j(n)$ is the learning weight, the intra-cluster density can be expressed using Formula (13):

$$\theta_j(n) = \theta_{jk}(n) + \eta \cdot \sum_{j=1}^{n} w_j(n) \tag{13}$$

where $\theta_{jk}(n)$ is the inter-class dissimilarity parameter.

Combining the results of intra-cluster tightness calculation, the storage space capacity of large-scale unstructured data is estimated, and the result can be calculated by Formula (14):

$$N^T = q_j^T + \int_0^j \theta_j(n)\, dn \tag{14}$$

where q_j^T denotes the amount of state characteristics of the distribution within the cluster. On the basis of the stored spatial distribution results, the spatial grid clustering method is used to obtain the point cloud data storage results.

$$Rw_j(n) = \theta_j(n) + N^T \cdot \frac{\sqrt{\eta}}{2} \tag{15}$$

The pseudocode of the whole implementation is shown below, which provides the basic framework and functional description of each function. These functions and the main control flow provide a simple framework to better understand the structure of the whole flow.

Algorithm 1.

```
import random
import numpy as np

def RANSAC_segmentation(point_cloud_data,
threshold):
    segmented_points = []
    max_iterations = 1000
    min_inliers = 50
    best_model = None
    best_inliers = []

    for iteration in range(max_iterations):
        sample = randomly_select_sample(point_cloud_
        data)
        model = fit_plane_model(sample)  # Assuming
        plane fitting as an example
        inliers = []

        for point in point_cloud_data:
            if distance(point, model) < threshold:
                inliers.append(point)

        if len(inliers) > min_inliers and
        len(inliers) > len(best_inliers):
            best_model = model
            best_inliers = inliers

    segmented_points = best_inliers
    return segmented_points

def randomly_select_sample(point_cloud_data,
sample_size=3):
    # Randomly select a sample of points from point
    cloud data
    return random.sample(point_cloud_data,
    sample_size)

def fit_plane_model(sample_points):
    # Fit a plane model to the given sample points using
    least squares fitting
    # Assuming a plane equation: ax + by + cz + d = 0
    # Return the plane parameters (a, b, c, d) or the plane
    equation
    # For simplicity, assuming numpy arrays for points
    points_matrix = np.array(sample_points)
    A = np.column_stack((points_matrix[:, 0],
    points_matrix[:, 1], np.ones(len(sample_points))))
    plane_parameters = np.linalg.lstsq(A,
    -points_matrix[:, 2], rcond=None)[0]
    return plane_parameters  # Assuming (a, b, c, d) for
    plane equation

def distance(point, model):
    # Calculate the distance between a point and a plane
    model
    # Assuming the plane model (a, b, c, d)
    return abs(model[0] * point[0] + model[1] *
    point[1] + model[2] * point[2] + model[3]) /
    np.sqrt(model[0]**2 +
            model[1]**2 + model[2]**2)
```

4. Experiments

4.1. Experimental Settings

The data collection for this study was performed using a ground-based 3D laser scanner, specifically the FARO® FocusS 70 Laser Scanner. This scanner offers a positioning accuracy of 2 mm/10 m and an accuracy error within ±1 mm, with a measurement speed of $1 \cdot 10^6$ s^{-1}.

For data processing, the Python programming language was employed alongside the Open3D point cloud data processing library. This combination facilitated the intelligent generation of building plans based on point cloud segmentation.

The data acquisition frequency was set at 120 Hz, with the sample data collected at a rate of 0.5 Hz. The data fusion clustering algorithm underwent 50 iterations, utilizing a discrimination threshold of 0.46 to distinguish between data attributes. These details provide a more comprehensive understanding of the equipment used, data acquisition frequency, processing tools, and specific parameter settings involved in the data acquisition and processing for this study.

The parameters of the RANSAC algorithm are set as follows. The maximum number of iterations is set at 1000, 50 samples are randomly selected as sampling points in each iteration, the threshold is set at 0.1 to determine whether the data points are suitable for the model, and the stopping condition is 0.95 within the model confidence or error range. For the line fitting model, the initial slope and intercept need to be determined according to the characteristics of the data set.

4.2. Analysis of Results

In this study, three alignment algorithms were compared with the proposed algorithm, namely Iterative Closest Point (ICP) [38], Normal Distributions Transform (NDT) [39], and Two-Stage Iterative Closest Point (TICP) [40].

ICP is a classic point cloud registration algorithm that minimizes the distance between two point clouds through iterative optimization. Its core idea is to align two point clouds to achieve the best spatial match. ICP performs well when the initial alignment for registration is poor or when noise is present.

The NDT algorithm models point clouds using Gaussian distribution functions and matches these functions to achieve registration. It is particularly suitable for point cloud data with significant noise and incomplete information and exhibits better robustness in complex environments compared to ICP.

