Accuracy Assessment Measures for Object Extraction from Remote Sensing Images

Liping Cai [1,2] 🆔 **, Wenzhong Shi [3,*], Zelang Miao [4,5] and Ming Hao [6]**

1   School of Geography and Tourism, Qufu Normal University, Rizhao 276826, China; cumtcailp@126.com
2   Key Laboratory of Coastal Zone Exploitation and Protection, Ministry of Land and Resource,
     Nanjing 210024, China
3   Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University,
     Hong Kong, China
4   School of Geosciences and Info-Physics, Central South University, Changsha 410012, China;
     zelang.miao@csu.edu.cn
5   Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring,
     Central South University, Changsha 410012, China
6   School of Environment Science and Spatial Informatics, China University of Mining and Technology,
     Xuzhou 221116, China; haomingcumt@gmail.com
*   Correspondence: lswzshi@polyu.edu.hk

Abstract: Object extraction from remote sensing images is critical for a wide range of applications, and object-oriented accuracy assessment plays a vital role in guaranteeing its quality. To evaluate object extraction accuracy, this paper presents several novel accuracy measures that differ from the norm. First, area-based and object number-based accuracy assessment measures are given based on a confusion matrix. Second, different accuracy assessment measures are provided by combining the similarities of multiple features. Third, to improve the reliability of the object extraction accuracy assessment results, two accuracy assessment measures based on object detail differences are designed. In contrast to existing measures, the presented method synergizes the feature similarity and distance difference, which considerably improves the reliability of object extraction evaluation. Encouraging results on two QuickBird images indicate the potential for further use of the presented algorithm.

Keywords: object-based image analysis; accuracy assessment; feature similarity; distance difference

1. Introduction

High spatial resolution satellite images are easily available thanks to advancements in modern sensor technology and have led to many applications in various fields, such as agriculture, forestry, and environmental protection [1–3]. Compared to medium/low resolution satellite images, high resolution satellite images contain richer information and clearer boundaries, making them attractive for object extraction [4–6]. The concept of the object, a group of pixels that share similar properties, was originally proposed in the 1970s [7], triggering a considerable amount of research in object-based image analysis (OBIA). Since its introduction, researchers have wondered how they may assess the results of OBIA.

Noise is inherent in satellite images, and thus, the accuracy of object extraction needs to be examined. This issue has received considerable critical attention [8–17]. Examples include the error matrix and confusion matrix, which are two typical methods for accuracy assessment. Despite their popularity, these methods ignore object features, making them unsuitable for OBIA accuracy evaluation. A direct solution [18] is to compute error and confusion matrixes on each object rather than at the pixel level. Although this simple solution can amend the shortage of pixel-level error and confusion matrixes

to a certain extent, it suffers from the issue of missing the object detail, leading to unreliable evaluation results. To enhance the reliability of this accuracy evaluation, a series of per-object accuracy assessment measures based on object features have been designed [8,19]. Among them, the similarity between evaluation and reference objects (e.g., area, size, shape, and location [19–24]) is commonly used. As the similarity measure can judge the correctness of extracted objects, it is possible to obtain numbers of correct, wrong, and missing objects for accuracy evaluation. Statistical values of object similarity (e.g., the weighted average value) are also commonly used measures to directly assess object extraction accuracy [8]. This method exploits the degree of overlap and position difference between evaluation and reference objects for evaluating object extraction accuracy [19,25–28]. Although these methods have resulted in significant improvements compared to pixel-wise accuracy measures, the effects of geometric information and difference in detail have not been comprehensively examined. Meanwhile, most existing evaluation methods were designed based on geometrical errors associated with objects. The influence of thematic error on object extraction accuracy evaluation, however, lacks a systematic understanding [29,30]. Thus, there is still significant room to improve the generalization ability of existing object-level accuracy measures.

This study investigates the use of four designed accuracy measures for object extraction evaluation. The proposed method systematically studies the influence of object characteristics on extraction accuracy, the aim being to present reliable accuracy measures for object extraction. The remainder of this paper is organized as follows. Section 2 introduces the methodology. The experimental results and discussion are given in Sections 3 and 4, respectively. Finally, Section 5 concludes the paper.

2. Methodology

This study supposes that remote sensing images have been pre-processed, and have undergone radiometric calibration and geometric correction. The objects are extracted from pre-processed images, and some of the objects are selected as evaluation samples. A unique reference object is matched for each evaluation object. The accuracy evaluation of object extraction is based on object matching.

2.1. Object Matching

Object extraction accuracy is evaluated by comparing the difference between the evaluated object and its reference data, and thus it is fundamental to match the reference and evaluated objects. To this end, this paper matches objects using the maximum overlap area algorithm due to its computation efficiency. The central idea of the maximum overlap area method is to compute the coincidence degree O_{ij} between two objects.

$$O_{ij} = \frac{1}{2} \left(\frac{A_{C,i} \cap A_{R,j}}{A_{C,i}} + \frac{A_{C,i} \cap A_{R,j}}{A_{R,j}} \right) \tag{1}$$

where $A_{C,i}$ denotes the area of the ith evaluated object, $A_{R,j}$ is the area of the jth reference object, and $A_{C,i} \cap A_{R,j}$ represents the intersection area. For an evaluated object and candidate reference objects, each coincidence degree will be computed. Two objects will be judged as being a matching pair if their coincidence degree is a maximum amongst all pairs.

