A Multifeature Fusion Framework Based on D-S Theory for Automatic Building Extraction From High-Resolution Remote Sensing Imagery

Xuedong Zhang, Xing Li, Jian Huang, Erzhu Li, Wei Liu, and Lianpeng Zhang

***Abstract*—Building information serves as a critical foundational dataset in the fields of urban planning, smart cities and surveying and mapping, and high-resolution remote sensing (HRRS) imagery has become a vital data source for extracting building information. However, automatically extracting building information from HRRS imagery using a single feature or method remains a challenging task. On one hand, buildings exhibit significant variations in terms of their size, color, geometric structures, and other aspects. On the other hand, there are also numerous features in the environment that bear spectral and morphological resemblances to buildings. In this article, we proposed a multifeature fusion framework based on Dempster–Shafer (D-S) theory that consists of two steps for automated building extraction from HRRS imagery. The initial D-S fusion step involves two branches: object-level and pixel-level feature fusion. Then the outcomes are further combined to derive the ultimate building confidence information. In the framework, we introduced a proportional consistency and centroid consistency index to convert pixel-level features to object-level features, thereby facilitating their fusion. In addition, we proposed an initialization module for the basic probability assignment formula, enabling the elimination of the impact of nonbuilding objects and simplifying the construction process of BPAF. The experimental results based on the Nanjing, WHU, and Washington datasets demonstrate the effectiveness of our method, the accuracy outperforms the other four advanced algorithms.**

***Index Terms*—Building extraction, Dempster–Shafer (D-S) theory, multifeature fusion, object-oriented, two-step decision fusion framework.**

I. Introduction

BUILDINGS are not only vital places for human production and living but also important artificial features in geographic databases [1], [2]. Accurate and timely building information is essential in diverse fields such as urban planning, land analysis, environmental monitoring, and autonomous driving. With the development of sensors and remote sensing platforms, high-resolution remote sensing (HRRS) images can provide rich and detailed building details and they have become crucial sources of data for building mapping. However, extracting building information from HRRS images is often difficult to achieve satisfactory results due to the low signal-to-noise ratio of HRRS images [3] and complex building morphology. Therefore, there is an urgent requirement to investigate accurate and effective methods for extraction of building footprint from HRRS images.

Previous research works have utilized various building features, such as spectrums, shapes, textures, and others, to extract building footprints in HRRS [4]. However, relying on a single building feature is insufficient for recognizing building footprints in diverse complex environments. Naturally, several studies have employed multiple features to hierarchically constraint candidates of buildings or simply fuse these features to filter out negative factors [5], [6]. While these methods have demonstrated impressive outcomes, striking a balance between the recall rate and precision rate of buildings remains challenging. Multiple building features often result in conflicting, complementary, and redundant information. The hierarchical extraction or linear fusion method based on multiple features is not sufficient to effectively process information. Effectively integrating different building features, leveraging redundant data to minimize detection inaccuracies, and fully utilizing complementary data are the key to accurate recognition of building footprints.

Dempster–Shafer (D-S) theory could handle incomplete, uncertain, or conflicting information from multiple sources, and provides a reasonable way to represent and reason with such information [7]. Therefore, we propose a multiple building features fusion framework based on D-S theory in this manuscript. The framework consists of two steps of fusion: the first step integrates object-level features and pixel-level features separately, while the second step further fuses the two types of fusion results from the first step. In the pixel-level features fusion stage, we propose the proportional consistency and centroid consistency index (PCCI) index, which represents the probability of an object belonging to a building by calculating the proportion of building pixels within the segmented object and the Euclidean distance between their centroids. The index efficiently converts

Manuscript received 23 September 2023; revised 23 November 2023, 24 January 2024, 21 March 2024, and 8 May 2024; accepted 18 June 2024. Date of publication 1 July 2024; date of current version 12 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42271465, in part by the Double carbon project of Jiangsu Normal University under Grant JSNUSTZX202202, and in part by the Postgraduate Research Innovation Program Project of Jiangsu Normal University under Grant 2022XKT0068. *(Corresponding author: Lianpeng Zhang.)*

Xuedong Zhang, Xing Li, Erzhu Li, Wei Liu, and Lianpeng Zhang are with the School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou 221116, China (e-mail: zhangxuedong@jsnu.edu.cn; lixing@jsnu.edu.cn; liezrs2018@jsnu.edu.cn; liuw@jsnu.edu.cn; zhanglp2000@126.com).

Jian Huang is with the Jiangsu Province Surveying and Mapping Engineering Institute, Nanjing 210013, China (e-mail: 304417348@qq.com).

Digital Object Identifier 10.1109/JSTARS.2024.3421278

pixel-level features into object-level features for subsequent object-based building extraction. For the initial assignment of probabilities to propositions in the D-S theory, we developed an initialization module for basic probability assignment formula (BPAF). This module combines feature constraints and fuzzy clustering methods to effectively mitigate the impact of nonbuilding objects and initialize the BPAF using fuzzy C-means clustering (FCM). In conclusion, our study contains the following main contributions.

1) A two-step fusion framework for building multiple features based on the D-S theory was proposed. The framework extracts building information from spectral, geometric, and texture features of buildings, and utilizes D-S theory to fuse information extracted from different features to solve the instability and uncertainty caused by using a single feature or method. The two-step fusion method can reduce the computational challenges brought by processing a large amount of evidence. The two branches, which are based on object-level and pixel-level features, can complement each other in achieving a high recall rate and precision rate in the analysis of buildings.
2) In the pixel level features fusion stage, we proposed a PCCI tailored to building segmentation objects. This index calculates the proportion of building pixels within the segmented object and the Euclidean distance between their centroids to represent the probability of the object belonging to a building. It serves as a bridge between building pixels and the segmentation objects, effectively integrating the strengths of both pixel-level and object-level building extraction methods. It is worth noting that the PCCI is not only applicable to this study but can also be used in scenarios where any pixel-level detection method is combined with segmentation objects.
3) This research constructed a BPAF initialization module that addresses the challenge of initializing the BPAF in the D-S theory. The module is constructed based on feature constraints and fuzzy clustering. Initially, this module utilizes shadow and vegetation features to restrict the object. Then, FCM clustering is employed to acquire object membership degrees as a substitute for the initial probability values. The module has the capability to mitigate the influence of nonbuilding objects on the extraction outcomes. Furthermore, it simplifies the BPAF construction process without any parameter tuning.

This article is structured as follows. Section II presents relevant research works of building extraction. Section III introduces the implementation process of a two-level decision fusion framework. In Section IV, we present the experimental results and provide relevant analysis. Section V discusses the method of threshold determination and the ablation experiment of the proposed method. Finally, Section VI concludes this article.

II. Related Work

Surveying the research of numerous scholars on building extraction over the past decades, these methods can be classified into two categories based on the need for labeled samples: machine learning based methods and methods based on physical rules [4]. In recent years, machine learning, particularly deep learning, has been extensively applied in the field of remote sensing image processing. Benefiting from the powerful feature extraction ability of deep neural networks, methods of deep learning have shown outstanding performance in building extraction. However, the majority of deep learning algorithms require a large number of annotated samples for model training. The labeling of samples is time-consuming and labor-intensive. The methods that use physical rules rely on the knowledge of domain experts instead of annotated samples for constructing models. The high interpretability of these approaches provides a clear understanding of how each feature contributes to building extraction. Naturally, when comparing deep learning-based methods with physical rule-based methods, the latter exhibit slightly lower accuracy in extracting buildings. Nonetheless, due to not requiring manual data annotation, these methods demonstrate superior extraction efficiency, rendering them the preferred option for specific application scenarios with less stringent precision requirements. Moreover, buildings extracted using physical rules-based methods can also serve as a source of samples for deep learning networks, which help to reduce the effort manually labeling the samples. Therefore, the method of extracting buildings based on physical rules from HRRS images remains of significant research importance. This study mainly reviews methods for extracting buildings based on physical rules.

Convolutional neural networks have attracted much attention due to their powerful feature-learning ability and have also made significant achievements in the field of building automation extraction. Classic semantic segmentation networks such as fully convolutional network (FCN) [8], SegNet [9], DeepLab series [10], etc., have the advantage of actively learning image features and obtaining rich global, contextual, and semantic information [11]. However, due to the complexity and diversity of HRRS image scenes, directly applying the network has poor accuracy. Considering that buildings appear in various sizes, Zhao et al. [12] proposed the PSPNet network to enhance multiscale information fusion and reduce local and global losses. Using pyramid pooling modules to capture and fuse multiscale features to improve segmentation accuracy. Wen et al. [13] designed a multiscale erosion network for building edge information and combined it with a semantic decoding module for constructing edge detection. Aiming at the geometric shape of building roofs, an adversarial shape regularization learning method is adopted to model the shape patterns of buildings [14], to learn vertex deformations and further refine the shape of buildings.

In recent years, with the development of deep learning, some advanced deep learning models have been applied in the field of remote sensing, such as the multiscale interactive fusion network and multistage self-guided separation network. Wang et al. [15] applied the proposed MIFNet network to the classification of hyperspectral and synthetic aperture radar (SAR) images. This model solves the problem of single data source feature extraction and multisource data feature fusion by designing a multiscale interactive information extraction (MIIE) module and a global dependency fusion module. To solve the problem of

discontinuous feature extraction caused by the limitations of traditional methods in their respective fields and the loss of multiscale information. Pang et al. [16] proposed a method called enhanced multiscale feature fusion network (EMFFN) for the classification of hyperspectral images. EMFFN extracts spectral and spatial multiscale features through two subnetworks, and integrates them together using a comprehensive loss function. Zhou et al. [17] proposed a multiscale cross-layer interaction and similarity refinement network to improve the network's representation ability. Multiscale feature interaction fusion network (MSIFNet) [18] is used for fusion tracking of RGBT (RGB and thermal infrared) images. The MSIFNet model should extract multiscale features through different convolutional branches and adaptively aggregate them through feature selection modules. At the same time, the transformer interaction fusion module was used to establish long-distance dependency relationships. Finally, a global feature fusion module was designed to adaptively adjust global information. Multistage self-guided separation network is also a popular method. Wang et al. [19] proposed a multistage self-guided separation network (MGSNet) for remote sensing scene classification. MGSNet utilizes a target background separation strategy to enhance the separation of target background information by extending feature attention across different network branches through comparative regularization. Solved the problem of insufficient feature differences and difficulty in effectively distinguishing different scene images in remote sensing image classification due to imbalanced changes between background and target. Xu et al. [20] proposed a multilayer feature fusion network for scene classification of high spatial resolution images. This method utilizes the VGGNet-16 model to extract features and improves classification accuracy through multilayer feature fusion.

