ABSTRACT

The presence of shadows in Remote Sensing images leads to misinterpretation of objects in several real-world applications, which includes Very High Resolution (VHR) image data from urban areas. Consequently, new methodologies are required to analyze urban data efficiently, due to the great variety of artifacts and shadows formed by elevated objects in the image. In this paper, a novel automatic shadow removal approach is proposed to recover missing information caused by shadows and other obstruction artifacts. First, an automated shadow detection method is applied by computing morphological operations between objects and their surroundings, which are combined with shadow spectral features extracted from a color space model, avoiding, this way, false detections in the shadow mask refined. Second, to recovering the missing information from the shadow guidance mask, an inpainting-inspired strategy is proposed, which unifies anisotropic diffusion, transport equation and texture synthesis into a robust and concise framework. The performance of our approach is evaluated by taking a WorldView-2 imagery, where it was found that the method achieves an overall accuracy on shadow detection up to 90%, in addition to a low rate of false detection (~20%). Moreover, the designed algorithm outperforms existing recovering techniques, providing high computational performance over VHR satellite images that could be suitable for object recognition, land-cover mapping, 3D reconstruction and, particularly, for developing countries where land use and land cover are rapidly changing with tall buildings/structures within urban areas.

KEYWORDS: Shadow detection, shadow removal, inpainting, high resolution imagery, remote sensing

INTRODUCTION

Remotely sensed (RS) data are very important information resources for several urban applications. In fact, new image processing tools and modern imaging machinery allow us to acquire high spatial and spectral features of the image, for example, to assess temporal changes of surface objects (Du et al., 2014). However, the combination of finer resolution, low sun elevation and tall buildings may lead to many undesirable artifacts (e.g., shadows), causing partial or even total loss of spectral features of the image, which can hamper the acquisition of reliable information from the analyzed image. Therefore, removing shadow effects from RS images is an important trait in real-world applications, being essential, for example, in automatic road extraction (Li et al., 2016) and land use classification (Movia et al., 2016). Despite the numerous studies that investigate the shadow presence in VHR images, the development of more effective and accurate shadow detection/removal methods still remains as an active field of study. Indeed, the use of well-known radiometric corrections is not enough to reach a good level of visual quality, since this kind of approach leads to correction methods that are very sensitive to sky illumination. Moreover, there are some disadvantages for misclassification and false detection with dark objects such as vegetation, water and dark materials (Adeline et al., 2013). Another aspect is that due to the large heterogeneity occasioned by dense urban areas, the shadows present distinct shapes and lengths, which demands the designing of specific methodologies to operate with spectral features and contextual correlations in a more accurate way.

A few image analysis approaches which address the reconstruction problem for missing areas in digital images are
found in the literature, being most of them intended to produce solutions for low- or medium-spatial-resolution images. In fact, these approaches produce solutions with missing pixels that are caused mainly due to cloud cover and sensor-specific problem. To deal with this issue, multitemporal data can be used to recover the cloud cover region by direct replacement of pixels based on their similarity w.r.t. other temporal frames. Gao and Gu (2017) proposed a multitemporal replacement method where the similar pixels are selected by tempo-spectral angle mapping (TSAM), in an attempt to recovering Landsat ETM+ SLC-off and cloud cover data. This technique can fully fill the gaps, however it tends to produce a low-quality result for cloud mask identification. For applications covering high spatial resolution images, the need for longer time series can be a hurdle due to its commercial acquisition. In addition, multitemporal replacement methods have no guarantees of temporal spectral surface, since they are prone to fail in the imminence of abrupt transformations of geographical objects (Li et al., 2016).

Still concerning the shadow obstruction problem, inpainting techniques that are inspired on progressive propagation of coherent content to the missing image areas have proved to be an interesting alternative for filling textured regions and enhancing data availability, especially for VHR data. Lorenzi et al. (2011) proposed the reconstruction of target regions of VHR images by propagating the spectrogeometrical information retrieved from outside the missing area. This strategy produces pleasant visual quality and it has the advantage of just utilizing the original data from the input image. However, its accuracy is sensitive to the size of missing area, becoming unfeasible to reconstruct cast shadows from dense urban areas that comprise many types of land cover categories.

