Optik 125 (2014) 741–744

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

A novel algorithm in buildings/shadow detection based on Harris detector

Bo Yu [a,b], Li Wang [a], Zheng Niu [a,*]

[a] The State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China
[b] Graduate University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 9 March 2013
Accepted 6 July 2013

Keywords:
Buildings/shadow detection
Harris corner detector
Algorithms

ABSTRACT

A novel method based on corner detectors is proposed in detecting shadow and buildings in this paper. Its most outstanding point is employing Harris corner detector in region-based detection, despite that Harris detector traditionally used to select pixels as final results. Different densities of buildings are generally influenced by different features for recognition. First time, images are self-grouped into two groups according to the distribution of buildings, and two specifical algorithms are ready for detection specifically. A region-based method is used in comparison with our algorithm, and the results indicate that the new idea works not only more robustly, but also more effectively. It is a fast and simple method, which needs average 3.28×10^{-5} s to run per square image.

1. Introduction

Buildings detection has been one of the most significant aspects in assessing the progress of investment programs and urban planning. Numerous algorithms have been proposed, and mostly, they are focused on spectral information and textural characteristics. Shadow is companying buildings everywhere and it is the main obstacle in recognizing buildings, especially in urban environment. Good shadow detection contributes to accurate recognition of buildings. Corner detectors are the methods that can detect points with specific features. Harris detector [1] has been widely used in corner detection [2,3] and image segmentation [4]. It is one of the most well-known algorithms in detecting feature points of interest, because of its robust in the variation of illumination, rotation and noise. Among the multiple corner detectors proposed now, Schmid, Mohr and Bauckhage have conducted several researches on evaluating the methods and concluded that Harris detector performed best [5].

To date, corner detectors are not commonly used in region based buildings or shadow detection, because it is designed for extracting pixels, which are intended to be final results. However, in this paper, corner detection is partly adopted and synthesized with the idea proposed in [6] to detect region-based shadow. The new algorithm is a simple, but efficient one. It is not only with high accuracy in detecting buildings, but also sensitive to shadow. The principle idea is discussed in Section 2 and the validation experiments are introduced in Section 3. Final part is the conclusion.

2. Proposed method

Commonly, researchers are trying to detect buildings regardless of their distribution density. Nevertheless, buildings with different densities are with different features. Densely distributed buildings tend to share little open space, and shadow is the main obstacle in identification. On the contrary, sparsely distributed buildings would have more open spaces, especially soil, which mainly suppresses recognition. Therefore, our proposed method is divided into two subparts after detecting shadow based on Harris corner detector and the idea proposed in [6].

Harris detector is a auto-correlated function, which measures the local variation by shifting in various orientations [7]. The general idea is convoluting the original image with a kernel, which is often intended to be Gaussian function. Because it has been proved to be the only possible scale-space kernel [8] by Koenderink [9] and Lindeberg [10]. Thereafter, the scale-space of the original image $I(x,y)$ can be generated as below:

$$S(x, y) = G(x, y) * I(x, y) \tag{1}$$

where $G(x,y)$ is the Gaussian function and * marks convolution operation.

Smoothed Gaussian filters are used in [6] and they are used in two directions, horizontal and vertical ones.

$$G_x(x, y) = \frac{-x}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) \tag{2}$$

* Corresponding author. Tel.: +86 15110093496.
E-mail address: niuz@irsa.ac.cn (Z. Niu).

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.07.043

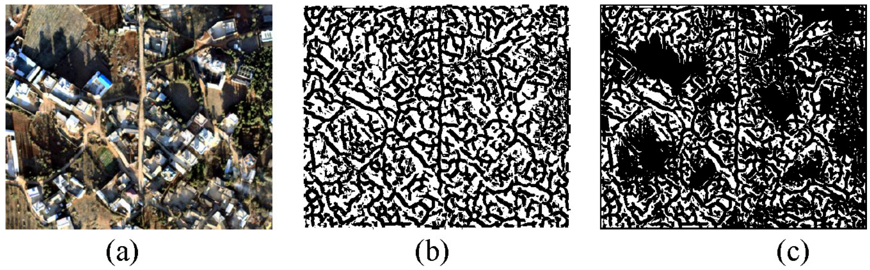

Fig. 1. Sample images for algorithm description: (a) original image, (b) feature points detected based on Harris corner detector and (c) binary magnitude image.