TICP is an improved version of ICP that enhances registration accuracy through a two-stage iterative process. The first stage involves global searching and coarse alignment, while the second stage focuses on local optimization. This two-stage approach improves the efficiency and accuracy of registration.

The alignment results, averaged over 50 iterations, for these four algorithms are presented in Figures 3 and 4. These figures depict the alignment accuracy and alignment time achieved by each algorithm.

In terms of alignment accuracy within the same order of magnitude, several algorithms were compared. Among these algorithms, the ICP algorithm was found to be the slowest in terms of computational speed. It was followed by the NDT algorithm, which exhibited slightly faster alignment but lower accuracy. In contrast, the TICP algorithm showcased notable speed enhancements compared to the ICP algorithm while preserving a comparable level of alignment accuracy. This indicates that the TICP algorithm is a more efficient alternative to the traditional ICP algorithm, striking a balance between speed and accuracy.

However, the proposed algorithm surpassed all the other algorithms in terms of computational speed while maintaining consistent alignment accuracy. It outperformed the TICP algorithm by reducing the alignment time by more than half. This highlights the potential of the proposed algorithm as a highly efficient solution for point cloud alignment tasks. Furthermore, when the two-point cloud models were far apart, all four algorithms exhibited alignment accuracy within the same order of magnitude and were roughly similar. This implies that regardless of the algorithm used, the accuracy of alignment remains comparable when dealing with distant point cloud models.

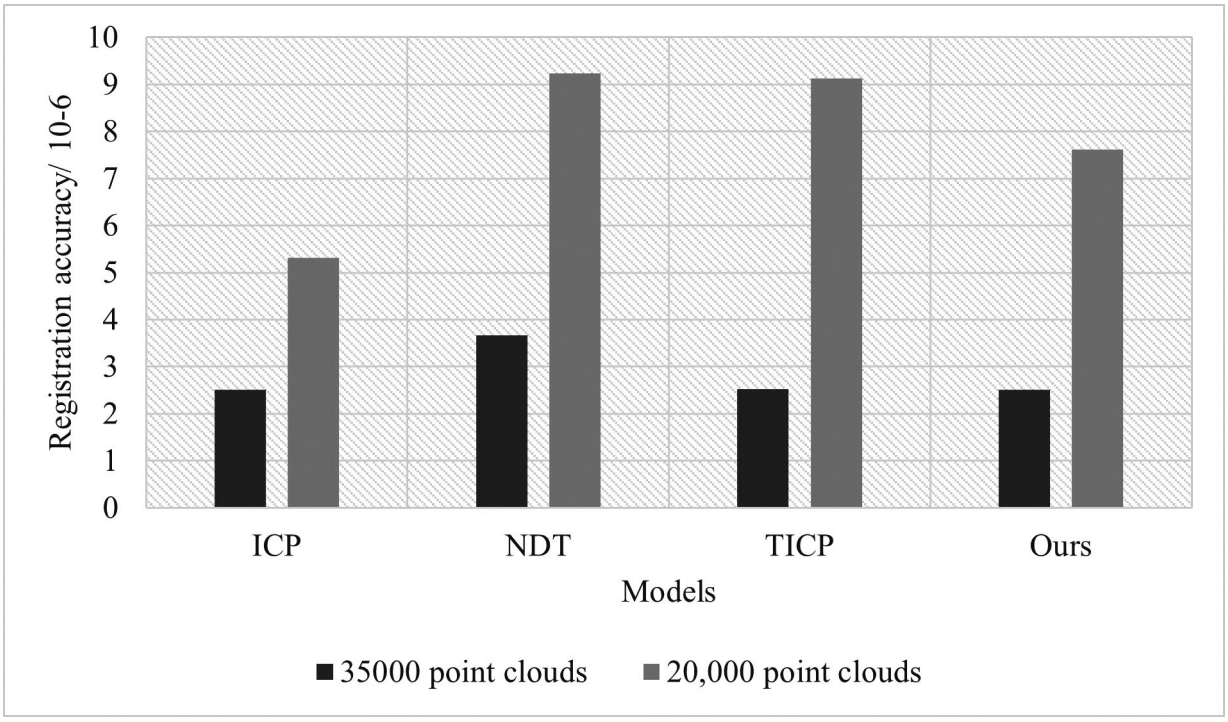

Figure 3. Algorithm comparison (Registration accuracy).

