

# Journal of Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

## Shadow detection using object area-based and morphological filtering for very high-resolution satellite imagery of urban areas

Samara Azevedo  
Erivaldo Silva  
Marilaine Colnago  
Rogério Negri  
Wallace Casaca

**SPIE.**

Samara Azevedo, Erivaldo Silva, Marilaine Colnago, Rogério Negri, Wallace Casaca, "Shadow detection using object area-based and morphological filtering for very high-resolution satellite imagery of urban areas," *J. Appl. Remote Sens.* **13**(3), 036506 (2019), doi: 10.1117/1.JRS.13.036506.

# Shadow detection using object area-based and morphological filtering for very high-resolution satellite imagery of urban areas

Samara Azevedo,<sup>a,\*</sup> Erivaldo Silva,<sup>b</sup> Marilaine Colnago,<sup>c</sup>  
Rogério Negri,<sup>d</sup> and Wallace Casaca<sup>c</sup>

<sup>a</sup>Federal University of Itajubá, Natural Resources Institute, Itajubá, Minas Gerais, Brazil

<sup>b</sup>São Paulo State University, Department of Cartography, Presidente Prudente, São Paulo, Brazil

<sup>c</sup>São Paulo State University, Department of Energy Engineering, Rosana, São Paulo, Brazil

<sup>d</sup>São Paulo State University, Department of Environmental Engineering,  
São José dos Campos, Brazil

**Abstract.** The presence of shadows in remote sensing images leads to misinterpretation of objects and a wrong discrimination of the targets of interest, therefore, limiting the use of several imaging applications. An automatic area-based approach for shadow detection is proposed, which combines spatial and spectral features into a unified and flexible approach. Potential shadow-pixels candidates are identified using morphological-based operators, in particular, black-top-hat transformations as well as area injunction strategies as computed by the well-established normalized saturation-value difference index. The obtained output is a shadow mask, refined in the last step of our method in order to reduce misclassified pixels. Experiments over a large dataset formed by more than 200 scenes of very high-resolution images covering the metropolitan urban area of São Paulo city are performed, where the images are collected from the WorldView-2 (WV-2) and Pléiades-1B (PL-1B) sensors. As verified by an extensive battery of tests, the proposed method provides a good level of discrimination between shadow and non-shadow pixels, with an overall accuracy up to 94.2%, for WV-2, and 90.84%, for PL-1B. Comparative results also attested that the designed approach is very competitive against representative state-of-the-art methods and it can be used for further shadow removal-dependent applications. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JRS.13.036506](https://doi.org/10.1117/1.JRS.13.036506)]

**Keywords:** shadow detection; morphological filtering; high-resolution imagery; urban remote sensing.

Paper 190334 received May 5, 2019; accepted for publication Jul. 19, 2019; published online Aug. 9, 2019.

## 1 Introduction

Very high-resolution (VHR) remotely sensed data are considered as one of the most important resources of urban mapping. Urban land-use maps are very useful for monitoring changes and phenomena in urban environments, including urban sprawl, urban climate and pollution, and urban runoff.<sup>1,2</sup> Although VHR remotely sensed images could provide reliable and genuine insights regarding a certain urban area, the combination of finer resolution, low sun elevation, and tall buildings are very potential candidates for generating occlusions by shadow projections. In fact, shadows often arise when the sunlight is blocked by tall objects present at the scene. They usually cause partial or even total loss of relevant spectral information.<sup>3,4</sup> Since shadow segments are very common in urban imagery, their detection is critically important, as identifying shadow areas is the first key step in many popular applications, e.g., for automatic road extraction,<sup>5</sup> land use classification,<sup>6-8</sup> and digital surface model (DSM) generation.<sup>9</sup>

Due to many circumstances in which shadows are unwelcome, there are currently different approaches to detect shadows in the specialized literature of remote sensing. Some of them use a

---

\*Address all correspondence to Samara Azevedo, E-mail: [samara\\_calcado@unifei.edu.br](mailto:samara_calcado@unifei.edu.br)

3-D model or a DSM with *a priori* knowledge of the capturing sensor and the illumination source parameters for shadows position estimation.<sup>10,11</sup> Methods from this category usually provide reliable shadow location, but they need to circumvent the issue of getting high-quality input as an available data source. Moreover, a mismatch between VHR imagery and 3-D data could hamper the direct application of a 3-D model, especially for dense urban areas, wherein 3-D city models may fail in capturing the variability of the scene as well its relevant details. The low accuracy of DSM compared with VHR imagery when used in practice and the requirement for a definitive determination of solar positions may also lead to miss-detection of shadow areas, as previously discussed in Refs. 4 and 12.

On the other hand, property-based methods are preferred, as they do not require *a priori* information about the scene or other specific repositories of external data. They try to exploit the patterns and the inner properties of the own image. Most property-based methods<sup>3,13–15</sup> rely on lower sensor radiance to capture shadows in an aerial or satellite image. From this basic premise, shadow areas can be effectively detected, for instance, using a histogram thresholding-based approach. However, the main drawback is determining an optimal threshold and they are prone to give poor results when a more accurate labeling between shadow and nonshadow pixels is required, as the misclassification of low-brightness nonshadows objects such as road and black roofs is observed under many circumstances for VHR imagery.<sup>4,16</sup>

In addition to low intensity, shadow pixels also present higher saturation in blue-violet wavelength with modest intensity due to Rayleigh effect of the atmospheric scattering, as reported in Ref. 17. For example, Ma et al.<sup>18</sup> took the difference between intensity and saturation color channels to improve the shadow segmentation of aerial and satellite images. Tsai<sup>19</sup> improved the results achieved in Ref. 17, wherein spectral ratio is computed between the hue and the intensity component of the image. They employ a successive thresholding scheme rather than a static threshold, as also conducted in Ref. 20. Another aspect addressing color space processing is that other nonconventional spaces such as C1C2C3<sup>6,21</sup> and CIELAB<sup>22</sup> have been successfully tested for shadow detection, as the detectors take advantage of the shadow properties in color spaces that are invariant with respect to (w.r.t.) radiance and chromaticity. In general, these color systems are not only invariant to lighting conditions but also to orientation of the object surface and viewing conditions. Despite the benefits of using invariant color models, radiometric corrections are not properly managed by methods purely inspired on color properties, hence making these methods very sensitive to sky illumination, misclassification, and false detection.<sup>4,6,22,23</sup>

Detecting shadows in VHR images is a challenging task and it remains an open problem for dense environments where the discrimination of the objects from the scene is very critical and difficult to be performed in practice, especially by nonsupervised paradigms.<sup>4,12</sup> A robust method should guarantee independency of the material reflectance and a low necessity of additional input data, being also capable of handling spectral features and contextual information simultaneously. Bearing this in mind, in this study, an automatic shadow detection method is presented. The proposed approach unifies morphological filtering, shadow properties, and invariant color models into a general approach for detecting shadows in VHR remotely sensed images. First, radiometric corrections are applied like a preprocessing step, proceeding with the pansharpening of the input data so as to enhance the spatial information for the multispectral images. Next, morphological attribute operators drive the area-based filtering of our method for shadow detection. These operators make use of multilevel analysis based on tree representations to capture a set of representative geometric features from the image.<sup>24,25</sup> In the last step, the output shadow mask is computed using masks intersections containing potential shadow pixels in order to rule out false positives definitively. From a comprehensive battery of tests involving a large dataset of shadow-damaged images and their ground-truth annotations, our method demonstrated a high level of accuracy in separating shadow and nonshadow areas, besides requiring no additional input except the original image. Moreover, comparative results indicated that the present method outperforms two well-established shadow detection methods under performance evaluation metrics popularly found in the literature.<sup>4</sup>

This paper is organized as follows. Section 2 describes the proposed framework for shadow detection. The dataset, performance measures followed by the experimental results are presented and discussed in Sec. 3. Finally, this paper concludes in Sec. 4.

