Article

Automatic Extraction of Built-Up Areas from Very High-Resolution Satellite Imagery Using Patch-Level Spatial Features and Gestalt Laws of Perceptual Grouping

Yixiang Chen [1], Zhiyong Lv [2], Bo Huang [3],*, Pengdong Zhang [1] and Yu Zhang [1]

[1] Department of Surveying and Geoinformatics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; chenyixiang@njupt.edu.cn (Y.C.); pengdong.zhang@njupt.edu.cn (P.Z.); 1317014316@njupt.edu.cn (Y.Z.)

[2] School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; zhiyonglyu@xaut.edu.cn

[3] Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong 999077, China

* Correspondence: bohuang@cuhk.edu.hk

check for updates

Abstract: Automatic extraction of built-up areas from very high-resolution (VHR) satellite images has received increasing attention in recent years. However, due to the complexity of spectral and spatial characteristics of built-up areas, it is still a challenging task to obtain their precise location and extent. In this study, a patch-based framework was proposed for unsupervised extraction of built-up areas from VHR imagery. First, a group of corner-constrained overlapping patches were defined to locate the candidate built-up areas. Second, for each patch, its salient textures and structural characteristics were represented as a feature vector using integrated high-frequency wavelet coefficients. Then, inspired by visual perception, a patch-level saliency model of built-up areas was constructed by incorporating Gestalt laws of proximity and similarity, which can effectively describe the spatial relationships between patches. Finally, built-up areas were extracted through thresholding and their boundaries were refined by morphological operations. The performance of the proposed method was evaluated on two VHR image datasets. The resulting average F-measure values were 0.8613 for the Google Earth dataset and 0.88 for the WorldView-2 dataset, respectively. Compared with existing models, the proposed method obtains better extraction results, which show more precise boundaries and preserve better shape integrity.

Keywords: high-resolution; satellite image; built-up area extraction; gestalt laws of grouping

1. Introduction

Built-up areas are important land-use types, especially in urban environments. The precise information about the locations, extents, and distributions of built-up areas is crucial for a wide range of applications [1,2], such as land use monitoring [3], population estimation and analysis [4,5], building detection [6], urban heat island analysis [7,8], and slum and poverty monitoring [9,10]. Rapidly developed high-resolution satellite sensors are currently providing images with a spatial resolution up to submeter (such as 0.6 m for Quickbird and 0.5 m for WorldView-2). Very high-resolution (VHR) satellite images have several advantages in describing the details of ground objects at a much finer scale. Thus, automatic extraction of built-up areas from VHR images has received increasing attention in recent years.

However, due to the uncertainty of object representation in feature space caused by spectral variability, the increased spatial resolution does not mean the improved accuracy for target classification [11,12], hence it poses a big challenge for built-up areas identification. To address this challenge, much work has been done by researchers to extract built-up areas from high-resolution satellite images. They achieved this aim mostly by utilizing spatial features such as textures and spatial structures [13–15]. Similar to the widely-known 'vegetation index', several indices, such as texture-derived built-up presence index (PanTex) [16,17], morphological building index (MBI) [18,19], and built-up areas saliency index (BASI) [20], have been proposed for the detection of built-up areas. However, these indices, which are derived by pixel-wise calculation in the neighborhood of each pixel, can only describe local textural patterns. They are not always effective in representing built-up areas, due to the spatial variability of textural patterns in VHR images. In addition, their basic assumption that there exists high local contrast between a building and its shadow is not always true, especially in complicated scenes.

Besides texture features, local geometrical structures, such as corners [21–23], edges [24,25], and line segments [26,27], are another kind of valuable spatial clues for built-up areas identification in VHR images, because, built-up areas are compound geographical objects containing various man-made structures and geometrical shapes (e.g., rectangular roof). Nevertheless, most of these methods achieve the final discrimination between built-up and non-built-up areas solely according to the distribution density of these local structures [28]. This is insufficient, especially in specific complex scenes where a large number of local structures can be detected in non-built-up classes (such as farmlands, roads, and trees). As a consequence, it often leads to false detection. In addition, due to the uneven spatial distribution of local structures inside a built-up area, missed detection can also occur in such locations that are with relatively sparse local features. This would make the extracted built-up area incomplete, e.g., with holes inside it.

Although the above-mentioned texture-based and local-structure-based approaches are effective in describing the local spatial patterns of built-up areas, the macro characteristics of built-up areas were usually overlooked or poorly handled. This is because these approaches used low-level pixel-wise processing for feature representation and discrimination, which cannot provide enough semantic information of built-up areas. Consequently, it often leads to an incorrect mapping from low-level features to high-level semantics of built-up areas. Therefore, false or missed detection easily occurs. On the other hand, in order to get more macro spatial patterns of built-up areas, a larger neighborhood around a central pixel is required, but this usually results in high computational costs.

To overcome these weaknesses of low-level pixel-based processing in spatial feature extraction and discrimination, mid-level image partition and feature representation have been adopted for built-up areas detection in recent years. State-of-the-art methods usually adopt superpixel-based [6,29,30] or block-based [31–33] processing strategy. In the superpixel-based approaches, the input image is over-segmented into homogeneous superpixels [34], which has shown many advantages over pixel-based algorithms; however, the over-segmentation would destroy the spatial patterns and semantics of built-up areas, and the quality of segmentation also has an important impact on the final detection results. The block-based approaches are very effective in modeling the complex textures and spatial patterns, especially for large-scale compound geographical objects, e.g., built-up areas. For example, Li et al. used multikernel learning to implement the block-level image interpretation, then fused the results using different block sizes to improve the detection of built-up areas in high-resolution satellite images [32]. In another study, block-based image features were represented in a multi-scale framework to capture the macro patterns of built-up areas [33]. Their work benefited from the use of block-based multi-scale information and training samples. Nevertheless, the hard subdivision of an image into non-overlapping blocks would also destroy the spatial structure of built-up areas. Moreover, it would generate a number of mixed blocks near the boundaries of built-up areas. This, thereby, resulted in large boundary errors in the detection results, especially when the block size was very large.

The objective of this study is to overcome the abovementioned shortcomings of existing methods in the representation, organization, and discrimination of spatial characteristics of built-up areas so as to improve the detection results. To this end, we address these issues from the perspective of visual perception. Inspired by perceptual grouping mechanism of human visual systems [35], we propose a patch-based framework for unsupervised extraction of built-up areas. In this framework, we employ Gestalt laws to describe the spatial relationships inside built-up areas. The major contributions of this paper are summarized as follows.

(1) The proposed framework takes corner-constrained overlapping image patches as primary calculation units, which can not only achieve the efficient locating of candidate built-up areas, but can also be of benefit to further spatial feature representation and semantic mapping of built-up areas. For each corner-constrained patch, in order to represent its salient spatial characteristics, a set of feature descriptors are constructed by integrating the low-level high-frequency wavelet coefficients, which can effectively characterize the local textural patterns of built-up areas.

(2) Inspired by visual perceptual mechanism, a patch-level saliency model is constructed by incorporating two Gestalt laws of proximity and similarity, which mean that elements tend to be perceived as a whole if they are close to each other and/or share similar visual features. For each patch, the proposed model measures its saliency according to both its own salient characteristics and its spatial relationship with surrounding patches. In this way, it can be enabled to capture the macro-patterns of built-up areas. To our knowledge, although Gestalt laws have been widely used in target recognition of natural images, little work has been done to apply them for built-up areas extraction in VHR images.

The rest of this paper is organized as follows. Section 2 presents the proposed approach. The experimental results and analysis are provided in Section 3. The discussion and conclusion are given in Sections 4 and 5, respectively.

2. Proposed Method

In this section, we introduce the details of our proposed method for automatic extraction of built-up areas from VHR satellite imagery. As illustrated in Figure 1, the proposed framework comprises the following steps. First, for an input VHR image, corners are detected and a group of corner-constrained patches are defined to locate the candidate built-up areas. Second, for each patch, its spatial characteristics are represented as a feature vector using integrated high-frequency wavelet coefficients. Then, the saliency of these patches is modeled by incorporating Gestalt laws of perceptual grouping. Finally, the extraction results are obtained by performing saliency map thresholding and refining the boundaries of built-up areas.
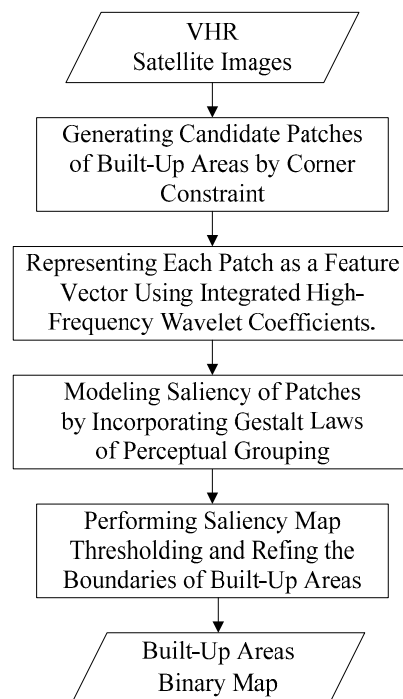
Figure 1. Flowchart of the proposed framework. VHR; very high-resolution.

