Urban Surface Water Mapping from VHR Images Based on Superpixel Segmentation and Target Detection

Qingwei Liu, Yugang Tian, Lihao Zhang, and Bo Chen

***Abstract*—Surface water is a fundamental resource in urban environments. Monitoring the spatio-temporal distribution of urban surface water from remotely sensed images is crucial for urban planning and management. Unfortunately, due to the limitation of spatial resolution, the method based on low/medium resolution images is difficult to extract small water bodies accurately. Recently, very high resolution (VHR) images have shown considerable potential for urban compositions mapping. However, fewer spectral bands, shadows, and high spectral heterogeneity of VHR images hinder the application of traditional methods. In this article, we proposed an urban surface water mapping method called sparse superpixel-based water extraction (SSWE) from VHR images. The method includes three steps: clustering water bodies into sparse targets at the object level by an improved scale-adaptive simple non-iterative clustering (SA-SNIC) superpixel segmentation; generating new bands with additional spectral, spatial, and derived features, to increase the dimensions of original data and enhance the separability between water bodies and background covers; and constructing a positive-negative constrained energy minimization multitarget sparse detector to highlight the water bodies while suppressing shadows. The proposed method was applied to GF-2 multispectral images of four cities in China. The results showed that SSWE achieved the highest accuracy compared with other methods, with an average OA of 98.91% and an average kappa coefficient of 0.942. Furthermore, the separability analysis also indicated that SSWE could effectively distinguish urban water bodies from shadows and other land covers. Stable results can be acquired by the suggested parameters and thresholds of SSWE.**

***Index Terms*—Sparse target detection, superpixel segmentation, urban surface water mapping, very high resolution (VHR) images.**

I. Introduction

AS ONE of the significant components of terrestrial water storage, surface water maintains diverse and complex societies and ecological systems [1]–[3]. Urban surface water is a vital indicator for flood monitoring [4], [5], urban hydrological evaluation [6]–[8], and urban heat island effect assessment [9], [10] and plays a significant role in urban planning. Given the background that the rapid urbanization process has caused dynamic changes in water bodies, it is of great significance to obtain the water body information and its spatio-temporal distribution quickly and accurately [11], [12].

Remote sensing technology has been widely applied for water bodies extraction in the past few decades. In general, these methods can be roughly grouped into three categories.

A. Spectral Mixture Analysis (SMA)

The methods in this class are dedicated to solving the spectral mixing effect of mixed land-water pixels [13]–[16], which assume that the spectrum of mixed pixels is composed of linear or nonlinear combinations of the spectrum of typical endmembers [8]. These methods are generally performed in two steps: endmember selection and abundance estimation, among which choosing appropriate endmembers is the critical challenge for large-area mapping [17], [18].

B. Classification

They include the unsupervised and the supervised classification, such as k-means [19], [20], support vector machine (SVM) [21], [22], decision tree [23], random forest [24], and deep learning [25], [26], etc. These classifiers are data-driven, and they learn and analyze the spectral and spatial features from different categories. Unsupervised classifiers used to be hindered by poor classification accuracy of complex optical images, while the performance of supervised classifiers depends on prior knowledge and training samples [27].

C. Index-based

These methods tend to highlight the characteristics that water can absorb most of the incident energy at near-infrared (NIR) and middle-infrared (MIR) wavelengths to construct a series of water indices [14]. Among them, the well-known normalized difference water index (NDWI) was first built with the reflectance of green and NIR bands [28]. And the modified NDWI (MNDWI) is constructed by replacing the NIR with the MIR band to suppress the signal in the built-up area [29]. Furthermore, the automated water extraction index (AWEI) is

Manuscript received December 24, 2021; revised April 13, 2022; accepted June 1, 2022. Date of publication June 10, 2022; date of current version July 13, 2022. This work was supported by the National Key Research and Development Program of China under Grant 2018YFB1004604. *(Corresponding author: Yugang Tian.)*

Qingwei Liu, Yugang Tian, and Lihao Zhang are with the School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China (e-mail: lqwei@cug.edu.cn; ygangtian@cug.edu.cn; zhanglihao@cug.edu.cn).

Bo Chen is with the State Key Laboratory of Earth Surface Processes and Resource Ecology, Academy of Disaster Reduction and Emergency Management, Beijing Normal University, Beijing 100875, China (e-mail: bochen@bnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3181720

designed as a linear combination of multiple bands to improve the accuracy of water extraction under the urban scenes over shadow and low albedo surfaces [30]. Excellent performance and simple implementation make the water index widely used, but the instability is reflected in the fluctuation of the binarization threshold when facing different backgrounds [21], [31], [32].

Although the methods mentioned above have been widely applied in natural regions for surface water extraction, most of them are only applicable to medium/low resolution satellites and broad-band multispectral sensors. Unlike rural scenes, the composition of artificial material in urban areas is more complex, which means higher requirements for the training samples for knowledge-based methods [11]. Moreover, the small water bodies with diverse morphologies often lead to underestimation [33]. The very high resolution (VHR) images have more detailed spatial information than the medium/low ones, which could serve as important data sources for accurate urban surface water extraction [25]. But as far as we know, few studies focus on urban surface water extraction using VHR images because it faces many new challenges: most VHR images such as IKONOS; QuickBird; Pleiades; Kompsat; and GaoFen-2 (GF-2), only have visible and NIR bands, limited spectral bands make some water indices inapplicable, such as MNDWI and AWEI, for lacking the SWIR band [21]. Shadows and low albedo surfaces with similar spectral features to water bodies are more conspicuous in VHR images, and low spectral separability may lead to overestimation of urban water bodies [34]. The high spectral heterogeneity of VHR images leads to the increase in intraclass spectral variance and the decrease in inter-class spectral separability, resulting in "salt and pepper" noises in water mapping [16], [35]. Overall, the above conventional methods using the medium/low images are not suitable for urban surface water extraction from VHR ones.

To address these issues, some researchers have designed rule-based methods to suppress the false alarms of building shadows based on the morphological features [morphological shadow index (MSI)] [36], texture features [gray-level co-occurrence matrix (GLCM)] [37], and the spatial adjacency between shadows and buildings [21]. Huang *et al.* [12] identified water types by extracting texture and geometrical features at the object level, and they found that the misclassified shadow at the pixel level can be correctly classified based on object characteristics. It has been suggested to use the object-based image analysis (OBIA) technique for VHR images, while subpixel/pixel-based classification methods are more suitable for medium/low-resolution images [38]. As a fundamental step of OBIA, segmentation aims to group pixels into homogenous objects to replace pixels as the basic analysis units [39], [40]. Thus, noisy pixels with mutation signals will be covered by objects, and more importantly, the additional spectral, textural, and geometrical features can be calculated for classification decisions at the object level. These newly added features can make up for the insufficient spectral bands and enhance the separability between water and backgrounds. However, the segmentation scale needs to be adjusted manually according to the spatial resolution and scene complexity in most OBIA methods, which will exert great influence on the computational efficiency and extraction accuracy [41].

An easily overlooked point is that water bodies have the characteristics of strong homogeneity and connectivity than most land covers. In terms of the segmentation strategy, more attention should be paid to the boundary adherence of the target of interest rather than the background covers. Superpixels, as a state-of-the-art segmentation paradigm, have been increasingly employed in the field of remote sensing for target detection, including high resolution images [42], hyperspectral images [43], and SAR images [44]. They segment images into meaningful atomic regions between pixel and object scales, which represent the image more abstractly than pixels. Superpixel-based segmentation is favored for its better compactness, faster speed, and fewer parameters than OBIA methods [45]. However, the superpixel algorithms adopt an over-segmentation strategy, which ignores the difference in the original scale of different land covers. Some researchers proposed to merge the superpixels into objects after the over-segmentation [44], [46], [47], but this actually increased the complexity of superpixel algorithms. To solve this problem, we propose a superpixel algorithm for multiscale segmentation without merging, which can adaptively adjust the segmentation scale according to the homogeneity of objects, so the object integrity of homogeneous objects such as water bodies can be maximally preserved, while complex objects like artificial covers are divided into pieces. Consequently, water bodies become sparsely distributed objects with a low probability after segmentation, which can be extracted more accurately by sparse target detectors.

The constrained energy minimization (CEM) algorithm is wildly used for the detection of sparse targets with a small size or low probability distribution in hyperspectral data [48]. Geng *et al.* pointed out that it is possible to apply CEM in the multispectral image by adding linearly irrelevant bands [49]. CEM is increasingly utilized for target detection in multispectral images, particularly for sparse mineral occurrence mapping [50], [51]. However, for large targets like water bodies, CEM will result in severe overestimation and omission errors [52]. To remedy this drawback, Ji *et al.* [52] proposed orthogonal subspace projection weighted CEM (OWCEM) to decrease the high contribution of water body pixels in the output energy of CEM by multiplying each pixel by a correlation weight coefficient. Yang *et al.* [53] also considered the relatively large proportion of water bodies, and regarded the noise instead of water as the target, as noise is sparsely distributed in water maps, which can be detected by CEM, and water bodies are therefore created by removing the detected noise from AWEI results. The above two kinds of schemes turn the problem of water bodies extraction into the sparse target detection, and this goal can also be achieved from the perspective of segmentation. Nevertheless, CEM and OWCEM are only single target detectors without eliminating undesired targets. Although the additional spectral bands are introduced in their studies, they both neglect rich spatial information, and the challenge from urban building shadows still remains.

In this article, we developed a novel urban surface water mapping framework for VHR images called sparse superpixel-based water extraction (SSWE). The main idea includes as follows.

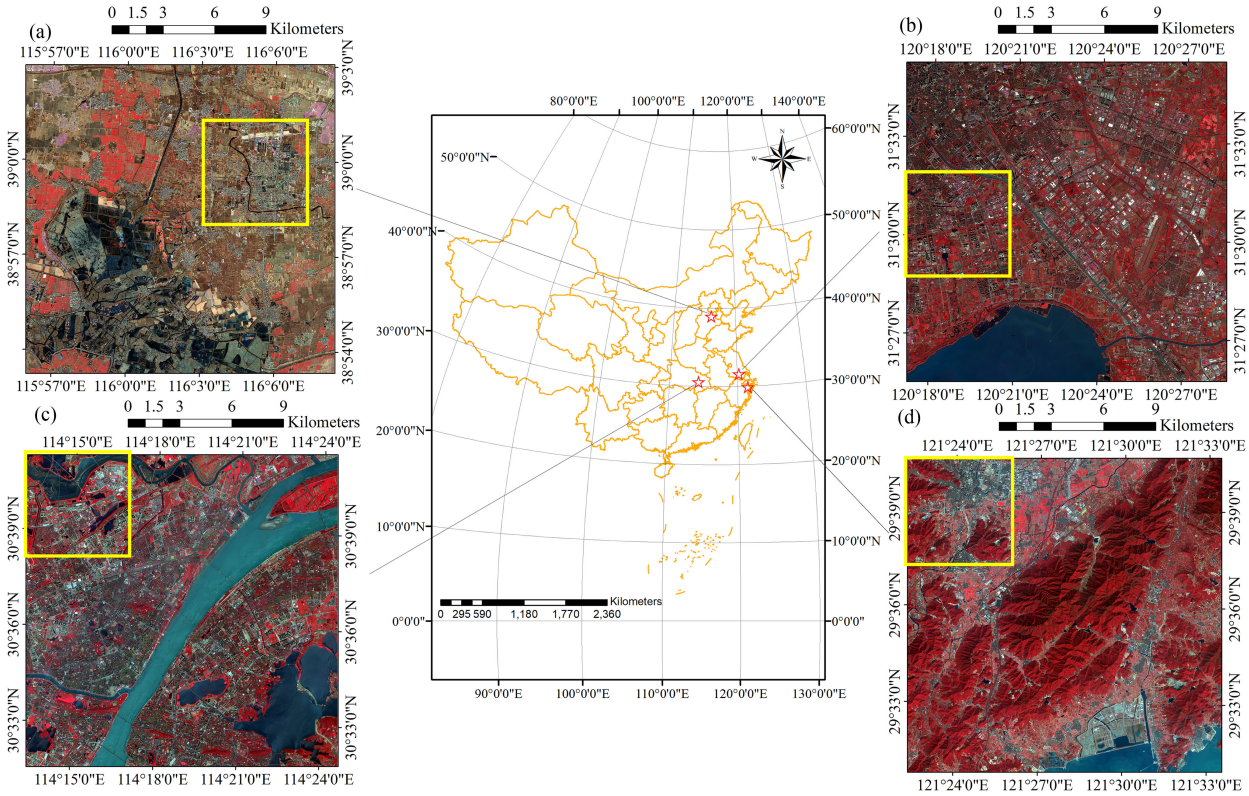

Fig. 1. Four study areas. (a) Xiong'an. (b) Wuxi. (c) Wuhan. (d) Ningbo. The images are shown with a pseudocolor, and the yellow rectangular regions in each image are selected for separability analysis.

1) We first improved the superpixel algorithm according to the characteristics of water bodies to obtain the scale-adaptive segmentation results, where water bodies maintain the integrity of objects and become sparsely distributed in object-level statistics without complex parameters and calculations.
2) Spatial-spectral information of VHR images, such as NDWI, MSI, and homogeneity, etc., were fully exploited to discriminate water against backgrounds and offer training samples for the next step.
3) A multitarget detector called positive-negative CEM (PNCEM) was designed to extract sparsely distributed water bodies, which can pass the positive targets while eliminating the negative targets.