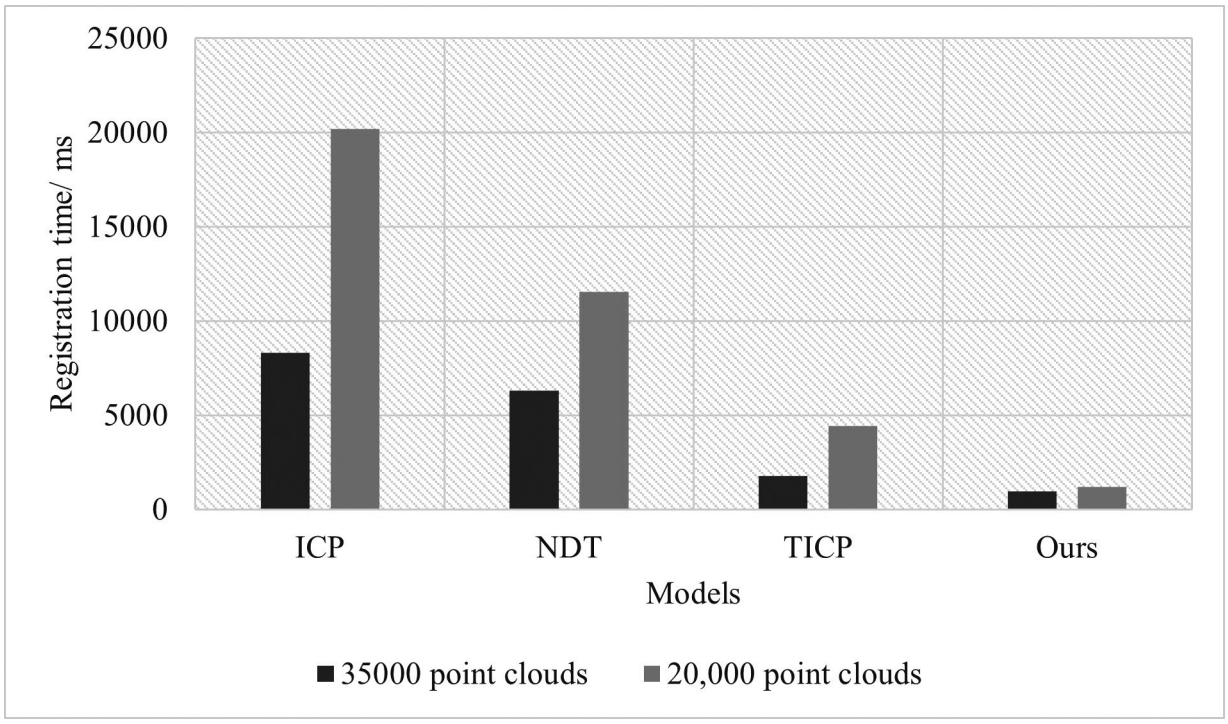

Figure 4. Algorithm comparison (Registration time).

In conclusion, the ICP algorithm provides high accuracy but consumes more time, while the NDT algorithm sacrifices some accuracy for faster alignment. The TICP algorithm achieves accuracy similar to the ICP algorithm with improved speed, and the proposed algorithm demonstrates accuracy comparable to the ICP algorithm while significantly reducing alignment time.

Additionally, the suggested method expedites the target extraction process, yet the segmented target outcomes might include abnormal noise. Nonetheless, the current algorithm adeptly eliminates such aberrant noise points during the point cloud segmentation process, thereby enhancing overall runtime efficiency.

To quantitatively evaluate the building filtering results, a comparative assessment was conducted on the point cloud set of the survey area using three metrics: correctness, completeness, and quality. The assessment results are presented in Figure 5, providing an overview of the performance in terms of these metrics.

In quantitative analysis, both the methods proposed in this paper and the comparative methods demonstrate a high level of completeness. This is attributed to the fact that both approaches perform building point cloud extraction through filtering operations. As buildings have distinct features, the probability of misclassifying building point clouds as non-building point clouds is extremely low in this process. Regarding correctness, the multi-attribute fusion of non-building point cloud filtering methods exhibits good detection performance. Compared to the traditional filtering method based on a single vegetation index, it shows improvements of 5.03% and 3.47% on the two datasets, respectively. Additionally, compared to the filtering method that fuses geometric information, it achieves improvements of 4.79% and 2.08%, respectively.

However, in areas with complex compositions of features, diverse vegetation types, and intricate housing structures, the correctness of the classification results cannot be guaranteed, even with the combination of color and scale information. Furthermore, the method proposed in this paper relies solely on the inherent characteristics of the point cloud for filtering. The reliability of color and local information obtained directly from the point cloud can be affected by issues such as illumination, occlusions, and uneven and sparse point cloud density. These limitations highlight the challenges of this method in complex environments.