2.2. Area-Based Accuracy Measures

Three area-based accuracy measures (i.e., correctness, completeness, and quality) are designed for OBIA evaluation. The purpose of area-based accuracy measures is to obtain stable accuracy measurements.

Correctness P_{AC} is defined as the ratio of correctly extracted area and the whole extracted area.

$$P_{AC} = \frac{A_C}{A_{DC}} \tag{2}$$

where A_{DC} is the area of the extracted object, and A_C is the correct part of A_{DC}. The range of correctness is from 0 to 1. If all the evaluated objects have their own fully corresponding reference

objects, then $P_{AC} = 1$. If there is no evaluated object from the same thematic class overlapping the reference object, then $P_{AC} = 0$.

The ratio of correctly extracted area A_C to the reference area A_{RC} is called the completeness P_{AR}.

$$P_{AR} = \frac{A_C}{A_{RC}} \tag{3}$$

The range of completeness is 0 to 1. If all reference objects have their own fully corresponding evaluated objects, then $P_{AR} = 1$. If there is no reference object from the same thematic class overlapping the evaluated object, then $P_{AR} = 0$.

Equations (2) and (3) show an interaction between correctness and completeness. For instance, a large A_{DC} leads to a small correctness value, while a small A_{RC} results in a large completeness value. To amend this issue, the quality P_{AL} is designed to balance correctness and completeness.

$$P_{AL} = \frac{A_C}{A_{DC} + A_{RC} - A_C} \tag{4}$$

The range of quality is 0 to 1. If the extraction results are exactly the same as the reference data, then $P_{AL} = 1$. If no thematic class evaluated object overlaps with the reference object, then $P_{AL} = 0$.

Figure 1 presents two cases to illustrate the advantage of area-based accuracy measures compared to the confusion matrix. The accuracy values of two cases computed by the confusion matrix will be significantly different, as the confusion matrix depends on total pixel number. In contrast, the evaluation results for two cases using area-based accuracy measures are equivalent, because the latter measurements rely only on the evaluation and reference objects and are independent of the total pixel number.
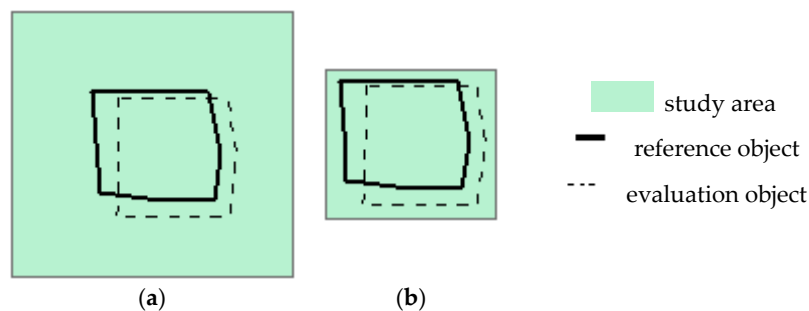
Figure 1. Schematic diagram of the influence of study area on object-based image analysis evaluation: (**a**) large study area; (**b**) small study area.

2.3. Number-Based Accuracy Measures

Three accuracy measures (i.e., correct, false, and missing rates), relying on counting the number of objects with different properties, are presented for testing OBIA performance. Specifically, the correct rate P_C, the false rate P_F, and the missing rate P_M are defined as

$$P_C = \frac{N_C}{N_C + N_F}, \ 0 \leq P_C \leq 1 \tag{5}$$

$$P_F = \frac{N_F}{N_C + N_F}, \ 0 \leq P_F \leq 1 \tag{6}$$

$$P_M = \frac{N_M}{N_C + N_M}, \ 0 \leq P_M \leq 1 \tag{7}$$

where N_C, N_F, and N_M represent the number of correct, false, and missed extracted objects, respectively. If all evaluated objects are correct, then $P_C = 1$ and $P_F = 0$. If all evaluated objects are incorrect,

then $P_C = 0$. If there is no false evaluated object, then $P_F = 1$. If all reference objects have their own correct evaluated objects, then $P_M = 0$. If no reference object corresponds correctly to the evaluated object, then $P_M = 1$.

The purpose of Equations (5)–(7) is to examine if the object is extracted correctly or falsely. To this end, if the proportion of correct pixels to total pixels for an object is larger than a given threshold, it is correctly extracted; otherwise, it is considered to be wrongly extracted.