$$G_y(x, y) = \frac{-y}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) \quad (3)$$

In our paper, we have a smoothing window of 7×7. x, y in Eqs. (2) and (3) represent the positions of every pixel in the window. Therefore, two kernels G_x and G_y focusing on two orientations are formed. Then, we can have $I_x(x,y)$ and $I_y(x,y)$ after convolution with the two kernels.

After every convolution, we have a T calculated as follows:

$$T(A) = Det(A) - \varepsilon Tr^2(A) \quad (4)$$

where A is a matrix composed of four elements

$$A = \begin{pmatrix} a_{xx} & a_{xy} \\ a_{xy} & a_{yy} \end{pmatrix} \quad (5)$$

and a_{xx} is the summary of all the $I_x{}^2(x_i,y_j)$, belonging to the smoothing window. a_{xy} is the total value of all the $I_x(x_i,y_j)I_y(x_i,y_j)$, and accordingly, a_{yy} is the sum of $I_y{}^2(x_i,y_j)$.

Harris and Stephens [1] believed that the corner points are selected by examining the maximum T of every 7×7 window after generating all the T in the image. And a sample image based on Harris corner detector is shown in Fig. 1(b).

To show the gradients among feature points, magnitude $M(x,y)$ is adopted from [6] and defined as:

$$M(x, y) = \sqrt{I_x^2(x, y) + I_y^2(x, y)} \quad (6)$$

In order to present the magnitudes more obvious, Otsu's method is applied [6] because of its automation and intelligence. No further subjective factors and manual work is required. Fig. 1(c) shows the result.

Obviously, shadow in Fig. 1(c) is entirely detected and maintained the original shape. Moreover, other feature points describing buildings or open spaces are kept quite well. Therefore, image *shadow*(x,y) can be generated by conducting a minus operation between inversed Harris corner image and inversed binary $M(x,y)$.

Since shadow can be generally extracted, building detection becomes much more accessible. HSV color space has been widely used in image segmentation, and H channel can be used to extract built-up areas for its robustness toward illumination and shadow. Then, we have initiatory built-up image *Built_up*(x,y) by

$$Built\_up(x, y) = H\_bi(x, y) - shadow(x, y) \quad (7)$$

where *H_bi* is the binary image obtained by Otsu method on H channel.

Further detection is based on *Built_up* image (Fig. 2(a)), and more features are considered according to the distribution of buildings.

Sparsely distributed buildings are heavily influenced by open spaces and soil. V channel in HSV is short for "value", which means brightness of a pixel. It is determined by the maximum normalized value by 255 in three channels. Since soil and open spaces mostly have low reflectivity compared with concrete in buildings, they tend to have low brightness values. Therefore, it is more likely for soil and open spaces to be distinguished from buildings. Binary V is displayed in Fig. 2(b).

From Fig. 2, we can see that binary V can well delineate open spaces, soil and roads, which are barriers for sparsely buildings detection. However, most obstacles can be extracted by subtracting binary V from built-up image.

The proposed method for sparsely distributed buildings is not a method for all. It does not suit densely distributed buildings. Because Otsu's method is adopted in the process of binarization, and it is performed according to its background. In the environment of multiple buildings, most pixels share high brightness value, and binaryzing would mistake some buildings with dark roofs as backgrounds. Therefore, quite many buildings would be missed. Since built-up image has rejected the most disturbing shadow from buildings, and it can well indicate the buildings exactly, there is no need for further procession.

Post procession for both densely and sparsely distributed buildings is necessary, because some small spots would ruin the neatness of the result and they should be omitted. It is done by thresholding the connected components, and labeling them as backgrounds.