The rest of this article is organized as follows. Section II displays four experimental areas and VHR satellite data. Section III introduces the detailed steps of the proposed SSWE and accuracy evaluation indices. Section IV describes specific results of water extraction, including accuracy analysis and separability analysis. Further, the parameters, input features, and thresholds of SSWE are discussed in Section V. Finally, Section VI concludes this article.

II. Study Area and Data

A. Study Area

Four cities (Xiong'an, Wuxi, Wuhan, and Ningbo) with different surface water abundances in China are selected to test the proposed SSWE method (see Fig. 1). Xiong'an (38° 58′N, 116°3′E) has more water resources than most of the other northern cities in China due to its proximity to Baiyangdian Lake. Wuxi (31°31′N, 120°23′E) orients in the north of Taihu Lake, and many canals are running through the city. Wuhan (30° 33′N, 114°17′E) is located in the middle and lower reaches of the Yangtze River and is known as the "city with hundreds of lakes". Ningbo (29° 35′N, 121°28′E), a coastal city in eastern China with high mountain ridges, has rich mountain shadows. Table I gives the details of the study areas. The main water types refer to rivers, lakes, ponds, and canals with complex features involving water quality and geometrical shapes. The main noise types include building and mountain shadows, dark roads, and low albedo surfaces. Different types of water and noise in each city bring challenges to water detection.

B. Data

In this article, GaoFen-2 (GF-2) satellite images are selected as the experimental data. In addition to a panchromatic camera with 0.8 m nadir resolution, GF-2 is also equipped with two 4 m multispectral scanners (PMS1, PMS2), containing visible and NIR bands. High positioning accuracy and fast attitude maneuvering capability effectively enhance the comprehensive observation effectiveness of GF-2 satellite. The high spatial resolution of GF-2 images makes it possible to extract delicate urban surface water bodies. GF-2 images can be downloaded from the China centre for resource satellite data and application.[1] All

[1][Online]. Available: http://www.cresda.com/EN/

TABLE I
DETAILS OF THE STUDY AREAS

Study area	Water types	Main noise	Image date	Water coverage	Image size
Xiong'an	Paddy field, Eutrophic ponds, Narrow turbid rivers, Lakes	Building shadows	January 1, 2020	10.00%	4500×4500
Wuxi	Clear lakes, Narrow eutrophic rivers, Clear canals, Turbid canals	Building shadows	November 4, 2014	12.45%	4500× 4500
Wuhan	Narrow turbid rivers, Clear rivers, Clear lakes, Narrow eutrophic rivers, Eutrophic lakes, Eutrophic Ponds	Dark roads Building shadows Low albedo features	September 1, 2016	22.33%	4500×4500
Ningbo	Narrow clear rivers, Harbors, Clear Ponds, Reservoirs, Sea	Mountain shadows Building shadows	march 1, 2017	4.63%	4500×4500

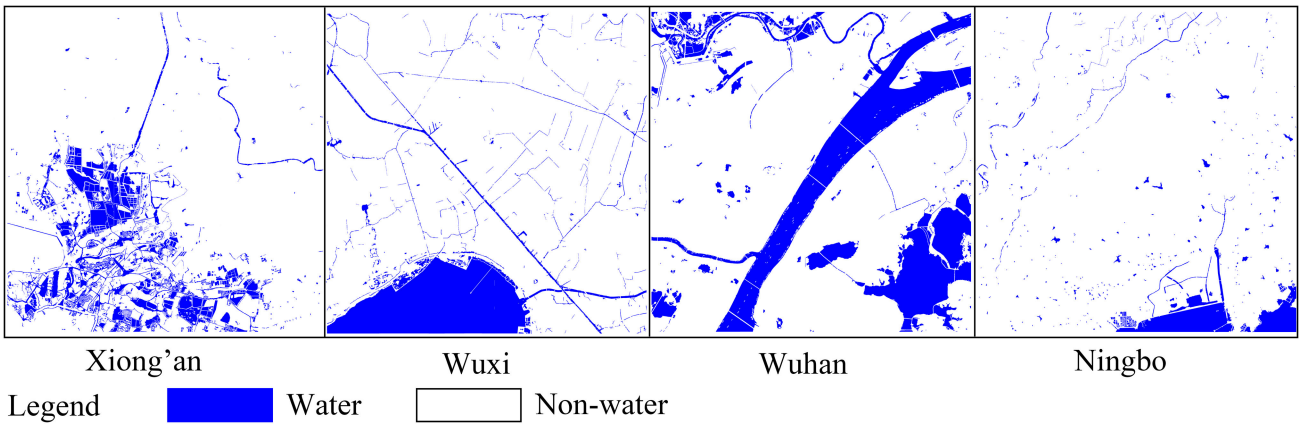

Fig. 2. Manually delineated reference water maps.

data we used in this article are level 1A products with complete radiometric calibration and geometric correction information.

Reference water maps are used to evaluate the accuracy of water extraction. Considering that the resolution of 0.8 m can clearly distinguish the boundary of the water body, the fusion image of GF-2 panchromatic and multispectral bands is used to assist in manually delineating water bodies (see Fig. 2).

C. Image Preprocessing

Before detecting water bodies, GF-2 images (level 1A) are preprocessed by ENVI 5.3, which contains orthorectification, radiometric calibration, atmospheric correction, and image fusion. Orthorectification is applied firstly to eliminate the deformation caused by terrain. Then, radiation calibration and atmospheric correction are performed to obtain the reflectance of ground objects. The Gram–Schmidt pan-sharpening algorithm in ENVI 5.3 was used to generate new fusion images.

III. METHOD

It is difficult to extract urban water bodies from medium/low resolution images for their diverse shapes, complex compositions, and narrow areas. Fortunately, the VHR images make it possible to map urban surface water accurately. However, the limited spectral bands, confusing shadows, and high heterogeneity of VHR images are newly prominent factors for water mapping. In this article, SSWE is proposed to suppress the noise caused by spectral heterogeneity under complex urban environments. First, the original superpixel algorithm is improved to satisfy the object-level segmentation results, which combines the advantages of superpixel segmentation and OBIA and ensures the integrity and the sparse distribution of the water bodies with stable parameters. Second, the additional spectral, spatial, and derived features, such as NDWI, MSI, and GLCM, are added to the original data as new bands to increase the separability of water bodies and other land covers, then water bodies become multi-vector sparse targets at the object level. Third, in order to overcome the extreme overestimation caused by shadows and low albedo objects, the PNCEM algorithm is designed with the input of the positive samples of water bodies and the negative samples of shadows. Finally, the OTSU [54] and empirical threshold are suggested in different scenes, and water bodies can be obtained by binary classification. The flowchart of SSWE is shown in Fig. 3.

A. Scale-Adaptive Simple Non-Iterative Clustering

Superpixel segmentation is an oversegmentation algorithm where superpixels are approximately equal in size and more diminutive than objects [55]. Currently, the simple linear iterative clustering (SLIC) algorithm performs well in the

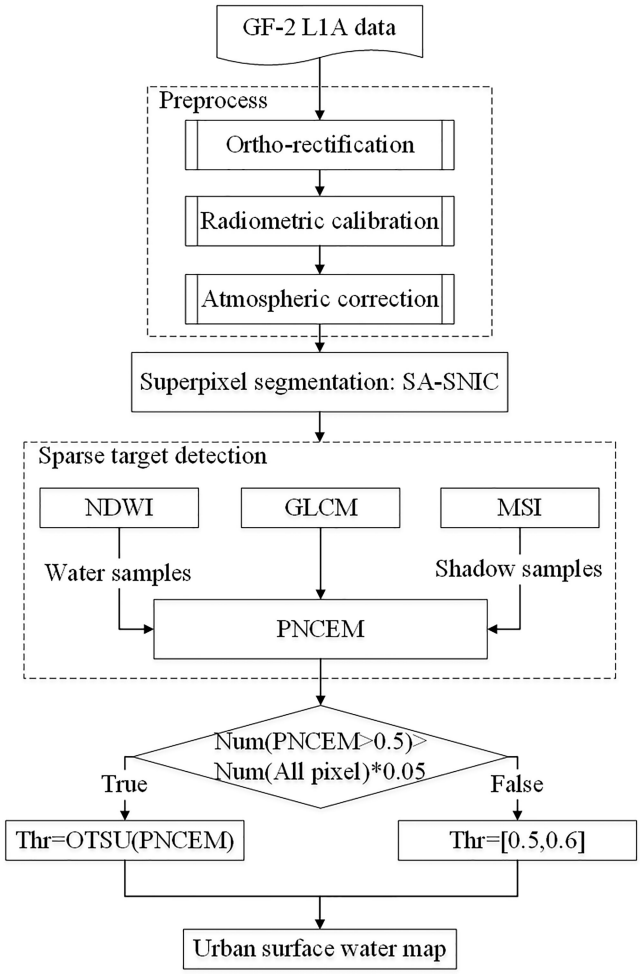

Fig. 3. Workflow of SSWE. Thr stands for threshold.

comprehensive evaluation of state-of-the-art superpixel algorithms [56]. And simple non-iterative clustering (SNIC) is an improved version of SLIC, which is superior to the latest technology in terms of segmentation quality and speed [57]. In order to preserve the complete boundary adherence of small objects, SNIC must increase the distribution density of initial seeds, which will lead to more detailed segmentation and more complex calculations. Furthermore, the result of oversegmentation often needs to be merged as an object, and the calculation burden will be doubled. To solve this problem, the improved scale-adaptive SNIC (SA-SNIC) is proposed to acquire compact boundaries without adding additional computational burden.

First, according to the standard SNIC algorithm, the five-dimensional normalized distance $d_{j,k}$ with spatial position $x = [x, y]^T$ and spectral position in CIELAB color $c = [l, a, b]^T$ in (1) is applied to measure the affinity of the pixel to the centroid of the superpixel

$$d_{j,k} = \sqrt{\frac{||x_j - x_k||_2^2}{s} + \frac{||c_j - c_k||_2^2}{m}} \tag{1}$$

where $[x_j,\ c_j]$ is the feature vectors of the *j*th candidate pixel, and $[x_k,\ c_k]$ is the centroid of the *k*th superpixel. The spatial normalizing factor s is the area of the *j*th cluster. The compactness factor m is set automatically or by users [57], which balances the compactness and boundary adherence of superpixels.

The core of the SNIC algorithm is the priority queue. In each iteration, the element with minimum d_{i,k_i} is popped from the priority queue, and the candidate pixel is labeled to the closest superpixel, then unlabeled neighborhoods of the candidate pixel are pushed into the priority queue. Therefore, the d_{i,k_i} of all elements becomes very large when the iteration times reach a certain number. Those candidate pixels are usually close to the edge of the object or some noise. If they are assigned as the labels of the nearest superpixels, the segmentation result will show a large object-level consistency error (OCE) and poor boundary adhesion [58]. In order to obtain homogeneous and complete clusters, we first reset the popped pixel with a distance beyond the constraint threshold Thr_{dis} to a new seed. Then, the distance from the unlabeled neighbors to the centroid of the new cluster is recalculated, which will be less than the distance to the original cluster centroid. Finally, a new superpixel will be generated in the local area. The flowchart of improved SA-SNIC is shown in Fig. 4, and the main steps are as follows.

1) Initial seeds are evenly distributed on the input image [see Fig. 4(a)]. They are sequentially assigned with different labels of superpixels, see in Fig. 4(b), different colors represent different superpixels. Then elements are constructed and pushed into the priority queue, which contains four factors: spatial position, spectral values, the label of the superpixel, and the distance to the centroid of the superpixel.
2) Top-most element with the smallest value of $d_{i,k}$ is popped from the priority queue. Fig. 4(c) intuitively shows the process, the pixel most similar to the neighboring superpixel is tagged with the same label as the superpixel, and the superpixel will be updated after adding pixels. If the 4 or 8 neighbors of the pixel have not been labeled, then new elements will be created and pushed into the right place of the priority queue.
3) When the $d_{i,k}$ of the top-most element e_i is greater than the threshold Thr_{dis}, element e_i will be reset as a new seed [see in Fig. 4(e)–(f)]. The distance of unlabeled neighbors of e_i to the centroid of the new cluster is recalculated and constructed as a new element and pushed into the right place of the priority queue.
4) Iterate steps (2) to (3) until the priority queue is emptied. After that, the segmentation result is shown in Fig. 4(g).
5) The superpixel with a size less than the constraint threshold Thr_{size} is labeled as background, as shown in the black patches in Fig. 4(h). Finally, the SA-SNIC segmentation algorithm is completed.

With the execution of SA-SNIC, labels are assigned at the head of the priority queue while new candidates are filled at the tail. When the queue is empty, all the pixels of the whole image are divided into different superpixels. Meanwhile, some noise may be reset as new seeds too, and many small superpixels are generated at the end of the clustering procedure. Combining small superpixels can reduce the workload of subsequent processing. After the operation is complete, all pixels are larger than the Thr_{size}, which is set by users.

