Shadow Detection and Reconstruction in High-Resolution Satellite Images via Morphological Filtering and Example-Based Learning

Huihui Song, Bo Huang, Associate Member, IEEE, and Kaihua Zhang

Abstract—The shadows in high-resolution satellite images are usually caused by the constraints of imaging conditions and the existence of high-rise objects, and this is particularly so in urban areas. To alleviate the shadow effects in high-resolution images for their further applications, this paper proposes a novel shadow detection algorithm based on the morphological filtering and a novel shadow reconstruction algorithm based on the example learning method. In the shadow detection stage, an initial shadow mask is generated by the thresholding method, and then, the noise and wrong shadow regions are removed by the morphological filtering method. The shadow reconstruction stage consists of two phases: the example-based learning phase and the inference phase. During the example-based learning phase, the shadow and the corresponding nonshadow pixels are first manually sampled from the study scene, and then, these samples form a shadow library and a nonshadow library, which are correlated by a Markov random field (MRF). During the inference phase, the underlying landcover pixels are reconstructed from the corresponding shadow pixels by adopting the Bayesian belief propagation algorithm to solve the MRF. Experimental results on QuickBird and WorldView-2 satellite images have demonstrated that the proposed shadow detection algorithm can generate accurate and continuous shadow masks and also that the estimated nonshadow regions from the proposed shadow reconstruction algorithm are highly compatible with their surrounding nonshadow regions. Finally, we examine the effects of the reconstructed image on the application of classification by comparing the classification maps of images before and after shadow reconstruction.

Index Terms—Example learning, Markov random field (MRF), morphological filtering, shadow detection, shadow reconstruction.

I. INTRODUCTION

W ith the technological developments in aerospace, an increasing number of Earth observation commercial satellites with high-resolution sensors have been launched, such as QuickBird (QB), IKONOS, and WorldView-2 (WV-2). The images obtained from these satellites have very high spatial resolution (VHSR), usually ranged from 0.5 to 4 m. At this resolution, details such as buildings and other infrastructures are easily visible. Therefore, these VHSR images have opened a new era for remote sensing applications, such as object detection [1], classification [2], object mapping [3], and change detection. In particular, VHSR images have attracted much attention from researchers studying urban areas, due to the existence of relatively small features, such as roads, buildings, bridges, and trees. Inevitably, tall standing objects (which mainly are buildings) among these small features cast long shadows in most of the captured VHSR images. On the one hand, these shadows may be utilized as a valuable cue for inferring 3-D scene information based on their position and shape, for example, for building detection and building height estimation [4]. On the other hand, the shadows cause partial or total loss of radiometric information in the affected areas, and consequently, they make tasks like image interpretation, object detection and recognition, and change detection more difficult or even impossible [5]. In this paper, we focus on the second aspect of shadows, i.e., to attenuate the problems caused by the loss of radiometric information in shadowed areas by compensating or reconstructing them. Generally, two steps are involved in this procedure: 1) shadow detection and 2) shadow reconstruction (compensation).

Regarding shadow detection in VHSR images, two main approaches are reported in the previous literature, namely, the model-based and the property-based. The former requires prior knowledge of scene or sensors, including, but not limited to, distribution of scene radiance and acquisition parameters like sun azimuth, sensor/camera localization, date, and the time of day of acquisition. Based on the prior information, the model-based approaches obtain good performance in detecting a particular type of objects like buildings and vehicles [6], [7]. However, these approaches are not general enough to deal with the great diversity of geometric structures which usually exist in VHSR satellite images of urban areas. The property-based approaches make use of certain shadow properties in images, such as brightness, spectral characteristics, and geometry. Because of their simplicity both in principle and implementation, the property-based approaches have been widely used in literature; they generally include four (usually interrelated) categories: 1) thresholding-based; 2) color-transformation-based; 3) region-growing-based; and 4) classification-based. In the
thresholding-based methods, the shadow and nonshadow pixels are determined according to a predefined threshold level, which usually can be set according to the bimodal distribution of image histogram, such as the method in [5]. In the color-transformation-based methods, the red–green–blue (RGB) color image is first transformed to a 3-D space, such as hue–intensity–saturation, hue–saturation–value (HSV), and YC_bC_r models; then, a new image is derived according to specific spectral properties of shadows in new space, such as shadow areas having lower intensity, higher hue values, and higher saturation [8]; finally, shadows are detected by thresholding the derived new image. The proposed shadow detection method in [9] first transformed the RGB image into HSV space and then derived a normalized saturation–value difference index (NSVDI) to identify shadows via thresholding. Several photometric invariant color models for shadow detection were compared in [10]. In the region-growing-based methods, the seed points are first selected, and then, each of the pixels is assigned to a segment according to their distance from those regions to which they could potentially be assigned. For example, the shadow detection method in [11] first transformed the RGB image into $c_1c_2c_3$ color space, and then, the region-growing process was applied to the $c_3$ component. The classification methods can also be employed for shadow detection because of the commonly possessed properties in shadowed areas, such as their lower intensity. Recently, the authors in [12] have proposed to utilize the support vector machine (SVM) classification method for shadow detection, in which a binary classification procedure was implemented in a supervised manner to derive a shadow-versus-nonshadow mask.