2.4. Feature Similarity-Based Accuracy Measures

The difference between object-based and pixel-based accuracy measures is the assessment unit. Compared to the pixel, the object consists of many similar pixels, and thus has more features. The object number-based accuracy measures consider feature difference, but omit the feature detail difference and degree of difference between evaluated and reference objects. As shown in Figure 2a, if the area coincidence rate was used as a criterion to judge similarity, two evaluated objects would be judged correctly. However, two evaluated objects would be judged incorrectly, if the maximum deviation distance was used as the distinguishing criterion. Thus, the correctly extracted object number cannot fully reflect the difference between two evaluated objects that have large geometrical differences. The area-based measures can reflect differences between objects, but neglect object features. Although object extraction results may have the same correct, false, and missing rates, different object features derived from satellite images can generate different object qualities. Figure 2b shows that the overlap areas between two evaluated objects and reference objects are the same, but their locations and geometrical characteristics differ. This indicates that using object number or object area independently cannot assess object extraction accurately. This issue can be solved to a certain extent by considering more object features.
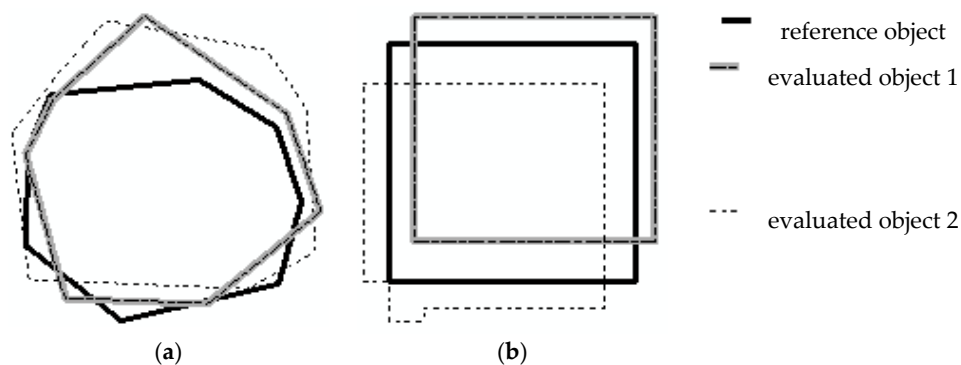
Figure 2. Schematic diagram of object area and number uncertainties: (**a**) feature difference; (**b**) feature detail difference.

Besides object number and object area, object geometric features can also reflect object difference, and thus can be used in complement to measure object extraction accuracy [31]. There are various geometric measures, and this paper selects typical measurements (e.g., area, perimeter, and barycenter) to design accuracy measures for OBIA.

The size difference reflects the basic similarity between two objects. Based on this observation, an object-based accuracy assessment method using size and size similarity S_M is defined as

$$S_M = \frac{\min(Size_C, Size_R)}{\max(Size_C, Size_R)} \tag{8}$$

where $Size_C$ denotes the size of the evaluated object and $Size_R$ denotes the size of the reference object. Standard geometric features, such as area, perimeter, and outer radius, can be used as assessment indices. The range of size similarity is 0 to 1. If all evaluated objects have the same size as that of the

reference object, then $S_M = 1$. If no evaluated object has the same size as that of the reference object, then $S_M = 0$.

Equation (8) ignores feature details of the object that may lead to inaccurate evaluation results. To tackle this issue, an improved size similarity S_F is presented.

$$S_F = 1 - \frac{|f_C - f_R|}{\min(f_C, f_R)} \tag{9}$$

where f_C is the feature value of the evaluated object, f_R is the feature value of the reference object, and $|f_C - f_R|$ is the feature difference between evaluated and reference objects. The features used in Equation (9) include area, perimeter, and diameter. The range of improved size similarity is 0 to 1. If all evaluated objects have the same size as that of the reference object, then $S_F = 1$. When the ratio between f_C and f_R exceeds 2 or is less than 0.5, then S_F is set to 0.

Equations (8) and (9) are relatively easy to implement, making them suitable for obtaining assessment results in near-real time. However, these two measures completely ignore the location difference that increases errors in the assessment results. To improve the measures in Equations (8) and (9), Tversky's feature contrast model [32], based on the feature similarity description, was proposed. This model measures S_O, the similarity of two objects, using the following equation:

$$S_O = \frac{f(C \cap R)}{f(C \cap R) + \alpha f(C - R) + \beta f(R - C)} \tag{10}$$

where $f(C \cap R)$ are common features of the evaluated object C and its reference object R, $f(C - R)$ denotes features that belong to the evaluated object C but not reference object R, and $f(R - C)$ stands for features that belong to the reference object R but not the evaluated object C, and α and β are weights for $f(C - R)$ and $f(R - C)$, respectively.

Equation (10) can measure the similarity of objects at the class or individual scale. Features in Equation (10) should be selected carefully, as some features (e.g., shape complexity, sphericity, and circularity), are challenging to describe using $f(C - R)$ and $f(R - C)$. To improve the generalization of the Tversky's feature contrast model, this paper defines an improved matching similarity as follows:

$$S_O = \frac{f_A(C \cap R)}{f_A(C \cap R) + \alpha f_A(C - R) + \beta f_A(R - C)} \tag{11}$$

where $f_A(C \cap R)$ represents features of the intersection area of C and R, $f_A(C - R)$ denotes features of the area of R to erase the evaluated object C, and $f_A(R - C)$ denotes the features of the area of the evaluated object C to erase reference object R. The improved model considers location differences and eases restrictions on feature selection. The range of S_O is 0 to 1. If the extracted and reference objects overlapped completely, then $S_O = 1$. If there is no overlap between the two objects, then $S_O = 0$.

Computing object similarity using a single feature will result in uncertain accuracy values. A natural solution is to apply multiple features to calculate object similarity. To this end, the object comprehensive similarity S' is defined as

$$S' = \begin{cases} 0 & if\ T_C \neq T_R \\ \frac{1}{N} \sum_{i=1}^{N} u_i S_i & otherwise \end{cases} \tag{12}$$

where T_C and T_R denote the classes of the evaluation and its matching reference objects, respectively, N is the number of features, S_i denotes the object similarity using the ith feature, and u_i is the weight of S_i. The feature weight is determined according to the real scenario, and the determination basis can be human subjectivity or feature applicability.