In general, the geometric structure of buildings differs significantly from other objects, making geometric feature-based a highly popular method [21]. In remote sensing images, buildings often exhibit diverse shapes, with rectangles or combinations of rectangles being the most common. Wang et al. [22] proposed a method using line segment detection. This method first uses a line detector to extract lines as candidate rectangular buildings, and then performs line connecting and contour closure through a recursive approach. Sun et al. [23] constructed a regular shape similarity index using the ratio of the object's area to its minimum bounding shape area and then utilized object-based image analysis methods to extract the targets. Ngo et al. [24] introduced a novel method for extracting rectangular buildings. This method merges regions with similar spectral characteristics based on their shadow positions and uses rectangularity to extract buildings. While shape-based extraction methods effectively detect rectangular buildings, they may overlook buildings with other shapes. Cote and Saeedi [25] utilized corner information extracted in the color-invariant space to generate polygonal roof outlines and refined building contours using the level set curve evolution method. Chen et al. [26] introduced multiresolution wavelet transform and local spatial autocorrelation statistics to model the spatial pattern of built-up areas, and used an adaptive threshold algorithm to achieve detection of built-up areas. The detection of built-up areas on ZY-3 and Quickbird full-color satellite images is very effective. Sirmacek and Unsalan [27] proposed a building detection probability framework. They first extracted corners using four different methods and used the corner information as the independent variable to estimate the variable kernel density function. Finally, they detected building locations through density estimation. Unlike the method used by Sirmacek and Unsalan, Munawar et al. [28] employ Gabor filters to detect local feature points and subsequently identify urban areas. Although corner matching methods can address issues arising from lighting and shape differences, their performance is poorer in densely built areas or regions with occluded corners.

Object-oriented methods for building extraction consider the spatial relationships and contextual information between pixels. Grouping pixels with similar features into objects can help prevent the occurrence of salt-and-pepper noise during building extraction, resulting in improved performance [4]. Several popular segmentation methods, such as GrabCut [29], MeanShift [30], and Watershed [31], can be applied in object-oriented building extraction techniques. Liasis and Stavrou [32] developed an optimized active contour level-set segmentation framework based on the RGB color space. They utilized color clustering to remove shadows and vegetation interference, thereby extracting buildings. Jiang et al. [33] proposed a semi-automatic method that first performed mean shift segmentation on the image and simultaneously obtained building boundaries through edge detection. Finally, manual merging of objects was performed to address over-segmentation. In order to address the issue of over-segmentation, studies in papers [34], [35] utilized a multiscale segmentation approach to choose the most suitable threshold for merging objects. Subsequently, the objects were discriminated and classified by combining the building feature database.

Another category of building extraction methods is based on building index indices. Building indices, like the built-up index (PanTex) [36] and built-up area significance index (BASI) [38], are constructed based on building texture features and can accurately identify urban or built-up regions. However, they lack the capability to extract fine building contours. Hu et al. [31] proposed a new spatial feature called object in correlative index (OCI), which considers the characteristics of image objects based on spectral similarity and constructs useful OCI to objectively describe spatial information. Huang and Zhang [39] proposed the morphological building index (MBI), which extracts building features such as contrast, size, and brightness through multiscale and multidirectional top-hat transforms. MBI performs well in urban scenes but often yields less satisfactory results in suburban or rural areas. In addition, there are many variants of MBI algorithms, such as MABI [40] and MFBI [41]. Liu et al. [42] proposed the perceptual building index (PBI) based on the presence of obvious corners around the built-up areas. PBI employs corner detectors to identify corners and junction points, subsequently propagating the significance of local structures throughout the image via spatial kernels, thereby calculating the building index. PBI exhibits independence from local image contrast, ensuring the detection of buildings with lower brightness does not go overlooked. Xia et al. [43] proposed the geometric building index (GBI) to reduce information redundancy caused by spatial voting in PBI. They utilized geometric

reasoning to extract the precise contour of individual buildings. However, the extraction results of GBI could be affected by nonbuilding rectangular objects.

Automatically extracting buildings using a single feature or method has inherent limitations and often fails to produce satisfactory results. A natural approach is to integrate multiple building features to comprehensively describe complex building structures. The common methods involve using multiple features for hierarchical constraint of building candidates or simple fusion to extract buildings [5], [6]. However, these approaches fail to fully exploit the interrelationships and complementarity among features. Due to the D-S theory's ability to effectively handle information from multiple sources, it has been applied in building change detection [1], [38], [44], [45], [46] and building extraction [3], [47]. The initialization of the BPAF in D-S theory greatly impacts the results. But determining how to allocate initial probabilities is a challenging aspect of applying the D-S theory. Previous experiments considered the relationships between different evidence and constructed new BPAFs based on experience [45], [46], [47]. This resulted in limitations in the applicability of the methods. Inspired by previous work [47], we propose a BPAF initialization method based on feature constraints and fuzzy clustering. We use FCM to replace the S-shaped membership function used in [47]. This avoids the need for parameter tuning. In addition, we apply constraints on shadows and vegetation to the obtained building features before assigning initial probabilities to different propositions. This effectively avoids the problem of nonbuilding objects having high rectangularity or low entropy values.

III. Proposed Methodology

This manuscript proposes a two-step D-S fusion framework based on multiple building features for detecting buildings in HRRS images. As depicted in Fig. 1, the framework primarily comprises the following three components.

1) Image segmentation.
2) First-level fusion of building multiple features.
3) Second-level fusion of building features.

In the first-level feature fusion stage, it is divided into two parts: object-level feature fusion and pixel-level feature fusion. In the object-level feature fusion part, first, we extract the rectangularity and entropy value of objects as two object-level features. Then, we construct a BPAF initialization module based on feature constraints and fuzzy clustering to assign an initial probability for the segmentation objects. Based on the D-S theory, we combine the attributes of rectangularity and entropy to calculate the probability that the segmented object belongs to buildings. In the pixel-level feature fusion part, first, we use the MBI and corner-based methods to obtain the building extraction outcomes at the pixel level. By calculating the proportion of building pixels in the objects and the distance between the centroid of buildings and the centroid of objects, we construct the PCCI index to convert pixel-level features into object-level features. In a similar manner to the object-level feature fusion component, we use the same BPAF initialization module and D-S fusion method to obtain the probability of objects belonging to buildings. In the second-level feature fusion stage, the probabilities of the objects generated from the two parts of the first-level fusion are fused again to obtain the final confidence information of the buildings, and the building extraction results are obtained using decision rules.

A. Image Segmentation

The multiscale morphological gradient reconstruction watershed transformation (MMGR-WT) algorithm [48] is employed to generate segmented images for object-oriented analysis. This segmentation method successfully preserves the integrity of object boundaries. The definition of a segmented image is as follows:

$$\mathrm{SI} = \mathrm{WT}(G(g, r_1, r_2)) \tag{1}$$

where SI represents the segmented image, WT represents watershed transformation, and represents the gradient image after multiscale morphological reconstruction. The calculation of G is as follows:

$$G(g, r_1, r_2) = \max\left(R^c_{r_1}(g), R^c_{r_1+1}(g), R^c_{r_1+2}(g), \ldots, R^c_{r_2}(g)\right) \tag{2}$$

where c represents morphological closing constructions. R denotes the newly generated gradient image obtained through the morphological closing reconstruction of the original gradient image g using a structuring element of size r_1. The term "max" denotes the operation of pixel-wise maximum value extraction. r_1 and r_2 represent the scale of the minimum and maximum structural elements, respectively.

The segmentation parameter r2 can be substituted with the error ε, as defined in (3). In our study, the value of ε is set as 0.00001, and the structural element's minimum scale r1 is set as 3

$$|G(g, r_1, r_2) - G(g, r_1, r_2 + 1)| < \varepsilon. \tag{3}$$

B. Object-Level Feature-Based D-S Fusion

In remote sensing imagery, buildings predominantly exhibit geometric structures that are either rectangular or similar to rectangles. In addition, the grayscale values of pixels within the building regions vary slightly, which means that individual buildings possess limited grayscale information, resulting in low entropy values. Therefore, we designed rectangularity and entropy as building features for segmentation objects. The initialization of the BPAF for each proposition has consistently posed a challenge when employing D-S theory for fusing two building features. To address this issue, we propose a BPAF initialization module that is founded on feature constraints and fuzzy clustering. The definition of BPAF can be simplified by substituting initial probability values with fuzzy cluster membership, while ensuring precision in extraction.

1) Extraction of Object-Level Features: Most buildings appear as rectangles in HRRS images. Thus, we can determine if an object is a building by assessing its resemblance to a rectangle. This study employs the concept of "rectangularity, which is defined as the ratio of the object's area to the area of

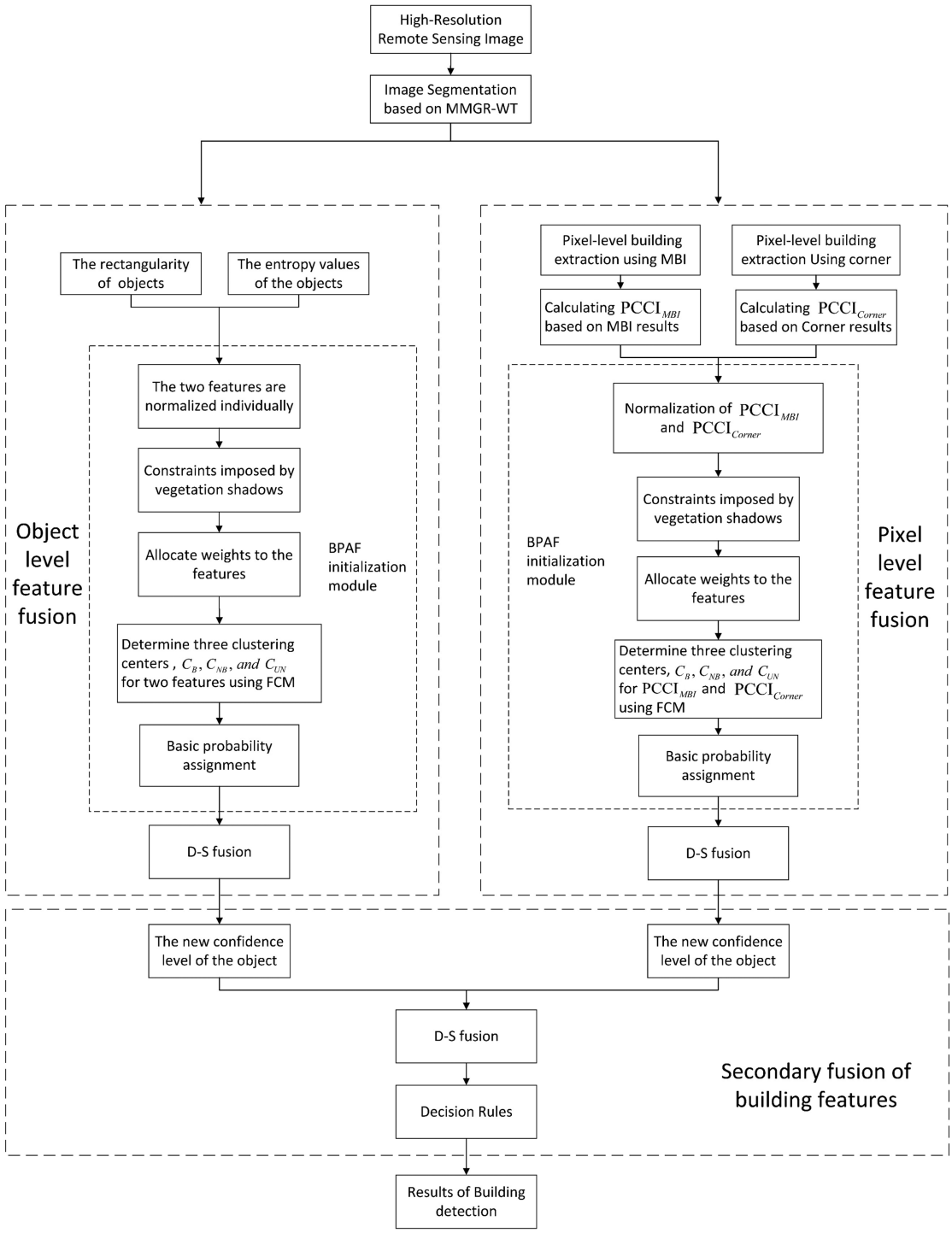

Fig. 1. Two-step D-S fusion framework for extracting buildings.

its minimum bounding rectangle, to quantify the likelihood of an object being a building. Equation (4) provides the definition of rectangularity

$$R^i = \mathrm{Nor}\left(\frac{{S_0}^i}{{S_{\mathrm{MER}}}^i}\right) \tag{4}$$

where R^i represents the rectangularity of object i, while S_0 and S_{MER} denote the area of object i and the minimum area of its enclosing rectangle, respectively. The symbol "Nor" represents the process of normalization calculation.