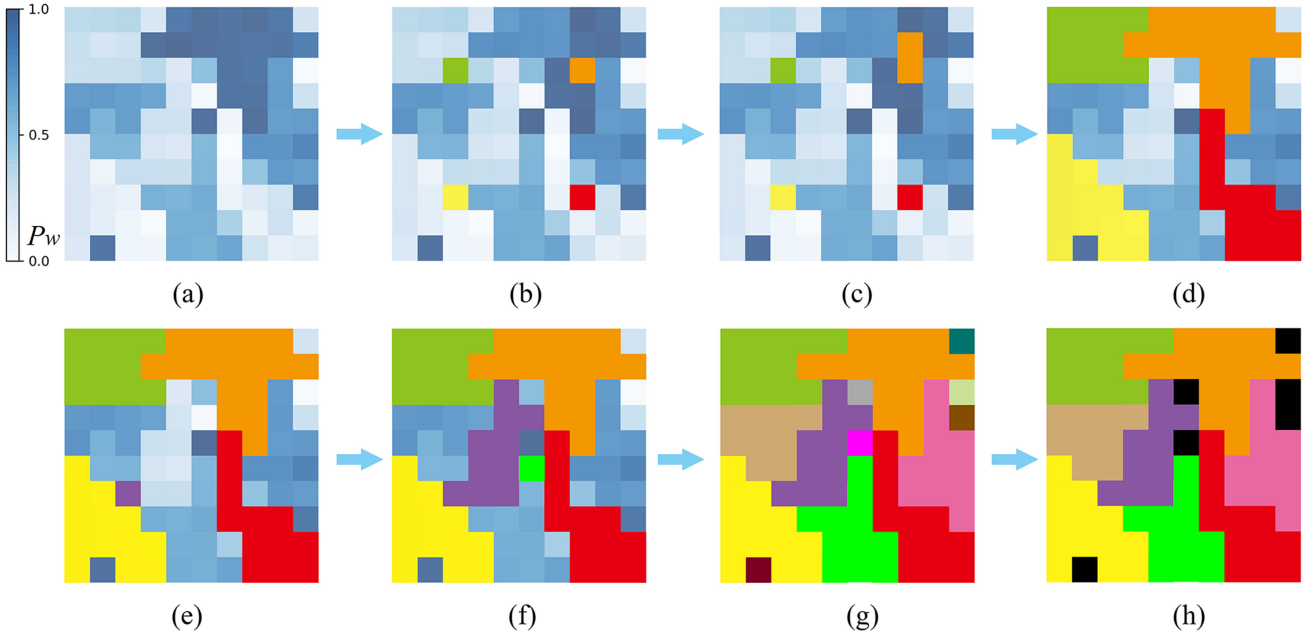

Fig. 4. Flowchart of the SA-SNIC segmentation. (a) Simplified input image and P_w represents the proportion of water in each pixel. (b)–(h) Intermediate results of the segmentation process, where different colors indicate different labels after segmentation, and black represents the background.

SA-SNIC has four input parameters: the initial number of superpixel seeds, compactness factor, and two newly introduced parameters: the distance threshold Thr_{dis} and the size threshold Thr_{size}. Since new seeds will be generated in the algorithm, the final number of superpixels obtained by segmentation will no longer be determined by the number of initial seeds but by Thr_{dis} and Thr_{size}. In practice, Thr_{size} depends on the minimum size of the detected target, so Thr_{dis} is the primary influence factor of SA-SNIC. These parameters are discussed in more detail in section V-A.

B. Band Expansion

Effective use of the rich spatial information of VHR images is an effective way to make up for their insufficient spectral information, which can help distinguish shadows and water bodies. Thus, spatial information should be fully exploited in urban surface water mapping. Considering that the water and shadow show a strong contrast from other land covers in NDWI, MSI, and GLCM, these features can be combined as new bands to increase class separability between water bodies and other noise. Meanwhile, more bands can help improve the performance of CEM [49].

1) Normalized Difference Water Index: NDWI has been widely used for water bodies extraction since it was proposed by McFeeters [28]. Its formula can be expressed as (2), and many studies have shown that NDWI can perform stably in various satellite images [59]

$$\text{NDWI} = \frac{\rho_{\text{Green}} - \rho_{\text{NIR}}}{\rho_{\text{Green}} + \rho_{\text{NIR}}} \tag{2}$$

where ρ_{Green} and ρ_{NIR} donate the reflectance of the green and NIR bands.

2) Morphological Shadow Index: MSI is an automatic shadow detector constructed by a sequence of morphological operators describing the spectral-spatial characteristics of shadows. The calculation equation of MSI is

$$\text{MSI} = \frac{\sum_{d,s} \text{DMP}_{\text{BTH}}(d,s)}{D \times S} \tag{3}$$

where d and s indicate the direction and length of the structural element, D and S are the total numbers of directions and scales, and $\text{DMP}_{\text{BTH}}(d,s)$ indicates the differential morphological profiles of black top-hat (BTH), which can highlight the dark structures within different directions and scales. In the urban scene, shadow usually presents sharp local contrast with adjacent buildings, which results in high MSI feature values in shadow areas [12], [60].

3) Gray-Level Co-Occurrence Matrix: GLCM is designed to describe texture features by considering spatial correlation statistic information at the gray level. Most water bodies show strong homogeneity, which is quite different from other background covers, including shadows. Therefore, we mainly consider the homogeneity, entropy, and angular second moment (ASM) features derived from GLCM.

In this article, a total of nine feature bands, including the original four spectral bands, NDWI, MSI, and three parts of GLCM, are used in SSWE. Each feature is normalized to [0, 1] to avoid weight imbalance. As shown in the subscene example in Fig. 5, NDWI [see Fig. 5(b)] and MSI [see Fig. 5(c)] can highlight the water and shadow, respectively. Homogeneity, entropy, and ASM feature present high contrast between water bodies and other background covers. Fig. 6 shows exhaustive distributions of different land covers in additional bands. There is a partial overlap of water and other surface covers in each

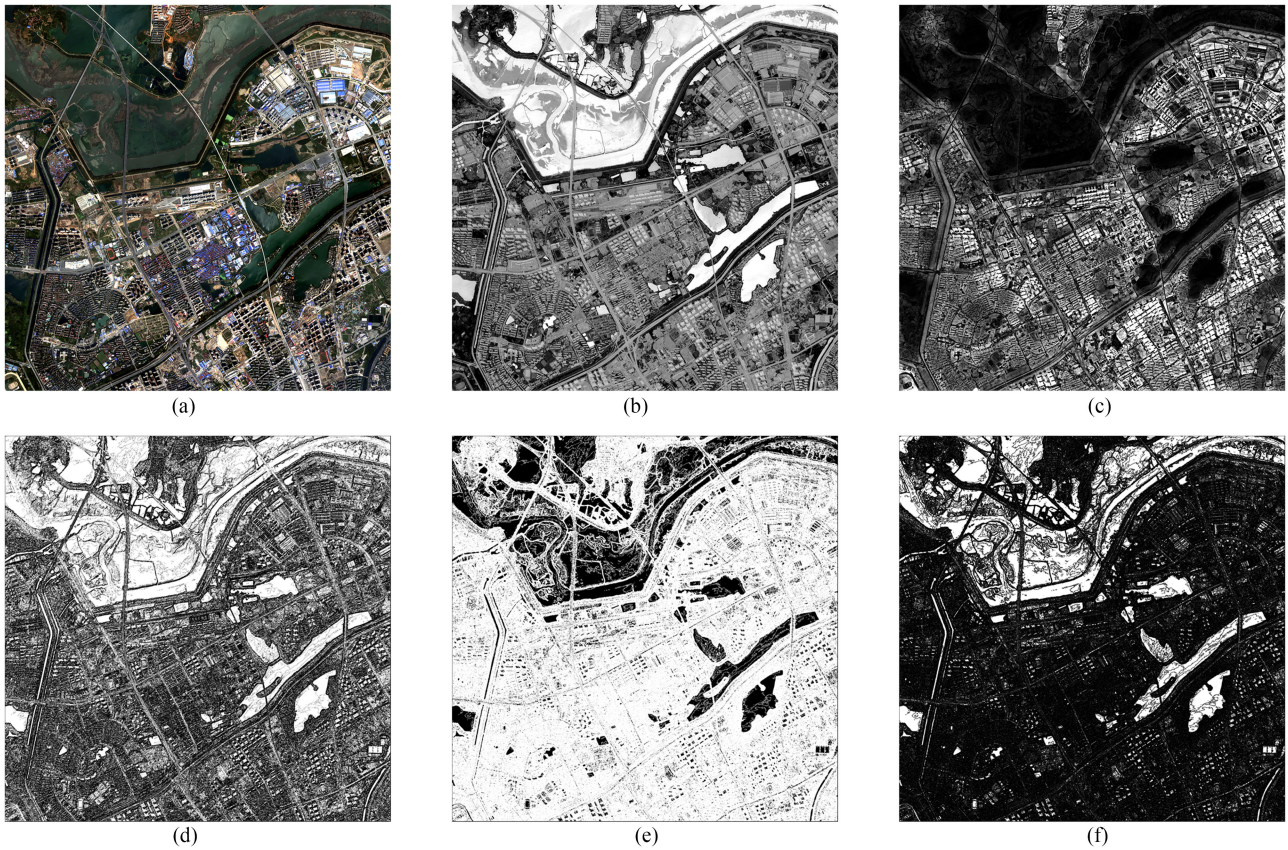

Fig. 5. Subscene example in Wuhan shows the differences in each feature between water and other background covers. (a) True-color image. (b) and (c) NDWI and MSI feature images. (d), (e), and (f) Homogeneity, entropy, and ASM features derived from GLCM.

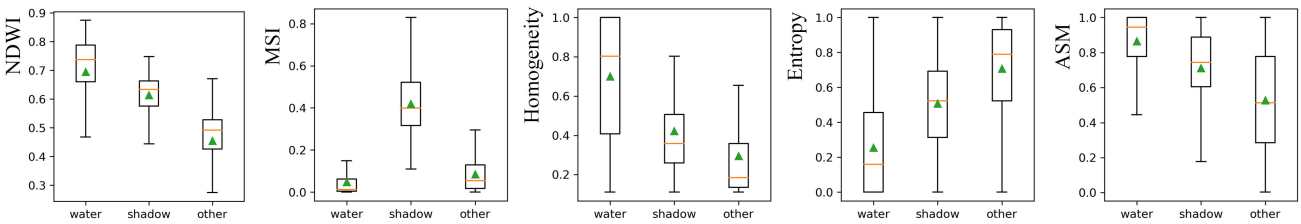

Fig. 6. Distributions of various features of different land cover types in Fig. 5 subscene. Each box plot uses horizontal lines to explain the location of the upper quartiles, median and lower quartiles, while two whiskers are the maximum and minimum values. The green triangle is the mean value of the data.

band, which indicates that the complete water body cannot be extracted only by a single band.

C. PNCEM

Sparse target detection algorithms are often applied in hyperspectral imagery, such as orthogonal subspace projection (OSP) [61], CEM [62], and matched filter (MF) [63]. The OSP algorithm requires complete knowledge of the image, including targets and background, which may be unrealistic in practice. CEM algorithm is a linear filter to minimize the output energy of the image. It focuses only on the desired target and has a better performance than OSP in eliminating unknown signals and suppressing noise [64]. MF detector has similar mathematical expressions as CEM, which needs the data to be centralized first. Adding linearly irrelevant bands has been proved to be beneficial to improve the performance of the CEM algorithm [49], which makes it possible for CEM to detect sparse targets on multispectral images. Nevertheless, CEM is a single target detection algorithm without the undesired target eliminated, which may cause extreme misclassification in urban surface water mapping. Despite introducing additional bands, shadows and other low albedo objects are still the main interferers. In order to resolve this dilemma, we design a multiple-target detector PNCEM, which can highlight the water bodies while suppressing the shadows. The main idea of PNCEM is described as follows.

Assuming $\{r_1, r_2, \cdots, r_N\}$ is a multispectral image with the total number of pixels N. Each pixel $r_i = (r_{i1}, r_{i2}, \cdots, r_{iL})^T$ is a vector with L bands, $1 < i < N$. Suppose d is the target

spectral signature to be detected with known prior information. It is under the condition of satisfying the following constraints

$$d^T w = \sum_{l=1}^{L} d_l w_l = 1 \tag{4}$$

where $w = (w_1, w_2, \ldots, w_L)^T$ is the L-dimension column vector consisting of filtering coefficient $\{w_1, w_2, \ldots, w_L\}$. Assume y_i is the output of the FIR filter of r_i

$$y_i = \sum_{l=1}^{L} w_l r_{il} = w^T r_i = r_i^T w \tag{5}$$

then, for all inputs $\{r_1, r_2, \ldots, r_N\}$, the average energy output by the filter is

$$E = \frac{1}{N} \sum_{i=1}^{N} y_i^2 = \frac{1}{N} \sum_{i=1}^{N} \left(r_i^T W\right)^T \left(r_i^T W\right)$$
$$= w^T \left(\frac{1}{N} \sum_{i=1}^{N} r_i r_i^T \right) w = w^T R w \tag{6}$$

where $R = \frac{1}{N} \sum_{i=1}^{N} r_i r_i^T$ is an autocorrelation matrix with $L \times L$ dimensions, CEM is defined as an optimization problem: $\min\{E\} = \min\{w^T R w\}$ subject to $d^T w = 1$.