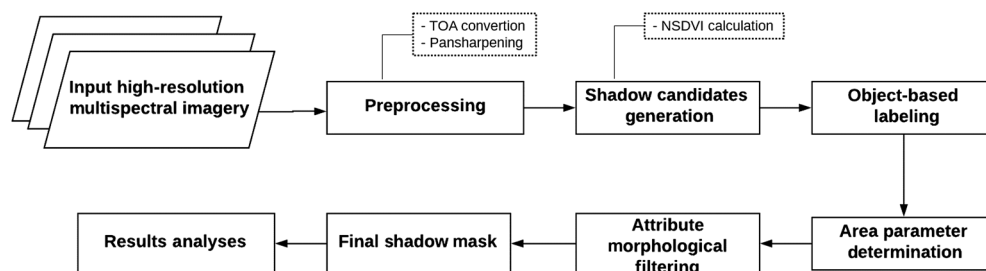
## 2 Proposed Method

Property- and physical-based methods have delivered the best performances among several shadow detection methods, as claimed in the survey paper.<sup>4</sup> Once these methods are considered as the most effective shadow detection approaches, a new framework for shadow detection is presented in this work, which is inspired by the property-based strategy. Our methodology (see Fig. 1) does not require any user intervention and comprises three major steps: data preprocessing, shadow candidates generation (which includes both the tasks of object-based labeling and area parameter estimation), and shadow mask refinement.

### 2.1 Preprocessing

Although shadows do not receive sunlight directly, they usually have sensor radiance registered due to atmospheric and environment scattered light contribution.<sup>4</sup> As a result, shadows can potentialize particular atmospheric effects. Due to this fact, the preprocessing step aims at reducing the atmospheric distortion caused to the multispectral WorldView-2 (WV-2) and Pléiades-1B (PL-1B) scenes, as detailed in Sec. 3.2. Visible and near-infrared original bands were converted from digital numbers (DN) to top-of-atmosphere (TOA) reflectance values using the metadata available from both sequences of scenes. Such a conversion was performed by means of the radiometric calibration module, available in the environment for visualizing images (ENVI) remote sensing tool. Although the data calibrated to TOA reflectance can reduce the scattered light atmospheric effect in the shadow pixels, these damaged areas still preserve some response due to environment contribution from the objects present at dense urban scene. Moreover, after preprocessing the shadow regions, they remain darker than their sunlit surrounding pixels, which potentialize the discrimination of the shadows.

After the previous conversion, the pansharpening of multispectral (MS) and panchromatic (PAN) bands was carried out for each dataset separately in order to acquire high-spatial resolution MS bands and to take advantage of the shadow properties in an invariant space color. Additionally, the improvement of spatial resolution becomes the MS data compatible with the PAN band used in the morphological context analysis. Two methods of “component substitution”—principal component analysis (PCA), and Gram–Schmidt (GS)<sup>26,27</sup> were analyzed and tested. No band limitations were observed in both methods. Technically speaking, these methods transform multivariate data as correlated variables into uncorrelated variables. The difference remains in the fact that PCA retains the maximum information into the first component, whereas GS distributes the average information among all the generated components. The PAN image is replaced with the first component from the PCA method, while for GS method, replaced by the first GS band. Finally, the inverse transformation is applied to obtain the pan-sharpened spectral bands, which provide an enhancement between the different types of land cover that are clearly seen when compared to the original MS image. PCA and GS pansharpening are performed in ENVI software so that no band registration was necessary, due to both datasets have georeferenced bands as primarily given by the same sensor.



**Fig. 1** Main steps of the proposed shadow detection approach.



### 2.1.1 Quality assessment of the pansharpened results

The most suitable fusion method was selected from an exhaustive quantitative and qualitative evaluation. The qualitative assessment involves visual inspections of the fused images with original MS bands, aiming at checking the improvement of information, visibility, texture, contrast, and detail preservation. Due to the limitation of visual inspections occasioned by human's subjective judgment, a quantitative evaluation was also carried out. The quantitative evaluation was conducted by metrics that compare the spectral similarity of the images merged with the original bands. The lack of a reference image is a disadvantage in this assessment since an ideal image with high spatial and high spectral resolution is the goal of the fusion. As the fused bands no longer presented the same spatial resolution of the original bands, a resampling process on the fused ones was necessary to make compatible and comparing the spectral similarity between them. The nearest neighbor interpolation method was selected for both, pansharpening and resampling the data to avoid new brightness values that do not correspond to the original scene.<sup>28</sup>

A MATLAB routine was implemented to run and calculate the metrics used in the quantitative evaluation. The quality measures chosen for assessment were the following: correlation coefficient (CC),<sup>29</sup> root-mean-square error (RMSE),<sup>30</sup> relative dimensionless global error in synthesis (ERGAS),<sup>31</sup> and universal image quality index (UIQI).<sup>32</sup> Table 1 summarizes the mathematical expressions of the adopted measures. The original bands were used as reference so that they are compared to the resampled pansharpened results.

These selected quality indicators provide the statistical spectral similarity and fidelity between the MS original and fused images. The lower ERGAS and RMSE values are, the higher the spectral quality and the lower the distortion caused by the fusion process, respectively. Regarding CC and UIQI indices, they vary from  $-1$  to  $1$ , and the ideal value  $1$  is obtained, if and only if, the original and fused images are identical for all the pixels. After processing both satellite datasets, their outputs were considered as input to the next step of our approach.

## 2.2 Shadow Candidate Generation

The next step of our method establishes the detection of dark pixels located in shadow areas. The preprocessed PAN and MS bands, as previously obtained, are the unique data source required for our approach. Also the method is fully automated and integrates the spectral characteristics of the shadows as part of specific invariant color models together with contextual analysis to exploit low responses in shadow areas, including complex environments covered by VHR remote sensing images as studied in this work.

**Table 1** Quality metrics for pansharpening assessment.

Function	Equation	Description
CC	$CC = \frac{\sigma_{r,f}}{\sqrt{\sigma_r \sigma_f}}$	where $\sigma_{r,f}$ is the covariance between the reference and fused image, $\sigma_r$ is the standard deviation of the reference image, and $\sigma_f$ is the standard deviation of the fused image
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N [I_F(i,j) - I_R(i,j)]^2}{M \times N}}$	where $I_R(i,j)$ represents pixels of the reference image, $I_F(i,j)$ denotes the pixels from fused image, and $M \times N$ is the image size
ERGAS	$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{RMSE_i^2}{\mu_{m_i}^2} \right)}$	where $h$ is the resolution of PAN, $l$ is the resolution of MS image, $\mu_m$ denotes mean radiance of each fused band, and $N$ represents the number of spectral bands.
UIQI	$UIQI = \frac{\sigma_{r,f}}{\sigma_r \sigma_f} \times \frac{2\mu_r \mu_f}{\mu_r^2 + \mu_f^2} \times \frac{2\sigma_r \sigma_f}{\sigma_r^2 + \sigma_f^2}$	where $\mu_r$ and $\mu_f$ are the mean of the reference and fused image, $\sigma_{r,f}$ is the covariance between the reference and fused image, $\sigma_r$ is the standard deviation of the reference image, and $\sigma_f$ gives the standard deviation of the fused image