Generally, the new method can be summarized in the following steps:

Step 1. Filter noise of original image with Gaussian filter.
Step 2. Sharpen image by histogram stretching.
Step 3. Reject plants by thresholding an image P and

$$P = \frac{2G}{R+B} \quad (8)$$

Step 4. Detect feature points from the image obtained in Step 3 by Harris corner detector.
Step 5. Obtain $M(x,y)$ according to Eq. (6).
Step 6. Transform original image into HSV and achieve built-up image in (7).
Step 7. Determine whether the image belongs to sparsely distributed buildings or densely one by calculating the average number of connective components of shadow image. Since

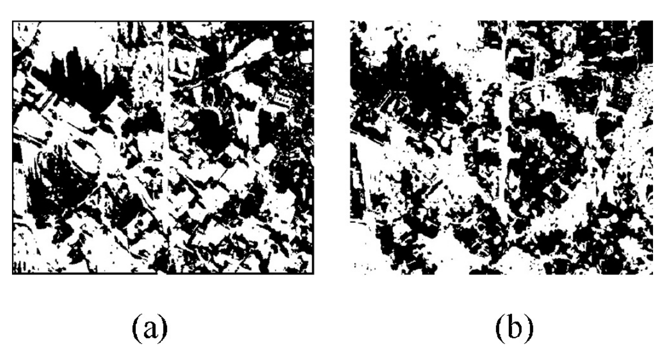

Fig. 2. Sample images indicating sparsely distributed buildings detection. (a) Built-up image and (b) binary V image.