After computing the similarity of each evaluated object, the overall accuracy S_{overall} for object extraction can be calculated by

$$S_{\text{overall}} = \sum_{j=1}^{M} w_j S'_j \tag{13}$$

where S'_j is the calculated similarity of the jth evaluated object, M is the number of evaluated objects, and w_j denotes the weight of jth evaluation object. The ratio of the area of the evaluated object to the sum of the area of all objects can be used as the object weight.

2.5. Distance-Based Accuracy Measures

The distance difference between two objects, which can be completely reflected by the boundary distribution, is an essential aspect to evaluate the similarity between objects. Particularly, Pratt introduced a figure of merit (F_{FOM}) model to evaluate the accuracy of image segmentation [33]. F_{FOM} is defined as follows:

$$F_{\text{FOM}} = \frac{1}{\max\{l_C, l_R\}} \sum_{i=1}^{l_C} \frac{1}{1 + d_i^2} \tag{14}$$

where l_C is the boundary pixel number of the evaluated object, l_R is the boundary pixel number of its matching reference object, and d_i is the distance of the ith boundary pixel from the evaluated object to the corresponding pixel on the reference object's boundary. Based on F_{FOM}, the shape similarity B_D is defined as follows:

$$B_D = \frac{1}{\max\{l_C, l_R\}} \sum_{i=1}^{l_C} \frac{1}{1 + \frac{d_i}{\max(r_C, r_R)}} \tag{15}$$

where r_C and r_R are the radii of the circumcircles for the evaluation and reference objects, respectively.

Generally, the boundary of the extracted object cannot be strictly the same as that of the reference due to the error propagation during image interpretation. This phenomenon reduces the feasibility of using Equation (15). To improve the flexibility of B_D, a tolerance is set to judge if the two objects are identical. The improved B_D is defined as

$$B_D = \frac{1}{\max\{l_C, l_R\}} \sum_{i=1}^{l_C} f_i, \quad f_i = \begin{cases} 0 & d_i \geq d_2 \\ \frac{1}{1 + \frac{d_i}{\max(r_C, r_R)}} & d_1 < d_i < d_2 \\ 1 & d_i \leq d_1 \end{cases} \tag{16}$$

where d_1 and d_2 are two thresholds. The value of d_1 represents the tolerance for random errors. If the distance d_i is less than the threshold d_1, the distance d_i can be tolerated. The value of the d_2 represents the unacceptable value of the error. If the distance d_i is larger than the threshold d_2, there is no reference object boundary pixel that matches the ith boundary pixel of the evaluation object. In combination with application purposes, the values of d_1 and d_2 are determined by the size of the object and the spatial resolution of the image. The range of shape similarity is from 0 to 1. If all distances from boundary pixels of the objects to matching reference object boundary are within the tolerance range, then $B_D = 1$. If all boundary pixels of the objects are incorrectly extracted, then $B_D = 0$.

Equation (16) compares all pixels in the extracted and reference objects that lead to precise evaluation result as well as low computation efficiency. This process can be simplified by choosing boundary pixels at an equal interval range. The simplified shape similarity B_L is defined as follows:

$$B_L = \left(1 - \frac{\sum_{i=1}^{k} |l_C(\theta_i) - l_R(\theta_i)|}{k \, Min(r_C, r_R)} \right) \left(1 - \frac{d_{C-R}}{2 Min(r_C, r_R)} \right) \tag{17}$$

where k is the direction number, $l_C(\theta_i)$ and $l_R(\theta_i)$ denote the distance between the evaluation and reference objects from the barycenter to the boundary along the direction θ_i respectively, and $\theta_i = \frac{i*2\pi}{k}$. If the evaluation or reference object is a concave polygon, where there may be many boundary points along the direction θ_i, $l_C(\theta_i)$ or $l_R(\theta_i)$ is replaced by the mean distance. d_{C-R} is the distance between the evaluated object barycenter and reference object barycenter, respectively. The range of B_L is from 0 to 1. If all sampling boundary pixels of the objects overlap with the matching reference object boundary, then $B_D = 1$. If all evaluated objects have no matching reference objects, then $B_D = 0$. Before calculating the distance difference, the evaluated object is shifted to the gravity center of the reference object (see Figure 3).
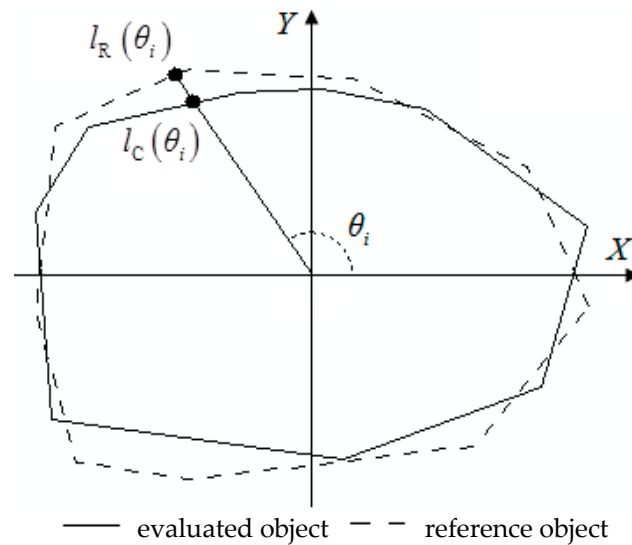

Figure 3. Illustration of Equation (17).