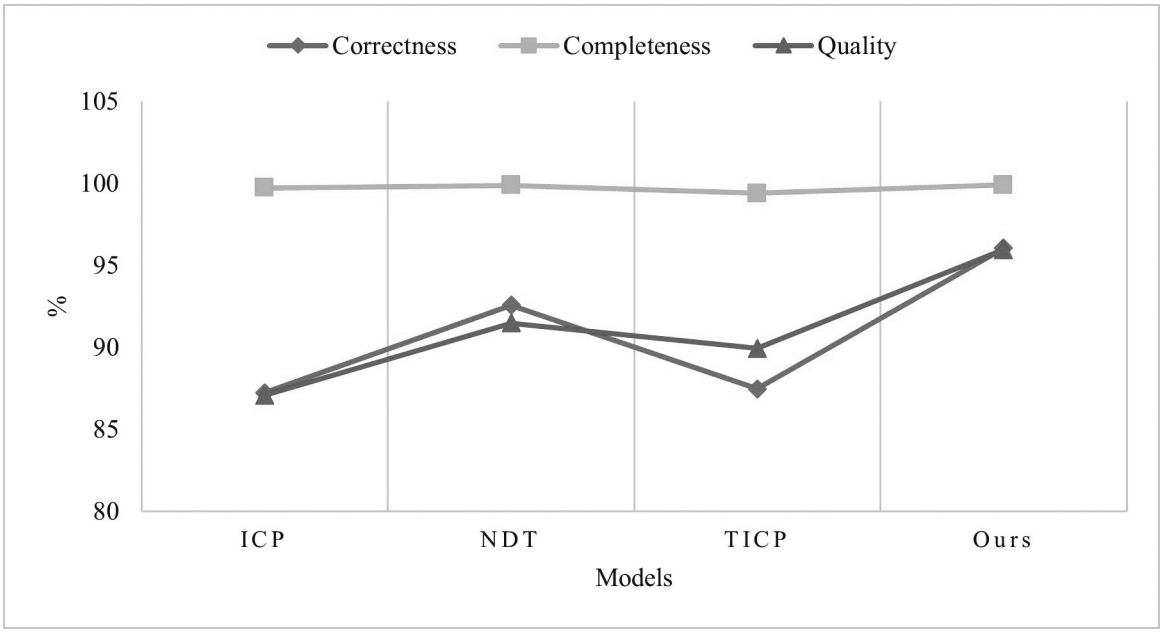

Figure 5. Building filtering results.

Figure 6 showcases an effective computational example of unstructured data distributed storage before and after comparing the application based on the experimental setup described earlier.

The analysis of Figure 6 demonstrates a clear trend: with each iteration increase, the effective computation in distributed data storage consistently improves, surpassing the pre-application baseline. This enhancement, facilitated by the utilization of linear fusion and binary programming for large-scale unstructured data reconstruction, showcases compelling numeric results. Notably, the maximum effective computation value reaches approximately 0.8, marking a significant leap beyond the performance benchmarks set by traditional algorithms. These results substantiate the method's efficacy, showcasing a notable 25% increase in effective computation compared to conventional approaches, underscoring its superior storage capabilities, especially for unstructured point cloud data.

4.3. Discussion

The experimental results demonstrate the limitations of traditional methods for processing point cloud data, which often rely on a single processing mode and fail to fully leverage distributed computing resources. In contrast, this paper introduces a novel job model through the design of a scheduling algorithm and a distributed, heterogeneous processing platform. This model enables multiple computing nodes to simultaneously execute processing tasks, leading to the efficient utilization of computational resources and improved overall processing efficiency. By assigning appropriate jobs to compute nodes, the system load is effectively reduced, further enhancing processing performance.

Furthermore, this paper focuses on addressing the issue of efficiency improvement in point cloud data segmentation. A new approach is adopted, involving the use of the RANSAC algorithm with a specified height threshold parameter for segmentation. By precisely setting