According to the Lagrange multiplier method, the solution to the linear constrained optimization mentioned above is called the CEM detector, the filtering coefficient is w_{CEM}

$$w_{\text{CEM}} = \frac{R^{-1} d}{d^T R^{-1} d}. \tag{7}$$

The positive samples $d=\{d_1, d_2, \ldots, d_p\}$ denote the desired-targets signature, while the negative samples $u = \{u_1, u_2, \ldots, u_p\}$ denote the undesired-targets signature, and they can be obtained automatically by NDWI and MSI in Section III-B. Then, the PNCEM detector is defined by

$$w_{\text{PNCEM}} = \frac{1}{p} \left(\sum_{i=1}^{p} \frac{R^{-1} d_i}{d_i^T R^{-1} d_i} - \sum_{i=1}^{p} \frac{R^{-1} u_i}{u_i^T R^{-1} u_i} \right). \tag{8}$$

Multiple samples for each type of target make PNCEM robust to noise, while the performance of CEM is sensitive to the prior knowledge of the single desired target signature [65]. And due to the value of the superpixel being averaged by all the pixel values contained, PNCEM based on superpixel becomes more robust.

D. Threshold Determination

The determination of the optimal threshold is a crucial factor for water bodies extraction from the gray-scale image. The greater the separability of water and background, the more stable the binarization threshold. However, a fixed threshold is difficult to be applied to all environments, which may result in the overestimation or underestimation of water bodies. In this article, the adaptive threshold OTSU is adopted to extract water bodies automatically.

In general, OTSU can achieve good performance when the percentages of target and background are balanced, but the problem will stand out in the special case that the proportion of water is much lower than the background [66]. When this happens, the histogram of PNCEM will be concentrated in a peaked distribution around 0, while the PNCEM values of water bodies are close to 1 (see Fig. 7). In that case, the efficiency of binary threshold OTSU will be limited with poor accuracy. Therefore, it is necessary to estimate the prior information on the proportion of water bodies. As shown in Fig. 8, the PNCEM values of various types of water bodies are greater than 0.5, while other land covers are near 0, especially the shadow is suppressed to be less than 0. All patches can be divided into rich/poor-water patches by calculating the percentage of pixels with a value greater than 0.5. If the percentage is higher than 5%, which is referred to the condition in [67], the OTSU is adopted as an optimal threshold for the rich-water patch. Otherwise, it means

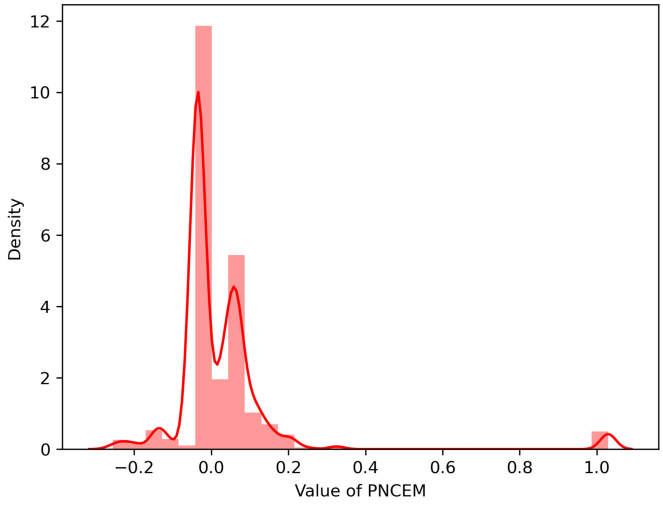

Fig. 7. Histogram of PNCEM in the poor-water patch.

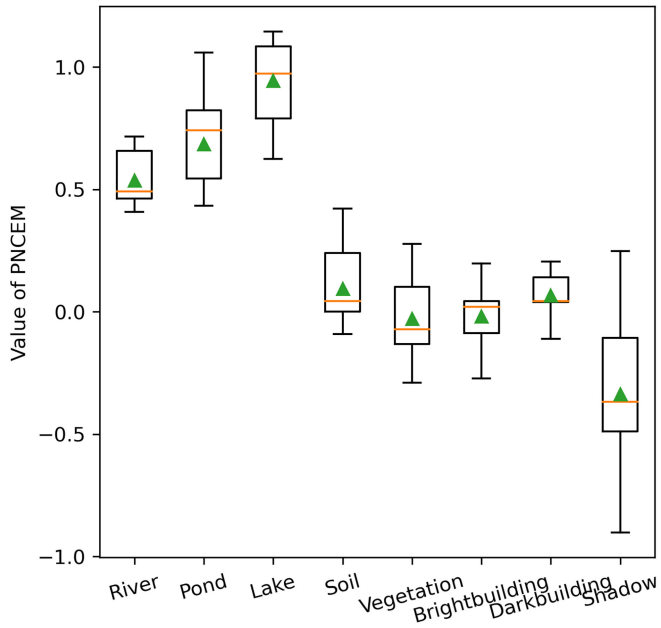

Fig. 8. Distribution of PNCEM values for different land covers.

TABLE II
CLASSES OF TRAINING SAMPLES FOR SUPPORT VECTOR MACHINE

	water	shadow	built-up	soil	vegetation
xiongan	697	189	1774	663	375
Wuxi	5092	278	8111	6568	3183
Wuhan	2697	377	3223	6431	9276
ningbo	673	262	1895	367	2122

that the proportion of water in the background is extremely uneven, and the empirical threshold of [0.5, 0.6] is suggested for the poor-water patch.

E. Accuracy Assessment

1) Extraction Accuracy: To evaluate the performance of SSWE, we compared the water bodies detection results with NDWI, SVM, and OWCEM [52]. Constant threshold zero and adaptive threshold OTSU are commonly used in NDWI [28], [35], [68], and we chose the one with higher accuracy in each study area. SVM has a high generalization performance and requires few training samples. The training samples of SVM were acquired by manually delineating the region of interest, including vegetation, soil, built-up, water, and shadow categories (see Table II). In this article, to avoid the influence caused by different input channels, the same nine spectral and spatial bands were utilized in OWCEM. OWCEM is a single desired target detector that only one target signature can be input, so we manually selected the most representative water sample in each study area for OWCEM. However, an appropriate threshold or threshold range is not suggested in OWCEM, so the optimum threshold with a balance between producer's accuracy (PA) and user's accuracy (UA) is adopted in this article. PA represents the probability of water ground truth reference being correctly classified, while UA represents the proportion of the categories classified as water that are correctly classified. Besides another two indicators of overall accuracy (OA) and the kappa coefficient are also introduced to describe the extraction accuracy

$$PA = \frac{TP}{TP + FN} \tag{9}$$

$$UA = \frac{TP}{TP + FP} \tag{10}$$

$$OA = \frac{TP + TN}{T} \tag{11}$$

$$Kappa = \frac{T \times (TP + TN) - \sum}{T \times T - \sum} \tag{12}$$

where $\sum = (TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN)$, and true positive (TP), false negative (FP), false positive (FN), and true negative (TN) are the statistics of the confusion matrix, which represent correctly detected, missing detected, false detected water pixels, and correctly detected nonwater pixels.

2) Separability Analysis: PNCEM is designed to highlight the difference between water and other urban surface covers, especially since the shadow is often confused with water. The threshold segmentation method relies on the separability between water and nonwater, while high separability results in accurate water mapping with a stable and suitable threshold. In this article, multiple targets CEM (MTCEM) [69], adaptive cosine estimator (ACE) [70], target-constrained interference-minimized filter (TCIMF) [71], OWCEM, multiple-targets inequality CEM (MTICEM) [69], and NDWI were selected for comparative analysis. It is worth noting that both ACE and OWCEM are single desired target detectors that only one target signature can be input, while MTCEM and MTICEM consider multiple target features in their filter design, and TCIMF can suppress the undesired targets. This comparative analysis was carried out at subscenes of each area in Fig. 1. The validation sets were obtained by randomly selecting pure samples from four experimental areas. Finally, we collected 2803 samples (water:928, shadow:759, others:1116) in Xiong'an, 4921 samples (water:1452, shadow:1892, others:1577) in Wuxi, 20109 samples (water:7316, shadow:7039, others:5754) in Wuhan, and 5405 samples (water:1668, shadow:1837, others:1900) in Ningbo.

Jeffries–Matusita (J–M) distance [72] is calculated to assess the separability among each urban composition in the above detectors. The larger the value of J–M distance, the higher the separability between two classes. In general, the J–M distance less than 1.00 represents poor separability, while greater than 1.38 indicates high separability. Formulas can be followed as:

$$\mathrm{JM} = 2\left(1 - e^{-B}\right) \tag{13}$$

$$B = \frac{1}{8}(\mu_i - \mu_j)^T \left[\frac{(C_i + C_j)}{2}\right]^{-1} (\mu_i - \mu_j) + \frac{1}{2}\ln\left[\frac{\left|\frac{C_i + C_j}{2}\right|}{\sqrt{|C_i|\,|C_j|}}\right] \tag{14}$$

where μ_i and μ_j are the mean vectors of two classes, and C_i and C_j are the covariance matrix, $|C_i|$ represents the determination of the covariance matrix C_i.

IV. EXPERIMENT RESULTS

A. Results of Water Mapping

In the experiment, four research sites are adopted to comparatively evaluate the performance of SSWE in different urban environments, and each GF-2 image is clipped into 3 × 3 even patches. The results are shown in Fig. 9. For better visual analysis and interpretation, the final water mappings are classified into four colors, and small areas in the yellow rectangle are utilized for detailed observation in Fig. 10.

According to the result, SSWE shows a great superiority over NDWI, SVM, and OWCEM in each study area. SSWE can remove the influence of building shadows, particularly in urban surface water detection. It can nearly extract the entire water bodies and has the same performance as SVM in a small river and ponds detection. For different types, such as eutrophic or turbid water, SSWE exhibits strong robustness due to its multiply

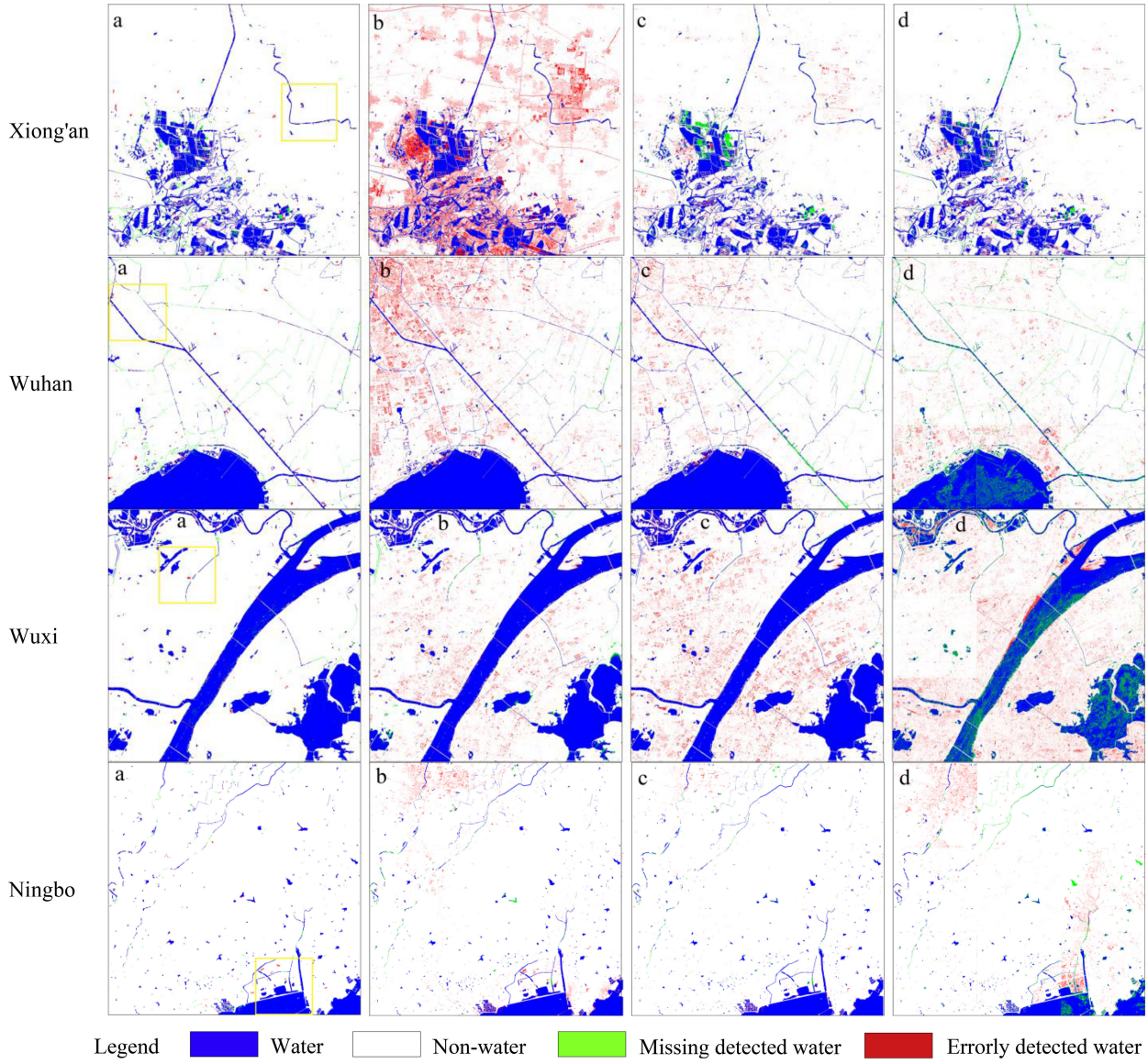

Fig. 9. Results of water bodies extraction in four study areas. (a) Proposed SSWE. (b) Normalized Difference Water Index. (c) Support vector machine. (d) Orthogonal subspace projection weighted CEM. The blue represents the correctly detected water pixels (TP), green represents the missing detected water pixels (FN), red represents the incorrectly detected water pixels (FP), and white is the correctly classified nonwater pixels (TN).

targets and object-level consideration. NDWI achieves significantly higher water extraction accuracy on natural surfaces than in artificial construction land (judging from the classification result of Ningbo), which is determined by the characteristics of the NIR band in NDWI. In general, the separability of NDWI results is poor, which leads to unstable thresholds and a large number of errors and omissions (judging from the classification result of Xiong'an and Wuxi). SVM is a kind of supervised classifier whose classification results are limited by sample selection. In the results of the SVM approach, only a small amount of water bodies remains undetected, but many shadows are mistakenly identified as water, especially in Wuhan and Wuxi. All methods but OWCEM perform well in large water bodies extraction, as OWCEM is a single target detector at the pixel level, it is sensitive to the target signature and cannot detect signatures similar to the target. Complex urban scenes and limited spectral characteristics result in the phenomenon of spectral heterogeneity being more prominent, which causes extreme commission and omission errors (in Wuhan, Wuxi, and Ningbo). Especially for OWCEM, different local desired signatures are selected in each patch, so there are apparent boundaries in the results of water mapping (such as Wuhan and Ningbo).