In order to reconstruct the detected shadowed areas, three algorithms were introduced in [13], namely, the gamma correction method, the linear-correlation method, and the histogram-matching method. The gamma correction method considered the shadow as a multiplicative source that corrupts the brightness of the underlying pixels and then built the relationship between shadow and nonshadow pixels with a power function. In the linear-correlation method, the shadow was modeled as a combination of additive and multiplicative noise, and then, the nonshadow pixels to the first order were restored by a linear function. In the histogram-matching method, the histogram of the shadowed region was matched to that of the nonshadow area of the same class in a window. In these algorithms, the parameters were first calibrated before shadow removal by extracting training data sets from the image. The limitations in these algorithms are that the estimated parameters can only be applied in a local region and that the shadows in the training phase and in the estimation phase should be captured under the same condition. The authors in [14] proposed a linear regression method to bridge nonshadow and shadow areas for each class in each band. Recently, another linear-regression-based method for shadow reconstruction has been proposed in [12], which assumed that both shadow and nonshadow pixels of each class follow a Gaussian distribution and then solved the linear regression parameters by the parametric estimation method. The problem with these linear regression methods is that they lost local variability for each class due to the implementation in a global manner. Clearly, all the aforementioned methods for shadow reconstruction are based on classification and thus need to determine the class of the shadowed areas before reconstruction. In the algorithm proposed in [12], the first step was to collect ground-truth region pairs for all classes, i.e., nonshadow classes and their shadow counterparts; then, these ground-truth regions were utilized for supervised classification in shadow and nonshadow classes separately. In the shadow reconstruction method proposed in this paper, a similar ground-truth collection procedure will be adopted but without the classification step.

In this paper, we propose an alternative shadow detection algorithm based on thresholding and morphological filtering, together with an alternative shadow reconstruction algorithm based on the example learning method and Markov random field (MRF). During the shadow detection procedure, the bimodal distributions of pixel values in the near-infrared (NIR) band and the panchromatic band are adopted for thresholding. During the shadow reconstruction procedure, we model the relationship between nonshadow and the corresponding shadow pixels and between neighboring nonshadow pixels by employing MRF.

II. PROBLEM FORMULATION AND MOTIVATION

Shadows occur when objects occlude direct light from a source of illumination, which is usually the sun. According to the principle of formation, shadows can be divided into cast shadow and self-shadow. Cast shadow is formulated by the projection of objects in the direction of the light source; self-shadow refers to the part of the object that is not illuminated. For a cast shadow, the part of it where direct light is completely blocked by an object is termed the umbra, while the part where direct light is partially blocked is termed the penumbra. Because of the existence of a penumbra, there will not be a definite boundary between shadowed and nonshadowed areas.

At present, most of the VHSR satellite sensors are designed with orbit type of sun synchronous and equatorial crossing time earlier in a day; this is because the atmosphere is generally clearer in the morning than later in the day. For example, the equatorial crossing times of QB, IKONOS, and WV-2 are 10:00 A.M., 10:30 A.M., and 10:30 A.M., respectively. This means that the solar elevation will never be high, irrespective of latitude and season. Thus, the problem of shadowing is particularly significant in high-resolution satellite images, and this will cause false image colors and further hinder the application of VHSR images, such as the generation of a classification map. Accordingly, we propose a new method of detecting and reconstructing shadows in VHSR satellite images. Because self-shadows usually have higher brightness than cast shadows, this paper focuses only on the cast shadows, as is the case in previous literature dealing with shadows of VHSR satellite images. With the variation of acquisition conditions and the height of erected objects, the penumbra cannot sometimes be neglected, particularly when the brightness of the surrounding shadowed areas is intense. The penumbra effect will be handled by shadow edge compensation in the proposed shadow detection algorithm.

For a given shadow region $R_s$ on a VHSR image, we seek to estimate the corresponding nonshadow region $R_n$. This
problem can be formulated as a maximum a posteriori problem, i.e., finding \( R_n \) that maximizes posterior probability \( P(R_n|R_s) \), which can be expressed as

\[
\hat{R}_n = \arg \max_{R_n} P(R_n|R_s).
\] (1)

According to the Bayesian theorem, \( P(R_n|R_s) \) can be expressed as

\[
P(R_n|R_s) = \frac{P(R_s, R_n)}{P(R_s)}
\] (2)

where \( P(R_s, R_n) \) is the joint probability of \( R_s \) and \( R_n \), and \( P(R_s) \) is a prior probability of \( R_s \). Since \( P(R_s) \) is a constant over nonshadow pixels, (1) can be equivalently written as

\[
\hat{R}_n = \arg \max_{R_n} P(R_s, R_n).
\] (3)