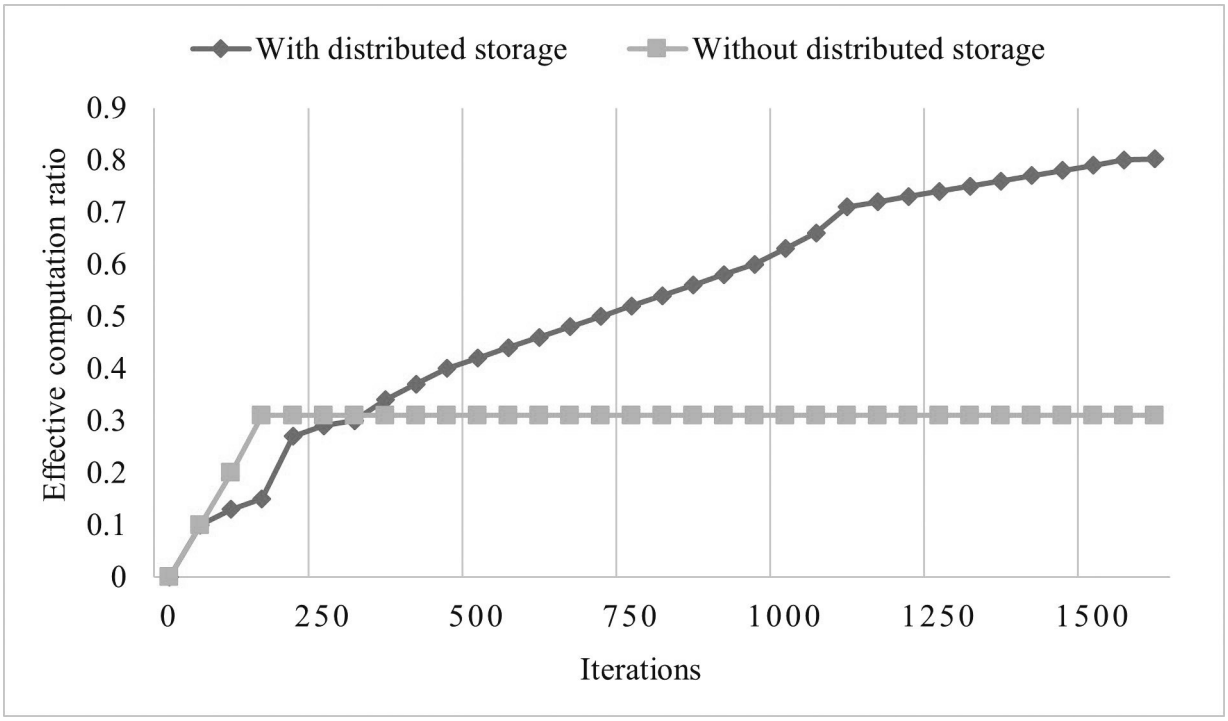

Figure 6. Effective calculation ratio of the model.

the segmentation threshold, the algorithm accurately segments the point cloud data into meaningful models. Additionally, the RANSAC algorithm with a distance threshold is employed to fit a plane and calculate the distance of points to the plane. This threshold-based filtering removes points outside the threshold, improving the segmentation of the measured target. The refined segmentation method not only elevates the segmentation model's quality but also minimizes the computation of invalid data in subsequent steps, significantly enhancing the overall processing efficiency.

The paper can indeed explore the idea of cost reduction in addition to operational model innovation and efficiency improvement. The scheduling algorithm and platform design of distributed heterogeneous processes enable the system to fully utilize different types of computing resources, including distributed computing nodes and heterogeneous processors. This flexible resource scheduling approach can effectively reduce the cost of the system, as there is no need for additional investment in expanding computational capacity. By optimizing algorithms and leveraging parallel computing, energy consumption and operation and maintenance costs can be reduced, further enhancing the economic feasibility of the system.

However, while the proposed system showcases significant advancements in processing efficiency and resource utilization, it is essential to acknowledge certain limitations within this approach. One notable constraint lies in the inherent complexity of integrating diverse computing nodes and heterogeneous processors. While the scheduling algorithm and distributed platform design optimize resource allocation, managing diverse hardware configurations and ensuring seamless coordination among various nodes could introduce complexities in system maintenance and compatibility.

Moreover, despite the emphasis on enhancing segmentation accuracy using the RANSAC algorithm with specified thresholds, determining universally optimal threshold parameters for different datasets or environments remains a challenge. The adaptability of these thresholds to diverse and dynamic point cloud datasets might pose challenges in achieving consistently accurate segmentation across varied scenarios. Furthermore, although the paper discusses cost reduction via improved resource utilization and optimized algorithms, conducting a comprehensive cost-benefit analysis, or providing a detailed breakdown of potential savings in real-world deployments would bolster the economic argument for implementing this system.

Future iterations or extensions of this research could delve deeper into addressing these limitations. Exploring methods to streamline the management of diverse computing nodes, devising adaptive thresholding techniques for improved segmentation across diverse datasets, and conducting thorough cost analyses in practical settings would significantly contribute to the holistic understanding and applicability of the proposed system.

To sum up, the paper can explore additional applications and benefits of the improved methodology. For example, it can incorporate multi-term historical 3D model data of ancient buildings to visually present the historical evolution of these structures or provide a platform for previewing ancient building renovation, planning, and design. This allows relevant workers to intuitively view the surrounding environment of ancient buildings and make informed planning adjustments and decisions. By integrating the 3D laser point cloud model of the ancient building's ground with tilt model data, a comprehensive and accurate model of the ancient structure can be obtained, meeting the centimeter-accuracy requirements for ancient building engineering mapping. This fusion model can serve as a 3D spatial database for the repair, planning, and design of ancient buildings. Additionally, modeling and printing technology can be employed to create 3D miniature models of ancient buildings, producing valuable cultural and creative products and enabling interactive experiences related to ancient buildings. By expanding the scope of applications and considering the economic aspects, the paper can provide a more comprehensive and impactful contribution to the field.