2.1. Generating Candidate Patches of Built-Up Areas by Corner Constraint

Selective mechanism of human visual systems would guide visual attention to the target with salient features, so as to quickly complete its locating, organization, and recognition [36,37]. Built-up areas are compound man-made objects. In a large geographical scene, they usually present distinctive textures and structural characteristics from the background. Among them, corners are significant cues for the discrimination of geometric primitives [38]. The spatial distribution of local corners in VHR images provides an indication of the presence of built-up areas. Therefore, local corner features can be employed to infer the potential locations of built-up areas.

Up to now, many corner detection methods have been developed, and they as a whole can be divided into two categories: Edge-based methods (e.g., CSS [39]) and grayscale-based methods (e.g., Harris [40], SUSAN [41], and FAST [42]). In contrast, the methods in the former category are usually weak in robustness, as they largely depend on the result of edge extraction or image segmentation. Thus, in this study, we employed a method in the latter category, specifically the well-known Harris corner detector [40], to extract corner points in images. It has been proved to be very effective in corner detection for VHR images [21–23]. Then, a patch-based approach, which takes corners as constraint, was introduced to locate the candidate built-up areas. More specifically, given a VHR image, we take each detected corner point as the center to define an $S*S$ ($S = 2r + 1$, and r is a radius parameter) rectangular neighborhood (i.e., image patch), which would then be used as the operation unit for further processing tasks. Generally speaking, the distribution of corners are aggregated in built-up areas; therefore, these image patches defined by corners are also densely distributed. In particular, with an appropriate value for radius r, they are able to spatially overlap with each other and cover the whole built-up areas.

To illustrate this process more intuitively, an example is presented in Figure 2, in which Figure 2a is a raw VHR image with a spatial resolution of 1 m, Figure 2b is the detected corners using Harris operator, and Figure 2c presents the generated candidate patches of built-up areas. Obviously, the union of all these patches forms a covering of built-up areas, and their potential locations have been obtained in an efficient manner. Nevertheless, image patches generated in this way do not all belong to built-up areas, because some non-built-up classes (such as roads, farmlands, and trees) may also

contain a large number of corners (as shown in Figure 2b). Therefore, the main task for the next phase is to first identify built-up patches from these candidate patches and then organize them into complete targets of built-up areas.

| (a) | (b) | (c) |

Figure 2. Candidate patches of built-up areas: (**a**) raw image; (**b**) detected corners (highlighted with yellow dots); and (**c**) corner-constrained overlapping image patches (highlighted with red rectangular boxes).

2.2. Representing Each Patch as a Feature Vector Using Integrated High-Frequency Wavelet Coefficients.

In candidate patches, the built-up patches may exhibit similar characteristics to the other ones, which increase the difficulty for the identification of built-up patches. In VHR images, built-up areas usually present far richer spatial characteristics such as textures and edges than other object classes. These spatial signatures provide an effective criterion for the discrimination of building areas. Spatial features can be represented either in spatial domain or in frequency domain. Frequency-domain-based methods such as Gabor and wavelet have shown superior performance in modeling the complex textures and spatial patterns of built-up areas [14,15].

Thus, in our work, we employed frequency-domain-based methods for feature representation. Specifically, the multi-resolution wavelet [43], which is able to capture the multi-scale spatial information of built-up areas [15], was selected to measure the local textures and structures for an input image. Then, by integrating low-level wavelet features, each corner-constrained patch can be represented as a feature vector. More specifically, let $f(x, y)$ stand for a grayscale image, using multi-resolution wavelet transformation (WT), the kth level decomposition can be performed as follows:

$$\begin{cases} [A_k, H_k, V_k, D_k] = WT(A_{k-1}) \\ A_0 = f \end{cases}, \tag{1}$$

where $k = 1, 2\cdots, L$ ($L \geq 1$ is the maximum number of levels for WT decomposition process); H_k, V_k, and D_k represent the high-frequency coefficients of horizontal, vertical, and diagonal details, respectively, and A_k is the low-frequency approximation output. Note that when $k = L$, it is at the coarsest resolution [44].

In VHR images, built-up features usually exhibit higher frequency responses than non-built-up classes due to their rich texture details and edge structures. Therefore, the high-frequency wavelet coefficients are more informative than the low-frequency ones for built-up areas discrimination [15]. With this consideration, the high-frequency coefficients are adopted to represent the textural intensity and distribution patterns for each corner-constrained patch. Furthermore, in order to reduce the feature

dimension, the maximization operator is employed to integrate the multi-directional high-frequency coefficients. Let I_k denote the integrated result in the kth level, then

$$I_k = \max(H_k, V_k, D_k), \tag{2}$$

where max stands for the maximization operation, which is implemented in a pixel-wise manner. The maximization integration is quite effective in capturing the information of directionality of building features due to their artificial spatial arrangement.

Suppose L-levels decomposition are implemented for an input VHR image. To represent each corner-constrained patch as a feature vector, we propose the following two feature descriptors at patch level by integrating the pixel-level high-frequency coefficients.

$$ave = [ave(I_1), \ldots, ave(I_k), \ldots, ave(I_L)], \tag{3}$$

$$var = [\text{var}(I_1), \ldots, \text{var}(I_k), \ldots, \text{var}(I_L)], \tag{4}$$

where *ave* and *var* denote the two descriptors, I_k is the kth-level high-frequency coefficients within a patch, and $ave(I_k)$ and $var(I_k)$ are the mean and variance of these coefficients, respectively.

In this study, the two descriptors are combined to describe the spatial features of each corner-constrained patch. Let X denote the feature vector representing such a patch, then

$$X = [ave, var], \tag{5}$$

where both *ave* and *var* can measure the textural responses of a patch at multiple scales, thereby providing more discriminative features for the detection of built-up areas. In particular, the descriptor *ave* is a representation of the texture intensity while *var* is useful in describing the distribution patterns. In addition, according to our previous research [15], three levels of decomposition are suitable for built-up areas discrimination in VHR images. Therefore, in this study, L is set to three.

2.3. Modeling Saliency of Patches by Incorporating Gestalt Laws of Perceptual Grouping

The saliency map of built-up areas, which can provide a measure to the feature conspicuity at every location in a visual field [45], is quite useful for unsupervised extraction of built-up areas. In fact, saliency modeling is a simulation of selective mechanism of visual attention in the computer community [45–47]. This mechanism would guide visual attention to the targets with salient features, so as to quickly complete their locating, organization, and identification. In this process, by imposing organization on visual elements, perceptual grouping function of human visual systems play a significant role in target attention and recognition. This has been confirmed by relevant studies in the domains of neuroscience and cognitive psychology [48,49]. The principles of perceptual grouping were described by Gestalt psychology [50]. They have proposed a set of laws suitable for perceptual organization, such as proximity, similarity, continuation, closure, and symmetry, which indicates that the target in a scene is an organized and non-random structure [35].

In this study, we model the saliency of each corner-constraint patch by incorporating the Gestalt clues of perceptual grouping, including the laws of proximity and similarity, which can be described as follow [47,51]:

(1)　Proximity means that elements tend to be perceived as a whole if they are close to each other.
(2)　Similarity means that elements tend to be grouped together if they share similar appearance of visual features.

Since built-up areas are composed of a group of spatially connected salient patches, the saliency of a patch is not only determined by its own characteristics, but also closely related to those of its surrounding patches. In other words, if a non-salient patch is surrounded by highly salient patches, it should be assigned a high saliency value. Conversely, a patch surrounded by those with low saliency

values is more likely to belong to the background, even if this patch is salient; therefore, it should be assigned a low saliency value.

With this consideration, for each patch, we measure its saliency according to both its own salient characteristics and its spatial relationship with surrounding patches. Assume the salient characteristics of the *i*th patch and its *j*th adjacent patch can be represented by feature vector X_i and X_j, respectively, then the dissimilarity between them can be measured using the following distance formula:

$$d_{ij} = \left(X_i - \overline{X}\right)\left(X_j - \overline{X}\right)^T,$$ (6)

where $\overline{X}$ is the mean of feature vectors of all image patches, while T represents the transposition of a vector. Note that by subtracting the mean vector, the dissimilarity measure has global characteristics, which are beneficial to capture global saliency for better extraction of built-up areas through thresholding.

In addition to similarity, spatial proximity also contributes significantly to the overall saliency of a patch. In our work, the binary weight matrix $w_{ij}(d)$ is used to measure the proximity between the *i*th patch and the *j*th one, which is defined as follows: If the distance of central points between the *j*th patch and the *i*th one is not more than d, then $w_{ij}(d) = 1$; otherwise, $w_{ij}(d) = 0$. Herein, d is a distance parameter representing the degree of spatial proximity, and its setting would be discussed in the following sections.