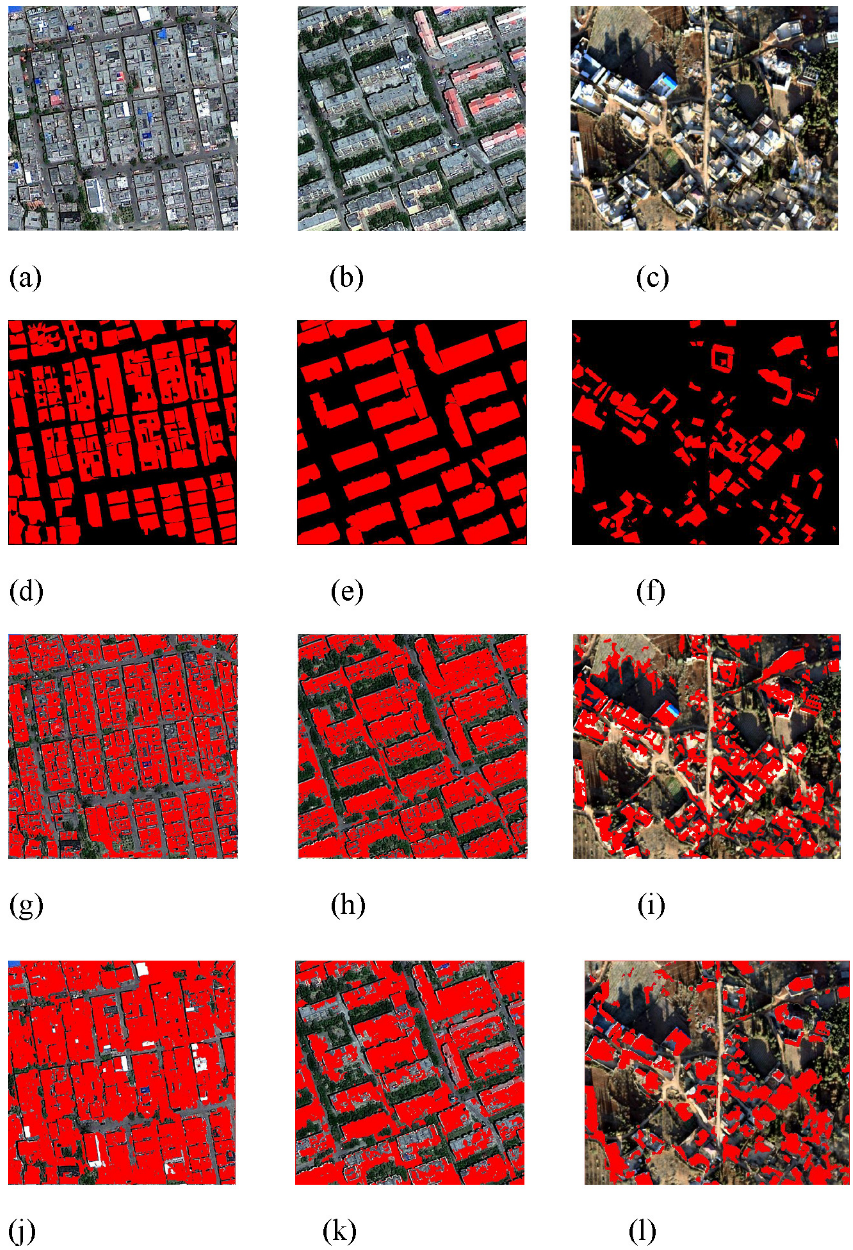

Fig. 3. Performances of BIS and the new method. The first row shows the original images of image 1, 2 and 3. (d), (e) and (f) are achieved by visual interpretation. Row three indicates the buildings detected by BIS, and the results are shown in (g), (h) and (i), respectively. Results generated by the new method are displayed in the final row, in (j), (k) and (l).

buildings are the main objects that produce shadow, densely distributed buildings would produce more average shadow connective components after divided by the area of image.

Step 8. If it is a densely distributed image, continue to Step 9. Or, subtract binary V from the built-up one.

Step 9. Post procession by removing small spots as noise to perform neatly.

3. Experiments and discussion

The new method is validated with three images captured from Google Earth, describing part of Wulumuqi, Xinjiang Province of China. Since it would be impossible to get updated statistical data from local government about the distribution of buildings, and visual interpretation on the original images is still recognized as reliable reference for comparing two techniques. However, it would seem non-practical to do visual interpretation on the image of a whole administrative area, three images are selected having dark roofs and similar spectral characteristics with roads or open spaces nearby. Enhancing the difficulty in detection makes the performances presented in this paper more reliable. Therefore, performances of our algorithm may not seem to be so well as in other papers, but if it can get good results in a relatively worse condition, higher accuracy would be easily accessible.

Among the images, one is with sparsely distributed buildings, the other two are with dense ones. It is compared with Berkeley Image Segmentation (BIS: http://www.imageseg.com) [9]. It is a object-oriented method based on region growing [10]. In BIS, every image is segmented automatically with its default settings, and various sub-regions are generated. Therefore, objects of interest can be extracted by manually generated rules. To show

Table 1
Statistical analysis of two methods processing three images.

Image	Method	OE (%)	CE (%)	OA (%)	Kappa
Image 1	BIS	11.24	34.76	68.1731	0.3356
	New method	18.55	33.00	73.7593	0.4800
Image2	BIS	27.59	31.42	72.4869	0.4475
	New method	32.34	26.98	74.0322	0.4726
Image3	BIS	43.69	49.93	78.1089	0.3880
	New method	31.53	48.51	84.1663	0.4922

the performance of BIS without other influences, buildings are detected based on visual interpretation with the help of Environment for Visualizing Images Feature Extraction (ENVI-EX) package (http://www.exelisvis.com/). The experimental results are shown in Fig. 3.

Original image 1 (Fig. 3(a)) is the most difficult for detection among the three images, because all the buildings and roads share similar spectral characteristics. If there is no shadow beside the buildings, it would be even hard for us to distinguish buildings from backgrounds. Moreover, compared with visual interpretation shown in Fig. 3(d), buildings detected by new method succeeded in being grouped in blocks, while the ones detected by BIS are mixed with roads and open spaces nearby. Nevertheless, the new method is sensitive to shadow, and the details in shadow are described in Fig. 3(g). However, BIS is a region-based method, and various pixels belonging to shadow are grouped into other features (see Fig. 3(j)).

From the experimental results shown in the second column of Fig. 3, we can see that BIS has missed many parts of buildings (Fig. 3(k)), while the new technique covers almost all the constructions. As for sparsely distributed buildings, the new method is more sensitive to roads and open spaces in comparison with BIS in Fig. 3l). Numerous pixels are mistaken as buildings and grouped into building parts.