The interior areas of a building are usually uniform and consistent, lacking noticeable texture or structural variations. Hence, this study utilizes object entropy values to quantify the extent of grayscale variation in the pixels within the building's interior regions. A lower entropy value of an object indicates a greater likelihood of it belonging to the building. The calculation of the entropy value of an object is performed as

follows:

$$H_0{}^i = -\sum p^i(x)\log_2(x) \tag{5}$$

where H_0 represents the entropy value of object i, x represents the grayscale value of the pixel, and $p^i(x)$ represents the probability of the pixel's grayscale value occurring. As the entropy value decreases, the possibility of the object being a building increases. To facilitate assigning credibility to each piece of evidence, the entropy value is first normalized. Then, one minus the entropy value is used to represent the probability of it being a building. The normalized entropy value is denoted as H^i

$$H^i = 1 - \mathrm{Nor}(H_0{}^i). \tag{6}$$

2) BPAF Initialization Module Based on Feature Constraints and Fuzzy Clustering: To address the difficulty in initializing BPAF, we propose a BPAF initialization method based on feature constraints and fuzzy clustering. By utilizing FCM algorithm to construct membership functions, the need for parameter adjustment is eliminated. In addition, before assigning the initial probability to different propositions, we consider the constraints imposed by shadows and vegetation on the derived building features. This avoids the impact of shadows and vegetation on the extraction results.

Shadows are frequently observed in HRRS images and exhibit lower brightness values. Considering the human eye's sensitivity to different wavelengths, the method described in [49] is used to calculate the brightness of the image, which serves as the first feature of shadows (7). In addition, normalized color space does not have an impact on either shadow areas or high-brightness areas. Thus, the difference between the original color space and the normalized color space in the image can be used to define the shadow feature [50], which functions as the secondary characteristic of shadows (8). Ultimately, shadow regions are extracted by employing comprehensive decision-making

$$\mathrm{SF1} = 0.46R + 0.5G + 0.04B \tag{7}$$

$$\mathrm{SF2} = (|r - R| + |g - G|)/2 \tag{8}$$

$$\text{shadow} = \begin{cases} 1, & \text{if}(\mathrm{SF2} - \mathrm{SF1}) > t \\ 0, & \text{other} \end{cases}. \tag{9}$$

R, G, and B are used to represent the respective red, green, and blue channels of RGB images. The variables r and g represent the normalized values of the red and green channels. The variable t is the threshold value, which is adaptively determined using the OTSU method in this article.

Vegetation demonstrates a pronounced reflectance in the green light spectrum, while its reflectance is comparatively low in the red and blue light spectra. This characteristic has led to the proposal of various vegetation indices by numerous scholars, including the excess red index, excess green index, and excess green minus red difference index (ExGR) [51], [52]. This study employs the frequently utilized ExGR index from the color indices for vegetation extraction

$$\mathrm{ExGR} = (2G - B - R) - (1.4r - g) \tag{10}$$

$$\text{Vegetation} = \begin{cases} 1, & \text{if}(\mathrm{ExGR} > 0) \\ 0, & \text{other} \end{cases} \tag{11}$$

where r and g represent the normalized values for red and green channels respectively.

According to (7)–(11), the binary images obtained for shadows and vegetation can be obtained. Next, calculate the ratio of the number of shadow and vegetation pixels in the segmented object to the total number of pixels in that object. Then, a multiplication constraint term is obtained by subtracting this proportion from 1, which is used to process R^i and H^i. The approach of constraining building features using a multiplication term demonstrates enhanced fault tolerance compared to the methods that directly eliminate shadow and vegetation regions. The calculation of the feature constraint is as follows:

$$R_1{}^i = R^i \times \left(1 - \frac{S^i \times V^i}{N^i \times N^i}\right) \tag{12}$$

$$H_1{}^i = H^i \times \left(1 - \frac{S^i \times V^i}{N^i \times N^i}\right). \tag{13}$$

N^i represents the total number of pixels in object i, while S^i and V^i represent the number of shadow pixels and vegetation pixels in object i. R_1 and H_1 represent the probability of object i being a building in two different sets of evidence. The R_1 and H_1 values calculated from the two different features have differences in their value distributions. To ensure equal contributions in the subsequent D-S fusion of the two features, it is necessary to assign weight to each of them. The updated probabilities R_2 and H_2 for objects belonging to buildings are defined as follows:

$$R_2{}^i = R_1{}^i \times \sum_i^K \frac{H_1^i}{R_1^i + H_1^i}$$

$$H_2{}^i = H_1{}^i \times \sum_i^K \frac{R_1^i}{R_1^i + H_1^i}. \tag{14}$$

Among them, K represents the total number of segmented objects in the image.

Let Θ represent the recognition framework in the context of D-S theory. 2^Θ represents the entire set of hypothetical events on the framework. Consider the function $m{:}2^\Theta \to [0,1]$ as the fundamental probability allocation function. Thus, the allocation function m satisfies

$$m(\emptyset) = 0$$

$$\sum_{A \in 2^\Theta} m(A) = 1 \tag{15}$$

where A is a focal element, denoted as a nonempty subset of 2^Θ, where $m(A)$ signifies its fundamental probability distribution.

In this study, the recognition framework Θ: {*B, NB, UN*} is defined, which represents the three propositions in this article. For any object, *B*, *NB*, and *UN* represent buildings, nonbuildings, and uncertain situations respectively. The nonempty subsets of 2^Θ are {*B*}, {*NB*}, {*UN*}. According to (14), we obtain the probabilities R_2 and H_2 of the object belonging to the building. For any object, if R_2 or H_2 value is higher, it is more likely to be classified as a building; otherwise, the possibility of it being a building is smaller. Naturally, we divide the values of R_2 or H_2 into three categories: larger values represent buildings, smaller

values represent nonbuildings, and moderate values represent uncertain situations. We use the FCM algorithm to determine the three cluster centers C_B, C_{UN}, and C_{NB} for R_2 and H_2, where each center corresponds to a building, uncertain situations, and a nonbuilding respectively. Simultaneously, the membership degrees U_B, U_{UN}, and U_{NB} of object i to a specific class can be obtained. For all objects, BPAF is defined as shown in the following equation:

$$\begin{aligned} m_n^i(\{B\}) &= U_B^i \\ m_n^i(\{UN\}) &= U_{UN}^i \\ m_n^i(\{NB\}) &= U_{NB}^i \end{aligned} \tag{16}$$

where n represents two distinct features: either that of a rectangle or entropy.

3) Features Fusion Based on D-S Theory: The D-S theory makes decisions by combining multiple pieces of evidence. Equation (16) can be used to calculate the initial basic probability allocation values for each proposition that the object belongs to. According to (17), we perform the first-level fusion on $m_n(\{B\})$, $m_n(\{UN\})$, and $m_n(\{NB\})$ to obtain the new probability values $m_1(\{B\})$, $m_1(\{UN\})$, and $m_1(\{NB\})$ for object i belonging to different propositions.

For recognizing $\forall$ AϵD on the frame Θ, the synthesis rule for n m-functions (mass functions) is defined as follows:

$$\begin{aligned} m(A) &= m_1 \oplus m_2 \oplus m3 \oplus \cdots \oplus m_n \\ &= \frac{1}{K} \sum_{A_1 \cap A_2 \cap \cdots \cap A_n = A} m_1(A_1) m_2(A_2) m_3(A_3) \cdots m_n(A_n). \end{aligned} \tag{17}$$

The variable K represents the degree of conflict between evidence. A higher K value indicates a stronger conflict among the evidence. K is defined as

$$K = \sum_{A_1 \cap A_2 \cap \cdots \cap A_n = \emptyset} m_1(A_1) m_2(A_2) m_3(A_3) \cdots m_n(A_n). \tag{18}$$

C. Pixel-Level Feature-Based D-S Fusion

Most buildings exhibit brighter values compared to the surrounding objects and possess higher local contrast with the environment in remote sensing images. Simultaneously, the presence of numerous corners surrounding the building renders local geometric features a potent criterion for extracting buildings. We utilize methods that are based on MBI [39], [53], and corner [27] to extract building features at the pixel level. The extraction techniques for pixel-level building features are relatively simple and efficient. However, it is widely acknowledged that pixel-based building extraction methods often encounter significant amounts of salt-and-pepper noise as well as holes, as illustrated in Fig. 2(b) and (c). Fortunately, object-based building extraction methods can overcome these challenges by aggregating pixels into homogeneous objects, leveraging contextual pixel information. Nonetheless, object-based algorithms tend to be more intricate, resulting in a longer extraction process. To tackle the above challenges, we propose a PCCI index as a bridge between

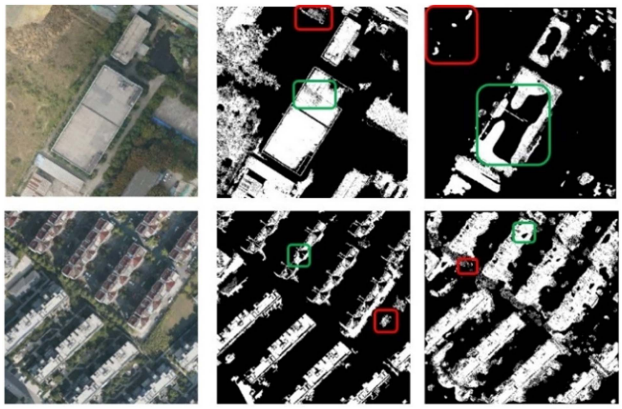

Fig. 2. Results of building detection based on pixel-level methods. The green box shows the phenomenon of holes, whereas the red box shows the noise of salt and pepper. (a) Original image. (b) MBI. (c) Corner.

building pixels and segmented objects, effectively integrating the advantages of both pixel-level and object-level building extraction methods.