B. *Results of Accuracy Analysis*

Table III gives the water extraction accuracy of the four approaches at different experimental sites. We can see that SSWE exhibits the highest values of OA, UA, and KAPPA in the four study areas, with an average KAPPA coefficient of 0.942, the average values of OA (98.91%) and UA (95.05%) indicate a low probability of incorrect extraction. On the contrary, NDWI always presents a higher PA value but lower UA value, especially

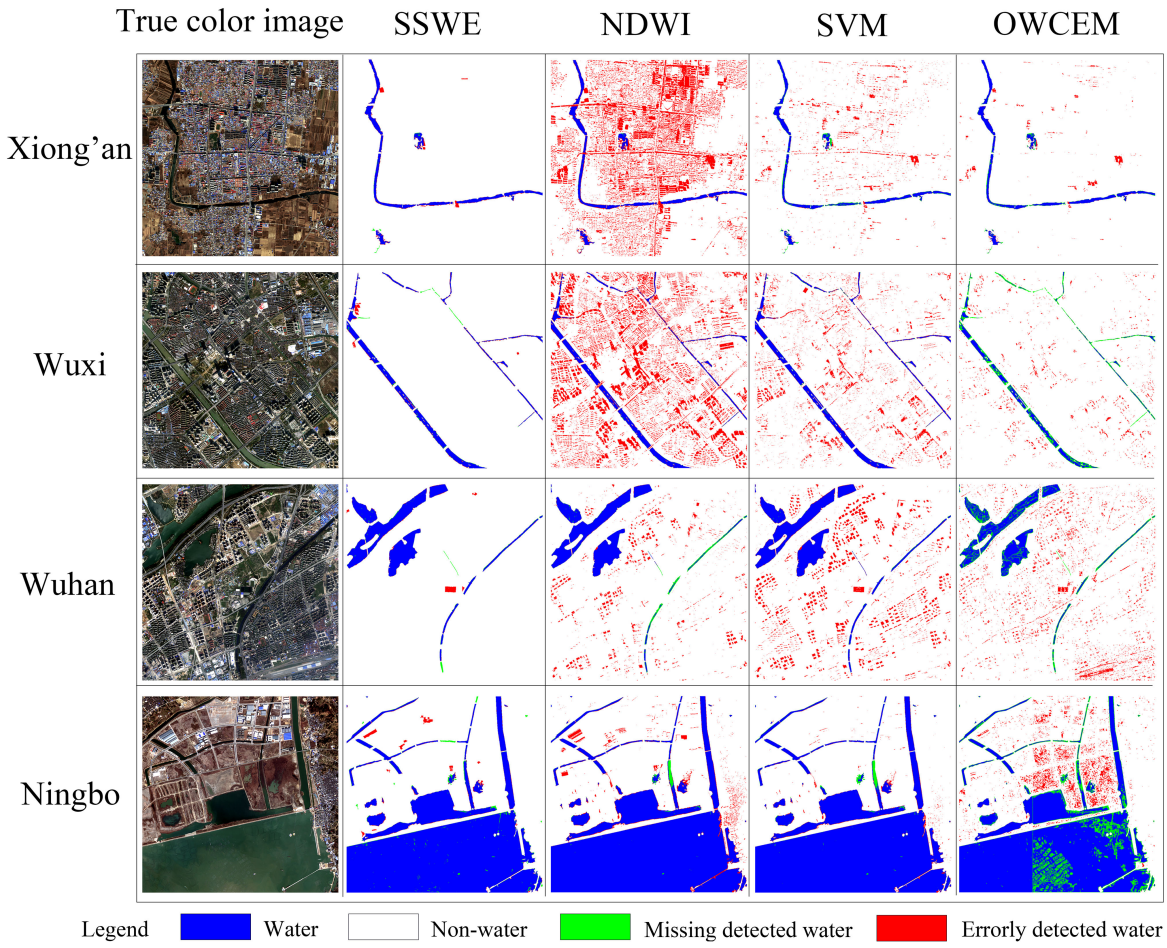

Fig. 10. Results of water bodies extraction of local areas in the yellow rectangle regions in Fig. 9.

in Xiong'an [PA (99.00%), UA (49.80%)]. The failure of NDWI in suppressing shadow noise of built-up areas results in a lot of false extractions. The performance of SVM is close to that of SSWE in Ningbo, where water bodies are clear rivers, reservoirs, and sea. Their Kappa coefficients differ by only 0.004. Although acceptable results can be achieved, the accuracy of SVM decreases in Wuhan and Wuxi with some types of turbid and eutrophic water, and the Kappa coefficients are 0.898 and 0.914. The instability of SVM is caused by the instability of manually selected samples in supervised classification. This phenomenon also appears in OWCEM and is even worse. OWCEM has the lowest accuracy among the three experimental areas, with an average KAPPA coefficient of 0.805, and only achieves good results in Xiong'an area.

C. Results of Separability Analysis

To demonstrate the advantage of the proposed PNCEM in distinguishing water bodies from shadows and other background covers, we have statistically analyzed the results of all detectors. The histograms of the results of each detector in different subscenes are plotted in Fig. 11. The J–M distance is used to measure the separability of water bodies from shadows and other background covers, as given in Tables IV and V. Considering that OWCEM is sensitive to the target signature, in this section, all sparse target detectors, including OWCEM, are at object-level, while only NDWI is based on pixel-level. All pixels contained in an object have the same gray value. Therefore, only the statistical histograms of NDWI have different gray values and are visually continuous in Fig. 11. The desired samples and undesired samples for sparse target detectors are automatically acquired by NDWI and MSI in Section III-C.

It can be seen from the comparative analysis that the JM_{ws} and JM_{wo} values of PNCEM are larger than the other detectors in Xiong'an and Ningbo. The MTCEM, ACE, TCIMF, OWCEM, and NDWI can hardly classify the water and shadow, while their JM_{ws} values are less than 1. MTCEM and ACE only focus on enhancing the desired targets without eliminating these undesired targets, this drawback is solved in TCIMF by mapping undesired targets into its null space. However, the performance of TCIMF will deteriorate sharply when the number of bands decreases. Consequently, it cannot even classify water and other background covers in VHR images, where its average JM_{wo} is 0.58. Noting that all methods, including PNCEM, show a poor separability between water and shadow in Wuxi, as some shadows are projected onto the water, which caused some confusion. Although a weighted autocorrelation matrix is introduced in OWCEM to minimize the output of background energy, it still ignores the strong similarity of undesirable goals. MTICEM replaces the equality constraints with inequality constraints to

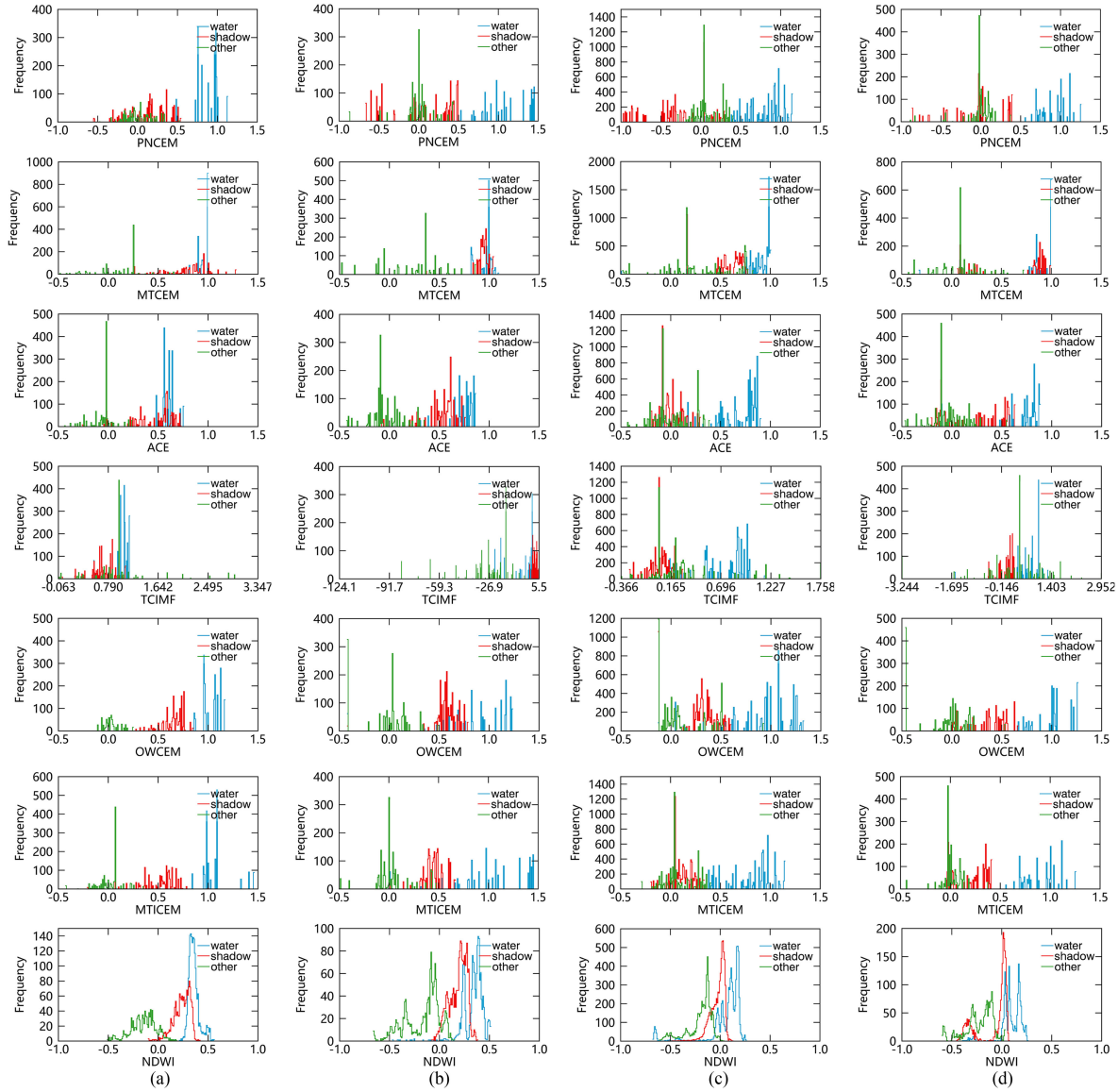

Fig. 11. Histograms of values of different detectors for each urban composition (water, shadow, and other background covers) in four subscenes in Fig. 1. (a) Xiong'an. (b) Wuxi. (c) Wuhan. (d) Ningbo. The seven rows, from top to bottom, are the PNCEM, MTCEM, ACE, TCIMF, OWCEM, MTICEM, and NDWI.

expand the solution space, and the advantages of MTICEM are significant in the excellent separability between water and other urban compositions, whose JM_{ws} and JM_{wo} values are close to that of PNCEM. However, it is time-consuming to solve the newly involved quadratic programming problem [73]. In summary, these results provide important insights that PNCEM has high separability and low time complexity, and it can overcome the limited number of bands of multispectral images. High separability of PNCEM leads to a stable binarization threshold and high accuracy results.

V. Discussion

A. Parameters Stability Analysis of SA-SNIC

The segmentation effect of water bodies is crucial to the final accuracy, so the different input parameters of SA-SNIC should be discussed first. Considering the various shapes of water bodies, the minimum value within the range of compactness factor recommended by SNIC is adopted to obtain better boundary adherence. The Thr_{size} is set to 50 based on the minimum size of the detected target. It can be seen from all curves in Fig. 12 that the number of superpixels increases slowly with the increase of the initial seeds number, while it decreases with the increase of the distance threshold Thr_{dis}. In the priority queue of SA-SNIC, new seeds are pushed after initial seeds are popped, so the number of superpixels is no longer determined by the number of initial seeds. However, large water bodies will be divided into multiple superpixels if the initial seeds are too many, which goes against the purpose of improving the efficiency of SNIC. So, it is suggested that the default number of initial seeds should not be too large. The higher the Thr_{dis} value, the fewer seeds will be newly generated at the cost of poorer boundary adherence. Therefore, the performance of SA-SNIC superpixel segmentation is mainly determined by the Thr_{dis}.