To do further assessment toward the two methods, kappa coefficient, overall accuracy (OA), ommision error (OE) and commission error (CE) are used to evaluate. Kappa coefficient is the most significant one, because it marks the robustness of an algorithm. If is is over 0.4, the algorithm is recognized as having good performance. OA is an overall assessment, indicating the general performance of the technique. OE and CE are the percentages of buildings being ommitted and other features mistaken as buildings. The stastics are shown in Table 1.

Kappa coefficients witness that new method performs better than BIS not only visually, but also statistically. The new technique has a robust kappa coefficient, concentrating at 0.48. Nevertheless, BIS's kappa coefficients range greatly from 0.33 to 0.44. Exactly, the new method is more robust and resists the variation of images, whether their buildings are densely distributed or not. Moreover, BIS obtains its buildings by thresholding its rules manually according to its image, while the new idea is an idea for all. From comparison between the two methods in OA, the new algorithm wins. However, both techniques obtain high errors in OE and CE. Many buildings are miss-classified into other features and multiple background information is mistaken as buildings. The possible reasons are that BIS is a region-based method, and various regions are generated by the initial parameters in spectral and textural features. When faced with buildings with similar spectral and textural information with its neighbors, it would be more difficult to group pixels belonging to the same object into one group. As for the new technique, more features should be added to separate buildings from backgrounds.

Efficiency is also a key assessment for an algorithm. Since BIS needs manual cooperation, which needs much time to consider, only running time of the new algorithm is monitored when processing the three images. For Image 1, it is 576×567 and needs 11.937 s to finish detection. Image 2 needs 9.751 s with 574×563 size. Finally, Image 3 is 479×402 and buildings can be detected within 6.116 s. To get the efficiency of the algorithm, average time processing per square meters is calculated to be 3.28×10^{-5} s. Commonly, it can be recognized as fast.

4. Conclusion

Multiple researches have been conducted in buildings recognition and corner detectors are rare in region-based detection. The new method in this paper is distinguished mainly from traditional ones with its first-time using Harris corner detector in shadow regions detection. And the images are grouped into two groups in accordance with the distribution of buildings. Both conditions are considered in the algorithm. Experiments show that the new idea is robust with the variation of images and performs better compared to region-based method, not only in efficiency but also in accuracy. Still, there is much work to do about the improvement of our algorithm in lowering miss-classified pixels.

Acknowledgements

The authors are appreciated for the help from the Major State Basic Research Development Program of China (2013CB733405), Major State Basic Research Development Program of China (2010CB950603), Public service sectors (meteorology) Special Fund Research (GYHY201006042), National Natural Science Foundation of China (41201345) and High resolution for earth observation major projects.

References

[1] C. Harris, M. Stephens, "A combined corner and edge detector," presented at the In: Proc. 4th Alvey Vision Conf., 1988.
[2] H. Moravec, Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover, Robotics Institute, Carnegie-Mellon University, 1980.
[3] L. Yi-bo, L. Jun-jun, Harris corner detection algorithm based on improved contourlet transform, Proc. Eng. 15 (0) (2011) 2239–2243.
[4] A. Kovacs, T. Sziranyi, Harris function based active contour external force for image segmentation, Pattern Recogn. Lett. 33 (9) (2012) 1180–1187.
[5] C. Schmid, R. Mohr, C. Bauckhage, Evalutation of interest point detectors, Int. J. Comput. Vision (Special issue on a special section on visual surveillance) 37 (2) (2000) 151–172.
[6] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1979) 62–66.
[7] K.G. Derpanis, The Harris Corner Detector, Department of Computer Science and Engineering, York University, 2004.
[8] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision (2004).
[9] N. Clinton, A. Holt, J. Scarborough, L. Yan, P. Gong, Accuracy assessment measures for object-based image segmentation goodness, Photogramm. Eng. Remote Sens. 76 (3) (2010) 289–299.
[10] U.C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, M. Heynen, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, ISPRS J. Photogramm. Remote Sens. 58 (2004) 239–258.