The research in [14] indicated that the darkness of shadow depends heavily on the surrounding conditions. Therefore, it is necessary to build multiple relationships for each nonshadow class and its shadowed counterpart in one scene. In this paper, this problem is treated by manually collecting nonshadow and shadow pixel pairs in different locations of one scene for each class during the training procedure. Because of the existence of human interaction, one requirement for the proposed shadow reconstruction algorithm is that the shadowed areas in the studied VHSR satellite image are not so dark, thereby enabling human eyes to distinguish the classes under shadowed pixels. In the case that the shadowed areas are so dark, one possible solution is to find a reference image with little or without shadows to assist in the training procedure. Another problem that was usually found in previous shadow reconstruction algorithms is that the reconstructed shadow area may not be consistent with its neighborhood nonshadow areas. In this paper, we handle this problem by utilizing the belief propagation method to keep the reconstructed shadow area consistent with its neighborhood nonshadow areas.

### III. Proposed Methodology

The flowchart in Fig. 1 shows the principal steps of the proposed methodology. The whole procedure includes shadow detection and shadow reconstruction stages by executing on a multispectral (MS) satellite image of an urban area scene. The main contribution of this paper in the shadow detection stage is that we combine thresholding and morphological filtering techniques by considering the spectral characteristics of different land-cover types. The shadow detection stage consists of three main steps: thresholding, morphological filtering, and edge compensation. First, a preliminary shadow mask is derived by the thresholding method according to the spectral characteristics of the MS image. Then, this shadow mask is elaborated by morphological operations to filter noise and the wrong shadow areas. Finally, the shadow edges are compensated considering the effects of penumbra and the surrounding conditions of shadows on VHSR images. The shadow reconstruction stage includes two main steps: example-based training and shadow reconstruction via Bayesian belief propagation (BBP). Before the training step, the nonshadow and shadow samples are first collected from the same image scene manually by visual judgment. Then, the training samples formulate a nonshadow library and a shadow library, which are correlated by an MRF. With the trained nonshadow and shadow libraries, the underlying nonshadow pixels can be reconstructed from the corresponding shadow pixels according to the derived shadow mask in the shadow detection stage.

#### A. Thresholding for Shadow Detection

Since the NIR spectrum has higher reflectivity than visual spectrum for many urban land-cover types, the digital number (DN) values of urban VHSR images are higher in the NIR band than in other bands. For shadow areas, the DN values in an NIR band drop in a higher degree because of the occlusion of direct sunlight. The study in [14] has shown that the DN ratio of light shadow and sunlight is lower in NIR band than in RGB bands. In this paper, we compare the ratios of nonshadow and shadow pixel values in different bands of QB and WV-2 images, which will be studied in our experimental part. We selected four common land-cover types in urban areas as the study objects hereof: vegetation, building, road, and cement. The nonshadow and the corresponding shadow samples (collected for training in step of building learning model in Section III-D) for the same objects (class) are first averaged, respectively, and then, the ratios of nonshadow and shadow values are derived for each land-cover type. The curves of ratio with respect to band number are shown in Fig. 2, from which we can observe that the nonshadow and shadow ratios in the NIR band are the highest for all land-cover types both on the QB image (band 4) and the WV-2 image (bands 7 and 8). These indicate that it is easier to distinguish shadow areas in the NIR band than in other bands. We thus execute the thresholding algorithm in the NIR band for shadow detection. Because there are mainly two features...
The threshold level can be determined by the bimodal histogram splitting method [15], modalities in the NIR histogram. The threshold level can then gap in their DN values, we assume that there are two main of interest, i.e., shadows and nonshadows, and there exists a gap in their DN values, we assume that there are two main modalities in the NIR histogram. The threshold level can then be determined by the bimodal histogram splitting method [15].

The threshold level \( T \) is set to the mean of the two peaks in the NIR histogram [5], which was found by experiments to give consistently accurate threshold levels in separating the shadow from the nonshadow regions. The shadow mask is then derived by the following formula:

\[
M_T = \begin{cases} 
1, & \text{if } \text{DN}_{\text{NIR}} > T \\
0, & \text{else}. 
\end{cases}
\]  

(4)

As the study scene may contain objects that have very low values in the NIR band but high values in other bands, such as water having relative high values in the green band, the derived shadow mask from the NIR band is further refined by the thresholding results in other bands (in this paper, we use the panchromatic band, which is approximated by the average of all other bands in the visual spectrum), i.e., if the detected shadow pixel in the NIR band is judged to be a nonshadow pixel in the panchromatic band, then this pixel is set to be a nonshadow pixel in the final shadow mask.

B. Morphological Filtering for Shadow Detection

An example of a shadow mask derived from the thresholding method is shown in Fig. 3(b), with black indicating shadow pixels and white indicating nonshadow pixels. From this, we can observe that there are two problems: 1) the existence of many small discontinuous shadow regions caused by the salt-and-pepper noise in VHSR images and 2) the wrong shadow regions caused by the low DNs of some objects both in the NIR and the panchromatic bands, particularly roads. In order to keep the shape of the detected shadow regions and to remove the noise, we adopt the morphological image processing method to enhance the detected shadow mask \( M_T \) derived from the thresholding method. Moreover, the morphological operations can also remove the wrong shadow regions with appropriate prior information.