Then, for the *i*th patch, we propose to employ the following formula to calculate its overall saliency value, which integrates the measures of both proximity and similarity between this patch and its neighboring patches.

$$v_i = \sum_{j=1}^{n} w_{ij}(d) * exp\left(-\frac{d_{ij}}{2\sigma^2}\right),$$ (7)

where v_i denotes the saliency value of the *i*th patch, n represents the total number of patches, and σ is an adjustable parameter (i.e., bandwidth) controlling the radial extent of Gaussian function, which would be tested in the experimental section. To calculate the spatial weight matrix $w_{ij}(d)$, the distance parameter d is set as follows:

$$d = 3\,\sigma,$$ (8)

This is inspired by three-sigma rule of Gaussian distribution, that is, for an approximately normal data set, the values within three standard deviation of the mean account for about 99.7% of the set [52]. According to our experience, this setting is quite effective for built-up areas extraction, and in particular, it can decrease by one parameter in the process of implementation of the algorithm.

In addition, Equation (7) can be further expressed as

$$v_i = exp\left(-\frac{d_{ii}}{2\sigma^2}\right) + \sum_{j=1,j\neq i}^{n} w_{ij}(d) * exp\left(-\frac{d_{ij}}{2\sigma^2}\right),$$ (9)

where the first term represents the feature saliency of the *i*th patch itself while the second term represents the saliency produced by its spatial relationship (i.e., proximity and similarity) with surrounding patches. It can be found by Equation (9) that for each patch, its surrounding patches dominate the overall saliency, which further demonstrate the fact that patches conforming to Gestalt laws of grouping have greater visual saliency, thereby making them easier to be correctly identified.

After each patch is assigned a saliency value, the saliency values of all patches are then normalized to be [0,1] in order to facilitate their comparison. Thus, a patch-level saliency map of built-up areas can be obtained.

2.4. Performing Saliency Map Thresholding and Refining the Boundaries of Built-Up Areas

Based on the saliency model above, a saliency value can be obtained for each corner-constrained image patch. The patches of built-up classes tend to show greater saliency values than other patches, because built-up areas have more salient textures and stronger spatial organization. Therefore, an appropriate saliency threshold could be used to distinguish between built-up patches and non-built-up patches. For threshold selection, although the trial-and-error method can be applied to determining the optimal value, it is very time-consuming and laborious. Hence, automatic threshold technique should be given priority so as to achieve the automatic extraction of built-up areas.

Previous studies have shown that Otsu method [53] is an automatic threshold technique, which, based on the standard of maximum inter-class variance, has shown superior performance in binary classification problems [54]. In this study, we adopt the Otsu method to binarize the saliency map of built-up areas. Before the implementation of this algorithm, all values in the saliency map are first transformed to be [0,255]. After thresholding, each built-up patch is labeled as 1, and the other patches are labeled as 0. Then, the union of all built-up patches constitutes the target of built-up areas.

The use of patch-based representation can also result in the serrated boundaries. However, compared with traditional methods based on non-overlapping blocks produced by hard partition of image, this phenomenon has been weakened by adopting corner-constraint patches in the proposed method. In order to further deal with this problem, morphological opening-and-closing operation [55] is employed to smooth the boundaries. One difficulty is the size selection of structuring elements in morphological operation. In this work, considering the size of corner-constraint patches, the size of structuring element (denoted by *SSE*) is defined as the radius of these patches, that is,

$$SSE = r \tag{10}$$

With this structuring element, the serrated boundaries can be well smoothed and refined, and accurate or approximately accurate boundaries can be obtained.

On the basis of the results shown in Figure 2, a further example is presented in Figure 3 to illustrate the process of built-up areas segmentation and boundary refinement. Figure 3a is a saliency map of built-up areas constructed using the proposed patch-based saliency model, Figure 3b is the segmentation result by thresholding the saliency map using the Otsu method, and Figure 3c shows the final extraction results. Thus, the built-up areas have been successfully extracted based on the proposed approach, and in particular, they preserve a complete shape and well-defined boundaries.
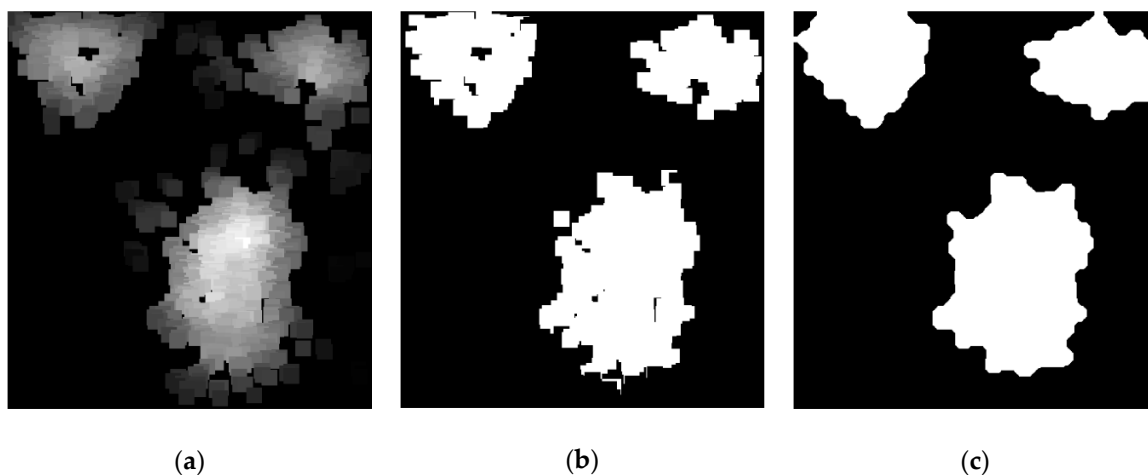

(**a**)   (**b**)   (**c**)

Figure 3. Built-up area segmentation and boundary refinement: (**a**) saliency map of built-up areas ($r = 9$, $\sigma = 15$); (**b**) segmentation result through Otsu's threshold; and (**c**) refinement result by morphological operators ($SSE = 9$).

3. Experiments and Analysis

3.1. Datasets and Evaluation Metrics

To evaluate the performance of the proposed method, several experiments were conducted on two VHR remotely sensed image datasets, which cover multiple test sites with various scene characteristics.

The first dataset contains six images downloaded from Google Earth. All of these images consist of three bands of RGB (i.e., red, green, and blue), with a spatial resolution of about 1 m. These images cover various geographical scenes, such as mountainous, industrial, and residential areas, and more detailed information about them is given in Table 1. Since the proposed method was designed for grayscale images, for each image in this dataset, we first converted its stacked RGB bands into a grayscale image and then implemented the proposed algorithm for built-up areas extraction.

Table 1. Information about the Google Earth dataset.

Image No.	Bands	Resolution	Size (Width × Height)	Scenes	Location
R1		0.91 m	1280 × 1280	Mountainous areas	Beijing, China
R2		0.91 m	1280 × 1024		
R3	RGB	0.99 m	1280 × 1280	Industrial areas	Dengfeng, China
R4		0.99 m	2048 × 1792	Urban suburbs and villages	
R5		1.0 m	1280 × 1280	Urban areas	Nanjing, China
R6		1.0 m	2048 × 1792		

The second one was from WorldView-2 satellite, which was launched on 6 October 2009. The dataset was constructed using a panchromatic image covering a certain region of Wuhan city, China. A total of 15 images taken from different sites of this region was used for performance evaluation. All of these images are with a resolution of 0.5 m, and their sizes are varying from 470 by 642 to 2274 by 1786 pixels, which cover different urban region scenes and contain various types of built-up areas so as to verify the robustness of our proposed method.

In order to evaluate the results of built-up areas extraction, we also constructed reference (ground-truth) datasets of the two datasets. For each test image, its reference data were obtained by visual interpretation. The built-up areas were first delineated manually to form vector polygons, which were then converted into a binary image. Although it is time-consuming and laborious to obtain these data, we try to ensure the reliability of results through multiple corrections by different interpreters. For the Google Earth reference dataset, it contains a total of 3,112,741 built-up pixels and 10,453,211 non-built-up pixels, while for the WorldView-2 reference dataset, the numbers of built-up and non-built-up samples (pixels) are 7,840,057 and 11,987,963, respectively.

The quantitative performance of the proposed approach on the two datasets is evaluated based on three widely used metrics: Precision P, recall R, and F-measure F_α as defined below, respectively [14,29,44].