1) Extraction of Pixel-Level Features: The fundamental idea of MBI is to utilize the local high contrast between buildings and the surrounding environment, as well as the structural characteristics of the buildings, to generate an index map of the buildings using morphological methods. After generating the building index map using MBI, we apply the OTSU algorithm adaptively to determine the segmentation threshold. We identify pixels with values greater than the threshold as buildings, while those with values less than or equal to the threshold are considered background. As a result, a binary image is generated. To evaluate the effectiveness of the OTSU algorithm, we conducted a comprehensive evaluation of various threshold segmentation methods, and their findings are detailed in Section V.

A method based on corner points and a variable kernel density function was proposed in [27] to estimate the position of buildings. The proposed method comprises two primary components: corner point extraction and estimation of the variable kernel density function. First, corner point information is extracted using the Harris corner detector [54], GMSR detector [55], Gabor filtering detector [56], and FAST detector [57]. The four different results of corner point extraction are represented by k_h, k_g, k_f, and k_s. Then, the four sets of corner point results are fused together to obtain $k = \{k_h, k_g, k_f, k_s\}$. This k is used as the observed value for the variable kernel density function to estimate it. In our study, adopting the parameter settings as in [27], the corner detection process is carried out. The building probability index map is obtained using a variable kernel density function, and a binary building image is acquired through the utilization of the OTSU method. Utilizing the corner points as a basis, the results of building extraction are achieved.

2) Calculation of the PCCI Index: The construction of this index is based on the following principle: the more building pixels there are in an object, and the more uniformly they are distributed within the object's internal boundary or the entire object, the higher the probability that the object represents a

building. Thus, PCCI is calculated by measuring the proportion of building pixels within an object (proportion consistency) and the distance between the centroid of building pixels and the centroid of the object (centroid consistency). The detailed steps for calculating PCCI are outlined below.

a) Calculate the proportion consistency index: Through MBI and corner points, we get the binary image of the building generated by two different methods. By combining the segmented images, we calculated the ratio of building pixels in object i to the total number of pixels within object i. This ratio is defined as the proportion consistency index

$$P_n^i = \frac{B_n^i}{N^i} \tag{19}$$

where B_n represents the total number of pixels of buildings in object i based on evidence n. N^i represents the total number of pixels in object i. A higher value of P_n in this equation implies a greater proportion of building pixels within the object, indicating a higher likelihood that the object is a building. In an ideal scenario of image segmentation and pixel-level building extraction, the number of building pixels within the object is equal to the total number of pixels in the object.

b) Calculate the proportion consistency index: The centroid consistency index is defined as the Euclidean distance between the centroid of object i and the centroid of building pixels in object i, as shown in (23)

$$C_n^i = \sqrt{(x^i - x_n^i)^2 + (y^i - y_n^i)^2} \tag{20}$$

where x and y represent the centroid coordinates of object i, whereas x_n and y_n represent the centroid coordinates of building pixels obtained based on evidence n within object i. The equation assesses the uniform distribution of building pixels within the internal edge region or the entire object. A lower value of C_n indicates a more uniformly distributed building pixels within the internal edge or the entire object, implying a higher likelihood of the object being a building. In an ideal situation, the centroid coordinates of the building pixels and the object would coincide.

c) Calculate the PCCI: To ensure that PCCI has a larger value only when the number of building pixels in the object is large and the distribution is relatively uniform, the definition of PCCI is shown in (21). Transforming the centroid consistency index into the multiplication of e^{-Cn} with the proportion consistency index is to facilitate the use of the PCCI index to indicate the probability of the object being a building, so that when P_n increases or e^{-Cn} increases, the probability of the object being a building also increases. Therefore, only when the values of P_n and e^{-Cn} are relatively large, that is, when PCCI has a larger value, the probability of the object belonging to a building will be relatively high

$$\mathrm{PCCI} = P_n^i \times e^{-C_n^i}. \tag{21}$$

The PCCI index can be calculated for any object using two different methods, resulting in $\mathrm{PCCI}_{\mathrm{MBI}}$ and $\mathrm{PCCI}_{\mathrm{Corner}}$. Normalization is applied to $\mathrm{PCCI}_{\mathrm{MBI}}$ and $\mathrm{PCCI}_{\mathrm{Corner}}$ to ensure comparability among different features. By employing the BPAF initialization method described in section of *Initialization for BPAF Based on Feature Constraints and Fuzzy Clustering*, the initial probabilities $m_n(\{B\})$, $m_n(\{UN\})$, and $m_n(\{NB\})$ are obtained. Subsequently, next, we perform a D-S fusion, resulting in new belief values $m_2(\{B\})$, $m_2(\{UN\})$, and $m_2(\{NB\})$ for object i.

D. Secondary D-S Fusion of Building Information

New degrees of Confidence in object i are obtained through pixel-level features fusion and object-level features fusion. The D-S fusion is applied once again to different degrees of belief, resulting in the secondary fused BPAF denoted as $m(\{B\})$, $m(\{UN\})$, and $m(\{NB\})$. Finally, (22) is utilized to achieve the ultimate outcome of building extraction

$$\text{Buildings} = \begin{cases} 1, & m^i(\{B\}) > m^i(\{UN\})\text{AND}m^i(\{B\}) \\ & \quad > m^i(\{NB\}) \\ 0, & \text{others} \end{cases}. \tag{22}$$

IV. Experiments and Analysis

A. Datasets and Evaluation Metrics

1) Datasets: In this study, we performed experiments using three separate datasets that encompassed various scenarios. The datasets utilized in our study are listed below.

The first dataset is aerial drone data from Jiangbei New Area, Nanjing, China, in October 2019. It consists of digital orthophotos with dimensions of 27337×21816 pixels, composed of three bands (RGB) with a spatial resolution of 0.3 m. The dataset covers an area of 53.67 km^2. To validate the efficacy of the building extraction methods, we constructed a reference dataset for buildings using visual interpretation and manual delineation methods. Due to computational performance and efficiency considerations, the original aerial images were resized, resulting in smaller images measuring 512×512 pixels. Not all cropped images contain buildings or have a very small proportion of area occupied by buildings. Therefore, only images with buildings and a building area larger than 2% were retained in the dataset. The Jiangbei dataset comprises 349 images of dimensions 512×512 pixels in addition to their reference data.

The second dataset was selected from the aerial imagery dataset of the WHU building dataset [58]. The image spatial resolution was downsampled from 0.075 to 0.3 m and cropped to a size of 512×512 pixels. This dataset comprises three subsets: a training set, a validation set, and a test set.

Considering the dataset's extensive size, the validation set with the least amount of data was chosen as the second dataset in this study. Similarly, only the images containing buildings and where the building area represents more than 2% of the cropped image were retained. The final dataset comprises 565 images and their respective label images.

The third dataset is an aerial image object recognition dataset obtained from the Washington area [59]. This dataset consists of two aerial orthoimages with a spatial resolution of 0.16 m. The covered area is predominantly urban, leading to a relatively dense configuration of buildings. Due to the large size of both images (5000×5000 pixels), we have cropped them into images

TABLE I
MBI PARAMETER SETTINGS ON DIFFERENT DATASETS

		Nanjing	WHU	Washington
Structure size	Minimum size	12	12	12
	Maximum size	292	292	452
	Interval size	70	70	110
Number of structural Directions		4	4	4

of size 500 × 500. We obtained a final set of 174 images and their corresponding label images by discarding those where the area occupied by buildings is less than 2% of the cropped image.

2) Evaluation Metrics: To quantitatively evaluate the final extraction results of buildings, five commonly used quantitative indicators were employed: overall accuracy (OA), precision (P), recall (R), and the harmonic mean of precision and recall (F1)

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{23}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{24}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{25}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{26}$$

where TP, FP, and FN represent true positive, false positive, and false negative, respectively. In this study, we assign equal importance to precision and recall for buildings. Thus, we evaluate our results using their harmonic mean F_1.

B. Experimental Results

1) Validation of the Validity of the PCCI: In the first step of the D-S fusion, we used the MBI method and the corner detection strategy to obtain building index maps, respectively. The parameter configurations for MBI across various datasets are presented in Table I. Then, we used OTSU for threshold segmentation to generate pixel-based binary images for building extraction. The PCCI of each object was calculated using the segmentation result of MMGR-WT. A higher PCCI of the object indicates a higher probability of being a building. Based on these, the building index maps were generated using an object-oriented approach. We used MBI-PCCI and Corner-PCCI to represent the combination of two different techniques and the PCCI. To assess the feasibility of the proposed PCCI, we employed a threshold range of [0,1] with a step size of 0.01 to partition the object-oriented building index maps. The highest F1 score was chosen as the resultant object-oriented building extraction. To ensure a fair comparison, we applied shadow and vegetation constraints to MBI and MBI-PCCI. Fig. 3 shows the building results extracted using PCCI indexes based on different methods. As we can see, the extraction results of MBI-PCCI and Corner-PCCI [see Fig. 3(d) and (e)] demonstrate a more comprehensive building extraction or fewer instances of holes, as compared to the results of MBI and corner extraction [see Fig. 3(b) and (c)]. Furthermore, these methods effectively minimize the impact of salt and pepper noise, leading to a substantial enhancement in both recall and precision rates. Accordingly, the PCCI index can function as a bridge connecting pixel-level building extraction methods with object segmentation, facilitating the shift from pixel-level to object-level processing and analysis. This index effectively combines the strengths of both pixel-level and object-level methods by integrating their respective advantages.