In this work, our goal is to introduce a novel approach that addresses the problem of shadow removal in VHR images from urban areas. Our technique makes use of only the original information taken from the optical imagery to accomplish the shadow detection and correction of obscured pixels that can be the key for cost reduction to commercial RS data of urban environment in which shadow occurrence is inevitable. Contextual and spectral data are properly combined to generate a refined shadow mask that is used as input for our shadow removal scheme. The proposed inpainting strategy unifies anisotropic diffusion, inner product-based filling order mechanism and exemplar-based completion into a robust and concise framework. In contrast to Casaca et al. (2014)-(2015), where the authors have inpainted targets over real-world images, in this work we propose the use of a shadow guidance mask combined with the inpainting process and multispectral bands as a specific and improved filling order mechanism to recovering large cast shadows regions. The algorithm was implemented via MATLAB and validated using pansharpened reflectance WorldView-2 (WV2) imagery. The conducted experiments confirm the effectiveness of the proposed approach, outperforming other shadow compensate methods such as histogram matching (HM) (Richards and Jia, 2012) in terms of quality of reconstitution as well as discontinuities treatment caused by penumbra regions.

STUDY AREA AND DATASET DESCRIPTION

The designed approach was experimented on WorldView-2 (WV2) VHR imagery acquired on July 06 2012 that encompasses the city of São Paulo, located in the southeast region of Brazil (Figure 1). The study site is the largest city in Brazil, with population about 11.9 million reaching almost 20 million people when metropolitan areas are also considered. Hence, it is a dense urban environment with a great variety of artificial and natural surface types that increase the complexity of urban image analysis. In addition, due to high resolution aspect, a great variety of shadow sizes and types are observed, impacting directly in the data processing of this imagery. WV2 images are orthorectified and georeferenced with basic level of radiometric correction. Panchromatic and four multispectral bands (visible blue, green, red and near-infrared), respectively, 50 cm and 2 m of spatial resolution are employed in order to exploit spatial context and spectral properties for shadow detection and removal steps.
Figure 1. Selected study site. (a) Location of São Paulo city in Brazil; (b) São Paulo’s metropolitan region that encompasses dense urban environment; (c) WV2 image (normal color composite of bands R3G2B1) shows the presence of shadows from different types of targets.

METHODOLOGY

Our methodology comprises three main steps: preprocessing, shadow detection, and shadow removal. The full approach is outlined in the flowchart illustrated in Figure 2. The SDC Morphology Toolbox and the MATLAB language were used to code the modulus of our approach, as detailed below.

Figure 2. Processing pipeline of the proposed methodology.

Preprocessing Step

Following Figure 2, the first step of our methodology consists in preprocessing all the WV2 imagery, which includes radiometric correction and pansharpening. The radiometric correction is computed for all WV2 bands (panchromatic-PAN and multispectral-MS) to convert digital numbers (DN) to top-of-atmosphere (TOA) reflectance data. In addition to the imagery, calibration directives and the parameters provided by the metadata for each band are also considered to properly perform this correction in the ENVI 5.2 software.

TOA reflectance data composed by high resolution PAN and lower resolution MS bands have been pansharpened by applying the Gram-Schmidt (GS) method (Maurer, 2013). This substitution method improves spatial details of MS images through the PAN image replacement by the first GS band. Once the WV2 imagery have been captured at the same time with the same sensor, no geometric corrections are needed and the GS method can be, hence, directly applied to the bands. Therefore, the inverse transformation is performed to obtain the pansharpened spectral bands, providing a substantial enhancement between different types of land cover.

Outputs are assessed through visual analysis and quantitative evaluations considering spectral and spatial...
similarities between the evaluated images. Correlation Coefficient (CC) and Universal Image Quality Index (UIQI) (Yuhendra et al., 2012) were computed by taking the original band as reference to compare it with the resampled pansharpened result. After applying the preprocessing step for the satellite images, the results are taken as input for our automatic shadow detection algorithm.