From the first shadow's particularity, which assumes that shadow pixels have the lowest sensor radiance values than their sunlit neighborhoods,<sup>4</sup> the morphological black-top-hat (BTH) transformation<sup>24</sup> was used to extract these dark features. In more mathematical words,

$$\text{BTH}(f) = \phi(f) - f, \quad (1)$$

where  $\phi$  represents the closing operator that simplifies the input image, i.e., it eliminates regional minima (darker objects than surrounding regions) from the scene. In many cases, morphological filtering operators have been used mainly to get rid of noisy and isolated regions,<sup>10,33,34</sup> and for smoothing the shape objects.<sup>22</sup> Conversely, BTH can also be applied to detect shadows effectively, as the technique works well to tag valleys. It consists of taking regional minima from the residuals of the arithmetic difference between closing transformation and the input image. As a well-established and versatile morphological filtering used by the image processing community, this operator considers a set of known shapes called structuring elements (SE). The SE can be defined based on the shapes and the length of the objects to be extracted. However, in the case of shapeless shadows of various sizes, closing the image is infeasible. Therefore, instead of only closing the image, the BTH area closing transformation was considered in our methodologic approach, i.e.,

$$\text{BTH}(f) = \phi_{\lambda}(f) - f, \quad (2)$$

where  $\phi_{\lambda}$  gives the area closing operator for a gray scale image, which is computed by

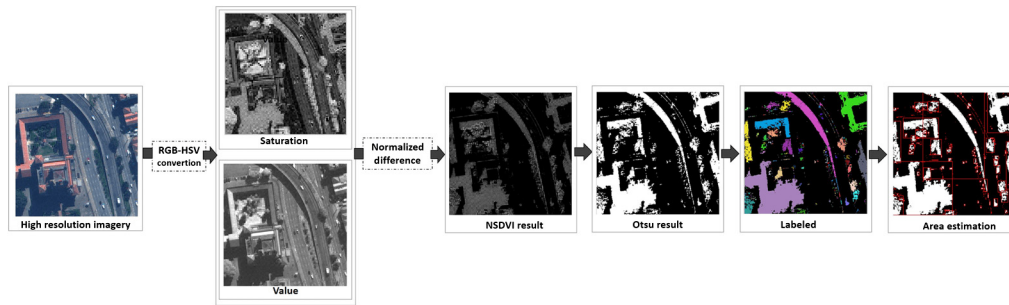
$$\phi_{\lambda} = \bigwedge_i \{\phi_{B_i}(f)\}. \quad (3)$$

This operator takes the infimum ( $\bigwedge$ ) of all closings with connected SE whose sizes, in number of pixels, are equal to  $\lambda$ .<sup>24</sup> In other words,  $\phi_{\lambda}$  removes connected dark components with areas smaller than the threshold  $\lambda$ , being recovered when the arithmetic difference regarding the input image  $f$  is achieved by the BTH. The principle of threshold superposition<sup>24</sup> is used for gray-scale images, i.e., the output is the sum of all filtered input threshold images, as a gray-scale image can be expressed as the sum of its binary thresholds. The use of attribute filtering allows for more flexibility when exploiting contextual relations and spatial information under morphological operators based on fixed SE. In addition, as a connected filter, the area closing operator only acts on the flat zones of the image (a set of connected iso-intensity pixels), avoiding shape distortion and the loss of spatial characteristics of the objects of interest.<sup>35</sup>

Azevedo et al.<sup>36</sup> set the area parameter empirically, making their approach semiautomatic instead of nonsupervised. This parameter depends on the resolution of the target images so that the connectivity of the image objects must be large enough when coping with VHR images. If the parameter is not accurately chosen, undesirable noise and over detection of nonshadow pixels can arise in the shadow mask. Therefore, in this study, the normalized saturation-value difference index (NSDVI)<sup>18</sup> was chosen and manipulated to drive the estimate of the shadow areas, thus delivering a fully automatic approach for shadow detection.

### 2.3 Object-Based Area Parameter Determination

Before obtaining a refined shadow mask, shadow and nonshadow pixels are presegmented by employing the NSDVI shadow index. Figure 2 shows the proposed automatic scheme to get the area parameter, as guided by NSDVI computation. First, the normal color composite from the preprocessed MS bands is converted to the HSV color space. HSV channels enable a better separability between chromaticity and illumination properties. As aforementioned, shadows appear highly saturated with low values on invariant color models due to the Rayleigh effect of atmospheric scattering.<sup>17</sup> Therefore, based on each HSV channel separately, a normalized difference between channels  $S$  and  $V$  is constructed as follows:



**Fig. 2** Area parameter estimation driven by shadow spectral index NSDVI.

$$\text{NSDVI} = \frac{S - V}{S + V}. \quad (4)$$

A binary NSDVI mask carrying shadows pixels is obtained by Otsu's method.<sup>37</sup> The Otsu's algorithm is applied to select the best threshold to segment pixels identified as shadow by the NSDVI while keeping our approach fully user independent. The result is a binary image where white pixels represent the areas identified as shadow by the index.

Next, the NSDVI output is labeled by computing the “mmlabel” function in the standard mode, which is a morphological operator as properly described in Ref. 38. This function creates a labeled image from the connect components present in the NSDVI mask. After generating the shadow segments in the NSDVI mask, the area is definitively calculated from these segments by applying the eight-connected “mmblob” function,<sup>38</sup> which measures the areas' sizes from the labeled image. From the NSDVI outcome, one must notice that there is an ambiguity between nonshadow pixels due to the occurrence of patterns with similar behavior on the HSV space color. However, as previously mentioned, the choice of the area parameter must be enough to extract all shadow pixels. Moreover, the overestimation observed in the NSDVI mask is suitable so that an effective coverage of large areas in the image is accomplished.

## 2.4 Definitive Shadow Mask Postprocessing

In the last step of our approach, the object-based area parameter determined in the previous step is provided as input to the BTH when applied to the PAN band over all the images. This is done for both datasets in order to get the definitive shadow masks from the processed images.

Once the images from WV2 dataset have been acquired as a resampled 8-bit scene, and due to the presence of spectral distortions that may be caused during the image fusion process, some vegetation areas may appear labeled as shadow in a few outputs of the WV2 dataset. To deal with these particular cases, we apply a postprocessing step in the WV2 subsets to eliminate such mislabeled pixels. More specifically, the normalized difference vegetation index (NDVI)<sup>39</sup> was applied to take off the vegetation areas misclassified as shadows in the NSDVI outputs. Next, the definitive WV2 shadow mask (SMask) is determined by computing the intersection between the shadow candidates (as generated in the previous step using the BTH), and the difference of NSDVI and NDVI masks, as properly given by the following equations:

$$\text{SMask} = \text{BTH}(f) \cap (\text{NSDVI} - \text{NDVI}), \quad (5)$$

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{R}}}{\rho_{\text{NIR}} + \rho_{\text{R}}}, \quad (6)$$

where  $\rho_{\text{NIR}}$  and  $\rho_{\text{R}}$  are TOA reflectance at near-infrared and red bands of the original images, respectively. It worth mentioning that the Otsu's method was also computed to promote an automatic thresholding to the NDVI results and, thus acquiring a binary mask of vegetation pixels. Finally, following Ref. 40, we have considered as nonshadow areas very small artifacts captured by the method so that we filter out from the SMask outliers less than 3 pixels.

### 3 Experimental Results and Discussion

This section is devoted to present and discuss the experimental results obtained by our approach in a large dataset of an urban environment that covers more than 200 scenes of VHR remotely sensed images. The metrics and the dataset considered in this study are presented in the following.