Mathematical morphology is a set- and lattice-theoretic methodology for image analysis, which aims to quantitatively describe the geometrical structure of image objects [16]. Thus, morphological filters, which are more suitable than linear filters for shape analysis, play a major role in geometry-based enhancement and detection. The basic morphological operations include erosion and dilation. With a structuring element, erosion shrinks the object, and dilation grows the object. When combining erosion and dilation, two new operations are generated, namely, opening (erosion followed by dilation) and closing (dilation followed by erosion), which keep the general shape of objects but possess different smoothing effects. Specifically, the opening removes small protrusions and thin connections, whereas the closing fills in small holes. In this paper, we adopt opening and closing operations to remove the noise and wrong shadows in \( M_T \) with a structuring element of \( 3 \times 3 \) ones.

Because the shape of roads is usually thin, the opening operation is applied to \( M_T \) once. Consequently, the road shadows are broken into discrete small regions, which can be deemed as noise in further steps. For illustration, the enhanced mask \( M_{TOO} \) of example \( M_T \) in Fig. 3(b) after this opening operation is shown in Fig. 3(c). To remove the noise in nonshadow areas, one opening operation with an eight-connected neighborhood constraint and area specifications is applied to \( M_{TO} \) based on the following steps: 1) determining the connected components; 2) computing the area of each component; and 3) removing small regions whose areas are smaller than a predefined threshold \( A_e \). The enhanced mask \( M_{TOO} \) of example \( M_{TO} \) in Fig. 3(c), after this step has been carried out, can be seen in Fig. 3(d). To remove the noise in shadow areas, the closing operation is applied to \( M_{TOO} \) with the same procedures as in the aforementioned opening operation. The enhanced shadow mask \( M_{TOO} \) of \( M_{TOO} \) in Fig. 3(d) after this step is shown Fig. 3(e). The two predefined thresholds in the aforementioned opening and closing operations should vary according to different spatial resolutions of study images. There are two solutions for this problem: 1) They can be determined empirically by experiments, and 2) they can be determined by some prior knowledge, for example, the smallest possible shadow area can be determined by the prior information about the high-rise objects in the study scene. We adopt the first strategy in this paper. Since the noise regions from broken roads are
larger and smaller accordingly, for example, larger than those in shadow areas, the predefined thresholds in the aforementioned opening and closing operations can be set larger and smaller accordingly, for example, $A_t = 80$ and $A_p = 30$ for deriving shadow masks in Fig. 3(d) and (e), respectively. For those thin shadows caused by bridges, we can first detect the bridges by adopting relevant methods [17] and then keep such shadows in the final shadow mask by considering their spatial locations with the detected bridges.

C. Shadow Edge Compensation

The width of the penumbra varies with the changes in the elevation angle of the sun and the height of the objects. For simplicity, the penumbra effect is tackled by compensating one pixel at the shadow edges. However, for shadow areas whose surroundings are high-brightness areas, their shadow edges are strongly affected, and therefore, the penumbra width needs to be extended in this case. We compensate these shadow edges by conducting the following steps: 1) growing both high-brightness areas and shadow areas along back-light and to-light directions, respectively, by a dilation operation, and 2) taking their intersecting regions as the compensated shadow edge regions. An example of a shadow mask after the aforementioned edge compensation steps is shown in Fig. 3(f), where the compensated areas are denoted by gray.

D. Learning Model of Shadow and Nonshadow Pixels Based on MRF

Freeman et al. in [18] proposed an example-based learning method for low-level vision problems via image/scene training. The basic idea of this method is to exploit the relationship of image and scene pairs in the training set to generate a learning model and to utilize the learning model to derive a scene from its image. This idea has been well applied to image superresolution tasks [19], [20] by building the relationship between a high-resolution image and its low-resolution version. Motivated by this application, we propose to build the relationship between shadow and nonshadow pixels via this example-based learning method. With light shadows in the VHSR image, we first extract shadow and nonshadow pixel samples in the study scene for training purposes. During this procedure, two conditions need to be satisfied: 1) The nonshadow and corresponding shadow samples for one land-cover type are as close as possible in spatial location to alleviate the effects of surrounding condition and illumination variation, and (2) the sampled shadow and nonshadow pixel pairs should include all land-cover types that may incur shadows in the study scene. With proper discrimination capability, this collection procedure is fast, and the time taken varies with the size of a study scene (a larger study scene costs longer time and vice versa).

Because of the complexity of imaging conditions and illumination variations in satellite sensors, it is difficult to build the relationship between shadow and nonshadow pixels with linear estimation methods. We thus relate shadow and nonshadow pixels by employing an MRF due to its good performance at modeling spatial relationships, such as in remote sensing applications of object identification [21] and image segmentation [22]. Then, we make the following Markov assumptions:

![Fig. 4. MRF model for shadow and nonshadow PVs. Each node in the network denotes a PV of shadow or nonshadow areas. Lines in the graph indicate statistical dependences between nodes.](image)

For each shadow and the underlying nonshadow pixels, they are assigned to one node of a Markov network; each node is statistically independent from other nodes except its direct neighbors.