The sample number of boundary pixel determines the assessment reliability as well as the computation efficiency. If the requirement for reliability is high, the number of samples should be appropriately increased. If the calculation speed needs to be increased, the number of samples should be appropriately reduced.

Considering object classes, the object extraction accuracy B_{overall} of the entire evaluation area can be calculated using Equation (18).

$$B_{\text{overall}} = \sum_{j=1}^{M} w_j B_j', \quad B_j' = \begin{cases} 0 & \text{if } T_{C,j} \neq T_{R,j} \\ B_j & \text{otherwise} \end{cases} \tag{18}$$

where $T_{C,j}$ and $T_{R,j}$ denote the classes for the jth evaluated object and its matched reference objects, B_j is the shape similarity of the jth evaluated object with its matched reference objects on one of the above shape similarities, M is the number of evaluated objects, and w_j denotes the weight of the jth evaluation object. The proportion of the area of the evaluated object to the total area of objects can be used as the object weight.

3. Experimental Results and Analysis

In this section, the performance of the presented method was validated on two object types (i.e., water and building), as they are representative of natural and artificial scenarios in general. Generally, the extraction performance of water is satisfactory because water has a distinct boundary feature compared to its surrounding pixels. Compared to water extraction, the building extraction results may contain more errors that lead to lower accuracy, due to the complex environment surrounding buildings. On the other hand, the building boundary is regular, while the water boundary

is irregular. The experiments were conducted on a PC with an Intel Core2Quad processor at a clock speed of 1.80 GHz. MATLAB® and ARCGIS® were utilized to produce experimental results.

3.1. Data Description

Two QuickBird images were selected to validate the proposed method. One image, with a spatial resolution of 2.4 m per pixel and an area of 1200 × 1200 pixels, was acquired on 16 July 2009 over Wuhan, China (see Figure 4a). The study area located on the outskirts of the city was mainly covered by water, farmland, roads, and buildings. Another pan-sharpened image, with a spatial resolution of 0.61 m per pixel and an area of 400 × 400 pixels, was acquired on 2 May 2005 over Xuzhou, China (see Figure 5a). The second study area, locating near the city center, was mainly covered by buildings, roads, water, bare land, and grassland. Figures 4b and 5b present two complete reference maps produced via manual interpretation.
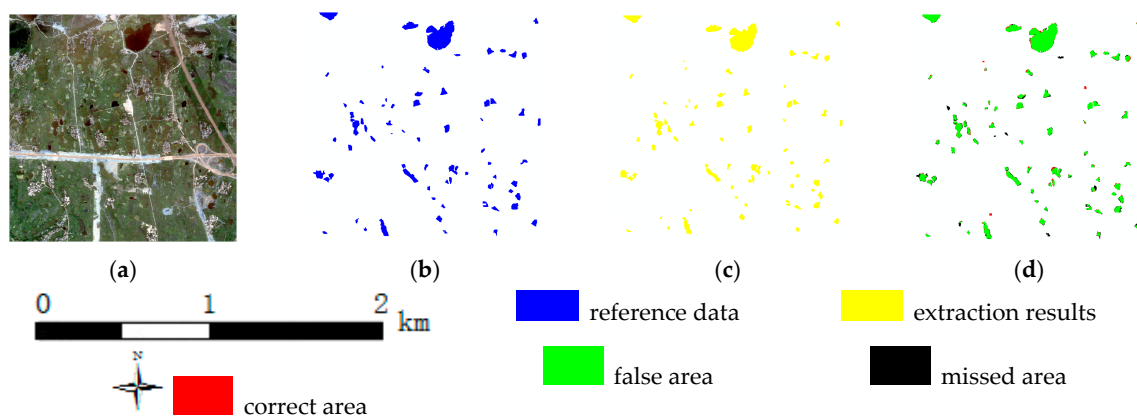
Figure 4. Water extraction results. Centre coordinates: 30°29′41″ N, 114°31′54″ E: (**a**) remote sensing image; (**b**) reference data; (**c**) extraction results; and (**d**) comparisons of extraction results with reference data.
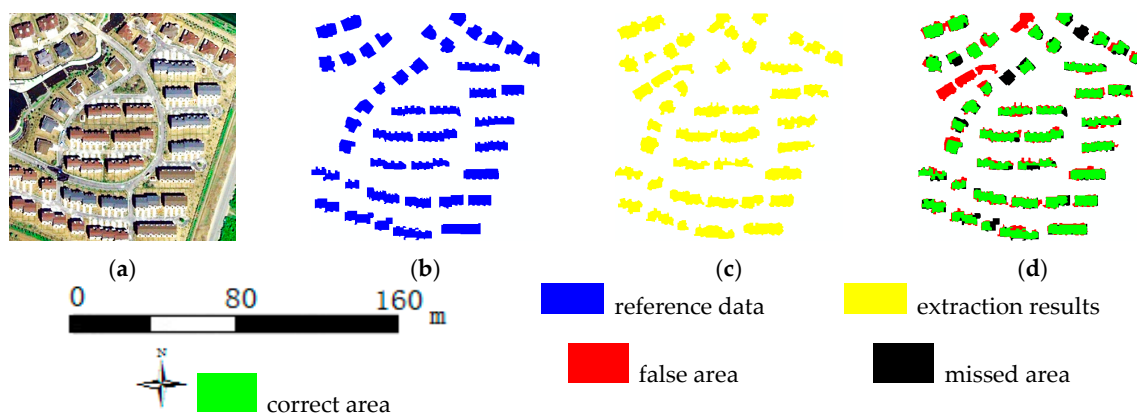
Figure 5. Building extraction results. Centre coordinates: 34°10′49″ N, 117°09′16″ E: (**a**) remote sensing image; (**b**) reference data; (**c**) extraction results; and (**d**) comparisons of extraction results with reference data.