#### 3.1 Performance Evaluation Metrics

In our experiments, the assessment of the results is performed by means of well-known metrics as established by Congalton,<sup>41</sup> more specifically producer's accuracy ( $P_s$ ), user's accuracy ( $U_s$ ), and overall accuracy ( $\tau$ ). These metrics have been adopted by most shadow detection approaches as benchmarks for RS imagery.<sup>4,19</sup> The metrics are computed as follows:

$$P_s = \frac{TP}{TP + FN}, \quad (7)$$

$$U_s = \frac{TP}{TP + FP}, \quad (8)$$

$$\tau = \frac{TP + TN}{TP + TN + FP + FN}, \quad (9)$$

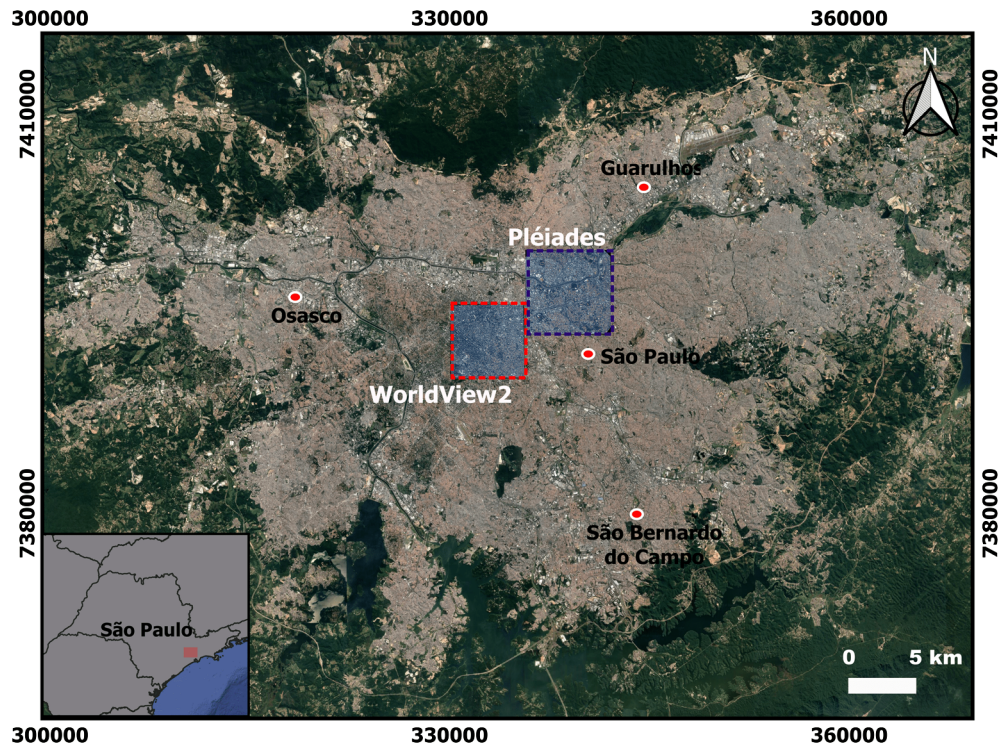
where TP (true positive) represents the number of shadow pixels correctly detected by our approach, FP (false positive) gives the number of nonshadow pixels labeled as shadow, FN (false negative) indicates the number of shadow pixels labeled as nonshadow, and TN (true negative) provides the nonshadow pixels correctly labeled as nonshadow by our approach. The indices are based on a pixel-by-pixel comparison involving the obtained detection results with a ground truth (GT) shadow mask created by visual inspection for each image of the studied dataset. To establish a valid comparison, a one-pixel acceptance buffer was created by a dilatation over detected and GT binary images to avoid inconsistency with coincident pixels. This tolerance is of paramount importance due to consuming and costly task of GT building to provide a benchmark mask which is not fully error-free. In summary, the overall accuracy refers to the percentage of shadow and nonshadows pixels correctly detected by the proposed approach, in which there is a trade-off between over- and under-detection as given by producer's and user's accuracies, respectively.

#### 3.2 Dataset

The validation of the proposed method is performed on WV-2 and PL-1B high-spatial satellite imagery. Both scenes cover an area of 25 km<sup>2</sup> from São Paulo city, which is characterized by typical urban land cover types, including high-rise buildings, residential and commercial buildings, roads, and vegetation such as grass and trees. The scenes' location of the studied area are shown in red (WV-2) and blue (PL-1B) polygons in Fig. 3. The dataset includes a PAN band and four traditional MS bands (red, green, blue, and near-infrared) for both satellite imagery with similar configuration (spatial resolution of the PAN image is 0.50 m while for the MS image is 2.00 m). The PL-1B image product was acquired with an original dynamic range of 12 bits, whereas WV-2 was resampled at 8 bits.

A database with 215 subsets from original scenes was created in order to assess the performance of the proposed methodology. The subsets were automatically generated with a fixed size of 300 × 300 pixels, which has improved the visual inspection of results while enabling the GT construction. The GTs for these images were created by an expert that manually detected shadow regions, whose process is time-consuming and limits the amount of available data. Each subset has an identification number (ID) used to automatically read the input data. This ID is related to the subset position at the scene based on image dimension (rows and columns). Therefore, since PAN and MS pansharpened have the same size and spatial resolution, the same number of



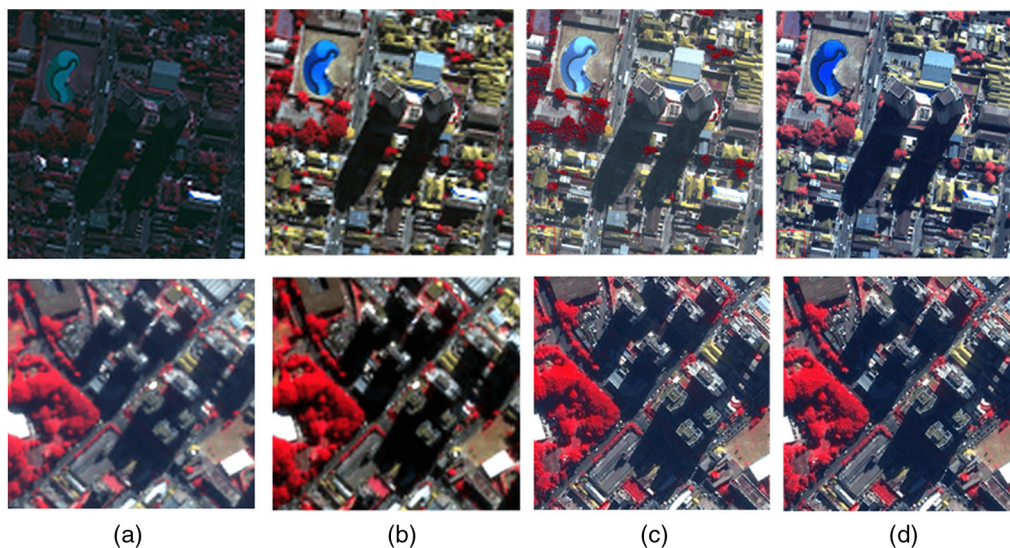


**Fig. 3** Location of WorldView2 (red dotted polygon) and Pléiades (blue dotted polygon) scenes in the study area from São Paulo city that were used for experimental evaluation and analysis.

subsets is obtained covering the same geographical location and guaranteeing the comparison between GT and detection results for quantitative evaluation.