5. Conclusion

In conclusion, this paper introduces an effective point cloud processing approach for traditional buildings involving segmentation, denoising, matching, and distributed storage modules. A key contribution is the RANSAC algorithm, which has a height threshold for improved segmentation quality and computational efficiency. The methodology not only extracts accurate building information but also addresses the intense storage and processing requirements of point cloud data through distributed control. Experiments demonstrate over 99% accuracy in segmentation, a 60% reduction in alignment time to 967ms, and a distributed computation efficiency of 0.8. Overall, the proposed techniques significantly advance point cloud-based modeling and digital archiving solutions for heritage structures.

However, limitations exist in terms of the robustness of segmentation in complex environments. Future efforts will incorporate color information analysis for enhanced performance across diverse scenarios. This study did not consider the similarity judgment of color information. Considering the low sensitivity of RGB color space, in the subsequent work, we will consider using linear transformation to transform RGB to HSV, LAB, and other color spaces, and explore more accurate color similarity calculation methods to improve the accuracy of color similarity calculation, so as to deal with the more complex problem of building monomer extraction.

Acknowledgement

This work was supported by the Technology Innovation Program Project of the Hunan Provincial Department of Science and Technology,"Xiangxi Tujia Brocade 3D Digital Museum and Digital Cultural and Creative Development "has achieved phased results; Project number: 2023NK4278.

This work was supported by the topic of the Education Department of Hunan Province, "Research on the Community of Rural Human Settlement Environment Governance under the Background of Urban Rural Integration - Taking Jianyan Village, Longshan County, Xiangxi Autonomous Prefecture as an Example", has achieved phased results;Project number: 23A0044.

References

[1] S. Sun and B. Wang, "Low-altitude UAV 3D Modeling Technology in the Application of Ancient Buildings Protection Situation Assessment", *Energy Procedia*, vol. 153, pp. 320–324, 2018. http://dx.doi.org/10.1016/j.egypro.2018.10.082

[2] Y. Yin and J. Antonio, "Application of 3D Laser Scanning Technology for Image Data Processing in the Protection of Ancient Building Sites Through Deep Learning", *Image and Vision Computing*, vol. 102, p. 103969, 2020. http://dx.doi.org/10.1016/j.imavis.2020.103969

[3] R. Verbeek and S. Overbeek, "A Critical Heuristics Approach for Approximating Fairness in Method Engineering", *International Journal of Information Technologies and Systems Approach (IJITSA)*, vol. 1, no. 1, pp. 1–17, 2022. http://dx.doi.org/10.4018/IJITSA.289995

[4] J. Li *et al.*, "An Experimental Study of the Damage Degrees to Ancient Building Timber Caused by Lightning Strikes", *Journal of Electrostatics*, vol. 90, pp. 23–30, 2017. http://dx.doi.org/10.1016/j.elstat.2017.08.009

[5] C. Taouche *et al.*, "Palmprint Recognition System Based on Multi-Block Local Line Directional Pattern and Feature Selection", *International Journal of Information Technologies and Systems Approach (IJITSA)*, vol. 15, no. 1, pp.1–26, 2022. http://dx.doi.org/10.4018/IJITSA.292042

[6] J. Cai *et al.*, "BIM Technology of Implicit and Explicit Parts of Historical Building Components Based on Point Cloud Data and Digital BIM Technology of Implicit and Explicit Parts of Historical Building Components Based on Point Cloud Data and Digital Radiographic Image: A Review", *Journal of Asian Architecture and Building Engineering*, 2023. https://doi.org/10.1080/13467581.2023.2215845

[7] U. Almac *et al.*, "Numerical Analysis of Historic Structural Elements Using 3D Point Cloud Data", *The Open Construction & Building Technology Journal*, vol. 10, no. 1, 2016. https://doi.org/10.2174/1874836801610010233

[8] P. V. V. Paiva *et al.*, "Historical Building Point Cloud Segmentation Combining Hierarchical Watershed Transform and Curvature Analysis", *Pattern Recognition Letters*, vol. 135, pp. 114–121, 2020. https://doi.org/10.1016/j.patrec.2020.04.010

[9] K. Kambatla *et al.*, "Trends in Big Data Analytics", *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573, 2014. https://doi.org/10.1016/j.jpdc.2014.01.003

[10] S. S. Ali and B. J. Choi, "State-of-the-art Artificial Intelligence Techniques for Distributed Smart Grids: A Review", *Electronics*, vol. 9, no. 6, pp. 1030, 2020. https://doi.org/10.3390/electronics9061030