TABLE III
URBAN SURFACE WATER EXTRACTION ACCURACY RESULTS OF FOUR ALGORITHMS IN DIFFERENT STUDY AREAS. THE BOLD VALUES ARE THE MAXIMUM VALUES IN EACH SITE

Study area	Methods	PA	UA	OA	KAPPA
Xiong'an	SSWE	90.47%	**91.35%**	**98.19%**	**0.899**
	NDWI	**99.00%**	49.80%	89.93%	0.611
	SVM	86.11%	86.91%	97.31%	0.850
	OWCEM	88.73%	88.73%	97.75%	0.875
Wuxi	SSWE	94.85%	**96.75%**	**98.96%**	**0.952**
	NDWI	**98.01%**	73.96%	95.46%	0.817
	SVM	96.20%	89.11%	98.06%	0.914
	OWCEM	78.87%	87.42%	95.96%	0.806
Wuhan	SSWE	97.70%	**97.52%**	**98.93%**	**0.969**
	NDWI	97.33%	91.83%	97.47%	0.929
	SVM	**98.91%**	86.40%	96.28%	0.898
	OWCEM	76.97%	82.30%	91.16%	0.739
Ningbo	SSWE	95.69%	**94.57%**	**99.55%**	**0.949**
	NDWI	96.06%	85.99%	99.09%	0.903
	SVM	**96.23%**	93.40%	99.51%	0.945
	OWCEM	80.68%	81.43%	98.25%	0.801
Average	SSWE	94.68%	**95.05%**	**98.91%**	**0.942**
	NDWI	**97.60%**	75.40%	95.49%	0.815
	SVM	94.36%	88.96%	97.79%	0.902
	OWCEM	81.31%	84.97%	95.78%	0.805

TABLE IV
JM_{WS} VALUES OF SEVEN DETECTORS IN FOUR EXPERIMENTAL AREAS. JM_{WS} REPRESENTS THE J–M DISTANCE BETWEEN WATER BODY AND SHADOW. THE MAXIMUM VALUES OF JM_{WS} IN EACH COLUMN ARE BOLDED

	Xiong'an	Wuxi	Wuhan	Ningbo	Average
PNCEM	**1.71**	1.16	1.74	**1.45**	**1.52**
MTCEM	0.62	0.23	0.39	0.14	0.35
ACE	0.53	0.27	**1.79**	0.72	0.83
TCIMF	0.89	0.88	1.57	0.84	1.05
OWCEM	0.73	0.64	1.09	0.71	0.79
MTICEM	1.50	**1.27**	1.72	1.33	1.46
NDWI	0.66	0.56	0.39	0.49	0.53

TABLE V
JM_{WO} VALUES OF SEVEN DETECTORS IN FOUR EXPERIMENTAL AREAS. JM_{WO} REPRESENTS THE J–M DISTANCE BETWEEN THE WATER BODY AND OTHER SURFACE COVERS. THE MAXIMUM VALUES OF JM_{WO} IN EACH COLUMN ARE BOLDED

	Xiong'an	Wuxi	Wuhan	Ningbo	Average
PNCEM	**2.00**	1.52	1.46	**1.69**	1.67
MTCEM	1.34	1.49	0.45	1.42	1.18
ACE	1.99	1.74	1.58	1.66	**1.74**
TCIMF	0.71	0.77	0.36	0.47	0.58
OWCEM	1.68	1.41	1.30	1.38	1.44
MTICEM	**2.00**	**1.61**	**1.65**	1.60	1.72
NDWI	1.94	**1.61**	0.65	1.21	1.35

Nevertheless, when the Thr_{dis} is minimal, many new seeds will be grouped into small superpixels and eventually incorporated into the background, so the curves of Thr_{25} and Thr_{100} are very close.

As shown in Fig. 13, SSWE has high accuracy within the range of Thr_{dis} less than 100. Under the condition of minimizing the space distance, the value of Thr_{dis} is determined by the CIELAB color space distance. The 100 of Thr_{dis} is corresponds

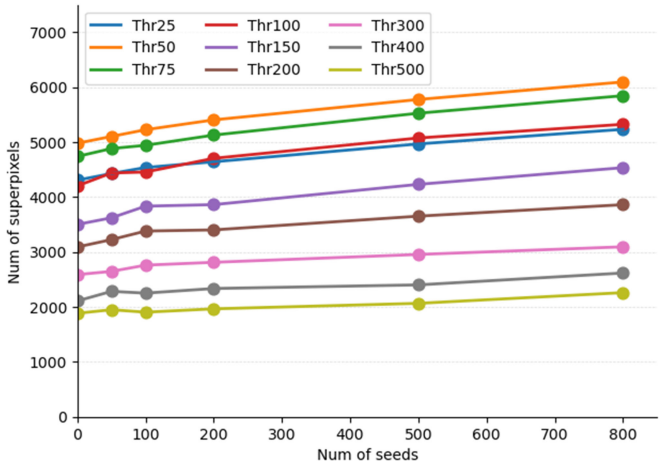

Fig. 12. Final number of superpixels generated by the different number of initial seeds and thresholds of distance.

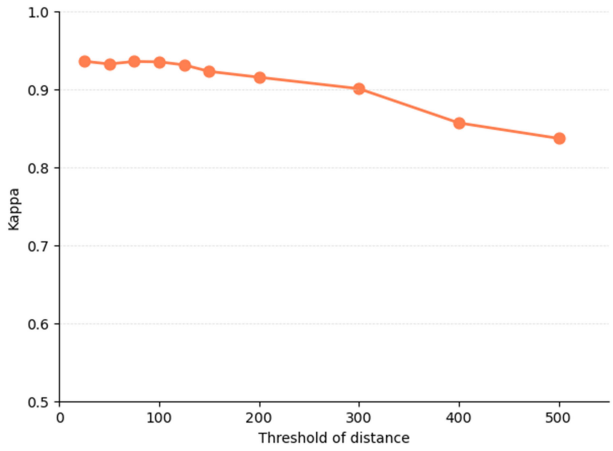

Fig. 13. Kappa coefficients of SSWE in a range of distance thresholds Thr_{dis} with the initial seed of 50.

to a maximum of 25 gray-scale differences in the NIR band (8 bytes), which is consistent with the visual difference of the boundary. When the value of Thr_{dis} is greater than 100, different land covers can easily be divided into the same superpixel. As shown in Fig. 14(a), the mixed land and water areas are most likely to be mixed with water bodies due to their similar spectrum and adjacent spatial distance. Besides, in the urban area, dark roads, dark buildings, and other low albedo features can also be confused with water bodies, resulting in poor performance of boundary adherence [see the red areas in Fig. 14(b)–(d)]. When the water superpixel contains other background covers, the average spectrum of the water superpixel differs from that of the water bodies, which lead to missed detection of water bodies [see the green areas in Fig. 14(a) and (b)]. The segmentation of shadow areas also faces the same problem, but its segmentation results do not interfere with the accuracy of water body extraction [see Fig. 14(c) and (d)]. In general, when Thr_{dis} is greater than 100, the higher value of the Thr_{dis} , the worse the accuracy of SSWE, so the recommended value of Thr_{dis} is around 100.

Fig. 15 shows the comparative segmentation results of two superpixel methods. The segmented boundary is depicted in

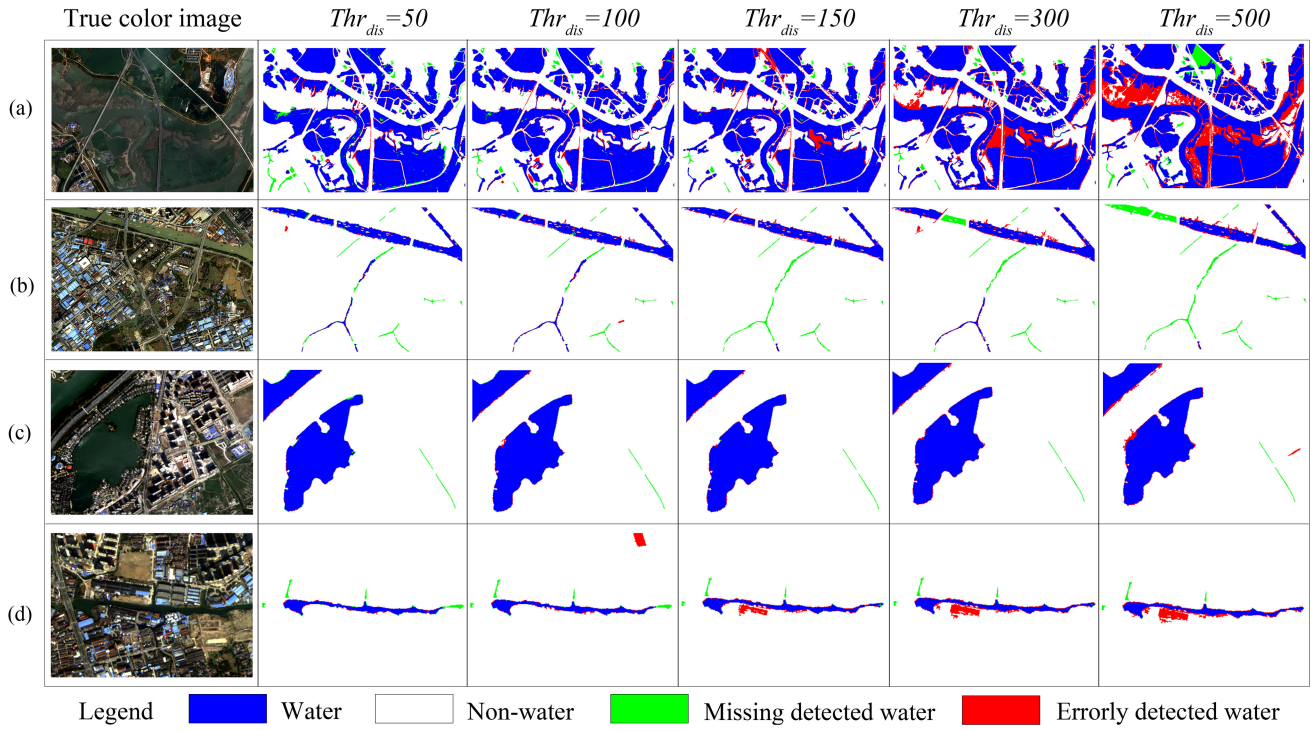

Fig. 14. Water bodies extraction results with different distance thresholds of Thr_{dis} in subimages.

red lines. It can be seen that superpixels in SNIC exhibit more compact with the approximate size, while complete water bodies are segmented into many parts with poor boundary adherence. But, water bodies in SA-SNIC can retain their integrity within various scales. Although the high spectral heterogeneity leads to fragmented segmentation results in urban surface land, it would not cause the misclassification of water bodies. On the contrary, the dense distribution of urban surface land makes water bodies become sparse at the object level, which just meets the statistical low probability requirements of sparse target detection. In general, SA-SNIC has the following advantages: good boundary adherence; no need to merge after the segmentation; and scale-adaptive segmentation. The SA-SNIC segmentation method can be applied not only to the water body, but also to other homogeneous features.

B. Contribution Analysis of Channels/features

The results of accuracy analysis show that SSWE can achieve good performance with the assistance of multiple features. But sometimes, certain features may not be easily accessible. Whether SSWE can still maintain good performance under the condition of missing some features is unclear. Therefore, we did ablation experiments in this section to evaluate the importance of each feature. The accuracy results are given in Table VI. Features are sequentially added as input bands into PNCEM in the order in which they are generated. In Fig. 16, VIS + NIR represents original spectral bands in GF-2. NDWI extracting positive samples is input first. +NDWI indicates NDWI is added as a new band to the existing bands. Then, MSI needs to be calculated to get negative samples. +MSI means adding MSI after NDWI has been added, where there are six bands in total. Finally, +GLCM represents a total of nine feature bands are added, including three components of the texture feature GLCM: homogeneity, entropy, and ASM.

As given in Table VI, the impact of each feature on the classification accuracy is positive, and the accuracy increases

TABLE VI
ACCURACY FOR DIFFERENT COMBINATIONS OF FEATURES USED BY SSWE IN EACH STUDY AREA. THE SYMBOL "√" INDICATES THAT THE FEATURE HAS BEEN ADDED, AND THE BOLD VALUES ARE THE MAXIMUM VALUES IN EACH SITE

Study area	Features				OA	KAPPA
	VIS+ NIR	NDWI	MSI	GLCM		
Xiong'an	√				94.89%	0.751
	√	√			95.95%	0.800
	√	√	√		96.09%	0.807
	√	√	√	√	**98.19%**	**0.899**
Wuxi	√				98.59%	0.936
	√	√			98.70%	0.942
	√	√	√		98.88%	0.948
	√	√	√	√	**98.96%**	**0.952**
Wuhan	√				98.47%	0.956
	√	√			98.69%	0.962
	√	√	√		98. 81%	0.966
	√	√	√	√	**98.93%**	**0.969**
Ningbo	√				99.05%	0.901
	√	√			99.29%	0.923
	√	√	√		99.34%	0.927
	√	√	√	√	**99.55%**	**0.949**

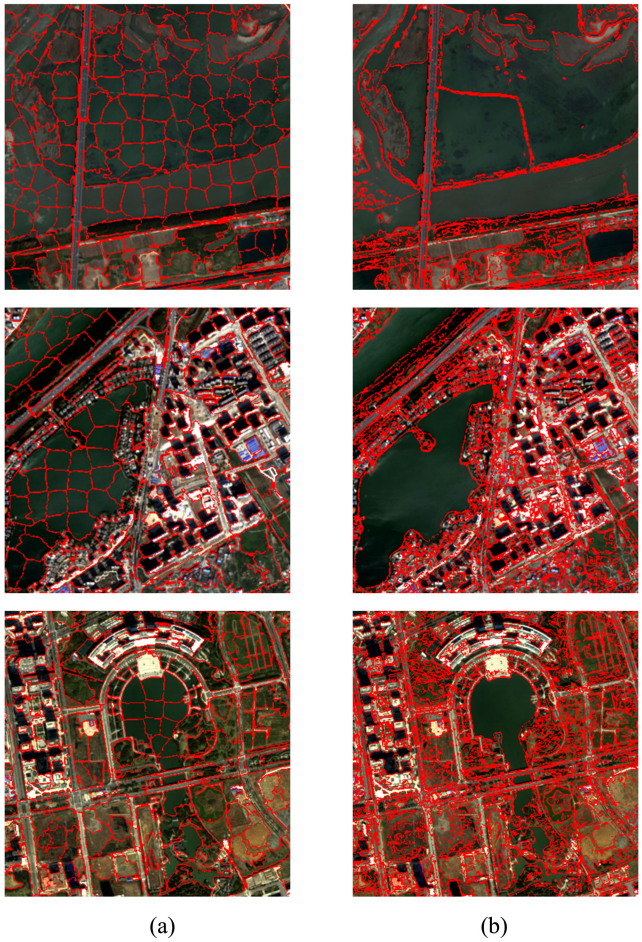

Fig. 15. Segmentation results of different superpixel algorithms. (a) SNIC. (b) SA-SNIC.