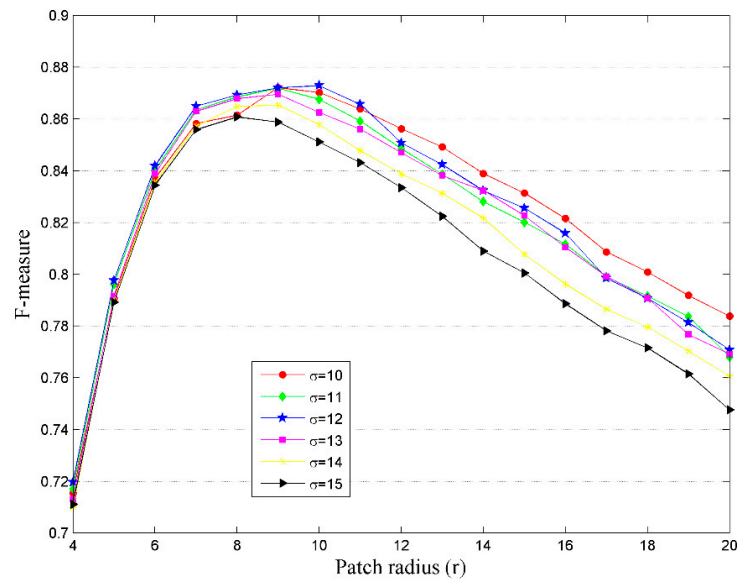$$P = \frac{TP}{TP + FP}, \tag{11}$$

$$R = \frac{TP}{TP + FN}, \tag{12}$$

$$F_\alpha = \frac{(1+\alpha)\cdot P\cdot R}{\alpha\cdot P + R}, \tag{13}$$

where TP, FP, and FN denote the numbers of true positive, false positive, and false negative detected pixels, respectively. The P and R indices are used to evaluate the detection results from the perspective of correctness and completeness, respectively. The F-measure (F_α) is an overall performance measure by the harmonic mean of P and R, where α is a positive parameter to balance the relative importance between the precision and the recall in performance evaluation. In our experiments, we set $\alpha = 1$, because precision and recall are equally important for the extraction of built-up areas [29].
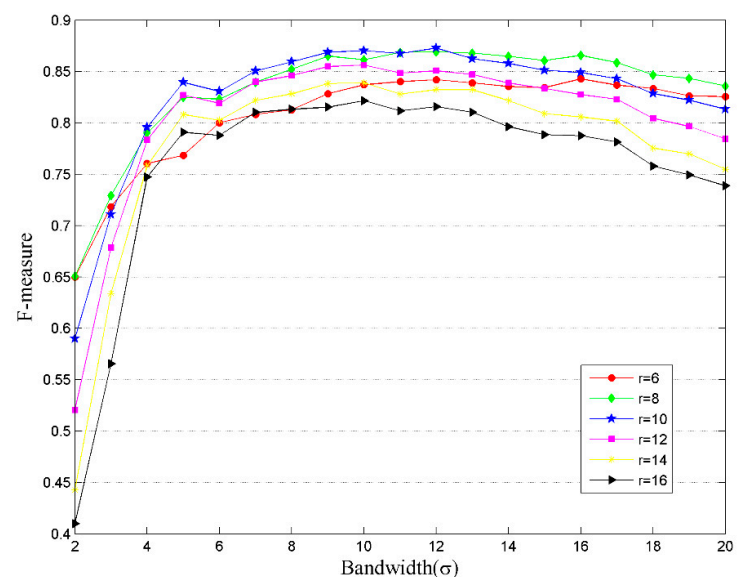
3.2. Tests on Parameter Values

The main parameters in our proposed method include the patch radius *r* and the bandwidth *σ* of Gaussian function in Equation (7). As previously mentioned, the first parameter is used to determine the primary calculation units while the second one is closely related to the influence range of an image patch; therefore, both of them play an important role in the final extracted results of built-up areas.

With different parameter settings, we conducted tests on one image from the first datasets by fixing one parameter and testing the other one each time. The resulting accuracy assessment curves are presented in Figure 4, where Figure 4a,b are the test results for parameter *r* and *σ*, respectively.

(**a**)

(**b**)

Figure 4. The result of parameter tests: (**a**) the change of F-measures as the patch radius *r* increases; (**b**) the change of F-measures as the bandwidth *σ* increases.

The patch radius has an important impact on the extraction results. For each fixed parameter σ (e.g., $\sigma = 10, 11, 12, 13, 14, 15$), as shown in Figure 4a, with the increase of patch radius r from a small value to a large one, the F-measure first increases rapidly to the peak of the curve, and then it begins to decline and maintains this trend. This can be explained by the fact that with a small patch radius, the built-up areas cannot be well covered by corner-constraint patches, whereas a large patch radius would result in more mixed image patches, especially in the boundary between built-up and non-built-up areas. These two situations would lead to the missed and false detection of built-up areas, respectively. For the test image, the appropriate value for parameter r can be determined according to the F-measure assessment results. The patch radius that makes the F-measure reach the maximum is the optimal. For example, when $\sigma = 12$, the optimal value for r is 10. Then this value (i.e., $r = 10$) can be used as an empirical value, and it can be directly or by tuning slightly applied to other images with the same resolution.

Likewise, with fixed patch radius r (e.g., $r = 6, 8, 10, 12, 14, 16$), the bandwidth parameter σ was tested on this image. As shown in Figure 4b, the F-measure curves also show a trend of first increasing and then decreasing as σ increases from two to 20 with a step equals to one. However, compared with the test results of parameter r, the F-measure curves of σ are less sensitive than those of r, because after a rapid increase to achieve high accuracies, the F-measure curves then increase slowly until they reach the maximums, followed by a gradually declining tendency. This can be explained by the fact that a larger value for σ means that proximity and similarity occur in a lager spatial neighborhood for each patch; hence, more adjacent patches contribute to the central patch for saliency calculation. However, when σ exceeds a certain threshold, mixed patches would increase in the neighborhood of a central patch, thereby resulting in decreased accuracy. With determined patch radius, the optimal or approximately optimal value for parameter σ can be easily selected by comparing the evaluation results with different parameter settings. For instance, when $r = 10$, the optimal value for σ is 12, because when $\sigma = 12$, the F-measure reached the peak of the curve.

3.3. Results and Performance Evaluation

In this section, the proposed built-up areas extraction algorithm was implemented on aforementioned two datasets so as to verify its effectiveness and robustness. Meanwhile, two state-of–the-art methods including the PanTex algorithm [16] and Hu's model [33] were also applied to the same datasets for performance comparison. In order to make a fair comparison, for the two methods, we used the optimal results (with maximum *F*-measure value) after trying different parameter values.

The experimental results for Google Earth images are provided in Figure 5, in which the first two columns are the original images and ground truths, respectively, while the last three columns correspond to the results of built-up area extraction using the proposed method, the PanTex procedure and Hu's model, respectively. By visual inspection, it can be found that good extraction results were obtained via the proposed method, which matched very well with the reference data. For each test image, the built-up areas were successfully identified, and their boundaries were well delineated. For the PanTex procedure and Hu's model, although they also achieved the detection of built-up areas, it seems that they did not perform as well as our method, because their results contain much more false or missed detection.

To quantitatively evaluate the quality of our method, Table 2 reports the statistical measures of accuracy assessment, in which the second and third columns are the used parameters, which make the F-measure reach the optimal or approximate optimal value for each image. The average F-measure value is 0.8613, while the average precision and recall are 0.8428 and 0.8828, respectively. This indicates that our proposed method achieves excellent performance for these VHR images.
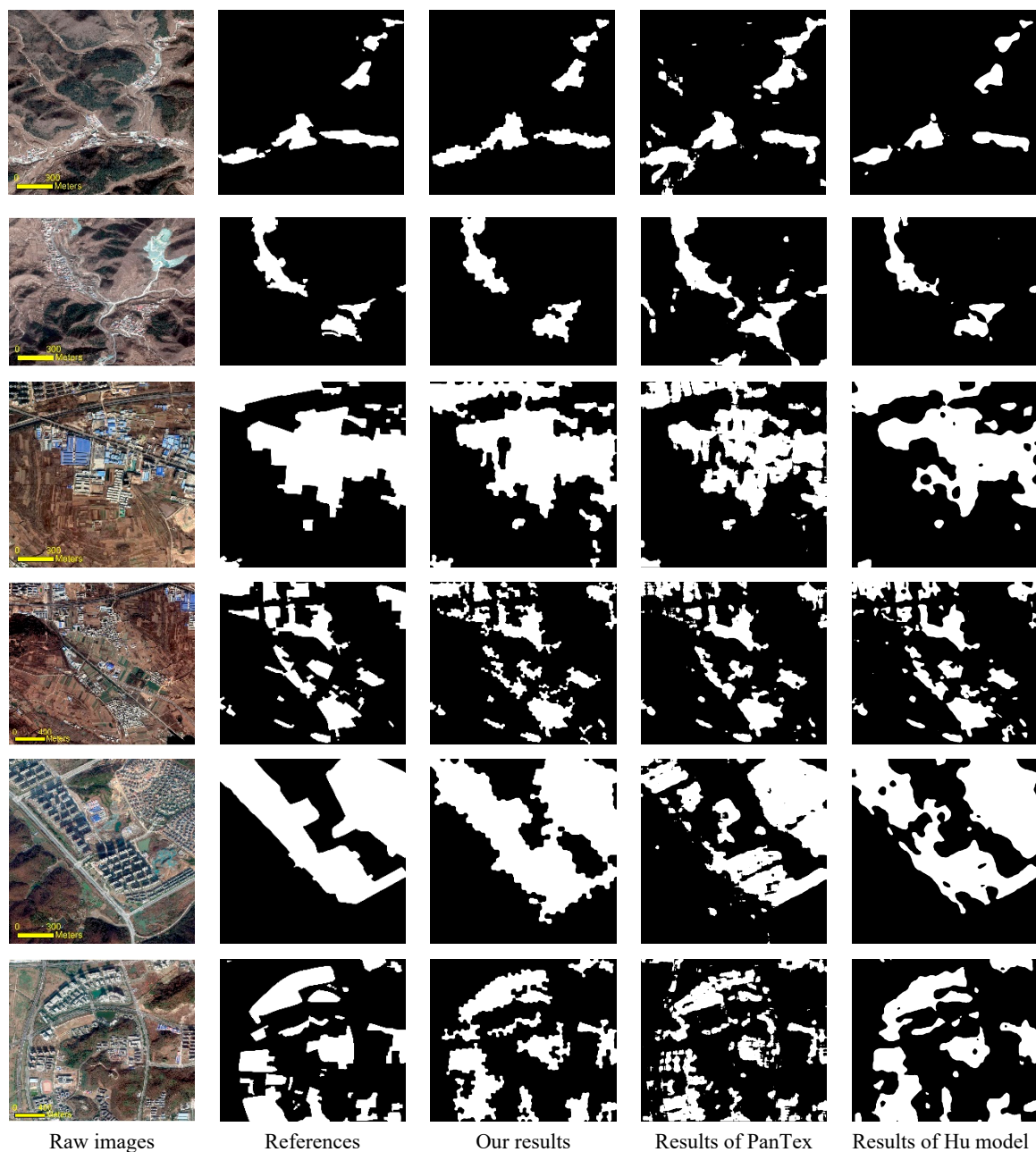
| Raw images | References | Our results | Results of PanTex | Results of Hu model |