TABLE II
BUILDING EXTRACTION RESULTS USING PCCI AND NOT USING PCCI IN THE NANJING DATASET

Method/Indicator	OA	P	R	F1
MBI [39]	0.7852	0.3405	0.5066	0.4072
MBI-PCCI	0.8369	0.4480	0.5154	0.4794
Corner [27]	0.8124	0.3726	0.4210	0.3953
Corner-PCCI	0.8549	0.5020	0.4976	0.4998

TABLE III
BUILDING EXTRACTION RESULTS USING PCCI AND NOT USING PCCI IN THE WHU DATASET

Method/Indicator	OA	P	R	F1
MBI [39]	0.7824	0.4612	0.5536	0.5032
MBI-PCCI	0.8041	0.5076	0.5493	0.5276
Corner [27]	0.6609	0.3213	0.6322	0.4260
Corner-PCCI	0.7150	0.3849	0.7213	0.5020

This study quantitatively evaluated the results of building extraction combined with the PCCI based on three datasets, as shown in Tables II–IV. The inclusion of the PCCI index in the method considerably enhanced the accuracy of building extraction in comparison to the method omitting the PCCI index. For three datasets, the F1 scores of building extraction using the MBI-PCCI method exhibited an increase of 7.22%, 2.44%, and 3.57% respectively, in contrast to the original MBI

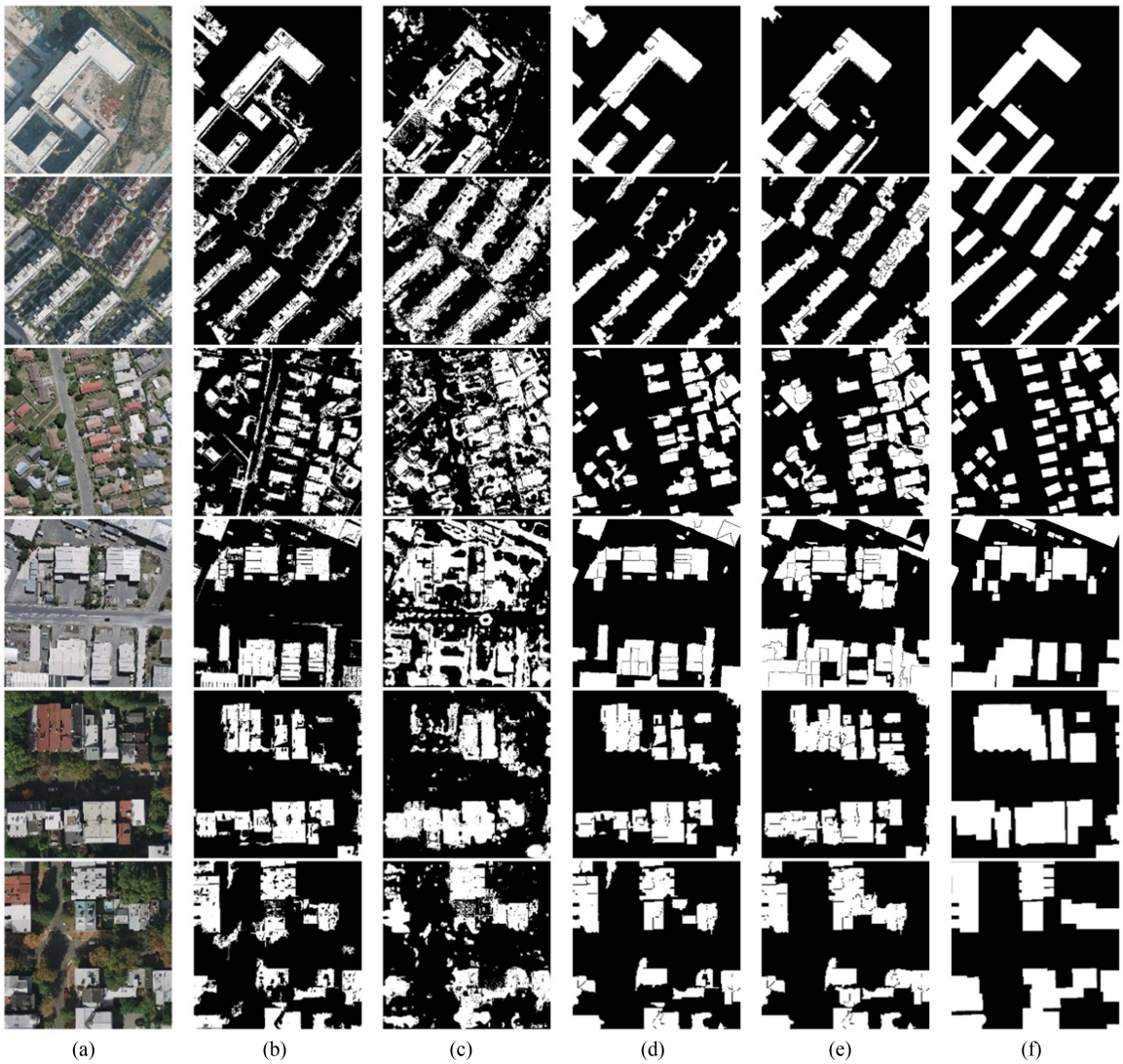

Fig. 3. Results of building extraction from three datasets, from top to bottom, are Nanjing, WHU, and Washington, with two images taken from each dataset. (a) Original image. (b) MBI method. (c) Corner method. (d) MBI-PCCI method. (e) Corner-PCCI method. (f) Reference map.

TABLE IV
BUILDING EXTRACTION RESULTS USING PCCI AND NOT USING PCCI IN THE WASHINGTON DATASET

Method/Indicator	OA	P	R	F1
MBI [39]	0.7316	0.6764	0.5166	0.5858
MBI-PCCI	0.7463	0.6875	0.5671	0.6215
Corner [27]	0.7137	0.6319	0.5281	0.5754
Corner-PCCI	0.7337	0.6225	0.6994	0.6586

method. The F1 scores of building extraction with the Corner-PCCI method showed even more significant improvements, with increases of 10.45%, 7.6%, and 8.32% in the three datasets respectively. Likewise, both the MBI-PCCI and Corner-PCCI methods displayed a prominent enhancement in the OA of building extraction results when compared to the original MBI and corner methods. These results align with the qualitative results, demonstrating that the integration of pixel-level methods with the PCCI significantly enhances the detection accuracy of buildings, thus providing compelling evidence of the efficacy of the PCCI.

The comparison of the extraction results between MBI-PCCI and Corner-PCCI reveals that the MBI-PCCI-based method achieves a higher precision rate, suggesting a reduced misclassification error. Nevertheless, the recall rate of the MBI-PCCI-based method is markedly lower than that of the corner-based method. In essence, the MBI-PCCI-based method exhibits reduced misclassification error, while the Corner-PCCI-based method demonstrates lower missed classification error. These findings effectively support the viability of integrating these two methods through D-S fusion and hierarchical fusion.

2) Results and Performance Evaluation: We conducted comprehensive and objective evaluations of our proposed method's overall performance by conducting comparative experiments using the following six advanced methods:

TABLE V
EVALUATION RESULTS IN NANJING DATASET

Method/Indicator	OA	P	R	F1
MBI [39]-PCCI	0.8369	0.4480	0.5154	0.4794
Corner [27]-PCCI	0.8549	0.5020	0.4976	0.4998
GBI [43]	0.8310	0.4400	0.5873	0.5031
Modified DeeplabV3+ [60]	0.8611	0.5440	0.2500	0.3233
CBRNet [61]	0.8373	0.4400	0.42910	0.4345
Major voting	0.8775	0.5970	0.4883	0.5372
Proposed method	**0.8795**	0.5931	0.5499	**0.5707**

Bold numbers represent the best results among different methods.

1) the object-oriented method combining MBI [39] and PCCI (MBI-PCCI),
2) the object-oriented method combining corner detection [27] and PCCI (Corner-PCCI),
3) the method based on the GBI [43], and
4) the majority voting method using rectangularity, entropy calculation, MBI-PCCI, and corner-PCCI (Major voting). The above-mentioned methods do not require any training data. We also compare our approach with deep learning-based approaches, including
5) Modified DeeplabV3+ [60], and
6) CBRNet [61].

To illustrate the rationality for utilizing the D-S theory, we introduce Method 4, which integrates four different techniques for building extraction and employs the majority voting method. Since methods 1, 2, and 4 ultimately yielded probability index maps of buildings, we performed a threshold segmentation on these maps to obtain the final binary building images. Consequently, we compared the F1 scores at different threshold intervals (threshold range: [0, 1], with an interval of 0.01) and chose the highest F1 score as the result. Notably, the index map requires normalization before threshold segmentation. By comparing with different methods, we can better evaluate the proposed method.

Fig. 4 displays the extraction results obtained from various methods. Correctly identified building pixels are highlighted in green, while undetected buildings are marked in blue, and erroneously detected buildings are represented in red. It depicts representative images from three datasets, with Image Fig. 4(a) representing the original image, and images Fig. 4(b)–(h) corresponding to MBI-PCCI, corner-PCCI, GBI, modified deeplabV3+, CBRNet, major voting, and the proposed method, respectively. Obviously, the method proposed demonstrates superior accuracy in detection results, with a lower incidence of false positives and false negatives compared to other methods. The MBI-PCCI method mistakenly identifies bright nonbuilding areas as buildings, while also missing some buildings that are relatively darker in comparison to the surrounding environment. Both the corner-PCCI and GBI methods exhibit a high number of false positives as they mistakenly identify areas with complex textures, such as parking lots and roads, as buildings. The major voting method exhibits a notable rate of false negatives for buildings. Notably, the modified deeplabV3+ method and CBRNet method exhibit exceptional performance in terms of minimizing both false positives and false negatives on the WHU dataset (rows 4–6 of Fig. 4). This is attributed to the fact that both models were trained using the WHU dataset. However, two network models demonstrate poor performance in building extraction when applied to the other two datasets. without manually annotated data, the proposed method exhibits superior detection results with fewer misses and false positives compared to other methods, as visually observed.

Furthermore, a quantitative evaluation of different methods on various datasets was performed, and the evaluation results are presented in Tables V–VII. The proposed method achieved the highest F1 score on both the Nanjing dataset and the Washington dataset. On the Wuhan University dataset, our method only ranked below two deep learning-based methods. Nevertheless, our method obtained the highest F1 score among the methods that did not employ training data. Compared to the Nanjing dataset and WHU dataset, the F1 score of the proposed method has significantly improved on the third dataset. This improvement can be attributed to the higher spatial resolution of Washington dataset, resulting in larger buildings in the images. In addition, Washington dataset predominantly covers urban areas characterized by simpler scenes and more densely populated buildings, thus reducing the complexity of building extraction.

The OA indicator reflects the proportion of correctly classified samples to all samples. The F1 score is a comprehensive evaluation indicator of precision and recall, which can effectively balance the accuracy and completeness of building extraction. Therefore, when assessing the effectiveness of building extraction, we comprehensively consider both OA and F1 indices of the chosen method.

The F1 scores and OA of all comparison methods on three datasets are presented in Tables V–VII. The results show that our method's F1 scores on the Nanjing and Washington datasets are 3.35% and 0.98% higher than the second-highest method,