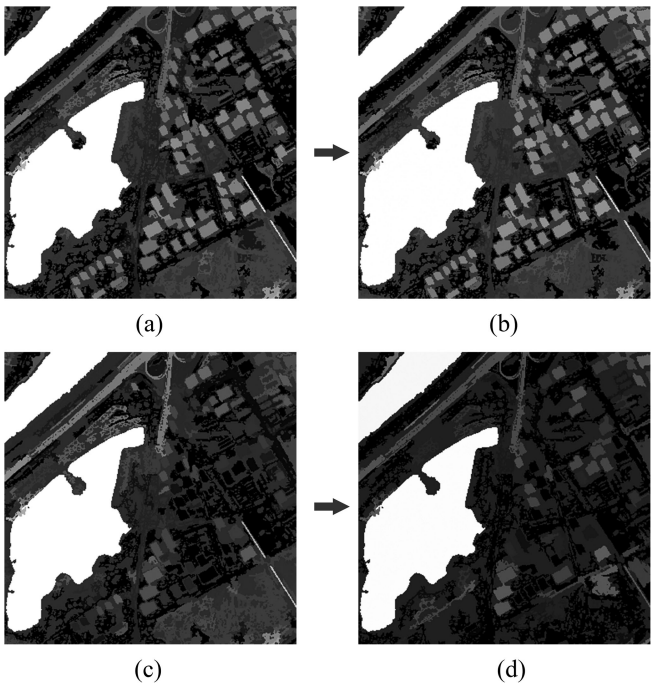

Fig. 16. (a) Filtering results of different combinations of features used by SSWE in the subscene of Wuhan. (b)–(d) All adding new features based on the previous step. (a) VIS+NIR. (b) +NDWI. (c) +MSI. (d) +GLCM.

TABLE VII
J–M DISTANCE FOR DIFFERENT COMBINATIONS OF FEATURES BETWEEN WATER BODIES, SHADOWS, AND OTHER LAND COVERS. THE MAXIMUM VALUES OF JM IN EACH ROW ARE BOLDED

	VIS+NIR	+NDWI	+MSI	+GLCM
JM_{ws}	1.31	1.29	1.54	**1.63**
JM_{wo}	0.77	0.91	1.10	**1.41**
JM_{so}	0.79	**0.84**	0.14	0.14

with the number of added bands. Typically, the effect of each feature on the accuracy improvement is affected by the surface coverage of the area. For example, due to the low proportion of the shadow areas, the accuracy is less improved after adding MSI in Xiong'an and Ningbo regions, where the kappa coefficient is improved by 0.007 and 0.004, respectively. The original four bands have achieved high accuracy in Wuhan and Wuxi regions, even higher than other methods, so the subsequent addition of new features does not improve accuracy significantly. Nevertheless, it can be seen from Fig. 16(c) and (d) that the suppression effect on the shadow is very obvious, and shadows gradually blend into the background after filtering. Especially after the introduction of MSI, the separability of water and shadow $\mathrm{JM_{ws}}$ increased sharply (see Table VII). Meanwhile, shadows start to mix with other land covers, where the $\mathrm{JM_{so}}$ value reaches 0.14. In general, NDWI can highlight water and shadows, MSI helps to discriminate shadows from other land covers, and water bodies are different from others due to their unique textural characteristics. It is worth noting that the added feature must contribute to distinguishing the target from other land covers to help improve the performance of SSWE, which is consistent with Figs. 5 and 6.

C. Threshold Stability Analysis of PNCEM

High separability of water and nonwater in gray-scale results will reduce the reliance on binarization thresholds. It is intuitively reflected in that the threshold obtained according to the separability index is usually close to the optimal threshold. The parts mentioned above have analyzed the advantages of SSWE in overall accuracy and separability, and here, we analyze the stability of the recommended threshold based on reference maps.

The result of each research site is clipped into 3×3 patches of the same size, and patches are numbered 01–09 from left to right and top to bottom. As shown in Fig. 17, in rich-water patches, almost all OTSU thresholds obtain high kappa coefficient values at four sites, and their values are close to the optimal values. The SSWE remains good performance in a range of thresholds near the OTSU value, which also confirms the high separability of the SSWE on the flip side. In poor-water patches, an empirical threshold is more suitable for water extraction. The positive samples taken by NDWI can cover most types of the water body, while some water types have not been exacted in rich-water patches. So, the average output values of water bodies in poor-water patches are slightly higher than that of rich-water patches. The empirical threshold is near 0.5, such as Xiong'an (0.6) and

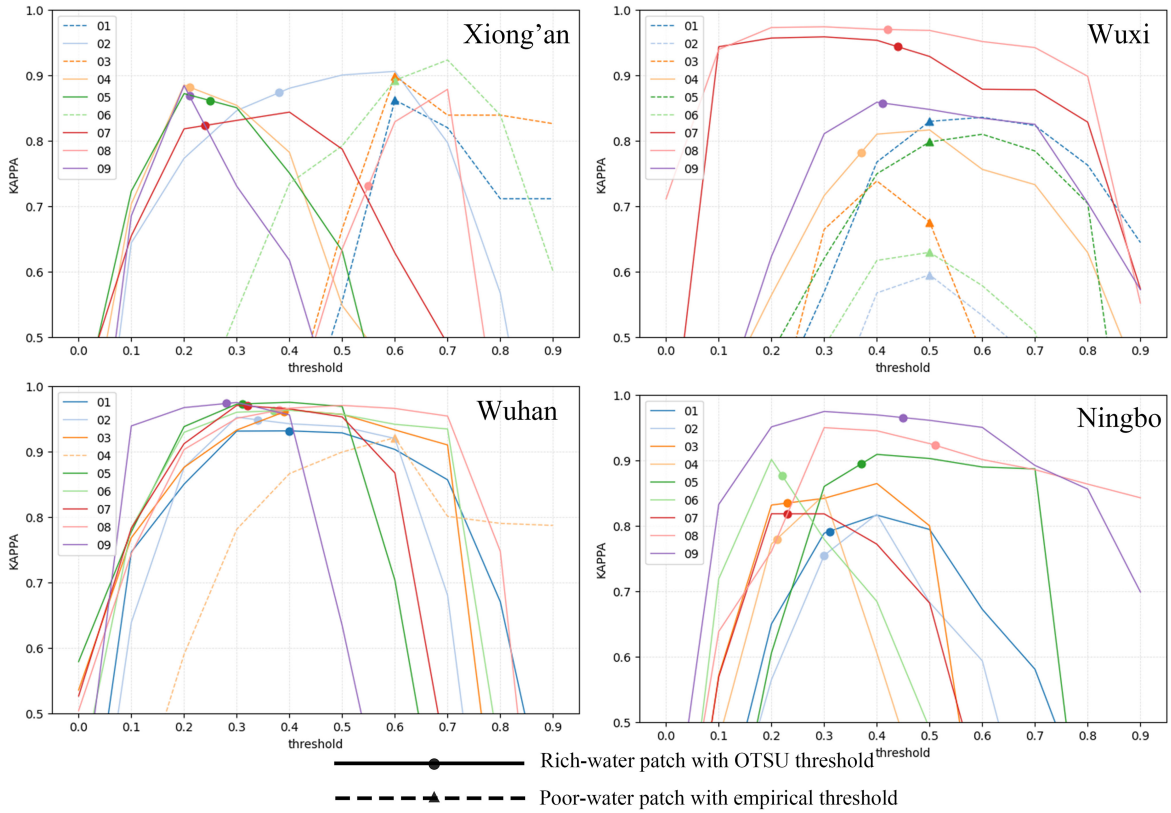

Fig. 17. Curve of SSWE accuracy at four experimental areas of the suggested threshold. Each image is clipped into 3 × 3 patches and numbered 01–09.

Wuxi (0.5), while OTSU thresholds in rich-water patches are less than the empirical threshold. However, the suggested thresholds do not always perform best, like 08-patch in Xiong'an, the kappa value of the OTSU threshold is 15% lower than that of the optimal threshold, which is affected by some mixed farmland of water and land. Although optimal kappa coefficients of the poor-water patches (02, 03, 06) in Wuxi are lower than 0.75, they have little effect on the accuracy of the whole image due to the small proportion of water bodies.

VI. Conclusion

The main purpose of this article is dedicated to solving the problem of misclassification of urban surface water mapping from VHR images. Based on the homogeneous characteristic of water bodies, the main idea of SSWE is to make large-scale water bodies become sparsely distributed targets at the object level by scale-adaptive segmentation, and then employ a sparse target detector to map the water bodies from the complex urban surface. To achieve this goal, the SNIC algorithm is improved to obtain object-oriented segmentation results, and band expansion increases the spectral, spatial, and derived information to separate water bodies from other land covers better. Finally, the PNCEM is designed to detect the distribution of sparse water objects. The experimental results indicate that: SSWE can obtain the highest accuracy of UA, OA, and kappa coefficient than NDWI, SVM, and OWCEM, while the PA of SSWE is slightly less than that of NDWI due to the severe overestimation of NDWI. For separability analysis among water bodies, shadows, and other land covers, PNCEM is superior to MTCEM, ACE, TCIMF, OWCEM, and NDWI. In addition, PNCEM is easier to implement compared with MTICEM. The stability analysis of SSWE further shows that the input parameters and suggested threshold are applicable. In general, SSWE exhibits good adaptability in complex urban scenes even with limited spectral bands, and its accuracy will increase with the addition of effective features. In further study, we will search for more unexplored features of VHR images, and we do believe that the strategy of SSWE would have great potential to map other compositions with high homogeneity from VHR images.

References

[1] G. H. Allen and T. M. Pavelsky, "Global extent of rivers and streams," *Science*, vol. 361, no. 6402, pp. 585–588, 2018, doi: 10.1126/science.aat0636.
[2] S. W. Cooley, J. C. Ryan, and L. C. Smith, "Human alteration of global surface water storage variability," *Nature*, vol. 591, no. 7848, pp. 78–81, 2021.
[3] X. Wang *et al.*, "Gainers and losers of surface and terrestrial water resources in China during 1989–2016," *Nature Commun.*, vol. 11, no. 1, Art. no. 3471, 2020.
[4] Y. Byun, Y. Han, and T. Chae, "Image fusion-based change detection for flood extent extraction using bi-temporal very high-resolution satellite images," *Remote Sens.*, vol. 7, no. 8, pp. 10347–10363, 2015.
[5] S. R. Proud, R. Fensholt, L. V. Rasmussen, and I. Sandholt, "Rapid response flood detection using the MSG geostationary satellite," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 4, pp. 536–544, 2011.
[6] Z. Shao, H. Fu, D. Li, O. Altan, and T. Cheng, "Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation," *Remote Sens. Environ.*, vol. 232, 2019, Art. no. 111338.

[7] Y. Tian, H. Chen, Q. Song, and K. Zheng, "A novel index for impervious surface area mapping: Development and validation," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1521, [Online]. Available: https://www.mdpi.com/2072-4292/10/10/1521

[8] L. Zhang, Y. Tian, and Q. Liu, "A novel urban composition index based on water-impervious surface-pervious surface (W-I-P) model for urban compositions mapping using landsat imagery," *Remote Sens.*, vol. 13, no. 1, 2021, Art. no. 3, [Online]. Available: https://www.mdpi.com/2072-4292/13/1/3

[9] X. Chen, H. Zhao, P. Li, and Z. Yin, "Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes," *Remote Sens. Environ.*, vol. 104, no. 2, pp. 133–146, 2006.

[10] R. C. Estoque and Y. Murayama, "Monitoring surface urban heat island formation in a tropical mountain city using landsat data (1987–2015)," *ISPRS J. Photogramm. Remote Sens.*, vol. 133, pp. 18–29, 2017.

[11] F. Chen, X. Chen, T. van de Voorde, D. Roberts, H. Jiang, and W. Xu, "Open water detection in urban environments using high spatial resolution remote sensing imagery," *Remote Sens. Environ.*, vol. 242, 2020, Art. no. 111706, doi: 10.1016/j.rse.2020.111706.