Figure 5. Extraction results for the Google Earth dataset. The first two columns are the raw images and references, while the last three columns correspond to the extraction results using the proposed method, the PanTex procedure and Hu's model, respectively.

Table 2. Parameter settings and accuracy evaluation for the Google Earth dataset.

Image No.	Parameter Settings		Accuracy Assessment		
	r	σ	P	R	F-Measure
R1	10	12	0.8251	0.9266	0.8729
R2	12	31	0.8691	0.8551	0.8620
R3	13	12	0.8466	0.9140	0.8790
R4	12	11	0.8510	0.8015	0.8255
R5	17	15	0.8536	0.9678	0.9071
R6	17	12	0.8116	0.8316	0.8215
Average	-	-	0.8428	0.8828	0.8613

Figure 6 presents the accuracy comparison of these methods based on F-measure. Overall, our method showed better performance than the other approaches. More specifically, for the images R1 and R2 (covering complex mountainous areas), the PanTex procedure performed poorly, resulting in low F-measure values (F < 0.75). By inspecting the corresponding result maps, it can be found that a number of irrelevant texture objects (e.g., ridges) were incorrectly extracted as built-up areas. By contrast, both our method and Hu's model showed stronger robustness to these interferences, thereby reducing false detection. For the images R3 and R5 (covering industrial and residential areas, respectively), several built-up areas were missed in the result maps of PanTex and Hu's model; however, the missed detection has been significantly improved by using the proposed method, which can better keep the shape and outline of the extracted built-up areas.
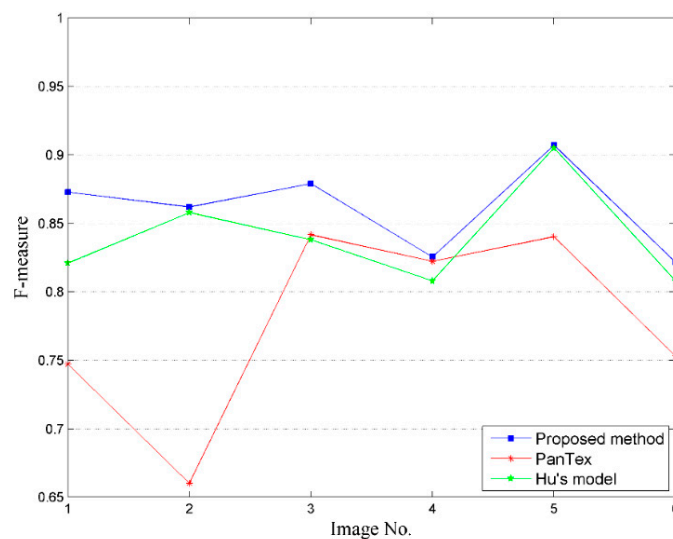

Figure 6. Comparisons of F-measure values with the PanTex procedure and Hu's model on the Google Earth dataset.

To further evaluate the performance of our method, the WorldView-2 datasets were used for built-up areas extraction. Experimental results using the proposed method and the other two approaches are listed in Figure 7. Although these image scenes contain complex textures and spatial patterns due to increased spatial resolution, good extraction results with well-defined boundaries were still obtained via the proposed method, which are quite close to the ground truths.
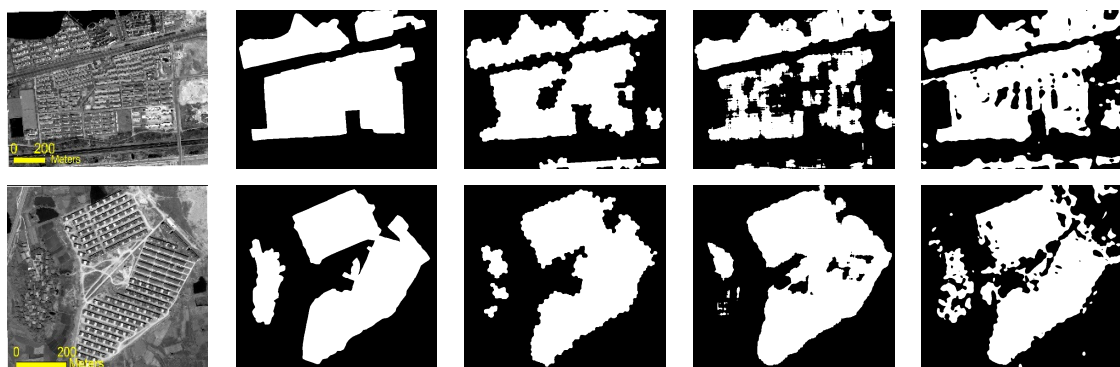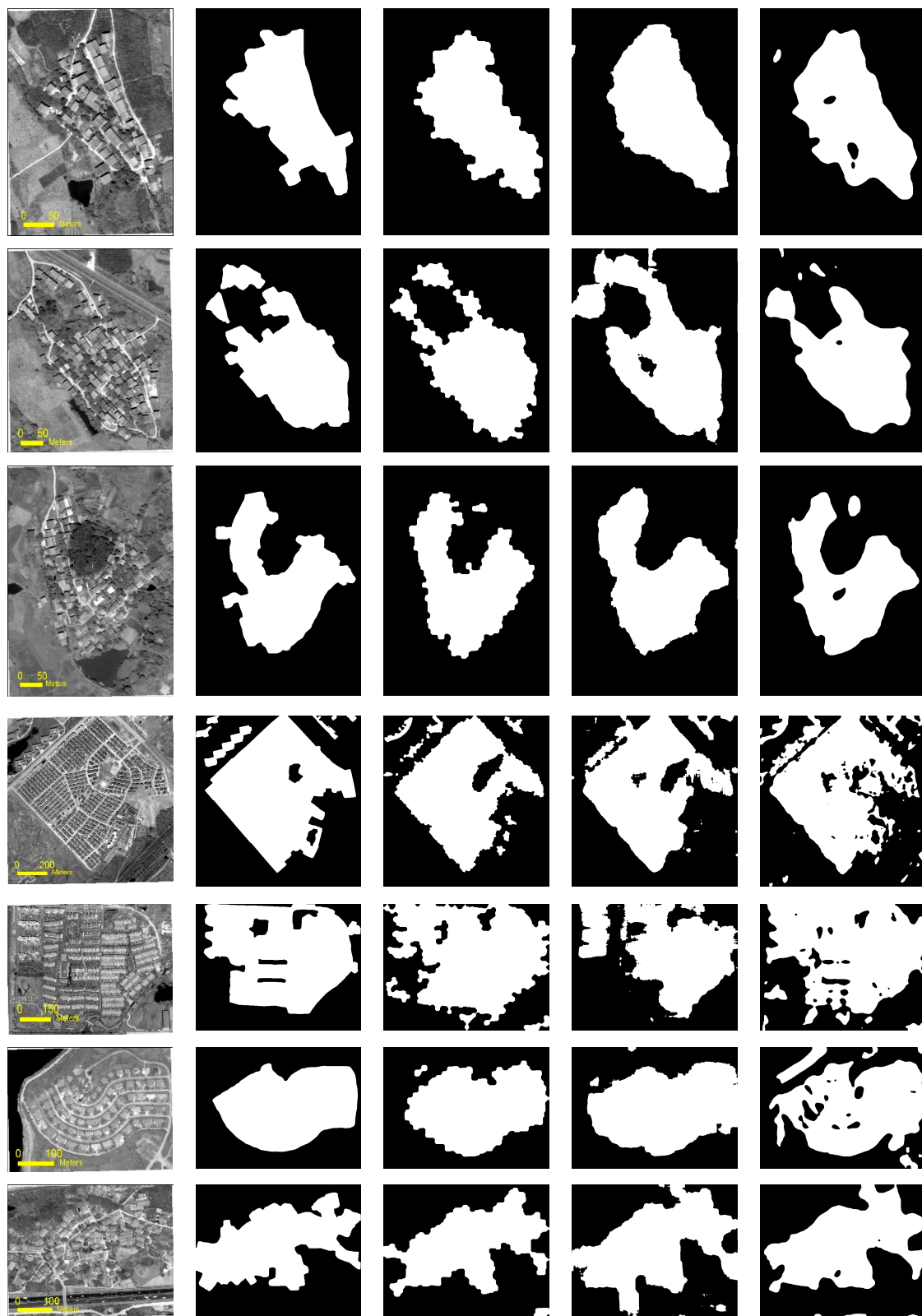