3.2. Object Extraction

Satellite images are processed to generate objects that will be subsequently used to verify the performance of OBIA assessment measures. To this end, an improved watershed segmentation method [34] was applied. The advantage of this method is that it integrates the spectral information, texture feature and spatial relationships, which in turn makes it able to produce objects whose sizes

are closer to the true sizes. Once the segmentation results were obtained, object features, including the geometric characteristic, modified normalized difference water index (WNDWI) and normalized difference vegetation index (NDVI), were computed. Finally, the decision tree [35] using object features as input was performed to classify the image into target and background classes. The classification results are shown in Figures 4c and 5c, respectively.

3.3. Evaluation of Object Extraction Accuracy Using Different Measures

The accuracy assessment is carried out by comparing extracted results with reference data, as shown in Figures 4d and 5d. In this paper, the accuracy assessment employs all objects rather than a fixed number of test samples.

The object extraction accuracies of two classes are firstly evaluated by the area-based measurement, and Table 1 reports the evaluation results. It can be seen that the water class performs better than the building class. The reason is that water is easily separated from background due to the relatively large spectral difference between water and the surrounding objects. However, both material change of the buildings rooves and the spectral similarity of buildings and their nearby objects decrease the extraction performance. The different performances of water and building classes is shown in Table 1, indicating that the area-based accuracy measure can assess OBIA accuracy in a straightforward and efficient manner.

Table 1. Object extraction evaluation in terms of the area-based measurement.

Class	Correctness (%)	Completeness (%)	Quality (%)
Water	93.55	89.09	83.94
Building	76.34	83.84	66.55

In the second experiment, an accuracy evaluation was conducted using the object number-based accuracy index. To this end, numbers of total, correct, incorrect, and missing objects need to be computed in advance. As in the first experiment, the object matching method is able to judge whether objects are extracted correctly. With such automatic object extraction methods, it is extremely difficult to achieve accurate results. To assess the precision characteristic based on the object number, different object matching thresholds are set to judge if the object is extracted correctly. Table 2 reports the evaluation results. The water class has a generally better performance than the building class, indicating that the object number-based index can reflect the OBIA performance. Table 2 also shows that the choice of threshold value has a profound impact on the number of correctly identified objects. If the threshold value is high, the correct number is low; conversely, a low threshold value leads to a high correct rate. Thus, the threshold can be taken as a guideline for users to measure the confidence level. That is, if objects with high confidence are required, the threshold should be set to a large value.

Table 2. Object extraction evaluation using the object number-based measurement.

Class	Threshold	Correct Number	Correct Rate (%)	False Rate (%)	Missing Rate (%)
	0.90	38	56.72	43.28	46.48
Water	0.85	55	82.09	17.91	22.54
	0.80	63	94.03	5.97	11.27
	0.90	4	9.30	90.70	90.48
Building	0.85	21	48.84	51.16	50.00
	0.80	31	72.09	27.91	26.19

Geometric features can effectively reflect the characteristics of an object. This experiment validates its potential in evaluating OBIA results. To this end, area and perimeter are selected to measure the feature similarity between the extracted and reference objects. In this experiment, weights of area

and perimeter are set as 0.67 and 0.33, respectively, by trial and error. The size similarity is calculated using Equations (8), (12), and (13), the improved size similarity is calculated using Equations (9), (12), and (13), and the matching similarity is calculated using Equations (11)–(13), as shown in Table 3. The evaluation results indicate similar trends in different similarity measures for the two experimental areas. Similar to other assessment measures, all the similarity accuracies for the Xuzhou area are lower than those for the Wuhan area. The size similarity and improved size similarity in terms of area are both lower than those in terms of perimeter. However, matching similarity in terms of area is higher than that in terms of perimeter. For both the area and perimeter as assessment measures, the size similarity is always higher than the improved size similarity and the matching similarity is the lowest. This difference stems from the size similarity and improved size similarity, which do not consider the positional differences. The matching similarity considers the positional difference between the evaluated object and the reference data, which better reflects the similarity of features. The similarities of objects, calculated using different methods and features, differ, and the accuracy (based on similarity) also contains great uncertainty. To obtain stable evaluation results, more features should be considered to calculate the similarity.

Table 3. Assessing quality in terms of similarity.

Class	Index	Size Similarity (%)	Improved Size Similarity (%)	Matching Similarity (%)
	Area	84.86	82.57	77.09
Water	Perimeter	89.12	87.88	58.92
	Area and perimeter	86.28	84.34	71.04
	Area	79.19	73.74	66.21
Building	Perimeter	81.16	78.55	55.55
	Area and perimeter	79.85	75.34	62.66

The Euclidean distance between the gravity centers is used to calculate the distance between the evaluated object and its matching reference object. Thresholds d_1 and d_2 are set to the width of a pixel and five pixels, respectively. The shape similarity is calculated using Equations (16) and (18), and the improved shape similarity is estimated using Equations (17) and (18), as shown in Table 4. In the two experimental areas, the similarity based on boundary distance is slightly higher than that based on boundary difference and barycenter distance. The precision of object extraction is relatively high in the Wuhan area, which can be attributed to the similarity based on boundary distance being very close to that based on boundary difference and barycenter distance. Both similarities are calculated by comparing the detailed differences in objects, which fully reflects the differences in objects and ensures that the assessment measures are more stable. Although they require a tedious calculation process, these two measures need to be considered when assessing the accuracy of object extraction requiring high precision.