[11] K. Tabassum *et al.*, "An Efficient Emergency Patient Monitoring Based on Mobile Ad Hoc Networks", *Journal of Organizational and End User Computing (JOEUC)*, vol. 34, no. 4, pp. 1–12, 2022. https://doi.org/10.4018/JOEUC.289435

[12] A. Savoli and M. Bhatt, "Chronic Patients' Emotions Toward Self-Managing Care IT: The Role of Health Centrality and Dependence on IT", *Journal of Organizational and End User Computing (JOEUC)*, vol. 34, no. 4, pp. 1–14, 2022. https://doi.org/10.4018/joeuc.288550

[13] A. Castañeda-Miranda and V. M. Castaño-Meneses, "Internet of Things for Smart Farming and Frost Intelligent Control in Greenhouses", *Computers and Electronics in Agriculture*, vol. 176, p. 105614, 2020. https://doi.org/10.1016/j.compag.2020.105614

[14] R. G. N. Ngassam *et al.*, "An Action Design Research to Facilitate the Adoption of Personal Health Records: The Case of Digital Allergy Cards", *Journal of Organizational and End User Computing (JOEUC)*, vol. 34, no. 4, pp. 1–18, 2022. https://doi.org/10.4018/JOEUC.288551

[15] S. Sicari *et al.*, "Security&privacy Issues and Challenges in NoSQL Databases", *Computer Networks*, p. 108828, 2022. https://doi.org/10.1016/j.comnet.2022.108828

[16] N. Maleki *et al.*, "Sofa: A Spark-oriented Fog Architecture", in *Proc. of the IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, 2019, no. 1, pp. 2792–2799. https://doi.org/10.1109/IECON.2019.8927065

[17] Y. Lei *et al.*, "BIM Based Cyber-physical Systems for Intelligent Disaster Prevention", *Journal of Industrial Information Integration*, vol. 20, p. 100171, 2020. https://doi.org/10.1016/j.jii.2020.100171

[18] X. Gao *et al.*, "Complete Scene Reconstruction by Merging Images and Laser Scans", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 10, pp. 3688–3701, 2019. https://doi.org/10.1109/TCSVT.2019.2943892

[19] S. Oesau *et al.*, "Indoor Scene Reconstruction Using Feature Sensitive Primitive Extraction and Graph-cut", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 90, pp. 68–82, 2014. https://doi.org/10.1016/j.isprsjprs.2014.02.004

[20] L. Yang *et al.*, "Efficient Plane Extraction Using Normal Estimation and RANSAC from 3D Point cloud", *Computer Standards & Interfaces*, vol. 82, p. 103608, 2020. https://doi.org/10.1016/j.csi.2021.103608

[21] Z. Zhou *et al.*, "Three-dimensional Reconstruction of Huizhou Landscape Combined with Multimedia Technology and Geographic Information System", *Mobile Information Systems*, vol. 2021, pp. 1–13, 2021. https://doi.org/10.1155/2021/9930692

[22] B. Quintana *et al.*, "Door Detection in 3D Coloured Point Clouds of Indoor Environments", *Automation in Construction*, vol. 85, pp. 146–166, 2018. https://doi.org/10.1016/j.autcon.2017.10.016

[23] L. Díaz-Vilariño *et al.*, "Indoor Modelling from SLAM-based Laser Scanner: Door Detection to Envelope Reconstruction", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 345–352, 2017. https://doi.org/10.5194/isprs-archives-XLII-2-W7-345-2017

[24] H. Kuang *et al.*, "Voxel-FPN: Multi-scale Voxel Feature Aggregation for 3D Object Detection from LIDAR Point Clouds", *Sensors*, vol. 20, no. 3, p. 704, 2020. https://doi.org/10.3390/s20030704

[25] D. Griffiths and J. Boehm, "A Review on Deep Learning Techniques for 3D Sensed Data Classification", *Remote Sensing*, vol. 11, no. 12, p. 1499, 2019. https://doi.org/10.3390/rs11121499

[26] T. Wang and L. Zhao, "Virtual Reality-based Digital Restoration Methods and Applications for Ancient Buildings", *Journal of Mathematics*, vol. 2022, pp. 1–10, 2022. https://doi.org/10.1155/2022/2305463

[27] M. Aricò and M. Lo Brutto, "From Scan-to-BIM to Heritage Building Information Modelling for an Ancient Arab-Norman Church", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 761–768, 2022. https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-761-2022

[28] W. Ali *et al.*, "A Feature Based Laser SLAM Using Rasterized Images of 3D Point Cloud", *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24422–24430, 2021. https://doi.org/10.1109/JSEN.2021.3113304