Figure 7. *Cont.*

Figure 7. *Cont.*

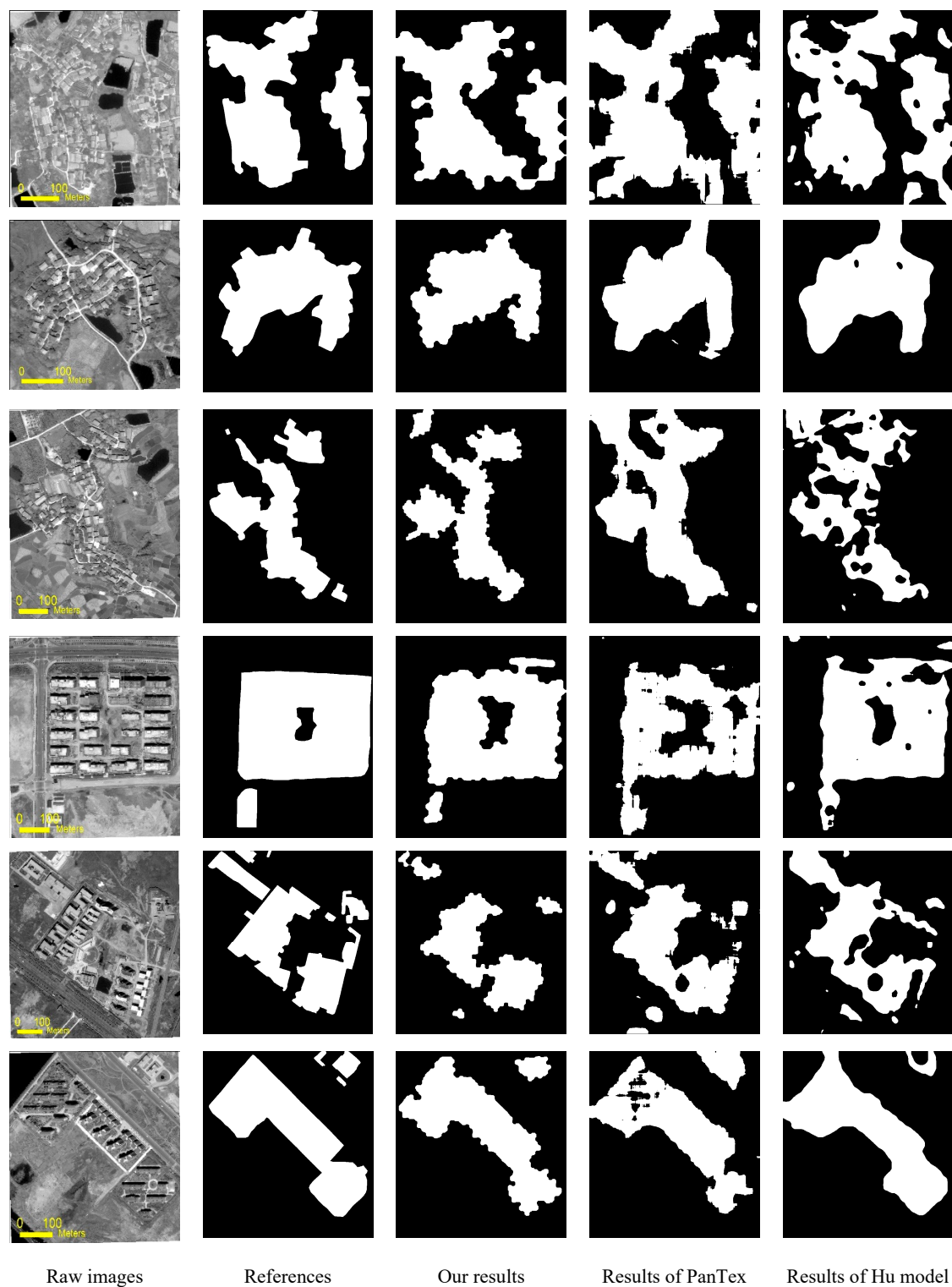| Raw images | References | Our results | Results of PanTex | Results of Hu model |

Figure 7. Extraction results for the WorldView-2 dataset. The first two columns are the raw images and references, while the last three columns correspond to the extraction results using the proposed method, the PanTex procedure and Hu's model, respectively.

Based on *P*, *R*, and F-measure, the quantitative evaluation results of our method on this dataset are shown in Table 3, in which the second and third columns are the corresponding parameters (optimal or approximate optimal). The average precision, recall, and F-measure values are 0.8899, 0.873, and

0.88, respectively. This indicates that our proposed method can effectively achieve the extraction of the built-up areas from this dataset with high correctness and completeness.

Table 3. Parameter settings and accuracy evaluation for the WorldView-2 dataset.

Image No.	Parameter Settings		Accuracy Assessment		
	r	σ	P	R	F-Measure
WV1	18	25	0.8664	0.8786	0.8725
WV2	12	20	0.9178	0.8837	0.9004
WV3	9	16	0.8823	0.9186	0.9001
WV4	9	20	0.9072	0.8824	0.8947
WV5	10	19	0.9110	0.9161	0.9135
WV6	12	15	0.9557	0.8564	0.9034
WV7	18	15	0.8707	0.9452	0.9065
WV8	14	42	0.9402	0.9373	0.9388
WV9	14	15	0.8796	0.8287	0.8534
WV10	20	36	0.7542	0.8530	0.8006
WV11	11	20	0.9222	0.8625	0.8914
WV12	10	18	0.8266	0.8631	0.8445
WV13	18	18	0.9200	0.8733	0.8960
WV14	14	20	0.8923	0.7068	0.7888
WV15	12	20	0.9021	0.8887	0.8954
Average	-	-	0.8899	0.8730	0.8800

The overall performance comparison is depicted in Figure 8. As can be seen, the proposed method achieves the highest F-measure values for almost all images in this dataset, which further demonstrates its superiority over the other two approaches. For example, for images of number one and number seven, which are covered by complex urban buildings with various types and spatial distributions, the resulting F-measure values of the PanTex procedure are relatively low, due to high missed detection rate as shown in Figure 7. Another obvious example is the image of number 12, in which built-up areas are surrounded by farmlands, resulting in great interference to built-up area detection, because some farmlands share similar spectral and spatial features to built-up areas. Consequently, neither the PanTex nor Hu's model performed well, and the former output an enlarged built-up area boundary while the latter produced high missed detection. By contrast, the proposed method performed best and obtain better extraction result.
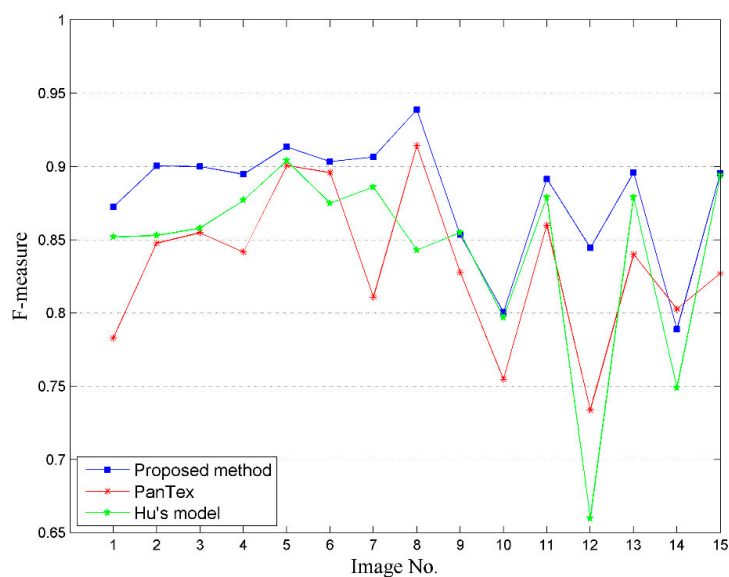
Figure 8. Comparisons of F-measure values with the PanTex procedure and Hu's model on the WorldView-2 dataset.

4. Discussion

The above experiments and comparisons demonstrate the effectiveness of our proposed method, and overall, it shows better performance than both the PanTex procedure and Hu' model in terms of *F*-measure values.