**Shadow detection algorithm**

Morphological filtering is accomplished to extract dark patterns from the image. These patterns can be viewed as valid shadow candidates, since shadow areas have lower brightness than their corresponding neighborhood pixels. Technically speaking, morphological attribute filtering is performed through Black-Top-Hat (BTH) transformation (Soille, 2003) so as to enhance the dark structures of the image. The BTH (Equation 1) works well to the task of detecting valleys of the image, especially when one uses the closing (\(\emptyset\)) operator followed by the arithmetic difference with the input image \((f)\).

\[
BTH(f) = \phi (f) - f
\]

Instead of employing a purely closing operator, we use in our approach the area-closing \((\emptyset_\lambda)\) operator for gray scale images, as defined in the following:

\[
\phi_\lambda = \bigwedge_i \{ \phi_i (f) \}
\]

where \(\wedge\) is the *infimum* of all closings with connected SEs \((B_i)\) whose sizes in number of pixels are equal to \(\lambda\). Thus, the output is the sum of all filtered input threshold images with dark structures w.r.t. parameter \(\lambda\). The use of \(\emptyset_\lambda\) allows for more flexibility when one intends to exploit contextual relations and spatial information from operators based on fixed SEs. Such a versatility becomes important when dealing with features like shadows, since this operator does not demand the use of prefixed sizes and shapes, besides also decreasing the processing complexity due to its direct use only on the flat zones of the image (a set of connected iso-intensity pixels).

The definition of area parameter is driven by the normalized saturation-value difference index (NSDVI) mask (Ma et al., 2008) obtained from the pansharpened image. NSDVI mask leads to an area overestimation due to wrong discrimination between vegetation and other features that share similar behavior as shadows in HSV color space. Otherwise, the object connectivity is strongly correlated with the spatial resolution of the images, and the area overestimation aims at guaranteeing that all shadows regions will be labeled by the area criterion. In fact, such a criterion is an important trait in this context, especially when coping with high resolution images. Additionally, the area selection criterion also keeps the methodology fully automated, contributing, this way, to support other applications that demand the use of shadow segmentation procedures.

An automatic binarization via Otsu (1979) method is applied to the processed image, generating a binary image that contains shadow candidate pixels. The results are then refined considering the intersection of shadow candidates obtained by the morphological filtering and the spectral normalized difference vegetation index (NDVI) mask. Indeed, this avoids false detections in addition to improving the accuracy of the algorithm. Finally, the detection results are assessed via the performance evaluation metrics as described by Adeline et al. (2013), which analyzes the shadow detection output with a manually built reference image, popularly called Ground Truth (GT). The next step of our methodology consists in removing the shadows previously detected, as detailed below.

**Local inpainting strategy**

Aiming at recovering shadow pixels, we propagate the valid content from the input image \(f\) into the occluded region within the shadow mask \(\Omega\). This procedure is performed by applying three fundamental approaches: anisotropic diffusion, an inner product-based filling order measure, and an exemplar-base completion scheme. First, a reference image \(u\) (namely here as cartoon image) is extracted from the input image \(f\) by numerically solving the anisotropic diffusion equation (Equation 3) proposed by Barcelos et al. (2003).

\[
\frac{\partial f^{(t)}}{\partial t} = g(\nabla f^{(t)}) | \text{div} \left( \frac{\nabla f^{(t)}}{|\nabla f^{(t)}|} \right) - (1-g)(f^{(t)} - f),
\]

where \(f^{(t)}\) corresponds to the scaled version of \(f\). \(g = g(\left| \nabla G_\sigma * f^{(t)} \right|)\) represents an edge detection function, and \(G\)
is the well-known gaussian kernel with a tuning parameter $\sigma$. As a result, $u$ is a smoothed image that holds geometric structures and homogeneous segments from $f$. An advantage when using $u$ instead of $f$ to guide the inpainting process is that the method promotes the filling of linear structures, being less affected by the high-frequencies of the image. Note also that the current inpainting methodology takes into account the multispectral dimension of the target image, which could be a grayscale image or RGB color image.

An inner product-based metric derived from the transport equation devised in (Bertalmío et al., 2000) is applied to $u$ so that the ordering of the non-shadow (valid) pixels in the fill front $\partial\Omega$ (the boundary of the shadow mask) will be traversed. In more mathematical words, the computation of the filling order of the pixels is accomplished by the following priority measure $\delta$:

$$
\delta(p_i) = R(p_i). C(p_i), \quad p_i \in \partial\Omega
$$

(4)

where $R(p)$ and $C(p)$ represent the Relevant and Biased Confidence terms, respectively, given by:

$$
R(p) = \left| \nabla(\Delta u) \cdot \vec{d}_p \right| = \frac{\nabla^2 u_p}{\nabla^2 u_p}
$$

(5)

$$
C(p) = \left\{ \sum_{q \in H_n(\partial\Omega - \partial\Omega)} C(q)^{\frac{1}{k}} \right\}^{\frac{1}{k}}
$$

(6)

with $|H_m(p)|$ being the size of a small square region $m \times m$ centered at pixel $p(x,y)$. While the relevance term $R$ computes the isophotes (pixels sharing the same gray level) from the fill front $\partial\Omega$, confidence term $C$ accounts for the filling coherence during the shadow region completion. More precisely, a label to each pixel in the fill front is properly computed and assigned.