### 3.3 Preprocessing Performance Analysis

By visually comparing the results from the preprocessing with the original false color composites (Fig. 4), it seems that the image targets can successfully be improved in terms of visual



**Fig. 4** The result of preprocessing candidate approaches when applied to WV-2 (upper) and PL-1B (lower) images: (a) original image in DN quantities, (b) TOA reflectance image, (c) the result of PCA fusion method, and (d) result of GS fusion method.



**Table 2** Quality-based metrics for the assessment of fused images.

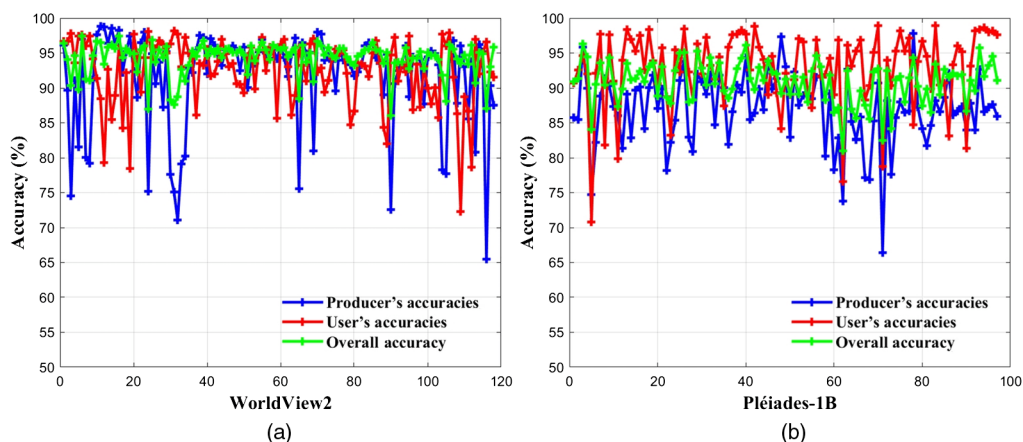
Quality metrics	WV-2		PL-1B	
	PCA	GS	PCA	GS
CC	0.8614	0.8643	0.8874	0.8859
RMSE	0.0025	0.0025	0.0269	0.0270
ERGAS	0.0920	0.0901	0.0567	0.0569
UIQI	0.8608	0.8642	0.8874	0.8858

aspects so that the distinction of image elements is enhanced. The original raw image depicted in Fig. 4(a) appears darker (upper WV-2 image). The presence of haze (lower, PL-1B image) is also observed. After the conversion of input data to TOA reflectance, an improvement w.r.t. visual quality is reached, as illustrated in Fig. 4(b). Relevant image elements are best distinguishable from the image in TOA quantities. Pansharpener results from the two methods are shown in Figs. 4(c) and 4(d), respectively. An insertion of spatial details can be noticed from both fusion methods. Moreover, fine details are more noticeable and a better discernment of individual elements than the original MS data is now given.

Still concerning the visual exploration of the results, one can see that there is a sharp color intensity due to a spectral distortion from PCA than GS fusion method in the WV-2 images. From the outputs of PL-1B data, visual comparisons lead to similar spectral characteristics. Based only in the visual aspects, it is not a straightforward task determining which one of the fusion methods provides better spatial quality images and noncolor distortion. Thus quantitative assessments are required. The results of the quality evaluation of the pansharpened data are shown in Table 2. From the quality indices, close values were delivered by the PCA and GS fusion methods for both datasets with lower RMSE and ERGAS values. This implies a higher spectral quality for the merged images while higher UIQI and CC scores indicate high correlation and spectral fidelity with respect to the original multispectral data. Therefore, based on performance evaluation as well as the visual inspection, GS pansharpened method was the one selected as input to our methodology for the goal of shadow detection in the next step.

### 3.4 Shadow Detection Performance Evaluation

Figure 5 shows the performance of the proposed shadow detection approach under the perspective of the metrics previously described in Sec. 3.1. The average values for quantitative

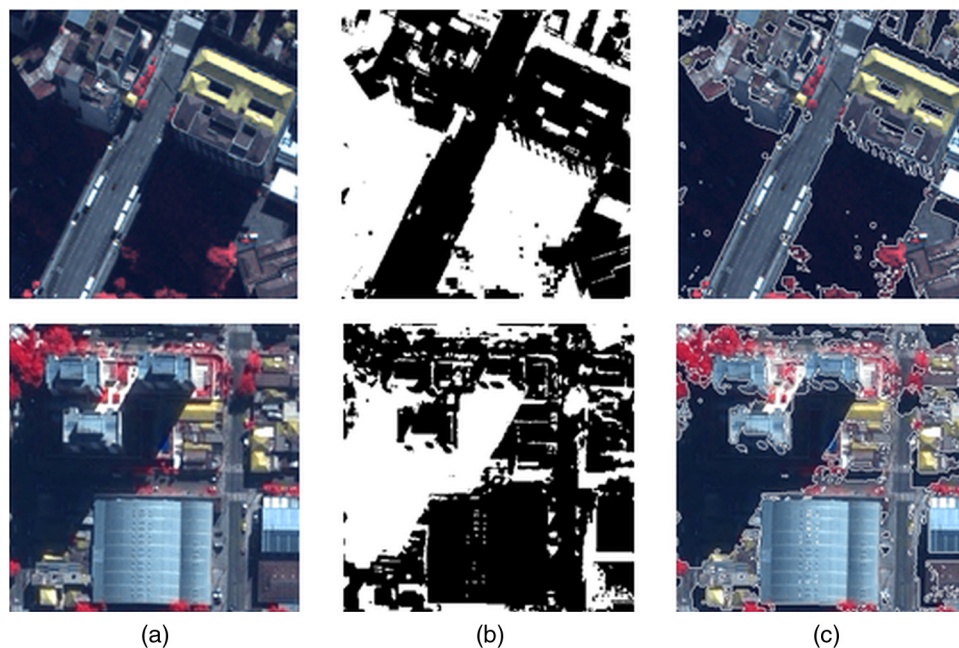
**Fig. 5** Shadow detection performance of the proposed method for (a) WV2 and (b) PL1B dataset.

**Table 3** Average of accuracy performance metrics.

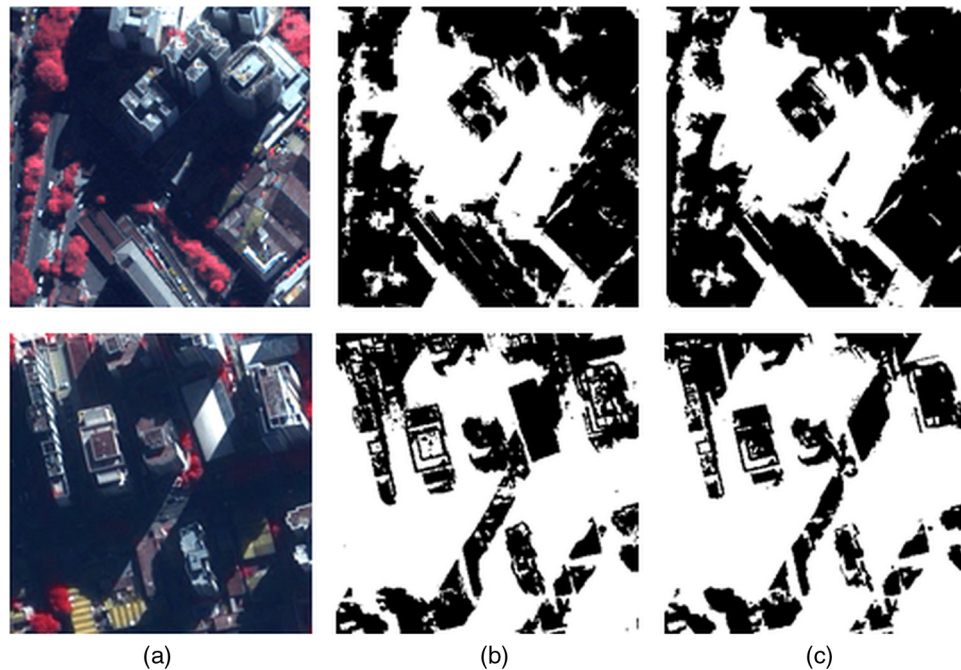
Dataset	Number of images	$P_S$ (%)	$U_S$ (%)	$\tau$ (%)
WV-2	118	$91.64 \pm 6.83$	$92.48 \pm 4.72$	$94.20 \pm 2.32$
PL-1B	97	$86.96 \pm 5.12$	$92.82 \pm 5.39$	$90.84 \pm 2.98$