In the proposed method, we adopted the corner-constrained overlapping image patches as the basic processing units, which play an important role in the efficient locating of candidate built-up areas. Compared with previously used non-overlapping superpixels or hard-partitioned blocks [29–33], the overlapping image patches have shown more compact spatial relationships in built-up areas. This would benefit to their grouping by incorporating more valuable spatial clues. To represent each patch, we considered two aspects of spatial characteristics: Its own textural patterns and its spatial relationships with surrounding patches, which were described by using the multi-resolution wavelet and Gestalt laws of proximity and similarity, respectively. The performance improvement of the proposed approach supports our argument that there are greater advantages using patch-based representation and incorporating Gestalt laws than those of the traditional methods using spatial information for built-up areas detection. On the other hand, this also confirms the previous findings that spatial information, such as textures [14–20], structures [13,21–27], and context [15,33], is very valuable for built-up areas identification from VHR images. Particularly, the potential of Gestalt laws in guiding target recognition is further verified, which can well describe the spatial relationships of built-up areas. It should be noted that apart from proximity and similarity, some other Gestalt laws (continuity, closure, figure and ground, background connectivity, etc.) could also be used to guild the detection of built-up areas. In addition, the proposed method adopted the Otsu algorithm to binarize the calculated saliency map in pursuit of automatic detection of built-up areas. Although satisfying extraction results have been obtained on the test datasets, for some images, the Otsu's optimal thresholds do not necessarily produce the best segmentation results. In other words, the extraction results could be further improved by optimizing the selection of thresholds.

In our experiment, it can be found that using the proposed method, built-up areas can be extracted from sub-images of an image. Therefore, for a large image, it can be divided into a series of small sub-images which may contain different parts of built-up areas, and then implement our method on each sub-image separately. The final extraction result can be obtained by uniting each individual one. In this way, if the image covers a very large area (e.g., a city), it may take a long time to complete the

extraction of built-up areas from the entire image. However, this calculation task can be accelerated by using parallel processing.

When implementing our method on a sub-image, many factors can affect the implementation time. Among them, the number of corners detected and the bandwidth parameter σ have significant impacts on the calculation time. This is because the former determines the number of the basic processing units (i.e., corner-constrained patches) while the latter determines the neighborhood size of a patch for its saliency calculation. Take the first image (1280 pixels by 1280 pixels) in the Google Earth dataset for example, in order to change the number of corners detected, we first used Gaussian filters with different parameters to obtain several smoothed images, which were then used as the inputs to implement the proposed method, respectively. The proposed method was executed on a computer with 4 GB RAM and Dual core 2.6 GHz Intel CPU, and the coding platform is MATLAB (R2010b). With different numbers of corners detected, the corresponding calculation time to complete the extraction of built-up areas is shown in Table 4. As can be seen, the increase (decrease) of the number of corners leads to a significant increase (decrease) in calculation time. The effect of bandwidth parameter σ on the calculation time is shown in Table 5. Obviously, the change of this parameter also has a great effect on the calculation time. For example, when σ = 3, it needs only 39.9 s to complete the extraction task, while when σ = 15, the implementation time is increased to 102.6 s. In contrast, for PanTex and Hu' model, which were also executed on the computer mentioned above but with C++ for coding, it took 115.7 s and 18.6 s to obtain the detection results (as shown in Figures 5 and 6), respectively. Besides the calculation time, Table 4 also reports the F-measure values under these conditions. It can be found that the difference in the number of corners detected could result in the variation in detection accuracy. In fact, not only the number of corners, their locations and spatial distributions could also affect the final detection of built-up areas. Therefore, the proposed method can be further improved by optimizing the detection of corners.

Table 4. The effect of the number of corners on the calculation time and the detection accuracy (*r* = 10, σ = 12).

Number of Corners	Calculation Time	F-measure
668	36.2 s	0.8332
967	43.3 s	0.8561
1299	56.1 s	0.8711
1429	64.1 s	0.8670
1582	74.6 s	0.8674
1617	77.4 s	0.8747
1738	82.6 s	0.8729

Table 5. The effect of the bandwidth σ on the calculation time (*r* = 10).

σ	Calculation Time
3	39.9 s
6	43.3 s
8	59.8 s
10	70.9 s
12	82.6 s
14	98.2 s
15	102.6 s

In addition, it should be pointed out that although the datasets contain images of various scenes, they only cover small areas compared with some practical applications at urban, regional, or even global scales. The improvement of the proposed method in built-up extraction is very relevant to the development of global built-up mapping products. Up to now, there are several relevant products [56], such as 500 m Map of Global Urban Extent (MOD500) [57], 30 m Global

Land Cover (GlobeLand30) [58], Global Human Settlement Layer (GHSL) [59], and Global Urban Footprint (GUF) [60,61]. The resolutions of GHSL and GUF are 38 m and 12 m, respectively. The availability of high-spatial-resolution images makes it possible to map this product at a finer resolution. For this purpose, the aforementioned PanTex has been successfully applied to obtain GHSL using VHR images [2,62]. Since our method has shown superior performance over the PanTex procedure at local scales, we think it also has the potential to be applied to develop built-up mapping products at global scales. In terms of computational costs, PanTex calculates local textural patterns in a pixel-wise manner, and for VHR images, they usually need a large neighborhood window for each pixel, which would increase the computational costs. This has also been confirmed by previous studies [29,33]. In contrast, our method adopts a patch-based processing strategy, which would require less calculation time than PanTex. In future studies, we will further improve our method for quick and effective built-up areas mapping at global scales.

5. Conclusions

In this paper, we proposed a novel framework for unsupervised extraction of built-up areas from VHR satellite imagery. The proposed method used a group of corner-constrained patches to locate the candidate built-up areas, and for each patch, its spatial characteristics were represented using integrated high-frequency wavelet coefficients. Then, a patch-level saliency model of built-up areas was constructed by incorporating Gestalt laws of grouping, and built-up areas were segmented by thresholding the saliency map, followed by boundary refinement to obtain the final extraction results. By using the proposed method, the extracted built-up areas had more precise boundaries and showed a better shape integrity than those of the other methods. Thus, the proposed patch-based representation strategy and saliency model can be used for built-up areas mapping in VHR satellite imagery. Furthermore, it can be concluded that visual cognition and Gestalt laws of grouping can play a significant role in guiding the automatic extraction of built-up areas.

In future studies, the following aspects are going to be focused on so as to further improve the current work. (1) The proposed patch-based strategy can be further improved by considering the type, size and distribution of buildings and optimizing the detection of corners. Thus, the candidate built-up areas would be better located. (2) Besides proximity and similarity, more Gestalt laws of grouping, such as continuity, closure, and background connectivity, would be considered and incorporated into our saliency model in order to further increase its robustness. (3) The possibility of applying the proposed method to large-scale built-up areas mapping will be investigated and evaluated so that relevant urban and regional applications can be served.

Author Contributions: Conceptualization, Y.C. and B.H.; methodology, Y.C. and Z.L.; software, Y.Z.; validation, Z.L.; formal analysis, P.Z.; writing—original draft preparation, Y.C.; writing—review and editing, B.H.; visualization, P.Z.; supervision, Y.C. and B.H.; project administration, Y.C.; funding acquisition, Y.C.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mboga, N.; Persello, C.; Bergado, J.R.; Stein, A. Detection of informal settlements from VHR images using convolutional neural networks. *Remote Sens.* **2017**, *9*, 1106. [CrossRef]
2. Pesaresi, M.; Guo, H.; Blaes, X.; Ehrlich, D.; Ferri, S.; Gueguen, L.; Halkia, M.; Kauffmann, M.; Kemper, T.; Lu, L.; et al. A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2012–2131. [CrossRef]

3. Wu, K.; Zhang, H. Land use dynamics, built-up land expansion patterns, and driving forces analysis of the fast-growing Hangzhou metropolitan area, eastern China (1978–2008). *Appl. Geogr.* **2012**, *34*, 137–145. [CrossRef]

4. Yang, X.; Jiang, G.M.; Luo, X.; Zheng, Z. Preliminary mapping of high-resolution rural population distribution based on imagery from Google Earth: A case study in the Lake Tai basin, eastern China. *Appl. Geogr.* **2012**, *32*, 221–227. [CrossRef]

5. Wania, A.; Kemper, T.; Tiede, D.; Zeil, P. Mapping recent built-up area changes in the city of Harare with high resolution satellite imagery. *Appl. Geogr.* **2014**, *46*, 35–44. [CrossRef]

6. You, Y.; Wang, S.; Ma, Y.; Chen, G.; Wang, B.; Shen, M.; Liu, W. Building detection from VHR remote sensing imagery based on the morphological building index. *Remote Sens.* **2018**, *10*, 1287. [CrossRef]

7. Meng, Q.; Zhang, L.; Sun, Z.; Meng, F.; Wang, L.; Sun, Y. Characterizing spatial and temporal trends of surface urban heat island effect in an urban main built-up area: A 12-year case study in Beijing, China. *Remote Sens. Environ.* **2018**, *204*, 826–837. [CrossRef]

8. Tran, D.X.; Pla, F.; Latorre-Carmona, P.; Myint, S.W.; Caetano, M.; Kieu, H.V. Characterizing the relationship between land use land cover change and land surface temperature. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 119–132. [CrossRef]