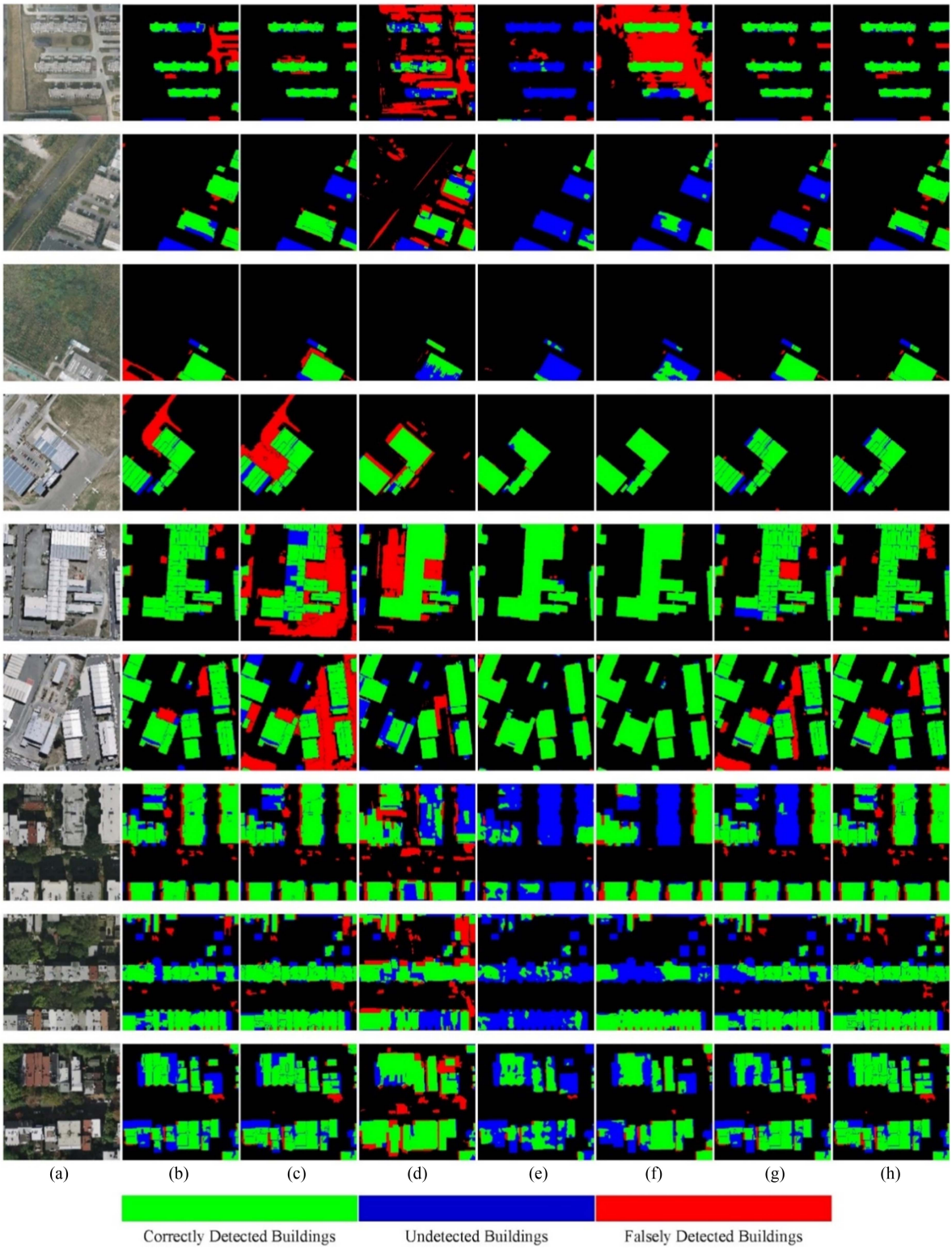

Fig. 4. Building detection results on different datasets using different methods, with 1–3 rows belonging to Nanjing dataset, 4–6 rows belonging to WHU dataset, and 7–9 rows belonging to Washington dataset. (a) original image. (b) MBI-PCCI. (c) Corner-PCCI. (d) GBI. (e) Modified DeeplabV3+. (f) CBRNet. (g) Major voting. (h) Proposed method.

TABLE VI
EVALUATION RESULTS IN WHU DATASET

Method/Indicator	OA	P	R	F1
MBI [39]-PCCI	0.8041	0.5076	0.54928	0.5276
Corner [27]-PCCI	0.7150	0.3829	0.7411	0.4833
GBI [43]	0.8187	0.5407	0.5939	0.5660
Modified DeeplabV3+ [60]	0.9570	0.8871	0.8714	0.8763
CBRNet [61]	**0.9839**	0.9615	0.9512	**0.9563**
Major voting	0.8320	0.5767	0.5870	0.5818
Proposed method	0.8366	0.5820	0.6356	0.6076

Bold numbers represent the best results among different methods.

TABLE VII
EVALUATION RESULTS IN WASHINGTON DATASET

Method/Indicator	OA	P	R	F1
MBI [39]-PCCI	0.7463	0.6880	0.5714	0.6074
Corner [27]-PCCI	0.7337	0.6253	0.6994	0.6459
GBI [43]	0.6882	0.5529	0.7900	0.6505
Modified DeeplabV3+ [60]	0.6971	0.8305	0.2247	0.3421
CBRNet [61]	0.7238	0.7045	0.4274	0.5320
Major voting	**0.7466**	0.7463	0.4702	0.5769
Proposed method	0.7382	0.6307	0.6928	**0.6603**

Bold numbers represent the best results among different methods.

respectively. On the Nanjing dataset, our method yielded the highest OA of 0.8795, and the Major voting method obtains the second-highest OA of 0.8775. Although the increase in OA is modest, just 0.2%, a closer observation of the F1 scores demonstrates that our method outperforms the Major voting method by 3.35%. On the Washington dataset, our method achieved the third-highest score with an OA value of 0.7382, which was 0.84% lower than the first-highest method (Major Voting). However, our method significantly enhances the F1 score by 8.34% in comparison to the Major voting method. We consider this tradeoff, sacrificing a slight improvement in OA for a substantial enhancement in F1 score, to be worthwhile. Both the CBRNet network model and the modified deeplabV3+ network model attained the top two positions in terms of F1 score on the WHU dataset. This success can be attributed to the fact that both models were trained directly on this specific dataset. Nevertheless, when two models were applied to the Nanjing dataset or the Washington dataset, they exhibited poor generalization capabilities. Specifically, the modified deeplabV3+ model and the CBRNet model recorded the lowest and second lowest F1 values, respectively, on both the Nanjing and the Washington datasets. Our method is optimal in the OA and F1 score on the Wuhan University dataset without training data.

Both the majority voting method and the method proposed in this article are based on MBI, corner points, rectangularity of objects, and entropy values of objects for the initial extraction of buildings. However, there is a distinction in the methodology employed in this article. It utilizes a multilevel D-S theory to integrate various approaches, thereby improving the modeling of uncertainty and facilitating the fusion of information from diverse evidence. The F1 scores of the proposed method consistently outperform those of the majority voting method across all three datasets, as illustrated in Tables V–VII. The F1 scores of the proposed method have exhibited respective improvements of 3.35%, 2.58%, and 8.34% compared to those of the majority voting method within each dataset. The difference in OA values between the two methods on three datasets is not significant. Without considering the experimental results of the modified deeplabV3+ model and CBRNet model, the majority voting method attains the second-highest F1 score in Nanjing dataset and WHU dataset but yields the poorest results in Washington dataset. This limitation emerges due to

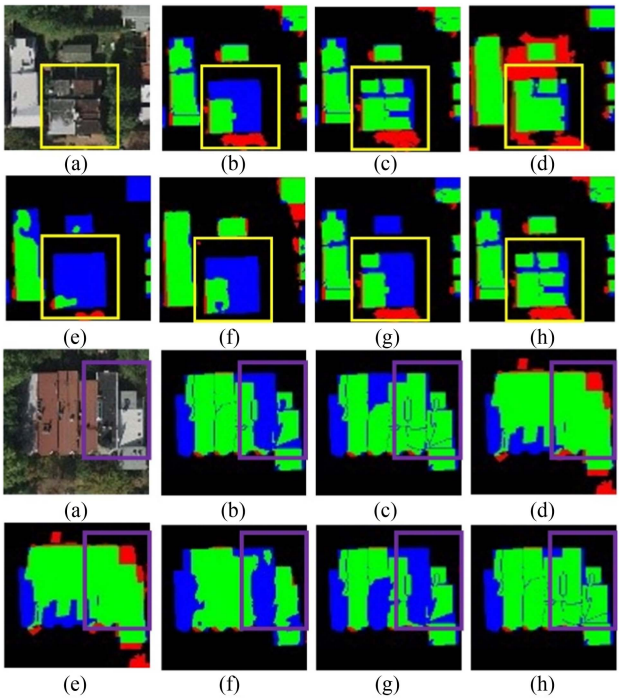

Fig. 5. Building detection results of patches. (a) Patch. (b)–(h) Results obtained in patch using methods 1–6 and the proposed method, respectively.

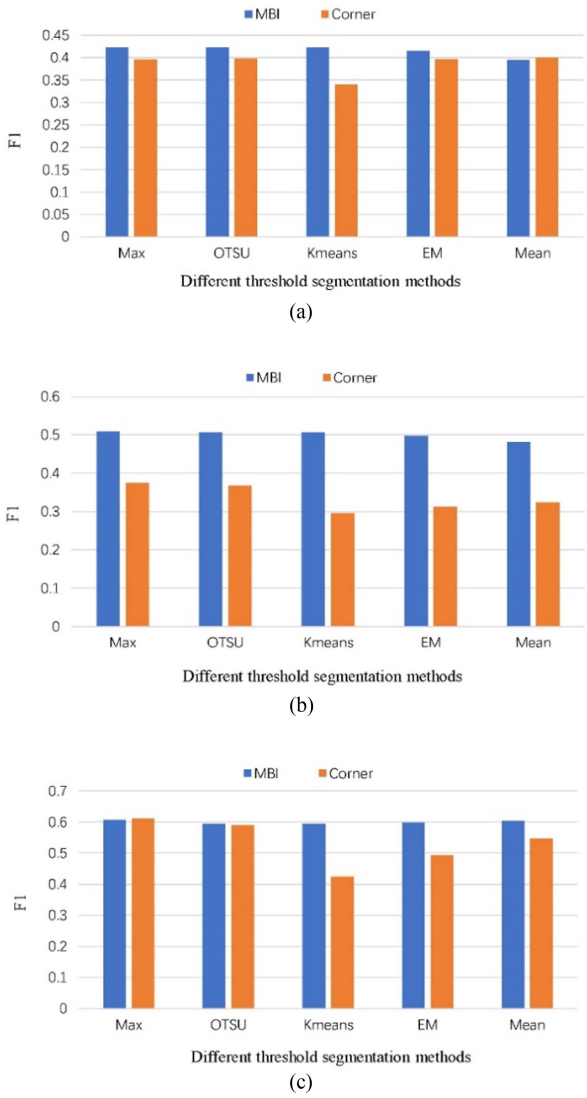

Fig. 6. F1 scores for different threshold methods on three datasets. (a)–(c) Nanjing dataset, WHU dataset, and Washington dataset, respectively.

the majority voting method's inability to effectively account for the uncertainty inherent in diverse evidence sources, hence leading to its overall instability. The observation suggests that the D-S theory approach outperforms major voting methods when it comes to integrating diverse evidence for building extraction.

The results of the representative patches are reported in Fig. 5. By observing the detection results of buildings within the yellow box, it can be found that methods 1, 4, 5, and 6 have caused missed detections of darker buildings. Methods 2, 3, and our method have significantly reduced missed detections, but method 3 may incorrectly identify some nonbuilding areas as buildings. Observing the results of the purple frame detection, it can be observed that in the same situation, methods 1, 5, and 6 cannot detect darker buildings. The buildings detected by methods 3 and 4 are relatively complete, but there are some false detections. Method 2 and the method proposed in this article have similar detection results in the purple frame, but in some areas, the missed detection situation of Method 2 is worse than that of the method proposed in this article.

Through visual analysis of representative patches, this method can detect different types of buildings and is insensitive to interference in most nonbuilding areas. This method is significantly superior to other comparative methods and consistent with the conclusions of quantitative analysis.