evaluation are listed in Table 3. From the obtained results, one may notice that an average overall accuracy over 90% was reached for both datasets, indicating a high performance for detected shadow pixels by the method. When applied to the WV-2 images, our approach presented a better shadow discrimination, reaching an average of 94.20% with low variance between indices, whereas PL-1B images reached up to 91%. In general, producer's and user's accuracies have demonstrated a good performance with an average over 90%, except for producer's accuracy in PL-1B images which reached up to 87%. The lower producer's accuracy means that pixels identified as shadow in GT and considered as true were not properly detected as shadow by our approach and have not matched with GT. On the other hand, high average user's accuracies up to 92% for both datasets attest how accurate the proposed method is in separating shadow from nonshadow. The capability of the present approach to avoid false detections is noticeable since most shadow detection methods misclassify dark pixels with shadows pixels.

Figure 6 shows the best shadow detection output obtained by the current method. The overall accuracy reached up to 97% and 96% for WV-2 and PL-1B, respectively. Large cast shadows caused by many high-rise buildings were well identified, as well as shadows from residential buildings, bridges, vegetation, and cars. Undetected shadow areas under high-reflectance objects are observed, which may explain the lower producer's accuracies, as discussed before. Additional results highlighting the satisfactory accuracy reached by the shadow detection method are depicted in Fig. 7. Notice that the current approach distinguishes size-independent shadows without requiring any user intervention, besides revealing high scores for the evaluation metrics.



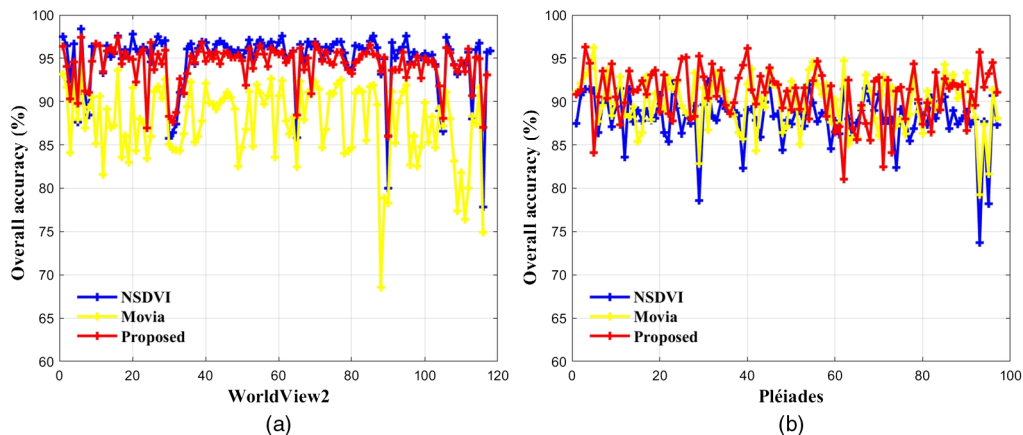
**Fig. 6** Visual results automatically took from the shadow detection method: (a) original false color composite WV-2\_12\_29 (upper) and PL-1B\_14\_44 (lower) images, (b) binary of shadow detection result, and (c) overlaid shadow detection result over the original image.



**Fig. 7** Shadow detection results from subsamples: (a) original false color composite WV-2\_29\_27 (upper) and PL-1B\_48\_40 (low) images, (b) binary of shadow detection result, and (c) GT.

### 3.5 Comparative Evaluation with State-of-the-Art Shadow Detection Methods

In order to compare the achieved results by the proposed method against state-of-the-art shadow detection approaches, we took the well-established NSDVI index and the framework proposed by Ref. 6. Both methods have been selected because they are property-based and fully automatic approaches, such as our approach. It is worth mentioning that the method given in Ref. 6 was implemented based on the original paper description, as no feedback after contacting the authors was obtained. Figure 8 presents the comparative evaluation of overall accuracy obtained by each image from the assessed methods when they are run to both datasets. One can verify that the proposed approach reached a high performance in both datasets while the remaining ones produced inconstant results. Although the satisfactory performance generated by the NSDVI outputs for the WV-2 images, they got worse when processing PL-1B images, diverging from the expected behavior as verified by the methods in Ref. 6.



**Fig. 8** Comparative results under the performance metrics collected for each image and method when assessed over (a) WV-2 and (b) PL-1B dataset.



**Table 4** Average of performance metrics achieved by each method from the tested dataset.

Method	WV-2			PL-1B		
	$P_S$ (%)	$U_S$ (%)	$\tau$ (%)	$P_S$ (%)	$U_S$ (%)	$\tau$ (%)
NSDVI	<b>95.51</b> $\pm$ 6.52	89.53 $\pm$ 5.78	<b>94.46</b> $\pm$ 3.38	80.19 $\pm$ 4.92	94.02 $\pm$ 5.82	88.13 $\pm$ 2.86
Movia et al.	86.67 $\pm$ 6.68	81.02 $\pm$ 12.23	87.53 $\pm$ 4.44	82.81 $\pm$ 3.31	<b>95.78</b> $\pm$ 4.44	89.80 $\pm$ 3.05
Proposed	91.64 $\pm$ 6.83	<b>92.48</b> $\pm$ 4.72	94.20 $\pm$ 2.32	<b>86.96</b> $\pm$ 5.12	92.82 $\pm$ 5.39	<b>90.84</b> $\pm$ 2.98

Table 4 presents the average and standard deviation of the performance evaluation metrics collected by each method, including the proposed one over both datasets. One can check from Table 4 that regardless of the dataset, the proposed shadow detection approach presented regularity and stable results. Even though NSDVI index resulted in the best overall accuracy for WV-2 images (up to 95%), the low user's accuracy indicates the presence of false positives which is a disadvantage in considering only this index for shadow segmentation since it may lead others image elements to be tagged as shadow. The performance values achieved by Ref. 6 attested good accuracy in the task of detecting only shadow pixels, especially for PL-1B images. However, it presents under-detection of shadow pixels as demonstrated by low producer's accuracy. On the other hand, our method remained stable enough in almost all the measurements, producing high scores especially for the PL-1B dataset.