9. Kit, O.; Lüdeke, M. Automated detection of slum area change in Hyderabad, India using multitemporal satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2013**, *83*, 130–137. [CrossRef]

10. Duque, J.C.; Patino, J.E.; Ruiz, L.A.; Pardo-Pascual, J.E. Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. *Landsc. Urban Plan.* **2015**, *135*, 11–21. [CrossRef]

11. Huang, X.; Zhang, L. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 257–272. [CrossRef]

12. Lv, Z.; Zhang, P.; Benediktsson, J.A. Automatic object-oriented, spectral-spatial feature extraction driven by Tobler's first law of geography for very high resolution aerial imagery classification. *Remote Sens.* **2017**, *9*, 285. [CrossRef]

13. Zhong, P.; Wang, R. Using combination of statistical models and multilevel structural information for detecting urban areas from a single gray-level image. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1469–1482. [CrossRef]

14. Zhang, L.; Li, A.; Zhang, Z.; Yang, K. Global and local saliency analysis for the extraction of residential areas in high-spatial-resolution remote sensing image. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3750–3763. [CrossRef]

15. Chen, Y.; Lv, Z.; Huang, B.; Jia, Y. Delineation of built-up areas from very high-resolution satellite imagery using multi-scale textures and spatial dependence. *Remote Sens.* **2018**, *10*, 1596. [CrossRef]

16. Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2008**, *1*, 180–192. [CrossRef]

17. Pesaresi, M.; Gerhardinger, A. Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 16–25. [CrossRef]

18. Huang, X.; Zhang, L.P. A Multidirectional and Multiscale Morphological Index for Automatic Building Extraction from Multispectral GeoEye-1 Imagery. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 721–732. [CrossRef]

19. Huang, X.; Zhang, L. Morphological Building/Shadow Index for Building Extraction from High-Resolution Imagery Over Urban Areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 161–172. [CrossRef]

20. Shao, Z.; Tian, Y.; Shen, X. BASI: A new index to extract built-up areas from high-resolution remote sensing images by visual attention model. *Remote Sens. Lett.* **2014**, *5*, 305–314. [CrossRef]

21. Tao, C.; Tan, Y.; Zou, Z.; Tian, J. Unsupervised detection of built-up areas from multiple high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1300–1304. [CrossRef]

22. Kovacs, A.; Szirányi, T. Improved Harris feature point set for orientation-sensitive urban-area detection in aerial images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 796–800. [CrossRef]

23. Chen, Y.; Qin, K.; Jiang, H.; Wu, T.; Zhang, Y. Built-up area extraction using data field from high-resolution satellite images. In Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing, Beijing, China, 10–15 July 2016; pp. 437–440.

24. Hu, X.; Shen, J.; Shan, J.; Pan, L. Local edge distributions for detection of salient structure textures and objects. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 466–470. [CrossRef]

25. Shi, H.; Chen, L.; Bi, F.; Chen, H.; Yu, Y. Accurate urban area detection in remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1948–1952. [CrossRef]

26. Ning, X.; Lin, X. An index based on joint density of corners and line segments for built-up area detection from high resolution satellite imagery. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 338. [CrossRef]

27. Zhang, Q.; Huang, X.; Zhang, G. Urban area extraction by regional and line segment feature fusion and urban morphology analysis. *Remote Sens.* **2017**, *9*, 663. [CrossRef]

28. Sirmacek, B.; Ünsalan, C. Urban area detection using local feature points and spatial voting. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 146–150. [CrossRef]

29. Li, Y.; Tan, Y.; Deng, J.; Wen, Q.; Tian, J. Cauchy graph embedding optimization for built-up areas detection from high-resolution remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2078–2096. [CrossRef]

30. Ma, L.; Du, B.; Chen, H.; Soomro, N.Q. Region-of-interest detection via superpixel-to-pixel saliency analysis for remote sensing image. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1752–1756. [CrossRef]

31. Weizman, L.; Goldberger, J. Urban-area segmentation using visual words. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 388–392. [CrossRef]

32. Li, Y.; Tan, Y.; Li, Y.; Qi, S.; Tian, J. Built-up area detection from satellite images using multikernel learning, multifieldintegrating, and multihypothesis voting. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1190–1194.

33. Hu, Z.; Li, Q.; Zhang, Q.; Wu, G. Representation of block-based image features in a multi-scale framework for built-up area detection. *Remote Sens.* **2016**, *8*, 155–174. [CrossRef]

34. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Susstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2281. [CrossRef] [PubMed]

35. Sarkar, S.; Boyer, K.L. Perceptual organization in computer vision: A review and a proposal for a classificatory structure. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 382–399. [CrossRef]

36. Watson, S.E.; Kramer, A.F. Object-based visual selective attention and perceptual organization. *Percept. Psychophys.* **1999**, *61*, 31–49. [CrossRef] [PubMed]

37. Scholl, B.J. Object and attention: The state of the art. *Cognition* **2001**, *80*, 1–46. [CrossRef]

38. Martinez-Fonte, L.; Gautama, S.; Philips, W.; Goeman, W. Evaluating corner detectors for the extraction of man-made structures in urban areas. In Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing, Seoul, Korea, 25–29 July 2005; pp. 237–240.

39. He, X.; Yung, N. Curvature scale space corner detector with adaptive threshold and dynamic region of support. In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 23–26 August 2004; pp. 791–794.

40. Harris, C.; Stephens, M. A combined corner and edge detector. In Proceedings of the 4th Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 147–151.

41. Smith, S.M.; Brady, J.M. SUSAN—A new approach to low level image processing. *Int. J. Comput. Vis.* **1997**, *23*, 45–78. [CrossRef]

42. Rosten, E.; Porter, R.; Drummond, T. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *32*, 105–119. [CrossRef]

43. Mallat, S.G. A theory for multi-resolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]

44. İmamoğlu, N.; Lin, W.; Fang, Y. A saliency detection model using low-level features based on wavelet transform. *IEEE Trans. Multimed.* **2013**, *15*, 96–105. [CrossRef]

45. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]

46. Itti, L.; Koch, C. Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2001**, *2*, 194–203. [CrossRef] [PubMed]

47. Yu, J.; Xia, G.; Gao, C.; Samal, A. A computational model for object-based visual saliency: Spreading attention along gestalt cues. *IEEE Trans. Multimed.* **2016**, *18*, 273–286. [CrossRef]

48. Wannig, A.; Stanisor, L.; Roelfsema, P.R. Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nat. Neurosci.* **2011**, *14*, 1243–1244. [CrossRef]

49. Chen, L. The topological approach to perceptual organization. *Vis. Cogn.* **2015**, *12*, 553–637. [CrossRef]

50. Turker, M.; Kok, E.H. Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS J. Photogramm. Remote Sens.* **2013**, *79*, 106–121. [CrossRef]

51. Yan, Y.; Ren, J.; Sun, G.; Zhao, H.; Han, J.; Li, X.; Marshall, S.; Zhan, J. Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement. *Pattern Recognit.* **2018**, *79*, 65–78. [CrossRef]

52. Pukelsheim, F. The three sigma rule. *Am. Stat.* **1994**, *48*, 88–91.

53. Otsu, N.A. Threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]

54. Lv, Z.; Liu, T.; Zhang, P.; Benediktsson, J.A.; Tao, L.; Zhang, X. Novel adaptive histogram trend similarity approach for land cover change detection by using bi-temporal very-high-resolution remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2019**. [CrossRef]

55. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson Education Limited: New York, NY, USA, 2018; pp. 648–656.

56. Liu, C.; Huang, X.; Zhu, Z.; Chen, H.; Tang, X.; Gong, J. Automatic extraction of built-up area from ZY3 multi-view satellite imagery: Analysis of 45 global cities. *Remote Sens. Environ.* **2019**, *226*, 51–73. [CrossRef]

57. Schneider, A.; Friedl, M.A.; Potere, D. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sens. Environ.* **2010**, *114*, 1733–1746. [CrossRef]

58. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [CrossRef]

59. Pesaresi, M.; Ehrlich, D.; Ferri, S.; Florczyk, A.; Freire, S.; Halkia, M.; Julea, A.; Kemper, T.; Soille, P.; Syrris, V. *Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014*; Publications Office of the European Union: Luxembourg, 2016.

60. Esch, T.; Marconcini, M.; Felbier, A.; Roth, A.; Heldens, W.; Huber, M.; Schwinger, M.; Taubenböck, H.; Müller, A.; Dech, S. Urban footprint processor—Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1617–1621. [CrossRef]

61. Esch, T.; Heldens, W.; Hirner, A.; Keil, M.; Marconcini, M.; Roth, A.; Zeidler, J.; Dech, S.; Strano, E. Breaking new ground in mapping human settlements from space—The global urban footprint. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 30–42. [CrossRef]

62. Pesaresi, M.; Ehrlich, D.; Caravaggi, I.; Kauffmann, M.; Louvrier, C. Toward global automatic built-up area recognition using optical VHR imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 923–934. [CrossRef]