Table 4. Assessing quality in terms of distance.

Class	Shape Similarity (%)	Improved Shape Similarity (%)
Water	85.50	85.90
Building	76.46	77.99

A comprehensive comparison of the object extraction for the two experimental areas is generated using Tables 1–4. According to the comprehensive comparison, we can conclude that the water extraction result is generally better than the building extraction result. The superior results are due to the greater spectral difference between water and its surrounding objects than that between buildings and their surrounding objects, especially as the interior pixels of water have high homogeneity and building structures are complex.

4. Discussion

This paper presents four kinds of accuracy measures based on different object characteristics, namely, area-based, object number-based, feature similarity-based, and distance-based accuracy measures. Since the study area is usually large, a suitable sampling method needs to be selected before assessing the accuracy. The accuracy assessment results using object area are similar to those at the pixel level. Accuracy assessment based on object number requires that the objects are first extracted correctly. The criteria used to check if the objects are correctly or falsely identified have a profound impact on the assessment results. Many characteristics can be used as the basis of assessment measures for object feature similarity. The selection of a base has a significant impact on the assessment results: unreasonable feature selection will lead to unreliable assessment results (such as when only perimeter or length is used, or equal weights are used for area and perimeter). Thus, feature selection and feature weight are essential to determine the optimal basis.

The computation of difference-based accuracy measures is relatively complex. Selecting boundary pixels at reasonable intervals can improve the computational efficiency while retaining the reliability of the assessment results. Both area- and object number-based accuracy measures ignore the object detail. By contrast, feature-based accuracy assessments do not directly judge an object to be correct or otherwise; this aspect can better reflect the feature difference between the extraction object and the reference object. The accuracy measures based on boundary/location difference consider details of the object and reflect the object local feature difference, resulting in more accurate and confident results. Each assessment measure has its own advantages and disadvantages. Thus, the measures should be selected carefully according to the requirements. Specifically, if the accuracy needs to be computed in near-real time, the area-based measure would be chosen. This is because the area-based accuracy measure is straightforward and does not require an object matching process. Despite some errors, considering the advantage of computational efficiency, area-based measures can be selected to obtain faster assessment results with fewer requirements. If object extraction accuracy and its confidence are required simultaneously, the object number-based accuracy measure is recommended, as the threshold of coincidence degree is related to the confidence level: the larger the threshold, the higher the confidence level. If object extraction accuracy is to be understood comprehensively, feature-based and distance-based accuracy measurements are advisable as they fully consider detailed information of the object, such as shape and size.

5. Conclusions

A series of factors influence the assessment of object extraction from remote sensing images, which makes a complete and general accuracy index difficult to obtain. To tackle this issue, this paper presents four novel assessment measures with different criteria. The designed measurements are highly generalizable and provide users with practical means to evaluate object extraction results according to their unique needs. The methods presented in this paper require static objects with clearly defined edges. Further investigation and experimentation into dynamic objects (e.g., moving clouds, cars, and ships) with fuzzy boundaries is strongly recommended. The accuracy for objects with indeterminate boundaries can be assessed by two means: (1) assuming that the determinate boundaries are assigned to an object with indeterminate boundaries; and (2) setting the tolerance for the uncertainty of object boundaries (for example, the accuracy can be obtained by calculating the shape similarity based on the tolerance of object boundary distance).

Author Contributions: Liping Cai proposed the study, conducted the experiments, interpreted the results, and wrote and revised the manuscript. Wenzhong Shi advised on the study design and the manuscript structure.

References

1. Guanter, L.; Segl, K.; Kaufmann, H. Simulation of optical remote-sensing scenes with application to the EnMAP hyperspectral mission. *TGRS* **2009**, *47*, 2340–2351. [CrossRef]

2. Bovensmann, H.; Buchwitz, M.; Burrows, J.P.; Reuter, M.; Krings, T.; Gerilowski, K.; Schneising, O.; Heymann, J.; Tretner, A.; Erzinger, J. A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications. *Atmos. Meas. Tech.* **2010**, *3*, 781–811. [CrossRef]

3. Pajares, G. Overview and current status of remote sensing applications based on Unmanned Aerial Vehicles (UAVs). *Photogramm. Eng. Remote Sens.* **2015**, *81*, 281–329. [CrossRef]

4. Yu, Q. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 799–811. [CrossRef]

5. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]

6. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [CrossRef]

7. Kettig, R.L.; Landgrebe, D.A. Classification of multispectral image data by extraction and classification of homogeneous objects. *ITGE* **1976**, *14*, 19–26. [CrossRef]

8. Zhan, Q.; Molenaar, M.; Tempfli, K.; Shi, W. Quality assessment for geo-spatial objects derived from remotely sensed data. *Int. J. Remote Sens.* **2005**, *26*, 2953–2974. [CrossRef]

9. Möller, M.; Lymburner, L.; Volk, M. The comparison index: A tool for assessing the accuracy of image segmentation. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 311–321. [CrossRef]