For a VHSR image with $B$-bands, a shadow pixel vector (PV) is denoted as $V_s \in R^B$ (i.e., the pixel values on all bands) and the underlying nonshadow PV as $V_n \in R^B$. We connect each nonshadow PV both to its corresponding shadow PV and to its spatial neighbors, as the example with four pixels shown in Fig. 4. Each nonshadow PV $V_n(i, j)$ is related to its corresponding shadow PV $V_s(i, j)$ and its neighbors $V_n(i + 1, j)$, $V_n(i, j + 1)$, $V_n(i - 1, j)$, and $V_n(i, j - 1)$. The relationship between these nodes can be represented by $\Theta(\cdot)$ and $\Psi(\cdot)$, respectively. $\Theta(\cdot)$ is the compatibility function of $V_n$ and its corresponding $V_s$, and $\Psi(\cdot)$ is the compatibility function of $V_n$ and its direct neighbors. Via the connection of this Markov network, the sampled shadow and nonshadow PVs are formed to build a shadow library $L_S$ and a nonshadow library $L_N$, respectively, as shown in the training stage in Fig. 1.

E. Shadow Reconstruction Via BBP

Given the shadow region $R_s = [V_s(1, 1), V_s(1, 2), \ldots, V_s(I, J)]$, we seek to estimate the underlying nonshadow region $R_n = [V_n(1, 1), V_n(1, 2), \ldots, V_n(I, J)]$, where $I$ and $J$ are the maximum indices of rows and columns, respectively. We can rewrite the joint probability $P(R_s, R_n)$ in (3) as

$$P(R_s, R_n) = P[V_s(1, 1), V_s(1, 2), \ldots, V_s(I, J), V_n(1, 1), V_n(1, 2), \ldots, V_n(I, J)]$$

(5)

where $V_n(i, j)$ is unknown. Since each nonshadow PV is only related to its underlying shadow PV and its direct neighbors, as shown in Fig. 4, it can be easily deduced from the Bayesian theory that the joint probability in (5) can be expressed as a product of conditionals in the form

$$P(R_s, R_n) = P[V_s(1, 1)] \times \prod_{(i, j)} P[V_s(i, j) | V_n(i, j)]$$

$$\times \prod_{(u, v) \in \Omega(i, j)} P[V_n(i, j) | V_n(u, v)]$$

(6)

where $\Omega(i, j)$ denotes the direct neighbors of $V_n(i, j)$. 

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
Let $\hat{V}_n(i, j)$ denote the predicted nonshadow PV in an MRF model and $V_n(i, j)$ its ideal underlying shadow PV in library $L_S$. We assume that the predicted $\hat{V}_n(i, j)$ is compatible with $V_n(i, j)$ if $\hat{V}_n(i, j)$ matches $V_n(i, j)$. We assume that the given shadow PV $V_n(i, j)$ differs from the selected shadow PV $\hat{V}_n(i, j)$ from $L_S$ by the Gaussian noise of covariance $\sigma_v$. Then, we can define the compatibility function $\Phi(\cdot)$ as follows:

$$\Phi[\hat{V}_n(i, j), V_n(i, j)] = \exp\left\{-\frac{|V_n(i, j) - \hat{V}_n(i, j)|^2}{2\sigma_v}\right\}. \quad (7)$$

When the distance between $V_n(i, j)$ and $\hat{V}_n(i, j)$ is small, $\Phi(\cdot)$ obtains a large value and vice versa. Similarly, by assuming that the predicted $\hat{V}_n(i, j)$ differs from its neighbors $\hat{V}_n(u, v)$ by the Gaussian noise of covariance $\sigma_h$, we can define the compatibility function $\Psi(\cdot)$ as follows:

$$\Psi[\hat{V}_n(i, j), \hat{V}_n(u, v)] = \exp\left\{-\frac{|\hat{V}_n(i, j) - \hat{V}_n(u, v)|^2}{2\sigma_h}\right\}. \quad (8)$$

When the distance between two neighboring nonshadow PVs is small, $\Psi(\cdot)$ obtains a large value and vice versa. The parameters $\sigma_v$ and $\sigma_h$ in (7) and (8) are set to the standard deviations of the shadow and nonshadow training samples, respectively.

Since $\Theta(\cdot)$ and $\Psi(\cdot)$ describe how two relevant PVs agree with each other, we can use them to represent the conditional probabilities in (6). Thus, we can convert the joint probability in (6) to an MRF representation as follows [22]:

$$P(R_s, R_n) = C \prod_{(u, v) \in \Omega(i, j)} \Psi[V_n(i, j), V_n(u, v)]$$

$$\times \prod_{(i, j)} \Phi[V_n(i, j), V_n(i, j)] \quad (9)$$

where $C$ is a normalization factor during the conversion.