[12] X. Huang, C. Xie, X. Fang, and L. Zhang, "Combining pixel- and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2097–2110, May 2015, doi: 10.1109/JSTARS.2015.2420713.

[13] W. Sun, B. Du, and S. Xiong, "Quantifying sub-pixel surface water coverage in urban environments using low-albedo fraction from landsat imagery," *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 428. [Online]. Available: https://www.mdpi.com/2072-4292/9/5/428

[14] H. Xie, X. Luo, X. Xu, H. Pan, and X. Tong, "Automated subpixel surface water mapping from heterogeneous urban environments using landsat 8 OLI imagery," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 584. [Online]. Available: https://www.mdpi.com/2072-4292/8/7/584

[15] L. Xiong *et al.*, "Subpixel surface water extraction (SSWE) using landsat 8 OLI data," *Water (Basel)*, vol. 10, no. 5, 2018, Art. no. 653. [Online]. Available: https://www.mdpi.com/2073-4441/10/5/653

[16] Y. Zhou, J. Luo, Z. Shen, X. Hu, and H. Yang, "Multi-scale water body extraction in urban environments from satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4301–4312, Oct. 2014, doi: 10.1109/JSTARS.2014.2360436.

[17] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sens. Environ.*, vol. 115, no. 7, pp. 1603–1616, 2011.

[18] J. Franke, D. A. Roberts, K. Halligan, and G. Menz, "Hierarchical multiple endmember spectral mixture analysis (MESMA) of hyperspectral imagery for urban environments," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1712–1723, 2009.

[19] X. Wang *et al.*, "A robust multi-band water index (MBWI) for automated extraction of surface water from landsat 8 OLI imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 68, pp. 73–91, 2018.

[20] Y. Zhang, X. Liu, Y. Zhang, X. Ling, and X. Huang, "Automatic and unsupervised water body extraction based on spectral-spatial features using GF-1 satellite imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 927–931, Jun. 2019, doi: 10.1109/LGRS.2018.2886422.

[21] F. Yao, C. Wang, D. Dong, J. Luo, Z. Shen, and K. Yang, "High-Resolution mapping of urban surface water using ZY-3 multi-spectral imagery," *Remote Sens.*, vol. 7, no. 9, pp. 12336–12355, 2015, [Online]. Available: https://www.mdpi.com/2072-4292/7/9/12336

[22] Z. Zeng, D. Wang, W. Tan, and J. Huang, "Extracting aquaculture ponds from natural water surfaces around inland lakes on medium resolution multispectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 80, pp. 13–25, 2019.

[23] N. Mueller *et al.*, "Water observations from space: Mapping surface water from 25years of landsat imagery across Australia," *Remote Sens. Environ.*, vol. 174, pp. 341–352, 2016.

[24] Y. Deng *et al.*, "Spatio-temporal change of lake water extent in Wuhan urban agglomeration based on landsat images from 1987 to 2015," *Remote Sens.*, vol. 9, no. 3, 2017, Art. no. 270, doi: 10.3390/rs9030270.

[25] Y. Chen, R. Fan, X. Yang, J. Wang, and A. Latif, "Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning," *Water (Basel)*, vol. 10, no. 5, 2018, Art. no. 585. [Online]. Available: https://www.mdpi.com/2073-4441/10/5/585

[26] W. Jiang *et al.*, "Multilayer perceptron neural network for surface water extraction in landsat 8 OLI satellite images," *Remote Sens.*, vol. 10, no. 5, 2018, Art. no. 755, doi: 10.3390/rs10050755.

[27] Y. Yang *et al.*, "Landsat 8 OLI image based terrestrial water extraction from heterogeneous backgrounds using a reflectance homogenization approach," *Remote Sens. Environ.*, vol. 171, pp. 14–32, 2015.

[28] S. K. McFeeters, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, 1996.

[29] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3025–3033, 2006.

[30] G. L. Feyisa, H. Meilby, R. Fensholt, and S. R. Proud, "Automated water extraction index: A new technique for surface water mapping using landsat imagery," *Remote Sens. Environ.*, vol. 140, pp. 23–35, 2014.

[31] Q. Guo, R. Pu, J. Li, and J. Cheng, "A weighted normalized difference water index for water extraction using landsat imagery," *Int. J. Remote Sens.*, vol. 38, no. 19, pp. 5430–5445, 2017.

[32] F. Zhang *et al.*, "A simple automated dynamic threshold extraction method for the classification of large water bodies from landsat-8 OLI water index images," *Int. J. Remote Sens.*, vol. 39, no. 11, pp. 3429–3451, 2018.

[33] S. Jin *et al.*, "River body extraction from sentinel-2A/B MSI images based on an adaptive multi-scale region growth method," *Remote Sens. Environ.*, vol. 255, 2021, Art. no. 112297.

[34] N. Su, Y. Zhang, S. Tian, Y. Yan, and X. Miao, "Shadow detection and removal for occluded object information recovery in Urban high-resolution panchromatic satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2568–2582, Jun. 2016.

[35] X. Yang and L. Chen, "Evaluation of automated urban surface water extraction from Sentinel-2A imagery using different water indices," *J. Appl. Remote Sens.*, vol. 11, no. 2, pp. 26016, 2017.

[36] 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, doi: 10.1109/JSTARS.2011.2168195.

[37] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans Syst Man Cybern*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[38] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, 2010.

[39] S. Ye, R. G. Pontius, and R. Rakshit, "A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 137–147, 2018.

[40] M. D. Hossain and D. Chen, "Segmentation for object-based image analysis (OBIA): A review of algorithms and challenges from remote sensing perspective," *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 115–134, 2019.

[41] O. Csillik, "Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels," *Remote Sens.*, vol. 9, no. 3, 2017, Art. no. 243, doi: 10.3390/rs9030243.

[42] X. Yang *et al.*, "Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data," *Remote Sens. Environ.*, vol. 244, 2020, Art. no. 111803.

[43] S. Li, L. Ni, X. Jia, L. Gao, B. Zhang, and M. Peng, "Multi-scale superpixel spectral–spatial classification of hyperspectral images," *Int. J. Remote Sens.*, vol. 37, no. 20, pp. 4905–4922, Oct. 2016.

[44] H. Yu, X. Zhang, S. Wang, and B. Hou, "Context-based hierarchical unequal merging for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 995–1009, Feb. 2013, doi: 10.1109/TGRS.2012.2203604.

[45] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to State-of-the-Art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.

[46] Z. Dong, M. Wang, and D. Li, "A high resolution remote sensing image segmentation method by combining superpixels with minimum spanning tree," *Acta Geodaetica et Cartographica Sinica*, vol. 46, no. 6, pp. 734–742, 2017.

[47] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Interactive image segmentation by maximal similarity based region merging," *Pattern Recognit.*, vol. 43, no. 2, pp. 445–456, 2010.

[48] L. Gao, B. Yang, Q. Du, and B. Zhang, "Adjusted spectral matched filter for target detection in hyperspectral imagery," *Remote Sens.*, vol. 7, no. 6, pp. 6611–6634, 2015.

[49] X. Geng, L. Ji, K. Sun, and Y. Zhao, "CEM: More bands, better performance," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1876–1880, Nov. 2014, doi: 10.1109/LGRS.2014.2312319.

[50] N. Mazhari, A. M. Shafaroudi, and M. Ghaderi, "Detecting and mapping different types of iron mineralization in Sangan mining region, NE Iran, using satellite image and airborne geophysical data," *Geosciences J.*, vol. 21, no. 1, pp. 137–148, 2017.
[51] N. Zhang and K. Zhou, "Identification of hydrothermal alteration zones of the Baogutu porphyry copper deposits in northwest China using ASTER data," *J. Appl. Remote Sens.*, vol. 11, no. 1, pp. 1–19, Feb. 2017.
[52] L. Ji, X. Geng, K. Sun, Y. Zhao, and P. Gong, "Target detection method for water mapping using landsat 8 OLI/TIRS imagery," *Water (Basel)*, vol. 7, no. 2, pp. 794–817, 2015, [Online]. Available: https://www.mdpi.com/2073-4441/7/2/794
[53] X. Yang, Q. Qin, P. Grussenmeyer, and M. Koehl, "Urban surface water body detection with suppressed built-up noise based on water indices from sentinel-2 MSI imagery," *Remote Sens. Environ.*, vol. 219, pp. 259–270, 2018, doi: 10.1016/j.rse.2018.09.016.
[54] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[55] Z. Gharibbafghi, J. Tian, and P. Reinartz, "Modified superpixel segmentation for digital surface model refinement and building extraction from satellite stereo imagery," *Remote Sens.*, vol. 10, no. 11, pp. 1824, 2018, [Online]. Available: https://www.mdpi.com/2072-4292/10/11/1824
[56] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Comput. Vis. Image Understanding*, vol. 166, pp. 1–27, 2018.
[57] R. Achanta and S. Süsstrunk, "Superpixels and polygons using simple Non-iterative clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4895–4904. doi: 10.1109/CVPR.2017.520.
[58] M. Polak, H. Zhang, and M. Pi, "An evaluation metric for image segmentation of multiple objects," *Image Vis. Comput.*, vol. 27, no. 8, pp. 1223–1227, 2009.
[59] Y. Lv, W. Gao, C. Yang, and Z. Fang, "A novel spatial–spectral extraction method for subpixel surface water," *Int. J. Remote Sens.*, vol. 41, no. 7, pp. 2477–2499, 2020.
[60] C. Xie, X. Huang, W. Zeng, and X. Fang, "A novel water index for urban high-resolution eight-band worldview-2 imagery," *Int. J. Digit. Earth*, vol. 9, no. 10, pp. 925–941, 2016.
[61] J. C. Harsanyi and C. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994, doi: 10.1109/36.298007.
[62] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sens. Environ.*, vol. 59, no. 1, pp. 64–76, 1997.
[63] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002, doi: 10.1109/79.974724.
[64] D. Qian, R. Hsuan, and I. C. Chein, "A comparative study for orthogonal subspace projection and constrained energy minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1525–1529, Jun. 2003, doi: 10.1109/TGRS.2003.813704.
[65] R. Marwaha, A. Kumar, P. L. N. Raju, and Y. V. N. K. Murthy, "Target detection algorithm for airborne thermal hyperspectral data," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 40, no. 8, 2014, Art. no. 827.
[66] S. U. Lee, S. Y. Chung, and R. H. Park, "A comparative performance study of several global thresholding techniques for segmentation," *Comput. Vis., Graph., Image Process.*, vol. 52, no. 2, pp. 171–190, 1990.
[67] L. Li, H. Su, Q. Du, and T. Wu, "A novel surface water index using local background information for long term and large-scale landsat images," *ISPRS J. Photogramm. Remote Sens.*, vol. 172, pp. 59–78, 2021.
[68] H. Xie, X. Luo, X. Xu, H. Pan, and X. Tong, "Evaluation of landsat 8 OLI imagery for unsupervised inland water extraction," *Int. J. Remote Sens.*, vol. 37, no. 8, pp. 1826–1844, 2016.
[69] L. Zhu, L. Wang, L. Ji, W. Yang, and X. Geng, "Multiple targets inequality constrained energy minimization for multispectral imagery," *Infrared Phys. Technol.*, vol. 110, no. 2, 2020, Art. no. 103465.
[70] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 427–438, Feb. 2005, doi: 10.1109/TSP.2004.840823.
[71] H. Ren and C.-I. Chang, "Target-constrained interference-minimized approach to subpixel target detection for hyperspectral images," *Opt. Eng.*, vol. 39, no. 12, pp. 3138–3145, 2000.
[72] P. H. Swain and S. M. Davis, "Remote sensing: The quantitative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-3, no. 6, pp. 713–714, Nov. 1981, doi: 10.1109/TPAMI.1981.4767177.
[73] J. Gondzio, "Interior point methods 25 years later," *Eur. J. Oper. Res.*, vol. 218, no. 3, pp. 587–601, 2012.

Qingwei Liu received the Bachelor's degree in remote sensing science and technology in 2019 from the China University of Geosciences, Wuhan, China, where he is currently working toward the M.S. degree with the School of Photogrammetry and Remote Sensing.

His research interests include remote sensing image processing and computer vision.

Yugang Tian received the B.E. degree in surveying and mapping engineering and the M.S. degree in geodesy and geomatics engineering from Wuhan University, Wuhan, China, in 2000 and 2003, respectively, and the Ph.D. degree in physical geography from Beijing Normal University, Beijing, China, in 2006.

Since 2009, he has been an Associate Professor with the School of Geography and Information Engineering, University of Geosciences, Wuhan, China. From 2014 to 2015, he was a Visiting Scholar with the Department of Civil and Environmental Engineering, Cornell University in the United States. His research interests include urban and environmental monitoring, image processing, and pattern cognition.

Lihao Zhang received the B.E. degree in remote sensing science and technology and the M.S. degree in photogrammetry and remote sensing from the China University of Geosciences, Wuhan, China, in 2018 and 2021, respectively.

His research interests include remote sensing image processing and application.

Bo Chen received the Graduate degree in GIS and disaster prevention from Beijing Normal University, Beijing, China, in 2008, and the Ph.D. degree in water resources and hydrology from the University of Iowa, Iowa City, IA, USA, in 2014.

In 2014, he was with Beijing Normal University, and he is currently an Associate Professor with the Faculty of Geographic Science. His scientific interest concern surface hydrology and flood risk management. His current research interests include understanding of hydrologic processes across scales through field experiment, exploratory data analysis and modeling, and the assessment of and dealing with flood risk.