V. Discussion

A. Segmentation Threshold Selection

When extracting buildings through MBI-PCCI and Corner-PCCI, we need to perform threshold segmentation on the index maps generated by MBI and corners to generate binary images of buildings. Subsequently, the PCCI is calculated using the binary images and segmentation maps to perform object-oriented building extraction. The OTSU method is employed in this process to dynamically determine the segmentation threshold. To validate the appropriateness of utilizing the OTSU algorithm in this study, we conducted a comparative analysis and evaluation with other threshold evaluation methods, which encompassed the K-means clustering algorithm (K-means), expectation maximization algorithm, mean method (Means), and the threshold that yields the highest F1 score across the entire dataset (Max, obtained through traversal).

Fig. 6 shows the results of various thresholding methods across three datasets. It is important to note that the F1 score presented here is computed using binary images obtained from pixel-based MBI and corner-point detection, without undergoing post-processing. The figure clearly shows that among the four methods for automatically determining segmentation

TABLE VIII
ABLATION STUDY RESULTS ON THREE DATASETS BASED ON PIXEL-LEVEL BRANCH

Baseline				Nanjing datasets	WHU datasets	Washington datasets
MBI	Corner	PCCI	BPAF initialization module	OA	OA	OA
√				0.7750	0.7818	0.7318
√		√		0.8096	0.8074	0.7339
√		√	√	0.8628	0.8103	0.7407
	√			0.7907	0.5493	0.6798
	√	√		0.8686	0.7212	0.7335
	√	√	√	0.8768	0.7981	0.7354
√	√	√	√	**0.8795**	**0.8286**	**0.7558**

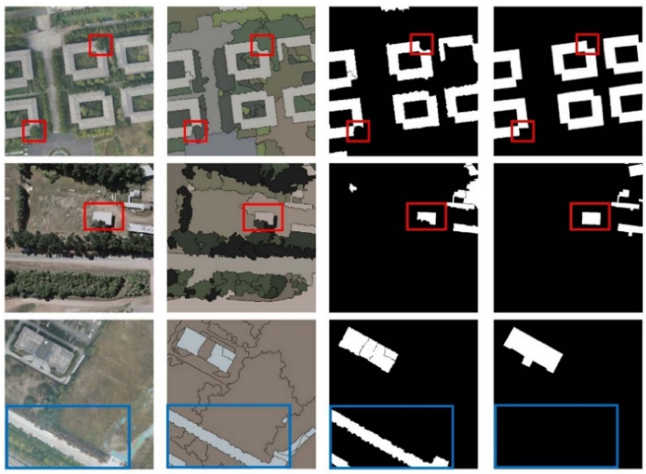

Fig. 7. Results of method proposed for building occluded areas and bright road areas. (a)–(d) Original image, segmentation image, results of our method, and real labels, respectively.

thresholds, the OTSU method consistently performs the best, closely trailing the Max method in most instances. Nonetheless, it is worth noting that in certain scenarios, the automatic thresholding methods can yield higher F1 scores than the Max method. This discrepancy is mainly attributed to the adaptive nature of the automatic thresholding methods, which determine thresholds for each image based on their distinct characteristics. In contrast, the Max method determines the threshold based on the entire dataset, potentially resulting in suboptimal segmentation thresholds for individual images. Therefore, using the Otsu method in this study to adaptively determine the threshold and generate binary images is reasonable.

B. Ablation Study

The framework proposed in this study comprises two branches, which are built upon pixel-level features and object-level features. In the pixel-level branch, we apply the proposed PCCI index and BPAF initialization module, while in the object-level branch, we solely utilize the BPAF initialization module. To better know the contribution of each component to the proposed framework, we conducted ablation experiments on different branches and three different datasets. It is worth mentioning that no post-processing methods were employed in any of the ablation experiments.

Table VIII shows the results of the ablation study conducted on the pixel-level features branch. We combined baseline models, MBI and Corner, with the PCCI and BPAF initialization modules to explore the contributions of each module. The OA of baseline MBI on three datasets is 0.7750, 0.7818, and 0.7318, respectively, while the OA of baseline Corner is 0.7907, 0.5493, and 0.6798. First, upon applying PCCI to both baselines, the "MBI+PCCI" yielded OAs of 0.8096, 0.8074, and 0.7339, while the "Corner+PCCI" achieved OAs of 0.8686, 0.7212, and 0.7335. These results confirm the effectiveness of PCCI in mitigating hole artifacts and salt-and-pepper noise, thus enhancing the precision of boundary delimitation. Second, the integration of the BPAF initialization module further improved building extraction performance. The addition of BPAF initialization module to "MBI+PCCI" resulted in OA increases of 5.32%, 0.29%, and 0.68%. Similarly, after integrating the BPAF initialization module on the baseline Corner, the OA improved by 0.82%, 7.69%, and 0.19% on the three datasets respectively. Finally, by integrating two baselines and two modules, the extraction performance of the building reached the optimal level, with OA improved to 0.8795, 0.8286, and 0.7558 on three datasets.

We evaluated the BPAF initialization module on the object-level branch, utilizing rectangularity and entropy as baselines. The experimental findings are outlined in Table IX. In addition to achieving the second-highest score on the Nanjing dataset, the method integrating rectangularity, entropy, and BPAF achieved the best results on the WHU and Washington datasets, with OA reaching 0.8034 and 0.7049, respectively. The results in Tables VIII and IX demonstrate the practicality and effectiveness of each component in the proposed method.

Without considering the secondary evidence fusion stage, we conducted a more detailed analysis of the extraction effects on buildings in the first layer through two distinct branches. Table X displays the extraction effects of the three datasets. As we have seen, the pixel-level branch demonstrates higher precision but

TABLE IX
ABLATION STUDY RESULTS ON THREE DATASETS BASED ON OBJECT-LEVEL BRANCH

Baseline		BPAF initialization module	Nanjing datasets	WHU datasets	Washington datasets
Rectangularity	Entropy		OA	OA	OA
√			0.7896	0.7212	0.6488
√		√	**0.8550**	0.8173	0.6872
	√		0.5980	0.5486	0.602
	√	√	0.8306	0.7882	0.6888
√	√	√	0.8365	**0.8034**	**0.7049**

TABLE X
BUILDING EXTRACTION RESULTS BY FUSING PIXEL LEVEL FEATURES AND OBJECT LEVEL FEATURES

Indicator/ Dataset	Nanjing dataset		WHU dataset		Washington dataset	
	object-level	pixel-level	object-level	pixel-level	object-level	pixel-level
OA	0.8365	0.8795	0.8034	0.8286	0.7049	0.7558
P	0.4525	0.6150	0.5046	0.5852	0.5772	0.7158
R	0.5809	0.4645	0.6993	0.4773	0.7351	0.5562
F1	0.5087	0.5292	0.5862	0.5258	0.6466	0.6259

lower recall. The object-level component compensates for this impact by increasing the recall rate.

Specifically, the pixel-level branch and object-level branch contribute to higher precision and recall rates in second-step evidence fusion. This proves that it is effective and reasonable to use the D-S theory to perform a hierarchical fusion of architectural features.

VI. CONCLUSION

Accurately and effectively extracting buildings from HRRS images using a single feature or method is challenging. To overcome this issue, our study proposes a fusion framework that integrates multiple building features. The framework fully utilizes the spectral and spatial geometric features of buildings and effectively combines the building information extracted from multiple features through a two-step decision-making process based on the D-S theory. The proposed approach better handles the uncertainties and correlations among multiple evidence and solves the problem of computational explosion when there is a lot of evidence, thus improving the quality and reliability of decisions. In the first-step fusion of pixel-level features, we propose a PCCI index that merges the pixel-level building extraction results with the segmentation of geographical objects. This index efficiently converts pixel-level building features into object-level features, thus facilitating subsequent object-based building analysis and ensuring greater consistency between the extracted building boundaries and human visual perception. To address the issue of initial probability allocation in D-S fusion, we construct a BPAF initialization module based on feature constraints and fuzzy clustering. The module diminishes the impact of nonbuilding objects on the extraction results, streamlines the construction process of BPAF, and eliminates the need for parameter adjustments. Experimental results on datasets from three different areas show that the proposed method outperforms the other four methods in both qualitative and quantitative evaluations.

The performance of the proposed method largely depends on the results of image segmentation. Images with undersegmentation often result in significant errors between the extracted boundaries of buildings and their actual boundaries. Segmentation algorithms encounter challenges in identifying obstructed regions of buildings that are concealed by shadows and vegetation, which can subsequently lead to the misclassification of neighboring objects. Consequently, the algorithm proposed in this article may result in false negatives in the occluded areas (see Fig. 7, rows 1–2). Model matching, corner assistance, and morphological methods [62] can be employed to rectify the problem of irregularly extracted building outlines due to occlusion. The algorithm in this manuscript mistakenly identifies some bright rectangular roads as buildings due to their similarity in spectra and geometric shapes (refer to row 3 in Fig. 7). Compared to roads, buildings have distinctive heights and exhibit shadow information around them. The impact of roads can be reduced by introducing height and shadow information. In the future, our primary focus will be on optimizing the boundaries after building extraction to enhance our current work.

REFERENCES

[1] Y. You et al., "Building detection from VHR remote sensing imagery based on the morphological building index," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1287.

[2] P. Weihao, X. Saibo, G. Hongyang, W. Yukun, W. Tao, and W. Chao, "Building change detection of high-resolution remote sensing images based on D-S evidence theory," *J. Electron. Meas. Instrum.*, vol. 36, no. 08, pp. 194–203, 2022.

[3] C. Wang, Y. Zhang, X. Chen, H. Jiang, M. Mukherjee, and S. Wang, "Automatic building detection from high-resolution remote sensing images based on joint optimization and decision fusion of morphological attribute profiles," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 357.

[4] J. Li, X. Huang, L. Tu, T. Zhang, and L. Wang, "A review of building detection from very high resolution optical remote sensing images," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 1199–1225, Dec. 2022.

[5] Y. Lin, B. Zhang, D. Wang, X. Chen, and J. Xu, "Hierarchical building extraction from high resolution remote sensing imagery based on multi-feature fusion," *J. Image Graph.*, vol. 22, no. 12, pp. 1798–1808, 2017.

[6] Y. Lin, B. Zhang, J. Xu, K. Hou, and X. Zhou, "Building extraction from high resolution remote sensing imagery with multi-feature and multi-scale," *Bull. Surveying Mapping*, vol. 12, pp. 53–57, 2017.

[7] G. Zhao, A. Chen, G. Lu, and W. Liu, "Data fusion algorithm based on fuzzy sets and D-S theory of evidence," *Tsinghua Sci. Technol.*, vol. 25, no. 1, pp. 12–19, Feb. 2020.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[10] L. C. Chen et al., "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[11] H. Zheng et al., "HFA-Net: High frequency attention siamese network for building change detection in VHR remote sensing images," *Pattern Recognit.*, vol. 129, 2022, Art. no. 108717.

[12] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6230–6239, doi: 10.1109/CVPR.2017.660.

[13] X. Wen et al., "Me-net: A multi-scale erosion network for crisp building edge detection from very high resolution remote sensing imagery," *Remote Sens.*, vol. 13, 2021, Art. no. 3826, doi: 10.3390/rs13193826.