10. Rutzinger, M.; Rottensteiner, F.; Pfeifer, N. A comparison of evaluation techniques for building extraction from airborne laser scanning. *J.-STARS* **2009**, *2*, 11–20. [CrossRef]

11. Hofmann, P.; Blaschke, T.; Strobl, J. Quantifying the robustness of fuzzy rule sets in object-based image analysis. *Int. J. Remote Sens.* **2011**, *32*, 7359–7381. [CrossRef]

12. Radoux, J.; Bogaert, P. Accounting for the area of polygon sampling units for the prediction of primary accuracy assessment indices. *Remote Sens. Environ.* **2014**, *142*, 9–19. [CrossRef]

13. Styers, D.M.; Moskal, L.M.; Richardson, J.J.; Halabisky, M.A. Evaluation of the contribution of LiDAR data and postclassification procedures to object-based classification accuracy. *J. Appl. Remote Sens.* **2014**, *8*, 083529. [CrossRef]

14. Whiteside, T.G.; Maier, S.W.; Boggs, G.S. Area-based and location-based validation of classified image objects. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *28*, 117–130. [CrossRef]

15. Shi, W.; Zhang, X.; Hao, M.; Shao, P.; Cai, L.; Lyu, X. Validation of land cover products using reliability evaluation methods. *Remote Sens.* **2015**, *7*, 7846–7864. [CrossRef]

16. Yang, J.; He, Y.; Caspersen, J.; Jones, T. A discrepancy measure for segmentation evaluation from the perspective of object recognition. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 186–192. [CrossRef]

17. Zhang, X.; Feng, X.; Xiao, P.; He, G.; Zhu, L. Segmentation quality evaluation using region-based precision and recall measures for remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 73–84. [CrossRef]

18. Maclean, M.G.; Congalton, R.G. Map accuracy assessment issues when using an object-oriented approach. In Proceedings of the ASPRS 2012 Annual Conference, Sacramento, CA, USA, 19–23 March 2012; pp. 19–23.

19. Clinton, N.; Holt, A.; Scarborough, J.; Yan, L.; Gong, P. Accuracy assessment measures for object-based image segmentation goodness. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 289–299. [CrossRef]

20. Montaghi, A.; Larsen, R.; Greve, M.H. Accuracy assessment measures for image segmentation goodness of the Land Parcel Identification System (LPIS) in Denmark. *Remote Sens. Lett.* **2013**, *4*, 946–955. [CrossRef]

21. Awrangjeb, M.; Fraser, C.S. An automatic and threshold-free performance evaluation system for building extraction techniques from airborne LiDAR data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4184–4198. [CrossRef]

22. Bonnet, S.; Gaulton, R.; Lehaire, F.; Lejeune, P. Canopy Gap Mapping from Airborne Laser Scanning: An Assessment of the Positional and Geometrical Accuracy. *Remote Sens.* **2015**, *7*, 11267–11294. [CrossRef]

23. Shahzad, N.; Ahmad, S.R.; Ashraf, S. An assessment of pan-sharpening algorithms for mapping mangrove ecosystems: A hybrid approach. *Int. J. Remote Sens.* **2017**, *38*, 1579–1599. [CrossRef]

24. Kuffer, M.; Pfeffer, K.; Sliuzas, R.; Baud, I. Extraction of slum areas from VHR imagery using GLCM variance. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1830–1840. [CrossRef]

25. Möller, M.; Birger, J.; Gidudu, A.; Gläßer, C. A framework for the geometric accuracy assessment of classified objects. *Int. J. Remote Sens.* **2013**, *34*, 8685–8698. [CrossRef]

26. Cheng, J.; Bo, Y.; Zhu, Y.; Ji, X. A novel method for assessing the segmentation quality of high-spatial resolution remote-sensing images. *Int. J. Remote Sens.* **2014**, *35*, 3816–3839. [CrossRef]

27. Eisank, C.; Smith, M.; Hillier, J. Assessment of multiresolution segmentation for delimiting drumlins in digital elevation models. *Geomorphology* **2014**, *214*, 452–464. [CrossRef] [PubMed]

28. Zhang, X.; Xiao, P.; Feng, X.; Feng, L.; Ye, N. Toward evaluating multiscale segmentations of high spatial resolution remote sensing images. *TGRS* **2015**, *53*, 3694–3706. [CrossRef]

29. Radoux, J.; Bogaert, P. Good Practices for Object-Based Accuracy Assessment. *Remote Sens.* **2017**, *9*, 646. [CrossRef]

30. Witharana, C.; Civco, D.L.; Meyer, T.H. Evaluation of data fusion and image segmentation in earth observation based rapid mapping workflows. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 1–18. [CrossRef]

31. Lizarazo, I. Accuracy assessment of object-based image classification: Another STEP. *Int. J. Remote Sens.* **2014**, *35*, 6135–6156. [CrossRef]

32. Tversky, A. Features of similarity. *Read. Cognit. Sci.* **1977**, *84*, 290–302. [CrossRef]

33. Pratt, W. *Introduction to Digital Image Processing*; CRC Press: Boca Raton, FL, USA, 2013.

34. Cai, L.; Shi, W.; He, P.; Miao, Z.; Hao, M.; Zhang, H. Fusion of multiple features to produce a segmentation algorithm for remote sensing images. *Remote Sens. Lett.* **2015**, *6*, 390–398. [CrossRef]

35. Friedl, M.A.; Brodley, C.E. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* **1997**, *61*, 399–409. [CrossRef]