In the inferring process, (9) is maximized to infer the unknown $V_n(i, j)$. The straightforward way is to consider all PVs in $L_S$ as candidates of $V_n(i, j)$. However, it may introduce a heavy computation load. Therefore, we choose a number of candidates from $L_N$ for the unknown $V_n(i, j)$ based on the similarity between input shadow PV $V_n(i, j)$ and the PVs in $L_S$. Specifically, the most similar $C$ PVs of $V_n(i, j)$ are found from $L_S$ ($C = 10$ in our experiments). Then, the corresponding $C$ nonshadow PVs from $L_N$ are considered as the candidates for the unknown $V_n(i, j)$. Fig. 5 illustrates the procedure of selecting candidates. To obtain the most $C$ similar candidates of $V_n(i, j)$, the following equation is used to calculate the similarity of two shadow PVs:

$$\text{Dist}[V_n(i, j), \hat{V}_n(i, j)] = \sqrt{|V_n(i, j) - \hat{V}_n(i, j)|^2}. \quad (10)$$

We adopt the BBP method [18] to efficiently find a local maximum of the posterior probability of (9) for the nonshadow PVs, given the shadow PVs. BBP is a message-passing algorithm for performing inference on the MRF. The advantage of BBP is that it accumulates local computation to avoid the global computation by recursively updating the local messages between adjacent unknown or known nodes. For each unknown node $V_n(i, j)$, BBP finds the optimal solution of the following formulation:

$$\hat{V}_n(i, j) = \arg \max_{V_n(i, j)} \left[ V_n(i, j), V_n(i, j) \right]$$

$$\times \prod_{(u, v) \in \Omega(i, j)} \Phi(V_n(i, j), V_n(i, j)) \quad (11)$$

where $\Omega(i, j)$ denotes the neighbors of $(i, j)$ and $m_{(u, v) \rightarrow (i, j)}[\hat{V}_n(i, j)]$ denotes the message that node $\hat{V}_n(u, v)$ sends to its neighboring node $\hat{V}_n(i, j)$. Among all candidates of $V_n(i, j)$, the $\hat{V}_n(i, j)$ that maximizes (11) individually is selected as the resultant PV. By setting the initialization of all messages to one, the messages in (11) are updated automatically through the following rules:

$$m_{(u, v) \rightarrow (i, j)}[\hat{V}_n(i, j)] = \max_{\hat{V}_n(u, v)} \left[ \hat{V}_n(i, j), \hat{V}_n(u, v) \right]$$

$$\Phi[\hat{V}_n(u, v), \hat{V}_n(u, v)] \times \prod_{(a, b) \in \Omega(u, v)} m_{(a, b) \rightarrow (u, v)}[\hat{V}_n(u, v)] \quad (12)$$

where $\Omega(u, v)$ denotes the neighbors of $(u, v)$ except $(i, j)$ and $m_{(a, b) \rightarrow (u, v)}[\hat{V}_n(u, v)]$ denotes the message that node $\hat{V}_n(a, b)$ sends to its neighboring node $\hat{V}_n(u, v)$ in the previous iteration. By setting the maximum number of allowed iterations as 20, the BBP algorithm converges after several iterations when the message $m_{(u, v) \rightarrow (i, j)}$ in the current iteration is equal to that in the previous iteration. To ensure that the reconstructed shadow region is consistent with its surrounding nonshadow regions, we include the surrounding nonshadow PVs into the optimization. That means the neighborhood $\Omega(i, j)$ for the compatibility function $\Psi(\cdot)$ in (9) includes both the reconstructed shadow PVs and the neighboring PVs in nonshadow areas. Under the similarity constraint defined by (8), the reconstructed shadow PVs are consistent with their surrounding nonshadow PVs. We achieve this purpose by implementing the BBP optimization procedure on the whole image scene.
IV. EXPERIMENTAL RESULTS AND COMPARISONS

A. Data Set Description

To evaluate the performance of the proposed method, we used three different images located in Hong Kong. The first two images are cut out of a standard-level MS QB image scene with sizes of $276 \times 297$ and $366 \times 424$, respectively. This MS QB scene was captured on January 7, 2007, with four spectral bands of red, green, blue, and NIR and a spatial resolution of 2.4 m. These two images with NIR–red–green as R–G–B composite are shown in Figs. 6(a) and 7(a), respectively, which both are located in suburban residential areas and are characterized by main shadow areas of vegetation, road, buildings, and cement. However, they differ in their shadow types, building densities, and road topologies. The third image is cut out of a level-2 MS WV-2 image scene with a size of $410 \times 442$. This MS WV-2 scene was captured on January 8, 2011, with eight spectral bands (coastal, blue, green, yellow, red, red edge, NIR1, and NIR2) and a spatial resolution of 2 m. The third WV-2 image is shown in Fig. 8(a) with NIR1–red–green as R–G–B composite, which is located in urban residential areas and is characterized by dense high-rise buildings and complex road topologies.