The next step is to determine the most similar patch of pixels from the dynamic region $\Lambda\Omega_p$ to be allocated in the shadow covered area. For this purpose, a cartoon-based metric is used to compare the fixed patch $H_n(p)$ with all candidate patches $H_n(q)$ inside $\Lambda\Omega_p$. More specifically, the optimal patch $H_n(q)$ is the one that minimizes the following distance (Equation 7) between $H_n(p)$ and $H_n(q)$:

$$
d(p,q) = \sqrt{\sum_{k=1}^{Lp} \frac{(p_k - q_k)^2}{\Delta U}}
$$

(7)

where:

$$
\|p\|_{\Delta U}^2 := \sqrt{p^T \Delta U p}
$$

(8)

$\Delta U$ is a diagonal matrix defined by the Laplacian of $u$ and $p = (I_{p1}, I_{p2}, ..., I_{pk})$ is the column vector containing the intensities of the given pixels on $H_n(p)$. Thus, for pixels that belong to the edges of the Laplacian of cartoon image $u$, Equation 7 assigns higher weights. Moreover, in our approach the multispectral bands are embedded into the distance calculation, which customizes the metric to be better evaluated in our context of remote sensing. Finally, concerning the computational effort, this strategy of searching candidate pixels only inside the dynamic region $\Lambda\Omega_p$ provides a gain in terms of avoiding pixel inspections in large areas of the image. Another important aspect is that this strategy avoids reallocating pixels that are distant from the shadowed areas, therefore improving the image coherence for those recovered areas.
RESULTS AND DISCUSSION

The proposed approach was applied on two subsample of 300 x 300 pixels from original WV2 scene in order to assess and visually check the results. Figure 3(a) shows the true color composition of original MS bands where one can notice the lack of details on urban targets. A visual comparison between the results after preprocessing (Figure 3(b)) demonstrates an insertion of spatial details by GS pansharpening, which enhances the discernment of individual elements not provided by original MS image. In addition, GS pansharpening have shown high correlation and spectral fidelity with respect to the original multispectral data, confirmed by high CC and UIQI, both with values of approximately 0.87.

The conversion into TOA reflectance also performed in the preprocessing leads to a satisfactory result in terms of haze separation, providing an improvement in the visual quality. Moreover, the dataset in reflectance quantities contributed to avoid inconsistencies in NSDVI and NDVI indices that are required for spectral mask generation.

Figure 3(c) depicts the binary results obtained from the shadow detection approach. By visual inspection, one can observe that a good discrimination of shadows regions was reached, which ranges from small targets, such as trees and vehicles, to complex urban environments. The effectiveness of detection is indeed confirmed by the result overlapped as red border on true color composite pansharpened image (Figure 3(d)). The approaches’ accuracy is numerically validated when one computes performance evaluation metrics for shadow detection, which is listed in Table 1. In fact, our automatic approach yields an overall accuracy of 90.27% and 86.83% for image 1 and image 2, respectively. A significant producer’s accuracy around 93% for both images shows how accurately the method detects the true shadowed, increasing the reliability of the method through higher agreement with GT. User’s accuracy indicates low rate of false detection (~20%), outstanding problems faced by other methodologies such as inefficiency in detecting high reflectance objects (cars) within shadow areas. Furthermore, an overestimation of false positive pixels could also occur due to the difficulty of GT manually generation, especially over dense vegetation and uncertainty pixels. Additionally, the automatic method demonstrated a low computational burden when detecting shadows of dense urban areas with a satisfactory level of discrimination.