[14] L. Ding, H. Tang, Y. Liu, Y. Shi, X. X. Zhu, and L. Bruzzone, "Adversarial shape learning for building extraction in VHR remote sensing images," *IEEE Trans. Image Process.*, vol. 31, pp. 678–690, 2022, doi: 10.1109/TIP.2021.3134455.

[15] J. Wang, W. Li, Y. Gao, M. Zhang, R. Tao, and Q. Du, "Hyperspectral and SAR image classification via multiscale interactive fusion network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10823–10837, Dec. 2023.

[16] Y. Pang, X. Zhao, L. Zhang, and H. Lu, "Multi-scale interactive network for salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9410–9419.

[17] W. Zhou, X. Fan, L. Yu, and J. Lei, "MISNet: Multiscale cross-layer interactive and similarity refinement network for scene parsing of aerial images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2025–2034, 2023.

[18] X. Xiao, X. Xiong, F. Meng, and Z. Chen, "Multi-scale feature interactive fusion network for RGBT tracking," *Sensing*, vol. 23, 2023, Art. no. 3410, doi: 10.3390/s23073410.

[19] J. Wang, W. Li, M. Zhang, R. Tao, and J. Chanussot, "Remote sensing scene classification via multi-stage self-guided separation network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5615312.

[20] K. Xu, H. Huang, Y. Li, and G. Shi, "Multilayer feature fusion network for scene classification in remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1894–1898, Nov. 2020.

[21] T. Partovi, R. Bahmanyar, T. Krauß, and P. Reinartz, "Building outline extraction using a heuristic approach based on generalization of line segments," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 3, pp. 933–947, Mar. 2017.

[22] J. Wang, X. Yang, X. Qin, X. Ye, and Q. Qin, "An efficient approach for automatic rectangular building extraction from very high resolution optical satellite imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 487–491, Mar. 2015.

[23] Z. Sun, H. Fang, M. Deng, A. Chen, P. Yue, and L. Di, "Regular shape similarity index: A novel index for accurate extraction of regular objects from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3737–3748, Jul. 2015.

[24] T. T. Ngo, C. Collet, and V. Mazet, "Automatic rectangular building detection from VHR aerial imagery using shadow and image segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 1483–1487.

[25] M. Cote and P. Saeedi, "Automatic rooftop extraction in nadir aerial imagery of suburban regions using corners and variational level set evolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 313–328, Jan. 2013.

[26] Y Chen et al., "Built-up area detection from high-resolution satellite images using multi-scale wavelet transform and local spatial statistics," *ISPRS Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 207–210, 2018.

[27] B. Sirmacek and C. Unsalan, "A probabilistic framework to detect buildings in aerial and satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 211–221, Jan. 2011.

[28] H. S. Munawar, R. Aggarwal, Z. Qadir, S. I. Khan, A. Z. Kouzani, and M. A. P. Mahmud, "A gabor filter-based protocol for automated image-based building detection," *Buildings*, vol. 11, no. 7, 2021, Art. no. 302.

[29] A. O. Ok, "Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts," *ISPRS J. Photogrammetry Remote Sens.*, vol. 86, pp. 21–40, 2013.

[30] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[31] Z. Hu, Q. Zou, and Q. Li, "Watershed superpixel," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 349–353.

[32] G. Liasis and S. Stavrou, "Building extraction in satellite images using active contours and colour features," *Int. J. Remote Sens.*, vol. 37, no. 5, pp. 1127–1153, 2016.

[33] N. Jiang, J. X. Zhang, H. T. Li, and X. G. Lin, "Semi-automatic building extraction from high resolution imagery based on segmentation," in *Proc. Int. Workshop Earth Observation Remote Sens. Appl.*, 2008, pp. 1–5.

[34] S. Fan, Y. Hu, and Z. Liu, "Research of information extraction of city building based on a new object-oriented method," *J. South China Norm. Univ., Nat. Sci. Ed.*, vol. 47, no. 6, pp. 91–97, 2015.

[35] D. Han, S. Yang, Q. Zhao, L. Han, Z. Yang, and S. Cui, "Object-oriented building information extraction from high resolution remote sensing satellite imagery," *J. Atmos. Environ. Opt.*, vol. 16, no. 4, pp. 358–364, 2021.

[36] M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A robust built-up area presence index by anisotropic rotation-invariant textural measure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 1, no. 3, pp. 180–192, Sep. 2008.

[37] P. Zhang, Z. Lv, and W. Shi, "Object-based spatial feature for classification of very high resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1572–1576, Nov. 2013.

[38] Z. Shao, Y. Tian, and X. Shen, "BASI: A new index to extract built-up areas from high-resolution remote sensing images by visual attention model," *Remote Sens. Lett.*, vol. 5, no. 4, pp. 305–314, Apr. 2014.

[39] X. Huang and L. Zhang, "Morphological building/shadow index for building extraction from high-resolution imagery over urban areas," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 161–172, Feb. 2012.

[40] W. Ma, Y. Wan, J. Li, S. Zhu, and M. Wang, "An automatic morphological attribute building extraction approach for satellite high spatial resolution imagery," *Remote Sens.*, vol. 11, no. 3, 2019, Art. no. 337.

[41] Q. Bi, K. Qin, H. Zhang, Y. Zhang, Z. Li, and K. Xu, "A multi-scale filtering building index for building extraction in very high-resolution satellite imagery," *Remote Sens.*, vol. 11, no. 5, 2019, Art. no. 482.

[42] G. Liu, G.-S. Xia, X. Huang, W. Yang, and L. Zhang, "A perception-inspired building index for automatic built-up area detection in high-resolution satellite images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 3132–3135.

[43] G.-S. Xia, J. Huang, N. Xue, Q. Lu, and X. Zhu, "GeoSay: A geometric saliency for extracting buildings in remote sensing images," *Comput. Vis. Image Understanding*, vol. 186, pp. 37–47, 2019.

[44] H. Luo, C. Liu, C. Wu, and X. Guo, "Urban change detection based on dempster–shafer theory for multitemporal very high-resolution imagery," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 980.

[45] Y. Han, A. Javed, S. Jung, and S. Liu, "Object-based change detection of very high resolution images by fusing pixel-based change detection results using weighted dempster–shafer theory," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 983.

[46] A. Javed, S. Jung, W. H. Lee, and Y. Han, "Object-based building change detection by fusing pixel-level change detection results generated from morphological building index," *Remote Sens.*, vol. 12, no. 18, 2020, Art. no. 2952.
[47] S. Zhang, Z. Sun, X. Wang, G. Chu, and S. Huang, "A method of building detection using DS evidence fusion theory and fuzzy set," *Remote Sens. Inf.*, vol. 35, no. 05, pp. 93–105, 2020.
[48] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng, and A. K. Nandi, "Superpixel-based fast fuzzy C-means clustering for color image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1753–1766, Sep. 2019.
[49] J. Liu, "Study on automatic recognition of high resolution optic remotely sensed images based on visual features," Ph.D. dissertation, Pattern Recognition and Intelligent System, Shanghai Jiao Tong Univ., Shanghai, China, 2011.
[50] L.-Q. Xu, J. L. Landabaso, and M. Pardàs, "Shadow removal with blob-based morphological reconstruction for error correction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, vol. 2, pp. ii/729–ii/732.
[51] Y. Wu, L. Zhao, H. Jiang, X. Guo, and F. Huang, "Image segmentation method for green crops using improved mean shift," *Trans. Chin. Soc. Agricultural Eng.*, vol. 30, no. 24, pp. 161–167, 2014.
[52] L. Wu, L. Xiong, and H. Peng, "Quantitative evaluation of in-field rapeseed image segmentation based on RGB vegetation indices," *J. Huazhong Agricultural Univ.*, vol. 38, no. 2, pp. 109–113, 2019.
[53] X. Huang, W. Yuan, J. Li, and L. Zhang, "A new building extraction postprocessing framework for high-spatial-resolution remote-sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 654–668, Feb. 2017.
[54] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 23.1–23.6.
[55] C. Uensalan, "Gradient-magnitude-based support regions in structural land use classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 546–550, Oct. 2006.
[56] V. Kyrki, J. K. Kamarainen, and H. Kalviainen, "Simple Gabor feature space for invariant object recognition," *Pattern Recognit. Lett.*, vol. 25, no. 3, pp. 311–318, Feb. 2004.
[57] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, Jan. 2010.
[58] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.
[59] K. Bradbury, "Washington, DC - aerial imagery object identification dataset for building and road detection, and building height estimation," figshare, Jul. 2016, doi: 10.6084/m9.figshare.3504320.v1.
[60] A. Jiwani, S. Ganguly, C. Ding, N. Zhou, and D. Chan, "A semantic segmentation network for urban-scale building footprint extraction using RGB satellite imagery," 2021, *arXiv:2104.01263*.
[61] H. Guo, B. Du, L. Zhang, and X. Su, "A coarse-to-fine boundary refinement network for building footprint extraction from remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 183, pp. 240–252, 2022.
[62] J. X. Chang, S. X. Wang, Y. W. Yang, and X. J. Gao, "Hierarchical optimization method of building contour in high-resolution remote sensing images," *Chin. J. Lasers*, vol. 47, no. 10, pp. 249–262, 2020.

Xuedong Zhang received the Bachelor's degree in surveying and mapping engineering from Shandong University of Technology, Zibo, China, in 2021. He is currently working toward the Master's degree in photogrammetry and remote sensing with Jiangsu Normal University, Xuzhou, China.

His research interest focuses on the high-resolution remote sensing.

Xing Li received the M.S. degree in cartography and geographic information engineering from the Shandong University of Science and Technology, Qingdao, China, in 2006, and the Ph.D. degree in physical geography from the East China Normal University, Shanghai, China, in 2010.

He is currently an Associate Professor with the School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou, China. His research interests include high-resolution image processing and computer vision in environmental remote sensing applications.

Jian Huang received the Master's degree in surveying and mapping engineering from the Wuhan University, Wuhan, China, in 2008.

He is a Senior Engineer with Jiangsu Province Surveying and Mapping Engineering Institute, Jiangsu, China. He is mainly engaged in the research and application of photogrammetry and remote sensing technology.

Erzhu Li received the M.S. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2014, and the Ph.D. degree in cartography and geographic information system from the Nanjing University, Nanjing, China, in 2017.

He is currently an Associate Professor with the School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou, China. His research interests include high-resolution image processing and computer vision in urban remote sensing applications.

Wei Liu received the M.S. and Ph.D. degrees in cartography and geographic information engineering from the China University of Mining and Technology, Xuzhou, China, in 2007 and 2010, respectively.

He is currently an Associate Professor with the School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou, China. His research interests include spatial data quality checking, high-resolution remote sensing image processing, and GIS development and applications.

Lianpeng Zhang received the M.S. degree in geodesy and surveying engineering from the Shandong University of Science and Technology, Taian, China, in 1989, and the Ph.D. degree in geodesy and surveying engineering from the Shandong University of Science and Technology, Qingdao, China, in 2003.

He is currently a Professor with the School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou, China. His research interests include high-resolution image processing and computer vision in urban remote sensing applications.