B. Shadow Detection Results

The shadow masks for the three images from our thresholding method are illustrated in Figs. 6(b), 7(b), and 8(b), respectively, from which we can see that most of the shadow regions are detected. However, there are many small shadow regions and some wrong shadow regions, which are mainly caused by the existence of noise and the low DNs of roads in the NIR band, respectively. The shadow masks after morphological filtering and edge compensation are shown in Figs. 6(c), 7(c), and 8(c), respectively, from which we can observe that the main and correct shadow regions are preserved and that the penumbras caused by high-brightness surroundings along the shadow direction are compensated, such as the gray areas in red polygons of Fig. 8(c).
TABLE I
FDR AND MDR OF THREE METHODS ON THE SECOND QB IMAGE

<table>
<thead>
<tr>
<th></th>
<th>Our</th>
<th>NSVDI</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FDR</strong></td>
<td>0.0642</td>
<td>0.0121</td>
<td>0.0668</td>
</tr>
<tr>
<td><strong>MDR</strong></td>
<td>0</td>
<td>0.1369</td>
<td>0</td>
</tr>
</tbody>
</table>

To demonstrate the superiority of the proposed shadow detection method, we compared it to the NSVDI method in [9], which is another thresholding-based shadow detection method. For the second QB image, we tuned the threshold level of NSVDI method to 0.5 to achieve the best result, and this is shown in Fig. 7(d). Comparing the results in Fig. 7(b) and (d), we can conclude that our method is more accurate in detecting shadow regions than NSVDI. For example, in the NSVDI mask, there are many noises in the detected shadow region, and the detected shadow shape is not accurate (e.g., the building shadow labeled by the red ellipse in Fig. 7(b) and (e)).

To demonstrate that the shadow detection part does not need training samples despite their availability, we compared it to the supervised shadow detection method based on binary SVM [12]. For this supervised method, we tuned the regularization and the kernel parameters of the SVM by cross-validation on the training samples (extracted for our shadow reconstruction stage). The shadow mask derived from this method is shown in Fig. 7(e). Comparing with Fig. 7(b) and (e), we can observe that this supervised method has similar performance with our thresholding method in shadow detection.

To further quantitatively evaluate the shadow detection results, we randomly selected 1153 nonshadow pixels and 621 shadow pixels by visual inspection from Fig. 7(a) as ground truth. Then, the false detection rate (FDR) and the missing detection rate (MDR) were calculated on the derived shadow masks from the proposed thresholding method, the NSVDI method, and the SVM method, respectively. From the comparison in Table I, we can conclude that the proposed thresholding method and the SVM method perform similarly while the NSVDI is worse than them at shadow detection.

C. Shadow Reconstruction Results

During the training stage for shadow reconstruction, we extracted shadow and nonshadow samples from the whole image scene and chose as many shadow types as possible. For the QB image scene, the dominant shadow land-cover types are vegetation, buildings, roads, cement, bare soil, and playground (or lawn). The sampled shadow/nonshadow pixel numbers for these land-cover types during our training stage are 289, 170, 217, 100, 93, and 62, respectively. For the WV-2 image scene, the dominant shadow land-cover types are the same as in QB, and the corresponding pixel numbers for the extracted shadow/nonshadow samples are 192, 159, 162, 72, 40, and 112, respectively. For each land-cover type, the training samples thereof consisted of within-class variations in aspects of illumination, color, etc. For example, the playground samples for the WV-2 image include colors of red, green, and purple. During the shadow reconstruction stage, for the compensated one pixel shadow edge, we took the average value of the nearest reconstructed pixel and the nearest nonshadow pixel along the shadow edge direction; for the compensated shadow pixels affected by high-brightness surroundings (because they still have some gap with samples in library $L_S$ in value), we set their values to be the same as those of the nearest reconstructed nonshadow pixels.

To demonstrate the benefit of using MRF and BBP, we show the results without (by only choosing the best candidate from the libraries) and with MRF and BBP, respectively. For the aforementioned three images, the reconstructed results with MRF and BBP are shown in Figs. 6(e), 7(f), and 8(d); for the first QB image, the result without MRF and BBP is shown in Fig. 6(d). Comparing these results, we can easily observe that the reconstructed shadow areas with MRF and BBP are smoother and more consistent with their surroundings. By visual inspection, we can see that there are more wrong reconstructions in the WV-2 image than in the QB images, such as the reconstructed playground area in red and the pink pixels in reconstructed roads in Fig. 8(d). This is caused mainly by the complex and volatile land-cover types in urban areas. To demonstrate the reconstruction accuracy, we took another WV-2 image located in the same area but captured on July 5, 2011, as the reference image. As shown in Fig. 8(e), there are few shadows in this reference image due to a different capture angle, and there are few land-cover type changes between these two images during this short period (from January to July). For the shadowed areas in green polygons in Fig. 8(a), the reconstructed ones and the actual land-cover types are marked by green polygons in Fig. 8(d) and (e), respectively, for comparison, from which we can observe that most of the shadow regions are reconstructed correctly.