![Figure 3](image-url)

**Figure 3.** Results of shadow detection for subset 1 (top) and subset 2 (bottom). (a) Original WV2 true color composite image; (b) Result obtained by the preprocessing; (c) Shadow detected after refined step; and (d) Shadow mask overlaid with original true color image.
Table 1. Performance evaluation of proposed shadow detection approach.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Producer’s accuracies (%)</th>
<th>User’s accuracies (%)</th>
<th>Overall accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image 1</td>
<td>93.51</td>
<td>81.72</td>
<td>90.27</td>
</tr>
<tr>
<td>Image 2</td>
<td>94.29</td>
<td>79.81</td>
<td>86.83</td>
</tr>
</tbody>
</table>

Shadow-effects mitigating was obtained by local inpainting method, where the results are displayed in Fig. 4. For this purpose, shadows refined map was used as seeds for the partition process. By analysis criterion, it was selected the most complex shadows regions considering the high-rise buildings of the scene (identified by red and yellow dotted lines). From the original images (Figure 4(a)), it is evident that our approach substantially improved useful areas of the image that were previously contaminated by shadows. The results produced by our inpainting strategy are provided in Figure 4(b), in which the outputs demonstrated good coherence for highway and cars recovery, especially from the region in the topmost area in Figure 4(b), highlighted by red dotted line. Unfortunately, from yellow dotted line region (Figure 4(a)-top), we noticed that the filling in some pixels mislead information and does not maintain the geometric configuration that shows discontinuous patches. From Figure 4(b)-bottom, a good coherence for the restored shadow pixels was observed in both dotted regions. Although some discontinuities and missing information such as the crosswalk have been observed in the road restoration, building rooftops was successfully recovered as well as some building contours.

In order to confirm the effectiveness of our methodology, we perform a comparison against the histogram matching method (HM) (Richards and Jia, 2012), which is the baseline solution for enhancing shadow pixels. The HM was applied in each pansharpened bands and a RGB true color final image was generated. Reconstructed shadow regions from both methods are presented in Fig. 5, where one can see that the results obtained by our method (Figure 5(a)) were more consistent with the surroundings than the HM (Figure 5(b)). Although local neighborhoods region pixels were used as reference for calculation of new intensity values using HM, an insertion of white edge was observed in the transition of light-shadow region, mainly due to penumbra pixels. Therefore, focusing on the task of removing inconsistencies caused by the penumbra, we have removed transition pixels and the HM was recalculated. In addition, a median filtering was applied to smooth out transition between surroundings pixels. Figure 5(c) shows the result of this reprocessing step. Note that our approach clearly repairs shadow region in the image and outperforms HM approach, since no improvements was observed even after the reprocessing stage.

Figure 4. Shadow removal results using inpainting strategy.

Figure 5. Shadow removal results using histogram matching method.
CONCLUSION

The proposed method was capable of detecting and removing shadow effects of high spatial resolution WV2 data from dense urban areas. The combination of contextual and spectral features show an overall accuracy up to 90% with reduction of omission errors and low rate of false detection (~20%). Moreover, the method avoids material dependence since it relies on reflectance pixels, surpassing the issues faced by other methodologies such as inefficiency in detecting high reflectance objects. Furthermore, the pansharpening provided an insertion of spatial details with high correlations into the original dataset, enhancing the discernment of individual elements that are of paramount importance for visual analysis conducted in other stages of the algorithm.

The shadow removal procedure as a new inpainting strategy has demonstrated a satisfactory coherence and showed up details priori obscured. The combination of anisotropic diffusion, inner product-based filling order mechanism and exemplar-based completion have resulted in a robust apparatus to properly repair shadows pixels, overcoming the HM technique in terms of good visual distinction and penumbra-effects mitigation. Large shadow areas were exploited and then reconstructed by the recovering inpainting strategy, demonstrating that highway and cars can be successfully recovered in the urban areas.

Despite the existence of some discontinuous patches and the tuning of several parameters, our approach clearly distinguishes and removes shadow pixels in a complex urban environment by only taking a single scene, yielding a good trade-off between visual quality and low computational cost. As future work, experiments with a great number of images will be conducted to better understand the behavior of the employed parameters under urban environments, besides assessing the results via unsupervised classifications.

ACKNOWLEDGMENT

This research is supported by FAPESP under grant number 2013/25257-4. One of the authors (Samara Calçado de Azevedo) is also thankful to FAPESP (grant number 2015/26743-5) for the awarded fellowship to visit Chapman University.

Figure 5. Comparison of shadow removal results from (a) the proposed method; (b) HM; (c) HM reprocessed.
REFERENCES