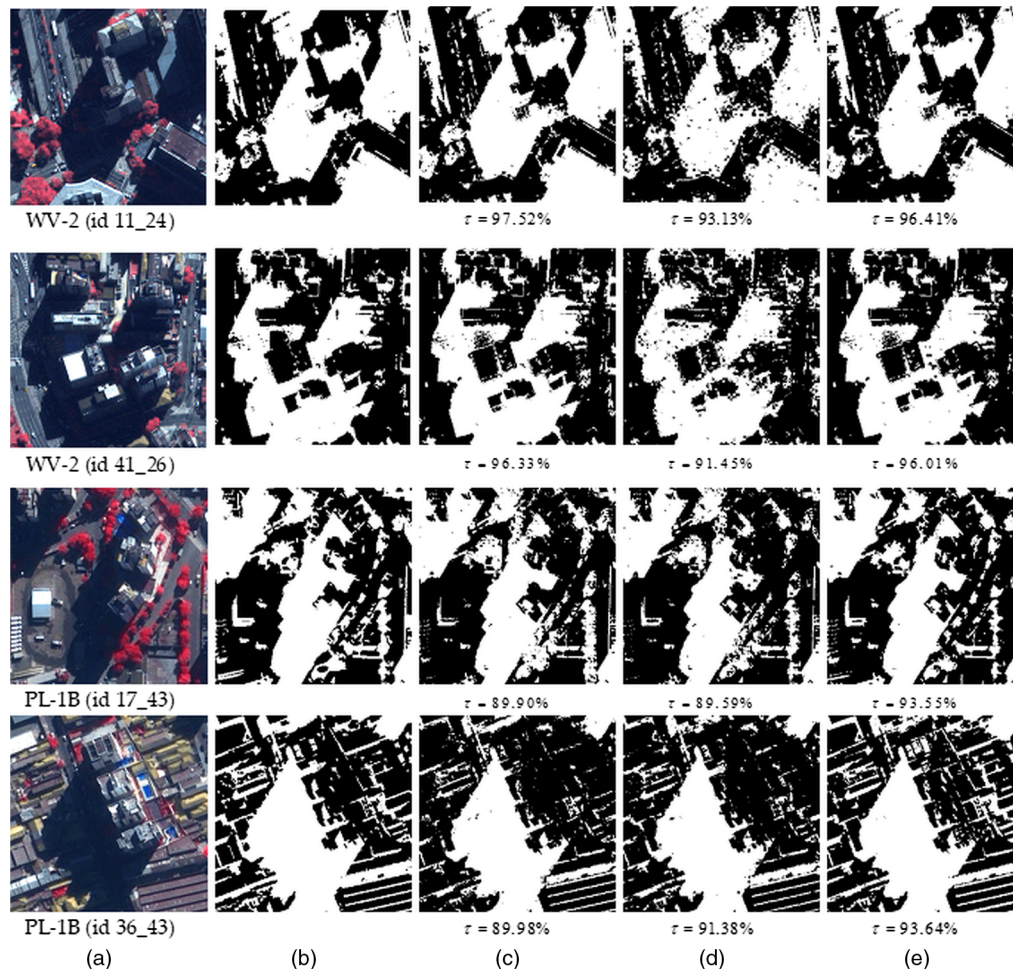
**Fig. 9** Visual comparative results of shadow detection methods assessed: (a) original false color composite images, (b) GT, (c) results of NSDVI, (d) Movia et al.,<sup>6</sup> and (e) proposed method.

Figure 9 brings a visual comparison among our approach and the other shadow detection methods. Considering the GT data [Fig. 9(b)], a more general inspection gives the impression that the results obtained by NSDVI [Fig. 9(c)], method in Ref. 6 [Fig. 9(d)], and our technique [Fig. 9(e)] are somewhat similar. But, looking the outputs one-by-one, the results achieved by Ref. 6 [Fig. 9(d)] demonstrated to be noised, in particular for WV-2 dataset in which the overall accuracy produced the lowest scores (93% and 91% for both scenes, respectively). On the other hand, the NSDVI [Fig. 9(c)] and the proposed method [Fig. 9(e)] for the same images reached up to 97% and 96% for the first image and a similar overall accuracy up to 96% for the secondary tested image.

Regarding PL-1B scenes, our method delivered the best overall accuracy, reaching up to 94% for both exposed scenes, whereas the other techniques reached the range around 90% to 91%. Based on experimental results, the main drawback of NSDVI and the method described in Ref. 6 is the over detection of high-reflectance pixels under shadow presence and the false labeling of the vegetation area. Although marking GT pixels over vegetation canopy is a meticulous task and may contain errors, when visually inspecting the NSDVI<sup>6</sup> results, one can conclude that there is confusion in labeling shadows (e.g., see the trees in PL-1B id 17\_43), leading to a reduction of the accuracy indices. In contrast, our approach leads to better visual and quantitative results (in overall), presenting high overall accuracy and low false detection rates even for complex urban environments such as the one evaluated in this study.

## 4 Conclusions

Shadow detection is a challenging task for many remote sensing applications, especially when one needs to deal with complex urban areas. In this paper, an area-based shadow detection method was proposed to automatically identify shadows in VHR multispectral data of dense urban regions. An area-based BTH transformation was applied to produce shadow candidates that are subsequently refined by a spectral vegetation index. The preprocessing step also contributed to enhancing the discrimination of individual elements with good correlation between the original bands and the low spectral distortion, allowing for the generation of spectral indices totally compatible with the PAN band.

When compared with the state-of-the-art methods such as presented in Refs. 6 and 18, and by employing a large dataset that comprises more than 200 subsets taken from WV-2 and PL-1B repositories, the proposed method achieved an average of overall accuracy up to 90% for both evaluated datasets. Low false detection and omission errors ( $\sim 7\%$ ) have been observed for the results of our approach, which was identified by the user's accuracy, while an over detection of high-reflectance pixels under shadow presence and confusion in vegetation areas was found by the other two specific methods from the literature. Overall, our approach reached high overall accuracy and low false detection rates, attesting its robustness when labeling shadow areas under vegetation and low-intensity objects such as cars and water, while other methods produced misclassification or noise.

In summary, the proposed method clearly distinguished shadow pixels from the rest of the image elements in a complex urban environment only by assuming the original data as available in the single scene. Moreover, the method is entirely user-independent and can succeed in many urban images wherein detecting shadows among several similar objects is of paramount importance.

## Acknowledgments

The authors would like to thank São Paulo Research Foundation (FAPESP, Grant No. 2013/25257-4) for the financial support given to the development of this work. The authors declare no conflicts of interest.

## References

1. A. A. Peeters, "GIS-based method for modeling urban-climate parameters using automated recognition of shadows cast by buildings," *Comput. Environ. Urban Syst.* **59**, 107–115 (2016).