[29] L. Guo *et al.*, "Extraction of Dense Urban Buildings from Photogrammetric and LiDAR Point Clouds", *IEEE Access*, vol. 9, pp. 111823–111832, 2021. https://doi.org/10.1109/ACCESS.2021.3102632

[30] D. Rückert *et al.*, "Adop: Approximate Differentiable One-pixel Point Rendering", *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1–14, 2022. https://doi.org/10.1145/3528223.3530122

[31] Q. Zhang *et al.*, "RegGeoNet: Learning Regular Representations for Large-Scale 3D Point Clouds", *International Journal of Computer Vision*, vol. 130, no. 12, pp. 3100–3122, 2022. https://doi.org/10.1007/s11263-022-01682-w

[32] H. Bao *et al.*, "DFCNN-based Semantic Recognition of Urban Functional Zones by Integrating Remote Sensing Data and POI Data", *Remote Sensing*, vol. 12, no. 7, p. 1088, 2022. https://doi.org/10.3390/rs12071088

[33] P. V. V. Paiva *et al.*, "Historical Building Point Cloud Segmentation Combining Hierarchical Watershed Transform and Curvature Analysis", *Pattern Recognition Letters*, vol. 135, pp. 114–121, 2020. https://doi.org/10.1016/j.patrec.2020.04.010

[34] H. Yuan *et al.*, "Line Laser Point Cloud Segmentation Based on the Combination of RANSAC and Region Growing", in *Proc. of the 2020 39th Chinese Control Conference (CCC). IEEE*, 2020, pp. 6324–6328. https://doi.org/10.23919/CCC50068.2020.9188506

[35] Y. Guo *et al.*, "Deep Learning for 3d Point Clouds: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4338–4364, 2020. https://doi.org/10.1109/TPAMI.2020.3005434

[36] J. Zhao *et al.*, "Automatically Modeling Piecewise Planar Furniture Shapes from Unorganized Point Cloud", *Computers & Graphics*, vol. 90, pp. 116–125, 2020. https://doi.org/10.1016/j.cag.2020.05.019

[37] P. Ram and K. Sinha, "Revisiting kd-tree for Nearest Neighbor Search", in *Proc. of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1378–1388. https://doi.org/10.1145/3292500.3330875

[38] Z. Zhou *et al.*, "Ndt-transformer: Large-scale 3d Point Cloud Localisation Using the Normal Distribution Transform Representation", in *Proc. of the 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE*, 2021, pp. 5654–5660. https://doi.org/10.13140/RG.2.2.28949.04325

[39] J. Huang *et al.*, "Point Cloud Registration Algorithm Based on ICP Algorithm and 3D-NDT Algorithm", *International Journal of Wireless and Mobile Computing*, vol. 22, no. 2, pp. 125–130, 2022. https://doi.org/10.1504/IJWMC.2022.123292

[40] W. Minhua *et al.*, "A Two-Stage Point Cloud Registration Method for Knee Joint Replacement Navigation", in *Proc. of the 2021 WRC Symposium on Advanced Robotics and Automation (WRC SARA). IEEE*, 2021, pp. 42–46. https://doi.org/10.1109/WRCSARA53879.2021.9612682

Received: November 2023
Revised: December 2023
Accepted: January 2024

Contact addresses:
Zhu Shen
College of Engineering and Design
Hunan Normal University
Changsha
China
e-mail: 310440106@qq.com

Ni Luo
College of Engineering and Design
Hunan Normal University
Changsha
China
e-mail: 1047008108@qq.com

Wei Wang*
College of Engineering and Design
Hunan Normal University
Changsha
China
e-mail: weeedesign2022@163.com
*Corresponding author

Bo Yang
College of Engineering and Design
Hunan Normal University
Changsha
China
e-mail: yangbo@mail.hunnu.edu.cn

Zhu Shen received his Master's degree from Hunan University, China, in 2006. He currently serves as an Associate Professor at the College of Engineering and Design, Hunan Normal University, Changsha, China. His research interests primarily revolve around historic preservation.

Ni Luo received his Master's degree from Hunan Normal University, China, in 2021. She currently serves as a research assistant at the College of Engineering and Design, Hunan Normal University, Changsha, China. Her research interests are mainly in the field of urban renewal.

Wei Wang graduated from Guangzhou Academy of Fine Arts, China, in 2009. He currently serves as an Associate professor, master tutor of Design and Art Design (MFA) at the College of Engineering and Design, Hunan Normal University, Changsha, China. His research interests include digital inheritance and innovative design of intangible cultural heritage.

Bo Yang received the doctor's degree from CenterSouth University, China, in 2013. Currently, he serves as the professor in College of Engineering and Design, Hunan Normal University, Changsha, China. His research interest include machine vision, robot and computational intelligence.