To demonstrate the superiority of our shadow reconstruction algorithm, we compared it with a recently proposed linear regression method in [12], which runs an SVM classification algorithm both for shadow and nonshadow regions, and then, it builds a linear regression function between shadow and nonshadow pixels for each class. For simplicity, we manually select three shadow areas with three kinds of land-cover types, i.e., vegetation type in the first QB image, building type in the second QB image, and road type in the WV-2 image, respectively. To solve the linear regression parameters of the method in [12] for these three land-cover types, we utilized the same training samples of shadow/nonshadow pixel pairs in our training stage. The selected vegetation, building, and road shadow regions are shown in yellow rectangles in Figs. 6(a), 7(a), and 8(a), respectively. The reconstructed shadow regions from both methods are shown in Figs. 6(f), 7(g) (in yellow polygons), and 8(f) (in yellow polygons), respectively, from which we can see that the results from our method are more consistent with the surroundings than the comparison algorithm. This mainly attributes to the propagation procedure of BBP between the reconstructed regions and the surrounding nonshadow regions in our algorithm.

To examine the sensitivity of the reconstruction results on the selected examples, we choose the road type on the WV-2 image and re-collect the road training samples with a half of the original sample size, i.e., the new road set has 81 shadow/nonshadow road pairs. For the road shadow region in the yellow rectangle in Fig. 8(a), the reconstructed shadow regions from the new example set and the original example set are shown in
constructed road and playground (or lawn) types. However, for vegetation and that there are few classification errors for reconstruction confusion matrix is shown in Table II, from which we can see that there are no classification errors for reconstructed images by visual inspection and utilizing the corresponding shadow masks. The pixel numbers of selected reconstructed shadow samples for vegetation, building, road, cement, and playground (or lawn) are 300, 150, 150, 150, and 100, respectively. For denotation convenience, these land-cover types are denoted by V, B, R, C, and P/L, respectively. The classification results before and after shadow reconstruction demonstrates that the image after shadow reconstruction is beneficial to improve the classification performance. However, there is also improvement space for shadow reconstruction problems in estimating land-cover types with continuous and the transitions between shadow and nonshadow regions are carefully considered. Compared to previous shadow reconstruction algorithms, the advantages of the proposed one are that the derived shadow region is continuous and correct shadow masks and that it does not need a classification step, it allows within-class variations for one land-cover type by only selecting those candidates that are most similar to the study pixel, and it keeps good compatibility between the reconstructed shadow regions and their surrounding nonshadow regions.

V. CONCLUSION

This paper has presented a new shadow detection algorithm and a new shadow reconstruction algorithm to deal with shadows on high-resolution satellite images. Combining the shadow properties and spectral characteristics of objects, we have proposed to utilize thresholding method and morphological filtering to detect shadows. To reconstruct the underlying scene pixels of shadows, we have developed a shadow reconstruction algorithm based on the example learning method and an MRF. The execution time of the reconstruction procedure for the aforementioned three images is about 1–3 min on a computer with 2.33-GHz CPU and 3.00-GB RAM. For larger scenes, the processing time should increase quickly with the growth of data amount. It is worth mentioning that we considered the effects of high-brightness areas on the adjoining shadow neighbors and integrated them into the shadow mask by employing morphological operations in the shadow detection algorithm. Another matter worth noting is that we considered the compatibility between the reconstructed shadow regions and their nonshadow surroundings and improved this issue by passing messages between them in BBP procedure.

The experimental results on the QB image and the WV-2 image indicate that the proposed shadow detection algorithm can derive continuous and correct shadow masks and that the reconstructed shadow regions from the proposed shadow reconstruction algorithm are consistent with their surroundings. The classification test on images before and after shadow reconstruction demonstrates that the image after shadow reconstruction is beneficial to improve the classification performance. However, there is also improvement space for shadow reconstruction problems in estimating land-cover types with similar spectral characteristics, such as cement and road. Compared to previous shadow detection algorithms, the advantages of the proposed one are that the derived shadow region is continuous and the transitions between shadow and nonshadow regions are carefully considered. Compared to previous shadow reconstruction algorithms, the superiorities of the proposed one are that it does not need a classification step, it allows within-class variations for one land-cover type by only selecting those candidates that are most similar to the study pixel, and it keeps good compatibility between the reconstructed shadow regions and their surrounding nonshadow regions.

### TABLE II
CLASSIFICATION CONFUSION MATRIX FOR SELECTED SAMPLES FROM THE RECONSTRUCTED SHADOW REGIONS OF THE THREE IMAGES

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>V</th>
<th>B</th>
<th>R</th>
<th>C</th>
<th>P/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>300</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>101</td>
<td>8</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>R</td>
<td>0</td>
<td>10</td>
<td>102</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>39</td>
<td>40</td>
<td>117</td>
<td>0</td>
</tr>
<tr>
<td>P/L</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>83</td>
</tr>
</tbody>
</table>

Fig. 9. Classification comparison before and after shadow reconstruction. (a) Image after shadow reconstruction. (b) Classification map before shadow reconstruction. (c) Classification map after shadow reconstruction.