2. B. Huang, B. Zhao, and S. Yimeng, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sens. Environ.* **214**, 73–86 (2018).
3. P. M. Dare, "Shadow analysis in high-resolution satellite imagery of urban areas," *Photogramm. Eng. Remote Sens.* **71**, 169–177 (2005).
4. K. Adeline et al., "Shadow detection in very high spatial resolution aerial images: a comparative study," *ISPRS J. Photogramm. Remote Sens.* **80**, 21–38 (2013).
5. M. Li et al., "Region-based urban road extraction from VHR satellite images using Binary Partition Tree," *Int. J. Appl. Earth Obs. Geoinf.* **44**, 217–225 (2016).
6. A. Movia, A. Beinat, and F. Crosilla, "Shadow detection and removal in RGB VHR images for land use unsupervised classification," *ISPRS J. Photogramm. Remote Sens.* **119**, 485–495 (2016).
7. R. G. Negri, E. A. Silva, and W. Casaca, "Inducing contextual classifications with Kernel functions into support vector machines," *IEEE Geosci. Remote Sens. Lett.* **15**, 962–966 (2018).
8. C. Mora et al., "Land cover classification using high-resolution aerial photography in adventdalen, Svalbard," *Geografiska Ann. Ser. A, Phys. Geogr.* **97**(3), 473–488 (2015).
9. R. A. Oliveira, A. M. Tommaselli, and E. Honkavaara, "Generating a hyperspectral digital surface model using a hyperspectral 2-D frame camera," *ISPRS J. Photogramm. Remote Sens.* **147**, 345–360 (2019).
10. K. Zhou, R. Lindenbergh, and B. Gorte, "Automatic shadow detection in urban very-high-resolution images using existing 3-D models for free training," *Remote Sens.* **11**(1), 72 (2019).
11. G. Tolt, M. Shimoni, and J. Ahlberg, "A shadow detection method for remote sensing images using VHR hyperspectral and LIDAR data," in *Proc. 2011 IEEE Int. Geosci. and Remote Sens. Symp. (IGARSS)*, Vancouver, Canada, pp. 4423–4426 (2011).
12. Q. Wang et al., "An automatic shadow detection method for VHR remote sensing orthoimagery," *Remote Sens.* **9**, 469 (2017).
13. Y. Chen et al., "Shadow information recovery in urban areas from very high resolution satellite imagery," *Int. J. Remote Sens.* **28**(15), 3249–3254 (2007).
14. F. Yamazaki, W. Liu, and M. Takasaki, "Characteristics of shadow and removal of its effects for remote sensing imagery," in *Proc. Int. Geosci. and Remote Sens. Symp. (IGARSS)*, Cape Town, South, Africa, Vol. 4, pp. 426–429 (2009).
15. T. Statella and E. A. Silva, "Shadows and clouds detection in high resolution images using mathematical morphology," in *Proc. Pecora 17—The Future of Land Imaging*, Denver, Colorado (2008).
16. W. Huang and M. Bu, "Detecting shadows in high-resolution remote-sensing images of urban areas using spectral and spatial features," *Int. J. Remote Sens.* **36**(24), 6224–6244 (2015).
17. A. M. Polidorio et al., "Automatic shadow segmentation in aerial images," in *Proc. XVI Brazilian Symp. Comput. Graphics and Image Process.*, Sao Carlos, Brazil, pp. 270–277 (2003).
18. H. Ma, Q. Qin, and X. Shen, "Shadow segmentation and compensation in high resolution satellite images," in *Geosci. and Remote Sens. Symp. (IGARSS)*, vol. 2. (2008).
19. V. J. D. Tsai, "A comparative study on shadow compensation of color aerial images in invariant color models," *IEEE Trans. Geosci. Remote Sens.* **44**(6), 1661–1671 (2006).
20. K. L. Chung, Y. R. Lin, and Y. H. Huang, "Efficient shadow detection of color aerial images based on successive thresholding scheme," *IEEE Trans. Geosci. Rem. Sens.* **47**(2), 671–682 (2009).
21. V. Arévalo, J. Gonzalez, and G. Ambrosio, "Shadow detection in colour high-resolution satellite images," *Int. J. Remote Sens.* **29**(7), 1945–1963 (2008).
22. G. F. Silva et al., "Near real-time shadow detection and removal in aerial motion imagery application," *ISPRS J. Photogramm. Remote Sens.* **140**, 104–121 (2018).
23. J. Tian et al., "New spectrum ratio properties and features for shadow detection," *Pattern Recognit.* **51**, 85–96 (2016).
24. P. Soille, *Morphological Image Analysis*, Springer-Verlag, Berlin (2004).

25. M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "Modeling structural information for building extraction with morphological attribute filters," *Proc. SPIE* **7477**, 747703 (2009).
26. Z. Wang et al., "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.* **43**, 1391–1402 (2005).
27. T. Maurer, "How to pan-sharpen images using the Gram-Schmidt pan-sharpen method—a recipe," in *Proc. Int. Arch. Photogramm., Remote Sens. and Spatial Inf. Sci.*, Germany (2013).
28. R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 3rd. ed., Academic Press, Burlington, Massachusetts (2007).
29. B. Aiazzi et al., "Quality assessment of pansharpening methods and products," *IEEE Geosci. Remote Sens. Soc. Newsletter* **161**, 10–18 (2011).
30. L. Alparone et al., "Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.* **45**, 3012–3021 (2007).
31. I. Alimuddin et al., "Assessment of pan-sharpening methods applied to image fusion of remotely sensed multi-band data," *Int. J. Appl. Earth Obs. Geoinf.* **18**, 165–175 (2012).
32. Z. Wang and A. C. A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.* **9**, 81–84 (2002).
33. H. Song, B. Huang, and K. Zhang, "Shadow detection and reconstruction in high-resolution satellite images via morphological filtering and example-based learning," *IEEE Trans. Geosci. Remote Sens.* **52**, 2545–2554 (2014).
34. L. Lorenzi, F. Melgani, and G. A. Mercier, "Complete processing chain for shadow detection and reconstruction in VHR images," *IEEE Trans. Geosci. Remote Sens.* **50**, 3440–3452 (2012).
35. D. M. Mura et al., "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.* **48**, 3747–3762 (2010).
36. S. C. Azevedo, E. A. Silva, and M. M. Pedrosa, "Shadow detection improvement using spectral indices and morphological operators in urban areas from high resolution," in *36th Int. Symp. Remote Sens. of Environ. (ISRSE)*, Berlin (2015).
37. N. Otsu, "A threshold selection method from gray level histogram," *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979).
38. SDC Information Systems, "Morphology Toolbox for Matlab 5 User's Guide," Naperville, Illinois (2001).
39. C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.* **8**, 127–150 (1979).
40. I. Huerta et al., "Chromatic shadow detection and tracking for moving foreground segmentation," *Image Vision Comput.* **41**, 42–53 (2015).
41. R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.* **37**, 35–46 (1991).

**Samara Azevedo** received her BS degree in cartographic engineering and her MSc and PhD degrees in cartographic sciences from São Paulo State University (UNESP) in 2011, 2014, and 2018, respectively. She is a professor of geomatics at the Natural Resource Institute of the Federal University of Itajubá. Her research interests include remote sensing, cartography, image processing, and GIS.

**Erivaldo Silva** received his BS degree in cartographic engineering from UNESP in 1985, his MSc degree in geosciences from the National Institute for Space Research (INPE) in 1989, and his PhD in transports engineering from the University of São Paulo (USP) in 1995. He is currently a full professor at UNESP and research group leader of image processing at UNESP. He has experience in geosciences, acting on the following subjects: mathematical morphology, cartography, cartographic products update, and features extraction.

**Marilaine Colnago** received her BSc degree in mathematics in 2009, her MSc degree in computational and applied mathematics from UNESP in 2012, and her PhD in computer science and computational mathematics from the USP in 2017. She is currently a professor at UNESP. Her

research interests lie in image processing (image and data clustering), computational mathematics, and numerical analysis.

**Rogério Negri** received his major's degree in mathematics in 2006 from UNESP, his master's degree in 2009, and her PhD in applied computation from the Brazilian National Institute for Space Research (INPE), Brazil, in 2013. He is currently a professor at the Institute of Science and Technology, UNESP, Brazil. He has experience in pattern recognition, radar image processing, geostatistics, and GIS.

**Wallace Casaca** received both his bachelor's and master's degrees in pure and applied mathematics from UNESP in 2008 and 2010, respectively. In 2014, he pursued his PhD in computer sciences and computational mathematics at the USP. His research interests include computer vision (image segmentation and classification), computational intelligence, remote sensing, information visualization, optimization, and numerical analysis.